Experimental comparison of repeatability metrics

Peter Gagliardi and Don C. Lawton

ABSTRACT

Time-lapse experiments were performed on the nrms repeatability (NRMS), predictability (PRED) and signal to distortion ratio (SDR) repeatability metrics, and the results studied in order to better understand their meaning. First, controlled time-shift, amplitude and additive noise perturbations were made to a baseline seismic trace. Time-shift had approximately linear effects on NRMS of about 15%/ms, subtle hyperbolic effects on PRED and a negligible effect on SDR. Amplitude tests showed that multiplication of the baseline trace by 0.9 resulted in an NRMS value of 10.5% and SDR value of $10^{2.04}$, while PRED remained unaffected by any amplitude change; analytic equations were found to relate amplitude changes to these metrics. Additive noise experiments revealed that NRMS and PRED are very sensitive to the strength and character of the noise, while SDR seems to be affected little by the noise character.

Second, all three metrics were calculated using a 2D walkaway vertical seismic profile (VSP) dataset from Violet Grove, Alberta, which consisted of three lines. For Lines 1, 2 and 3, NRMS values were 60.6%, 61.4% and 45.2% for horizontal components, and 46.3%, 42.6% and 41.4% for the vertical component. PRED was 0.73, 0.72 and 0.83 for the horizontal components, and 0.82, 0.83 and 0.87 for the vertical component. Finally, SDR was $10^{0.38}$, $10^{0.29}$ and $10^{0.70}$ for the horizontal components and $10^{0.74}$, $10^{0.85}$ and $10^{0.79}$ for the vertical component. The trends of these metrics, while similar, do not always agree with each other, and should be used in tandem to better understand the repeatability metrics of time-lapse data.

INTRODUCTION

Time-lapse geophysics has become a widely used tool for applications such as reservoir monitoring and CO_2 sequestration studies; its goal is to detect changes in the subsurface, and therefore must include at least 2 separate seismic experiments. The first is generally referred to as a "baseline" survey, while subsequent experiments are called "monitor" surveys. The ideal end result of a time-lapse experiment is to have a reliable measure of changes in the subsurface, while minimising or eliminating changes due to any other factors, which can be grouped together as "4D noise". Some sources of 4D noise include source and receiver mispositioning, changes in the weathering layer, changes in source character and changes in the ambient noise.

In order to determine the amount of change between a baseline and monitor survey, it is useful to define some quantitative metrics. Two commonly used metrics are called NRMS repeatability and predictability (Kragh and Christie, 2002). These provide a measure of the trace-by-trace "repeatability"; that is, the overall similarity of a monitor trace to a baseline trace. However, these do not necessarily distinguish between 4D noise and meaningful changes in the subsurface, leading to some ambiguity in their interpretation. Cantillo (2011) provides a new metric, called signal to distortion ratio, which the author suggests is more meaningful than either of the metrics provided by

Kragh and Christie (2002). The purpose of this study is to further understand the results provided by these three metrics; this will be done using artificially perturbed traces as well as analysis of a case study.

STUDY AREA

The VSP data used in this study was taken from the Pembina CO_2 enhanced oil recovery project; this oilfield is about 100 km southwest of Edmonton, Alberta, and the Cardium in this area is the largest conventional oil pool discovered in Western Canada (Hitchon, 2009). The VSP consisted of eight 3-component geophones placed every 20 m, starting at 1498 m depth, in the observation well 07-11-048-09W5 near Violet Grove, Alberta (Hitchon, 2009). The baseline (Phase I) dataset was acquired in March 2005, and the monitor (Phase III) dataset was acquired in March 2007. Three surface seismic lines were common between them: Line 1, which runs North-South, and Lines 2 and 3, which run East-West (Figure 1). A previous repeatability study was performed on these data by Gagliardi and Lawton (2010), using NRMS repeatability and predictability values only; this study will draw comparisons between those metrics and the SDR metric applied to the same data.

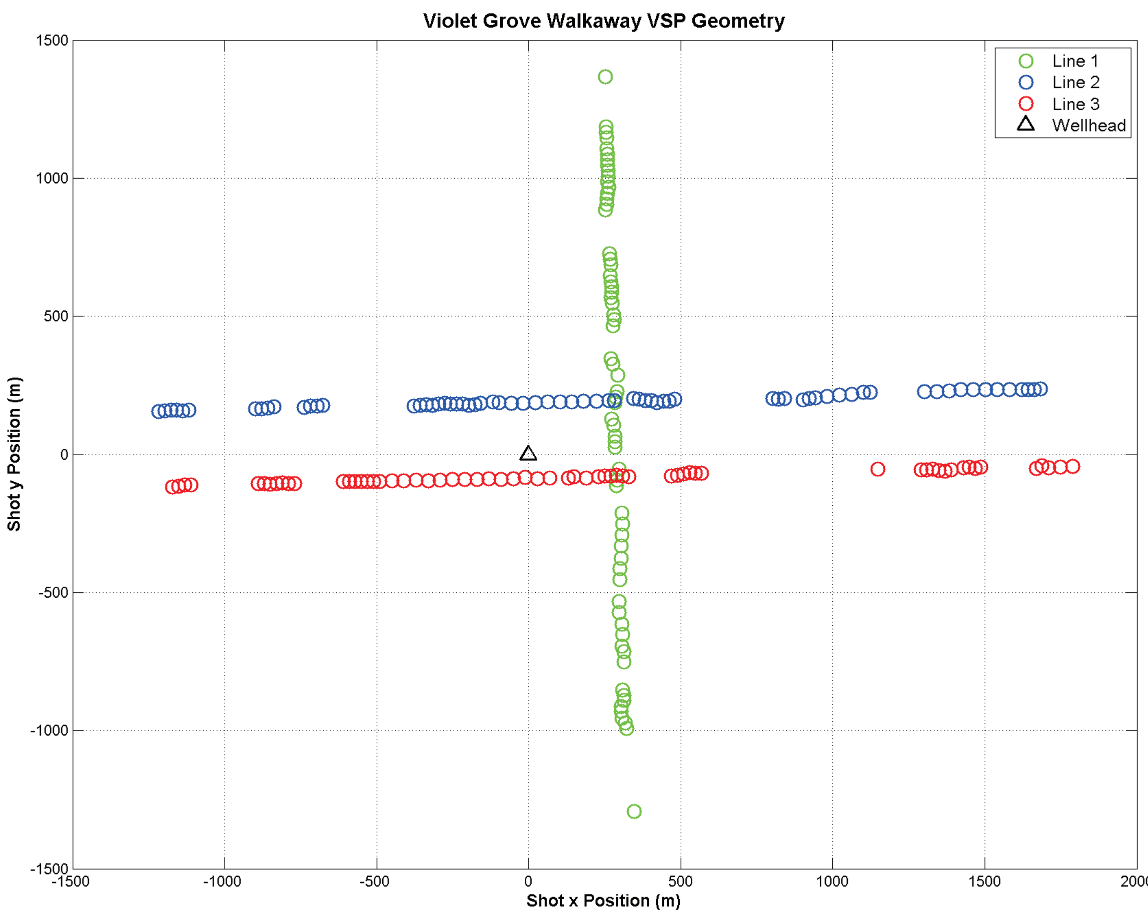

FIG. 1. Surface geometry for Violet Grove walkaway VSP used in this study.

