PROCEEDINGS, Thirty-Seventh Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, January 30 - February 1, 2012
SGP-TR-194

EFFECT OF GEOLOGICAL DATA QUALITY ON UNCERTAINTIES IN GEOLOGICAL MODELS AND SUBSURFACE FLOW FIELDS

J. Florian WELLMANN and Klaus REGENAUER-LIEB

Western Australian Geothermal Centre of Excellence
CSIRO Earth Science and Resource Engineering
26 Dick Perry Avenue
Kensington, WA, 6151, Australia
e-mail: florian.wellmann@csiro.au

ABSTRACT

Geological data at the depth of geothermal resources and reservoirs are rare and of varying quality. In order to evaluate how the structural geological model and simulations of the hydrothermal flow field are affected by the geological data quality, meaningful measures are required to characterize these systems. We present here system-based thermodynamic measures to classify uncertainties in geological models and in geothermal flow fields. Information entropy is proposed to evaluate uncertainties in geological models, and thermal entropy production is proposed to analyze uncertainties related to hydrothermal flow. As these measures have a fundamental theoretical basis and are related to the internal state of the system, they can be interpreted quantitatively and, consequently, give uncertainties a meaning.

Information entropy values are directly related to the state of uncertainty of a geological model. For a point within the model, information entropy is a measure of the minimum number of geological units that could occur at its location. If the information entropy is zero, only one unit is possible and no uncertainty exists. If the information entropy value is greater than zero, at least two units are probable. If it increases above 1, three units can occur. In general the measure provides a weight of probability for different states. An advantage of the method is that it gives an entropy measure for the state of the entire model and therefore lends itself as a robust measure to quantitatively compare uncertainties in difference models.

In a similar sense, the thermal entropy production provides a quantitative measure of the thermo-dynamic state of a hydrothermal system. When the entropy production is zero, the system must be in a conductive steady state for a closed system. If the entropy production is larger than zero, the system can be in a convective or transient conductive state. For higher values of entropy production, the convective units show higher complexities and, hence, the uncertainty of the hydrothermal field increases. Moreover, the average model entropy production gives a measure of the convective vigor that can be expected in the system. This is directly related to the efficiency of heat transfer over the system. The measure is therefore not only useful for comparison of different models, but also has a quantitative meaning for the productivity of heat that can be harvested from a particular setting.

We present an application of both measures for a complex case study to investigate the influence of geological data quality on the uncertainty of geological model and geothermal flow predictions. This analysis has only been possible due to a newly developed workflow that integrates geological modeling and geothermal flow simulations.

Our application to the realistic case study confirms the key hypothesis that the geological uncertainty and the flow uncertainties can be subsumed in two interrelated measures: (a) information entropy, and (b) the spread of the thermal entropy. Since the measures provide single values that characterize uncertainties, they provide a promising path for physically based data compression in data intensive geothermal modeling.

INTRODUCTION

Structural geological models are commonly used to determine the property distribution for flow simulations. Geological models always contain uncertainties, so an important question is: *If the geological model is uncertain, how uncertain are the simulated flow fields?*

Many standard methods exist to evaluate uncertainties of specific observables in reservoir simulations, for example temperatures at an observation point or oil production in a well

(e.g. Finsterle, 2004; Suzuki, 2008; Bundschuh, 2010). However, uncertainties at observation points are not necessarily suitable to analyze the uncertainty within the whole system. We test here the hypothesis that system measures based on the concept of entropy can be applied to analyze and compare system-wide uncertainties in geological models and in simulated hydrothermal flow fields. These measures can then be applied to evaluate the correlation of uncertainties in both systems.

We will first outline the concepts of information entropy, used here to quantify uncertainties in a set of structural geological models, and thermal entropy production, applied to evaluate the thermodynamic state of a simulated flow field. We will then outline a specifically developed workflow that enables the automatic simulation of a set of hydrothermal flow fields from stochastically generated realizations of structural geological models. Applying this workflow, we will then use both system-based measures to evaluate uncertainties in a concrete example of a geothermal resource-scale study in the North Perth Basin, Western Australia.

