MANUSCRIPT SUBMITTED TO SCIENCE AND TECHNOLOGY FOR THE BUILT ENVIRONMENT

Optimal discretization of geothermal boreholes for the calculation of *g*-functions

Massimo Cimmino[a], Jonathan C. Cook[b] and José A. Isiordia Farrera[a]

[a]Département de gnie mcanique, Polytechnique Montréal, Case postale 6079, Succursale "Centre-ville", Montréal, Qubec, H3C 3A7, Canada; [b]Independent researcher.

ARTICLE HISTORY
Compiled December 1, 2023

ABSTRACT
The effect of the discretization of geothermal boreholes on the accuracy of *g*-function evaluations is studied. A data set of 557,056 bore field configurations covering a large range of geometrical parameters is generated using `pygfunction`. A non-uniform discretization of borehole segments geometrically expanding in length towards the middle of the borehole is proposed. A non-uniform discretization is shown to achieve better accuracy than a uniform discretization. The non-uniform discretization is optimized to minimize the maximum absolute percentage error over the entire data set. The discretization is optimized for each bore field configuration, and an artifical neural network (ANN) is trained to predict the optimal discretization given only geometrical and thermal parameters of the boreholes, excluding the borehole positions. Thermal parameters that quantify the bore field temperature distribution are introduced as inputs to the ANN. The maximum absolute percentage error using a uniform discretization is 99.0% in the worst studied case of a dense rectangular field of $N_b = 1116$ boreholes with lengths of 418.8 m and spacings of 3.14 m and 3.18 m along rows and columns, while only 1% of the cases feature an error above 26.7%. The error is reduced to 3.6% using the global optimal discretization and 3.3% using the ANN.

KEYWORDS
g-Functions; Thermal response factors; Finite line source; Geothermal boreholes; Ground-source heat pumps

1. Introduction

The simulation of ground-source heat pump (GSHP) systems relies on models of the transient heat transfer in the soil to predict the variation of the heat carrier fluid temperature during the systems' operation. The returning fluid temperature from the bore field affects the efficiency of the heat pump, and imposes limits on the heat that can be extracted to maintain safe and efficient operation of a system. Dimensionless step-response functions, often called *g*-functions or thermal response factors, give the variation of the effective temperature at the walls of the geothermal boreholes that comprise the bore field, given a known total heat extraction (or injection) rate. *g*-Functions depend on the configuration and size of the boreholes. Simulations can be

CONTACT M. Cimmino. Email: massimo.cimmino@polymtl.ca

performed iteratively to determine a suitable configuration and size for the boreholes. Given the central part of g-functions in the sizing of geothermal bore fields, their accurate evaluation is crucial to ensure proper operation of GSHP systems and attain cost-effective designs.

Eskilson (1987) developed the superposition borehole model (SBM) to evaluate g-functions of fields of geothermal boreholes. Each borehole is modelled individually in a radial-axial grid, and a finite difference method is applied to solve the transient heat conduction around the boreholes. The total temperature variation at the borehole walls is obtained by spatial superposition of the different grids. At each timestep, the profiles of heat extraction rate per unit borehole length required to maintain a uniform temperature along the walls of the boreholes and a constant total heat extraction rate are calculated. The boundary condition of uniform temperature is a key element of the method and is meant to represent boreholes connected in parallel that receive the heat carrier fluid at the same temperature.

g-Functions are not necessarily associated with the boundary condition of uniform temperature. Methods that rely on analytical solutions were introduced to evaluate g-functions, and are now often preferred due to the simplicity of their implementation and their reduced computational costs. The finite line source (FLS) solution allows to evaluate the temperature variation in the soil around a line that extracts heat at a constant and uniform rate (i.e. using a boundary condition of uniform heat extraction rate). Eskilson (1987) used the FLS solution to evaluate the g-function of a single borehole. The FLS Solution was later reintroduced by Zeng, Diao, and Fang (2002), who proposed the use of the average (over the length) temperature variation and spatial superposition to consider heat interactions between boreholes. Single integral formulations were later proposed by Lamarche and Beauchamp (2007) and Claesson and Javed (2011). Cimmino (2021) developed an approximation of the FLS solution, based on an approximation of the error function, that enables an analytical solution of the integral and removes the need for numerical integration. Fossa (2011) showed that the g-functions calculated using the FLS solution and a boundary condition of uniform heat extraction rate overpredict g-functions calculated using a boundary condition of uniform temperature.

This issue was addressed by Cimmino, Bernier, and Adams (2013) and Cimmino and Bernier (2014). The authors split the boreholes into segments and modelled each of the segments using a finite line source. A system of equations is obtained by spatial and temporal superposition of the FLS solution, considering equal temperatures at all segments and a constant total heat extraction rate. This method resulted in the calculation of g-functions that mimicked Eskilson's. Cimmino and Bernier (2014) suggested the use of $n_s = 12$ equal segment lengths based on the analysis of a field of 100 boreholes. Lazzarotto (2016) extended the method to fields of inclined boreholes, building upon prior solutions of the FLS for inclined line sources (Cui, Yang, & Fang, 2006; Lamarche, 2011; Marcotte & Pasquier, 2009). Lamarche (2017) extended the FLS solution to consider piecewise linear variations of the heat extraction rate along the length of boreholes and decreased the number of segments to $n_s = 8$. Ongoing research on the calculation of g-functions using analytical methods is conducted mainly in the pursuit of two objectives : (1) to decrease the computational cost for the evaluation of g-functions, and (2) to increase the complexity of the modelled physical processes.

Cimmino (2018a) proposed a method to reduce the number of evaluations of the FLS solution by identifying similarity conditions in the FLS solution where its solution takes the same value for multiple pairs of segments along the length of two boreholes.

The number of evaluations is further reduced by grouping radial distances for the evaluation of the FLS solution, under a tolerance on the radial distance. Dusseault, Pasquier, and Marcotte (2018) used Chebychev polynomials to estimate the FLS solution and eliminate the requirement for numerical integration. Dusseault and Pasquier (2019) used this method to generate a data set of g-functions and train an artificial neural network (ANN) to evaluate g-functions. Nguyen and Pasquier (2021) interpolated the FLS solution along the radial coordinate and in time to reduce the number of evaluations of the FLS solution. The system of equations is then solved using an iterative algorithm. Prieto and Cimmino (2021) developed the equivalent borehole method, which consists in grouping boreholes that are expected to share similar temperatures and heat extraction rates. Each group is then modelled using a single equivalent borehole, and the size of the system of equations is thereby significantly reduced – and so is the calculation time.

Marcotte and Pasquier (2014) calculated g-functions of fields of parallel- and series-connected boreholes by considering the borehole resistances and energy balances to account for the influence of hydraulic connections on the thermal response of the bore field. They used $n_s = 1$ segment per borehole in their model. Cimmino (2015) coupled the FLS solution to a thermal circuit model of the interior of a borehole to evaluate g-functions of boreholes using a boundary condition of equal inlet fluid temperature while considering both the variations of borehole wall temperatures and heat extraction rates along the boreholes. Cimmino (2019) extended the method to model series-connections between boreholes. The concepts of effective borehole wall temperature and effective bore field resistance were introduced so that the evaluated g-functions may be used in the same manner as the original g-functions of Eskilson (1987). Molina-Giraldo, Blum, Zhu, Bayer, and Fang (2011) proposed a moving finite line source (MFLS) solution to consider groundwater advection around boreholes. Li and Lai (2012) extended the FLS solution to consider anisotropic soil thermal properties. Abdelaziz, Ozudogru, Olgun, and Martin (2014) developed a multi-layer FLS method to account for horizontal soil layers with different thermal properties. Hu (2017) and Erol and Franois (2018) developed multi-layer FLS methods that account for horizontal layers with different thermal properties and groundwater advection.

Cimmino (2018b) developed `pygfunction`, an open-source Python package for the calculation of g-functions. `pygfunction` implements different boundary conditions (Cimmino, 2019; Cimmino & Bernier, 2014) and solution algorithms (Cimmino, 2018a; Prieto & Cimmino, 2021). The latest developments to `pygfunction` were presented in the precursory conference paper (Cimmino & Cook, 2022). The conference paper presented the evolution of the accuracy, computational time and memory usage of the package from version 1.0 to the current version 2.2. A 4 orders of magnitude reduction in calculation time was obtained for large irregular bore field configurations, with the g-function of a field of 400 randomly-positioned boreholes evaluated in 0.245 seconds.

The accuracy of g-functions is dependent on the number of segments used in the discretization of the boreholes. Figure 1 shows the g-function of a rectangular field of 25×25 boreholes ($N_b = 625$ boreholes), evaluated with $n_s = 1$, 12 and 96 segments of equal length. As shown, g-functions calculated using only 1 segment can significantly overpredict the g-function. Using a uniform discretization of 12 segments can also produce large errors in such cases were the number of boreholes (N_b) is large, the spacing (B) is low, and the buried depth (D) is low. The recommendation of Cimmino and Bernier (2014) of using a uniform discretization of 12 segments, which was inferred from the analysis of a smaller field of 100 boreholes with fixed geometrical parameters, may not always be sufficient.

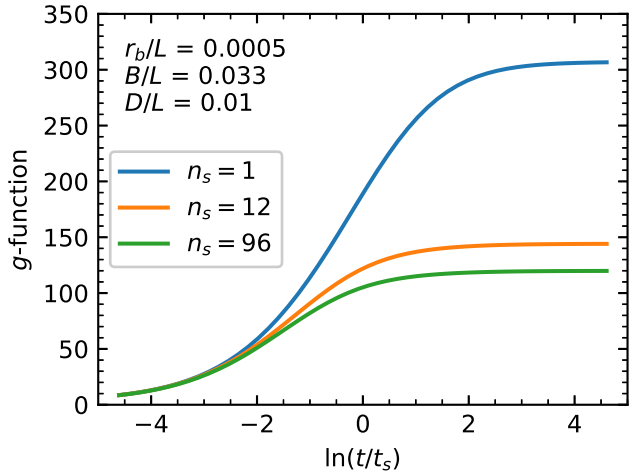

Figure 1. g-Functions of a field of 25 × 25 boreholes ($N_b = 625$) calculated with 1, 12 and 96 segments

Cook (2021) proposed a non-uniform discretization of the boreholes to evaluate g-functions and showed that using a non-uniform discretization of 8 segments along the boreholes, rather than a uniform discretization of 12 segments, decreases the mean percent error on the g-functions and the calculation time. The reduction in the error is explained by the smaller segments concentrated towards the ends of the boreholes, which allows a better representation of the sharp variations in heat extraction rates at these locations. The reduction in calculation time (and memory usage) is due to the decrease in the number of segment pairs, which decreases the number of evaluations of the finite line source solution and the size of the system of equations that is solved. Cook (2021) also showed that the number of segments required in a uniform discretization to achieve an acceptable accuracy increased with the number of boreholes and with the density of boreholes within the bore field. West and Spitler (2022) proposed optimal discretizations of 3 segments to obtain quick evaluations of g-functions in the context of iterative design of irregular bore fields. They suggest using predefined borehole lengths and discretizations associated with each of the lengths to evaluate g-functions, and to use interpolation to evaluate g-functions outside of these lengths. In their analysis of 837 bore fields ranging from 50 to 1000 boreholes, the maximum error on the g-functions was less than 5% and the corresponding error on the design length of the boreholes was less than 1%. The decrease in the number of segments achieves an order of magnitude improvement in calculation time. Although their proposed discretization is only dependent on the length of the boreholes, the authors noted that the optimal discretization should also be dependent on the field shape, field spacing, the number of boreholes, as well as other factors.

