Stanford Exploration Project, Report 80, May 15, 2001, pages 1–508

Stolt without artifacts? — dropping the Jacobian

Stewart A. Levin[1]

ABSTRACT

In recent SEP articles (Popovici et al., 1993; Lin et al., 1993; Blondel and Muir, 1993) the SEP "in-time" group derived and tested a Stolt migration and modeling method that employed the discrete Fourier transform (DFT) for nonuniform spacing, the *slow Fourier transform*, to avoid interpolation artifacts. In an attempt to better evaluate their work, this paper approaches the subject afresh, first considering a change of variables in a continuous integral and its relation to discrete theory and matrix adjoints. Along the way, I spin off yet another Stolt migration method that permits one to still employ the fast Fourier transform (FFT) and avoid (traditional) interpolation, and that does not require any Jacobian scaling. Finally, I return to the DFT-based method and conclude that, under classical assumptions of geophysical time series analysis, the alternate interpolation it implies is indeed more correct than the sinc interpolation Harlan (1982) recommends.

INTRODUCTION

Artifacts in Stolt migration are easily recognizable by their global appearance. This is because the method works in the F-K domain — local errors in F-K are transformed into global errors in T-X. The source of most such artifacts arises from the choice of interpolation method used for the Stolt mapping in the F-K space. It is therefore important to understand this mapping or change of variable thoroughly. Consider the integral

$$\int_a^b f(x)\,dx \tag{1}$$

of a function f of a real variable x. Using elementary calculus, if x is a monotonic function of y, we can change variables in the integral to y, as follows

$$\int_{\tilde a}^{\tilde b} \tilde f(y)\frac{dx}{dy}\,dy \quad , \tag{2}$$

where $\tilde f(y) \equiv f(x(y))$. The derivative $\frac{dx}{dy}$ of the change of variables is called the *Jacobian* of the transformation $x \to y$. It has a geometric interpretation in terms of lengths and area (or mass). Figure 1 represents the function $f(x)$ over some small interval of length Δx. The area under its graph is then approximately $f(x)\,\Delta x$. After the change of variables to y, the function

[1]**email:** not available

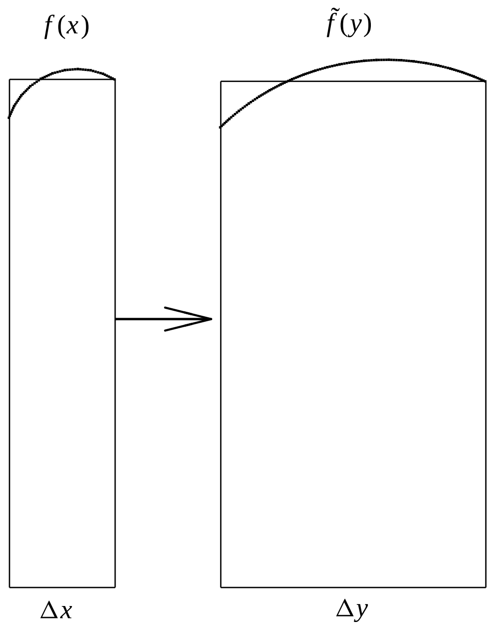

Figure 1: Change of variable from $x \to y$ for continuous integration.

$\tilde{f}(y)$ has the same height as $f(x)$, but the width of the small interval is now $\Delta y \approx \Delta x\, dy/dx$. In order to preserve the area under the original graph of $f(x)$, we need to rescale $\tilde{f}(y)$ by the reciprocal of the interval scaling dy/dx. Let us consider now a discrete or sampled function modeled in the continuous domain as a set of point masses (Dirac delta functions) at discrete intervals as in Fig. 2. A change of variable $x \to y$ on the interval $[a,b]$ moves each of the point masses to a corresponding location in the interval $[\tilde{a},\tilde{b}]$. Integration, that is, summing the point masses, produces the same total mass in both settings *without* the need to rescale by the Jacobian.[2] The world of seismic recording lies between these two cases. On the one hand, the earth's response is essentially continuous. On the other hand, our recordings, being digital, are either assumed or forced to be band-limited. The band-limited assumption requires us to replace the (infinite-bandwidth) point impulses with their band-limited equivalents: sinc functions. The sections that follow explain how to apply both analogues of integration to the method of Stolt migration and explore their consequences.

STOLT MIGRATION — CONTINUOUS AND DISCRETE

Stolt migration (Stolt, 1978) is based upon the constant-velocity imaging formula

$$M(k_x,\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\tau k_\tau} P(k_x,\omega)\, d\omega \quad , \tag{3}$$

[2]Harlan uses the formula $\delta(f(x)) = \sum \left| f'(x_n) \right|^{-1} \delta(x_n)$, where x_n are the zeros of $f(x)$, thus including the Jacobian. This formulation is also correct because of the distinction between shifting and change of variable. See Appendix A for a detailed discussion.

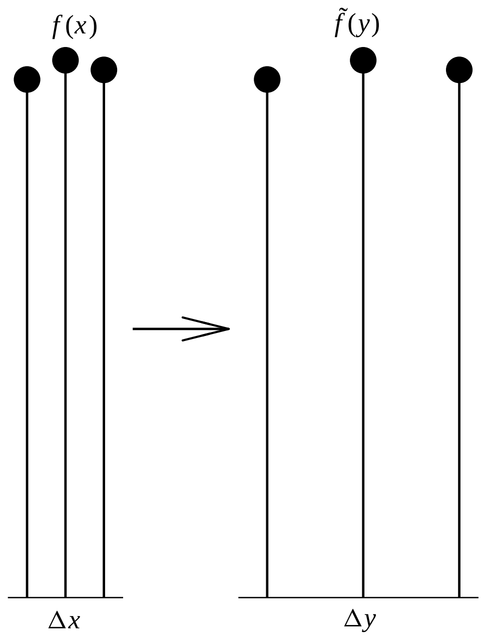

Figure 2: Change of variable from $x \to y$ for discrete integration.

where ω is temporal frequency, τ is vertical traveltime, k_x is spatial wavenumber, and k_τ is given by the dispersion relation

$$k_\tau = \mathrm{sgn}(\omega)\sqrt{\omega^2 - v^2 k_x^2} \quad , \tag{4}$$

which characterizes one-way wave propagation in a medium with constant velocity v. Stolt's fast migration method changes the variable of integration in equation **??** from $\omega \to k_\tau$, turning it into the inverse Fourier transform

$$M(k_x,\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\tau k_\tau}\,\tilde{P}(k_x,k_\tau)\,dk_\tau \tag{5}$$

with

$$\tilde{P}(k_x,k_\tau) \equiv P(k_x,\omega(k_\tau))\frac{d\omega}{dk_\tau} \ . \tag{6}$$

It is noteworthy that the Stolt equation **??** differs from the earlier phase-shift equation **??** in that it images without evanescent energy (Stolt, 1984). From equation **??** we find that $d\omega/dk_\tau = k_\tau/\omega$. For discrete data, we first limit the continuous integral **??** to the temporal frequency band $[-\pi/\Delta t, \pi/\Delta t]$. Then we use a discrete Fourier transform to evaluate the integral for our digital data. The change of variable $\omega \to k_\tau$ in **??** then calls for some interpolation in order to use the fast Fourier transform. Modulated-sinc interpolation has been shown to be appropriate for this purpose (Harlan, 1982).

