Meaningful and Meaningless Statements in Landscape Ecology and Environmental Sustainability

by

Fred S. Roberts

DIMACS Center

Rutgers University

Piscataway, NJ 08854 USA

froberts@dimacs.rutgers.edu

Acknowledgements: The author gratefully acknowledges the support of the National Science Foundation under grant number DMS-0829652 to Rutgers University. A number of ideas and some of the examples and language in this paper are borrowed from my papers Roberts (1994,2012), which explore meaningful and meaningless statements in operations research and in epidemiology and public health, respectively. The author gratefully and thankfully acknowledges the many stimulating and fruitful scientific interchanges with Boris Mirkin over a period of many years, and wishes him many years of continued good health and success.

1 Introduction

The growing population and increasing pressures for development lead to challenges to life on our planet. Increasingly, we are seeing how human activities affect the natural environment, including systems that sustain life: climate, healthy air and water, arable land to grow food, etc. There is growing interest (and urgency) in understanding how changes in human activities might lead to long-term sustainability of critical environmental systems. Of particular interest are large ecological systems that affect climate, air and water, etc. *Landscape Ecology* is concerned with such systems. Understanding the challenges facing our planet requires us to summarize data, understand claims, and investigate hypotheses. To be useful, these summaries, claims, and hypotheses are often stated using metrics of various kinds, using a variety of scales of measurement. The modern theory of measurement shows us that we have to be careful using scales of measurement and that sometimes statements using such scales can be meaningless – in a very precise sense. We will summarize the theory of meaningful and meaningless statements in measurement and apply it to statements in landscape ecology and environmental sustainability.

The modern theory of measurement was developed in part to deal with measurement in the social and biological sciences, where scales are not as readily defined as in the physical sciences. Extensive work has been done to understand scales measuring utility, noise, intelligence, etc. The theory of measurement was developed as an interdisciplinary subject, aiming at putting the foundations of measurement on a firm mathematical foundation. The theory traces its roots to work of Helmholtz (1887/1930), and was widely formalized in the 20th century in such books as Krantz, Luce, Suppes, and Tverksy (1971), Luce, Krantz, Suppes, and Tversky (1990), Pfanzagl (1968), Roberts (1979/2009), and Suppes, Krantz, Luce, and Tversky (1989). Measurement theory is now beginning to be applied in a wide variety of new areas. Traditional concepts of measurement theory are not well known in the landscape ecology world or in new investigations in environmental sustainability. They are finding new applications there and, in turn, problems of landscape ecology and environmental sustainability are providing new challenges for measurement theory.

We will seek to answer questions such as the following:

- Is it meaningful to say that the biodiversity of an ecosystem has increased by 10%?
- Is the average health of forests in South Africa higher than the average health of forests in Kenya?
- For measuring the health of grasslands using vegetation indices such as leaf area index or normalized difference vegetation index, which optical instrument is best?

All of these questions have something to do with measurement. In the next section, we provide a brief introduction to the theory of measurement. Then, in Section 3, we formalize the concept of meaningful statement. The rest of the paper describes a variety of meaningful and meaningless statements, starting with measures of biodiversity, scales of average forest health, and vegetation index.

2 Scales of Measurement

Measurement has something to do with numbers. In the theory of measurement, we think of starting with a set A of objects that we want to measure. We shall think of a *scale of measurement* as a function f that assigns a real number $f(a)$ to each element a of A. More generally, we can think of $f(a)$ as belonging to another set B. The "representational theory of measurement" gives conditions under which a function is an *acceptable scale* of measurement. For an exposition of this theory, see for example Krantz, Luce, Suppes, and Tversky (1971) or Roberts (1979/2009). Following ideas of Stevens (1946, 1951, 1959), we speak of an *admissible transformation* as a function that sends one acceptable scale into another, for example Centigrade into Fahrenheit and kilograms into pounds. In most cases, we can think of an admissible transformation as defined on the range of the scale of measurement. Suppose that f is an acceptable scale on A, taking values in B. Then a function ϕ that takes $f(a)$ into $(\phi \circ f)(a)$ is called an admissible transformation if $(\phi \circ f)(a)$ is again an acceptable scale. For example, $\phi(x) = (9/5)x + 32$ is the tranformation that takes Centigrade into Fahrenheit and $\phi(x) = 2.2x$ is the transformation that takes kilograms into pounds. Stevens classified scales into types according to the associated class of admissible transformations. For instance, the class of admissible transformations of the form $\phi(x) = \alpha x$, $\alpha > 0$, defines the class of scales known as *ratio scales.* Thus, a scale f is a ratio scale if and only if every transformation $\phi(x) = \alpha x$, $\alpha > 0$, is admissible and every admissible transformation is of the form $\phi(x) = \alpha x$, $\alpha > 0$. Such transformations change the unit. Mass is an example of a ratio scale, where admissible transformations take kilograms into pounds, ounces into milligrams, grams into kilograms, etc. Time intervals are another example: we can change from years to days, from days to minutes, etc. Length is another example, with changes from meters to yards, inches to kilometers, meters to millimeters, etc. Volume is another example and so is temperature on the Kelvin scale, where there is an "absolute zero.

