# ANALYTICAL TOOLS AND PROCEDURES FOR GEOINFORMATION PROCESSING

J.A.R. BLAIS
**Department of Surveying Engineering**
**The University of Calgary**
Calgary, Alberta, Canada

**ABSTRACT**

With the rapidly evolving possibilities in computations and visualization as a consequence of tremendous changes in computer technology, the processing of geoinformation can be increasingly automated with adaptive procedures in varied applications. Selecting appropriate methodologies usually involves dealing with incomplete information while minimizing the necessary assumptions. One general approach which has proven successful in very different applications is using information theory. Practical applications of information theory in spectrum estimation, adaptive filter design and inverse problems will illustrate the potential for other similar applications. General conclusions and recommendations will be included.

## 1. INTRODUCTION

Fundamental changes are taking place in positioning, mapping and related fields. Angle and distance measurements are being supplemented and replaced by the recording of observations from satellites, accelerometer systems and gyroscopes along with frequency standards for digital signal processing. Analog photographs are rapidly becoming obsolete with digital imaging systems and mapping operations integrated into geographical and more generally spatial information systems. Visualization systems will soon replace conventional topographical maps with terrain rendering and spatial information display systems. These fundamental changes have far reaching implications for the analytical tools and procedures which are necessary in the data processing of observations and measurements.

The advent of modern computer technology with the ever increasing computing power and availability also has profound implications in the selection of tools and procedures. In particular, algorithms are becoming more sophisticated and adaptive procedures are most often desired in view of automating the processing of observations and measurements. The increasing sophistication has led to dealing with patterns, trends and the like in order to have adaptive procedures. These tendencies are really leading to information processing in general and even knowledge processing in specialized application areas.

Among the outstanding questions in this field of geoinformation processing are the identification and recognition of patterns, the quantification of information contents for adaptive processing and quality control, and inference procedures that can transcend the data processing requirements. In other words, what are the new requirements in terms of tools and procedures for geoinformation processing that will lead to the knowledge-based systems of the near future?

The following discussion will first consider the mathematical differences between data, information and knowledge, and then the relationships between accuracy, uncertainty and semantics. These concepts have important implications for the understanding of what is needed in categorizing and characterizing the analytical tools and procedures for geoinformation processing. As distinct from data processing, information processing deals with adaptive methods of filtering, pattern extraction and processing, dealing with incomplete information for decision support and other applications.

A brief overview of information theory which has proven very appropriate for quantifying information contents and categorizing patterns and structures will be given. Examples from three areas of applications will be discussed to illustrate the use of information theoretical principles in those contexts. General conclusions and recommendations will be included as guidelines for other areas of potential applications.

## 2. FROM DATA TO INFORMATION

Observations and measurements are well known to provide little information unless they are properly designed and adequately carried out. This clearly illustrates that data do not always imply substantial or significant information and hence proper strategies and analyses are required. Among the principal objectives of geoinformation processing, the extraction of useful information from available data ranks very high and much research and development efforts have been invested in various application fields. However, fundamental questions remain about the nature of information and its quantification or measure to decide on optimal data and information processing strategies for all kinds of applications.

In the case of distance and angle data processing for positioning purposes, the analysis is definitely simpler than when dealing with pattern identification and recognition in digital image processing. Accuracy and reliability in survey networks are well understood but the analogous concepts with digital imagery and spatial information are definitely more complex. With the current emphasis on spatial information systems (SIS), these questions are becoming more and more important and further investigations are definitely warranted.

Data are observations or measurements that are collected in order to extract some required or expected information and knowledge. Data processing using conventional algorithmic methods is well understood in physical science and engineering. Accuracy considerations are usually taken into account in the data processing to reflect the quality of observation and measurement procedures, the numerical techniques and related operations. The derived quantities are then categorized in terms of accuracy and reliability.

Information usually refers to patterns, features and the like that are normally extracted from observation and measurement data through processing. Explicit interpretations are not normally included in such patterns and tendencies as these can easily be context or application dependent. For instance, the estimation of a linear or quadratic trend between two data sequences does not necessarily include any interpretation of the inferences for the variables in question. Hence mathematical information can be considered as abstractions or derivations from data without including any semantics. Information processing in the mathematical sense is logical pattern and similar processing that would take any uncertainty or incompleteness of the information into consideration.

