Computers & Geosciences 103 (2017) 36–50

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Statistical modeling of geopressured geothermal reservoirs

Esmail Ansari[a,*], Richard Hughes[a], Christopher D. White[b]

[a] Louisiana State University, Baton Rouge, LA 70803, United States
[b] Tulane University, New Orleans, LA 70118, United States

ARTICLE INFO

Keywords:
Predictive model
Statistical model
Experimental design
Geothermal reservoir
Inspectional analysis
Dimensional analysis

ABSTRACT

Identifying attractive candidate reservoirs for producing geothermal energy requires predictive models. In this work, inspectional analysis and statistical modeling are used to create simple predictive models for a line drive design. Inspectional analysis on the partial differential equations governing this design yields a minimum number of fifteen dimensionless groups required to describe the physics of the system. These dimensionless groups are explained and confirmed using models with similar dimensionless groups but different dimensional parameters. This study models dimensionless production temperature and thermal recovery factor as the responses of a numerical model. These responses are obtained by a Box-Behnken experimental design. An uncertainty plot is used to segment the dimensionless time and develop a model for each segment. The important dimensionless numbers for each segment of the dimensionless time are identified using the Boosting method. These selected numbers are used in the regression models. The developed models are reduced to have a minimum number of predictors and interactions. The reduced final models are then presented and assessed using testing runs. Finally, applications of these models are offered. The presented workflow is generic and can be used to translate the output of a numerical simulator into simple predictive models in other research areas involving numerical simulation.

1. Introduction

Developing geopressured-geothermal reservoirs reduces global warming and secures energy needs. Louisiana's geopressured geothermal reservoirs have been examined previously (Bassiouni, 1980; McMullan and Bassiouni, 1984). However, these resources have not been developed, and uncertainty associated with their development persists. Identifying attractive candidates in a database of reservoirs for producing geothermal energy requires quick and simple models because simulating each case before development is time consuming and expensive. One approach for creating these screening models is to translate the output of a simulator into simple models with general applicability at all scales by combining inspectional analysis with statistical modeling.

In Inspectional Analysis (IA), the differential equations of a system are coupled with initial and boundary condition equations to describe the physics of the system. The system of equations is then transformed into dimensionless form to obtain a set of dimensionless groups (Shook et al., 1992; Jin et al., 2010). The necessary groups are then selected and used in statistical modeling.

In statistical or predictive modeling, a model is fit to the runs obtained by experimental design (Wood et al., 2008; Anbar and Akin, 2011; Mishra et al., 2015). Numerical simulation is used as virtual experimentation to create a database of observations. These observations are used to create the proxy models for prediction. This work uses regression analysis for creating the models for a geopressured-geothermal reservoir developed under a regular line drive design (Ansari and Hughes, 2016).

Wood et al. (2008) develop predictive models for CO_2 flooding and storage in Gulf Coast reservoirs using the dimensionless groups found by Shook et al. (1992). They use a combination of Box-Behnken and factorial experimental design and use linear regression for modeling. Anbar and Akin (2011) present a proxy model which can predict CO_2 storage capacity of deep saline carbonate aquifers. They use Latin Hypercube experimental design and linear regression models. Jin (2013) uses inspectional analysis, Box-Behnken design and linear regression to create models for a downhole water loop to improve well performance. He uses analysis of variance (ANOVA) for selecting the important dimensionless numbers. Schuetter et al. (2014) compare the use of Box-Behnken sampling and quadratic polynomial regression with Latin Hypercube sampling and Kriging, Multivariate Adaptive Regression Spline (MARS) and Additivity Variance Stabilization (AVAS) techniques for geological CO_2 sequestration. They conclude that the model developed using Box-Behnken and quadratic polyno-

* Corresponding author.
E-mail addresses: eansar2@lsu.edu (E. Ansari), rghughes@lsu.edu (R. Hughes), cwhite18@tulane.edu (C.D. White).

http://dx.doi.org/10.1016/j.cageo.2017.02.015
Received 3 March 2016; Received in revised form 6 August 2016; Accepted 16 February 2017
Available online 21 February 2017
0098-3004/

Nomenclature

Subscripts

1 *or* 2	Scale factor number
avg	Average
D	Dimensionless
f	Fluid
i	Initial
inj	Injection
ins	insulation
prod	production
r	Rock
ref	Reference
t	Total
ϕ	Pore

Greek

ϕ	Porosity (–)
α, α_X, α_Y	Dip angle (°)
β	Thermal expansivity (K^{-1})
ρ	Density (kg. m^{-3})
$\boldsymbol{\tau}$	Geothermal gradient (K. m^{-1}))
π	Scale group (–)
λ	Thermal conductivity tensor (W. m^{-1}. K^{-1})
λ	Thermal conductivity (W. m^{-1}. K^{-1})
κ	Thermal diffusivity (m^2. s^{-1})
μ	Viscosity (Pa.s)

Other symbols

$\forall$	For all
$\nabla.$	Divergence
∇	Gradient

Superscripts

*	Denotes a scale factor

Roman

c	Compressibility (Pa^{-1})
M	Matrix/fluid heat capacity ratio (–)
PV	Pore volume (ϕLWH, m^3)
p	Pressure (Pa)
q	Flow rate (m^3. s^{-1})
T	Temperature (K)
t	Time (s)
U	Internal energy (J)
u_T	Injection/Production velocity (m. s^{-1})
$\mathbf{u}$	Interstitial fluid velocity (m. s^{-1})
W	Reservoir width (m)
C_P	Isobaric specific heat capacity (J. kg^{-1}. K^{-1})
C_v	Volumetric specific heat capacity (J. kg^{-1}. K^{-1})
$\mathbf{g}$	Gravity vector (m. s^{-2})
H	Reservoir thickness (m)
$\mathcal{H}$	Enthalpy (J)
$\mathbf{K}$	Permeability tensor (m^2)
k_x, k_y, k_z	Directional permeabilities (m^2)
L	Reservoir length (m)

mials performs the best. Ganesh and Mishra (2014) give a 2D simplified dimensionless proxy model to characterize CO_2 injectivity in semi-confined layered saline aquifers. Kang et al. (2016) develop a statistical kriging model for a CO_2-capture-enabled power station and use the model as an efficient proxy to optimize the power station process. Ansari (2016) provides a statistical model for high angle wells in three dimensional geothermal systems.

The line drive design has been used in studies such as waterflooding and CO_2 flooding (Shook et al., 1992; Wood et al., 2008) because the simple injector producer pair can be a proxy for many different alternative patterns. A regular design means injecting cooler water at the up-dip side of the reservoir and producing hot geofluid from the down-dip portion of the reservoir (Fig. 1) where the temperature is higher due to the geothermal gradient (Plaksina et al., 2011; Ansari and Hughes, 2016).

This study develops a set of statistical models that can be used to quickly estimate production temperature and produced energy from geopressured-geothermal reservoirs. These models can screen databases of reservoirs to select the most attractive candidates for evaluating geothermal energy production.

This paper is divided into three sections: mathematical formulation, numerical simulation and statistical modeling. Following these three sections, applications of the models and conclusions are presented.

2. Mathematical formulation

The mathematical formulation section proceeds as follows: the governing equations describing a line drive system are derived from the fundamental partial differential equations for a geothermal system. Inspectional analysis is used to obtain the dimensionless groups (provided in Appendix A). These dimensionless groups are explained and validated by considering models with different dimensional parameters but similar dimensionless numbers. Validation is performed numerically to confirm the dimensionless groups.

The general continuity equation for single phase flow in porous media states that the divergence of the mass transport determines the change in the water mass in the medium (Eq. (1)).

$$\frac{\partial}{\partial t}(\phi\rho_f) + \nabla.\ (\rho_f \mathbf{u}) = 0 \tag{1}$$

The continuity equation can be expanded to yield Eq. (2):

$$\rho_f \frac{\partial \phi}{\partial t} + \phi \frac{\partial \rho_f}{\partial t} + \nabla\rho_f.\ \mathbf{u} + \rho_f \nabla.\ \mathbf{u} = 0 \tag{2}$$

Rock and fluid characteristics (Eqs. (3)–(6)) can be substituted into the continuity equation by applying the chain rule in time (Eqs. (7) and (8)) and space (Eqs. (9) and (10)) to get a physically useful form (Eq. (11)).

$$c_f = \frac{1}{\rho_f}\left(\frac{\partial \rho_f}{\partial p}\right) \tag{3}$$

$$c_\phi = \frac{1}{\phi}\left(\frac{\partial \phi}{\partial p}\right) \tag{4}$$

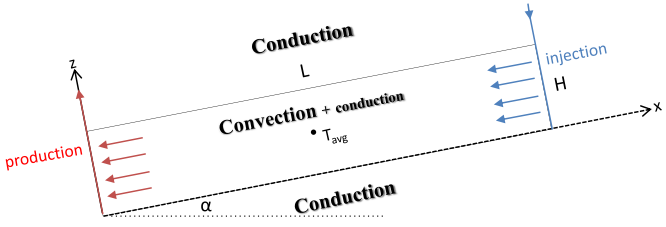


Fig. 1. Regular heat extraction design. Similar design is used in studies such as waterflooding and CO_2 flooding (Shook et al., 1992; Wood et al., 2008).

$$\beta_f = -\frac{1}{\rho_f}\left(\frac{\partial \rho_f}{\partial T}\right) \tag{5}$$

$$\beta_\phi = -\frac{1}{\phi}\left(\frac{\partial \phi}{\partial T}\right) \tag{6}$$

$$\frac{\partial \rho_f}{\partial t} = \frac{\partial \rho_f}{\partial p}.\frac{\partial p}{\partial t} + \frac{\partial \rho_f}{\partial T}.\frac{\partial T}{\partial t} = \rho_f c_f \frac{\partial p}{\partial t} - \rho_f \beta_f \frac{\partial T}{\partial t} \tag{7}$$

$$\frac{\partial \phi}{\partial t} = \frac{\partial \phi}{\partial p}.\frac{\partial p}{\partial t} + \frac{\partial \phi}{\partial T}.\frac{\partial T}{\partial t} = \phi c_\phi \frac{\partial p}{\partial t} - \phi \beta_\phi \frac{\partial T}{\partial t} \tag{8}$$

$$\nabla \rho_f = \rho_f c_f \nabla p - \rho_f \beta_f \nabla T \tag{9}$$

$$\nabla \phi = \phi c_\phi \nabla p - \phi \beta_\phi \nabla T \tag{10}$$

$$\phi \rho_f \left(c_\phi \frac{\partial p}{\partial t} - \beta_\phi \frac{\partial T}{\partial t}\right) + \phi \rho_f \left(c_f \frac{\partial p}{\partial t} - \beta_f \frac{\partial T}{\partial t}\right) + \rho_f (c_f \nabla p - \beta_f \nabla T)\cdot \mathbf{u} + \rho_f \nabla \cdot \mathbf{u} = 0 \tag{11}$$

Dividing Eq. (11) by ρ_f and substituting $c_t = c_f + c_\phi$ and $\beta_t = \beta_f + \beta_\phi$ yields Eq. (12):

$$(\phi c_t)\frac{\partial p}{\partial t} - (\phi \beta_t)\frac{\partial T}{\partial t} + (c_f \nabla p - \beta_f \nabla T)\cdot \mathbf{u} + \nabla \cdot \mathbf{u} = 0 \tag{12}$$

