Symposium no. 65 Paper no. 2009 Presentation: oral

Entropy based triangle for designing sustainable soil management

OLESCHKO Klaudia (1)**, KORVIN Gabor** (2) **and FIGUEROA Benjamín** (3)

(1) Instituto de Geología, Universidad Nacional Autónoma de México (UNAM), Aportado Postal 70-296, Ciudad Universitaria, C.P. 04510, Coyoacan, D.F., México
(2) King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
(3) Colegio de Postgraduados, km. 35.5 Carretera México-Texcoc, Montecillos, Estado de México, C.P., 56230, México

Abstract

The concept "sustainability" as used in the present research means the degree of equilibrium among three principal conflictive aspects of Agriculture: the ecological, the economic and the social. The new quantitative approach based on a simple mathematical model has been designed, verified and calibrated, in order to evaluate the sustainability of the agroecosystem´ management in terms of entropy, using the triangle of Moebius as the point of departure. The designed triangle is called "Triangle of the Agro-ecosystem Sustainability" or TRISA. The edapho-ecological and economical dimensionalities of the system are taken as the triangle´s base, the social dimensionality corresponds to its superior apex. Under "dimensionality" a complex of measures is understood, each one characterizing a specific qualitative or quantitative attribute of the studding system, situated inside the metric space, selected by the researcher. The exact number of variables needed for sustainability estimation depends on the research objectives. A principal property of the TRISA is the boundaries´ permeability along the sides from each triangle apex to an other. This permeability is measurable in terms of system resistance to the information, that flows from the physical medium to the economical and social, or vice verse. The Shannon entropy of a probability distribution constitutes the center of sustainability triangle. The identification of the basic patterns and structures, operating on each level of organization, is the first step en TRISA characterization. All selected soil-ecology, economy, and social variables are transformed to dimensionless variables $p_{soil-ecol}$, p_{econ} and p_{soc} , normalized between $[0,1]$, and united in a ternary diagram. The total Shannon entropy (H_{total}) of the complete probability distribution $\{p_j^{(soil-ecol)}, p_k^{(econ)}, p_l^{(soc)}\}_{j,k,l}$ is estimated. The relative weights of information which the soil-ecological, economic and social variables contribute to H may be precisely calculated and used for the unbiased analysis of the management sustainability. The proposed method and algorithm are also useful for any other complex ecologic-economic-social systems, requiring quantitative evaluation. We present a few examples of the management sustainability analysis realized in Mexico, for the case of contrasting systems: traditional-, minimum- and zero tillage - applied to Faeozems, Andosols and Vertisols-during more then 10 years. Some theoretical alternatives for sustainable management of each one of the mentioned soils are derived from: (1) Direct physical and chemical measurements, realized *in situ* by different no-

invasive techniques (such as Ground Penetrating Radar, Time Domain Reflectometry, Ultrasonic Penetrometer, Permeameter Guelph etc.); and, (2). Theoretical models and computer simulations.

Keywords: entropy based triangle, sustainable soil management, agro-ecosystem, soil ecology variable, social variable

Introduction

Over recent years, three main types of sustainability: social, economic and environmental, were defined and measured by social scientists, economists and biophysics, in terms of numerous, broadly diversified variables (Daly and Cobb, 1989; Karlen *et al.*, 1997; Goodland, 2002). These variables, derived from natural processes as well as from socioeconomic dynamics, have been studied at widely different spatial-temporal scales, and their interpretation requires a high degree of technical expertise. The problem that arises concerns the impossibility to compare these data banks and to identify each specific trend in the multifaceted sustainability dynamics of the diverse systems. To solve this problem, we design the user-friendly, quantitative approach to measure sustainability in any kind of complex multilevel systems, which can be described by three meaningfully related variables at each level. We assume that there is a regular hierarchy of interdependencies between the variables determining the state of a system. For instance, the soil management sustainability means that the system is achieving some physical, chemical and biological qualities which in turn should be measured by further numerous indicators. Our program calculates the relative entropies of three selected parameters, using the original algorithm, and represents them in a triangular diagram. Therefore, the triangular diagram was accepted as basis for multiscale data representation. Ternary diagrams are broadly used in science to describe multicomponent systems: for instance, for particle size classification (texture triangle, Bullock *et al.*, 1985), sand-stone composition representation (Pettijohn triangle, 1957), or to visualize the development of mankind´s energy consumption (Klepper, 2001).

