Research article

Geoenergy

https://doi.org/10.1144/geoenergy2023-010 | Vol. 1 | 2023 | geoenergy2023-010

# Temperature uncertainty modelling with proxy structural data as geostatistical constraints for well siting: an example applied to Granite Springs Valley, NV, USA

**Whitney Trainor-Guitton[1]*, Drew Siler[2] and Bridget Ayling[3]**

[1] National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden CO 80401, USA
[2] U.S. Geological Survey, Geology, Minerals, Energy, and Geophysics Science Center, Moffett Field, California 94035, USA
[3] Great Basin Center for Geothermal Energy, Nevada Bureau of Mines and Geology, University of Nevada-Reno, Reno, NV
WT, 0000-0002-5726-3886
* Correspondence: Whitney.TrainorGuitton@nrel.gov
Present addresses: DS, Geologica, Reno, NV, USA; BA, Contact Energy Ltd, New Zealand

**Abstract:** Utilizing existing temperature and structural geology information around Granite Springs Valley, Nevada, we build 3D stochastic temperature models with the aims of evaluating the 3D uncertainty of temperature and choosing between candidate exploration well locations. The data used to support the modelling are measured temperatures and structural proxies from 3D geologic modelling (distance to fault, distance to fault intersections and terminations, Coulomb stress change and dilation tendency), the latter considered 'secondary' data. Two stochastic geostatistical techniques are explored for incorporating the structural proxies: cosimulation and local varying mean.

With both the cosimulation and local varying mean methods, many equally-likely temperature models (i.e. realizations) are produced, from which temperature probability profiles are calculated at candidate well locations. To aid in choosing between the candidate locations, two quantities summarize the temperature probabilities: $V_{\text{prior}}$ and entropy. $V_{\text{prior}}$ quantifies the likelihood for economic temperatures at each candidate location, whereas entropy identifies where new information has the most potential to reduce uncertainty.

In general, the cosimulation realizations have smoother spatial structure, and extrapolate high temperatures at candidate locations that are located along the direction of the longest spatial correlation, which are down dip from existing temperature logs. The smooth realizations result in tight temperature probability profiles that are easier to interpret, but they have unrealistic temperature reversals in some locations because of the dipping ellipsoid shape created and that the cosimulation technique does not enforce a conductive geothermal gradient as a baseline (i.e. linearly increasing temperature with depth). The local varying mean results produce realizations with more realistic geothermal gradients, with temperatures increasing downward since a depth-temperature relationship is included. However, because they have much noisier spatial nature compared to cosimulation, it is harder to interpret the temperature probability profiles. The different local varying mean results allow the geologist to determine which proxy (e.g. dilation v. distance to fault termination) should be used given the specific geothermal system. In general, $V_{\text{prior}}$ from local varying mean results identify locations that are close to high values for the structural proxies: areas with higher probabilities for higher temperatures. The entropy results identify where uncertainty is greatest and therefore new drilling information could be most useful. Though these techniques provide useful information, even when applied to areas of sparse data, our comparison of these two techniques demonstrates the need for new geothermal geostatistics techniques that combine the advantages of these two methods and that are tailored to the spatial uncertainty issues inherent in geothermal exploration.

**Received** 15 March 2023; **revised** 3 November 2023; **accepted** 9 November 2023

Prospecting for hidden geothermal resources is challenging. In areas where no surficial evidence indicates that heat, fluids, and permeability may coexist at depths that can be economically drilled, many inferences must be made. Geothermal play fairway analysis has provided a means for combining diverse regional data types and geologic observations into likelihoods of geothermal resources at the regional scale, allowing for discovery of potential and previously unknown hidden geothermal resources. Generally, play fairway analysis has been very successful in identifying hidden prospects via 2D regional analysis (Faulds *et al.* 2015*b*, 2017; Craig *et al.* 2021).

After a geothermal prospect is identified, the next stage is discovery, where the goal is to delineate how hot and large the resource is by refining the conceptual geothermal model; ultimately, these models are then used to determine the electric power capacity (Cumming 2016*a*, *b*; Ayling and Hinz 2020; Craig *et al.* 2021). This stage includes well-siting decisions to prove the resource and/or refine the hypotheses around the system (e.g. reservoir geometry, hydrothermal upflow/outflow pathways, and spatial distribution of subsurface temperatures). Determining where to collect deeper temperature observations, such as drilling thermal gradient holes (TGH), can be challenging: TGHs drilled at finite locations may confirm that economic temperatures exist *and/or* provide evidence to either prove or disprove current conceptual models of the geologic structure or geothermal processes. Herein lies an important theme of the work presented: temperature measurements provide a direct measure of heat albeit at finite 3D locations, whereas geologic modelling can give spatially extensive estimates of proxy measurements for temperature.

Various geological and geophysical datasets are often used as proxies for one or more of the three components needed for conventional geothermal systems: fluids, permeability, and heat. Fault mapping can reveal possible permeability pathways for deeper, hotter fluids to travel towards the surface (Faulds *et al.* 2010, 2015*a*). Three-dimensional mapping of these geologic features brings more geologic realism to the representation and

understanding of the subsurface geologic complexity and its control on permeability (Witter *et al.* 2019; Siler and Pepin 2021).

Geothermal conceptual modelling has not historically employed 3D stochastic modelling, which would provide robust probabilistic results to the subsurface temperature distribution, while also adhering to spatial correlations observed in the data. Whereas earth scientists in mineral and petroleum applications have utilized geostatistics to generate spatial models quantitatively and probabilistically, geostatistics has a long history of spatial interpolation of sparsely sampled earth science data (Deutsch and Journel 1998). Geostatistical interpolation often uses the variogram, which (1) quantifies the spatial correlation of the property of interest (e.g. porosity, permeability, temperature) and (2) can be used to inform spatial interpolation, known as Kriging. (Williams and DeAngelo 2011), one of the first published studies using Kriging for geothermal applications, compared 2D Kriging to spline estimations for both heat, flow and temperature at discrete depths.

In this study, we apply a 3D stochastic geostatistical technique that utilizes both direct and proxy measurements of temperature to help inform exploration decision-making, where the information available is representative of the discovery stage: 2-meter temperature probes and a handful of deeper temperature gradient measurements exist along with 3D models of temperature proxies, where we define proxies geologic properties that may allow favourable flow properties and thus higher temperatures (after Siler *et al.* 2018). Specifically, we identify four derivative products from 3D geologic modelling: distance to fault, distance fault terminations/intersections, Coulomb stress change and dilation changes. Specifically, we explore these geostatistical methods with the Granite Springs Valley geothermal system located in Nevada, USA that hosts a shallow thermal anomaly initially defined by 47 shallow (15–550 m) TGH datasets from exploration efforts in the 1970s. More recently, Granite Springs Valley emerged from the Nevada Play Fairway project (2016–19; (Nevada Bureau of Mines and Geology 2017)) as a particularly prospective blind geothermal resource (Faulds *et al.* 2017, 2019, 2021). As part of this effort, opaline silica sinter, silicified lacustrine sediments, and travertine were discovered cropping out discontinuously along a ~6 km (north–south) profile adjacent to the thermal anomaly (Fig. 1). Associated geophysical (gravity, magnetotellurics, and seismic) and geological studies informed an initial six-well thermal gradient drilling campaign in 2018. Maximum temperatures of 96°C at 150–250 m depth and 77.7°C at 150 m depth were encountered at the northern and southern extents of the surface geothermal deposits, respectively. However, additional drilling is required to test updated resource conceptual models and evaluate the most likely hydrothermal upflow location for the system. At Granite Springs Valley, nine candidate locations for new TGH wells have been identified that would test the proposed isotherms of the different conceptual models (e.g. P10/optimistic, P50, and P90/pessimistic) (Ayling *et al.* 2022).

We present and compare two quantitative workflows, cosimulation and local varying mean, to aid in choosing future drilling sites for TGHs at Granite Springs Valley (Fig. 1) and ideally for other geothermal prospects at the discovery stage. Both workflows combine directly observed temperatures and inferred but spatially extensive geologic proxies to temperature. These workflows represent the first stochastic temperature modelling conducted at Granite Springs Valley to date, which utilizes calculated and spatially extensive proxies related to the inferred location of faults and stress/strain-related proxies for permeability. Specifically, the methodology utilizes many equally probable 3D temperature models to calculate the probability of temperature at any location in the 3D subsurface. These temperature probabilities are then summarized into two quantities: prior value ($V_{\text{prior}}$) and entropy ($H$). $V_{\text{prior}}$ and entropy are calculated from these results and are

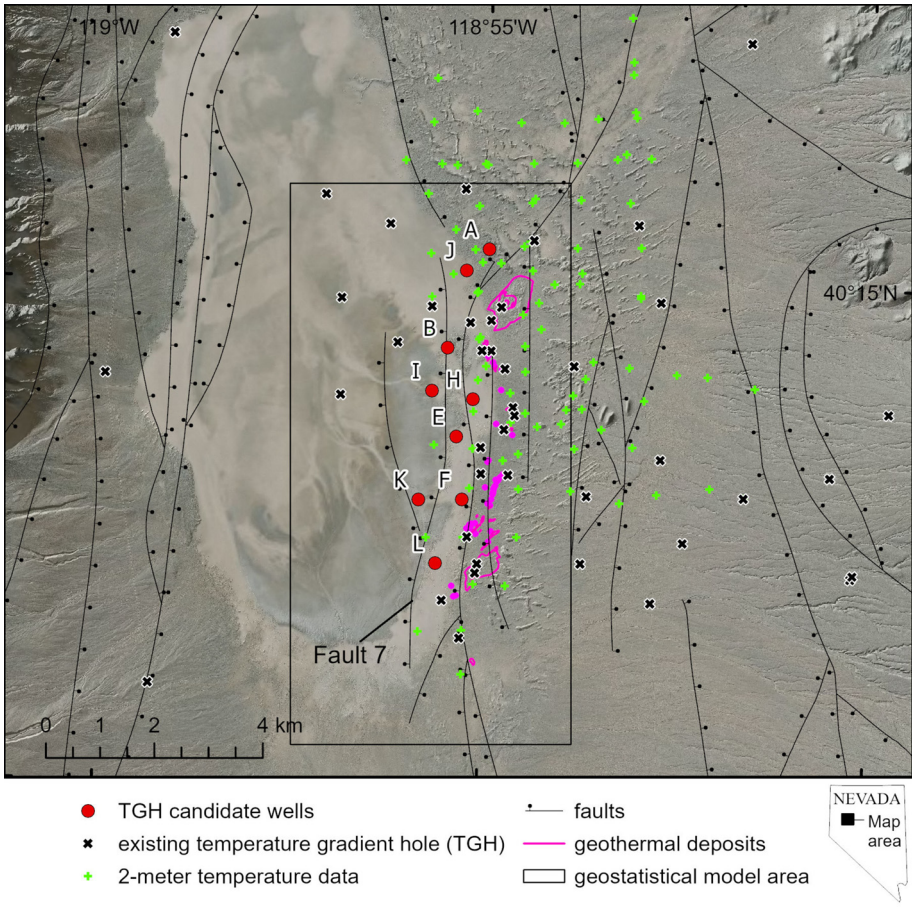


**Fig. 1.** Map of northern Granite Springs Valley geothermal area. The local 2m temperature data (green) and thermal gradient holes (TGH) temperature data locations are shown (black 'X'), though only temperature data within the geostatistical model area are used in the analyses. Also shown are the candidate locations for new TGH (red with letters).

useful to balance the two objectives of (1) confirming high temperatures and (2) improving knowledge of the subsurface with the proposed new TGH locations at Granite Springs Valley. Thus, the paper begins by providing two example calculations of $V_{\text{prior}}$ and entropy, then describes the two stochastic geostatistical techniques. The results section describes these modelling techniques applied to Granite Springs Valley. We conclude the paper with discussion of the results and potential improvements to the workflow for future application to other greenfield prospects.

## Geostatistical methodology

In this section, we introduce $V_{\text{prior}}$ and entropy to motivate the stochastic geostatistical modelling. We then briefly describe Kriging, the variogram, and stochastic simulation. Lastly, we introduce the two methods that allow geologic proxies to be included in the stochastic temperature modelling: cosimulation and local varying mean.

### $V_{\text{prior}}$

The prior value ($V_{\text{prior}}$) identifies the action that results in the highest utility given many uncertain scenarios. For the geothermal drilling example, we can make a simplifying assumption that our decision outcome is determined mostly by temperature ($T$). Thus, a $V_{\text{prior}}$ expression could be

$$V_{\text{prior}}(\boldsymbol{u}) = \max_{a}\left[\sum_{i=1}^{Ntemp_{\text{bins}}} \Pr\left(T = t_i(\boldsymbol{u})\right) v_a(T = t_i(\boldsymbol{u}))\right]; \quad a = \{\text{drill},\ \text{don't drill}\} \tag{1}$$

where the sum represents a weighted average (e.g. expected value) over the temperature bins ($t_i$), $Ntemp_{\text{bins}}$ is the number of temperature classes (eight here), $\boldsymbol{u}$ is position vector, and $v_a$ represents the utility outcome (often in turns of monetary units) of some decision action $a$, which could be to drill a TGH or not. The weights for each temperature bin are expressed with the probability Pr(.). To explore our chances of drilling success (which may be defined by intersecting hot temperatures), we want to obtain the probability of the temperature classes occurring in the 3D subsurface, utilizing all the current information available to us today.

There are two examples of these probabilities in Table 1. The first demonstrates a location where all the temperature bins have the same probability: in other words, this example represents a noninformative or uniform probability. The second example has 0 probability for the lowest temperature bin and the temperature bin of 100–110°C has doubled to 0.25 from 0.125. In our real-world application of these methods to Granite Springs Valley, these probabilities are informed by the stochastic models. The value outcomes in the last row of Table 1 are nominal but represent how hotter temperatures in general represent a more profitable outcome. For the examples shown, if we assume drilling a TGH costs $1 000 000, reaching a maximum temperature in that well of <40°C, would result in a complete loss of that investment. If progressively higher temperatures are measured, the monetary loss decreases. If 70–80°C is measured, there is a relatively small monetary gain from the drilling investment. If progressively higher temperatures are measured, progressively larger monetary gains are made. If we use these two examples to calculate $V_{\text{prior}}$, we can compare how these probabilities change our expected outcome: how well we expect to do if we drill a well at these two locations.

