

UNDERSTANDING THE IMPACT OF BOUNDARY AND INITIAL CONDITION ERRORS ON THE SOLUTION TO A THERMAL DIFFUSIVITY INVERSE PROBLEM

X. FU¹, B. LEVENTHAL²

¹ Department of Mathematics, Clarkson University, fux@clarkson.edu; ² Department of Mathematics, Clarkson University, leventbc@clarkson.edu

ABSTRACT

In this work, we consider simulation of heat flow in the shallow subsurface. As sunlight heats up the surface of soil, the thermal energy received dissipates downward into the ground. This process can be modeled using a partial differential equation known as the heat equation. The spatial distribution of soil thermal conductivities is a key factor in the modeling process. Prior to this study, temperature profiles were recorded at different depths at various times. This work is motivated by trying to match these temperature profiles using a simulation-based approach in the context of an inverse problem. Specifically we determine soil thermal conductivities using derivative-free optimization to minimize the nonlinear-least square errors between simulation and data profiles. We also study how errors in the initial and boundary conditions propagate over time using numerical approach.

1. INTRODUCTION

The one dimensional heat equation is commonly used to model the temperature distribution of a linear heat-conducting system. Knowing the spatially distributed thermal conductivity as well as its initial and boundary temperature, one can approximate the temperature at any given position at any given time in the system. In this study, we are solving the inverse of this problem. Given the initial and boundary condition as well as the temperature profiles of the system at all positions and time, our goal is to estimate the thermal conductivities of the linear system.

We will use subsurface temperature data that have been monitored and logged in time and space at a meteorology station. In the given data set, soil temperature is measured at 0cm (top surface), 1cm, 5cm, 10cm, 15cm, 20cm, 25cm and 30cm below the surface and the measurement is taken every 5 minutes for 7 days.

The proposed work will:

- identify, using the data, the best-fit thermal conductivities in the subsurface system;
- analyze how choices made on various aspects of the inverse solution procedure impact the estimation results;
- study the impact of errors in data on the accuracy of estimation.

This paper is outlined in two main parts as follows. In the first part, we began by introducing the general framework for the identification process of thermal conductivities. Next, we discussed implementation details and challenges of the numerical estimation process and

Date: October 17, 2011.

presented results. The second part consists of a sensitivity study of the heat model for the heterogeneous systems. The sensitivity study evaluates how inevitable errors from data impacts the estimation results, thus assisting in determining the reliability of the approximated parameters.

2. METHODOLOGY

2.1. Homogeneous Heat Transport Model. The basic model that describes one dimensional, transient heat transport in homogeneous system is shown in Eq.(1):

$$\frac{\partial T}{\partial t} = K \frac{\partial^2 T}{\partial z^2}, \quad (1)$$

where temperature, denoted as T , is a function of time t and space z . The thermal diffusivity is usually referred to as K [L/T] and is related to thermal conductivity by:

$$K = \frac{\hat{k}}{\rho c}, \quad (2)$$

where \hat{k} is the thermal conductivity, ρ is the density and c is the heat capacity.

To solve this system, one also needs an initial condition, $T(t = 0, z)$, a top boundary condition, $T(t, z = 0)$, and a bottom boundary condition, $T(t, z \rightarrow \infty)$. This is a second order partial differential equation with respect to time and space. The analytic solution of this heat fl w model has been thoroughly studied since the 1800s [2]; its solution, given a variety of initial and boundary conditions, can be found in current literature such as [1] and [5]. For instance, if we consider the following boundary conditions:

$$T(t, 0) = T_a + A \sin(\omega t + \phi) \quad (3)$$

and

$$\lim_{z \rightarrow \infty} T(t, z) = T_a \quad (4)$$

an analytic solution to Eq.(1) is given by

$$T(t, z) = T_a + A e^{-\frac{z}{D}} \sin[\omega t - \frac{z}{D} + \phi], D = \sqrt{\frac{2K}{\omega}}. \quad (5)$$

Eq.(3) indicates a sinusoidal temperature pattern at the top of the soil column where the sunlight directly impacts. Eq.(4) implies that when it is sufficientl deep below the surface, the soil temperature converges to the average subsurface temperature T_a and is independent of time.

For a homogeneous system, the thermal conductivity of the soil column is a constant value K throughout the domain. In our study of a heterogeneous system however, the thermal conductivity can vary by depth and is a mapping from the seven spatial intervals in Z to seven values of K : $\{K_i \mid i \in Z\}$, where $Z = \{(0 \sim 1), (1 \sim 5), (5 \sim 10), (10 \sim 15), (15 \sim 20), (20 \sim 25), (25 \sim 30)cm\}$.

2.2. Heterogeneous Heat Transport Model. Heat conduction in a heterogeneous soil system can be described using a slight variation of Eq.(1):

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial z} [K \frac{\partial T}{\partial z}] = K \frac{\partial^2 T}{\partial z^2} + \frac{\partial K}{\partial z} \frac{\partial T}{\partial z}, \quad (6)$$

where $K = K(z)$ is a function of spatial variable z . Finite differences were used to discretize Eq.(6) and a centered difference is used for $\frac{\partial T}{\partial z}$. The numerical solver has been validated using

a test problem where a forcing term $f(z, t)$ is added to the right side and the function $K(z)$ is known. Since we only consider K values at seven locations, we use a spline over space to create a continuous function where K is specified at all locations. The derivative of the continuous spline function is then used to account for the $\frac{\partial K}{\partial z}$ term in Eq.(6).

2.3. Estimating Thermal Conductivities. In this project, discrete values of the temperature at 5 min time intervals are already known while the thermal conductivities K are unknown. Figure 1 plots the given temperature data over time and space. The temperatures range from about 13 to 67 °C in the first 24 hours.

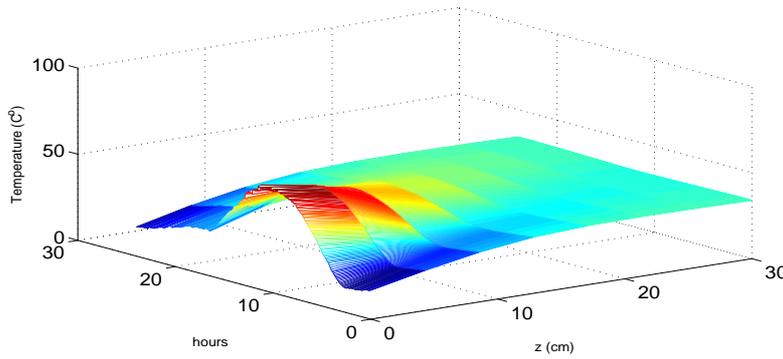


FIGURE 1. Temperature data

We cast this problem to an optimization frame where the least-squares error (LSE) between data and simulated temperature profile is minimized. Specifically, the logged data T^{obs} is a $N_t \times N_z$ matrix where N_t is the number of time points and N_z is the number of spatial nodes; the simulated temperature profile $\hat{T}(K)$ is obtained by numerically solving the heat equation where the boundary conditions come from data and K is a guessed vector of conductivity values at each iteration of the optimization scheme. We pose this idea mathematically as:

$$\min_{K \in \Omega} J(K) = \frac{1}{2} \sum_{i=1}^N w_i (\hat{T}_i(K) - T_i^{obs})^2, \quad (7)$$

where Ω represents reasonable bound constraints on K , w_i are algorithmic weights, and N represents the number of data points that we compare at.