The present work provides a comprehensive study of the impact of the discretization on the accuracy of g-functions using a broad range of geometrical parameters for the bore fields and large bore fields of up to 1600 boreholes. It is shown that bore field configuration-dependent optimal non-uniform discretizations can be evaluated from an artificial neural network by the introduction of thermal density parameters based on the steady-state finite line source solution in addition to the geometrical parameters of the boreholes.

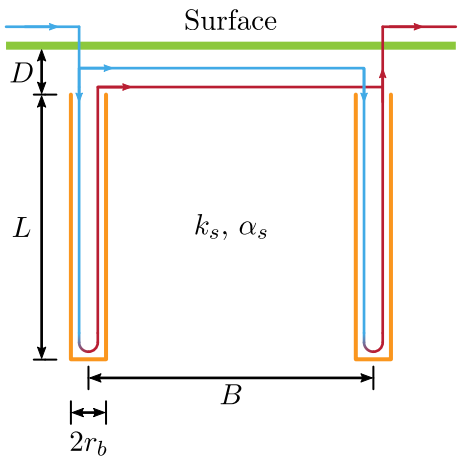

Figure 2. Geometry of a field of $N_b = 2$ boreholes

2. Methodology

g-Functions of bore fields are calculated using the method of equivalent boreholes (Prieto & Cimmino, 2021). A data set of 557,056 reference g-functions with varied configurations and geometrical parameters is calculated using finely discretized boreholes. A single vertical discretization is applied to all boreholes in a particular field. Various options for the discretization of boreholes are introduced and the resulting error on the g-functions are compared. An optimal (under constraints) non-uniform 8 segment discretization is determined for each bore field by minimizing the maximum absolute percentage error. A single globally-optimal (under constraints) non-uniform 8 segment discretization is then evaluated to minimize the maximum absolute percentage error on the entire data set. Finally, a neural-network is trained to estimate the optimal non-uniform discretization of any individual bore field based on geometrical and thermal parameters of the boreholes.

The geometry of a field of $N_b = 2$ boreholes is presented on figure 2. Boreholes have the same length L, radius r_b, and are buried at a depth D from the ground surface. The soil has a homogeneous and isotropic thermal conductivity k_s, thermal diffusivity α_s, and is initially at a uniform undisturbed temperature T_0. All boreholes are connected in parallel and receive the heat carrier fluid at the same inlet temperature. The 2 boreholes represented on figure 2 are spaced at a distance B, which corresponds to the radial distance between their horizontal coordinates (x_i, y_i).

2.1. g-Functions

2.1.1. Definition and use

g-Functions are step-response functions that give the relation between the time-variation of the effective borehole wall temperature to a unit constant total heat extraction rate in a bore field. For time-varying heat extraction rates, the time-variation of the effective borehole wall temperature is evaluated by temporal superposition using

a convolution product between the heat extraction rates and the g-function:

$$T_b^*(t) = T_0 - \frac{1}{2\pi k_s} \int_0^t Q'(t) \frac{dg}{d\tau}(t-\tau) d\tau \tag{1}$$

where T_b^* is the effective borehole wall temperature, Q' is the average heat extraction rate per unit borehole length and g is the g-function.

The effective borehole wall temperature is an effective temperature for heat exchange between the heat carrier fluid and the ground. It is representative – but not necessarily equal to – of the average temperature at the wall of the boreholes (i.e. the interface between the boreholes and the soil). Its mathematical definition is tied to that of the effective borehole thermal resistance:

$$Q'(t) = \frac{T_b^*(t) - \bar{T}_f(t)}{R_b^*} \tag{2a}$$

$$\bar{T}_f = \frac{1}{2}\left(T_{f,in} + T_{f,out}\right) \tag{2b}$$

where $\bar{T}_f$ is the mean fluid temperature, $T_{f,in}$ is the inlet fluid temperature, $T_{f,out}$ is the outlet fluid temperature, and R_b^* is the effective borehole thermal resistance. The effective borehole thermal resistance is evaluated from the fluid temperature profile and heat extraction rate considering a uniform borehole wall temperature (Hellström, 1991). If boreholes are not connected in parallel, the effective borehole thermal resistance is replaced by the effective bore field thermal resistance (Cimmino, 2019).

From equation 1, dimensionless variables can be introduced for the effective borehole wall temperature and heat extraction rate per unit borehole length:

$$\theta_b^*(t) = \frac{T_b^*(t) - T_0}{-Q'_0/2\pi k_s} \tag{3a}$$

$$\phi'(t) = \frac{Q'(t)}{Q'_0} \tag{3b}$$

where θ_b^* is the dimensionless effective borehole wall temperature, ϕ' is the normalized heat extraction rate per unit borehole length, and Q'_0 is an arbitrary reference heat extraction rate per unit borehole length.

Any temperature and heat extraction rate per unit length can be nondimensionalized in the same manner as presented in equation 3. The dimensionless time $\tau = t/t_s$ is evaluated from the bore field characteristic time $t_s = L^2/9\alpha_s$. Dimensionless geometrical parameters are obtained by normalization with regards to the borehole length: r_b/L the dimensionless radius, D/L the dimensionless buried depth, and B/L the dimensionless borehole spacing. While the dimensionless borehole spacing B/L is often used for the presentation of g-functions, it does not mean that bore fields with irregular spacing cannot be modelled using g-functions. In this case, dimensionless borehole coordinates $(x^*, y^*) = (x/L, y/L)$ become relevant in the g-function definition.

In simulations with discrete timesteps, the convolution product in equation 1 is replaced with a summation. Assuming piece-wise constant heat extraction rates per

unit borehole length:

$$T_b^*(t_k) = T_0 - \frac{1}{2\pi k_s} \sum_{p=1}^{k} Q'(t_p) \left(g(t_k - t_{p-1}) - g(t_k - t_p) \right) \tag{4a}$$

$$\theta_b^*(t_k) = \sum_{p=1}^{k} \phi'(t_p) \left(g(t_k - t_{p-1}) - g(t_k - t_p) \right) \tag{4b}$$

where $Q'(t_p)$ is a constant heat extraction rate per unit borehole length prevailing over the interval $t_{p-1} < t \leq t_p$, and $\phi'(t_p)$ is its dimensionless counterpart. Once the effective borehole wall temperature is known, the inlet and outlet fluid temperatures can be evaluated from equation 2.

2.1.2. Method of equivalent boreholes

The method of equivalent boreholes, introduced by Prieto and Cimmino (2021), is used in this paper to evaluate g-functions. The method is summarized, as some elements are relevant to the artificial neural network model that will be used to estimate optimal borehole discretizations. For an in-depth presentation of the method, the reader is referred to the work of Prieto and Cimmino (2021).

The method relies on the spatial and temporal superposition of the finite line source (FLS) solution to evaluate the temperature variation in the soil caused by heat extraction at the boreholes. Each borehole is discretized into a series of n_s segments, and each borehole segment is represented by a finite line source of equal length located at its axis. In previous approaches, the temperature variation at each segment caused by the heat extraction at all other segments, individually, are evaluated using segment-to-segment thermal response factors derived from the FLS solution (Cimmino, 2015, 2018a, 2019; Cimmino & Bernier, 2014; Lamarche, 2017; Lazzarotto, 2016; Lazzarotto & Björk, 2016). In the method of equivalent boreholes, boreholes are classified into groups that are expected to share similar temperatures and heat extraction rates, and group-to-group thermal response factors are instead derived from the FLS solution. This is a generalization of the concept of bore field symmetries considered by Eskilson (1987) in their numerical model and by Cimmino et al. (2013) in their analytical model based on the FLS solution.

Figure 3 illustrates group-to-group thermal response factors for the interaction between groups of borehole segments. Borehole j, located in the center, is the sole member of borehole group $\mathcal{J}$. Boreholes i_1 and i_2, the only members of borehole group $\mathcal{I}$ are located on each side of borehole j. The group-to-group thermal response factor between two segments u and v of boreholes in groups $\mathcal{I}$ and $\mathcal{J}$, respectively, is given by the average of the total (superimposed) temperature variation on boreholes in group $\mathcal{I}$ caused by all boreholes in group $\mathcal{J}$. Group-to-group thermal response factors can

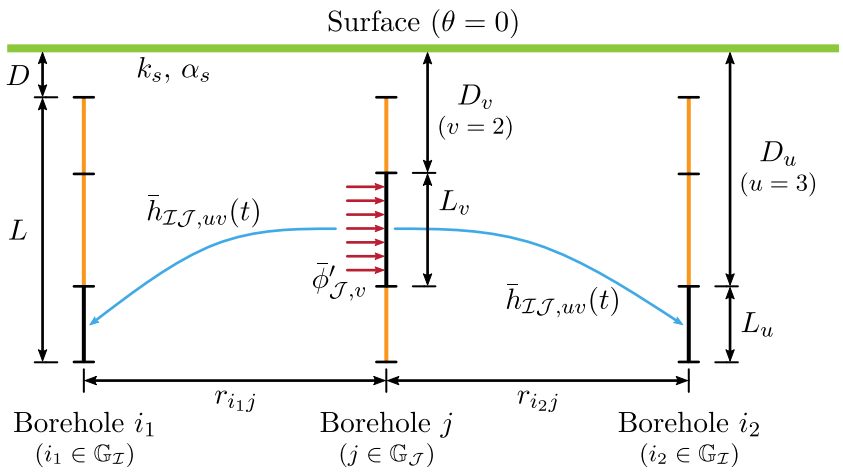

Figure 3. Group-to-group thermal response factors

then be calculated from the finite line source solution:

$$\bar{h}_{\mathcal{IJ},uv}(t) = \frac{1}{2L_u n_{b,\mathcal{I}}} \int_{\frac{1}{\sqrt{4\alpha_s t}}}^{\infty} \sum_{i\in\mathbb{G}_\mathcal{I}} \sum_{j\in\mathbb{G}_\mathcal{J}} \frac{1}{s^2} \exp(-r_{ij}^2 s^2) F_{uv}(s) ds \tag{5a}$$

$$\begin{aligned} F_{uv}(s) = {} & \operatorname{erfint}((D_u - D_v + L_u)s) - \operatorname{erfint}((D_u - D_v)s) \\ & + \operatorname{erfint}((D_u - D_v - L_v)s) - \operatorname{erfint}((D_u - D_v + L_u - L_v)s) \\ & + \operatorname{erfint}((D_u + D_v + L_u)s) - \operatorname{erfint}((D_u + D_v)s) \\ & + \operatorname{erfint}((D_u + D_v + L_v)s) - \operatorname{erfint}((D_u + D_v + L_u + L_v)s) \end{aligned} \tag{5b}$$

$$\operatorname{erfint}(x) = \int_0^x \operatorname{erf}(x')dx' = x\operatorname{erf}(x) - \frac{1}{\sqrt{\pi}}(1 - \exp(-x^2)) \tag{5c}$$

where $\bar{h}_{\mathcal{IJ},uv}$ is the group-to-group thermal response factor that gives the relation between the temperature variation of a segment u of an equivalent borehole $\mathcal{I}$ caused by heat extracted along segments v of boreholes in group $\mathcal{J}$, $r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ is the radial distance between boreholes i and j (with $r_{ii} = r_b$), $n_{b,\mathcal{I}}$ is the number of boreholes in group $\mathcal{I}$, L_u is the length of segment u, and D_u is the buried depth of segment u. The summation over the set of boreholes in group $\mathcal{J}$, $\mathbb{G}_\mathcal{J}$, represents the spatial superposition of the FLS. The summation over the set of boreholes in group $\mathcal{I}$, $\mathbb{G}_\mathcal{I}$, serves to calculate the average.

