An Entropy-Based Measure of Continuity for Weighting Multiple Training Images

Daniel A. Silva and Clayton V. Deutsch

Abstract
The geological modeling of domains in a mining deposit is an important stage within the assessment of mineral resources and reserves. One of the main challenges consists of representing and characterizing the contacts between the geological units. The idea of using two training images, one from a smooth interpretation and a second from stochastic simulation, has been proposed. An important challenge of this idea is determining the weight that each training image should receive. A calibration approach is proposed in this paper that uses a measure of entropy. The reference entropy is calculated from the drillholes. Then, the entropy is calculated, in the same orientation as the drillholes, from realizations using different weights to the training images. The correct weighting is when the realization matches the drillholes.

Introduction
A sound understanding of the geological formation processes that created the mineral deposit allows improved geological models and resource estimates. A manual interpretation perhaps aided by distance function interpolation provides a smooth continuous model of the different rock types. A purely geostatistical procedure using sequential simulation provides a relatively discontinuous model of the different rock types. Combining these two models provides an opportunity to create realizations with the correct amount of continuity/variability. The present work aims to determine how much weight each model should receive by utilizing a measure of entropy to represent the continuity of a model.

The continuity along the geological contact is understood as the degree of randomness, that is, whether the boundaries between the units are sufficiently smoothed or have a pattern with many random changes. Putting into a geostatistical perspective the interpretation of the continuity of the contact this may be depicted by a scale of randomness from zero to one. Where we set zero to be the geologist's interpretation, generated by digitalization of the geology by means of cross sections and planviews, which is completely smoothed and deterministic; and on the other hand, with a value of one to the a relatively randomized contact characterized by a geostatistical stochastic method such as sequential indicator simulation (SIS).

Geological modeling leads to different degrees of complexity ranging from gradational changes of the rock properties from one geological unit to another, convoluted shapes, structural elements and the continuity of the contacts - directly related with the scale of the geological phenomena. In practice, there is a combination of larger scale deterministic features and shorter scale more random features. The concept of entropy (Journel and Deutsch, 1993) was introduced as an alternative to measure the continuity of the contacts in the dataset. Calculating the entropy along the drillholes provides a measure of the spatial disorder of the geology and setting up the grade of randomness. The current proposal aims to calibrate the mixing between two training images (the deterministic geological one and a more random geostatistical one).

Entropy as Measure of Continuity
The concept of entropy was introduced in geostatistics by (Christakos, 1990) as a measure of the uncertainty of a probability density function. In the case of discrete variable, it is defined as:

$$H = -\sum_{k=1}^{n} p_k \ln(p_k)$$

With p_k as the probability of occurrence of k possible outcomes of the random function with a probability law $\mathcal{P}$and where $\sum_{k=1}^{n} p_k = 1$.

Like the variance, the entropy is a summary statistic that does not deliver any information respect to the shape of the distribution. Furthermore, this statistics has showed that the one of the most utilized geostatistical methods for simulating mining deposits or petroleum reservoirs, such as Sequential Gaussian Simulation does not lead to connected extreme values, conversely the median values are well correlated. Thus, the entropy may be understood in terms of spatial configuration of the regionalized phenomena as a measure of the degree of disorder of the natural variables.

An example is used to depict the usefulness of this measure for the concept of continuity. Figure (1) shows two training image with significant differences between them in terms of continuity. Let's suppose the template as shown Figure (2.B). With two categorical values on the training image and with a template of three by one a total of 8 (2^3) configurations are possible. After scanning the training images with the template and classifying the configurations found at each location, the calculation of the probabilities of equation (1) are possible. The entropy of training image A is less than the entropy of training image B, which is expected given the visual degree of spatial disorder between the training images. The consequences of using a different template were studied by Larrondo (2003). An increment of the dimensionality and number of categorical variables increases the measure of entropy; however, this does not alter the ranking of the two training images.