For this project, instead of using a traditional LSE calculation as described in Eq.(7), a scaled LSE calculation is used for most part of the study (except for Sec. 4.4) for fair comparisons across trials. We denote the scaled LSE with **sLSE**, which is calculated as:

$$sLSE = \frac{1}{2} \frac{\sum_{i=1}^N w_i (\hat{T}_i(K) - T_i^{obs})^2}{\sum_{i=1}^N (T_i^{obs})^2}, N_t = 1, \quad (8)$$

where N_t is the number of time points. For the majority of this study, error is calculated only at a specific time, meaning N_t is set to 1. In Sec.4.4, N_t is larger than 1 and a new error calculation is introduced. Here, the normal LSE is scaled by the summation of squared data values.

2.4. Parameter Identification Using Optimization. Figure 2 shows the structure of the optimization scheme for a heterogeneous system. At each iteration, the optimizer will, based on the bounds Ω and previous function evaluations, pick a set of six K values as input for the numerical solver for the heat equation that outputs the temperature profile. The simulated profile is then compared to actual data to obtain the error at current iteration. The optimization terminates when the error becomes sufficiently small and the current set of K values is a potential optimal solution. Both the genetic algorithm ([4]) and the implicit filtering ([3]) are used as optimizers in this process.

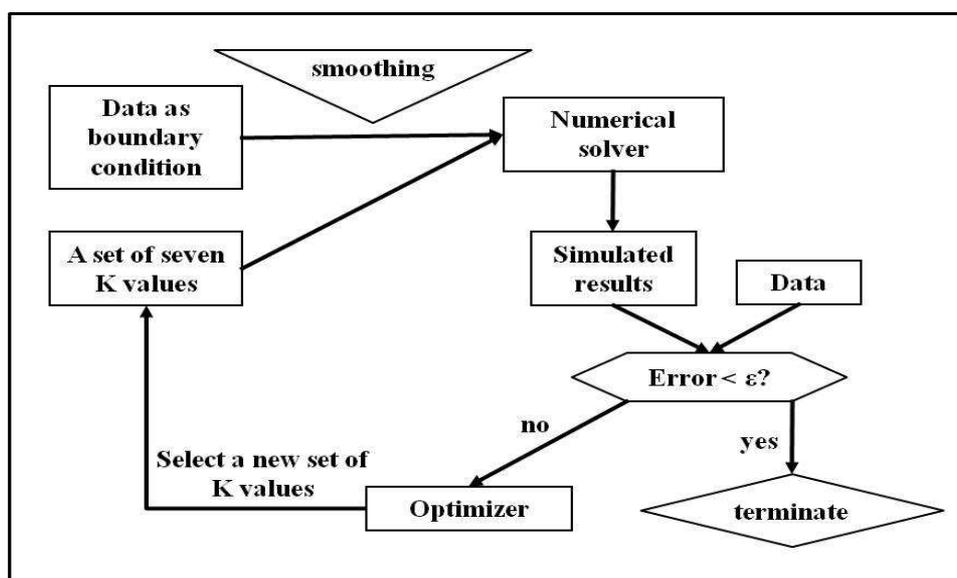


FIGURE 2. Structure of the Optimization Scheme

3. PRELIMINARY RESULTS

With the working simulation tool in hand, we fit the data profile in all layers at the 24th hour by optimizing the set of conductivities at six layers described in Sec. 2.1. Since the sunlight strength oscillates on a cycle of 24 hours, the data over each 24-hour interval is assumed to be similar. Therefore, to save computation time of the study, we only used data from the first day. In this first attempt, we also assumed that the conductivity is a linear function over space in each layer. Thus, to create a continuous function $K(z)$, we simply used a linear spline curve-fitting technique. We enforced 70cm to be the bottom depth and assumed that the conductivity at layer 0-1cm and 30cm-70cm (bottom) are constant and are equal to K_{1-5cm} and $K_{25-30cm}$ respectively. In our simulation, $\Delta z = 0.1cm$ and $\Delta t = 0.1$ minutes. Table 1 shows the optimal conductivity values obtained for each depth using genetic algorithm when the results are optimized at the 24th hour data. The last two rows show the sLSE (see Eq.8 for definition of sLSE) and the maximum temperature difference (E) at the final time. These results are only preliminary and are shown here for comparison with later results by an enhanced simulator.

TABLE 1. Preliminary optimization results

Depth	K (cm/min)
1 cm	3.2809e-002
5 cm	3.1440e-001
10 cm	2.6486e-002
15 cm	3.9509e-001
20 cm	1.9201e-001
25 cm	9.8476e-003
30 cm	2.6704e-001
Temp ($^{\circ}C$)	35.33
sLSE	5.9633e-005
E ($^{\circ}C$)	0.895

4. EXPERIMENTATION ON IMPLEMENTATION DETAILS

The above results inspired a series of studies in regards to the choices made in the original modeling and optimization process to understand how these choices impact the optimal solution and its simulation fitting at intermediate time points. Below we summarized the numerical experiments that have contributed to the improvements of the fitting results:

Section 4.1: Assumptions on Conductivities

In this study, we explored how different implementations on the heterogeneity of the system influence the stability of the simulator. Specifically, we tested the simulator with K vectors that are 1) linearly connected; 2) cubic splined; 3) interpolated using a hybrid method. The third method proved to be most suitable for this problem.

Section 4.2: Using 30cm as Bottom Boundary Condition

In our first attempt, the simulator used 70cm as the bottom depth and assumed a constant temperature as the bottom boundary condition. In this experiment, we leveled up the bottom depth to 30cm, where the bottom boundary condition is available from recorded data.

Section 4.3: Top Surface Data Elimination

The micro environment can change constantly around the top surface sensor due to human activities and natural impacts such as rainfalls. Such change can introduce bad noises into the data at 0cm and 1cm. In this study, we avoided using data at 0cm and 1cm and only optimize from 5cm to 30cm.

Section 4.4: Guided Optimization: Error Calculation Over All Time Points

In our first attempt, the optimizer only minimizes the error at the final time of simulation. However, the objective of this work is to determine a set of thermal conductivities that simulates temperature profile that fits the data throughout the time domain. In this study, we re-designed the error calculation formula so that it incorporates fitting error at other times as well.

These details have proven to be effective in our effort of tailoring the simulator and optimization framework to its best ability to make accurate estimations on the thermal conductivities. The following sections are dedicated to the specifics of the changes and their results.

4.1. Assumptions on Conductivities. Initially, to model the heterogeneity of the system, seven conductivity values are connected using linear functions and thus the derivative is piece-wise constant. Although several optimization runs produced reasonable LSE, the solution showed sensitivity to sharp changes in conductivity values. To smooth out the sharp transitions in K , we replaced the linear interpolation with cubic spline. However, since there are only seven values over a domain of 30cm, cubic spline over the sparse data resulted in concavity in the curve that produced unrealistic negative K values. Finally, we implemented a moving average method to smooth out the curve while avoiding unwanted concavity. This is done by the following three steps:

- (1) apply linear spline to the seven K values to obtain a denser value set;
- (2) use moving average to filter out the noises in the denser set and obtain a smooth set;
- (3) apply cubic spline to the smoothed value set.