Band-limited spikes

Instead of interpolating between discretely sampled values of ω, let us apply a band-limited version of point-mass remapping in the Stolt integral. Each sample of the digital series represents a circularly-wrapped sinc function centered about that sample. Following the point-mass recipe for a change of variable $x \to y$ discussed in the introduction, each such sinc function at a point x should be shifted up or down to the new location y. Adding them all together (at the discrete sample points) produces the uniformly sampled integrand without recourse to Jacobian rescaling. This is then the promised Stolt migration method: replace each input $P(k_x, \omega)$ sample with a corresponding (tapered) sinc function centered at the output location (k_x, k_τ) and then apply an inverse 2-D fast Fourier transform.

Matrix formalism

Let us see what this Stolt variant looks like in matrix form. Treating k_x as a parameter, we can represent Stolt migration as a matrix mapping of input temporal frequency ω to output vertical wavenumber k_τ in which the nonzero coefficients are concentrated about the hyperbolic trajectory implied by equation **??**. Claerbout (1985) illustrates related mappings in normal moveout correction. The band-limited spike method tells us to place a sinc function vertically in each column of the matrix centered about the k_τ corresponding to that column's value of ω, as illustrated in Figure 3. Conventional Stolt migration has us place an appropriate sinc

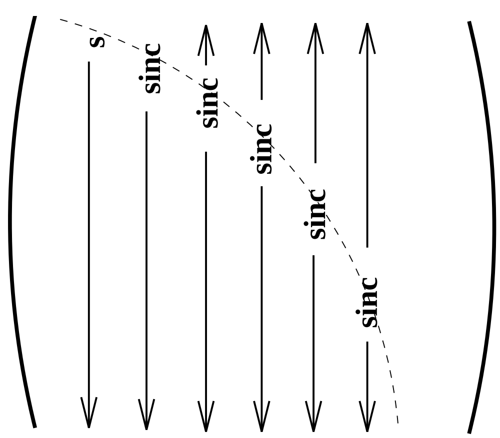

Figure 3: Matrix representation of the new Stolt migration method, placing sinc functions in each column along the NMO trajectory. stew1-newmigmatrix [NR]

function in each row of the matrix at the ω corresponding to that row's k_τ and then scale the row by the Jacobian k_τ/ω, as sketched in Figure 4.

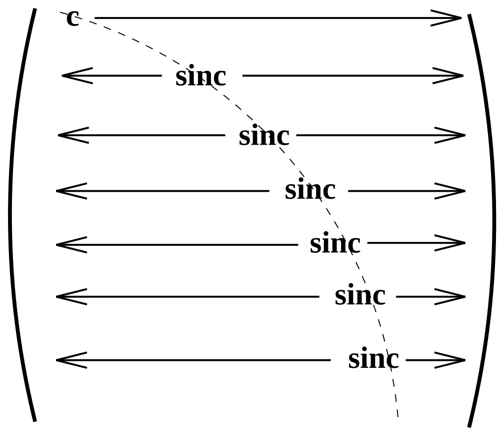

Figure 4: Matrix representation of the conventional Stolt migration method, placing sinc functions in each row along the NMO trajectory and scaling each row by the k_τ/ω Jacobian. stew1-oldmigmatrix [NR]

DISCUSSION

We have now approached discrete Stolt migration from a number of perspectives. How different are the results? To answer this, I have reprinted the following table from Levin (1986) attributed to Harlan (1983):

	Operator	**Transpose**	
Migration	$NMO\cos\theta$	$INMO$	(7)
Diffraction	$INMO\sec\theta$	NMO	

where *NMO* represents the Stolt frequency downshift $\omega \to k_\tau$ and $\cos\theta$ is the Jacobian ratio k_τ/ω. *INMO* is the inverse *NMO* that zero fills the evanescent region. Notice that Harlan's migration definition is just what we have described as the matrix representation of conventional Stolt migration — each row is some *NMO* remapping plus a Jacobian scaling. Its transpose, more properly its adjoint, is also a remapping (*INMO*) and does not have any separate rescaling. Our new Stolt migration is indeed $INMO^T$, equivalent, therefore, to conventional Stolt migration in the continuous limit. To verify this, one notes that the rows of *INMO* are sinc functions centered at the k_τ corresponding to the row's value of ω. Our variant Stolt migration places sinc functions centered at the k_τ corresponding to the column's value of ω and hence is the adjoint of Harlan's adjoint to Stolt migration and therefore is Stolt migration. Further, this formulation is modestly superior to conventional Stolt migration because it does not have to apply a Jacobian scaling. This is not because of the trivial potential savings of one less multiplication in the algorithm. It is because the new equation requires discretization of only the continuous change of frequency variable given by equation **??**, whereas the conventional method also discretizes the derivative of the change of variable.

Fast or slow transform?

In SEP-79, the in-time group (op. cit.) derived and tested Stolt migration by the slow Fourier transform. With this method, the function $\tilde{P}(k_x, k_\tau)$ is computed by evaluating the discrete Fourier transform of $P(k_x, \omega)$ directly at $\omega(k_\tau)$ instead of interpolating from the uniformly spaced output of the FFT. Is this approach really better? To analyze the difference between the two methods, I will start from the usual assumption of geophysical time series analysis that our data is a digitized version of a band-limited continuous function. (Figure 5.) More specifically, we apply, in either order, a time window and a uniform sampling to the signal. The time windowing convolves the spectrum with the sinc function transform of the windowing function, producing leakage outside the original frequency band. (Figure 6.) The effect of sampling is then to convolve the spectrum with a comb, replicating a folded copy of the spectrum periodically. (Figure 7.) If we performed the two commutative operations in reverse order, then sampling would replicate the original band-limited spectrum, and windowing would convolve with the sinc function. The spectrum, of course, remains periodic, but may be modified by leakage or wraparound in the convolution. Note that a band-limited impulse in the continu-

