LA-UR-17-25281

Title: Sequential Design of Experiments

Author(s): Anderson-Cook, Christine Michaela

Intended for: Report

Issued: 2017-06-30

CCSI^2

Carbon Capture Simulation for Industry Impact

Sequential Design of Experiments

Christine Anderson-Cook
Los Alamos National Laboratory

Prepared for
U.S. Department of Energy
National Energy Technology Laboratory

5/30/2017

Revision Log

Revision	Date	Revised By:	Description
0	5/30/2017	Joel D. Kress	Original

Disclaimer

Table of Contents

1.0 Executive Summary

A sequential design of experiments strategy is being developed and implemented that allows for adaptive learning based on incoming results as the experiment is being run. The plan is to incorporate these strategies for the NCCC and TCM experimental campaigns to be run in the coming months.

This strategy for experimentation has the advantages of allowing new data collected during the experiment to inform future experimental runs based on their projected utility for a particular goal. For example, the current effort for the MEA capture system at NCCC plans to focus on maximally improving the quality of prediction of CO_2 capture efficiency as measured by the width of the confidence interval for the underlying response surface that is modeled as a function of 1) Flue Gas Flowrate [1000-3000] kg/hr; 2) CO2 weight fraction [0.125-0.175]; 3) Lean solvent loading [0.1-0.3], and; 4) Lean solvent flowrate [3000-12000] kg/hr.

The process for developing a plan for sequential design of experiments involves the following steps:

1. Identify one or more criteria over which to optimize. Common choices are (a) improving the precision (or reducing the uncertainty) in the estimation of model parameters, (b) improving the precision of prediction for new observation in the design region, or (c) quantifying the discrepancy between the model and data. These criteria have the characteristics that they will change as new data are included in the analysis, and the impact of new data at different locations in the design space can be evaluated for their impact on the criteria. If more than one criterion is going to be used, then identify how they will be combined into a utility function.
2. Develop a working model of the process that can be used to calculate the criteria values based on currently available knowledge and data.
3. Define the inputs that will be manipulated during the experiment, and the ranges of interest for these factors. Some inputs may be held constant during the experiment to simplify experimentation or to allow for adequate exploration of the most important factors of interest.
4. Identify candidate input factor locations that are being considered for new experiments. This can be a grid of input combinations or continuous regions in the design space. If there are combinations of the factors that will not yield results or that are not of interest, these regions of the design space should be excluded from consideration when optimizing which runs to select.
5. Develop a working model of the process that is able to receive new data (input combination where it was collected and the responses obtained) and incorporate them to update the calculated criteria values.
6. Based on the time required to set-up and run the experiments as well as the computational time required to process new data and update the working model, develop a plan for the size of the sequential design batches. For example, if the experimenter can run 2 runs per day with measured responses are available immediately, and computationally updating the model take 1 day, then it may make sense to aim for a batch size of 4, with the next data to be collected being identified every 2 days.
7. Identify the initial batch of experiments to be run at the beginning of the experiments based on the model developed in step 2. This involves examining the utility of new data at each candidate location, and comparing which locations have the highest anticipated utility. Since

the methodology is flexible enough to accommodate different criteria over which to optimize, the identification of the optimal batch requires customized statistical software (It is highly desirable to also have a plan for the second and later batches in case there are problems with the implementation of the sequential portion of the design of experiments plan).

8. Run the first batch of experimental runs, update the model developed in step 5 with the new results, and generate the next batch of experimental runs.
9. For the duration of the experiment, repeat steps 7 and 8 for subsequent batches based on the updated model after incorporating the newly obtained data.