Entropy Calculation of Drillholes and Training Images

The proposed approach consists of evaluating the degree of mixing between two training images to calibrate the measure of continuity from the data configuration. In the case of scarce spatial information such as drillholes, the entropy can only be calculated along the drillholes. Figure 2.A, 2.B and 2.C show a scheme of a template with two, three and four cells, respectively. The key idea is to calculate the probability of every template configuration along each drillhole and evaluating the entropy for every single drillhole separately. Lately, these are averaged to attain the entropy of the spatial data configuration depending on the orientation and geology, it is represented by equation (2).

$$H_{drillhole} = -\frac{1}{nd}\sum_{i=1}^{nd}\sum_{j=1}^{k} p_{i,k} ln(p_{i,k}) \qquad (2)$$

Where nd is the number of drillholes and k the number of possible template configuration. In order to make possible the calibration of mixing between the training images a comparison of $H_{drillhole}$ with the entropy of the training images is required. Figure (3) shows the regular sampling procedure on 2-D training image through different drillholes orientation. At each of the synthetic drillholes, the entropy is calculate using (2) and finally the entropy of the training image is the average of the entropy for every synthetic drillhole. Equation (3) summarizes the process:

$$H_{Training\ Image} = \frac{1}{nd}\frac{1}{nsd}\sum_{i=1}^{nd}\sum_{l=1}^{nsd}\sum_{k=1}^{k} p_{i,l,k} ln(p_{i,l,k}) \qquad (3)$$

Where nd is the number of drillholes and nsd the number of replicates-orientated drillholes. Let's illustrate the entropy calculation with an example. Assuming a 2-D spatial drillhole configuration such as that given on Figure (4.A). Two training images both related but with marked differences in terms of continuity as those showed on Figure (4.B and C) are considered. A sampling process over the two training images was realized using the drillhole configuration of Figure (4.A). Thus on the one part, using the equation (2) the entropy for the drillholes was calculated with a one dimensional template of three cells, Figure (2.B). On the other hand, the calculation of entropy of the training image, using equation (3), considered the configuration of each drillhole on the space replicating the orientation as showed Figure (3) for each of the ten drillholes sampled. The values of the entropy for both: the drillholes and training images are showed on each plot.

Calibration of the Mixing of Training Images

A way to relate the entropy of whatever configuration of drillholes and the entropy of a training image has been proposed. As a consequence, we may calibrate the degree of mixing between two training images by calculating the 1-D orientated entropy of the drillholes and comparing to the entropy of the training images. An example allows us to check the correlation between the entropy of the drillholes and the training images with the result of calibrating the continuity. A novel approach was proposed to mixing the conditional probabilities - relies on the Linear Opinion Pool Paradigm - of two or more training images

(Silva and Deutsch, 2012) and getting simulated categorical attributes that extract the intrinsic properties of each training images. The importance of each training image is based on the mixing of the training images. Using these training images - Figure (4.B and C) - a total of 100 simulations were completed. For instance, a simulation with a proportion of 0.3 is composed by 30% of training (A) and 70% (B). Each of the 100 simulations changes the proportioning composition of the two training images (A and B) at a rate of 0.01 increasing from zero to one.

By Equation (3) the entropy of the training image is calculated taking into account the orientation of the drillhole configuration and plotted in Figure (5). The curve shows the reduction in the spatial disorder when the proportion of the simulations goes up to one or we give less weight to the more “continuous” training image. In the case of drillholes, the entropy is evaluated once the sampling at each training image has been performed using equation (2) and plotted versus the entropy of training images going through the 100 simulations such as shown on Figure (6). There is a high correlation of 0.94 that shows that the entropy along the drillholes is accurately reflected by the entropy of the training images.

The calibration is performed using the correlation of Figure (6). Once the entropy of the drillholes is calculated it will have several values for the entropy given that value. Using Figure (5) the values of the entropies of the training images are related with the proportion for each training images used up. For instance, an entropy drillhole of 1.07 with a tolerance interval of $\pm$ 0.01 has associated the set of training images’ entropies $\{1.10,1.08,1.09,1.09,1.00,1.03\}$ and, in turn, these values are linked with the combining proportions $\{9\%,26\%,27\%,28\%,41\%,43\%\}$. Finally, the measure of continuity given by the drillhole entropy calculation delivers us an interval of possible mixing proportion between a minimum of 9% and a maximum of 43% with an expected value of 29%. Using the expected value, the mixing of training images (4.B) and of (4.C), with a proportion of 29% and 71%, will set out the training image to be utilized by whatever MPS algorithm and reproducing the continuity of the boundaries.

Conclusions

An approach has been proposed to quantify the continuity of geological units by using the calculation of entropy based on a 1-D template. The high correlation between the entropy evaluated from drillholes and those calculated for the training images makes the approach suitable to determine the weight that should be given to multiple training images in modeling. The drillholes must fairly represent the study area and there must be sufficient data for reliable statistics.

