Noise alignment in trim statics

Charles P. Ursenbach and John C. Bancroft

ABSTRACT

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. A quantitative expression for the cross-correlation coefficient of purely random traces is obtained which appears to be valid within certain limits. These results are of particular value in that they allow one to predict what choice of cross-correlation parameters are likely to result in spurious alignment of noise.

INTRODUCTION

Individual traces possess small (and in some cases large) time shifts relative to each other. Statics calculations attempt to correct for these errors. A statics procedure typically begins as follows: First, traces with the same midpoint are stacked together to form a reference or pilot trace. Second, the individual traces are scaled to a common rms amplitude. Third, a time cross-correlation function is calculated between each individual trace and the reference. Finally, the time shift corresponding to the largest cross-correlation value is identified for each trace. This process of picking the time shifts is the focus of this paper.

When calculating the cross-correlations, it is important to keep in mind the role of the maximum allowable shift (which we will denote t_{max}). If too small of a t_{max} is chosen, then the algorithm will not be allowed to explore enough of the cross-correlation function to properly correct the traces. If too large of a t_{max} is chosen, incorrect alignment can occur through processes such as cycle skipping (Yilmaz, 1987; Cox 1999).

The purpose of this paper is to study the ability of the time-pick algorithm to spuriously align noise in individual traces in order to reproduce a reference trace. To do this we synthesize traces out of pure random noise and apply the cross-correlation procedures. From this we are able to provide quantitative guidelines that will help to prevent noise alignment with real data.

DESCRIPTION OF CALCULATION

The study is based on several series of calculations. Each calculation involves generating M reference traces, and, for each reference trace generating n individual traces to be stacked. A correlation function is calculated between each reference trace and each one of the traces in its stack, and, within a given maximum allowable shift (t_{max}), the optimal shift is located which gives the maximum cross-correlation. The traces are then stacked using their optimal shifts, and the stacked trace is then cross-correlated with the reference trace to obtain an estimate of the similarity of the two.

The output of the calculation consists of the cross-correlation coefficient (ccc) between the stacked trace and the reference trace averaged over the M reference traces. We will denote this <ccc>, where <…> indicates an average over the M reference traces. <ccc> = 0 implies that no signal has been generated, while <ccc> = 1 implies that the reference trace has been perfectly reproduced. The signal to noise ratio (SNR) may be estimated as

$$\mathrm{SNR} = <ccc> / (1-<ccc>). \tag{1}$$

The following algorithm generated each of the M reference traces for each calculation. Each point in the .002s sample rate trace was assigned a random number between −1 and 1. The trace was cubed in order to thin out the number of major peaks while preserving the sign. This result was convolved with an Ormsby wavelet of duration v. The trace was then multiplied by a window function which has a value of 1 from $-w/2$ to $w/2$, and 0 elsewhere. This defines a correlation window of duration w. Because the reference trace is truncated to zero outside of the window, we are able to use an FFT procedure to carry out the cross-correlation more efficiently.

Each reference trace was associated with a stack of n traces. These were generated in the same fashion as the reference trace, but with a length of $w+2t_{max}$. For zero time shift, the center of the reference trace was assumed to coincide with the center of the traces to be stacked. We were thus able to model displacements of up to $\pm t_{max}$. These traces are intended to model random noise traces.

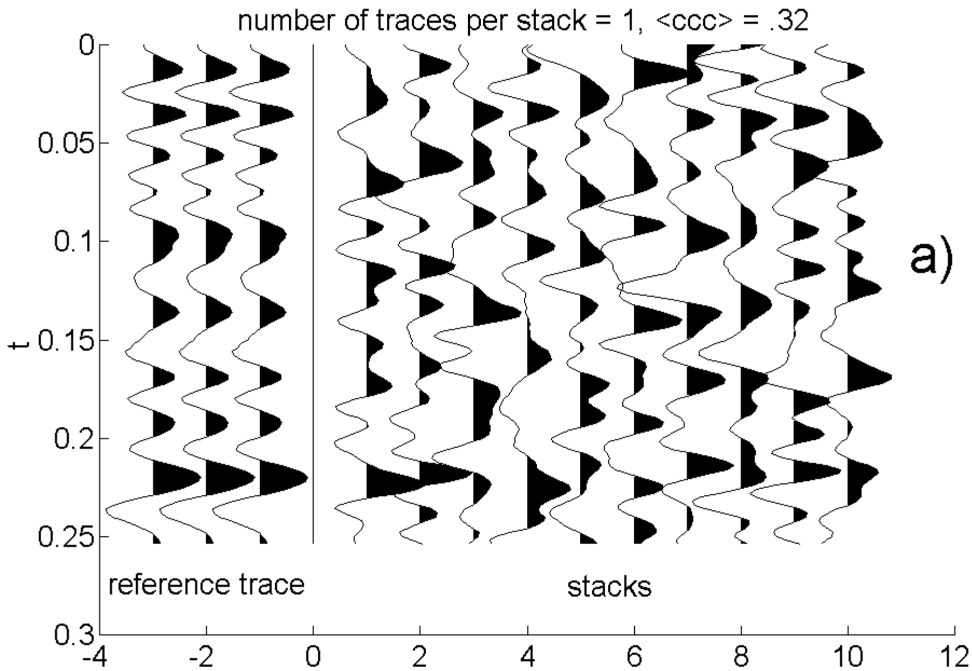
The variables that are adjusted for each calculation are the window size (w, in seconds), the maximum allowable shift (t_{max}, in seconds), the number of traces to be stacked (n), and the wavelet duration (v, in seconds). Other variables that we hold fixed are the number of reference traces (M = 10)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

