Optimal survey design for marine borehole seismics
Darrell Coles*, Schlumberger-Doll Research, Yi Yang, WesternGeco, Hugues Djikpesse, Michael Prange, Schlumberger-Doll Research, and Konstantin Osypov, WesternGeco

Summary
Optimal experimental design is the theory and practice of maximizing the expected information in data in order to minimize the post-analysis uncertainty in parameters constrained by those data. In this article, a borehole seismic survey is optimized in the Gulf of Mexico with the \( D_{N} \)-criterion, a nonlinear design objective function amenable to efficient optimization. Several potential applications are proposed, including: (1) identifying the maximum offset for 3D vertical seismic profiling (VSP); (2) an annular-spiral 3D VSP geometry to reduce acquisition time/cost; (3) a ‘Z-geometry’ for presurvey data borehole seismic acquisition; (4) optimizing source vessel placement for offset/azimuth checkshots; (5) identifying optimal data for near-real-time quality control and inversion; and (6) optimally controlling data set decimation to maximize the efficient and accurate analysis of large data sets.

Introduction
Geological model uncertainty is an important source of information in hydrocarbon exploration. Quantified descriptions of uncertainty are particularly important because they can be used formally to identify and characterize risks and to anticipate potential difficulties throughout the reservoir life cycle, all of which being factors that vitally inform decision-making. Failing to properly account for uncertainty is to disregard the value of information (Bratvold et al., 2009) it brings to the decision-making process (Osypov et al., 2011). Consequently, it is also valuable to minimize this earth model uncertainty because so doing minimizes the risk associated with potentially large investments.

One way to minimize risk in this context is to apply statistical experimental design (SED) – the theory and practice of optimizing experiments to maximize the expected information in data observations. Nonlinear SED is a special branch dedicated to nonlinear data-model relationships, in which the information content of data varies with the earth model, unlike in linear SED (Atkinson et al., 2007, p. 248). It is important to address nonlinearity because most data-model relationships (theoretical functions, hereafter) in hydrocarbon exploration are nonlinear and affect model uncertainty in complicated ways (cf. Guest and Curtis, 2011; Coles and Curtis, 2011b; Winterfors and Curtis, 2008).

The information-dependence of data on models may be accommodated by representing the earth model, the expected data arising from it, and measurement errors, with probability distributions. However, probabilistic approaches typically involve high-dimensional integrals that are expensive to compute and thus limit the size of the experiments that can be optimized (Coles and Curtis, 2011). The \( D_{N} \)-criterion was introduced as a possible solution to the nonlinear problem (ibid). It is a nonlinear design objective function that can be maximized using very efficient algorithms from linearized design theory. This makes the \( D_{N} \)-criterion capable of optimizing large-scale experiments, a feature common to only a few nonlinear design methods (Coles and Prange, 2011; Coles and Curtis, 2011b; Coles, 2011; Guest and Curtis, 2011; Haber et al., 2008). Here, we sketch a derivation of the \( D_{N} \)-criterion which elucidates how it works while eliminating a restrictive assumption from the original derivation (Coles and Curtis, 2011b).

Theory
A common goal in survey design is to identify an experiment that maximally discriminates between two or more theoretical functions that are thought to model some observed data, and this is typically done by optimizing a hypothesis test (Fedorov, 1972, p. 226). This idea may be modified to handle competing parameterizations of the same theoretical function as follows. Without loss of generality, one parameterization may be treated as ‘true’ (null hypothesis) and another is treated as some ‘alternative’ (alternative hypothesis). The object in optimizing a discriminating survey is to maximize the odds that the alternative hypothesis is rejected, which ensures that the model parameters most likely to explain an observed data set are, in fact, the true ones.

To develop the idea, let

\[
d(\mathbf{m}, \hat{\mathbf{z}}) = g(\mathbf{m}, \mathbf{z}) + \varepsilon(\mathbf{m}, \hat{\mathbf{z}}) \tag{1}
\]

be a known mathematical model, where \( d \) is a vector of data observations made at observation points \( \mathbf{z} \), \( \mathbf{m} \) is the vector of model parameters, \( g \) is a known theoretical function relating \( d \) and \( \mathbf{m} \), and \( \varepsilon \) is a vector of stochastic measurement errors.

We assume that \( \mathbf{m} \) has a known prior distribution, \( \rho(\mathbf{m}) \), which characterizes the state of knowledge about \( \mathbf{m} \) before any new data are acquired, and likewise, \( \varepsilon \) has a known distribution.

