Bayesian inference of structural error in inverse models of thermal response tests

Wonjun Choi[1,*], Kathrin Menberg[2,3], Hideki Kikumoto[1], Yeonsook Heo[4], Ruchi Choudhary[3], Ryozo Ooka[1]

[1] Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
[2] Institute of Applied Geosciences, Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, 76131 Karlsruhe, Germany
[3] Energy Efficient Cities Initiative, Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK
[4] Department of Architecture, University of Cambridge, 1-5 Scroope Terrace, Cambridge CB2 1PX, UK

*Corresponding author.
Tel.: +81-3-5452-6434, Fax: +81-3-5452-6432, E-mail: wonjun@iis.u-tokyo.ac.jp

Abstract

For the design of ground-source heat pumps (GSHPs), two design parameters, namely the ground thermal conductivity and borehole thermal resistance are estimated by interpreting thermal response test (TRT) data using a physical model. In most cases, the parameters are fitted to the measured data assuming that the chosen model can fully reproduce the actual physical response. However, two significant sources of error make the estimation uncertain: random error from experiments and structural bias error that describes the discrepancy between the model and actual physical phenomena. Generally, these two error sources are not evaluated separately. As a result, the suitability of selected models to correctly infer parameters from TRTs are not well understood. In this study, the Bayesian calibration framework proposed by Kennedy and O'Hagan is employed to estimate the GSHP design parameters and quantify the random and structural errors in the inference. The calibration framework enables us to examine structural errors in the commonly used infinite line source model arising due to the conditions in which the TRT takes place. Two in situ TRT datasets were used: TRT1, influenced by contextual disturbances from the outdoor environment, and TRT2, influenced by a strong groundwater flow caused by heavy rainfall. We show that the Bayesian calibration framework is able to quantify the structural errors in the TRT interpretation and therefore can yield more accurate estimates of design parameters with full quantification of uncertainties.

Keywords: Ground-source heat pump (GSHP); Thermal response test (TRT); Bayesian calibration; Structural biased error; Uncertainty assessment; Parameter estimation; Groundwater flow

Nomenclature

C	volumetric heat capacity (J/(m^3·K))
E	expectation
Ei	exponential integral
I_{sol}	global solar irradiance (W/m^2)
n_x	number of scenario parameters
n_θ	number of calibration parameters
p	probability distribution
q	heat rate per unit length of BHE (W/m)
$\bar{q}$	averaged heat rate per unit length of BHE (W/m)
r_b	radius of borehole (m)
R_b	borehole thermal resistance (m·K/W)
t	time or elapsed time after heat injection (s)
T	temperature (°C)
$\bar{T}$	mean fluid temperature (°C)
T_{DB}	dry-bulb temperature (°C)
v	precision parameter of covariance function
$\dot{V}$	volumetric flow rate (m^3/s)

x	scenario variable
y_c	simulation output
y_f	field measurement data
y	augmented observation vector, $y = \left({y_f}^T, {y_c}^T\right)^T$

Subscripts

b	model inadequacy (bias) function
c	computer data
f	field data
in	inlet
n	time step
N	total number of data (time steps)
out	outlet
s	soil or ground
0	initial
δ	bias (model inadequacy) function
η	computation model, GP emulator

Superscripts

T	transpose of vector or matrix

Greek letters

α	thermal diffusivity (m^2/s)
β	correlation parameter of covariance function
δ	model bias
ε	observation (random) error
ε_n	numerical (random) error
ζ	real (unobservable) physical process
η	model emulator term
θ	calibration parameters
θ^*	randomly generated parameters by LHS method
σ	standard deviation
λ_{eff}	effective thermal conductivity (W/(m·K))
Σ	covariance function of Gaussian process

Acronyms, abbreviations

BC	Bayesian calibration
BI	Bayesian inference
CI	credible interval
GP	Gaussian process
LHS	Latin hypercube sampling
MAP	maximum a posteriori
MCMC	Markov chain Monte Carlo
PM	posterior mean
PPDF	posterior probability density function

(All bold characters in the manuscript denote a vector or matrix.)

1. Introduction

Ground-source heat pump (GSHP) systems that utilize the shallow part of the ground as a heat source or sink have witnessed widespread use in recent years. The ground is not only a spatially inhomogeneous composite medium but also a porous medium. Therefore, subsurface heat transfer involves conduction and advection (e.g., forced convection by groundwater flow and natural convection). Identifying and measuring dominant heat transfer processes in the subsurface, where the ground heat exchangers (GHEs) are installed, is difficult and expensive compared to quantifying them for the load side of the GSHP, where the energy is supplied. This intrinsic nature of geothermal applications leads to significant uncertainties in the design and operation of GSHPs.

Research on uncertainty quantification of GSHP system performance follows the framework of the ISO's Guide to the Expression of Uncertainty in Measurement (GUM) [1]. The GUM framework emphasizes uncertainties associated with sensor data. GUM has been used to quantify uncertainties for various GSHP configurations and operation strategies. Notable studies include uncertainty analysis of the thermodynamic performance of a GSHP [2,3], uncertainty in performance of a hybrid GSHP combined with a solar thermal collector [4], and uncertainty in evaluating the energy balance of a GSHP's load and source sides [5].

In GSHP systems, the uncertainty associated with ground-related parameters is particularly important. These include the ground thermal conductivity and borehole thermal resistance. Both parameters have a significant impact on the design length of the GHE: Incorrect estimation can lead to a large increase in initial costs due to oversizing, or system failure during operation due to undersizing [6]. The GUM framework has also been used to quantify the uncertainties in these GSHP design parameters [7,8]. However, current work only considers sensor error as the main source of uncertainty and not the estimation process as a whole. The design parameters of a GSHP system are typically estimated via an inverse model using measured temperatures from thermal response tests (TRTs).

In inverse problems, such as in the inference of GSHP design parameters, the first task is to match the experimental conditions to the assumptions and boundary conditions made in the physical model (e.g., analytical or numerical). It involves selecting or developing a physical model that best represents the experiment and thus enables accurate inference of relevant parameters. However, often the physical model only partially represents the actual physical phenomena being measured. This may be due to lack of information on the system of interest or simplifications necessary in the modelling process.

A closer examination of the commonly used forward model for interpreting TRTs further highlights instances where experimental conditions do not match the physical model; in the uncertainty quantification literature, this is termed "model inadequacy." For instance, the commonly used infinite line source (ILS) model [9,10] and infinite cylindrical source (ICS) model [9] for interpreting TRTs assume that a TRT is conducted under the following conditions: the ground surface is adiabatic and the heat flux from the source is constant. However, at an actual TRT site where the TRT setup is fully exposed to the outdoor environment, such assumptions are usually violated by the fluctuation of the supply voltage [11,12], the heat exchange between the aboveground TRT setup and the outdoor environment [13,14], and heat transfer in the ground surface [15].

This mismatch between the model assumptions and the experimental conditions are well acknowledged and many studies have investigated this issue. For example, related to the unstable power rate issue, Shonder and Beck [11] developed a parameter estimation method that includes a one-dimensional numerical model as a forward model to consider the fluctuating power input. Hu et al. [12] proposed a data processing method that uses the Gaussian kernel regression method to eliminate the high frequency noise in the heat rate. Witte et al. [16] tried to solve the unstable power issue by using a special TRT apparatus equipped with a water-to-air heat pump, buffer tank, regulating valves, and control components. Because the apparatus could maintain a constant heat rate by mechanical control, very stable estimation behavior was achieved. Additionally, efforts have been made to study the effect of aboveground thermal disturbance caused by the heat exchange between the aboveground TRT setup and the outdoor environment because it can result in inaccurate inference of ground thermal properties [17]. The disturbance is generally perceived as a part to be removed, and thus studies have been conducted to remove the disturbed portion from the measured heat rate [18,19]. However, this approach leads to more complex inverse problems that require an additional measurement during TRT related to the radiative, convective, and conductive heat transfers in the experimental setup. On the contrary, Choi and Ooka [20] considered the fluctuating heat rate as known experimental boundary conditions that contain all contextual disturbances during TRTs. By applying a parameter estimation method that combines the quasi-Newton method with the temporal superposed ILS model, they showed that the estimation can be very stable without an explicit analysis of disturbances.

A third common instance where model inadequacy becomes important is when the TRT may be significantly influenced

by groundwater flow. Some analytical models (particularly so-called moving line source (MLS) models) have been developed that can account for the effects of groundwater based on the moving point source [9] and moving infinite line source [9,21]. For example, the moving infinite line source model on transient condition [22], the moving finite line source model for homogeneous groundwater flow [23], and the moving finite line source model for inhomogeneous groundwater flow [24] are representative analytical models that can consider the advection effect by subsurface groundwater flow. However, inverse application of these models for the estimation of GSHP design parameters poses several difficulties: First, in the estimation using an MLS model, two inter-dependent parameters, the ground thermal conductivity and Darcy velocity must be estimated simultaneously. A more fundamental problem in the estimation using an MLS model is the absence of information about the subsurface. When TRT is conducted in the field, one generally has very limited information (if any) on groundwater flow, groundwater temperature, temporal changes in groundwater velocity, and the spatial distribution of hydraulic properties. However, all these features need to be provided to perform the estimation using an MLS model by applying appropriate assumptions on the Darcy velocity. Therefore, the ILS is still the favored model for inverse estimation of ground-related GSHP design parameters. However, the ground thermal conductivity in a conduction-only model may be overfitted to the temperature response data without proper consideration of the external disturbances and groundwater flow. This work, thus tests and proposes a new framework for inverse estimation of ground thermal properties that enables explicit consideration of model inadequacy along with full quantification of uncertainties.