References

Christakos, G., 1990, A Bayesian/Maximum-Entropy View to the Spatial Estimation Problem: Math Geol., v22. N. 7, p. 763-777.

Journel, A.G. and Deutsch, C.V., 1993. Entropy and Spatial Disorder. Mathematical Geology, Vol. 25, No. 3, pp. 329-355.

Larrondo P.F., 2003. Entropy of Random Gaussian Functions and Consequences in Geostatistics Technical Report, Centre for Computational Geostatistics, University of Alberta. September 2003.

Silva D.A., 2012. Multiple Point Statistics with Multiple Training Images in Geostatistics Technical Report, Centre for Computational Geostatistics, University of Alberta. September 2012.

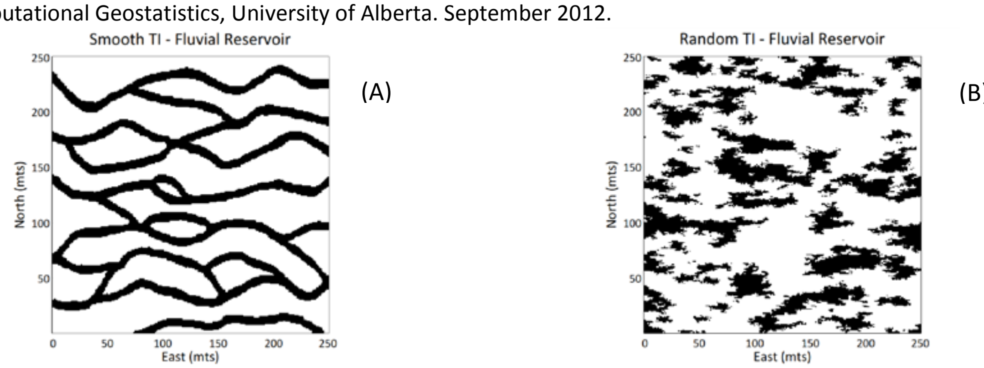

Figure 1: Two training images with extreme difference in terms of continuity. On the left training image from a geological interpretation and on the right a training image product of a Sequential Indicator Simulation.

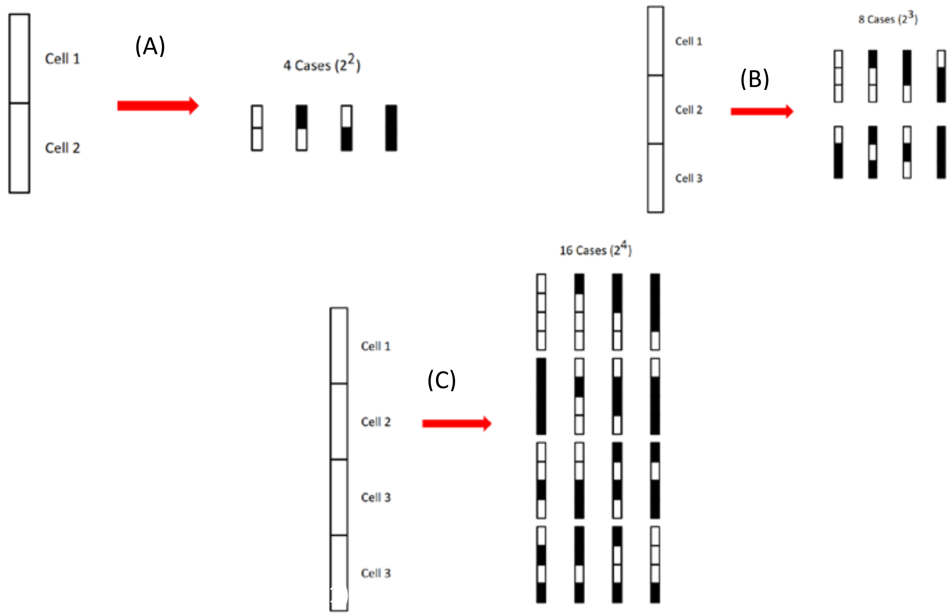

Figure 2: On the left, 1-D template with three locations and on the right eight possible configurations for the template with two geological units.

