On Maximizing Spatial Entropy

R. Mohan Srivastava[1] and Oy Leuangthong[2]

Abstract: The “maximum spatial entropy” concept is often invoked in discussions of geostatistical simulation, following the intellectual tradition described by Jaynes (1957): that maximum entropy models are better because they assume less according to Shannon’s interpretation of entropy as an information measure. This paper explores whether or not multi-Gaussian models maximize spatial entropy. Although the answer to this has been assumed to be “Yes”, there has been no specific demonstration of this for geostatistical simulations; it is assumed to be an extension of the proof by Patil et al. (1975) that, in the class of all unbounded distributions, the Gaussian distribution maximizes Shannon’s entropy if the variance is fixed (or, for multivariate distributions, if the covariance is fixed). A test using a small indicator simulation problem indicates that the highest spatial entropies are not achieved with a multiGaussian procedure and that simulated annealing can produce realizations with higher spatial entropy. “Maximum diversity” and “maximum richness” criteria are discussed as alternatives to “maximum spatial entropy”; these alternatives, which are used in other disciplines that use spatial simulations, may be more relevant to the practice of conditional simulation.

Introduction

“Any method involving the notion of entropy … will doubtless seem to many far-fetched, and may repel beginners as obscure and difficult of comprehension” – W. Gibbs (1873)

“Entropy, whose meaning varies substantially, at times even within the same domain of intellectual endeavor, is probably the most unfortunate concept in the history of science” –N. Georgescu-Roegen (1971)

Entropy, which is often invoked as a concept important for prediction, including spatial prediction, is an elusive concept whose specific meaning is clouded by many different usages. Some of these are very clear and precise; others

---

[1] FSS Canada, #1100 - 120 Eglinton Ave. East, Toronto, Ontario, Canada, M4P 1E2, MoSrivastava@fssconsultants.ca

[2] SRK Consulting (Canada) Inc., 25 Adelaide St. East, Suite 2100, Toronto, Ontario, Canada, M5C 3A1, oleuangthong@srk.com

are so vague that they seem almost mystical. This ambiguity is due, in part, to the fact that the term "entropy" is used for two different but related concepts that both pertain to disorder or randomness. The 19th century physicist and mathematician Rudolf Clausius introduced the term to describe disorder in isolated systems, part of his work on the Second Law of Thermodynamics. The 20th century mathematician Claude Shannon chose the same term to describe the information content of messages, part of his work on a theory of communications. The potential for confusion was apparent to von Neumann, who encouraged Shannon to use the term entropy, pointing out that "it will give you a great edge in debates because nobody really knows what entropy is anyway" (Denbigh and Denbigh, 1985).

When used in the context of statistical inference and prediction, entropy usually refers to the definition proposed by Shannon in the 1940s. The longer historical tradition and more widespread cultural usage of thermodynamic entropy, however, has also caused the term to be invoked when statisticians seek to elevate some inference and prediction procedures to the status of universal natural laws akin to the Second Law of Thermodynamics. Some hold the belief that the best procedures are those that explicitly maximize entropy, as opposed, for example, to minimizing squared error or maximizing likelihood.

Entropy has surfaced in geostatistics, as "spatial entropy", and is used with even less clarity than in other applied statistical disciplines. The various geostatistical usages all turn on notions of disorder. Some geostatisticians assign a specific mathematical definition to spatial entropy, adapting Shannon's original definition to a spatial context; others use the term more loosely, intending only to point in a semi-quantitative way to a lack of structure or order.

While geostatistics shares with other applied statistical disciplines a range of vague to specific meanings for entropy, it distinguishes itself by sustaining a much more vigorous debate on the issue of whether or not entropy is good or bad, with utterly opposing views on whether this is a desirable property in an earth science model. Some geostatisticians explicitly embrace the notion that the best models maximize entropy because they minimize the incorporation of undocumented information or unstated assumptions; others reject maximum entropy as a guiding star, arguing that it is completely divorced from reality, the very opposite of how earth science phenomena actually behave (e.g. Gomez-Hernandez and Wen, 1998).