Thus far, uncertainty analyses for TRTs have ignored this consideration of model inadequacy and have only analyzed the intrinsic random errors of measurement on the basis of frequentist statistics. This has crucial limitations in two respects. First, such analyses are based on the assumption that a model itself can perfectly reproduce actual physical phenomena; however, it is already known that no model is perfect [25]. Furthermore, the main factors of uncertainty in the estimation of GSHP design parameters are the lack of knowledge about the unknown physical phenomena influencing the TRT rather than the intrinsic random error of measurement. Therefore, investigating the model inadequacy is essential for comprehensive uncertainty analysis and constitutes the main objective of this paper. We show that by doing so, one can obtain a more accurate estimation with reduced uncertainty and gain insights for improving the model to better match the actual phenomena. Second, we argue that frequentist techniques applied thus far for uncertainty analysis of ground thermal properties may not be suitable because they are suitable for events that can be repeated an infinite number of times under identical conditions [26]. However, TRTs cannot usually be conducted an infinite number of times and the environmental conditions differ each time; they are a one-time experiment. Therefore, this paper argues that Bayesian techniques are more suitable for uncertainty analysis in this context – especially when experimental data is limited.

Kennedy and O'Hagan [27] presented a Bayesian statistical framework (hereafter, KOH framework) for the calibration of computer models. Their framework facilitates the inference of the uncertain input parameters in computer models together with the quantification of the model inadequacy through a term called the model bias function. In this paper, we apply the KOH framework to infer the inadequacy of the ILS model along with the two unknown GSHP design parameters: effective ground thermal conductivity and borehole thermal resistance. Two different TRT datasets are selected for this purpose; TRT1 is affected by contextual disturbance and TRT2 is affected by a strong groundwater flow caused by heavy rainfall. The Bayesian calibration for TRT1 were conducted using two different forms of the ILS model to compare the magnitude of uncertainties and difference in the estimation results: 1) constant heat rate form; and 2) temporal superposition applied form for the consideration of the variable heat rate. The Bayesian calibration for TRT2 examines the situation when there is a significant discrepancy between the ILS conduction only model and the actual subsurface heat transfer owing to groundwater flow. These two TRT datasets thus include a representative model inadequacy that we typically encounter in TRT estimations. By applying the Bayesian calibration framework to the two TRTs, we show improved inference of ground thermal properties and their uncertainty range when model inadequacy is considered.

2. Methodology

2.1 Two datasets from thermal response tests

The experimental system for the TRTs was constructed at Inage Ward, Chiba, Japan, in January 2014. Except for the top ground layer of clay and loam (to a depth of 8 m), the site mainly comprises a fine sand having a porosity of 0.35 and hydraulic conductivity of 2.1×10^{-4} m/s. Two boreholes with a diameter of 165 mm and a depth of 52 m were drilled, and a single high-density polyethylene U-tube (outer and inner diameters of 34 and 27 mm, respectively) was inserted into each borehole. The two BHEs had the same effective length of 50 m and the same geometry, but their filling materials were different; one BHE was grouted with Portland cement mixed with 20% silica sand and the other BHE was filled with gravel having a grain size range of 8–15 mm. The TRT apparatus can generate up to 7 kW of heat. Well-calibrated Pt-100 sensors

and an electromagnetic flow meter were installed to measure the fluid temperature and flow rate, respectively. The hydraulic circuit that connects the BHE and TRT apparatus and the entire apparatus were insulated with 10-mm-thick polyethylene foam with a thermal conductivity of 0.04 W/(m·K). Moreover, the surface of the insulation was covered with white tape to increase radiation reflectance. The entire aboveground TRT setup was exposed to the outdoor environment and an installation photo can be found in Ref. [20].

Using this experimental setup, more than 30 TRTs were carried out from January 2014 to February 2017 (details regarding the experimental setup can be found in Ref. [28]). Two TRTs were chosen for this study, namely TRT1 and TRT2. TRT1 and TRT2 were conducted using the cement-grouted BHE and gravel-backfilled BHE, respectively. The experimental conditions of the two TRTs are summarized in Table 1, and the fluid temperatures, dry-bulb temperatures, and heat injection rates are shown in Fig. 1. The TRTs have the following characteristics:

- TRT1: It was subject to strong contextual disturbance due to sunny weather during the TRT; the diurnal air temperature difference was ~7–11 °C and the amount of solar irradiance was ~1000 W/m^2 (presented later in Fig. 6(a)). These are the cause of the fluctuating heat injection rate across the TRT period as shown in Fig. 1(a).
- TRT2: Nine days before the start of this TRT, there was heavy rainfall for five days owing to a stationary weather front. During these rainy days, the maximum hourly and daily precipitations were 25 mm/h and 117 mm/day, respectively, and the average rainfall for five days was 56 mm/day (presented later in Fig. 10). The heavy rainfall led to an unusual strong groundwater flow. Moreover, there was rainfall on the second and third days of the TRT period.

As aforementioned, neither external contextual disturbances (high solar irradiance and air temperature in this case) nor groundwater flow are represented in the ILS model. Hence, the two TRTs serve well to test the KOH framework for correctly quantifying these model inadequacies.

Table 1. Experimental conditions of two in situ TRTs ($\bar{q}$: average heat injection rate, $\dot{V}$: volumetric flow rate, T_0: initial ground temperature).

Experiment name	BHE type	Start time	Duration [h]	$\bar{q}$ [W/m]	$\dot{V}$ [L/min]	T_0 [°C]
TRT1	Cement-grouted	23:00 Jul 9, 2015	96	46.00	20.13	17.2
TRT2	Gravel-backfilled	00:00 Sep 16, 2015	96	44.91	20.03	17.0

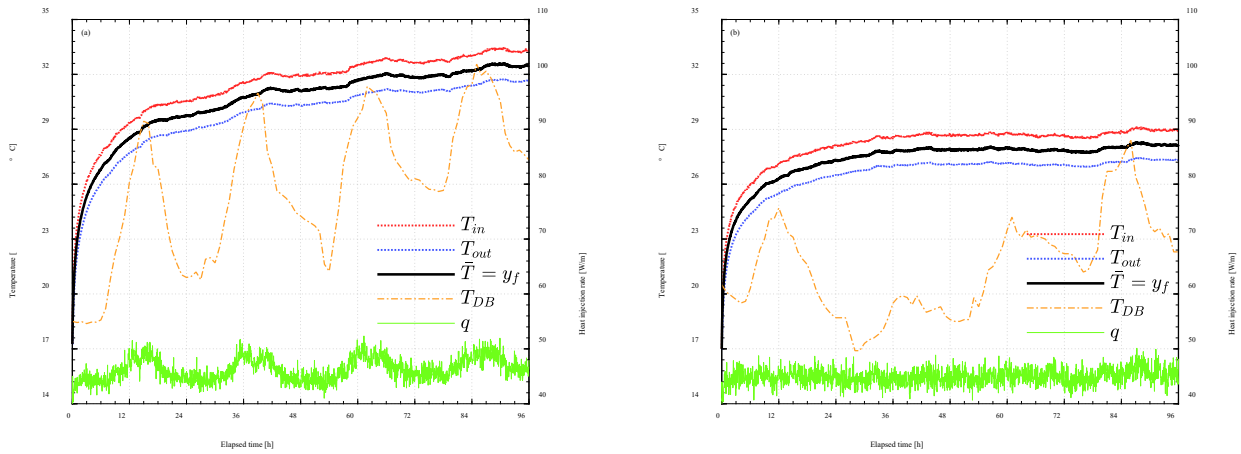

Fig. 1. Measured fluid temperature, dry-bulb temperature, and heat injection rate of two in situ thermal response tests: (a) TRT1 (dominated by conduction with contextual disturbances) and (b) TRT2 (affected by a strong advection caused by a heavy rainfall).

2.2 Computational model of the thermal response test

To infer the parameters from the TRT data, a physical model (computation model) is defined. The temperature response of a borehole heat exchanger (BHE) can be modeled using analytical or numerical methods. Each model has its own advantages and disadvantages in terms of complexity, accuracy, and computation costs, but most models assume that the mode of subsurface heat transfer is conduction only. Although some models such as the moving line source [9,21,22] and

moving finite line source [23,24] consider advection effects of groundwater flow, it is difficult to use them for inverse problems because of limitations in available information about the subsurface conditions, such as the permeability, hydraulic head, porosity, and hydraulic conductivity; in general, there are many unknown parameters to be estimated. Therefore, a computational model based on conduction only is used as the forward model here, which indirectly signifies that there is a high probability of structural error in the inverse estimation of parameters due to the inconsistency between the physical phenomenon and the chosen forward model.

Two different forms of the ILS model were used: the original form that uses a constant heat rate $\bar{q}$ and the temporal superposed form that considers a variable heat rate. They are expressed with respect to the mean fluid temperature $\bar{T}$ by Eqs. (1) and (2), respectively:

$$\bar{T} = \frac{\bar{q}}{4\pi\lambda_{eff}} \mathrm{Ei}\left(\frac{C_s r_b^2}{4\lambda_{eff} t}\right) + R_b \bar{q} + T_0 \tag{1}$$

$$\bar{T} = \sum_{n=1}^{N} \left[\frac{(q_n - q_{n-1})}{4\pi\lambda_{eff}} \mathrm{Ei}\left(\frac{C_s r_b^2}{4\lambda_{eff}(t_N - t_{n-1})}\right)\right] + R_b q_N + T_0 \tag{2}$$

Hereafter, Eqs. (1) and (2) are referred to as the constant model and the variable model, respectively.

