Geothermics 70 (2017) 62–84

Contents lists available at ScienceDirect

Geothermics

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

Machine learning for creation of generalized lumped parameter tank models of low temperature geothermal reservoir systems

Y. Li[a,b,*], E. Júlíusson[b,c], H. Pálsson[d], H. Stefánsson[b], Á. Valfells[b]

[a] Key Laboratory of Efficient Utilization of Low and Medium Grade Energy, Ministry of Education, Tianjin University, 92 Weijin Road, Nankai District, Tianjin 300072, China
[b] School of Science and Engineering, Reykjavík University, Menntavegi 1, IS-105 Reykjavík, Iceland
[c] Landsvirkjun, Háaleitisbraut 68, IS-103 Reykjavík, Iceland
[d] Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Hjardarhagi 2-6, IS-107 Reykjavik, Iceland

ARTICLE INFO

Keywords:
Geothermal reservoir modeling
Lumped-parameter models
Machine learning
Complexity analysis
Information criterion
Complexity reduction algorithm
Switch-back method
Model selection
K-means clustering algorithm

ABSTRACT

Lumped parameter tank models have gained renewed interest in recent years as an alternative tool for geothermal reservoir analysis and production planning. The models can be structured in various ways regarding the number of tanks, connections between the tanks and the parameters representing the physical properties of the geothermal system. It usually requires a time consuming and difficult process of trials and errors to manually decide the optimal configuration of a tank model. Inspired by recent development in the use of machine learning methods, we propose a method for automatically generating accurate and computationally feasible generalized tank models for isothermal, single phase, reservoirs. This is an extension of earlier work on complexity reduction of generalized tank models (Li et al., 2016). Here, a recursive "switch-back" method is constructed to maximize prediction accuracy of the model. It is also shown how the *K*-means clustering algorithm can be used to aggregate production wells in generalized tank models. One synthetic example and one field application from t Reykir geothermal fields in Iceland are used to illustrate the effectiveness of these methods.

1. Introduction

Geothermal energy is an attractive source for heating and electrical power production by extracting heat from the earth (Erdogdu, 2009). It is a renewable energy resource that, if carefully managed, can be harnessed in a sustainable manner (Axelsson, 2010; SigurÐardottir, 2013; Shortall et al., 2015). To do this it is of vital importance to understand how drawdown in the geothermal reservoir will change under a certain production rate.

One way to understand the behavior of a geothermal field is to describe it as an abstract system shown in Fig. 1, that has inputs and outputs that correspond to production and drawdown respectively. The structure and parameters of the model describing the geothermal field are then chosen so that the calculated response fits the experimental data. Once a model structure has been chosen, a series of measured data (the *training data*) is used to find the parameters of the model. Once the model parameters have been found, a second data series (the *validation data*), independent of the training data, is fed into the model to inspect the predictive capability of the model. If the user is satisfied with the outcome of validation, the model may be used to predict the system response for specific production profiles. Various models have been introduced to represent the real geothermal system such that the errors in fitting the calculated drawdown to both the training data and validation data are within an acceptable range. Generally, a complex model (a model with more parameters) is more likely to lead to a better fit to the training error, while it is not necessary that it will generate a better fit to the validation data, since over-fitting may occur when information is insufficient compared with the complexity of the model (Bishop, 2006; Li et al., 2016).

A common approach to model a geothermal system is to use fully discretized numerical methods in which the geothermal system is divided into tens of thousands of consecutive cells and to introduce mass and energy balance equations to simulate the behavior of the system (Wu and Guan, 2009). Although exhaustive in its nature, these detailed numerical models lead to considerable computational cost. Furthermore, in practical applications, it is always the case that only limited information about the system being modelled is available, which is certainly a disadvantage for using this kind of model, since a successful numerical model is heavily dependent on sufficient amount of data. (Türeyen et al., 2014)

* Corresponding author at: Key Laboratory of Efficient Utilization of Low and Medium Grade Energy, Ministry of Education, Tianjin University, 92 Weijin Road, Nankai District, Tianjin 300072, China.
E-mail address: yuxi15@ru.is (Y. Li).

http://dx.doi.org/10.1016/j.geothermics.2017.05.009
Received 10 February 2017; Received in revised form 19 May 2017; Accepted 22 May 2017

