G Model
GEOT-1339; No. of Pages 13

ARTICLE IN PRESS

Geothermics xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Geothermics

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

Play fairway analysis of geothermal resources across the state of Hawaii: 2. Resource probability mapping

Garrett Ito[a,*], Neil Frazer[a], Nicole Lautze[b], Donald Thomas[b], Nicholas Hinz[c], David Waller[a], Robert Whittier[d], Erin Wallin[b]

[a] Department of Geology and Geophysics, SOEST, University of Hawaii, Manoa, Honolulu, HI 96822, USA
[b] Hawaiian Institute for Geophysics and Planetology, SOEST, University of Hawaii, Manoa, Honolulu, HI 96822, USA
[c] Nevada Bureau of Mines and Geology, Mackay School of Earth Sciences and Engineering, University of Nevada, Reno, USA
[d] Safe Drinking Water Branch, Hawaii State Department of Health, USA

ARTICLE INFO

Article history:
Received 28 May 2016
Received in revised form 7 October 2016
Accepted 4 November 2016
Available online xxx

Keywords:
Volcanic islands
Resource prospecting
Data analysis
Geology
Geophysics
Geochemistry
Probability and statistics

ABSTRACT

We develop a new geostatistical method to combine evidence provided by diverse geological data sets and produce maps of geothermal resource probability. The application is to the State of Hawaii, and the data sets include the locations and ages of mapped volcanic centers, gravity and magnetotelluric measurements, groundwater temperature and geochemistry, ground surface deformation, seismicity, water table elevation, and groundwater recharge. Using the basic principles of Bayesian statistics, these data and expert knowledge about the effects and importance of the data are used to compute the probabilities of the primary resource qualities of elevated subsurface heat, reservoir permeability, and reservoir fluid content. The product of these marginal probabilities estimates the joint probability of all three qualities and hence the probability of a successful geothermal prospect at each map point. An analogous set of algorithms is used to quantify the confidence in the probability at each point. Not surprisingly, we find that successful geothermal prospects are most probable on the active volcanoes of Hawaii Island, including the area of Hawaii's single geothermal energy plant. Probability decreases primarily with shield volcano age, being relatively moderate in select locations on Maui and Lanai, relatively low on Oahu, and minimal on Kauai. Future exploration efforts should consider these results as well as the practical, societal, and economic conditions that influence development viability. The difficulties of interisland power transmission mean that even areas with moderate to low probabilities are worth investigating on islands with population centers.

1. Introduction

We conducted an assessment of geothermal resource potential across the state of Hawaii, updating the last assessment which was done three decades ago (Thomas, 1985). The overall goal is to identify the *plays*, or probable areas for geothermal energy development in the *fairway*, of the Hawaiian volcanic island chain. The first of three manuscripts (Lautze et al., 2016a) summarizes the geologic conditions that support geothermal resources in Hawaii and the datasets selected to provide evidence for these conditions. The third paper (Lautze et al., 2016b) describes the essential practical and economic criteria needed to assess *development viability* and, with the results of the geologic considerations presented in this paper, recommends a prioritized list of sites for future exploration. This manuscript—the second paper in the series—describes our methods and the results of processing the various geoscientific datasets into probabilities of geothermal resources across the state.

Methods used to map the spatial distributions of geothermal resource potential can be categorized as knowledge-driven or data-driven (Bonham-Carter, 1994). Knowledge-driven, or deterministic, models rely on the judgment of experts to assign the relative importance of different data types to resource potential. These methods are needed especially in the reconnaissance phase of exploration when few or no resources have been found (e.g. Prol-Ledesma, 2000). Techniques of combining the evidence provided by the data include Boolean operators (Noorollahi et al., 2008; Yousefi et al., 2010), index quantification and weighting (Noorollahi et al., 2008; Tüfekçi et al., 2010; Trumpy et al., 2015), and quantification with *fuzzy* or continuous functions (Prol-Ledesma, 2000; Siler et al., 2016). In contrast, data-driven models incorporate *data* on