2.3 Bayesian calibration with the KOH framework

Overview of Bayesian calibration framework

The Bayesian calibration framework developed by Kennedy & O'Hagan [27] enables us to make inferences about the unknown calibration (model) parameters θ while considering different errors terms, such as the random observation error and structural model inadequacy or model bias. In this study, the calibration parameters θ consist of the effective thermal conductivity λ_{eff} and the borehole thermal resistance R_b. The approach is based on Bayes' theorem (Eq. (3)), which relates the probability *p* of an event (or a specific parameter value θ) given evidence (or data) $p(\theta|y)$ to the probability of the event $p(\theta)$ and the likelihood $p(y|\theta)$:

$$p(\theta|y) \propto p(\theta)p(y|\theta) \tag{3}$$

Based on this relation, prior beliefs about parameter values $p(\theta)$ are updated based on observed data y and quantified in the form of posterior probability distributions. According to the KOH framework, the relation between observed data (field measurements) y_f and computer simulation outputs y_c can be expressed as

$$y_f(x) = \zeta(x) + \varepsilon = y_c(x, \theta) + \delta(x) + \varepsilon + \varepsilon_n \tag{4}$$

where ζ is the true physical process that cannot be observed and depends on the state variable *x*, which often represents an input to the computational model and can also be regarded as the scenario in which the experiments are conducted (e.g., elapsed time from heat injection and heat rate in TRT); ε represents the measurement errors related to the field observations; $\delta(x)$ represents the model inadequacy in the form of model bias function representing the discrepancy between the model and the true process along the dimensions of the state variables; and ε_n is the random numerical error term originating from the simulation model. Here, the KOH framework differs significantly from the frequently used GUM framework [1], which only takes into account the uncertainty regarding measured quantities and does not allow for explicit, inverse quantification of different sources of uncertainty.

When the computational load of the model is high, simulations with varying parameter values can be performed only a limited number of times. On the basis of the approach adopted in previous studies [27,29,30], we use an emulator denoted by $\eta(x, \theta)$ instead of using the computational model directly in order to improve the computational efficiency (Eq. (5)).

$$y_f(x) = \eta(x, \theta) + \delta(x) + \varepsilon + \varepsilon_n \tag{5}$$

For the generation of training data for the emulator, a finite number of parameter sets θ^* are generated from the predefined parameter space using the Latin hypercube sampling (LHS) method [31,32], and the ILS model is executed using the parameter sets as input values under scenarios x. The emulator is constructed using the simulation output. The field observations y_f and computational model outputs y_c are combined into the augmented observation vector $y = \left(y_f{}^T, y_c{}^T\right)^T$.

The model emulator and the model bias function are both modeled using the Gaussian processes. GP models are generalizations of nonlinear multivariate regression models that relate individual input parameters with the model outcome by mean and covariance functions [33]. The GP models in this study were defined using a zero mean function corresponding to the standardized model outputs and parameters, while the covariance functions for the emulator Σ_η and the model bias Σ_δ are specified according to Eqs. (6) and (7) [34]:

$$\Sigma_{\eta(i,j)} = \frac{1}{\nu_\eta}\exp\left[-\sum_{k=1}^{n_x}\beta_{\eta,k}\left(x_{ik}-x_{jk}\right)^2 - \sum_{k'=1}^{n_\theta}\beta_{\eta,n_x+k'}\left(\theta_{ik'}-\theta_{jk'}\right)^2\right] \tag{6}$$

$$\Sigma_{\delta(i,j)} = \frac{1}{\nu_\delta}\exp\left[-\sum_{k=1}^{n_x}\beta_{\delta,k}\left(x_{ik}-x_{jk}\right)^2\right] \tag{7}$$

This formulation introduces several uncertain hyperparameters to the Bayesian calibration process: ν determines the precision of the corresponding covariance function and thus the magnitude of the emulator term $\eta(x,\theta)$ and the model discrepancy term $\delta(x)$; two sets of parameters β, based on the number n_x of x and the number n_θ of θ, determine the correlation strength of the corresponding covariance function and thus the smoothness of the emulator output and the model discrepancy function over x, respectively [34]. The random error terms for the field observation error and the numerical error are represented by the covariances Σ_ε and Σ_{ε_n}, respectively. These error terms are assumed to be independent and identically distributed, and their covariances are specified by additional precision hyperparameters, ν_ε and ν_{ε_n}.

The covariance function of the augmented dataset y used for calibration, which includes both field measurements and computer model outputs, is then specified as Eq. (8):

$$\Sigma_y = \Sigma_\eta + \begin{pmatrix}\Sigma_\delta & 0\\ 0 & 0\end{pmatrix} + \begin{pmatrix}\Sigma_\varepsilon & 0\\ 0 & \Sigma_{\varepsilon_n}\end{pmatrix} \tag{8}$$

Prior probability distributions for the uncertain parameters and hyperparameters are assigned on the basis of suggestions made in previous studies [30,34]. Gamma distributions are assigned to the precision parameters; ν_η: Gamma(10, 10), ν_δ: Gamma(10, 0.3), ν_ε: Gamma(10, 0.01), and ν_{ε_n}: Gamma(10, 0.001). These priors reflect our belief that the emulator accounts for the majority of the variation in the standardized field and model responses, while the model bias and observation error make smaller contributions and the numerical error is believed to be very small. For the smoothness parameter β, we follow the re-parameterization suggested by Guillas et al. [30] and specify the priors similarly to reflect our belief of a rather smooth emulator and model discrepancy functions (β_η: Beta(1, 0.5), β_δ: Beta(1, 0.4)). The prior distribution for the unknown model parameters θ is specified using a triangular distribution, the details of which are described in Section 3.2.

Finally, the posterior density of the parameters in this study is explicitly expressed as follows:

$$p\left(\theta,\nu_\eta,\beta_\eta,\nu_\delta,\beta_\delta,\nu_\varepsilon,\nu_{\varepsilon_n}|y\right) \propto \underbrace{p\left(y|\theta,\nu_\eta,\beta_\eta,\nu_\delta,\beta_\delta,\nu_\varepsilon,\nu_{\varepsilon_n}\right)}_{\text{likelihood}}\underbrace{p(\theta)p\left(\nu_\eta\right)p\left(\beta_\eta\right)p(\nu_\delta)p(\beta_\delta)p(\nu_\varepsilon)p\left(\nu_{\varepsilon_n}\right)}_{\text{prior}} \tag{9}$$

The posterior probabilistic distribution of the unknown model parameters θ (and the hyperparameters) are obtained by repeated evaluations of the emulator $\eta(x,\theta)$ for different θ across x, with iterative sample draws from the prior distributions. We use the random-walk Markov chain Monte Carlo method for sampling with 15,000 iterations and the Metropolis-Hastings criterion [35–37] to accept or reject new sample draws based on the change in the posterior density [38]. After obtaining the posterior samples, we visually inspect the trace plots of all the (hyper-) parameters after a burn-in

period of 5000 samples to check for convergence to the target distribution. Then, we statistically evaluate the posterior samples to predict the mean (expectation) and the variance of the emulator term and the outcome of the calibrated model. By subtracting the emulator results from the model outcome, we can infer the model bias function over the range of x. The schematic for the Bayesian calibration framework is shown in Fig. 2.

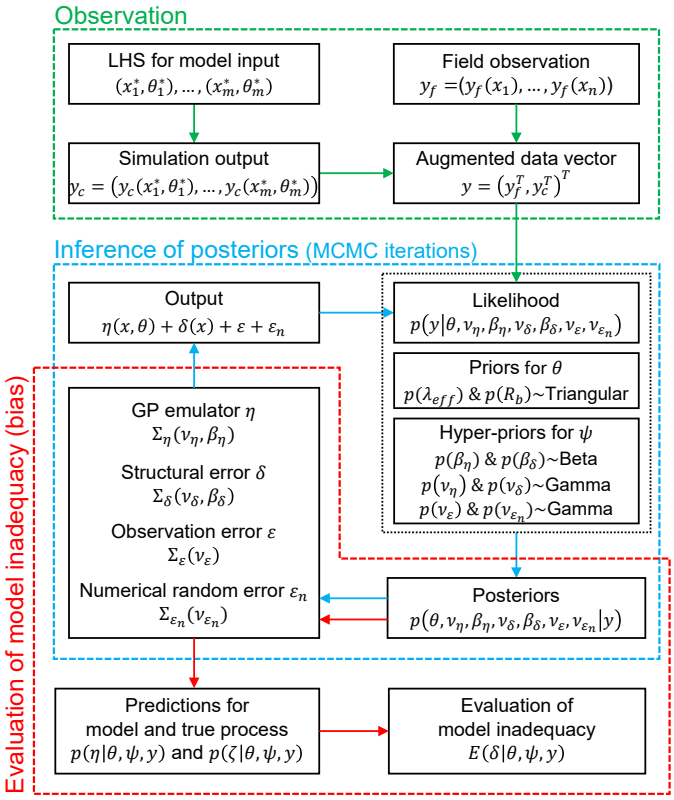

Fig. 2. Schematic of Bayesian calibration with the Kennedy & O'Hagan framework.

Combined dataset

The combined field experiment and computer model data denoted by the augmented dataset y were used for the calibration. The field data consist of the measured outputs y_f, which represent the measured mean fluid temperatures, and the values of the field scenario variables x_f. For the evaluation of the constant heat rate model (Eq. (1)), the elapsed time and two weather parameters (ambient air temperature and solar radiation) were selected as scenario variables x_f, because they are known to play an important role in the perturbation of the temperature response [17–19,39–42]. Because the average heat rate does not change with time when the constant heat rate model is used, the heat rate is not considered as the scenario variable. By contrast, when the variable heat rate model (Eq. (2)) was used for calibration, the elapsed time and variable heat rate, which are required as model inputs, were selected as the scenario variables x_f.

Note that augmented dataset y also includes information from the computer model. The computer model data consist of the calculated mean fluid temperature y_c at given scenario values x_c, which are set to be identical to x_f in this study, and at a given set of values of the calibration parameters θ. As the computer model used for the TRT includes only two unknown parameters, we choose the effective thermal conductivity λ_{eff} and the borehole resistance R_b as the calibration parameters.

For the training of the GP emulator, we use the model outputs y_c of two computer models (Eqs. (1)) and (2)). The Latin hypercube sampling (LHS) method was used to create 40 pseudo-random sets from a pre-defined parameter space of model input (calibration) parameters. The parameter ranges of λ_{eff} and R_b were set to 1.7–2.3 W/(m·K) and 0.13–0.17 m·K/W, respectively, which covers most of past TRT results over 3 years. We use triangular distributions to define priors of calibration parameters; in the order of the upper and lower limits and mode, λ_{eff}:Tri(1.7, 2.3, 1.9) and R_b:Tri(0.13, 0.17, 0.15).