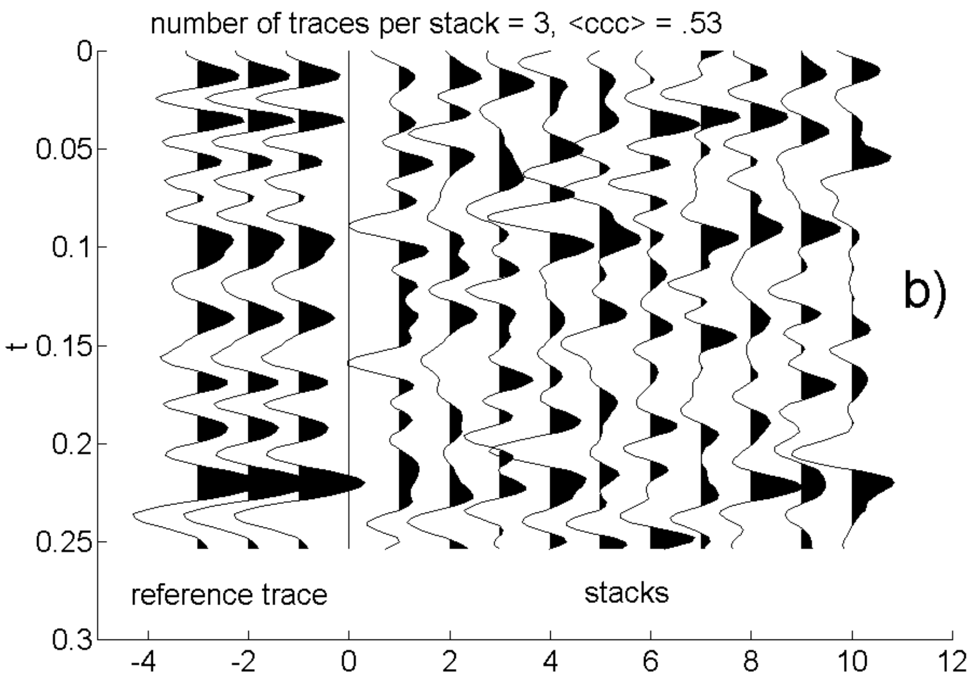
We present the data as plots of <ccc> as it varies with combinations of w, v, n, and t_{max}. We also present some representative collections of reference traces and corresponding synthetic stacked data.