This paper addresses one specific claim made about entropy in geostatistics: that conditional simulations based on an assumption of multivariate Gaussianity provide maximum entropy realizations of spatial phenomena. To some, this is a good thing because they want to create realizations that contain no additional structure or order beyond that imposed by the user's parameter choices and by the available conditioning data. To others, this is a bad thing because they want to create realizations that portray the structure and order they observe in real geology. Without taking a position on whether or not maximum entropy realizations are desirable, this paper simply tests whether the basic assertion is true: do multivariate Gaussian simulations produce realizations that maximize spatial entropy? The answer, it turns out, appears to be "No".

The paper then explores the issue of whether or not entropy is a useful yardstick for what matters in the practice of conditional simulation, closing with a discussion of alternatives to Shannon's definition of entropy that are being used in other disciplines that use spatial simulations. These alternatives, which include measures of "diversity" and "richness", open up new directions for the theory and practice of conditional simulation.

Shannon's entropy

Shannon (1948) proposed entropy as a measure of the information content of a message, using the following formal definition:

$$Entropy = H(f_1, \ldots, f_n) = -\sum_{i} f_i \log f_i$$

with the f_i values being the frequencies of the n symbols used to communicate the message. Using a written phrase as an example, the message "gray goats eat grass" consists of 20 characters (including the blank spaces) and uses 9 symbols: a, e, g, o, r, s, t, y and the blank space symbol. It has an entropy of

$$H = -\sum \text{[illegible]} = 2.085$$

As originally proposed, entropy is a univariate characteristic of a complete set of frequencies (or of a probability distribution) that depends only on the frequencies of occurrence of the various symbols. When applied to a distribution of data values, it does not take into account the data values; it treats each data value as a symbol, as a letter in an alphabet.

Entropy is largest when every symbol occurs with the same frequency, when the histogram is flat. It decreases when some symbols start to occur more frequently than others, when the histograms has modes. It reaches zero when there is only one symbol in the entire message, when the histogram has a single mode with a frequency of 1.

Information, meaning and order

Shannon's entropy increases when a message contains less information. In this framework, "information" has nothing to do with what we conventionally understand as "meaning". The message "aaaa egggo rrs sstty" also has an entropy of 2.085 because it has the same frequency for each of the nine symbols as the more meaningful "gray goats eat grass". Shortly, when we discuss "spatial entropy", and how this is often interpreted as a lack of geological structure, we will see that, in a spatial setting, the difference between "information" and "meaning" becomes blurred.

The meaningless message in the previous paragraph also serves to show that Shannon's entropy has nothing to do with the order or position of the symbols; scrambling the message by shuffling the letters leaves the entropy of the message unchanged. Spatial order and position will become relevant when the classical definition of entropy is adapted to a spatial setting.

The maximum entropy principle

The maximum entropy (MAXENT) principle, first presented in two seminal papers by Jaynes (1957a, 1957b), states that when we make inferences based on incomplete information, we should draw them from that probability distribution that has the maximum entropy permitted by the available information. To many, this sounds like a sound approach to creating spatial simulations that aim to explore the space of uncertainty; rather than over-constraining realizations with unstated assumptions or hidden information, it seems sensible to allow realizations to acquire only as much structure and order … as much "information" … as the data and the explicit parameter choices permit.

Univariate distributions with maximum entropy

It has been proven (Patil et al., 1975) that, for a given variance, the univariate distribution that maximizes Shannon's entropy depends on the nature of the distribution's upper and lower bounds. The maximum-entropy distribution is:

- uniform, if the distribution is bounded by a minimum and maximum,
- exponential, if the distribution is bounded on only one side, and
- Gaussian, if the distribution is not bounded on either side.

Spatial entropy

The first attempt to adapt Shannon's concept of entropy to a spatial setting appears to be Batty (1974) who proposed a definition of "spatial entropy" for the problem of aggregating spatial data into discrete regions, using entropy as a yardstick for measuring the information content of the aggregated map. More recently, geostatisticians have proposed other definitions for "spatial entropy".

What all of the various "spatial entropy" definitions have in common is Shannon's basic equation. Where they differ is in how they define the members of the population … the “letters” or “symbols” whose frequencies of occurrence drive the calculation of entropy.

