Framework for Entropy-based Map Evaluation

Jan T. Bjørke

ABSTRACT. The automation of map design is a challenging task for both researchers and designers of spatial information systems. A main problem in automation is the quantification and formalization of the properties of the process to be automated. This article contributes to the formalization of some steps in the processes involved in map design and demonstrates how the Shannon information theory (Shannon and Weaver 1964) can be used to compute an evaluation index of a map, i.e., a parameter which measures the efficiency of the map. Throughout this article, the term "information" is mostly used in a narrow sense and the application of information theory is restricted to the syntactic level of cartographic communication. Information sources for map entropy computations are identified and elaborated on. A special class of map information sources are defined and termed "orthogonal map information sources". Further, a strategy to consider spatial properties of a map in entropy computations is presented. At the end of the article, some examples demonstrate how the channel capacity and other entropy related measures can be computed and used to control automated processes for map design or map generalization.

KEYWORDS: Information theory, automated map design, map evaluation, automated cartographic generalization

Introduction

One of the main problems in the automation of cartographic design is quantifying the efficiency of a map. In recent years there has been considerable activity in related fields such as the representation of cartographic knowledge. The collection of articles in Buttenfield and McMaster (1991) covers some of this research and gives valuable insight to the problem of automation in cartographic design. There is still a lack of fundamental theory, however, on how to quantify the efficiency of cartographic communication. For a time in the 1970s the mathematical theory of communication (Shannon and Weaver 1964), normally termed "information theory," inspired several research articles in cartography. The Shannon information theory also became popular in several other disciplines (for example engineering, psychology, and biology) in the first 20 years after its introduction, but it never reached a high level of sophistication outside electronic communications. Moles (1966) points out that the most obvious failure of the theory in its simplest form, when applied to psychology, is that it appears an atomistic theory which tends to explain reality by decomposing it into simple elements. In the 1970s there were several critics of the cartographic relevance of the communication paradigm of Shannon and Weaver. Head (1991) points out that how to quantify the information itself was never fully understood:

> It came to be recognized, however, that map readers often seemed to get things from map reading that were not consciously designed-in by the cartographer, and this made measurement of information loss a fuzzy business (Head 1991, 238).

Neumann (1994, 26) notes the following criticism made in the 1970s:

> The communication concept had one weak point—the use of information theory was mechanically conditioned by the application of Shannon's theory of communication (Shannon and Weaver 1964). Consequently, it was criticized by Salichtchev (1973), Robinson and Petchenik (1976), and other authors in the 1970s. The critics were particular to point out that the conventional process of communication, accompanied with losses in transmitted information, could not be used as a model of the cartographic process which, in contrast, produced an increase in the amount of information.

Jan T. Bjørke is a professor in the Department of Surveying and Mapping at Norwegian University of Science and Technology, N-7034 Trondheim, Norway. E-mail: bjorke@iko.unit.no .

The type of criticism cited reveals the limitations of Shannon information theory and the problems that arise when attempting to apply it to map evaluation. The present article shows, however, that when Shannon information theory is applied to the syntactic component of a map, the theory can be used in map design and cartographic generalization with a potentially significant degree of success.

I will comment on the criticism of the cartographic relevance of information theory on the basis of the syntactic, semantic, and pragmatic aspects of communication. Klir and Folger (1988, 188) describe these three aspects:

1. *the syntactic aspect*, i.e., the relationship among the signs that are employed in the communication;
2. *the semantic aspect*, i.e., the relationship between the signs and the entities which they represent, that is, the designation of the meaning of the signs; and
3. *the pragmatic aspect*, i.e., the relationship between the signs and their application.

Shannon and Weaver (1964) distinguish these aspects of communication. In their terminology the three aspects are termed levels of communication problems and are given the abbreviations: level A (syntactic), level B (semantic), and level C (pragmatic). Shannon and Weaver emphasize that at level A they use "information" in a special sense that must not be confused with its ordinary usage.

> In particular, information must not be confused with meaning (1964, 8).

> To be somewhat more definite, the amount of information is defined, in the simplest cases, to be measured by the logarithm of the number of available choices (1964, 9).

It may be that some of the earlier criticism of the application of information theory would be met if we were more distinct about the three levels of communication problems, and if we evaluated the relevance of information theory to each of the levels, specifically. It is hoped that this article will demonstrate that information theory provides a sound basis for map evaluation when applied to the syntactic level of cartographic communication. Any attempt to apply information theory to the semantic level or the pragmatic level of cartographic communication, on the other hand, will meet the problems pointed out by Head (1991).

Robinson and Petchenik (1976, 41) correctly point out that the positional factor of a map must be considered if information theory is to be applied to cartography. The cartographic application of information theory of the 1960s and 1970s did not emphasize the positional component of a map and this is probably what brought information theory into discredit. This article demonstrates how different types of spatial information sources may consider the spatial component of a map in entropy computations.

Knöpfli (1983) explains the difference between aerial photos and maps in terms of information theory and shows that the amount of information in aerial photos, as well as maps, can be reduced by misinterpretations of the relevant messages. He nicely demonstrates the effect of distorted (noisy) information transmission and sets up two steps in order to reduce the loss of relevant information (1983, 207). These two steps are:

1. omit the irrelevant characteristics, and
2. strengthen the relevant characteristics.

These rules can be reformulated to:

1. not overloading the map with information (in this context information has the narrow meaning as earlier defined), and
2. maintaining a sufficient "visual distance" between the map symbols to make them distinguishable (in this context "visual distance" can be Euclidian distance in the map plane or distance defined in the domain of the visual variables such as color, shape and size).

Even if these rules are simplistic and general, they are very important considerations in map design. Information theory offers a mathematical basis which takes into account these rules, since the idea is to compute the efficiency of any kind of communication as the difference between the variation within the message and the amount of potential misinterpretation. This difference is termed *useful information* and can be expressed as:

$$R = H(X) - H_Y(X) \tag{1}$$

The mathematical basis for Equation (1) is explained in the Appendix. Shannon and Weaver (1964) term $H(X)$ the *entropy* of the information source and $H_Y(X)$ the *equivocation* of the information source. The *capacity* C of a noisy channel corresponds to the maximum rate of transmission and is defined as:

$$C = \max\left[H(X) - H_Y(X)\right] \tag{2}$$

A cartographic interpretation of Equation (2) can be done as follows:

rule (1) "not overloading the map with information" is considered by the max-operator, whereas
rule (2) "maintaining a sufficient visual distance between the map symbols to make them distinguishable", is considered by the expression $-H_Y(X)$.

The present article is concerned with the application of Shannon information theory to cartography, but restricted to communication problems at level A, i.e., *the syntactic aspect* of cartographic communication. Since levels B and C make use of the signals at level A, an efficient coding at level A is obviously a basis for efficiency at the other two levels. Map evaluation at the syntactic level should be an important step in any map design process. If information theory is to be applied to levels B and C, on the other hand, we must consider the purpose of the map and the meaning of the map symbols. The problems that raises are outside the scope of this article.

Three issues are addressed here. The first is the identification of the information sources of a map, i.e., what are the events and characteristics which the entropy computations should be based on? The second is how the spatial component in entropy computations should be considered. The last relates to the computation of the channel capacity in Equation (2). At the end of the article some examples demonstrate the application of information theory to map design. Finally, some important properties of Shannon entropy are summarized in the Appendix.

Previous Attempts at Applying Communication Theory to Cartography

Up to now there are few presented articles which demonstrate the utility of Shannon information theory in cartographic communication. If we turn to digital signal engineering, on the other hand, we find that information theory has become a sub-discipline to which entire journals and symposia are dedicated. A short survey of articles dealing with information theory in cartography follows.

Sukhov (1967) proposes an atomistic method to compute the entropy of a map. This is based on a method which breaks a map into discrete elements. A statistical sampling method is used for selecting typical unit areas from the map for measuring the entropy. The method is applied to different subsystems of the map, i.e., to different themes such as hydrography, relief, and roads. Finally, the map entropy is computed as the sum of the entropies of its different subsystems. Sukhov distinguishes the significance of the correlation between the subsystems (1967, 214). The subsystems of the study were weakly correlated (1967, 214), which gave Sukhov the basis for using the joint entropy computation in Equation (15) (Appendix). Sukhov's contribution gives insight into the significance of correlation in the computation of the joint entropy of different information sources.

Two articles by Knöpfli explain some features of cartographic generalization in terms of Shannon entropy. The first (Knöpfli 1980) demonstrates that some information can be derived from the structure of what is termed "the embedding space" using inductive reasoning. For example, if a city is located on both sides of a river, we can conjecture that there must be a bridge between the two parts of the city. Since the no-bridge case would be very unusual, its information value is very high. The information that there is a bridge has a lower information value than the information that there is no bridge. This example demonstrates that spatial correlation and spatial context should be considered in the entropy computations. In the second article (Knöpfli 1983), the difference between aerial photographs and maps are explained in terms of information theory. The article demonstrates very clearly that the scatter of the relevant messages (noise) leads to loss of information.

