# Experimental Design for $CO_2$ Sequestration Monitoring with Gravity Data

*Alison E. Malcolm*, and Paul Barbone^*
**Memorial University of Newfoundland, ^Boston University*

## Summary

The information required from geophysical data for $CO_2$ sequestration is quite different than that required for oil and gas exploration and development. It is particularly important to choose cost effective experimental setups – and those setups need not constrain all of the details of the subsurface, they need only constrain the existence and perhaps extent of leakage. We use the simple example of gravity data to test various experimental design methods from information theory. We find that simple techniques are not particularly effective while more complicated techniques quickly become computationally intractable. We show an asymptotic approximation that is promising for estimating optimal locations of multiple sensors.

## Introduction

The goal of experimental design is to effectively choose locations for sensors to maximally constrain the resulting quantities of interest. For example, here we search for the best locations of gravimeters to constrain the mass of an underground spherical body. Although this is a highly simplified problem, it represents the idea of recovering the mass of CO2 remaining in the subsurface over time. Most experimental design methods involve optimization to choose the ideal sensor locations. We refer readers to Curtis (2004), Ajo-Franklin (2009), Curtis and Arnold (2018), and Strutz and Curtis (2024) for details on experimental design. We test two of these methods for our simple gravity problem, the inner-product method (Curtis (2004)), and the Expected Information Gain method (Arnold and Curtis, 2018), for which we use an asymptotic formulation to substantially reduce computation time.

## Theory

Gravity data are straightforward to model and fit nicely into a sparse acquisition structure (i.e. only a few stations). Defining $m$ to be a vector of masses defined at locations $(x, z)$ in two dimensions, and data $d$ recorded at locations $(x_{obs}, z_{obs})$ we have that

$$d(x_{obs}, z_{obs}) = \boldsymbol{A}(x_{obs}, z_{obs}, x, z)\, m(x, z),$$

where $A$ is a well-defined mapping from model to data (based on $F = \frac{G\, m_1 m_2}{r^2}$ see e.g. Dentith and Mudge (2014)).

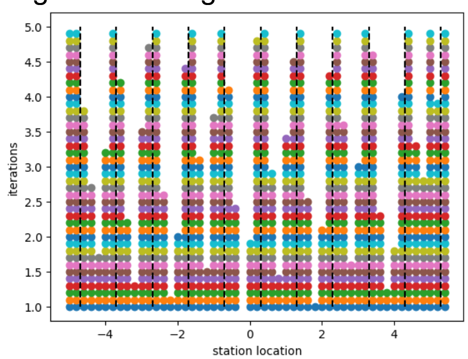
The matrix $\boldsymbol{A}$ defined in equation (1) defines the relationship between the model and the experimental data. We can plan experiments based only on this matrix, without needing to model or acquire the data $d$.

Method 1: Inner Products. Following Curtis (2004) one way that we can to understand experimental design is to recall that each row of the $\boldsymbol{A}$ matrix is an equation relating model parameters to data points. If these equations are similar, then the system in equation (1) will be ill-conditioned and the resulting parameter recovery poor. By comparing the rows, we can determine which data to exclude. This method is quick to run and easy to implement, though it will become challenging for large matrices.

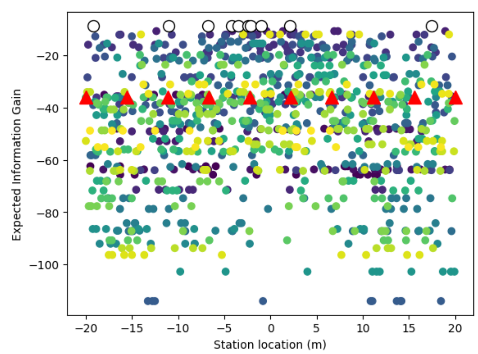
Method 2: Expected Information Gain. Following Curtis and Arnold (2018), another commonly used measure is the so-called 'Expected Information Gain' or EIG. The EIG is a statistical measure of how much information is available in a particular experimental design. Unlike the inner-product method, this method is quite slow to compute as it involves sampling several distributions, requiring thousands of evaluations of the forward model. Here we show results of an asymptotic approximation to the EIG useful for situations in which we have many stations.

## Results, Observations, Conclusions

Figure 1 compares the results of the two methods. Note that the inner-product method gives a result that depends strongly on the chosen model grid, whereas the EIG method does not. By contrast, the EIG method is assuming a fixed anomaly location whereas the inner-product method allows for the anomaly to be in any part of the domain. The EIG method can be extended to this more general configuration.



(a)



(b)

Figure 1: (a) The chosen stations for each iteration using the inner-product formulation. This method suggests a uniform grid, and the spacing of stations depends strongly on the chosen model grid locations (black dashed lines). (b) The asymptotic approximation of the Expected Information Gain (EIG) for randomly chosen station configurations. This measure suggests clusters of stations near the peak, and in the tails of the data. The best dataset is shown with white circles, while the uniform sampling geometry is shown as red triangles. The colours of other configurations are not meaningful – they are used only to help connect stations in a single configuration.

## Novel/Additive Information

The novelty in this work involves the asymptotic approximation, which makes the calculations feasible for a large number of stations.

## Acknowledgements

This work is funded by the Natural Science and Engineering Research Council, through the Discovery Grant Program.

### References

Ajo-Franklin, Jonathan B. "Optimal experiment design for time-lapse traveltime tomography." *Geophysics* 74.4 (2009): Q27-Q40.
Curtis, Andrew. "Theory of model-based geophysical survey and experimental design: part 1—linear problems." *The Leading Edge* 23.10 (2004): 997-1004.
Curtis, Andrew, and Richard Arnold. "Interrogation Theory." *Geophysical Journal International* 214.3 (2018): 1830-1846.
Dentith, Michael, and Stephen T. Mudge. *Geophysics for the mineral exploration geoscientist*. Cambridge University Press, 2014.
Strutz, Dominik, and Andrew Curtis. "Variational Bayesian experimental design for geophysical applications: seismic source location, amplitude versus offset inversion, and estimating CO2 saturations in a subsurface reservoir." *Geophysical Journal International* 236.3 (2024): 1309-1331.