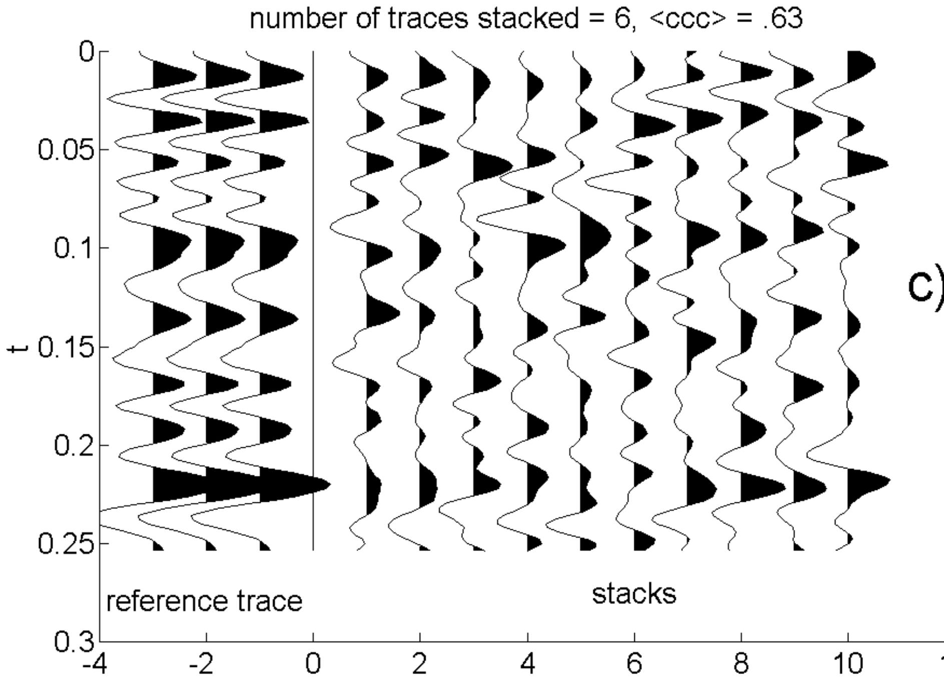
RESULTS OF CALCULATION

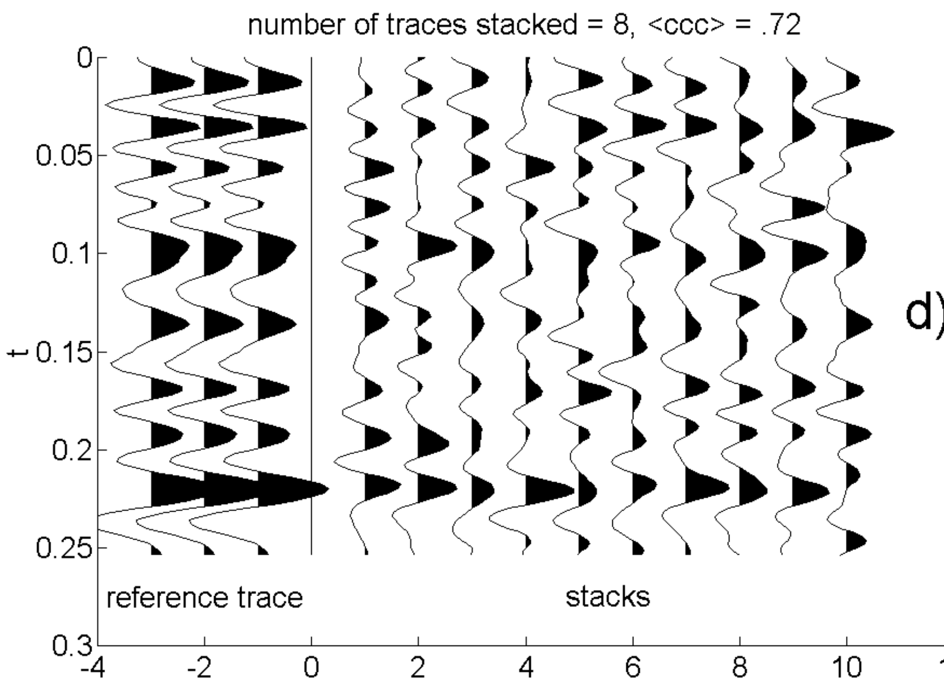
Degree of alignment for representative ccc values

Figures 1a through 1e each illustrate ten different stacks produced by the correlation procedure with a given reference trace. These particular calculations used parameters of w = .256s, t_{max} = .032s, and v = .08s. Values of n vary from n = 1 in Figure 1a to n = 50 in 1e.


number of traces per stack = 1, <ccc> = .32
a)
t
reference trace
stacks
0
0.05
0.1
0.15
0.2
0.25
0.3
-4
-2
0
2
4
6
8
10
12


number of traces per stack = 3, <ccc> = .53
b)
t
reference trace
stacks
0
0.05
0.1
0.15
0.2
0.25
0.3
-4
-2
0
2
4
6
8
10
12


number of traces stacked = 6, <ccc> = .63
t
0
0.05
0.1
0.15
0.2
0.25
0.3
-4
-2
0
2
4
6
8
10
12
reference trace
stacks
c)


number of traces stacked = 8, <ccc> = .72
t
0
0.05
0.1
0.15
0.2
0.25
0.3
-4
-2
0
2
4
6
8
10
12
reference trace
stacks
d)

