PROCEEDINGS, 42nd Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 13-15, 2017
SGP-TR-212

Data Fusion for Hydrothermal Reservoir Characterization through use of Bayesian Statistical Inference and MCMC Maximum Likelihood Models

Cari D. Covell[1], Ágúst Valfells[1], María S. Guðjónsdóttir[1], Hlynur Stefánsson[1], Egill Júlíusson[2], Halldór Pálsson[3], Birgir Hrafnkelsson[3]

[1]Reykjavík University, Menntavegur 1, 101 Reykjavík, Iceland

[2]Landsvirkjun, Háaleitisbraut 68, 103 Reykjavík, Iceland

[3]University of Iceland, Sæmundargata 2, 101 Reykjavík, Iceland

cari14@ru.is

Keywords: data fusion, hydrothermal reservoir characterization, Bayesian statistical inference, Markov Chain Monte Carlo (MCMC), maximum likelihood.

ABSTRACT

Geothermal operators model reservoirs using several sets of data that come from various sources and are often related. Many different models of the reservoir can fit some subset of the available data. The challenge at hand is to find those models that best fit all or most of the available data. A new data fusion and inversion methodology is introduced for implementation into a proposed modelling tool for hydrothermal systems. Algorithms are expected to be developed for joint inversion Bayesian inference of data from geothermal exploration, and for model likelihoods using parallel tempering (PT) Markov Chain Monte Carlo (MCMC) to characterize Icelandic hydrothermal systems. The project builds on a methodology from the National Information and Communications Technology of Australia (NICTA, current name Data61) where a similar approach was developed for Enhanced Geothermal Systems (EGS). Major differences in analyses of hydrothermal systems from EGS systems include permeability response and fluid flow through a naturally made fracture network. The aim for using Bayesian inference and MCMC likelihood methods is to interpret probabilistic models for quantification of uncertainty in low and high temperature hydrothermal systems. The purpose of this research is to develop a tool that can improve the accuracy of reservoir models that are vital to reducing risk of geothermal projects, and to support better decision making for exploration and development of the hydrothermal reservoirs.

1. INTRODUCTION

Geothermal exploration is based on collection of large amounts of data from geophysical, geochemical, and geological surveys which are used as input for constructing models of the geothermal reservoir. Model construction of the geothermal reservoir based on measurements is however a nontrivial task. Reservoir properties must be inferred from the measured data to obtain observed values and compare to measured results. The methodology proposed intends to combine geophysical and geological surveys through a data fusion inversion process to create conceptual and numerical reservoir models.

The background for the methodology that will be developed streamlines data fusion and inversion from a variety of geological and geophysical inputs from multiple exploration surveys. Candidate models will be used to simulate results of geothermal exploration and operation. The Bayesian inference approach will be used to invert the data and provide probabilistic distributions for the most likely numerical reservoir models created using a parallel tempering (PT) Markov Chain Monte Carlo (MCMC) approach. The simulated values will then be compared to measured values from the field, where uncertainties in each generated model will be quantified and ranked in confidence intervals. The methodology is implemented for Icelandic hydrothermal systems, due to availability of survey data.

The modelling tool specific to hydrothermal systems will address several research questions that include the following:

- Can a fully integrated digital tool be created to handle the variety of geophysical and geological inputs?
- How confidently can reservoir models be constructed to characterize the most important features of geothermal reservoirs (location of fracture networks, hot rock, etc.)?
- How does the dynamic nature of a hydrothermal system affect the applicability of existing data when compared to present conditions?
- What risks can be evaluated in development of geothermal fields?
- Can the value of exploration data be quantified?