Table 2 shows the $V_{\text{prior}}$ calculation and how the change in probability between the two examples affects our expected outcome. For Example 2, in which there is zero chance to reach temperatures lower than 40°C and a relatively high probability to

**Table 1.** Discrete temperature bins for which probabilities are decision outcomes are calculated (demonstration values)

| Temperature (C°) class | $T \leq 40$ | $40 < T \leq 60$ | $60 < T \leq 70$ | $70 < T \leq 80$ | $80 < T \leq 90$ | $90 < T \leq 100$ | $100 < T \leq 110$ | $T > 110$ |
|---|---|---|---|---|---|---|---|---|
| **Example 1** Probability at $\Pr(T = t_i(x_1, y_1, z_1))$ | 0.125 | 0.125 | 0.125 | 0.125 | 0.125 | 0.125 | 0.125 | 0.125 |
| **Example 2** Probability at $\Pr(T = t_i(x_2, y_2, z_2))$ | *0* | 0.125 | 0.125 | 0.125 | 0.125 | 0.125 | *0.25* | 0.125 |
| Temperature Based Outcome $v_a(T = t_i)$ | $-1 000 000 | −$662 500 | −$325 000 | $12 500 | $350 000 | $687 500 | $1 025 000 | $1 362 500 |

**Table 2.** Example $V_{\text{prior}}$ calculations for the probability examples of Table 1

| | Weighted average calculation | $V_{\text{prior}}$ | Entropy |
|---|---|---|---|
| **Example 1** Probability at $\Pr(T = t_i(\boldsymbol{u} = x_1, y_1, z_1))$ | 0.125*$ − 1 000 000 + 0.125* − $662 500 + 0.125* − $325 000 + 0.125*$12 500 + 0.125 * $350 000 + 0.125 *$687 500 + 0.125*$1 025 000 + 0.125*$1 362 500 | $181 250 | 3.0 |
| **Example 2** Probability at $\Pr(T = t_i(\boldsymbol{u} = x_2, y_2, z_2))$ | ***0****$ − 1 000 000 + 0.125* − $662 500 + 0.125* − $325 000 + 0.125*$12 500 + 0.125 * $350 000 + 0.125 *$687 500 + ***0.25****$1 025 000 + 0.125*$1 362 500 | $434 375 | 2.75 |

reach 100–110°C, the expected outcome is $434 375, whereas the expected outcome Example 1, which has equal probability for each temperature bin is $181 205. This demonstrates that representing the probabilities as realistically as possible in the 3D subsurface is critically important.

## Entropy

The other quantity explored to identify where to drill new wells is entropy, which addresses a very different drilling objective from Vprior. Entropy quantifies statistical disorder, and specifically in this study how the temperature probabilities are distributed among the temperature classes. The following equation is the Shannon entropy (Shannon 1948) value at each location, $u$, on the basis of the class probabilities $\Pr(T = t_i)$:

$$H(u) = -\sum_{i=1}^{Ntemp_{\text{bins}}} \Pr(T = t_i(u)) \log_2 \Pr(T = t_i(u)) \qquad (2)$$

The Entropy for the two previous examples is shown in the last column of Table 2. Example 1 has the higher Entropy relative to Example 2: the more the probabilities are equivalent, the higher the entropy. Entropy is reduced (e.g. statistical disorder reduced) when one class has a higher probability relative to the other classes. Entropy doesn't give a unique value for different probability combinations: if the 0.25 changed from the $100°C < T \leq 110°C$ temperature class to another, it would still be 2.75 for that combination. For these two examples, the temperature uncertainty is higher in Example 1 relative to Example 2, but the chance that newly acquired data will constrain the modelling is also higher.

If we use Vprior to site new wells, the drilling objective is to find economic temperatures. Conversely, by targeting locations in the model that have high Entropy locations, we aim to reduce disorder: by collecting new data in those areas and adding the newly acquired temperature data to the model the Entropy would then be reduced, and future models would be better constrained. In other words, areas of higher Entropy are locations where new information has the most potential to reduce uncertainty.

## Variograms & stochastic 3D geostatistics

In order to make spatial decisions about where to drill using $V_{\text{prior}}$ and entropy in the 3D subsurface, we must explicitly represent all the locations ($\boldsymbol{u}$) and define their respective temperature class probabilities $\Pr(T = t_i(\boldsymbol{u}))$. Stochastic geostatistics provides methods for making these spatial inferences. The language here will speak specifically to temperature data, but the equations are general and applicable to any spatial data type.

Kriging is sometimes referred to as geologically based spatial interpolation, because through the variogram, modelers can impose which direction(s) have higher spatial correlation. Experimental variograms are calculated from observed data, but when data are very sparse, geologic modelers can supplement this with their geologic knowledge of the area (e.g. depositional trends, structural tendencies, or stress orientations). Kriging interpolation uses the variogram/covariance to determine how much weight to put on each observed data point when calculating the value at each unsampled location. So unlike inverse distance weighting, Kriging will give more weight to data that are in the direction with lower dissimilarity, or higher spatial continuity (Webster and Oliver 2008), and importantly, this allows for data that are aligned with geologic trends to have more impact on estimated values. Stochastic simulation techniques combine the Kriging interpolation methods, with Monte Carlo sampling to produce many different, but equally likely models that match the observed measurements and reproduce the variogram.

For our purposes, the variogram is a statistical measurement that quantifies how the temperature changes in space, and particularly how it changes in different directions. The 3D variogram captures this by comparing pairs of data, keeping track of the magnitude and direction of the separation of the pairs of data. Formally, this comparison is a squared difference of two data points separated by a distance vector lag $\boldsymbol{h}$ seen in the definition of the (semi-)variogram

$$\gamma(\boldsymbol{h}) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [T(\boldsymbol{u}_i) - T(\boldsymbol{u}_i + h)]^2 \qquad (3)$$

where $\boldsymbol{T}$ represents the property being modelled (here temperature), $\boldsymbol{u_i}$ is position vector, and $\boldsymbol{N}$ is the number of pairs of data found at distance lag $\boldsymbol{h}$, which has a magnitude and angle (Remy *et al.* 2011). The division and summation are an averaging of the difference for all data pairs that fit that same distance lag $\boldsymbol{h}$. The variogram, $\gamma$, is a measure of 'dissimilarity.' In general, we expect this dissimilarity to increase with increasing lag distance because temperatures (among other geologic parameters) will differ more at larger distances as shown in Figure 1. When variograms are used for properties such as porosity in 'layer cake' geology, we expect variograms calculated in horizontal direction to change more slowly than in the vertical direction. This means the spatial correlation is higher horizontally than vertically. However, this may not be the case in geothermal environments where the heat source is originating from depths and will be different between areas of convective and conductive heat transfer.

Figure 2 depicts a generic experimental variogram where the points represent the quantity in equation (3) at discrete $h$ separation distances for data pairs. With sparse data, there will be much fewer variogram measurements, especially if it's a directional variogram (e.g. horizontal or vertical) and the azimuth and dip between data are limited, which means there would be many fewer dots than what is seen in Figure 2. Variogram models are chosen to reasonably fit the experimental variogram, but also to provide dissimilarity values at all distances $h$ (instead of just the 'dots' of Fig. 2) and directions that allows the geomodeler to levy local resource knowledge into the variogram model (Gringarten and Deutsch 2001; Chilès and Delfiner 2012). During the modelling process one assigns a range (distance at which maximum dissimilarity or conversely zero correlation is reached), a sill (the maximum dissimilarity), and a nugget effect (dissimilarity at infinitesimal distances), all of which or noted in Figure 2. Figure 2 is a depiction of a one -dimensional variogram. In two-and three-dimensions, the variogram is an ellipse and an ellipsoid, respectively, where the radii are defined by the ranges in the different directions. The directions with the longer and longest radii are the directions with maximum spatial continuity. If the radii are the same, the property is isotropic.

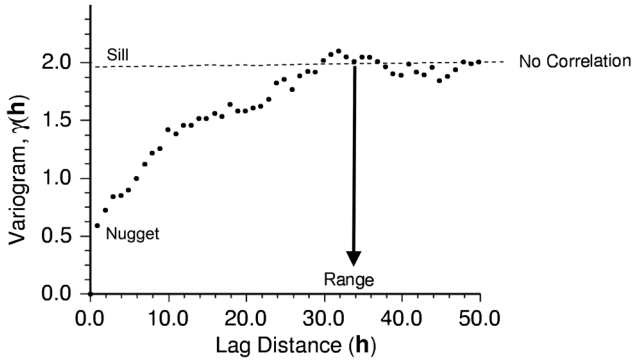


**Fig. 2.** Example variogram with separation distance on the *x*-axis and dissimilarity on the *y*-axis.

The variogram is initially calculated instead of its converse, spatial correlation, as the variogram models (as opposed to covariance models) ensure more stable mathematics for Kriging (Isaaks and Srivastava 1989). However, after a variogram model $\gamma(h)$ is fit and chosen, the spatial covariances $C(h)$ are calculated as follows:

$$C(\boldsymbol{h}) = C(0) - \gamma(\boldsymbol{h}) \tag{4}$$

where $C(\boldsymbol{h})$ is the covariance at a lag vector $h$ which includes distance and direction, $C(0)$ is the variance of the data (e.g. the sill of the variogram), and $\gamma(\boldsymbol{h})$ is the variogram defined at a lag distance $h$. Therefore, higher covariance means that two data separated by some distance and direction are more correlated.

Kriging interpolation uses the variogram/covariance to determine how much weight to put on each observed data point when calculating the value at each unsampled location. This is eloquently summarized by the matrix equation that solves for the Kriging weights $\lambda$

$$\lambda_{\alpha_i 0} = C^{-1}_{\alpha_i \alpha_j} c_{\alpha_i 0} \quad \forall\alpha\text{:primary data}, \quad 0\text{:unsampled location} \tag{5}$$

where $C$ is the data-to-data covariance matrix (the covariances between each of the conditioning data), $c$ is the data-to-unknown covariance array (the covariances between each of the conditioning data and the location being estimated), and $\lambda$ is the array of weights of each of the conditioning data. $\alpha_{i/j}$ index the locations of the primary data, therefore equation (5) could represent a large system of equations depending on how many data are present. The important take-away is that data points with high spatial covariance to the unknown location ($c_{\alpha_i 0}$) will get a higher weight ($\lambda_{i0}$), where the directional information of the variogram (e.g. spatial anisotropy) is embedded in the value. For our purposes, kriging uses these weights to make an interpolated value of temperature ($T^*$) at unsampled location $\boldsymbol{u}$, that is a weighted mean

$$T^*_{SK}(\boldsymbol{u}) - m = \sum_{\alpha=1}^{n(u)} \lambda_\alpha [T_1(\boldsymbol{u_\alpha}) - m] \tag{6}$$

where $\lambda_\alpha$ is the column matrix of the $n(u)$ kriging weights, $T_1(u_\alpha)$ are the neighbouring data values, where 1 indicates primary data. Equation (6) is the simple Kriging ($SK$) equation (Remy *et al.* 2011); $[T(u_\alpha) - m]$ represents the residual data values by removing a known mean $m$. Kriging works in residuals to ensure unbiased estimations. Kriging is a deterministic estimation technique, meaning it only produces one interpolation model. Kriging provides an unbiased, conservative estimate: the average of the errors will be as small as possible and will not over or under predict systematically. Here the errors are the difference in the estimated (predicted) value and the true value, which is of course unknown, but is handled as a random variable as is customary in statistical inference.

Kriging is deterministic and will therefore not provide the probabilities that we need for $V_{\text{prior}}$. Thus, we use a stochastic method that produces many possible 3D temperature models, which, like Kriging, reproduce the temperature measurements at their locations. Stochastic simulation uses the Kriging weights (equation 5) or estimate (equation 6) to define a conditional cumulative density functions (ccdf) and draws a Monte Carlo sample in order to make many equally probable interpolation models (Deutsch and Journel 1998; Goovaerts 2001; Hansen *et al.* 2006; Remy *et al.* 2011). To produce a single model (sometime referred to as a realization), stochastic simulation populates a user-defined grid volume by defining a random path around the grid. As defined in (Remy *et al.* 2011), a simulation algorithm is summarized as follows, where text in italic will be explained in the next section :

(1) Define an order to visiting each node **u** of the grid via a random seed.
(2) For each node ($\mathbf{u}_i$):
   a. Get the conditioning data consisting of neighbouring original data and previously simulated values *(and the secondary data)*
   b. Get the local Gaussian ccdf for the primary variable, with mean equal to the *(co)*kriging estimate and variance equal to the *(co)*kriging variance
   c. Draw a value from that Gaussian ccdf and add it to the data set

In the first step, an order for when each cell is visited is established. Step 2 is repeated using the order established in Step 1. Step 2a uses a 'search neighbourhood' to define which data and previous simulated values will be used for location $\mathbf{u}_i$, e.g. which are within certain distances. Step 2b generates the ccdf by solving for location-specific Kriging-based weights $\lambda_{i0}$ using the data defined in Step 2a. The Gaussian ccdf is defined by the Kriging estimate and variance as the mean and variance respectively. The Monte Carlo draw (step 2c) results in the simulated temperature at location $\mathbf{u}_i$. Steps 1 and 2 are repeated $N$ times to make $N$ realizations, where the difference is the order in which unknown/unsampled locations are visited. The end result is $N$ 3D models that, taken together represent our uncertainty of any of the temperature classes occurring at any location $\mathbf{u}$:($x_1$, $y_1$, $z_1$) in the 3D subsurface.

### Normal space, data space and histogram transformation

Variogram modelling and stochastic simulation are performed in a transformed space where all the temperature data are turned into Gaussian scores. This is similar to scaling performed in machine learning. The stochastic temperature models are first created in these spaces, then transformed back into the 'temperature data space.' During this step, the same histogram as the observed temperatures can be reproduced and/or one can choose to extrapolate from the lowest and highest temperatures observed (Remy *et al.* 2011). Therefore, the variables $T_i(\boldsymbol{u_\alpha})$ should be assumed to be the normal scores of the data provided.