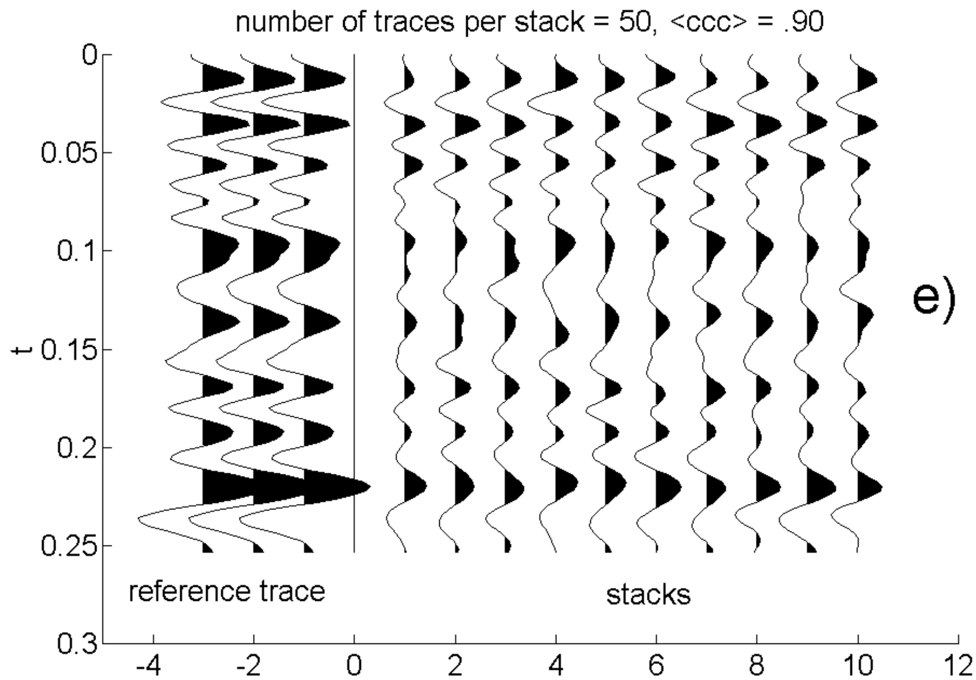


Figure 1. In each part, a)-e), of this figure, n randomly generated traces are each correlated with the given reference trace (which is repeated three times for visibility). The n traces are then stacked, and the cross-correlation coefficient (ccc) is calculated between the reference and the stack. This process is repeated for nine more sets of n traces to give a total of ten stacks (shown in each part) and ten ccc values, which are averaged to obtain <ccc>. It is clear that as <ccc> increases, the reproduction of the reference trace improves as well. All calculations used the parameters w = .256s, t_{max} = .032s, and v = .08s. The SNR for each set of stacks as estimated from Eq.(1) is a) 0.47 b) 1.1 c) 1.7 d) 2.6 e) 9.0.

It is clear that for a large enough n the random noise can reproduce the reference signal. As a rough guide we might say that at ccc = .9 there is good alignment, at ccc = .7 there is moderate alignment, and at ccc = .5-.6 there is emerging alignment. For ccc = .3 there is essentially no alignment. We now turn to delineating the dependence of the ccc on n, w, and t_{max} for values of v between .02s and .16s.

ccc dependence on w and n

In this section we establish that ccc varies as the square root of w/n, and that w/n and t_{max} are variables independent of each other. To show this we begin by plotting values of <ccc> against w and n.

In Figure 2 the dependence of <ccc> on n is given for v = .08 and for a selection of values of w and t_{max}. In Figure 3 the dependence of <ccc> on w is given for v = .08 and for a selection of values of n and t_{max}.

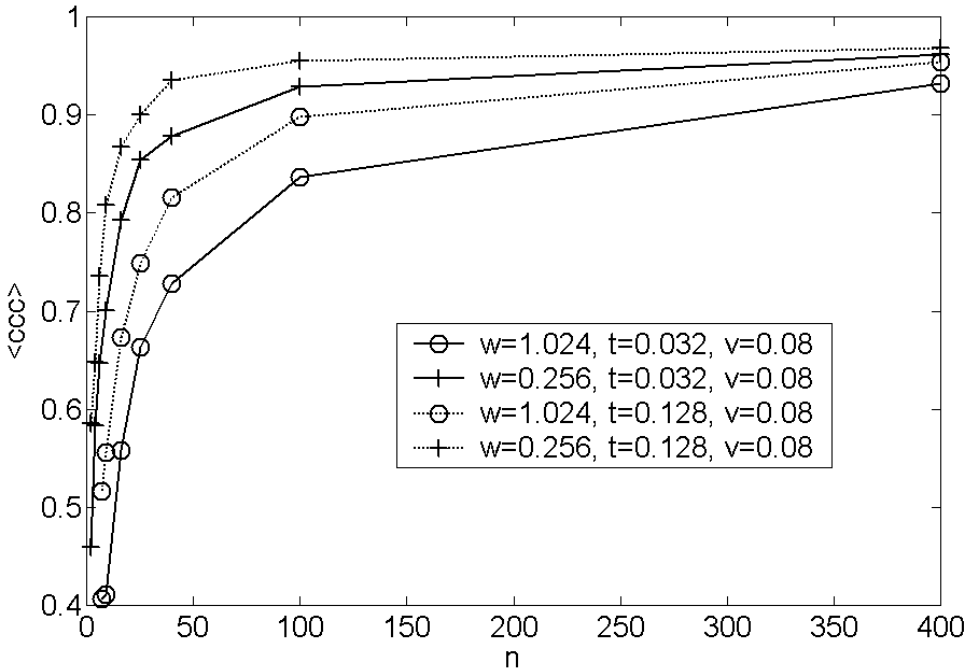


Figure 2. The variation with number of stacked traces (n) of the averaged cross-correlation coefficient (<ccc>) between a reference trace and stacked, correlated random noise. Note that in the plots t is the same as t_{max}.