REPEATABILITY METRICS

NRMS repeatability (NRMS)

NRMS repeatability is defined as (Kragh and Christie, 2002)

$$NRMS = \frac{2\sqrt{\sum_{t_1}^{t_2}(b_t - m_t)^2/N}}{\sqrt{\sum_{t_1}^{t_2}(b_t)^2/N} + \sqrt{\sum_{t_1}^{t_2}(m_t)^2/N}}, \tag{1}$$

where b and m are the baseline and monitor traces, t_1 and t_2 are the start and end times of the desired window, and N represents the total number of samples per trace within the window. The values for NRMS repeatability are generally given in percent, and range from 0% to 200%, where lower values represent more repeatable traces; it is also interesting to note that complete noise should theoretically be calculated as $\sqrt{2}$ (roughly 141%) (Kragh and Christie, 2002).

Predictability (PRED)

Predictability is defined as (Kragh and Christie, 2002)

$$PRED = \frac{\left(\sum_{-n}^{+n} b \otimes m\right)^2}{\left(\sum_{-n}^{+n} b \otimes b\right)\left(\sum_{-n}^{+n} m \otimes m\right)}, \tag{2}$$

where b and m are the baseline and monitor traces windowed from t_1 to t_2, $\otimes$ is the crosscorrelation operator, and the sum is performed over lags $-n$ to $+n$. For this study, only the zero lag values are considered (i.e. n=0). It is pointed out in Cantillo (2011) that the number of lags used in the summation can have an important effect on the predictability value; thus, it is difficult to gain a good understanding of this metric without consistency in the number of lags used. The values for predictability range from 0 to 1, where higher values represent more repeatable traces (Kragh and Christie, 2002).

Signal to distortion ratio (SDR)

Signal to distortion ratio is defined as (Cantillo, 2011)

$$SDR = \frac{\sum_{t_1}^{t_2} b_t^2}{\sum_{t_1}^{t_2} d_t^2} = \frac{\max(b \otimes m)^2}{1 - \max(b \otimes m)^2}, \tag{3}$$

given the model

$$m = \delta_t * b + d. \tag{4}$$

Here, b, m and $\otimes$ are defined as above, δ_t is a delta function, and d (the "distortion") encompasses all changes between the baseline and monitor traces, once time-shifts have been removed. Cantillo (2011) suggests that this metric, along with time-shift measurements, would be more suitable for repeatability studies than the currently used NRMS and PRED. Throughout this study, the $\log_{10}$ of SDR was favoured, as it allowed for more useful trace by trace comparisons.

CONTROLLED EXPERIMENT

Experiment parameters

In order to test the separate effects of time-shift, amplitude difference and noise, a trace was chosen from the baseline survey (Figure 2). Copies of this trace were then perturbed in several ways:

1) The traces were time-shifted by values ranging from -5 ms to +5 ms, incrementing by 0.1 ms with and without resampling of the initial trace.

2) The traces were multiplied by constants ranging from 0.5 to 1.5, incrementing by 0.01.

3) Noise, which was extracted from the first 500 ms of both the baseline and monitor traces, was multiplied by ratios of the maximum noise to maximum signal of the baseline trace, ranging from 0 to 0.5 in increments of 0.02.

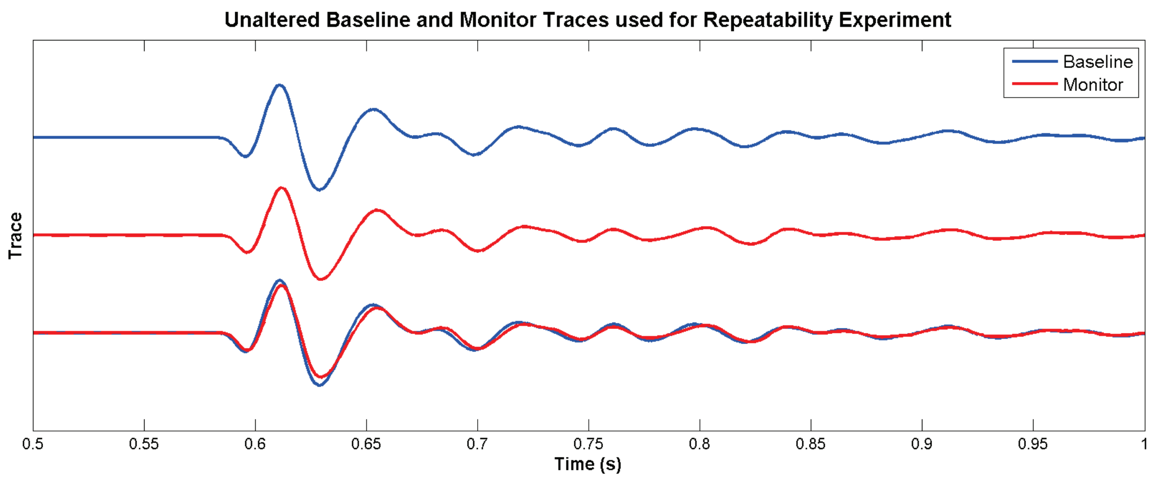

FIG. 2. Traces used for controlled repeatability experiment; baseline is shown in blue and monitor is shown in red.

Finally, it should be noted that the NRMS, PRED and $\log_{10}$ SDR values calculated between the original baseline and monitor traces were 24.9%, 0.96 and 1.49 respectively.

Time-shift tests

Figure 3 shows the results of the time-shift tests. First, it should be noted that the sample rate (relative to the time-shift) has a subtle effect on NRMS and PRED, and a much more noticeable effect on SDR. However, even at a 1 ms sample rate, the SDR value remains consistently above 10^4 – this suggests time-shift does not have a largely detrimental effect on this metric, which could be inferred from Equation 3. NRMS appears to have a linear dependence on time-shift; the result of a linear regression on these values produces a slope of roughly 15%/ms. PRED seems to have a hyperbolic trend, only changing by about 0.03 with a 1 ms time-shift. All three metrics appear to be roughly symmetric between positive and negative time-shift values.