> It is always claimed that aerial photographs contain much more information than maps. Since I have dealt with the production of topographic maps from aerial photos for years, I am familiar with the advantages and disadvantages of both products and have never agreed with this assertion (Knöpfli 1983, 177).

Bjørke and Aasgaard (1990) propose information theory as a part of the concept that they call "cartographic zoom." This is a real-time concept which aims to generate map versions adjusted to the dynamic change of map scale on a computer screen. Information theory is described as a tool to measure the amount of information on a map and it is proposed that this be integrated into a subsystem which controls the number of map

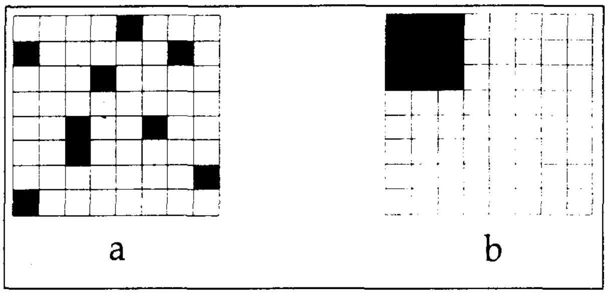


Figure 1. Different patterns of binary images. The patterns made by the two images are different, but the number of black pixels is equal in both.

symbols and their visibility. They emphasize that they use the term "information" in a narrow sense, and that their use of "information" has no connection with "meaning" (1990, 346). It seems clear, therefore, that their application of information theory is restricted to the syntactic level of information, i.e., level A according to the terminology of Shannon and Weaver.

Bjørke (1992) demonstrates how information theory can be used to control the generalization process in the two cases:

1. the selection of the number of classes in choropleth raster maps, and
2. the selection of parameter values in automated line generalization.

In both cases the channel capacity of the maps was computed. In the first case the borders between the raster elements (pixels) were selected as events for the entropy computation. An investigation of some subjects gave the probability that the different gray values were misinterpreted. Then the channel capacity of a random and a correlated choropleth map was computed. From this computation an optimum number of classes was derived for the two maps. In the second case, the angular change of the line to be generalized served as a basis for the entropy computation. Based on a model of the minimum separable distance between the events, an optimum value of the line generalization parameter was derived. Bjørke and Midtbø (1993) go further and apply information theory to contouring from digital elevation models (DEMs). In this case the underlying terrain model, not the contour lines themselves, was simplified and an optimum generalization parameter value was derived. An information theory method to compute an optimum contour interval is also proposed loosely in this article.

Bjørke (1994) introduces the concept of different types of entropies in a map and proposes a model for map design based on information theory. At the same time Neumann presents an article where the topological entropy of a map is focused (1994). The topological entropy of Neumann is computed from dual graphs (Region Adjacency Graphs). Bjørke (1994) defines an arrangement-entropy which also has a topological aspect, but the entropies of Neumann and Bjørke are different. The present article adopts the ideas of Bjørke et al. (1990; 1992; 1993; 1994). Earlier findings will be substantially deepened, and new ideas and perspectives are added.

Properties of the Syntactic Component of a Map

Spatial Correlation

When applying Shannon information theory in cartography, we face the problem of how to deal with spatial correlation. An aspect of the problem is demonstrated in Figure 1. Both the images in the figure consist of nine black and 55 white pixels, but the patterns in the two images are different. If we calculate the entropy of the two images on the basis of counting the number of black and white pixels, the entropy is computed as:

$$H_a(X) = H_b(X) = -\frac{9}{64}\cdot\log_2\frac{9}{64} - \frac{55}{64}\cdot\log_2\frac{55}{64} = 0.586$$

where

$\frac{9}{64}$ is the probability of finding black pixels, and

$\frac{55}{64}$ is the probability of finding white pixels.

To an observer, the pattern of image (b) in Figure 1 looks more ordered than the pattern of image (a), but we have computed identical values for their entropies. The reason is that the events in the message are spatially correlated and we have not modeled that correlation. The spatial correlation between neighboring pixels of an image can be taken care of by replacing the values of the pixels by their differences (Gatrell 1977; Weber 1980; Bjørke 1992 and 1993). Based on this idea, the previous entropy computation will be reformulated. If two neighboring pixels have the same color, we define their difference to be positive. Otherwise, if the pixels have different colors, their difference is defined as negative. According to this strategy, the entropy of a binary image can be defined as:

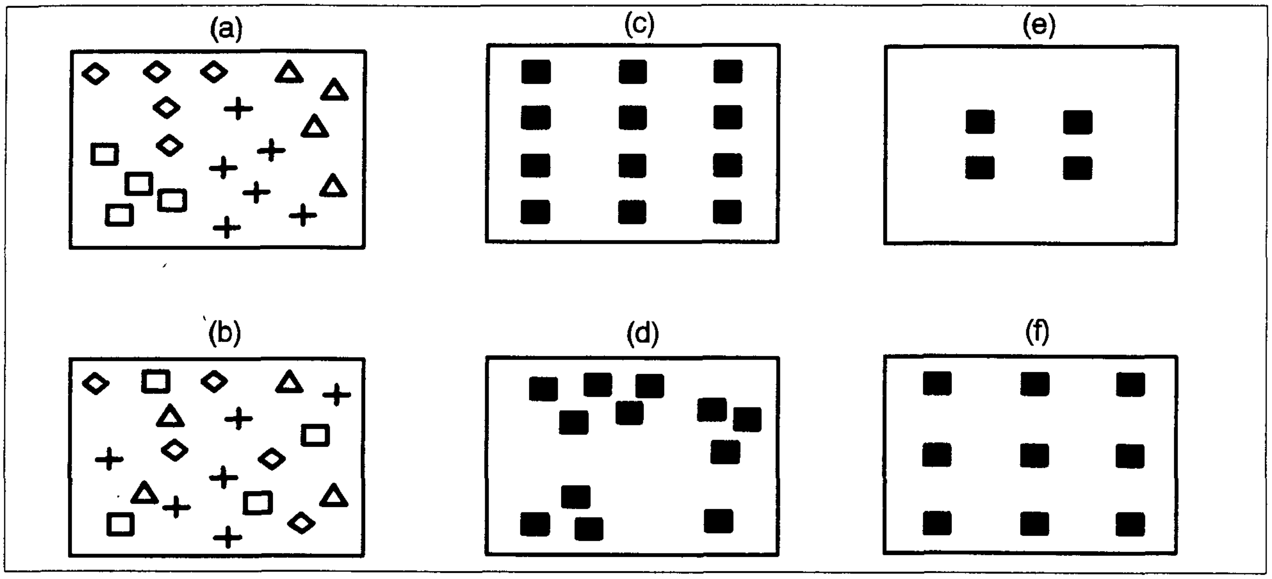


Figure 2. Entropies of the map plane. Image (a) and image (b) demonstrate topological entropy, image (c) and (d) demonstrate the concept of metrical entropy, whereas image (e) and (f) demonstrate positional entropy.

$$H(X) = -p^+ \cdot \log_2 p^+ - p^- \cdot \log_2 p^- \quad (3)$$

where

p^+ is the probability of (*black,black*) and (*white,white*) neighbors, while

p^- is the probability of (*black,white*) and (*white,black*) neighbors.

Applying this technique to the images of Figure 1, we get $H_a(X) = 0.825$ and $H_b(X) = 0.301$. Image (b) now has lower entropy than image (a) which puts the images into a sequence corresponding to our visual judgment. The difference technique described above, in the context of point symbol maps, is elaborated on in the next section. The technique is also demonstrated in the choropleth map example later in this article.

Gatrell (1977) proposes computing the entropy of a binary image as a weighted mean value of the entropy at the different orders of neighborhood. The computation can be done by applying Equation (3) to the different orders of neighborhood. We can set up the equation:

$$H(X) = \sum_{k=0}^{n} w(k) \cdot H(X)_k \quad (4)$$

where

$w(k)$ is a weight function, and

k is the order of neighborhood.

Equation (4) has some conformity with the joint entropy in Equation (15) (Appendix). If $w(1) = w(2) = \cdots = w(k) = 1$ and the different levels are independent in the probabilistic sense, the equation corresponds to the joint entropy of the k information sources. The weight function is used to control the size of the neighborhood to be evaluated. A high value of k corresponds to a global neighborhood, while a small value corresponds to a local neighborhood.

Map Information Sources

Information theory deals with variation. Therefore, when applying the theory to cartography we should carefully identify the elements which make up the variation of a map. As stated earlier, this article deals mainly with communication problems at the syntactic level of cartographic communication. Therefore, our identification of information sources concerns only the syntactic properties of the map. For the coming discussion we need a definition of the terms map entity and map information source.