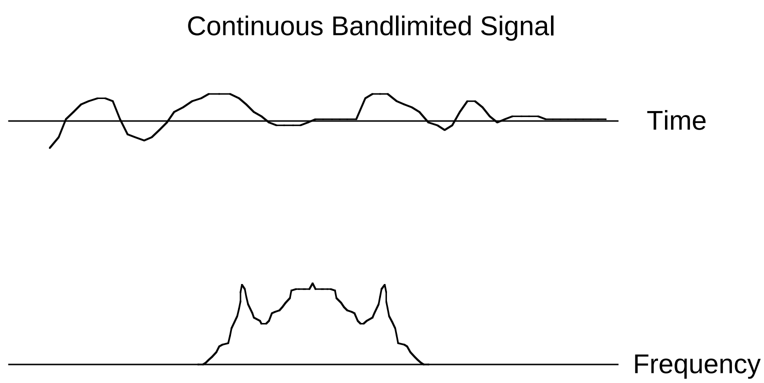

Figure 5: Continuous band-limited signal to be processed. stew1-SEP-sem01 [NR]

ous time function is a sinc function. If we sample the sinc function at its peak, the remaining samples are at zero crossings and its digitized representation is a single spike. If we sample it off its peak, we encounter nonzero sidelobes in the time series. First, let us consider shifting the original continuous-time function by T and then sampling. Using Fourier transform theory, this is accomplished by multiplying the spectrum by $e^{i\omega T}$ and then convolving with the sampling comb. (Figures 8 and 9.) If, on the other hand, we digitize first and then multiply the spectrum by $e^{i\omega T}$, we will not get the correct result. (Figure 10.) In fact we only obtain the correct result when $e^{i\omega T}$ has the same periodicity as the spectrum, that is, when T is an integral number of sample periods. The proper thing to do in the spectral domain is to multiply the Fourier transform of the discrete data by a replicated copy of $e^{i\omega T}$ over the fundamental band. This is then discrete convolution in the temporal domain with a sampled version of a shifted sinc function, in other words, sinc interpolation. Figure 11 summarizes the two flows we have just compared. We now turn from the time to the frequency domain and consider shifting the Fourier transform of a band-limited function by Ω. Again Fourier transform theory says to multiply by an exponential, this time by $e^{-i\Omega t}$ in the temporal domain. Sampling, that is,

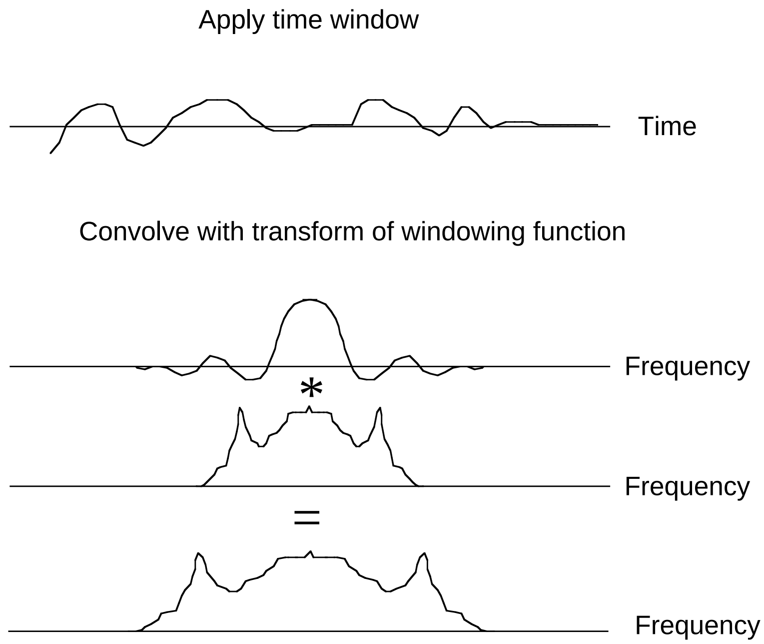

Figure 6: Windowing the continuous band-limited signal of Figure 5. stew1-SEP-sem02 [NR]

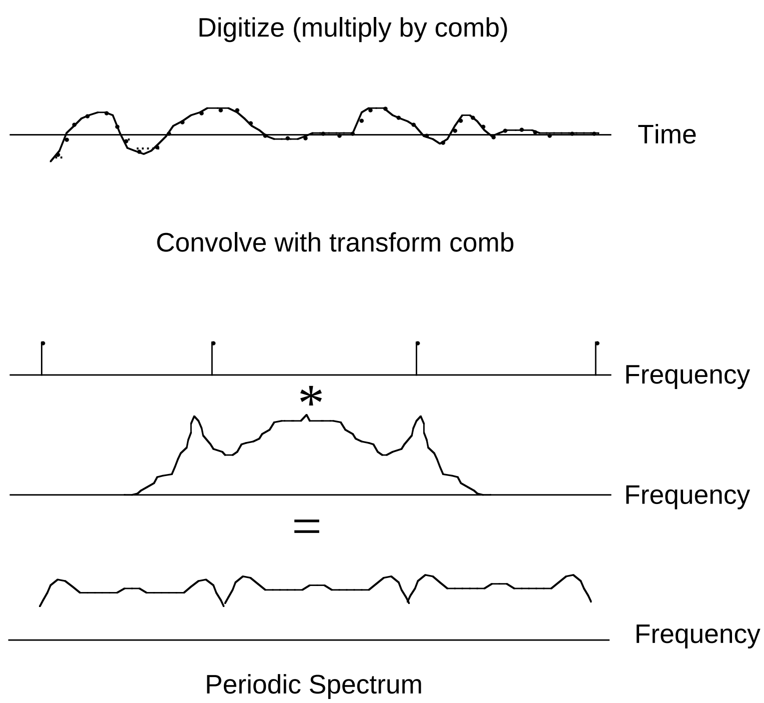

Figure 7: Sampling of the time-windowed signal of Figure 6. stew1-SEP-sem03 [NR]

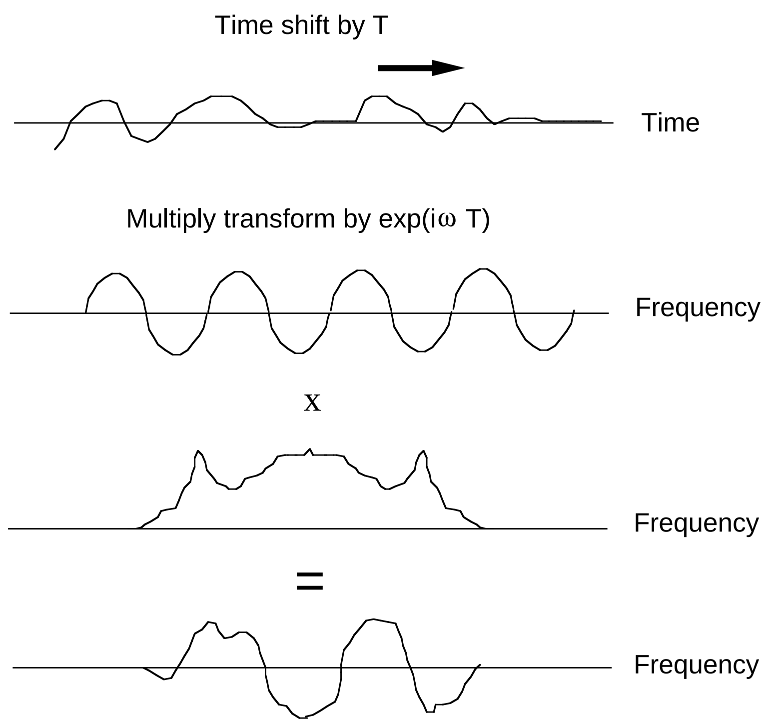

Figure 8: Time shift of a continuous band-limited signal. stew1-SEP-sem05 [NR]