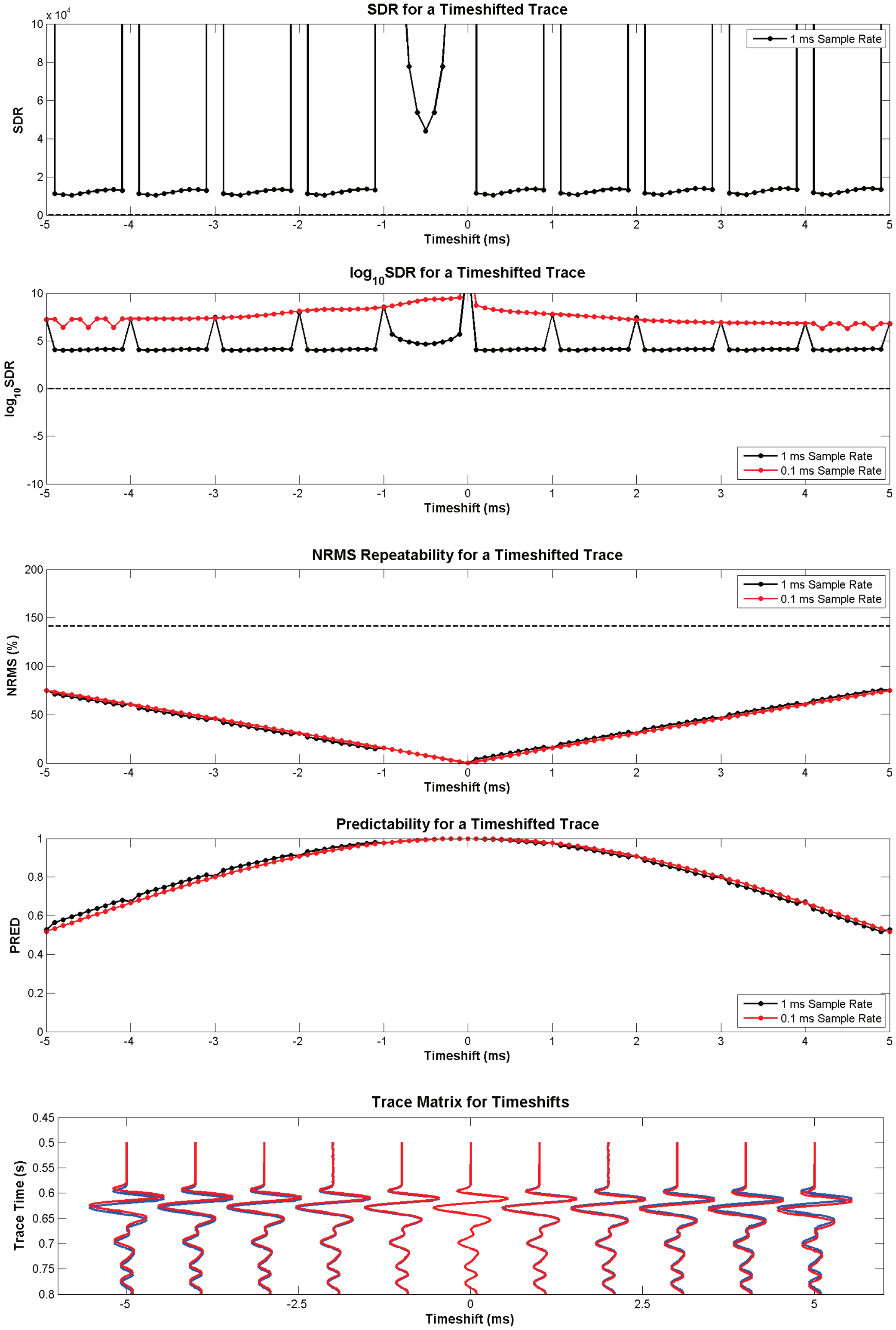

FIG. 3. From top to bottom: SDR, $\log_{10}$ SDR, NRMS repeatability and predictability for time-shift experiment, showing traces left at the original sample rate (black) and traces resampled to 0.1 ms (red). Bottommost panel is a wiggle display of the time-shifted traces (red) overlapping the original trace (blue).

Amplitude tests

The results of the amplitude tests are shown in Figure 4. Again, the NRMS values appear to have a linear relationship to the amplitude ratio, though the curve is no longer symmetric. The value becomes 9.5% when the ratio of monitor to baseline amplitude is 1.1, and 10.5% when the ratio is 0.9. With the same amplitude ratios, the SDR becomes $10^{1.95}$ and $10^{2.04}$ respectively. Finally, PRED remains unchanged regardless of the amplitude ratio. The behaviour of all three of these metrics in the presence of an amplitude perturbation can be found analytically, if *m* is replaced with *Ab*, where *A* is a scalar; after performing this substitution in Equations 1 – 3 and simplifying, we find that

$$NRMS(A) = \frac{2\sqrt{\sum_{t_1}^{t_2}(b_t - Ab_t)^2/N}}{\sqrt{\sum_{t_1}^{t_2}(b_t)^2/N} + \sqrt{\sum_{t_1}^{t_2}(Ab_t)^2/N}} = \frac{\left(2\sqrt{(1-A)^2}\right)\left(\sqrt{\sum_{t_1}^{t_2}(b_t)^2/N}\right)}{\left(1+\sqrt{A^2}\right)\left(\sqrt{\sum_{t_1}^{t_2}(b_t)^2/N}\right)} = \frac{2|1-A|}{1+|A|}, \quad (5)$$

$$PRED(A) = \frac{\left(\sum_{-n}^{+n} b \otimes Ab\right)^2}{\left(\sum_{-n}^{+n} b \otimes b\right)\left(\sum_{-n}^{+n} Ab \otimes Ab\right)} = \frac{A^2\left(\sum_{-n}^{+n} b \otimes b\right)^2}{A * A\left(\sum_{-n}^{+n} b \otimes b\right)\left(\sum_{-n}^{+n} b \otimes b\right)} = \frac{A^2}{A^2} = 1, \quad (6)$$

and

$$SDR(A) = \frac{\max(b \otimes Ab)^2}{1 - \max(b \otimes Ab)^2} = \frac{A^2 \max(b \otimes b)^2}{1 - A^2 \max(b \otimes b)^2} = \frac{A^2}{1 - A^2}. \quad (7)$$

These relationships agree with what is seen in Figure 4.

Noise tests

Finally, repeatability after the addition of additive noise is shown in Figure 5; the two curves represent two different types of noise: noise taken from the baseline trace (black curve) and noise taken from the monitor trace (red curve). The noise taken from the monitor was dominated by a 60 Hz cable signal, and can be considered as non-random. The x-axis is calculated as the ratio of the maximum amplitudes of the noise and the baseline; that is,

$$\text{Amplitude Ratio} = \frac{\max(A_{noise})}{\max(A_b)}. \quad (8)$$

As a benchmark, we may consider that the monitor (non-random) noise curve crosses SDR=1 at a value of about 0.15; Figure 6 shows a zoomed in view of the amplitude ratio interval 0-0.15. Interestingly, $\log_{10}$ SDR is very similar for both types of noise added, whereas NRMS and PRED show a clear separation between the two noise curves. When the ratio is at 0.1, NRMS, PRED and SDR become 37.3%, 0.92 and $10^{0.81}$ for baseline (random) noise and 56.3%, 0.74 and $10^{0.39}$ for monitor (non-random) noise. These figures are quite large when considering how this appears visually, as shown in the third trace in the bottom two panels of Figure 5.