Definition 1
A map entity can be a map symbol, a part of a map symbol, groups of map symbols, an attribute of a map symbol, or a derived characteristic of a map which can serve as an entity for entropy computations.

Definition 2
A map information source, denoted by (X, C), is an object which contains a set X of map entities and a characteristic C of them which make up their variation.

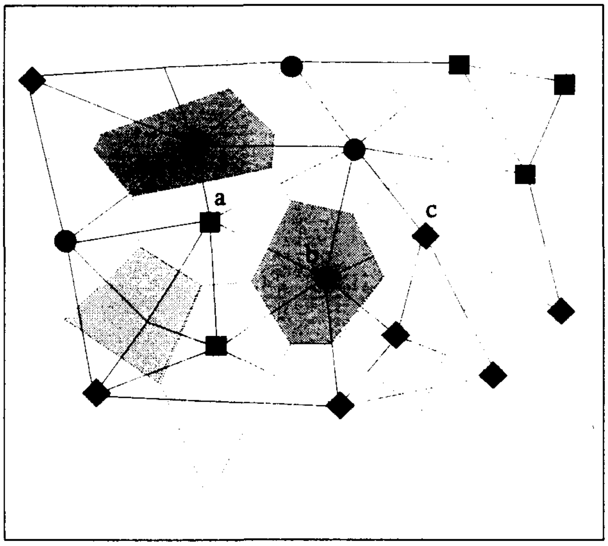


Figure 3. A point symbol map and its Thiessen polygons.

The visual variables identified by Bertin (1981) will serve as a basis for our classification of map information sources. According to Bertin, the variables which are used to manipulate the map symbols are: X,Y (the two dimensions of the plane), size, value, texture, color, orientation, and shape. Bertin operates with two components of the map plane, the X and Y co-ordinates. In entropy computations it is more appropriate to distinguish between three components of the map plane as illustrated in Figure 2. A first entropy is derived from images (a) and (b) in Figure 2. From a visual point of view it is clear that in Figure 2, map (a) is more ordered than map (b), but the number of different map symbols and the (X,Y) positions occupied by the set of symbols are equal in both the maps. The entropy of the kind considered in images (a) and (b), will be termed *topological* entropy.

Definition 3
The topological entropy of a map considers the topological arrangement of the map entities.

A second entropy which can be derived from images (c) and (d) in Figure 2 is the *metrical* entropy of a map.

Definition 4
The metrical entropy of a map considers the variation of the distance between the map entities. The distance is measured according to some metric.

A third type of entropy which can be derived from images (e) and (f) in Figure 2, is *positional* entropy.

Definition 5
The positional entropy of a map considers all the occurrences of the map entities as unique events. In the special case that all the map events are equally probable

$$H(X) = \log_2 n$$

where
n *is the number of entities.*

The term positional entropy is motivated from its relation to the number of positions occupied by the map entities. If we assume that each map entity occupies one position, the positional entropy is simply computed from counting the number of map entities. Our definition of topological entropy and metrical entropy correspond to the definitions of Bjørke (1994) while the definition of positional entropy corresponds to his definition of density entropy.

The computation of topological entropy and metrical entropy of the point symbol maps in Figure 2 requires a spatial concept. There may be several strategies which can be applied to this, but I propose a method similar to that used for the binary image case in Figure 1. Imagine a point symbol and some neighboring symbols. We will define the *visual area* of a point symbol as its Thiessen polygon. Since a Delaunay triangulation is the dual of a set of Thiessen polygons (Lee and Schachter 1980), we will base the neighborhood definition in a point symbol map on a Delaunay triangulation. This idea is demonstrated in Figure 3. A Thiessen polygon is constructed around each map symbol. Therefore, the map symbols are nodes in a network created by a Delaunay triangulation. Given two nodes in the network created by the Delaunay triangulation, the order of the neighborhood is computed by counting the number of edges on the shortest path, by the number of links between the points considered. For example, point (b) is a 1st order neighbor of point (a), whereas point (c) is a 2nd order neighbor of point (a). Since we have a strategy to define neighbors, we can apply the difference technique of the binary image in Figure 2. The topological entropy is based on computing the probability of different types of binary relations between the map symbols. In Figure 2, for example, we get the set E of entities (relations):

$$E = \begin{Bmatrix} E_{11} & E_{12} & E_{13} & E_{14} \\ E_{21} & E_{22} & E_{23} & E_{24} \\ E_{31} & E_{32} & E_{33} & E_{34} \\ E_{41} & E_{42} & E_{43} & E_{44} \end{Bmatrix} \quad (5)$$

where the indices 1,2,3,4 represent the four different map symbols in images (a) and (b).

The definition of the entities in Equation (5) is more complete than the definitions in Equation (3), since in Equation (5) the symmetry *(black,white)*, *(white,black)* and *(black,black)*, *(white,white)* is regarded as distinct events. If the 0-th order neighborhood is only considered, we get the subset:

$$E^0 = \{E_{11}, E_{22}, E_{33}, E_{44}\} \qquad (6)$$

which corresponds to the selection of entities proposed in Knöpfli (1983). Applying the method considered to different orders of neighborhood, we get a set of entropies. A mean value for the set can be computed as a weighted sum of the entropies at different orders of neighborhood (Equation (4)). For the metrical entropy of the maps in Figure 2, we can simply calculate the Euclidian distance between the neighboring map symbols and apply the distance *differences* rather than the distance values themselves as entities. As with the topological entropy, the metrical entropy can also be computed at different orders of neighborhood.

Equation (5) shows a relation between topological entropy and the visual variable shape. In this case the differential variable shape does distinguish between the sixteen elements of set E. To be more definite, the visual variables: size, value, texture, color, orientation and shape belong to the attribute domain of the map. A class name for a specific group of map information sources will be introduced.

Definition 6
Map information sources are orthogonal if none of the information sources can be derived from combining some of the other information sources.

Definition 7
The topological, metrical and positional entropies have orthogonal map information sources; which are information sources of the spatial domain of a map.

Definition 8
The visual variables as: size, value, texture, colour, orientation and shape have orthogonal map information sources; which are information sources of the attribute domain of a map.

Similarity Grade, Transition Probability and Equivocation

If the map user is uncertain about the map symbols actually received, this uncertainty is defined as equivocation (Knöpfli 1983). Knöpfli clearly shows that the "visual distance" between the map symbols is important for the perception of the symbols, i.e., at a small visual distance there is a chance that one symbol is interpreted as another symbol. For example, if two lines A and B are very close to each other, it may be difficult to visually separate the one line from the other. Therefore, some parts of line A may be interpreted as line B. Another example is that if two symbols have similar colors, the color of one symbol may be interpreted as the color of the other symbol. If the map designer planned to distinguish between the two colors, the similarity in color may cause confusion for the map reader. The perceived similarity between map symbols calls for a definition:

Definition 9
Let x and y be map entities. A function $\mu(y,x)$ which defines the grade of perceived similarity between x and y will be termed the similarity function. The similarity is measured on the interval [0,1] of real numbers. If x and y are clearly separable, the similarity grade is 0. If x and y are completely unseparable, the similarity grade is 1. Generally, $\mu(y,x) \neq \mu(x,y)$.

The computation of the similarity function for a particular map information source is not a trivial task because the perceived similarity between map entities may be influenced by several types of phenomena. For example, Gilmartin (1981) shows that the perceived size of a circle may be biased by its map context and Robinson et al. (1995, 398) point out that the perceived size of a line is influenced by its background color. Methods of computing the similarity function and of computing which perceptual phenomena to consider are mainly outside the scope of this article.

In equivocation computations we need to know the transition probabilities, i.e., the conditional probabilities in Equations (11) or (12) (Appendix). Similarity grade and transition probability are related to each other, but they are different. Our definition of the similarity function corresponds to the definition of the membership function in fuzzy set theory (for an explanation of fuzzy set theory, see, e.g., Klir and Folger (1988)). In fuzzy set theory the membership function assigns a value to the members of the set. The membership function by which a set A is defined has the form

$$\mu_A : X \rightarrow [0, 1]$$

where
[0,1] denotes the interval of real numbers from 0 to 1, inclusive.

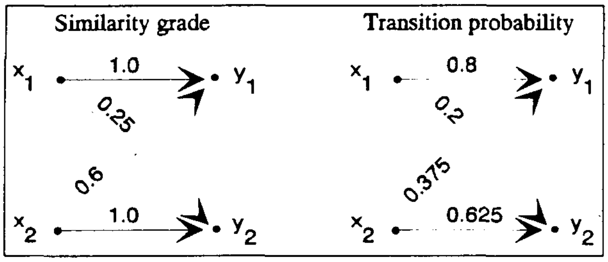


Figure 4. Similarity grade and transition probability (conditional probability).