An equivalent borehole is representative of an average borehole within its group. For the purpose of evaluating thermal interactions between boreholes, it is assumed that all boreholes in a group share the same heat extraction rates per unit borehole length as the equivalent borehole. The spatial and temporal superposition of group-to-group thermal response factors for equivalent boreholes is thus given by:

$$\bar{\theta}_{b,\mathcal{I},u}(t_k) = \sum_{\mathcal{J}=1}^{N_g} \sum_{v=1}^{n_s} \sum_{p=1}^{k} \bar{\phi}'_{\mathcal{J},v}(t_p) \left(\bar{h}_{\mathcal{IJ},uv}(t_k - t_{p-1}) - \bar{h}_{\mathcal{IJ},uv}(t_k - t_p)\right) \tag{6}$$

where $\bar{\theta}_{b,\mathcal{I},u}$ is the dimensionless borehole wall temperature along segment u of equivalent borehole $\mathcal{I}$, $\bar{\phi}_{\mathcal{J},v}$ is the normalized heat extraction rate per unit borehole length

along segment v of equivalent borehole $\mathcal{J}$, and N_g is the number of borehole groups.

Boundary conditions are necessary to complete the model and evaluate the unit-step response. A boundary condition of uniform temperature is imposed, as introduced by Eskilson (1987). Other, more detailed, boundary conditions can be applied using the method of equivalent boreholes, e.g. an equal inlet fluid temperature into the boreholes. The boundary condition of uniform temperature is still used here to simplify the model and focus on the effect of the borehole discretization on the accuracy of g-function evaluations. The boundary condition of uniform temperature is reasonable, given the following assumptions:

- Boreholes are connected in parallel and receive the fluid at the same inlet fluid temperature;
- The fluid mass flow rate is sufficiently high so that the difference between inlet and outlet temperatures is small and the mean fluid temperature is approximately uniform along the borehole length;
- The borehole thermal resistance is relatively small, and the borehole wall temperature is relatively close to the mean fluid temperature.

Cimmino (2015) showed that the difference between g-functions evaluated with a boundary condition of uniform temperature and with a boundary condition of equal inlet fluid temperature depends strongly on the dimensionless borehole thermal resistance ($\Omega_b = 2\pi k_s R_b$), and weakly on the fluid mass flow rate and short-circuit thermal resistance (R_a). The uniform temperature boundary condition underpredicted the value of equal inlet fluid temperature g-functions by less than 4.3% when $\ln(\Omega_b) < 0.23$ for a rectangular field of 10 by 10 boreholes. Cimmino (2019) later showed that the fluid mass flow rate may have a strong impact on g-functions for fields of boreholes connected in series.

The borehole wall temperature at all segments are thus considered equal, and the total heat extraction rate in the bore field is constant:

$$\theta_b^*(t_k) = \bar{\theta}_{b,\mathcal{I},u}(t_k) \tag{7a}$$

$$1 = \sum_{\mathcal{J}=1}^{N_g} \sum_{v=1}^{n_s} \frac{n_{b,\mathcal{J}}}{N_b} \frac{L_v}{L} \bar{\phi}'_{\mathcal{J},v}(t_p) \tag{7b}$$

Equations 6 and 7 yield, for each timestep, a linear system of $N_g n_s + 1$ equations with an equal number of unknowns, θ_b^* and $\bar{\phi}_{\mathcal{J},v}$. The solution of the system of equations gives the dimensionless effective borehole wall temperature θ_b^*, equal to the g-function ($g(t_k) = \theta_b^*(t_k)$).

2.1.3. Steady-state single-segment temperatures

Borehole groups are identified from classification using a hierarchical agglomerative clustering algorithm. The distance metric corresponds to the absolute difference between steady-state borehole wall temperatures evaluated using one segment per borehole and a uniform heat extraction rate per unit borehole length equal for all boreholes.

The steady-state borehole-to-borehole thermal response factor, h_{ij}^{∞}, is given by:

$$h_{ij}^{\infty} = \frac{1}{2L} \sum_{m=1}^{8} (-1)^{m+1} \left\{ w_m \ln \left[\left(w_m + \sqrt{w_m^2 + r_{ij}^2} \right) / r_{ij} \right] - \sqrt{w_m^2 + r_{ij}^2} \right\} \tag{8a}$$

$$w_m = \{L, 0, -L, 0, 2D + L, 2D, 2D + L, 2D + 2L\} \tag{8b}$$

Details of the solution to the steady-state borehole-to-borehole thermal response factor are provided in appendix.

The steady-state borehole wall temperature of any borehole, h_i^{∞}, is obtained from the spatial superposition of borehole-to-borehole thermal response factors:

$$h_i^{\infty} = \sum_{j=1}^{N_b} h_{ij}^{\infty} \tag{9}$$

The linkage criterion between groups $\mathcal{I}$ and $\mathcal{J}$ is given by the maximum absolute difference in steady-state borehole wall temperatures between any two boreholes i and j of the two groups:

$$\mathcal{L}(\mathcal{I}, \mathcal{J}) = \max_{i \in \mathbb{G}_{\mathcal{I}}, j \in \mathbb{G}_{\mathcal{J}}} \mathcal{M}(i, j) \tag{10a}$$

$$\mathcal{M}(i, j) = \left| h_i^{\infty} - h_j^{\infty} \right| \tag{10b}$$

The minimum number of groups is equal to the number of branches on the clustering dendrogram intersected by an horizontal cut at the mid-segment of the branch representing the maximum dissimilarity (i.e. the longest branch). A precision increment of 1 is added to the minimum number of groups for the calculation of g-functions. The steady-state borehole wall temperatures will be used later as thermal parameters of the bore fields in the development of the neural network model.

2.2. Selected bore fields

A large data set of g-functions is generated for five regular bore field configurations, shown on figure 4. Each of the bore field types, with the exception of the circle field, has 6 geometrical parameters : N_x and N_y the numbers of columns and rows, B_x and B_y the spacings along the columns and rows, L the borehole length, and D the borehole buried depth. Since the circle fields are axisymmetric, only one number $N = N_b$ and only one spacing B are required, which decreases the number of geometrical parameters to 4. The borehole radius is kept constant at $r_b = 0.075$ m. Note that despite this all dimensionless geometrical parameters identified under equation 3 are still varied.

The ranges of variation of the parameters are presented in table 1. Sets of parameters are generated using a Sobol sequence (Sobol, 1967; Stephen & Kuo, 2008) using the algorithm implemented in the `SciPy` (Virtanen et al., 2020) Python package. Rounding is applied to the N_x, N_y and N parameters to obtain integer-valued parameters. 2^{17} (= 131,072) configurations are generated for each of the rectangular, box-shaped, U-shape and L-shaped fields. For the circle field, the number of configurations is reduced to 2^{15} (= 32,768) due to the reduced number of parameters. The data set is then comprised of a total of 557,056 bore field configurations covering a

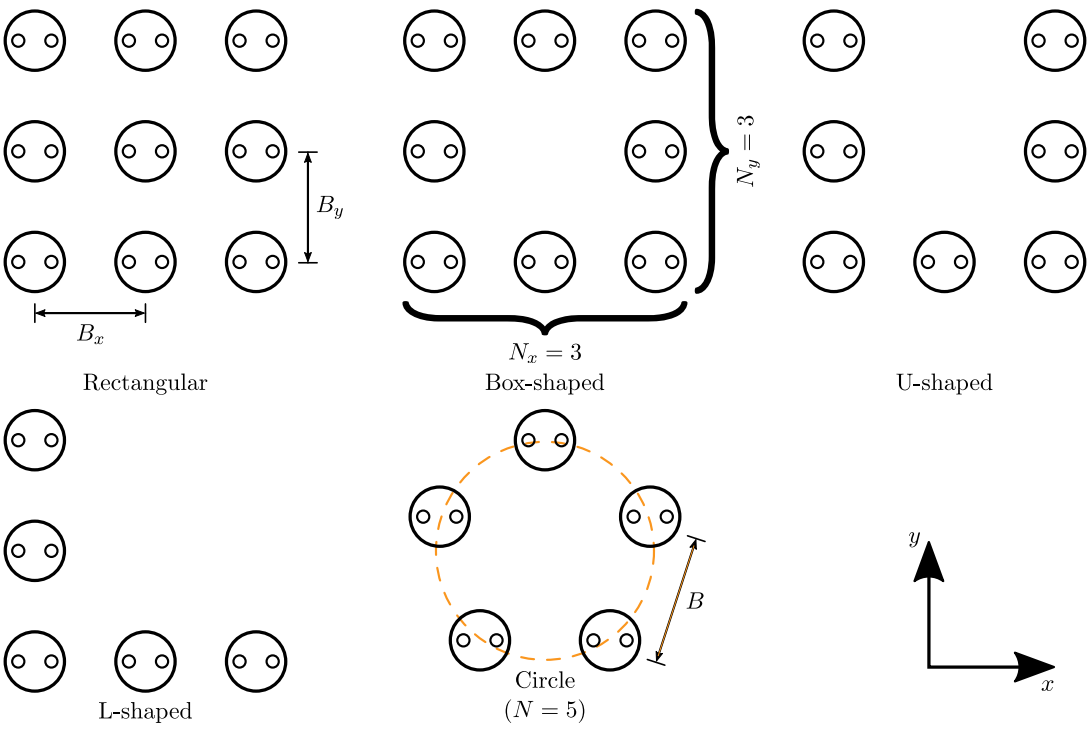

Figure 4. Rectangular, box-shaped, U-shaped, L-shaped and circle fields

Table 1. Parameters and bounds of the Sobol sequence

Parameters	Units	Lower bound	Upper bound
$N_{\{x,y\}}$	-	1	40
$B_{\{x,y\}}$	m	2.5	20
L	m	10	500
D	m	0.5	10

large range of geometrical parameters. The number of boreholes varies from 1 to 1600 in the rectangular bore field configurations.

2.3. Borehole discretization

The purpose of the discretization of boreholes into segments is to model the variation of the heat extraction rates along the length of the boreholes. The numerical model of Eskilson (1987) considered a non-uniform discretization where the lengths of the segments increased at a constant multiplying factor towards the middle of the boreholes. The multiplying factor was fixed at $\gamma = \sqrt{2}$ and the length of the segments was adjusted according to the total number of segments. The maximum number of segments considered was $n_s = 12$, presumably due to computational limits of the time. The choice of the multiplying factor is reasonable but not necessarily optimal with regards to the accuracy of the g-function calculation. Also, the length of the smallest segments located at the ends of the boreholes becomes very small when the number

of segments increases above 12.

An alternative to choosing and applying the multiplying factor, γ, is to choose the length of the smallest segments (i.e. at both ends of the borehole) and find the multiplying factor that satisfies the required number of segments. This discretization method was introduced by Cook (2021) and implemented into the `pygfunction` (Cimmino & Cook, 2022) Python package. The advantages of this discretization are that the smallest segment length is controlled and will not inadvertently become too small, and that the impact of the parameter on the discretization is more intuitive to the user.