Step 1 inserts values among the original seven points to help guide the construction of cubic spline in Step 3; this way, unwanted concavity can be avoided. The reason to use cubic spline in the last step is to obtain the polynomial coefficient to differentiate the spline function for the $\frac{\partial K}{\partial z}$ term in Eq.(6), which are now quadratic and are smooth at the spline knots (the sensor locations). Figure 3 shows a comparison among the different methods of handling the K values. Here, we used a generic K vector for demonstration purpose. As seen from the graph, the hybrid moving average method agree at most locations with the linear spline but have softer transitions; meanwhile, the moving average does not create negative K values as the cubic spline does. The stability of the simulation tool was improved after the implementation of the moving spline method and thereby the simulation results are improved.

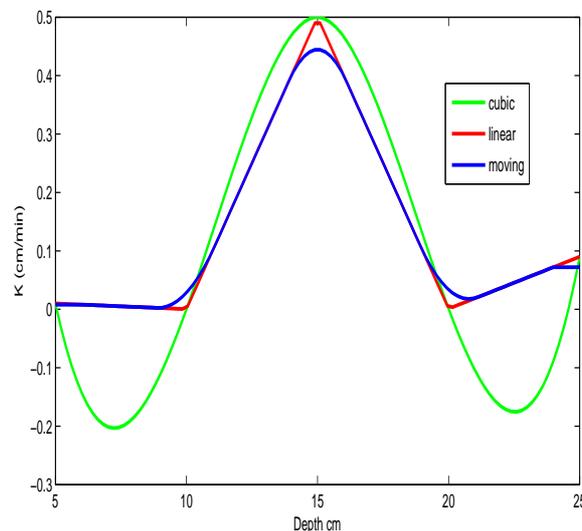


FIGURE 3. Comparison of methods for incorporating thermal conductivities

4.2. Using 30cm as Bottom Boundary Condition. The purpose of this study is to remove the influence of the missing bottom boundary condition and initial data on the optimization results. Here we consider a five layer system where we only identify the conductivity values

at 5 (1 ~ 5), 10 (5 ~ 10), 15 (10 ~ 15), 20 (15 ~ 20) and 25 (20 ~ 25) cm layer. The bottom depth is set at 30cm (instead of 70cm) and a spline of the data at 30cm is used as the bottom boundary condition. We also run optimization experiments to fit the final temperature data profile at 6, 12, 18 and 24th hour to observe the variation in the optimal conductivity sets.

Table 2 shows the optimal set of K values when fitting at four different final times. Although the sLSE for the 24th hour trial (4.4196e-004) is larger than that from the preliminary result (5.9633e-005 as shown in Table 1), the sLSE from the 12th hour trial shows a significant improvement of one magnitude less from our best error so far.

TABLE 2. Best-fit K values at 6, 12, 18, and 24 hours using 30cm as the bottom boundary condition

Final Time	K_5	K_{10}	K_{15}	K_{20}	K_{25}	sLSE
6hr	0.0821	0.4872	0.3459	0.4998	0.4998	7.3285e-005
12hr	0.1314	0.2592	0.1682	0.3842	0.2822	2.5603e-006
18 hr	0.0924	0.5000	0.00001	0.5000	0.375	1.3531e-003
24 hr	0.007	0.0004	0.4999	1.3431e-05	0.5000	4.4196e-004

4.3. Bad Data Elimination and Scaled Error Calculation. Top surface sensors are exposed to constantly changing environment; thus the data recorded in those locations are susceptible to bad noises. Such noise could severely corrupt the simulation by introducing unwanted trends into the system. In this study, we avoid using the data at 0cm and 1cm and only consider data from 5-30cm. To accommodate for this change, the K values at 0cm and 1cm are eliminated from the K vector; only K values at 5, 10, 15, 20, 25, and 30cm are being searched for. In addition, the top boundary condition changed to the data measured at 5cm instead of at 0cm.

With the updated simulator, same experiments as described in Section 4.2 (results shown in Table 2) are conducted. Table 3 shows the new results:

TABLE 3. Best-fit K values at 6, 12, 18, and 24 hours using the updated simulator

Final Time (trial)	K_5	K_{10}	K_{15}	K_{20}	K_{25}	K_{30}	sLSE
6hr	0.2308	0.1415	0.3300	0.2662	1.3652	2.0000	3.3559e-007
12hr	0.1722	0.2041	0.2253	0.3968	0.0553	0.0625	1.2309e-006
18 hr	0.2990	0.2335	0.2171	0.3553	0.4738	0.6681	8.1094e-008
24 hr	0.1563	0.1538	0.1646	0.3704	0.1582	0.3372	2.4496e-007

The stability of the new results have significantly improved as shown in the table. The sLSE for all four trials are on a similar scale and are reasonably small. Also notice that the sLSE in Table 3 is around two magnitudes smaller compared to results shown in Table 2. On the other hand, however, the K values from the 18 and 24 hour trial are more consistent across the row in comparison to those from 6 and 12 hour. Such oscillation in the first two trials could be caused by noises in data. An alternative approach to analyze the quality of a K set estimation is to examine the fitting of its simulation at other times other than the final time of optimization. For instance, Figure 4 and Table 4 show the fitting at 6, 12, 18, and 24 hours using the K values from the third trial (18 hour as final time).

Figure 4 and Table 4 have both indicated that the fitting at 18 hour is significantly better than fitting at other hours. Using the K values from the 18 hour trial to predict the profile at a future time, 24th hour, is particularly challenging as indicated by the large sLSE. Since the optimization in the 18 hour trial is guided towards minimizing the error at that hour, the observed biased behaviour is reasonable; however such temporal bias should be eliminated since we are trying to fit the profile at all times.

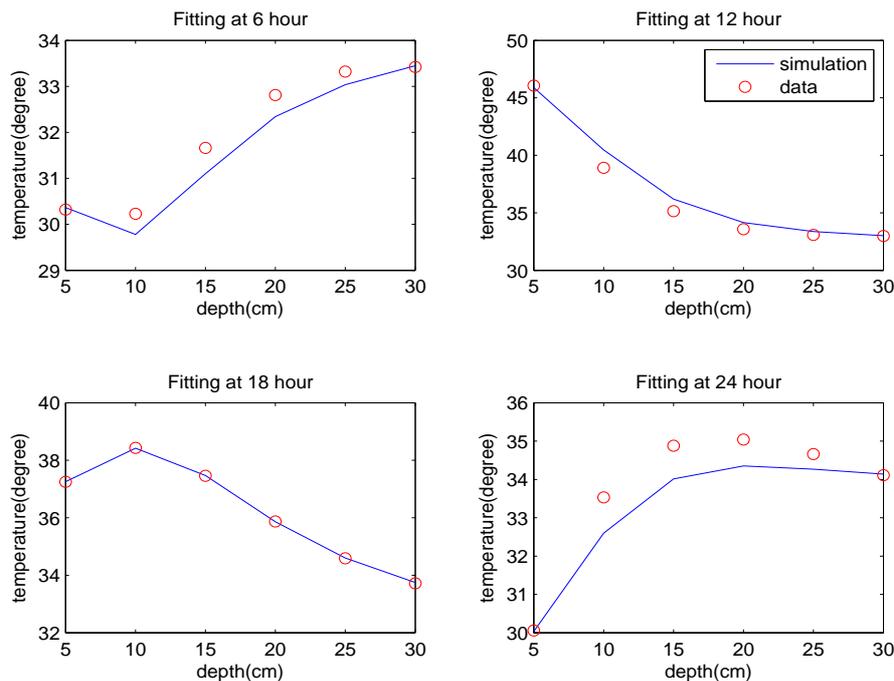


FIGURE 4. Profile fitting at different times using optimal K values from the 18 hour trial