The grade of membership of an element x in A is written as $\mu_A(x)$. Sometimes $\mu_A(x)$ is termed the possibility that x is a member of A. The concept of possibility and probability are both used to represent and manipulate imprecision or uncertainty. In everyday speech the terms possibility and probability are sometimes used interchangeably. There is a fundamental difference between possibility and probability, however. For example, the probabilities must sum to 1 whereas the possibility values are not restricted in such a way.

Consider a set X of map entities and a set Y of perceived map entities and the relation $R(X,Y)$ defined by the Cartesian product $X \times Y$. Formally, $X \times Y = \{(x,y) \mid x \in X \text{ and } y \in X\}$. Let every tuple (x,y) of the relation be assigned a similarity grade, i.e., the tuples may have varying degrees of membership within the relation. Since a map entity should be similar to itself by the maximum rate of similarity, the tuple (x,y) is given the membership grade 1 if $x=y$. From $R(X,Y)$ we select the similarity class $x \times Y = \{(x,y) \mid y \in Y\}$ and compute the class sum $\Sigma_{y \in Y} \mu(y \mid x$. Our definition of similarity class has the properties of similarity class in fuzzy set theory. In fuzzy set theory a similarity class is defined as a fuzzy set in which the membership grade of any particular element represents the similarity of that element to the element x (Klir and Folger 1988, 83). The mapping from grade of similarity to transition probabilities (conditional probabilities) can be done for each similarity class as:

$$p(y \mid x) = \frac{\mu(y \mid x)}{\Sigma_{y \in Y} \mu(y \mid x)} \text{ for each } y \in Y \qquad (7)$$

where the notation $\mu(y \mid x)$ is equivalent to the notation $\mu(y,x)$ and should be read as: the grade of membership of the relation from x to y.

Applying Equation (7) to all the classes of X ensures that:

$$\sum_{y \in Y} p(y \mid x) = \text{ for each } 1x \in X$$

For example, consider Figure 4 and the two map symbols x_1 and x_2. The grades of similarity $\mu(y_1 \mid x_1) =$ and $\mu(y_2 \mid x_1) = 0.2$ are mapped to conditional probabilities as:

$$p(y_1 \mid x_1) = \frac{1}{1+0.25} = 0.80 \text{ and } p(y_2 \mid x_1) = \frac{0.25}{1+0.25} = 0.2$$

Equation (7) has, in the worst case, the computational effort $T = O(n^2)$ when applied to all similarity classes of X (n is the number of elements of X). Usually, the conflict between map entity x and its neighboring map entities is limited to the neighbors inside a small region around x. Therefore, we can substitute Y in Equation (7) with

$$Y_{s(x)} = \{y \in Y \mid y \text{ inside } s(x)\}$$

where
$s(x)$ defines a search region around x.

The computation of entropy, equivocation and useful information will be demonstrated from the transition probabilities in Figure 4. We assume that the probabilities of the two map entities x_1 and x_2 are $p(x_1) = 0.3$ and $p(x_2) = 0.7$. Then the probabilities of the perceived map entities y_1 and y_2 are computed as:

$$p(y_1) = 0.3 \cdot 0.8 + 0.7 \cdot 0.375 = 0.503 \text{ and}$$
$$p(y_2) = 0.3 \cdot 0.2 + 0.7 \cdot 0.625 = 0.497$$

As a control, $p(y_1) + p(y_2) = 1$. . The entropy of the perceived map entities is computed from Equation (9) (Appendix):

$$H(Y) = -0.503 \log_2 0.503 - 0.497 \log_2 0.497 = 0.99997$$

The equivocation is computed from Equation (12) (Appendix). The summations will be broken into small steps which makes it easier to interpret the equation. The uncertainty in the perceived entities when entity x_1 is sent is:

$$H(Y \mid x_1) = -0.8 \log_2 0.8 - 0.2 \log_2 0.2 = 0.72193$$

The uncertainty in the perceived entities when entity x_2 is sent is:

$$H(Y \mid x_2) = 0.375 \log_2 0.375 - 0.625 \log_2 0.625 = 0.95443$$

The mean uncertainty in the perceived map is:

$$H(Y \mid X) = 0.3 \cdot 0.72193 + 0.7 \cdot 0.95443 = 0.88468$$

Finally, the useful information is computed from Equation (17) (Appendix):

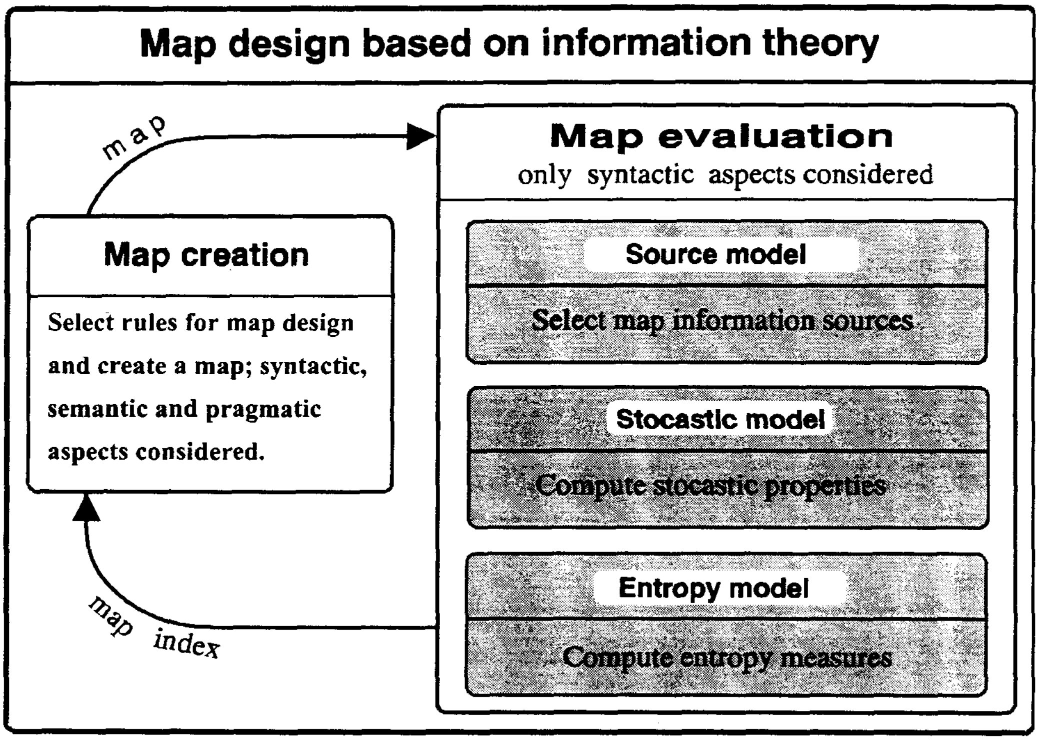


Figure 5. Map design based on information theory.

$$R = H(Y) - H(Y|X) = 0.99997 - 0.88468 = 0.11529$$

Since we have a noisy channel, the entropy of the information source is different from the entropy of the received signals, i.e., $H(X) \neq H(Y)$. The entropy $H(X)$ of the information source is $-0.3\log_2 0.3 - 0.7\log_2 0.7 = 0.88129$ whereas the entropy $H(Y)$ of the received signals is 0.99997.

Map Design Based on Information Theory

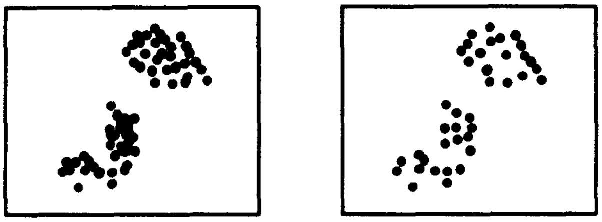
Bjørke (1994) presents a conceptual model for a map design process which incorporates information theory. The model has two main parts: a *map creation* process and a *map evaluation* process. The map creation process creates maps based on knowledge about cartographic design while the map evaluation process evaluates syntactic aspects of the maps based on information theory. The map evaluation process is composed of three operational areas. The three areas first outlined in Bjørke (1994) are renamed:

1. source model,
2. stochastic model, and
3. entropy model.

The source model describes which map information sources are to be selected, while the stochastic model describes their stochastic properties as spatial correlation and transition probabilities. Finally, the entropy model uses the source model and the stochastic model to compute different entropy measures as: R, $H(Y)$ and $H(Y|X)$. The map design process considered is presented as the data flow diagram in Figure 5. The diagram emphasizes that the map evaluation process considers only syntactic aspects of a map. Despite Shannon and Weaver's (1964, 26) assumption that information theory can be applied to all three levels of communication problems, the proposed map evaluation process is limited to only syntactic aspects of map information within the scope of this article.