There are two new and distinctive features of our ternary diagram. First of all, the system variables (which can be scalars or vectors) are expressed in terms of their relative Shannon entropy with respect to the total entropy of the system, so that all variables become mutually commensurable. The second novelty in our approach is the fractal subdivision of the initial “mother-triangle”. The Sierpinski gasket model (popularized in Mandelbrot´s 1977 book) has been found an efficient tool for multiscale sustainability analysis. We suppose that this kind of Sierpinski –gasket analysis of point distribution provides a way for an objective fractal cluster analysis, which represents the co-existence groups of variables at different hierarchical levels in any ternary ABC subsystem of the whole mother-triangle, which we propose to call TRISA: Triangle of the Sustainability of Agro-ecosystem.

In the present paper, the designed algorithm is described and the first steps of the TRISA construction are illustrated. Further examples, on a series of soil management systems, compared for diverse areas of Mexico, will be given in forthcoming publications.

Algorithm description

In order to circumvent the problem of data incompatibility we propose.

(1) To transform the economy, social, ecology variables of interest, which are generally multicomponent vectors, to dimensionless scalar variables $p_{econ}, p_{soc}, p_{ecol}$ normalized between [0,1] and

(2) To assure that $p_{econ}\ 2\ p_{soc}\ 2\ p_{ecol}\ |\ 1$ always, because otherwise we cannot work with the $(p_{econ}, p_{soc}, p_{ecol})$ ternary diagram.

We present an example of how to proceed in case of soil management sustainability analysis. Any other *complex economic-social-ecologic system* could be treated along very similar lines.

Steps of the algorithm

The designed algorithm may be applied for comparison among countries, agricultural zones of the same country, or various producers of the same agricultural area of interest. Select some soil management systems, which should be compared, for instance traditional, conservation and zero tillage. In order to simplify discussion, the following algorithm description will be formulated in terms of countries comparison.

Step 1) Design a number $N_{econ} - 15\ 4\ 20$ of possible classes of economy, derived from some specific soil management (each variable used to characterize the agro-ecosystem economical sustainability may be chosen) where the value of economy variable can belong: $ECON_1, ECON_2, 3, ECON_{N_{econ}}$.

If the economy of the country is characterized by an m_{econ}-dimensional vector whose i-th component lies in the interval $[\min_{econ,i}, \max_{econ,i}]; (i\ |\ 1,2, 3, m_{econ})$ then each class should be a convex domain of the m_{econ}-dimensional Euclidean space such that $\overset{m_{econ}}{\underset{i|1}{8}} ECON_i\ |\ [\min_{econ,1}, \max_{econ,1}]\ \Delta\ 3\ \Delta [\min_{econ,m_{econ}}, \max_{econ,m_{econ}}] \leq R^{m_{econ}}$ and $i \prod j \heartsuit\ ECON_i\ 1\ ECON_j\ |\ \therefore$.

The classes $ECON_1, ECON_2, 3, ECON_{N_{econ}}$ can be arranged in increasing order of merit, so that according to some economic criterion $ECON_2$ is "better" than $ECON_1$, etc. In a similar way the selected social variables should be divided to a number $N_{soc} - 15\ 4\ 20$ possible classes $SOC_1, SOC_2, 3, SOC_{N_{soc}}$, arranged in increasing order of merit; and the possible ecologic parameters similarly classified to a number $N_{ecol} - 15\ 4\ 20$ groups in $ECOL_1, ECOL_2, 3, ECOL_{N_{ecol}}$. In case of scalar variables, if we have *N* countries, and E_{max} and E_{min} are the maximum and minimum values of the parameter *E*, and N_{econ} is the number of classes, then the number *j* (*j*=1,2,...,N_{econ}) of a class the *E* value of the *i*-th country belongs to (*i*=1,2,...*N*) can simply be defined as:

$$j = \mathrm{Round}((E_i-E_{min})*(N_{econ}-1)/(E_{max}-E_{min}))+1,$$

where E_i is the value of the selected parameter of the *i*-th country. Here *E* is any of the scalar variables ECON, SOC or ECOL.

Step 2) Use the known statistics of *N* (*N* about 100 or more) countries (agricultural arias, producers etc.) for some year of interest, to prepare empirical histograms for the distribution of the variables ECON, SOC, ECOL among the classes defined in Step 1. Make the selection of variables representative so that different economic, social systems, geographic conditions, etc. be included.

Step 3) Find a meaningful and objective *well-being function W* to characterize the stage of development of a country (for instance Gross Agricultural Product/area of cultivated land, etc.). Let the well-being function of the *i*-th country be $W_i (i = 1, 2, \ldots, N)$. If country *i* belongs to economy class $ECON_j$, social class SOC_k, ecology class $ECOL_l$, then define.

$$\begin{aligned} econ_i &= \frac{j}{N_{econ}}; 0 \le econ_i \le 1 \\ soc_i &= \frac{k}{N_{soc}}; 0 \le soc_i \le 1 \\ ecol_i &= \frac{l}{N_{ecol}}; 0 \le ecol_i \le 1 \end{aligned} \tag{1a}$$

Step 4) Fit *W* linearly as

$$W_i \approx \varsigma \cdot econ_i + \sigma \cdot soc_i + \tau \cdot ecol_i \tag{1b}$$

where the coefficients ς, σ, τ are determined as to be optimal in the least mean squares sense:

$$\sum_{i=1}^{N} (W_i - \varsigma \cdot econ_i - \sigma \cdot soc_i - \tau \cdot ecol_i)^2 = min \tag{1c}$$

Step 5) The histograms constructed in Step 2 define three probability distributions. For the case of economy, for example, as there are *N* countries and N_{econ} economic classes, if there are $N_1^{(econ)}, N_2^{(econ)}, \ldots$ countries in classes $ECON_1, ECON_2, \ldots$, such that $\sum_{j=1}^{N_{econ}} N_j^{(econ)} = 1$, denoting

$$p_j^{(econ)} = \frac{N_j^{(econ)}}{N} \tag{2a}$$

we get a complete probability distribution $\left\{ p_j^{(econ)}, j = 1, \ldots, N_{econ}, \sum_{j=1}^{N_{econ}} p_j^{(econ)} = 1 \right\}$.

We similarly define the complete probability distributions

$$\left\{ p_k^{(soc)} = \frac{N_k^{(soc)}}{N}, k = 1, \ldots, N_{soc}; \sum_{k=1}^{N_{soc}} p_k^{(soc)} = 1 \right\} \tag{2b}$$

and

$$\left\{ p_k^{(ecol)} = \frac{N_k^{(ecol)}}{N}, k = 1, \ldots, N_{ecol}; \sum_{k=1}^{N_{eecol}} p_k^{(ecol)} = 1 \right\} \tag{2c}$$

The set of probabilities $p_j^{(econ)} \cdot p_k^{(soc)} \cdot p_l^{(ecol)}$, corresponding to the event that a given country falls to the *j*-th economic, *k*-th social and *l*-th ecologic class, also form a complete distribution $\sum_{j=1}^{N_{econ}} \sum_{k=1}^{N_{soc}} \sum_{l=1}^{N_{ecol}} p_j^{econ} p_k^{soc} p_l^{ecol} = 1$.