Knowledge would then refer to context dependent information or interpretations that are common in reasoning like processing. In other words, spatial information patterns and trends generally have different interpretations and implications for different classes of users of the information. For example, a linear or quadratic trend between two variables can be the object of numerous interpretations. Knowledge processing involves facts, rules and procedures that often come from learning experience in some specific context or environment. Knowledge acquisition, representation and processing are among the outstanding research topics in knowledge based system design and implementation.

In terms of abstraction and complexity, information processing problems range from data processing algorithmic problems to knowledge related questions. The sequel will only consider the selection and analysis of appropriate analytical tools and procedures for information processing purposes. Open problems in geoinformation processing abound and further research and development work is required for wide ranging applications. Such experience with information processing should be useful in contemplating the problems of knowledge acquisition, representation and processing.

## 3. OVERVIEW OF INFORMATION THEORY

From a theoretical perspective, the origins of information theory go back to the foundations of probability theory as dealing with uncertain or incomplete information is at the very basis of probabilistic considerations. Measuring or quantifying information contents is fundamental in formulating optimal solutions for estimation and inference problems. Depending upon the specific requirements, some information measures and related discrimination functions may be more appropriate than others.

Information measures are often expressed in terms of frequencies of occurrence of errors or grey levels as these provide a general approach to information contents without necessarily any interpretation or evaluation of the implications. Various information measures have been suggested and used in different application contexts. For digital image processing and related applications, the Shannon-Wiener entropy H[p] in terms of discrete frequencies or probabilities $p = [p_1, p_2, \ldots, p_n]$ is perhaps the best known and most appropriate for the intended applications. Explicitly, the Shannon-Wiener entropy H[p] is defined by

$$H[p] = H[p_1, p_2, \ldots, p_n] = -\sum_{k=1}^{n} p_k \log p_k$$

and the corresponding relative entropy in case of a general background or reference probability distribution $q = [q_1, q_2, \ldots, q_n]$,

$$H[p;q] = H[p_1, p_2, \ldots, p_n; q_1, q_2, \ldots, q_n] = -\sum_{k=1}^{n} p_k \log (p_k/q_k)$$

where the summation signs are replaced by integral signs in applications with continuous probabilities. The logarithms used in these definitions are assumed to have the appropriate base (usually 2) or else a multiplicative constant should be included. When the background or reference probability distribution is uniform, then the relative entropy reduces to the absolute entropy.

For practical applications, information measures need to be coordinate system independent and least sensitive to additive noise in the data. The Shannon-Wiener relative entropy has been shown to satisfy these conditions in practice [Blais and Boulianne, 1988]. Furthermore, the relative entropy measure is known to be unaffected by any orthogonal transformation (e.g., a rotation) of digital image data where the normalized grey level frequencies are interpreted as probability distribution frequencies [Andrews, 1970]. The latter is especially important in the context of digital image processing using Fourier and other orthogonal transforms which preserve the energy associated with the grey levels.

For a continuous random variable with a Gaussian probability distribution, the Shannon-Wiener entropy is proportional to the logarithm of the variance in one dimension, and the logarithm of the covariance matrix in higher dimensions [e.g., Blais, 1991a]. This is not a surprising result as a Gaussian probability distribution is fully specified by its first two moments and hence the Shannon-Wiener entropy can be expected to be expressible in terms of the second moment. Obviously, the situation is different with other probability distribution functions which can only be specified fully by their higher statistical moments.

It is important to realize that no interpretation nor any semantics are included in the preceding definitions and discussions. Mathematically, the analysis of a probability distribution does not require any interpretation of the inferences as these can be very different in different application contexts. On the other hand, the appropriateness and implications of using one information measure in a specific context may very well include semantics and valuations for reasoning-like processing as in expert systems.

The preceding concepts from information theory are very useful in estimation and inverse problems where the available observational and other information is often incomplete for the desired solution. Considering the available information for maximum exploitation without making any unnecessary assumptions about what is not known is precisely the maximum information or maximum entropy approach. Explicitly, the maximum entropy principle states:

When making inferences based on incomplete information, the estimates should be based on probability distributions corresponding to the maximum entropy permitted by the available information.