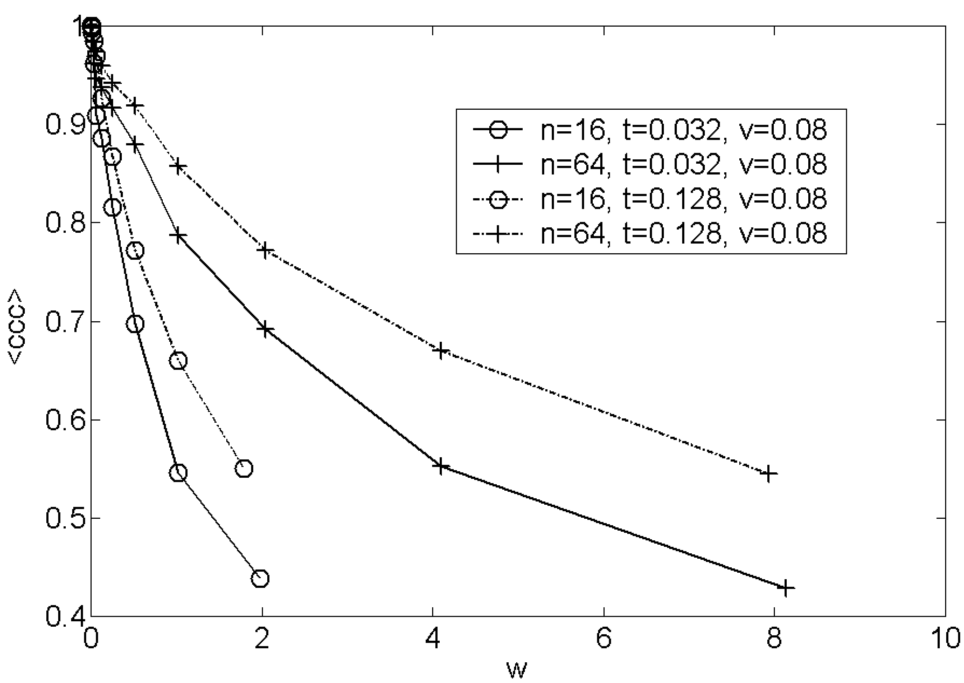


Figure 3. The variation with correlation window (w, in seconds) of the averaged cross-correlation coefficient (<ccc>) between a reference trace and stacked, correlated random noise. Note that 1–<ccc> appears to vary as $\sqrt{w}$.

After considering the behaviour of the data in these two figures, it is appropriate to recast them in Figures 4 and 5, where the same information is given but in the form of (1−<ccc>) vs. $\sqrt{(w/v/n)}$, where v is included to scale w to a unitless quantity. In Figure 4 the value of w is fixed in each plot and only n varies, while w varies and n is fixed in Figure 5. At least two aspects of these graphs are noteworthy. First, a linear approach to the origin is apparent both for large n and for small w. Second, the lines are grouped according to the value of t_{max}, not only within each graph, but between graphs as well.

It is also apparent that the linearity degrades slightly for the very largest values of n and very smallest values of w. In the case of w this is because the traces are in numerical form, and <ccc> must take on a value of unity for w = dt. Thus in Figure 5 the curve tends to zero slightly before the origin.