TABLE 4. sLSE of fitting at different times using optimal K values from the 18 hour trial

Fitting at	6 hour	12 hour	18 hour	24 hour
sLSE	6.6541e-005	2.4030e-004	8.0971e-008	1.6450e-004

4.4. Guided Optimization: Error Calculation Over All Time Points. To resolve the biased optimization described in the previous section, we revised the error calculation so that it evaluates the fitting at all discrete time nodes:

$$nLSE = \frac{1}{2} \frac{\sum_{i=1}^N w_i (\hat{T}_i(K) - T_i^{obs})^2}{\sum_{i=1}^N (T_i^{obs})^2}, N_t = 289, \tag{9}$$

N_t is the number of time points. The number of time steps in the current simulator is 14400 over a span of 24 hours; however, to save data storage, we set $N_t = 289$ to match the first 24 hour data, which has 289 time nodes (measure every five minutes); this is done by saving the simulation result at every 50th time step. With the new error calculation, denoted as **nLSE**, the optimizer is guided to find a solution that fits the temperature profile at all times, other than just at the final time.

Table 5 shows the updated results of the same experiments as conducted in Section 4.2 and Section 4.3 (results shown in Table 2 and Table 3 respectively). In the updated results, the oscillation of solutions among different trials has significantly decreased; the difference between solutions from the 18 hour trial and 24 hour trial can be considered trivial. Furthermore, the nLSE decreases monotonically as the final time increases.

TABLE 5. Using nLSE calculation, results when optimized to different final time

Final time	K_5	K_{10}	K_{15}	K_{20}	K_{25}	K_{30}	nLSE
6 hour	0.1000	0.2250	0.1000	0.2875	0.2250	0.2875	1.1634e-005
12 hour	0.1669	0.1533	0.1916	0.3008	0.3079	0.5992	7.5479e-006
18 hour	0.1718	0.1637	0.2178	0.2169	0.3267	0.2651	5.3053e-006
24 hour	0.1724	0.1579	0.2151	0.1988	0.2783	0.2242	4.0694e-006

To examine how well each solution from Table 5 fits the profile at other times, we ran simulations over 24 hours using each of the four sets of K values from the table and compared the profile at 6, 12, 18 and 24 hours. Table 6 presents the results. For example, the entry in row 2 and column 5 is the sLSE of the solution at the 24th hour when simulated using the K values from the 12 hour final time trial. For each trial (row), the sLSE in bold is the smallest among all errors in that row. Notice that the minimum sLSE does not necessarily occur at the same time as the final time of the optimization. For instance, the smallest sLSE for the 6 hour final time trial occurred at the 24th hour. This is a strong argument for an unbiased optimization.

TABLE 6. Fitting errors (sLSE) comparison when using optimal Ks from different trials

Trial	$sLSE_6$	$sLSE_{12}$	$sLSE_{18}$	$sLSE_{24}$
K_6	1.2249e-005	2.2495e-005	2.5854e-005	6.8053e-006
K_{12}	2.8981e-006	5.5453e-006	2.1316e-005	2.0589e-005
K_{18}	5.3504e-006	1.4656e-006	8.4758e-007	3.4899e-006
K_{24}	4.1154e-006	2.6162e-006	7.3536e-007	9.5003e-008

Since the 24 hour trial has the smallest nLSE overall, we selected its corresponding optimal K values to examine the fitting at other times graphically. Figure 5 shows the fitting at all four hours; the fittings are greatly improved compared to Figure 4. Through numerical calculation, the maximum error in this trial is found to be $1.1^\circ C$, which is within the range of machine error as provided by the collaborator. This value is also significantly smaller than $6.2^\circ C$ as recorded in Table 1 using the older version of the simulator.

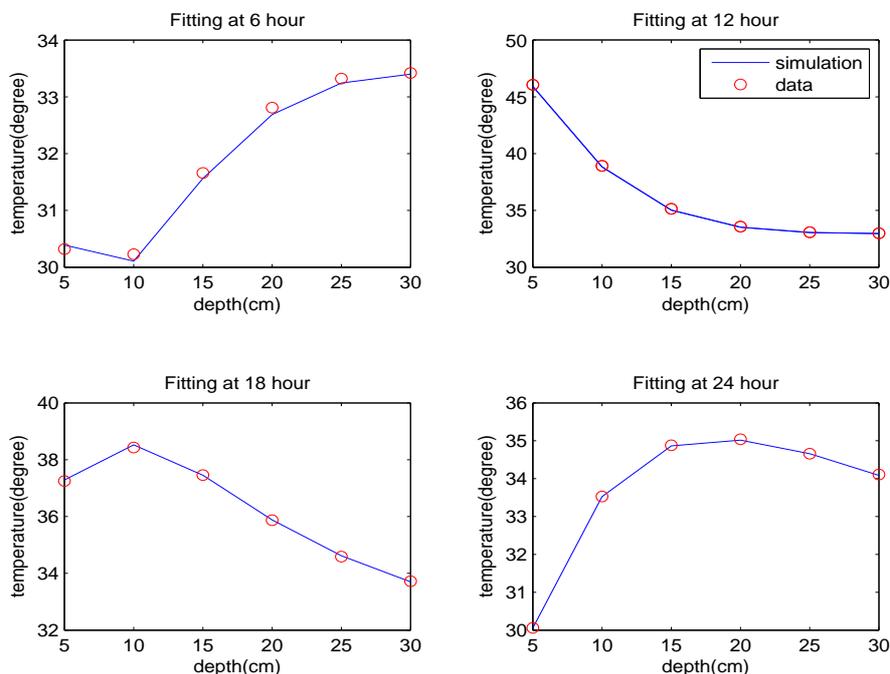


FIGURE 5. Under unbiased optimization: profil fitting at different times using optimal K values from the 24 hour trial

4.5. Summary of Implementation Experiments. To this end, we have completed a series experiments that help us with the choices made to tailor the simulator and optimization set up to its best functionality. To summarize the features in the fina setup of our simulator:

- Top Boundary Condition: data at 5cm
- Bottom Boundary Condition: data at 30cm
- K values interpolated using hybrid method
- Objective function: Error calculated using Eq.9

With the current setup, we are able to predict a set of K values that simulates reasonably fittin temperature profile at all times.

As will be discussed in the next two sections, measurement errors from the sensor are introduced to the simulator through the boundary and initial conditions; such error can not be eliminated using numerical scheme, however it will dissipate out of the system over time. In the fift row of Table 6, the sLSE decreases monotonically over time and the error at 24th hour can be seen as insignificant This dissipation behavior coincides with the analytical study of error propagation and sensitivity analysis of the simulator discussed in the following section.