A second important type of scale is an *interval scale*, where the class of admissible transformations is the class of transformations of the form $\phi(x) = \alpha x + \beta$, $\alpha > 0$. Here, we can change not only the unit but also the zero point. Temperature as in Centigrade to Fahrenheit is an example of an interval scale. So is time on the calendar, where we set a zero point and can change it. For example, this is the year 2013, starting from a given year as 0.

We say a scale is an *ordinal scale* if the admissible transformations are the (strictly) monotone increasing transformations. Grades of leather, wool, etc. define ordinal scale. The Mohs scale of hardness is another ordinal scale. On this scale, every mineral gets a number between 1 and 10, but the only significance of these numbers is that a mineral with a higher number "scratches" a mineral with a lower number, and so we can use any 10 numbers rather than 1, 2, ..., 10 as long as we keep the principle that a mineral assigned a higher number "scratches" one assigned a lower number. Some people feel that "preference" judgments, which lead to numbers called "utilities" in economics, only define an ordinal scale, while some think utilities define an interval scale under certain circumstances. Subjective judgments of quality of vegetation probably also only define an ordinal scale.

We say that we have an *absolute scale* if the only admissible transformation is the identity. Counting defines an absolute scale. For definitions of some other scale types, see Roberts (1979/2009).

3 Meaningful Statements

In measurement theory, we speak of a statement as being *meaningful* if its truth or falsity is not an artifact of the particular scale values used. The following definition is due to Suppes (1959) and Suppes and Zinnes (1963):

Definition: *A statement involving numerical scales is meaningful if its truth or falsity is unchanged after any (or all) of the scales is transformed (independently?) by an admissible transformation.*

A slightly more informal definition is the following:

Alternate Definition: *A statement involving numerical scales is meaningful if its truth or falsity is unchanged after any (or all) of the scales is (independently?) replaced by another acceptable scale.*

In some practical examples, for instance those involving preference judgments under the "semiorder" model, it is possible to have two scales where one cannot go from one to the other by an admissible transformation, so one has to use this alternate definition. (See Roberts, 1979/2000, Roberts, 1994.) There is a long literature of more sophisticated approaches to meaningfulness to avoid situations where either of the above definitions may run into trouble, but we will avoid those complications here. Our emphasis is on the notion of "invariance" of truth value. Our motivation is that scales used in practice might be somewhat arbitrary, involving choices about zero points or units or the like. We would not want conclusions or decisions to be different if the arbitrary choices made are changed in some "admissible" way.

To start, let us consider the following statement:

Statement S: "The duration of the most recent drought in a given ecological reserve was three times the duration of the previous drought."

Is this meaningful? We have a ratio scale (time intervals) and we consider the statement:.

$$f(a) = 3f(b). \tag{1}$$

This is meaningful if f is a ratio scale. For, an admissible transformation is $\phi(x) = \alpha x, \alpha > 0$. We want Equation (1) to hold iff

$$(\phi \circ f)(a) = 3(\phi \circ f)(b). \tag{2}$$

But Equation (2) becomes

$$\alpha f(a) = 3\alpha f(b) \tag{3}$$

and (1) iff (3) since $\alpha > 0$. Thus, the statment S is meaningful.

Consider next the statement:

Statement T "The high temperature in a given ecological reserve in 2012 was 2 per cent higher than it was in 1912."

Is this meaningful? This is the statement

$$f(a) = 1.02f(b).$$

This is meaningless. It could be true with Fahrenheit and false with Centigrade, or vice versa. In general, for ratio scales, it is meaningful to compare ratios:

$$f(a)/f(b) > f(c)/f(d).$$

For interval scales, it is meaningful to compare intervals:

$$f(a) - f(b) > f(c) - f(d).$$

For ordinal scales, it is meaningful to compare size:

$$f(a) > f(b).$$

Sometimes in ecology, we try to weigh samples. We might have two equal size baskets, one containing feathers and one containing (elephant) tusks. Consider the claim:

Statement W "The total weight of my basket of feathers is 1000 times that of my basket of tusks."