A discriminating test commonly used in experimental design is the log-likelihood-ratio test (Fedorov, 1972, p. 249), which expresses the odds ratio of the null and alternative hypothesis.
Denoting the true model and its corresponding data by \( \mathbf{m}_0 \) and \( \mathbf{g}(\mathbf{m}_0) + \varepsilon \), respectively, and denoting an alternative model \( \mathbf{m}_1 \), the log-likelihood-ratio in question is

\[
\ln \Lambda = \ln \frac{L[\mathbf{g}(\mathbf{m}_0) + \varepsilon | \mathbf{m}_0]}{L[\mathbf{g}(\mathbf{m}_1) + \varepsilon | \mathbf{m}_1]},
\]

(2)

where \( L \) is the data likelihood function (dependence on \( \xi \) is suppressed for ease of notation). Maximizing \( \Lambda \) with respect to \( \xi \) maximizes the odds that the true model, \( \mathbf{m}_0 \), will be accepted and the alternative model, \( \mathbf{m}_1 \), rejected.

The log-likelihood ratio in expression 2 is defined for a single \( \mathbf{m}_0 \) and \( \mathbf{m}_1 \) and may be generalized for all probable model parameterizations by averaging it over \( \mathbf{m}_0, \mathbf{m}_1 \sim \rho(\mathbf{m}) \), which gives

\[
\mathbb{E}_e \ln \Lambda = \int \pi(\mathbf{m}_0, \mathbf{m}_1) \ln \frac{L[\mathbf{g}(\mathbf{m}_0) + \varepsilon | \mathbf{m}_0]}{L[\mathbf{g}(\mathbf{m}_1) + \varepsilon | \mathbf{m}_1]} \, d\mathbf{m}_0 \, d\mathbf{m}_1,
\]

(3)

where \( \mathbb{E}_e \) is the expectation operator over the joint distribution of \( \mathbf{m}_0 \) and \( \mathbf{m}_1 \). \( \pi(\mathbf{m}_0, \mathbf{m}_1) = \rho(\mathbf{m}_0) \rho(\mathbf{m}_1) \).

When \( \varepsilon \) is Gaussian with zero mean and covariance \( \Sigma_\varepsilon \), it is easy to show that equation 3 simplifies to

\[
\mathbb{E}_e \ln \Lambda = \text{tr} \left( \Sigma_{\varepsilon}^{-1/2} \Sigma_\varepsilon \Sigma_{\varepsilon}^{-1/2} \right),
\]

(4)

where \( \text{tr} \) is the trace operator and

\[
\Sigma_{\delta} = \frac{1}{2} \mathbb{E}_e \left[ \mathbf{g}(\mathbf{m}_0) - \mathbf{g}(\mathbf{m}_1) \right] \left[ \mathbf{g}(\mathbf{m}_0) - \mathbf{g}(\mathbf{m}_1) \right]^T.
\]

(5)

Maximizing the average log-likelihood ratio in expression 4 should therefore maximize the odds that the true parameterization is accepted over all probable alternatives.

The trace in expression 4, however, does not penalize zero eigenvalues. All eigenvalues must be nonzero for the data-model system to be well determined. Recalling that the trace equals the sum of eigenvalues, we replace it with the sum of log eigenvalues, which is given by the log of the determinant. This sum goes to negative infinity for any \( \xi \) that causes \( \Sigma_\varepsilon^{-1/2} \Sigma_\varepsilon \Sigma_\varepsilon^{-1/2} \) to be singular, which is equivalent to having an undetermined system. Thus, we arrive at the DN-criterion,

\[
\Phi = \ln \det \left( \Sigma_{\varepsilon}^{-1/2} \Sigma_\varepsilon \Sigma_{\varepsilon}^{-1/2} \right) = \ln \left| \frac{\Sigma_{\delta}}{\Sigma_{\varepsilon}} \right|.
\]

(6)

This derivation avoids the assumption that \( \mathbf{g}(\mathbf{m}_0) - \mathbf{g}(\mathbf{m}_1) \) is multivariate Gaussian (multinormal), a strongly limiting assumption from the original derivation (Coles and Curtis, 2011b).