Nomenclature

Greek letters

μ_k	The center for each cluster

Subscripts

∞	The infinite recharge source
i	The i-th tank

Superscripts

$(swch)$	The model after using switch-back method
(m)	The best model among one iteration
(0)	The model at initiating stage
0	The initial estimate in the corresponding model
K	Storage capacity(ms2)
σ	Conductance between tanks(ms)
ρf	Density of the geothermal fluid(kg/m3)
g	Gravitational acceleration(m/s2)
$\dot{m}$	the rate of production(kg/s
T	The total number of tanks
N	The number of production tanks
k	The number of observation tanks
$\vec{h}$	Vector of drawdown
$\overrightarrow{\sigma^T}$	Vector of conductance between tanks and infinite recharge resource
$\overrightarrow{m^T}$	Vector of production rate
$\vec{\theta}$	Vector of parameters
$\overrightarrow{he}$	Vector of measured data
K	Matrix of storage capacity
S	Matrix of conductance
t	Time(s)
Δτ	The time interval between the two data points(s)
$\mathscr{T}$	The transform between input and output
$\mathscr{M}$	The lumped parameter model chosen
M	The number of data points of a data set
xm	The location vector for each well
r_{mk}	The binary indicator variable
KMM	The number of clusters
para	The number of parameters of a model
μk	The center for each cluster
∞	The infinite recharge source
i	The i-th tank
$(swch)$	The model after using switch-back method
(m)	The best model among one iteration
(0)	The model at initiating stage
0	The initial estimate in the corresponding model

Lumped parameter models introduced by Grant et al. (2011), Axelsson (1989), Alkan and Satman (1990), in which the reservoir is represented by tanks and average properties are assigned to these tanks, have gained increasing interest in reservoir modeling (Axelsson 1991; Hjartarson et al., 2002; Thorvaldsson et al., 2010; SigurÐardottir et al., 2010, 2015). Compared with their numerical counterparts, lumped parameter models are much simpler, thus require less computation time for parameter estimation (Türeyen and Akyapı, 2011). Studies show that based on long data sets, this alternative method is quite reliable for liquid dominated geothermal systems and is able to generate a reasonable prediction of drawdown as a function of production (Axelsson et al., 2005). Tank models that only take mass conservation into consideration are referred to as *isothermal lumped parameter models*, while models that consider both mass and energy conservation are referred to as *non-isothermal lumped parameter models* (Onur et al., 2008). With the isothermal flow assumption, the changes in temperature with time cannot be accounted for, while using non-isothermal lumped parameter models, both the drawdown and temperature response can be modelled. Isothermal lumped parameter models are the topic of this work. Fig. 2 shows a typical isothermal lumped parameter system made up of two tanks and an external recharge source. Tank number 1 represents the wells in the geothermal field, the mass flow from this tank is the sum of production from all wells in the field. The drawdown in tank 1 is typically measured from the fluid level in an observation well. Tank number 2 represents the larger field. Each tank has a characteristic capacity, K, which is a measure of how drawdown in the tank changes as a function of net mass flow from it. Between the tanks are links that allow fluid to flow from the higher pressure tank to the lower pressure tank. The links are characterized by a conductance, σ, that is a constant of proportionality between the difference in drawdown of linked tanks and the flow through the link between them. The model in Fig. 2 is an *open system*, in the sense that it is connected to an external recharge source of infinite capacity. A system that is not connected to an external recharge source is *closed*.

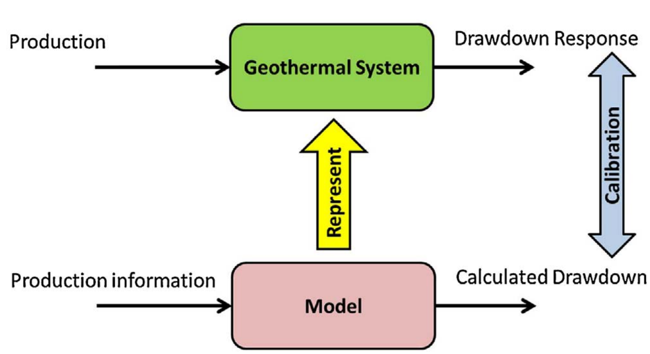

Fig. 1. Schematic graph of reservoir modeling. This figure shows that the state of the geothermal system will change under certain activity. The ultimate goal of reservoir modeling is to make a prediction of future response under certain production conditions. Modeling such a system lies in finding a mathematical representation such that the calculated response coincides with observed response as well as possible future responses to production.

In lumped parameter models, the parameters are determined in a process called history matching, in which an appropriate optimization method is applied to minimize the least-squares error between measured and calculated data series. But before history matching, a proper lumped parameter model has to be chosen. Six standard models have been used in previous works. They are the one tank closed and open models, the two tank closed and open models and the three tank closed and open models (Sarak et al., 2005). In these models the production wells are typically amalgamated into a single tank, while the other tanks represent the area surrounding the wells. Deciding which model to use remains an art in some sense. Tureyen et al. (2014) presented a simple procedure based on the inspection of the RMS and confidence intervals of the estimated parameters for a chosen lumped parameter model to identify the appropriate model among the set of lumped parameter tank models. Statistical confidence intervals prove to be useful for a quantitative evaluation of model discrimination and assessment of uncertainty in the estimated parameters. However, in order to identify the appropriate model, the RMS value and confidence intervals have to be calculated for each model in the model set. This may turn out to be a computationally expensive process. In addition, it may also be time consuming to check whether modeling the reservoir as a