## Secondary information: indirect indication

Geostatistical methods allow for the inclusion of secondary data that may also constrain the quantity of interest (here temperature) in addition to primary data. The secondary data are typically used to capture larger spatial trends of the main property (Remy *et al.* 2011; Chilès and Delfiner 2012). Commonly, the secondary data are geophysical or remote-sensing data (Gloaguen *et al.* 2005, 2012; Hansen *et al.* 2006). Recently, other less conventional data have been used as secondary data, such as geologic interpretations of lithofacies (Boyd *et al.* 2020). In the Results: stochastic temperature

modeling at Granite Springs Valley section, we will describe in detail the four quantities used as secondary data: distance to fault, distance to fault intersection (Siler and Faulds 2013; Siler and Pepin 2021), Coulomb shear stress change, and dilation (Siler *et al.* 2018). We explore two possible methods for including secondary information: cosimulation and local varying mean (LVM). Cosimulation represents the secondary data as discrete or finite data locations and incorporates them via variogram analysis, similar to Kriging. LVM uses secondary information to describe how the mean of the primary data (e.g. temperature data) changes in space.

### Cosimulation

Like stochastic simulation for Kriging, cokriging can be used to produce stochastic results. This is called cosimulation. The theory and Kriging expressions are flexible to allow the data $T(u_\alpha)$ and the unknown $T^*_{SK}(\boldsymbol{u})$ to be different attributes. Commonly, the estimation of porosity has been informed by ancillary data such as seismic amplitude data . Cokriging provides an extension of Kriging where data of different attributes can be used to make estimations of a property of interest, in our case temperature. The simple cokriging equation for an unsampled temperature location $\boldsymbol{T^*_1(u)}$ is as follows

$$\boldsymbol{T^*_1}(\boldsymbol{u}) - m_1 = \sum_{\alpha=1}^{n_1(u)} \lambda_\alpha [T_1(\boldsymbol{u_\alpha}) - m_1] + \sum_{\beta=1}^{n_2(u)} \lambda_\beta [T_2(\boldsymbol{u'_\beta}) - m_2] \quad (7)$$

where $n_1(u)$ denotes the number of neighbouring temperature data $T_1(\boldsymbol{u})$, and $n_2(u)$ is the number of secondary data $T_2(\boldsymbol{u'_\beta})$. The $\lambda_\alpha$ and $\lambda_\beta$ are the corresponding cokriging weights. The means of the two respective properties are represented by $m_1$ and $m_2$.

The cosimulation algorithm is identical to what is presented above in theEntropy subsection. The only difference is that in Step 2a, one would add 'and the secondary data' for conditioning data and change kriging to 'cokriging' in Step 2b, as is denoted in red text.

The cokriging theory and development originally required the variograms of the primary and secondary data, as well as cross-variograms between the primary and secondary data (Zhu and Journel 1993). While modelling the secondary variogram was relatively easy, since it was usually more spatially abundant, the cross variogram is at least or more challenging than modelling the primary variogram alone. To make cokriging more assessable, the Markov–Bayes approach was developed, which essentially estimates the cross-variogram by scaling the secondary variogram with the correlation coefficient between the primary and secondary data (Ma and Journel 1999).

In general, three factors control how much the secondary data affect simulated cokriging value. First is the correlation coefficient between the primary and secondary data, as this will directly affect the Kriging weights and in turn the ccdf's and estimate. With a higher correlation, there will be an increase in the influence of the secondary data. The second factor is controlled by the nugget of the variogram: as the nugget is increased this denotes that the primary data is 'noisier'. Thus, the secondary data can have a larger influence. For temperature modelling, one doesn't expect a large nugget effect (e.g. large variations at very small spatial scales) due to the physical process of heat conduction. Finally, sample density of both the primary and secondary data influences the cokriging results. The secondary data have more influence when primary data are under sampled, but the density of the secondary data is also important.

**Principle component analysis.** The cokriging theory and equations have the flexibility to extend to many secondary data types. However, the computational resources used herein, namely the open-source Stanford Geostatistical Modelling Software (SGEMS) only allow for one secondary variable (Remy *et al.* 2011). Therefore, we use the dimensionality reduction method, Principal Component Analysis (PCA), to represent the information from two proxies into one variable. PCA rotates the attributes to find a space that best represent the variability of multiple parameters (De Lathauwer *et al.* 2000).

### Local varying mean

Another method for stochastically simulating temperature data while also incorporating secondary information is simulation with local varying mean (LVM). The LVM method utilizes a secondary data set to capture systematic variation in the mean of the property. Typically, the LVM values are derived from aerially extensive remote-sensing data such as seismic reflection. Thus, LVM should (1) have a trend and (2) this trend should smoothly vary. Equation (8) provides the LVM equation

$$T^*_{\mathrm{LVM}}(\boldsymbol{u}) - m^*(\boldsymbol{u}) = \sum_{\alpha=1}^{n(u)} \lambda^{SK}_\alpha(u)[T(\boldsymbol{u}_\alpha) - m^*(\boldsymbol{u}_\alpha)] \quad (8)$$

where the locally varying mean that is available at all locations is denoted as $m^*(\boldsymbol{u})$. Typically, the secondary variable is transformed through linear regression into the local mean of the primary attribute. This differs from the cosimulation method, which integrates the information through covariances (via variogram).

Therefore, the secondary data $T_2(\boldsymbol{u'_\beta})$ may need to (1) be interpolated to be represented at all locations $T_2(\boldsymbol{u})$ and (2) be regressed to the parameter of the primary data using collocated primary data

$$t_1(\boldsymbol{u_\alpha}) = at_2(\boldsymbol{u_\alpha}) + b \quad (9)$$

which regression parameters, $a$ and $b$, can be used on the interpolated secondary data to define the local varying mean $m^*(\boldsymbol{u})$

$$m^*(\boldsymbol{u}) = at_2(\boldsymbol{u}) + b \quad (10)$$

Finally, the $m^*(\boldsymbol{u})$ is turned into the normal scores consistent with the normal scores of the data.

## Results: stochastic temperature modelling at Granite Springs Valley

This section presents the results of the described methods applied to Granite Springs Valley, NV. First, the model volume, the temperature (primary) data, and their respective 3D variogram calculations are described. Next the secondary data originating from the geologic conceptual model and their geomechanical/stress calculations and the Principal Component Analysis calculations to combine them are described. This information is used as a secondary constraint in the 3D stochastic temperature modelling, which is performed in the open source SGEMS software (Remy *et al.* 2011). Finally, the 3D modelling results, the 1D probability profiles extracted at the candidate well locations, and their respective $V_{\mathrm{prior}}$ and entropy measurements are presented.

### Existing temperature data & 3D variogram

A 5175 m (easting) by 10 350 m (northing) by 800 m (depth) volume in the Granite Springs Valley area was the focus of 3D geostatistical modelling. This volume centers around the existing temperature observations (Fig. 3). Given the elevation of Granite Springs Valley (~1200 m), the model domain reaches 300 m of elevation. This represents the possible depths for the next TG drilling campaign. Geostatistical interpolation is bounded by what was understood to be the principal western fault, therefore no temperatures west of this fault were used or interpolated, assuming that statistical stationarity is not valid across the fault. Also shown in

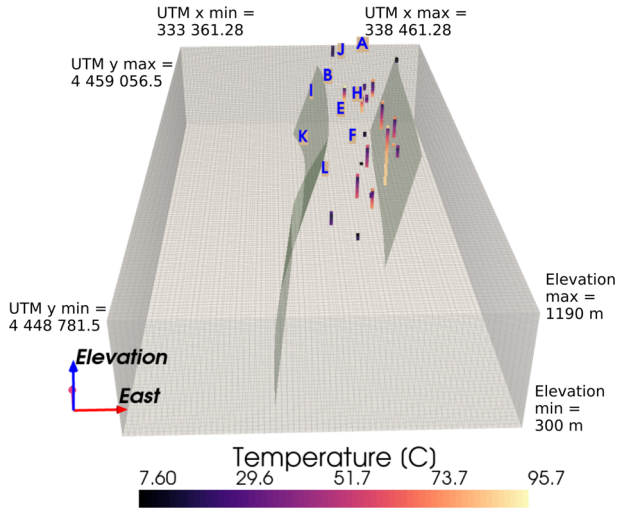


**Fig. 3.** Geostatistical modelling volume with 2 faults from the Granite Springs Valley conceptual model, the temperature data projected onto well traces, and the 9 candidate well locations (surface location labelled with blue letters).

Figure 3 is what was interpreted as the main eastern fault in the area. The grid size is 69 × 138 × 90 cells such that each cell represents 75 × 75 × 10 m. The higher depth resolution represents the importance of capturing temperature trends in depth. The figures that follow with the temperature modelling results represent the same volume (UTM and elevation) as Figure 3.

We utilize both the existing 2-meter temperature and the vertical temperature data for the 3D variogram calculation and subsequent geostatistical model generation. The left plot in Figure 4 is the histogram of the observed temperature data, where the peak around 20°C is from the 2-meter measurements, the most common data type. The hottest temperatures are from the deep vertical logs that reach up to 550 m depth (around 600 m above sea level). Given the occurrence of opaline sinter (suggestive of reservoir temperatures >175°C) and geothermometry suggesting reservoir temperatures of ~140°C (refer to Ayling *et al.* 2022), it was deemed reasonable to extrapolate to 140°C during the data transformation stage such that each 3D model will have a temperature distribution that closely resembles the histogram on the right in Figure 4. Extrapolation was done using the simulated normal scores in the resulting stochastic simulations but increasing the 100th percentile from 95C to 140C (please see histogram transformations in Remy *et al.* 2011 which gives options for the shape of the extrapolation).

Figure 5 contains four directional variograms of the temperature data: north–south, vertical, NE (N45E), and east–west. The 'X' marks depict the temperature data (empirical) variograms. In all but the vertical variogram, there is also a continuous black line depicting the modelled variogram in that direction. Variograms are insightful in determining which direction the data have more correlation or change the slowest; this is captured partially by the range: the distance ($x$-axis) at which the variogram asymptotes to the highest dissimilarity value (sill value on $y$-axis). The roughly interpreted value of the range for the four directions is depicted in Figure 5, except for the vertical direction, which is discussed below. The data variograms are influenced by the configuration in which the data are sampled. For instance, as seen in Figure 3, most measurements are collected along the north–south direction. Therefore, it is imperative to combine the raw variogram calculations with geologic knowledge of the area, which is done by fine-tuning the 3D variogram model that ultimately goes into the geostatistical modelling.

The empirical and modelled variograms were calculated in 3D using the SGEMS 3d variogram functionality (Remy *et al.* 2011), so Figure 5 contains 1d slides of the 3D ellipsoid. Also note that the dissimilarity is in terms of the normal scores ($y$-axes). The modelled variograms are fit by adjusting the direction and magnitude of the maximum, median and minimum ranges of the 3D ellipsoid.

The absence of a sill in the vertical variogram (Fig. 5a) indicates there is a trend in the data. As explained in (Gringarten and Deutsch 2001), the trend should be removed before fitting the variogram. This was done, but then the range was found to be only 100 meters. Our modelling goals are to give highest spatial continuity down dip of the faults, since we hypothesize these are facilitating heat transferring up to the shallow subsurface.

The main western fault, seen in Figure 3, is a possible controlling geologic feature for temperature, so we use this to help describe the 3D ellipsoid that is the 3D variogram. The mean strike of the fault is azimuth = 000 (north–south), the mean dip is 60°, and the mean dip direction is azimuth = 270 (due west); these angles are represented in Figure 6. We expect geothermal fluids to be upwelling up-dip in the fractured rock (high porosity and permeability) near faults, so we use the dip direction (azimuth = 270, clockwise from north) and dip (60° from $XY$ plane) of the western fault to define the direction of the major axis. Given the absence of a sill in the vertical variogram, the range for the vertical variogram is chosen to fit the observed range in the north–south direction (the one with the most data). The major radius along the fault strike and dip is set at 2700 m, which when projected into the $XY$ plane is approximately 1300 m (= 2700*cosine(60°)), matching roughly what is observed in the north–south variogram (Fig. 5b). After some initial testing in the geostatistical realizations, the medium axis is the fault strike direction [azimuth = 000 (north–south), dip = 0° (horizontal)], and minimum axis is orthogonal to the fault plane (azimuth = 270,

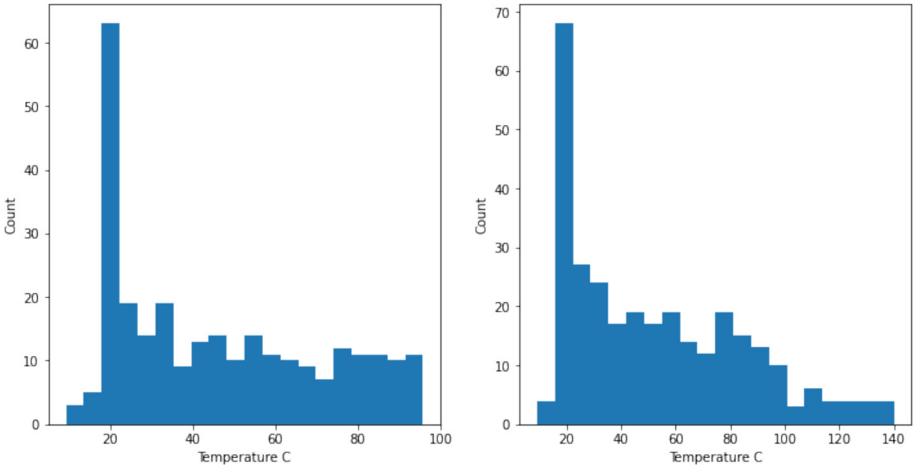


**Fig. 4.** Left: distribution of observed Granite Springs Valley temperatures in the 2m and deeper well logs. Right: the extrapolation used to put highest temperatures at 140°C in all 3D temperature realizations.