Is this statement meaningful? Yes, since it involves ratio scales and is presumably false no matter what unit is used to measure weight. The point is that meaningfulness is different from truth. It has to do with what kinds of assertions it makes sense to make, which assertions are not accidents of the particular choice of scale (units, zero points) in use.

4 Biodiversity

Next we ask if it is meaningful to say that the biodiversity of an ecosystem has increased by 10%. Evidence about the health of ecosystems is often obtained by measuring the biodiversity. Loss of biodiversity is considered an indicator of declining health of an ecosystem and there is great concern that climate change and other environmental stressors - natural and man-made - are leading to such a loss. One way of measuring progress in controlling the unwanted environmental effects of human activities – effects of human systems on natural systems – is to determine the extent to which the loss of biodiversity has been controlled. An index of biodiversity allows us to set specific goals and measure progress toward them. The 1992 Convention on Biological Diversity (CBD) (http://www.biodiv.org) set the goal that, by 2010, we should achieve a significant reduction of the current state of biodiversity loss at the global, regional, and national level (UNEP 2002). How can we tell if we have achieved this goal? We need to be able to measure biodiversity.

There have been hundreds of papers attempting to define biodiversity precisely. Traditional approaches consider two basic determinants of biodiversity: *Richness* is the number of species and *evenness* is the extent to which species are equally distributed (Magurran, 1991). These concepts assume that all species are equal, that all individuals are equal (we disregard differences in size, health, etc.), and that spatial distribution is irrelevant. These may not be appropriate assumptions. We shall concentrate here on the notion of evenness, which is based on ideas going back in the economic literature to the work of Gini (1909,1912) on measures of even income distribution and of Dalton (1920) on measures of inequality.

Some measures of biodiversity or evenness go back to work in communication theory, in particular the work of Shannon (1948) on entropy in information theory.

Let S be the number of species in an ecosystem and x_i be the number of individuals of species i found (the *abundance* of species i). In some cases, x_i is not a number, but some measure of biomass, e.g., grams per square meter. The vector $\mathbf{x} = (x_1, x_2, \ldots, x_S)$ is called the *abundance vector* and we seek a measure of evenness $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_S)$. We shall take $f(\mathbf{x})$ to be low if very even, high if very uneven. Finally, let a_i be the proportion of the population represented by species i, i.e., $a_i = x_i / \sum_j x_j$. In the literature, there are many proposed measures of evenness. We give a few examples. The *Simpson index* (Simpson, 1949) is given by $\lambda = \sum_i a_i{}^2$. It measures the probability that any two individuals drawn at random from an infinite population will belong to the same species. The *Shannon-Wiener Diversity Index* is given by $-\sum_i a_i ln(a_i)$. In information theory, the negative of this index is called the Shannon entropy. The Shannon entropy if maximized if each x_i is the same, so the Shannon-Wiener Diversity Index is minimized in this case.

Let us consider the statement that the biodiversity of an ecosystem has increased by 10% as the following:

Statement E "The evenness of an ecosystem has increased by 10%."

If x_i is the number of individuals of species i, then we have an absolute scale and the only admissible transformation of scale is the identiy, so a_i does not change and neither does either of the indices of evenness we are looking at. So, the statement is meaningful. However, what if x_i is the biomass of species i, for example kilograms of i per square meter? Both mass and length are ratio scales, so we can change for example from kilograms per square meter to grams per square centimeter, and so on. What happens if we multiply mass by a constant α and length by a constant β? Let y_i be the new abundance value and b_i be the new abundance proportion for species i. We have

$$y_i = (\alpha/\beta^2)x_i, \tag{4}$$

so

$$b_i = y_i / \sum_j y_j = (\alpha/\beta^2)x_i / \sum_j (\alpha/\beta^2)x_j = a_i. \tag{5}$$

It follows that neither the Simpson Index nor the Shannon Index changes after we change units, and so the Statement E is meaningful.

5 Averaging Judgments of Forest Health

Suppose we study two groups of forests, one in South Africa and one in Kenya. Let $f(a)$ be the health of forest a as judged by an "expert" on a subjective forest health scale using values 1 to 5 or 1 to 6, as is sometimes done. Suppose that data suggests that the average health of the forests in South Africa is higher than that of the forests in Kenya. Is this meaningful? Let $a_1, a_2, \ldots a_n$ be forests in the South African group and $b_1, b_2, \ldots, b_m$ be forests in the Kenyan group. Note that m could be different from n. Then we are (probably) asserting that

$$\frac{1}{n}\sum_{i=1}^{n} f(a_i) > \frac{1}{m}\sum_{i=1}^{m} f(b_i). \tag{6}$$

We are comparing arithmetic means. The statement (6) is meaningful if and only if under admissible transformation ϕ, (6) holds if and only if