The relationship between the length of the smallest segments and the multiplying factor is given by the polynomial equations:

$$0 = \begin{cases} (1-2\delta_1) - \delta_1 \cdot \gamma + 2\delta_1 \cdot \gamma^{n_s/2} & \text{if } n_s \text{ is even} \\ (1-2\delta_1) - \delta_1 \cdot \gamma + \delta_1 \cdot \gamma^{(n_s-1)/2} + 2\delta_1 \cdot \gamma^{(n_s-1)/2+1} & \text{if } n_s \text{ is odd} \end{cases} \tag{11}$$

where $\delta_1 = L_1/L$ is hereby called the end-length-ratio.

The positive polynomial root of equation 11 gives the multiplying factor for any chosen end-length-ratio and number of segments. Following the solution of equation 11, the individual segment length and buried depth are given by:

$$L_u = \begin{cases} L\delta_1\gamma^{u-1} & \text{if } u \leq n_s/2 \\ L\delta_1\gamma^{n_s-u} & \text{otherwise} \end{cases} \tag{12a}$$

$$D_u = D + \sum_{v=1}^{u} L_v \tag{12b}$$

2.3.1. Uniform, baseline and reference discretizations

Different values of the end-length-ratio will be compared with regards to the accuracy of the evaluation of the g-functions. A *uniform discretization* of $n_s = 12$ segments, as suggested by Cimmino and Bernier (2014), will be used to quantify the accuracy of a uniform discretization. This corresponds to an end-length-ratio $\delta_1 = \frac{1}{12} = 0.0833$ and multiplying factor $\gamma = 1$. The uniform discretization is first compared to a *baseline non-uniform discretization* of $n_s = 8$ segments with an end-length-ratio $\delta_1 = 0.02$. This end-length-ratio was adopted by Cook (2021) and implemented into `pygfunction` as a default option for discretization (Cimmino & Cook, 2022). The accuracy is quantified using the maximum absolute percentage error ($\varepsilon_{\max}$), evaluated based on the difference between the g-function and a reference g-function calculated using a *reference non-uniform discretization* of $n_s = 42$ segments with an end-length-ratio modulated proportionally to the baseline non-uniform discretization ($\delta_1 = 0.02 \cdot \frac{8}{42} = 0.0038$).

The reference non-uniform discretization is chosen based on a grid-independence analysis of non-uniform discretizations with end-length-ratios modulated proportionally to the baseline non-uniform discretization $\delta_1 = 0.02 \cdot \frac{8}{n_s}$. Figure 5 shows the variation of the maximum absolute percentage error of the g-function for three rectangular fields as a function of the number of segments. The number of boreholes in the x and y dimension for the fields are: 10 × 10 boreholes ($N_b = 100$), 25 × 25 boreholes ($N_b = 625$) and 40 × 40 boreholes ($N_b = 1600$). The bore fields consist of boreholes of length $L = 150$ m, buried depth $D = 4$ m and radius $r_b = 0.075$ m, with equal spacing $B_x = B_y = 7.5$ m along the columns and rows. The error is calculated

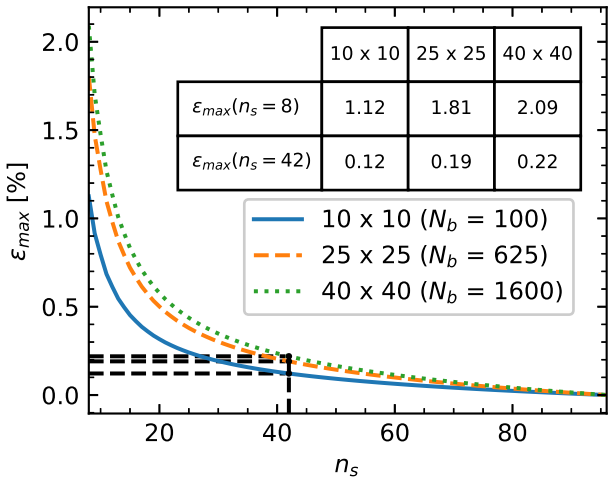

Figure 5. Maximum absolute percentage error of the g-function for three rectangular fields, containing 100, 625 and 1600 boreholes, as a function of the number of segments

with regards to a g-function evaluated with $n_s = 96$ segments and an end-length-ratio $\delta_1 = 0.02 \cdot \frac{8}{96} = 0.0017$. The error is reduced for each of the fields as the number of segments is increased. The number of boreholes and the proximity impacts the number of segments required to obtain an adequately converged grid independent solution. The largest field containing 1600 boreholes has the largest maximum error at each number of segments, n_s. The chosen reference non-uniform discretization of $n_s = 42$ segments presents a maximum absolute percentage error for the borefield containing 1600 boreholes of 0.22%, compared to 2.09% for the baseline non-uniform discretization of $n_s = 8$ segments. The choice of the reference discretization is also a compromise on the calculation time to allow the evaluation of reference g-functions for the entire data set; the calculation time is 9.2 seconds using $n_s = 42$ segments for the 1600 borehole rectangular field, and 184.2 seconds using $n_s = 96$ segments.

While the method of similarities (Cimmino, 2018a) provides a more accurate evaluation of g-functions, the method of equivalent boreholes (Prieto & Cimmino, 2021) is still used here to calculate reference g-functions. Using the method of equivalent boreholes for the reference g-functions decouples the errors caused by the method of equivalent boreholes to ensure focus is only on the discretization. The error thus reaches zero on figure 5 for $n_s = 96$ segments for each of the borehole fields.

Figure 6 shows the uniform discretization and baseline non-uniform discretizations. It can be seen that the non-uniform discretization has a finer discretization at both ends of the borehole, which better captures the expected sharp variations of heat extraction rates at these locations even with fewer segments. A third non-uniform discretization is shown on figure 6(c). This discretization corresponds to an end-length-ratio $\delta_1 = 0.005525$ for $n_s = 8$ segments (the two segments closest to each end are barely visible). This discretization results from the minimization of the maximum absolute percentage error over the entire data set, as will be presented in section 2.3.3, but is still illustrated here.

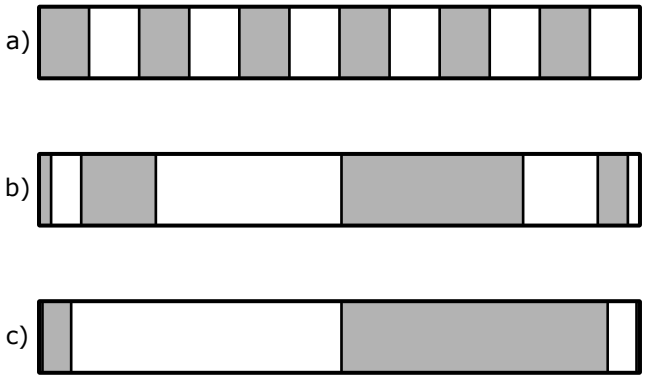

Figure 6. Discretization of boreholes : a) Uniform discretization of 12 segments, b) Baseline non-uniform discretization of 8 segments with $\delta_1 = 0.02$, and c) Globally optimal non-uniform discretization of 8 segments with with $\delta_1 = 0.005525$

2.3.2. Optimal discretization for each bore field

An optimal value for the end-length-ratio can be obtained for each bore field by minimizing the maximum absolute percentage error with respect to the reference g-function calculated with $n_s = 42$ segments. The optimization problem is given by:

$$\begin{cases} \text{Minimize:} & \varepsilon_{\max}(\delta_1^{(n)}) = \max\limits_{k} \frac{\left|g^{(n)}(t_k) - g_{ref}^{(n)}(t_k)\right|}{g_{ref}^{(n)}(t_k)} \cdot 100\% \\ \text{s.t.:} & \delta_{\min} \leq \delta_1^{(n)} \leq \frac{1}{n_s} \end{cases} \tag{13}$$

where $\varepsilon_{\max}$ is the maximum absolute percentage error, $\delta_1^{(n)}$ is the end-length-ratio that minimizes the error for a bore field n, $g^{(n)}$ is the g-function calculated with the non-uniform discretization given by $\delta_1^{(n)}$, $g_{ref}^{(n)}$ is the reference g-function for the same bore field, and $\delta_{\min} = 10^{-5}$ is a lower bound to avoid the case $\delta_1^{(n)} \leq 0$. The upper bound, $1/n_s$, ensures that the length of the segments increases towards the middle of the boreholes.

The optimization problem is solved using the Nelder-Mead algorithm implemented in `SciPy` (Gao & Han, 2012; Virtanen et al., 2020). The *optimal non-uniform discretization* (different for each bore field) stemming from the solution of the optimization problem will be used as a measure of the maximum achievable accuracy using $n_s = 8$ segments for the discretization scheme presented in section 2.3.

2.3.3. Optimal discretization for all bore fields

The identification of a *global optimal non-uniform discretization* using a single value for the end-length-ratio is relevant to revise the value adopted by Cook (2021) and implemented into `pygfunction` (Cimmino & Cook, 2022) supported by an analysis of the error on the g-function. A similar optimization problem can be formulated:

$$\begin{cases} \text{Minimize:} & \max(\varepsilon_{\max})(\delta_1) = \max\limits_{n,k} \frac{\left|g^{(n)}(t_k) - g_{ref}^{(n)}(t_k)\right|}{g_{ref}^{(n)}(t_k)} \cdot 100\% \\ \text{s.t.:} & \delta_{\min} \leq \delta_1 \leq \frac{1}{n_s} \end{cases} \tag{14}$$

where δ_1 is now the end-length-ratio that minimizes the maximum absolute percentage error for the entire data set of g-functions.

The optimization problem is solved using the same Nelder-Mead algorithm implemented in `SciPy` (Gao & Han, 2012; Virtanen et al., 2020).

2.4. Estimation of the optimal discretization using an artificial neural network

A fully connected feed forward multi-layer perceptron artificial neural network (ANN) is constructed using the Keras module in the TensorFlow (Abadi et al., 2015) Python package. The network is trained using a combination of nine geometrical and thermal input parameters and one optimal end-length-ratio output. The inputs utilized are the dimensionless borehole radius, r_b/L, the dimensionless buried depth, D/L, the steady-state borehole wall temperatures of three equivalent boreholes ($N_g = 3$), $\bar{h}^{\infty}_{\mathcal{I}}$, a ratio of the number of boreholes in each group over the total number of boreholes in the field, $n_{b,\mathcal{I}}/N_b$ and the total number of boreholes in the field, N_b. The three steady-state borehole wall temperatures and each associated ratio of the number of boreholes in a group to the total in the field are a measurement of the temperature distribution in a borehole field. They provide an indirect quantification of the bore field density, as denser bore fields are expected to have larger steady-state temperatures. The boreholes belonging to each of the three groups are obtained by cutting the dendrogram resulting from the hierarchical agglomerative clustering in the equivalent borehole method. The number of groups is fixed at $N_g = 3$ to train the ANN but the number of groups used in the calculation of the g-functions may be different. This does not require an additional clustering of the boreholes, as the dendrogram can simply be cut more than once to produce any number of groups.

The steady-state borehole wall temperatures of the equivalent boreholes are given by:

$$\bar{h}^{\infty}_{\mathcal{I}} = \frac{1}{n_{b,\mathcal{I}}} \sum_{i \in \mathbb{G}_{\mathcal{I}}} h^{\infty}_i \tag{15}$$

The Scikit-Learn (Pedregosa et al., 2011) Python package is utilized to split the borehole fields into 80% training and 20% test data sets for each borehole field category. All of the training data is combined into one set, where the test data is kept separate to later enable the quantification of error for each category. The input parameters are normalized into a standard score. The mean, μ, and standard deviation calculated, σ, for the standard score are stored. The feature input vector, X, to the ANN will always need to be normalized with the same mean and standard deviation as calculated with the training set.