This principle was proposed independently by Kullback and Liebler [1951], Jaynes [1957] and Ingarden [1963]. It has been justified in terms of combinatorial arguments, axiomatic inference, objectivity, consistency and reliability of the estimation process [Jaynes, 1982 and 1983].

Applications of this maximum information principle are wide ranging in physical science and engineering. Some applications in model identification, digital image processing and spatial information systems are discussed in Blais [1991a and b]. The following discussions will concentrate on applications in spectrum estimation, adaptive filter design and inverse problems to illustrate the applicability of information theory and the principle of maximum entropy.

## 4. APPLICATIONS IN SPECTRUM ESTIMATION

Estimates of power spectral density functions are required for numerous applications in digital signal and image processing. Filter design often relies on the analysis of the spectral analyses of data sequences and arrays. The estimation of the spectrum of one-dimensional data sequences is relatively straightforward and the analysis of the estimates does not usually present any problems. The situation is however quite different in two and higher dimensions where the implications of difficulties in factorization and positive definiteness of autocovariance functions can imply serious difficulties.

Given a sample autocovariance sequence of finite length, the spectrum estimation problem involves the extension of this sequence for the Fourier transformation to estimate the spectrum of the process. Well known approaches to the spectrum estimation problem include the periodogram and correlogram methods, the parametric modeling techniques of autoregressive and moving average formulations, and the maximum entropy approach which is based on information theory.

When using Fourier based methods, the extension of the autocovariance function is implied by the periodicity of the Fourier transform. This situation is usually quite appropriate in noise dominated sequences although the spectral resolutions are affected by the well known leakage and aliasing effects that are unavoidable with Fourier transforms. With proper analysis of the available information and constraints for the application context, the periodogram and correlogram approaches to spectrum estimation are generally acceptable, but not necessarily optimal at least in terms of resolution.

With the parametric modeling approaches, the extension of the autocovariance function is implied by the autoregressive, moving average, autoregressive-moving-average or variation of these models. Some constraints may also be required to ensure that the extension of the autocovariance function is fully compatible with the observations of the physical process. It is important to note that the autoregressive modeling approach in one dimension fully agrees with the maximum information or entropy of the underlying stochastic process (see, e.g., Blais [1992b] for details).

In two and higher dimensions, these classical Fourier based and parametric methods often lead to complications and ambiguities. More specifically, extensions of the sample autocovariance functions need to be compatible with causality and other physical requirements of the observed process. The nonuniqueness of the extension solution implies that the estimated spectrum needs to be constrained to correspond to the application requirements.

Numerous researchers have investigated the use of maximum entropy approaches for spectrum estimation in two and higher dimensions. Among them are Burg [1975], Pendrell [1979], Wernecke and D'Addario [1977], Lim and Malik [1981], and Lang and McClellan [1982]. The approach of Lim and Malik [1981] is especially appealing with a recursive strategy using fast Fourier transforms and the dual of the sample autocovariance function. The latter has been studied and further discussions can be found in Blais [1992b] with a variation of the Lim and Malik [1981] approach having been proposed and experimented with in Blais and Zhou [1990 and 1992].

The maximum entropy approach has intrinsic features which are most interesting in the sense that the sample autocovariance function is extended in an optimal manner without using artificial constraints or models. The implemented conditions in this extension are simply the positive definiteness for realizability of the physical process and correlation matching for known lags. In other words, only the implications of the observed process are used in the estimation of the spectrum. Among the characteristics of the spectrum estimates are the resolution features, the consistency and reliability of the results.

## 5. APPLICATIONS IN ADAPTIVE FILTER DESIGN

In digital signal and array processing, filters are designed to restore the information by removing the noise or correcting for some degradation function. In several applications of signal and array processing, the underlying process is not stationary with the implication that the filters need to be adaptive to meet the expectations. Adaptiveness in filter design means that the filter parameters change whenever the conditions in the applications warrant it. In other words, the filters are self calibrating in their implementation.

The adaptability of a mean or median filter in digital image processing simply implies a variable mask or template over which the mean and median operations are carried out. For instance, under smooth texture conditions, a smaller mask may be sufficient while under rougher conditions, the mask may need to be larger for reliability and other similar requirements. Other filter applications may have directional dependence and hence the detection of optimal directions may be necessary for adaptability to different conditions.