* Corresponding author.
E-mail address: gito@hawaii.edu (G. Ito).

http://dx.doi.org/10.1016/j.geothermics.2016.11.004
0375-6505/

Please cite this article in press as: Ito, G., et al., Play fairway analysis of geothermal resources across the state of Hawaii: 2. Resource probability mapping. Geothermics (2016), http://dx.doi.org/10.1016/j.geothermics.2016.11.004

known resource locations or training sites to relate observational evidence to resource potential. These methods use statistical techniques including weight-of-evidence (Bonham-Carter et al., 1989; Coolbaugh and Bedell, 2006; Coolbaugh et al., 2007), logistic regression (Coolbaugh et al., 2002, 2005), and evidence belief functions (Carranza and Hale, 2003). Data-driven methods have been used even more extensively in mineral resource exploration (e.g., Porwal and Kreuzer, 2010). In this context, the method developed here is knowledge-driven, uses continuous quantities for the influence of different data types, but like some of the data-driven techniques, the core algorithm is based on the principles of Bayesian statistics. Unlike earlier methods used, that produced measures of resource "favorability", our method predicts relative probabilities.

A successful geothermal prospect must have all of three primary qualities: elevated heat (H), elevated permeability (P), and adequate fluid (F) to deliver the heat. Table 1 lists the data types used for Hawaii as indicators of each of quality, and summarizes the evidence each data type provides as discussed in detail by Lautze et al. (2016a). Section 2 of this paper reviews the theory of our method, our approach to eliciting expert knowledge, and the algorithm by which this knowledge and the data are combined to compute probabilities. Section 3 details the specific parameters and functions used for each data type and their individual effects on the marginal probabilities of the three resource qualities. Section 4 presents the resulting resource probabilities and their associated confidence measures for the main Hawaiian Islands. Finally, we close with a discussion of the strengths and weaknesses of our method, and the role our results could play in Hawaii's exploratory decision-making process.

2. Methods of data processing and modeling probability and confidence

2.1. Overview

The first building block of our method is a generalized linear model (e.g., MCullah and Nelder, 1983) in which the evidence provided by each data type is weighted and summed in the logistic link function (e.g., Bonham-Carter et al., 1989),

$$Pr(\mathbf{x}) = \left[1 + \exp\left(-w_0 - \sum_{i=1}^{m} w_i z_i(\mathbf{x})\right)\right]^{-1}. \quad (1)$$

Here $\Pr(\mathbf{x})$ is the probability of just one of the resource qualities (elevated heat H, permeability P, fluid F) at location $\mathbf{x}$ on the map. A similar equation is used for each of the two other qualities. In this equation $z_i(\mathbf{x})$ is a transformed and scaled (explained in Sections 2.2 and 2.3 below), real-number, form of data type i; w_i is a weight that reflects the relative importance of data type i to the quality of interest; and m is the number of data types present at location $\mathbf{x}$. This equation implicitly includes a reference probability, or prior probability $\Pr_0$, represented on the right-hand side by the quantity w_0. We refer to Eq. (1) as the "voter equation" because it allows each data type to influence the outcome (positively or negatively) depending on its weight w_i.

