Design Optimization of the Spacing Parameters in The Canadian Deep Geological Repository Containing High Level Radioactive Nuclear Waste

by

Shankai Lin

A thesis
presented to the University of Waterloo in
fulfillment of the
thesis requirement for the degree of
Master of Applied Science in
System Design Engineering

Waterloo, Ontario, Canada, 2018

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The Canadian Deep geological repository (DGR) is the facility designed to isolate and contain the highly radioactive nuclear waste at approximately 500m underground for approximately one million years. The reliability and safety of the DGR depends on the performance of the designed engineering barrier systems (EBS) as well as the environments at the site. In this thesis, surrogate models are utilized for optimizing the spacing of the current design of DGR by incorporating the thermal numerical model developed in finite element software COMSOL, which is hard to use for this purpose due to the long CPU time it takes.

The temperature rise caused by the radioactive waste buried inside the repository is analyzed by COMSOL. The result indicates that the temperature peak will likely happen within the first 100 years. Other than the material properties, the peak temperature also depends on the design of spacing between container and placement rooms. Increasing spacing reduces the peak temperature and thus increases the probability of meeting the temperature constraint, however, it increases the cost and the overall footprint of the DGR. Optimizing the design directly using the simulation is computationally expensive due to long simulation times, for example, a single simulation may take minutes to hours. As a result, surrogate based optimization is used to solve the optimization problem as well as computing the probability of meeting the temperature constraint for a given spacing design.

Using the surrogate-based optimization, the optimal design points based only on temperature constraint can be found with 19 COMSOL evaluations. The surrogate's mean absolute error (MAE) at these design points are 1.1 °C comparing to COMSOL results. Similarly, for the cost and multi-objective optimization problems, the surrogate converged with 16 and 25 COMSOL evaluations, the MAE for these optimal points are 0.4°C and 0.091°C. For these design problems, there are a total of 310 possible design combinations by varying the spacer width and the room spacing, the surrogate-based optimization greatly reduces the number of simulations comparing to an exhaustive simulation approach where all 310 points are evaluated for finding optimal designs.

For the uncertainty analysis, three variables are fit to the model, the possible input combination will be 10^3 if 10 values are sampled for each variable. The resulting surrogate model converged with 75 evaluations, the number of simulations is an order of magnitude less than the exhaustive simulation approach. The surrogate model estimates the temperature of the sample design model with MAE of 0.6°C. Comparing the surrogate prediction results to the results of the 320 known COMSOL evaluations

from other fitting process yields a MAE of 1.0°C. In these surrogate model experiments, the reduction in the total number of simulations are significant while the surrogate only slightly sacrifices the accuracy comparing to original model evaluation.

Acknowledgements

I would like to express my sincere gratitude for professor K.Ponnamabalam for providing me the opportunity to pursue a MASc. degree in System design engineering. Professor Ponnambalam's advice and support is the main reason for the completion of this thesis.

I would like to thank my reader professor Andrea Scott and Dipanjan Basu for their time, effort and suggestions on this thesis.

I would also like to thank members of NWMO for providing the information and recommendation for our research especially Ruiping Guo for the advice on the numerical modelling work, Peter Keech for directing the research and Jeff Binns for the suggestion on this thesis.

Last but not least, I would like to thank my families and friends for the support over the years and my colleagues in Design Optimization under Uncertainties Group for the resourceful discussion on our research and the good time we had together.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

DGR	Deep Geological Repository
EBS	Engineered barrier systems
HCB	Highly compacted bentonite
UFC	Used fuel canister
SBO	Surrogated based optimization
RBF	Radial basis function
DOE	Design of experiment
LHS	Latin Hypercube Sampling
GA	Genetic Algorithm
PDE	Partial Differential Equation
MAE	Mean Absolute Error
MSE	Mean Square Error
RMSE	Root Mean Square Error
Temp.	Temperature
RS	Room Spacing
S	Spacer Width

List of Symbols

$\boldsymbol{x}$	input variables
$\boldsymbol{x_i^L, x_i^U}$	lower and upper bound for the i^{th} input variable
$\boldsymbol{X}$	matrix of the input variables
$\boldsymbol{f}$	functional surrogate
$\boldsymbol{\mathcal{F}}$	RBF or polynomial functions/matrices
$\boldsymbol{\pi}$	independent random permutation for integer in LHS
$\boldsymbol{U}$	uniformly generated coordinates in LHS
$\boldsymbol{\lambda_j}$	RBF parameter for the j^{th} data point
$\boldsymbol{\gamma}$	lower order polynomial parameter for RBF
$\boldsymbol{\phi}$	radial basis functions
$\boldsymbol{\Phi}$	matrix of radial basis function
$\|\ldots\|$	Euclidean distance
$\boldsymbol{v_j}$	j^{th} polynomial basis function
$\boldsymbol{\beta_j}$	polynomial parameters for the jth basis function
$\boldsymbol{\rho}$	material density in kg/m^3
$\boldsymbol{k}$	thermal conductivity in W/(m*°C)
$\boldsymbol{c_P}$	specific heat capacity under constant pressure in J/°kg*°C
$\boldsymbol{T}$	temperature
$\boldsymbol{t}$	time
$\boldsymbol{x, y, z}$	coordinate of the axis
$\boldsymbol{Q}$	heat generation

Chapter 1
Introduction

Deep Geological Repository (DGR) is designed to be a permanent storage facility for used nuclear fuel. The facility will be constructed approximately 500m underground and have multiple engineered barriers to contain the highly radioactive waste for the entire design lifetime. Since the radioactivity of the used fuel decays as a function of the half-life of the radioactive element, the time it takes for the used fuel bundle to reach natural uranium ore radioactivity level will be over one million years [1].

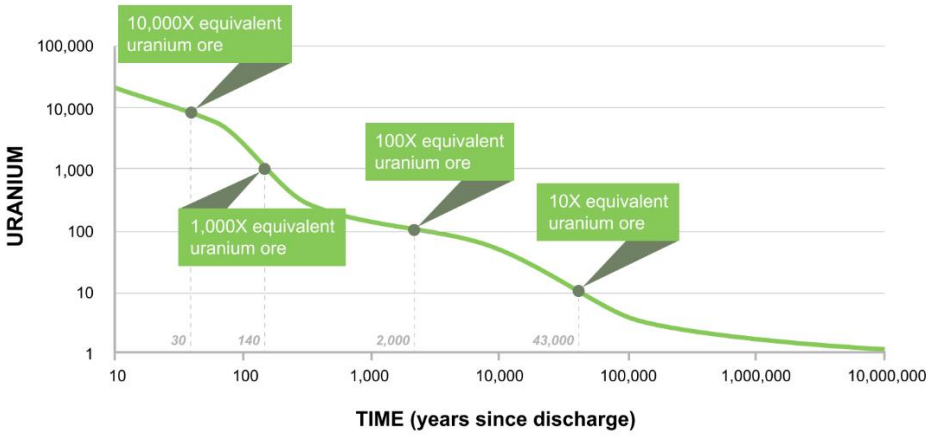


Figure 1.1 Radioactivity of Used Fuel Bundles in Comparison to Natural Uranium [1]

To ensure the safety of people and the environment, the DGR will have to isolate and contain the used fuel for a very long time frame. Figure 1.2 shows a schematic design of the DGR and the engineered barrier systems (EBS). In this design, 48 used CANDU fuel bundles are sealed in a copper-coated steel used fuel container (UFC). This container is then placed inside a highly compacted bentonite (HCB) buffer box and the whole assembly is then emplaced approximately 500m below ground in a stable geological formation. After the placement of the UFC and HCB box assembly, the remaining gaps along the outside of the emplacement room is then filled with a gapfill material (GFM) consisting of granulated bentonite to seal the placement rooms.

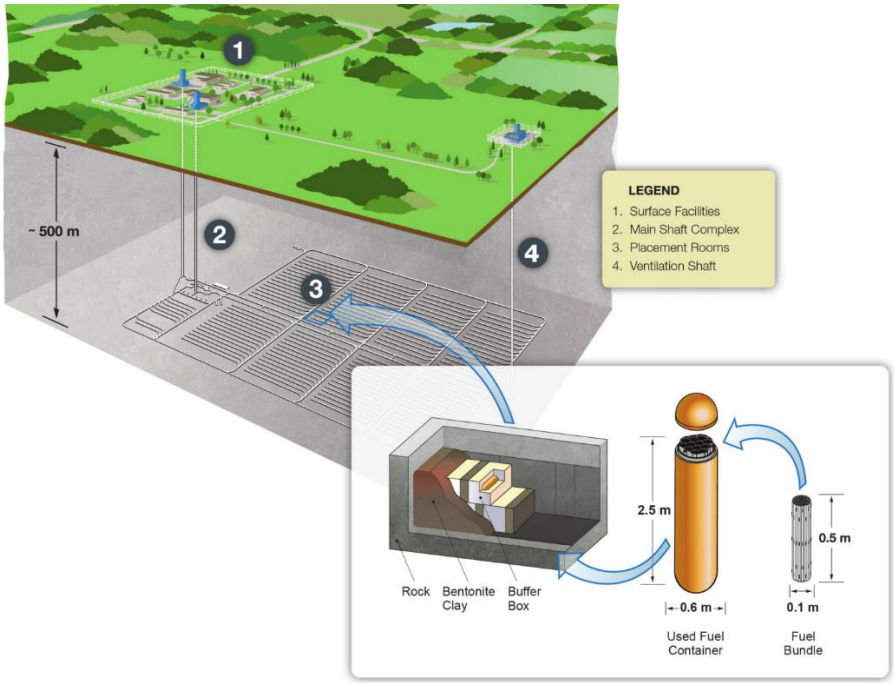


Figure 1.2 Schematic Design of Deep Geological Repository [2]

The reliability of the DGR design depends on many factors, one mode of failure is the corrosion of the 3 mm copper coated layer of the container. The corrosion behavior of the copper coating depends on various chemical and physical parameters such as the composition of the host site groundwater, the transport of oxidants to and corrosion products away from the UFC and the temperature, among others. Of particular interest for the performance of the EBS is the temperature produced by the radioactive decay of the used fuel. The temperature of the DGR will affect the kinetics of chemical reactions and mass transport.

The thermal behavior of the DGR can be simulated using a finite element model. However, to properly simulate the thermal behavior within the placement room requires a detailed near-field and far-field model [3], [4]. Without using cluster computing, a single analysis takes minutes to hours to complete on numerical software such as COMSOL. As a result of that, simulating and studying different configurations of the container placement and material uncertainties will not be computationally feasible if hundreds or thousands of simulations are required to be run.

Surrogate modeling is one way of improving the computational speed where the computation/ simulation time is too long using the original model. Surrogate-based optimization (SBO) has been

considered an effective way for design problems involving time consuming simulations. The optimization using the original model can be replaced by a surrogate that is computationally cheap and analytically tractable. The surrogate model is developed through an iterative process from the data generated by the original model [5]. In the field of aerodynamic design, uncertainty analysis, hydrology or other engineering problems [6]–[9] where high dimensional problems and computationally expensive models appear, surrogate modeling can be a useful tool to reduce the necessary time for solving a given problem.

In this thesis, the surrogate model is applied to the design optimization of spacing parameters of the container and placement room as well as the uncertainty analysis of input variables in the DGR. The chapters in this thesis are as follows. Chapter 2 presents reviews of some of the literature and methodologies that will be used later in this thesis. Chapter 3 presents details of the numerical model constructed for simulation of the temperature behavior of the DGR. Chapter 4 focuses on the surrogate modeling experiment, algorithm and methodologies used in each of the experiment. Chapter 5 presents summary and conclusion of the thesis and suggestions for future work.

Chapter 2
Literature Review

2.1 Introduction

This section presents review of the reference model used for numerical simulation and the methodologies used for developing the surrogate model. Section 2.2 will review the numerical model, while section 2.3 will focus on the surrogate-based method. The numerical model developed by Guo [3], [4] offers the basis for the numerical modeling work of this thesis. As an initial input of the surrogate model, the numerical model results provide the data that the surrogate model builds on. The surrogate-based approach contains few key components for building the algorithm; the review in this chapter will focus on the methodologies used in some of the applications and the theories behind each component for building the surrogate model.

2.2 Thermal Response Model

The DGR has significantly different scales in geometry where the overall repository is in the scale of kilometers while the individual components are in the scale of meters or millimeters. This characteristic makes modeling the behavior of the complete repository while including the details within the placement room difficult. The plan view of the DGR is shown in Figure 2.1, this layout is used in the sixth case study conducted by NWMO. The DGR has a total of 8 panels each consists of 36 placement rooms with 20m spacing between each room. For crystalline rock each room is 308m in length 2.2m in height and 3.2m in width. For the case in sedimentary rock, the configuration is similar, but the length of each repository room will be 342m instead. The far-field numerical model which is used to study the repository behavior in the rock mass of granite is based on the configuration shown in Figure 2.1.

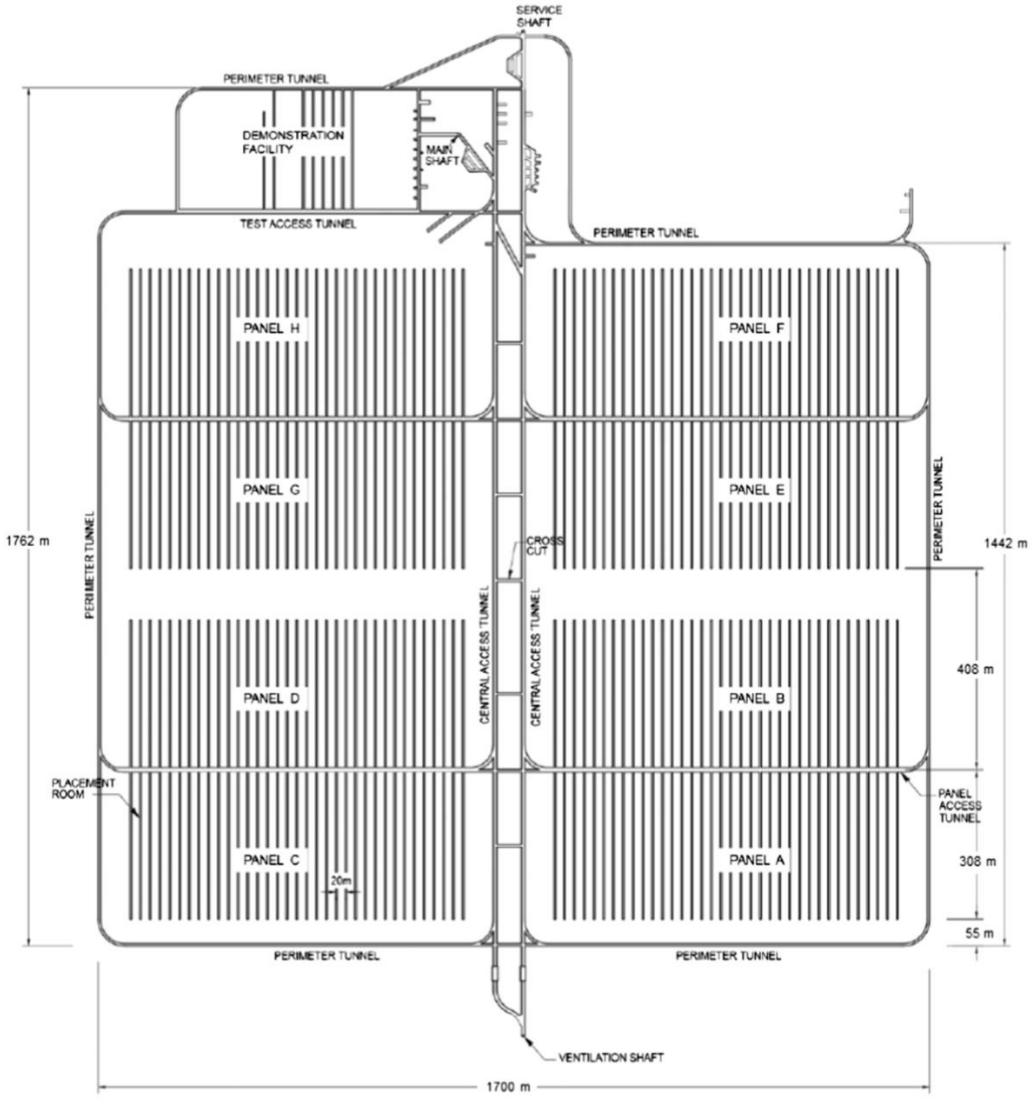


Figure 2.1 Plan View of Crystalline Underground Layout [4]

2.2.1 Geometry Construction

Guo [3], [4] modeled this geometry by building a far-field models and detailed near-field model shown in Figure 2.2 and Figure 2.3. The far-field model simulates a quarter of the repository in a field of granite which is 10,000m in depth and 6000m by 6000m in the horizontal dimension. The repository is represented by two planes that are 2m in height and 281.5m by 720m in horizontal dimension and it is assumed to be granite instead of the material that represents container and bentonite box. Since the far field model is intended to be used to study the DGR's heat generation within a large area of rock mass, the details within the repository does not influence the result much.

To examine the behavior of container, bentonite box and gap fill, a more detailed near-field model is needed. In the near-field model, the symmetry condition on the container and placement room are used. As shown in Figure 2.2, the adiabatic condition is assumed on the four vertical surfaces of the geometry while the isothermal boundary is applied on the top and bottom boundary.

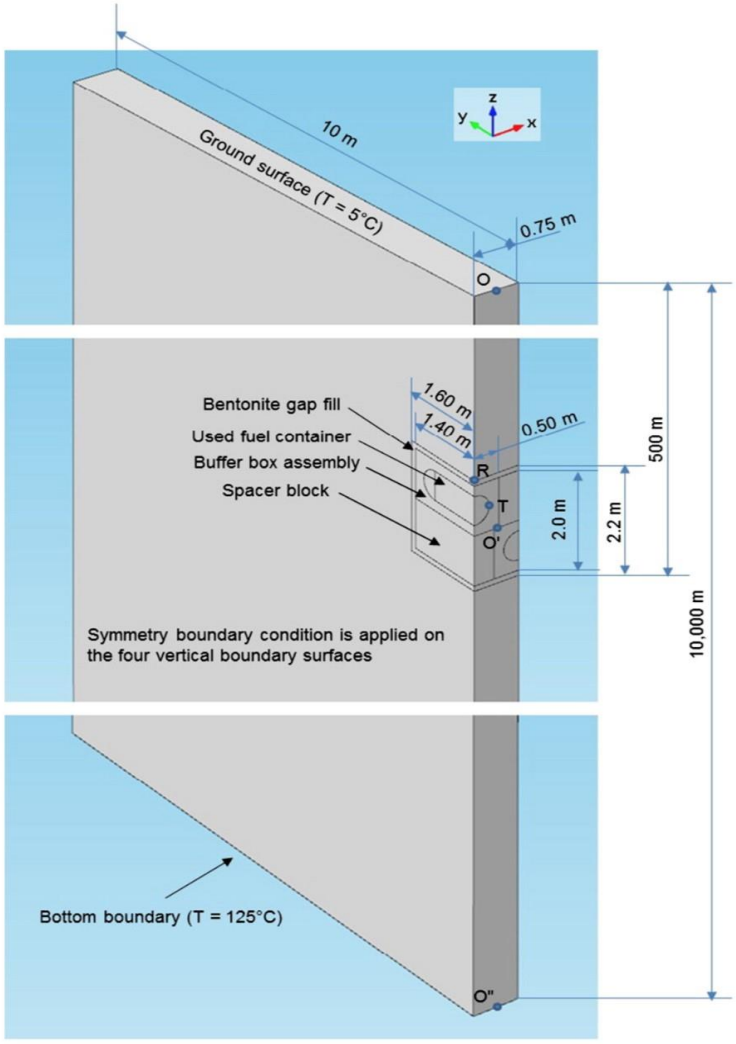


Figure 2.2 Details of Near Field Geometry Used in Simulation [4]

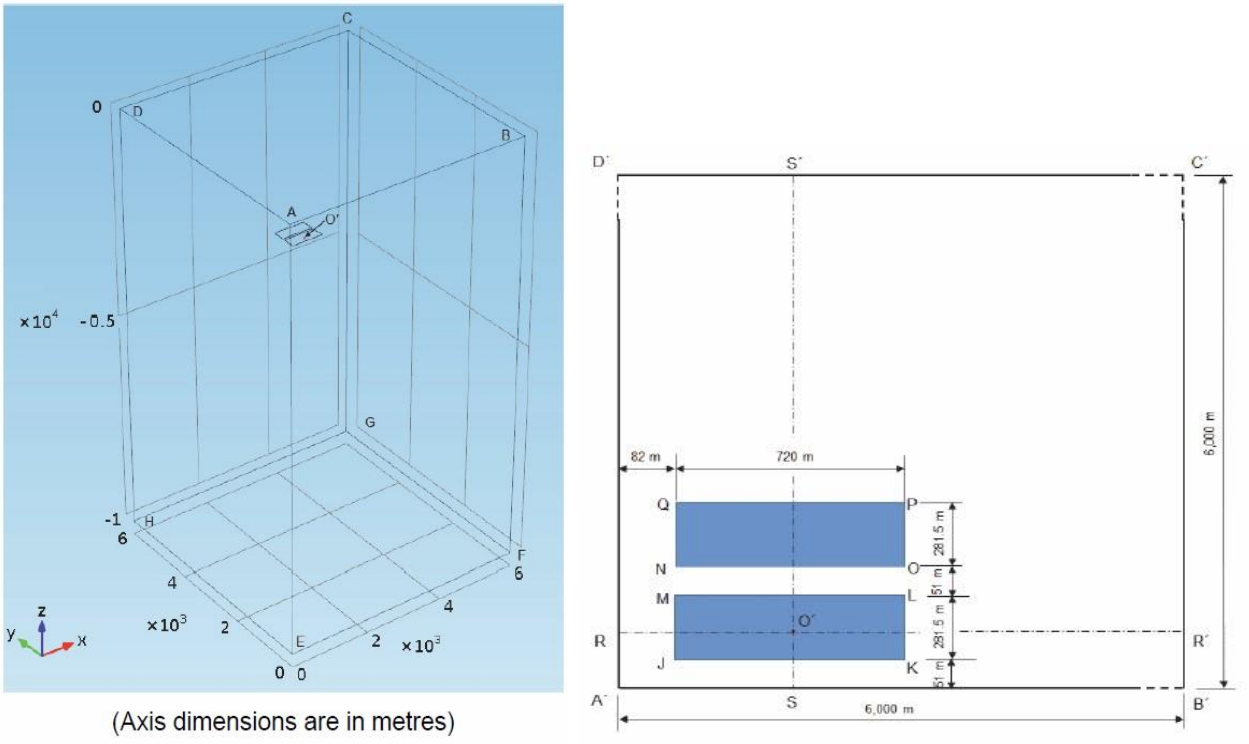


Figure 2.3 Details of Far-field Geometry used in Simulation [3]

2.2.2 Simulation Results

Based on these assumptions in material properties and other boundary condition inputs, the COMSOL solution showed that the maximum temperature within the placement room occurs at year 45, and the maximum temperature is 83.7°C as shown in Figure 2.4. The Figure 2.4 shows the temperature within the repository at the level of container center.

The maximum temperature shown in Figure 2.4 provides reasonable estimate of the placement room temperature at year 45 given the assumptions. However, due to the fact that the adiabatic condition is set on the four vertical surfaces for the symmetry assumptions, the detailed near-field model assumes that the repository is an infinite array of container-bentonite box assembly and the temperature is overestimated as time progresses, which happens after the first 100 years. The overestimation is due to the heat reflection from the adiabatic surface outside of the placement room. Therefore, to remove the effect of the overestimation, the far-field model is needed for studying the actual temperature progression within the placement room.

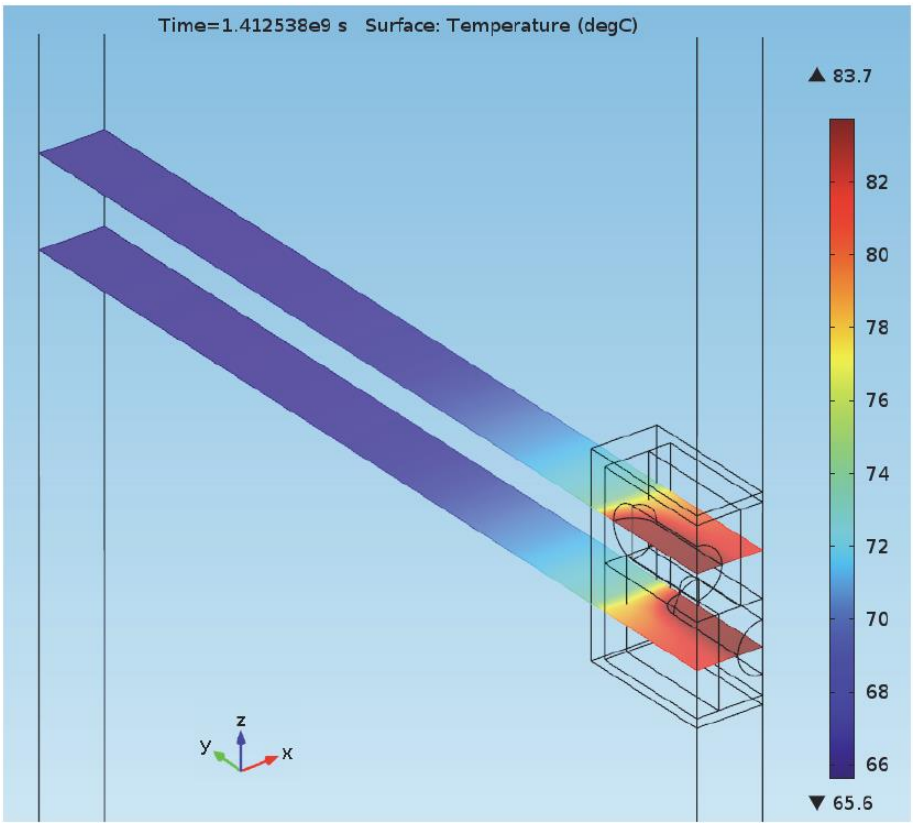


Figure 2.4 Temperature along Horizontal Cross-Section through Container Axis at Year 45 [3]

2.2.3 Overestimation and Correction

The figure shown in Figure 2.5 explains the mechanism of overestimation of the temperature. The near-field model not only takes into account the heat from the repository (Q1) but also the areas beyond the finite repository (Q2), which is caused by the model assumption that the container is in an infinite array. To correct the temperature overestimation shown in Figure 2.5, Guo [3], [4] proposed to calculate the differences between the far-field model and an equivalent simplified near-field model as shown in Figure 2.6 to correct this effect. The difference found from the comparison between the far-field model and equivalent simplified near field model can be used to correct the simulation result from detailed near-field model shown in Figure 2.2.