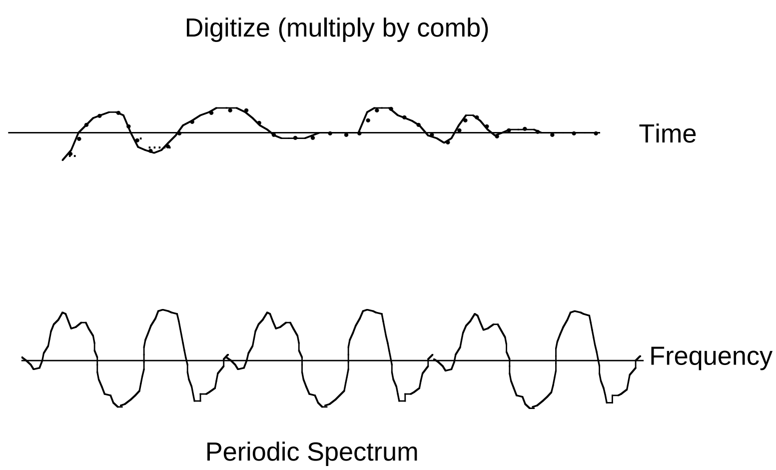

Figure 9: Sampling after the time shift of the continuous signal in Figure 8. stew1-SEP-sem06 [NR]

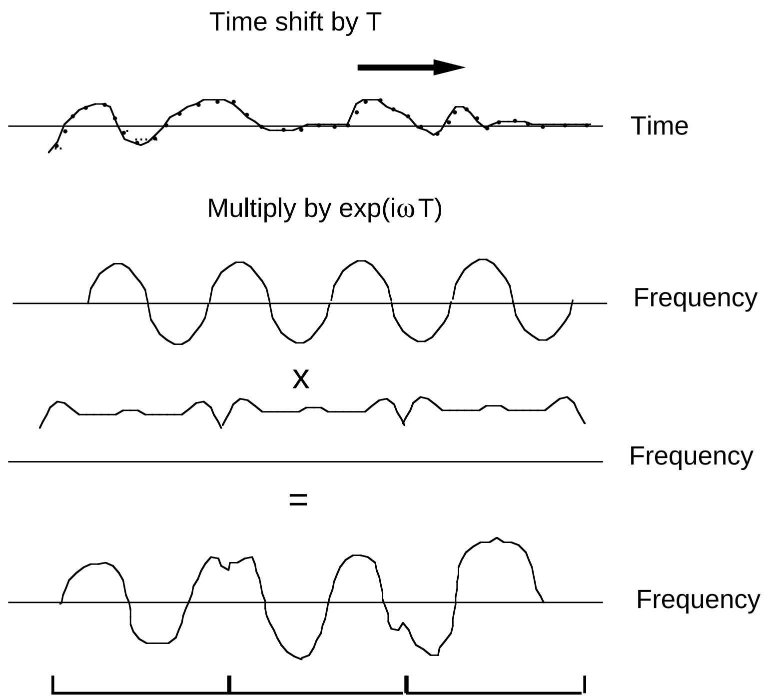

Figure 10: Time shift of the sampled signal for comparison with Figure 9. stew1-SEP-sem04 [NR]

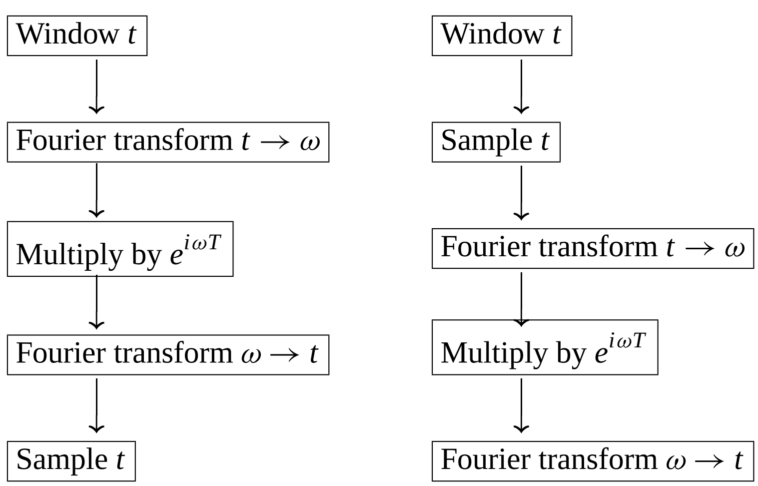

Figure 11: Shifting and discretizing on the time axis.

multiplying by a comb, commutes with the shifting exponential. Thus we can view the Fourier domain result either as a periodic replication of a shifted copy of the original spectrum or as a shifting of the periodically replicated band-limited spectrum. (Figure 12.) In Stolt migration,

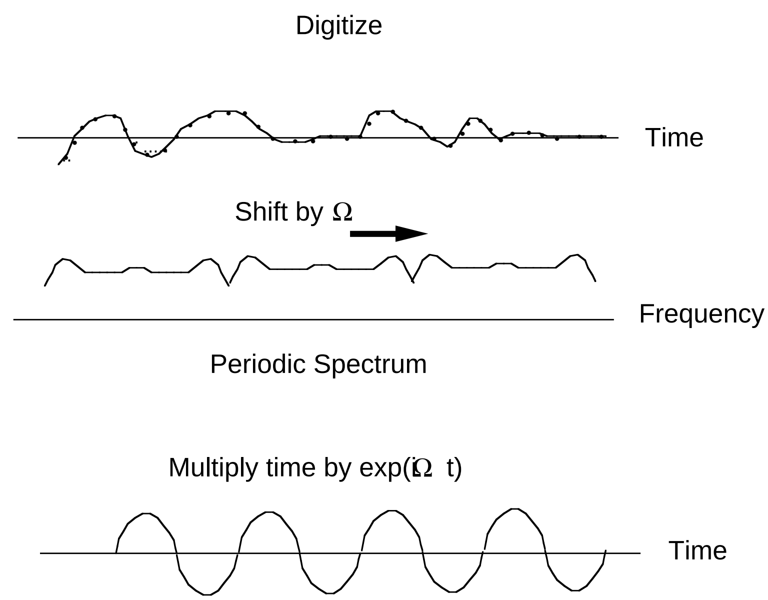

Figure 12: Frequency shift of a sampled signal. The sampling and shifting commute. stew1-SEP-sem08 [NR]