The method proposed by Journel and Deutsch (1993) consists of defining all possible configurations of a discrete variable on a template that consists of a fixed arrangement of cells on a regular grid. In an 0/1 indicator simulation, for example, where there are 2^{16} configurations of 0s and 1s in a 4x4 template, we can use each of these 2^{16} configurations as a unique “letter” in the alphabet, calculate the frequency of occurrence of each, and then apply Shannon’s formula to calculate the entropy.

While linking the entropy calculation to a specific template produces a clear definition of the set of configurations whose frequencies need to be calculated, it does so at the expense of making spatial entropy dependent on the choice of template; for each new template, there will be a different entropy.

At the same time that there have been attempts to make the calculation of spatial entropy clear, geostatisticians have also used the term in a more qualitative way to describe the visual disorder of an image. In many geostatistical articles, “high spatial entropy” is used to point to a lack of coherence of specific features like faults with high permeability zones, veins with high gold values, or shales and clay drapes that serve as barriers to flow in aquifers. When used this way, it is hard to reconcile “spatial entropy” with Shannon’s entropy from information and communication theory, where order is not relevant to the calculation.

The problem, in part, stems from the wish to interpret entropy as a measure of meaning. When geologists see simulations of earth-science phenomenon, their eyes interpret connected structures in a way that is meaningful to them, in the same way that a native English speaker sees the word “grass” and finds it meaningful. As discussed earlier, the information content of “gray goats eat grass”, as measured by its entropy, is the same as the information content of “aaaa egggo rrs sstty”, even though one looks like English and the other one doesn’t.

In geostatistics, we have yet to reconcile our attempts to adapt Shannon’s entropy to our spatial problems with the fact that the original concept does not propose to measure whether or not something is meaningful, or whether or not it looks right.

Maximum entropy in realizations of spatial phenomena

In the geostatistical literature, it has often been stated, and generally assumed to be true, that the conditional simulation procedures that will produce realizations with the largest spatial entropy are those that are based on an assumption of multivariate Gaussianity, such as sequential Gaussian simulation (sGs) or Cholesky decomposition of the spatial covariance matrix.

This claim can be tested by the following numerical experiment:

1. Select variogram model parameters and conditioning data for a conditional simulation exercise.
2. Use a multi-Gaussian simulation technique to create realizations.
3. Use simulated annealing to create realizations using the same variogram model parameters and the same conditioning data, and including in the objective function the maximization of spatial entropy.
4. Compare the spatial entropies from Steps 2 and 3.

If the realizations in Step 3 have spatial entropies that are similar to those in Step 2, or smaller, this establishes that, for the given test conditions (conditioning data and variogram model parameters), multiGaussian realizations do, indeed, have spatial entropies as large as those that can be created by a non-Gaussian technique, annealing, that specifically aims to maximize spatial entropy.

If the realizations in Step 3 have spatial entropies larger than those in Step 2, this establishes that, for the given test conditions (conditioning data and variogram model parameters), multi-Gaussian realizations are not, in fact, maximum entropy realizations.

The test study

Using a directionally isotropic linear variogram model with no nugget effect and a range of 10, we will produce realizations of a 0/1 indicator variable on a 25x25 grid, honoring the data points shown in Figure 1.

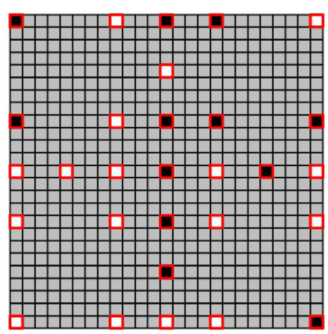


Figure 1. Conditioning data used for the test study.

The multi-Gaussian simulation procedure will use Cholesky decomposition (Davis, 1987) to create conditional realizations of a N(0,1) field that will be truncated at its median, with values above this threshold being assigned an indicator value of 1, and values below this threshold being assigned an indicator of 0.

The annealing procedure uses an objective function that includes:

1. The absolute differences between model and experimental variograms for all separation distances less than the range.
2. The simulated spatial entropy, H_{sim}, calculated using the 2^4 possible 0/1 configurations that can be observed in a 2x2 moving window.