2. BACKGROUND

The geophysical inverse problem is designed to infer subsurface properties given measurements acquired in geothermal exploration. Inversion problems in geophysics are typically high dimensional, in that there are an infinite number of plausible geological/geophysical models to consider, and under constrained, in that many different models may produce the same measurements on the surface (Li and Oldenburg 1998). Geophysical data comes in several forms of magnetic, electromagnetic, gravity, and seismic surveys. The most common collection methods include Direct Current (DC), Magnetotellurics (MT), and Transient Electromagnetic Method (TEM). However, a single type of sensor may not represent the entire dimensionality of the problem. For example, gravity sensors alone provide poor depth resolution but can be very useful when fused with complementary sources of information such as seismic surveys (McCalman et. al. 2014). Additionally, an important part of geophysical inversion is that a single most likely solution will fail to account for uncertainty associated with the result (Kearey et. al. 2002). The Bayesian inversion approach addresses the uncertainty and allows for multi-modal information to constrain the solutions (McCalman et. al. 2014).

In Iceland, recent work in geophysical inversions have been performed for 1-D surface measurements using joint MT and TEM data, where TEM is used to correct static shift. Inversions have been performed for 2-D and 3-D surface and subsurface measurements only using MT data. Examples include the Hengill area (Árnason et. al. 2010), Krýsuvík area of Reykjanes Peninsula (Irabaruta 2011), area around Eyjafjallajökull glacier (Barkaou 2011), Þeistareykir Geothermal Area (Karlsdóttir et. al. 2012), and Námafjall area (Uddin 2012). All joint inversion methods in the examples were done using conventional algorithms for one, two, or three candidate models, therefore quantification of uncertainty in high dimensional space is not part of the inversion process.

The problem under consideration for hydrothermal systems is a nonlinear geophysical inversion problem which involves sampling from multi-modal density functions. The problem is characterized by the marginal likelihoods from the Bayesian inference and typically leads to diverse but good solutions which fulfill the conditions given by the data (McCalman et. al. 2014). In the nonlinear case, the probability density function (pdf) through Bayesian inference involves difficult computation of the model space (Malinverno 2002). Therefore, Markov Chain Monte Carlo (MCMC) algorithms are used to efficiently evaluate integrals in high dimensional space (Malinverno 2002).

3. METHODS

A modelling tool is proposed to combine geophysical and geological surveys through a data fusion and joint inversion process in order to create reservoir models and assess geothermal development risks and decision making. Figure 1 shows a flowchart describing the proposed methodology. The steps include: (i) creating candidate geological models, (ii) fusing geophysical data to feed into the forward model through a Bayesian statistical approach, (iii) finding model likelihoods through Markov Chain Monte Carlo (MCMC) processes, and (iv) characterizing hydrothermal systems to validate field measurements and allow risk assessment for better decision making. These steps are explained in more detail in the following subsections.

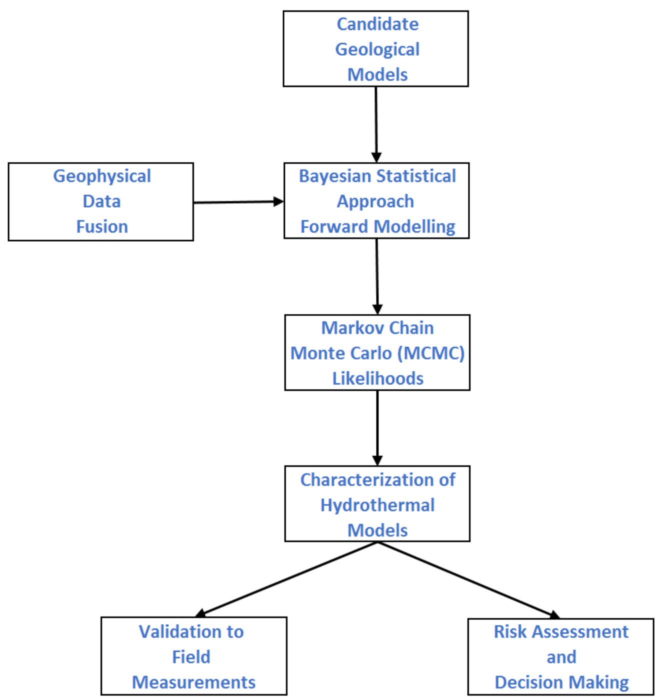

Figure 1: Flow chart representing the proposed methodology of Bayesian inversion and MCMC likelihood models for hydrothermal reservoir characterization.