SYSTEM-BASED MEASURES

Information entropy

We apply information entropy as a quantitative measure for the quality of a structural geological model, as introduced in Wellmann (2011). Information entropy is based on the Shannon Entropy model (Shannon, 1948), originally defined to evaluate the amount of missing information in a transmitted message. We apply the concept in the context of structural geological models to describe the information entropy for a point in time and space:

$$H(x,t) = \sum_{m=1}^{M} p_m(x,t) \log p_m(x,t) \qquad (1)$$

The quantitative interpretation of the measure is straightforward and can be illustrated with a simple example of the probability of encountering several geological units in a borehole (fig. 1). The data in this figure is derived from stochastic realizations of a structural geological model where the exact position of a structural boundary at depth is uncertain. The left graph in fig. 1 shows the probability of encountering one specific geological unit at a potential drilling site at depth. At around sea level near the top of the figure, the probability of encountering this formation is low, according to the stochastically generated set of structural models. Below -1000 meters, the unit is present in all realizations and the probability of finding it is 1. Below -3000, the probability decreases and below -6000 meters, the unit was not observed in any realizations and the probability is accordingly 0 again.

The probability representation for the occurrence of one unit is a practical way to visualize uncertainties for a single unit; however, it quickly becomes confusing when more than one unit are analysed (middle graph in fig. 1).

Calculating the information entropy from all combined probabilities provides a clear picture of uncertainties in the model (right graph in fig. 1). Where one unit has a probability of 1, H is zero: there is no uncertainty (horizontal label “A”). For every value of H > 0, at least two units can occur. If two values are exactly equally probable, then H = 1 (label “B”), and accordingly for higher values: at label “C”, 4 units are equally probable.

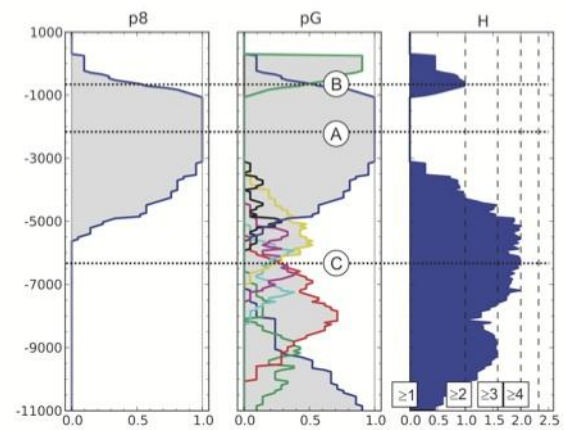

Figure 1: Example of the application of information entropy to evaluate uncertainties in a hypothetical drill hole (ordinate axis is altitude in metres). Left: probability to encounter one specific geological unit. Middle: probability for all geological units in the mode. Right graph: information entropy at depth, combining the probability occurrences of all units into one meaningful measure (from Wellmann, 2012).

The extension of the concept, illustrated on a vertical transect above, into a spatial context is straightforward (Goodchild, 1994). As an example, we consider a geological map, consisting of three geological units, where the exact position of the boundary between the three units is uncertain (fig. 2a). The uncertainty about the position of the boundaries can be transferred into probability maps for each of the units (fig. 2b). Here we show discrete maps with probability values assigned to grid cells for each of the geological units. Eq. (1) can then be used to calculate the information entropy of each of

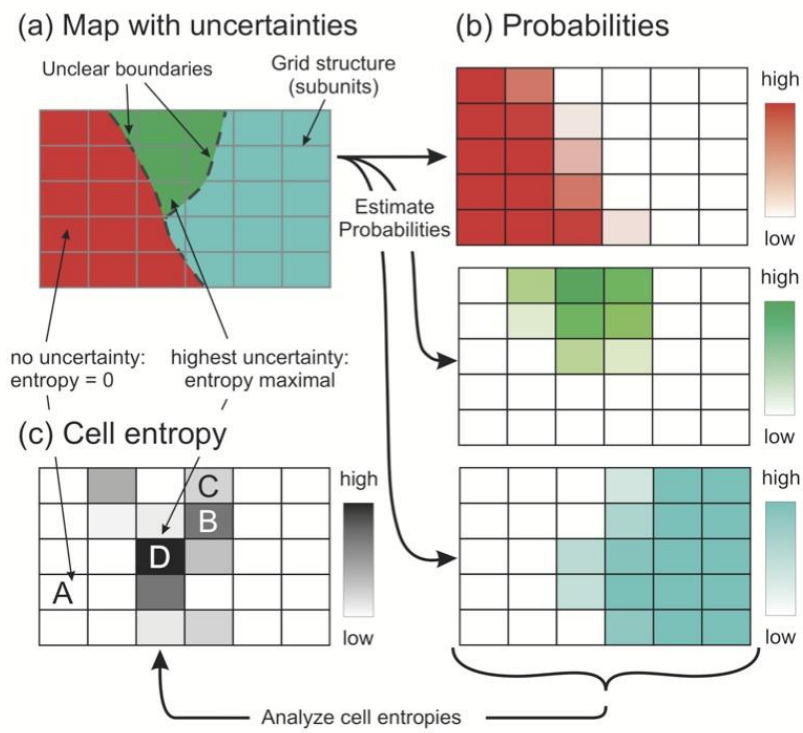

Figure 2: Information entropy in a spatial context: (a) map of three geological units with uncertain boundaries; (b) discreet probability maps for each of the units; (c) information entropy for each cell, derived from the cell probabilities: the entropy is zero when one unit has a probability of 1 at this cell (e.g. label "A"); the value increases when more than one unit is probable and highest when all units are almost equally probable (label "D"). Figure from Wellmann (2011a).