The adaptability of a spectral filter such as an inverse filter would require a variable transfer function while an adaptable Wiener filter would mean a variable transfer function or spectral density function for the signal. In such applications, the average information content often plays an important role as optimal information extraction is the usual objective of the filtering. The question of deciding on an appropriate measure for the information content is very much dependent on the application context and the specific objectives of the operations.

The problem of optimal filter design is essentially one of model identification strategies and information theory is well known for its applicability in these areas [e.g., Blais, 1987 and 1991b]. The observational and related information can usually be analyzed in terms of information contents to infer a most appropriate model for the application. A number of researchers from Kullback and Liebler [1951] to Landau [1987] and others have studied the use of information theory for these applications. A number of distributional and related model identification results can be found in Blais [1987 and 1991b].

There are still several open questions in model identification concerning the consistency and asymptotic efficiency of the selected models, especially in multivariate applications and implementations with limited data samples and missing observations. Research on these and related questions is continuing, especially for digital image and array processing.

## 6. APPLICATIONS IN INVERSE PROBLEMS

Inverse problems are among the most common problems encountered in the physical sciences. With only limited observational and other information, inverse problems often present tremendous difficulties to the scientists who want reliable answers that are justifiable and appropriate.

There are different classes of inverse problems depending on the nature of the problems and the information involved. First, there are mathematical inverse problems such as Cauchy contour integration and the inversion of integral transforms in purely analytical terms. Second, there are the inverse problems exemplified by object reconstruction and tomography which involve geometrical analysis as well as estimation considerations. Third, there are the geophysical inverse problems which involve much physics and geology for analysis and interpretation of the results.

One implication of the preceding observations is that the study of inverse problems is clearly more than a simple extension and generalization of estimation theory. The perspective used in the following is that inverse problems are problems with incomplete information so that much of the experienced complications are actually due to the missing information and the implications thereof. One approach which has been successful in numerous applications is using information theory and related considerations. The advantages of this general approach will be discussed in the following with examples of applications.

In strictly mathematical terms, the problem can be formulated as follows:

$$u = KU$$

where U describes a true state vector and u is the observed state vector or the perceived signal or image after having been corrupted or modified by mechanisms of observation or measurement. The direct problem is primarily one of deductive prediction, i.e. given prior knowledge of U and the operator K, deduce or estimate u. The corresponding inverse problem, i.e. given the observed or measured u and a specific operator K, estimate the true vector U. In practice, it often occurs that K is also poorly defined or even understood.

The general solution to any inverse problem can be described in terms of Bayes' theorem which involves probabilistic measures of the available information. Assuming that the observed or measured vector u is a function of components $U_n$ of the true state vector U with the probability $(u|U_n)$ known, then Bayes' theorem implies

$$\mu(U_n|u,I) = \frac{\mu(u|U_n)\mu(U_n|I)}{\sum_m \mu(u|U_m)\mu(U_m|I)}$$

where I denotes the available prior information. In cases where prior information is known to be uniform, then

$$\mu(U_n|u,I) \propto \mu(u|U_n)$$

which implies a straightforward solution.

The preceding Bayes' theorem shows how to combine partial information in a mathematically rigorous manner. Then the principle of maximum information or entropy can be used to arrive at optimal frequencies taking into consideration additional modeling and other constraints required for the solution of the inverse problem. This methodology which is based on Bayes' theorem is only one of the possibilities to formulate the desired solution for a given inverse problem. Tarantola and Valette [1982] offer another strategy based on an extension of Bayes' theorem.

The implementation of the preceding approach to solving practical inverse problems is sometimes difficult as the available observations or measurements and prior information have uncertainties which cannot easily be quantified or measured. This is where information theory can help in providing better insight into the situation. Further discussions of these questions can be found in Blais [1992a].

## 7. CONCLUSIONS AND GENERAL RECOMMENDATIONS

With the changes taking place in data and geoinformation processing, the analytical tools and procedures need to be more sophisticated and adaptive in their implementations. Information theory provides insight and methodologies for analyzing problems characterized by incomplete observational and prior information.