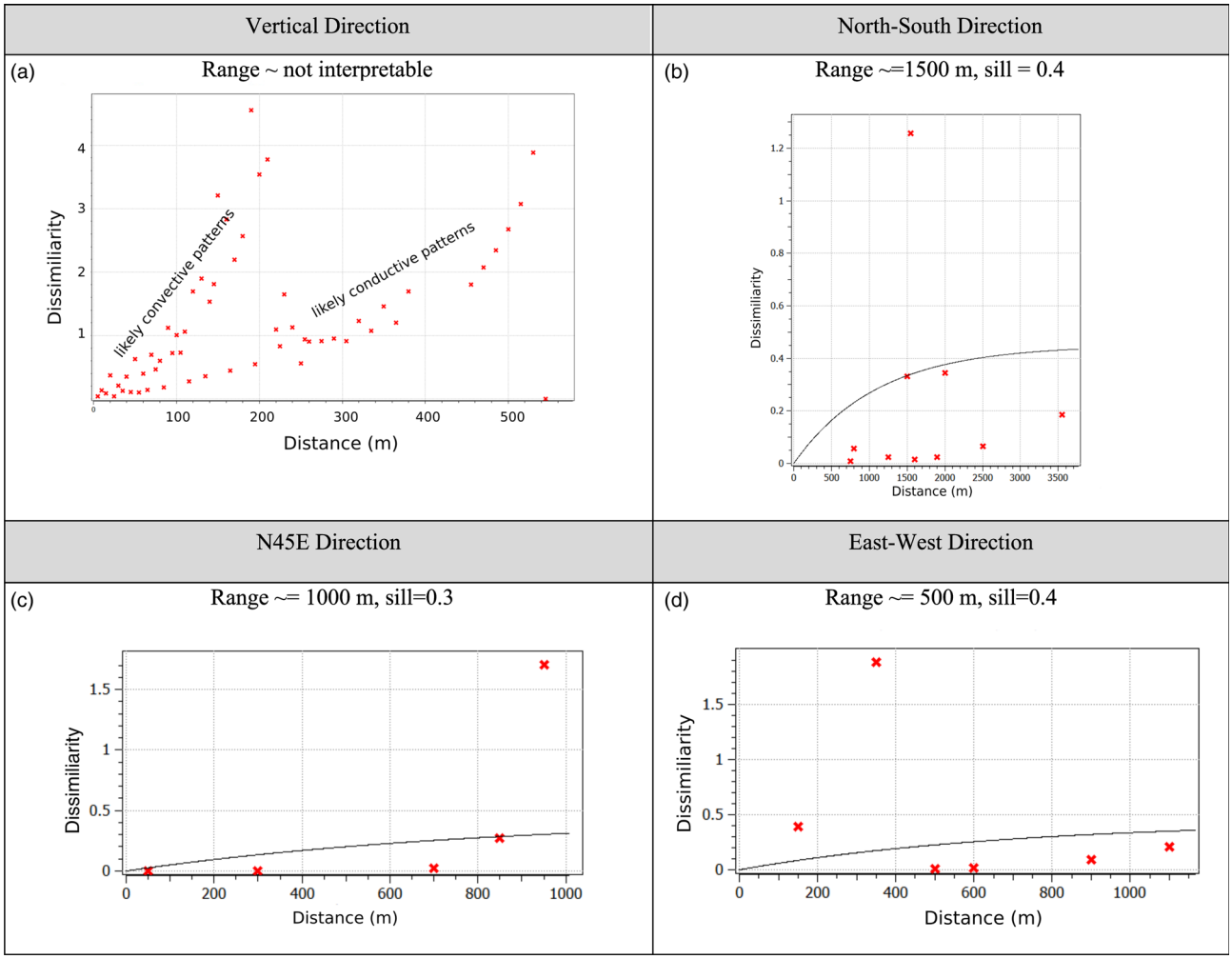


**Fig. 5.** Four directional variograms informed by the 16 vertical temperature logs and 2m temperatures: (**a**) vertical temperature variogram. The conduction v. convection trends are clear in the two spatial dissimilarity trends. (**b**) North–south direction with range about 1500 m. (**c**) 45 degree range ~1000 m, and (**d**) east–west range is less than 500 m.

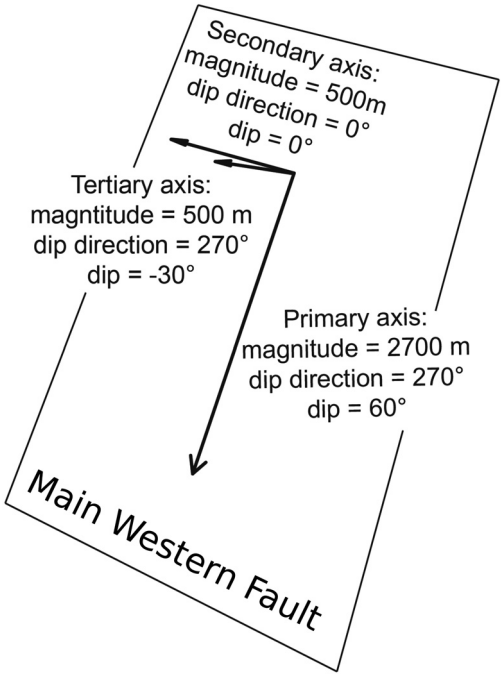


**Fig. 6.** Conceptual cartoon showing main western fault and the radii of the primary variogram.

dip −30 normal to the fault plane). The radii for the medium and minimum axes are set to 500 m, which is consistent with the angles seen in Figure 5.

Table 3 summarizes the 3D variogram that was fit to the measured temperature data. In the table are the three radii of the ellipsoid and the direction in which the maximum direction of correlation is oriented (max radius). Also, the table shows that a Gaussian variogram is chosen, which describes a property that changes slowly at small spatial scales, which is appropriate for the property of temperature given the physical processes of diffusion. Other variogram types (e.g. exponential or spherical) would impose more abrupt changes at very short spatial scales. The variogram model has no nugget effect, again to prevent these discontinuities at short distances in the interpolated models.

Figure 7 displays two of the 50 models created using only temperature data along with the variogram described in Table 3. There are five east–west ($X_{min}$ to $X_{max}$) slices shown; these provide

**Table 3.** The parameters (direction, ranges, and type) of the primary variogram (fit to the existing measured temperature data)

| Gaussian, nugget = 0 | MAX | MEDIUM | MIN |
|---|---|---|---|
| Ranges (meters) | 2730 | 500 | 500 |
| | Azimuth | Dip | Rake |
| Angles (degrees) | 270 | −60 | 0 |

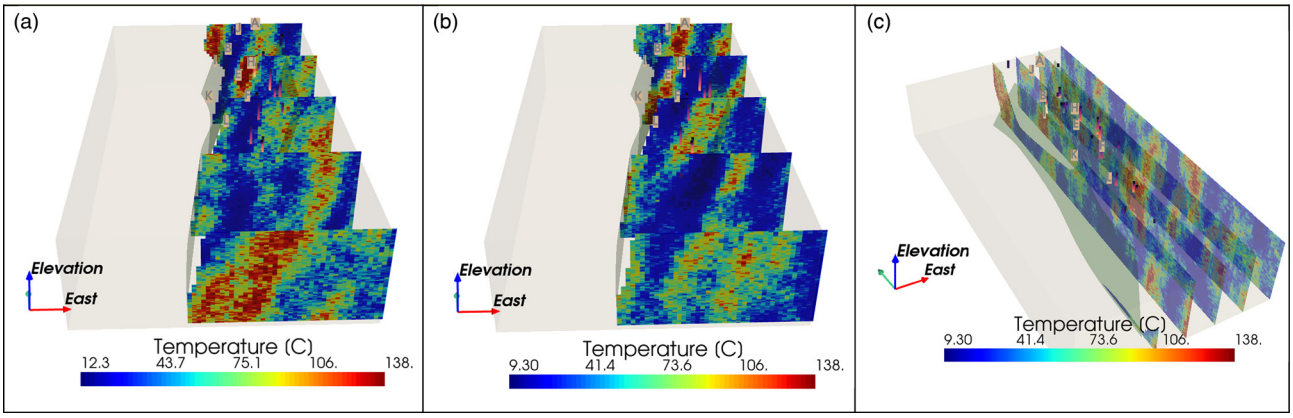


**Fig. 7.** Two sample realizations of the subsurface temperature distribution generated *only* with temperature data. (**a**) Realization 25, (**b**) realization 49 and (**c**) *y*-slices of realization 25.

visual confirmation of the modelled 3D variogram ellipsoid, as one can observe the 60° dip. The southern-most slice (foreground slice) differs the most between the two realizations, which is expected since there are no subsurface temperature data in that region. The middle three sections have similar locations of higher temperature. In general, these stochastic realizations do not have a smooth, continuous character of temperature as we would expect in a geothermal setting, although the Gaussian variogram model with no nugget effect and a relatively long range was given. More constraints are needed to enforce more geothermal realism into the 3D temperature models.

## Secondary data: temperature proxies from geologic modelling

Conceptual geologic modelling is an increasingly important step in discovering and/or understanding geothermal fluid circulation (Siler *et al.* 2018; Siler and Pepin 2021). At Granite Springs Valley, 3D geologic analyses are used to produce four quantities: distance to fault, distance to fault intersection (Siler and Faulds 2013; Siler and Pepin 2021), Coulomb shear stress change, and dilation (Siler *et al.* 2018). Faults and fault intersections are important structures that can produce permeability, which then can allow for pathways for geothermal upwelling, so the distance to faults and distance to fault intersection are reasonable proxies for high temperatures, i.e. they are likely loci of convective heat transport. The dilation and coulomb shear stress change proxies are modelled stress and strain effects of fault slip. Areas that undergo high dilation and high coulomb shear stress change are prospective areas for geothermal processes, since they represent long-lived areas of permeability generation and maintenance (Siler *et al.* 2018). These are shown in Figure 8.

Secondary data are incorporated via the Markov–Bayes method, in which correlation coefficients between the four proxy parameters and the measured temperatures are calculated. Only proxy values that are within 100 m of a measured temperature are used to compute the correlation matrix. A weighted correlation coefficient is used to give more weight to deeper measurements, where the

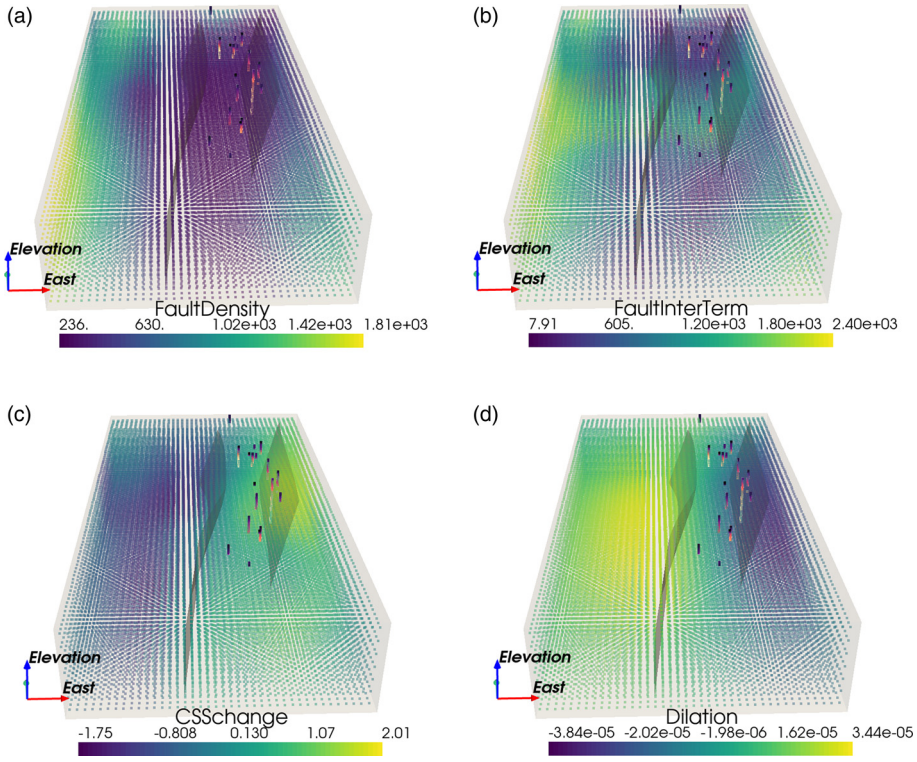


**Fig. 8.** Temperature-proxies (**a**) distance to fault, (**b**) distance to fault intersections/terminations, (**c**) Coulomb shear stress (CSS) and (**d**) dilation tendency.

**Table 4.** Correlation coefficient table for the four secondary data (proxies for temperature) and the measured temperature (last column)

| | Distance to fault | Distance to fault intersections | Coulomb stress change | Dilation | Measured temperature |
|---|---|---|---|---|---|
| Distance to fault | 1 | 0.63 | −0.34 | 0.18 | 0.29 |
| Distance to fault intersections | 0.63 | 1 | −0.31 | 0.22 | 0.21 |
| Coulomb stress change | −0.34 | −0.31 | 1 | −0.86 | 0.35 |
| Dilation | 0.18 | 0.22 | −0.86 | 1 | −0.56 |
| Measured temperature | 0.29 | 0.29 | 0.35 | −0.56 | 1 |

weights are normalized by the depth of each temperature measurement. The weighted correlations are shown in Table 4. The last row and column contain the correlations of the temperatures and the four proxies, none of which are very strong. The two fault proxies (distance to fault and fault intersections) are somewhat correlated at 0.63, and the two stress-related proxies have a high anti-correlation of −0.86.

These correlation coefficients are one way to check how these proxy parameters compare to relatively 'collocated' temperature measurements (here defined as within 100 m), but they are not a definitive way to assess which one of these four proxies is best to use. These collocated relationships are highly influenced by where the temperatures were collected. In general, the temperature measurements are further from the fault intersection values than they are for the distance to fault. Therefore, in general, sampling bias or lack of sampling of the sparse temperature data will affect how we perceive the relationship of these proxies to temperature.

## PCA results

The principal component analysis (PCA) is performed twice: on the two fault proxies (distance and distance to termination) and the two stress proxies (Coulomb stress and dilation). The explained variance ratios of 0.8 and 0.9 by the first principal component for the fault and stress proxies, respectively, indicate that PCA is successful at dimensionality reduction. We continue with the magnitude of these two PCA1 (first component of two different pairs of proxies) as our secondary data. We refer to these from now on as PCA1 Fault (which combines the distance to faults and distance to intersections and terminations) and PCA1 Stress (which combines the Coulomb stress change and dilation tendency), which are used to produce two separate cosimulation results.