The ANN, shown in figure 7, contains an input layer, three hidden layers and one output layer. The three hidden layers have 64, 128 and 64 neurons, respectively. Each of the three hidden layers utilize a rectified linear unit (ReLU) activation function. A linear activation function is used on the output layer. The Adam method (Kingma & Ba, 2017), a stochastic optimization algorithm, performs the hyper parameter tuning while working to minimize the mean square error. An NVIDIA Quadro M2000 graphics card performed the training. The amount of memory consumed during training is limited by constraining the number of samples used for each forward and backward propagation through the network. The entirety of the training data set is fed to the

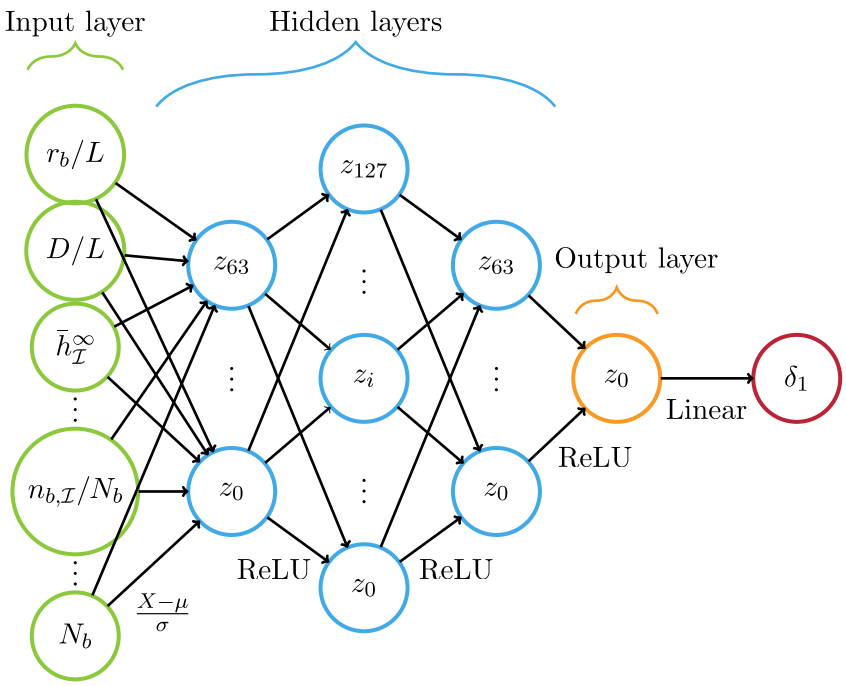

Figure 7. The multi-layer perception containing normalization on the input layer, 3 hidden layers with 64, 128 and 64 neurons using ReLU activation functions, and an output layer with a linear activation function.

Adam optimization method a total of 200 times. TensorFlow refers to this as the number of epochs.

3. Results

3.1. Borehole discretization

g-Functions are calculated for all generated bore field configurations using the four discretization approaches described in section 2.3. All g-functions are calculated using Python and the implementation of the method of equivalent boreholes available in the `pygfunction` package (Cimmino & Cook, 2022). g-Functions are evaluated at five timesteps, $t/t_s = \{0.01, 0.1, 1, 10, 100\}$ ($\ln(t/t_s) = \{-4.61, -2.30, 0, 2.30, 4.61\}$). This range effectively reaches steady-state, as shown for example on figure 1. The error on the g-function is reported using the maximum absolute percentage error:

$$\varepsilon_{\max} = \max_k \frac{\left| g^{(n)}(t_k) - g^{(n)}_{ref}(t_k) \right|}{g^{(n)}_{ref}(t_k)} \cdot 100\% \tag{16}$$

where $g^{(n)}_{ref}$ is the g-function evaluated using the reference non-uniform discretization of $n_s = 42$ segments with an end-length-ratio $\delta_1 = 0.02 \cdot \frac{8}{42} = 0.0038$.

The distribution of the maximum absolute percentage error is shown on figure 8 for all four discretizations. The uniform discretization of $n_s = 12$ segments (figure 8(a)) exhibits larger errors than the non-uniform discretizations and most importantly a significant maximum error of 99.0% in the worst bore field configuration. In this case,

5% of the g-functions have maximum errors above 13.3% and 1% of the g-functions have maximum errors above 26.7%. This runs contrary to the recommendation of Cimmino and Bernier (2014). However, the authors based their recommendation solely on the accuracy of the g-function of a rectangular field of 10 $\times$ 10 boreholes (N_b = 100) with parameters $r_b/H = 0.0005$, $B/H = 0.05$ and $D/H = 0.027$. The baseline non-uniform discretization of $n_s = 8$ (figure 8(b)) successfully decreases the error distribution despite the lowered number of segments. A significant maximum error of 25.7% still remains, with 5% of the g-functions having maximum errors above 2.1% and 1% of the g-functions having maximum errors above 4.0%. The error distributions of the optimal and global optimal non-uniform discretizations (figure 8(c-d)) show that a careful selection of discretization parameters can limit the maximum error. A constant end-length-ratio $\delta_1 = 0.005525$ limits the maximum error to 3.6%. This is at the cost of an increase in the distribution of the error when compared to the optimal non-uniform discretization: the median error increases from 0.6% to 0.9%.

Figure 9 compares the maximum absolute percentage error of the different discretizations against the baseline non-uniform discretization. It is shown that the maximum absolute percentage error of the uniform discretization is always greater than that of the baseline non-uniform discretization, and that the maximum absolute percentage error of the optimal non-uniform discretization is always lower than that of the baseline non-uniform discretization. The global optimal non-uniform discretization has a lower error than the baseline non-uniform discretization for 26.3% of the bore fields.

Figure 10 presents the g-functions for the bore fields associated with the maximum error for all four discretizations. The maximum absolute percentage error is maximum for one of the rectangular configurations for all four discretizations. Figure 10(a) shows the g-function of a rectangular field of 36 $\times$ 31 boreholes (N_b = 1116) with L = 418.8 m, D = 0.63 m, r_b = 0.075 m, B_x = 3.14 m and B_y = 3.18 m, associated with the maximum error for both of the uniform and baseline non-uniform discretizations. Figure 10(b) shows the g-function of a rectangular field of 40 $\times$ 36 boreholes (N_b = 1440) with L = 427.2 m, D = 0.51 m, r_b = 0.075 m, B_x = 17.42 m and B_y = 12.41 m, associated with the maximum error for the optimal non-uniform discretization. Figure 10(c) shows the g-function of a rectangular field of 31 $\times$ 40 boreholes (N_b = 1240) with L = 399.5 m, D = 0.54 m, r_b = 0.075 m, B_x = 3.74 m and B_y = 4.77 m, associated with the maximum error for the global optimal non-uniform discretization. For all three bore fields, the uniform discretization has the largest error, and the baseline non-uniform discretization has the second largest error. It is shown that the maximum error is obtained for boreholes with large length (L) and low buried depth (D). Even though g-functions take larger values when the spacing is small due to an increased intensity in thermal interactions between boreholes, the spacing does not have to be small to produce larger errors on the g-functions (e.g. as shown on figure 10(b)).

Figure 11 shows the dependence of the maximum absolute percentage error on the number of boreholes (N_b) and the dimensionless borehole spacing. For all field configurations except for the circle fields, the minimum of the spacings along the columns and rows of the bore fields is used to evaluate the dimensionless borehole spacing ($\min(B_x, B_y)$). For the baseline non-uniform discretization of $n_s = 8$ segments (figure 11(a)), its is shown that the error generally increases with increasing number of boreholes, especially for fields of less than 200 boreholes, and also increases with decreasing dimensionless borehole spacing. The dimensionless borehole spacing has a more prominent impact on the maximum absolute percentage error. Figure 11(a)

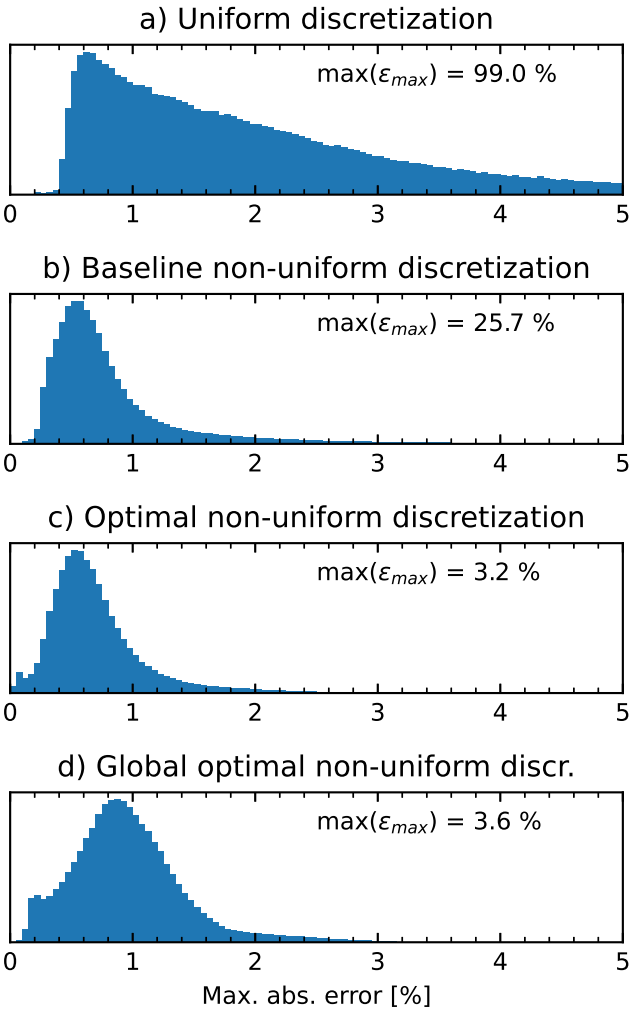

Figure 8. Distributions of the maximum absolute percentage error : a) Uniform discretization of 12 segments, b) Baseline non-uniform discretization of 8 segments with $\delta_1 = 0.02$, c) Optimal non-uniform discretization of 8 segments with optimal values of $\delta_1^{(n)}$, and d) Global optimal non-uniform discretization of 8 segments with $\delta_1 = 0.005525$