$$\frac{1}{n}\sum_{i=1}^{n}(\phi \circ f)(a_i) > \frac{1}{m}\sum_{i=1}^{m}(\phi \circ f)(b_i) \tag{7}$$

holds. If forest health defines a ratio scale, then (7) is the same as

$$\frac{1}{n}\sum_{i=1}^{n} \alpha f(a_i) > \frac{1}{m}\sum_{i=1}^{m} \alpha f(b_i), \tag{8}$$

for some positive α. Certainly (6) holds if and only if (8) does, so (6) is meaningful. This kind of comparison would work if we were simply comparing biomass of forests.

Note that (6) is still meaningful if f is an interval scale. For instance, we could be comparing utility or worth of a forest (e.g., in terms of "ecosystem services") $f(a)$. Some economists think that in some cases, utility defines an interval scale. It is meaningful to assert that the average health of the first group is higher than the average health of the second group. To see why, note that (6) is equivalent to

$$\frac{1}{n}\sum_{i=1}^{n}[\alpha f(a_i) + \beta] > \frac{1}{m}\sum_{i=1}^{m}[\alpha f(b_i) + \beta],$$

where $\alpha > 0$.

However, (6) is easily seen to be meaningless if f is just an ordinal scale. To show that comparison of arithmetic means can be meaningless for ordinal scales, note that we are asking experts for a subjective judgment of forest health. Suppose that $f(a)$ is measured on a 5-point scale: 5=very healthy, 4=healthy, 3=neutral, 2=unhealthy, 1=very unhealthy. In such a scale, the numbers may not mean anything; only their order matters. Suppose that group 1 has three members with scores of 5, 3, and 1, for an average of 3, while group 2 has three members with scores of 4, 4, and 2 for an average of 3.33. Then the average score in group 2 is higher than the average score in group 1. On the other hand, suppose we consider the admissible transformation ϕ defined by $\phi(5) = 100$, $\phi(4) = 75$, $\phi(3) = 65$, $\phi(2) = 40$, $\phi(1) = 30$. Then after transformation, members of group 1 have scores of 100, 65, 30, with an average of 65, while those in group 2 have scores of 75, 75, 40, with an average of 63.33. Now, group 1 has a higher average score. Which group had a higher average score? The answer clearly depends on which version of the scale is used. Of course, one can argue against this kind of example. As Suppes (1979) remarks in the case of a similar example having to do with grading apples in four ordered categories, "surely there is something quite unnatural about this transformation" ϕ. He suggests that "there is a strong natural tendency to treat the ordered categories as being equally spaced." However, if we require this, then the scale is not an ordinal scale according to our definition. Not every strictly monotone increasing transformation is admissible. Moreover, there is no reason, given the nature of the categories, to feel that this is demanded in our example. In any case, the argument is not with the precept that we have stated, but with the question of whether the five point scale we have given is indeed an ordinal scale as we have defined it. To complete this example, let us simply remark that comparison of medians rather than arithmetic means is meaningful with ordinal scales: The statement that one group has a higher median than another group is preserved under admissible transformation.

Let us return to forest health, but now suppose that each of n observers is asked to rate each of a collection of forests as to their relative health. Alternatively, suppose we rate forests on different criteria or against different benchmarks. (A similar analysis applies with performance ratings, importance ratings, etc.) Let $f_i(a)$ be the rating of forest a by expert i (or under criterion i). Is it meaningful to assert that the average rating of forest a is higher than the average rating of forest b? A similar question arises in expert-judged ratings of health of individual species, quality of water in a stream, severity of pollution, etc.

We are now considering the statement

$$\frac{1}{n}\sum_{i=1}^{n} f_i(a) > \frac{1}{n}\sum_{i=1}^{n} f_i(b). \tag{9}$$

Note in contrast to statement (6) that we have the same number of terms in each sum and that the subscript is now on the scale value f rather than on the alternative a or b. If each f_i is a ratio scale, we then ask whether or not (9) is equivalent to

$$\frac{1}{n}\sum_{i=1}^{n} \alpha f_i(a) > \frac{1}{n}\sum_{i=1}^{n} \alpha f_i(b),$$

$\alpha > 0$. This is clearly the case.