The spatial entropy of the realizations will be calculated using the definition proposed by Journel & Deutsch (1993), with the 2^4 possible 0/1 configurations defining the 16 outcomes in the population. This quantity can be calculated empirically from the realizations produced by the two simulation procedures.

Four examples of the realizations created by the multi-Gaussian procedure are shown in Figure 2; these span the range from the one with the lowest entropy (1.553) to the one with the highest entropy (2.270).

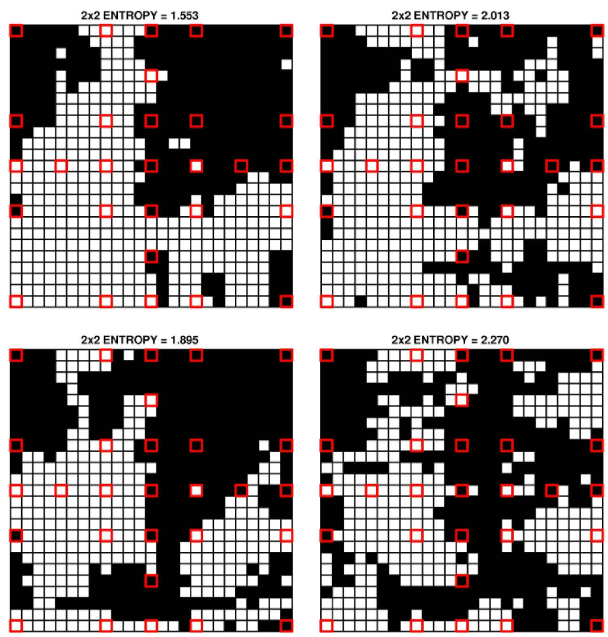


Figure 2. Four examples of realizations created using Cholesky decomposition.

Figure 3 shows four examples from the annealing procedure, spanning the range from the one with the lowest entropy (2.212) to the one with the highest (2.326).

Figure 4 shows side-by-side boxplots of the distribution of entropies for the two procedures. The annealing procedure produced realizations with entropies that were generally much higher than those produced by the multi-Gaussian method.

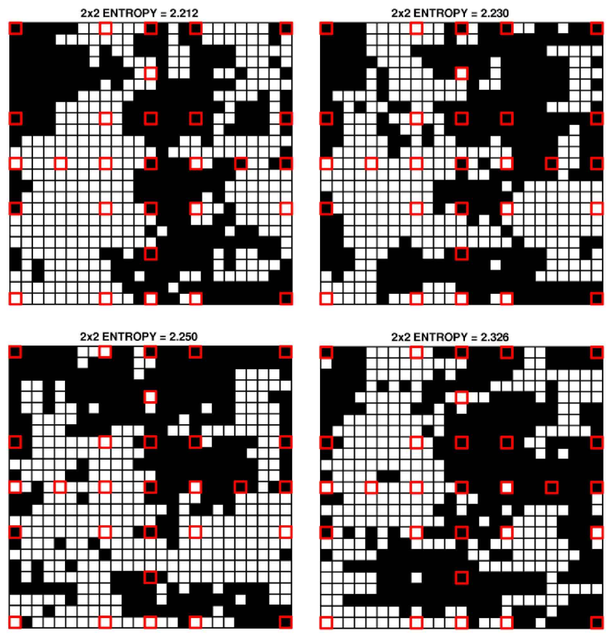
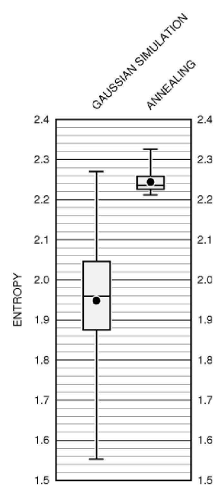


Figure 3. Four examples of realizations created using annealing.



	GAUSSIAN SIMULATION	ANNEALING
Number	100	100
Mean	1.948	2.244
Standard deviation	0.134	0.026
Maximum	2.270	2.326
75th percentile	2.046	2.258
Median	1.959	2.236
25th percentile	1.875	2.226
Minimum	1.553	2.212

Figure 4. Comparison of the distribution of entropies from 100 realizations created by a multiGaussian simulation procedure and 100 realizations created by annealing procedure.

