Preventing Noise Alignment in Cross-correlation

Charles Ursenbach and John Bancroft

Summary

It has been shown (e.g., Cox, 1999, Section 7.9.2) that one difficulty of the cross-correlation procedure in trim statics is that it is possible to have spurious reproduction of signal. This phenomenon is quantified and is shown, in the limits of large fold, small correlation window, and large maximum allowable shift, to behave as a simple function of these variables for physically reasonable wavelet lengths. These results allow one to predict what choices of cross-correlation parameters are likely to result in spurious alignment of noise. Quantities are then defined that may be easily determined for real data. Their behaviour under conditions of signal alignment and noise alignment are elucidated. It is shown that these quantities, and particularly their trends with increasing window size or maximum allowable shift (two easily controlled cross-correlation parameters), can help to indicate whether growing signal is a result of the desired signal alignment, or simply constructed from random background noise.

Introduction

Trim statics are generally carried out with a very small maximum allowable shift. Even when cross-correlation is carried out as the first step in a residual statics calculation, the maximum shift is on the order of a wavelength. Occasionally, it is unclear whether apparent structure is real, or simply the result of very large residual statics. In these cases one might wish to employ a larger than usual maximum allowable shift. There is a well-known danger in such a procedure that one can appear to be aligning signal when in fact random noise is being aligned instead. This is illustrated in Figure 1, in which a stack of band-limited random noise has been trimmed using a large maximum allowable shift, so that the stack closely mimics the displayed pilot trace used in the trim statics.

The purpose of this study is, first, to develop a quantitative understanding of how noise is aligned, and, secondly, to use this understanding to develop tools which allow one to safely choose the largest possible maximum allowable shift.

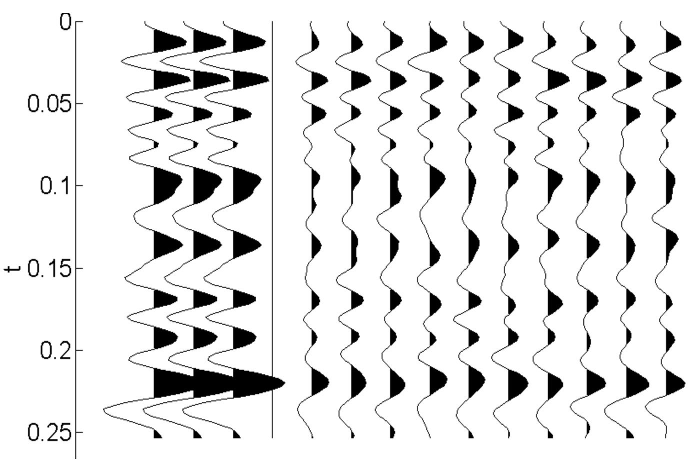

Figure 1. A pilot trace (repeated three times for visibility) and a trimmed stack. Each of the 10 stacked traces consists of a stack of 50 random noise data traces that have been cross-correlated against the displayed pilot trace.

Theory and Method

To generate a random noise data trace, every sample is assigned a random number between –1 and 1. The trace is then cubed, and this result is convolved with an Ormsby wavelet of duration *v*. A synthetic pilot trace was produced in this manner. To create a synthetic *n*-fold CMP stack of *M* traces, a total of *nM* data traces were generated by the above procedure. When these traces are stacked, the result represents the NMO corrected CMP stack in a typical trim statics procedure. Trim statics are then applied as usual, calculating cross-correlation functions between individual data traces and the pilot trace. The individual traces are shifted by the time delay corresponding to maximum cross-correlation amplitude, and then restacked. The output of principal interest in this study consists of the cross-correlation coefficient (ccc) between traces of the new stack and the pilot trace. This quantity is averaged over the *M* calculations (which all use the same trim parameters) and is denoted <ccc>. <ccc> = 0 implies that no signal has been generated, while <ccc> = 1 implies that the pilot trace has been perfectly reproduced. For Figure 1, <ccc> = 0.90, congruent with the obvious visual similarity between the pilot trace and the stack. The signal to noise ratio (SNR) of the final stack may also be estimated as

$$\mathrm{SNR} = \langle ccc \rangle / (1-\langle ccc \rangle). \qquad (1)$$

The variables that are adjusted for each calculation are the window size (w, in seconds), the maximum allowable shift (t_{max}, in seconds), the number of traces to be stacked (n), and the wavelet duration (v, in seconds). Other variables that we hold fixed are the number of calculations used in the averaging (M = 10) and the wavelet parameters (f1 = 0.4/v, f2 = 0.8/v, f3 = 4.0/v, f4 = 5.6/v).