In Figure 9, the PCA1 Fault (left side) diverges from the PCA1 Stress (right side) in magnitude in the NE region of our model volume. However, there is also a relatively high density of temperature measurements in this area, which have more influence on the interpolated values than the secondary data. None of the candidate well locations are in the area where the fault and stress proxies diverge the most (Fig. 3).

As described in the Variograms & stochastic 3D geostatistics section, the Markov–Bayes for cosimulation requires the correlation coefficient between the primary data (here the temperature) and the secondary (the geologic proxies) in order replace the cross variogram. Table 4 demonstrates the calculated correlation. However, a secondary variogram is still needed. Figure 10 displays the empirical variogram in discrete 'X' marks (from the PCA analysis of the geologic proxies) with the modelled variogram in black. The parameters of that modelled secondary variogram are shown in Table 5. Note that as should be expected the range is longer than the primary variogram: the proxies have a longer spatial correlation than the temperature itself. Also, because it is the first component of the PCA transformation, the sill or contribution amount is quite low since they are standardized between −1 and 1.

## Local varying mean (LVM) calculations

As described in the Local varying mean subsection, LVM is derived from a secondary source, but must be in terms of the primary data, temperature, and must be exhaustively available within the simulation area (unlike secondary data in cosimulation, which can be at limited locations). Interpolations of the four geologic proxies were used to calculate each LVM in terms of temperature using the four regression expressions. We explain later how the spatial nature of the LVM results can partially be explained by the fact that we are using an interpolated version of the proxies.

Additionally, to develop an LVM in terms of temperature from the geologic proxies, four different linear regressions are performed using the same 'collocated' temperature measurements (equations 5 and 6). The resulting coefficient, intercept, and R^2 score for the four geologic proxies are shown in Table 6. The regression was done including the depth of the temperature measurement, but converting

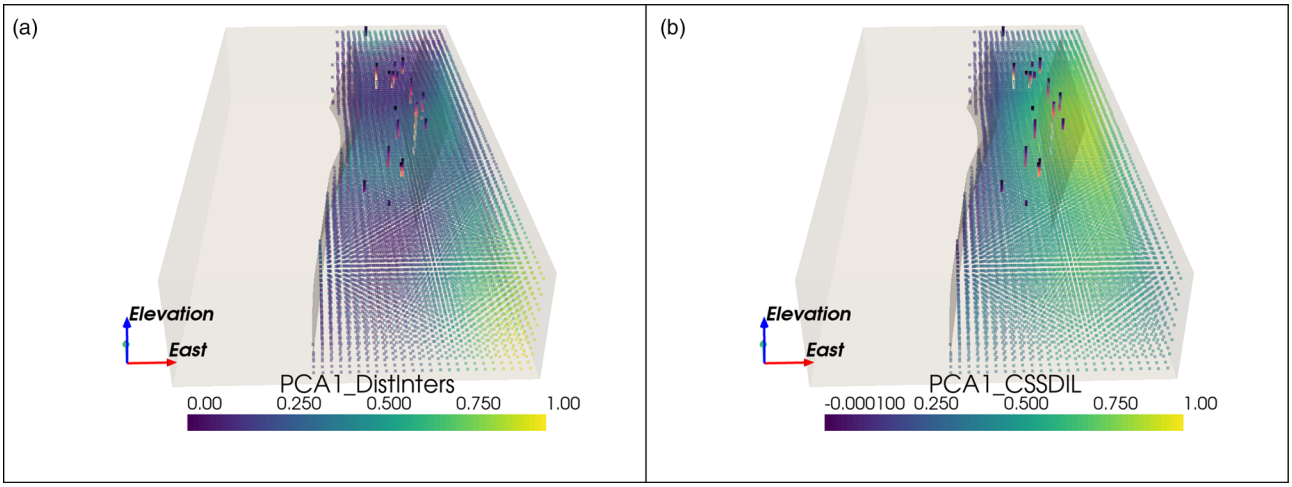


**Fig. 9.** The first principal component for the (**a**) fault proxies and (**b**) stress proxies.

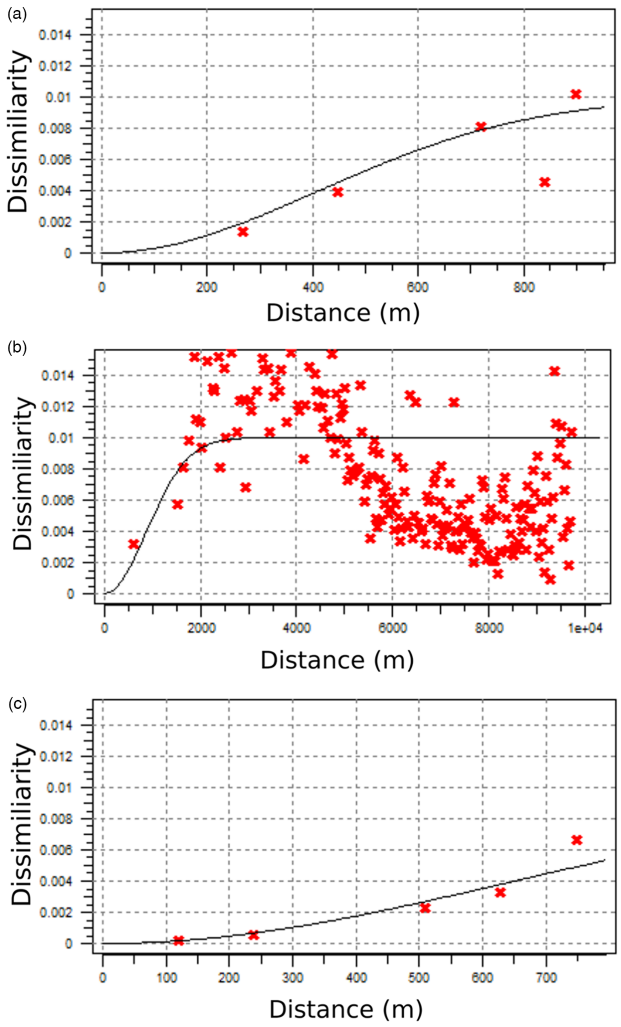


**Fig. 10.** Empirical (red x's) and model (solid black lines) variograms of PCA1 for normalized distance to fault & fault terminations/intersections (**a**) azimuth 270, dip −60, (**b**) north–southouth direction and (**c**) vertical direction.

them into normal scores first and then fitting:

$$t_1(\boldsymbol{u_\alpha}) = at_2(\boldsymbol{u_\alpha}) + b\,\text{depth}(\boldsymbol{u_\alpha}) + c \tag{11}$$

where $t_2$ represents the one of the four different geologic proxies at each location, $a$ and $b$ are the regression coefficients, and $c$ is the intercept. Also included in Table 6 are the statistical significance of each coefficient ($P$-values), where we see that temperature is not significant for Coulomb stress change, nor Dilation ($P$-value > 0). The LVM is then defined by

$$m^*(\boldsymbol{u}) = at_2(\boldsymbol{u}) + b\,\text{depth}(\boldsymbol{u}) + c \tag{12}$$

This should be considered an advantage of the LVM work, as it can provide a constraint to force the temperatures to increase with depth (i.e. a realistic scenario in a conduction based geothermal setting), whereas cosimulation alone will not. The LVM using distance to fault (DF, $x$-axis) is shown Figure 11. The points represent the colocated temperature measurements used in the regression at their depth and distance to fault locations.

## Example 3D realizations

This subsection provides the same five east–west ($X_{\min}$ to $X_{\max}$) slices for two of the 50 3D realizations as shown in Figure 7 for the temperature-only simulation. The slice views confirm the 60° dip of

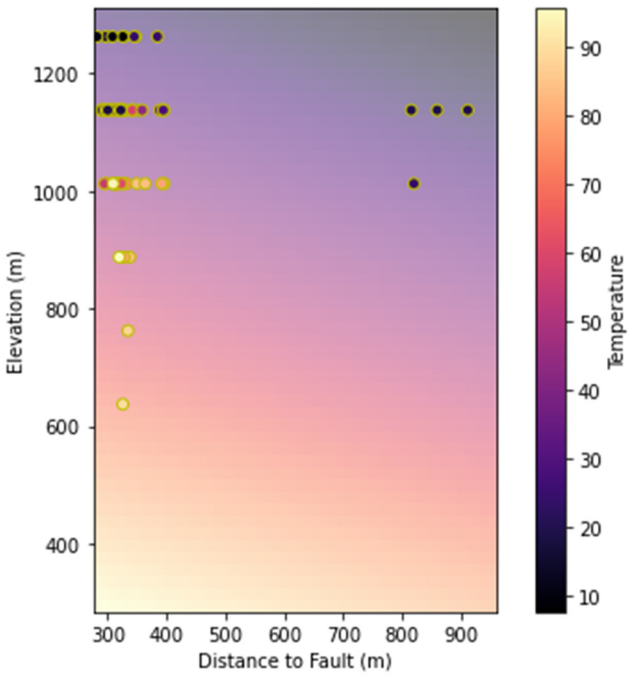


**Fig. 11.** Visual results of the LVM using the distance to fault (DF) secondary data and depth information, with the collocated temperature measurements shown in with scatter points: $LVM = -0.4\ DF + 0.13\ Depth + 204$.

the primary axis of the 3D variogram. Realizations 25 and 49 are shown for the following three groups: temperature data and PCA1 Fault in cosimulation, temperature data and PCA1 Stress in cosimulation, and finally temperature data with LVM defined with the four different proxies. As a reminder, the difference between the two realizations within each group is the order in which each grid cell is visited as outlined in the Entropy section and referred to as the 'random seed' in the geostatistical parlance. Therefore, the previously simulated values that affect all subsequent temperature draws will differ for each realization.

Figure 12 contains the models created with the temperature data *and* PCA1 Fault component proxies. The high temperature values are centred around the measured hot temperatures and are in shapes given the 3D variogram ellipsoids' orientation and size. The two realizations are consistent with each other, demonstrating that the secondary data has a constraining effect. The differences between the realizations are more pronounced the further away they are from temperature measurements.

Figure 13 has the same respective slices and realizations of Figure 12 but using the PCA1 for the Stress proxies. The minimum and maximum difference for all 50 realizations created by the PCA1 Fault and PCA1 Stress is on the order of $10^{-3}$°C. The differences in the realizations created from the different PCA1 component are highly influenced by the histogram transformation step. The histogram transformation takes a ranking of each simulated temperature and transforms it into a temperature according to the given distribution; this ends up squeezing the values into smaller spread. Differences between Fault and Stress PCA1 realizations

**Table 5.** Secondary variogram: PCA1 of distance to fault and fault intersections/terminations

| Gaussian sill = 0.01 | MAX | MEDIUM | MIN |
|---|---|---|---|
| Ranges (meters) | 3120 | 2200 | 877 |
| | Azimuth | Dip | Rake |
| Angles (degrees) | 270 | −60 | 0 |

**Table 6.** Linear regression parameters for solving for temperature with proxies and depth including their statistical significance (*P*-values)

| | Distance to fault/ *P*-values | Distance to terminations/intersections/ *P*-values | Coulomb stress change/ *P*-values | Dilation/*P*-values |
|---|---|---|---|---|
| Proxy coefficient (*a*) | −0.4/0.0 | −0.03/0 | −3.85/0.54 | $-1.72 \times 10^{+05}$/0.491 |
| Depth coefficient (*b*) | 0.13/0.016 | 0.16/0 | 0.14/0 | 0.12/0 |
| Intercept (*c*) | 204/0.0 | 262.5/0 | 198.9/0 | 181.8/0 |
| $R^2$ score | 0.62 | 0.68 | 0.57 | 0.56 |

could also be very small because the PCA are normalized between −1 and 1.

The same cross sections and realizations shown for the cosimulations (Figs 12, 13) are shown for the LVM simulations in Figure 14, where each row shows the respective LVM derived from the four different proxies: distance to fault, distance to fault intersections and terminations, Coulomb stress, and dilation tendency. The slice views again demonstrate the influence of the 60° dip of the primary axis of the 3D variogram. Compared to the cosimulation results, the LVM realizations place the hotter temperatures deeper in the subsurface, because of the depth being one of two variables used in the LVM linear regression. This is a more realistic scenario for a geothermal system (i.e. temperature generally increase with increasing depth), relative to the cosimulation results. Visually the same realizations across the different proxies (columns of Fig. 14) have subtle differences that are controlled by the locations of faults or the estimated stress properties.

## Vertical profiles at candidate well locations: temperature probability given 50 3D realizations

The 50 stochastic temperature models within each simulation group are extracted at the 9 candidate well locations, and the probability of

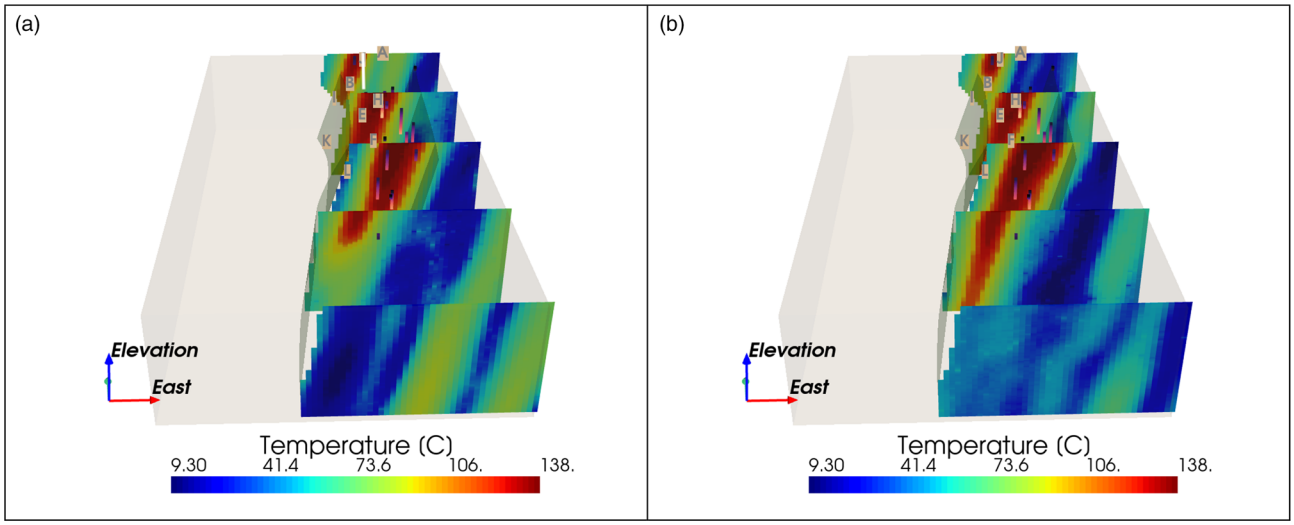


**Fig. 12.** Two example realizations using the PCA 1 of fault proxies: (**a**) realization 25 with hypothetical well J depicted in northernmost slice as white line, (**b**) realization 49.