3.1 Candidate Geological Models

Candidate geological models are created to determine the location and properties of geobodies. In Iceland, geological field exploration includes identifying basalt masses, magmatic intrusions, and alteration layers. Once location is identified, the model is populated with rock properties, i.e. porosity, permeability, and density, taken from the IMAGE project database (Reinsch et. al. 2016). Conventionally, one data set of rock properties is used to populate one candidate geological model. The initial step of creating candidate geological models will address uncertainties within conventional assumptions by fusing multiple datasets within the IMAGE database. Therefore, a few immensely different geological models (ex. 5-10), and subsequent stochastic versions (ex. 1000) are generated and used as candidate models in the proceeding forward modelling algorithms.

3.2 Bayesian Statistical Approach Forward Modelling

Forward modelling is performed on the candidate models and compared to measured geophysical data. The fit to the data is then used to infer e.g. the resistivity values of the rock, density of the reservoir rock, and fluid that influences the gravitational field. Conventionally, inversions are performed separately and only one candidate model is chosen for simplicity.

A critical property of a geophysical inversion is that a single 'most likely' answer will fail to quantify the uncertainty associated with results (Kearey et. al. 2002). The Bayesian approach directly assesses uncertainty and allows observations to be fused from different types of sensors (McCalman et. al. 2014). Bayesian inference is the posterior density function (pdf) of the unknown parameters defined by a likelihood function and prior distribution (Bishop 2006). The general form of Bayes' rule is:

$$P(g|D) = \frac{P(D|g)P(g)}{\int P(D|g)P(g)dg}$$

where $D = D_k$, $k = 1 \ldots K$ is the set of observed quantities from K distinct sensors that respond to J spatially distributed properties on g model functions of spatial position x in the form $g(x) = [g_1(x), \ldots, g_J(x)]$ (McCalman et. al. 2014). For inversions, the prior and likelihood terms of the Bayes' Rule are split into four modelling components: (i) a prior belief of the geological structures (ii) a joint prior over rock properties (iii) geophysical forward simulations, and (iv) probabilistic likelihood models. Table 1 shows each modelling component and the associated notation used to perform the Bayesian inversion. Figure 2 shows a graphical representation of the Bayesian inversion algorithm.

Table 1: Modelling Component and Associated Notation for Bayesian Inversion (derived from McCalman et. al. 2014).

<table>
<tr><th>Modelling Component</th><th colspan="2">Associated Notation</th></tr>
<tr><td rowspan="2">(i) Prior belief of geological structures</td><td>$P(\alpha)$</td><td>$P(c_l)$</td></tr>
<tr><td>Structural prior of underlying 3D volume</td><td>Distribution of rock types given layers $l = 1 \ldots L$</td></tr>
<tr><td rowspan="2">(ii) Joint prior over rock properties</td><td>$P(g)$</td><td>$P(\rho_l)$</td></tr>
<tr><td>Previously defined candidate geological model</td><td>Defined for each layer (rock type) $\rho_l = \rho_1 \ldots \rho_L$</td></tr>
<tr><td rowspan="2">(iii) Geophysical forward simulations</td><td colspan="2">D_k</td></tr>
<tr><td colspan="2">An ideal observation for the forward models f_k for the k^{th} sensor type mapping geology g defined as $f_k(g)$</td></tr>
<tr><td rowspan="2">(iv) Probablistic likelihood models</td><td colspan="2">$P(\alpha, c, \rho, D|\mu_\alpha, \Sigma_\alpha, \{\mu_r, \Sigma_r\}, \{\theta_l\},\{a_k, b_k\}) =$
$P(\alpha|\mu_\alpha, \Sigma_\alpha)P(c|\{\theta_l\})\Pi^L_{l=1}P(\rho_l|\{\mu_r, \Sigma_r\})\Pi^K_{k=1}P(D_k|g, a_k, b_k)$</td></tr>
<tr><td colspan="2">The joint inversion notation of Bayes' Rule, where μ and Σ are the mean and covariance, θ is a hyperparameter of the forward model, and a_k and b_k are the shape and scale parameters, respecivly.</td></tr>
</table>