By carrying out a number of calculations and analyzing their behavior the following expression was empirically derived for $t_{max} > 0.06$s, $w/n < 0.03$s, and $0.02\text{s} < v < 0.16$s:

$$ccc \approx 1 - 0.18[1 + 2/\sqrt{2(t_{\max}/v)+1}]\sqrt{(w/v)/n} \quad (2)$$

The values of t_{max} and w are scaled by v, so that the result is dimensionless. Although Equation (2) is empirical, it is reasonable for 1–ccc to depend on $1/\sqrt{n}$, as a non-zero t_{max} essentially imbues the trace with an effective signal, and this signal is enhanced as $\sqrt{n}$. However, if the correlation window size is doubled, there are twice as many points to align, and twice as many traces are required to produce the same result. This rationalizes the observed w/n dependence. For a trace of random points, not convolved with a wavelet, shifting a trace with displacements of $\pm\, t_{max}$ relative to the pilot trace would be analogous to comparing the pilot trace to $(2t_{max}/dt)+1$ different traces (dt is the sample rate). Convolving with a wavelet results in an auto-correlation length of ~v, so that shifting the trace by $\pm\, t_{max}$ would be similar to comparing the pilot trace to $(2t_{max}/v)+1$ different traces. Thus the observed t_{max} dependence is rationalized by analogy to the dependence on n. Some of the numerical parameters, such as 0.18, might be replaced by functions of v, but the dependence appears to be weak, and Equation (2) was found to be reasonably accurate.

One immediate use of Equation (2) is to confirm that a chosen set of cross-correlation parameters will not result in significant noise alignment (i.e., if Equation (2) yields a value of, say, ccc < 0.5, or SNR < 1, then the chosen parameters are reasonable).

A second use, as will be shown next, is to discern after a cross-correlation procedure whether signal or noise has been preferentially aligned. This allows one to use parameters which predict ccc > 0.5, and to tell afterward whether the choice was justified. To do this it is useful to define two other quantities, in addition to ccc. The first is the *amplitude ratio*. This is the average of the ratio of zero-time autocorrelations of the stack traces and the pilot trace.

$$\text{amplitude ratio} = \langle C_{stack}(0) / C_{pilot}(0) \rangle \quad (3)$$

This quantity is useful because two traces that are simply proportional to each other (i.e., not having the same absolute amplitude, but identical in all other respects) will have a cross-correlation coefficient of 1. The amplitude ratio will, in such a case, be sensitive to the differing amplitudes, which will decrease in the case of noise alignment as seen in Figure 1. For this quantity to be meaningful of course the traces should be scaled to a common rms amplitude.

The second quantity to define is the *relative shift*. This is the average of the average absolute value shift of the data traces, divided by one half of the maximum allowable shift. It may be written

$$\text{relative shift} = \langle \text{avg}(|t_{opt}|)\rangle / (t_{max}/2). \quad (4)$$

Dividing by $t_{max}/2$ normalizes this quantity to unity when t_{opt} is chosen randomly from the allowed shift interval, as is the case when random noise is being aligned. This quantity can take on values between 0 and 2, but when t_{max} is much larger than the width of actual signal statics, the relative shift will tend to zero for properly aligned signal.

To test the utility of these measures, synthetic data was created to model mixed signal and noise traces. Random noise pilot traces were created as before, but data traces for the stack were created from a mixture of the pilot trace (displaced by a static chosen from a Gaussian distribution) and an independently generated noise trace. Because synthetic data was employed it was possible to generate a fourth quantity, a realignment measure, which was 0 for uncorrelated stacks, and varied from –1 for perfect signal alignment to 1 for pure noise alignment. It employed information obtained from creating the synthetic data and of course is not available with real data. It was used simply to test the validity of conclusions.