Three areas of applications, i.e. spectrum estimation, adaptive filter design and inverse problems have been selected to illustrate the usefulness and applicability of information theory in geomatics. The discussions have been in terms of methodologies to solve such problems with references to other publications for specifics on the formulation and implementation of the solutions in practical environments. The emphasis on the solution strategies is justifiable in terms of the rapidly changing application contexts and the anticipated requirements of the near future activities in applied science and engineering.

## 8. ACKNOWLEDGMENTS

The author wishes to acknowledge the sponsorship of the Natural Sciences and Engineering Research Council of Canada in the form of an operating grant on the applications of information theory, and research funding for the development of analytical tools in geomatics from Energy, Mines and Resources Canada.

## REFERENCES

Andrews, H.C., 1970. Computer Techniques in Image Processing. Academic Press.

Blais, J.A.R., 1992a. Generalized Inverse Problems and Information Theory. Presentation at the Canadian Geophysical Union and American Geophysical Union Meeting in Montreal, Quebec.

Blais, J.A.R., 1992b. Spectrum Estimation Using Maximum Entropy. Manuscript under preparation, Department of Surveying Engineering, The University of Calgary, Calgary.

Blais, J.A.R., 1991a. On Some Practical Applications of Information Theory in Geomatics. CISM Journal ACSGC, Vol.45, No.2, pp.239-247.

Blais, J.A.R., 1991b. On Some Model Identification Strategies Using Information Theory. Manuscripta Geodaetica, Vol.16, No.5, pp.326-332.

Blais, J.A.R., 1987. Information Theory and Optimal Estimation. Manuscripta Geodaetica, Vol.12, No.4, pp.238-244.

Blais, J.A.R. and M. Boulianne, 1988. Comparative Analysis of Information Measures for Digital Image Processing. Archives of Congress of the ISPRS, Commission III, Working Group 4, in Kyoto, Japan, Vol.27, part B8, pp.34-44.

Blais, J.A.R. and W. Zhou, 1992. Spectral Analysis Applications in Digital Image Processing. Archives of Congress of the ISPRS, Commission III, in Washington, D.C., in press.

Blais, J.A.R. and W. Zhou, 1990. Spectral Analysis of Digital Images for Filtering and Restoration. Proceedings of the ISPRS Commission III Symposium in Wuhan, P.R. China, Volume 28, Part 3/2, pp.48-63.

Burg, J.P., 1975. Maximum Entropy Spectral Analysis. Ph.D. Dissertation, Stanford University, Stanford.

Ingarden, R.S., 1963. Information Theory and Variational Principles in Statistical Theories. Bull. Acad. Polon. Sci. Ser. Sci. Math. Astronom. Phys. Vol.11, pp.541-547.

Jaynes, E.T., 1983. Papers on Probability, Statistics and Statistical Physics. Reidel.

Jaynes, E.T., 1982. On the Rationale of Maximum Entropy Methods. Proceedings of the IEEE, Vol.70, No.9, pp.939-952.

Jaynes, E.T., 1957. Information Theory and Statistical Mechanics. Physics Reviews, Vol.106, pp.620-630; Vol.108, pp.171-182.

Kullback, S. and R.A. Liebler, 1951. On Information and Sufficiency. Annals of Mathematical Statistics, Vol.22, pp.79-86.

Landau, H.J., 1987. Maximum Entropy and the Moment Problem. Bulletin of the American Mathematical Society, Vol.16, No.1, pp.47-77.

Lang, S.W. and J.H. McClelland, 1982. Multi-Dimensional MEM Spectral Estimation. IEEE Trans. Acoust., Speech and Signal Processing, ASSP-30, pp.880-887.

Lim, J.S. and N.A. Malik, 1981. A New Algorithm for Two-Dimensional Maximum Entropy Power Spectrum Estimation. IEEE Trans. Acoust., Speech and Signal Processing, Vol.ASSP-29, No.3, pp.401-412.

Pendrell, J.V., 1979. The Maximum Entropy Principle in Two-Dimensional Analysis. Ph.D. Dissertation, York University, Toronto.

Tarantola, A and B. Valette, 1982. Inverse Problems = Quest for Information. Journal of Geophysics, Vol.50, pp. 159-170.

Wernecke, S.J. and L.R. D'Addario, 1977. Maximum Entropy Image Reconstruction. IEEE Trans. Computer, Vol.C-26, pp.351-358.