Using the 40 input parameter combinations from the LHS, measured heat rates (averaged heat rate for Eq. (1) and

variable heat rate for Eq. (2)), and computer models, we generated 40 sets of mean fluid temperature for the elapsed time of 16 h to 96 h at 1-h intervals. Because the ILS model with the steady-state R_b assumption cannot accurately predict the early-time fluid temperature, we only consider the measured and modeled temperature data from the elapsed time of 16 h. Fig. 3 shows the range of the modeled fluid temperatures using 40 random parameter sets with the measured mean fluid temperatures.

From the experience, we know the groundwater flow at the experimental site is usually very weak. Thus, selecting the ILS model as the forward model, which only considers the subsurface heat conduction, is a reasonable choice. However, unlike Fig. 3(a) and (b), in Fig. 3(c), the generated learning data (gray dots) using the 40 pairs of parameter sets θ^* and the measured heat rate deviate severely from the experimental response data. This means that a very unusual thermal response was obtained because of some unexpected effects, presumably unusual strong groundwater flow and, consequently, the selected forward model cannot reproduce the physical phenomenon properly. In this case, the use of a general deterministic estimation method or a Bayesian inference that overfits the unknown parameters without considering the possible model bias is problematic. The effect of the model bias on the estimation results in comparison to conventional estimation methods only with random errors will be discussed in Section 4.2.

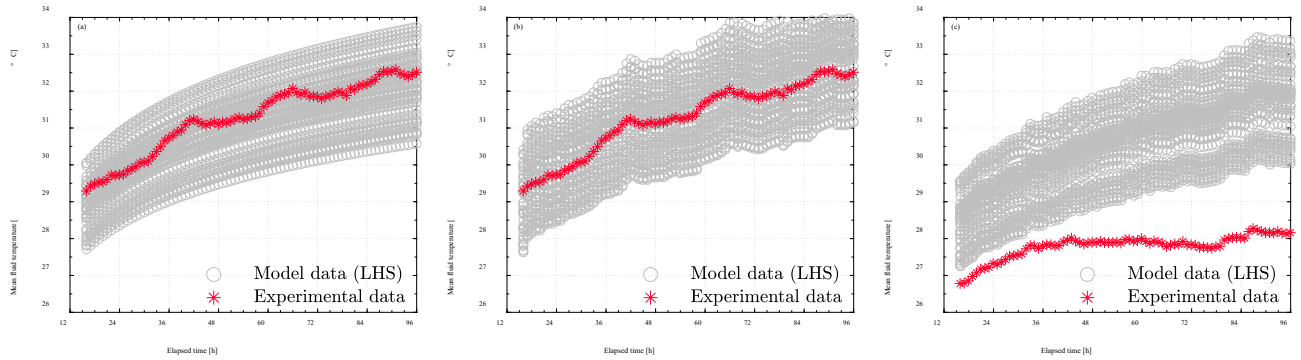

Fig. 3. Range covered by modeled mean fluid temperature with 40 LHS samples of λ_{eff} and R_b ((a): constant model TRT1, (b): variable model TRT1, (c): variable model TRT2)

4. Results and discussion

4.1 TRT1: Conduction-dominated TRT data with contextual disturbances (comparison between constant heat rate model with three scenario variables and variable heat rate model with two scenario variables)

We apply the KOH framework to the conduction-dominated TRT1 data for calibration of the constant heat rate model and the variable heat rate model. On the basis of the inferred joint posterior distribution constructed from 10,000 samples, we plot predictions over x of the emulator term $\eta(x,\theta)$ (computer model output) for the two models. Fig. 4 shows the expected prediction results based on the posterior mean (PM) and credible intervals (CIs) of two emulators. We observe that the variable heat rate emulator shows better prediction performance than the constant heat rate emulator. In the case of the constant model, some field data points are out of the 2σ credible interval of the computer model, whereas in the case of the variable model, most field data are covered by the credible interval of 1σ.

As stated in Section 3.3, the constant model considers three scenario parameters x: elapsed time t, dry-bulb temperature T_{DB}, and solar irradiance I_{sol}. By contrast, the variable model uses two scenario parameters: elapsed time and variable heat rate. For the sake of comparison, we show both model emulators over the elapsed time (Fig. 4(a) and (b)). In fact, setting the elapsed time as the only scenario parameter for the constant model emulator provides results that are very similar to those obtained with three scenario parameters because the computer model (Eq. (1)) does not consider the outdoor temperature and solar irradiance directly. However, we have this on-site information and the knowledge that these two parameters affect the measured temperature y_f [17,42]. Therefore, they are included as additional state variables in the calibration settings to capture model discrepancy as the function of contextual disturbances.

The hyperparameters β_η in the emulator function represent the correlation strength of the scenario parameters. Thus, the posteriors of β_η can provide additional insights into the relevance of the considered scenario parameters, which is not possible with the traditional approach for uncertainty assessment; the maximum a posteriori (MAP) estimates of the emulator hyperparameters are $\beta_{\eta,t} = 2.4$, $\beta_{\eta,T_{DB}} = 3.9 \times 10^{-4}$, and $\beta_{\eta,I_{sol}} = 1.5 \times 10^{-4}$. From the results, we can conclude

that the dry-bulb temperature and solar irradiance are more relevant to the fluid temperature fluctuation than the elapsed time.

The precision hyperparameters ν (ν_η, ν_ε, ν_{ε_n}, and ν_δ) represent the magnitude of the covariance functions. For example, ν_η describes the covariance in the combined field and model outputs explained by the emulator term. A small value for ν implies that the corresponding component of the statistical model (Eq. (5)) absorbs a large part of the variance in the model output, and accordingly explains a large amount of the variation in y. To prevent the algorithm from putting too much focus on the variance of a single component, which would most likely represent an unrealistic solution, we constrain some of the posterior distributions by defining upper and lower boundaries. We set $\nu_{\varepsilon_n} \leq 2 \times 10^5$, as we expect a very small numerical error. For ν_η, we set $\nu_\eta \geq 0.3$, which implies that a part of the variance in the model must be explained by other components. Inspecting the posterior distribution of ν_η reveals that most posterior realizations are indeed close to the set boundary. The reason for very low ν_η values is probably related to the information contained in the data used for the calibration (e.g., strong correlations between individual input parameters x and θ, redundant information, and noisy data for very similar intervals of x). At the same time, we observed very high MAP values for the measurement and numerical error terms corresponding to very small error (constant model: $\nu_\varepsilon = 465$, $\nu_{\varepsilon_n} = 20{,}000$, $\nu_\delta = 22$; variable model: $\nu_\varepsilon = 570$, $\nu_{\varepsilon_n} = 20{,}000$, $\nu_\delta = 33$).

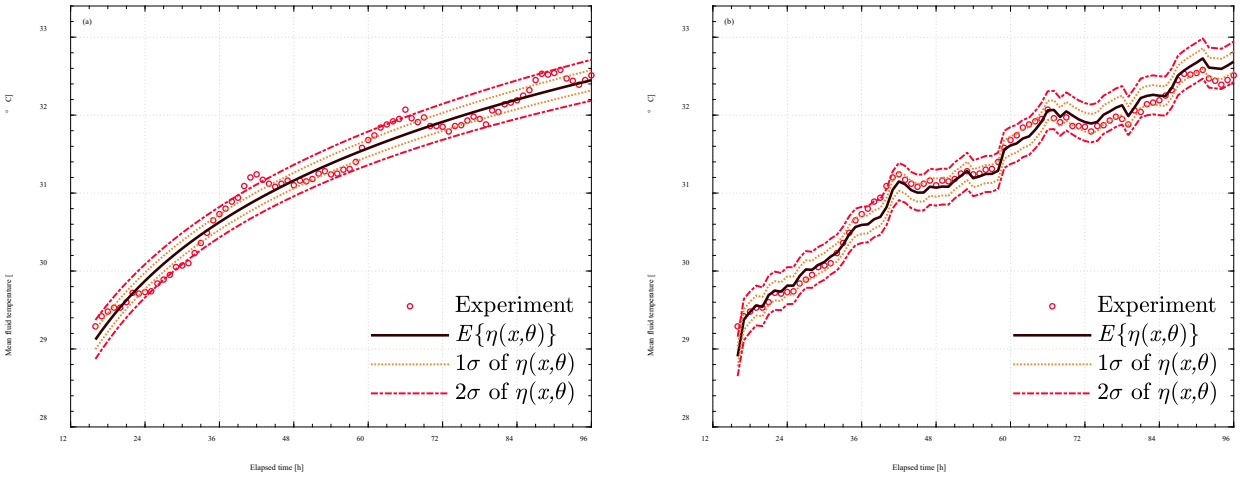

Fig. 4. Mean fluid temperature predicted by the emulator of the simulation model $E\{\eta(x,\theta)\}$ and uncertainty ranges plotted over the elapsed time ((a) constant model, and (b) variable model) based on the inferred posterior distributions.

From the posterior probability density functions of the unknown calibration parameter θ, we then infer the unknown model parameters. Fig. 5 shows a comparison between the prior distributions and the posteriors from the constant and variable heat rate models.

Because the priors $p(\theta)$ were set based on the results from past TRTs, they nearly cover the lower and upper bounds of the PPDFs of the two parameters in both the constant and the variable models. The mode of λ_{eff} for the variable model is higher than that of the constant model (Fig. 5(a)). In addition, the PPDF range of the variable model is narrower than that of the constant model. This means that by using the variable heat rate model and the Bayesian inference framework, we can reduce the uncertainty about this unknown model parameter because the additional information contained in the more complex model structure leads to better reproducibility of the actual thermal response. The PPDFs of R_b show very similar results. The PPDF of variable model is shifted to lower values and the uncertainty range of R_b in the variable model is again narrower than that in the constant model.