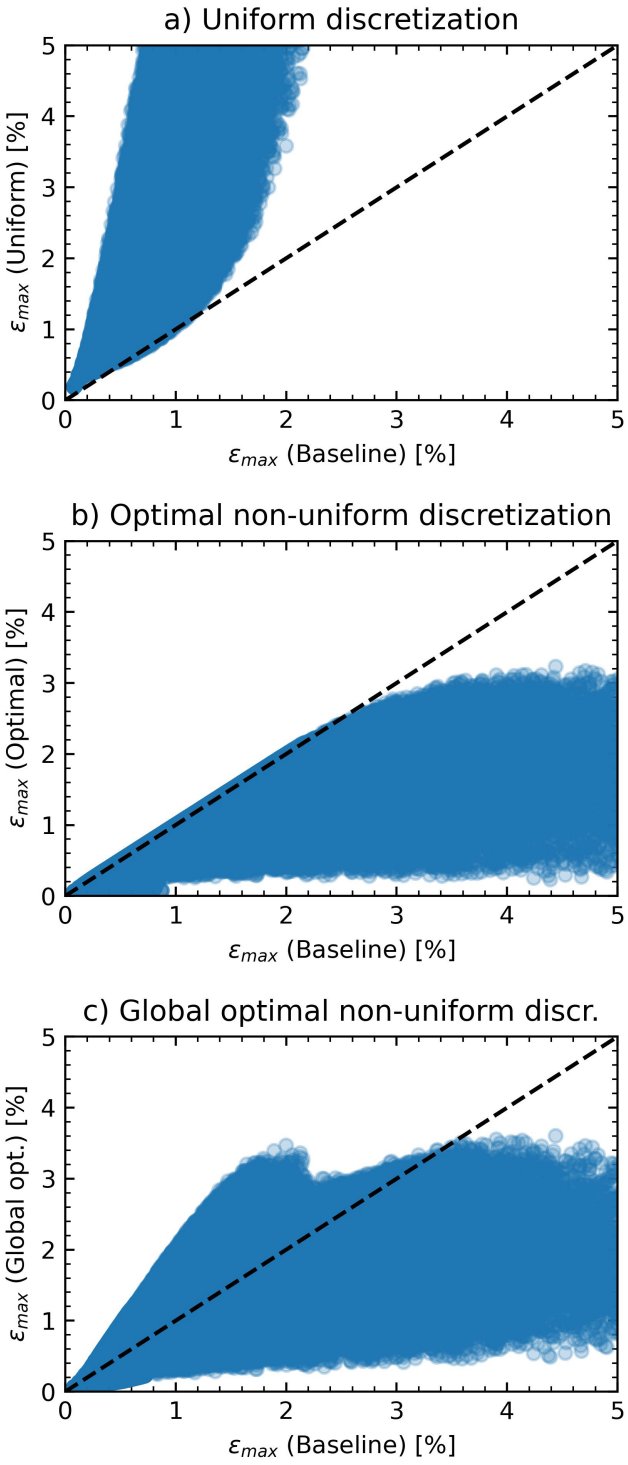

Figure 9. Comparisons of the maximum absolute error against the baseline non-uniform discretization of 8 segments with $\delta_1 = 0.02$: a) Uniform discretization of 12 segments, b) Optimal non-uniform discretization of 8 segments with optimal values of δ_1, and c) Global optimal non-uniform discretization of 8 segments with with $\delta_1 = 0.005525$

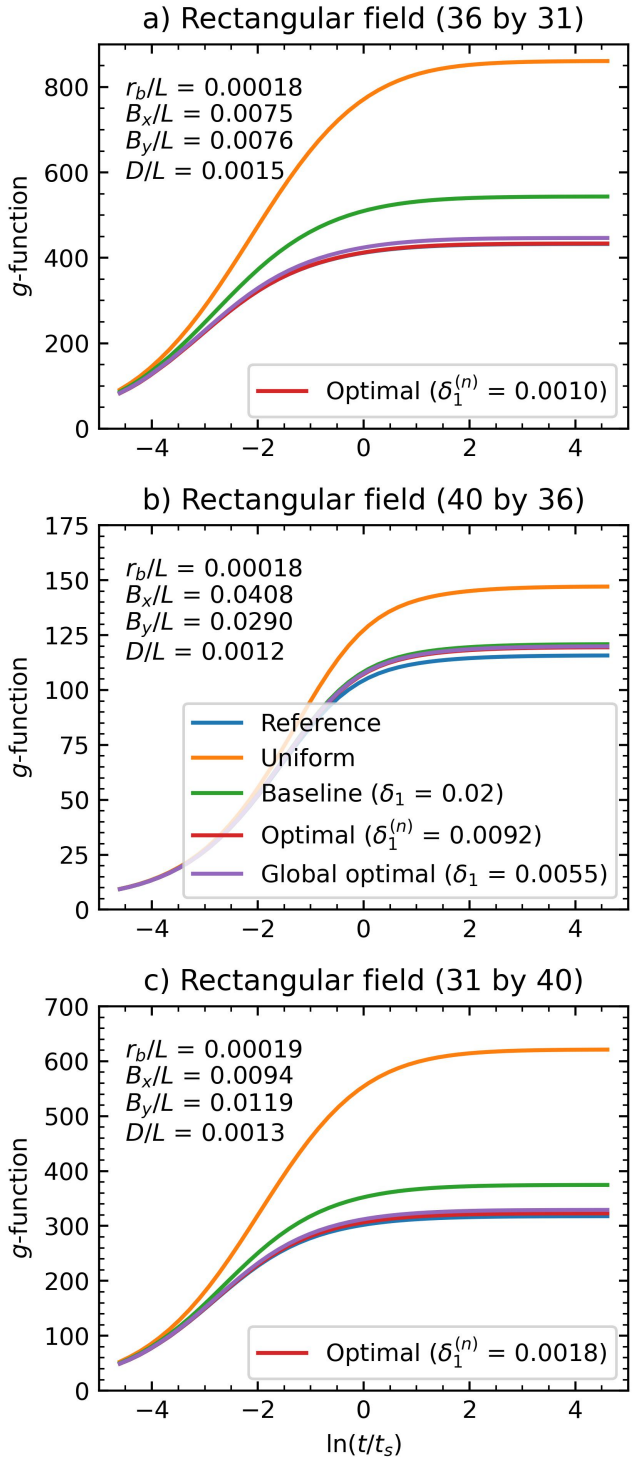

Figure 10. g-Functions of fields with the maximum absolute percentage error : a) Uniform discretization of 12 segments and baseline non-uniform discretization of 8 segments with $\delta_1 = 0.02$, b) Optimal non-uniform discretization of 8 segments with optimal values of $\delta_1^{(n)}$, and c) Global optimal non-uniform discretization of 8 segments with $\delta_1 = 0.005525$

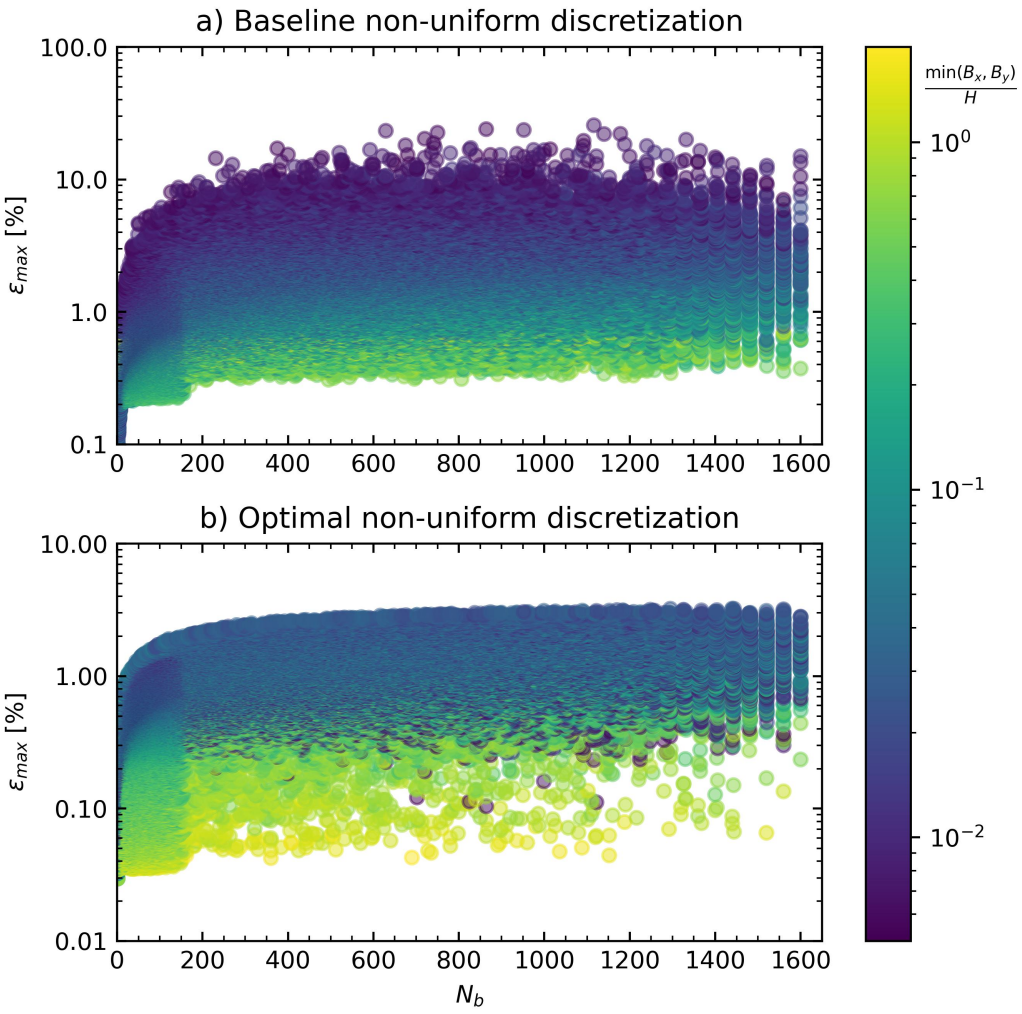

Figure 11. Impact of the number of boreholes and the dimensionless borehole spacing on the maximum absolute percentage error: a) Baseline non-uniform discretization of 8 segments with $\delta_1 = 0.02$, b) Optimal non-uniform discretization of 8 segments with optimal values of $\delta_1^{(n)}$

shows a discontinuity in the error at $N_b = 156$ which simply corresponds to the upper limit in the number of boreholes considered for box-shaped fields and thus changes the composition of the data set below this threshold. Other discontinuities caused by the upper limits on the number of boreholes for U-shaped ($N_b = 118$), L-shaped ($N_b = 79$) and circle ($N_b = 40$) fields are not as apparent. As shown on figure 11(b), the optimal non-uniform discretization lowers the maximum absolute percentage error overall and retains the general trend of the error increasing with decreasing dimensionless borehole spacing. However, the optimization seems able to further reduce the error for very low dimensionless borehole spacing.

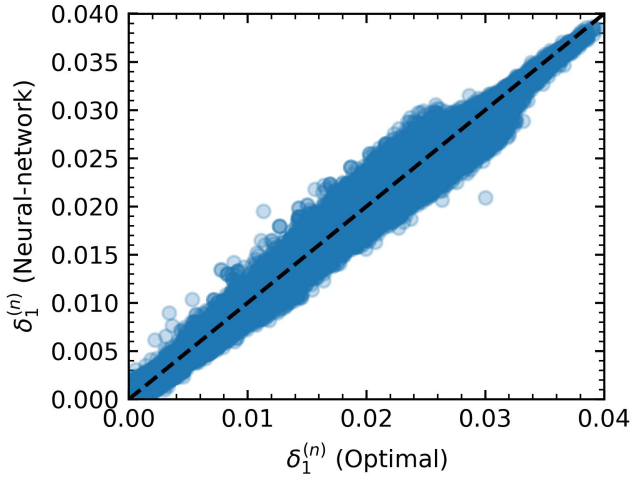

Figure 12. Comparison of the values of $\delta_1^{(n)}$ estimated by the ANN against the optimal values of $\delta_1^{(n)}$

3.2. Estimation of the optimal discretization using an artificial neural network

Figure 12 presents a comparison of the end-length-ratio predicted by the trained artificial neural network model against the optimal end-length-ratio ($\delta_1^{(n)}$) that minimizes the maximum absolute percentage error for each bore field. It is shown that the neural network is able to capture the trend in variation of the end-length-ratio. The root mean square errors on the end-length-ratios on the testing sets are 0.0006, 0.0006, 0.0008, 0.0005 and 0.0006 for the rectangular, box-shaped, U-shaped, L-shaped and circle fields, respectively. This further shows that the combination of geometrical parameters r_b/L, D/L, N_b, and the thermal parameters $\bar{h}_{\mathcal{I}}^{\infty}$, $n_{b,\mathcal{I}}/N_b$, are sufficient to capture the heat transfer characteristics of the bore fields. It should be noted that the geometrical parameters alone are not able to capture the trend in end-length-ratio, and that using $N_g = 2$ borehole groups is insufficient.