5. SENSITIVITY STUDY

The study above identifie a potential set of thermal conductivities for the specifi site where the temperature data is taken. However, errors in this estimation process can be introduced in two ways: 1) errors in temperature measurement, and 2) system errors from numerical methods.

These two types of errors combined decide the accuracy of our estimation. System error is often inevitable and hard to predict due to the nature of numerical estimation; however, errors from data can be determined via specification of the measuring infrastructure (e.g the temperature sensor used in the study). A sensitivity study is where we analyze how the known error from initial setup of the system impacts the final results. It provides guidance on how much one should trust the estimation. Specifically, we study the impact of various types of errors in the initial condition and the speed with which they dissipate in the final solution.

5.1. On Homogeneous Model. Since the heat equation for a homogeneous system is linear, the error analysis can be done using analytic methods. Here, we consider Eq.(5) with a simplified case where $A = 1$, $T_a = 0$, $\omega = 1$, $\phi = 0$, and $K = 1$. The resulting analytical solution is:

$$T(t, z) = e^{-z/\sqrt{2}} \sin(t - z/\sqrt{2}). \quad (10)$$

Setting $t = 0$ in Eq.(10), we obtained the initial condition for the simplified case:

$$T(0, z) = e^{-z/\sqrt{2}} \sin(-z/\sqrt{2}) \quad (11)$$

We then introduce an error in the initial condition described in Eq.(11) with an additive term ε to get

$$T(0, z) = e^{-z/\sqrt{2}} \sin(-z/\sqrt{2}) + \varepsilon. \quad (12)$$

Let T_{true} be the true solution where it is obtained using the correct initial condition; let T_{error} be the solution of the PDE with perturbed initial condition where error ε is added. Let E be the error between the two solutions: $T_{error} = T_{true} + E$. Thus, the error E will satisfy the following boundary-value problem:

$$\begin{aligned} \frac{\partial E}{\partial t} &= \frac{\partial^2 E}{\partial z^2} \\ E(t, 0) &= 0; E(t, 30) = 0; \\ E(0, z) &= \varepsilon; \end{aligned}$$

The solution of the above system can be described analytically as:

$$E = \sum_{n=1}^{\infty} b_n \sin(\mu_n z) \exp(-\mu_n^2 t), \quad (13)$$

where $\mu_n = \frac{n\pi}{30}$, and

$$b_n = \frac{2}{n\pi} \cdot (1 - \cos(n\pi)) \cdot \varepsilon$$

It is obvious from Eq.(13) that the maximum error over time domain always occurs at $t = 0$ and is:

$$E = \sum_{n=1}^{\infty} b_n \sin(\mu_n z) \quad (14)$$

Eq.(13) also says that error at each spatial node decays over time exponentially. Figure 6 shows the error over time and space, when the initial error $\varepsilon = 0.05$.

The above study shows that, in a homogeneous transport system, error introduced through initial condition dissipates over time exponentially and the maximum error over time occurs in the very beginning.

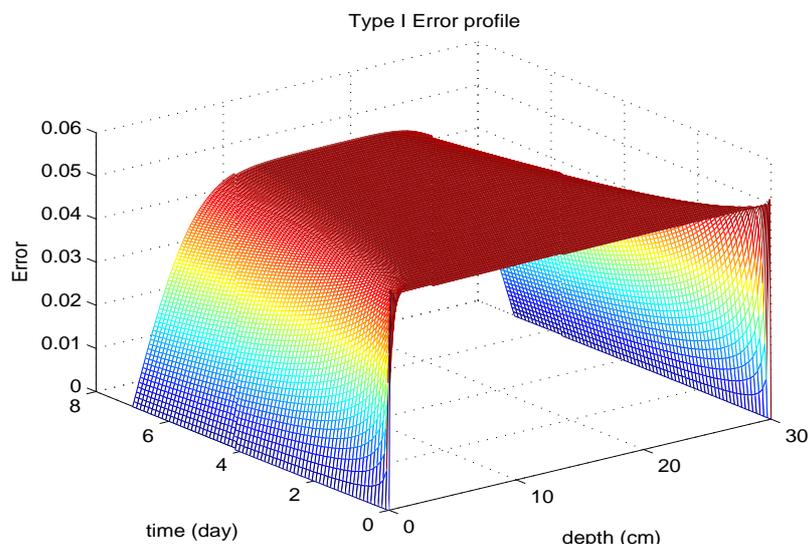


FIGURE 6. Type I error over time and space, using analytic method

5.2. Sensitivity Study on Heterogeneous Model. Error propagation in homogeneous model has shown high predictability from the study above. In reality, however, most of the soil environment is heterogeneous and thus a similar set of studies are done for a heterogeneous system. However, since the heterogeneous heat equation is no longer linear, it is more challenging to study the error propagation pattern using analytic method.

In this section, we present a series of numerical experiments to gain insight to the error propagation pattern in a heterogeneous heat model. To calculate the estimation error, we compare true solution to solutions with artificial perturbed initial conditions. Since there is no analytical solutions for a heterogeneous model like the one described in Eq.(5), instead, the true solution is the numerical results of a specific scenario that we defined. We chose a non-uniform five layer soil structure described in Section 2.1. The five corresponding conductivity values are 0.1314, 0.2592, 0.1682, 0.3842, and 0.2822 (cm/min). The entire set up is identical to the 12-hour final-time study using 30cm as bottom depth and data as boundary and initial conditions. More details are given in Section 4.2 and the optimal results are also shown in Table 2. For this specific study, we set the simulation time as seven days to make full use of the seven day data as boundary condition.

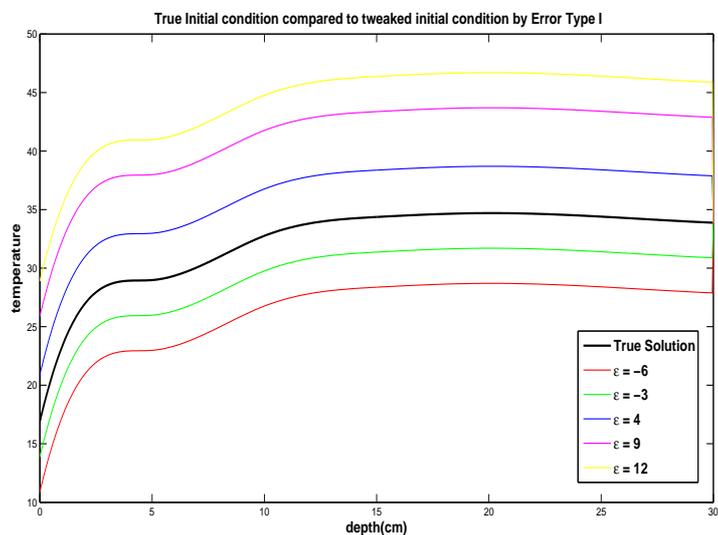
5.3. Type I Error. In this experiment, we introduce an error in the initial condition described in real data with an additive term ϵ to get:

$$T(z, 0) = DATA(z, 0) + \epsilon. \quad (15)$$

Here, $DATA(z, 0)$ is the first row of the data set, describing initial temperature profile. The ϵ s are chosen to be $\pm 4, -6, -3, 9, 12$, approximately $\pm 10\%$ to $\pm 30\%$ of the mean temperature of $40^\circ C$. Figure 7 compares the tweaked initial conditions with the original initial condition.