these cells. The map of cell entropies (fig. 2c) provides a clear representation of the spatial uncertainties, similar to the borehole example of figure 1. Cells where one unit has a probability of 1 (e.g. at label "A") have a cell information entropy of zero because no uncertainty exists at this point. Cells where more than one unit are probable have a higher information entropy (label "B" and "C") and the highest entropy exists where all three units are almost equally probable (label "D"), reflecting the highest uncertainty at this point.

The previous examples highlight how information entropy can be used to visualize uncertainties in a meaningful way. However, in order to quantify the overall uncertainties in the whole model space, an additional measure is required.

As an extension of the concept of information entropy at a point, the total information entropy for the whole model space can be defined as a sum:

$$\begin{aligned} H_T(t) &= -\frac{1}{N}\sum_{x=1}^{N} H(x,t) \\ &= -\frac{1}{N}\sum_{x=1}^{N}\sum_{m=1}^{M} p_m(x,t)\log p_m(x,t) \end{aligned} \tag{2}$$

In accordance with the cell entropy, the total information entropy provides a single measure of uncertainties in the entire model: if the geological units are exactly known everywhere, the entropy is zero. For an increasing number of uncertain cells, the total information entropy of the model increases, capturing the increase in uncertainty.

Thermal entropy production

Thermodynamic measures can be applied to describe the state and to predict the response of a system, without having to know all detailed processes within the system. The thermodynamic measure of entropy production is related to dissipative heat processes within a system. The entropy of a diabatic system changes if heat is supplied or removed from the system. The entropy production, the change of

entropy, is defined as the ratio between the change in heat Q and the temperature T (e.g. Callen, 1985):

$$\dot{S} \equiv \frac{\delta Q}{T} \tag{3}$$

Entropy is produced due to reversible and irreversible processes. If we only consider the entropy production due to thermal dissipation in a slow moving fluid in a permeable matrix, the entropy production $\dot{S}$ is reduced to the heat that is supplied normal to the boundary A by the heat flux q_h at a temperature T (Ozawa et al., 2003):

$$\dot{S} = \oint_A \frac{1}{T} q_h \cdot n \, \mathrm{d}A \tag{4}$$

In a conductive system in steady state, all heat fluxes are balanced and no thermal entropy is produced. The situation is different in an advective system. Here, advective heat transport leads locally to an increase in entropy production.

A simple convection system is presented in fig. 3. The black and white background image shows a typical convection temperature profile where darker colours correspond here to hotter temperatures. The white arrows indicate dominating fluid flow, disturbing the temperature field. In the central part of this model, colder fluids are transported downwards by this fluid flow (downwelling zone). These colder fluid parcels induce locally a conductive temperature flux from the adjacent warmer areas (subfigure in fig. 3) and this flux results in a non-zero entropy production of the system, even if it is in steady state. The analogue behaviour exists in the upwelling zone.

Entropy production can therefore be related to the heat transport mechanisms within a system: if a system is in steady state, the entropy production is zero if heat transport is purely conductive. As soon as convection sets in, the entropy production is greater than zero and increasing for more vigorous convective systems. In fact, the entropy production is directly related to the Nusselt number, a measure of the heat transfer (Regenauer-Lieb, 2010).

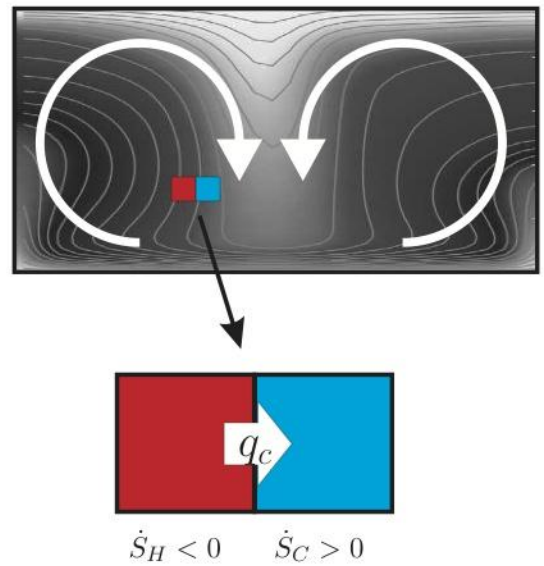

Figure 3: entropy production in a convective system: the advective transport of relatively cold fluid parcels downwards induces locally a conductive heat transport, resulting in a non-zero entropy production.