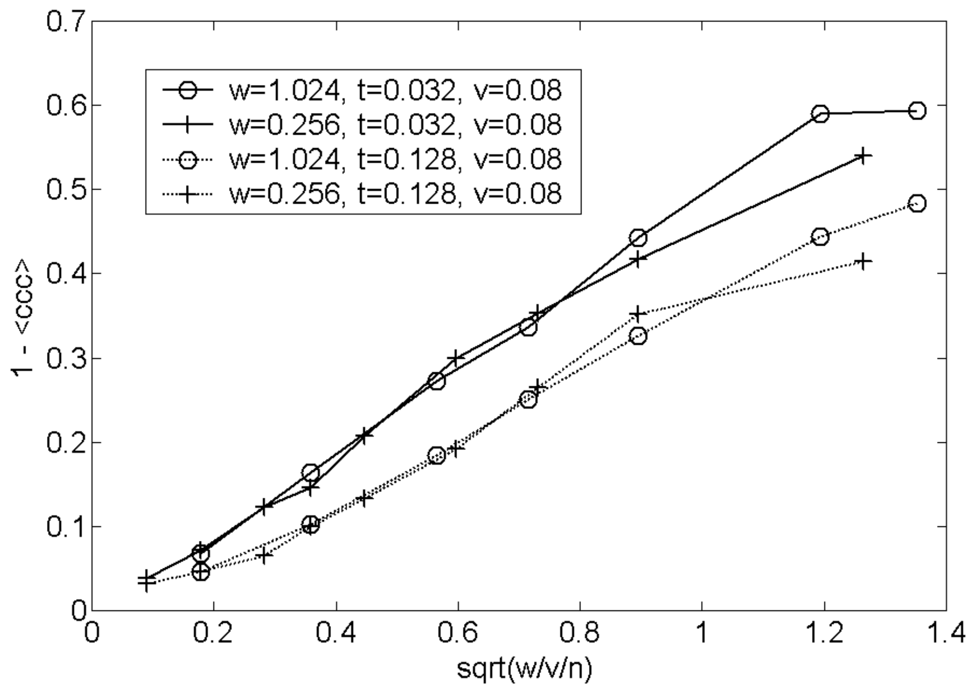


Figure 4. The variation with number of stacked traces (n) of the averaged cross-correlation coefficient (<ccc>) between a reference trace and stacked, correlated random noise. This is the same data as in Figure 2, but 1–<ccc> is plotted instead of <ccc>, and $1/\sqrt{n}$ instead of n. Furthermore, n is scaled by v/w. Note the linear behaviour for large n, and that plots are grouped by their value of t (same as t_{max}).

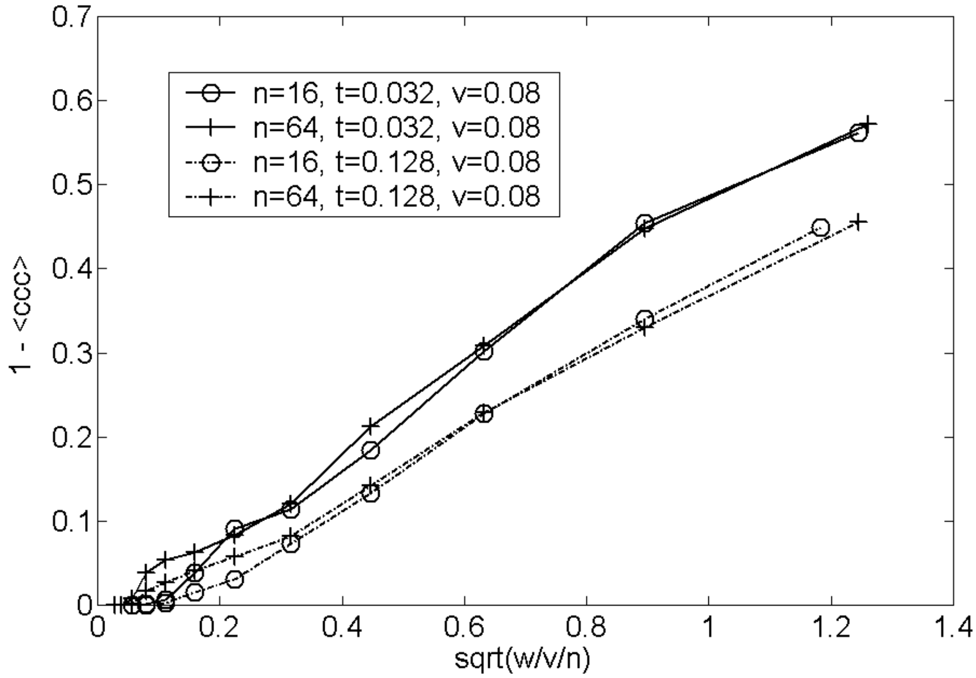


Figure 5. The variation with correlation window (w) of the averaged cross-correlation coefficient (<ccc>) between a reference trace and stacked, correlated random noise. This is the same data as in Figure 3, but 1–<ccc> is plotted instead of <ccc>, and √w instead of w. Furthermore, w is scaled by 1/(nv). Note that lines are grouped by their value of t (same as t_{max}) in this Figure, and also that lines in Figure 4 with a given value of t are similar in the linear region to lines in this Figure with the same value of t.