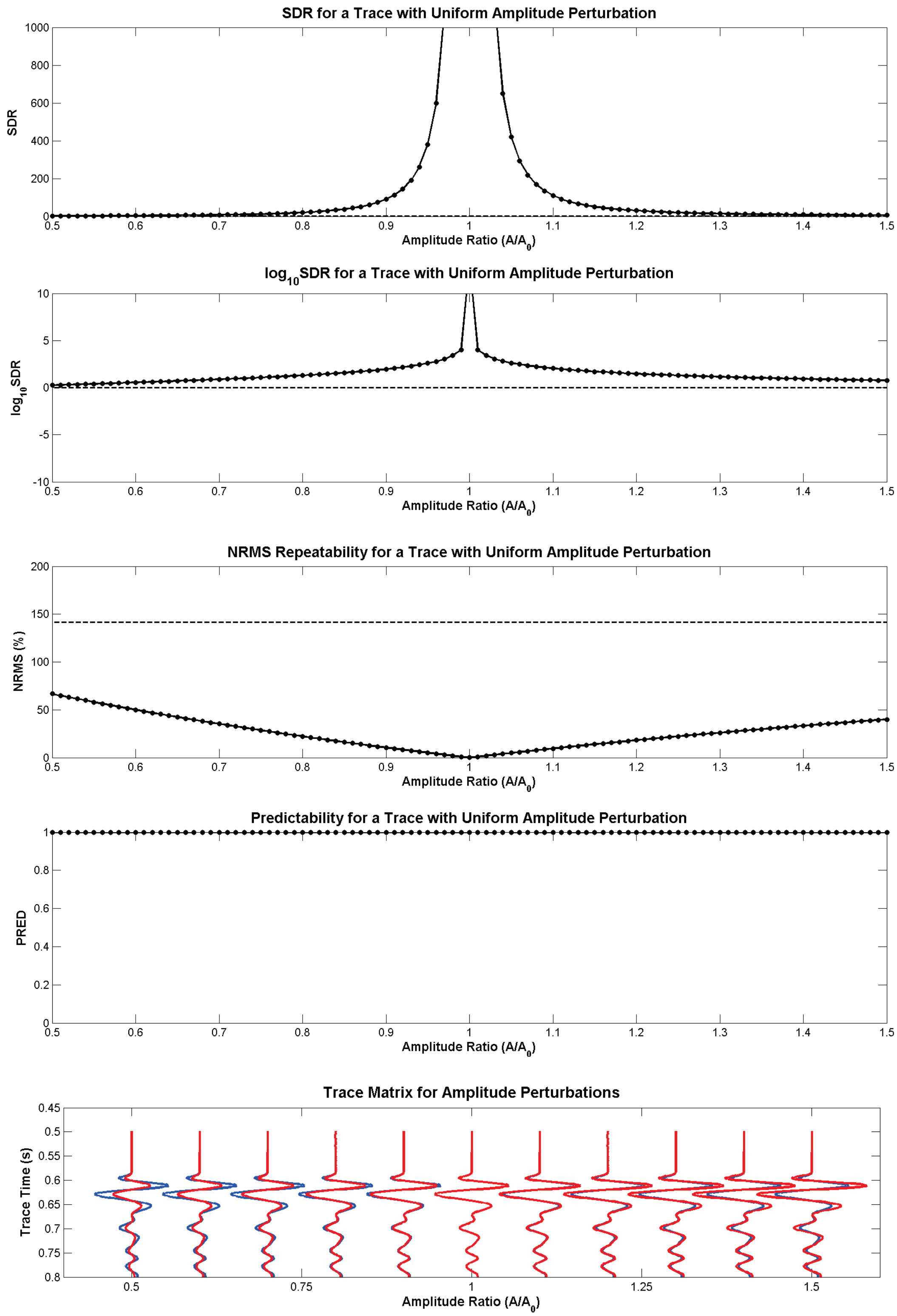

FIG. 4. From top to bottom: SDR, $\log_{10}$ SDR, NRMS repeatability and predictability for amplitude experiment. Bottommost panel is a wiggle display of the amplitude modified traces (red) overlapping the original trace (blue).