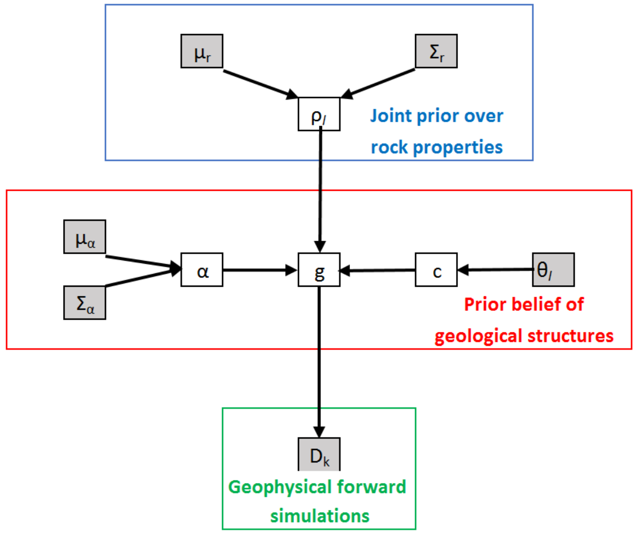

Figure 2: Graphical model illustration of the probabilistic fusion procedure implemented using Bayes' Rule.

3.3 Markov Chain Monte Carlo (MCMC) Likelihoods

MCMC methods are based on Monte Carlo sampling with a random walk process based on a probability distribution represented by a Markov chain and hence the current state is the only relevant information for the choice of the next state (Malinverno 2002). A standard technique to sample multiple-dimensional distributions is the Metropolis Hastings (MH) algorithm. MH aids in the difficult calculation of the normalization factor by using a proportion between the function to the desired probability distribution, and either accepts the candidate if $x \geq 1$ or rejects the candidate if $x < 1$ (McCalman et. al. 2014).

However, the trade-off between a sufficiently broad distribution to bridge the regions of low probabilities between modes and a tolerable acceptance rate of proposed states is difficult to achieve with standard MCMC techniques (McCalman et. al. 2014) (Brooks et. al. 2011). Hence, a meta-algorithm known as parallel tempering (PT) can be used to aid with convergence by utilizing multiple coupled MH chains. PT runs multiple MH chains that are updated using MCMC simultaneously at different sampling energies (McCalman et. al. 2014). The acceptance probability function is modified to provide a mechanism for chains at higher energies to more easily move between modes, while low energy chains enable more precise local sampling (McCalman et. al. 2014). PT also incorporates an interchange procedure whereby two chains at different energies can swap location in the state space in order to satisfy balanced constraint of MCMC (McCalman et. al. 2014).

Convergence is monitored using the potential scale reduction factor, where multiple chains at identical energy levels run until measurements approach unity (McCalman et. al. 2014). A different reduction factor is required for each dimension of the state space. Once the chains are deemed to have converged, the samples can be combined to create a long single chain from the stationary distribution, and the chain is accepted or rejected until there is recovery of true posterior samples (McCalman et. al. 2014).

3.4 Characterization of Hydrothermal Models for Validation and Risk Assessment

The hydrothermal reservoir models are created after running MCMC to show the probable distribution of confidence for the computed values from Bayesian joint inversion. The computed values would include reservoir characteristics such as temperature and pressure. In addition, the likelihood models show information about faults, fracture geometry and orientation, and permeable fluid flow paths. For a particular case study, the models are validated by comparing computed values to measured values after field development, and are analyzed for modelling error. The risk assessment is then based on the sensitivity of the parameters obtained from geological and geophysical exploration data.