Despite the observed differences between the predicted and optimal end-length ratios, the end-length-ratios predicted by the ANN provide good accuracy on the evaluation of g-functions. Figure 13 compares the maximum absolute percentage error of the discretization predicted by the ANN against the optimal non-uniform discretization. The error follows the same trend. The maximum value of the maximum absolute percentage error from the ANN is 3.3%, which is close to the maximum value of 3.2% from the optimal non-uniform discretization. As expected, the maximum absolute percentage error from the ANN is always greater than that of the optimal non-uniform discretization. The maximum absolute percentage error of the ANN discretization is lower than that of the global optimal non-uniform discretization for 99.1% of the cases.

4. Conclusions

The effect of the borehole discretization on the accuracy of g-functions is presented. It is shown that a uniform discretization of $n_s = 12$ segments, as proposed in previous work (Cimmino & Bernier, 2014), can lead to significant error (up to 99.0% for the

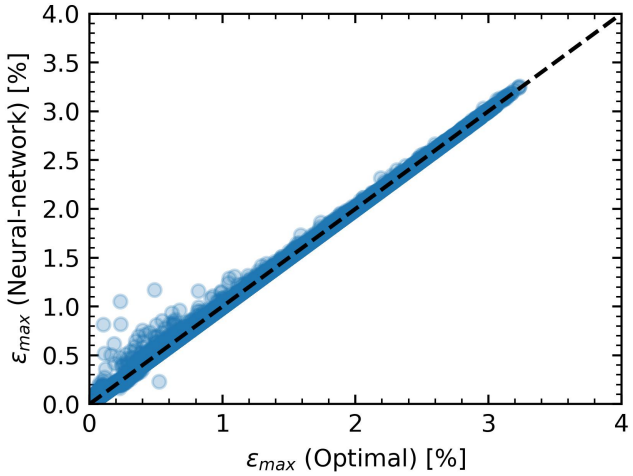

Figure 13. Comparison of the maximum absolute percentage error of the values of $\delta_1^{(n)}$ estimated by the neural-network model against the optimal values of $\delta_1^{(n)}$

selected bore fields). A non-uniform discretization of $n_s = 8$ segments decreases the error on the g-functions, despite using fewer segments. A careful selection of the discretization parameter (i.e. the end-length-ratio) is required to limit the errors on the g-function, which can also become significantly large (up to 25.7% for the selected bore fields using a baseline discretization). The maximum value of the maximum absolute percentage error is 3.2% for the optimal non-uniform discretization.

Optimal and global optimal non-uniform discretizations are obtained by minimizing the maximum absolute percentage error on each bore field of a data set in the former case, or the maximum value of the maximum absolute percentage error over the entire data set in the latter case. The recommended value of the end-length-ratio for the global optimal non-uniform discretization of 8 segments is $\delta_1 = 0.005525$. In both cases, the evaluation of the end-length-ratios necessitates the solution of an optimization problem and thus requires repeated evaluation of g-functions for all of the bore fields. This task is only now possible due to recent advancements in the calculation of g-functions (Prieto & Cimmino, 2021, for instance) and the development of open-source computation packages such as `pygfunction` (Cimmino & Cook, 2022). The total number of considered bore fields is 557,056, bringing the total number of g-functions considered in the analysis to 3,342,336 (i.e. for all considered discretizations) and neglecting repeated evaluation of g-functions for the optimization.

An artificial neural network is trained to predict the optimal value of the end-length-ratio based on geometrical parameters of the boreholes (excluding their position) and on new thermal parameters, namely steady-state borehole wall temperatures, obtained from the equivalent borehole method using $N_g = 3$ borehole groups. The ANN gives adequate prediction of the end-length-ratios, leading to a maximum value of the maximum absolute percentage error on the g-functions of 3.3%. The adequate prediction of end-length-ratios from the addition of the thermal parameters indicates that these added parameters may be representative of the heat transfer characteristics of the bore fields. The accuracy of the prediction of end-length-ratios may be improved by increasing the number of borehole groups N_g. Alternatively, in future work, the g-function could be directly evaluated using these thermal parameters without relying on geo-

metrical parameters of the bore field (i.e. the borehole positions) to train the ANN. The global optimal non-uniform discretization of 8 segments will be implemented into `pygfunction`, as well as the ANN to evaluate the optimal discretization for the equivalent borehole method. The methodology presented in this paper can be repeated to obtain recommendations for the global optimal non-uniform discretization for any number of segments, and also to train an ANN to calculate the optimal discretization.

The analysis is limited to fields of boreholes of equal length connected in parallel and modelled using a boundary condition of uniform borehole wall temperature. The extension to a boundary condition of equal inlet fluid temperature could be achieved by including the effective borehole thermal resistance in the parameters of the Sobol sequence to generate the bore fields. The global optimal end-length-ratio would then be a function of the dimensionless borehole thermal resistance. This would greatly increase the number of bore fields for which the g-functions need to be calculated. However, as evidenced in this work and the precursory conference publication (Cimmino & Cook, 2022), g-function calculation methods have sufficiently progressed to enable large data sets of g-functions.

The error on g-functions is related to errors on the evaluation of fluid temperatures and heat pump COP during simulation, and to errors on the size of the bore fields during design. However, g-functions correspond to the ground temperature variation caused by a constant heat extraction (or injection) from the bore field. Meanwhile, fluid temperatures, heat pump COP and bore field sizing also depend on the ground load profile. The errors on g-functions are thus not equal to errors on these quantities. Further study into the impact of g-function errors is needed in the future.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems.* Retrieved from `https://www.tensorflow.org/` (Software available from tensorflow.org)

Abdelaziz, S. L., Ozudogru, T. Y., Olgun, C. G., & Martin, J. R. (2014). Multilayer finite line source model for vertical heat exchangers. *Geothermics*, *51*, 406-416.

Cimmino, M. (2015). The effects of borehole thermal resistances and fluid flow rate on the g-functions of geothermal bore fields. *International Journal of Heat and Mass Transfer*, *91*, 1119-1127.

Cimmino, M. (2018a). Fast calculation of the g-functions of geothermal borehole fields using similarities in the evaluation of the finite line source solution. *Journal of Building Performance Simulation*, *11*(6), 655-668.

Cimmino, M. (2018b). pygfunction: an open-source toolbox for the evaluation of thermal response factors for geothermal borehole fields [Conference Proceedings]. In *Proceedings of eSim 2018, the 10th conference of IBPSA-Canada* (pp. 492–501).

Cimmino, M. (2019). Semi-analytical method for g-function calculation of bore fields with series- and parallel-connected boreholes. *Science and Technology for the Built Environment*, *25*(8), 1007-1022.

Cimmino, M. (2021). An approximation of the finite line source solution to model thermal interactions between geothermal boreholes. *International Communications in Heat and Mass Transfer*, *127*, 105496.

Cimmino, M., & Bernier, M. (2014). A semi-analytical method to generate g-functions for geothermal bore fields. *International Journal of Heat and Mass Transfer*, *70*, 641-650.

Cimmino, M., Bernier, M., & Adams, F. (2013). A contribution towards the determination of g-functions using the finite line source. *Applied Thermal Engineering*, *51*(1), 401-412.

Cimmino, M., & Cook, J. C. (2022). pygfunction 2.2 : New features and improvements in accuracy and computational efficiency [Conference Proceedings]. In *Proceedings of the IGSHPA Research Track* (p. 45-52).

Claesson, J., & Javed, S. (2011). An analytical method to calculate borehole fluid temperatures for time-scales from minutes to decades. *ASHRAE Transactions*, *118*(2), ML-11-C034.

Cook, J. C. (2021). *Development of computer programs for fast computation of g-functions and automated ground heat exchanger design* (M.Sc. Thesis, Oklahoma State University, Stillwater, Oklahoma). Retrieved from `https://shareok.org/handle/11244/335489`

Cui, P., Yang, H., & Fang, Z. (2006). Heat transfer analysis of ground heat exchangers with inclined boreholes. *Applied Thermal Engineering*, *26*(11), 1169-1175.

Dusseault, B., & Pasquier, P. (2019). Efficient g-function approximation with artificial neural networks for a varying number of boreholes on a regular or irregular layout. *Science and Technology for the Built Environment*, *25*(8), 1023-1035.

Dusseault, B., Pasquier, P., & Marcotte, D. (2018). A block matrix formulation for efficient g-function construction. *Renewable Energy*, *121*, 249-260.

Erol, S., & Franois, B. (2018). Multilayer analytical model for vertical ground heat exchanger with groundwater flow. *Geothermics*, *71*, 294-305.

Eskilson, P. (1987). *Thermal analysis of heat extraction boreholes* (Ph.D. Thesis). University of Lund.

Fossa, M. (2011). The temperature penalty approach to the design of borehole heat exchangers for heat pump applications. *Energy and Buildings*, *43*(6), 1473-1479.

Gao, F., & Han, L. (2012). Implementing the nelder-mead simplex algorithm with adaptive parameters. *omputational Optimization and Applications*, *51*(1), 259–277.

Hellström, G. (1991). *Ground heat storage: thermal analyses of duct storage system* (Ph.D. Thesis). University of Lund.

Hu, J. (2017). An improved analytical model for vertical borehole ground heat exchanger with multiple-layer substrates and groundwater flow. *Applied Energy*, *202*, 537-549.

Kingma, D. P., & Ba, J. (2017). *Adam: A method for stochastic optimization.*

Lamarche, L. (2011). Analytical g-function for inclined boreholes in ground-source heat pump systems. *Geothermics*, *40*(4), 241-249.

Lamarche, L. (2017). g-function generation using a piecewise-linear profile applied to ground heat exchangers. *International Journal of Heat and Mass Transfer*, *115*, 354-360.

Lamarche, L., & Beauchamp, B. (2007). A new contribution to the finite line-source model for geothermal boreholes. *Energy and Buildings*, *39*(2), 188-198.

Lazzarotto, A. (2016). A methodology for the calculation of response functions for geothermal fields with arbitrarily oriented boreholes - part 1. *Renewable Energy*, *86*, 1380-1393.

Lazzarotto, A., & Björk, F. (2016). A methodology for the calculation of response functions for geothermal fields with arbitrarily oriented boreholes - part 2. *Renewable Energy*, *86*, 1353-1361.

Li, M., & Lai, A. C. (2012). Heat-source solutions to heat conduction in anisotropic media with application to pile and borehole ground heat exchangers. *Applied Energy*, *96*, 451-458.

Marcotte, D., & Pasquier, P. (2009). The effect of borehole inclination on fluid and ground temperature for glhe systems. *Geothermics*, *38*(4), 392-398.