However, we have perhaps gone too quickly. What if f_1, f_2, ..., f_n have independent units? In this case, we want to allow independent admissible transformations of the f_i. Thus, we must consider

$$\frac{1}{n}\sum_{i=1}^{n} \alpha_i f_i(a) > \frac{1}{n}\sum_{i=1}^{n} \alpha_i f_i(b), \tag{10}$$

all $\alpha_i > 0$. It is easy to find $\alpha_i's$ for which (9) holds but (10) fails. Thus, (9) is meaningless. Does it make sense to consider different α_i? It certainly does in some contexts. Consider the case where the alternatives are animals in an ecosystem and one expert measures their health in terms of their weight gain while a second measures it in terms of their height gain.

The conclusion is that we need to be careful when comparing arithmetic mean ratings, even when we are using ratio scales. Norman Dalkey [personal communication] was the first person to point out to the author that, in many cases, it is safer to use geometric means, a conclusion which by now is "folklore." For consider the comparison

$$\sqrt[n]{\prod_{i=1}^{n} f_i(a)} \quad > \quad \sqrt[n]{\prod_{i=1}^{n} f_i(b)}. \tag{11}$$

If all $\alpha_i > 0$, then (11) holds if and only if

$$\sqrt[n]{\prod_{i=1}^{n} \alpha_i f_i(a)} \quad > \quad \sqrt[n]{\prod_{i=1}^{n} \alpha_i f_i(b)}.$$

Thus, if each f_i is a ratio scale, then even if experts change the units of their rating scales independently, the comparison of geometric means is meaningful even though the comparison of arithmetic means is not. An example of an application of this observation is the use of the geometric mean by Roberts (1972, 1973). The problem arose in a study of air pollution and energy use in commuter transportation. A preliminary step in the model building involved the choice of the most important variables to consider in the model. Each member of a panel of experts estimated the relative importance of variables using a procedure called magnitude estimation. (Here, the most important variable is given a score of 100, a variable judged half as important is given a score of 50, and so on.) There is a strong body of opinion that magnitude estimation leads to a ratio scale, much of it going back to Stevens. (See the discussion in Roberts (1979/2009, pp. 179-180).) How then should we choose the most important variables? By the discussion above, it is "safer" to combine the experts' importance ratings by using geometric means and then to choose the most important variables as those having the highest geometric mean relative importance ratings, than it is to do this by using arithmetic means. That is why Roberts (1972, 1973) used geometric means.

6 Evaluation of Alternative Optical Instruments for Measuring Vegetation Indices

Various indices have been developed to characterize type, amount, and condition of vegetation present. Remote sensing is often used for this purpose. Among the indices of interest are the leaf area index and the normalized difference vegetation index, both based on spectral reflectance (Jackson and Huete, 1991). Recent developments have provided a variety of new types of optical remote sensing equipment for estimatng reflectance characteristics and thus calculating indices (see e.g., van Wijk and Williams, 2005). What if we want to compare alternative remote sensing devices that are candidates for this use? How might we do it?

One common procedure for comparing alternative instruments, machines, treatments, etc. is the following. A number of instruments are compared on different criteria/benchmarks. Their scores on each criterion are normalized relative to the score of one of the instruments. The normalized scores of an instrument are combined by some averaging procedure and average scores are compared. If the averaging is the arithmetic mean, then consider the statement:

Statement N "One instrument has a higher arithmetic mean normalized score than another instrument."

Table 1: Score of Instrument i on Criterion j

Instrument/Criterion	A	B	C	D	E
I	417	83	66	39,449	772
II	244	70	153	33,527	368
III	134	70	135	66,000	369

Table 2: Normalizing Relative to Instrument I

Instrument/Criterion	A	B	C	D	E	Arithmetic Mean	Geometric Mean
I	1.00	1.00	1.00	1.00	1.00	1.00	1.00
II	.59	.84	2.32	.85	.48	1.01	.86
III	.32	.85	2.05	1.67	.45	1.07	.84

Statement N is meaningless: The instrument to which scores are normalized can determine which has the higher arithmetic mean. Similar methods are used in comparing performance of alternative computer systems or other types of machinery. To illustrate, consider a number of potential criteria for optical instruments for measuring vegetation indices: Accuracy on cloudy days, accuracy with low-stature vegetation, accuracy for extremely diverse forests, ease of use, reliability, etc.

Table 1 shows three instruments I, II, III and five criteria A, B, C, D, E, with the i, j entry giving the score of the ith treatment on the jth criterion. Table 2 shows the score of each instrument normalized relative to treatment I, i.e., by dividing by instrument I's score. Thus, for example, the 1,2 entry is 83/83 = 1, while the 2,2 entry is 70/83 = .84. The arithmetic means of the normalized scores in each row are also shown in Table 2. We conclude that instrument III is best.