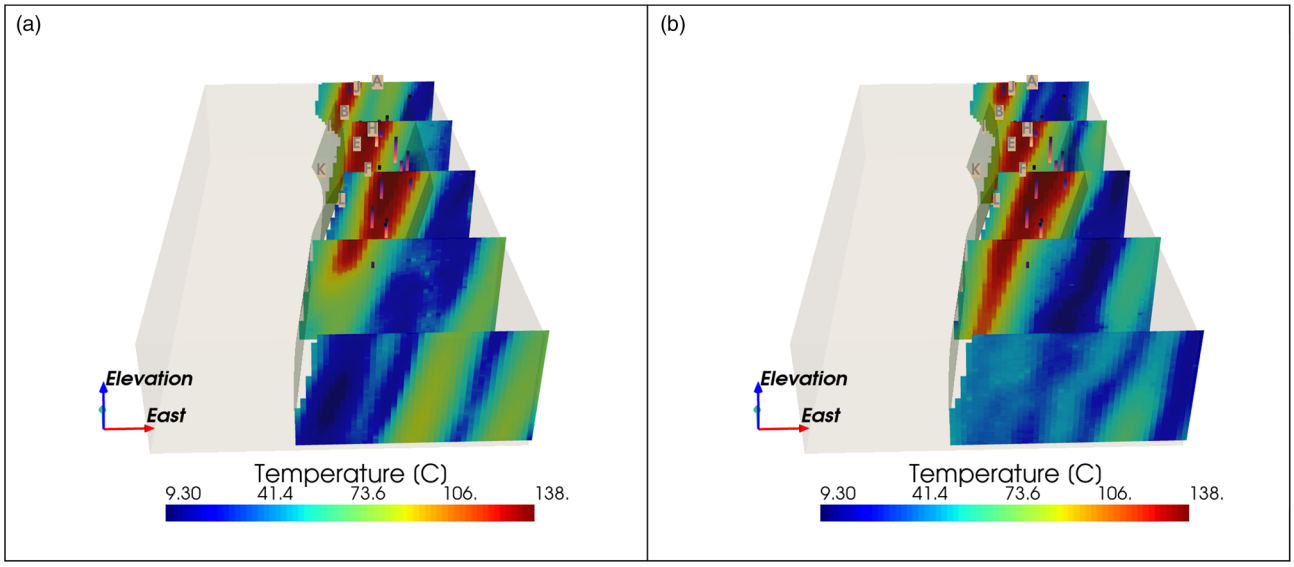


**Fig. 13.** Two example realizations using the PCA 1 of stress proxies: (**a**) realization 25, (**b**) realization 49.

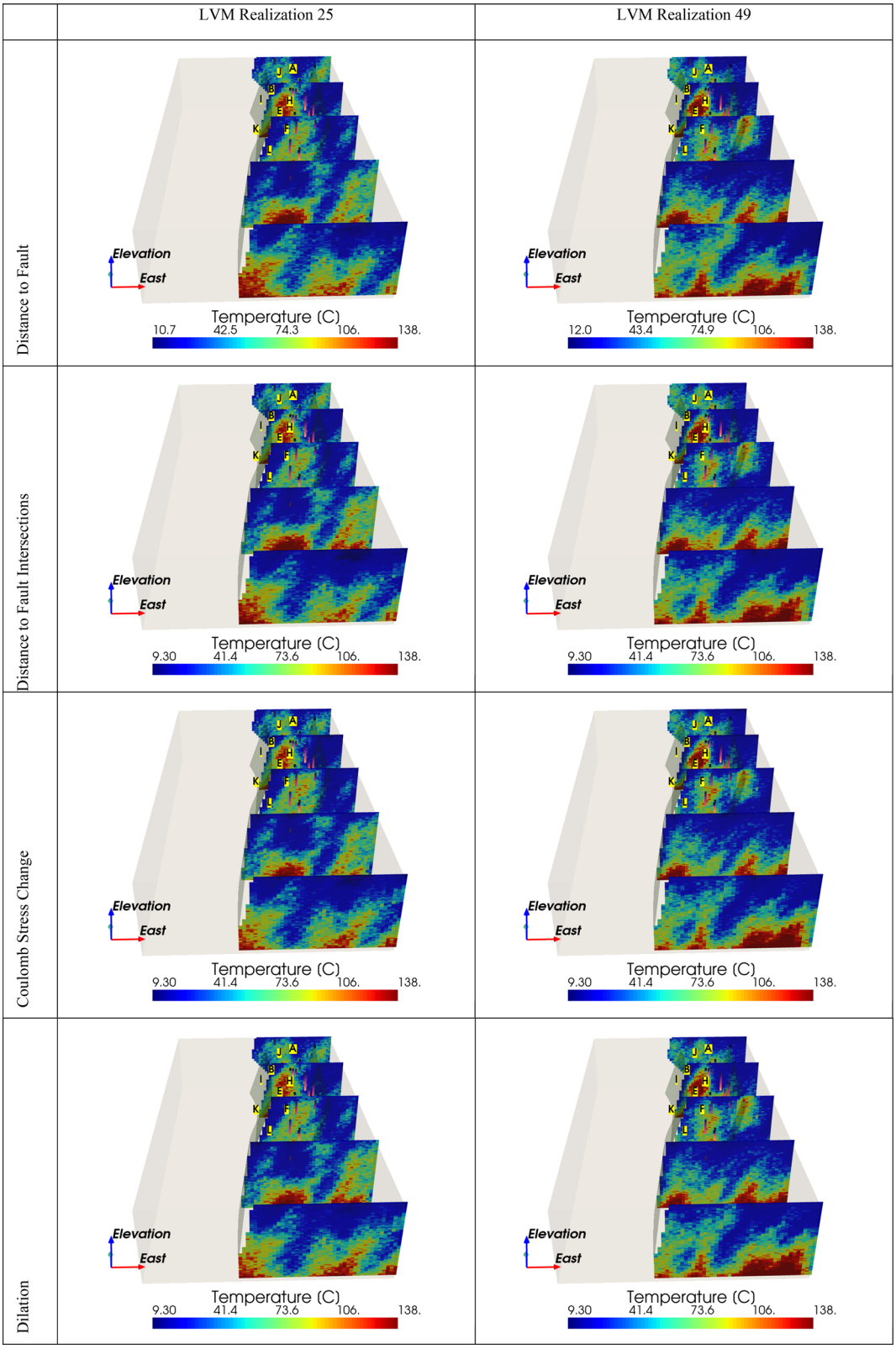


**Fig. 14.** Two example realizations of LVM results using regressions with the four different proxies: distance to fault (top row), distance to fault intersection (2nd row), Coulomb stress (3rd row), and dilation tendency (last row).

temperature bins is calculated for each elevation. These are shown in Figure 15 for the fault-based proxy set of cosimulation; the probabilities for the stress-based are extremely similar because the simulated temperatures have magnitude difference of $10^{-3}$°C as explained earlier and therefore not shown. The colour scales are set to minimum of 0 and maximum of 30% probability. Candidates I and K are only defined at the very bottom elevations because the well heads are on the west side of the western fault but intersect it at depth, so the upper sections of the wells are outside of the model volume. Also plotted for reference are the temperature profiles for realization 25 (magenta) and 49 (orange), which are the realizations show on Figures 13, 14 and 16. The two temperature profiles track close to each other, which explains why the probabilities are relatively tight. A dashed vertical line is plotted at 100°C. Candidates B, E, H, and L seem to have a pattern toward higher temperatures at depth, while A, F, and J do not. This can be

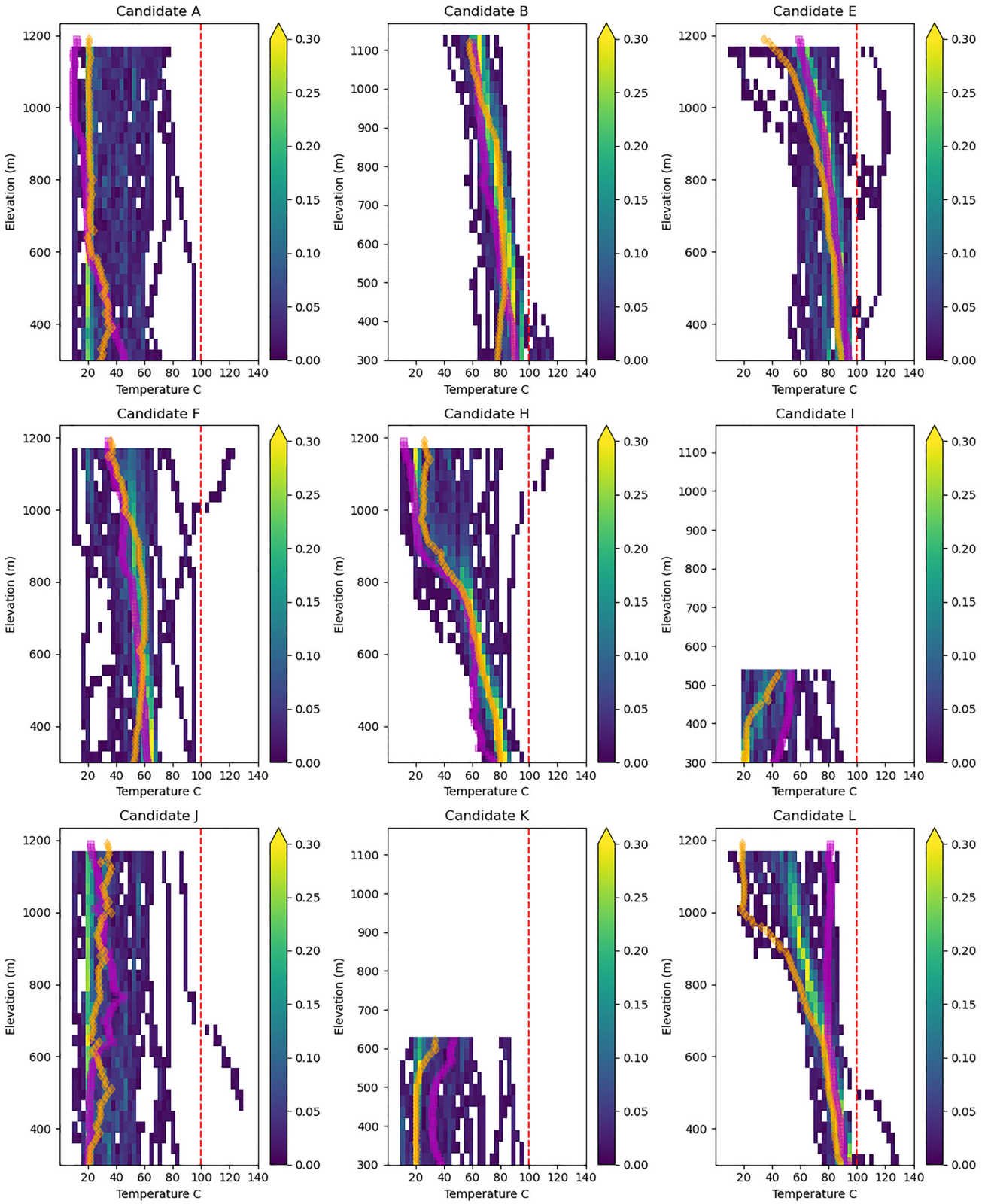


**Fig. 15.** The temperature probability profiles at the 9 candidate well locations for the cosimulation with fault-based proxies. Probability colour scale is 0–0.3 for all profiles. Magenta and orange lines are from realization 25 and 49, respectively. Left: PCA1 faults as secondary data; Right: PCA1 stress as secondary data.

explained by the hypothetical wells sampling the realizations vertically. For example, in Figure 12a, at the northernmost slice, a white line is depicting hypothetical well J. The white line intersects the high temperatures in the centre of the ellipsoid pattern then transverses in depth to the lower temperatures. This generally occurs at A and F as well. The probability patterns are very similar between the two proxy types, which is expected given how similar the realizations are (Figs 12, 13).