An automated system based on the proposed map design model is a stepwise procedure. The map creation process generates different maps and thereafter the map evaluation process computes entropy measures for the maps. The information measures (map indexes) are sent to the map creation process, which enables it to draw conclusions about how to alter the map design in order to get more efficient maps (this is elaborated on in the examples at the end of this article). The process cycle of map creation and map evaluation terminates when the map index requirements are met.

McMaster and Shea (1992) describe a map generalization model which breaks the generalization process down into three operational areas:



(a) (b)

Figure 6. Two dot maps with different number of entities per dot.

1. why to generalize,
2. when to generalize, and
3. how to generalize.

Information theory cannot show how to generalize a map, but it can be applied for a better understanding of why and when to generalize. The three sub-processes of map evaluation in Figure 5 mostly cover the second operational area (when to generalize) of McMaster and Shea.

The relation between the two main processes in Figure 5 can be elaborated on in the context of a map evaluation method described by Morrison (1984). Morrison analyses the symbolization used on general-purpose atlas reference maps from a semiotics point of view. The simplest definition of semiotics is perhaps "the study of sign systems" (Head 1991, 240). In order to systematically evaluate the maps, Morrison (1984) states their purpose and concentrates on the semantic and the pragmatic levels of map communication. The application of information theory in the proposed map design process (Figure 5) can coexist with Morrison's evaluation strategy. Since the level A evaluations of the information theory method and Morrison's method evaluate different components of a map, they should not be set against each other. With reference to the proposed map design model, Morrison's evaluation method should be applied in the map creation process. Accordingly, information theory evaluations of the syntactic map component together with the map creation process as a whole take into account all three levels of communication problems; syntactic, semantic, and pragmatic.

Examples

The examples given here demonstrate the application of information theory to map design. These examples raise several research issues which relate to map perception. But, in order to keep the focus on the application of information theory, detailed discussions of map perception are outside the scope of this article.

Abbreviations (Top, Met, Pos) are used for topological, metrical, and positional entropies respectively (*Definitions 3, 4,* and *5*). A map information source (*Definition 2*) for a topological entropy will be written as (X,Top) for example. The different entropy measures $R(X)$, $H(X)$ and $H(Y|X)$, which are used in the examples, are explained and elaborated on in the Appendix.

Dot Map

Dot maps are often used to show the spatial distribution of discrete geographical point entities. The traditional design rules of dot maps include:

1. selection of the dot size, and
2. selection of the number of events per dot.

Figure 6 shows two dot maps and demonstrates the significance of the second design rule, since map (a) has a lower number of entities per dot than map (b). The evaluation model proposed does not consider the spatial correlation of the dots. Therefore, a rather simple model can be set up.

Design Goal

We assume that process map creation (Figure 5) has set up the following design goals:

1. make the number q of events per dot as small as possible, i.e., as many dots as acceptable to visual perception; and
2. the preferable dot diameter should be S_0.

Source Model

Select the map information source:

$$(X, \mathrm{Pos})$$
$$X = \{x \mid \text{element } x \text{ is a dot}\}$$

i.e., X is the set of all dots on the map.

If aspects of spatial correlation are to be considered in the map evaluation process, an information source for the metrical entropy can be selected as a second information source. In that way we can compute a map index from two orthogonal information sources (*Definition 6*):

1. metrical, and
2. positional.

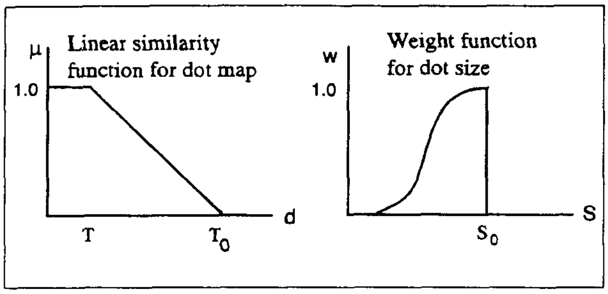


Figure 7. Functions for dot map design.

Stochastic Model
The dots are assumed to be equally probable, i.e.,

$$p(x) = \frac{1}{N_x} \text{ for each } x \in X$$

where
N_x is the number of elements in X.

The transition probabilities are not as easily derived. Let us assume that we can set up a model so that the visual separation between two neighboring dots x and y is a function of the distance $d(x,y)$ between them and the dot size S. Further, let us assume that visual separation and visual similarity are inverse quantities. Hence, the similarity function (*Definition 9*) can be defined as $\mu(x,y) = f(S, d(x,y))$. An example of a linear similarity function is given by Figure 7. In the figure the grade of similarity $\mu(x,y) = 1$ if $d(x,y) \leq T$, i.e., when the dots are so close to each other that they cannot be separated. If $d(x,y) \geq T_0$ the dots are clearly separable and $\mu(x,y) = 0$. When the similarity function is defined, the transition probabilities which we need for the entropy computations can be derived from Equation (7). The design of the similarity function should consider the resolution and the type of the output media, the color of the dots and other parameters related to map perception. A more detailed discussion of this specific topic is outside of the scope of this article.

Entropy Model
We assume a map creation process $M(q,S)$ which produces dot maps by varying the number q of events per dot and varying the dot size S. The first statement of the design goal can be modeled by $\max[R(X) \mid M(q,S)]$, which is the maximum value of the useful information of map information source under the constraint that the different map alternatives are produced by $M(q,S)$. The second statement of the design goal can be satisfied by a weighted entropy computation. Hence, the map index K can be computed from $K = w(S) \cdot R(X)$ where $w(S)$ is a weight function which takes the dot size as a variable. An example of a weight function is given in Figure 7. From the figure, we can see that the weight has its maximum value at the preferred dot size S_0, i.e., $w(S_0) = 1$. There is no good reason to consider a dot size greater than S_0. Therefore, the weight function is designed so that $w(S) = 0$ when $S > S_0$.

Selection Criterion
Select the pair (q,S) which corresponds to the maximum value of the map index, i.e.,:

$$K_{\max} = \max[w(S) \cdot R(X \mid M(q,S))]$$

where
$R(X)$ is computed from Equation (17) (Appendix).

In order to reduce the computational effort of the process cycle, some strategy to eliminate maps which are not candidates for the best solution should be implemented.

Contour Map

An information theory approach to the selection of an appropriate contour interval in contour maps is proposed by Bjørke (1993). The following example expands this proposal.

Design Goal
Make the contour interval e as small as possible, i.e., as many contour lines as are acceptable to visual perception.

Source Model
Select the map information source:

$$(X, \text{Pos})$$
$$X = \{x \mid h(x) = ne \wedge H_{\min} \leq ne \leq H_{\max} \wedge n \in I\}$$

where
$h(x)$ is the height value of contour line x, and
$H_{\min}$ and $H_{\max}$ are the minimum height and maximum height on the map, respectively.

The model says that X is the set of contour lines of a map with contour interval e.

Stochastic Model
It seems reasonable that a long contour line should have a higher probability than a short contour line. Therefore, we will select the model:

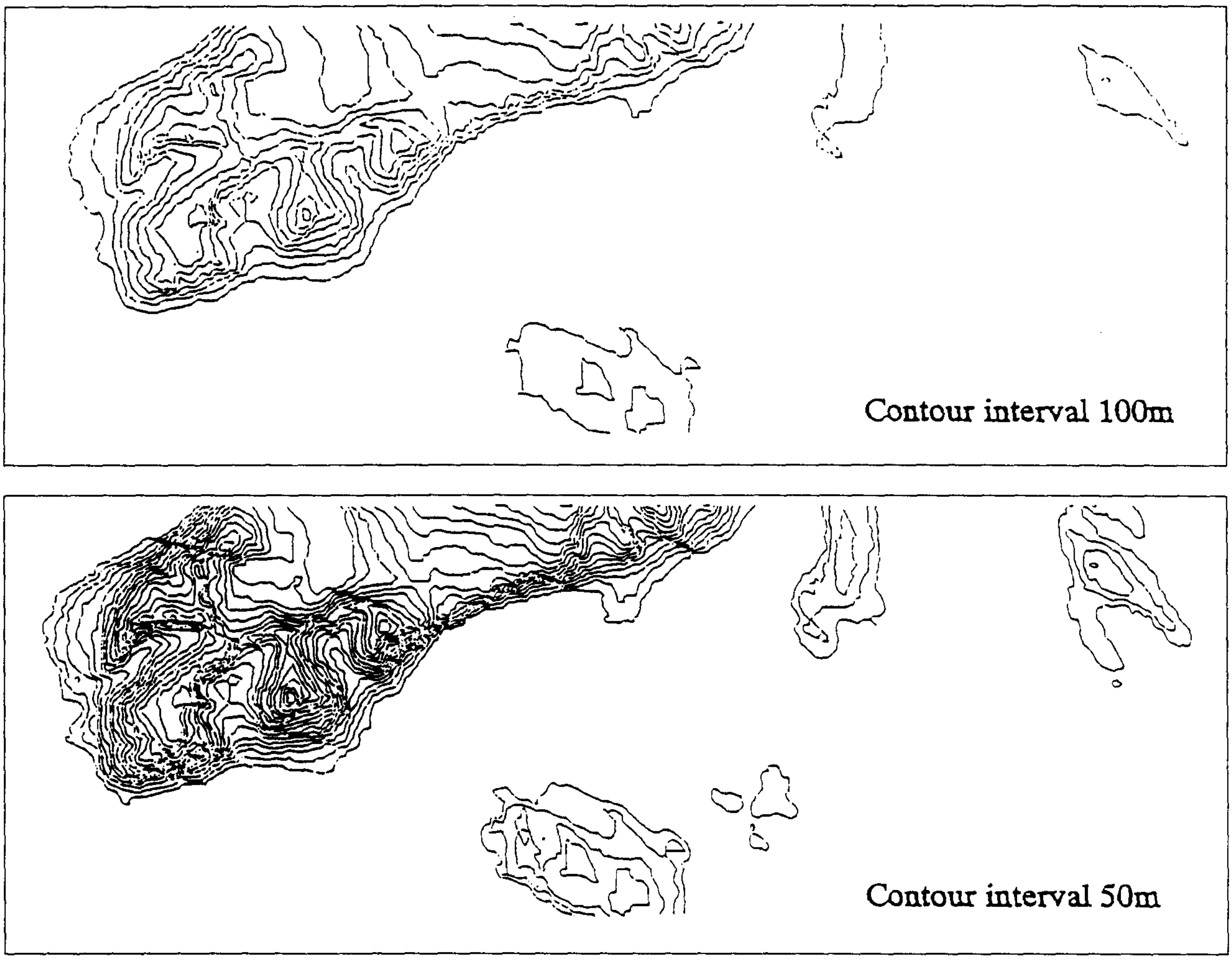


Figure 8. Some maps created in the contour map experiment. Map scale 1:120 000.

$$p(x) = \frac{l(x)}{\Sigma_{x \in X}\, l(x)} \text{ for each } x \in X$$

where

$l(x)$ is the length of contour line x, and

Σ computes the total length of all the contour lines.

We assume that the similarity between neighboring contour lines can be modeled by a function of the type used in the dot map example, i.e., the similarity between two lines is zero when the distance between the lines is greater than T_0. With the exception of parallel lines, the distance between two contour lines will vary. Therefore, the similarity between two contour lines can be computed as a mean value for different sections of the lines.

Entropy Model

We select a map process $M(e)$ which creates maps with varying contour intervals under the constraint $e \in E$. The constraint can, for example, limit e to integer values only or to real values which are easy to remember. The design goal will be evaluated against the useful information of the maps, i.e., the map index is computed as $K=R(X)$.

Selection Criterion

Select the contour interval which corresponds to

$$K_{\max} = \max\,[R(X \mid M(e))]$$

An experiment based on the model above was carried out on a digital terrain model (DTM) of a small part of Norway. Table 1 summarizes the experiment and shows the map index at different contour intervals for some selected maps. Two of the maps generated from process $M(e)$ are shown in Figure 8; e=100m and e=50m. The computations in Table 1 assume the map scale 1:120,000; and a linear similarity function (Figure 7) with T=0.1mm and T_0=0.4mm. In this experiment, the map index reached the maximum value 3.275 at contour interval 49m. One should note that the

contour interval m	map index K=R(X)	entropy H(Y)	equivocation H(Y|X)
150	2.254	2.254	0.000
125	2.519	2.520	0.001
100	2.833	2.854	0.021
75	3.150	3.292	0.142
60	3.269	3.628	0.359
55	3.273	3.757	0.484
49	**3.275**	**3.918**	**0.643**
48	3.260	3.951	0.691
46	3.244	4.013	0.769

Table 1. Map index at different contour intervals. Map scale 1:120,000

selection of parameter values in the similarity function has great influence on the equivocation computation. A more detailed discussion of this issue related to map perception is outside the scope of this article.

The experiment demonstrates a property of the evaluation method. At a high value of *e* the contour lines can easily be separated, but there are few of them. On the other hand, at a low value of *e* we have the opposite situation. Our model considers this property of the maps and makes a balanced selection between grade of entropy and grade of equivocation. At the optimum value of *e*, we have in Table 1 the equivocation 0.643 and the entropy 3.918, which corresponds to the maximum value of K=3.918-0.643=3.275. Hence, a property of our selection criteria is that the optimum choice is not necessarily a map with zero equivocation.

Line Generalization

An information theory approach to the selection of appropriate parameters in line generalization algorithms is presented in Bjørke (1992). The approach selects a source model based on angular change. Saga (1994) discusses this approach and demonstrates that it is too simplistic to base the selection of generalization parameters on angular change only. Saga (1994) shows that structural information should be considered as well. The complexity of line generalization and the fact that a number of fundamental problems are still unsolved are pointed out by several authors (Li and Openshaw 1993; Wang and Muller 1993).

The present example deals with line simplification. "Simplification is necessary to eliminate unwanted details (such as small wobbles along lines) that would be difficult or impossible to perceive after scale reduction" (Wang and Muller 1993, 105). The problem we will put into focus is that of how to set up a map evaluation model that can assist us in the selection of an appropriate grade of simplification. Our example will not be connected to a specific line simplification algorithm, since in principle any simplification algorithm can serve as a basis for the map creation process (Figure 5). The model is not complete since it still is under investigation. It is hoped that it can be looked upon as an innovative framework for further research in this area. We assume the following design goal:

Design Goal

Keep as much of the variation along the line as is acceptable to visual perception.

Source Model

How to compute entropy measures for a line is not a trivial task. Nevertheless, we will select the two information sources:

$$(A, Met)$$
$$(X, Pos)$$

where

A = $\{\alpha \in \Re \mid$ element α is a break angle of the digitized line$\}$

X = $\{x \mid$ element x is a δ-circle of the line$\}$

The elements of X are some derived entities, which we term δ-circles. The concept of δ-circles is demonstrated in Figure 9. The circles are of equal size and distributed along the line according to the following rules:

1. the circle centers are located on the line,
2. the distance d between the circle centers is constant when measured along the line, and
3. the diameter δ of the circles is equal to d.

Our δ-circles have some similarity to ∈-circles in Perkal (1966), since both are used to represent some minimum quantities. The diameter of the δ-circles will influence the values of the entropy measures. An appropriate circle size is supposed to consider visual limitations as least perceivable winding etc. The size of δ-circles as a specific topic is an issue for further research.

Stochastic Model

$$\mathbf{P}_{\alpha} = (p(\alpha) \mid 0 \le \alpha \le 2\pi)$$

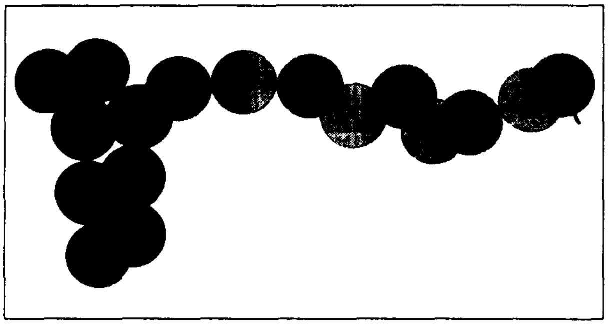


Figure 9. δ-circles of a line.

where

$\mathbf{P}_\alpha$ is a probability distribution

$$p(x) = \frac{1}{N_x} \text{ for each } x \in X$$

where

N_x is the number of δ-circles of the line.

The transition probabilities of the two information sources can be computed using a strategy similar to that of the dot map example.

Entropy Model
We select a map process $M(t)$ which creates different versions of lines by varying the generalization parameter t. The map index is computed as a weighted sum for the R-values of the two map information sources:

$$R = w_\alpha \cdot R(A) + w_x \cdot R(X)$$

where

w represents a weight.

Since α is a continuous varying variable, the entropy computation is based on:

$$H(\alpha) = -\int_0^{2\pi} p(\alpha)\log_2 p(\alpha) d\alpha \approx -\sum_{i=1}^{n} p(A_i)\log_2 p(A_i)$$

where the approximation of the integral is based on dividing the continuous domain in n discrete classes, i.e., $A = \cup_{i=1}^{n} A_i$.

Selection Criterion
We assume that the design goal is met at the maximum value of the map index:

$$K_{\max} = \max\left[w_\alpha \cdot R(A \mid M(t)) + w_x \cdot R(X \mid M(t))\right]$$

The scope of the present model is not to give a complete set of constraints to control the complex line generalization process, but rather to demonstrate properties of information theory. Therefore, the information theory model presented calls for further research in order to achieve a successful cartographic adaptation.

Choropleth Map

A statistical surface can be visualized in several ways. One such method is a choropleth representation (Robinson et al. 1995). The traditional design rules of choropleth maps include:

1. selection of the number of classes, and
2. determination of class limits.