However, let us now normalize relative to Instrument II, obtaining the normalized scores of Table 3. Based on the arithmetic mean normalized scores of each row shown in Table 3, we now conclude that Instrument I is best. So, the conclusion that a given instrument is best by taking arithmetic mean of normalized scores is meaningless in this case: Statement N is meaningless.

The numbers in this example are taken from Fleming and Wallace (1986), with data from Heath (1984),

Table 3: Normalizing Relative to Instrument II

Instrument/Criterion	A	B	C	D	E	Arithmetic Mean	Geometric Mean
I	1.71	1.19	.43	1.18	2.10	1.32	1.17
II	1.00	1.00	1.00	1.00	1.00	1.00	1.00
III	.55	1.00	1.88	1.97	1.08	1.07	.99

and represent actual scores of alternative "instruments" in a computing machine application.

Sometimes, geometric mean is helpful. The geometric mean normalized scores of each row are shown in Tables 2 and 3. Note that in each case, we conclude that Instrument I is best. In this situation, it is easy to show that the conclusion that a given instrument has highest geometric mean normalized score is a meaningful conclusion. It is even meaningful to assert something like: A given instrument has geometric mean normalized score 20 per cent higher than another instrument.

Fleming and Wallace give general conditions under which comparing geometric means of normalized scores is meaningful. We have now given several examples where comparing geometric means leads to meaningful conclusions while comparing arithmetic means does not. However, there are situations where comparing arithmetic means leads to meaningful conclusions and comparing geometric means does not. It is a research area in measurement theory, with a long history and large literature, to determine what averaging procedures make sense in what situations. For some further details on this topic, and in particular for an example where arithmetic mean comparison is meaningful while geometric mean is not, see see Roberts (2012).

The message from measurement theory is: Do not perform arithmetic operations on data without paying attention to whether the conclusions you get are meaningful.

7 Optimization Problems in Landscape Ecology

Raster datasets represent geographic features by dividing the world into discrete sqaure or rectangular cells laid out in a grid. Each cell has a characteristic value that is used to represent some characteristic of that location. As noted by Zettenberg (2009), a GIS raster can be seen as a network, with grid cells as nodes and a link (edge) from each cell to its vertical, horizontal, and diagonal neighbors. The links might have weights or costs on them. According to Zettenberg, the least-cost path in between two nodes can represent a "geodesic path" between two points "(approximated by grid cells) on a projected surface. .. Even though the straight line Euclidean distance is a lot shorter, it may be *functionally* shorter for example to follow a detour along a preferred habitat." As Zettenberg also says, within the raster, the "cost-distance value at any point (i.e., grid cell) is the least-cost distance from that point to the closest specified source point." Sometimes we seek "patches" made up of cells with cost-distance value below some threshold that corresponds to some ecologically relevant value. The problem of finding the shortest distance between two nodes in a network (where the "length" of a path is the sum of weights on edges in it) is a widely studied problem in operations research and there are very efficient algorithms for solving it. The shortest path problem occurs widely in practice. In the US, just one agency of the US Department of Transportation in the federal government has applied algorithms to solve this problem literally billions of times a year (Goldman, 1981).

Consider a simple network with nodes x, y and z and edges from x to y with weight 2, y to z with weight 4, and x to z with weight 15. What is the shortest path from x to z in this network? The shortest path is the path that goes from x to y to z, with a total "length" of 6. The alternative path that goes directly from x to z has total "length" 15. Is the conclusion that x to y to z is the shortest path a meaningful conclusion?

The conclusion is meaningful if the weights on edges define a ratio scale, as they do if they are physical distances or monetary amounts. However, what if they define an interval scale? This could happen if the weights are utilities or values, rather than dollar amounts or physical lengths. As noted earlier, utilities might be defined on interval scales. If the weights define an interval scale, consider the admissible transformation $\phi(x) = 3x + 100$. Now the weights change to 106 on the edge from x to y, 112 on the edge from y to z, and 145 on the edge from x to z. We conclude that going directly from x to z is the shortest path. The original conclusion was meaningless.

The shortest path problem can be formulated as a linear programming problem. Thus, the conclusion that A is the solution to a linear programming problem can be meaningless if cost parameters are measured on an interval scale. Note that linear programming is widely used in landscape ecology as well as in other areas of application. For example, it is used to determine optimal inventories of equipment, assignments of researchers to projects, optimization of the size of an ecological reserve, amount to invest in preventive treatments, etc.