For the two-dimensional system shown in Fig. 1, the general continuity equation for single phase flow in porous media then reduces to Eq. (13).

$$\phi\left(c_t \frac{\partial p}{\partial t} - \beta_t \frac{\partial T}{\partial t}\right) + u_x\left(c_f \frac{\partial p}{\partial x} - \beta_f \frac{\partial T}{\partial x}\right) + u_z\left(c_f \frac{\partial p}{\partial z} - \beta_f \frac{\partial T}{\partial z}\right) + \frac{\partial u_x}{\partial x} + \frac{\partial u_z}{\partial z} = 0 \tag{13}$$

The general form of the Darcy equation describes the velocity field in the reservoir (Eq. (14)); in which $\mathbf{Z}$ is a unit vector normal to the horizontal and $\mathbf{k}$ is the permeability tensor.

$$\mathbf{u} = -\frac{\mathbf{k}}{\mu}(\nabla P + \rho_f g \mathbf{Z}) \tag{14}$$

For the given system, the gravity vector follows along the x and z axes reducing the Darcy equation to

$$u_x = -k_x\left(\frac{\partial p}{\partial x} - \rho g \sin\alpha\right) \tag{15}$$

$$u_z = -k_z\left(\frac{\partial p}{\partial z} - \rho g \cos\alpha\right) \tag{16}$$

The general form of the energy equation in porous media (Eq. (17)) describes convection and conduction in the reservoir (Grant, 2013).

$$\frac{\partial}{\partial t}\left((1-\phi)\rho_r U_r + \phi \rho_f U_f\right) + \nabla\cdot(\rho_f \mathbf{u}\,\mathcal{H} - \lambda \nabla T) = 0 \tag{17}$$

For liquids and solids, $C_v \approx C_p$ and we can assume $U = \mathcal{H} = C_p T$ (Al-Khoury, 2011). Assuming constant matrix conductivity ($\lambda = \lambda_m = \lambda_f \phi + \lambda_r(1-\phi)$) and expanding the convection term yields Eq. (18):

$$\left((1-\phi)\rho_r C_{pr}\frac{\partial T}{\partial t} + \phi \rho_f C_{pf}\frac{\partial T}{\partial t}\right) + \nabla \rho_f \mathbf{u}\,\mathcal{H} + \rho_f \mathbf{u}.\nabla \mathcal{H} + \rho_f \mathcal{H}\nabla.\,\mathbf{u} - \lambda_m \nabla^2 T = 0 \tag{18}$$

Substituting Eqs. (7) and (8) (the chain rule in time) into Eq. (2), multiplying by $\mathcal{H}$ and rearranging, yields Eq. (19).

$$\nabla \rho_f.\,\mathbf{u}\,\mathcal{H} + \rho_f \mathcal{H}\nabla.\,\mathbf{u} = -C_{pf} T\phi\left(\rho_f c_f \frac{\partial p}{\partial t} - \rho_f \beta_f \frac{\partial T}{\partial t}\right) - C_{pf} T \rho_f \left(\phi c_\phi \frac{\partial p}{\partial t} - \phi\beta\phi \frac{\partial T}{\partial t}\right) \tag{19}$$

Substituting Eq. (19) into Eq. (18), defining matrix heat capacity $\rho_m C_{pm} = \phi \rho_f C_{pf} + (1-\phi)\rho_r C_{pr}$, matrix/fluid heat capacity ratio $M = (\rho_m C_{pm})/(\rho_f C_{pf})$, thermal diffusivity $\kappa = \lambda_m/(\rho_f C_{pf})$, total compressibility $c_t = c_f + c_\phi$ and total thermal expansivity $\beta_t = \beta_f + \beta_\phi$, yields Eq. (20):

$$M\frac{\partial T}{\partial t} + \mathbf{u}.\nabla T - T\phi\left(c_t \frac{\partial p}{\partial t} - \beta_t \frac{\partial T}{\partial t}\right) - \kappa \nabla^2 T = 0 \tag{20}$$

The energy equation for two-dimensions is then

$$M\frac{\partial T}{\partial t} + u_x \frac{\partial T}{\partial x} + u_z \frac{\partial T}{\partial z} - T\phi\left(c_t \frac{\partial p}{\partial t} - \beta_t \frac{\partial T}{\partial t}\right) - \kappa\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial z^2}\right) = 0 \tag{21}$$

For boundary conditions in this development, we assume that all produced mass is returned to the reservoir by reinjection of cooled geofluid and overburden and underburden layers are impermeable. Thus there is no flow across the overburden and underburden (Eqs. (22) and (23)). The only heat that transfers across the overburden and underburden is due to conduction (Eqs. (24) and (25), in which $\kappa'_{ob} = \lambda_{ob}/(\rho_{ob} C_{p,ob})$ and $\kappa'_{ub} = \lambda_{ub}/(\rho_{ub} C_{p,ub})$). The length of the reservoir is assumed to be much larger than the thickness and so it can further be assumed that there is no heat conduction across the model laterals (Eqs. (26) and (27)).

$$u_z = 0 \quad at\ z = 0,\ \forall\ x,\ t \tag{22}$$

$$u_z = 0 \quad at\ z = H,\ \forall\ x,\ t \tag{23}$$

$$\frac{\partial T}{\partial t} = \kappa'_{ob}\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial z^2}\right) \tag{24}$$

$$\frac{\partial T}{\partial t} = \kappa'_{ub}\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial z^2}\right) \tag{25}$$

$$\frac{\partial T}{\partial x} = 0 \quad At\ x = 0,\ \forall\ z,\ t \tag{26}$$

$$\frac{\partial T}{\partial x} = 0 \quad At\ x = L,\ \forall\ z,\ t \tag{27}$$

T_{avg} indicates the average temperature of the reservoir. The initial condition of the reservoir contains both the average reservoir temperature and the linear geothermal gradient in the zone of interest. Temperature gradient normal to the ground (i.e. $\boldsymbol{\tau} = \frac{\partial T}{\partial Z}$) can be calculated using well data. For applying IA to the boundary heat gain/loss process, it is assumed that there is an estimate of the initial reservoir temperature in the legacy data for the reservoirs (see Bassiouni, 1980 and John et al., 1998 for the legacy data used here) or this temperature can be estimated using the region's geothermal gradient. A semi-analytical thermal boundary condition developed by Vinsome and Westerveld (1980), which is available in CMG STARS (STARS manual, 2011), is used for numerically modeling the heat gain/loss by overburden and underburden boundary conditions.

For the injection completion, the boundary conditions are:

$$u_x = -u_T \quad At\ x = L,\ \forall\ z,\ \forall\ t \tag{28}$$

$$T = T_{inj} \quad At\ x = L,\ \forall\ z,\ \forall\ t \tag{29}$$

$$q_{inj} = \int_0^H u_T W\, dz = -q_{prod} \tag{30}$$

in which W is the reservoir width. The sign of u_T is the same for both injection and production but the sign of q differs. A negative sign for u_T is chosen because the velocity vector is in the $-x$ direction (see Fig. 1). For the production completion, the boundary conditions are:

$$u_x = -u_T \quad At\ x = 0,\ \forall\ z,\ \forall\ t \tag{31}$$

$$q_{prod} = -\int_0^H u_T W\, dz \tag{32}$$

The initial conditions for the system are:

$$p = p_i \quad At\ t = 0, \forall\ x, z \tag{33}$$

$$T = T_{avg} + (\tau\cos\alpha)\left(\frac{H}{2} - z\right) + (\tau\sin\alpha)\left(\frac{L}{2} - x\right) \quad At\ t = 0, \forall\ x, z \tag{34}$$

where T_{avg} is the average reservoir temperature at the mid point of the reservoir.

2.1. Dimensionless groups

Appendix A provides the derivation of the dimensionless groups using inspectional analysis which are:

$$\pi_1 = \frac{\phi c_t}{c_f} \qquad \pi_2 = \frac{\phi\beta_t}{\beta_f}$$

$$\pi_3 = M = \frac{\rho_m C_{pm}}{\rho_f C_{pf}} \qquad \pi_4 = c_f p_1^* = \frac{c_f \mu q L}{k_x WH}$$

$$\pi_5 = \beta_f T_1^* = \beta_f T_{avg} \qquad \pi_6 = R_L = \frac{L}{H}\sqrt{\frac{k_z}{k_x}}$$

$$\pi_7 = \frac{L}{H}\tan\alpha \qquad \pi_8 = \frac{L}{H}$$

$$\pi_9 = N_\alpha = \frac{k_x \rho_f g \sin\alpha WH}{q\mu} \qquad \pi_{10} = Pe = \frac{u_T H}{\kappa} = \frac{q}{\kappa W}$$

$$\pi_{11} = \frac{\kappa'_{ob}}{\kappa} \qquad \pi_{12} = \frac{\kappa'_{ub}}{\kappa}$$

$$\pi_{13} = c_f p_i \qquad \pi_{14} = \frac{T_{inj}}{T_{avg}}$$

$$\pi_{15} = \frac{\tau\sin\alpha L}{T_{avg}}$$

Because the focus of the study is to model thermal recovery factor and production temperature, the energy equation is used for scaling the time Appendix A. The dimensionless time for this system is then:

$$t_D = \frac{1}{M}\frac{qt}{LWH} \tag{35}$$

in which $M = (\rho_m C_{pm})/(\rho_f C_{pf})$. Note that if the dimensionless time was defined based on the momentum equation, it would be $t_{D_{hyd}} = \frac{\phi C_f}{C_t}\frac{qt}{LWH}$.

All of these dimensionless groups are needed for transforming the dimensional model into dimensionless representation; however, their form can be heuristically manipulated (multiplied or divided) to get other desirable dimensionless groups.

Five dimensionless groups are identical to those published by previous researchers. They are: π_3 representing matrix to fluid heat capacity ratio (Phillips, 2009), π_6 representing an effective aspect ratio (Shook et al., 1992; Wood et al., 2008), π_7 representing a dip angle group (Shook et al., 1992; Wood et al., 2008), $\frac{\pi_4\pi_9\pi_{15}}{\pi_7}$ representing the Buoyancy number (Shook et al., 1992; Wood et al., 2008) and π_{10} representing the thermal Peclet number (Phillips, 2009).

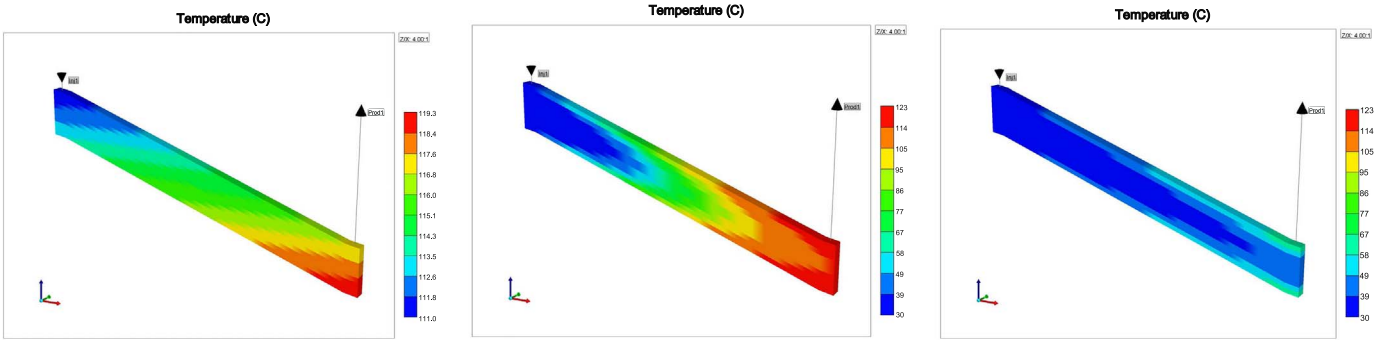
The meaning of other dimensionless groups can be discerned from their derivation or their format: π_1 is the ratio of total compressibility to fluid compressibility, π_2 is the ratio of total expansivity to fluid expansivity and π_5 is fluid expansion due to average reservoir temperature. The π_{11} and π_{12} terms show the ratio of heat conduction across the boundary to heat conduction within the reservoir. The π_{13} term describes fluid compression as a result of reservoir pressure. The π_{14} term scales the injection temperature to the average reservoir temperature and π_{15} scales the temperature difference across the reservoir to the average reservoir temperature and represents the temperature distribution in the reservoir.