In order to obtain a measure of the state of the entire system, we apply the average specific entropy production, calculated as the average value of the specific entropy productions in subsystems, scaled by the mass:

$$\langle \dot{s} \rangle = \frac{1}{V} \int_V \frac{\dot{S}}{m} \, \mathrm{d}V \tag{5}$$

This measure provides an insight into the state of the entire system and classifies it with a single number. As an example, in figure 4, the development of entropy production over time is presented for the onset of convection in a simple homogeneous system. Plotted here is the average entropy production of the entire system. Until convection sets in, the entropy production is zero. It then increases to a maximum value (at 3750 years), but then decreases again to a finite non-zero value when the convective system reaches a steady state.

This example indicates that thermal entropy production can be used to classify the hydrothermal state of the system with one meaningful value. We will use this value below to evaluate how a system reacts to changes in the constraining parameters.

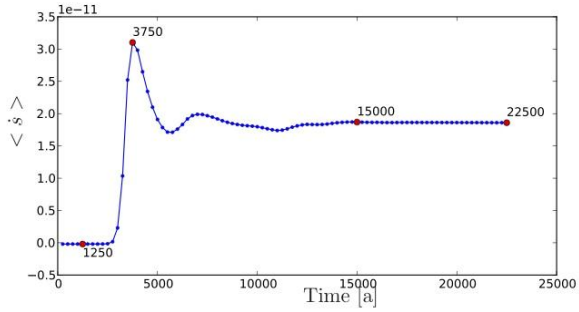

Figure 4: Average specific entropy production during the onset of convection in a simple system; From an initial conductive steady state, the system goes through a phase of high entropy production until it reaches a convective equilibrium state with a finite non-zero entropy production.

STRUCTURAL UNCERTAINTY WORKFLOW

Our aim is to evaluate the influence of uncertainties in structural geological models on simulated flow predictions for reasonably complex and realistic geological scenarios. We developed a method that enables stochastic geological model generation and the direct simulation of hydrothermal flow fields for all realizations of the geological model.

The standard procedure to simulate a hydrothermal flow field for a realistic geological setting is outlined in figure 5: starting from discrete geological input data (e.g. well observations of surface contacts or picks of seismic reflectors), a continuous geological model is constructed. This model is then transferred into a discrete version for the numerical process simulation. Then, relevant properties have to be assigned to the geological units (e.g. thermal conductivity, permeability) and boundary conditions have to be assigned. Based on this information, the hydrothermal simulation can be performed.

A problem with this standard procedure is that several of the steps shown in figure 5 require manual interaction. Specifically the steps of construction of the geological model ("Step 1"), and the discretization or meshing ("Step 2") are not automated in standard methods.

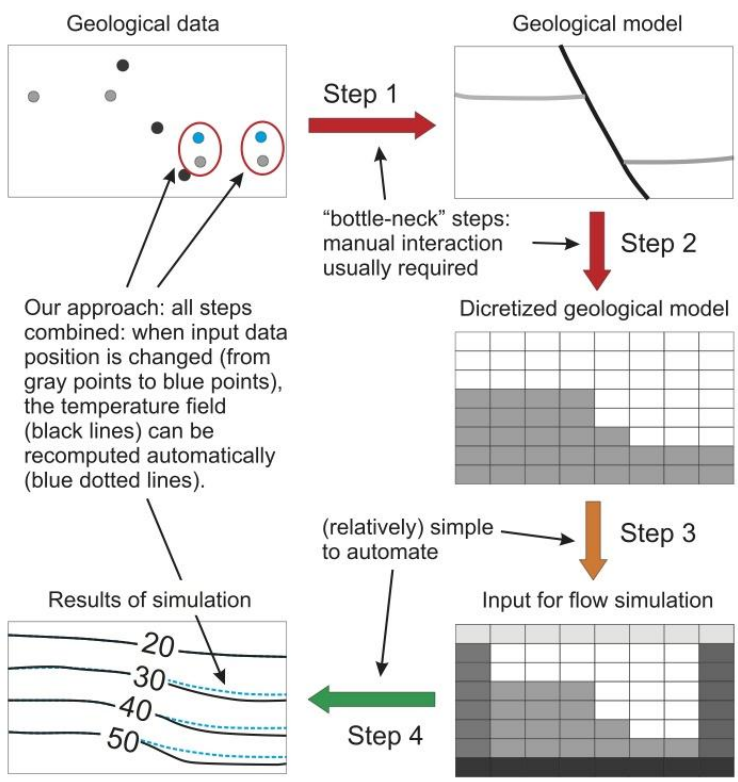

Figure 5: workflow from geological input data (here: surface contact points between different geological units) to simulated flow fields (here: tempterature) (Wellmann, 2011b).