The emphasis of the methodology is based on previous data fusion work for Enhanced Geothermal Systems (EGS) by the National Information and Communications Technology of Australia (NICTA) in 2014. NICTA has developed an open source software called Obsidian, which performs data fusion inversion on geological and geophysical data to provide Bayesian probabilistic outputs using MCMC likelihood (Beardsmore 2014). Key objectives of initializing Obsidian involved validating its predictions with conventional modelling methods using exploration data, and to infer relevant geological properties through reservoir characterization prior to drilling and development.

4. DISCUSSION

4.1 Unique attributes for the hydrothermal case

The methodology will address several gaps in knowledge about hydrothermal systems. Optimization of the Bayesian inference will include multiple hyperparameters in the data, meaning some data may not consist of a single value to describe one point in space. The algorithms lead to solving a non-linear high-dimensional problem, which can create more complicated models yet very detailed and realistic analyses. Probabilistic distributions of the parameters in the hydrothermal reservoir models will be quantified to show measurement error, therefore leading to a potential need to explore alternative data collection methods for minimizing the errors. Once the technical challenges of fusing geophysical data are overcome, knowledge about fracture networks and earthquake susceptible areas from fissure swarms will be imperative for understanding fluid flow paths.

The differences from NICTA's EGS examples lie in the nature of reservoir characterization. Hydrothermal systems call attention to conductive fluid behavior, phase change, geochemical properties, more dynamic pressure drops and boiling, and permeability response. The probable distributions associated with drilling risk will also be a separate issue for analysis, whereby the models provide an interface for well optimization when developing a geothermal area.

4.2 Market Value for Modelling Tool

Today, there are 13.8 GW of installed power worldwide from geothermal energy (Rocco 2016). Figure 4 shows that installed power has increased substantially over the last several years and is directly correlated with energy produced. Prior to utilization for power generation, geothermal reservoirs need to be explored and characterized to predict response to production. The models are used to locate the wells which are then drilled. During power generation, new wells may also need to be drilled to compensate others. The geothermal industry worldwide could benefit from the results of a modelling tool in the form of a software, in terms of quantifying risk reduction.

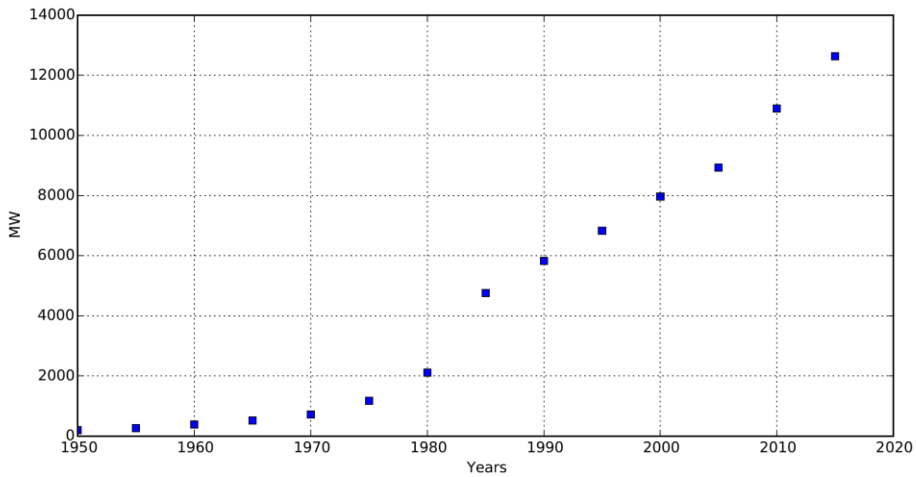

Figure 4: Worldwide installed geothermal capacity from 1950 up to 2015.