Fig. 2 shows an example of the two-dimensional line drive model for a base case simulation in CMG STARS (STARS manual, 2011). The model has 25×1×10 grid blocks with each grid block being 120 m×100 m×10.5 m in the x and y and z directions, respectively. This base model has a dip angle of 5° with an average reservoir temperature of 115 °C assigned to the middle of the model. The geothermal gradient is 24 °C/km. The injection temperature is 30 °C and the injection and production rates are 1,250 m^3/day. The number of grid blocks used for the simulations in this work is fixed at 250 and grid block sizes are calculated based on the distance between the injector and the producer. This distance is one of the parameters varied in the experimental design (Table B2). These values were chosen from the legacy data (Bassiouni, 1980; John et al., 1998).

At the start of the simulation, the temperature drop in the production well is due to the geothermal gradient in the model (Fig. 2a). The reservoir rock cools down as the front slowly propagates through the model and moves towards the production well. When breakthrough happens, there is a sharp decrease in the production temperature (Fig. 2b) and finally at late times, the model temperature asymptotically approaches the injection temperature (Fig. 2c). These observations are typical for the cases simulated and inform some of the choices in the regression modeling to follow.

2.2. Validation

For any given similar flow problem (i.e. the same configuration and boundary conditions), matching the dimensionless numbers yields similar dimensionless results (e.g. thermal recovery factor or dimensionless production temperature) between scales. For validating the dimensionless numbers, the parameters in the groups are varied but the groups are held constant. Matching the response for the cases where the parameters are changed but the dimensionless group values



(a) Grid system at initial conditions (b) Breakthrough conditions (c) Late time conditions

Fig. 2. (a) Reservoir temperature at initial condition (t_D=0) with an average temperature of 115 °C assigned to the middle of the reservoir. Temperature gradient is 24 °C/m and dip angle is 5°. (b) Thermal front at breakthrough t_D=0.6 (ca. 25 years). (c) At late times (t_D=3), the model temperature asymptotically approaches the injection temperature. Vertical axis is exaggerated four times in both figures.

remain constant, suggests that the dimensionless numbers adequately scale the system.

Fifteen different reservoir models were considered such that their geometrical dimensional properties are different but their dimensionless groups are identical (for details see Appendix B). The dimensionless production temperature was then plotted versus dimensionless time. The dimensional result vs. time for these fifteen models does not show any pattern (Fig. 3a). The dimensionless results are plotted versus dimensionless time for the same set of models (Fig. 3b). The models with identical dimensionless numbers (same color) show the same dimensionless temperatures.

3. Numerical simulation

Experimental design is used for sampling the parameter space. Experimental design is an efficient method for sampling and calculating the response with minimum number of runs (Montgomery et al., 2008). Instead of changing the parameters one-at-a-time, by which, interactions cannot be obtained and a large number of simulation runs is needed, experimental design changes the parameters systematically to obtain a smaller set of designed simulation runs, enough to reveal effects and interactions. We used Box-Behnken design (Box et al., 2005). This design needs fewer runs compared with its three-level counterparts (e.g. central composite designs). Central Composite Designs are complete designs which are used to fit a full quadratic model including all interaction terms. Having n factors, Central Composite Design (CCD) includes all the corners of a n dimensional cube when all the factors are scaled between −1 and +1. Similar to CCD, Box-Behnken designs can also be used to fit a full quadratic response surface but with fewer simulation runs. This design also has three levels of factors and is rotatable but yields poorer predictions especially in the corners of the cube because it does not consider the corner points. Because Box-Behnken design requires fewer simulation runs, it will be used to fit a full quadratic response surface including all interaction terms to the detailed simulation results (Box et al., 2005). Fig. 4 shows Circumscribed Central Composite and the Box-Behnken designs.

The numerical simulation proceeds as follows: A Box-Behnken design samples the parameter space and generates models to be simulated using the CMG STARS software (STARS manual, 2011).

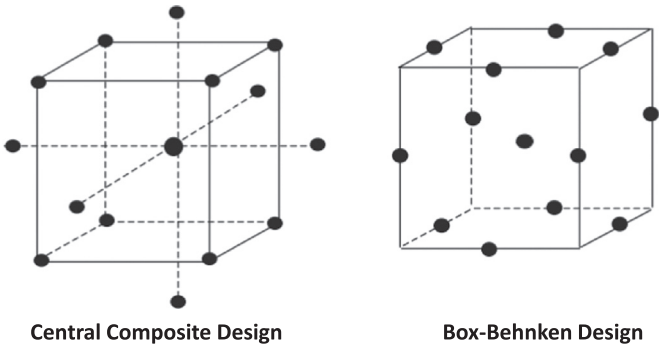


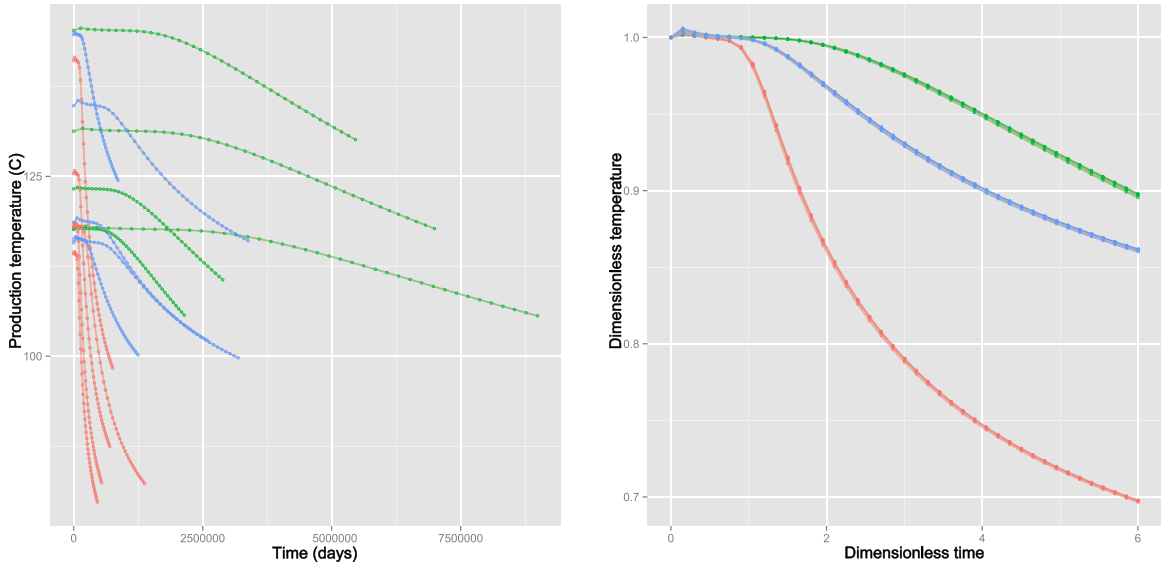
Fig. 4. Two types of experimental design. The Box-Behnken design is used in this work (from Kalla, 2005).

The response for each model is output at specific dimensionless times associated with its parameters. The output response is used for calculating the dimensionless response.

Twenty parameters were used in the Box-Behnken design to sample the parameter space and model the dimensionless production temperature and thermal recovery factor (see Appendix A Table B2). The parameter space distributions were obtained from the data collected during the US Department of Energy sponsored "Wells-of-Opportunity" and "Design Wells" programs (Bassiouni, 1980; John et al., 1998). These runs were divided into 20 parallel runs per batch submission and required ca. 8 minutes to complete on a standard PC workstation. For each run, the π-groups were noted and the dimensionless response (i.e. dimensionless production temperature or energy recovery factor) were obtained as a function of dimensionless time. The set of experimental design runs then makes up the database of sample responses (Fig. 5).

4. Statistical modeling

The statistical modeling section proceeds as follows: the entire database is shown in a single violin plot and is divided into piecewise linear segments. Before creating the regression models for each segment, an algorithm known as Boosting (James et al., 2013; Kuhn and Johnson, 2013; Hastie et al., 2009) is used to find the important



(a) Dimensional results (b) Dimensionless results

Fig. 3. (a) Models with the same dimensionless numbers (i.e. same color) show different production temperature in dimensional space. (b) Models with the same dimensionless numbers show similar dimensionless production temperature.

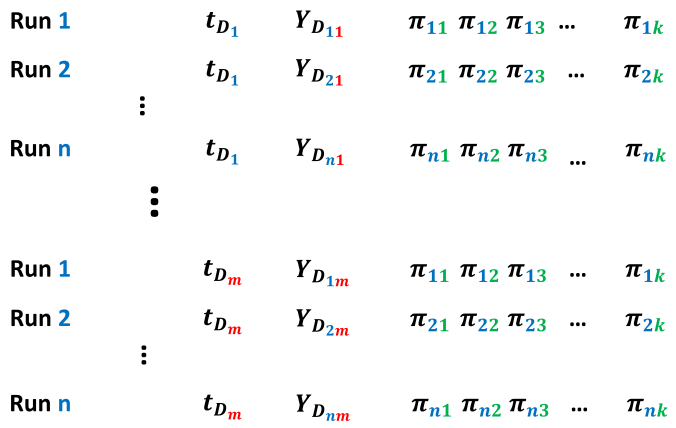


Fig. 5. Configuration of the database. n stands for the number of simulation runs obtained from the Box-Behnken design (blue color), m stands for the number of dimensionless time (red color), k stands for the number of dimensionless numbers (green color), Y_D stands for the dimensionless response. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

dimensionless numbers for predicting the response. Then, a model selection procedure known as Best Subset Selection (James et al., 2013; Kuhn and Johnson, 2013; Hastie et al., 2009) is used to compare possible sub-models and select the best one. The reduced final models are then presented and assessed using testing runs. Finally, limitations of these models are discussed.

Table 1
Ranges of the dimensionless numbers used in the sensitivity analysis.