To overcome this limitation, we developed a workflow that integrates all aspects into one framework. We apply several commercial codes for geological modeling and hydrothermal flow simulations and our own programs for meshing and input file generation:

- We apply a geological potential-field method (implemented in the software GeoModeller, Calcagno, 2009) to automate the geological model construction;
- Mesh discretization and input file generation for subsequent hydrothermal simulations is automated with the Python modules PyTOUGH and PySHEMAT (Wellmann, 2011);
- Hydrothermal flow simulations are performed with SHEMAT (Clauser, 2003) or TOUGH2 (Pruess, 1991).

All of these steps are automated and integrated into one Python framework. With this workflow, it is directly possible to evaluate how a change in the geological input data set influences the simulated flow fields (see fig. 5). Combining this workflow with a stochastic geological modeling method (Wellmann, 2010), we obtain a method to generate multiple geological model realizations and the subsequent flow simulations for these models automatically.

CASE STUDY: NORTH PERTH BASIN, WESTERN AUSTRALIA

We present an application of both measures, information entropy and thermal entropy production, for a complex case study to investigate the influence of geological data quality on the uncertainty of geological model and flow predictions. Our case study is a geothermal resource area in the North Perth Basin, Western Australia (fig. 6, left). The geological setting is a half-graben structure with a deep sedimentary basin (> 10 km). As seismic resolution is poor in the region and few wells penetrate deep into the basin, two main structural uncertainties exist in this region (Mory, 2006):

1) The exact position of the geological surfaces at depth is not well known;
2) Additional faults might exist in the basin.

Based on these two main types of uncertainty, we evaluate two uncertainty scenarios: in the first case (fig. 6, standard scenario), only the position of geological surfaces at depth is considered uncertain. In the second case (fig. 6, fault scenario), additional normal faults are introduced, in accordance with the general tectonic setting.

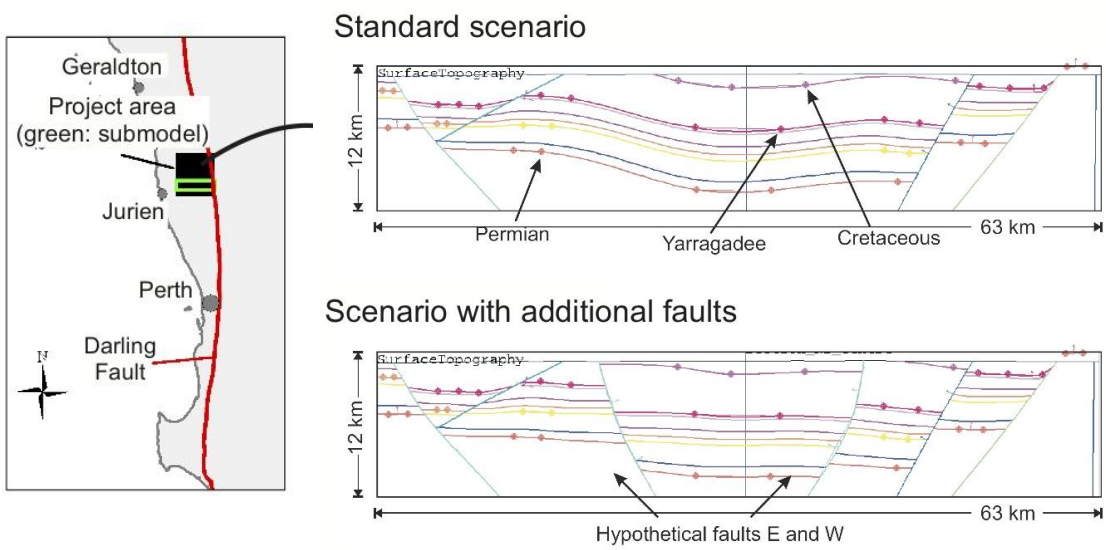

Figure 6: Left figure: Location of the model area in the North Perth Basin, Australia; the simulated hydrothermal model is a full 3-D submodel (green) of a regional scale geological model; Right: vertical section through the model showing geological structures and control points for the standard scenario and a scenario with additional faults in the basin.