Step 6) Recall the meaning of the Shannon entropy of a probability distribution Rényi, 1987): Suppose $f(p)$ is a measure of the information that we obtain by observing that an event *A* of probability *p* occurs. If *B* is another event, independent of *A* and of probability *B*, the probability of a simultaneous occurrence of *A* and *B* is *pq* and the information obtained is *f(p+q)*, that is for independent events

$$f(p \cdot q) = f(p) + f(q). \tag{3}$$

By Cauchy´s functional equation (Aczel, 1966) the only continuous, non-trivial solution of Eq. (3) is

$$f(x) = c \cdot \log x \tag{4}$$

where *c is* an arbitrary constant, and "log" means the natural logarithm. The expected value of the information obtained by observing a random outcome from a set of independent events $\left\{ A_i : P(A_i) = p_i; i = 1, \ldots, N; \sum_{i=1}^{1} p_i = 1 \right\}$ is

$$H(p_1, p_2, \ldots, p_n) = -c \cdot \sum_{i=1}^{N} p_i \log p_i \tag{4}$$

where *c* is a suitably selected positive constant. With *c = 1* we get the Shannon entropy $-\sum_{i=1}^{N} p_i \log p_i$.

Step 7) The total Shannon entropy of the complete probability distribution $\left\{ p_j^{(econ)} \cdot p_k^{(soc)} \cdot p_l^{(ecol)} \right\}_{j,k,l}$ is

$$H_{total} = -\sum_{j=1}^{N_{econ}} \sum_{k=1}^{N_{soc}} \sum_{l=1}^{N_{ecol}} p_j^{(econ)} \cdot p_k^{(soc)} \cdot p_l^{(ecol)} \cdot \log\left(p_j^{(econ)} \cdot p_k^{(soc)} \cdot p_l^{(ecol)} \right)$$

$$= -\sum_{j=1}^{N_{econ}} \sum_{k=1}^{N_{soc}} \sum_{l=1}^{N_{ecol}} p_j^{(econ)} \cdot p_k^{(soc)} \cdot p_l^{(ecol)} \cdot \left(\log p_j^{(econ)} + \log p_k^{(soc)} + p_l^{(ecol)} \right) \tag{5}$$

If we observe an "event" (a country) $A = \left\{ ECON_j, SOC_k, ECOL_l \right\}$, this contributes a partial entropy

$$H_{jkl}\left(p_j^{(econ)} \cdot p_k^{(soc)} \cdot p_l^{(ecol)} \right) = -p_j^{(econ)} \cdot p_k^{(soc)} \cdot p_l^{(ecol)} \cdot \log\left(p_j^{(econ)} \cdot p_k^{(soc)} \cdot p_l^{(ecol)} \right) \tag{4}$$

to the total entropy H_{total}. The relative weights of information contributes by the economic, social, and ecologic variables to H_{jkl} are as follows:

$$h_{j,relative}^{(econ)} \stackrel{def.}{=} \frac{-p_j^{(econ)} \cdot p_k^{(soc)} \cdot p_l^{(ecol)} \cdot \log p_j^{(econ)}}{-p_j^{(econ)} \cdot p_k^{(soc)} \cdot p_l^{(ecol)} \cdot \left(\log p_j^{(econ)} + \log p_k^{(soc)} + \log p_l^{(ecol)}\right)} = \frac{\log p_j^{(econ)}}{\log p_j^{(econ)} + \log p_k^{(soc)} + \log p_l^{(ecol)}} \quad (5a)$$

and, similarly,

$$h_{k,relative}^{(soc)} = \frac{\log p_k^{(soc)}}{\log p_j^{(econ)} + \log p_k^{(soc)} + \log p_l^{(ecol)}} \quad (5b)$$

$$h_{l,relative}^{(ecol)} = \frac{\log p_l^{(ecol)}}{\log p_j^{(econ)} + \log p_k^{(soc)} + \log p_l^{(ecol)}} \quad (5c)$$