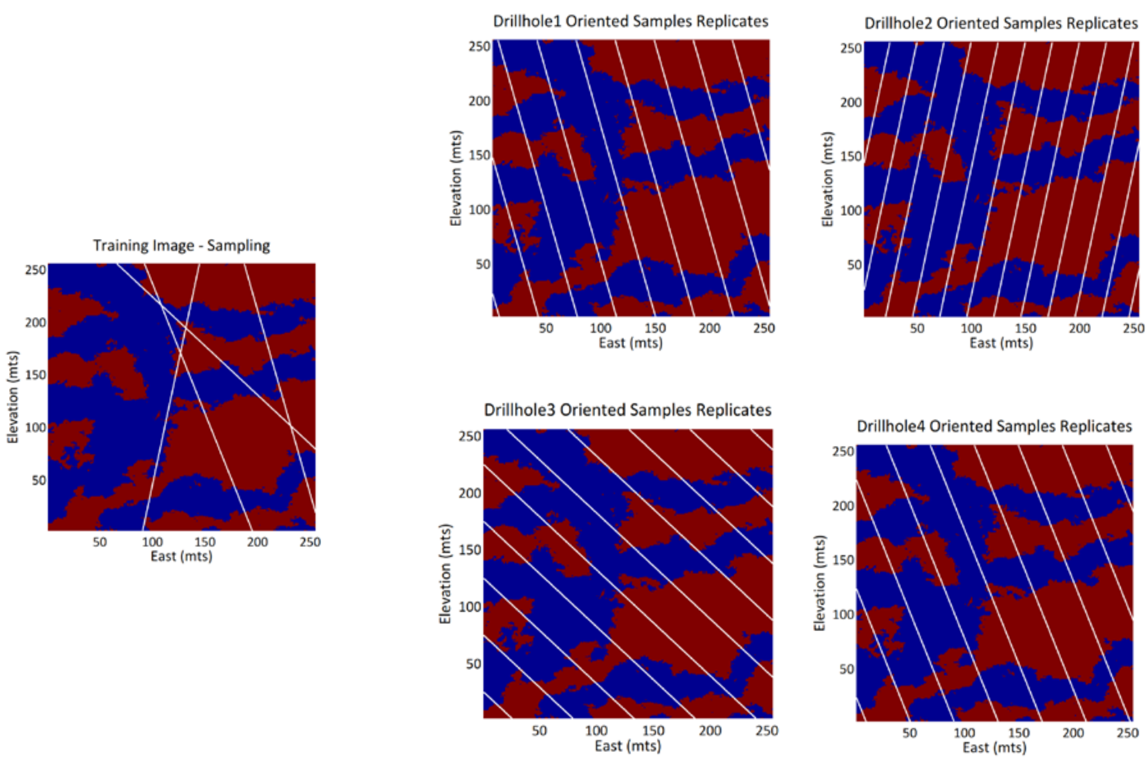

Figure 3: On the graph of the left, four sampled drillholes. Each of them is replicated several times according with their orientation, figures on the right.