Bjørke (1992) presents an information theoretic approach to compute an optimum number of classes in choropleth maps. Based on this proposal, the following model is set up:

Design Goal
Select as many classes as are acceptable to visual perception, i.e., seek an optimal solution for how much variation of the statistical surface can be portrayed on the map.

Source Model
Assume a raster map. Select the map information source:

$$\overset{(X,\text{Top})}{X = \{x_{lr} \mid (l,r) \in H^2}$$

$$\mid \wedge (\text{element } x \text{ is an edge between two adjacent pixels})\}$$

where

x_{lr} represents the color of the left-hand pixel and the right-hand pixel of x

H is the set of different colors of the map, and

H^2 is the Cartesian product $H \times H$.

If the map has black and white pixels only, we have: $H = \{black, white\} = \{b, w\}$. The set of entities in this case: $X = \{x_{bw}, x_{bb}, x_{wb}, x_{ww}\}$.

Stochastic Model

$$p(x_{lr}) = \frac{N(l,r)}{\sum_{(l,r)\in C^2} N(l,r)} \text{ for each } (l,r) \in H^2$$

where

$N(l,r)$ is the number of edges with the color attribute (l,r)

Entropy Model
Consider the map process $M(h)$ which generates choropleth maps with a different number h of classes. The map index to be computed is:

The correlated map			
Class no.	**H(Y)**	**H(Y|X)**	***R***
3	2.04	0.16	1.88
4	2.71	0.51	2.20
5	**3.32**	**0.98**	**2.34**
6	3.92	1.71	2.21

The random map			
Class no.	**H(Y)**	**H(Y|X)**	***R***
3	2.48	0.16	2.32
4	**3.24**	**0.63**	**2.61**
5	3.84	1.28	2.56
6	4.33	2.29	2.05

Table 2. Map statistics at different class numbers. The bold face numbers indicate the level of the channel capacity.

$$K = R(X \mid M(h))$$

Selection Criterion
Select the number of classes which corresponds to $K_{max} = \max[R(X \mid M(h))]$. In Bjørke (1992), the transition probabilities were estimated from an investigation in which 30 subjects were asked to distinguish between different gray values on some test plates. Based on the transition probabilities from the investigation above, entropy measures are computed for two choropleth maps; one map has a correlated spatial distribution of the classes, while the other map has a random spatial distribution of its classes. The entropy, equivocation, and the useful information are computed at different class numbers. Table 2 shows the results of the computation. The table demonstrates that the correlated map gets its maximum value of $R=2.34$ in five classes while the random map gets its maximum value of $R=2.61$ in four classes.

Area Elimination

Elimination routines can be used to simplify area features. The criteria may be

1. minimum feature size, or
2. proximity to neighboring features (Robinson et al. 1995, 466).

Assume a map with equally sized area features (Figure 10). Due to exaggeration, as a part of map generalization, the map symbols may overlap or may be very close to each other. This is often a problem in small scale maps, and is the case for house symbols in the 1:50,000 topographic maps from the Norwegian Mapping Authority.

Design Goal

1. Keep as much of the variation as is acceptable to visual perception;
2. eliminate features by proximity to neighboring features.

Source Model
Select the map information source:

$$(X, \text{Pos})$$
$$X = \{x \mid \text{element } x \text{ is an area feature}\}$$

Stochastic Model
Since the features are assumed to be of equal size, their probabilities are modeled as:

$$p(x) = \frac{1}{N_x} \text{ for each } x \in X$$

where
N_x is the number of features.

The transition probabilities can be derived in a similar fashion as in the dot map example.

Entropy Model
The second requirement of the design goal can be met by eliminating the feature x which has the greatest local equivocation, i.e., x corresponds to $\max_{x \in X}[H(Y \mid x)]$. The first design requirement can be met by maximizing $R(X)$. Therefore, two map indexes should be sent to the map creation process, one index which corresponds to the local equivocation and another index corresponding to the useful information of the map as a whole.

Selection Criterion
Eliminate the feature which corresponds to

$$K_x = \max_{x \in X}[H(Y \mid x)]$$

Select the map which corresponds to

$$K_{max} = R(X \mid M(x))$$

where
$M(x)$ is a process which eliminates from the map.

The computation of $H(Y \mid x)$ is based on the last Σ in Equation (12) (Appendix): $H(Y \mid x) = \sum_{y \in Y} p(y \mid x)\log_2 p(y \mid x)$. For each time an area feature is eliminated from the map, a new

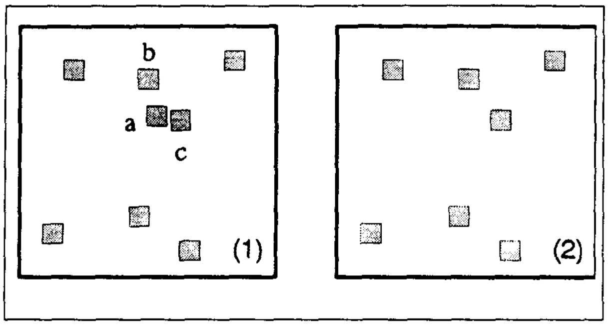


Figure 10. Simplification by area elimination. Feature a is a candidate for elimination in map (1). This feature is eliminated in map (2).

candidate to be eliminated should be computed. The process terminates when $R(X)$ receives its maximum value.

The computation of a candidate to be eliminated will be illustrated. Related to Figure 10, assume the following similarities: $\mu(a \mid a) = \mu(b \mid b) = \mu(c \mid c) = 1$, $\mu(a \mid b) = \mu(b \mid a) = 0.1$ and $\mu(a \mid c) = \mu(c \mid a) = 0.4$. All other similarities are assumed to be zero. The corresponding transition probabilities are computed from Equation (7):

$$p(a|a) = 0.667,\ p(b|a) = 0.067,\ p(c|a) = 0.266,$$
$$p(b|b) = 0.909,\ p(a|b) = 0.091,\ p(c|c) = 0.714,$$
and
$$p(a|c) = 0.286.$$

The local equivocations are computed as:

$$H(Y|a) = -0.667\log_2 0.667 - 0.067\log_2 0.067$$
$$-0.266\log_2 0.266 = 1.16$$

Similarly, $H(Y|b) = 0.44$ and $H(Y|c) = 0.86$ give the priority list for feature elimination: (a,c,b), i.e., feature a is to be eliminated since it generates a higher local equivocation than c and b.

Conclusions

This article describes a possible role for information theory in automated map design or automated map generalization and demonstrates how different entropy measures can be used as control parameters in optimization at the syntactic level of cartographic communication. This entropy-based map design methodology has two main components: map creation, and map evaluation.

The presented model breaks the entropy-based map evaluation down into three operational areas: source model, stochastic model, and entropy model. Several map examples demonstrate how these three operational areas are used to structure the design of map evaluation models. These map examples: dot map, contour map, line generalization, choropleth map, and area elimination, show that information theory has a potential for automated information systems. But they also demonstrate that definition of the similarity functions is not a trivial task. The application of information theory calls for further research into map perception by users.

The successful application of information theory in map evaluation is likely to be based on a successful solution of the following steps:

1. the selection of map information sources,
2. the modeling of spatial correlation, and
3. the modeling of similarity.

AKNOWLEDGEMENTS

The author wishes to thank the three anonymous reviewers for their comments and suggestions on an earlier draft of this article. The author is also grateful to Kjell O. Roksvåg and Henrik Thorenfeldt who prepared the computer program for the contour line example.

REFERENCES

Bertin, J. 1981. *Graphics and Graphic Information Processing*. Berlin: Walter de Gruyter.

Bjørke, J. T., and R. Aasgaard. 1990. Cartographic zoom. *Proceedings of the 4th International Symposium on Spatial Data Handling*, Zurich. Columbus, Ohio: International Geographical Union (IGU), The Ohio State University, 345-53.

Bjørke, J. T. 1992. Towards a formal basis for cartographic generalization. In J. T. Bjørke ed. *Proceedings of Neste generasjon GIS*. Trondheim, Norway: The Norwegian Institute of Technology, Department of Surveying and Mapping, 101-09.

Bjørke, J. T., and T. Midtbø. 1993. Generalization of digital surface models. *Proceedings of the 16th International Cartographic Conference*, Cologne, 363-371.

Bjørke, J. T. 1994. Information theory: The implications for automated map design. In M. Konecny, ed. *Supplement to proceedings of the conference Europe in Transition: The Context of geographic information systems*. Brno: Laboratory on Geoinformatics and Cartography, Department of Geography, Faculty of Science, Masaryk University, 8-19.

Buttenfield, B., and R. B. McMaster. 1991. *Map generalization: Making rules for knowledge representation*. U.K.: Longman.

Gatrell, A. G. 1977. Complexity and redundancy in binary maps. *Geographical Analysis* IX:29-41.