our aim is to apply the NMO mapping of equation **??** in the Fourier domain. Thus we need to shift the Fourier transform. Therefore, the proper prescription is multiplication of the sampled time series by $e^{-i\Omega t}$ followed by Fourier transformation. This is what the in-time group has advocated — evaluating the discrete Fourier transform (DFT) at $\omega - \Omega$. The frequency domain equivalent is convolution with the Fourier transform of the sampled function $e^{-i\Omega t}$. Popovici *et al.* (1993) give a formula for this in terms of the fast Fourier transform (FFT) of the original function. The discussion is not, however, complete until we include the effect of sampling on the frequency axis, i.e. using the inverse FFT algorithm. The general theory says this operation is multiplication by a comb function and hence is equivalent to convolving with a comb back in the time domain. The convolution has the effect of periodically replicating the *output* time series, a result of no consequence to us. (Figure 13.) The procedure has only one flaw: using the discrete Fourier transform to compute the sinusoidal components of the input time series for an arbitrary frequency. The inverse DFT is a *synthesis* operator; only under special conditions is it also the inverse of an *analysis* operator. That is, the DFT itself does not necessarily produce an exact representation of its input in terms of sinusoidal components. The interpolation formula the in-time group really wants to use is derived from

$$[DFT^{-1}]^*[FFT^{-1}] \tag{8}$$

and results in an implicitly-defined interpolation scheme in which the complex conjugate of their coefficients applied to the output of the Stolt mapping will produce the FFT'ed input to the Stolt mapping. On the other hand, if we apply the same idea to the input-oriented Stolt

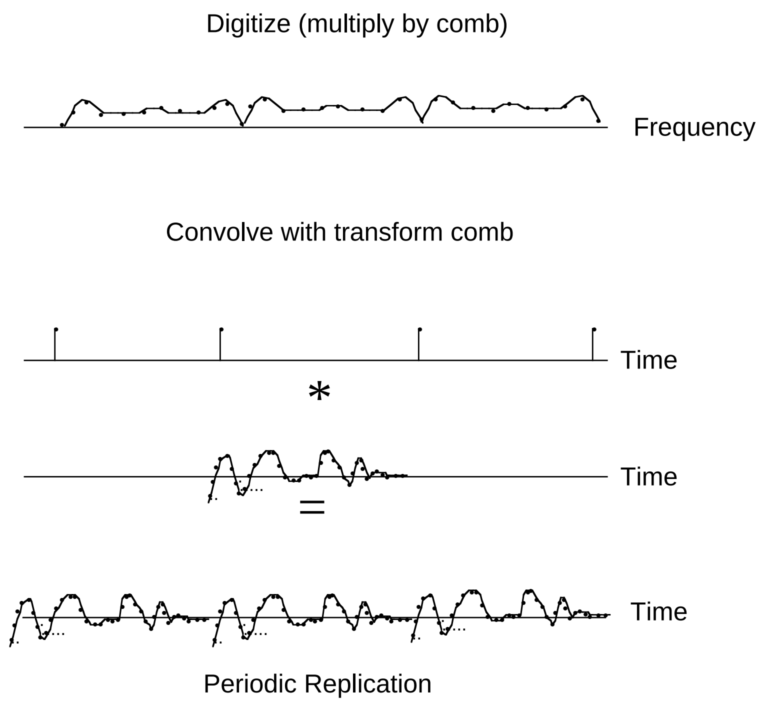

Figure 13: Sampling the spectrum replicates on the time axis. stew1-SEP-sem12 [NR]

mapping I have proposed, the interpolation is based on

$$[FFT][DFT]^* \tag{9}$$

and results in the conjugate coefficients being applied to the input to the Stolt mapping. So far I have been unsuccessful in tracking down a name in the literature for this interpolation.

CONCLUSIONS

I have described another correct variant of Stolt migration with interesting features. This method is implicit in Harlan's earlier work (Harlan, 1983). Like conventional Stolt migration, it sandwiches a frequency domain change of variable between fast Fourier transforms. Unlike conventional Stolt migration, it does not impose a Jacobian rescaling after frequency domain interpolation. I have further compared the Fourier domain interpolation schemes proposed in (Harlan, 1982) and (Popovici et al., 1993) and determined the latter as the more appropriate one to use, albeit not perfect.

ACKNOWLEDGMENTS

Thanks to Patrick Blondel, Biondo Biondi, Dave Nichols, Jun Ji, Mihai Popovici, and Francis Muir for clarifying my thoughts on this subject.

REFERENCES

Blondel, P., and Muir, F., 1993, Parallel computing exhumes slow Fourier transform in Stolt migration: SEP–**79**, 265–268.

Claerbout, J. F., 1985, What is the transpose operation?: SEP–**42**, 113–128.

Claerbout, J. F., 1992, Earth Soundings Analysis: Processing versus Inversion: Blackwell Scientific Publications.

Harlan, W. S., 1982, Avoiding interpolation artifacts in Stolt migration: SEP–**30**, 103–110.

Harlan, W. S., 1983, Linear properties of Stolt migration and diffraction: SEP–**35**, 181–184.

Levin, S., 1986, Test your migration IQ: SEP–**48**, 147–160.

Lin, J., Teng, L., and Muir, F., 1993, Comparison of different interpolation methods for Stolt migration: SEP–**79**, 255–260.

Popovici, A. M., Blondel, P., and Muir, F., 1993, Interpolation in Stolt migration: SEP–**79**, 261–264.

Stolt, R. H., 1978, Migration by Fourier transform: Geophysics, **43**, no. 1, 23–48.

Stolt, R. H., 1984, Comment on “A simple exact method of three-dimensional migration - Theory," by Jakubowicz, H., et al. (GPR-31-01-0034-0056): Geophys. Prosp., **32**, no. 2, 347–349.