It is easy to show that \( |\Sigma_{\varepsilon}| = |V_\varepsilon||R_\varepsilon| \), where \( V_\varepsilon = \text{diag}(\Sigma_{\varepsilon}) \) is the diagonal variance matrix and \( R_\varepsilon \) is the correlation matrix of the theoretical data. Therefore, as a rule, \( \Phi \) must generally increase with the variance, and decrease with the correlation ( \( |R_\varepsilon| \) decreases as correlation increases), of the theoretical data. This means that it is easier to discriminate between model parameterizations if the predicted data vary greatly from parameterization to parameterization, accounting for data noise (this is the likelihood objective). It also means that it is easier to discriminate between parameterizations if their predicted data are expected to be uncorrelated (this is the degrees of freedom or objective).

Several optimization algorithms exist that may be used to efficiently maximize \( \Phi \), including the sequential Construction, Exchange, and Decimation Algorithms (cf. Coles and Curtis, 2011a). These algorithms are greedy in that a solution is optimized through a sequence of locally optimal updates in the hope that the result is close to the global optimum. Coles and Curtis (2011a) showed evidence that some greedy techniques may approach quite close to global optima.

3D VSP optimization

We demonstrate D_N-optimization for 3D VSP at a real field located on the Walker Ridge in the Gulf of Mexico. Several exploratory wells have been drilled in a site of interest and have shown signs of hydrocarbons. A reservoir has been conjectured beneath a salt trap at this site and is the impetus for this demonstration.

An existing borehole was chosen which transects the salt trap and passes through the inferred reservoir beneath. This borehole was ’equipped’ with a virtual string of forty equispaced geophones encompassing the reservoir interval (Yi Yang, personal communication). Receiver spacing was 22.9 m, and the top receiver was placed 9053 m below sea surface. A virtual carpet of candidate shots was placed at the sea surface, centered on the wellhead, with 91.4-m shot spacing in each cardinal direction. From this candidate set, an optimal subset of shots was to be selected. The carpet covered ~325 km² and comprised 40000 candidate shots (Figure 1).

Figure 1 A borehole (white line) is centered in the target region and a virtual string of 40 geophones (black dots) is placed over the inferred reservoir interval. Blue-white-red indicates P-wave velocities for one of the 500 models. The candidate shot carpet is exemplified by orange dots. The horizontal surface (blue-green-red) depicts the salt trap horizon. Earth model is from Osypov et al., 2011.

The earth model was previously characterized by a structural uncertainty workflow (Osypov et al., 2010; Osypov et al.,
D<sub>N</sub>-criterion for nonlinear survey design

(2011) which described knowledge of the anisotropic model parameters by a multinormal distribution based on previous inversion of wide-azimuth (WAZ) surface seismic data. From this multinormal distribution, five hundred VTI earth models were randomly sampled, creating an ensemble characterization of the current state of model uncertainty. The 500 models were 3D meshes of the elastic properties \( V_p, \varepsilon, \) and \( \delta \). The model cubes were also centered on the wellhead, with areal extent identical to the candidate shot carpet and extending 15 km in depth from the sea surface. The model ensemble was used as prior information by the \( D_N \)-optimizer.

Because \( D_N \)-optimization operates in data space, it was necessary to compute the P-wave traveltimes for all combinations of shots, receivers, and models. In all, 800 million traveltimes = 40 (receivers) x 40000 (shots) x 500 (models) were computed.

Before looking at the optimal borehole survey results, let us examine the expected data variability and correlation over the ensemble of 500 anisotropic earth models. The \( D_N \)-criterion should select shot locations where there is large expected data variability and low data correlation.

Figure 2 shows in map view the traveltime variability of the candidate shot carpet for a selection of geophones. The color at any position signifies the variability of the traveltime (over the ensemble of 500 models) for a candidate shot located there with respect to the noted receiver. Each receiver exhibits a crude annulus of high variability that increases with geophone depth. Notice the radial ‘fingers’ and ring-like features that appear, disappear, or persist at different receiver depths. These are probably due to model structures and anisotropy. Most notably, the greatest traveltime variability occurs in the northwest quadrant, associated with upper and middle receivers, and in the southwest quadrant, for deeper receivers.

Figure 3a shows in map view the spatial correlation of the traveltime at one candidate shot location with respect to the entire shot carpet. Notice the preferential alignment of correlation along a SW-NE direction, which is probably a manifestation of model anisotropy. Furthermore, notice the strong and widespread distribution of correlation, which indicates that neighboring shots in a fairly large area will provide little complementary information relative to a selected shot. This does not mean that nearby shots should be ignored, only that they reinforce the information provided by the selected shot. Figure 3b shows spatial correlations for a shot positioned roughly 1 km WNW of the shot in Figure 3a. Notice the reduction in correlation; the shallow receivers show little or no spatial correlation at all, and while the deeper receivers exhibit a larger area of correlation, the correlation magnitude is less than in Figure 3a.
The $D_N$-criterion would, therefore, favor the shot in Figure 3b over that in Figure 3a because it has low spatial correlation and high traveltime variability (Figure 2).