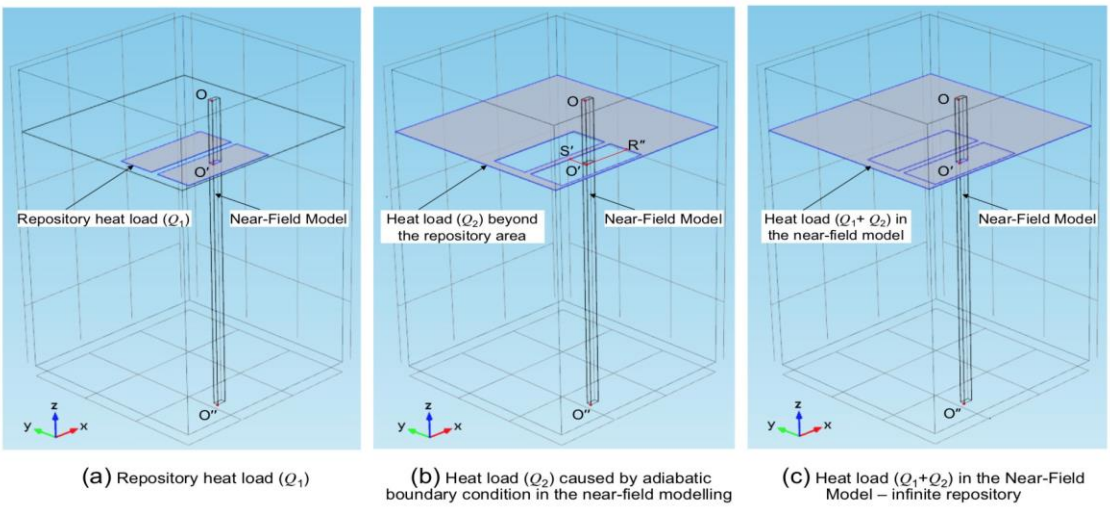


Figure 2.5 Heat Load on the Near-Field Model [4]

The simplified near-field model shown in Figure 2.6 is a section of the far-field model where the repository level is represented by a 2m height placement room with the rest of the dimension equals to the detailed near-field model shown in Figure 2.2. This simplified near-field model also only has one material which is the granite, same as the far-field model to make it a comparable model.

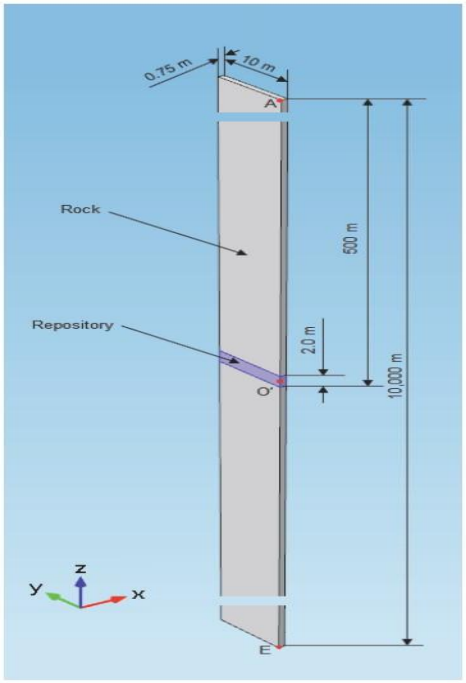


Figure 2.6 Simplified Near-Field Model [3]

Figure 2.7 shows the comparison between the simplified near-field model (Infinite repository) and the far-field model (Finite Repository). The difference line (red dashed line) shows that only after 100 years did the differences become observable and after 1000 years the differences become significant enough to alter the result of two models.

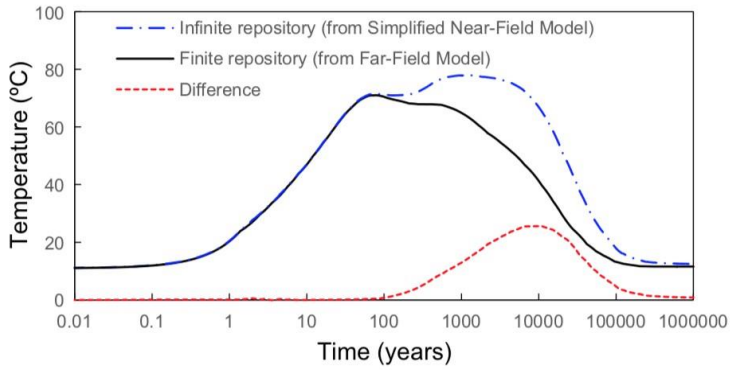


Figure 2.7 Difference between Finite and Infinite Repository Model [4]

The difference line shown in Figure 2.7, can be used to correct the temperature of the detailed near-field model. In addition, the overestimation of temperature also depends on the spacing parameters which will be discussed in Chapter 3. Figure 2.8 shows the example of using the proposed method to modify the temperature for the points in container surface (T) and tunnel center (C). T_m and C_m corresponds to the modified results of these two points. The overestimation shown here is dealing with a specific spacing configuration, when the spacing varies the overestimation will vary. This will be discussed in Chapter 3 for spacing variation.

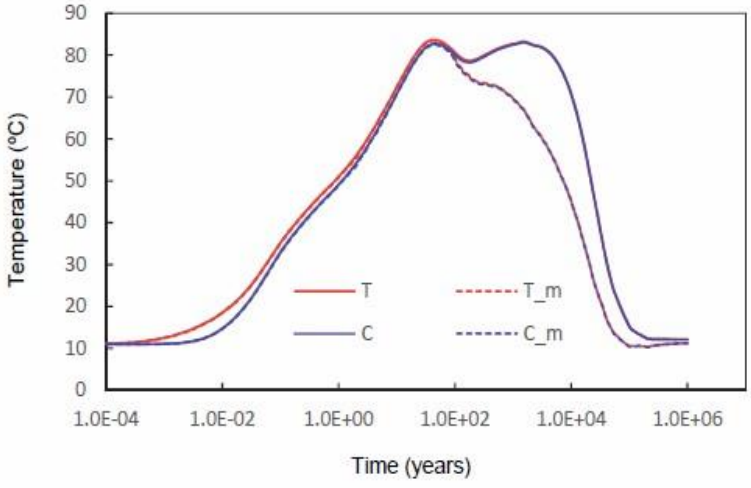


Figure 2.8 Modified Temperature [3]

Guo's model is constructed in the numerical software COMSOL, which will also be the software used for the numerical simulation in this thesis. The geometry and assumptions made in this model will also be applied to the numerical model developed in this thesis with slight alteration to the geometry and material properties.

2.3 Surrogate Modeling

2.3.1 Overview

In engineering problems, the optimization of an objective function usually involves solving multiple computationally heavy simulations which individually will take minutes to hours or even days to be solved. In addition, these simulation models are sometimes black-box like models where analytical version of the model or model derivative is not available. Therefore, surrogate model is used to develop an approximation for the computationally expensive model result for optimization.

The computationally heavy model is often referred to as high-fidelity model while the surrogate model is referred to as low-fidelity model. Surrogate model can also be called metamodel or response surface model [10]. The surrogate model is often analytically tractable with computationally cheap functions such as polynomial, radial basis function, neural networks, etc. The algorithm used in surrogate-based model efficiently finds the optimal solution without the need to exhaustively evaluate all the possible points in the sampling space [5], [6]. The benefit of surrogate-based model is significant in the case of high dimensional problem where the expensive model evaluation across the entire sampling space is infeasible.

The surrogate-based algorithm can be formulated as the following process [5]:

$$\boldsymbol{x^{(i+1)} = min_x f^{(i)}(x)} \tag{2.1}$$

The algorithm first generates a series of points where the result will first be evaluated and then fit with surrogate function f, the optimal x for this surrogate model will be the point to be evaluated next using the high-fidelity model, if the error between the result predicted by the surrogate and expensive function evaluation meets the convergence criteria the algorithm stops, and the optimal result is found. Otherwise, a new surrogate is fit, and the process repeats until the convergence or stopping criteria is met. The general process of the surrogate modeling is shown in Figure 2.9.

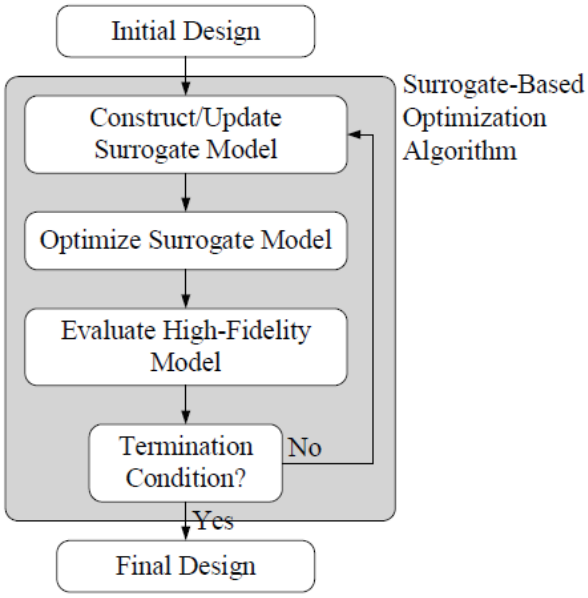


Figure 2.9 General Process of Surrogate Based Method [5]

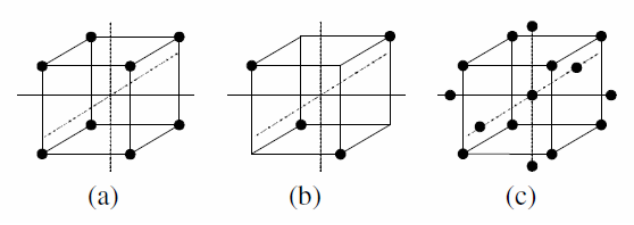
The surrogate model can be either a physical model or a functional surrogate model [5]. In this thesis, the focus will be using a functional based surrogate for the COMSOL simulation result. The following sections provides a brief overview of the methodologies used in developing the surrogate model in this thesis.

2.3.2 Experiment Design

When the surrogate model is first being constructed, a Design of Experiment (DOE) method will be applied to allocate the sampling points uniformly within the sampling space. The DOE is intended to maximize the information acquired in the sampling space that can be used to construct the functional surrogate model. The ideal samples are as spread out as possible within the sampling space in order to capture the global trend of the high-fidelity model.

The classical DOE includes full factorial design, fractional factorial design and central composite design, Box-Behnken Design, etc. The classical DOE mostly deals with a problem that can be physically measured with an error term ε, therefore, sampling the point near or at the boundary of the sampling domain and introduce duplicate sampling may improve the approximation of the true trend when the random error exists in the high-fidelity model [11]. The drawback of the classical DOE is that the number of sampling points increases exponentially with the dimension of the input parameter[11].

For example, sampling points of central composite method and full factorial method scales with 2^n as the dimension n increases. This method becomes infeasible if the experiment is for high dimensional problems.



a. full factorial, b. fractional factorial, c. central composite

Figure 2.10 Example of Classical DOE [5]

The modern DOE methods deal with the experiment that does not produce the error term ε such as computer simulation. The most common modern DOE method is Latin Hypercube Sampling (LHS)[12]. The LHS subdivides each of the range into p bins, for n dimensional variables it yields a total of p^n bins in the sampling space. The sampling point is randomly selected so that each sampling point is in one of the bins and for all the one-dimensional projection of the p samples and bins there is exactly one sample. With the same number of samples, the LHS will have less error in estimating the means than the other modern DOE method such as Monte Carlo sampling[11]. The LHS also has the benefit of configuring any number of sampling points without strictly following the specific multiples or power of n. The LHS algorithm can be described as [11]:

$$x_j^{(i)} = \frac{\pi_j^{(i)} + U_j^{(i)}}{k} \qquad \textbf{(2.2)}$$

$$for\ 1 \leq j \leq\ n\ and\ 1 \leq i \leq k$$

where k is the number of samples, i denotes the sample number and j denotes the dimensional index, U is the uniform random variable on [0, 1] and π is the independent random permutation of integer of 0 to k-1. For example, the π for a two-dimensional sampling creates a integer combination of (0,0) which indicates the sample point is in the bin at lower left corner, and U value of (0.1, 0.6) denotes the location of the sample point within this particular bin, Figure 2.11 shows an example of sampling 2 dimensional space using LHS. If U in this equation is replaced with the value of 0.5 instead of uniform sampling, the sample point is always in the middle of the bin. This slight modification is called lattice sampling.

Other modern DOE method may include orthogonal array sampling, quasi Monte Carlo sampling and Hammersley sampling [5].

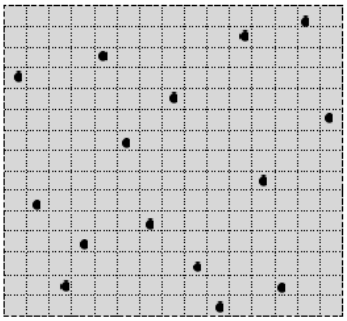


Figure 2.11 Example of LHS sampling for 2D Space [5]

2.3.3 Surrogate Model Functions

The surrogate model function can use a large variety of function that is analytically tractable and computationally cheap. The following are a few surrogate functions that are used in the application at Chapter 4. The surrogate modeling can not only use a single function but also an ensemble of different type of functions to improve the model accuracy as mentioned and applied in some of the literature[13], [14]. Other forms of functions that are not mentioned below include Neural Network, Kriging, support vector regression, etc.

Radial Basis Function (RBF)

The radial basis function [15] utilizes the Euclidean distance between a point and all the data points of the known measurements to interpolate the value of a given point. The radial basis function has the following general form:

$$\boldsymbol{\mathcal{F}}(\boldsymbol{x}) = \sum_{\boldsymbol{j=1}}^{\boldsymbol{N}} \boldsymbol{\lambda_j \phi}(\|\boldsymbol{x} - \boldsymbol{x_j}\|) \qquad \textbf{(2.3)}$$

where, ϕ is the radial symmetric function, λ_j is the length j vector of model parameter, x_j is the j^{th} known center of the radial basis function and j denotes number of the known data point. The radial symmetric function can take on one of the forms shown in Figure 2.12 where r is the Euclidean norm of the data points.

RBF Name	Mathematical Expression
Multiquadric	$\phi(r) = (1+(\varepsilon r)^2)^{1/2}$
Inverse multiquadric	$\phi(r) = (1+(\varepsilon r)^2)^{-1/2}$
Gaussian	$\phi(r) = e^{-(\varepsilon r)^2}$
Thin plate spline	$\phi(r) = r^2 \log(r)$
Cubic	$\phi(r) = r^3$
Wendland's	$\phi(r) = (1-\varepsilon r)_+^4 (4\varepsilon r+1)$

Figure 2.12 Common RBF Forms [15]

The model parameter can be solved by inverting the radial symmetric function as:

$$\boldsymbol{\lambda} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\boldsymbol{\mathcal{F}}(\boldsymbol{x}) \tag{2.4}$$

where ϕ is a j by j matrix of the following form:

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\phi}(\|\boldsymbol{x_1} - \boldsymbol{x_1}\|) & \cdots & \boldsymbol{\phi}(\|\boldsymbol{x_1} - \boldsymbol{x_N}\|) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\phi}(\|\boldsymbol{x_N} - \boldsymbol{x_1}\|) & \cdots & \boldsymbol{\phi}(\|\boldsymbol{x_N} - \boldsymbol{x_N}\|) \end{bmatrix} \tag{2.5}$$

Sometimes, in the case such as cubic RBF, the location of the data point will result in singular matrix so an alternative way of solving for the model parameter is proposed by Misha et.al [15] to introduce a low order polynomial term. For example, a 2D RBF is given as:

$$\boldsymbol{\mathcal{F}}(\boldsymbol{x}) = \sum_{\boldsymbol{j=1}}^{\boldsymbol{N}} \boldsymbol{\lambda_j}\boldsymbol{\phi}\left(\left\|\boldsymbol{x} - \boldsymbol{x_j}\right\|\right) + \boldsymbol{C_{N+1}} + \boldsymbol{C_{N+2}}\boldsymbol{x} + \boldsymbol{C_{N+3}}\boldsymbol{y} \tag{2.6}$$

where $x = (x, y) \in \mathbb{R}^2$, and another three conditions is:

$$\sum_{k=1}^{N} C_k = 0, \qquad \sum_{k=1}^{N} C_k x_k = 0, \qquad \sum_{k=1}^{N} C_k y_k = 0$$

Therefore, the resulting system of equation has the form as:

$$\begin{bmatrix} \boldsymbol{\Phi} & \boldsymbol{p} \\ \boldsymbol{p^T} & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mathcal{F}} \\ \boldsymbol{0} \end{bmatrix} \tag{2.7}$$

where p = [1 x y] and 0 is vector of length 3 and O a zero matrix of size 3 x 3, and γ is the additional model parameter corresponding to the lower order polynomial.

Using the above formulation can result in a non-singular system to obtain the value of model parameters. The RBFs used later on in the experiment for fitting are cubic, Thin plate spline and linear RBF.

Polynomial Regression

The polynomial regression assumes the relationship of linear combination of different order terms and model parameters estimated by the data points available. The polynomial has the following form as:

$$\mathcal{F}(x^i) = \sum_{j=1}^{N} \beta_j v_j(x^i) \tag{2.8}$$

Or in matrix form:

$$\mathcal{F} = X\beta \tag{2.9}$$

where v_j is the polynomial basis function for an input x^i, and β is the model parameters.

The model parameter β can be solved as linear least square problem using pseudo inverse as:

$$\beta = \left(X^T X\right)^{-1} X^T \mathcal{F} \tag{2.10}$$

The most commonly used first and second order polynomials can be written as:

$$\mathcal{F}(x) = \beta_0 + \sum_{j=1}^{n} \beta_j x_j \tag{2.11}$$

$$\mathcal{F}(x) = \beta_0 + \sum_{j=1}^{n} \beta_j x_j + \sum_{i=1}^{n} \sum_{j \le i}^{n} \beta_{ij} x_i x_j \tag{2.12}$$

The polynomial can also be in third order with an additional cubic term. The forms used later on in the experiment include Linear, Quadratic, and Cubic Polynomials.

2.3.4 Sampling Point and Optimization

After the DOE points have been evaluated and the initial surrogate model is fitted with data, the surrogate needs to evaluate new sample point using the high-fidelity model to determine the accuracy of the surrogate or to validate the prediction made by surrogate. In this thesis, the sample point selection process follows some of the ideas of the algorithm developed in MATSuMoTo toolbox [13], [16], [17] for discrete optimization problem. The most popular way of sampling for design optimization problem is to use the minimum point of objective function as the next evaluation point for model evaluation.

The minimum point is found by solving the constraint optimization problem using Genetic Algorithm (GA) by MATLAB function *ga()*. The optimization problem can be formulated as:

$$min_x f(x) \tag{2.13}$$

$$s.t. \quad C_j(x) \le 0, \qquad \forall j = 1, \dots, m \tag{2.14}$$

$$x_i^L \leq x \leq x_i^U, \forall i = 1, \ldots, k \quad \textbf{(2.15)}$$

Where $x_i^L \; and \; x_i^U$ are lower and upper bound of the variable x, and f is the objective function and C_j is the constraint function.

The minimum point sampling will select the best point found on the current surrogate model as the next sampling point to do the high-fidelity model evaluation. If the result from the high-fidelity model result does not converge with the result predicted by the surrogate model, the surrogate model is updated with the newly obtained value from the high-fidelity model. As a result, the surrogate model will become increasingly accurate near the optimal point of the objective functions until the surrogate model result converges with the high-fidelity model result.

Since the experiment problem for design optimization in this thesis consider the input variables as discrete, the genetic algorithm is used for having the option of applying integrality constraint to the input variables.

The general GA optimization has a few key components, chromosome encoding, fitness function, selection, recombination and evolution [18]. The fitness function is generally the objective function the algorithm is trying to optimize, in each generation, a randomly generated combination of input **C**(chromosome) is tested using the fitness function and the best input will have a higher probability of being selected(selection) and combined(recombine) with other selected input to form the next generation of inputs. The process repeats until the optimal point is found or the stopping criteria are met. The algorithm can be summarized in the following steps:

1. Randomly generate the source population P of input/chromosome(**C**)
2. Calculate the fitness of function F(**C**) for each chromosome in the population
3. Create an empty successor population and then repeat the following steps until P chromosome is created
    a. Using fitness function to select the two chromosome **C**1 and **C**2 from the original population
    b. Apply one-point crossover of **C**1 and **C**2 to generate a child chromosome **C**
    c. Apply uniform mutation to **C** to produce **C**' based on mutation probability
    d. Add **C**' to the successor population

4. Repeat the process from 2 using the successor population until the stopping criterion is met

The GA is a heuristic search and optimization technique that can be applied for a wide range of application and it is ideal for solving discrete optimization problem seen in this thesis. For the case of multi-objective optimization, the GA solves for the pareto front which is also known as the non-dominated solution set. The definition of a dominated solution has to meet the following two definition[19]:

1. For solution $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$, if $\boldsymbol{x}^1$ is no worse than $\boldsymbol{x}^2$ in all the objective function
2. The solution $\boldsymbol{x}^1$ is strictly better than $\boldsymbol{x}^2$ in at least one objective

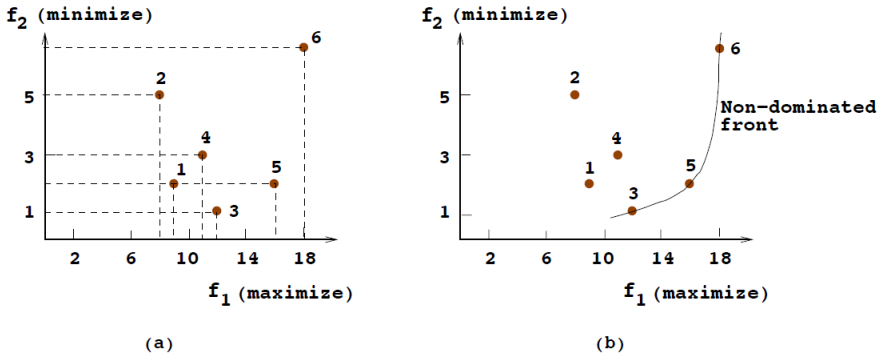


Figure 2.13 Comparison of Dominated and Non-Dominated Solutions [19]

In Figure 2.13, (a) shows the solutions that are dominated to each other and (b) shows the set of solutions that are non-dominated to each other. For example, point 1 dominates point 4 and point 5 but points 3, 5 and 6 do not dominate each other. All points that are not dominated by any other member of the set are called non-dominated set (Pareto Set). The boundary defined by the set of all point mapped from pareto set is called the pareto front (non-dominated front). The *gamultiobj()* function in MATLAB is designed to solve for the pareto set.

For design optimization using surrogate modeling, minimum point sampling is often used, however, other ways of sampling are also common for optimization problem including approximation model management optimization, space mapping and manifold mapping [5]. The MATSuMo toolbox includes a local sampling approach which is by perturbing the local optimal point and a global sampling approach which is the uniformly sampling within the entire variable domain [13].

2.4 Summary

The finite element model shown in section 2.2 indicates that the maximum temperature will occur within the first 100 years and at year 45 to be specific. The maximum temperature based on the current assumptions and the spacing configuration is 83.7 °C. Even though the temperature simulated by finite element model is overestimated after the first 100 years, a correction can be made using the results from the two different models and the result within the first 100 years is considered accurate.

The surrogate modeling approach utilizes fitting ability of interpolating function to reduce the necessary computationally heavy simulations. Combining with the modern design of experiment method, the surrogate can solve a given optimization problem with much less simulation and evaluation. The genetic algorithm provides the tool for solving discrete value optimization problem which is defined by the spacing optimization problem later in section 4.2.

The following Chapter 2 will develop and analyze the thermal numerical model for the deep geological repository similar to the one shown in section 2.2 which will be used as the basis for the surrogate model in Chapter 4.