Group	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9
Low	0.204	0.142	0.463	0.00163	0.0602	5.46	0.000	11.51	0
High	0.967	0.305	0.738	0.0598	0.1738	90.63	15.98	120.85	0.724

Group	π_{10}	π_{11}	π_{12}	π_{13}	π_{14}	π_{15}
Low	96.16	0.264	0.264	0.028	0.173	0
High	615.48	0.482	0.482	0.044	0.897	0.1325

4.1. Modeling production temperature

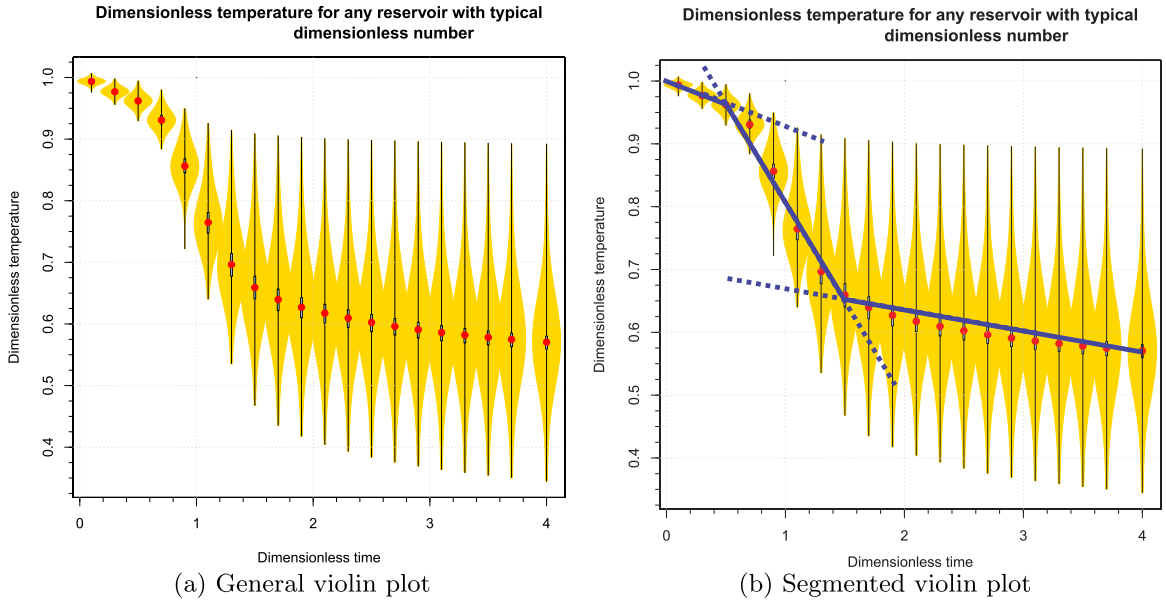
The entire database can be plotted in a single violin plot for each response. A violin plot was created for the dimensionless number ranges given in Table 1. Violin plots are useful for answering generic uncertain questions with minimum data (White, 2013). In this work, they were also used for segmenting the dimensionless time.

A violin plot (Fig. 6a) is a combination of a box plot and a rotated kernel density plot (i.e. an estimation for the probability density function). In the violin plot, the middle dot shows the median (which is identical to the mode and maximum likelihood estimator of the mean in a normal distribution), the white boxes indicate the lower to upper quartile and the thin black lines are called whiskers and show the range of values at each dimensionless time. Violin plots can also be updated when additional certain data are obtained.

Violin plots are used to show the statistical spread of the response variable as a function of time. For this work, the are also used to determine how many piecewise regression models are needed and to specify their dimensionless time ranges since it will be shown that dimensionless time is the key parameter in the modeled system. As Fig. 6b indicates three regression models can predict the median points (i.e. red points): $0 < t_D \leq 0.5$, $0.5 < t_D \leq 1.5$ and $1.5 < t_D \leq 4$. Each segment can be considered as a line which goes through the median points. Note that for the system under consideration, $t_D > 4$ is well past the effective life of a geothermal power plant.

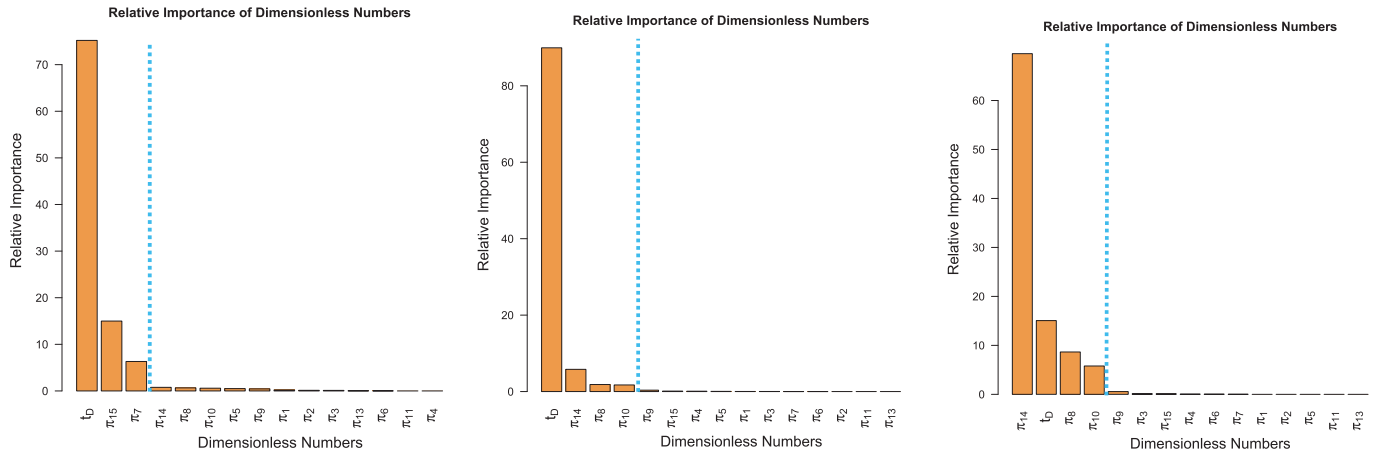
In order to determine the relative importance of the predictors, a robust tree-based algorithm, known as Boosting is used. The Boosting method creates an ensemble of independent regression trees and updates them sequentially. Each tree is fit to the current step residual and the final model is the average of all the trees weighted by a shrinkage parameter which controls the rate at which Boosting learns (James et al., 2013; Hastie et al., 2009; Kuhn and Johnson, 2013). There were 10,000 trees with an interaction depth of 4 and shrinkage parameter of 0.01 in the Boosting algorithm for this work. Finding the important dimensionless numbers is also useful in the validation process. Reservoirs with identical important dimensionless numbers show similar behavior in the dimensionless space.

Table 1 shows the ranges of the dimensionless numbers used in the sensitivity analysis of the dimensionless thermal recovery response. Dimensionless numbers can be rank-ordered based on their relative importance in predicting the response. Fig. 7 shows important



(a) General violin plot (b) Segmented violin plot

Fig. 6. (a) Violin plot for fining dimensionless temperature of uncertain reservoirs. This violin plot can be used as a rule of thumb for answering uncertain questions with minimum data. (b) Three ranges for segmenting the model are considered: $0 < t_D \leq 0.5$, $0.5 < t_D \leq 1.5$ and $1.5 < t_D \leq 4$. The blue lines are sketched manually.



(a) Finding important dimensionless numbers for the range $0 < t_D \leq 0.5$

(b) Finding important dimensionless numbers for the range $0.5 < t_D \leq 1.5$

(c) Finding important dimensionless numbers for the range $1.5 < t_D \leq 4$

Fig. 7. Important dimensionless numbers for each range of the dimensionless time. These dimensionless numbers change smoothly and continuously during dimensionless time. Because modeling every dimensionless time is redundant, the sensitivity analysis is performed over a range of dimensionless time. The cyan line shows the cut off for the dimensionless numbers with less than 1% importance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

dimensionless numbers for each range of the dimensionless time. These dimensionless numbers change smoothly and continuously during dimensionless time. Dimensionless time is the most important factor in determining dimensionless temperature for the first two segments, $0 < t_D \leq 1.5$, while temperature ratio (π_{14}) is more important for the late-time response ($1.5 < t_D \leq 4$). Peclet number (π_{10}) and aspect ratio (π_8) are other important dimensionless numbers in predicting the response at late times. The dimensionless numbers which do not affect the response can be safely disregarded for prediction. Dimensionless numbers with less than 1% relative importance were ignored which is indicated by terms to the right of the dashed lines in Fig. 7.

Once the important dimensionless numbers were found, regression was used to create response surfaces. In this work, dimensionless production temperature and thermal recovery factor are defined as the response variables. The dimensionless production temperature is the ratio of the producing fluid temperature to the initial average reservoir temperature. A model for dimensionless production temperature can be used to calculate the rate of energy production versus time. The energy recovery factor is defined as the ratio of the produced energy to the total energy in place before exploiting the reservoir (Muffler and Cataldi, 1978; Nathenson, 1975; Williams et al., 2008). Cumulative produced energy can be directly calculated using energy recovery factor.

The model for dimensionless production temperature was fit at eighty values for dimensionless time with a step size of t_D=0.05 (i.e. 0, 0.05, 0.1,..., 4). Segmented (piecewise) regression was used to model dimensionless production temperature using 761 numerical simulation models. Since dimensionless time is the dominant parameter affecting the response, the violin plot (Fig. 6b) was used for segmenting the model. The segments can then be tested using trial and error to determine whether they are the best ranges for segmenting. The violin plot shows that three lines can go through the maximum likelihood points (i.e. red points) and can sufficiently model the entire violin. The first segment ($0 \leq t_D \leq 0.5$) has 11 values for dimensionless time (761×11 samples points), the second segment ($0.5 < t_D \leq 1.5$) has twenty values (761×20 points) and the third segment ($1.5 < t_D \leq 4$) has fifty values for dimensionless time (761×50 points). In Fig. 8, the black area shows the presence of a predictor and the white area shows its absence. The model with the largest R^2 value was found by an "Best Subset Selection" algorithm (James et al., 2013) and was selected as the best model (Fig. 8). Thus, the best model for the first segment (a) has an R^2 value of 0.8853, the best model for the second segment (b) has an R^2 value of 0.9651 and the best model for the third segment (c) has an R^2 value of 0.9558. These selected models for dimensionless production temperature are presented in Box 1.

The purpose of the model selection for this work is to obtain a simple model with a minimum number of predictors and interactions. Choosing a simpler model trades off a small increase in the model bias. Introducing more predictors improves the R^2. The optimum model is both simple (i.e. has fewer terms) and representative (i.e. close to the best model). The choice of the final model is based on the modeler's selection and how the person trades off increasing terms with increasing R^2 value. These reduced final models are then assessed using testing runs.

Box 1.Proposed model for dimensionless production temperature.