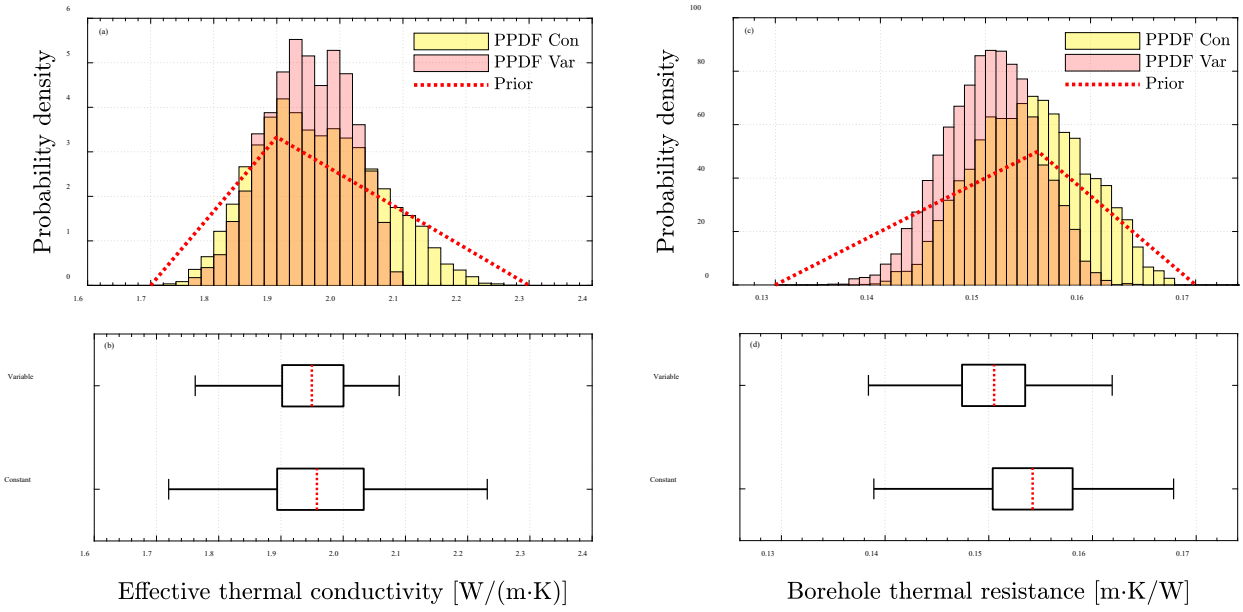

Fig. 5. Prior and posterior distributions of effective thermal conductivity and borehole thermal resistance with the constant and variable heat rate models: (a) prior and PPDFs of λ_{eff}, (b) box plots of λ_{eff}, (c) prior and PPDFs of R_b, and (d) box plots of R_b. The lower and upper ends of a box represent the first and third quartiles, respectively, while the red band inside the box denotes the median. The left and right ends of a whisker represent $Q_1 - 1.5(Q_3 - Q_1)$ and $Q_3 + 1.5(Q_3 - Q_1)$, respectively. Outliers are excluded from the plots.

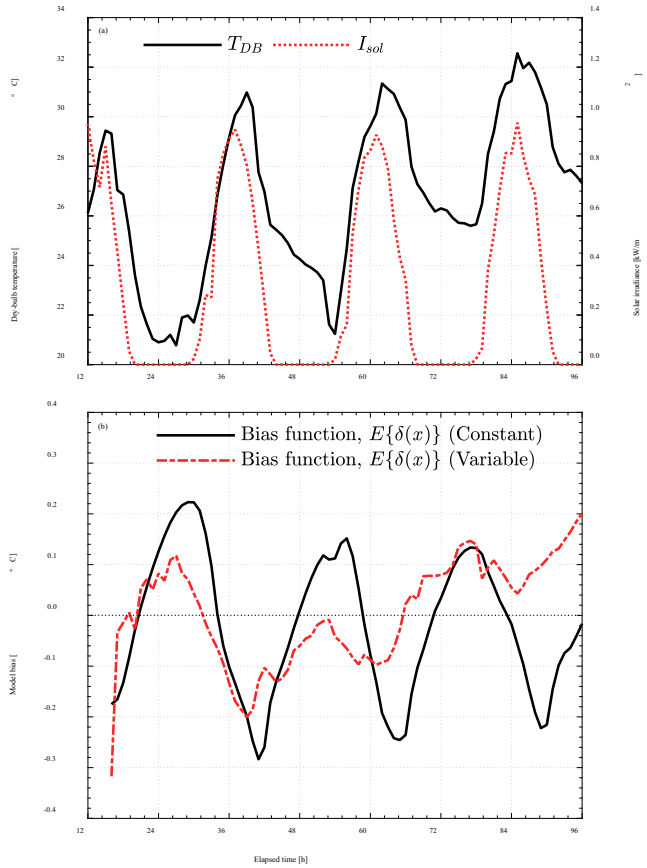

Fig. 6. (a) Measured dry-bulb temperature and solar irradiance during TRT1 and (b) inferred model bias functions (i.e., difference between the measured mean fluid temperature and modeled mean fluid temperature) over time for constant and variable heat rate models.

Based on the separation of structural and random error in the applied approach, we can now examine the model bias function associated with the two computer models for the first time. We expect that the function for the constant model will explain the additional heat injection due to high solar irradiance during daytime hours. Fig. 6(a) shows the dry-bulb temperature and solar irradiance during TRT1 and Fig. 6(b) shows the bias functions of the constant and variable models that represent most of the difference between the measured mean fluid temperature and modeled mean fluid temperature. We can observe an apparent time-lagged coupling between the weather data and the bias function in the case of the constant model. This is confirmed by the posterior mode values of hyperparameter $\beta_{\delta,k}$ for the constant model, which represents the contribution of a certain scenario parameter to the bias function. The smaller a correlation parameter, the greater is the sensitivity to the bias function. The PMs of $\beta_{\delta,t}$, $\beta_{\delta,T_{DB}}$, and $\beta_{\delta,I_{sol}}$ are 36.5, 1.4, and 0.4, respectively. This confirms that the solar irradiance is the most critical bias factor (i.e., cause of model inadequacy in the constant model) among the scenario parameters. The results are consistent with those of Refs. [17,42] in which the contextual disturbances in the aboveground TRT setup were analytically modeled and the relative sensitivity of disturbance factors to the temperature perturbation was determined.

By contrast, in the variable heat rate model, no certain periodic pattern or correlation against the weather data was observed. The absolute bias is less than 0.2 K except for the first hour, which is close to the measurement accuracy of the Pt-100 sensors. From the results, we can confirm that the variable model is better suited to reproduce the temperature response data, as was shown in [20]. Moreover, the model bias function of the variable model shows no regular pattern over both scenario parameters (elapsed time and variable heat rate), which means that the remaining model inadequacy is unlikely to be related to a conduction-related physical effect. If the model bias was related to the physics, we might speculate that very weak intermittent subsurface advection could be one possible cause of the result. The remaining model bias here might also be related to the GP emulator that we use, which absorbs any small structural error in the data into the model bias, regardless of the underlying physical process or original error source. However, inspecting the model bias function of the variable heat rate model for other TRT datasets might be useful for investigating other possible structural effects.

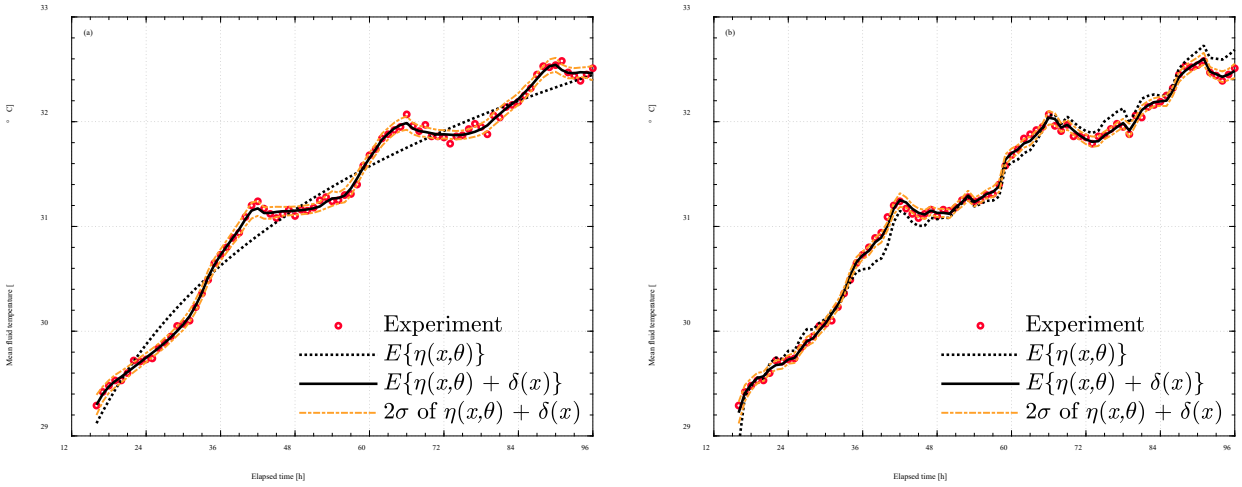

Fig. 7. Measured temperature data, predicted fluid temperature by model emulator ($E\{\eta(x,\theta)\}$), calibrated model output ($E\{\eta(x,\theta) + \delta(x)\}$), and 2σ range of calibrated model ((a) constant model, (b) variable model).

Finally, we can use the samples from the joint posterior to make predictions of the model outcome under consideration of uncertainty, which represents a significant improvement to common deterministic approaches. Fig. 7 shows that the calibrated model outcome of both the constant and the variable models closely follows the measured data, and most measured data points are within the 2σ intervals. For the constant model, we clearly see the effect of the inferred model bias function when comparing the model prediction and the calibrated model output lines, which compensates the effect of the solar irradiance and outdoor dry-bulb temperature. For the variable heat rate model, the difference between the emulator and the model outcome shows that there are some deviations in the elapsed time between ~36 and 40 h and from 72 h onwards, which are correctly compensated by the model bias function (see also Fig. 6(b)).

Because TRT1 dataset was also used in Ref. [43], in which Bayesian inference was conducted by considering only the

random error term, it would be interesting to compare the inferred posteriors of λ_{eff} and R_b of the previous and current studies. The comparison and discussion between two results are in Appendix A.