The aim of the study is to evaluate for these scenarios how decreasing accuracy in the initial geological data affects the simulated hydrothermal flow fields. A Gaussian error was assumed for all data points in fig. 6. Standard deviations are increasing with depth, from 50m to 1000m for the high accuracy data case, to 150m to 2000m for the low accuracy data case (see table 1).

Table 1: standard deviations assigned to the geological surface contact points in meters for the scenarios with (1 A-C) and without (2 A-C) faults (fig. 6); The sub-scenarios A-C simulate the effect of decreasing geological data quality.

Scenario	Shallow	Deep	Add. faults
1 A	50	300	n/a
1 B	100	600	n/a
1 C	150	900	n/a
2 A	50	300	1000
2 B	100	600	1500
2 C	150	900	2000

For each of these scenarios, 20 realizations of the structural geological model were created, and the coupled hydrothermal flow field was simulated for each of these structural realizations with the workflow presented above. The geological model is discretized into a regular grid with 250 x 25 x 200 cells. The model has a size of 63km x 15km x 12km. Flow boundary conditions are no-flow at all boundaries. Thermal boundary conditions are basal heat flux, fixed annual mean temperature at the top and no flow at lateral boundaries. For more details about the hydrothermal simulation itself, see Wellmann (2011b).

For each of these scenarios, the information entropy of the structural geological model and thermal entropy productions of each of the realizations for a simulation time of 100,000 years are calculated.

The cell information entropy for the low data accuracy case of both scenarios is presented in figure 7. The increase of uncertainties in the structural model with depth is clearly visible, as well as the influence of the additional faults.

Figure 8 shows the development of the average specific thermal entropy production over a simulation time of 100,000 years. Presented here are the results for the scenario with additional faults: the high data accuracy (Scenario 2A) and the low data accuracy (Scenario 2C) case. Initially, all models are in a conductive steady state and entropy production is accordingly zero. Then, convection sets in and the entropy production increases to a maximal value. After approximately 60,000 years, the hydrothermal systems appear to reach a finite entropy production states. It is clearly visible that the average specific entropy production values show a higher variability for the cases of low data accuracy. This behavior is specifically obvious in the initial phase where convection sets in (around 10,000 years).

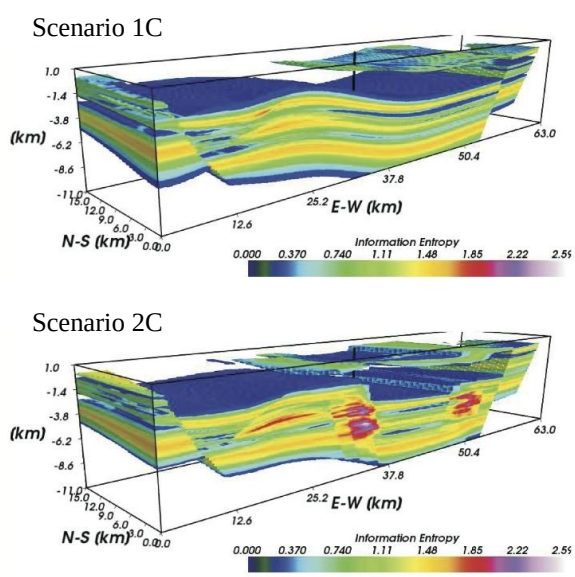

Figure 7: visualization of cell information entropies for both scenarios and the case of low data accuracy (1C and 2C), filtered view for values of H > 0.

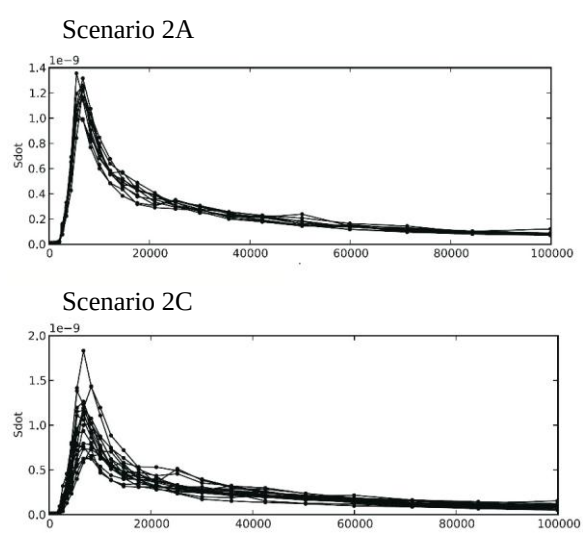

Figure 8: average specific thermal entropy production during the equilibration phase of hydrothermal simulations for scenarios 2A (high data quality) and 2C (low data quality).

