## Modeling Uncertainty in the Earth Sciences

**Matlab on modeling response uncertainty**

**The context**

In this homework you will assess the uncertainty of the cumulative oil production in an actual offshore reservoir. The reservoir concerns a turbidite reservoir with complex geological formation of channel sand in a shale matrix. A geologist has constructed 72 Earth models by varying the following geological parameters

* The channel thickness
* The channel-width to thickness ratio
* The channel sinuosity
* The total proportion of channel sand

In order to calculate the cumulative oil, flow simulation is required. However, each flow simulation takes 12 hours to complete, so the reservoir engineer cannot afford to run flow simulation on each of the 72 models. Your task is two-fold

1. Select a limited number of models to run flow simulation on
2. Rank the geological parameter in term of their impact on the cumulative oil production

Instead, the reservoir engineer has a “quick-and-dirty” (*qad*) flow simulator that takes only 30 minutes to run, however, the results from this flow simulator are only an approximation of the actual physical simulation that takes 12 hours.

**What data is available?**

Load the file “homework5.mat” in matlab. You will see it contains:

1. A 72 × 4 table “Parameters” containing the values of the 4 parameters used to generate the 72 Earth models (you don’t need the actual Earth model, so we don’t give it to you). For each column you have in the following order: the channel thickness, width-thickness ratio, channel sinuosity and % of sand for each model and whether that parameter value was “H” (High), “M” (Medium”) or “L” (Low). Note the actual value is not important to you in this homework, just whether they are high, medium or low.
2. A 72 × 22 table ‘qad\_simulation’ with the response calculated from the *qad* simulation. The 22 refers to the 22 consecutive time steps at which the response is calculated. What the actual time step is, is not important here, so call the first time “1” and the last time “22”. You cannot use the *qad* simulation to predict oil production.

In addition, we also provide the true flow simulation results, i.e. those that took 12 hours to simulate. In reality you do not have those, but we give all of them to you so you can “flow simulate” those few models that you select. The true flow simulations are in a different table of size 72 × 39, named ‘True\_simulation’ so in this case we have 39 time steps.

**Task 1: Uncertainty Assessment**

1. Using the Euclidean distances between the *qad* responses, make an MDS map in 2D and 3D. How many positive eigenvalues do you get?
2. Now do a kernel transformation, and make the same MDS plot after kernel transformation. How many eigenvalues do you get? What is the difference between the two plots?
3. Repeat 1 and 2 for the true response, giving you the true unknown MDS map. Does it compare well with the proxy response?
4. Perform k-medoid clustering (go back to the proxy response). Experiment with the number of clusters. What do you think is a good number to retain?
5. Plot the true responses of the selected Earth models. Does this compare well with the total set of true responses?
6. Calculate the P25 and P75 of the cumulative oil production from the selected set and for the complete set. Compare.

Report with sufficient amount of plots to demonstrate your answer.

**Task 2: Sensitivity analysis**

Determine the most sensitive geological parameters by estimating their “effects”

1. Using only the selected model responses
2. Using all model responses

Does this ranking make sense to you? Why?

**Task 3: Response surface analysis**

Perform a response surface analysis using a central composite design with =1 (see Figure 10.4). Note that you have *k*=4 factors and *s*=3 levels. Call the low level ‘-1’, the mid-level ‘0’ and the high level ‘+1’.

1. How many response evaluations does this design call for?
2. For those responses available in the database of 72 responses (not all responses in the design are available to you), create a response surface with those set of design points. Choose whether you would use linear, interaction and/or quadratic terms.
3. Make a Pareto plot with the effect estimate of the parameters
4. Create a response surface with all 72 responses.
5. Do both response surfaces compare well?

You can use the matlab function regstats to create response surfaces