Chapter 3
Numerical Simulation

3.1 Introduction

The numerical simulation provides the high-fidelity model for surrogate modeling as well as the base case scenario of the temperature behavior of current design. In this thesis, the high-fidelity model is being constructed in COMSOL, a numerical Partial Differential Equations (PDE) Solver that solves the physically based PDE using finite element methods. The numerical simulation provides the base case scenario for determining the maximum temperature within the repository given the current design. This chapter is organized as follows. The section 3.2 describes the geometry construction used in the simulation and its comparison with Guo's model [3]. Section 3.3 lists the assumptions used and the base case input parameters for material properties and boundary conditions. Section 3.4 presents the results from the simulation including the base case model result, computational time and model validation. The results also include the discussion of the effect of spacing parameters and sensitivity analysis of the material properties and boundary condition assumptions.

The model used for the simulation has multiple components to reflect the reasonable details of the DGR design. The followings are components used for the simulation:

- **Container:** The container in this simulation is assumed to be a solid cylinder made of steel. A heat generation is assigned to the cylinder corresponding to the heat generation of the used fuel within the container. The container is simplified and does not model the 3mm copper coating.
- **Highly Compacted Bentonite (HCB) Box:** The HCB box is the crucial component of the engineering barrier that creates the anaerobic condition to prevent the copper layer being oxidized and fluid flow from reaching the container surface. The characteristics of the HCB box depends on the bentonite composition, temperature and saturations etc [20]. The HCB provides swelling pressure when it becomes saturated, and the swelling pressure has been studied in many situations[20]–[24].
- **Spacer Box:** The spacer block is also made from HCB, it is a smaller HCB box that used to create separation for the bentonite boxes.

- **Rock Mass:** The host rock surrounding the DGR. Depends on the final decision on the site, it is either sedimentary or crystalline rock. The rock is only considered to be homogeneous throughout the modeling area.
- **Gap Fill:** The spacing between the bentonite box and the placement room will be filled with gap fill which is consisting of granulated bentonite.
- **Asphalt:** The asphalt is the flooring material in the room, the exact properties are not decided yet.

The purpose of the numerical simulation is to construct a base case result for maximum temperature occurring within the placement room for current assumptions and inputs.

3.2 Geometry Construction

Based on Guo's near field model [4] shown in Figure 2.2 and the updated dimension and design given in data clearance form [25] for the seventh case study, a near field model is created to study the temperature behavior of the DGR. The geometry is based on the simplified configuration shown in Figure 3.1 and **Error! Reference source not found.** Figure 3.2 which is the seventh case study design used by NWMO.

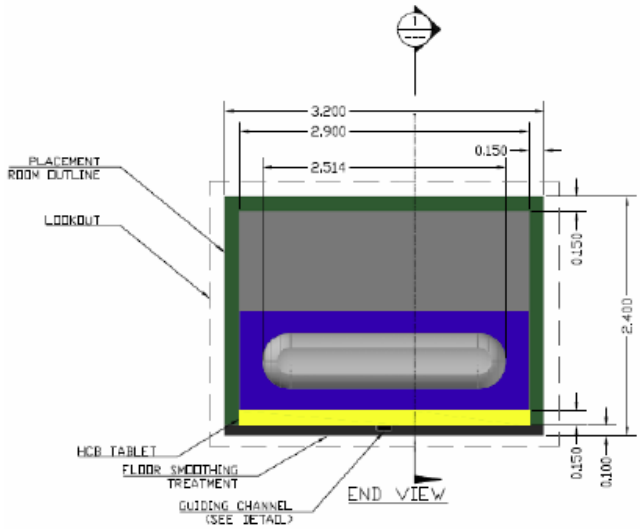


Figure 3.1 Cross-Sectional View of Placement Room (End View)

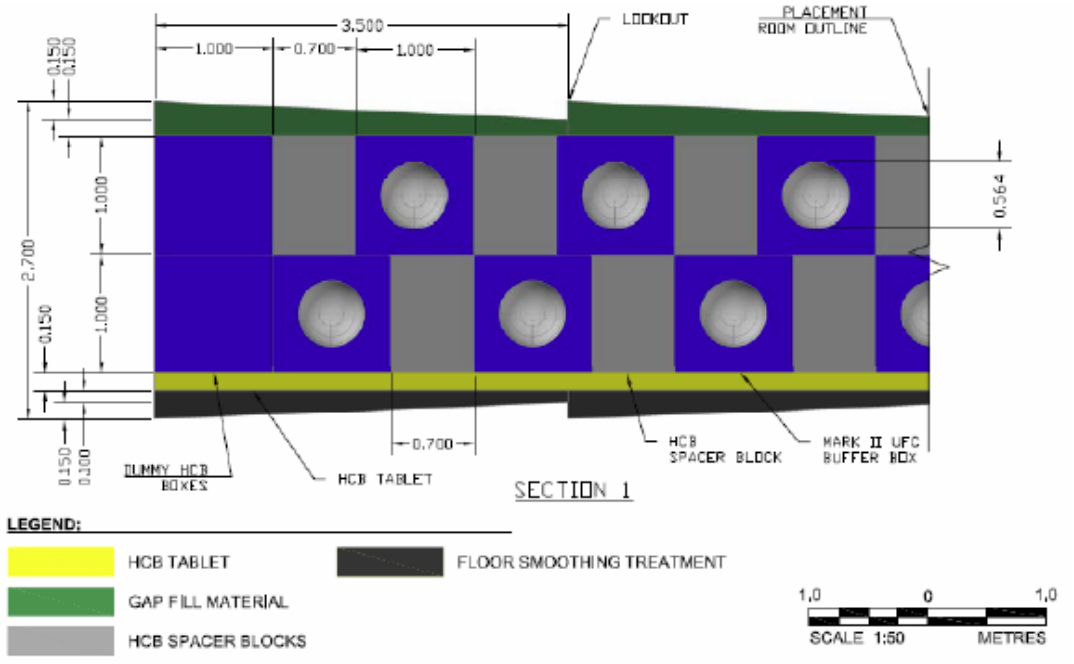


Figure 3.2 Cross-Sectional View of Placement Room (Side View)

Some modification of the above configuration is made to simplify the geometry construction for the simulations. The zig-zag shape of the lookout section is not being modeled and the first two dummy HCB boxes are not considered in this model. Similar to the simulation run by Guo, these simplification in geometry will reduce the solution time while the result will not be changed significantly [3]. The geometry shown in Figure 3.3 considered the symmetry in the center of the room as well as on the centerline of the container. The placement of the spacer block is assumed to be in the center of the bentonite box instead of the offset placement shown in Figure 3.2. The room spacing is 25m center to center compare to 20m in Guo's model. The spacer box width is 0.7m instead of 0.5m in Guo's model. A comparison of the geometry is shown in Table 3-1.

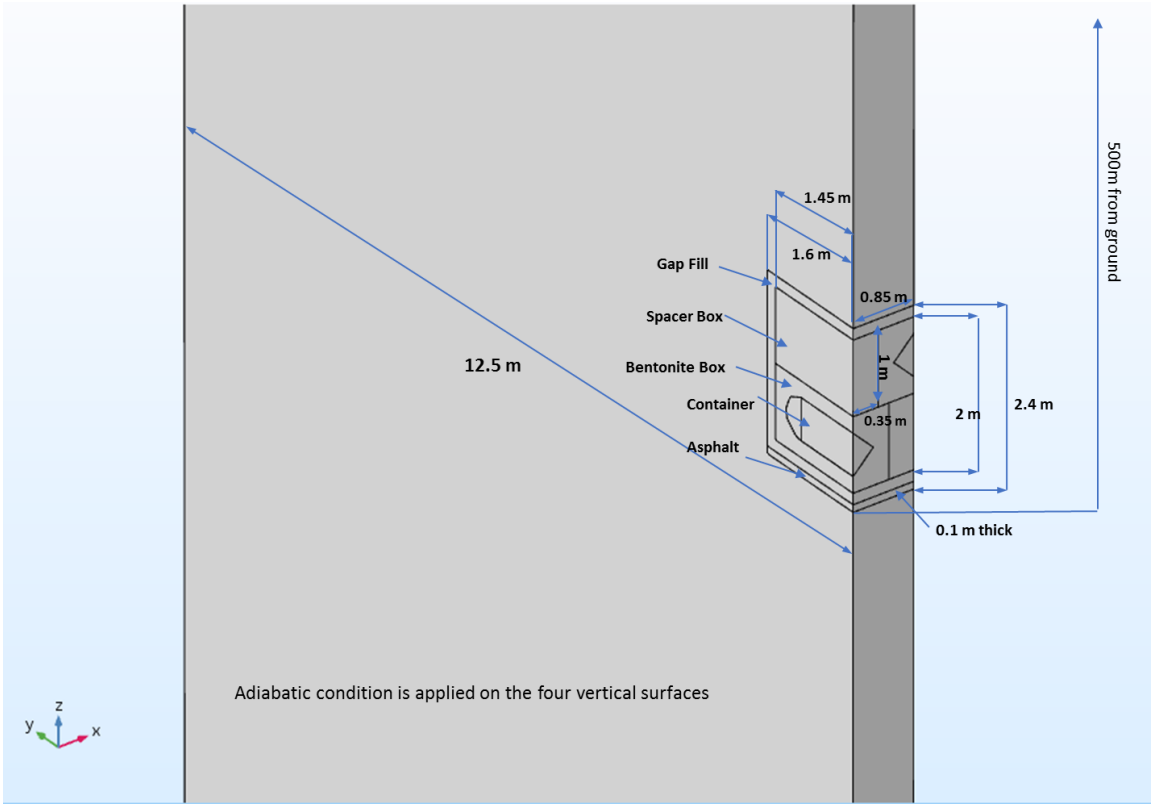


Figure 3.3 Geometry for Simulation

The geometry shown in Figure 3.3 reflects a section within the repository. Similar to the modeling assumption mentioned in section 2.2, it represents an infinite array of repository which tends to overestimate the temperature in the period after first hundred years of the simulation. The difference in geometry is due to the small design change in the seventh case study comparing to the sixth case study design used by Guo [3], [4]. The addition in flooring material asphalt, a dimension change in the gap fill and spacer block width contribute to the difference in geometry. The modeling result from this geometry will be compared with the Guo's result before the correction shown in Figure 2.8. The model validation with respect to Guo's model will be presented in section 3.4.1.2.

Table 3-1 Geometry Differences Between Current Model and Guo's Model

	Guo's Model	Current Model
Room Spacing (Room Center to Center)	20 m	25m
Spacer Block Width	0.5m	0.7m

Dimension of the placement room in model (D x H x W)	1.6 x 2.2 x 0.75	1.6 x 2.4 x 0.85
Model Depth	10,000 m	2,000 m
Rock Mass	Crystalline	Crystalline

The geometry used in this model is to produce a comparable model with Guo so that it can provide meaningful evaluation on the placement room temperature for current design and later on in Chapter 4 for surrogate model development. Since the maximum temperature always happens within the first 100 years, the overestimation of temperature occurs after this period will not affect the estimation in maximum temperature. Section 3.4 will discuss the simulation result validation and interpretation of maximum temperature of this model.

3.3 Model Input & Assumptions

The input data for the thermal modeling consists of assumptions on material properties and the assumptions on initial and boundary conditions. The material properties data mainly consists of density, heat capacity and heat conductivity as shown in equation 3.1.

$$\rho c_P \frac{\partial T}{\partial t} - \frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) - \frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right) - \frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right) = Q \qquad (3.1)$$

Where Q stands for the heat generation from the used fuel for each unit of container, k is the thermal conductivity of the material, C_p is the heat capacity under constant pressure and $\frac{\partial T}{\partial x}$ is the thermal gradient of the medium at a specific point. The initial condition of the temperature depends on the assumptions of the surface temperature as well as the geothermal gradient.

In this simulation, the material properties are not time dependent, and the material property is assumed to be homogeneous within each material. The following table summarizes the material properties used in the simulation.

Table 3-2 Material Properties Assumptions for Current Model

	k (W/(m*°C))	C_p (J/°kg*°C)	ρ (kg/m^3)
Rock (Crystalline)	3	845	2700
Container	60.5	434	7800

Bentonite Box	1	1240	1955
Spacer	2	2276	1060
Gap Fill	0.5	870	1439
Asphalt	0.75	920	1960

The material input used in this case refers to the value used in Guo's model in order to benchmark the result for validation purpose.

For the boundary conditions, the following assumptions are used [3]:

- Symmetry (Adiabatic) condition is applied to the four vertical surfaces
- Constant Surface temperature (Isothermal) of 5 °C at ground level
- Geothermal Gradient of 0.012 °C/m

The surface boundary condition as well as geothermal gradient determines the initial temperature at a specific depth within the model. For example, at the depth of 500m, which is the depth of the repository, the initial temperature will be:

$$5\,°C\ +\ 0.012°C/m * 500m = 11°C$$

And 29 °C at the bottom surface of 2000m depth.

The heat generation Q is the most important factor that determines the temperature within the repository. Since the decay of the heat generation follows exponential function, the simulation time stepping needs to follow logarithmic order to produce accurate results. Therefore, both the simulation time stepping as well as the heat generation time steps are all in 10^n seconds. The heat generation due to the decay of used fuel bundles are shown in Figure 3.4. The year 0 corresponds to a used fuel life of 30 years old, which means that the used fuel has been removed from the reactor for cooling for exactly 30 years. To make the figure more easily understand, the x-axis of the figure converts seconds to years.

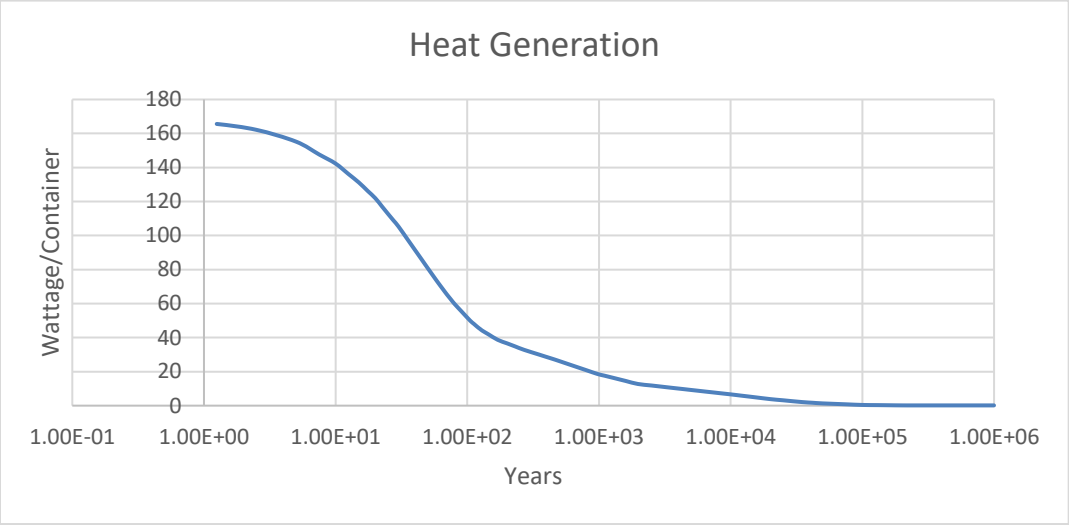


Figure 3.4 Heat Generation from Used Fuel Bundle in a Single Container

The above figure shows the heat generation for a single container, however, since we only consider half container due to symmetry shown in Figure 3.3, the value used in the model will be half of the value shown above and time step will be in seconds to match the time stepping of the simulation. The exact input for the heat generation can be found in **Appendix A**

The meshing of the model uses tetrahedral elements with concentrated meshes inside the placement room where high thermal gradient is likely to occur. As shown in Figure 3.5, there are in total 179,208 tetrahedral elements in the model. The exact number of elements may vary in different configuration of the geometry however, the magnitude of the elements will be similar.

The last assumption that could be overlooked is that all the used fuel container will be present inside the room at the same time once the construction of the whole repository is completed, and the simulation starts at that time. In reality, the used fuel container and the construction will likely be progressive, and the container will be placed inside the repository when the construction is in progress. This assumption will ensure that the simulation will produce maximum temperature possible, in reality, the actual temperature may be lowered than the model result if the container is placed progressively.

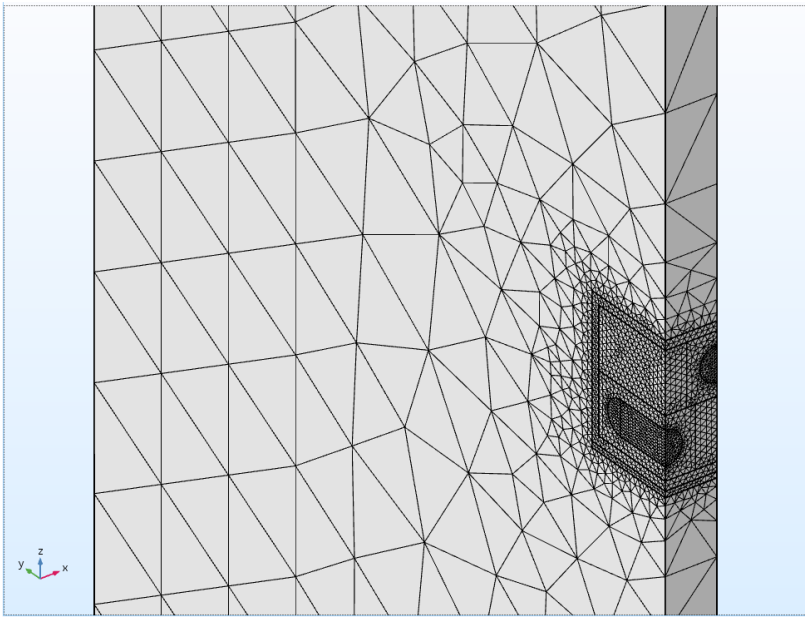


Figure 3.5 Finite Element Meshing of Geometry Near Placement Room

The simulation time for the base model is 10^0 to $10^{13.5}$ seconds which corresponds to 1 seconds to 1 million years. The time stepping is $10^{0.01}$ seconds. The relative tolerance for the simulation is $1e^{-8}$. The configuration of the solver and its effect on computational speed will be discussed in section 3.4.1 and Table 3-4.

3.4 Simulation Results

3.4.1 Base Model Result and Validation

3.4.1.1 Base Model Simulation Results for the Maximum Temperature

The temperature behavior of the DGR will be determined by two simulation results. The maximum temperature inside the placement room for each time step and the maximum temperature for a given placement room configuration over the entire simulation period.

The Figure 3.4 is the maximum temperature for each time steps within the placement room. Based on the geometry shown in Figure 3.3, the maximum temperature occurs within the first hundred years. In addition, in this configuration, the second peak of the maximum temperature is lower than the first peak.

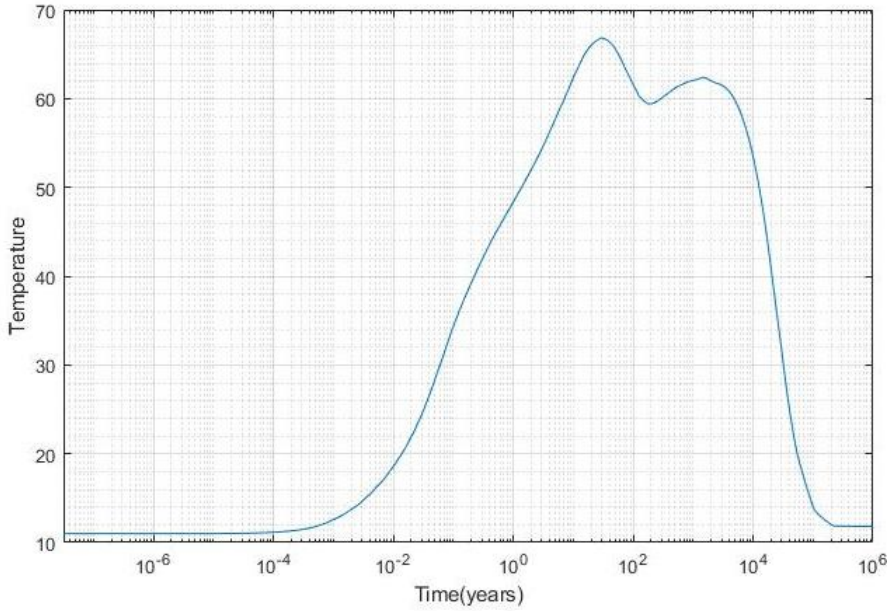


Figure 3.6 Maximum Temperature within Placement Room Over Time

The maximum temperature occurs at around year 47 (1.479^9 seconds). The following Figure 3.7 shows the temperature of the cross-section at the center of each container at year 47. As shown in Figure 3.7, the temperature is the highest near the center of the container and reduces as the distance from the container increases. In comparison with Guo's model, the maximum temperature occurs in almost identical time with maximum temperature around 66 °C. In Guo's model, the maximum temperature is 83.7 °C. The lowered temperature is mainly due to the differences in spacing between room and spacer block width. Since in the current model, the spacing between room assumes a spacing of room center to center of 25m and spacer block width of 0.7m, which differs from Guo's model of 20m and 0.5m.

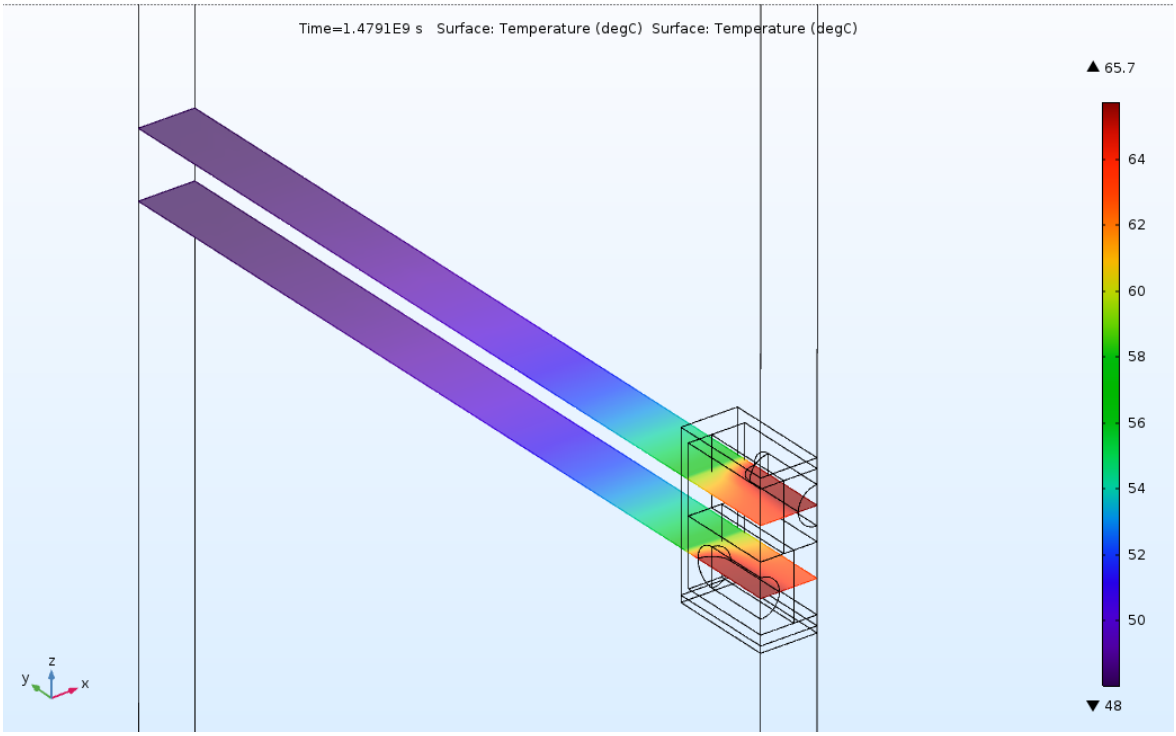


Figure 3.7 Maximum Temperature Horizontal Cross-Section at Container Center

3.4.1.2 Model Validation

In order to verify the results obtained in this model, an equivalent model is constructed with comparable dimensions to Guo's model. Since the temperature is sensitive to spacing, the equivalent model will have the same room spacing of 20m and spacer block dimension of 0.5m. The Table 3-3 compares some of the key inputs between Guo's model, Current Model and Equivalent Model to compare with Guo's Model.

Table 3-3 Key Input Comparison Between Three Model

	Guo's Model	**Equivalent Current Model**	**Current Model**
Room Spacing	20 m	20m	25m
Spacer Block Dimension	0.5m	0.5m	0.7m

Dimension of the placement room (D x H x W)	1.6 x 2.2 x 0.75	1.6 x 2.4 x 0.75	1.6 x 2.4 x 0.85
Simulation Period	$1\text{-}1^{13.5}$ Seconds with 0.01 step size	$1\text{-}1^{13.5}$ Seconds with 0.01 step size	$1\text{-}1^{13.5}$ Seconds with 0.01 step size
Relative Tolerance	1e-8	1e-8	1e-8

The simulation results for the three models are shown below in Figure 3.8. The maximum temperature within the placement room between Guo's model and the equivalent current model is almost identical with slight differences in temperature. Therefore, it can conclude that the difference in temperature profile is due to the differences in spacing configuration. The current model is therefore comparable to the model constructed by Guo. Another observation can be drawn from the results is that increasing room spacing will change the height of the second peak in the temperature profile. Since the second peak is mainly caused by the adiabatic boundary condition and the reflection of heat from the far side boundary, the closer the boundary is the higher the temperature will be for the second peak. Therefore, the temperature overestimation will depend on the room spacing used in different model configurations.