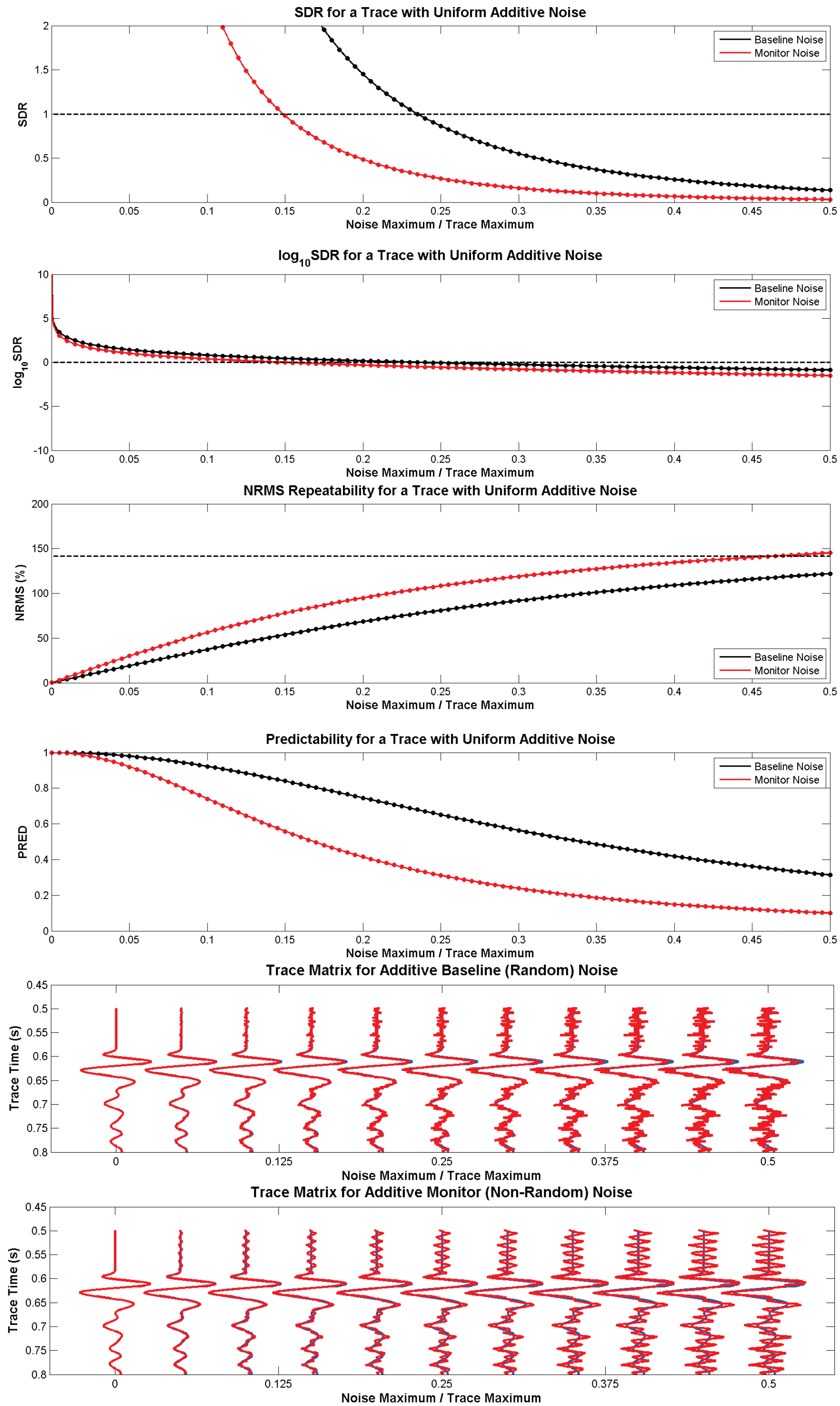

FIG. 5. From top to bottom: SDR, $\log_{10}$ SDR, NRMS repeatability and predictability for additive noise experiment, showing addition of baseline (black) and monitor (red) noise. Two bottommost panels are wiggle displays of the noisy traces (red) overlapping the original trace (blue).

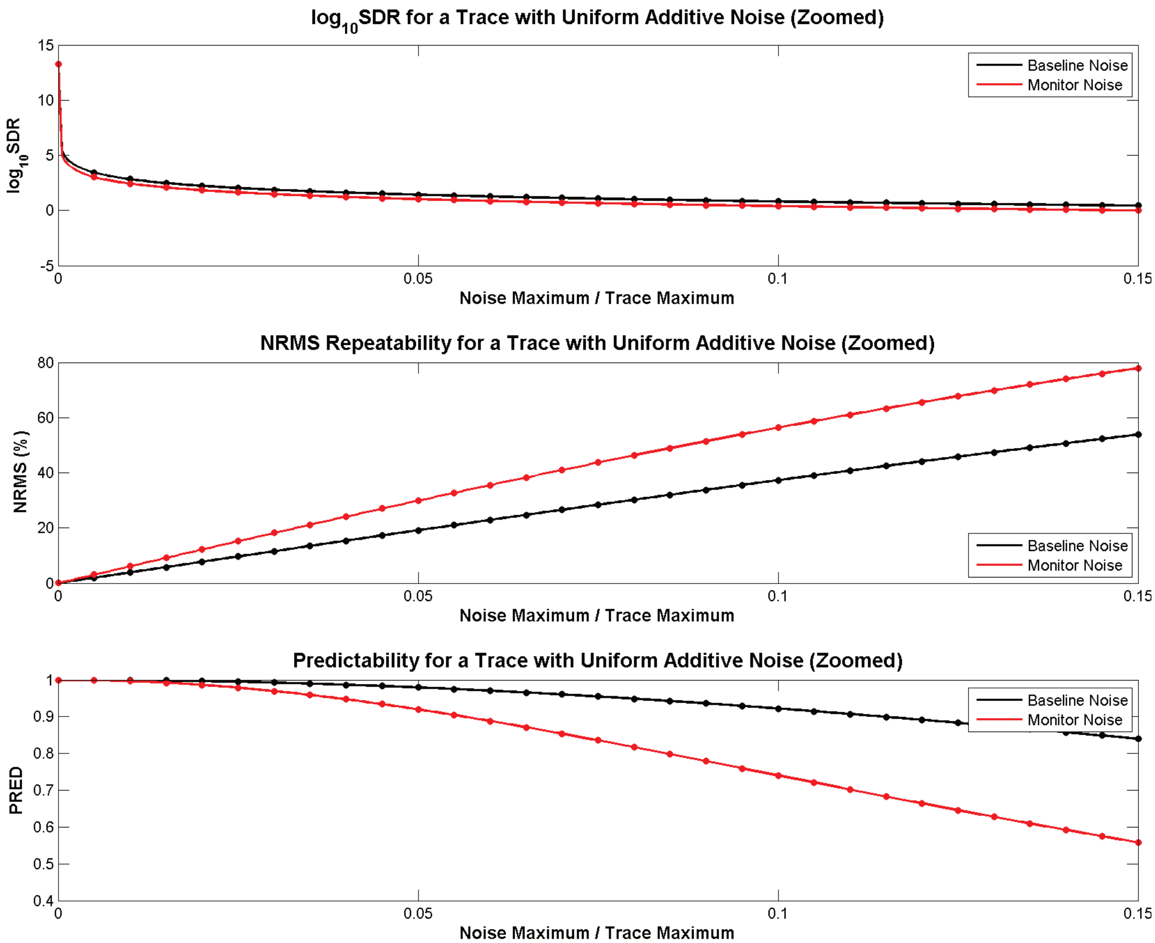

FIG. 6. $\log_{10}$ SDR (top), NRMS repeatability (middle) and predictability (bottom) for additive noise experiment, zoomed in on the interval between 0 and 0.15.

FIELD EXPERIMENT

A repeatability study previously done on the Violet Grove dataset by Gagliardi and Lawton (2010) examined the NRMS and PRED metrics. The results presented here include the results of that study, with the addition of the SDR metric. Figures 7 – 9 include plots of $\log_{10}$ SDR plotted alongside NRMS, as well as alongside PRED, for receiver 3. These plots reveal that $\log_{10}$ SDR produces trends that are nearly the mirror image of NRMS. It also appears very similar to PRED, though differences are more noticeable than with NRMS. Tables 1 – 3 summarise the results numerically, on a receiver by receiver basis. Averages calculated in these tables do not include values from components that were problematic in Phase III; these were: z-component of Receiver 2, x-component of Receiver 4 and y-component of Receiver 6 (Gagliardi and Lawton, 2010).