The vertical probability profiles for the four different LVM results are shown in Figure 16 with the same colour scale used in Figure 15. The temperature profiles for realizations 25 and 49 are plotted again, which unlike the cosimulation cases are different from each other and have a lot of discontinuity at short depth scales. This contributes to the probabilities being lower and spread out, but there is a slight increase through depth, due to depth being included in the regression calculation. The LVM based on distance to fault seems

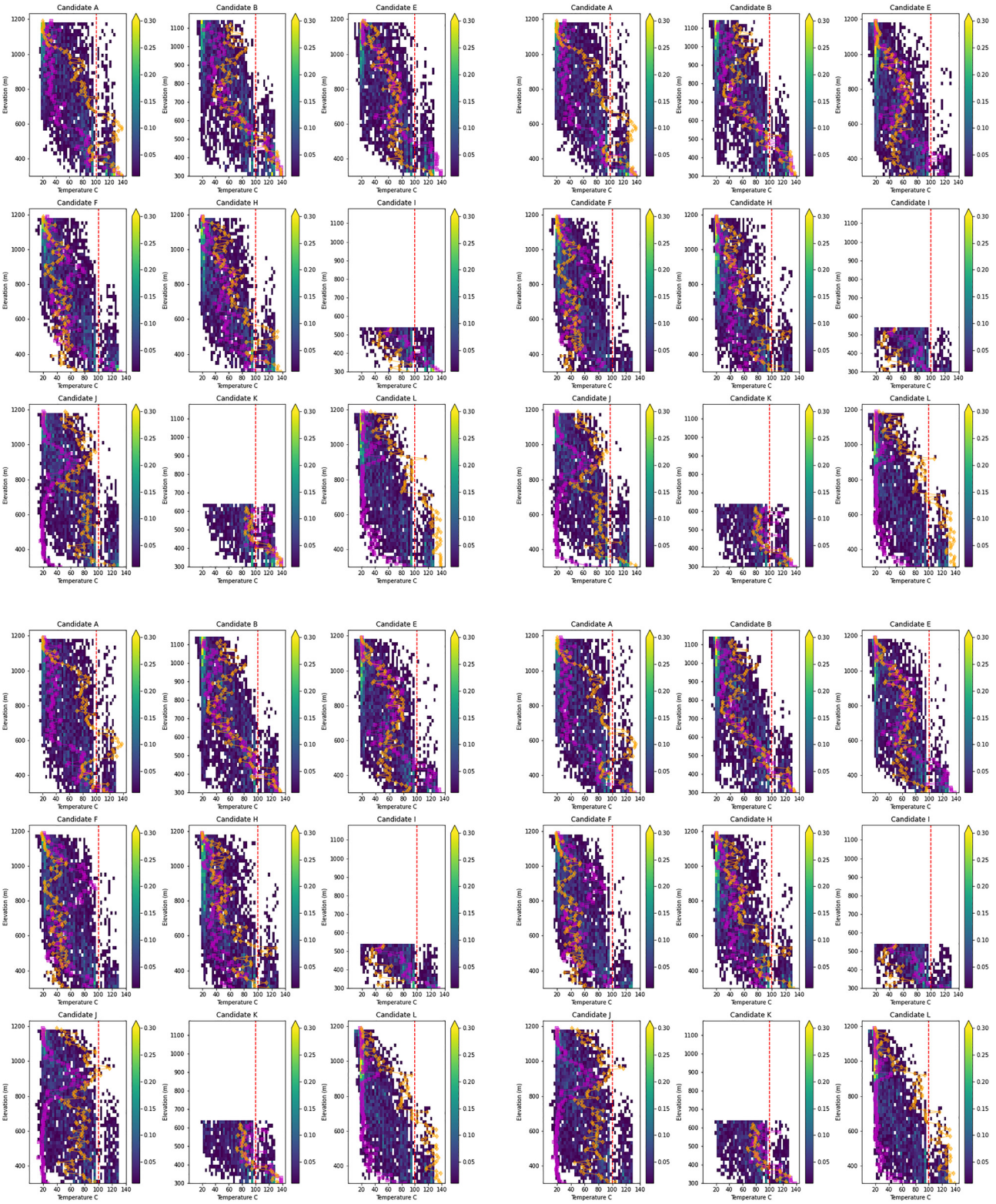


**Fig. 16.** LVM results: vertical probability profiles at 9 candidate wells. Magenta and orange lines are from realization 25 and 49, respectively. Top row: fault-based proxies. Bottom row: stress-based proxies.

to have slightly narrower probability range than the distance to fault intersections. In general, it's hard to distinguish the differences between the four LVM probability profiles, but $V_{\text{prior}}$ and entropy synthesize these into one number, making it easier to compare them.

## $V_{\text{prior}}$ & entropy

Each of the candidate temperature probability profiles from above are summarized into two quantities: $V_{\text{prior}}$ and entropy. The prior value is the expected outcome (weighted average) where higher temperatures have higher dollar amount ($), and entropy measures how distributed the probabilities are through the temperature bins (e.g. statistical disorder). First the cosimulation results are summarized, then the LVM results are presented for these two quantities. In order for an unbiased comparison of each profile, only the deepest ~250 m are used, which coincides with the simulated temperatures of Candidates I and K.

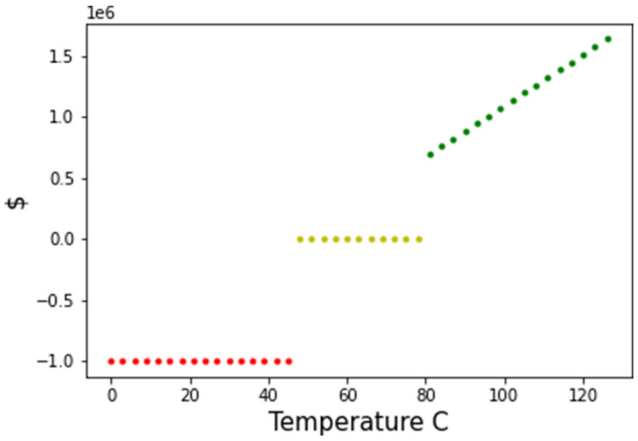


**Fig. 17.** Value outcomes for drilling for economic temperatures, where temperatures <45°C result in loss, <80°C is neutral, and >80°C is an increasing profit.

$V_{prior}$ gives the likelihood of a high temperature outcome within the Granite Springs Valley. $V_{prior}$ is calculated by using the probabilities of temperature and the nominal monetary amounts show in Figure 17 in equation (1) ($v_a(T)$). The important take away is that any utility function can be used and tailored for specific decision scenarios. Here we explain what the example monetary amounts could represent for our example utility function (Fig. 17). Temperatures less than 45°C result in a monetary loss. In this case the temperatures would be so low that the developer might abandon the prospect, hence a total monetary loss. If temperatures between 45°C and 80°C are encountered, these temperatures may not be high enough to utilize for electricity generation, but they may be indicative of local convective hydrothermal circulation, though not in that particular TGH. Therefore, this temperature outcome may have value in selecting the next drilling location(s). In this case the monetary result would be neutral since the data gained have some value. If temperatures greater than 80°C are encountered, there are profits to be made from geothermal development, and increasingly high temperature provides an increasing profit. Again, we emphasize that other values-temperature functions could be used to represent different risk attitudes and would change the $V_{prior}$ results shown in this section.

Table 7 contains the $V_{prior}$ values for all the candidate well locations for the PCA1 Fault. The rankings are denoted in parentheses. Candidates B, E, and L have the highest potential for high temperatures, e.g. these candidate locations have the top three $V_{prior}$ values and ranking. This can be explained by their orientation and relative proximity to measured temperatures greater than 90°C as shown in the last column of Table 7. Kriging system is such that in general the larger weights ($\lambda_i$) will be placed on the data values ($t(u)$) that align with the variogram's longest radius. The PCA1 Stress results are identical to those in Table 7; this is expected given the small differences in the stochastic models (Fig. 15), and the further averaging that occurs in the $V_{prior}$ calculation. Also of note is that four candidates have positive $V_{prior}$'s; those with negative $V_{prior}$ indicates they have a majority of temperatures <40°C across the realizations.

Table 8 contains the $V_{prior}$ results for the four different LVM groups, alongside the $V_{prior}$ from cosimulation; each are colour coded independently with dark green identifying higher values (in terms of $) and red as lower. Recall that $V_{prior}$ (and entropy) are calculated only using the deepest 250 m, where Candidates I and K are defined, and LVM constrains the temperature means to be hotter deeper. Therefore, in general, $V_{prior}$ is higher for the LVM than cosimulation results. Candidates A and B, on average, have the highest $V_{prior}$ of the four LVM, however, we can see they are different depending on which of the four proxies are used. For cosimulation, Candidates B, E, and L have the highest likelihoods for hot temperatures.

All candidate well locations are fairly close to a fault at these deeper depths, thus they are all green using 'distance to fault' as an LVM. Candidate J is ranked lowest for all LVMs except fault

**Table 7.** Cosimulation $V_{prior}$ and entropy results with PCA1 Fault/Stress proxies as secondary data

| | $V_{prior} \times 10^{+06}$ (rank) | Entropy (rank) | Map distance (meters) to nearest >90°C temperature | Azimuth to nearest >90°C temperature measurement |
|---|---|---|---|---|
| Candidate A | −9.37 (8) | 3.80 (1) | 1396 | 14 |
| Candidate B | 7.26 (1) | 2.68 (9) | 630 | 227 |
| Candidate E | 6.31 (3) | 3.33 (6) | 2124 | 262 |
| Candidate F | −1.27 (5) | 3.46 (5) | 3264 | 267 |
| Candidate H | 0.74 (4) | 3.16 (7) | 1416 | 178 |
| Candidate I | −5.99 (6) | 3.66 (3) | 1447 | 240 |
| Candidate J | −9.81 (9) | 3.72 (2) | 965 | 94 |
| Candidate K | −8.67 (7) | 3.48 (4) | 3398 | 254 |
| Candidate L | 6.37 (2) | 2.95 (8) | 4482 | 261 |

**Table 8.** $V_{prior}$ ($10 \times 10^{+6}$) values for 9 candidate thermal gradient well locations for cosimulation (left) and four LVM (4 right most columns)

| Candidate | Location | Cosim $V_{prior}$ | LVM fault distance | LVM fault terminations/intersections | LVM Coulomb stress | LVM dilations |
|---|---|---|---|---|---|---|
| A | north | −9.37 | 9.55 | 9.96 | 5.21 | 7.09 |
| B | on fault | 7.26 | 9.99 | 8.47 | 5.98 | 6.53 |
| E | east middle | 6.31 | 10.2 | 2.15 | 5.45 | 6.72 |
| F | east middle | −1.27 | 9.12 | 3.63 | 4.42 | 6.17 |
| H | east middle | 0.743 | 9.43 | 4.79 | 3.76 | 6.76 |
| I | west | −5.99 | 9.32 | 3.14 | 5.38 | 4.12 |
| J | north | −9.81 | 6.29 | 7.73 | 3.2 | 3.99 |
| K | west | −8.67 | 9.68 | 6.06 | 6.13 | 4.68 |
| L | south west | 6.37 | 8.48 | 7.14 | 5.45 | 4.78 |

terminations/intersections. Therefore, Candidate J is quite close to this geologic proxy and not the other three. Conversely, Candidate E is in the top three highest $V_{prior}$ for all except fault terminations/intersections. LVM results show relatively higher $V_{prior}$ for the western most wells (A, B, and K). If you consider Candidate J with intersections/terminations, it follows this pattern as well. It appears LVM, therefore, projects relatively high temperature down-dip (west) along the fault system despite wells A and K generally being location farther from the input temperature data (Table 7).

Entropy gives an indication of where new information could reduce statistical disorder the most. Entropy is calculated at each depth and averaged over the deepest 250 m of the profile (equation 2). Table 9 contains the entropy values and rankings for the cosimulation (again the same for both Fault and Stress proxies for the same reason in $V_{prior}$) and LVM results on independent colour scales (pink as highest and blue as lowest). For cosimulation, Candidate B has the lowest entropy due to its relatively close location to existing temperature measurements (Table 7). A, J, and I have the top three highest entropy. Candidates A and J are on the northern outskirts of the measured temperature logs while I is located due west (Fig. 3), therefore it is expected that they have higher entropies.

The LVM entropy values in Table 9 are higher than the cosimulation entropy values which can be attributed to the noisier spatial character of the LVM realizations. In the LVM results, Candidates J, L, and H have the highest entropies; H is surprising given that it is surrounded by temperature measurements. Therefore, this speaks of the proxy values being non-constraining or uninformative. The lowest are those using the distance to fault, which was the highest $V_{prior}$ results of the four LVM.

Comparing Tables 8 and 9, high entropy is never associated with high $V_{prior}$, since high entropy means there are equivalent probabilities in the lower and upper temperature classes, thus reducing weighting for $V_{prior}$ greatly. A combination of low entropy (dark blue) and low $V_{prior}$ (orange to red) should probably be ruled out for new wells as they represent more certain cooler temperature locations (e.g. locations E and F). If considering the LVM fault distance results, Candidate I could represent this scenario. Conversely, Candidate H seems to be an interesting location for providing more information since it has relatively high entropy for all LVM cases, including cosimulation.

## Discussion

The clear advantage of LVM is that it produces more physically realistic representations by imposing a realistic temperature/depth relationship, and it removes the decision of how to represent the spatial density of measurements since a temperature/depth value is needed at every grid location (exhaustively) and thus require interpolation onto the simulation grid. Conversely, the high spatial frequencies in the LVM realizations produced are not as physically realistic, and therefore favour the two cosimulation realization sets. Additionally, detailed studies should be performed to understand if anomalous convective fluid upwellings could be systematically overshadowed by the geothermal gradient imposed by including depth in the regression.

Because cosimulation uses specific locations, the density at which these data are represented in the simulation grid are very consequential. This is a recognized challenge and both Koch *et al.* (2014) and Boyd *et al.* (2020) tested the sensitivity of the final uncertainty measures (like entropy) to different decimation factors of the secondary data. We performed some sensitivity studies to understand this sensitivity with different densities. In general, the $V_{prior}$ and entropy values and rankings were similar until the secondary data were represented at the highest density (e.g. interpolated). In other words, when the secondary data are exhaustive, similar to how they are represented in the LVM realizations, the cosimulations have higher spatial frequencies (noisier characteristics) and therefore have higher entropy values like the LVM results. Although we found that the secondary data density was inconsequential to $V_{prior}$ and entropy for most scenarios, the sensitivity study provided an additional observation to the LVM results that exhaustive data can create the high spatial frequencies, despite the fact that a Gaussian variogram with no nugget effect was imposed. Like the LVM probabilities, cosimulation with exhaustive secondary data have spread out probability profiles. If cosimulation is pursued for future geothermal drilling decisions, methods must be developed to better choose the secondary data density.

Cokriging, the basis of both cosimulation and LVM, shares all the contributions and limitations of Kriging. Kriging provides a linear, least-squared error regression, combining data of diverse types accounting for their redundancy in space and their spatial correlation, as quantified by the variograms. The linear limitation of kriging may be more serious in cokriging, since cokriging would ignore any non-linear relation between two different attributes. We performed PCA transformation, which is considered a linear, orthogonal transformation. However, a non-linear transformation may be more appropriate for the four proxies given. Further analysis is needed to reveal if such non-linear dependence exists between the temperature data and the proxies. Additionally, combining all four proxies together and using their resulting PCA of these could be another avenue.

The ‘decision of stationarity’ is necessary for these geostatistical methods (Journel and Huijbregts 1978). This is a precarious assumption in geothermal where temperature patterns are changing from convective, outflow and conductive processes. Future geothermal geostatistical methods should consider how these geologic proxies may aide in determining which variogram is appropriate at different locations within our geostatistical simulation grid.