Another very important practical combinatorial optimization problem is the minimum spanning tree problem. Given a connected, weighted graph or network, we ask for the spanning tree with total sum of costs or weights as small as possible. (A *spanning tree* is a tree that includes all the nodes of the network.) This problem has applications in the planning of large-scale transportation, communication, and distribution networks, among other things. Minimum spanning trees arise in landscape ecology in the following way. Following Urban and Keitt (2001), consider a landscape of habitat patches. Build a graph whose nodes are the patches, with an edge between patches if there is some "ecological flux" between them, e.g., via dispersal or material flow. Put weights on the edges to reflect flow rates or dispersal probabilities. Next, the patches are rated in terms of their "importance." We consider patterns of habitat loss and degradation. In a simplified model, we remove patches in entirety one at a time, i.e., remove available habitat gradually, one patch at a time. This amounts to removing one node from the graph at a time. We study preservation of species by asking how much habitat must be removed before that species become extinct (at least in the system being modeled).

Urban and Keitt studied the following patch-removal algorithm: Find a minimum spanning tree that has a "leaf" (node with only one neighbor) of smallest importance and remove the patch corresponding to that leaf. Then repeat the process on the remaining graph. Urban and Keitt studied this process for the Mexican Spotted Owl. In 1993, this subspecies was listed as threatened under the Endangered Species Act in the US. Habitat distribution for this species is highly fragmented in the US Southwest. By using this patch-removal algorithm, Urban and Keitt found in simulation models that the Mexican Spotted Owl population actually increased until nearly all the habitat was removed. By way of contrast, if patches were removed in random order, the owl population declined dramatically as habitat was removed. Urban and Keitt explain their algorithm by noting that the spanning tree "maintains the integrity of the landscape by not only providing large core populations, but also by providing dispersal routes between core habitats."

It is natural to ask if the conclusion that a given set of edges defines a minimum spanning tree is meaningful. (In Urban and Keitt's work, determining the scale type of the edge-weights is a rather complex issue.) It is surprising to observe that even if the weights on the edges define only an ordinal scale, then the conclusion is meaningful. This is not a priori obvious. However, it follows from the fact that the well-known algorithm known as Kruskal's algorithm or the greedy algorithm gives a solution. In Kruskal's algorithm (Kruskal, 1956, Papadimitriou and Steiglitz, 1982), we order edges in increasing order of weight and then examine edges in this order, including an edge if it does not form a cycle with edges previously included. We stop when all nodes are included. Since any admissible transformation will not change the order in which edges are examined in this algorithm, the same solution will be produced.

Many practical decision making problems in landscape ecology, environmental sustainability, and other fields involve the search for an optimal solution as in the shortest path and minimum spanning tree problems. Little attention is paid to the possibility that the conclusion that a particular solution is optimal may be an accident of the way that things are measured. For the beginnings of the theory of meaningfulness of conclusions in combinatorial optimization, see Mahadev, Pekeč, and Roberts (1998), Pekeč (1996a, 1996b), and Roberts (1990, 1994, 1999).

There is much more analysis of a similar nature in the field of landscape ecology or the study of sustainable environments that can be done with the principles of measurement theory. The issues involved present challenges both for theory and for application.

References

[1] Dalton, H. (1920). The measurement of inequality of incomes. *The Economic Journal, 30* 348-361.

[2] Fleming, P.J., & Wallace, J.J. (1986). How not to lie with statistics: The correct way to summarize benchmark results. *Communications of the ACM, 29*, 218-221.

[3] Gini, C. (1909). Il diverso accrescimento delle classi sociali e la concentrazione della richezza. *Giornale degli Economist, serie II*, 37.

[4] Gini, C. (1912). Variabilite mutabilita. *Studi Economicoaguridic della Facotta di Giurisprudenza dell Univ. di Cagliari III. Parte II.*

[5] Goldman, A.J. (1981). *Discrete mathematics in government.* Lecture presented at SIAM Symposium on Applications of Discrete Mathematics, Troy, NY, June 1981.

[6] Heath, J.L. (1984). Re-evaluation of RISC I. *Comput. Archit. News, 12*, 3-10.

[7] Jackson, R.D., & Huete, A.R. (1991). Interpreting vegetation indices. *Preventive Veterinary Medicine, 11*, 185-200.

[8] Krantz, D.H., Luce, R.D., Suppes, P, & Tversky, A. (1971). *Foundations of measurement.* (Vol. I). New York: Academic Press.

[9] Kruskal, J.B. (1956) On the shortest spanning tree of a graph and the traveling salesman problem. *Proc. Amer. Math. Soc., 7*, 48-50.

[10] Luce, R.D., Krantz, D.H., Suppes, P., & Tversky, A. (1990). *Foundations of measurement.* (Vol. III). New York: Academic Press.

[11] Magurran, A.E. (1991). *Ecological Diversity and its Measurement.* London: Chapman & Hall.