5.3.1. Numerical Experiments. We simulate the solutions using the perturbed initial conditions for seven days. Error is calculated using infinity norm of the differences of two matrix. Figure 8

FIGURE 7. True initial condition compared to tweaked initial condition, Experiment I for Heterogeneous Model



plots, in log scale, the infinity-norm error over time domain of 10800 minutes (7 days). Parallel and linear patterns are shown in this graph. Also notice that the curve for $\epsilon = 4$ and $\epsilon = -4$ overlaps, indicating that the sign of error is insignificant to the propagation pattern. Due to the symmetry, we omitted $\epsilon = -4$ in later experiments.

FIGURE 8. Infinit norm error over time, Experiment I, Heterogeneous Model

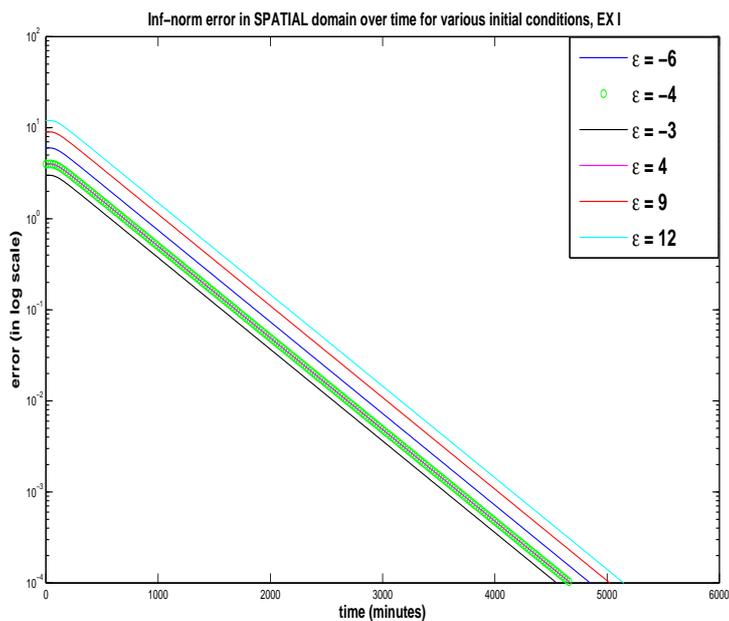
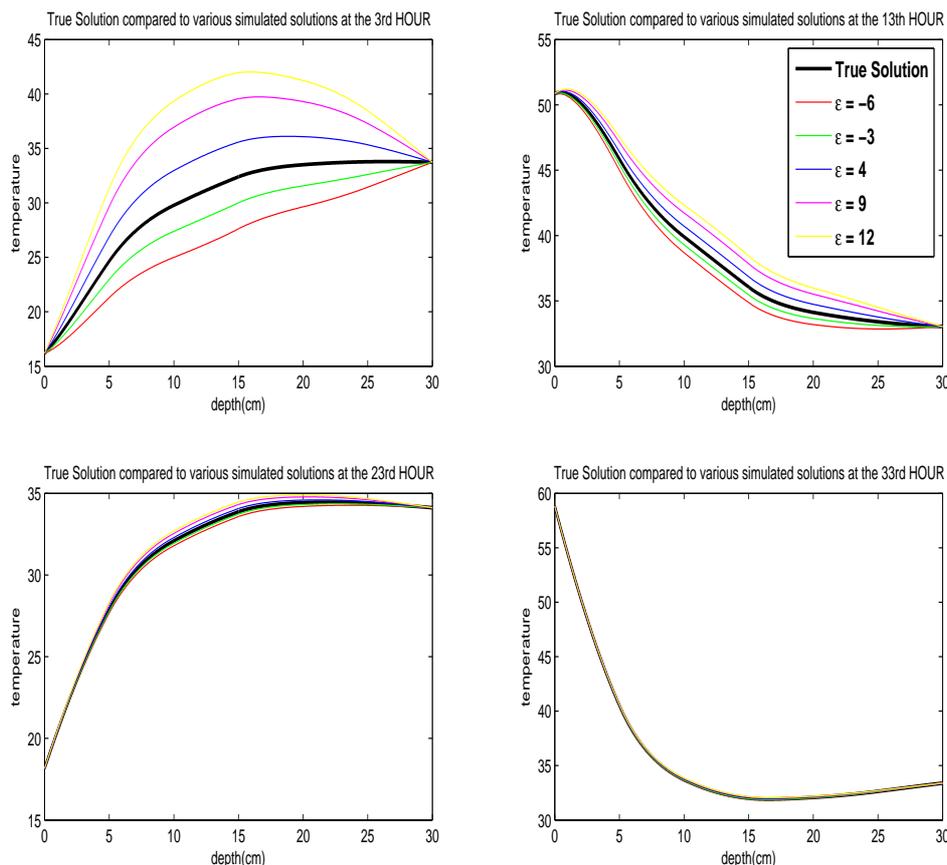


Figure 9 demonstrates how the simulations with tweaked initial conditions gradually approach the true solution. Notice that the difference becomes sufficiently small at the 33rd hour and the curves visually overlap.

FIGURE 9. True solution compared to simulated solutions at various time for each ϵ , Experiment I



5.3.2. *Analysis and Error Function for Type I Error.* As seen in Figure 8, the error plots of $\epsilon = -4$ and $\epsilon = 4$ overlap. This symmetry implies that the error propagation pattern is insensitive to the sign of the error, thus we assume that:

$$E_{\epsilon}(t) = E_{-\epsilon}(t) \tag{16}$$

As discussed in the previous section, the rate of decrease of log error is constant in respect to different ϵ values as indicated by the linear and parallel portion of the curves in Figure 8. Thus, the error function, $E_{\epsilon}(t)$, has the same negative slope for all ϵ . Using linear fit technique, we calculated the rate of decreasing as:

$$\frac{d\log(E_{\epsilon})}{dt} \approx -0.0010 \tag{17}$$

The unit of this rate is $\frac{\log(Error)}{time}$, since the linearity is shown in the log plot of error.

Next, we construct the line that describes the log of the error over time for different ϵ . Figure 8 shows that the curve stays flat at the beginning. Further data analysis confirms that the error stays constant until the 100th minute for all ϵ tested in Experiment I. This implies that the linear error equation for each ϵ always passes through the point $(100, \log(|\epsilon|))$. Now, we can formalize our analysis with the following:

$$\log(E_\epsilon(t)) = -0.0010 \times (t - 100) + \log(|\epsilon|), \quad (18)$$

where

$$\begin{aligned} E_\epsilon(t) &= 10^{-0.0010 \times (t-100) + \log(|\epsilon|)} \\ &= |\epsilon| \times 10^{-0.0010 \times (t-100)}. \end{aligned} \quad (19)$$

To validate the above analysis, we present the following plots:

FIGURE 10. Heterogeneous Model Experiment I: $\log(\text{error})$ over time

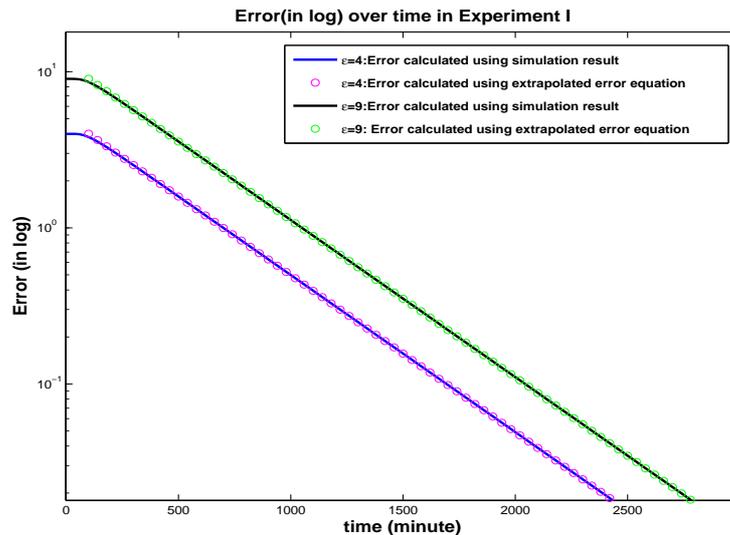


Figure 10 shows the original curve of $\log(Error)$ over time for $\epsilon = 4$ and $\epsilon = 9$ in solid line; then the same curves are generated using Eq.(22) with the two different E s. Both graphs show good fitting between data and the predictions made by the equations. Further validation was done to confirm the correctness of Eq.(22); details are not given here due to its similarity to what was done in the homogeneous study. Figure 11 is the infinity norm error plotted over time using both data and Eq.(22).

5.4. Type II error. In this experiment, the initial temperature is set to a constant:

$$T(z, 0) = T_0, \quad (20)$$

We ran five sets of experiments and for each set chose $T_0 = 20, 30, 40, 50, 60^\circ C$ respectively; here, the T_0 s are chosen to be within the range of real data ($13^\circ C$ to $67^\circ C$). Figure 12 plots the tweaked initial conditions and the original initial condition.

FIGURE 11. Heterogeneous Model Experiment I: maximum error over time

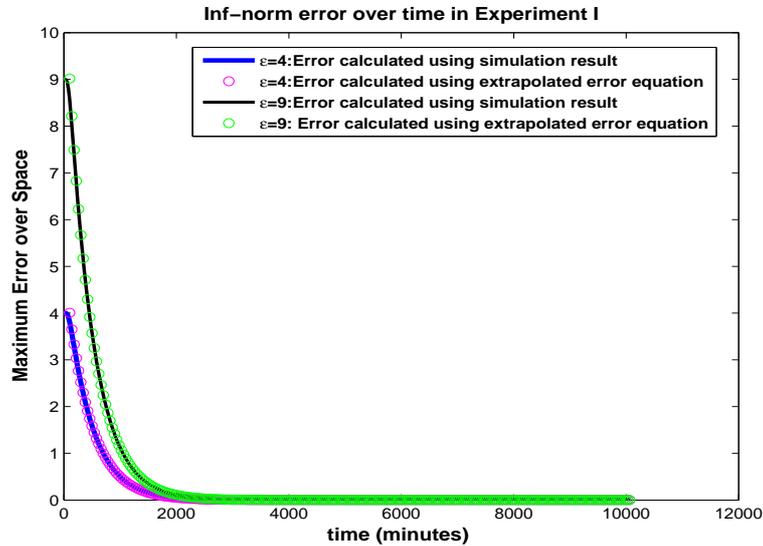
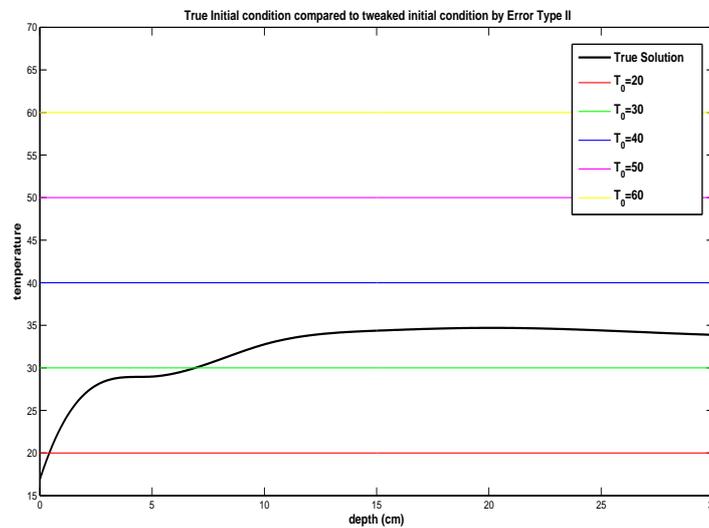


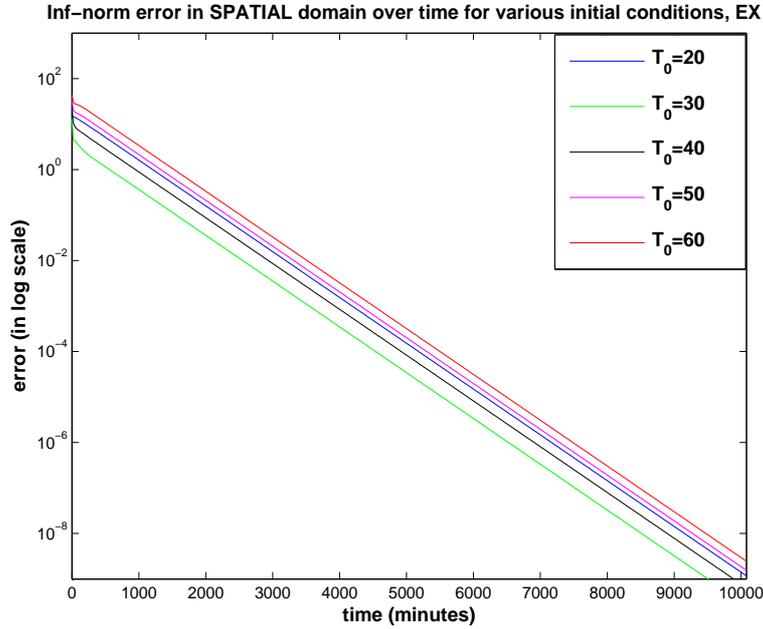
FIGURE 12. True initial condition compared to tweaked initial condition, Experiment II for Heterogeneous Model



5.4.1. *Numerical Experiments.* Figure 13 plots, in log scale, the infinity-norm error over time domain for the Type II Error.

Figure 14 shows how the impact of artificial tweaked initial condition eventually dissipates over time and the difference between true and ill-conditioned solutions becomes trivial in the end. More specifically, the curves overlap after the 33rd hour. These results are similar to what was shown for Type I Error.