The linear behaviour of these graphs can be readily rationalized. Allowing a non-zero t_{max} for random noise traces endows them with an effective non-zero signal-to-noise ratio. Noise in averaged samples generally decreases as $1/\sqrt{n}$, and this is precisely what is observed here. However, if the correlation window size is doubled, there are twice as many points to align with, and thus twice as many traces are required to produce the same result. On the basis of these plots then it seems reasonable to write the large-n, small-w limit of <ccc> as

$$1 - \langle ccc \rangle = f(t_{\max}, v)\sqrt{(w/v)/n} \tag{2}$$

where, according to the above figures, f is a decreasing function of t_{max}, as would be expected. We next delineate this function.

ccc dependence on t_{max}

Can we make some guesses as to the likely form of f? For a trace of random points, not convolved with a wavelet, displacing a trace over an interval of $2t_{max}$ relative to the reference trace would be analogous to comparing the reference 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 over an interval of $2t_{max}$ would be similar to comparing the reference trace to $(2t_{max}/v)+1$ different traces.

Thus, by analogy to the <ccc> dependence on n, it is reasonable to attempt a plot of (1−<ccc>) vs. $1/\sqrt{(2t_{max}/v+1)}$. This exercise is carried out in Figure 6.

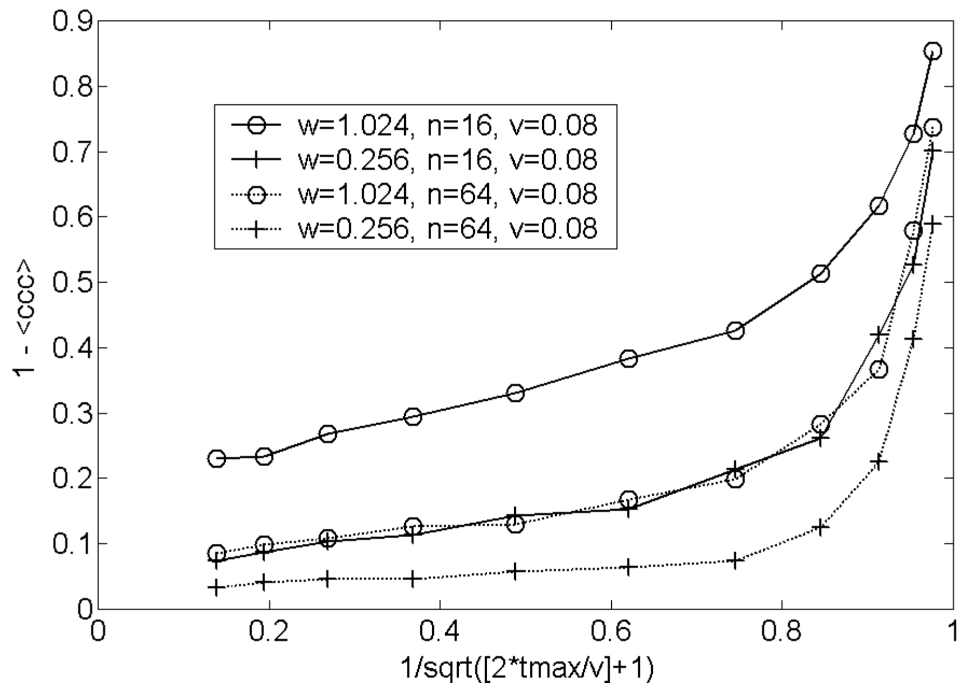


Figure 6. The variation with maximum allowable shift (t_{max}) of the averaged cross-correlation coefficient (<ccc>) between a reference trace and stacked, correlated random noise. Note that the behaviour is linear for large t_{max}, but that the plots do not approach the origin.

The behaviour is indeed linear, but does not intersect closely to the origin. As expected, the lines are grouped according to their value of w/n. One could further specify Eq. (2) in its large-t_{max} limit as

$$1-<ccc>=[A+B/\sqrt{(2t_{max}/v)+1}]\sqrt{(w/v)/n} \tag{3}$$

where A and B are unknown constants, which may depend on v, but not on w or n.

A and B can be evaluated by obtaining the slope and intercept of the linear part of each line in Figure 6, and dividing those values by $\sqrt{(w/v/n)}$, then averaging the results. This gives A = .14±.05 and B = .36±.08 for v = .08s. Note that A and B are both dimensionless.

