Uncertainty quantification for land seismic acquisition

Iga Pawelec & Paul Sava, Center for Wave Phenomena, Colorado School of Mines

SUMMARY

Noise is an inherent feature of seismic data, especially in land acquisition. Due to the presence of noise, observations can be treated as random variables with associated uncertainty. An accurate estimate of data uncertainty is important for data interpretation and also for imaging and tomography. For challenging inverse problem, like full waveform inversion, an accurate estimate of data uncertainty can lead to robust qualitative estimates of posterior uncertainty, and also guide the selection of key parameters, e.g. the regularization strength. Uncertainty estimation can be a challenging task for seismic data because not enough repeat measurements are available to estimate reliable statistics. Based on a land seismic field experiment with repeated shots, we quantify seismic data uncertainty (with standard deviation as a proxy) and the short-term repeatability of reflection data. We find that the uncertainty for coherent seismic events in data is proportional to their amplitude while the distributions characterizing the main events differ from Gaussians.

INTRODUCTION

Seismic data provide a wealth of information about geological structures, fluid content and physical properties of the subsurface. The information is captured by the traveltime and the amplitude of seismic reflections. To make valid inferences from data, it is important to understand their uncertainty. Although wave propagation is deterministic in nature, i.e., if the medium is known, traveltime and amplitude can be uniquely determined, the signal recorded during field acquisition is a superposition of wave phenomena triggered by the seismic source together with the ambient noise.

Repeatability is an important concept for time lapse monitoring, where one seeks to quantify the change in seismic amplitudes related to changes in elastic moduli (Hughes et al., 1998; Landrø, 2001; Lumley, 2001). Much effort goes into acquisition design to find a setup that maximizes repeatability between the surveys (Poggiagliolmi et al., 1998; Naess, 2006; Houck, 2007; Bakulin et al., 2014). However, acquisition geometry repeatability is not the only factor affecting the data (Landrø, 1999). Ambient noise, which may differ between the surveys, and changes to the survey site, such as new sediments on a sea floor or different soil saturation on land, also leave an imprint on data.

In this work, we seek to evaluate the uncertainty of seismic data, especially with regard to repeatability that is not caused by changes of the medium. We exploit data from a field experiment with 100 repeated measurements at the same location in the span of 1 hour. Primary sources of uncertainty for this scenario are the instrumentation noise (uncertainty in source signature, sensitivity and coupling of geophones), ambient noise (traffic, background seismicity, wind) and small changes to the experimental environment caused by ground compaction due to multiple repetitions of a Vibroseis source at the same location. We describe the uncertainty in terms of mean amplitude and standard deviation. The ratio of standard deviation to the absolute value of a mean amplitude quantifies the relative repeatability of a seismic amplitude. Standard deviation is a measure of dispersion. When compared to the mean amplitude, it represents percent changes in amplitude values through experiments. Throughout this experiment, we pay special attention to the reflection signal and quantify its amplitude and kinematic repeatability.

To infer the properties of the subsurface from seismic data, one needs to solve an inverse problem. This can be accomplished within either the probabilistic Bayesian framework (Tarantola, 2005), which relies on combining prior information about model and data with theoretical relationship between them to achieve a refined, posterior state of information or the deterministic framework that aims to minimize a specific objective function. Information about data uncertainty may be incorporated as data prior in the Bayesian framework (e.g., Osypov et al. (2008) solve a tomographic problem in the Bayesian framework to quantify the posterior uncertainty) or in the regularization term of deterministic inversion. The data regularization term is especially important for strongly non-linear problems, such as full waveform inversion (Tarantola and Valette, 1982; Pratt, 1999; Virieux and Operto, 2009).

The data prior, captured by the data probability distribution, can be particularly problematic to characterize, due to time and cost involved in seismic acquisition. We usually do not acquire multiple measurements to statistically analyze data and quantify their uncertainty. Instead, noise is usually assumed to be Gaussian, with fixed standard deviation. It is a pragmatic choice in the absence of additional information, as multivariate Gaussian distributions are well studied (Anderson et al., 1958), their parametrization is easy to interpret in terms of probability and they provide an accurate description of many random processes. However, if the true distribution significantly differs from Gaussian, assuming Gaussianity may lead to wrong conclusions about the solution of an inverse problem.