APPENDIX A

AN EXPLICATION OF NMO

As usual, it was about two years after Jon Claerbout published an idea (Claerbout, 1992) that I grasped its true significance. In section 5.3 of PVI, Jon notes that NMO is a linear mapping and distinguishes between looping over input and output space for numerical implementation. Let us first return to the example given in the introduction of monotonic remapping of point masses. There I argued that the change of variable in this case uses no Jacobian to compensate for loss of mass, unlike continuous integration. If we model this remapping as integration against δ-functions, the result is given as the limit as $\epsilon \rightarrow 0$ of integration against test functions, δ_ϵ, of width ϵ and unit area. In this case we write

$$\sum_{a}^{b} f(x_j) = \lim_{\epsilon \rightarrow 0} \int_a^b \sum \delta_\epsilon (x - x_j)\, f(x)\, dx \quad . \tag{A-1}$$

Changing variable from $x \rightarrow y$, converts the formula to

$$\sum_{\tilde{a}}^{\tilde{b}} \tilde{f}(y_j) = \lim_{\epsilon \rightarrow 0} \int_{\tilde{a}}^{\tilde{b}} \sum \delta_\epsilon (y - y_j)\, \tilde{f}(y) \frac{dx}{dy}\, dy \quad . \tag{A-2}$$

Comparing equation **??** with **??**, reveals that

$$\delta(y(x)) = \frac{\delta(x - y^{-1}(0))}{y'(y^{-1}(0))} \quad , \tag{A-3}$$

which informally means that one should rescale by the Jacobian of the mapping. This prescription agrees with Harlan's discussion (1983) of linear properties of migration and diffraction. Common sense tells us, however, that all this rigamarole is aimed toward one goal — to preserve mass under change of variable. How then is this apparently contradictory conclusion reconciled with our simple, mass-preserving, point mass remapping? The answer lies in the difference between *shifting* and *distorting*. To understand this distinction, we need to take a first look at input versus output-oriented processing. In our simple approach, we take each input point mass and place it in its corresponding output location. More precisely, we shift it up or down to its output location. Since shifting is an area preserving, invertible operation, no Jacobian is involved. This is input-oriented processing. In the change of variable approach, we reach back from each output location $y(x)$ and grab the corresponding input sample $f(x)$. Doing so distorts the input by local stretching or squeezing, requiring us to correct the distortion by Jacobian rescaling. Following Claerbout, let us now look at these two approaches in a discrete matrix representation. Assume that sampling is sufficiently fine that nearest-neighbor interpolation suffices. In the input-oriented approach, each input sample corresponds to a column in the matrix containing a single 1 at the output index to which we want to shift the input sample. In the output-oriented formulation, each output sample corresponds to a row of the matrix containing a single 1 at the appropriate input index. Both of these mappings have problems: one when the NMO curve is flat and the other when it is steep. What we really

$$
\begin{pmatrix}
1 & 1 & 1 & 1 & & & & & & & \\
 & & & & 1 & 1 & & & & & \\
 & & & & & & 1 & & & & \\
 & & & & & & & 1 & & & \\
 & & & & & & & & & & \\
 & & & & & & & & 1 & & \\
 & & & & & & & & & & \\
 & & & & & & & & & 1 & \\
 & & & & & & & & & & \\
 & & & & & & & & & & 1
\end{pmatrix}
$$

Figure A-1: Input-oriented nearest-neighbor NMO. This form places a 1 in each column of the matrix.

$$
\begin{pmatrix}
1 & & & & & & & & & \\
 & & & 1 & & & & & & \\
 & & & & & 1 & & & & \\
 & & & & & & 1 & & & \\
 & & & & & & & 1 & & \\
 & & & & & & & 1 & & \\
 & & & & & & & & 1 & \\
 & & & & & & & & 1 & \\
 & & & & & & & & & 1 \\
 & & & & & & & & & 1
\end{pmatrix}
$$

Figure A-2: Output-oriented nearest-neighbor NMO. This form places a 1 in each row of the matrix.

want is to have a "footprint" in both input and output spaces (Jon Claerbout, pers. comm.). I believe another way to express this is that the interpolation coefficients should generally be laid neither vertically nor horizontally within the matrix. Instead they might be aligned locally tangent (or perpendicular?) to the NMO curve within the matrix. I have not yet pursued this line far enough to report success or failure. Francis Muir suggests splitting up the NMO into a cascade of two pieces, each applying half the NMO. Half the NMO means NMO with a velocity $\sqrt{2}$ higher than normal. There are a number of ways this could be done by combining two of the following operators:

$$
\begin{array}{l}
NMO \\
INMO^T \\
INMO^{-1} \\
[NMO^T]^{-1} \\
[NMO^{-1}]^T
\end{array}
$$

An obvious choice is to do half input-oriented plus half output-oriented processing, that is, $NMO\ INMO^T$ or $INMO^T\ NMO$. These choices are illustrated in Figures A-3 through A-6. As we see from the results of processing the constant-amplitude synthetic, output-oriented nearest-neighbor normal moveout has the best response. This is also apparent on the corresponding seismic traces. Unfortunately, NMO differs from windowed processing — we cannot directly apply the constant-amplitude synthetic results as a weighting function to correct the amplitude "glitches" on the data trace. The reason is clear upon inspecting the data traces — the anomalous trace amplitudes are nowhere near two or more times larger than surrounding values. The proper weights are frequency-dependent. This fact does not mean the idea of splitting up the NMO into pieces is a bad one — but it does strongly suggest that good combinations will be implicit, not explicit, ones. This, too, remains to be explored.