For $0 \leq t_D \leq 0.5$, we have:

$$T_D = 0.9936 - 0.0267t_D - 0.0516\pi_{15} + 0.0002\pi_7 - 0.8434t_D\pi_{15} + 0.0029t_D\pi_7 + 0.0045\pi_{15}\pi_7 \quad (36)$$

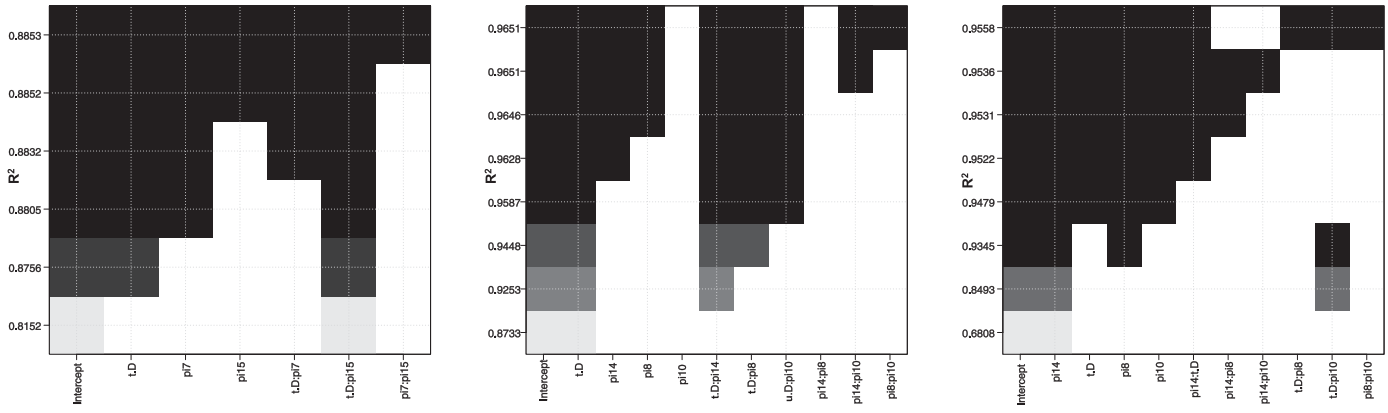
for $0.5 < t_D \leq 1.5$, we have:

$$T_D = 1.3429 - 0.6041t_D - 0.4467\pi_{14} - 0.001\pi_8 + 0.6101t_D\pi_{14} + 0.0022t_D\pi_8 - 0.0003t_D\pi_{10} + 0.0003\pi_{10}\pi_{14} - 0.0001\pi_8\pi_{10} \quad (37)$$

and for $1.5 < t_D \leq 4$, we have:

$$T_D = 0.4528 - 0.0587t_D + 0.4880\pi_{14} - 0.0004\pi_{10} + 0.0037\pi_8 + 0.0729t_D\pi_{14} - 0.003\pi_{14}\pi_8 - 0.0003t_D\pi_8 + 0.0004\pi_{14}\pi_{10} \quad (38)$$

Fig. 9 shows the assessment of the reduced final model (Box 1). In each sub-figure, 711 runs were used for training and 50 runs for testing. This means that 7,821 (i.e. 711×11) training samples (blue points) and 550 (i.e. 50×11) testing samples (red points) were used for the first segment, 14,220 training samples and 1,000 testing samples were used for the second segment and 35,550 training samples and 2,500 testing samples were used for the third segment.



(a) $0 < t_D \leq 0.5$. (b) $0.5 < t_D \leq 1.5$ (c) $1.5 < t_D \leq 4$.

Fig. 8. Reducing the models for production temperature at each range of dimensionless time.

4.2. Modeling thermal recovery factor

The same procedure that was used to model dimensionless production temperature was used to develop a thermal recovery factor model. First, violin plots are presented and the segmentation ranges for the dimensionless time are heuristically found. The important dimensionless numbers in each range of the dimensionless time are found and then the optimum models consisting of these important dimensionless numbers are selected. Finally, the simplified models are presented and assessed.

Again, violin plots were used to determine the number of segments and their ranges (Fig. 10(a)). Fig. 10(b) indicates two piecewise regression models can predict the maximum likelihood points (i.e. red points). The ranges of the first and second model are $0 < t_D \leq 1$ and $1 < t_D \leq 4$. Each of these segments can be considered as a line which goes through the maximum likelihood points.

A sensitivity analysis is performed to find the important dimensionless numbers in predicting the response (Fig. 11). For $0 \leq t_D \leq 1$, dimensionless time is the only predictor that is needed and for $1 < t_D \leq 4$, dimensionless time, temperature ratio (π_{14}) and Peclet number (π_{10}) are needed for predicting the recovery factor. Note that thermal recovery factor may have values greater than one because the reservoir may recharge heat from its cap/base rock.

These important dimensionless numbers were used for developing a model for the thermal recovery factor. Fig. 12 compares the R^2 values of all subset regression models. An R^2 value of 0.9876 was chosen for the best model. The simplified models for energy recovery factor are highlighted in Box 2.

Box 2.Proposed model for thermal recovery factor.

For $0 \leq t_D \leq 1$, we have:

$$RF = 1.083 t_D \tag{39}$$

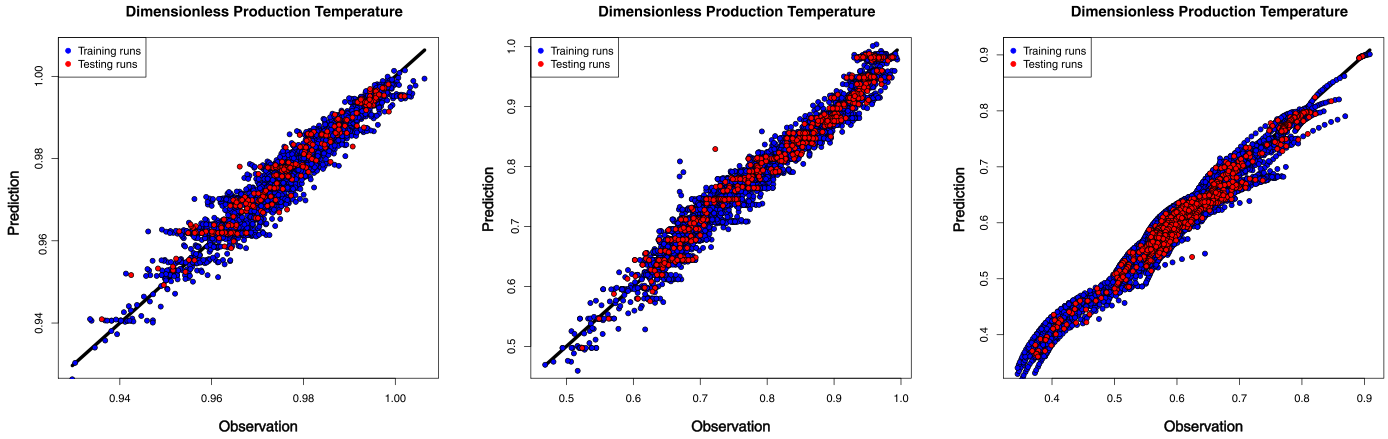
and for $1 < t_D \leq 4$, we have:

$$RF = 1.01 + 0.0405 t_D - 0.829 \pi_{14} - 0.002 \pi_8 + 1.01 t_D \pi_{14} + 0.0022 t_D \pi_8 - 0.0071 \pi_{14} \pi_8 \tag{40}$$

Finally, the reduced models (Box 2) are assessed using testing runs (Fig. 13). Similar to dimensionless temperature, in each sub-Fig. 711 runs are used for training and 50 runs for testing. The assessment plots show a good match between the model prediction and observation values (i.e. simulation results) and indicate that the fitted models are adequate for predicting thermal recovery factor.

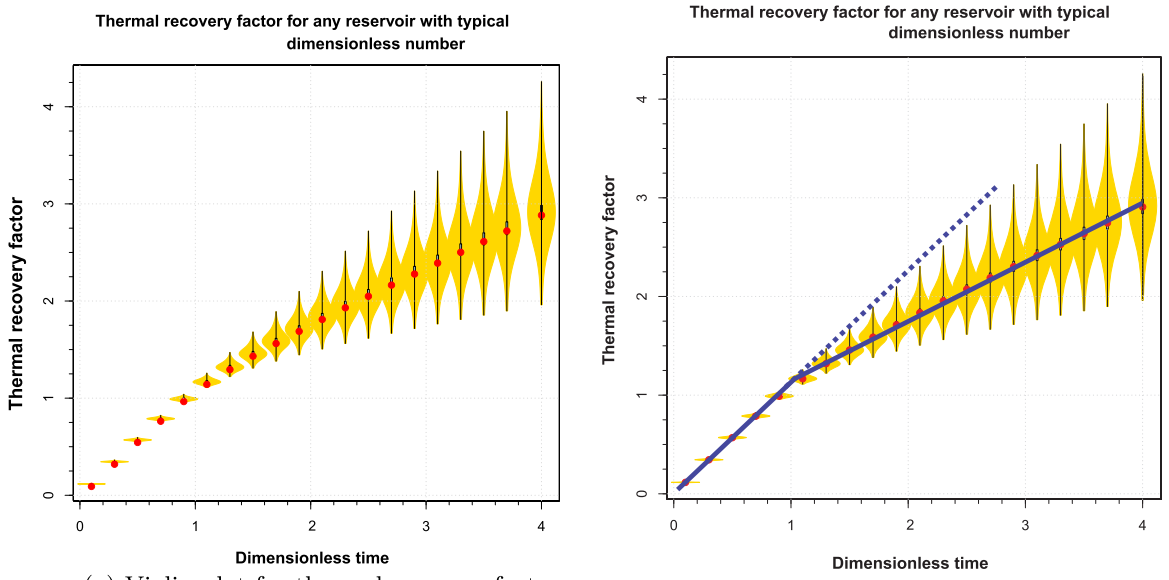
4.3. Limitations

The presented workflow is not a replacement for full numerical simulation and history matching of a reservoir. Full numerical simula-



(a) $0 \leq t_D \leq 0.5$ (b) $0.5 < t_D \leq 1.5$ (c) $1.5 < t_D \leq 4$

Fig. 9. Assessing the models for the dimensionless production temperature. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)



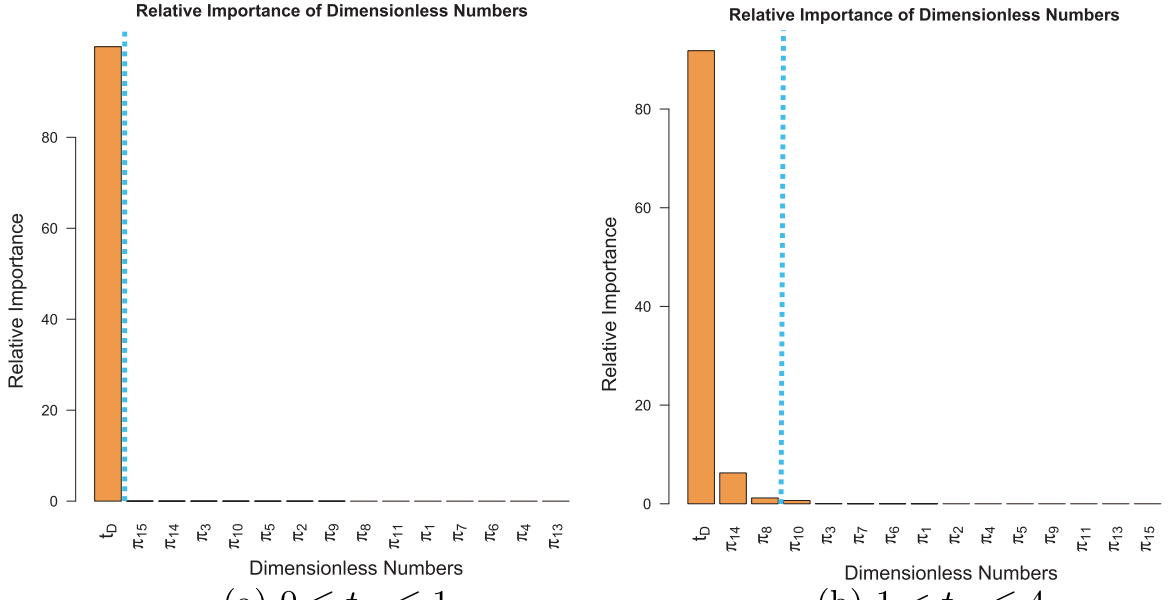
(a) Violin plot for thermal recovery factor. (b) Segmented violin plot.