In order to assess these assumptions, we conduct a field experiment with repeated seismic measurements, which allows to evaluate distributions and the associated statis-

tics for every data point. Since our experiment is conducted within a short period over a non-producing area, changes in elastic moduli of a medium are not expected except in the immediate vicinity of the seismic source, thus allowing us to quantify the data uncertainty caused by uncontrolled environmental causes.

METHODOLOGY

We analyze the acquired seismic data with two main objectives: quantifying the uncertainty and assessing the short-term repeatability of reflection data.

To meet the first objective, we use an empirical PDF-based approach. We describe our data as a volume with dimensions t, x and e - time, offset, and experiment index, where $Ne = 100$. At every (t, x) point, there are 100 amplitude values that we use to create a PDF. Given this PDF, mean and variance can be computed at every t and x:

$$\bar{A}(t,x) = \sum_e A(t,x) f(A(t,x)), \tag{1}$$

$$\sigma^2(t,x) = \sum_e (A(t,x) - \bar{A}(t,x))^2 f(A(t,x)), \tag{2}$$

where $f(A(t,x))$ is the estimated PDF. The standard deviation σ is a proxy for the uncertainty. We assess the repeatability of reflections by tracking their time and amplitude on a high fidelity seismic trace with easily identifiable events. Trace fidelity F, which helps to identify reliable channels and avoid those badly affected by different types of noise, is computed as a sum of standard deviations for all samples of the trace:

$$F(x) = \sum_t \sigma(t,x). \tag{3}$$

The typical repeatability metrics used in time lapse seismic, normalized RMS and predictability (Kragh and Christie, 2002) are not feasible for the experiment with a shot repeated 100 times since they look at changes between two surveys. Here, our objective is to quantify uncertainty, not compare individual shots (each shot-shot pair would have their own NRMS or predictability). An alternative metric could be standard deviation, but because it may depend on individual sensors' amplitude response, we propose to measure amplitude repeatability as the ratio of standard deviation to absolute mean amplitude:

$$R(t,x) = \frac{\sigma(t,x)}{|\bar{A}(t,x)|} \cdot 100. \tag{4}$$

Perfect repeatability would correspond to $\sigma = 0$ and $R = 0$, which implies zero uncertainty.

EXPERIMENT

The experimental data were acquired in Pagosa Springs during the Colorado School of Mines Geophysics Field Camp. We used split-spread acquisition, with a Vibroseis source in the middle, and 120 stations on both sides of the source. Station spacing is 10 meters, and each station consists of a group of 6 geophones spread symmetrically around the station. In order to test repeatability, we repeated one shot (10 s long non-linear 4-128 Hz upsweep) 100 times within 1 hour. An example of a shot record for repeatability experiment is shown on Figure 1. Some records were affected by noise from the sporadic traffic. After sweep correlation, traffic noise is very strong and has spike-like character. Thus, affected samples appear as extreme outliers in data and bias the derivation of distributions. To mitigate the effect of traffic noise on uncertainty analysis, we apply a median filter along the experiment axis, as only some shots among the 100 experiments were affected by traffic noise. Figure 2 shows a receiver gather before and after filtering. Only median filtering is performed on data since our objective is to look at their uncertainty with as small a processing imprint as possible.

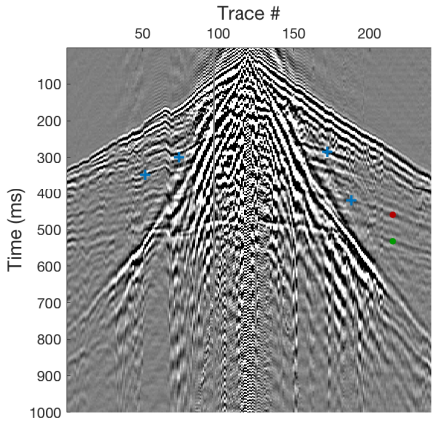


Figure 1: An example of a shot recording for 100 experiments. The red and green dots indicate points for which reflection amplitude and position are analyzed.