4.2 TRT2: Rainfall and groundwater flow affected TRT data

Results from the conduction-dominated TRT1 data showed that Bayesian calibration leads to reliable PPDFs for the calibration parameters and can correctly infer the model bias function caused by ignoring the effect of disturbance. We now test the methodology using the TRT2 data.

Although TRT2 was conducted at the same site with a similar heat injection rate, as shown in Fig. 1(b) and Table 1, compared to TRT1, an unusual time evolution of the fluid temperature is observed. Instead of a logarithmically increasing trend of the temperature level, the mean fluid temperature remains around 27–28 °C and even decreases slightly over a short period of elapsed time around 72 h. We speculate that this is due to subsurface advection by groundwater flow.

We apply the KOH framework to the TRT2 data to examine whether the methodology can correctly identify the model bias and infer reliable posteriors for the calibration parameters. For Bayesian calibration of TRT2, the variable heat rate model was selected as the physical model, and the elapsed time and heat rate were selected as scenario parameters x. The same distributions as those used for TRT1 were set for λ_{eff} and R_b and all (hyper-)parameters. As stated in Section 4.1, 15,000 MCMC samplings were conducted, and the first 5,000 samples were not considered to construct the posterior distributions.

Based on the obtained posterior distributions, predictions were made for the emulator term. Fig. 8 shows a significant over-prediction of the fluid temperature over time, which indicates a significant effect of some physical process that is not captured by the emulator. Even the credible interval of 2σ predicted temperatures cannot cover the measured temperature.

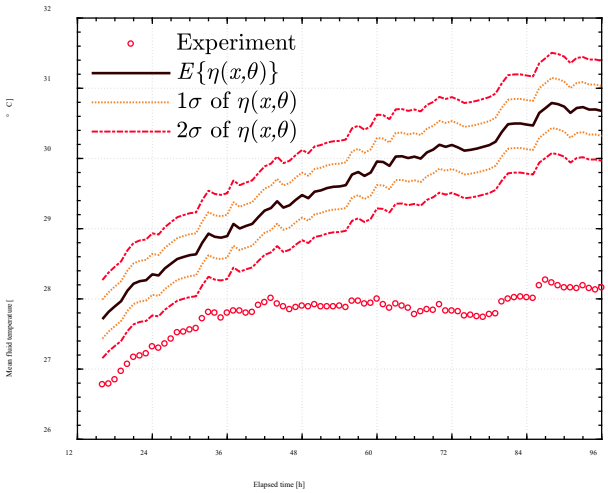

Fig. 8. Mean fluid temperature predicted by emulator $E\{\eta(x,\theta)\}$, prediction uncertainty ranges, and measured mean fluid temperature (TRT2).

Fig. 9 shows the posteriors of the calibration parameters λ_{eff} and R_b from the TRT2 data. Significant differences are observed compared to the results of TRT1. The posterior distributions are skewed to the upper and lower bounds of the prior distributions (i.e., 2.3 W/(m·K) and 0.13 m·K/W), respectively, which suggest a larger thermal conductivity and lower borehole resistance. These shifts in the two calibration parameters confirm our assumption that the significant difference is due to advection effects. Such high λ_{eff} is a typical result from advection-affected TRTs, as the groundwater flow in the subsurface carries a significant part of the injected heat away, which results in a nonlinear increase in the thermal conductivity because the diffusion (conduction) term alone cannot properly represent the convection mode of subsurface heat transfer.

The difference between the posteriors of λ_{eff} and R_b of TRT1 and TRT2 indicates that the model bias term of TRT2 compensates any remaining effect caused by the additional physical process in the subsurface. Ideally, the posteriors of λ_{eff} and R_b of the two TRTs should be very similar, and the bias function should fully account for any discrepancy.

On the other hand, the standard interpretation method for TRTs that uses the semi-log linear regression leads to a significantly larger λ_{eff} of 5.97 W/(m·K). and a slightly larger R_b of 0.177 m·K/W compared to the TRT results of the

past 3 years at the same site. This large λ_{eff} is due to the overfitting of the ILS model to the measured data, whereas the large R_b is due to the positive correlation between λ_{eff} and R_b when two parameters are estimated simultaneously as discussed in Ref. [43]. Therefore, whilst the model bias function does not fully capture the convection mode of subsurface heat transfer, it does prevent overfitting between the measured data and computer model outputs, yielding improved estimates of parameter values.

It is necessary to investigate the cause of the unusual temperature response and estimates. Fig. 10 shows the rainfall and dry-bulb temperature from 14 days before the starting time of TRT2. The pink shaded area in Fig. 10 represents the period during which the TRT2 was conducted. Although there was rainfall during the TRT period, it is highly unlikely that the rain penetrated the top 8 m formation, which consists of loam and clay, and significantly affected the temperature response because the hydraulic conductivity of loam and clay is very low [44]. From approximately nine days before the start of the TRT (Sep. 7, see Fig. 10), there was heavy rainfall for 5 days because of a stationary weather front. During these rainy days, the maximum hourly and daily precipitations were 25 mm/h and 117 mm/day, respectively, and the average precipitation was 56 mm/day. Based on the regional characteristics of the test site located in a coastal area, it is likely that the rainfall would cause a regional increase in the groundwater level and groundwater flow velocity in the permeable aquifer, which would take effect during the TRT period. Previous studies in similar geomorphological and hydrogeological settings have found time delays between the beginning of rainfall events and peaks in groundwater level between 40 hours and ~100 days depending on the specific aquifer characteristics, especially the hydraulic storage capacity [45,46]. Given the local aquifer thickness and hydraulic parameters (see [28]), a delay time of ~9 days seems to be reasonable for our test site (Fig. 10). Based on this assumption, the rainfall for 9 days before TRT2, the predicted and measured temperature responses, and the bias function are shown in Fig. 11.

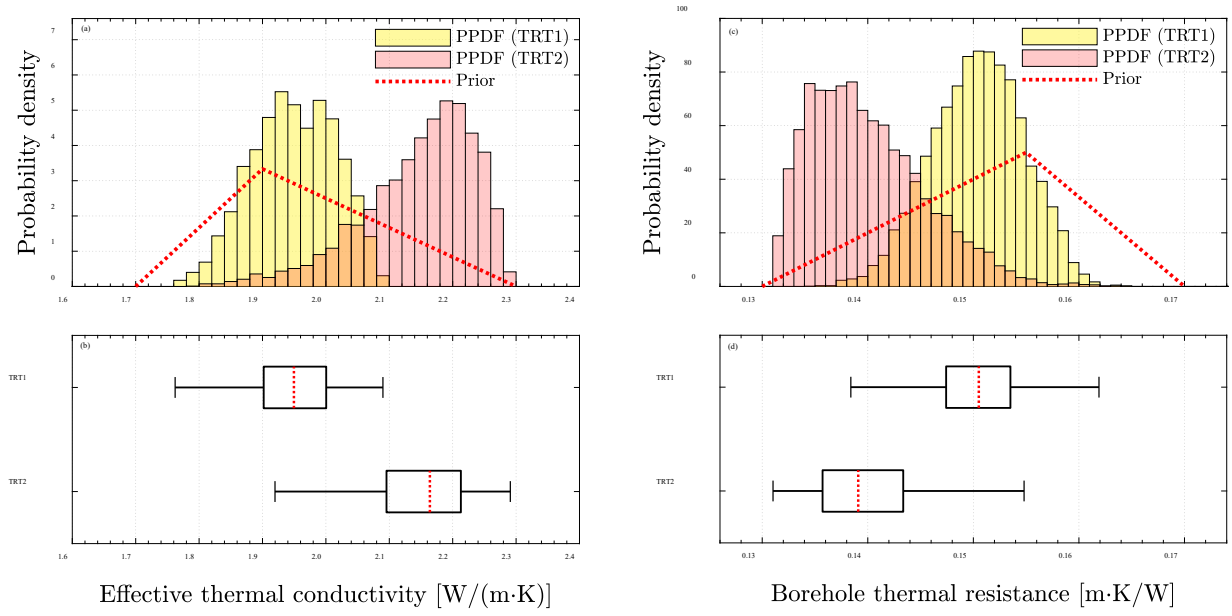

Fig. 9. Posterior and prior probability distribution functions of thermal conductivity ((a) and (b)) and borehole thermal resistance ((b) and (c)) for conduction dominant TRT1 and advection-affected TRT2.

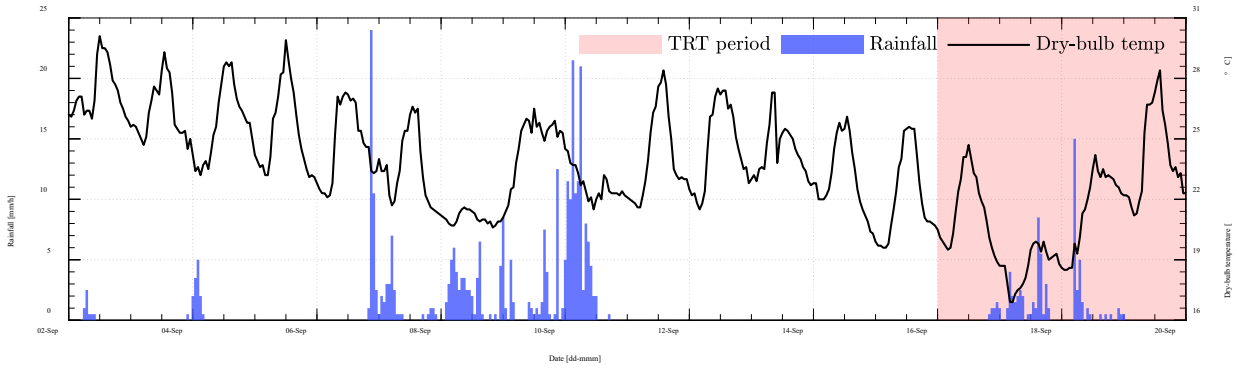

Fig. 10. Rainfall and dry-bulb temperature before and during TRT2.