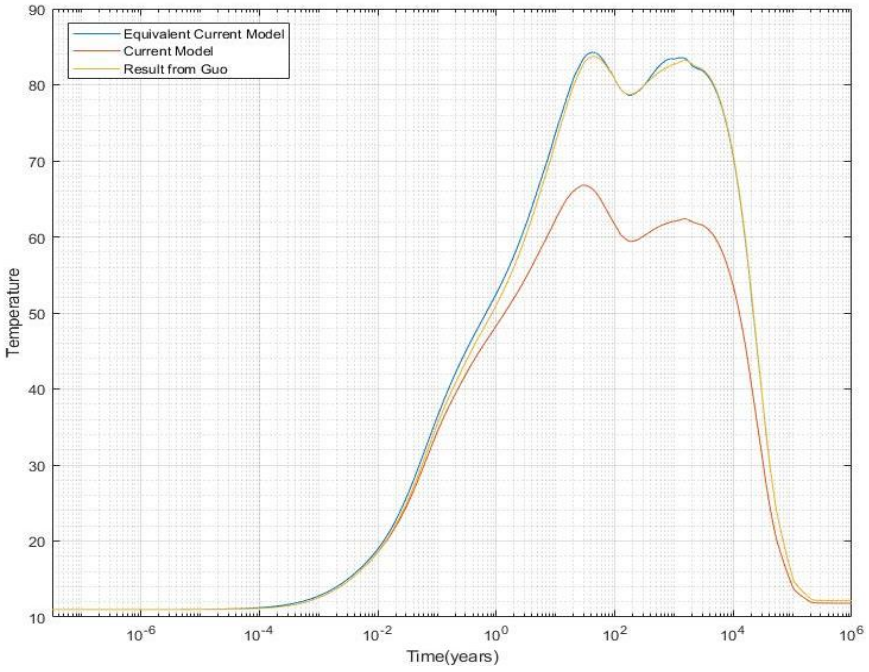


Figure 3.8 Comparison Between Benchmark Models and Base Model

The model validation shows that even though the spacing parameters differ in the current model and Guo's model, the timing of the maximum temperature is similar in both models. In addition, the magnitude of the maximum temperature is only a function of spacing if material properties are kept the same. This result indicates that this model will provide similar and consistent temperature profile if the spacing parameters vary and the maximum temperature can be interpreted as a function of spacing.

3.4.1.3 Computational Speed

Finally, the simulation time for the model is important when developing the surrogate model as it requires large number of simulations. The simulation is run on single socket computer with 8 cores AMD Ryzen 2700 CPU at 4.1 GHz with 16 GB of memory. Besides the processor speed, the simulation time is also sensitive to the COMSOL solver configuration, relative tolerance, time stepping and model size.

The following table summarizes the computation speed for different model and solver configuration, time stepping size, relative tolerance and solver configuration.

Table 3-4 Solution Time for different Model Setup and Solver Configuration

	Base Model	**Model B**	**Model C**	**Model D**	**Model E**
Model Size	0.7 x 25	0.5 x 20	0.7 x 25	0.7 x 25	0.7 x 25
Time Steps	$1\text{-}1^{13.5}$ Seconds 0.01 step	$1\text{-}1^{13.5}$ Seconds 0.01 step	$1\text{-}1^{13.5}$ Seconds 0.05 step	$1\text{-}1^{10}$ Seconds 0.05 step	$1\text{-}1^{10}$ Seconds 0.05 step
Relative Tolerance	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-05
Solver Type	Direct	Direct	Direct	Direct	Direct
Solution Time	1062 s	1115 s	827 s	270	269 s

Although reducing the mesh element will reduce the computational time, in this study, the mesh density will be kept similar to the density as in Guo's model. The result from higher tolerance does not affect the accuracy of the simulation. For based model, Model B and Model D, the simulation provides the same maximum temperature result to the one hundredth of a degree Celsius. Similar result is also obtained for model D, E and F.

Since the first peak occurs in the first hundred years will likely represents the maximum temperature reached within the repository, the simulation period from 1 to 1^{10} seconds (320 years) will be

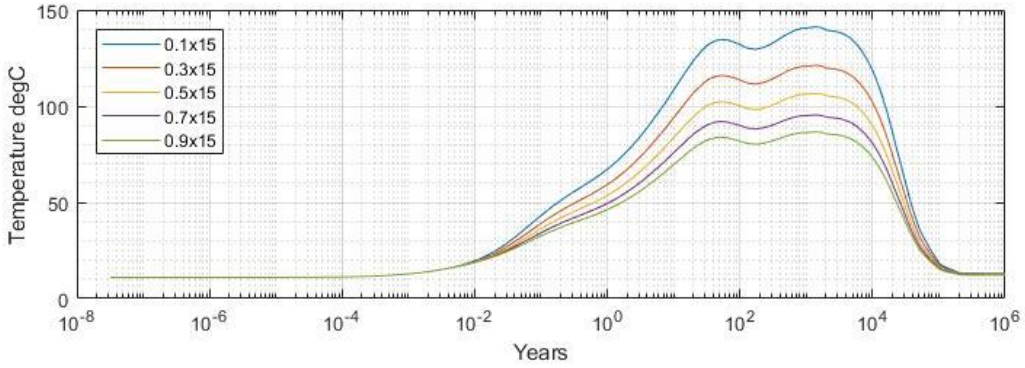
representative enough to determine whether the repository will reach the temperature constraint within its lifetime.

3.4.2 Spacing Variation and Sensitivity Analysis

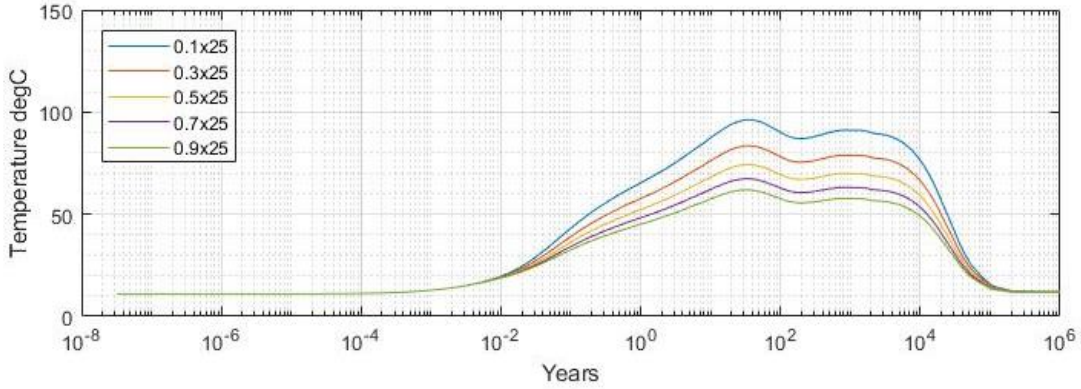
Given data on the heat generation, boundary condition and material properties, the maximum temperature in the placement room depends mostly on the spacing between the room and the spacer block width. The spacing variation explores the relationship between maximum temperature and the two spacing parameters, room spacing and the spacer block width. The sensitivity analysis explores changes of input data has on the temperature behavior of the repository.

3.4.2.1 Spacing Variation

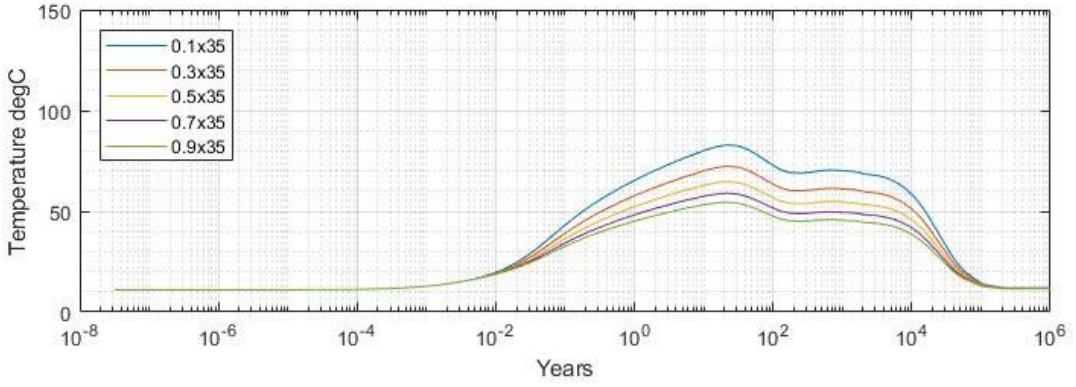
The spacing variation varies the spacer block dimension between 0.1, 0.3, 0.5, 0.7 and 0.9 meters, as well as varying room spacing between 15, 25 and 35 meters. As shown in **Error! Reference source not found.**, the label "0.1 x 15" represents spacer block dimension of 0.1m and room spacing of 15m. The results show that the temperature is more sensitive to the spacing between room than the spacer block width. In addition, the second peak in the temperature profile also depends on the geometry and it depends more heavily on the spacing between rooms than the spacer width. Whether the second peak will be higher than the first one shows dependence on the room spacing, however, the behavior of the first peak is more consistent throughout these simulations. Taking the advantage of this behavior and that the most important result is the magnitude of the first peak, the simulation does not need to run for the complete 10^6 years to determine the maximum temperature occurs during the lifetime of the repository. Setting the simulation period to 1^{10} seconds (320 years), will eliminate the need to correct the temperature overestimation while also having long enough simulation period to determine the maximum temperature within the lifetime of the repository. The following surrogate construction in Chapter 4 will therefore use this shorter simulation periods.



(a) Temperature Profile of Diffferent Spacer Width Configurations with Room Spacing of 15m



(b) Temperature Profile of Diffferent Spacer Width Configurations with Room Spacing of 25m



(c) Temperature Profile of Diffferent Spacer Width Configurations with Room Spacing of 35m

Figure 3.9 Temperature Profile for Different Spacing Configuration

The spacing variation result will be used to determine the optimal spacing configuration for a given set of assumption so that the temperature constraint can be met. Since the design space is relatively large for the time takes to run each simulation, the study of the optimization of spacing design will be detailed in Chapter 4 where surrogate modeling is ideal for solving this optimization problem.

3.4.2.2 Sensitivity Analysis

The sensitivity analysis explores the effect on the temperature profile by varying the input parameters of the model. In this study, the input parameters is varied by ±50% to examine how an individual input variation will affect the temperature profile from 1 to 1^{10} seconds (about 320 years). There are a total of 20 input parameters that are included in the analysis. The parameters include material properties data and boundary condition data for the surface temperature and the geothermal gradient. The base model used in this sensitivity analysis has the spacer width of 0.7m and room spacing of 25m same as the current design spacing.

The material properties include the density, heat capacity under constant pressure and thermal conductivity for the materials used in the base model. The parameter name used in the simulation and result figure presents in **Error! Reference source not found.** and Appendix B is shown in the parenthesis in Table 3-5.

Table 3-5 Values for Sensitivity Analysis Parameters

Material Properties Names		**-50%**	**Base**	**+50%**
k (W/(m*°C))	Rock (k_rock)	1.5	3	4.5
	Container (k_Con)	30.25	60.5	90.8
	Bentonite Box (k_BB)	0.5	1	1.5
	Spacer (k_SP)	1	2	3
	Gap Fill (k_GF)	0.25	0.5	0.8
	Asphalt (k_ASP)	0.375	0.75	1.1
C_p (J/°kg*°C)	Rock (cp_rock)	422.5	845	1267.5
	Container (cp_Con)	217	434	651
	Bentonite Box (cp_BB)	620	1240	1860
	Spacer (cp_SP)	1138	2276	3414
	Gap Fill (cp_GF)	435	870	1305
	Asphalt (cp_ASP)	460	920	1380
ρ (kg/m^3)	Rock (rho_rock)	1350	2700	4050
	Container (rho_Con)	3900	7800	11700.0

	Bentonite Box (rho_BB)	977.5	1955	2932.5
	Spacer (rho_SP)	530	1060	1590
	Gap Fill (rho_GF)	719.5	1439	2158.5
	Asphalt (rho_ASP)	980	1960	2940
Geothermal Gradient (gradient) (°C/m)		0.006	0.012	0.018
Surface Temperature (surftemp) (°C)		2.5	5	7.5

Although some of the input may varies less than ±50% in practice and some inputs may likely vary more than ±50%, the sensitivity analysis shown here is mainly to examine the effect of change but not a reflection of real variation of input parameters. In the uncertainty analysis presents in Section 4.4.3, similar methodology is used where a predetermined variation is applied to the based case material data for rock properties.

The followings are the sample simulation results over the simulation periods and its change from base case temperature profile over the time steps for all the parameters listed above. The rest of the simulation results can be found in **Appendix B.**

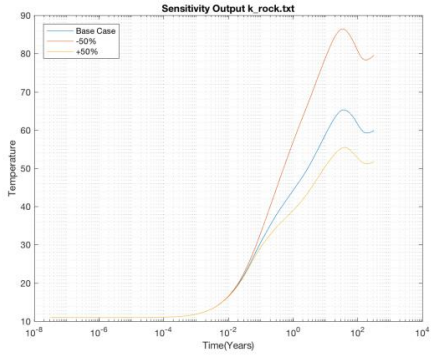
As shown in **Error! Reference source not found.**, the figure (a) shows the temperature profile of v arying thermal conductivity of rock by ±50% compared to the base case scenario in blue. It shows that by reducing the thermal conductivity of rock by 50% the temperature will increase significantly overtime, while increasing by 50% will reduce the temperature but less significantly. The figure (b) shows the percentage change from the base case. For example, the blue line in the top right corner represents that lowering the thermal conductivity will results in an increase of temperature around 30-40% of the original temperature over the 300 years simulation period. The temperature will diverge further from the original case as the simulation goes. The top right figure also shows that the temperature reducing effect is less significant than the temperature increasing effect.

On the other hand, the effect of container thermal conductivity change shown in the lower two figures indicates that the thermal conductivity of the container will not affect the temperature significantly over the entire period.

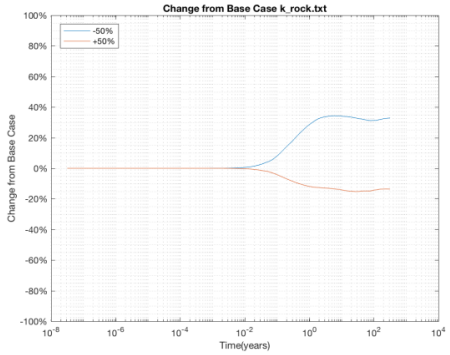
The parameters that shows the highest influence on the temperature profiles by percentage change in temperature:

- **Thermal conductivity** of rock, Bentonite Box, and Gap Fill,
- **Specific heat capacity** of rock, container, and Bentonite Box

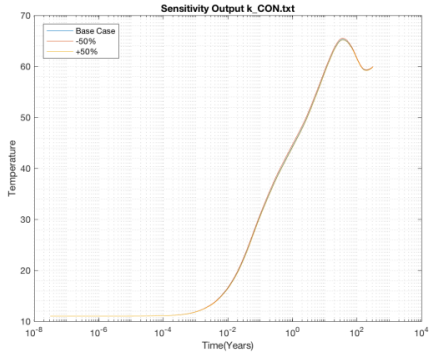
- **Density** of rock, container, and Bentonite Box
- Both geothermal gradient and surface temperature has significant effect in temperature at the beginning but reduces over the simulation period.



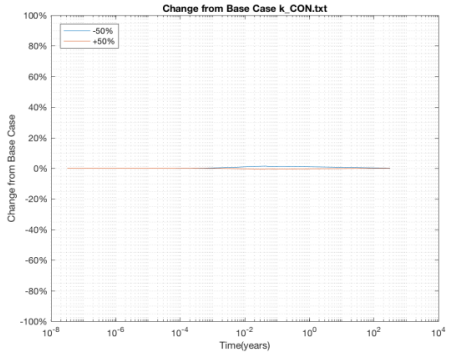
(a) Temp. Sensitivity for k_rock



(b) Temp. Percentage Change of k_rock



(c) Temp. Sensitivity of k_CON



(d) Temp. Percentage Change of k_CON

Figure 3.10 Sample Results from Sensitivity Analysis

The following is the summary of results for sensitivity analysis grouped by different types of influence on temperature profile for the simulate period:

- **Parameters that have little or no influence**:
    - Thermal conductivity of container, Thermal conductivity of spacer block, Thermal conductivity of asphalt.
    - Heat capacity of spacer block, Heat capacity of gap fill, Heat capacity of asphalt.
    - Density of spacer block, Density of gapfill, Density of asphalt.
- **Parameters resulted in increasing divergence in temperature**:

    - Thermal conductivity in rock.
    - Heat capacity in rock.
    - Density of rock.
- **Parameters resulted in reducing influence in temperature**:
    - Geothermal Gradient.
    - Surface temperature.
- **Parameters that only have influence in a specific time period**:
    - Thermal conductivity of bentonite box, Thermal conductivity of gap fill.
    - Heat capacity of container, Heat capacity of bentonite box.
    - Density of container, Density of bentonite box.

The sensitivity analysis shows that the surrounding material properties will have the most significant impact on the container temperature, with rock being the most significant factor. Therefore, in the design process, the thermal properties of the material that surrounds the used fuel container will be the most important factor to consider.

3.5 Summary

Based on the results from the numerical simulation, the newly constructed model produced similar results as the model reviewed in Chapter 2. Based on the current design and the input assumptions used in the modeling process, the maximum temperature is 66°C in year 47 which matches the timing the model shown in Chapter 2.

With different modeling setups, the simulation times may vary significantly. Using the optimal modeling setup can reduce the computational time significantly for the purpose of surrogate modeling in the following sections. The spacing variation study reveals that the temperature profile will vary significantly after the first 1000 years, since the second peak depends on the spacing parameters. However, it also shows that the peak temperature almost always happen within the first 100 years and the peak depends on the two spacing parameters in the model. The sensitivity analysis reveals that the rock properties has the most influence over the temperature. In addition, it also shows that different parameters influence the temperature differently over time. These results can be used as the basis for the siting process or design for the DGR. The current model is used later to optimize the spacing between rooms and between containers and model uncertainties of temperature profiles caused by different input parameters used in this model.

In the following Chapter 4, the numerical model developed in this chapter will be utilized for design optimization using the surrogate-based method.

Chapter 4
Surrogate Based Method for Design Optimization

4.1 Introduction

Based on the finite element model shown in Chapter 3, the maximum temperature occurring within the life of the repository can be simulated. For different design variables such as room spacing and spacer width and input data such as material properties, corresponding maximum temperatures can be obtained by simulating the desired design for given input data. The finite element model is treated as the high-fidelity model that is used to train and fit the surrogate model. Each simulation will take around 270 seconds to complete as shown in Table 3-4. Even though the Model E is the fastest model to be solved, doing design or optimization problem directly using this simulation will be time consuming. Approximation and predictive power of the surface response function such as RBF and polynomials, reduce the necessary simulation times for optimal design. The cost of designing directly using the high-fidelity method increases exponentially as the number of input parameters increases linearly [7]. In this case, where there are a large number of possible input parameters, the benefit of utilizing surrogate model for design optimization and uncertainty analysis is significant. This chapter will focus on the development of a surrogate model fitting algorithm and finding optimal designs to meet the maximum temperature constraints and also how to use surrogate models for uncertainty analysis for a given design. The section 4.2, defines the problem that the surrogate modeling will be solving. Section 4.3 presents details of the algorithms and key components used for solving the problem and section 4.4 presents results from the solution obtained by surrogate modeling.

4.2 Experiment Problem

The description of the problems that will be solved using the surrogate-based method is presented here. The spacing design problem considers only the temperature constraint to design the spacing parameters. The cost optimization problem takes cost as objective function while temperature as constraint. The multi-objective problem optimizes multiple objective functions including the cost and the DGR footprint. Unlike a typical surrogate optimization problem where the objective function is obtained by expensive model evaluation, in this case, the expensive model evaluation is to obtain constraint functions. The objective function in this case is a function of the spacing input which can be evaluated

very easily. The last problem is the uncertainty analysis where a surrogate model is fitted to the 3 material properties of rock and a Monte Carlo Simulation is run using this surrogate.

4.2.1 Spacing Design Problem

As shown in the finite element model, the maximum temperature will occur within 100 years of depositing the used fuel container. Since the heat generation of used fuel container decays as time progresses, and the temperature after the first peak will not likely be higher than this for the entire lifetime of the DGR. In the geometry of the numerical model, there are many dimensions that could be changed, the two most important dimension variables are the spacing between each placement room and the spacing between each container. The spacing between each container is determined by the spacer block width. Figure 4.1 shows the design parameters of room spacing and spacer width, since the geometry is symmetric, the diagram shown here is half of the design spacing.

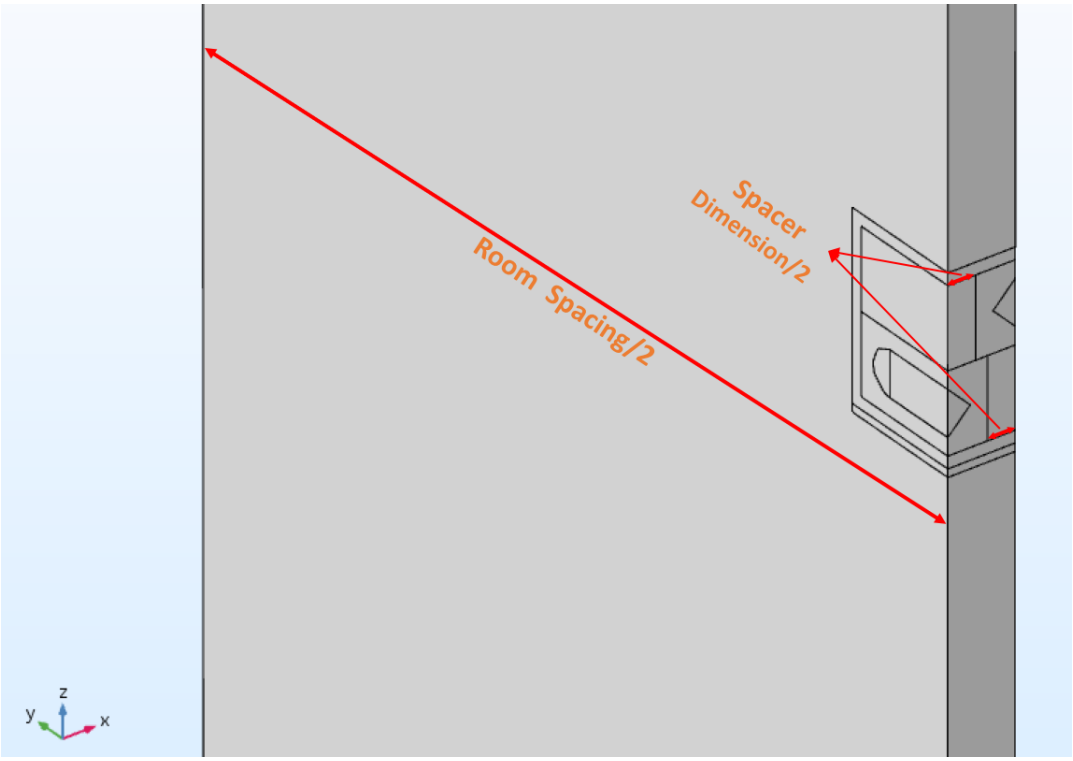


Figure 4.1 Spacing Design Parameters

The room spacing also decides how big the overall footprint of the repository is as shown in the overall DGR layout in Figure 2.1, the larger the room spacing is, the larger the footprint and the excavation is needed. But bigger spacing will reduce the maximum temperature of the repository as shown in the

sensitivity study in Appendix B. By increasing the spacer width, the maximum temperature can also decrease but it will need more material for building spacer block and less container can be fitted into each room which will result in the need of building more rooms.

Without considering the cost or material consumption, the best spacing is the one that just meets the maximum temperature constraint, in this experiment, the temperature constraint of 100°C and is used to determine the optimal design.

4.2.2 Cost/Multi-objective Optimizations

The cost formulation for the construction of the DGR at current stage is only an estimate of the actual cost. To formulate the cost estimation for different configuration of designs, the unit cost of the excavation volume is derived from the 2016 lifecycle cost estimate update summary report [2].

The cost formulation only considers the excavation cost arising from excavating the placement room and the material cost of manufacturing the spacer block since cost from access tunnel excavation or auxiliary support facilities will be determined by the overall design and layout of the DGR which depends on the condition of a specific site. From the information provided by the cost estimation report, the following is the calculation for estimating 1 m^3 of underground excavation for the placement rooms [2].