The mean amplitude and standard deviation maps are shown on Figures 3a and 3b. The standard deviation map is presented in the logarithmic scale. Compared to the shot record (Figure 1), the mean amplitude map is less noisy, but all seismic events are consistent. The biggest uncertainty corresponds to the region directly under the seismic source. In general, uncertainty decreases with increasing offset. Furthermore, we note higher uncertainty for waves traveling in the shallow parts of the subsurface (surface waves and head waves) and through the air. Uncertainty related to reflections is small compared to that of other events and does not stand out. Some patterns visible on the standard deviation maps are artifacts from imperfect traffic noise

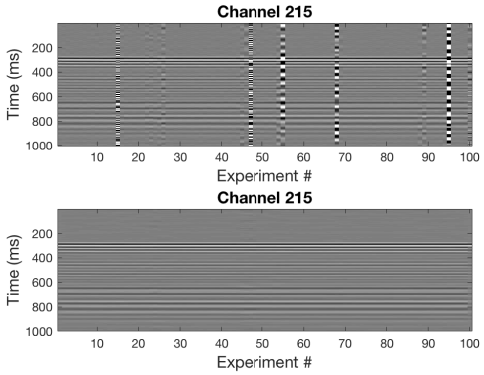


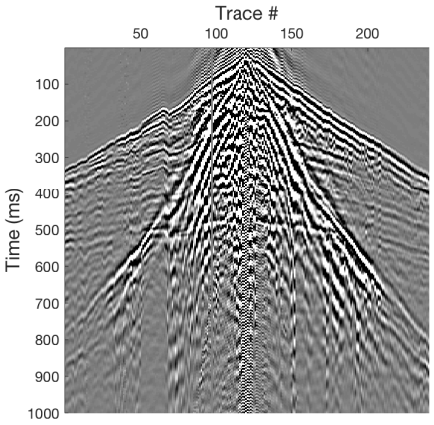
Figure 2: Receiver gather with car noise (top) and after median filtering along the experiment axis (bottom).

attenuation (note the scattered pattern on some channels on the left side of the source), indicate dead traces (black line), or indicate consistently noisy channel (purple line).

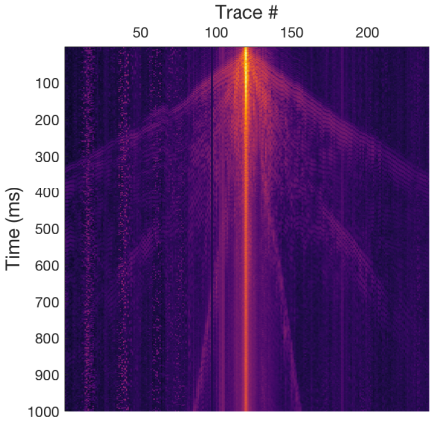
Probability density functions for data points indicated by blue crosses on Figure 1 are shown in Figure 4. The amplitude range is different for different points, but the bin size is constant. The shape of the distributions differs from Gaussians, especially on Figures 4c and 4d, implying that noise in data, when observed within one hour, is not of Gaussian character.

With the aid of the trace fidelity plot (Figure 5) and shot record, we select a good trace with strong reflection signal. The red and green dots on Figure 1 indicate reflections of interest. Figure 6 shows the raw trace, its mean and its standard deviation as a function of time. Again, we note that mean is a less noisy version of a raw trace, having similar effect to the stacking procedure. The largest uncertainty corresponds to the strongest events.

For all experiments, we track the events marked by the red and green dots. The kinematics of the reflections, shown on Figure 7, are highly consistent, differing only by a maximum of 3ms for the shallower event and 1ms for the deeper event. The reflection amplitudes, also shown on Figure 7, are not as consistent; one may observe fluctuations that are visibly correlated for the two events. Some of the fluctuations can be explained by the imperfect removal of traffic noise (large amplitude decrease for shot 53), but in general, changing amplitude is due to uncontrollable environmental conditions. The solid black line indicates the mean amplitude, while dashed lines mark ± one standard deviation. Repeatability R, as defined in equation 4, is 5.4% for shallower event and 4% for the deeper event. This gives an accurate idea about the magnitude of changes expected in the observed reflection amplitudes that are not re-



(a)



(b)

Figure 3: (a) Mean and (b) standard deviation for 100 experiments, computed using equations 1 and 2.

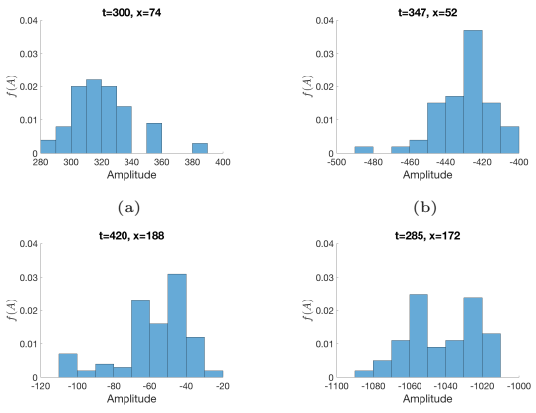


Figure 4: PDF's for data points indicated by blue crosses on Figure 1. Distribution shapes are not Gaussians.