In Fig. 11 (b) and (c), the difference between the predicted and measured temperatures at the elapsed time of 16 h is ~1 °C. It should be noted that the predictions by the emulator were made using λ_{eff} and R_b inferred by the calibrated TRT2 model. Compared to TRT1, the temperature of TRT2 at the elapsed time of 16 h is ~3 °C lower (see Fig. 1). Thus, it can be seen that TRT2 is strongly influenced by the groundwater flow from its early stages. The magnitude of the bias increases with time (Fig. 11 (c)), and this increasing trend is similar to the previous trend of rainfall (Fig. 11 (a)). As shown in Fig. 12, when the emulator and bias function are considered simultaneously (e.g., calibrated model), the resultant temperature response closely follows the measured temperature. Thus, from the viewpoint of prediction, the calibration was successful. However, we focus not only on using the calibrated model for future predictions, but also on discovering the structural error in the parameter inference. Therefore, this point merits further discussion.

Unlike for TRT1, a clear model inadequacy was captured by the bias function in TRT2 (Fig. 11 (c)). However, it cannot be said that this bias function completely captures the model inadequacy. Compared with previous TRT results, the inferred parameters in the KOH framework shown in Fig. 9 are unreasonable, because some part of the model inadequacy that should be accounted for by the bias function is compensated by parameter overfitting with convergence to the upper and lower bounds of the parameter space. Although completely capturing the bias in the TRT is not possible by applying the KOH framework in the form presented in this paper, an important result is that by modeling the bias error term, we can determine that there is a structural error between the measured temperature response and the physical model, and infer the approximate shape of the model bias function. Moreover, this approach gives us insights into the right selection of physical model for parameter estimation and the modification of models or development of alternative models.

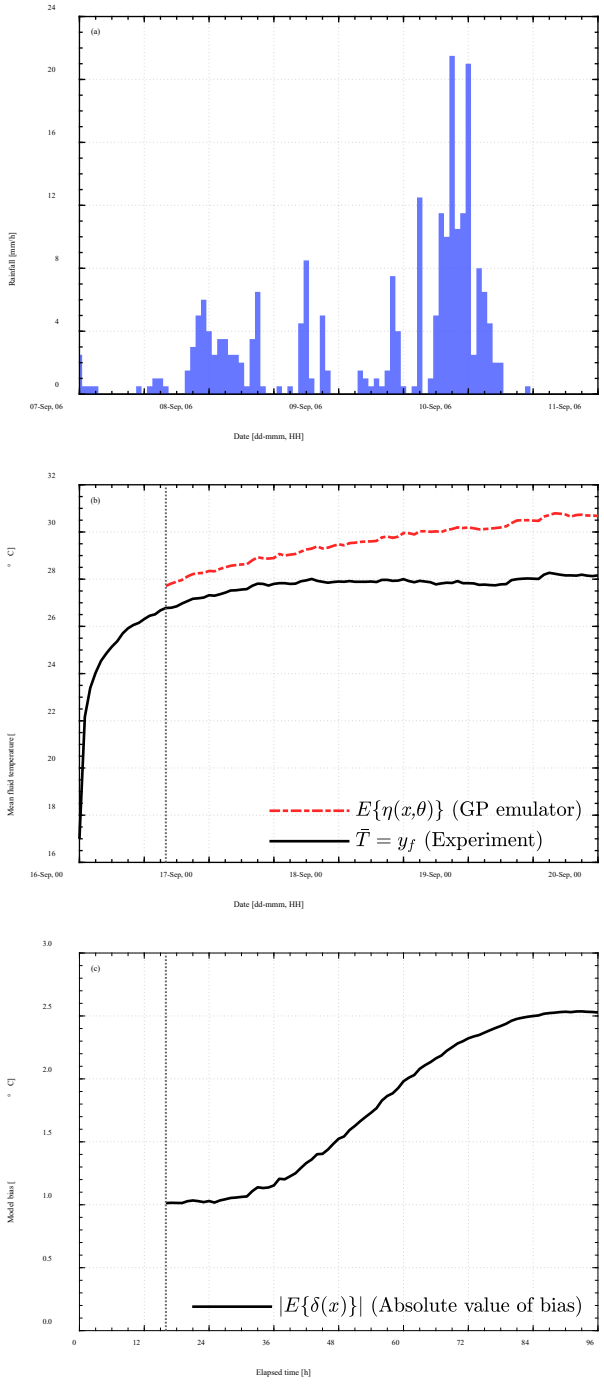

Fig. 11. (a) Rainfall 9 days before TRT2, (b) temperature responses of emulator and experiment, and (c) bias function. The bias function originally has negative values here because it is defined by measured value minus predicted value. For the purpose of comparison with the rainfall, the bias function is expressed as an absolute value. The vertical dotted line in figures (b) and (c) is the indicator for the elapsed time of 16 h.

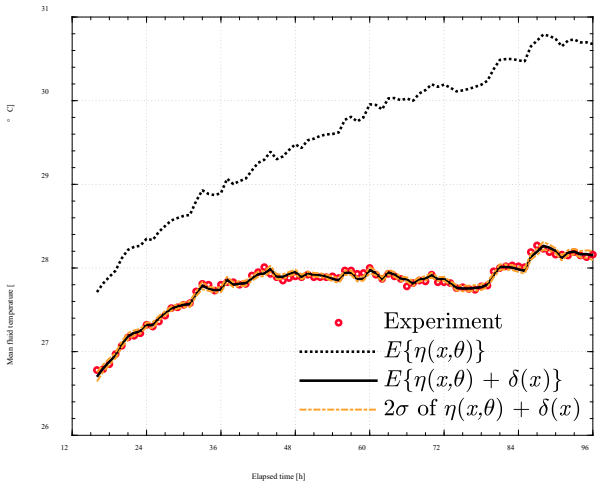

Fig. 12. Measured temperature data, predicted fluid temperature by model emulator ($E\{\eta(x,\theta)\}$), calibrated model output ($E\{\eta(x,\theta) + \delta(x)\}$), and 2σ range of calibrated model (TRT2).

6. Conclusions

We applied the Kennedy and O'Hagan framework for Bayesian calibration for the first time to two different TRT datasets in order to estimate the design parameters for ground heat exchangers and examine structural errors in the models used for interpreting TRTs. As the forward model for the Gaussian process emulator, the infinite line source model was used in two different forms: the constant heat rate model and the variable heat rate model.

The difference between the Kennedy and O'Hagan method and other approaches (e.g., Bayesian inference, GUM framework) is that it explicitly accounts for different error sources, namely the random errors from the measurements and computer model, and the structural error, which we call the model bias function, and which is linked to the physical inadequacy of the computer model.

The Bayesian calibration framework successfully identifies the structural model inadequacy in the constant heat rate model. In TRT1 data with the constant model, periodic variations in the model bias with time were clearly captured. Moreover, in the bias term of the Gaussian process model, a high correlation between the model bias and the solar irradiance and dry-bulb temperature could be assessed by evaluating the posterior distributions of the hyperparameters of the Gaussian process model.

The variable heat rate model, which implicitly incorporates the causes of the model bias in the constant model, enables us to infer narrower uncertainty ranges for λ_{eff} and R_b than those inferred using the constant model. Thus, we have stronger confidence about the estimated λ_{eff} and R_b. At the same time, the irregular shape and small magnitude of the model bias function of the variable heat rate model suggests that all major physical effects are considered in the model. Therefore, from the results of the Bayesian calibration, we gain clear insights into the model itself.

For the TRT2 dataset, which was affected by heavy rainfall and the resultant groundwater flow, the clear advantage of the Kennedy and O'Hagan framework is that it considers a model bias term that can capture the structural error in the model. The inferred posteriors of λ_{eff} and R_b differ significantly from the values estimated using the standard semi-log plot method. Moreover, the inferred posteriors of TRT2 also differ significantly from the TRT1 results and past TRT results. This means that the obtained model bias function does not fully compensate the effect of advection, and the part not considered by the bias function still leads to parameter overfitting, albeit less significantly than with the standard method. To improve the methodology, employing a model bias that allows more complex specifications or fine-tuning of the prior distributions for the Gaussian process model bias can be considered in the future. Alternatively, different analytical solutions, such as the moving line source models, which consider groundwater flow, could be employed for case studies where sufficient prior knowledge about the additionally required model parameters is available.

Acknowledgements

This work was supported by the Japan Society for the Promotion of Science (JSPS) (KAKENHI, grant numbers 26709041 and P16074).

References

[1] ISO. Guide to the Expression of Uncertainty in Measurement. Geneva, Switzerland: International Organization for Standardization; 1995.

[2] Bakirci K, Colak D. Effect of a superheating and sub-cooling heat exchanger to the performance of a ground source heat pump system. Energy 2012;44:996–1004. doi:10.1016/j.energy.2012.04.049.

[3] Shang Y, Dong M, Li S. Intermittent experimental study of a vertical ground source heat pump system. Appl Energy 2014;136:628–35. doi:10.1016/j.apenergy.2014.09.072.

[4] Xi C, Hongxing Y, Lin L, Jinggang W, Wei L. Experimental studies on a ground coupled heat pump with solar thermal collectors for space heating. Energy 2011;36:5292–300. doi:10.1016/j.energy.2011.06.037.

[5] Naili N, Hazami M, Attar I, Farhat A. In-field performance analysis of ground source cooling system with horizontal ground heat exchanger in Tunisia. Energy 2013;61:319–31. doi:10.1016/j.energy.2013.08.054.

[6] Dehkordi SE, Schincariol RA. Effect of thermal-hydrogeological and borehole heat exchanger properties on performance and impact of vertical closed-loop geothermal heat pump systems. Hydrogeol J 2013;22:189–203. doi:10.1007/s10040-013-1060-6.

[7] Bujok P, Grycz D, Klempa M, Kunz A, Porzer M, Pytlik A, et al. Assessment of the influence of shortening the duration of TRT (thermal response test) on the precision of measured values. Energy 2014;64:120–9. doi:10.1016/j.energy.2013.11.079.