Table 4-1 Excavation Cost Estimation

Total Excavation Volume of one Panel			Total Cost of The Excavation	
Height (m)	Width (m)	Total Length (m)	Total Cost	Cost/m3
3.2	2.2	8498	$ 54,299,355	$ 908

Since the cost of highly compacted bentonite does not reflect directly in the cost estimation report, the cost is assumed to be the cost estimated from the cost of normal bentonite. The cost of the bentonite is obtained from Statista [26] as average bentonite cost in United States for 2017. The bentonite cost is $75/ton, since the highly compacted bentonite requires processing, for the purpose of this experiment, it is assumed to be twice of the regular bentonite at $150/ton in the cost formulation. Even though the room spacing and spacer width both determines the design, without considering the cost of excavating

access tunnel, perimeter tunnel, or other supporting facility, the cost is only a function of spacer width. In addition, the cost estimation also uses the following assumptions:

- The total number of used fuel containers is 108,833
- There are in total 34 dummy Bentonite boxes that occupies unsuitable position for container in each room
- Each room is assumed to be 308m for crystalline rock scenario, and 341.6m for sedimentary rock scenario
- Each room will have 22m of unusable length for room sealing at the entrance of the each room
- Each placement room will have constant cross section as 7.68 m^2 (3.2m x 2.4m)
- The spacer block made out of HCB will have a dry density of 1.7 ton/m^3
- Each spacer block is attached to a 1m width bentonite box
- The bentonite box dimension is 2.9 x 1 x 1 m
- HCB cost is \$150/ton, and excavation cost is \$908/m^3

With the above assumptions, the cost for a specific spacing design is formulated as:

$$\boldsymbol{Cost} = Total\ Rooms\ *\ Per\ Room\ Volume * Excavation\ Cost + Total\ Spacer\ Volume * HCB\ Density\ *\ Bentonite\ Cost$$

where

$$Total\ Rooms\ = \frac{108{,}833}{(Room\ Length\ - 22)/(1 + Spacer\ Dimension) - 34}$$

and

$$Per\ Room\ Volumn\ =\ 7.68\ *\ Room\ Length$$

$$Total\ Spacer\ Volumn =\ 1 *\ 2.9\ *\ Spacer\ \text{Width} *\ 108{,}833$$

This cost formulation is a highly simplified version of the real cost that will occur with respect to design parameters discussed here. However, the purpose of formulating the cost is to be able to develop the

surrogate model that could optimize the objective function, the cost function is needed for the implementation of the methodology. The cost optimization problem can be formulated as:

$$\boldsymbol{min_s\ Cost(S)} \tag{4.1}$$

$$s.t.$$

$$\boldsymbol{f(S, RS)\ -\ tempearture\ constraint\ \leq\ 0} \tag{4.2}$$

$$\boldsymbol{RS^L \leq\ RS \leq RS^u\ ,\ \ S^L \leq\ S \leq S^u} \tag{4.3}$$

where Cost is the cost objectives as a function of spacer width, and surrogate model $\boldsymbol{f}$ as a function of spacer width (S) and room spacing (RS).

As an extension of cost optimization, the algorithm can also optimize multiple objectives such as the footprint of the placement rooms and the cost. The footprint of the placement room is the total area taken by all the rooms, it will be affected by the overall layout of the DGR, but for the purpose of this problem, the footprint is formulated as a function of total number of rooms and room spacing. It is assumed the room is arranged in a single array of rooms with equal center to center room spacing between them.

$$\boldsymbol{Footprint}\ =\ Total\ Room\ *\ Room\ Spacing\ *\ Room\ Length$$

The multi-objective optimization will be solving the two design objectives, footprint and cost with temperature constraint function obtained by surrogate modeling. The temperature constraint of 100°C will be used. Similarly, the multi-objective optimization can be formulated as:

$$\boldsymbol{min_{RS,S}\ [Footprint(RS, S); Cost(S)]} \tag{4.4}$$

$$s.t.$$

$$\boldsymbol{f(S, RS)\ -\ tempearture\ constraint\ \leq\ 0} \tag{4.5}$$

$$\boldsymbol{RS^L \leq\ RS \leq RS^u\ ,\ \ S^L \leq\ S \leq S^u} \tag{4.6}$$

Figure 4.2 shows the visualization of the objective function given in equation 4.1 and 4.4 respectively, only the footprint is a function of both room spacing and spacer width. As shown in Figure 4.2, the cost function is only a function of spacer width and the cost does not change with increasing room spacing, while for the objective function of the footprint, it varies with spacer width as well as room spacing.

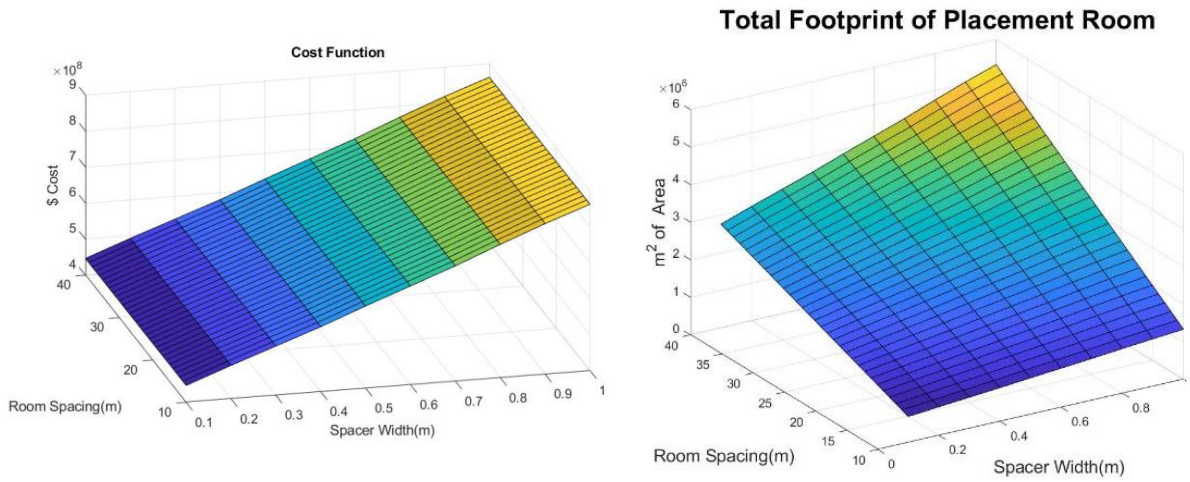


Figure 4.2 Visualization of Objective Functions

4.2.3 Uncertainty Analysis

Since the materials such as highly compacted bentonite may not be homogeneous, and the material properties may differ from design values when it is implemented. The uncertainties in the temperature will arise from the variation of these material properties. In addition, the thermal properties of rock and bentonite will also vary as the temperature changes in the DGR's lifetime [27]–[30]. More importantly, the rock composition is generally non-homogeneous which also creates uncertainties in temperature even if the rock type is known.

Therefore, surrogate model can be used to develop an interpolation model that can be used to produce probabilistic estimation of a given uncertainty in the input data. In this experiment, the surrogate model accounts for the uncertainties of the thermal conductivity, heat capacity and density of the rock is developed. The model have 3 variables for illustration of methodology, other variables can be considered for study in the future. Running exhaustive simulation requires a total of n^3 simulation, where n is the number of values taken for each variable. In this high dimensional application surrogate model can greatly reduce the necessary simulations for developing such model.

The experiment will first develop a surrogate model that is accurate enough in the entire sampling space. Then a monte carlo simulation is run using the surrogate model developed. Since the surrogate model is computationally cheap to run, a realization in the order of 10^4 can be simulated with negligible time.

4.3 Surrogate Based Algorithm

The surrogate model process is an iterative process shown in section 2.3. The algorithm used for the experimental problems shown above can be summarized as follows:

1. Generate initial experiment points for the input variables
2. Check if the points have been evaluated from previous simulation.
    a. If yes, store previous result and skip the simulation
    b. If not, run simulation on the experiment point using COMSOL
3. Fit the surrogate model using the points available for maximum temperature as a function of input variables
4. Find unique new sample points for evaluation (3 different ways of sampling for different types of problems)
    a. For spacing design problem, find a feasible design point that meets the temperature constraint within tolerance, if no new sample point can be found, increase the tolerance by 20%
    b. For cost optimization or multi-objective optimization, find the minimum point to evaluate, if no new point can be found the optimal cost point is found
    c. For uncertainty analysis, randomly generate sample points within the sampling space to fit the surrogate model for the entire domain
5. Evaluate the new sample points using COMSOL
6. Calculate the prediction error between surrogate model prediction and COMSOL simulation
7. Stop if prediction error within range or maximum iteration reached, otherwise go back to step 3

The followings are descriptions of initial experiment design, surrogate fitting, sample point selection and error calculation used in the algorithm and the problems described in section 4.2.

4.3.1 Initial Experiment Design

The initial design experiment is done by Latin Hypercube Sampling method (LHS) using MATLAB function *lhsdesign()*. The integer constraint is applied after the point has been generated. Since the LHS function in MATLAB generates normalized point in [0,1], the sampling points generated need to be scaled using the upper and lower limit of each of the variable as:

$$Scaled\ Sample\ Points\ =\ (Upper\ limit\ -\ Lower\ limit) \times S_n\ + Lower\ Limit$$

where S_n are the normalized sampling points generated by MATLAB.

The MATLAB default setting of *lhsdesign()*, maximizes the minimum distance between points, it can also be set to reduce the correlation between each point if needed.

After the sample points have been properly scaled to the scale of the variables, the sample point is first being rounded off to only keep 1 decimal place and then applied the integer constraint for the variable that only takes integer values such as Room Spacing. Since running simulation with one hundredth of a difference will only provide result that is not significantly different, taking sample points that is one tenth different will reduce the possible sampling points and reduce the computation time.

4.3.2 Surrogate Model fitting

The fitting of surrogate follows the solution process shown in section 2.3.3 for polynomial and RBF function. There are few forms of polynomial and radial basis function to choose from for fitting a surrogate model. All of the functions can be solved implicitly using the pseudoinverse in MATLAB. The following are the list of the functions available to choose from. The surrogate function is adopted from MATSuTo [13].

Table 4-2 Functional Surrogate Used in the Algorithm

Polynomial Functions	RBF Functions
Linear	Linear
Quadratic	Thin Plate Spline
Cubic	Cubic
Reduced Quadratic	
Reduced Cubic	

The surrogate function chosen for a specific application needs to be determined by the order of the true response surface, if a lower order surrogate is used to fit a higher order response surface, the surrogate will not provide an accurate solution and will fail to converge. In addition, if the fitting data has different magnitude of scale, bounding the data to be between 0 and 1 before fitting can increase the accuracy of the model.

4.3.3 Sample Point Selection

For the three different types of problems there are three types of sample selection methods. Each sampling method can be used in different applications and each one will result in different surrogate model and results.

Feasible Design Points Sampling

The feasible points sampling found the new sample points by determining all the feasible points that meet the temperature constraint within a tolerance. This sampling method can be summarized in following steps:

1. Generate exhaustive samples of design in the design space.
2. Calculate the temperature of these sample points using the most updated surrogate model.
3. Find out the design points that results in temperature within a tolerance(0.5°C), i.e ±0.5 °C of 100 °C temperature constraint.
4. Discard previous evaluated points, if no new sample point is found increase the temperature tolerance by 20% and repeat the process again.

This sampling method will result in sampling the points in the neighborhood of the design points that just meet the temperature constraint. However, this method does not consider the objective functions and the temperature prediction is only accurate near these feasible design points. This sampling method can be used to find out all the feasible design points in terms of meeting the temperature constraint. This sampling method is designed to solve the problem shown in section 4.2.1.

Optimal Point Sampling

The optimal point sampling method can be used to find out the optimal design points that both meets the temperature constraint and minimize the objective functions. As shown in section 4.2.2, the surrogate model is used to compute the constraint. The objective function can either be cost minimization or could include multi-objectives. The process of finding the new sample point is:

1. Using MatLab function *ga()* or *gamultiobj()* to find the optimal design points that meets the constraint.
2. Determine if the sample point has evaluated before.

a. If sample point has already been evaluated, optimize again to find a new point, if no new sample point can be found, the optimal design is found.

b. If sample point has not been evaluated, apply appropriate rounding and apply the integer constraint and return the sample points.

The optimal point sampling will minimize the objective function while meeting the temperature constraint. For cost optimization, the genetic algorithm finds the optimal point that minimizes the cost function. While for multi-objectives optimization, the genetic algorithm finds the pareto front of the objective space [19], [31].

Even though the objective function can be directly optimized without running the simulation when the temperature constraint is not considered, the surrogate-based optimization using the surrogate as constraint function reduces the necessary simulation required to find the design point that is within the predetermined temperature constraint.

Global Sampling

The global sampling method uniformly samples a predetermined number of points within the entire variable space. This sampling method can be used to develop a surrogate model that will make prediction within the entire domain of a variable. This method of sampling requires many more simulation runs in order to reach convergence since the surrogate has a larger domain to fit. This sampling method will be useful if the surrogate is required to develop good level of fitting for the entire sampling domain, which is ideal for uncertainty analysis problem mentioned in section 4.2.3. It is also being implemented in MATSuMo toolbox [13]. The process of the sampling the variable domain can be summarized as:

$$Global\ Sample = U * Sample\ Range + Lower\ Sample\ Bound$$

where, U is the uniform random variable in [0, 1]

Alternatively, the global sampling can also utilize the LHS method mentioned earlier for initial sampling point selection.

4.3.4 Model Validation and Correction

Once the prediction result is made by the surrogate model and the new sample point is selected by a corresponding method, the result is compared with the simulation result evaluated at the new sample points. The accuracy of the surrogate model prediction is determined by calculating the error with the

simulation result. The error calculation can choose from one of the three ways, Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Square Error (MSE).

The MAE is calculated as:

$$\boldsymbol{MAE} = \frac{\boldsymbol{\Sigma_{i=1}^{n}|y_i - \hat{y}_i|}}{\boldsymbol{n}} \quad \textbf{(4.7)}$$

where y is the simulation result and $\hat{y}$ is the surrogate prediction result.

The MSE is:

$$\boldsymbol{MSE} = \frac{\boldsymbol{\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\boldsymbol{n}} \quad \textbf{(4.8)}$$

And the RMSE is:

$$\boldsymbol{RMSE} = \sqrt{\frac{\boldsymbol{\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\boldsymbol{n}}} \quad \textbf{(4.9)}$$

The surrogate modeling process stops when the maximum iteration is reached for fitting and sampling new points, or when the error of temperature prediction made by the surrogate model is within a tolerance. In the case of design optimization problem, the error tolerance is set to 0.5°C and the error term is MAE which can be easily interpreted as the temperature difference between simulation result and surrogate prediction.

If the current surrogate model does not provide prediction within the reasonable error tolerance, a new surrogate will be fitted using the newly generated point from the simulation. As a result, the surrogate updates and corrects itself as more points are evaluated. The surrogate model will converge to the error tolerance if the surrogate model becomes accurate enough for a given error criteria or when it is running out of new point to evaluate which also indicates that the surrogate has found the optimal solution, or it has evaluated all the possible points and fitted itself correctly in the interested domains.

4.4 Surrogate Modeling Result

As mentioned above, there are three types of problems to be solved using the surrogate-based method, the spacing design problem, optimization problem and the uncertainty analysis problem. The following result only shows one solution for each of the experiment problem. For each solution, the initial DOE point, initial fitted surrogate and final surrogate will be shown. For all the problems, the initial design point is set to 10 points, and other than the room spacing being integer, the rest of the variables will

round to 1 digit after decimal point. The convergence criteria is 0.5°C for error between prediction result and simulation result.

4.4.1 Spacing Design Problem

For the spacing design problem, the Cubic RBF is used as the functional surrogate and the MAE is used as the error term to determine the convergence condition. The initial DOE point used for evaluation is shown in Figure 4.3. The design space for the LHS sampling is bound by [10, 40] of room spacing and [0.1, 1] for Spacer Width.

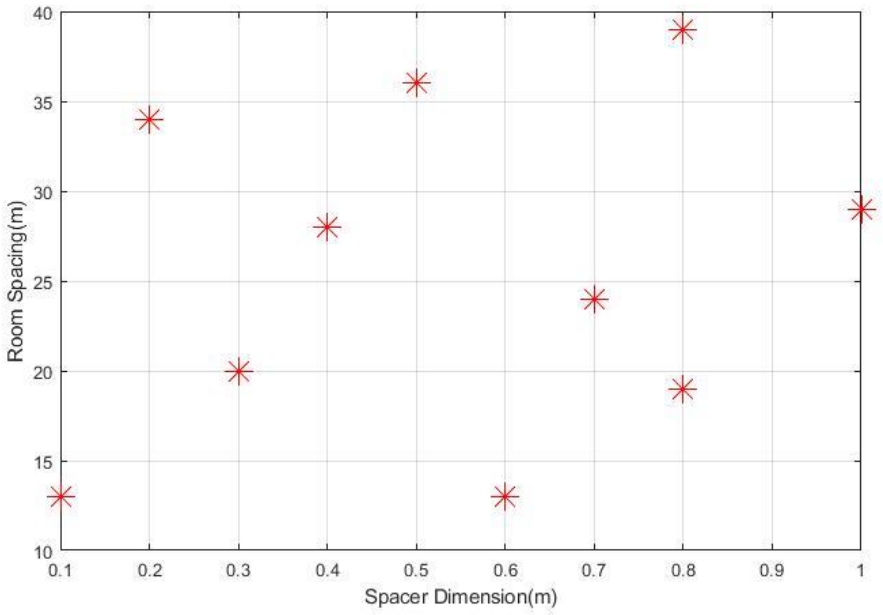


Figure 4.3 Initial DOE Point for Spacing Design Problem

After the initial 10 DOE points have been evaluated, the maximum temperature for each simulation is taken as the output of the simulation. A cubic RBF is then fitted with the output generated from the simulation. The resulting surrogate model is shown in Figure 4.4. The fitted surrogate represents the maximum temperature as a function of the room spacing and spacer width. This initial surrogate is then used to determine all the feasible design point that can meet the temperature constraint of 100°C. There are in total of 310 (31 x 10) possible sampling points in the entire domain. The algorithm using the RBF can calculate these points and determine the feasible points in a fraction of a seconds without the need to run the simulation. The design point that is within the limit of temperature constraint is then selected as the next evaluation point for COMSOL simulation.

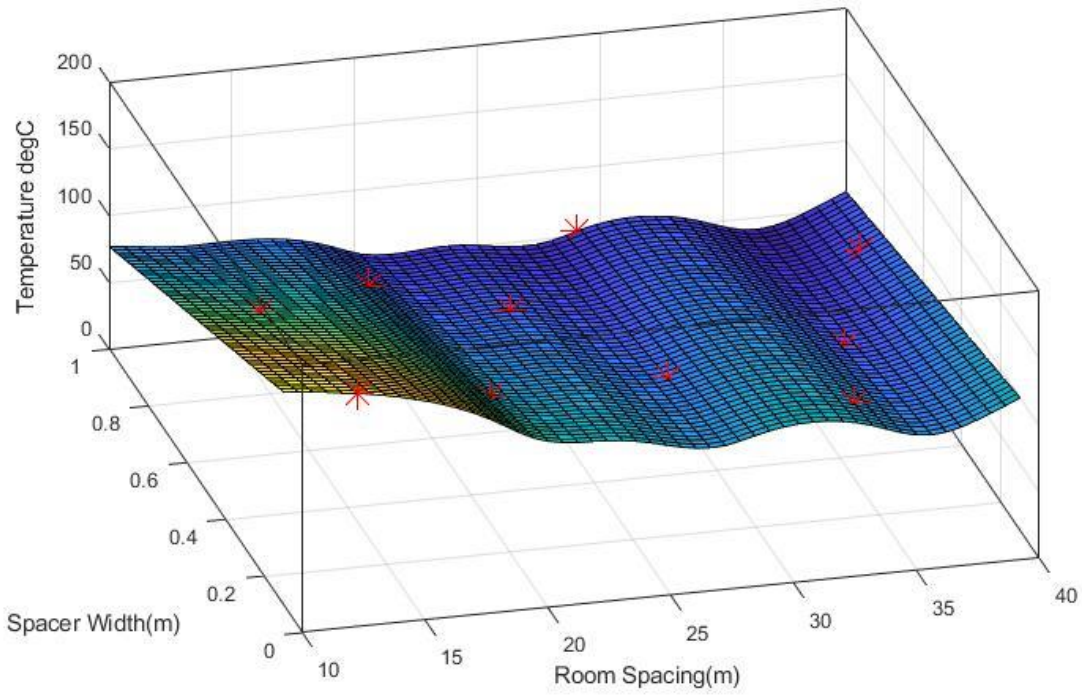


Figure 4.4 Initial Surrogate Fitting for Spacing Design Problem (Cubic RBF)

Following the iterative steps shown in section 4.3, the surrogate model can reach the convergence criteria within 0.5℃ of MAE error with 4 iterations, the final MAE error is 0.26°C. The resulting surrogate model is shown in Figure 4.5. In addition to the 10 points evaluated initially, the surrogate converges after evaluating 9 additional points. The error progression for the 4 iterations is shown in Figure 4.6.

The Figure 4.5 shows the 9 points evaluated after the final surrogate is fitted, it can be observed that the additional evaluated points(green) are in the neighborhood of 100℃, which indicates that the algorithm is searching for the design points that is close to the temperature constraint. The 4 points in red circle is the point being selected as the feasible design point after the surrogate model has converged.

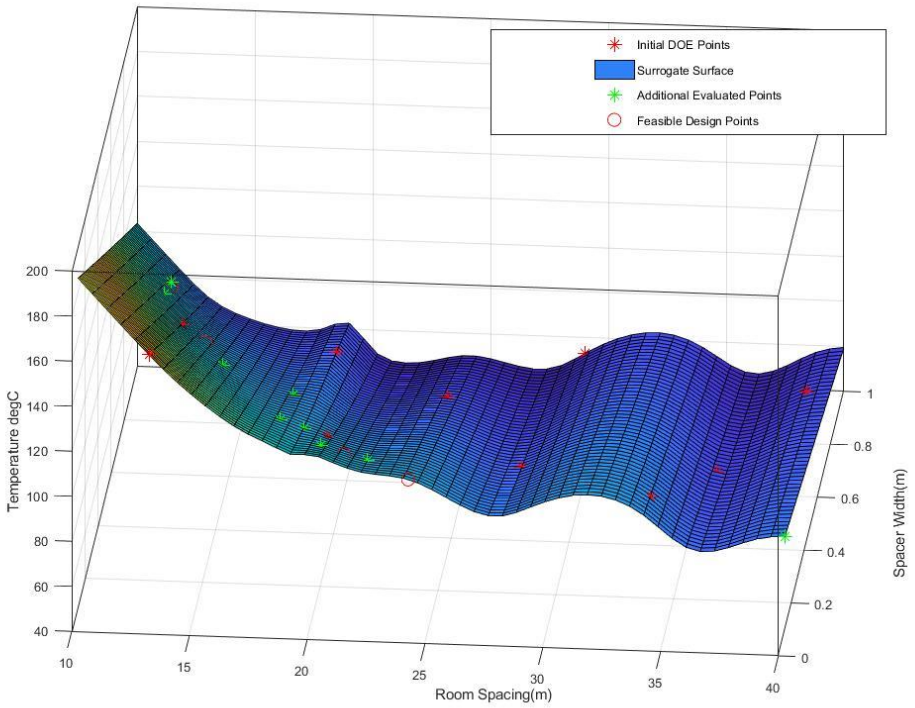


Figure 4.5 Final Surrogate Model for Spacing Design Problem (Cubic RBF)

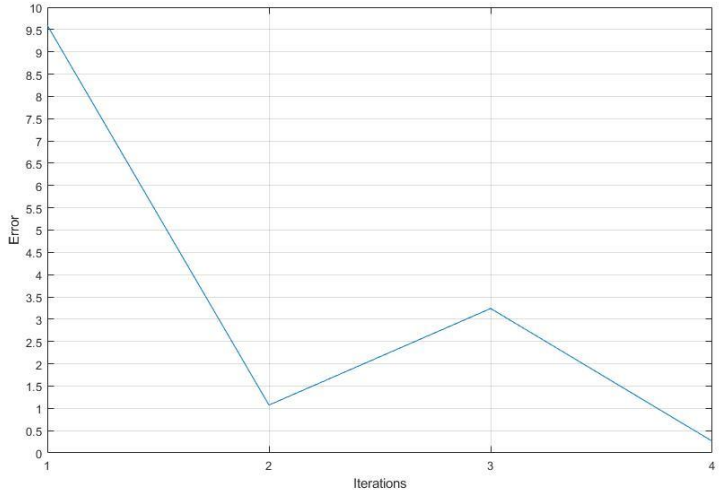


Figure 4.6 Error Progression for Spacing Design Problem

The final surrogate surface is smoother than the initial surrogate near the area around 100℃ and optimal design points since more points are fitted in that neighborhood.

The four final design points are found by setting a temperature tolerance and temperature constraint. In this Case, setting the tolerance of 1℃ and constraint of 100℃ returns 4 optimal design points. The design points and its COMSOL simulation result are shown in Table 4-3.

Table 4-3 Spacing Design Problem Solutions

Design Point		Surrogate Result (℃)	COMSOL Result (℃)	Absolute Error (℃)
S(m)	RS(m)			
0.1	24	100.8	100.4	0.4
0.2	21	99.1	99.4	0.3
0.6	14	99.3	100.5	1.2
0.8	12	99.5	101.8	2.4
MAE				1.1

As shown above, the design point temperature estimation by the surrogate matches closely with the result obtained by the COMSOL simulations. Although there is still discrepancy of results, the surrogate model has provided reasonably accurate results for the problems given. With tighter convergence criterion, the error can be improved further.

4.4.2 Design Optimization Problem

There are mainly two optimizations problems to be solved, one is the cost optimization, and the other is the multi-objective optimization as outlined in section 4.2.2. Since the cost function is a function of spacer width, without incorporating the room spacing into the cost function the algorithm will finds the smallest spacer width as optimal design. Therefore, using the multi-objective optimization will incorporate the room spacing parameters into the objective function which may result in a solution that does not decide only on the spacer width.