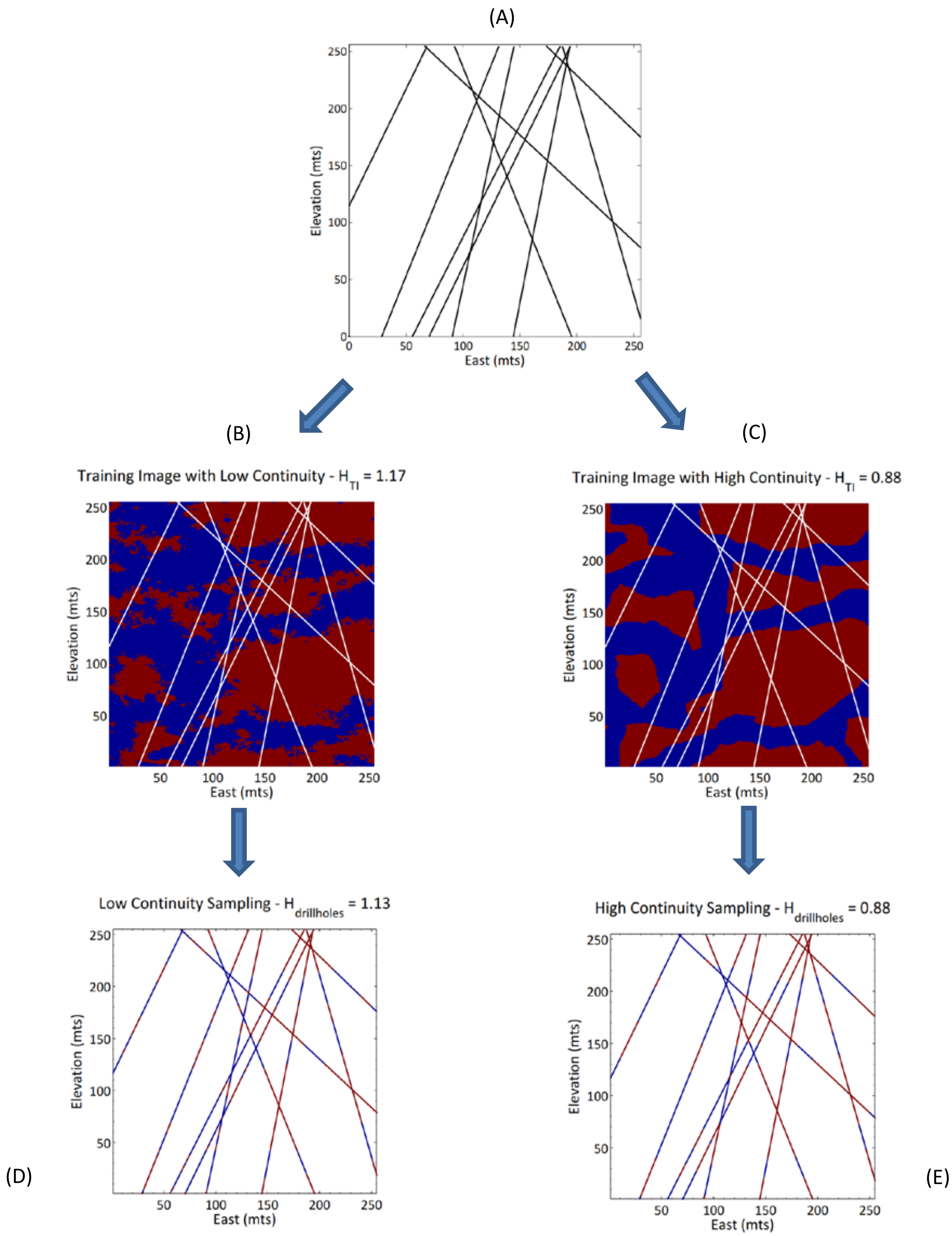

Figure 4: Sampling process over two training images, which are utilized to assess the entropy of training images and drillholes.

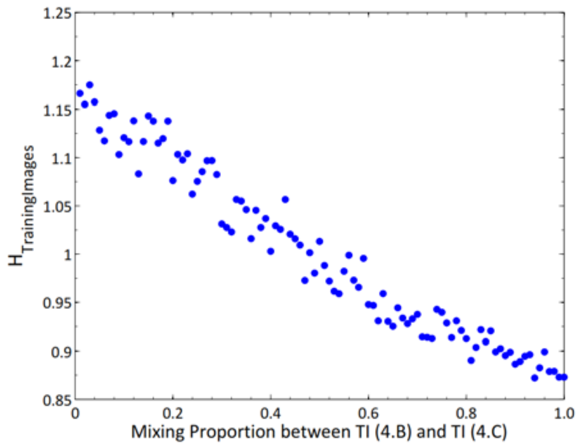

Figure 5: Entropy of 100 training images which, in turn, are brought about by the mixing between training images of Figure 4.B and 4.C.

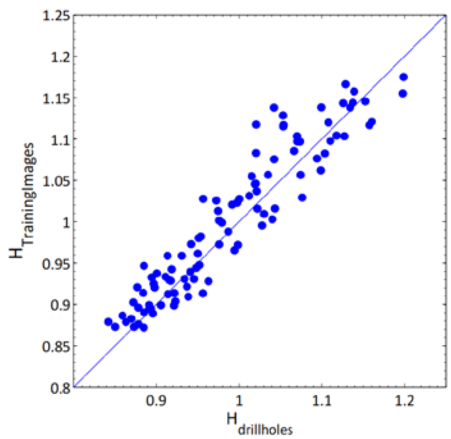

Figure 6: Entropy of training images against entropy of drillholes.