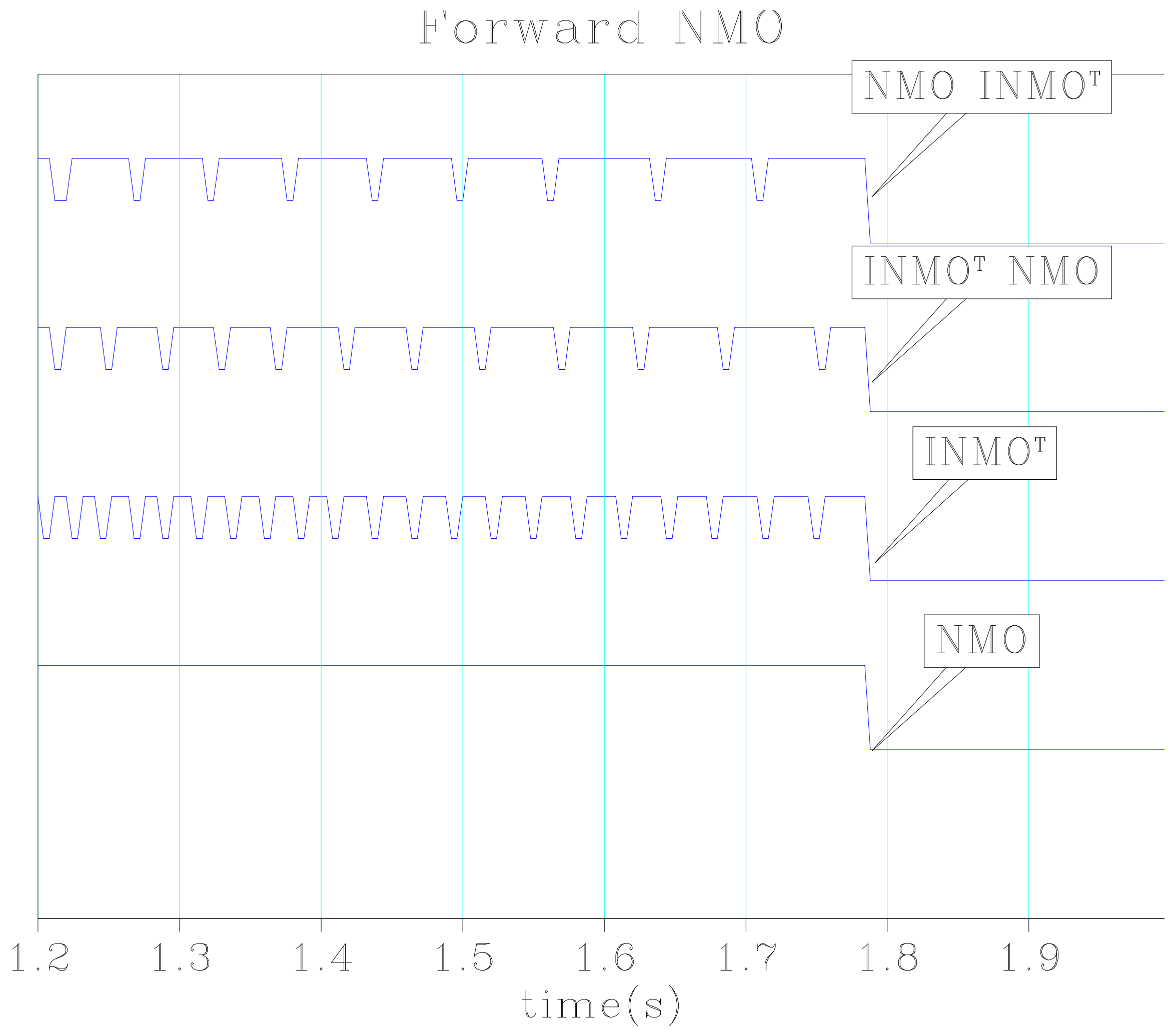

Figure A-3: A variety of NMO combinations applied to a constant-amplitude synthetic. stew1-synf [ER]

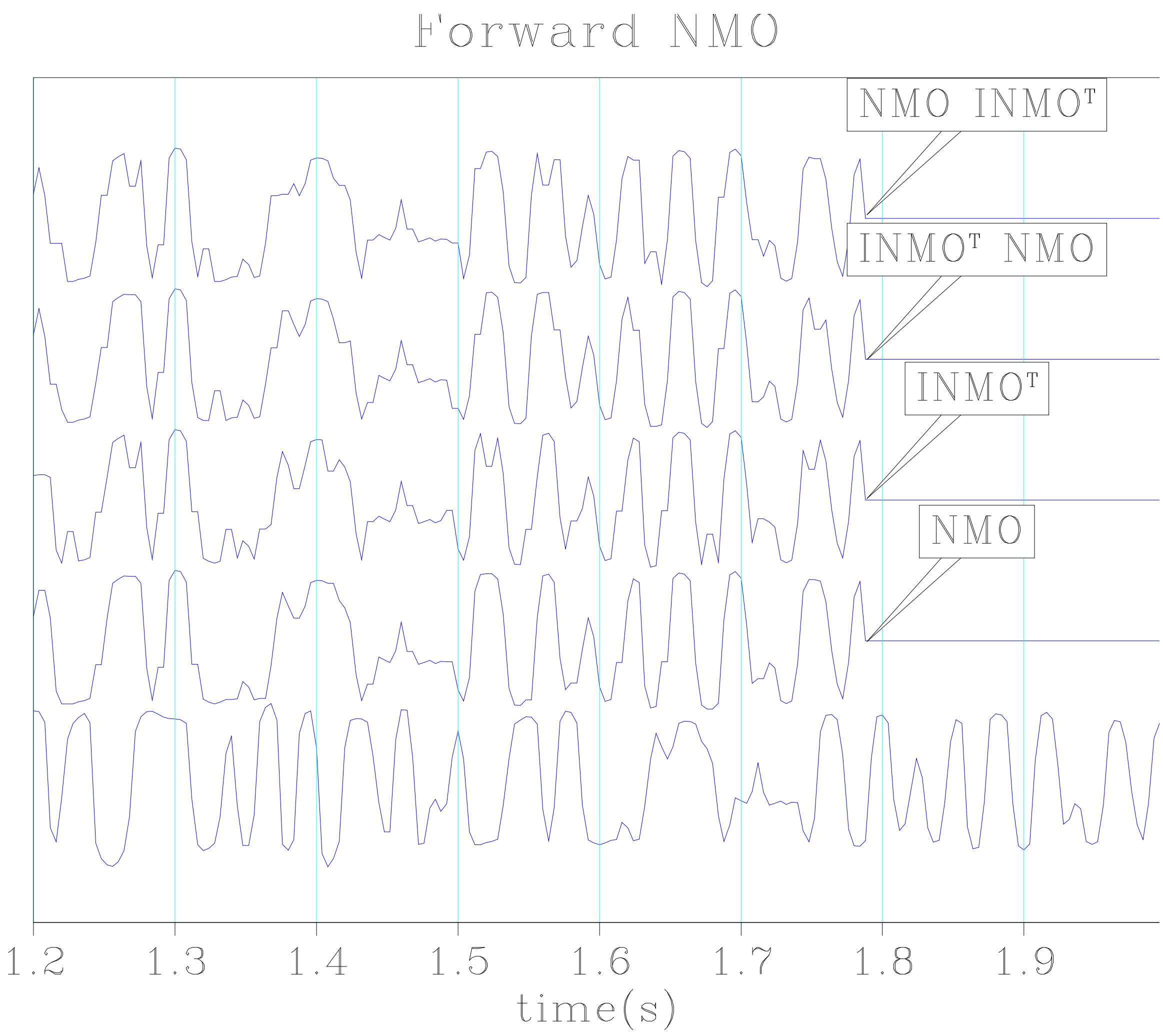

Figure A-4: A variety of NMO combinations applied to a field data trace. stew1-fldf [ER]