4.4.2.1 Cost Optimization

For the cost objective function formulated in section 4.2.2, genetic algorithm with integral constraint is used to solve the optimization problem. This experiment used the cubic polynomial other than the cubic RBF to examine the polynomial's ability for fitting the response surface. The initial 10 DOE points

generated for simulation, and the surrogate model fitted for these initial points are shown in Figure 4.7 and Figure 4.8.

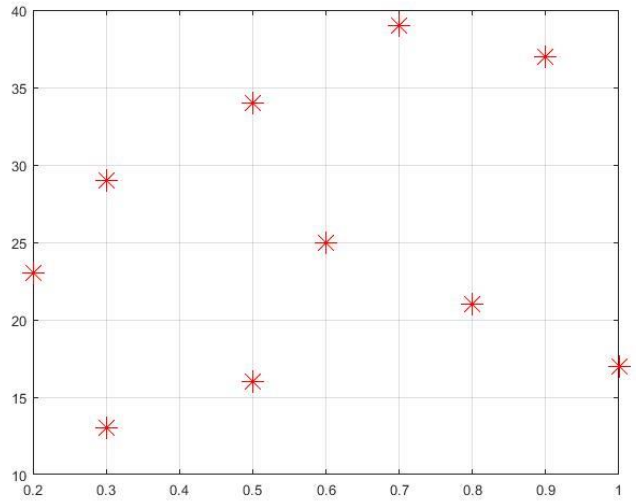


Figure 4.7 Initial DOE Point for Cost Optimization Problem

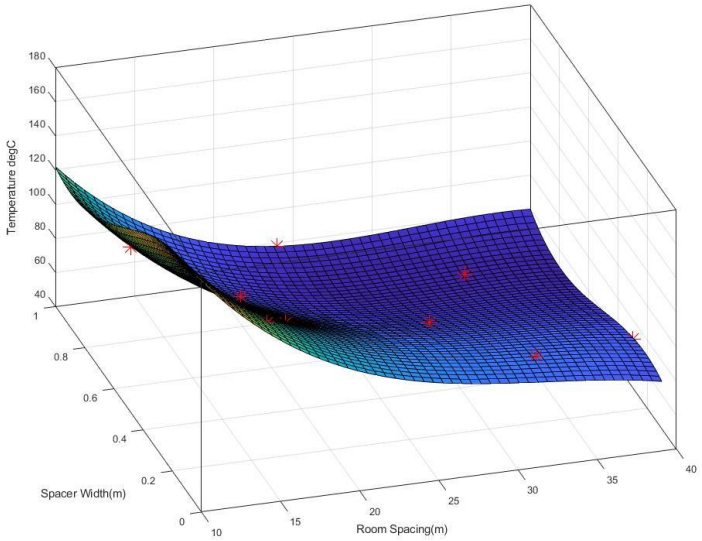


Figure 4.8 Initial Surrogate Model for Cost Optimization Problem (Cubic Polynomial)

The genetic algorithm then takes the surrogate temperature prediction as the constraint for the optimization and solves for the minimum cost point of the design iteratively improves the surrogate accuracy. The optimization function is formulated as:

$$\boldsymbol{min_s\ Cost(S)} \tag{4.10}$$

$$s.t.$$

$$\boldsymbol{f(S, RS)\ -\ tempearture\ constraint\ \leq\ 0} \tag{4.11}$$

$$\boldsymbol{RS^L \leq\ RS \leq RS^u\ ,\ \ S^L \leq\ S \leq S^u} \tag{4.12}$$

The final surrogate model and the error progression is shown in Figure 4.9 and Figure 4.10. As shown in these figures, the algorithm takes 8 iterations and 6 additional evaluation points to reach the convergence criterion, the final error is 0.49°C. The optimal result in this case all has spacer width of 0.1 m since the cost function is a function of the spacer width only, and the optimal design is all the points that have the minimum spacer width that meets the temperature constraint.

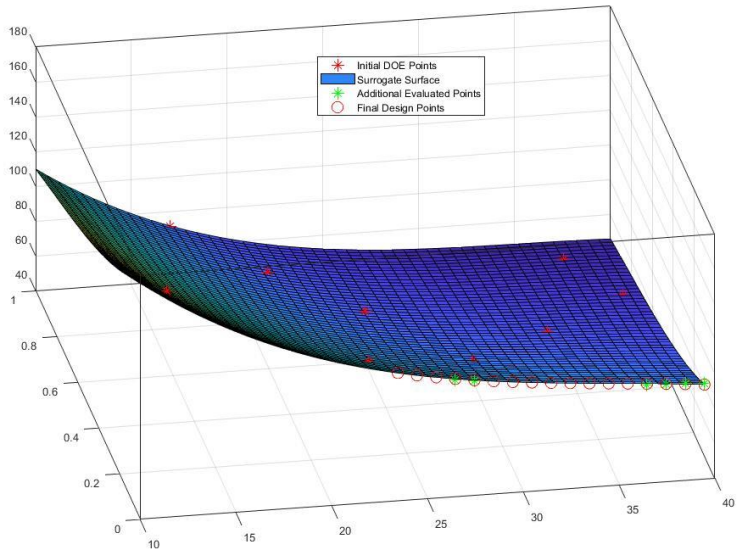


Figure 4.9 Final Surrogate for Cost Optimization Problem (Cubic Polynomial)

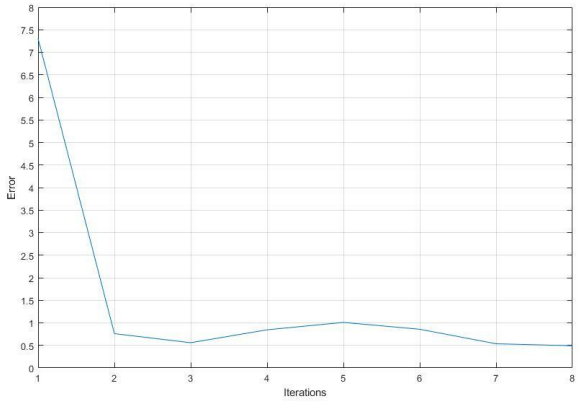


Figure 4.10 Error Progression for Cost Optimization Problem

Since all the optimal points have the same spacer width, the minimum cost is the same for all the design points. With exhaustive search using the GA, there are 17 final cost optimal points, which all have the same spacer width of 0.1m. To compare the accuracy of the surrogate model, the comparison between the temperature prediction and COMSOL simulation result is calculated for the optimal points selected. The minimum cost for these design points are all $448,740,938, which means based on the objective function, it is indifferent to any of these points.

Table 4-4 Final Optimal Cost Design Point

Design Point		Surrogate Result (°C)	COMSOL Result (°C)	Absolute Error (°C)
S(m)	RS(m)			
0.1	24	99.8	100.4	0.59
0.1	25	97.7	98.3	0.66
0.1	26	95.7	95.4	0.40
0.1	27	94.0	93.7	0.34
0.1	28	92.5	92.2	0.36
0.1	29	91.1	91.4	0.26
0.1	30	89.9	90.1	0.15
0.1	31	88.8	88.9	0.06
0.1	32	87.8	87.3	0.45
0.1	33	86.9	86.4	0.47
0.1	34	86.0	85.6	0.39
0.1	35	85.1	84.8	0.27
0.1	36	84.3	83.5	0.80
0.1	37	83.4	82.8	0.58
0.1	38	82.5	82.3	0.22
0.1	39	81.5	81.8	0.25

0.1	40	80.5	81.3	0.81
MAE				0.4

Additional three surrogate processes are run to test the algorithm's dependence on the initial sampling points, each of these processes is initialized with different sampling points. As shown in Table 4-5, the starting error is the error of the surrogate after it is fitted with the initial sampling points, and error after convergence is the final error between the surrogate prediction and all the points evaluated for fitting the surrogate. The starting error depends on the initial sampling points, since these points will dictate the fitting of the initial surrogate. If the initial error is small, the convergence process will be faster. However, the total number of final design points does not change with different initial samplings or the surrogate model fitted. For this case, all 17 points are the same as shown in Table 4-4. The difference in error after convergence are caused by the differences in the points evaluated by COMSOL during the fitting process. The difference in point evaluation resulted in slightly different surrogate models and error at convergence each time. However, using these slightly different surrogates, the algorithm is still able to return the same final designs in all three cases.

Table 4-5 Sensitivity on Initial Sampling Points

	Total Design Points	Total Iteration	Total Points Fitted	Starting Error (°C)	Error after Convergence (°C)
Run 1	17	4	13	7.02	0.44
Run 2	17	2	11	0.91	0.10
Run 3	17	4	13	11.60	0.38

4.4.2.2 Multi-Objective Optimization

The similar process applies to the multi-objective optimization, the MATLAB function *gamultiobj()* is used for solving the multi-objective optimization problem. The surrogate function used in this case is Cubic RBF. The initial DOE points and the initial fitted model is shown in Figure 4.11 and Figure 4.12.

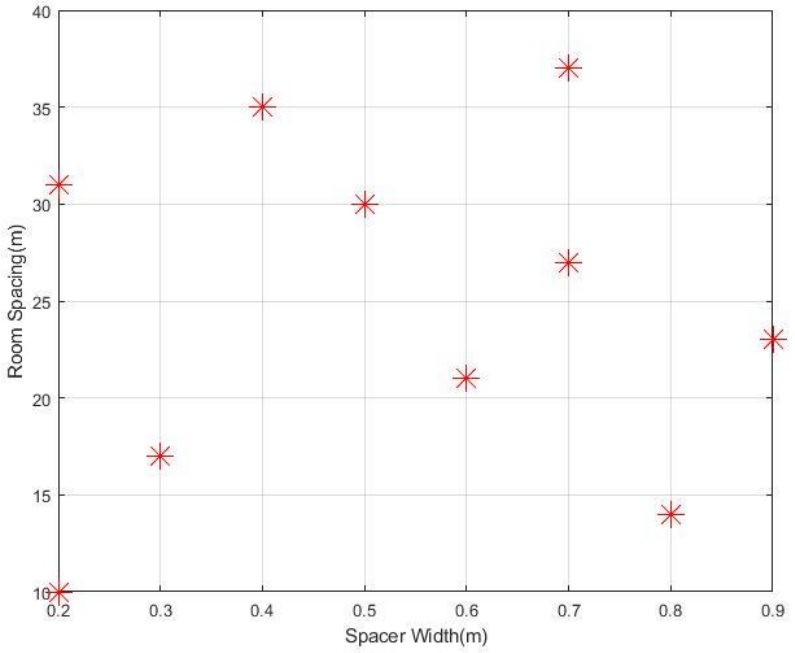


Figure 4.11 Initial DOE Point for Multi-Objective Problem

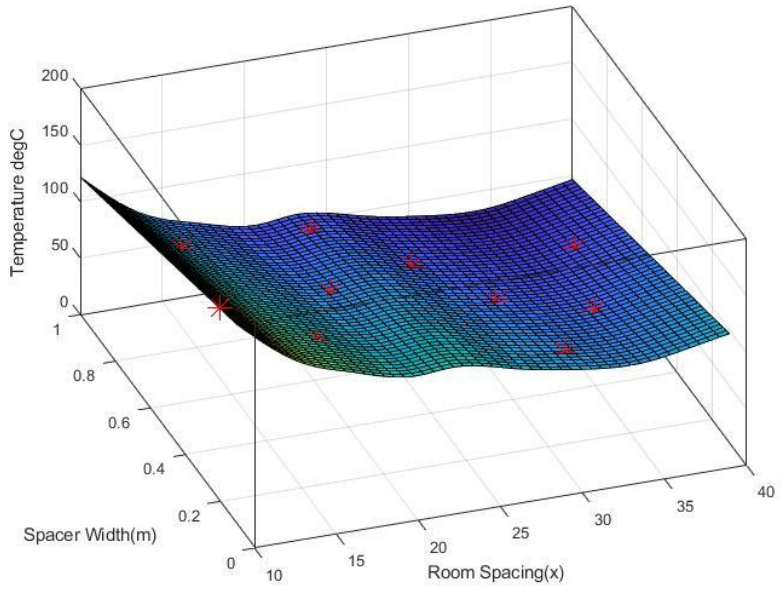


Figure 4.12 Initial Surrogate for Multi-Objective Problem (Cubic RBF)

After the initial surrogate is fitted, the algorithm took 3 iterations and an additional 15 COMSOL evaluations to reach the convergence criteria, the final MAE is 0.091°C. The final surrogate model in Figure 4.13 shows that the optimal points obtained by the algorithm are all in a close neighborhood. Unlike the optimal point in cost optimization, in this case, the optimal configuration is not restricted with only having spacer width of 0.1m due to the inclusion of room spacing in the objective functions.

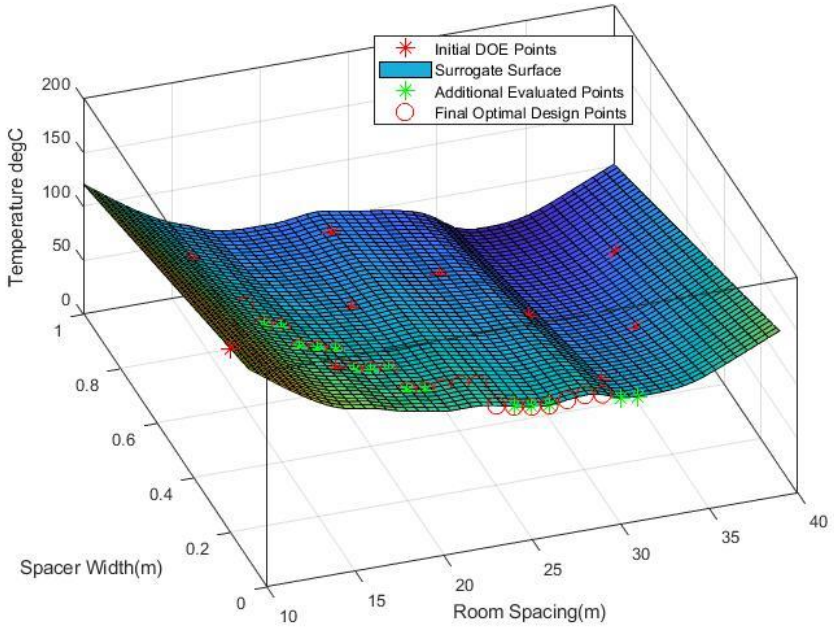


Figure 4.13 Final Surrogate for Multi-Objective Problem (Cubic RBF)

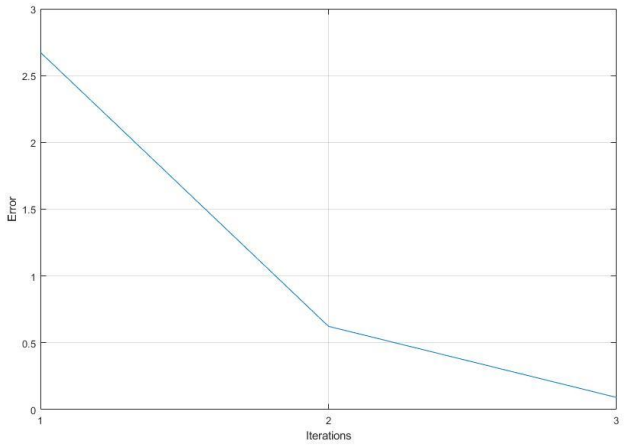


Figure 4.14 Error Progression for Multi-Objective Problem

The optimal solutions results are shown in Table 4-6, the optimal solutions shown here are the solution of the optimal design points after exhaustive search using *gamultiobj().* Since each multi-objective optimization will return different pareto front, solving the optimization multiple times will results in solution in different pareto fronts. The table below shows all the unique 21 points found using the *gamultiobj().*

Table 4-6 Possible Multi-Objectives Optimal Design Points

Design Point		Objectives		Surrogate	COMSOL	Absolute
S(m)	RS(m)	Cost ($)	Footprint (m^2)	Result (°C)	Result (°C)	Error (°C)
0.1	**24**	**4.5E+08**	**1.7E+06**	**102.2**	**100.4**	**1.8**
0.1	25	4.5E+08	1.7E+06	98.3	98.3	0.0
0.1	26	4.5E+08	1.8E+06	95.4	95.4	0.0
0.1	27	4.5E+08	1.9E+06	93.7	93.7	0.0
0.1	28	4.5E+08	1.9E+06	95.6	92.2	3.4
0.1	***29***	***4.5E+08***	***2.0E+06***	***97.3***	***91.4***	***5.9***
0.1	30	4.5E+08	2.1E+06	94.9	90.1	4.9
0.2	20	4.9E+08	1.5E+06	102.6	102.6	0.0
0.2	21	4.9E+08	1.6E+06	99.4	99.4	0.0
0.2	22	4.9E+08	1.7E+06	99.2	96.6	2.6
0.2	23	4.9E+08	1.7E+06	99.4	94.1	5.3

0.2	***24***	***4.9E+08***	***1.8E+06***	***97.0***	***91.9***	***5.1***
0.3	18	5.4E+08	1.5E+06	101.7	101.7	0.0
0.3	19	5.4E+08	1.6E+06	98.1	98.1	0.0
0.3	20	5.4E+08	1.7E+06	96.3	94.9	1.4
0.4	**16**	**5.8E+08**	**1.4E+06**	**103.0**	**103.0**	**0.0**
0.4	17	5.8E+08	1.5E+06	98.7	98.7	0.0
0.4	18	5.8E+08	1.6E+06	94.9	94.9	0.0
0.5	15	6.3E+08	1.5E+06	101.3	101.3	0.0
0.5	16	6.3E+08	1.6E+06	96.7	96.7	0.0
0.6	15	6.7E+08	1.6E+06	95.1	95.4	0.4
				MAE		1.5

Figure 4.15 shows the result of the pareto front from a single *gamultiobj()* optimization, after applying the integral constraint and combining multiple optimization results, the above Table 4-6 is obtained.

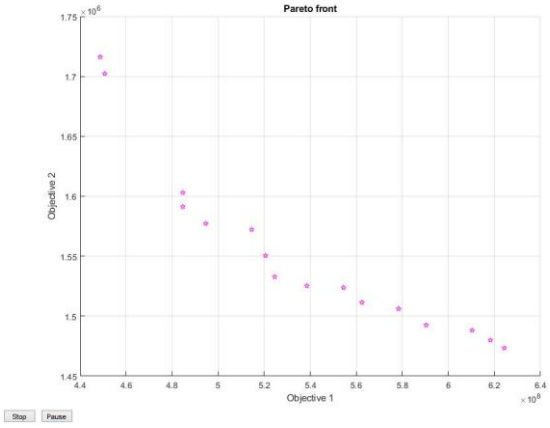


Figure 4.15 Pareto Front of Genetic Algorithm Multi-Objective Optimization

The objective values on all the design points from Table 4-6 is shown in Figure 4.16. For these optimal design points the Pareto front can also be found, and it is labelled in red As a result, these 4 points will be the best design points considering the objectives function. These points are bolded in Table 4-6. Taking in to consideration of the temperature constraint, the best design points are the bold italic point. The two points (0.1, 29) and (0.2, 24) are the points where COMSOL result meets the temperature constraint of 100°C.

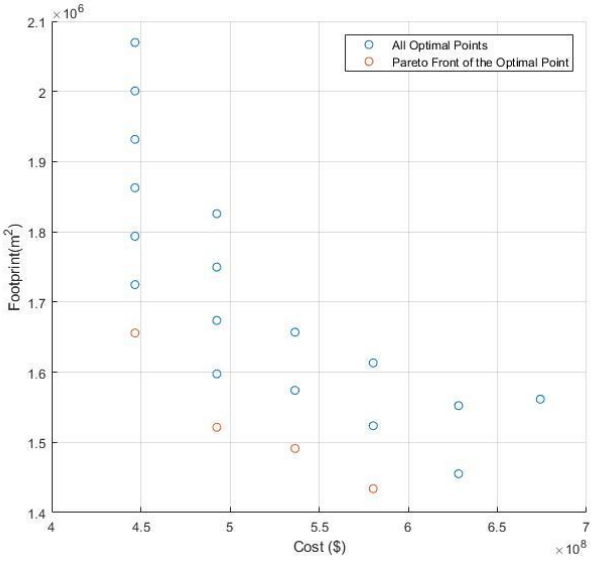


Figure 4.16 Pareto Front of the Optimal Design Points

Similarly, three additional surrogate optimization processes are run to determine the dependence on initial sampling points. Different from the previous case, in this case the total design points may change based on the initial sampling points. Since the objective function covers both the room spacing and spacer width parameter, as a result, a larger area of the surrogate is explored. If the initial sampling point is not the same near the design points, the curve fitted will be different resulting in different number of design points. However, even though the starting error varies slightly among the three cases, the error after convergence are close to each other.

Table 4-7 Sensitivity on Initial Sampling Points.

	Total Design Points	Total Iteration	Total Points Fitted	Starting Error (°C)	Error after Convergence (°C)
Run 1	24	2	20	5.36	0.19
Run 2	24	4	28	3.98	0.14
Run 3	28	5	26	8.05	0.20

The results show that the initial sampling points will slightly affect the starting error of the surrogate, speed of convergence and the final design points. The error after convergence show that despite these slight differences, the model is able to converge to a similar level of accuracy in the end.

4.4.3 Uncertainty Analysis

The uncertainty analysis aims to produce a surrogate model that is reasonably accurate for the entire variable domain. In this experiment, the surrogate model is to be fit with the 3 material properties of the rock using a sample design case of 0.5m spacer width and 25m center to center room spacing. Once the surrogate model is developed, Monte Carlo simulation will be run on the surrogate given a predetermine spacing configurations and normal random variables on the thermal properties of the rock. The cubic RBF is used in this experiment. Since it is hard to visualize higher dimensions, the visualization of the fitted surface is not shown here. The evaluated points will be shown using scatter plot matrix.

The upper and lower bound of the spacing variables remains the same, and the following bound is applied to the three properties of the rock.

- Density of Rock [2000,4000]
- Thermal Conductivity of Rock [1,6]
- Thermal Capacity of Rock [400,1300]

The value assigned here aims to include the possible range of the thermal properties of the common rock type according to the range of values from some of the technical reports on thermal properties of rock [27], [28]. The Figure 4.17 shows the first 15 initial design points used to generate the first surrogate models as visualized in 2D space.

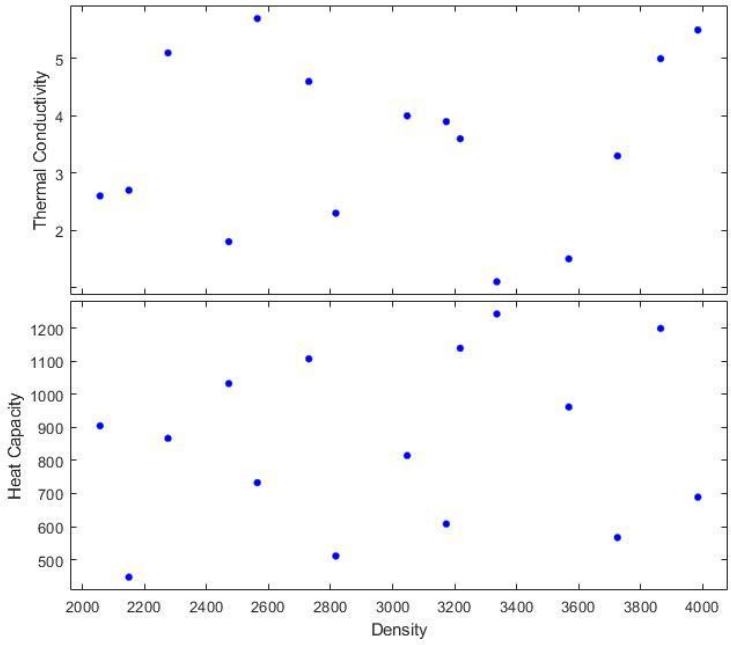


Figure 4.17 Initial DOE Points for Uncertainty Analysis

In this case, since the scale of the parameter is very different from one another, the input data points are first normalized to the scale of 1 using the upper bound of sampling for each of the parameters.

After the initial model is fitted, the algorithm samples uniformly in the global variable domain as mentioned in section 4.3.3 until the surrogate model meets the convergence criteria. In each iteration, 5 unique sampling points are generated. As a result, the surrogate model converged after 13 iterations and a total of 75 simulations are evaluated and fitted to the surrogate.

In this experiment, since the sampling space is larger than previous experiments, a smaller error tolerance of 0.1 is used and the error calculation used is RMSE. The error progression is shown in Figure 4.18 the final RMSE error is 0.0626.

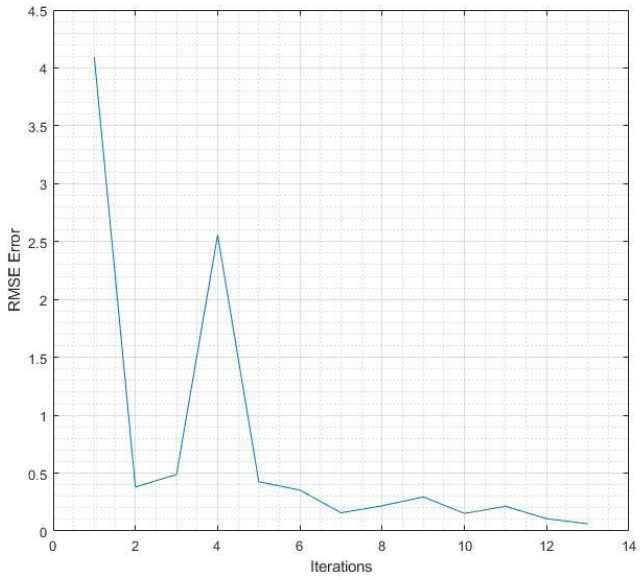


Figure 4.18 Error Progression for Uncertainty Analysis Problem

The Figure 4.19 shows all the points evaluated before the surrogate model is converged.