We would like to emphasize there are many worthy scientific and economic motivations for drilling geothermal exploration wells, such as defining edges of the system and trying to retrieve the best geochemistry samples. The parameterization and uncertainty model would look very different for these two goals. Modelling of just the boundaries of the system or the geochemistry differences would

**Table 9.** Entropy values for 9 candidate thermal gradiant well locations for cosimulation (left) and four LVM (4 right most columns)

| Candidate | Location | Cosim entropy | LVM fault distance | LVM fault terminations/intersections | LVM Coulomb stress | LVM dilations |
|---|---|---|---|---|---|---|
| A | north | 3.80 | 4.03 | 3.99 | 4.42 | 4.23 |
| B | on fault | 2.68 | 4.06 | 4.27 | 4.49 | 4.45 |
| E | east middle | 3.33 | 4.04 | 4.48 | 4.38 | 4.24 |
| F | east middle | 3.46 | 4.11 | 4.46 | 4.36 | 4.30 |
| H | east middle | 3.16 | 4.16 | 4.50 | 4.54 | 4.31 |
| I | west | 3.66 | 4.03 | 4.32 | 4.26 | 4.30 |
| J | north | 3.72 | 4.42 | 4.24 | 4.57 | 4.58 |
| K | west | 3.48 | 4.05 | 4.34 | 4.28 | 4.42 |
| L | south west | 2.95 | 4.27 | 4.42 | 4.46 | 4.51 |

need to be evaluated instead of continuous temperature. We believe that although the parameters would be different, the basic elements of this workflow could still be transferable for different drilling objectives.

## Conclusions

The aim of this study was to use 3D uncertainty modelling of temperature to guide the decision for where to place new temperature gradient wells, with the dual and not necessarily mutually exclusive purposes of confirming economic temperatures and/or reducing uncertainty about the conceptual understanding of the geothermal resource. For the current stage of exploration at Granite Springs Valley, the focus is on understanding the uncertainty of temperature in the subsurface, given the limited direct observations.

Two metrics were used to summarize the 3D stochastic temperature results address these two separate goals. $V_{\text{prior}}$ was defined to place a higher value on hotter temperatures (Fig. 17). Using high $V_{\text{prior}}$ to determine new wells sites addresses the goal to confirm economic temperatures. Higher entropy locations can be ideal targets for placing wells with the purpose of increasing understanding about the geothermal resource.

To produce these 3D temperature models, we performed variogram analysis of the measured data in combination with the 3D geological model to determine the orientation and radii of the 3D ellipsoid, where the major direction of spatial correlation is in line with the western fault in the vicinity of the temperature measurements. Sequential simulation was utilized to produce 50 different models (realizations) that reproduced the measured temperatures at their respective 3D locations and also accounted for the spatial correlations represented by the 3D variogram ellipsoid.

The innovation explored in this paper was the inclusion of secondary data, where the secondary data are derived from the 3D geologic modelling. The four temperature proxies were distance to fault, distance to fault terminations and intersections, Coulomb stress change, and dilation tendency. Two methods were explored for including this information: cosimulation (stochastic cokriging) and LVM. In practice, the LVM method fixes the mean temperature in a local neighbourhood, while cosimulation treats the secondary data as secondary constraints at specific locations.

For cosimulation, first the correlation coefficients between the observed temperature data and four proxies were calculated to eliminate the need to calculate the cross variogram. The fault and stress proxies had very high correlations with each other, respectively, but none had higher magnitude than −0.56 (dilation) with the measured temperatures. It was decided to combine the two fault and two stress proxies via PCA for cosimulation. Cosimulation results using fault v. stress PCA1 produced nearly identical temperature models after their histograms were transformed, and therefore their $V_{\text{prior}}$ and entropy summaries were equivalent. If the candidate locations were further away from the cluster of known temperatures, the influence of the secondary data will be more pronounced.

The LVM method requires an exhaustive representation of the mean temperature. The higher spatial frequencies again produce higher entropy values, but by including the depth in the linear regression and thus into the LVM, the temperature profiles generally increase with depth and therefore the $V_{\text{prior}}$'s are all positive. The $V_{\text{prior}}$ values and rankings across the four LVM groups can vary given the location of candidate's location to the geologic proxies. With the exception of Candidates I and J, the LVM $V_{\text{prior}}$ results show higher temperatures down dip of the western fault. But the $V_{\text{prior}}$ for Candidate J is high using the fault intersection/terminations. The cosimulation results are more influenced by the primary data and thus the primary variogram and less on the patterns of the geologic proxies.

An improvement to this workflow would include increasing the depth of the model. Depending on the decision objectives, it may be more useful to represent production depths, and investigate where the highest (extrapolated temperatures) are put in the 3D subsurface, but the extrapolation methods would only be useful if guided by the secondary data proxies. Although these depths are far from measured temperatures, the secondary data could guide the simulated temperatures.

**Acknowledgements** This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding was provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Geothermal Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

**Author contributions** **WT-G**: conceptualization (lead), data curation (lead), formal analysis (lead), funding acquisition (supporting), methodology (lead), project administration (equal), resources (lead), software (lead), validation (lead), visualization (lead), writing – original draft (lead), writing – review & editing (equal); **DS**: conceptualization (supporting), data curation (supporting), formal analysis (supporting), writing – review & editing (equal); **BA**: conceptualization (supporting), formal analysis (supporting), funding acquisition (lead), supervision (supporting), writing – review & editing (supporting)

**Funding** This work is part of the INGENIOUS project funded by the U.S. Department of Energy Geothermal Technologies Office (award number DE-EE0009254 to the University of Nevada, Reno).

**Competing interests** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Data availability** The datasets generated during and/or analysed during the current study are available in the Geothermal Data Repository repository, https://gdr.openei.org/submissions/969. The software used for modelling and visualization are open source (Remy *et al.* 2011; Sullivan and Trainor-Guitton 2019).

## References

Ayling, B. and Hinz, N. 2020. Developing a conceptual model and power capacity estimates for a low-temperature geothermal prospect with two chemically and thermally distinct reservoir compartments, Hawthorne, Nevada, USA. *Geothermics*, **87**, https://doi.org/10.1016/j.geothermics.2020.101870

Ayling, B., Kirby, S., Hardwick, C., Kleeber, E. and Trainor-Guitton, W. 2022. INGENIOUS Phase 1 (budget period 1) progress report.

Boyd, D.L., Walton, G. and Trainor-Guitton, W. 2020. Geostatistical estimation of Ground Class prior to and during excavation for the Caldecott Tunnel Fourth Bore project. *Tunnelling and Underground Space Technology*, **100**, 103391, https://doi.org/10.1016/j.tust.2020.103391

Chilès, J.-P. and Delfiner, P. 2012. *Geostatistics: Modeling Spatial Uncertainty*. 2nd edn, Wiley.

Craig, J.W., Faulds, J.E. *et al.* 2021. Discovery and analysis of a blind geothermal system in Southeastern Gabbs valley, Western Nevada, USA. *Geothermics*, **97**, https://doi.org/10.1016/j.geothermics.2021.102177

Cumming, W. 2016*a*. Resource capacity estimation using lognormal power density from producing fields and area from resource conceptual models; advantages, pitfalls and remedies. *In:* Proceedings of the 41st Workshop on Geothermal Reservoir Engineering Stanford University, 22–24 February, Stanford, California.

Cumming, W. 2016*b*. Resource Conceptual Models of Volcano-Hosted Geothermal Reservoirs for Exploration Well Targeting and Resource Capacity Assessment: Construction, Pitfalls and Challenges. *Geothermal Research Council Transactions*, **40**, 623–638.

De Lathauwer, L., De Moor, B. and Vandewalle, J. 2000. A multi- linear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, **21**, 1253–1278, https://doi.org/10.1137/S0895479896305696

Deutsch, C.V. and Journel, A.G. 1998. *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, New York.

Faulds, J.E., Coolbaugh, M.F., Benoit, D., Oppliger, G., Perkins, M., Moeck, I. and Drakos, P. 2010. Structural controls of geothermal activity in the northern Hot Springs Mountains, western Nevada: the tale of three geothermal systems (Brady's, Desert Peak, and Desert Queen). *Geothermal Resources Council Transactions*, **34**, 675–683, https://publications.mygeoenergynow.org/grc/1028722.pdf

Faulds, J.E., Hinz, N.H. *et al.* 2015*a*. Discovering Blind Geothermal Systems in the Great Basin Region: An Integrated Geologic and Geophysical Approach for Establishing Geothermal Play Fairways.

Faulds, J.E., Hinz, N.H. *et al.* 2015*b*. Integrated geologic and geophysical approach for establishing geothermal play fairways and discovering blind geothermal systems in the Great Basin Region, Western USA: a progress report. *Geothermal Research Council Transactions*, **39**, 691–700.

Faulds, J.E., Hinz, N.H. *et al.* 2017. Progress report on the Nevada play fairway project: integrated geological, geochemical, and geophysical analyses of possible new geothermal systems in the Great Basin Region. *In:* Proceedings, 42nd Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California.

Faulds, J., Hinz, N. *et al.* 2019. Vectoring into potential blind geothermal systems in the granite Springs Valley Area, Western Nevada: application of the play fairway analysis at multiple scales. *In:* Proceedings, 44th Workshop on Geothermal Reservoir Engineering Stanford University, 11–13 February, Stanford, California, 2019 SGP-TR-214.

Faulds, J.E., Hinz, N.H. *et al.* 2021. Discovering Blind Geothermal Systems in the Great Basin Region: An Integrated Geologic and Geophysical Approach for Establishing Geothermal Play Fairways: All Phases. *Technical Report*, https://doi.org/10.2172/1724080

Gloaguen, E., Marcotte, D., Chouteau, M. and Perroud, H. 2005. Borehole radar velocity inversion using cokriging and cosimulation. *Journal of Applied Geophysics*, **57**, 242–259, https://doi.org/10.1016/j.jappgeo.2005.01.001

Gloaguen, E., Lefebvre, R., Ballard, J.M., Paradis, D., Tremblay, L. and Michaud, Y. 2012. Inference of the two dimensional GPR velocity field using collocated cokriging of Direct Push permittivity and conductivity logs and GPR profiles. *Journal of Applied Geophysics*, **78**, 94–101, https://doi.org/10.1016/j.jappgeo.2011.10.015

Goovaerts, P. 2001. Geostatistical modelling of uncertainty in soil science. *Geoderma*, **103**, 3–26, https://doi.org/10.1016/S0016-7061(01)00067-2

Gringarten, E. and Deutsch, C.V. 2001. Teacher's aide variogram interpretation and modeling. *Mathematical Geology*, **33**, 507–534, https://doi.org/10.1023/a:1011093014141

Hansen, T.M., Journel, A.G., Tarantola, A. and Mosegaard, K. 2006. Linear inverse Gaussian theory and geostatistics. *Geophysics*, **71**, https://doi.org/10.1190/1.2345195

Isaaks, E. and Srivastava, R. 1989. *An Introduction to Applied Geostatistics*. Oxford Univ. Press, New York.

Journel, A. and Huijbregts, C. 1978. *Mining Geostatistics*. Academic Press, London.

Koch, J., He, X., Jensen, K.H. and Refsgaard, J.C. 2014. Challenges in conditioning a stochastic geological model of a heterogeneous glacial aquifer to a comprehensive soft data set. *Hydrology and Earth System Sciences*, **18**, 2907–2923, https://doi.org/10.5194/hess-18-2907-2014

Ma, X. and Journel, A.G. 1999. An expanded GSLIB cokriging program allowing for two Markov models. *Computers and Geosciences*, **25**, 627–639, https://doi.org/10.1016/S0098-3004(99)00009-6

Nevada Bureau of Mines and Geology 2017. Granite Springs Valley, Nevada Play Fairway Analysis - Well data and Temperature Survey [data set]. Retrieved from: Geothermal Data Repository, https://doi.org/10.15121/1452721

Remy, N., Boucher, A. and Wu, J. 2011. *Applied Geostatistics with SGeMS: A User's Guide*. Cambridge University Press.

Shannon, C.E. 1948. A mathematical theory of communication. *The Bell System Technical Journal*, **27**, 379–423, https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Siler, D.L. and Faulds, J.E. 2013. Three-dimensional geothermal fairway mapping: examples from the Western Great Basin, USA. *Geothermal Research Council Transactions*, **37**, 327–332, https://publications.mygeoenergynow.org/grc/1030589.pdf

Siler, D.L. and Pepin, J.D. 2021. 3-D geologic controls of hydrothermal fluid flow at brady geothermal field, Nevada, USA. *Geothermics*, **94**, 102112, https://doi.org/10.1016/j.geothermics.2021.102112

Siler, D.L., Hinz, N.H. and Faulds, J.E. 2018. Stress concentrations at structural discontinuities in active fault zones in the western United States: implications for permeability and fluid flow in geothermal fields. *Bulletin of the Geological Society of America*, **130**, 1273–1288, https://doi.org/10.1130/B31729.1

Sullivan, C. and Trainor-Guitton, W. 2019. PVGeo: an open-source Python package for geoscientific visualization in VTK and ParaView. *Journal of Open Source Software*, **4**, 1451, https://doi.org/10.21105/joss.01451

Webster, R. and Oliver, M.A. 2008. *Geostatistics for Environmental Scientists*. 2nd edn., 1–315.

Williams, C. and DeAngelo, C. 2011. Evaluation of approaches and associated uncertainties in the estimation of temperatures in the upper crust of the western United States. *Geothermal Research Council Transactions*, **35**, 1599–1606, https://publications.mygeoenergynow.org/grc/1029460.pdf

Witter, J.B., Trainor-Guitton, W.J. and Siler, D.L. 2019. Uncertainty and risk evaluation during the exploration stage of geothermal development: a review. *Geothermics*, **78**, 233–242, https://doi.org/10.1016/j.geothermics.2018.12.011

Zhu, H. and Journel, A.G. 1993. Formating and integrating soft data: stochastic imaging via the Markov-Bayes Algorithm. *In:* Soares, A. (ed.) *Geostatistics Troia '92*. Kluwer Acad., Norwell, Mass., 1–11.