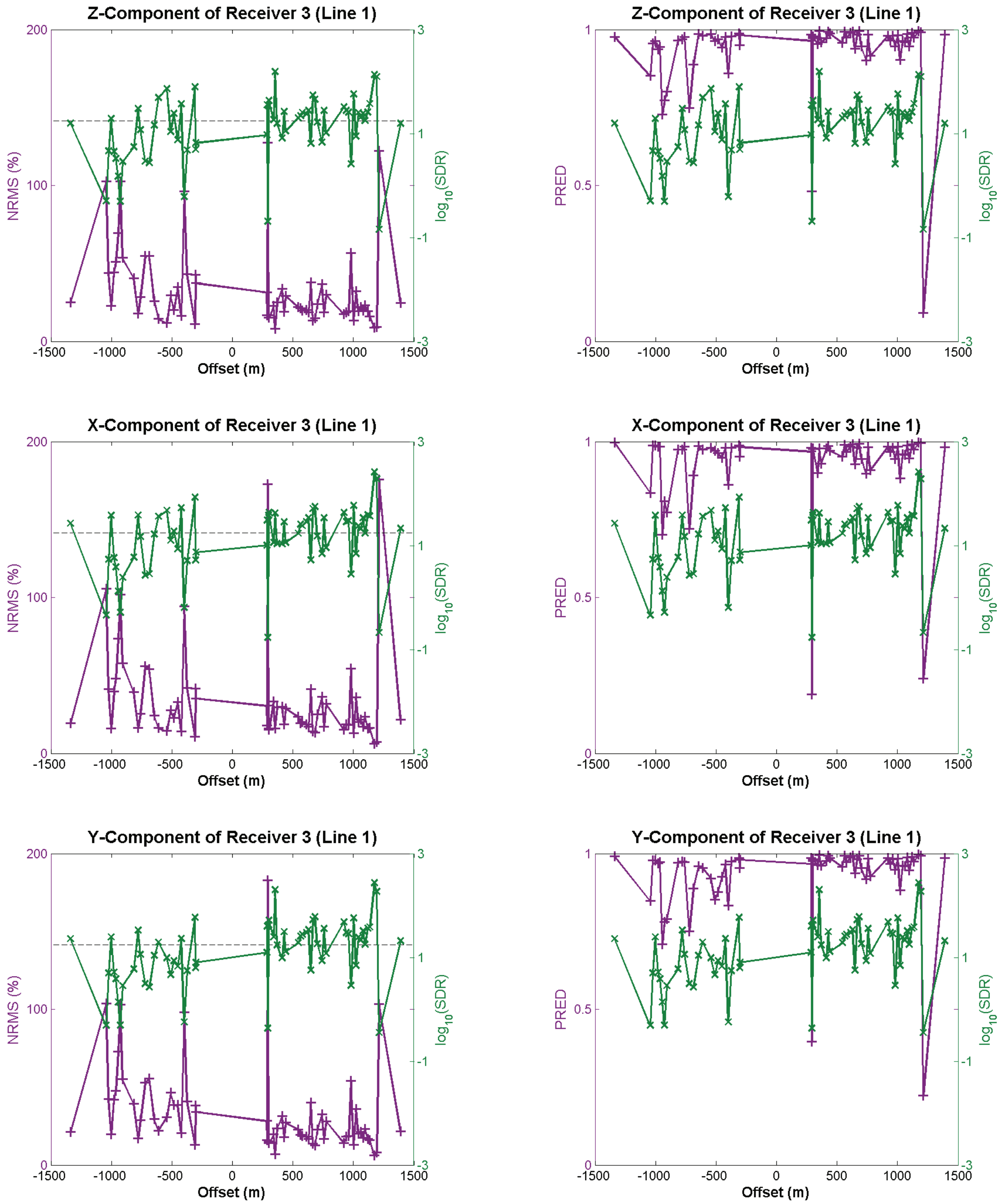

FIG. 7. Comparison of NRMS repeatability (left), predictability (right) and $\log_{10}$ SDR of receiver 3 for Line 1. The z-component is shown at the top, x-component is shown in the middle, and y-component is shown at the bottom. The dashed line in NRMS plots indicates the theoretical noise line.

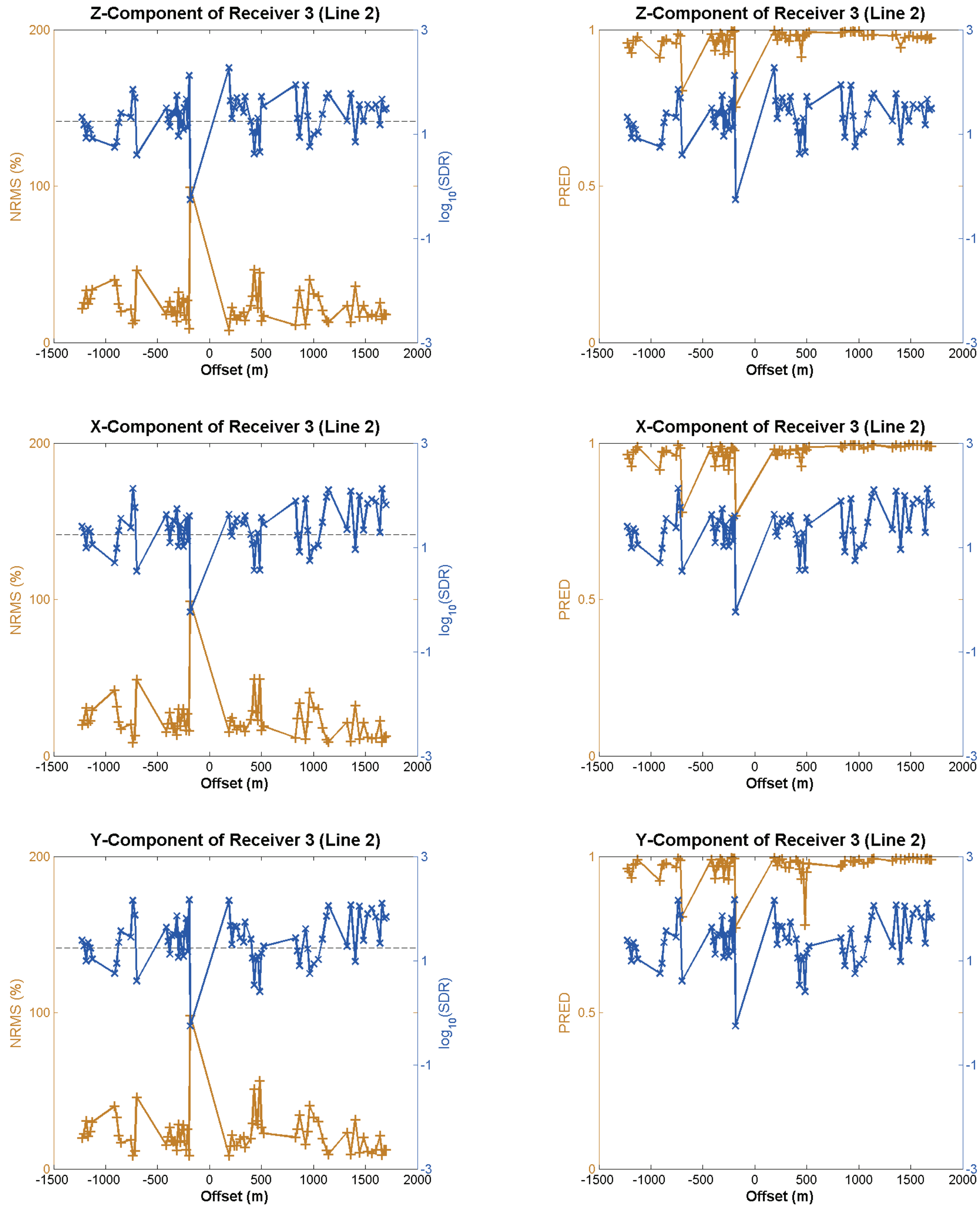

FIG. 8. Comparison of NRMS repeatability (left), predictability (right) and $\log_{10}$ SDR of receiver 3 for Line 2. The z-component is shown at the top, x-component is shown in the middle, and y-component is shown at the bottom. The dashed line in NRMS plots indicates the theoretical noise line.