We now apply the total information entropy (eq. 2) and the spread (25th to 75th percentile) of the average specific entropy production (eq. 5) of the final simulation time step as measures of the system uncertainties. Results are presented in figure 9. A comparison suggests that the overall uncertainties for both scenarios are similar, and that uncertainties increase with reduced data quality. It is interesting to note that, from scenario 2A to 2B, the uncertainty in the geological model increases, whereas the flow field variation does not significantly increase. A possible interpretation is that the additional faults in the basin have a stabilizing effect on the flow fields (e.g. Garibaldi, 2010).

DISCUSSION

The application of information entropy and thermal entropy production as system-based measures shows that they can be used as meaningful measures to interpret and compare uncertainties in subsurface geology and simulated flow fields. With both concepts, it is possible to derive a scalar value that is related to the state of uncertainty within each system. The total information entropy of a geological model represents the average uncertainty for the occurrence of a geological unit at each point in the model. The average specific thermal entropy production is directly related to the dominating heat transfer mechanisms in the hydrothermal model.

We applied the measures here in a case study to compare the influence of geological data accuracy on geological model uncertainty and flow field variability. The results show that both the geological model and the simulated subsurface flow fields are affected by a varying quality of geological input data. In fact, in this case uncertainties in the geological model and in the simulated flow fields respond in a similar way to decreasing accuracy of the geological input data. However, the difference in the flow field variation response for the two different geological scenarios indicates that geological structures have an important controlling effect on the flow behavior.

The ability to describe the state of uncertainty in a system with a meaningful scalar value provides a way forward for physically based data compression as geological modeling and geothermal simulations are becoming more data-intensive. In future work, we will explore the possibility to integrate these values in stochastic inversion methods.

The comparison of geological uncertainties and flow field variation as a function of geological data quality has only been possible with the development of a dedicated workflow combining geological modeling and geothermal flow simulations. Even though sophisticated methods already exist that combine

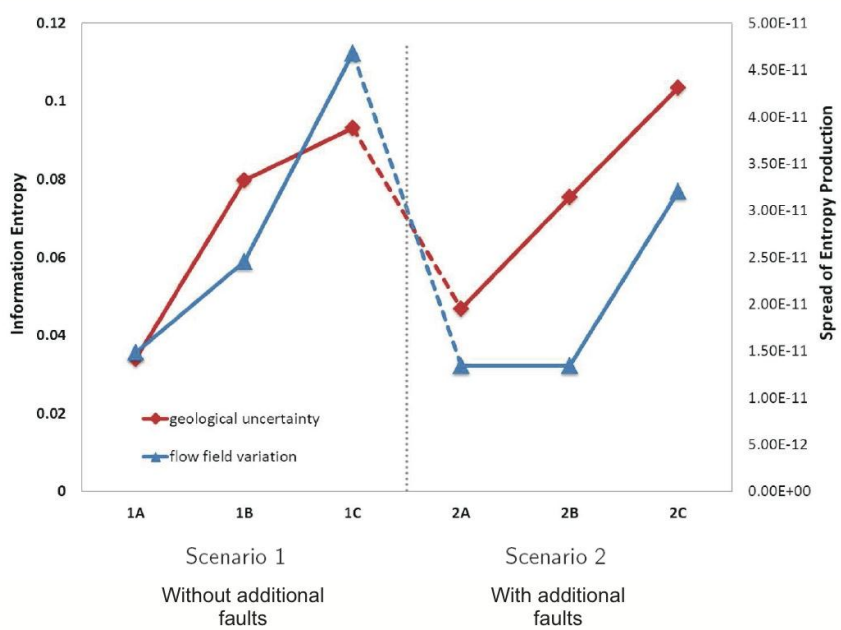

Figure 9: comparison of geological uncertainty, evaluated with total information entropy, and flow field variation, determined from the spread of average specific entropy production for the different geological scenarios and decreasing data quality.

geological models with flow simulations (Suzuki, 2008), the methods presented here are, to the best of our knowledge, the first approach to integrate the accuracy of the geological input data itself. This is an important step forward as it enables the consideration of the quality of the initial geological data, even for complex full 3-D geological settings.

It is worth noting that the system-based measures are applicable in a much broader context. We applied them here in the context of stochastic structural geological modelling and data intensive simulations. However, they could similarly be used to only compare results of two specific scenarios, or to analyse probability fields derived with different methods, for example geostatistical simulations. We envisage a wide range of possible applications of the measures in the context of geological models and hydrothermal simulations.