Obviously, all these three relative weights of information are a) dimensionless and between 0 and 1; and b) their sum is 1

$$h_{j,rel}^{(econ)} + h_{k,rel}^{(soc)} + h_{l,rel}^{(ecol)} = 1 \quad (6)$$

Consequently, these variables can be used as coordinates along the sides of the ECON-SOC-ECOL equilateral triangle (instead of the original, incommensurable variables ECON, SOC, and ECOL) to plot the contoured elevation map of any function *F(ECON,SOC,ECOL)* of the original variables inside the triangle $\left(h_{j,rel}^{(econ)}, h_{k,rel}^{(soc)}, h_{l,rel}^{(ecol)}\right)$

Step 8) Final step: Now, we shall construct two ternary plots for the relative entropy dynamics, for the well-being function or indeed, for any function F(ECON, SOC, ECOL). As there are N_{econ} possible economy classes, N_{soc} possible social classes, N_{ecol} possible ecology classes, there are altogether $N_{econ} \cdot N_{soc} \cdot N_{ecol}$ functional values to be plotted inside both ternary diagrams. The combination of variables $\left(ECON \subset ECON_j; SOC \subset SOC_k; ECOL \subset ECOL_l\right)$ will correspond to the ternary coordinates $\left(h_{j,rel}^{(econ)}, h_{k,rel}^{(soc)}, h_{l,rel}^{(ecol)}\right)$ along the sides of the triangle, where $h_{j,rel}^{(econ)}, h_{k,rel}^{(soc)}, h_{l,rel}^{(ecol)}$ can be computed using Eqs. 2a-2c, 5a-5b.

In the ternary diagram for entropy we plot the relative entropy of the event $A = \left(ECON_j, SOC_k, ECOL_l\right)$ with respect to the total entropy:

$$H_{rel}\left(ECON_j, SOC_k, ECOL_l\right) = \frac{-p_j^{(econ)} \cdot p_k^{(soc)} \cdot p_l^{(ecol)}}{H_{total}} \quad (7)$$

where H_{total} is given by Eq. 5.

In the ternary diagram for well-being function, if for the *i*-th country the parameters are $\left(ECON_j, SOC_k, ECOL_l\right)$ we plot, at the point $\left(h_{j,rel}^{(econ)}, h_{k,rel}^{(soc)}, h_{l,rel}^{(ecol)}\right)$ inside the triangle, instead of the original W_i, the smoothed value

$$W_{smoothed} - \varsigma \left(\frac{j}{N_{econ}} \right. 2\ \sigma \left(\frac{k}{N_{soc}} \right. 2\ \Upsilon \left(\frac{l}{N_{ecol}} \right. \qquad \text{(see Eqs. 1a-1c)}$$

instead of the original W_i, to eliminate random fluctuations.

Example for TRISA construction

The example of five iterations of TRISA construction is presented in Figure 1 Only a perfect regular hierarchy is shown, when every variable depends on exactly three more refined variables. However, in general case the variables may depend on different numbers of more refined parameters, and the construction may achieve as many refinements (iterations) as necessary. At the 0^{th} level of construction, the state of a system is determined by three variables only, namely by ECOL0, ECON0 and SOC0. After the first iteration, the mother-triangle is divided in 4 equal parts, but only 3 of these will contain the further system variables which can only be observed at higher-and higher resolution. The triangle at bottom right contains social variables, at bottom left economic variables, and the triangle at the top contains ecological variables. Therefore, at the 1^{th} level of construction ECOL0, for instance, depends on three ecological variables, chosen by researcher: ECOL11, ECOL12, ECOL13, and their mutual information content is displayed by the first entropy-ecological triangle (on top of Figure 1). The researcher define how many levels of hierarchy will be studied. When the construction has been finished (Figure 1), the map of experimental points clusters is clearly observed. Another original algorithm gives information about each single point of the diagram from any level of TRISA. Alternatively, the functions of the variables in any subtriangle can be presented in form of contour elevation map, or can be calibrated to color or gray scale.