The very lowest entropy in the set of 100 realizations produced by annealing (the 2.212 shown as the upper-left example in Figure 3), is at the 99th percentile of the entropies produced by Cholesky decomposition. The only multi-Gaussian realization that had a higher entropy was the one with the very largest entropy (the 2.270 shown as the lower-right example in Figure 2).

Discussion

The fact that the annealing procedure systematically produces realizations with higher entropy challenges the generally-held belief that multiGaussian models have the highest spatial entropy.

The test example links the calculation of entropy to a specific template, a 2x2 window that has 16 possible 0/1 configurations. Other templates were tested, but the result was always the same: simulated annealing produces realizations with entropies very high on the distribution of entropies produced by Cholesky decomposition, with most of the annealing realizations having entropies higher than the maximum observed in the multiGaussian realizations.

The difference between the annealing entropies and the multiGaussian entropies diminishes as the amount of conditioning data decreases. It is possible that the failure of the multiGaussian approach to achieve a high entropy is related to inconsistencies between the conditioning data and the multiGaussian model. In the limit, with abundant conditioning data, the multivariate distribution is not going to have many of the properties of the “pure” multiGaussian model.

Finally, the possibility that the multiGaussian model might not be able to claim the “maximum spatial entropy” prize should lead to a reflection on what that prize is worth; should we value maximum spatial entropy models?

In ecological statistics, there is rich literature on indexes of species diversity. These include Shannon’s entropy, along with measures like the Simpson index, and the richness index. Geostatistical simulation is able to produce realizations that are consistent with the available data but that show us a rich diversity of alternatives, some of which we may never have imagined. With this in mind, it is worth considering the possibility of explicitly creating simulation algorithms that do something other than maximize spatial entropy. A procedure that, for example, maximized the diversity of the realizations, without associating any sense of probability distribution to them, might better serve the purpose of risk analysis. If our technical and economic analysis had the benefit of seeing some truly bizarre (but still data-consistent) realizations, would it really matter if they were the 1st percentile, or the 0.1th percentile? Even without a probability distribution associated to our realizations, our plans can still benefit from the awareness of some previously unimagined possibilities that are plausible, given the available data.

Bibliography

Batty, M., 1974, “Spatial Entropy”, *Geographical Analysis,* v.6, n.1, p.1-31.

Davis, M.W., 1987, “Production of conditional simulations via the LU triangular decomposition of the covariance matrix”, *Mathematical Geology,* v.19, n.2, p.91-98.

Denbigh, K.G. and Denbigh, J.S., 1985, *Entropy in Relation to Incomplete Knowledge*, Cambridge University Press, Cambridge, UK.

Deutsch, C.V. and Journel, A.G., 1994, “The Application of Simulated Annealing to Stochastic Reservoir Modeling”, *SPE Advanced Technology Series,* v.2, n.2, p.222-227.

Georgescu-Roegen, N., 1971, *The Entropy Law and the Economic Process,* Harvard University Press, Cambridge, Massachusetts.

Gómez-Hernández, J.J., and Wen, X.H., 1998, “To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology”, *Advances in Water Resources,* v.21, n.1, p.47-61.

Jaynes, E.T., 1957a, “Information Theory and Statistical Mechanics”, *Physical Review Series II*, v.106, n.4, p.620-630.

Jaynes, E.T., 1957b, “Information Theory and Statistical Mechanics II”, *Physical Review Series II*, v.108, n.2, p.171-190.

Jaynes, E.T., 1982, “On the Rationale of Maximum Entropy Methods”, *Proceedings of the IEEE*, v.70, n.9, p.939-952.

Journel, A.G., and Deutsch, C.V., 1993, “Entropy and Spatial Disorder”, *Mathematical Geology*, v.25, p. 329-355.

Patil, G. P., Kotz, S., Ord, J.K. (eds.), 1975, *Statistical Distributions in Scientific Work,* Volume 3, D. Reidel Publishing Company, Dordrecht, 460p.

Shannon, C.E., 1948, “A Mathematical Theory of Communication”, *Bell System Technical Journal,* v.27, p. 379-423 and 623-656.