ACKNOWLEDGEMENTS

This work was enabled by an International Postgraduate Research Scholarship (IPRS) from the Australian Government and an OCE postdoctoral fellowship of CSIRO. The Western Australian Geothermal Centre of Excellence (WAGCoE) is a joint center between Curtin, UWA, CSIRO and funded by the Western Australian State Government. We thank Lynn B. Reid for very valuable comments that improved the manuscript.

REFERENCES

Bundschuh, J., and Arriaga, S. (2010). Introduction to the numerical modeling of groundwater and geothermal systems: Fundamentals of mass, energy and solute transport in poroelastic rocks. Multiphysics Modeling (Vol. 2). Leiden: CRC Press/Balkema.

Calcagno, P., Chil es, J.-P., Courrioux, G., & Guillen, A. (2008). Geological modelling from field data and geological knowledge: Part I. Modelling method coupling 3D potential-field interpolation and geological rules: Recent Advances in Computational Geodynamics: Theory, Numerics and Applications. *Physics of the Earth and Planetary Interiors*, ***171*** (1-4), 147–157.

Callen, B. H. (1985). Thermodynamics and an Introduction to Thermostatistics (Vol. 66). New York: Wiley.

Clauser, C., & Bartels, J. (2003). Numerical simulation of reactive flow in hot aquifers: SHEMAT and processing SHEMAT. Berlin: Springer.

Finsterle, S. (2004). Multiphase inverse modeling: Review and iTOUGH2 applications. *Vadose Zone Journal*, ***3*** (3), 747.

Garibaldi, C., Guillou-Frottier, L., Lardeaux, M. J., Bonte, D., Lopez, S., Bouchot, V., & Ledru, P. (2010). Thermal anomalies and geological structures in the Provence basin: Implications for hydrothermal circulations at depth. *Bulletin de la Societe Geologique de France*, ***181*** (4), 363.

Goodchild, F. M., Chih-Chang, L., & Leung, Y. (1994). Visualizing fuzzy maps. In M. H. Hearnshaw & J. D. Unwin (Eds.), *Visualization in Geographical Information Systems* (pp. 158–167). New York: John Wiley & Sons.

Mory, J. A., & Iasky, P. R. (1996). *Stratigraphy and structure of the onshore northern Perth Basin, Western Australia*. Report (Vol. 46). Perth: Western Australia Geological Survey.

Ozawa, H., Ohmura, A., Lorenz, D. R., & Pujol, T. (2003). The second law of thermodynamics and the global climate system: a review of the maximum entropy production principle. *Rev. Geophys*, ***41*** (4), 1018.

Pruess, K. (1991). TOUGH2: A general-purpose numerical simulator for multiphase fluid and heat flow. Technical Report, Lawrence Berkeley Lab., CA (United States), LBL-29400

Regenauer-Lieb, K., Karrech, A., Chua, T. H., Horowitz, G. F., & Yuen, D. (2010). Time-dependent, irreversible entropy production and geodynamics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, ***368*** (1910), 285.

Suzuki, S., Caumon, G., and Caers, J. (2008). Dynamic data integration for structural modeling: model screening approach using a distance-based model parameterization. *Computational Geosciences*, **12** (1), 105–119

Wellmann, J. F., Horowitz, F. G., & Regenauer-Lieb, K. (2010). Influence of Geological Structures on Fluid and Heat Flow Fields: Proceedings, Thirty-Fifth Workshop on Geothermal Reservoir Engineering Stanford University, Stanford, California, February 1-3, 2010 SGP-TR-188.

Wellmann, J. F., & Regenauer-Lieb, K. (2011a). Uncertainties have a meaning: Information entropy as a quality measure for 3-D geological models. *Tectonophysics*, in press, doi:10.1016/ j.tecto.2011.05.001

Wellmann, F. J., Reid, B. L., & Regenauer-Lieb, K. (2011b). From outcrop to flow fields: combining geological modeling and coupled THMC simulations. In J. Liu (Ed.), *GeoPROC 2011: Cross Boundaries Through THMC Integration*,

6-9 July. Presented at the GeoPROC 2011: Cross Boundaries Through THMC Integration, 6-9 July.

Wellmann, J. F., Croucher, A., & Regenauer-Lieb, K. (2011). Python scripting libraries for subsurface fluid and heat flow simulations with TOUGH2 and SHEMAT. *Computers & Geosciences*, in press, doi:10.1016/j.cageo.2011.10.011

Wellmann, J. F. (2012). Uncertainties have a meaning: Quantitative Interpretation of the Relationship between Subsurface Flow and Geological Data Quality (PhD Thesis.). University of Western Australia, Perth.