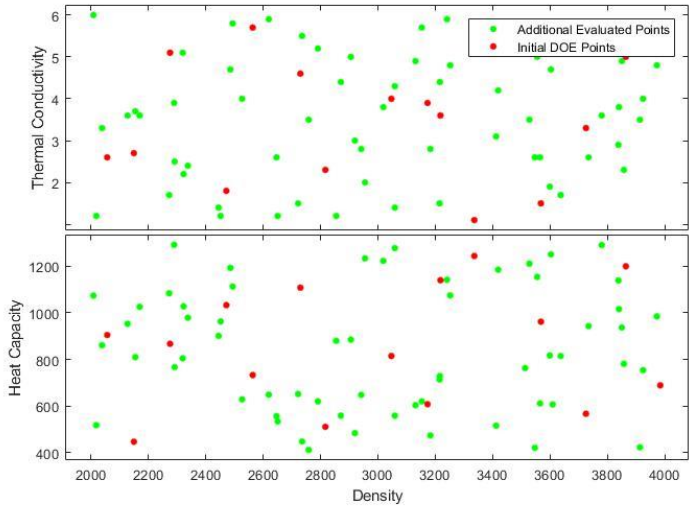


Figure 4.19 Evaluation Points for Uncertainty Analysis

Before the final surrogate is used to determine the uncertainty of the design, the result predicted by the surrogate are compared with the results from the simulation for the sample design case using the input as (2700, 3, 845) representing the rock density, thermal conductivity and heat capacity. The following

table shows the comparison of the surrogate prediction and simulation result. There are in total 320 known simulation results evaluated from all the surrogate fitting processes. The surrogate model shown here is fitted with 75 of these 320 points which are the points evaluated in this fitting process. The error shown here indicates that the surrogate has yet to generalize the entire variable domains to the predetermined convergence error. However, the result of the prediction is within a reasonable range.

Table 4-8 Uncertainty Analysis Model Validation

Comparison Data Group	Surrogate (°C)	COMSOL (°C)	MAE Error (°C)
Current Fitted point			0.0626
320 Known Simulation Results			1.0
Sample Case Results	72.4	73.0	0.6

Even though the surrogate shown above has yet to show perfect generalization for the entire variable domain, the Monte Carlo simulation run using such model still provide reasonable results for analyzing the uncertainty caused by rock properties. The Monte Carlo simulation will be run on the sample design where the room spacing is 25m and spacer width is 0.5m. In this experiment, the random variable will be generated from a normal distribution with the mean value equals the base case input shown in Table 3-2 and a Coefficient of Variation of 10%. Using the surrogate model, a large realization (10^4) can be run in a fraction of a second. The followings are the mean and standard deviation result using different realizations.

Table 4-9 Monte Carlo Simulation Results

MC Realizations	Mean (°C)	Standard Deviation (°C)
1000	72.69	2.31
5000	72.61	2.20
10000	72.61	2.24
50000	72.58	2.22
100000	72.59	2.22
500000	72.60	2.22

Based on the 50,000 realizations, Figure 4.20 shows the histogram of the MC results. All of the realization is below the temperature constraint of 100°C, the highest probability of realization is between 72-74°C.

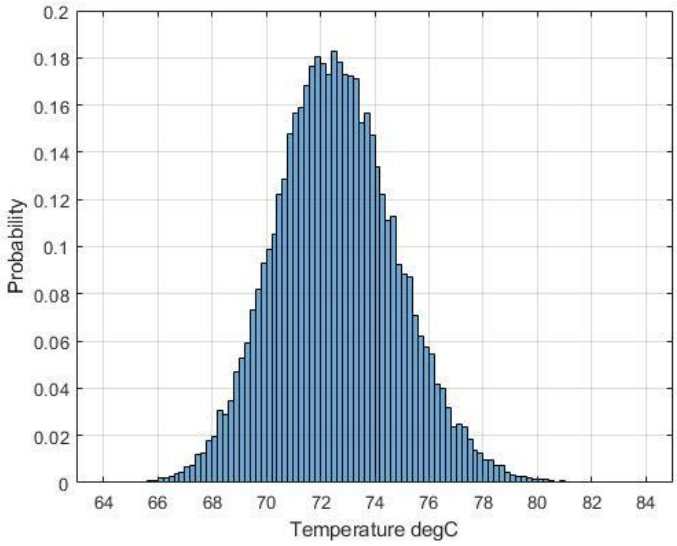


Figure 4.20 Histogram of 50,000 MC Realization

Figure 4.21 shows the empirical cumulative distribution function of the MC realization. For a temperature constraint of 100°C, the probability of meeting the constraint is 100%, which implies that the probability of failure is 0% based on the above assumptions and surrogate models.

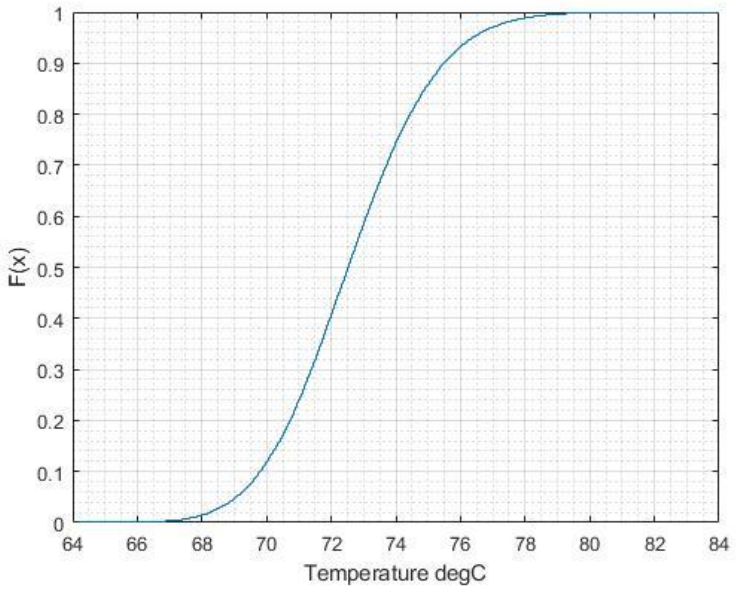


Figure 4.21 CDF for 50,000 MC Realization

Figure 4.22 shows a comparison of the MC histogram with a normal distribution. In comparison, the MC realization follows the normal distribution tightly with very slight kurtosis to the left.

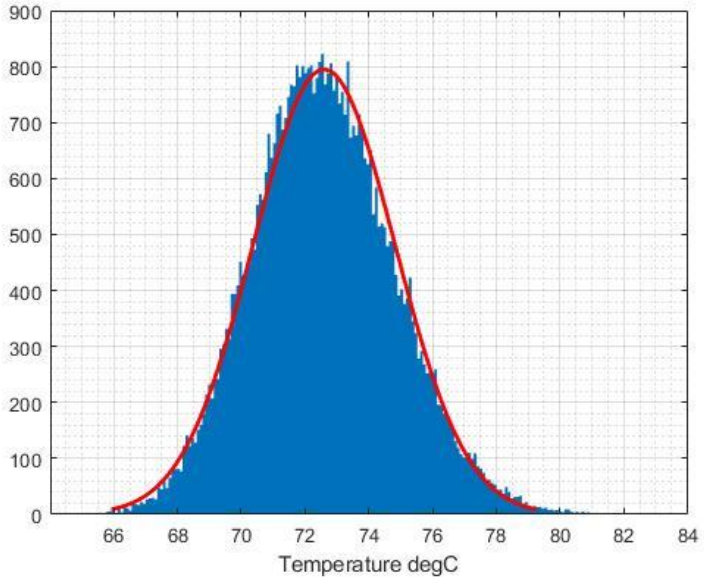


Figure 4.22 Histogram & Normal Distribution for 50,000 MC realization

The uncertainty analysis here is a demonstration of the possible application of surrogate model. In this case, if 10 values are sampled for the three variables shown here, an exhaustive run on all the possible

combination of these three variables will take 1000 COMSOL simulations. Using the surrogate model only 75 simulations are run in this experiment. The reduction in computational time is significant. With the use of surrogate model, the uncertainty analysis using Monte Carlo Simulation can be run in a fraction of a second. On the other hand, running such simulation in COMSOL without cluster computer is likely infeasible. In the future, when site-specific data and the coefficient of variation becomes available, the surrogate-based method can be employed similarly in determining the reliability of the design that will meet the predetermined standard.

4.5 Summary

The experiment shown here displays that the surrogate model does reduce the number of necessary simulations for solving optimization and uncertainty analysis problems. In general, the surrogate function can produce accurate enough prediction near the sampling space of the solution. In the case of spacing design and design optimization, the surrogate model can find the solution with around 20 simulations. However, as the sampling domain increases, the amount of necessary expensive simulation increases, and the accuracy of surrogate prediction decreases, without an appropriate higher order surrogate function for more complex problem, the surrogate model prediction may not be as accurate.

In the problem of design optimization where only a certain local area needs to be evaluated, the surrogate-based method provides the optimal design with reasonably accurate temperature prediction using a small number of COMSOL evaluations. Due to the fact that the true response surface on the maximum temperature vs. spacing parameter is a smooth concave function, the surrogate function used here can be fit to the surface easily and produce accurate prediction. In the case of uncertainty analysis, a higher dimensional sampling space result in a less accurate prediction made by the surrogate. Since the true response surface of this problem may not be a smooth concave function similar to the spacing design problem, the surrogate model in this experiment requires more sampling points. However, in this experiment, the surrogate model still demonstrates its ability in significantly reducing the necessary simulations with only a slight sacrifice in its prediction accuracy.

In short, the surrogate performs the best when the fitting is only required in a localized area and when the surrogate function is capable of fitting the true response surface. When the sampling space increases, or the complexity of the problem increases, the accuracy of the surrogate will deteriorate if it is not appropriately implemented. However, in any of the three experiments shown above, the surrogate-based method is proven to provide meaningful result with relatively few expensive model evaluations.

Chapter 5
Conclusion

The earlier chapters explored the application of surrogate-based approach in solving the optimization problem of DGR spacing as well as examining the thermal behavior given a DGR design.

As shown in Chapter 2 and Chapter 3, the finite element models with a proper set up and assumptions can provide a reasonable estimation of the temperature behavior of the repository. Given that the assumptions of material and boundary input data as well as geometry are reasonably accurate, the numerical model can be a useful tool for examining and predicting the behavior of the DGR design. Based on the material properties assumptions and design configuration, the numerical model predicts that the maximum temperature will be 66°C in year 47 after the used fuel is placed into the repository. This prediction is in line with the similar model shown in the literature review. However, as shown in the sensitivity analysis, since the simulation period and the desired lifetime is so long, a slight change in the initial condition will have a significant effect on the simulation outcome. The sensitivity analysis shows that different input assumptions will affect the temperature behavior differently. Some inputs have a significant influence on the temperature profile, such as rock properties. The sensitivity analysis also shows that some inputs' effect on temperature will diminish over time such as geothermal gradient and surface temperature, while some will influence the temperature more as time progress such as density and heat capacity of rock. Therefore, in the design and siting process of the DGR the more influential material and its properties should be examined carefully to ensure the design's reliability over its long lifetime. Even though the finite element model is a simplified version of the reality, it still provides meaningful results and guidance for the thermal behavior of the repository.

The experimental problems solved by the surrogate modeling approach demonstrates that for a computationally heavy problem like the one shown here, the surrogate-based approach can substantially reduce the necessary simulations compared to an exhaustive simulation approach. Using the surrogate-based optimization, the design problem can be solved with a small number of simulations in comparison to its sampling space. However, the surrogate model needs to be carefully designed for different types of problems since the surrogate function, sampling domain and techniques and convergence criteria all have a significant influence on the accuracy of the solution and computational time needed. For high dimensional problems, the functional selection, sampling region and convergence criteria all play an important role on the accuracy of the surrogate since a low dimensional surrogate will not be capable

of fitting a high dimensional function properly. The surrogate-based approach also shows the versatility for a large variety of problems when the problem involves heavy computation and multidimensionality, utilizing the approximation and predictive ability of the surrogate function, the solution time can be improved by orders of magnitudes.

In the future, the surrogate approach can be utilized for probabilistic design optimization of the spacing when site-specific information becomes available. With a combination of first or second order reliability method, the surrogate model can be used to reduce the necessary computation for solving a more complete reliability-based design problem. For the purpose of fitting the surrogate function to a high dimensional space, using surrogate function such as neural network or convolution neural network may improve the accuracy when facing high dimensional problems. The surrogate model can also incorporate other module of physics such as mass transport and chemical reactions to provide a better probabilistic analysis of the DGR design. To further speed up the solution time, parallel computing can also be employed within the algorithm.

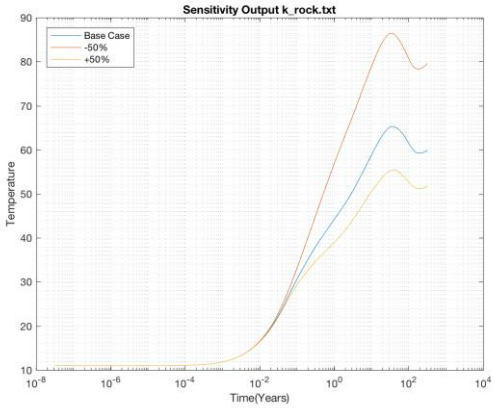
Appendix A

Heat Generation Input used in Simulations [3]

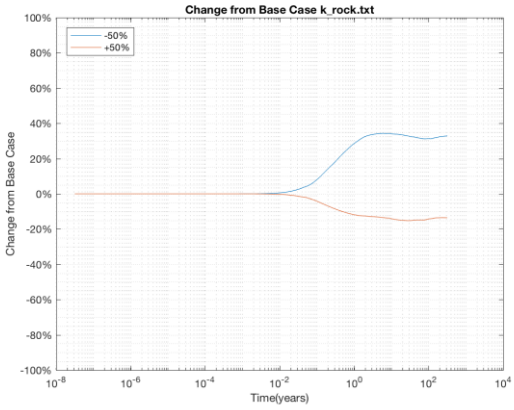
Seconds	Wattage	Seconds	Wattage	Seconds	Wattage
0	84.546	**2.84E+09**	27.77027	**1.57E+11**	4.62
19710000	83.67975	**3.31E+09**	24.9942	**2.05E+11**	4.119514
39420000	82.8135	**3.53E+09**	24.05636	**2.52E+11**	3.732302
78840000	81.081	**3.78E+09**	23.0538	**3.14E+11**	3.32178
1.58E+08	77.616	**3.94E+09**	22.53543	**3.78E+11**	2.948311
2.37E+08	73.83246	**4.1E+09**	22.0374	**4.73E+11**	2.498632
3.15E+08	71.148	**4.57E+09**	20.94664	**6.30E+11**	1.92192
3.94E+08	68.09692	**5.36E+09**	19.3578	**7.88E+11**	1.571803
4.73E+08	65.604	**6.31E+09**	18.31908	**1.10E+12**	1.04874
5.52E+08	63.12843	**7.25E+09**	17.42581	**1.32E+12**	0.849573
6.31E+08	60.984	**8.51E+09**	16.401	**1.58E+12**	0.66066
7.1E+08	58.5454	**1.1E+10**	15.01672	**2.37E+12**	0.393099
7.88E+08	56.364	**1.32E+10**	14.04418	**3.15E+12**	0.203742
8.67E+08	54.43188	**1.48E+10**	13.4442	**4.10E+12**	0.155432
9.46E+08	52.668	**1.73E+10**	12.55219	**5.05E+12**	0.117243
1.1E+09	48.97941	**2.05E+10**	11.60418	**6.31E+12**	0.07623
1.26E+09	45.7842	**2.52E+10**	10.42586	**7.88E+12**	0.070224
1.42E+09	42.966	**3.06E+10**	9.3324	**1.58E+13**	0.068376
1.58E+09	40.425	**3.47E+10**	8.807518	**3.15E+13**	0.068376
1.73E+09	38.17891	**4.1E+10**	8.11034		
1.89E+09	36.1284	**5.05E+10**	7.243785		
2.05E+09	34.3292	**6.21E+10**	6.3756		
2.21E+09	32.6634	**7.88E+10**	5.92359		
2.37E+09	31.23117	**1.07E+11**	5.340244		
2.52E+09	29.8914	**1.26E+11**	5.031921		