Fig. 10. (a) Violin plot for thermal recovery factor. (b) Segmented violin plot for thermal recovery factor. Two ranges for segmenting the model are considered: $0 < t_D \leq 1$ and $1 < t_D \leq 4$. The blue lines are sketched manually. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tion and history matching of a developed reservoir is necessary for forecasting the future and to design the initial field implementation. It is logical that for a flow scenario with different boundary conditions (e.g. horizontal wells instead of vertical), the workflow should be repeated since the boundary and initial conditions drive many of the parameter choices in the scaling. For example the behavior of the line drive system is dependent on the time to thermal breakthrough and the spread in the violin plots show the impact of this dependency. Nevertheless, the presented workflow can be applied to large models because the simulation of models are completely independent from each other. The line drive pattern is a reasonable proxy for other pattern types and reinjection of produced fluid is a necessary requirement for the low enthalpy systems for which these models were developed. Rapid growth in cloud computing makes this workflow useful. The entire set of runs can be effectively distributed on a cluster of nodes and the simulation results can be retrieved to create the described database for statistical analysis. If grid and boundary conditions are chosen wisely, simulating the entire set of necessary runs should take approximately the same amount of time of a single detailed reservoir simulation run (plus a little time for retrieving and assembling the database). In the following, regression model applications are presented. These applications were developed to fall within the statistics of the data from Bassiouni (1980) and John et al. (1998). Ansari (2016) presents the successful application of the models to additional examples that fall outside these parameter ranges.

5. Applications

In the following some applications of the developed models are discussed. A reservoir with some unknown data is described in



(a) $0 \leq t_D \leq 1$ (b) $1 < t_D \leq 4$

Fig. 11. Finding the important dimensionless numbers for predicting thermal recovery factor for each range of the dimensionless time. The cyan line shows the cut off for the dimensionless numbers with less than 1% importance.

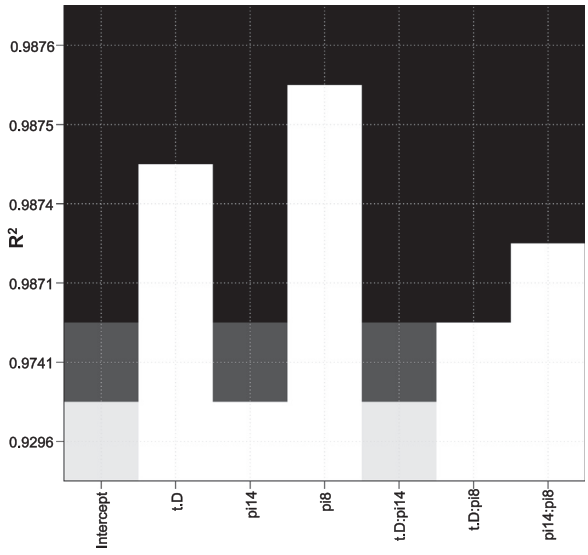


Fig. 12. Comparing the R^2 values of the models for thermal recovery factor for the range $1 < t_D \leq 4$.

Question A and the uncertainty in its temperature and produced energy is evaluated. Questions 2 and 3 use the developed models to calculate these values for the same reservoir.

Example 1. Ten wells (five injectors and five producers) are completed uniformly using line drive patterns in a geopressured-geothermal zone. The producers can provide a rate of 1,000 m^3/day, all of which is injected back into the reservoir. Only the area, thickness and porosity are known to be ca. 25 km^2 (12,500 m×2,000 m), 30 m and 0.2 respectively. Other information about the reservoir (e.g average temperature, dip angle, temperature gradient, etc.) are unknown but they are within the statistics of the data in the area (Table 1). The goal is to obtain the dimensionless temperature and thermal recovery factor after 30 years of production.

Answer. Typical values for rock and water volumetric heat capacities are 2×10^6 and 4×10^6 J/m^3C and the drainage area available for a single line drive pattern is 60 acres (242,811 m^2):

$$\rho_m C_{pm} = \phi\, \rho_w C_{pw} + (1-\phi)\rho_r C_{pr} = 0.2\times 4\times 10^6 + 0.8\times 2\times 10^6 = 2.4\times 10^6 \text{ J/m}^3\,^\circ\text{C}$$

$$M = \frac{\rho_m C_{pm}}{\rho_f C_{pf}} = \frac{2.4\times 10^6}{4\times 10^6} = 0.6 \tag{41}$$

$$t_D = \frac{1}{M}\frac{qt}{LWH} = \frac{1}{0.6}\frac{1000\ \frac{\text{m}^3}{\text{day}}\times 30\times 365\ \text{days}}{60\times 4046.86\times 30\ \text{m}^3} = 2.5 \tag{42}$$

Using Fig. 6, maximum, median and minimum dimensionless temperature values for this reservoir after 30 years of production are ca. 0.39, 0.61 and 0.9 respectively.

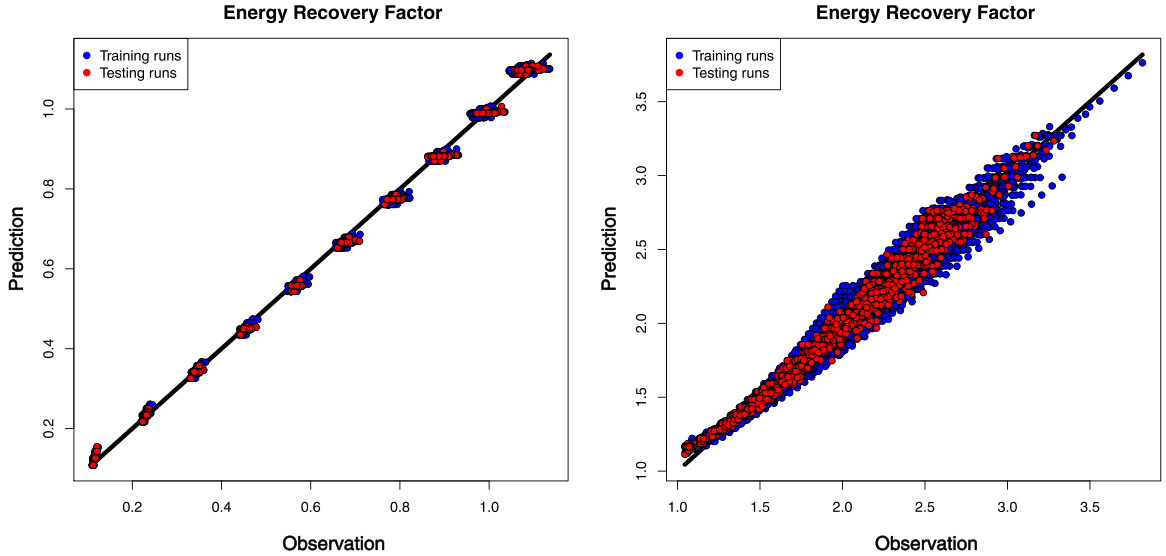
Using Fig. 10(a), the maximum, median and minimum thermal recovery factor values for this reservoir after 30 years of production are ca. 2.6, 2.1 and 1.6 respectively. The recovery factor can be multiplied by the initial energy in the system to obtain the energy that can be generated over the life of a plant. If this value is within the bounds of an economic project, detailed modeling and simulation could be used to assess the reservoir; if not, the area could be eliminated from consideration.

Example 2. For the reservoir in the Question 1, if the average temperature, injection temperature and the rock's thermal capacity are 120 °C, 40 °C and 2×10^6 J/m^3 °C respectively; find the rate of energy production and net cumulative energy that could be produced from this system in 30 years.

Answer. The maximum, median and minimum production temperature after 30 years of production are: 108, 73.2 and 46.8 C and water heat capacity at these temperatures is ca. 4.1×10^6 J/m^3 °C. Thus using Eq. (43), the rate of energy production would be 2.79×10^{11} (3.2 MW), 1.36×10^{11} (1.57 MW) and 2.79×10^{10} J/day (0.32 MW) respectively.

$$\dot{H} = q\rho_f C_{pf} (T_{prod} - T_{inj}) \tag{43}$$

Now that we have temperature, we can find accurate water thermal capacity values at 120 °C and 40 °C are 3.99×10^6 J/m^3 °C and 4.12×10^6 J/m^3 °C. The energy in place is:



(a) $0 \leq t_D \leq 1$ (b) $1 \leq t_D \leq 4$

Fig. 13. Assessing the models for the thermal recovery factor.

$$\rho_m C_{pm} = \phi \rho_w C_{pw} + (1-\phi)\rho_r C_{pr} = 0.2 \times 3.99 \times 10^6 + 0.8 \times 2 \times 10^6 = 2.398 \times 10^6 \text{ J/m}^3\,^\circ\text{C}$$

$$H_{in\ place} = \rho_m C_{pm} V (T_p - T_{ref}) = 2.398 \times 10^6 \times 60 \times 4047 \times 30 \times (120-25) = 1.66 \times 10^{15} \text{ J}$$

The produced enthalpy is:

$$H_{pord} = RF \times H_{in\ place}$$

Thus the maximum, median and minimum produced energy values are 4.3×10^{15}, 3.49×10^{15}, 2.66×10^{15} J:

The enthalpy that is reinjected back into the reservoir during 30 years is:

$$H_{inj} = q_i t \rho_m C_{pm} (T_i - T_{ref}) = 1000 \times 4.12 \times 10^6 \times 30 \times 365.24 \times (40-25) = 6.77 \times 10^{14} \text{ J}$$

The net produced energy is then:

$$H_{net} = H_{prod} - H_{inj}$$

The maximum, median and minimum net produced energy values from a single line drive pattern would be 3.62×10^{15}, 2.81×10^{15} and 1.98×10^{15} J.

Example 3. For the reservoir in the Question 1, assume a length and width of 1215 m and 200 m and a thermal conductivity of 2.5×10^5 for the reservoir rock. Calculate the dimensionless production temperature and energy recovery factor using the developed models.

Answer.

$$\kappa = \frac{\lambda_m}{\rho_f C_{pf}} = \frac{2.5 \times 10^5}{4 \times 10^6} = 0.0625 \tag{44}$$

$$\pi_{10} = Pe = \frac{u_T H}{\kappa} = \frac{q}{\kappa W} = \frac{1000}{0.0625 \times 200} = 80 \tag{45}$$

$$\pi_{14} = \frac{40}{120} = 0.333 \tag{46}$$

$$\pi_8 = \frac{L}{H} = \frac{1215}{30} = 40.5 \tag{47}$$

$$T_D = 0.4528 - 0.0587 t_D + 0.4880\pi_{14} - 0.0004\pi_{10} + 0.0037\pi_8 + 0.0729 t_D\pi_{14} - 0.003\pi_{14}\pi_8 - 0.0003 t_D\pi_8 + 0.0004\pi_{14}\pi_{10} = 0.587$$

$$RF = 1.01 + 0.0405 t_D - 0.829\pi_{14} - 0.002\pi_8 + 1.01 t_D\pi_{14} + 0.0022 t_D\pi_8 - 0.0071\pi_{14}\pi_8 = 1.72$$

Thus, the estimated production temperature after 30 years is 70.4 °C and the total produced energy and net produced energy after 30 years are 2.86×10^{15} and 2.18×10^{15}, respectively.