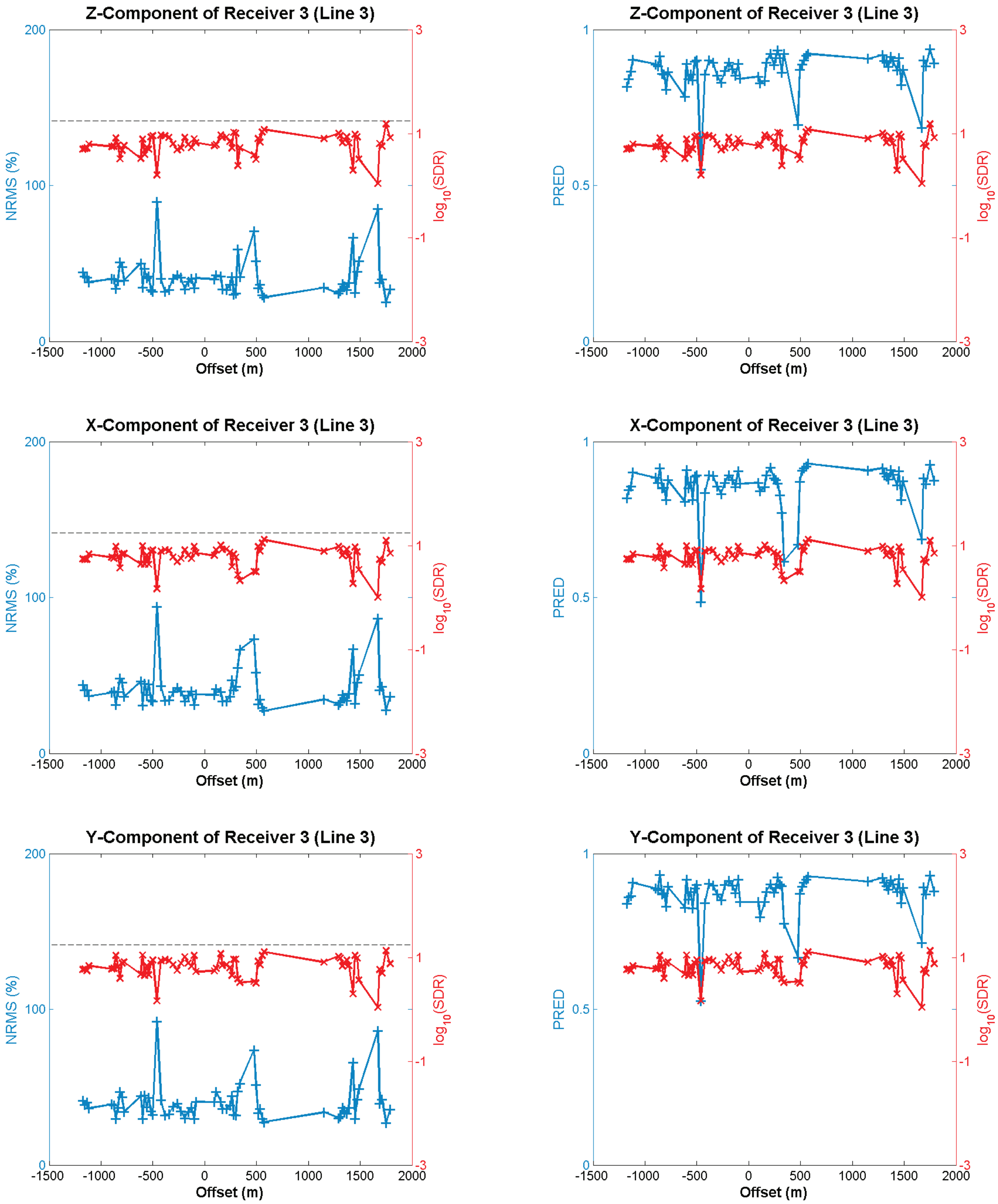

FIG. 9. Comparison of NRMS repeatability (left), predictability (right) and $\log_{10}$ SDR of receiver 3 for Line 3. The z-component is shown at the top, x-component is shown in the middle, and y-component is shown at the bottom. The dashed line in NRMS plots indicates the theoretical noise line.

Table 1. Summary of repeatability metrics of Violet Grove z-component data. Average does not include Receiver 2.

	Line 1			Line 2			Line 3		
Receiver	Average NRMS	Average PRED	Average log SDR	Average NRMS	Average PRED	Average log SDR	Average NRMS	Average PRED	Average log SDR
1	50.1%	0.82	0.60	49.5%	0.79	0.60	44.2%	0.86	0.74
2	155.6%	0.00	-5.27	151.9%	0.00	-5.10	192.0%	0.00	-5.87
3	33.2%	0.93	1.09	24.5%	0.96	1.32	41.2%	0.87	0.79
4	43.5%	0.85	0.77	39.9%	0.86	0.86	44.3%	0.84	0.75
5	95.9%	0.40	-0.42	108.4%	0.34	-0.68	42.4%	0.86	0.75
6	34.2%	0.93	1.05	25.4%	0.96	1.30	39.6%	0.88	0.81
7	36.5%	0.91	0.93	26.0%	0.95	1.23	40.8%	0.87	0.80
8	30.6%	0.93	1.18	24.5%	0.96	1.35	37.1%	0.90	0.89
Average	**46.3%**	**0.82**	**0.74**	**42.6%**	**0.83**	**0.85**	**41.4%**	**0.87**	**0.79**

Table 2. Summary of repeatability metrics of Violet Grove x-component data. Average does not include Receiver 4.