The general behavior of the voter equation can be understood through a qualitative discussion. Suppose z_1 is a quantity representing the gravity anomaly at location $\mathbf{x}$, and z_2 represents a measure of electrical resistivity beneath the ground at $\mathbf{x}$. Because high positive values of gravity are interpreted as indicating dense intrusive source rock (and z_1 is positive when the gravity anomaly is relatively high), the associated weight w_1 will be positive. In contrast, unusually low resistivity (indicated by a negative value of z_2) is associated with hot rock and therefore w_2 will be negative. Thus, a large positive value of the sum $\Sigma = w_0 + w_1 z_1 + w_2 z_2$ indicates a high favorability of elevated heat. Clearly as more data types contribute positively to the sum, the sum increases monotonically. However, if there are five strong positive data contributions of elevated heat from five different data types for example, then adding a sixth positive contribution does not provide much new information. This aspect is taken into account with the logistic link, or *expit* function, $Pr = \exp\mathrm{it}(\Sigma) = e^{\Sigma}/[1 + e^{\Sigma}] = [1 + e^{-\Sigma}]^{-1}$ (Eq. (1)), which spans 0–1 as does a true probability. In another location the sum $\sum$ could be large and negative, in which case the probability of heat will be small. In yet another location where there are no data, the data votes will be zero, but the probability will not be; it will equal the prior probability $Pr_0 = \mathrm{expit}(w_0) = [1 + e^{-w_0}]^{-1}$. The probabilities of elevated permeability and fluid are computed in the same way.

Using the marginal probabilities of all three resource qualities ($\Pr_H$, $\Pr_P$, $\Pr_F$), we approximate the probability of a viable resource $\Pr_R$ by the product of the marginals,

$$\Pr_R(\mathbf{x}) = \Pr_H(\mathbf{x})\Pr_P(\mathbf{x})\Pr_F(\mathbf{x}). \quad (2)$$

This equation is the second building block of our method; like Eq. (1), it is based on a conditional independence assumption that has a long record of surprising robustness in Bayesian learning (e.g., Domingos and Pazzani, 1997; Porwal et al., 2006). We refer to Eq. (2) as the "veto equation" because if any one quality has a low probability, so will the probability of a viable resource. The output probabilities are evaluated at each 200 m × 200 m cell of the model grid, the centers of which define $\mathbf{x}$. The calculations were performed primarily and displayed entirely using Generic Mapping Tools (GMT) (Wessel et al., 2013). Some of the calculations, prior to visualization, were done using Matlab® (www.mathworks.com).

2.2. Specifics: expert elicitation and defining weights (w_i)

The voter Eq. (1) requires converting the starting data value D_i to its processed form z_i, and relating the importance of the data, quantified by its weight w_i, to the probability of a given resource quality. In this knowledge-driven, reconnaissance application, we use *expert elicitation* (e.g., O'Hagan et al., 2006; O'Leary et al., 2009). As such, the prospecting algorithm incorporates the expertise of our research team, and is thus able to "think" like an expert with years of experience. To understand how we do this, consider first the baseline probability value $\Pr_0$ for a given resource quality (H, P, or F). We ask the expert for the probability of that quality at an unknown location. The expert knows only that the location is in Hawaii, and is free to solve the question in any way he or she wishes. We then set the expert's estimated probability $\Pr_0$ equal to $\mathrm{expit}(w_0)$ and solve for w_0, using the inverse function,

$$w_0 = \mathrm{logit}(Pr_0) \equiv \ln\left(Pr_0/(1 - Pr_0)\right). \quad (3)$$

To incorporate input from multiple experts, we weight by years of experience and take the weighted average of their respective values of w_0.

Now consider how to elicit the effects of the first data type D_1 (e.g., gravity) on the probability of a resource quality, for example heat $\Pr_H$. We seek to define z_1 and w_1 so that with only that data type appearing in the sum of the voter Eq. (1), the resulting values of $\Pr_H$ at one or two values of D_1 are consistent with the expert's intuition. (I) First, we give each expert in our team a particularly promising data value in either its starting D_1^+ or processed z_1^+ form (whichever is more intuitive to the expert), and ask them to estimate the corresponding probability Pr_H^+. (II) Second, we then ask the expert to estimate the value, D_1^- for which the data has no effect on probability. Question (I) is used to establish the *location* property—i.e., what $\Pr_H$ is at a given D_1 (or z_1) —for the dependence of $\Pr_H$ on D_1 alone. With the answer to question (I), question