Appendix B

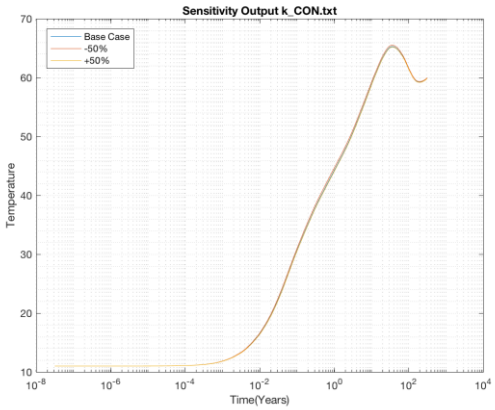
Complete Sensitivity Analysis Results



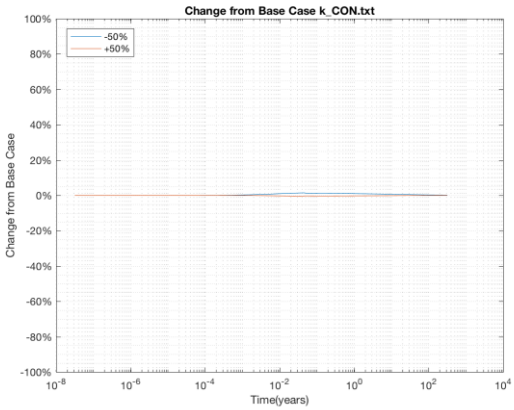
(a) Temp. Sensitivity for k_rock



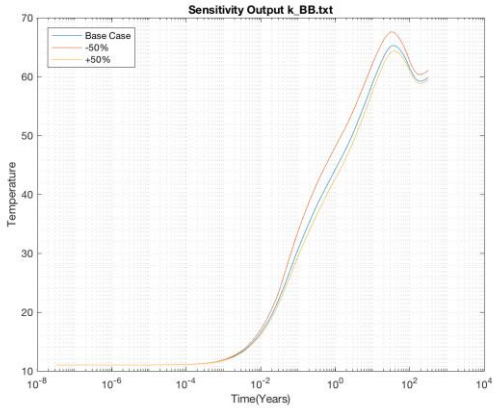
(b) Temp. Percentage Change of k_rock



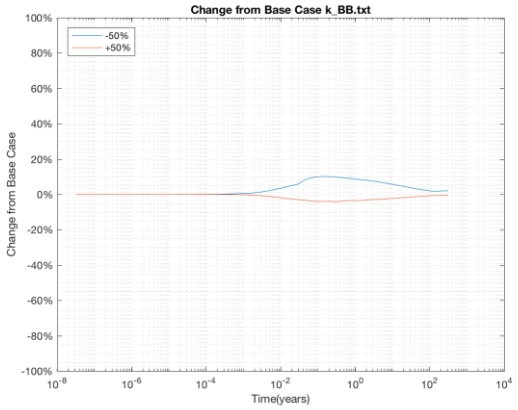
(c) Temp. Sensitivity for k_CON



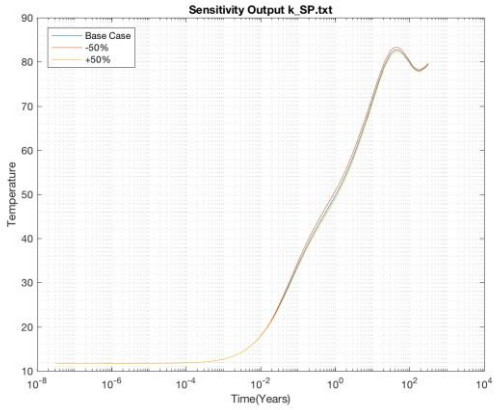
(d) Temp. Percentage Change of k_CON



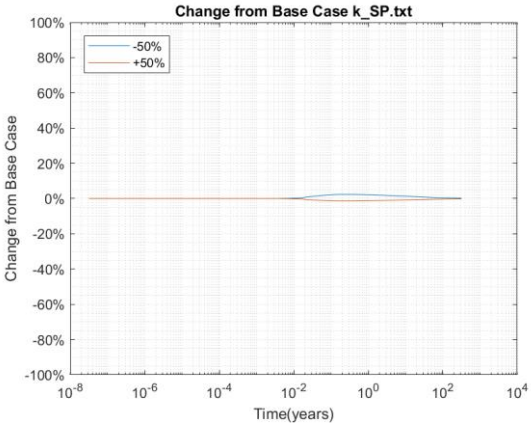
(e) Temp. Sensitivity for k_BB



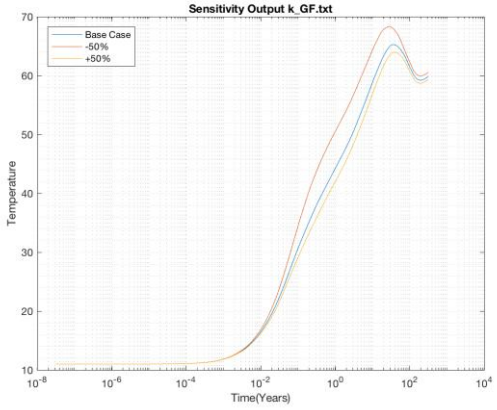
(f) Temp. Percentage Change of k_BB



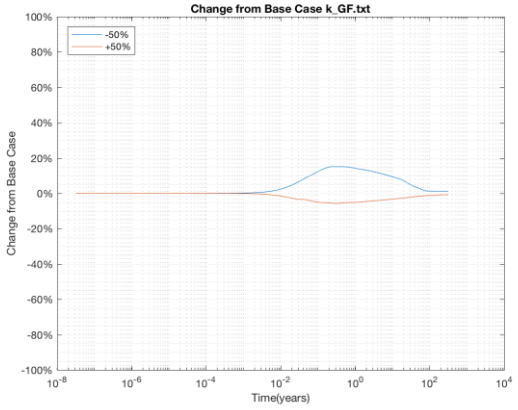
(g) Temp. Sensitivity for k_SP



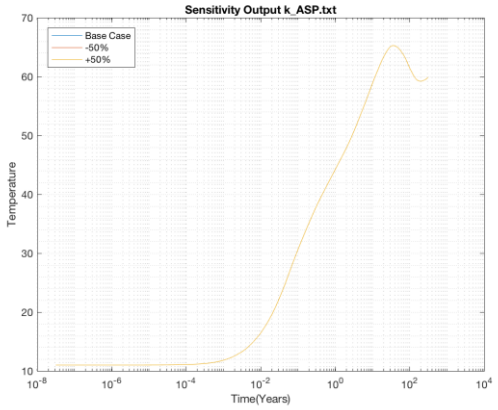
(h) Temp. Percentage Change of k_SP



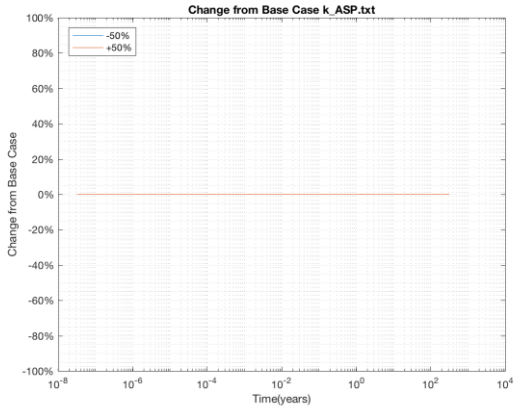
(i) Temp. Sensitivity for k_GF



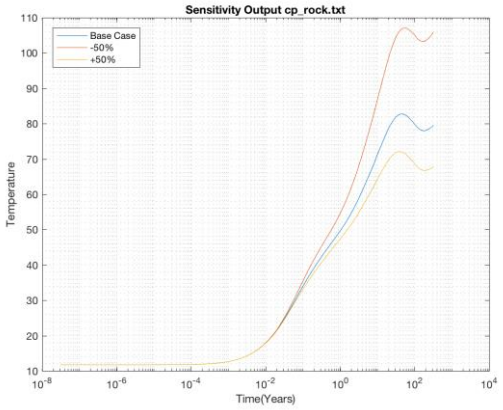
(j) Temp. Percentage Change of k_GF



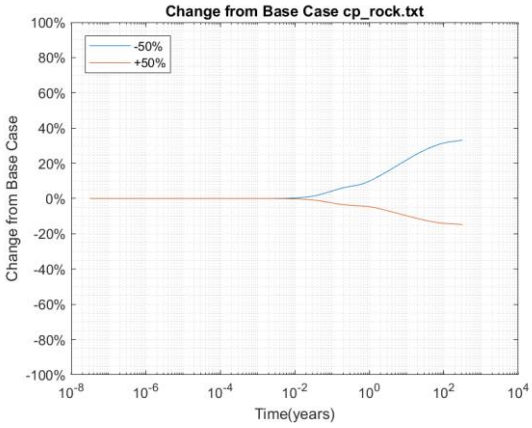
(k) Temp. Sensitivity for k_ASP



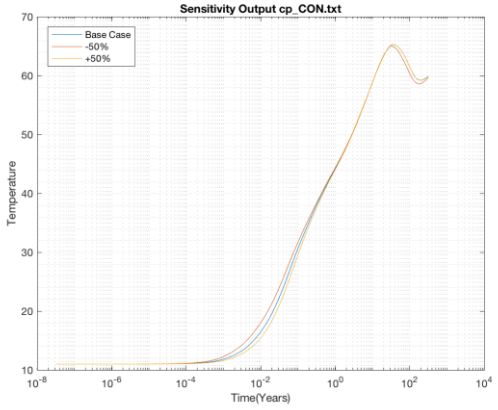
(l) Temp. Percentage Change of k_ASP



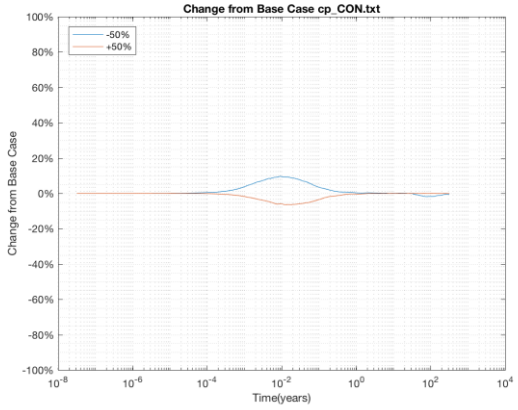
(m) Temp. Sensitivity for cp_rock



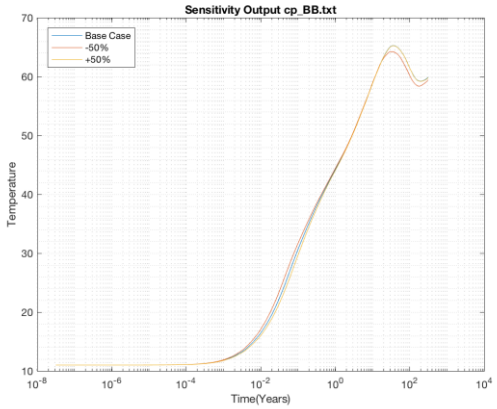
(n) Temp. Percentage Change of cp_rock



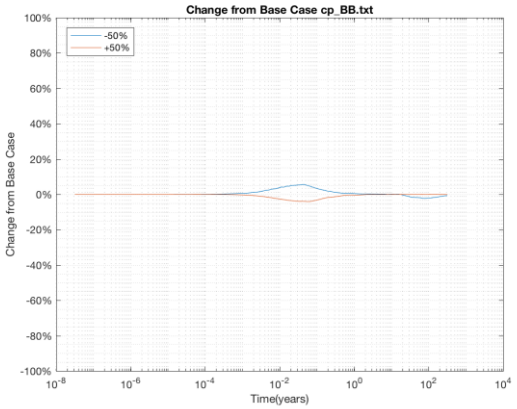
(o) Temp. Sensitivity for cp_CON



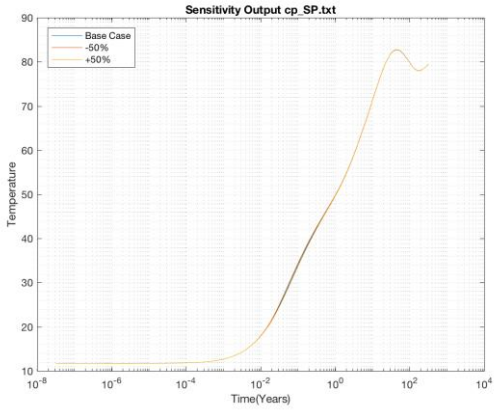
(p) Temp. Percentage Change of cp_rock



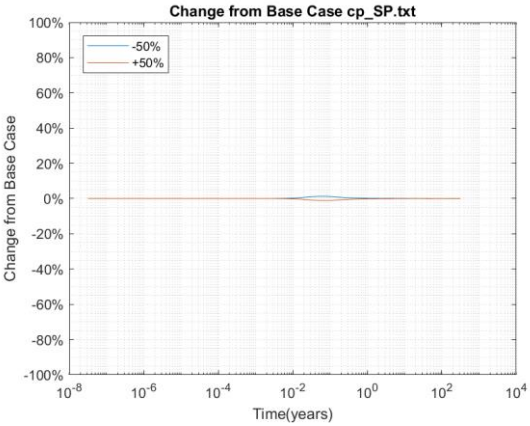
(q) Temp. Sensitivity for cp_BB



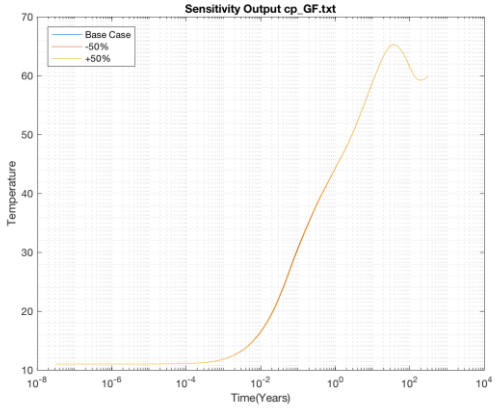
(r) Temp. Percentage Change of cp_BB



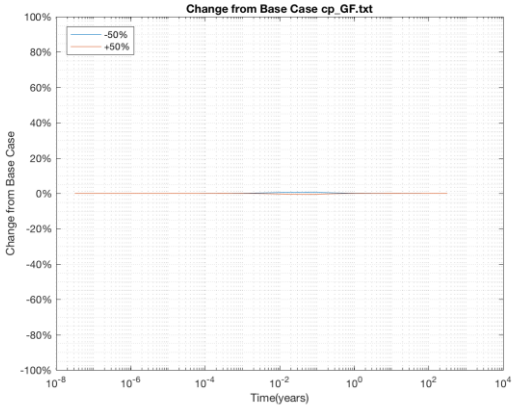
(s) Temp. Sensitivity for cp_SP



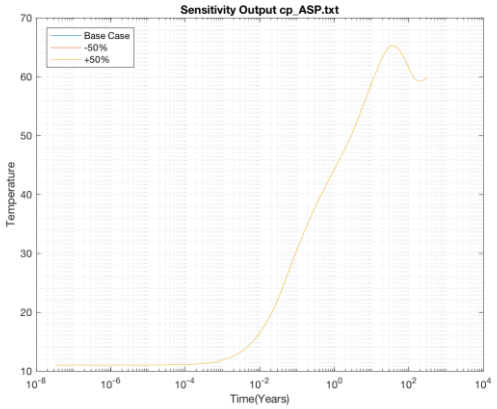
(t) Temp. Percentage Change of cp_SP



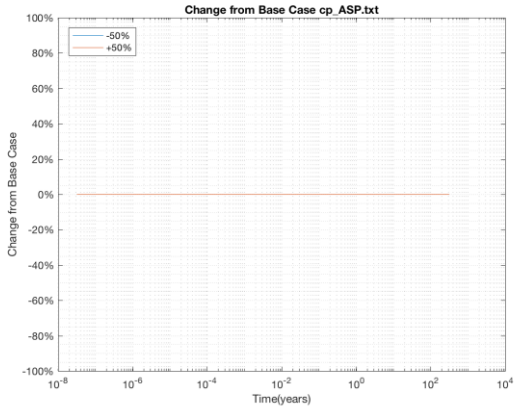
(u) Temp. Sensitivity for cp_GF



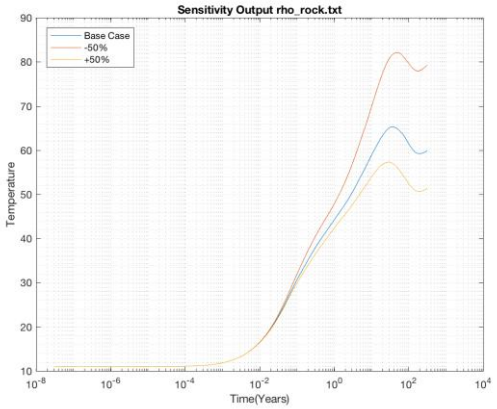
(v) Temp. Percentage Change of cp_GF



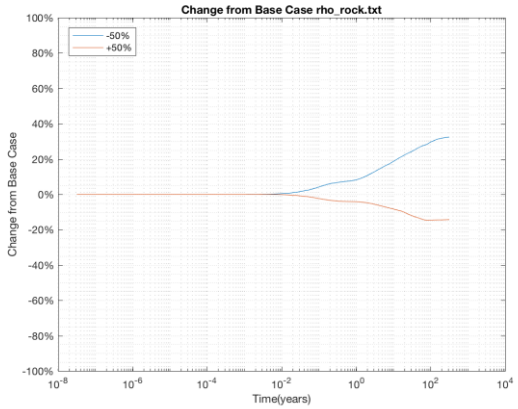
(w) Temp. Sensitivity for cp_ASP



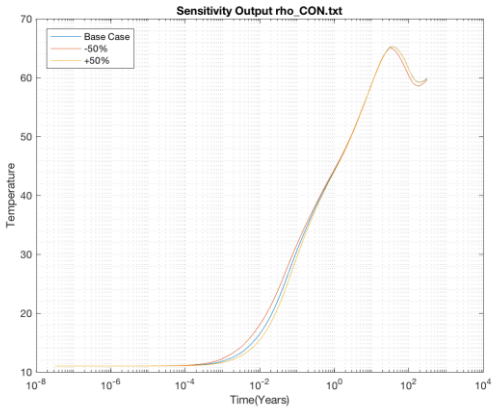
(x) Temp. Percentage Change of cp_ASP



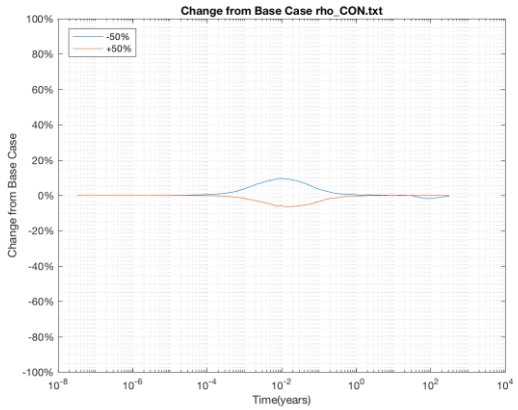
(y) Temp. Sensitivity for rho_rock



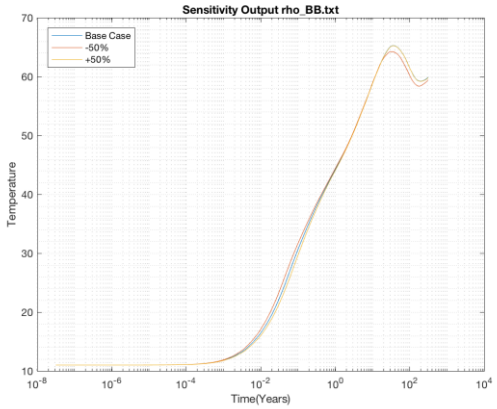
(z) Temp. Percentage Change of rho_rock



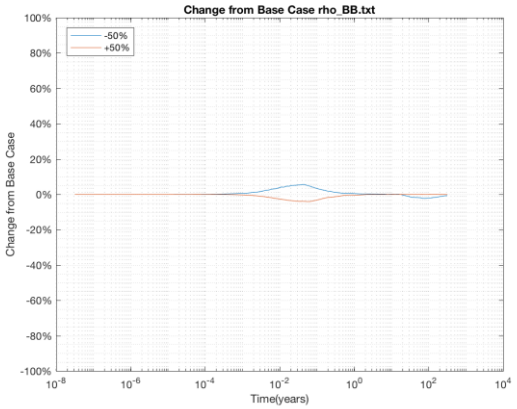
(aa) Temp. Sensitivity for rho_CON



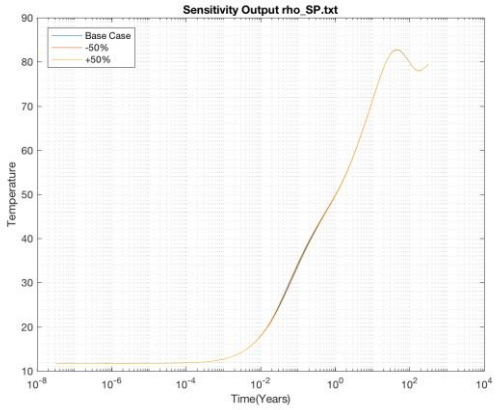
(bb) Temp. Percentage Change of rho_CON



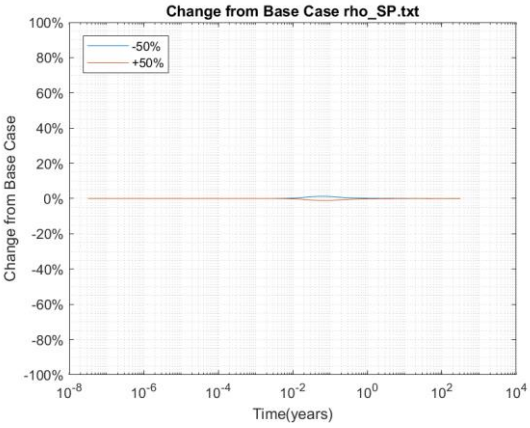
(cc) Temp. Sensitivity for rho_BB



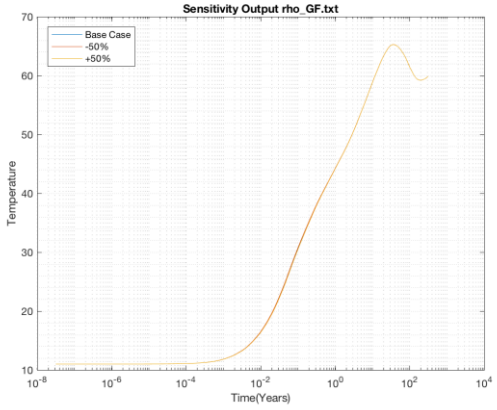
(dd) Temp. Percentage Change of rho_BB



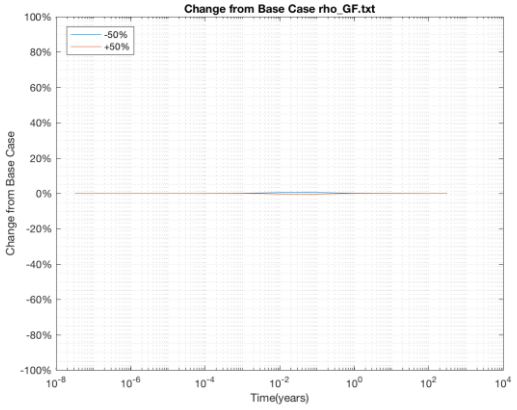
(ee) Temp. Sensitivity for rho_SP



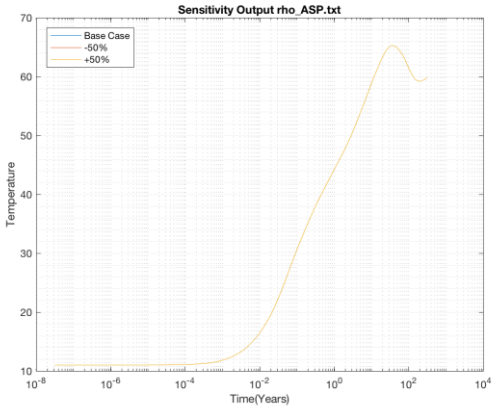
(ff) Temp. Percentage Change of rho_SP



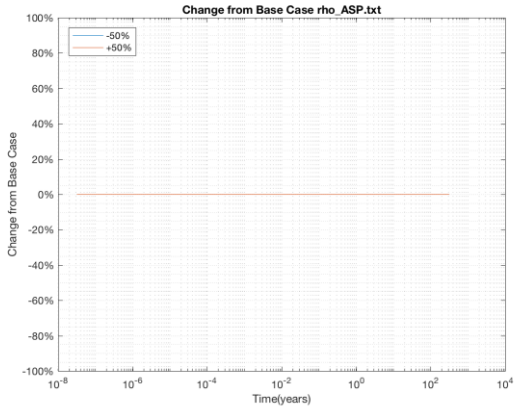
(gg) Temp. Sensitivity for rho_GF



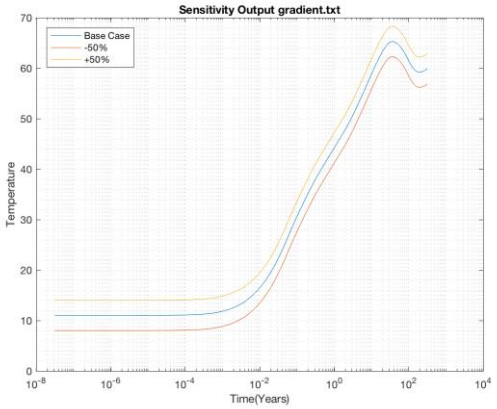
(hh) Temp. Percentage Change of rho_GF



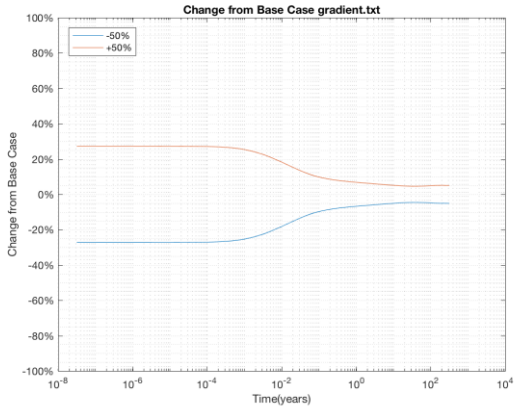
(ii) Temp. Sensitivity for rho_ASP



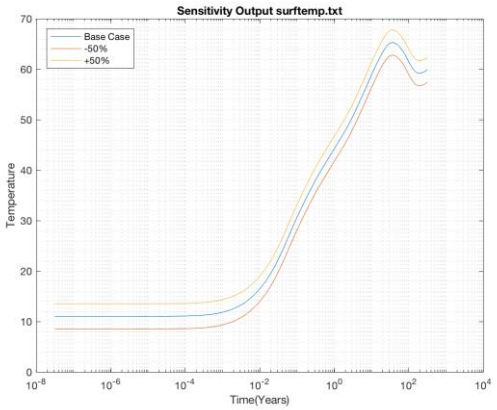
(jj) Temp. Percentage Change of rho_ASP



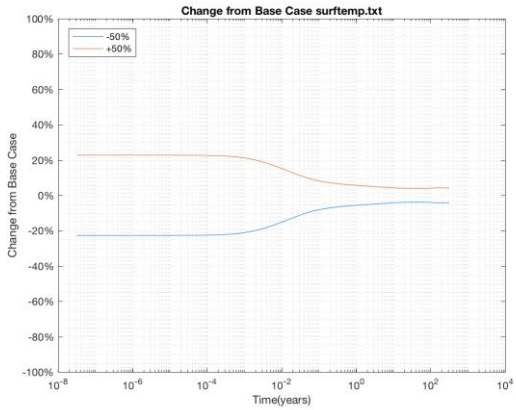
(kk) Temp. Sensitivity for gradient



(ll) Temp. Percentage Change of gradient



(mm) Temp. Sensitivity for surftemp



(nn) Temp. Percentage Change of surftemp

References

[1] Nuclear Waste Management Organization, "What is Used Nuclear Fuel?," 2016. [Online]. Available:https://www.nwmo.ca/~/media/Site/Files/PDFs/2016/11/10/12/38/EN_Backgrounder_UsedNuclearFuel_LowRes.ashx?la=en. [Accessed: 01-Aug-2018].

[2] D. Heimlich, "APM 2016 Lifecycle Cost Estimate Update Summary Report," Nuclear Waste Management Organization, Apr. 2016.

[3] R. Guo, "Thermal Response of a Mark II Conceptual Deep Geological Repository in Crystalline Rock," NWMO, Technical Report, 2016.

[4] R. Guo, "Thermal response of a Canadian conceptual deep geological repository in crystalline rock and a method to correct the influence of the near-field adiabatic boundary condition," *Eng. Geol.*, vol. 218, pp. 50–62, Feb. 2017.

[5] S. Koziel, D. E. Ciaurri, and L. Leifsson, "Surrogate-based methods," in *Computational optimization, methods and algorithms*, Springer, 2011, pp. 33–59.

[6] "Max-Min Surrogate-Assisted Evolutionary Algorithm for Robust Design." .

[7] R. Tripathy and I. Bilionis, "Deep UQ: Learning deep neural network surrogate models for high dimensional uncertainty quantification," *ArXiv Prepr. ArXiv180200850*, 2018.

[8] M. Kamali, "Calibration of hydrologic models using distributed surrogate model optimization techniques: A WATCLASS Case Study," 2009.

[9] U. Abubakar, S. Sriramula, and N. C. Renton, "A Hybrid Method for Stochastic Performance Modeling and Optimization of Chemical Engineering Processes," *Chem. Eng. Commun.*, vol. 202, no. 2, pp. 217–231, Feb. 2015.

[10] Y. Jin, "Surrogate-assisted evolutionary computation: Recent advances and future challenges," *Swarm Evol. Comput.*, vol. 1, no. 2, pp. 61–70, Jun. 2011.

[11] A. Giunta, S. Wojtkiewicz, and M. Eldred, "Overview of Modern Design of Experiments Methods for Computational Simulations (Invited)," 2003.

[12] M. D. Mckay, R. J. Beckman, and W. J. Conover, "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, vol. 42, no. 1, p. 55, 1979.

[13] J. Mueller, "MATSuMoTo: The MATLAB Surrogate Model Toolbox For Computationally Expensive Black-Box Global Optimization Problems," *ArXiv14044261 Math*, Apr. 2014.

[14] J. Müller and R. Piché, "Mixture surrogate models based on Dempster-Shafer theory for global optimization problems," *J. Glob. Optim.*, vol. 51, no. 1, pp. 79–104, Sep. 2011.

[15] P. K. Mishra, S. K. Nath, M. K. Sen, and G. E. Fasshauer, "Hybrid Gaussian-cubic radial basis functions for scattered data interpolation," *Comput. Geosci.*, May 2018.

[16] J. Muller, C. A. Shoemaker, and R. Piche, "SO-MI: A Surrogate Model Algorithm for Computationally Expensive Nonlinear Mixed-Integer Black-Box Global Optimization Problems Online Supplement," p. 33.

[17] J. Müller, C. A. Shoemaker, and R. Piché, "SO-I: a surrogate model algorithm for expensive nonlinear integer programming problems including global optimization applications," *J. Glob. Optim.*, vol. 59, no. 4, pp. 865–889, Aug. 2014.

[18] J. McCall, "Genetic algorithms for modelling and optimisation," *J. Comput. Appl. Math.*, vol. 184, no. 1, pp. 205–222, Dec. 2005.

[19] K. Deb, "Multi-Objective Optimization Using Evolutionary Algorithms: An Introduction," p. 24.

[20] M. Y. Fattah and A. H. S. Al-Lami, "Behavior and characteristics of compacted expansive unsaturated bentonite-sand mixture," *J. Rock Mech. Geotech. Eng.*, vol. 8, no. 5, pp. 629–639, Oct. 2016.

[21] W. M. Ye, M. Wan, B. Chen, Y. G. Chen, Y. J. Cui, and J. Wang, “Temperature effects on the swelling pressure and saturated hydraulic conductivity of the compacted GMZ01 bentonite,” *Environ. Earth Sci.*, vol. 68, no. 1, pp. 281–288, Jan. 2013.
[22] W. M. Ye, Y. He, Y. G. Chen, B. Chen, and Y. J. Cui, “Thermochemical effects on the smectite alteration of GMZ bentonite for deep geological repository,” *Environ. Earth Sci.*, vol. 75, no. 10, May 2016.
[23] D. Sun, W. Sun, and L. Fang, “Swelling characteristics of Gaomiaozi bentonite and its prediction,” *J. Rock Mech. Geotech. Eng.*, vol. 6, no. 2, pp. 113–118, Apr. 2014.
[24] W. Sun, Z. Wei, D. Sun, S. Liu, B. Fatahi, and X. Wang, “Evaluation of the swelling characteristics of bentonite–sand mixtures,” *Eng. Geol.*, vol. 199, pp. 1–11, Dec. 2015.
[25] “2017 Data Clearance Form.” Nuclear Waste Management Organizations, 2017.
[26] “Average bentonite price United States 2017 | Statistic,” *Statista*. [Online]. Available: https://www.statista.com/statistics/248186/average-bentonite-price/. [Accessed: 09-Aug-2018].
[27] O. Stephansson and P. Gipper, “2015:01 Rock Mechanics - Thermal properties and thermal modelling of the rock in a repository of spent nuclear fuel at Forsmark,” p. 59.
[28] P. Hartlieb, M. Toifl, F. Kuchar, R. Meisels, and T. Antretter, “Thermo-physical properties of selected hard rocks and their relation to microwave-assisted comminution,” *Miner. Eng.*, vol. 91, pp. 34–41, May 2016.
[29] P. Abootalebi and G. Siemens, “Thermal properties of engineered barriers for a Canadian deep geological repository,” *Can. Geotech. J.*, vol. 55, no. 6, pp. 759–776, Jun. 2018.
[30] L. Eppelbaum, I. Kutasov, and A. Pilchin, “Thermal Properties of Rocks and Density of Fluids,” in *Applied Geothermics*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 99–149.
[31] P. D. Justesen, “Multi-objective Optimization using Evolutionary Algorithms,” p. 36.