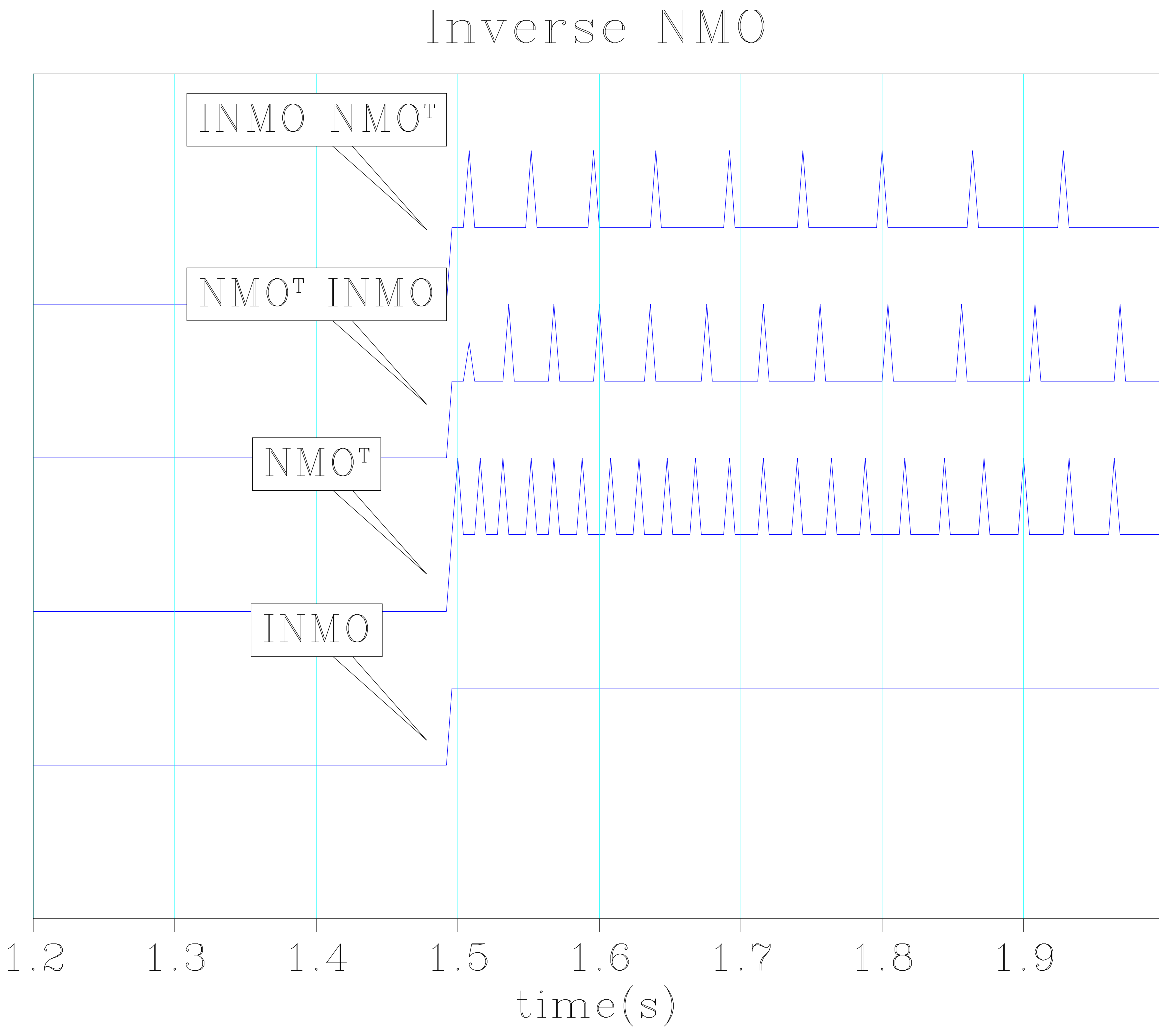

Figure A-5: A variety of inverse NMO combinations applied to a constant-amplitude synthetic. stew1-syni [ER]

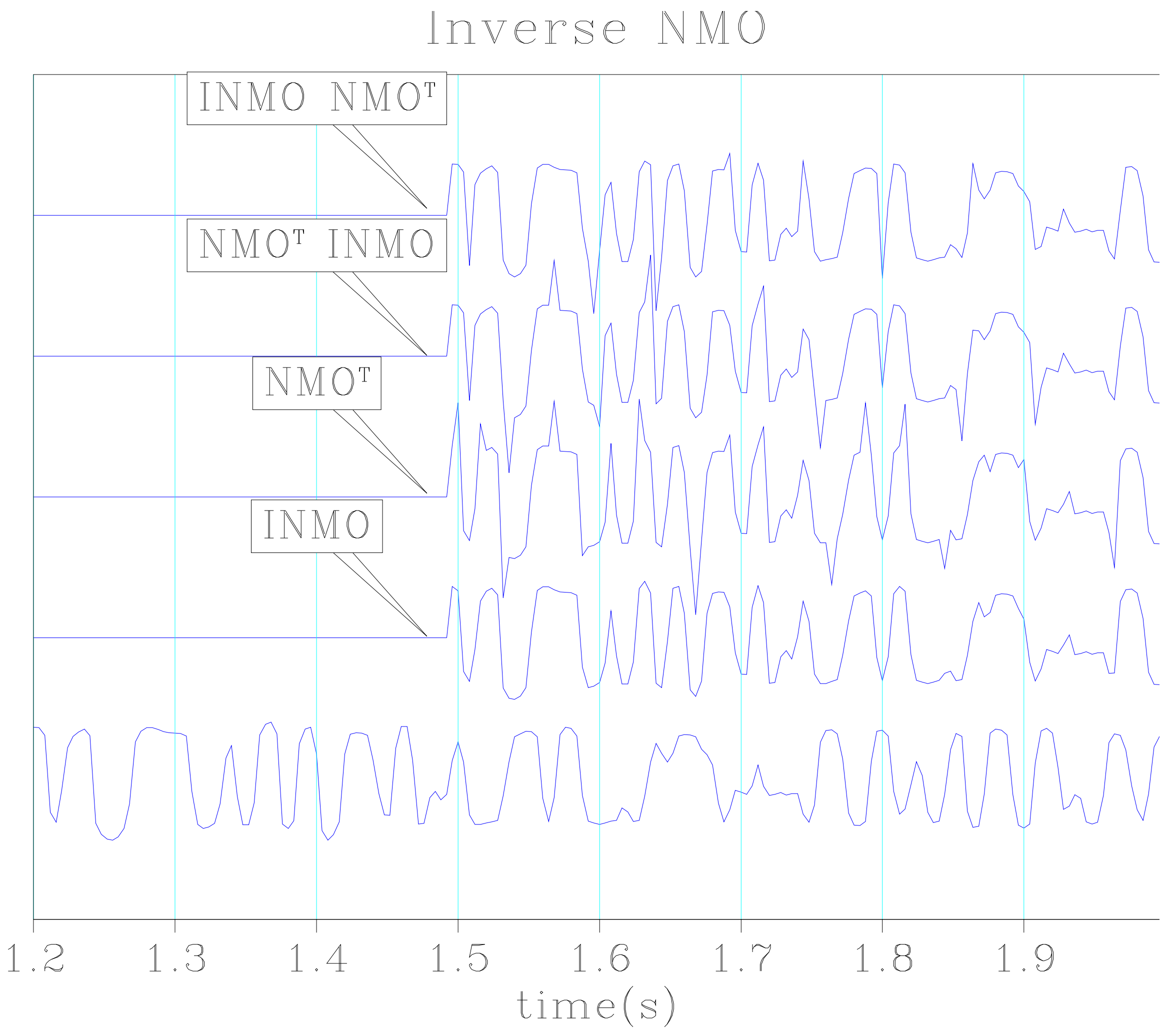

Figure A-6: A variety of inverse NMO combinations applied to a field data trace. stew1-fldi [ER]