Gilmartin, P. P. 1981. Influences of map context on circle perception. *Annals of the Association of American Geographers* 71(2): 253-58.

Head, C. G. 1991. Mapping as language or semiotic system: review and comment. In D. M. Mark and A. U. Frank eds.

Cognitive and linguistic aspects of space. Dordrecht: Kluwer, 273-62.
Klir, G. J., and T. A. Folger. 1988. *Fuzzy sets, uncertainty, and information*. New York: Prentice Hall.
Knöpfli, R. 1980. Informationstheoretische betrachtungen zur verringerung von ungevissheit durch struktur. *International Yearbook of Cartography*, 129-41.
Knöpfli, R. 1983. Communication theory and generalization. In D. R. F. Taylor ed. *Graphic communication and design in contemporary cartography*. New York: Wiley, 177-218.
Lee, D. T., and B. J. Schachter. 1980. Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer and Information Sciences* 9(3): 219-42.
Li, Z., and S. Openshaw. 1993. A natural principle for the objective generalization of digital maps. *Cartography and Geographic Information Systems* 20(1): 19-29.
Morrison, J. L. 1984. Applied cartographic communication: Map symbolization for atlases. *Cartographica* 21(1): 44-84.
Moles, A. A. 1966. *Information theory and esthetic perception*. Urbana: University of Illinois Press.
McMaster, R. B., and S. K. Shea. 1992. *Generalization in digital cartography*. Washington D.C.: Association of American Geographers.
Neumann, J. 1994. The topological information content of a map: An attempt at a rehabilitation of information theory in cartography. *Cartographica* 31(1): 26-34.
Perkal, J. 1966. An attempt at objective generalization. In J. Nystuen, ed., W. Jackowski, trans., *Discussion Paper 10*. Ann Arbor, Mich.: Michigan Inter-University Community of Mathematical Geographers.
Robinson, A. H., and B. B. Petchenik 1976. *The nature of maps*. Essays toward understanding maps and mapping. Chicago: The University of Chicago Press.
Robinson, A. H., J. L. Morrison, P. C. Muehrcke, A. J. Kimerling, and S. C. Guptill. 1995. *Elements of cartography*, sixth edition. New York: Wiley.
Saga, S. A. 1994. Structural information in line generalization. Unpublished manuscript. Department of Surveying, Agricultural University of Norway.
Salichtchev, K. A. 1973. Some reflections on the subject and method of cartography after the Sixth International Cartographic Conference. *Canadian Cartographer* 10:106-11.
Sukhov, V. I. 1967. Information capacity of a map: entropy. *Geodesy and Aerophotography* X(4): 212-15.
Shannon, C. E., and W. Weaver. 1964. *The mathematical theory of communication*. Urbana: The University of Illinois Press.
Wang, Z., and J. C. Muller. 1993. Complex coastline generalization. *Cartography and Geographic Information Systems* 20(2): 96-106.
Weber, W. 1980. Map generalization—An information science approach. In H. Opheim ed. *Contributions to map generalization proceedings*. Oslo, Norway: Norwegian Computing Center, report no. 679, 31-52. ■

Appendix

Properties of Shannon Entropy

The principles of Shannon entropy are presented in several textbooks such as Shannon (1964) and Klir and Folger (1988). This appendix briefly reviews some concepts of Shannon entropy necessary for the development of the theoretical basis of this article.

Given two sets, X and Y, we can recognize three types of entropies.

1. Simple Entropies

The first type of entropies are two *simple entropies* based on marginal probability distribution,

$$H(X) = \sum_{x \in X} p(x)\log_2 \frac{1}{p(x)} = -\sum_{x \in X} p(x)\log_2 p(x) \tag{8}$$

$$H(Y) = \sum_{y \in Y} p(y)\log_2 \frac{1}{p(y)} = -\sum_{y \in Y} p(y)\log_2 p(y) \tag{9}$$

The maximum entropy is obtained when all events are equally probable, i.e.

$$H(p_1, p_2, \ldots, p_n) \le H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) = \log_2 n$$

If the information source is continuous, the entropy computation can be expressed as:

$$H(X) = -\int_{-\infty}^{+\infty} p(x)\log_2 p(x)dx$$

2. Joint Entropy

The second type is a *joint entropy* defined in terms of the joint probability distribution on $X \times Y$,

$$H(X,Y) = -\sum_{(x,y) \in X \times Y} p(x,y)\log_2 p(x,y) \tag{10}$$

3. Conditional Entropies

Third we can identify two *conditional entropies* defined in terms of weighted averages of local conditional entropies:

$$H(X \mid Y) = -\sum_{y \in Y} p(y) \sum_{x \in X} p(x \mid y)\log_2 p(x \mid y) \tag{11}$$

$$H(Y \mid X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y \mid x)\log_2 p(y \mid x) \tag{12}$$

Based on the relation

it can be shown that:

(13)

which can be generalized to

$$H(X_1, X_2, X_3, \ldots, X_n) = H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_1, X_2) + \cdots + H(X_n \mid X_1, X_2, \ldots, X_{n-1}) \quad (14)$$

It can also be shown that

$$H(X_1, X_2, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i) \quad (15)$$

The equality holds if, and only if, the elements from the n sets are independent in the probabilistic sense. The property of Shannon entropy which follows from Equation (15) is termed the *subadditive property*. From the rules of probability, two sets X and Y are defined as independent if $p(x,y) = p(x) \cdot p(y)$ for each $x \in X$ and each $y \in Y$. If the sets X and Y are independent, their joint entropy is:

$$\begin{aligned} H(X,Y) = {} & H(p(x_1)p(y_1), p(x_1)p(y_2), \ldots, p(x_1)p(y_s), \mid \\ & p(x_2)p(y_1), p(x_2)p(y_2), \ldots, p(x_2)p(y_s), \ldots \\ & \mid \ldots, p(x_n)p(y_1), p(x_n)p(y_2), \ldots, p(x_n)p(y_s)) \\ = {} & H(p(x_1), p(x_2), \ldots, p(x_n),) + H(p(y_1), p(y_2), \ldots, p(y_s),) \\ = {} & H(X) + H(Y) \end{aligned}$$

This property is termed the *additive property* of Shannon entropy.

If the communication channel is noisy, it is not in general possible to reconstruct the original message with certainty by any operation on the received signals. The information loss in a noisy channel is termed *equivocation* (Shannon and Weaver 1964) and is expressed as a conditional entropy. Let X and Y denote the set of input signals and the set of received signals respectively. The *useful information* R is obtained by subtracting from the source entropy the average rate of conditional entropy (equivocation).

$$\begin{aligned} R &= H(X) - H(X \mid Y) && (16) \\ &= H(Y) - H(Y \mid X) && (17) \\ &= H(X) + H(Y) - H(X,Y) && (18) \end{aligned}$$

where

$H(X \mid Y)$ is the equivocation of the information source when the received signals are known, and

$H(Y \mid X)$ is the equivocation of the received signals when the signals sent are known.

The first expression measures the amount of information sent less the uncertainty of what was sent. The second measures the amount of received information less the part of this which is due to noise. The third is the sum of the entropy of the signals sent and the entropy of the signals received less the joint entropy. The *capacity* of a noisy channel corresponds to the maximum rate of the transmission and is defined as:

$$C = \max(R) \quad (19)$$

Equation (18) follows when combining Equation (13) and Equation (16) or when combining Equation (13) and Equation (17). The symmetry of Equation (16) and Equation (17) can easily be verified as follows:

Theorem

$$H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$$

Proof

$$\begin{aligned} H(X) - H(X \mid Y) &= H(X) + \sum_{y \in Y} p(y) \sum_{x \in X} p(x \mid y) \log_2 p(x \mid y) \\ &= H(X) + \sum_{y \in Y} p(y) \sum_{x \in X} \frac{p(x)p(y|x)}{p(y)} \log_2 \frac{p(x)p(y \mid x)}{p(y)} \\ &= H(X) + \sum_{y \in Y} \sum_{x \in X} p(x)p(y|x) \log_2 \frac{p(x)p(y|x)}{p(y)} \\ &= H(X) + \sum_{y \in Y} \sum_{x \in X} p(x)p(y \mid x) \log_2 p(x) \\ &\quad + \sum_{y \in Y} \sum_{x \in X} p(x)p(y \mid x) \log_2 p(y \mid x) \\ &\quad - \sum_{y \in Y} \sum_{x \in X} p(x)p(y \mid x) \log_2 p(y) \\ &= H(X) + \sum_{x \in X} p(x) \log_2 p(x) \\ &\quad + \sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2 p(y|x) \\ &\quad - \sum_{y \in Y} p(y) \log_2 p(y) \\ &= H(X) - H(X) - H(Y \mid X) + H(Y) \\ &= H(Y) - H(Y \mid X) \end{aligned}$$

which completes the proof. ■