Example 4. The Gueydan Dome geothermal reservoir has a volume of 8.58×10^8 m^3 and porosity of 0.2. If four production and four injection wells were used to develop this reservoir and the rate for each well is 2,000 m^3/day. Average reservoir temperature is 140 °C, injection temperature is assumed to be 70 °C and reservoir rock thermal capacity is 1.97×10^6 J/m^3 °C. What would be the estimated recovery factor from this reservoir in 30 years? (assume reservoir dip angle is zero)

Answer. Water thermal capacity at 140 °C is ca. 3.956×10^6 J/m^3 °C. The energy in place is:

$$\rho_m C_{pm} = \phi \rho_w C_{pw} + (1-\phi)\rho_r C_{pr} = 0.2 \times 3.956 \times 10^6 + 0.8 \times 1.97 \times 10^6 = 2.37 \times 10^6 \text{ J/m}^3\,^\circ\text{C} \tag{48}$$

$$M = \frac{\rho_m C_{pm}}{\rho_f C_{pf}} = \frac{2.37 \times 10^6}{3.956 \times 10^6} = 0.6 \tag{49}$$

$$t_D = \frac{1}{0.6}\frac{qt}{LWH} = \frac{1}{0.6}\frac{8000\ \frac{\text{m}^3}{\text{day}} \times 30 \text{ years}}{8.58 \times 10^8 \text{ m}^3} = 0.17 \tag{50}$$

$$\pi_7 = \frac{L}{H}\tan(\alpha) = 0 \tag{51}$$

$$\pi_{15} = \frac{\tau \sin(\alpha) H}{T_{avg}} = 0 \tag{52}$$

$$T_D = 0.9936 - 0.0267 t_D - 0.0516\pi_{15} + 0.0002\pi_7 - 0.8434 t_D\pi_{15} + 0.0029 t_D\pi_7 + 0.0045\pi_{15}\pi_7 = 0.989 \tag{53}$$

$$RF = 1.083 t_D = 0.18 \tag{54}$$

$$H_{in\ place} = \rho_m C_{pm} V (T_p - T_{ref}) = 2.37 \times 10^6 \times 8.58 \times 10^8 \times (140-25) = 2.34 \times 10^{17} \text{ J} \tag{55}$$

The produced enthalpy is:

$$H_{pord} = RF \times H_{in\ place} = 4.22 \times 10^{16} \tag{56}$$

Water thermal capacity at 70 C is 4.09×10^6 J/m^3 °C. The enthalpy that is injected back into the reservoir during 30 years is:

$$H_{inj} = q_i \rho_m C_{pm} (T_i - T_{ref}) = 8000 \times 4.09 \times 10^6 \times 30 \times 365.24 \times (70-25) = 1.61 \times 10^{16} \text{ J} \tag{57}$$

Thus the net produced energy is:

$$H_{net} = H_{prod} - H_{inj} = 2.61 \times 10^{16} \text{ J} \tag{58}$$

Ansari et al. (2014) simulated an energy recovery value for this reservoir at 2.72×10^{16} J.

6. Conclusion

Analytical solutions are not always available and interpreting simulator results is difficult. Statistics can be used to interpret numerical simulation results and create simple predictive models. In this work, dimensionless production temperature and thermal recovery factor for a regular line drive design were modeled using inspectional analysis and statistical modeling. Inspectional analysis produced a minimum number of fifteen dimensionless groups necessary to describe a line drive design. Then, experimental design, Boosting and regression were used to create the models for different segments of the dimensionless time. The final models were reduced, tested and their applications were presented. The workflow presented here is generic and can be used to interpret and translate the output of a simulator into predictive models in other research areas involving numerical simulation.

Acknowledgements

The authors gratefully acknowledge financial support for this work from the US Department of Energy under grant DE-EE0005125. We thank Computer Modeling Group for providing reservoir simulation software. We also thank the members of the LSU Geothermal team for their comments, suggestions and ideas supporting our efforts. The R source codes and simulation results for this manuscript are available upon request to the authors.

Appendix A. Inspectional analysis

After the physical process is formulated, we use scale factors and follow the work of previous researchers to non-dimensionalize the equations (Shook et al., 1992; Novakovic, 2002; Wood et al., 2008; Jin et al., 2010). The equations listed in the main text involve twenty three parameters (ϕ, c_t, c_f, β_t, β_f, M, k_x, k_z, L, H, κ, κ'_{ob}, κ'_{ub}, ρ_f, p_i, T_{avg}, T_{inj}, τ, α, q, μ) and seven variables (x, z, u_x, u_z, p, T, t) to be scaled using Eq. (A.1). In these linear combinations, the quantities with an asterisk "*" are called scale factors and subscripts "1" and "2" and "D" indicate the multiplicative and additive scale factors and dimensionless variables, respectively.

$$x = x_1^* x_D + x_2^* z = z_1^* z_D + z_2^* u_x = u_{x1}^* u_{xD} + u_{x2}^* u_z = u_{z1}^* u_{zD} + u_{z2}^* p = p_1^* p_D + p_2^* T = T_1^* T_D + T_2^* t = t_1^* t_D + t_2^* \tag{A.1}$$

Substituting scaled variables into the equations transforms all the equations into the dimensionless space.

A.1. Continuity equation

$$\frac{\partial p_D}{\partial t_D} - \frac{\beta_t}{c_t}\frac{T_1^*}{p_1^*}\frac{\partial T_D}{\partial t_D} + (u_{x1}^* u_{xD} + u_{x2}^*)\left(\frac{c_f}{\phi c_t}\frac{t_1^*}{x_1^*}\frac{\partial p_D}{\partial x_D} - \frac{\beta_f}{\phi c_t}\frac{T_1^* t_1^*}{p_1^* x_1^*}\frac{\partial T_D}{\partial x_D}\right) + (u_{z1}^* u_{zD} + u_{z2}^*)\left(\frac{c_f}{\phi c_t}\frac{t_1^*}{z_1^*}\frac{\partial p_D}{\partial z_D} - \frac{\beta_f}{\phi c_t}\frac{T_1^* t_1^*}{p_1^* z_1^*}\frac{\partial T_D}{\partial z_D}\right) + \frac{1}{\phi c_t}\frac{u_{x1}^* t_1^*}{p_1^* x_1^*}\frac{\partial u_{xD}}{\partial x_D} + \frac{1}{\phi c_t}\frac{u_{z1}^* t_1^*}{p_1^* z_1^*}\frac{\partial u_{zD}}{\partial z_D} = 0 \tag{A.2}$$

A.2. Darcy equation

$$u_{xD} = -\frac{k_x}{u_{x1}^* \mu}\left(\frac{p_1^*}{x_1^*}\frac{\partial p_D}{\partial x_D} - \rho_f g \sin\alpha\right) - \frac{u_{x2}^*}{u_{x1}^*} \tag{A.3}$$

$$u_{zD} = -\frac{k_z}{u_{z1}^* \mu}\left(\frac{p_1^*}{z_1^*}\frac{\partial p_D}{\partial z_D} - \rho_f g \cos\alpha\right) - \frac{u_{z2}^*}{u_{z1}^*} \tag{A.4}$$

A.3. Energy equation

$$\frac{\partial T_D}{\partial t_D} + (u_{x1}^* u_{xD} + u_{x2}^*)\frac{t_1^*}{M x_1^*}\frac{\partial T_D}{\partial x_D} + (u_{z1}^* u_{zD} + u_{z2}^*)\frac{t_1^*}{M z_1^*}\frac{\partial T_D}{\partial z_D} + (T_1^* T_D + T_2^*)\left(\frac{\phi c_t}{M}\frac{p_1^*}{T_1^*}\frac{\partial p_D}{\partial t_D} - \frac{\phi\beta_t}{M}\frac{\partial T_D}{\partial t_D}\right) - \kappa\left(\frac{t_1^*}{M {x_1^*}^2}\frac{\partial^2 T_D}{\partial x_D^2} + \frac{t_1^*}{M {z_1^*}^2}\frac{\partial^2 T_D}{\partial z_D^2}\right) = 0 \tag{A.5}$$

A.4. Reservoir boundary condition

$$\frac{T_1^*}{t_1^*}\frac{\partial T_D}{\partial t_D} = \kappa'_{ob}\left(\frac{T_1^*}{{x_1^*}^2}\frac{\partial^2 T_D}{\partial x_D^2} + \frac{T_1^*}{{z_1^*}^2}\frac{\partial^2 T_D}{\partial z_D^2}\right) \tag{A.6}$$

$$\frac{T_1^*}{t_1^*}\frac{\partial T_D}{\partial t_D} = \kappa'_{ub}\left(\frac{T_1^*}{{x_1^*}^2}\frac{\partial^2 T_D}{\partial x_D^2} + \frac{T_1^*}{{z_1^*}^2}\frac{\partial^2 T_D}{\partial z_D^2}\right) \tag{A.7}$$

$$u_{zD} = -\frac{u_{z2}^*}{u_{z1}^*} \quad At\ z_D = -\frac{z_2^*}{z_1^*},\ \forall\ x_D,\ t_D \tag{A.8}$$

$$u_{zD} = -\frac{u_{z2}^*}{u_{z1}^*} \quad At\ z_D = \frac{H - z_2^*}{z_1^*},\ \forall\ x_D,\ t_D \tag{A.9}$$

$$\frac{\partial T_D}{\partial x_D} = 0 \quad At\ x_D = 0,\ \forall\ z_D,\ t_D \tag{A.10}$$

$$\frac{\partial T_D}{\partial x_D} = 0 \quad At\ x_D = \frac{L - x_2^*}{x_1^*},\ \forall\ z_D,\ t_D \tag{A.11}$$

A.5. Injection

$$u_{xD} = \frac{u_T}{u_{x1}^*} \quad At\ x_D = \frac{L - x_2^*}{x_1^*},\ \forall\ z_D,\ \forall\ t_D \tag{A.12}$$

$$T_D = \frac{T_{inj}}{T_1^*} \quad At\ x_D = -\frac{x_2^*}{x_1^*},\ \forall\ z_D,\ \forall\ t_D \tag{A.13}$$

$$q_{inj} = u_T WH = -\,q_{prod} \tag{A.14}$$

A.6. Production

$$u_{xD} = \frac{u_T}{u_{x1}^*} \quad At\ x_D = \frac{x_2^*}{x_1^*},\ \forall\ z_D,\ \forall\ t_D \tag{A.15}$$

A.7. Initial conditions

$$p_D = \frac{p_i}{p_1^*} \quad At \quad t_D = -\,\frac{t_2^*}{t_1^*} \quad \forall\ x_D,\ z_D \tag{A.16}$$

$$T_D = \frac{T_{avg}}{T_1^*} + \frac{\tau\cos\alpha}{T_1^*}\left(\frac{H}{2} - z_1^* z_D - z_2^*\right) + \frac{\tau\sin\alpha}{T_1^*}\left(\frac{L}{2} - x_1^* x_D - x_2^*\right) \quad \forall\ x_D,\ z_D,\ t \tag{A.17}$$

A.8. Dimensionless groups

The form of the equations should not change by transforming to dimensionless space. Many of these scaling groups can be set to zero or one (Eq. (A.18)). The remaining scaling groups that cannot be assigned any number defines the dimensionless groups required to describe the system. These remaining groups should also be analyzed for dependency to eliminate redundant dimensionless numbers and further be heuristically manipulated to achieve the most succinct form of the minimum dimensionless numbers. We further note that the dimensionless time is defined based on the energy equation.

$$u_{x2}^* = u_{z2}^* = z_2^* = x_2^* = t_2^* = T_2^* = p_2^* = 0 \tag{A.18}$$

$$x_1^* = L \qquad z_1^* = H$$

$$\frac{u_T}{u_{x1}^*} = 1 \Rightarrow u_{x1}^* = u_T \qquad \frac{u_{x1}^*}{x_1^*} = \frac{u_{z1}^*}{z_1^*} \Rightarrow u_{z1}^* = \frac{H}{L}u_T$$

$$T_1^* = T_{avg} \qquad t_1^* = M\frac{x_1^*}{u_{x1}^*} = M\frac{L}{u_t} = M\frac{PV}{\phi q}$$

$$p_1^* = \frac{u_T L\mu}{k_x}$$

Note that if the dimensionless time was defined based on the momentum equation, it was $t_{D_{hyd}} = \frac{\phi C_f}{C_t}\frac{qt}{LWH}$.