Figure 4 shows a $D_N$-optimal survey of 1500 shots. The optimization was carried out with respect to the complete borehole string because the whole string would be used to observe any one shot in a real situation. Thus, $60000 = 40$ (receivers) x 1500 (shots) shot-receiver pairs were actually optimized. The optimized shots roughly occupy an annulus between 6 and 9 km radial distance from the wellhead, which was already anticipated in our discussion of Figure 2. This suggests that shots closer than 6 km offset are relatively uninformative. One explanation for this is that the earth model parameters sensitive to shots within this region have already been constrained by the WAZ surface seismic data. The most informative regions are now associated with long offsets and non-vertical incidence. The clustering of optimal shots in Figure 4, particularly in the northwest quadrant, was also anticipated in Figure 2 and Figure 3, where coincident regions of high expected data variability and low data correlation were seen.

**Discussion and conclusion**

Based on the result in Figure 4, we can identify a maximum radius for a spiral 3D VSP, which would be around 9 km from the wellhead. The ability to systematically recommend a maximum radius for spiral VSP is useful because it ensures that the most informative data are collected while minimizing acquisition costs.

In Figure 4 we can also identify a minimum radius for spiral 3D VSP. Shots closer than 6 km to the wellhead are uninformative and can be disregarded for tomography. We may use this fact to prescribe an annular-spiral 3D VSP that covers only the annular region between 6 and 9 km from the wellhead. By foregoing offsets less than 6 km, the acquisition time can be reduced by more than 40% compared with a complete spiral from the wellhead to 9 km.

$D_N$-optimization could also be used to design a presurvey acquisition geometry in which the expected measurement noise would be optimally characterized to ensure maximum data quality during actual data acquisition. In this example, the geometry could be a ‘Z-survey’, a walkaway-azimuthal VSP in which the top bar of the ‘Z’ would be an azimuthal segment through the hot region in the northwest, the diagonal part would be a walkaway traversing the wellhead, and the bottom bar (which might be reversed in this case) would be a second azimuthal segment through the hot region in the south. Such a geometry would ensure that the most informative data are acquired, ensuring good coverage for the characterization of noise and data quality.

Another use for $D_N$-optimization would be to produce real-time ‘information maps’ for ‘steering’ an acquisition vessel or to place it in the vicinity where maximally-informative shots are expected to occur. Such an approach could be used, for example, to identify far-offset checkshot positions to optimally constrain look-ahead models in real-time drilling.

$D_N$-optimization could also be used for rapid post-acquisition quality control. The most informative shots occur in the aforesaid annulus, and they identify the best data to be inverted to render a maximally-informative quick-look at the model, which would be emerging as data are acquired.

One last idea would be to use $D_N$-optimization to decimate an existing data set that is too large to be practically analyzed *in toto* or that needs to be rapidly analyzed for decision-making purposes. The idea would be to systematically find a suitably small subset of the data that maximally constrains the earth model – or any targeted subset thereof. This has broad potential application, including for tomography, least-squares migration, full-waveform inversion, and reverse-time migration for local seismic imaging, to name a few.

Apart from the several notable contributions above, $D_N$-optimization can handle industrial-scale nonlinear design problems. Though a modest body of work exists on nonlinear design, much remains to be done to make it practical for large problems. The ability to probabilistically optimize experiments for the nonlinear case is essential because posterior model distributions are complicated by nonlinearity and $D_N$-optimization accomplishes this while still being computationally feasible for real-world problems.
EDITED REFERENCES

Note: This reference list is a copy-edited version of the reference list submitted by the author. Reference lists for the 2012 SEG Technical Program Expanded Abstracts have been copy edited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES


Coles, D., and A. Curtis, 2011b, Efficient nonlinear Bayesian survey design by D<sub>N</sub>-optimization: Geophysics, 76, no. 2, Q1–Q8.


Fedorov, V.V., 1972, Theory of optimal experiments: Academic Press.

Guest, T., and A. Curtis, 2009, Iteratively constructive sequential design of experiments and surveys with nonlinear parameter-data relationships: Journal of Geophysical Research, 114, B04307.


Guest, T., and A. Curtis, 2010, Optimal trace selection for AVA processing of shale-sand reservoirs: Geophysics, 75, no. 4, C37–C47.