4.3 Challenges of Model Development and Implementation

The design challenges are associated with the data fusion and model characterization. Measured data typically involves some or all aspects of missing data, skewed data, and outliers. The data is also stored in different file formats and/or not compatible with today's technology, and must be converted in time consuming processes. Additionally, analyzing characteristics of a reservoir in more than 3 dimensions will be difficult considering the many factors in play when looking at a hydrothermal system.

The technical challenges are associated with developing and running the software. Identifying, securing, and financing an appropriate cluster computing facility takes time to acquire. Additionally, there is a need to find ample capacity to store data in one location. Selection of efficient computational algorithms will also be necessary to run Bayesian inversion and MCMC faster, given the large amounts of information as input.

5. ACKNOWLEDGEMENTS

The proposed methodology and modelling tool is work to be carried out under the Operations Research in Subsurface Modelling (ORSM) research group consisting of the authors listed in this paper from Reykjavik University, Landsvirkjun, and the University of Iceland. This research is supported through collaboration with Knútur Árnason of Iceland GeoSurvey (ISOR) on geophysical data collection and inversion, Reykjavik Energy on exploration data collection, and Landsvirkjun on assistance with model development. Graeme Beardsmore, Technical Director of Hot Dry Rocks Pty Ltd. is also acknowledged to assist on the Bayesian inference and MCMC likelihood algorithms.

REFERENCES

Árnason, Knútur, Hjálmar Eysteinsson, and Gylfi Páll Hersir. "Joint 1D inversion of TEM and MT data and 3D inversion of MT data in the Hengill area, SW Iceland." Geothermics 39.1 (2010): 13-34.

Barkaou, A-E. “Joint 1D inversion of TEM and MT resistivity data with an example from the area around the Eyjafjallajökull glacier, S Iceland” UNU-GTP Training Programme, Reports 2011 Number 9. (2011).

Beardsmore, Graeme, "Data Fusion and Machine Learning for Geothermal Target Exploration and Characterisation." (2014).

Bertani, Ruggero. “Geothermal Power Generation in the World 2010-2014 Update Report.” Proceedings World Geothermal Congress 2015. (2015).

Bishop, Christopher M. "Pattern recognition." Machine Learning 128 (2006).

Brooks, Steve, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng, eds. Handbook of Markov Chain Monte Carlo. Chapman and Hall/CRC, (2011).

Irabaruta, Constantin. "Joint 1-D inversion of TEM and MT resistivity data, comparison with mineral alteration and temperature in drillholes-case study: Krýsuvík area, SW-Iceland." (2011).

Karlsdóttir, Ragna, et al. "Þeistareykir geothermal area, Northern Iceland: 3D inversion of MT and TEM data." (2012).

Kearey, P., M. Brooks, and I. Hill. "An Introduction to Geophysical Exploration.,(Blackwell Science: Oxford)." (2002).

Li, Yaoguo, and Douglas W. Oldenburg. "3-D inversion of magnetic data." Geophysics 61.2 (1996): 394-408.

Li, Yaoguo, and Douglas W. Oldenburg. "3-D inversion of gravity data." Geophysics 63.1 (1998): 109-119.

Malinverno, Alberto. "Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem." Geophysical Journal International 151, no. 3 (2002): 675-688.

McCalman, Lachlan, Simon T. O’Callaghan, Alistair Reid, Darren Shen, Simon Carter, Lars Krieger, Graeme Beardsmore, Edwin V. Bonilla, and Fabio T. Ramos. "Distributed Bayesian Geophysical Inversions." In Thirty-Ninth Stanford Geothermal Workshop, pp. 1-11. (2014).

Reinsch, T. et al. “IMAGE Integrated Methods for Advanced Geothermal Exploration”. IMAGE-D3.03-2016.11.04 (2016).

Rocco, Anthony. “2016 Geothermal Power: International Market Update”. Geothermal Energy Association. (2016).

Uddin, Mohammad Zohir. "1D joint inversion of TEM and MT resistivity data, with an application of soundings from the Námafjall high-temperature geothermal area, NE-Iceland." UNU-GTP Training Programme, Reports 2012 Number 35. (2012).