Example

In Figures 2 and 3 below, cross-correlation results are given for synthetic data in which data traces are 50% pilot trace, offset by a Gaussian distribution whose width is equal to the wavelet length (v = 0.08s), and 50% independent random noise. In Figure 2, where t_{max} is the independent variable, the large window samples have aligned signal and the small window samples have aligned noise, according to the realignment measure. This is manifested clearly in the ccc, which follows Equation (2) for small w, and in the relative shift, which vanishes at large t_{max} for large w. In Figure 3, where w is the independent variable, it is clear from the realignment that signal is being aligned for sufficiently large t_{max}. This is manifested by the amplitude ratio in which lines are grouped by their value of t_{max}. When noise is being aligned, the plots tend to group according to the value of n. An example of this is shown in Figure 4, for which the initial data is 33% pilot trace and 67% random noise. These are therefore useful measures when attempting to use the largest possible maximum allowable shift in cross-correlation.

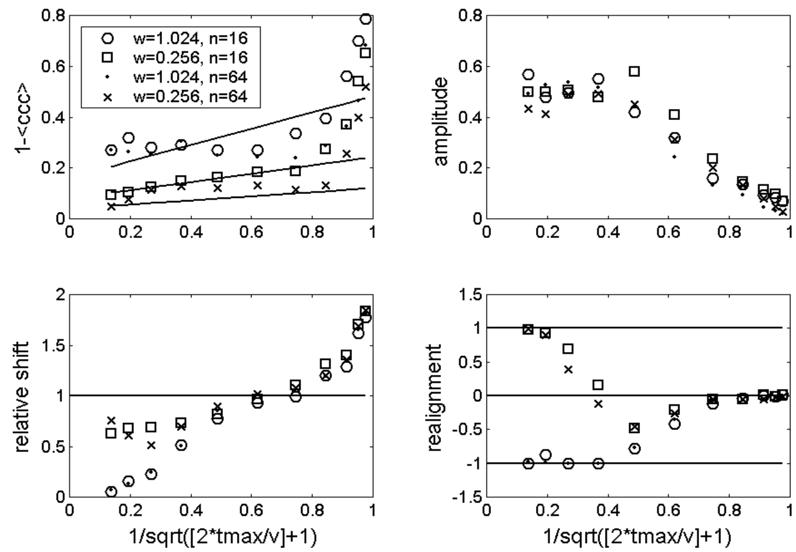

Figure 2. The dependence on t_{max} of quantities defined in the text. Each point represents a calculation for one set of parameters. The solid lines represent Equation (2).

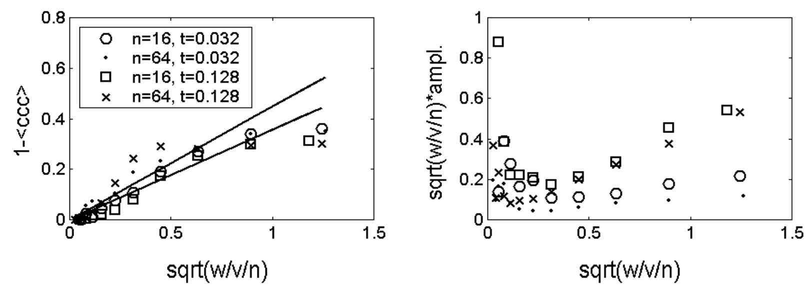

Figure 3. The dependence on *w* of quantities defined in the text.

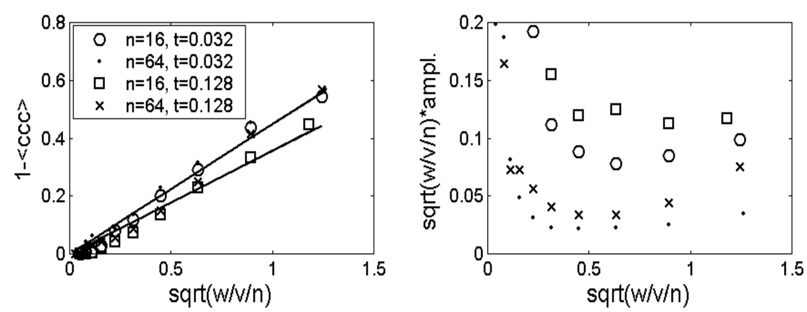

Figure 4. The same as Figure 3 except for lower signal-to-noise in the initial data.

References

Cox, M., 1999. Static Corrections for Seismic Reflection Surveys. SEG, Tulsa, OK.

Ursenbach, C.P., and Bancroft, J.C., 2000. Noise alignment in trim statics. CREWES Research Report, Volume 12.

Ursenbach, C.P., 2000. Distinguishing noise alignment from signal alignment in statics calculations. CREWES Research Report, Volume 12.