## Modeling Uncertainty in the Earth Sciences

Course project

**Part I: spatial uncertainty and sensitivity analysis**

**Background**

In this task you will be studying the concept of “entropy” of an image. In probability theory, entropy of a discrete random variable *X* is defined as



Clearly the highest entropy is obtained when all probabilities are the same (the most randomness) and the lowest entropy is when one category has probability 1 and the rest zero (deterministic). One can also define the entropy of a 2D image (or Earth model). To do so, we use a window (or template) that records the frequency of certain patterns in the image. A pattern is simply the pixel values recorded within a given window. We will use only square windows here. If only one pattern exists, then we have low entropy, while if a large number of patterns exists each with somewhat equal frequency then entropy (randomness) is high. Since the frequency of patterns is dependent on the window size used to define and extract patterns from the image, as a result, the entropy is also function of the window size:



where **PAT**(w) is the random vector describing the variation of patterns observed with window size w × w, **pat***i* are the various possible outcomes of patterns. A simple matlab code does the calculation as follows:

max\_size = 21; % maximum window size to be used

entropy = zeros(1,(max\_size-1)/2); % only odd values of w will be used

for w = 1:(max\_size-1)/2

plogp = entropyfilt(EarthModel, ones(2\*w+1));

entropy(w) = mean(plogp(:));

end

**Tasks**

1. Using the Boolean simulation tool in SGEMS, create 120 Earth models on a 100 x 100 grid using the following model uncertainty specifications
   1. The object is a simple rectangle
   2. The length (x-direction) varies uniformly between 5 and 15 pixels
   3. The width (y-direction) varies uniformly between 3 and 6 pixels
   4. Three total proportions (of cubes) exist p=0.2, p=0.25, p=0.3 each with equal probability of occurrence
   5. Two spatial trends exist: 1) no trend and 2) EW trend, each with equal probability
2. Document how you created these models and show a few with plots.
3. Calculate the entropy *E*(w) and display all curves on one plot.
4. Using the difference between entropy of the 120 Earth models create an MDS plot, with as color indicator *E*(11). Make sure the distance you chose reflects correctly the difference between the entropy functions (hint: don’t use the default pdist).
5. Perform a sensitivity analysis on what input parameter influences entropy most.
6. Explain the result of this sensitivity analysis. Does this make sense?

**Part II: Value of Information**

**Description**

You are an engineer working for the local water district responsible for a water recharge project. Your main task is to prevent saline water intrusion into an aquifer whose water is used for drinking water for the local residents. You are in the process of assessing whether drilling wells to inject fresh water is needed to prevent sea-water water intrusion into the aquifer. Figure 1 describes the setting of the problem. You have money to drill a maximum of two wells. The cost of drilling one well is $30000. However, only two site locations are available because of restricted access in a suburban area. You have determined that such wells need to be drilled in that part of the subsurface formation that contains clean sand. Otherwise any injection would not be effective. It has been determined that doing nothing will result in a cost to desalinate the water of $60000. However with the successful drilling (success means=it is drilled in sand) of each additional injection well, that cost will be divided by 3. If you decide to go for two wells, than they will have to be drilled at the same time.

The subsurface is known to consist of clean sand with clay barriers in between; however there is much uncertainty about the nature of these barriers. A preliminary geological report shows that the nature of the sand/clay heterogeneity can be roughly described as follows:

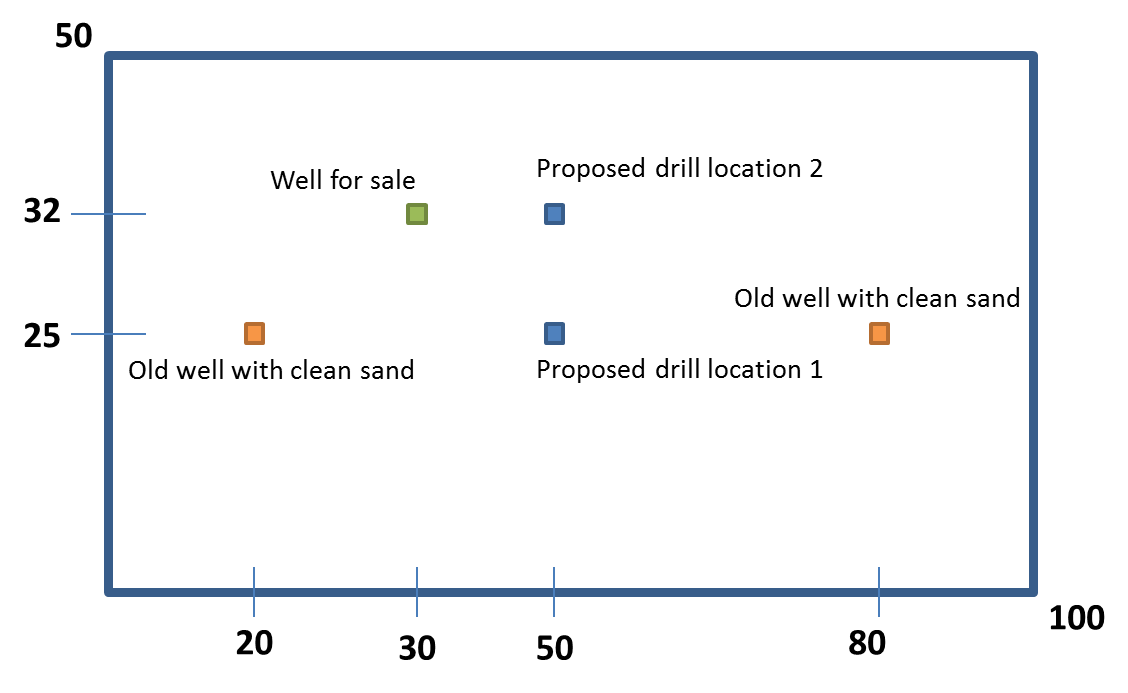
* Two possible scenarios exist for the sand bodies, each with equal probability: they could be sand point bars of elliptical shapes deposited in a fluvial systems, or they could be sand flats of cuboid shape. In each case the width and length of these objects are the same, namely width=5 and length=25. They are elongated along the EW.
* The proportion *PROP* of objects is uncertain, namely

P(*PROP*=0.3)=0.7 and P(*PROP*=0.4)=0.3

* The formation is thin enough to be modeled as a 2D Earth model

Next to geological information you have two old wells available, both of them are drilled in sand, see Figure 1.

**The question is simply**: How many wells would you drill and if you decide to drill only one well, what location?

****

**Figure 1: Description of the study area and problem**

**Value of information question**

A consultant approaches you at a local conference and mentions that additional studies have been done at your site, namely an additional well has been drilled and logged and the data from that well-log would reveal the true nature of the formation at the drilled location. What would be the maximum price you would pay for this data? Report clearly the various steps and decisions you took in answering this question.

However, a few months later, while you are still working on this study, the same consultant contacts you again and that he/she learned he/she was overly optimistic about the fact that the drilling would reveal the true nature of the deposit at that location. In fact, the consultant said he/she had done some investigations on the logging companies record and found that in cases when the true formation is sand, the interpretation from the well-logging is only 80% of the time correct, while the interpretation of shale is only 70% of the time correct. Would you still buy the data? Document how you would solve this problem.