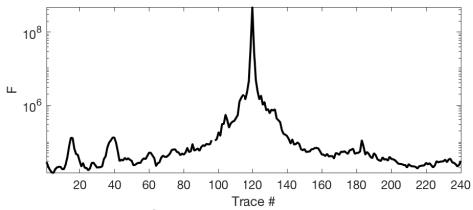


Figure 5: Trace fidelity computed using equation 3.

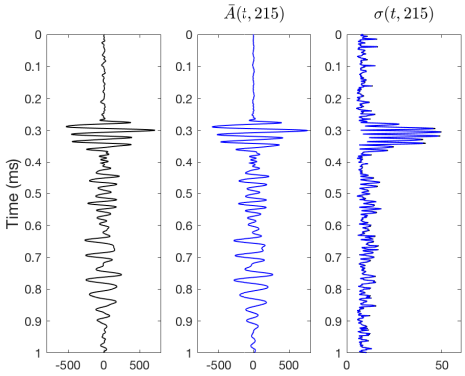


Figure 6: From left to right: a trace 215 extracted from experiment 2, its mean and standard deviation as functions of time.

lated to changes in elastic moduli of a medium. This parameter can also feed into subsequent processing, e.g. through FWI.

DISCUSSION AND CONCLUSIONS

Our analysis shows that data uncertainty is variable and depends on the sample location in time and space. The most uncertain data region is directly below the seismic source. This is related to the experimental conditions, since the coupling between the shaker's plate and the dirt road changes as the ground compacts. It is an unavoidable pitfall of repeated experiments in the field: elastic properties in close vicinity of the source change as the source is activated over time. The biggest change in amplitude likely caused by compaction can be observed at the beginning of the experiments, as depicted by Figure 7. However, such changes have minor effect on data as the offset increases.

High energy events, such as surface waves and head waves, also have high uncertainty. Traveling in the very shallow subsurface, they are the most affected by environmental factors. Assuming no changes to the medium, the traveltimes of waves do not change. However, amplitude changes because of noise, which may cause the apparent shift in observed traveltimes, as is the case for the reflection traveltime shown on Figure 7. Due to the short time in which experiment was performed, the noise

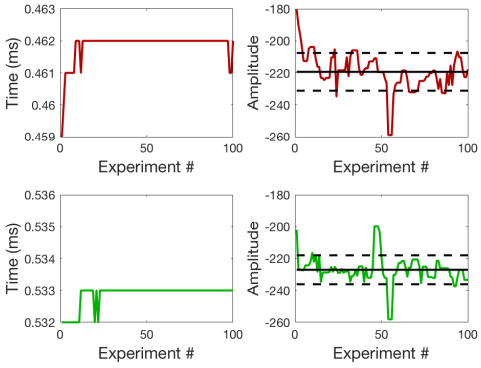


Figure 7: Reflection time uncertainty (left) and amplitude uncertainty (right) for events indicated on Figure 1.

causing amplitude distortion is not related to changes in the temperature or soil saturation. The uncertainty is approximately proportional to the magnitude of a seismic event. Therefore, the noise is not a simple random variable with fixed standard deviation and should not be modeled as such. We interpret that uncertainty is mostly related to sensors noise.

Seismic reflections, whose energy is orders of magnitude smaller than the energy of the surface waves, also have much smaller uncertainty. The kinematics of reflections are highly repeatable, while amplitudes fluctuate, albeit consistently. Therefore, inversions that use the amplitude information can potentially be more uncertain than those relying on traveltimes only.

We have shown that seismic data distributions are not Gaussian. Nor are they consistent for different $A(t, x)$ samples - note different shapes in Figure 4. The question that springs to mind is: how would using the true distributions affect the results of inversion? Data prior is often treated as a normalization constant in the Bayesian framework (Ely et al., 2018). Incorporating the information about uncertainty may improve posterior estimates. For the deterministic framework, uncertainty information is a viable alternative to the discrepancy principle in determining optimal data misfit, or can be used to pick data regularization term.

ACKNOWLEDGEMENTS

We acknowledge Dawson and Sercel for providing equipment and technical support at the CSM Geophysical Field Camp. SEG provided partial financial support of the Field Camp. We also acknowledge undergraduate students who helped with data acquisition.