[8] Witte HJL. Error analysis of thermal response tests. Appl Energy 2013;109:302–11. doi:10.1016/j.apenergy.2012.11.060.

[9] Carslaw HS, Jaeger JC. Conduction of Heat in Solids. 2nd ed. UK: Oxford University Press; 1959.

[10] Ingersoll LR, Zobel OJ, Ingersoll AC. Heat conduction, with engineering and geological applications. McGraw Hill Book Company Inc.; 1948.

[11] Shonder JA, Beck J. Field test of a new method for determining soil formation thermal conductivity and borehole resistance. ASHRAE Trans 2000;106:843–50.

[12] Hu P, Meng Q, Sun Q, Zhu N, Guan C. A method and case study of thermal response test with unstable heat rate. Energy Build 2012;48:199–205. doi:10.1016/j.enbuild.2012.01.036.

[13] Sharqawy MH, Mokheimer EM, Habib MA, Badr HM, Said SA, Al-Shayea NA. Energy, exergy and uncertainty analyses of the thermal response test for a ground heat exchanger. Int J Energy Res 2009;33:582–92. doi:10.1002/er.1496.

[14] Li S, Dong K, Wang J, Zhang X. Long Term Coupled Simulation for Ground Source Heat Pump and Underground Heat Exchangers. Energy Build 2015. doi:10.1016/j.enbuild.2015.05.041.

[15] Signorelli S. Geoscientific investigations for the use of shallow low-enthalpy systems. SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH, 2004.

[16] Witte HJL, Van Gelder GJ, Spitler JD. In situ measurement of ground thermal conductivity: A Dutch perspective. ASHRAE Trans 2002;108:263–72.

[17] Choi W, Ooka R. Effect of disturbance on thermal response test, part 2: Numerical study of applicability and limitation of infinite line source model for interpretation under disturbance from outdoor environment. Renew Energy 2016;85:1090–105. doi:10.1016/j.renene.2015.07.049.

[18] Bandos T V, Montero Á, Fernández de Córdoba P, Urchueguía JF. Improving parameter estimates obtained from thermal response tests: Effect of ambient air temperature variations. Geothermics 2011;40:136–43. doi:10.1016/j.geothermics.2011.02.003.

[19] Roth P, Georgiev A, Busso A, Barraza E. First in situ determination of ground and borehole thermal properties in Latin America. Renew Energy 2004;29:1947–63. doi:10.1016/j.renene.2004.02.014.

[20] Choi W, Ooka R. Interpretation of disturbed data in thermal response tests using the infinite line source model and numerical parameter estimation method. Appl Energy 2015;148:476–88. doi:10.1016/j.apenergy.2015.03.097.

[21] Sutton MG, Nutter DW, Couvillion RJ. A ground resistance for vertical bore heat exchangers with groundwater flow. J Energy Resour Technol Asme 2003;125:183–9. doi:Doi 10.1151/1.1591203.

[22] Diao N, Li Q, Fang Z. Heat transfer in ground heat exchangers with groundwater advection. Int J Therm Sci 2004;43:1203–11. doi:10.1016/j.ijthermalsci.2004.04.009.

[23] Molina-Giraldo N, Blum P, Zhu K, Bayer P, Fang Z. A moving finite line source model to simulate borehole heat exchangers with groundwater advection. Int J Therm Sci 2011;50:2506–13. doi:10.1016/j.ijthermalsci.2011.06.012.

[24] Hu J. An improved analytical model for vertical borehole ground heat exchanger with multiple-layer substrates and groundwater flow. Appl Energy 2017;202:537–49. doi:10.1016/j.apenergy.2017.05.152.

[25] Box GE. Robustness in the Strategy of Scientific Model Building. vol. 1. Academic Press; 1979.
[26] O'Hagan A. Bayesian analysis of computer code outputs: A tutorial. Reliab Eng Syst Saf 2006;91:1290–300. doi:10.1016/j.ress.2005.11.025.
[27] Kennedy MC, O'Hagan A. Bayesian Calibration of Computer Models. J R Stat Soc Ser B (Statistical Methodol 2001;63:425–64. doi:10.1111/1467-9868.00294.
[28] Choi W, Ooka R. Effect of natural convection on thermal response test conducted in saturated porous formation: Comparison of gravel-backfilled and cement-grouted borehole heat exchangers. Renew Energy 2016;96:891–903. doi:10.1016/j.renene.2016.05.040.
[29] Heo Y, Choudhary R, Augenbroe G a. Calibration of building energy models for retrofit analysis under uncertainty. Energy Build 2012;47:550–60. doi:10.1016/j.enbuild.2011.12.029.
[30] Guillas S, Rougier J, Maute A, Richmond AD, Linkletter CD. Bayesian calibration of the Thermosphere-Ionosphere Electrodynamics General Circulation Model (TIE-GCM). Geosci Model Dev Discuss 2009;2:485–506. doi:10.5194/gmdd-2-485-2009.
[31] Iman RL, Helton JC, Campbell JE. An Approach to Sensitivity Analysis of Computer Models: Part I - Introduction, Variable Selection and Preliminary Variable Assessment. J Qual Technol 1981;13:174–83.
[32] McKay MD, Beckman RJ, Conover WJ. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics 1979;21:239. doi:10.2307/1268522.
[33] Rasmussen CE, Williams CKI. Gaussian processes for machine learning. 2006.
[34] Higdon D, Kennedy M, Cavendish JC, Cafeo JA, Ryne RD. Combining Field Data and Computer Simulations for Calibration and Prediction. SIAM J Sci Comput 2004;26:448–66. doi:10.1137/S1064827503426693.
[35] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. J Chem Phys 1953;21:1087–92. doi:10.1063/1.1699114.
[36] Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 1970;57:97–109. doi:10.1093/biomet/57.1.97.
[37] Chib S, Greenberg E. Understanding the Metropolis-Hastings Algorithm. Am Stat 1995;49:327. doi:10.2307/2684568.
[38] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. CRC Press; 2014.
[39] Signorelli S, Bassetti S, Pahud D, Kohl T. Numerical evaluation of thermal response tests. Geothermics 2007;36:141–66. doi:10.1016/j.geothermics.2006.10.006.
[40] Borinaga-Treviño R, Norambuena-Contreras J, Castro-Fresno D. How to correct the ambient temperature influence on the thermal response test results. Appl Therm Eng 2015;82:39–47. doi:10.1016/j.applthermaleng.2015.02.050.
[41] Abdelaziz SL, Olgun CG, Martin JR. Counterbalancing ambient interference on thermal conductivity tests for energy piles. Geothermics 2015;56:45–59. doi:10.1016/j.geothermics.2015.03.005.
[42] Choi W, Ooka R. Effect of disturbance on thermal response test, part 1: Development of disturbance analytical model, parametric study, and sensitivity analysis. Renew Energy 2016;85:306–18. doi:10.1016/j.renene.2015.06.042.
[43] Choi W, Kikumoto H, Choudhary R, Ooka R. Bayesian inference for thermal response test parameter estimation and uncertainty assessment. Appl Energy 2018;209:306–21. doi:10.1016/j.apenergy.2017.10.034.
[44] Bear J. Dynamics of fluids in porous media. Dover Publications; 1988.
[45] Allen DM, Whitfield PH, Werner A. Groundwater level responses in temperate mountainous terrain: Regime classification, and linkages to climate and streamflow. Hydrol Process 2010;24:3392–412. doi:10.1002/hyp.7757.
[46] Rodhe A, Seibert J. Groundwater dynamics in a till hillslope: flow directions, gradients and delay. Hydrol Process 2011;25:1899–909. doi:10.1002/hyp.7946.

Appendix A. Comparison between KOH Bayesian calibration and Bayesian inference

In our previous work [43], we developed a Bayesian inference technique for the TRT parameter estimation in which the inference was conducted by considering only the random error term. Thus, it would be interesting to compare the inferred posteriors of λ_{eff} and R_b from the previous and current studies. Hereafter, we refer to the previous and current frameworks as Bayesian inference (BI) and Bayesian calibration (BC), respectively. We recall that the most significant difference between the two studies is the distinction between random error terms and a structured model bias function, which should prevent the inferred model parameters from overfitting (i.e., prevent absorption of model bias in the posteriors of inferred parameters θ). Because the number of MCMC in the previous study was 5×10^5 and no informative prior was assigned, 1.5×10^4 samplings with the same prior were newly conducted in this study to enable a fair comparison.

The posterior mean, maximum a posteriori, and 95% credible interval of Bayesian inference and calibration are summarized in Table A1. In should be noted that the listed BC values in Table A1 are from PPDFs shown in Fig. 5. Compared to the mode of Bayesian inference, a slightly lower mode for λ_{eff} is obtained from Bayesian calibration, and the R_b values are nearly identical.

We can observe differences between the PM for calibration parameters from BI and BC. This difference is most likely caused by the absorption of the model discrepancy into the PPDF of the calibration parameters for BI, which only accounts for random error. This finding is also supported by the narrower 95% CIs of PPDFs for BI than those for BC, which signifies the overfitting of the calibration parameters. In addition, the unknown hyperparameters of the covariance function introduce a larger uncertainty to the joint posterior distribution in BC and accordingly also to the marginal distributions of the calibration parameters.

Table A1. Characteristics of posterior distributions of calibration parameters obtained by Bayesian calibration (BC) with constant and variable models, and by Bayesian inference (BI) with constant model.

Estimator	Method and model	Thermal conductivity [W/(m·K)]	Thermal resistance [m·K/W]
Posterior mean	BC constant	1.97	0.154
	BC variable	1.95	0.150
	BI constant	1.92	0.147
Maximum a posteriori	BC constant	1.91	0.155
	BC variable	1.93	0.151
	BI constant	1.91	0.147
95% credible interval	BC constant	1.80–2.16	0.144–0.164
	BC variable	1.82–2.07	0.142–0.158
	BI constant	1.87–2.00	0.144–0.152