Marcotte, D., & Pasquier, P. (2014). Unit-response function for ground heat exchanger with parallel, series or mixed borehole arrangement. *Renewable Energy*, *68*, 14-24.

Molina-Giraldo, N., Blum, P., Zhu, K., Bayer, P., & Fang, Z. (2011). A moving finite line source model to simulate borehole heat exchangers with groundwater advection. *International Journal of Thermal Sciences*, *50*(12), 2506-2513.

Ng, E. W., & Geller, M. (1969). A table of integrals of the error functions. *Journal of Research of the National Bureau of Standards – B*, *73*(1), 1–20.

Nguyen, A., & Pasquier, P. (2021). A successive flux estimation method for rapid g-function construction of small to large-scale ground heat exchanger. *Renewable Energy*, *165*, 359-368.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay,

E. (2011). Scikit-learn: Machine Learning in Python . *Journal of Machine Learning Research*, *12*, 2825–2830.

Prieto, C., & Cimmino, M. (2021). Thermal interactions in large irregular fields of geothermal boreholes: the method of equivalent boreholes. *Journal of Building Performance Simulation*, *14*(4), 446–460.

Sobol, I. (1967). The distribution of points in a cube and the accurate evaluation of integrals. *Vychisl. Mat. i Mater. Phys.*, *7*, 784–802.

Stephen, J., & Kuo, F. Y. (2008). Constructing sobol sequences with better two-dimensional projections. *SIAM Journal on Scientific Computing*, *30*(5), 2635–2654.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... Van Mulbregt, P. (2020). Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, *17*(3), 261-272.

West, T. N., & Spitler, J. D. (2022, 17–21 October). Approximate g-functions for selection of borehole field configurations used with ground-source heat pump systems [Conference Proceedings]. In *Proceedings of the european geothermal congress 2022.* Berlin, Germany.

Zeng, H. Y., Diao, N. R., & Fang, Z. H. (2002). A finite line-source model for boreholes in geothermal heat exchangers. *Heat Transfer - Asian Research*, *31*(7), 558-567.

Appendix A. Steady-state finite line source solution

The finite line source solution gives the average temperature change over a vertical line segment i of length L_i buried at a distance D_i from an isothermal surface due to heat extracted from another vertical line segment j of length L_j buried at a distance D_j from the isothermal surface and at located at a radial distance r_{ij} from line segment i. The geometry of the problem is shown on Figure A1. The finite line source solution is given by (Cimmino & Bernier, 2014; Claesson & Javed, 2011):

$$h_{ij}(t) = \frac{1}{2L_i} \int_{\frac{1}{\sqrt{4\alpha_s t}}}^{\infty} \frac{1}{s^2} \exp(-r_{ij}^2 s^2) F_{ij}(s) ds \tag{A1a}$$

$$\begin{aligned} F_{ij}(s) = &\operatorname{erfint}((D_i - D_j + L_i)s) - \operatorname{erfint}((D_i - D_j)s) \\ &+ \operatorname{erfint}((D_i - D_j - L_j)s) - \operatorname{erfint}((D_i - D_j + L_i - L_j)s) \\ &+ \operatorname{erfint}((D_i + D_j + L_i)s) - \operatorname{erfint}((D_i + D_j)s) \\ &+ \operatorname{erfint}((D_i + D_j + L_j)s) - \operatorname{erfint}((D_i + D_j + L_i + L_j)s) \end{aligned} \tag{A1b}$$

$$\operatorname{erfint}(x) = \int_0^x \operatorname{erf}(x')dx' = x \operatorname{erf}(x) - \frac{1}{\sqrt{\pi}}(1 - \exp(-x^2) \tag{A1c}$$

where α_s is the ground thermal diffusivity.

The steady-state finite line source solution was presented by Prieto and Cimmino (2021) and is used in the present work. However, the development of the solution was not presented by Prieto and Cimmino (2021). It is convenient to adopt a compact form of the finite line source solution to simplify the development:

$$h_{ij}(t) = \frac{1}{2L_i} \sum_{m=1}^{8} (-1)^{m+1} \int_{\frac{1}{\sqrt{4\alpha_s t}}}^{\infty} \frac{1}{s^2} \exp(-r_{ij}^2 s^2) \operatorname{erfint}(w_m s) ds \tag{A2a}$$

$$\begin{aligned} w_m = \{&D_i - D_j + L_i, D_i - D_j, D_i - D_j - L_j, D_i - D_j + L_i - L_j, \\ &D_i + D_j + L_i, D_i + D_j, D_i + D_j + L_j, D_i + D_j + L_i + L_j\} \end{aligned} \tag{A2b}$$

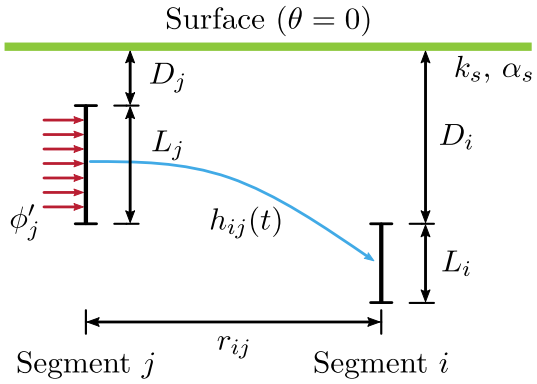

Figure A1. Finite line source solution

The integral of the error function (erfint $(\cdot)$) is substituted by its definition from equation A1b for all terms of the summation over m:

$$h_{ij}(t) = \sum_{m=1}^{8} (-1)^{m+1} h_{ij}^{(m)}(t) \tag{A3a}$$

$$\begin{aligned} h_{ij}^{(m)}(t) = \frac{1}{2L_i} \int_{\frac{1}{\sqrt{4\alpha_s t}}}^{\infty} \Bigg\{ & \frac{w_m}{s} \exp\left(-r_{ij}^2 s^2\right) \operatorname{erf}\left(w_m s\right) \\ & + \frac{-1}{s^2\sqrt{\pi}} \left[\exp\left(-r_{ij}^2 s^2\right) - \exp\left(-\left(r_{ij}^2 + w_m^2\right) s^2\right)\right] \Bigg\} ds \end{aligned} \tag{A3b}$$

The steady-state solution is obtained by solving the integral at the limit $t \to \infty$:

$$h_{ij}^{\infty} = \lim_{t\to\infty} \sum_{m=1}^{8} (-1)^{m+1} h_{ij}^{(m)}(t) = \sum_{m=1}^{8} (-1)^{m+1} h_{ij}^{\infty,(m)} \tag{A4a}$$

$$\begin{aligned} h_{ij}^{\infty,(m)} = \frac{1}{2L_i} \lim_{a\to 0^+} \int_{a}^{\infty} \Bigg\{ & \frac{w_m}{s} \exp\left(-r_{ij}^2 s^2\right) \operatorname{erf}\left(w_m s\right) \\ & + \frac{-1}{s^2\sqrt{\pi}} \left[\exp\left(-r_{ij}^2 s^2\right) - \exp\left(-\left(r_{ij}^2 + w_m^2\right) s^2\right)\right] \Bigg\} ds \end{aligned} \tag{A4b}$$

The integral can be separated into two parts. The first part involves the multiplication of the error function with an exponential divided by the integration variable s. The second part involves a difference of two exponentials divided by the square of the integration variable s^2. The first part has a known solution (Ng & Geller, 1969):

$$\int_0^{\infty} \operatorname{erf}\left(ax\right) \exp\left(-b^2 x^2\right) \frac{dx}{x} = \ln \frac{a + \sqrt{a^2 + b^2}}{b} \tag{A5}$$

with $\Re(b^2) > 0$.

This gives:

$$\begin{aligned}&\frac{1}{2L_i}\lim_{a\to 0^+}\int_a^\infty \frac{w_m}{s}\exp\left(-r_{ij}^2 s^2\right)\operatorname{erf}\left(w_m s\right)ds = \\ &\frac{1}{2L_i}w_m \ln\left[\left(w_m+\sqrt{w_m^2+r_{ij}^2}\right)/r_{ij}\right]\end{aligned} \tag{A6}$$

The indefinite integral of the second part is given by:

$$\int \exp\left(-a^2x^2\right)\frac{dx}{x^2} = -\frac{\exp\left(-a^2x^2\right)}{x} - \sqrt{\pi}a\operatorname{erf}\left(ax\right) \tag{A7}$$

This gives:

$$\begin{aligned}&\frac{1}{2L_i}\lim_{a\to 0^+}\int_a^\infty \frac{-1}{s^2\sqrt{\pi}}\left[\exp\left(-r_{ij}^2s^2\right)-\exp\left(-\left(r_{ij}^2+w_m^2\right)s^2\right)\right]ds = \\ &\frac{1}{2L_i}\lim_{\substack{a\to 0^+\\ b\to\infty}}\left(\frac{\exp\left(-r_{ij}^2s^2\right)-\exp\left(-\left(r_{ij}^2+w_m^2\right)s^2\right)}{s\sqrt{\pi}}+r_{ij}\operatorname{erf}\left(r_{ij}s\right)-\sqrt{r_{ij}^2+w_m^2}\operatorname{erf}\left(\sqrt{r_{ij}^2+w_m^2}s\right)\right)_a^b\end{aligned} \tag{A8}$$

The solution of the second part of the integral involves the following limits:

$$\lim_{s\to\infty}\frac{\exp\left(-r_{ij}^2s^2\right)-\exp\left(-\left(r_{ij}^2+w_m^2\right)s^2\right)}{s\sqrt{\pi}} = 0 \tag{A9a}$$

$$\lim_{s\to 0^+} r_{ij}\operatorname{erf}\left(r_{ij}s\right)-\sqrt{r_{ij}^2+w_m^2}\operatorname{erf}\left(\sqrt{r_{ij}^2+w_m^2}s\right) = 0 \tag{A9b}$$

$$\lim_{s\to\infty} r_{ij}\operatorname{erf}\left(r_{ij}s\right)-\sqrt{r_{ij}^2+w_m^2}\operatorname{erf}\left(\sqrt{r_{ij}^2+w_m^2}s\right) = r_{ij}-\sqrt{r_{ij}^2+w_m^2} \tag{A9c}$$

The remaining limit is indeterminate and requires the use of L'Hôpital's rule:

$$\begin{aligned}&\lim_{s\to 0^+}\frac{\exp\left(-r_{ij}^2s^2\right)-\exp\left(-\left(r_{ij}^2+w_m^2\right)s^2\right)}{s\sqrt{\pi}} \\ &= \lim_{s\to 0^+}\frac{2r_{ij}^2s\exp\left(-r_{ij}^2s^2\right)-2\left(r_{ij}^2+w_m^2\right)s\exp\left(-\left(r_{ij}^2+w_m^2\right)s^2\right)}{\sqrt{\pi}} \\ &= 0\end{aligned} \tag{A10}$$

Inserting the limits (equations A9 and A10) into equation A8 and substituting equations A6 and A8 into equation A4:

$$h_{ij}^\infty = \frac{1}{2L_i}\sum_{m=1}^{8}(-1)^{m+1}\left\{w_m\ln\left[\left(w_m+\sqrt{w_m^2+r_{ij}^2}\right)/r_{ij}\right]+r_{ij}-\sqrt{w_m^2+r_{ij}^2}\right\} \tag{A11}$$

with $\sum_{m=1}^{8}(-1)^{m+1}r_{ij}=0$. This results corresponds to equation 8.