[12] Mahadev, N.V.R., Pekeč, A., & Roberts, F.S. (1998) On the meaningfulness of optimal solutions to scheduling problems: Can an optimal solution be non-optimal? *Operations Research, 46 suppl*, S120-S134.

[13] Papadimitriou, C.H., & Steiglitz, K. (1982). *Combinatorial optimization: Algorithms and complexity*, Englewood Cliffs, NJ: Prentice-Hall

[14] Pekeč, A.. (1996a). *Limitations on conclusions from combinatorial optimization*, Ph.D. Thesis, Department of Mathematics, Rutgers University.

[15] Pekeč, A. (1996b). Scalings in linear programming: Necessary and sufficient conditions for invariance. Center for Basic Research in Computer Science (BRICS), technical report RS-96-50.

[16] Pfanzagl, J. (1968). *Theory of measurement.* New York: Wiley.

[17] Roberts, F.S. (1972). Building an energy demand signed digraph I: Choosing the nodes. Rept. $927/1 - NSF$. April. Santa Monica, CA: The RAND Corporation.

[18] Roberts, F.S. (1973) Building and analyzing an energy demand signed digraph," *Envir. & Plan., 5*, 199-221.

[19] Roberts, F.S. (1979). Measurement theory, with applications to decisionmaking, utility, and the social sciences, Reading, MA: Addison-Wesley. Digital Reprinting (2009). Cambridge, UK: Cambridge University Press.

[20] Roberts, F.S. (1990). Meaningfulness of conclusions from combinatorial optimization. *Discr. Appl. Math., 29*, 221-241.

[21] Roberts, F.S. (1994). Limitations on conclusions using scales of measurement. In S.M. Pollock, M.H. Rothkopf, & A. Barnett (Eds.), *Operations research and the public sector*, Vol. 6 in *Handbooks in operations research and management science.* (621-671). Amsterdam: North-Holland.

[22] Roberts, F.S. (1999). Meaningless statements. In *Contemporary trends in discrete mathematics*, DIMACS Series, (Vol. 49, 257-274). Providence, RI: American Mathematical Society.

[23] Roberts, F.S. (2012). Meaningful and meaningless statements in epidemiology and public health. In B. Berglund, G.B. Rossi, J. Townsend, & L. Pendrills (Eds.), *Measurements with persons.* (75-95). Taylor and Francis.

[24] Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379-423.

[25] Simpson, E.H. (1949). Measurement of diversity. *Nature, 163*, 688.

[26] Stevens, S.S. (1946). On the theory of scales of measurement. *Science 103*, 677-680.

[27] Stevens, S.S., (1951). Mathematics, measurement, and psychophysics. In S.S. Stevens (Ed.), *Handbook of experimental psychology* (1-49). New York: Wiley.

[28] Stevens, S.S. (1959). Measurement, psychophysics, and utility. In C.W. Churchman & P. Ratoosh (Eds.), *Measurement: Definitions and theories*, (18-63). New York: Wiley.

[29] Stevens, S.S. (1968). Ratio scales of opinion. In D.K. Whitla (Ed.), *Handbook of Measurement and assessment in behavioral sciences*, Reading, MA: Addison-Wesley.

[30] Suppes, P. (1959). Measurement, empirical meaningfulness and three-valued logic. In C.W. Churchman & P. Ratoosh (Eds.), *Measurement: Definitions and theories* (129-143). New York: Wiley.

[31] Suppes, P. (1979). Replies. In R.J. Bogdan (Ed.), *Patrick Suppes* (207-232). Dordrecht, Holland: Reidel.

[32] Suppes, P., Krantz, D.H., Luce, R.D., & Tversky, A. (1989) *Foundations of measurement.* (Vol. II). New York: Academic Press.

[33] Suppes, P., & Zinnes, J. (1963) Basic measurement theory. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, (Vol. 1, 1-76). New York: Wiley.

[34] UNEP (2002). Report of the sixth meeting of the conference of the parties to the Convention on Biological Diversity (UNEP/CBD/COP/6/20).

[35] Urban, D, & Keitt, T. (2001). Landscape connectivity: A graph-theoretic perspective. *Ecology, 82*, 1205-1218.

[36] van Wijk, M.T., & Williams, M. (2005). Optical instruments for measuring leaf area index in low vegetion: Application in Arctic ecosystems. *Ecological Applications, 15*, 1462-1470.

[37] Zettenberg, A. (2009), *Network based tools and indicators for landscape ecological assessments, planning, and design*, Licentiate Thesis, KTH-Environmental Management and Assessement Research Group, Department of Land and Water Resources Engineering, Royal Institute of Technology (KTH), Stockholm, Sweden.