	Line 1			Line 2			Line 3		
Receiver	Average NRMS	Average PRED	Average log SDR	Average NRMS	Average PRED	Average log SDR	Average NRMS	Average PRED	Average log SDR
1	38.2%	0.92	0.97	26.8%	0.96	1.22	42.4%	0.86	0.77
2	43.0%	0.85	0.86	29.2%	0.94	1.14	40.1%	0.87	0.82
3	34.4%	0.93	1.12	23.7%	0.97	1.37	41.7%	0.86	0.78
4	148.7%	0.00	-5.16	153.5%	0.00	-5.19	178.6%	0.00	-5.05
5	157.6%	0.09	-2.46	155.8%	0.10	-2.29	50.1%	0.79	0.58
6	34.9%	0.93	1.09	24.0%	0.97	1.35	41.0%	0.86	0.78
7	51.8%	0.78	0.57	83.0%	0.55	-0.17	46.5%	0.83	0.75
8	43.2%	0.86	0.80	31.6%	0.93	1.07	44.5%	0.84	0.72
Average	**57.6%**	**0.77**	**0.42**	**53.4%**	**0.77**	**0.53**	**43.8%**	**0.84**	**0.74**

Table 3. Summary of repeatability metrics of Violet Grove y-component data. Average does not include Receiver 6.

	Line 1			Line 2			Line 3		
Receiver	Average NRMS	Average PRED	Average log SDR	Average NRMS	Average PRED	Average log SDR	Average NRMS	Average PRED	Average log SDR
1	46.5%	0.84	0.74	48.6%	0.78	0.67	48.7%	0.81	0.65
2	136.7%	0.14	-1.43	167.5%	0.04	-2.64	57.7%	0.73	0.42
3	34.0%	0.93	1.11	23.4%	0.96	1.38	40.2%	0.87	0.81
4	35.1%	0.93	1.08	24.8%	0.96	1.32	42.6%	0.86	0.76
5	110.9%	0.29	-0.81	157.3%	0.09	-2.43	54.2%	0.76	0.48
6	196.8%	0.00	-8.27	197.1%	0.00	-8.39	164.6%	0.00	-4.55
7	44.0%	0.87	0.78	29.3%	0.94	1.12	42.2%	0.85	0.75
8	37.4%	0.91	0.97	34.5%	0.91	0.96	40.7%	0.87	0.78
Average	**63.5%**	**0.70**	**0.35**	**69.3%**	**0.67**	**0.05**	**46.6%**	**0.82**	**0.67**

DISCUSSION

The results presented in this study show that all three repeatability metrics can be very sensitive to effects of time-shifts, amplitude differences and random noise. Time-shifts and amplitude differences can, in general, be minimized reasonably well on a trace by trace basis, and noise can be reduced through processes such as stacking and frequency filtering. However, even subtle changes in these three categories will cause large changes in the repeatability calculations; this is emphasised when looking at the visual trace comparisons.

Both the controlled and field experiments show that there are differences and similarities in the way NRMS, PRED and SDR behave. NRMS appears to have a nearly linear dependence on time-shifts and amplitude perturbations, and is easily disturbed by strength and character of noise. PRED is insensitive to amplitude changes and is largely unaffected by small changes in time-shift, though it too is sensitive to noise strength and character. Noise character seems to have little influence on SDR, however, and it is also insensitive to time-shifts. In amplitude and noise tests, $\log_{10}$ SDR changes rapidly when a trace moves from no perturbation to slight perturbation, but beyond that SDR shows a nearly linear response. Examination of the field experiment shows remarkable similarity between NRMS and SDR and PRED and SDR. Nevertheless, it can also be seen that the three repeatability metrics do not always show the same trend; for example, the z-component shows that Line 3 has an NRMS value 1.2% better than Line 2, while PRED is better by 0.04 and SDR is worse by $10^{0.06}$. Perhaps through further examination of these parameters, the different results they provide can be used to interpret the main contributions to 4D noise.

CONCLUSIONS

- Residual time-shift between a noise-free baseline and monitor trace will cause NRMS to change almost linearly by 15%/ms; the effect is much more subtle on PRED, only changing it by about 0.03 for 1 ms difference. Time-shift has essentially no effect on SDR, which is intended from its definition. It should also be noted that there is a slight jitter that arises from time-shifts smaller than the trace sample rate.

- Amplitude perturbations between a noise-free baseline and monitor trace showed that NRMS changed by 10.5% and SDR is $10^{2.04}$ when the amplitude ratio of the two traces was 0.9; PRED remained unaffected by amplitude changes. Relationships of amplitude change to each of these three metrics were also determined analytically.

- When additive noise was introduced to the monitor trace, $\log_{10}$ SDR showed results that were very similar for both the baseline (random) and monitor (non-random) noise; while there was a large drop in its value when a slight amount of noise was added, further addition of noise resulted in a nearly linear response with a gentle slope. NRMS and PRED produced curves that were easily distinguishable between the types of noise added. NRMS appeared to have a linear response while the noise strength was low; PRED changed little with low noise strength, and was more sensitive to the type of noise added.

- Repeatability analysis of the Violet Grove horizontal component data yielded NRMS values of 60.6%, 61.4% and 45.2%, PRED values of 0.73, 0.72 and 0.83, and SDR values of $10^{0.38}$, $10^{0.29}$ and $10^{0.70}$ for Lines 1, 2 and 3 respectively.

- Repeatability analysis of the Violet Grove vertical component data yielded NRMS values of 46.3%, 42.6% and 41.4%, PRED values of 0.82, 0.83 and 0.87, and SDR values of $10^{0.74}$, $10^{0.85}$ and $10^{0.79}$ for Lines 1, 2 and 3 respectively.

- While at a high level all three of these metrics produced similar trends in the field data example, they appear to behave differently depending on the type of 4D noise, and should thus be used together to better understand the repeatability of time-lapse seismic.

FUTURE WORK

Further controlled experiments using these three repeatability metrics could be performed in order to understand them better; this could include the examination of other perturbations, as well as combinations of several parameters. Testing could be performed with modelled data, allowing for simulated changes in the geological information in addition to 4D noise. Analytic work, similar to that done with the amplitude experiments, could also prove valuable in understanding these metrics. Finally, more in depth analysis of field examples using these repeatability metrics would help us understand how each can be used as an interpretation tool for time-lapse geophysics.

ACKNOWLEDGEMENTS

We would like to thank Penn West and Schlumberger for the data used in this study, NRCan and AERI for funding used in the Pembina CO_2 monitoring project, and CREWES Sponsors for continued support of our research.

REFERENCES

Cantillo, J., 2011, A quantitative discussion on time-lapse repeatability and its metrics: SEG Expanded Abstracts, **30,** 4160-4164.

Gagliardi, P. and Lawton, D. C., 2010, Borehole geophone repeatability experiment: CREWES Research Report, **22,** 22.1-22.23.

Hitchon, B., 2009, Pembina Cardium CO_2 Monitoring Pilot: A CO_2-EOR Project, Alberta, Canada: Final Report: Geoscience Publishing, Limited.

Kragh, E. and Christie, P., 2002, Seismic repeatability, normalized rms, and predictability: The Leading Edge, **21,** 640-647.