The use of scaling factors defines many of the dimensionless groups in Eqs. (A.2)–(A.17). The sixteen remaining dimensionless numbers are no longer arbitrary and are specific to the described problem. They are:

$$\pi_1 = \frac{\phi c_t}{c_f} \qquad \pi_2 = \frac{\phi\beta_t}{\beta_f}$$

$$\pi_3 = M \qquad \pi_4 = c_f p_1^* = \frac{c_f \mu q L}{k_x WH}$$

$$\pi_5 = \beta_f T_1^* = \beta_f T_{avg} \qquad \pi_6 = \frac{L}{H}\sqrt{\frac{k_z}{k_x}}$$

$$\pi_7 = \frac{L}{H}\tan\alpha \qquad \pi_8 = \frac{L}{H}$$

$$\pi_9 = \frac{k_x \rho_f g\sin(\alpha) WH}{q\mu} \qquad \pi_{10} = \frac{q}{\kappa W}$$

$$\pi_{11} = \frac{\kappa'_{ob}}{\kappa} \qquad \pi_{12} = \frac{\kappa'_{ub}}{\kappa}$$

$$\pi_{13} = c_f p_i \qquad \pi_{14} = \frac{T_{inj}}{T_{avg}}$$

$$\pi_{15} = \frac{\tau\sin(\alpha) L}{T_{avg}}$$

Appendix B. Tables

Tables B1 and B2

Table B1
Parameters and dimensionless numbers of the fifteen reservoirs used in the line drive validation study.

Color	T_{avg}	T_{inj}	Length	Width	Thickness	q	π_7	π_8	π_9	π_{10}	π_{14}	π_{15}
Green	117.7231	41.20308	8370	77.6923	62	67.03419	25	135	11765	20	0.35	1.75
Green	145.2924	50.85235	13365	88.15585	99	76.0959	25	135	11765	20	0.35	1.75
Green	117.6128	41.16448	17145	77.60251	127	66.95646	25	135	11765	20	0.35	1.75
Green	131.2866	45.9503	15120	72.38119	112	62.47178	25	135	11765	20	0.35	1.75
Green	123.2541	43.13895	9720	64.9294	72	56.03065	25	135	11765	20	0.35	1.75
Blue	144.7008	95.79194	9180	132.3458	68	342.7211	30	135	11765	60	0.662	1.75
Blue	115.7777	76.64485	17685	68.68726	131	177.7817	30	135	11765	60	0.662	1.75
Blue	118.6059	78.51708	15930	66.26117	118	171.5188	30	135	11765	60	0.662	1.75
Blue	116.238	76.94953	11070	100.5957	82	260.3737	30	135	11765	60	0.662	1.75
Blue	134.8069	89.24215	18225	75.42822	135	195.3134	30	135	11765	60	0.662	1.75
Red	118.0974	59.75727	7480	145.0222	88	625.6456	25	85	11765	100	0.506	1.75
Red	125.385	63.44482	8500	114.6313	100	494.628	25	85	11765	100	0.506	1.75
Red	141.1314	71.4125	8840	145.9744	104	630.0176	25	85	11765	100	0.506	1.75
Red	117.9554	59.68541	11900	145.9291	140	629.5552	25	85	11765	100	0.506	1.75
Red	114.2298	57.80028	6885	142.9264	81	616.5229	25	85	11765	100	0.506	1.75

Table B2
Parameters used in the Box-Behnken experimental design for the line drive system. Level 0 indicates a normal score transform mean of the parameter; level ± 1 indicates values that are ± 2 standard deviations away from the normal score mean values.

Level	T_{avg} (C)	T_{inj} (C)	τ (C/m)	Length (m)	Thickness (m)	Dip angle	K_z/K_x	Poro.
−1	100.3	30	0.024	2000	33.1	0	0.1	0.14
0	133.7	60	0.0255	3000	103.4	5	0.5	0.22
+1	167.2	90	0.027	4000	173.8	10	1	0.30

Level	Q (m^3/day)	β_r (C^{-1})	β_f (C^{-1})	Cp_r (J/m^3. C)	Cp_{ob} (J/m^3. C)	Cp_f (J/m^3. C)
−1	500	0.5×10^{-5}	6×10^{-4}	1.5×10^6	1.5×10^6	3.9×10^6
0	1250	1×10^{-5}	8×10^{-4}	2×10^6	2×10^6	4×10^6
+1	2000	1.5×10^{-5}	10×10^{-4}	2.5×10^6	2.5×10^6	4.1×10^6

Level	k_r (J/m. day. C)	k_{ob} (J/m. day. C)	k_f (J/m. day. C)	C_r (kpa^{-1})	C_f (kpa^{-1})	Perm. (md)
−1	1.5×10^5	1.5×10^5	5.5×10^4	1×10^{-6}	3.5×10^{-7}	19
0	2×10^5	2×10^5	5.9×10^4	2.5×10^{-7}	4.5×10^{-7}	127
+1	2.5×10^5	2.5×10^5	6.3×10^4	5×10^{-7}	5.5×10^{-7}	235

Appendix C. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.cageo.2017.02.015.

References

Al-Khoury, R., 2011. Computational Modeling of Shallow Geothermal Systems. CRC Press.

Anbar, S., Akin, S., 2011. Development of a linear predictive model for carbon dioxide sequestration in deep saline carbonate aquifers. Comput. Geosci. 37 (11), 1802–1815.

Ansari, E., Hughes, R., 2016. Response surface method for assessing energy production from geopressured geothermal reservoirs. Geothermal Energy 4 (1), 15.

Ansari, E., 2016. Mathematical Scaling and Statistical Modeling of Geopressured Geothermal Reservoirs. Ph.D. thesis.

Ansari, E., Hughes, R., White, C. D., 2014. Well placement optimization for maximum energy recovery from hot saline aquifers. In: 39th Workshop on Geothermal Reservoir Engineering, Stanford University. SGP-TR-202.

Bassiouni, Z., 1980. Evaluation of potential geopressure geothermal test sites in southern Louisiana. Tech. rep., Louisiana State Univ., Baton Rouge (USA). Dept. of Petroleum Engineering.

Box, G.E., Hunter, J.S., Hunter, W.G., 2005. Statistics for Experimenters: Design, Innovation, and Discovery vol. 2. Wiley Online Library.

Ganesh, P.R., Mishra, S., 2014. Reduced physics modeling of CO_2 injectivity. Energy Procedia 63, 3116–3125.

Grant, M., 2013. Geothermal Reservoir Engineering. Access Online via Elsevier.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning vol. 2. Springer.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning. Springer.

Jin, L., Wojtanowicz, A.K., Afonja, G., Li, W., 2010. Scaling analysis of wells with downhole water loop completion for bottomwater control. Journal of Canadian Petroleum Technology 49 (11), 81–90.

Jin, L., 2013. A Feasibility Study of Multi-functional Wells for Water Coning Control and Disposal. Ph.D. thesis, Louisiana State University.

John, C., Maciasz, G., Harder, B., 1998. Gulf Coast geopressured-geothermal program summary report compilation. Volume 2-A: Resource description, program history, wells tested, university and company based research, site restoration. Tech. Rep., Louisiana State Univ., Basin Research Inst., Baton Rouge, LA (United States).

Kalla, S., 2005. Use of orthogonal arrays, quasi-monte carlo sampling, and kriging response models for reservoir simulation with many varying factors. Master's thesis, Louisiana State University.

Kang, C.A., Brandt, A.R., Durlofsky, L.J., 2016. A new carbon capture proxy model for optimizing the design and time-varying operation of a coal-natural gas power station. Int. J. Greenh. Gas Control 48, 234–252.

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer.

McMullan, J.H., Bassiouni, Z., 1984. Rediction of maximum flow rates from geopressured aquifers. J. Petrol. Technol. 36 (3), 503–509.

Mishra, S., Ganesh, P. R., Schuetter, J., He, J., Jin, Z., Durlofsky, L.J., et al., 2015. Developing and validating simplified predictive models for CO_2 geologic sequestration. In: SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Montgomery, D.C., Montgomery, D.C., Montgomery, D.C., 2008. Design and Analysis of Experiments vol. 7. Wiley, New York.

Muffler, P., Cataldi, R., 1978. Methods for regional assessment of geothermal resources.

Geothermics 7 (2), 53–89.
Nathenson, M., 1975. Physical factors determining the fraction of stored energy recoverable from hydrothermal convection systems and conduction-dominated areas. Tech. Rep., Geological Survey, Menlo Park, Calif.(USA).
Novakovic, D., 2002. Numerical reservoir characterization using dimensionless scale numbers with application in upscaling. Ph.D. Thesis, Louisiana State University.
Phillips, O.M., 2009. Geological Fluid Dynamics: Sub-surface Flow and Reactions. Cambridge University Press.
Plaksina, T., White, C., Nunn, J., Gray, T., 2011. Effects of coupled convection and CO_2 injection in simulation of geopressured geothermal reservoirs. In: 36th Workshop on Geothermal Reservoir Engineering, Stanford University. pp. 146–154.
Schuetter, J., Ganesh, P.R., Mooney, D., et al., 2014. Building statistical proxy models for CO_2 geologic sequestration. Energy Procedia 63, 3702–3714.
Shook, M., Li, D., Lake, L.W., 1992. Scaling immiscible flow through permeable media by inspectional analysis. In Situ; (United States) 16(4).
STARS manual, 2011. Advanced Process and Thermal Reservoir Simulator. Computational Modeling Group, Ltd., Calgary, AB, Canada.
Vinsome, P., Westerveld, J., 1980. A simple method for predicting cap and base rock heat losses in'thermal reservoir simulators. J. Can. Petrol. Technol. 19 (3).
White, C.D., 2013. Statistical Reservoir Modeling, PETE 7285 course notes.
Williams, C.F., Reed, M.J., Mariner, R.H., 2008. A review of methods applied by the us geological survey in the assessment of identified geothermal resources. US Geol. Survey Open-File Report 1296, 27.
Wood, D.J., Lake, L.W., Johns, R.T., Nunez, V., et al., 2008. A screening model for CO_2 flooding and storage in Gulf Coast reservoirs based on dimensionless groups. SPE Reserv. Eval. Eng. 11 (03), 513–520.