FIGURE 13. Infinit norm error over time, Experiment II, Heterogeneous Model



5.4.2. *Analysis and Error Function for Type II Error.* The analysis for Type II Error is similar to the analysis given in previous session. Surprisingly, the rate of decrease of log error is the same for Type II as it is for Type I:

$$\frac{d\log(E_{T_0})}{dt} \approx -0.0010 \quad (21)$$

Next, using the same technique, we are able to construct the error function for Type II as the following:

$$\begin{aligned} E_{T_0}(t) &= 10^{-0.0010 \times (t-100) + \log(|\delta T_0|)} \\ &= |\Delta T_0| \times 10^{-0.0010 \times (t-100)}. \end{aligned} \quad (22)$$

where ΔT_0 is the temperature difference at the first time step, or the initial error:

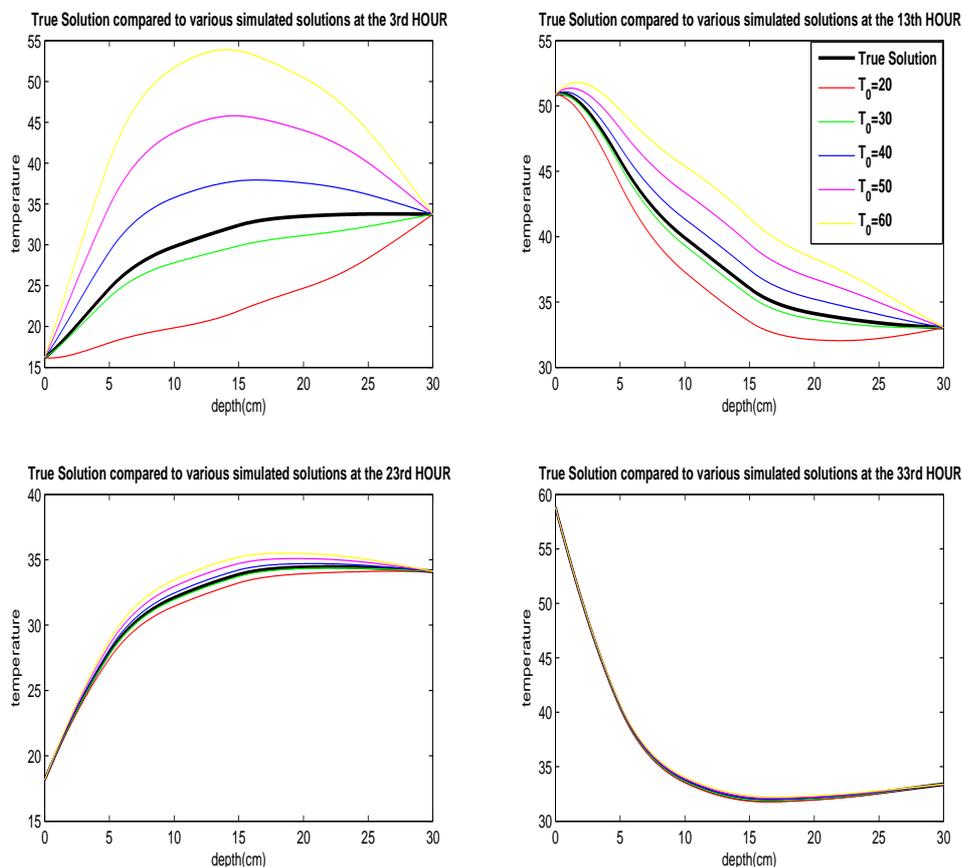
$$\Delta T_0 = T_0 - DATA(0). \quad (23)$$

Validation of the above analysis is done similarly to what was shown in Type I error. Details will not be included here.

6. CONCLUSION

For this work, we are given a data set that describes a subsurface soil temperature profile. To inversely map out the thermal conductivities that correspond to the temperature data, we posed the work under an optimization framework where the objective is to minimize the error between simulation and data. The simulator solves the 1-D heat equation with heterogeneous K values. Specifically, the boundary conditions are implemented as Dirichlet conditions with values from data; the initial condition is implemented using recorded data as well.

FIGURE 14. True solution compared to simulated solutions at various time for each T_0 , Experiment II

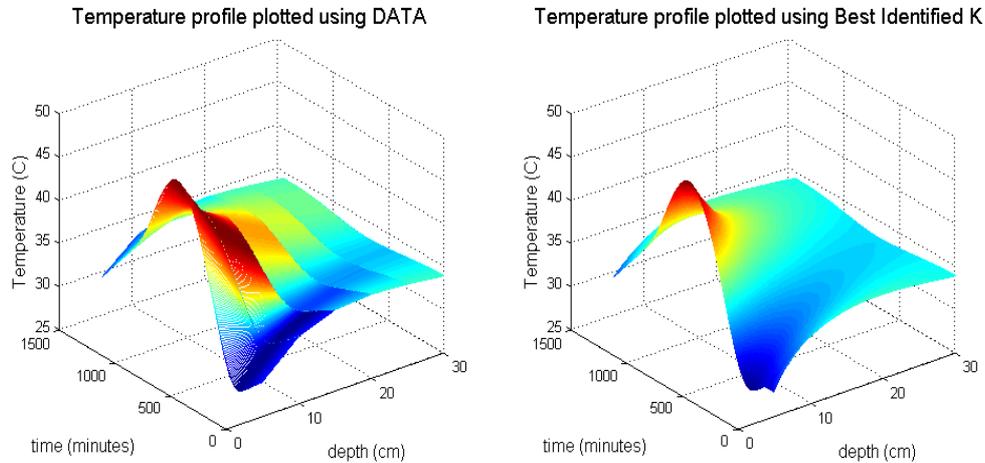


A stable simulator is the key to a good conductivity estimation. The challenges of building a stable solver in this work come from the uncertainties within the given data set. Due to noises and unknown structures in the data, implementation details in the simulator can have significant impact on the results. The numerical experimentations in Section 4 gave us the opportunity to evaluate how choices on the simulation and optimization process impact the final solutions. Through these studies, we are able to decide on the optimal set up for a stable solver. With this solver, we are able to identify a set of thermal conductivities that simulates well-fitting temperature profiles compared to the data at all times. The best identified optimal set of K values are shown in Table 7 (see also Table 5). Figure 15 compares the data profile with the simulated temperature profile using the best identified conductivities.

TABLE 7. Best identified K values

K_5	K_{10}	K_{15}	K_{20}	K_{25}	K_{30}	nLSE(Error)
0.1724	0.1579	0.2151	0.1988	0.2783	0.2242	4.0694e-006

FIGURE 15. Profil Comparison



The oscillation among these K values are small, indicating a potential homogeneous subsurface system.

To support the numerical and optimization work with data, we also conducted a sensitivity analysis on the simulator and its error propagation pattern. Both the optimization results from data and the analytical study confirm that the initial error dissipates out of the system over time. We are also able to quantify the amount of time it takes for error to become insignificant in a system, if given the initial error.

REFERENCES

- [1] D. Powers, *Boundary Value Problems and Partial Differential Equations*, chapter 2, ELSEVIER Academic Press, 2006.
- [2] T.N. Narasimhan, *The dichotomous history of diffusion*, Physics Today 62, 48, 2009.
- [3] C. T. Kelley, *MATLAB/FORTRAN software for Iterative Methods for Optimization* http://www4.ncsu.edu/~ctk/matlab_darts.html (online distribution), 1998
- [4] D. E. Goldberg *Genetic Algorithms in Search, Optimization and Machine Learning* Addison-Wesley Longman Publishing Co., Inc., 1989
- [5] J. R. Cannon *The One-Dimensional Heat Equation*, Encyclopedia of Mathematics and Its Applications. Vol 32, Addison-Wesley Publishing, 1984