



Uncertainties have a meaning: Information entropy as a quality measure for 3-D geological models

J. Florian Wellmann ^{a,*}, Klaus Regenauer-Lieb ^{a,b}

^a Western Australian Geothermal Centre of Excellence, School of Earth and Environment, The University of Western Australia, 35 Stirling Hwy, Crawley, WA-6009, Australia

^b CSIRO, ESRE, PO Box 1130, Bentley, WA-6102, Australia

ARTICLE INFO

Article history:

Received 14 January 2011

Received in revised form 8 April 2011

Accepted 2 May 2011

Available online 8 May 2011

Keywords:

Information entropy

3-D geological modeling

Uncertainty

Simulation

Fuzziness

Visualization

ABSTRACT

Analyzing, visualizing and communicating uncertainties are important issues as geological models can never be fully determined. To date, there exists no general approach to quantify uncertainties in geological modeling. We propose here to use information entropy as an objective measure to compare and evaluate model and observational results. Information entropy was introduced in the 50s and defines a scalar value at every location in the model for predictability. We show that this method not only provides a quantitative insight into model uncertainties but, due to the underlying concept of information entropy, can be related to questions of data integration (i.e. how is the model quality interconnected with the used input data) and model evolution (i.e. does new data – or a changed geological hypothesis – optimize the model). In other words information entropy is a powerful measure to be used for data assimilation and inversion.

As a first test of feasibility, we present the application of the new method to the visualization of uncertainties in geological models, here understood as structural representations of the subsurface. Applying the concept of information entropy on a suite of simulated models, we can clearly identify (a) uncertain regions within the model, even for complex geometries; (b) the overall uncertainty of a geological unit, which is, for example, of great relevance in any type of resource estimation; (c) a mean entropy for the whole model, important to track model changes with one overall measure. These results cannot easily be obtained with existing standard methods. The results suggest that information entropy is a powerful method to visualize uncertainties in geological models, and to classify the indefiniteness of single units and the mean entropy of a model quantitatively. Due to the relationship of this measure to the missing information, we expect the method to have a great potential in many types of geoscientific data assimilation problems – beyond pure visualization.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Visualizing and analyzing uncertainties in 3-D structural geological models are widely recognized as important issues (e.g. Bond et al., 2007; Caumon et al., 2009; Jessell et al., 2010; Jones et al., 2004). We propose here an information entropy method as a quantitative measure for the quality of a geological model and its geological sub-units, based on the Shannon Entropy model (Shannon, 1948). Goodchild et al. (1994) already used this method to visualize uncertainties in map applications in a 2D context for fuzzy sets resulting from poorly constrained data. We extend it here into the third dimension and combine it with an uncertainty simulation approach that is applicable to complex 3-D geological settings (Wellmann et al., 2010). In the context presented here, we propose that information entropy is a sound method to visualize uncertainties in any complex 3-D setting. But due to its underlying concept of

information, we expect that it can provide significant insights into the used and missing information to constrain a model.

In the following, we will briefly review the concept of information entropy and will show how this measure can be used to evaluate uncertainty at one position within a model domain. We will then apply this measure to visualize a 3-D uncertainty field, calculated with a simulation method for geological uncertainty and discuss the feasibility of this method and the possible extension beyond visualization.

2. Materials and methods

2.1. Visualizing uncertainty

A variety of different methods exists to evaluate the quality of a geological model. These range from simple uncertainty measures, determined from data uncertainties, over geostatistical evaluations (e.g. Chilès and Delfiner, 1999) to model simulations (Suzuki et al., 2008). A challenge is, in many cases, the communication of the model quality. Several authors have developed methods to visualize uncertainties in a geographic reference frame (e.g. MacEachren et al.,

* Corresponding author. Tel.: +61 8 6488 7230; fax: +61 8 6488 1037.

E-mail address: florian.wellmann@uwa.edu.au (J.F. Wellmann).

2005) or in a stratigraphic grid (Viard et al., 2007). The problem with most of these visualizations is that they are restricted to specific geological settings (i.e. most cannot handle complex three-dimensional settings like domes or overturned folds), that many estimations are subjective and ‘ad-hoc’ approaches (Jones et al., 2004) or only suitable to represent specific types of uncertainties (see Bond et al., 2007, and references therein for a discussion of several approaches).

Our aim of visualizing the uncertainty in a structural geological model is to find a measure that predicts the accuracy of the model at every location in the model space, for any type of complex geological model. To achieve this, we subdivide the whole model space into a regular raster with equal cell sizes (also referred to as voxels). We can then evaluate the accuracy of every voxel, everywhere within a unit – rather than only the accuracy of a polygon/surface boundary (see Leung et al., 1993 for a discussion of raster vs. polygon estimation).

Within each of the raster cells, we want to find a measure of uncertainty fulfilling the following criteria:

1. The value should be 0 when no uncertainty exists, i.e. there is only one possible outcome (i.e. geological unit) i with probability $p_i = 1$;
2. The value should be maximal when all n possible outcomes are equally likely: $p_i = \frac{1}{n} \forall i \in n$. Or, formulated for the case of a geological model: there is absolutely no reason to expect a specific geological unit at this location;
3. If more outcomes are equally likely, the measure should have a higher value (“monotonicity”);
4. The measure should fulfill the criteria of expansibility, i.e. the value should not change when an additional outcome with probability 0 is added;
5. The measure should be independent of the order of results (“symmetry”).