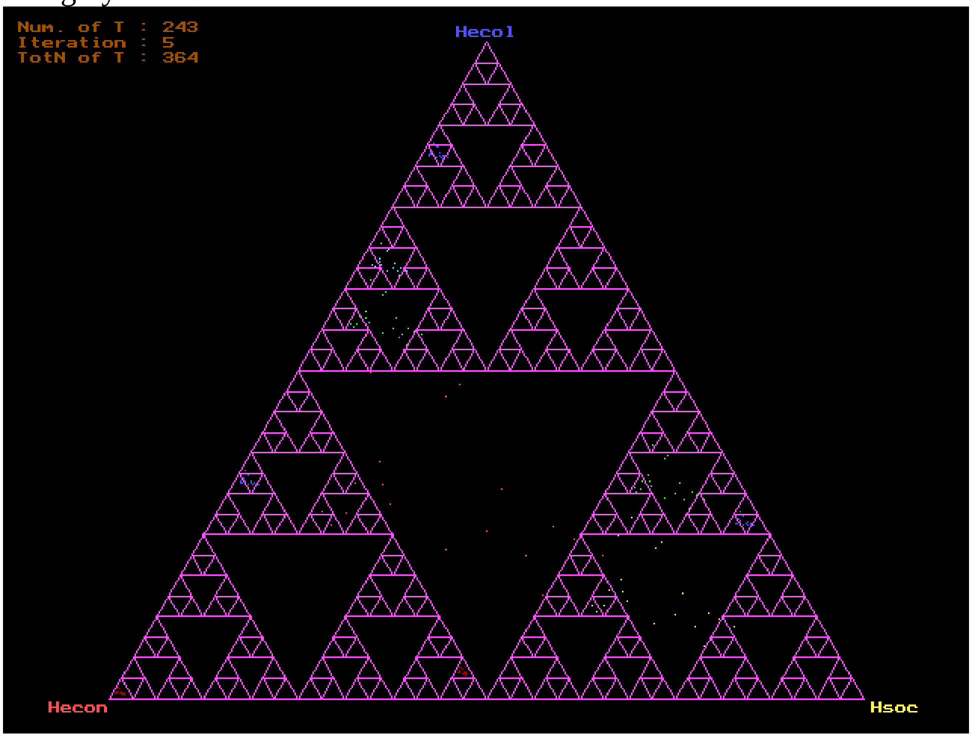

Figure 1 TRISA constructed with original data, using 5 iteration of algorithm.

Conclusion

We propose a new algorithm for multilevel ternary analysis of a perfect regular system hierarchy, when every variable depends on exactly three more refined variables. This is a schematic topological way to represent the interdependence of selected indicators of sustainability. The kind of Sierpinski-gasket analysis of point distributions in a multi-level ternary diagram, called TRISA, provides a way for an objective fractal cluster analysis of soil management sustainability. TRISA seems to be useful for all researches which require collaboration among diverse disciplines of science.

References

Aczel, J. 1966. Lectures on Functional Equations and Their Applications. Academic Press, New York and London.

Bullock, P., N. Fedoroff, A. Jongerius, G. Stoops, T. Tursina and U. Babel. 1985. Handbook for soils thin section description. Waine Research Publications. 152 p.

Daly, H.E. and J.B. Cobb. 1989. For the Common Good. Beacon Press, Boston. 443-507.

Goodland, R. 2002. Sustainability: Human, Social, Economic and Environmental. Encyclopedia of Global Environmental Change. John Wiley and Sons, Ltd.

Karlen, D.L., M.J. Mausbach, J.W. Doran, R.G. Cline, R.F. Harris and G.E. Schuman. 1997. Soil Quality: A Concept, Definition, and Framework for evaluation. Soil Sci. Soc. Am. J. 61:4-10.

Mandelbrot, B.B. 1977. Fractals: Form, Chance, and Dimension. Freeman, San Francisco.

Rényi, A. 1987. A Diary on Information Theory. J. Wiley and Sons, Chicester.