How sensitive are A and B to the value of v? We have repeated the calculation above for data obtained using v = .02s and v = .16s, with results for (A,B) of (.21±.03,.38±.09) and (.12±.06,.31±.12) respectively. From this it appears that both A and B decrease with v, but on the other hand, taking the error limits into consideration, all of the above would be consistent with A = .18 and any B in the range .29 to .43. We conclude that in the range of physically relevant wavelets, a convenient expression for the cross-correlation coefficient is

$$ccc \approx 1 - 0.18\left(1 + \frac{2}{\sqrt{(2t_{\max}/v)+1}}\right)\sqrt{\frac{w/v}{n}} \qquad (4)$$

From the regions of Figures 4 through 6 where linear behaviour is observed to begin (and from similar data for v = .02s, .16s not shown), we surmise that this expression should be applicable at least for t_{max} >0.06s and w/n<0.03s, and for .02s<v<.16s.

ccc dependence on v

In the previous section we assumed that A and B of Eq. (3) are roughly independent of v. Figure 7 shows the behaviour of <ccc> with varying v. It is apparent that <ccc> changes but little with v, and compares favourably with behaviour predicted by Eq.(4). For instance, two features of the data points that are consistent with the Eq.(4) are, first, tany change that <ccc> does experience is a monotonic increase with v, and second, that v is independent of w/n, as the two plots in Figure 7 with the same value of this ratio overlie each other. Improvements to agreement to agreement with Eq.(4) would probably come through refining the dependence on w/n.

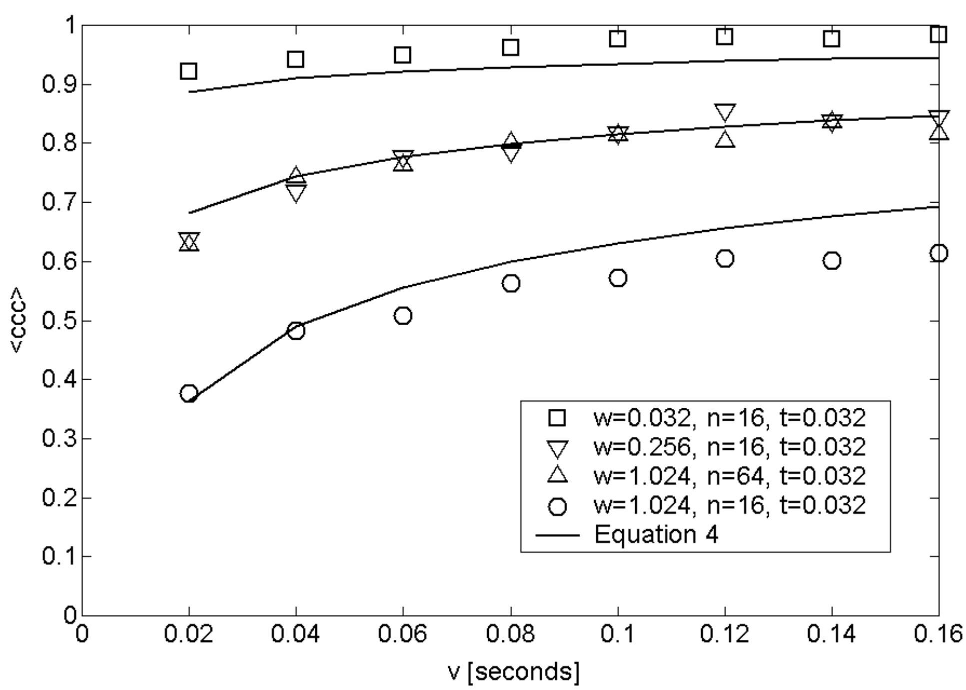


Figure 7. The variation of <ccc> with the wavelet length, v. Values derived from the cross-correlation procedure are given as points, while values from Eq.(4) are given as corresponding solid lines. Note that <ccc> varies little with v. Any increase is monotonic, in agreement with Eq.(4), and the two plots with the same value of w/n lie over top of each other, also consistent with Eq.(4).

APPLICATION

In carrying out trim statics corrections, such as in the first step of a residual statics calculation, Eq. (4) should be applied to any contemplated set of correlation parameters. If a value of $ccc \geq 0.6$ is obtained, then it is possible that the correlation procedure will result in the spurious alignment of noise, unless there is a good signal-to-noise ratio in the individual traces. Conversely, if $ccc \leq 0.5$ is obtained, then spurious alignment is unlikely, even for noisy data.

The present study was carried out with the extreme case of random noise data. Real data has some degree of signal, so there would be a competition between signal and noise for alignment. Thus it is conceivable that cross-correlation *could* be valid even if Eq.(4) predicts a value of $ccc > 0.6$. A further paper (Ursenbach, 2000, this volume) builds on the results of this study to consider the separate problem of distinguishing between alignment of signal and noise.

CONCLUSIONS

In the large-n, small-w, and large-t_{max} limits, the correlation procedure of trim statics time shift selection is shown to be capable of aligning random noise traces to reproduce a given reference trace. The functional dependence on these parameters is shown to follow a reasonable and simple form for physically sensible v values. This result can be of considerable practical help in choosing parameters for cross-correlation procedures.

REFERENCES

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

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

Yilmaz, O.,1987, Seismic Data Processing: SEG, Tulsa, OK.