The information entropy, first defined by Shannon (1948), fulfills all of these properties (see Klir et al., 1988; Yager, 1995, for further details). We apply it here as a method to visualize uncertainties and discuss some potential applications beyond pure visualization.

2.2. Information entropy

The concept of information entropy was first defined by Shannon (1948) in a study performed to identify the amount of information required to transmit English text. The underlying idea was that, given the probabilities of letters occurring in the English alphabet, it is possible to derive a measure describing the *missing information* to determine the full text of a partially transmitted message, where information is understood as the information required to identify the message, not the information of the message itself. Based on several theoretical considerations, Shannon derived the following equation to classify a measure of the missing information, often referred to as information entropy:

$$H = - \sum_i^N p_i \log p_i. \quad (1)$$

The information entropy H is defined as the sum of all products of probabilities p for each possible outcome i of N total possible outcomes with its logarithm. The minimum value is 0, because $\log 1 = 0$ and $\lim_{x \rightarrow 0} (x \log x) = 0$ (possible to prove with L'Hopital's theorem, see Ben-Naim, 2008). The logarithm can be taken with any base, depending on the applied unit of information. We will use the logarithm with base “2” in the following examples and discussion, as it relates to the information unit of one bit (see below).

2.2.1. Information entropy of a 1 bit system with two possible outcomes

As an example, we will briefly examine the information entropy of a very simple system with only two exclusive outcomes A and B , with $P(A)$ and $P(B) = P(A)$. We can, for example, consider the example of a

coin flip where the outcome is either *head* or *number* (1-bit system with two possible outcomes). The information entropy quantifies the *amount of missing information* to classify this system.

If the coin is fair, i.e. outcome of head and tail is equal, then the entropy is highest. Whereas, when the coin is unfair, the entropy is lower, because one outcome is more probable than the other. In the extreme case (an extremely unfair coin) where the outcome is always the same, the information entropy is 0, because the outcome is known. Using Eq. (1), this simple example is presented in Fig. 1.

2.2.2. Information entropy in an evolutionary spatial context

In a spatial or modeling context, we can interpret the information entropy of a model subregion (e.g. a cell) as the *amount of missing information* with respect to the discrete properties of the cell. In the following, we consider the membership to a specific geological unit as this property. This is possible, as we consider geological units as exclusive events, i.e. a cell belongs either to the geological unit 1, or unit 2, ... M .

For each discrete subregion, we can describe the information entropy as

$$H(x, t) = - \sum_{m=1}^M p_m(x, t) \log p_m(x, t) \quad (2)$$

where x denotes the location of the subregion and M the number of possible (exclusive) members the subregion can contain. t could be physical time or any other parameter describing the evolution of a model.

This measure can then be used to visualize the uncertainty of cells in a straight-forward way. Let us, for example, consider a map with discrete subdivisions into a regular grid (Fig. 2a) where each cell can contain one of three possible members and where we can derive probabilities for each possible outcome in each cell (Fig. 2b). Applying Eq. (2) for each cell with position x , we obtain a map of information entropies (Fig. 2c).

We can see that the entropy measure fulfills the criteria described above (Section 2.1), specifically:

- The entropy is 0 for cells where one member has the probability 1 and all others are 0 (e.g. cell A).

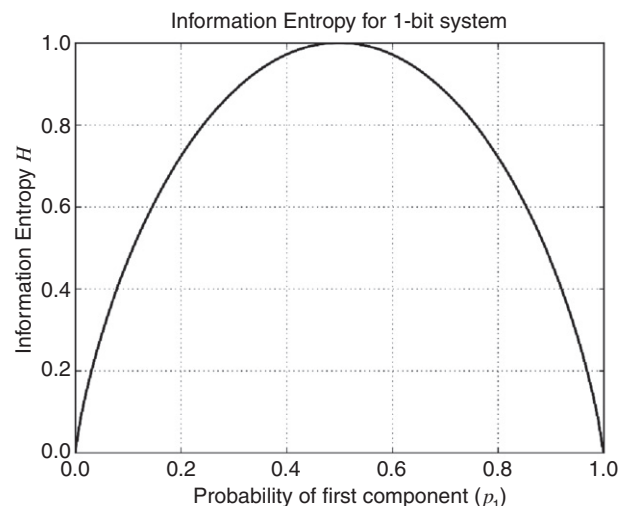


Fig. 1. Shannon entropy for 1 bit system (two possible outcomes) as a function of the probability of the first outcome. When the probability of one outcome is close to 0 or 1, the entropy is low; whereas when the probability of both outcomes is 0.5, the entropy is maximal and equal to 1 if the information unit is 1 bit and the logarithm to base 2 is taken.

Download English Version:

<https://daneshyari.com/en/article/4692910>

Download Persian Version:

<https://daneshyari.com/article/4692910>

[Daneshyari.com](https://daneshyari.com)