

# **What is a good index? Problems with statistically based indicators and the Malmquist index as alternative.**

**G. Whittaker<sup>1</sup>, S. Lautenbach<sup>2,3</sup>, M. Volk<sup>2</sup>**

<sup>1</sup> *ARS, USDA, National Forage Seed Production Research Center, 3450 SW Campus Way, Corvallis, OR 97331-7102, USA, [jerry.whittaker@ars.usda.gov](mailto:jerry.whittaker@ars.usda.gov)*

<sup>2</sup> *UFZ – Helmholtz Centre for Environmental Research, Department of, Computational Landscape Ecology, Leipzig, Germany*

<sup>3</sup> *Department of Urban Planning and Real Estate Management, University Bonn, Germany*

**Abstract:** Conventional multivariate statistical methods have been used for decades to calculate environmental indicators. These methods generally work fine if they are used in a situation where the method can be tailored to the data. But there is some skepticism that the methods might fail in the context of skewed data distributions or spatial auto-correlation. Further, the indicators developed by statistical approaches can not be used to compare different regions or time periods that had not been covered by the input data. The aim of the paper is to demonstrate some of the shortcomings and to identify how the Malmquist index might be used as an alternative. The paper presents the results of an exhaustive review in the field of environment, hydrology and water quality which identified the most commonly used approaches. Then principal component analysis as representative of these approaches and the Malmquist index are challenged with simulated time series data to demonstrate the failure of statistical methods in two of the most common problems faced in construction of a water quality index.

**Keywords:** index; malmquist

## **1. Introduction**

Over several decades there have been many attempts to produce satisfactory indexing systems. One of the main reasons for producing such indexing systems is to provide scientists and managers with easily assimilable information about environmental quality issues - such as for instance water quality. Water quality is a complex issue, influenced by numerous factors. As a consequence, a large number different approaches to the calculation of a water quality index (WQI) have been developed that reflect diverse motivations in hydrology, environmental sciences, biology, etc., or variable reduction for monitoring. As Beamonte Cordoba et al. (2010) state, water quality is intrinsically tied to the different intended uses of the water; different uses require different criteria. Water quality is one of the most important factors that must be considered when evaluating the sustainable development of a given region. Concerns over how to soundly measure water quality have increased in recent years, also with regard to the implementation of related acts and directives to improve water quality (Jarvie et al., 1997; Ullrich and Volk, 2010).

Many physical and chemical characteristics can be used to evaluate water quality or the degree of water pollution. Therefore, it is not possible in practice to clearly define water quality by separately examining the behavior of every individual variable (either spatially or temporally). The (difficult) alternative consists of integrating the values of a set of physical and chemical variables into a unique value (i.e., an overall or global index). The techniques to construct the indices include cluster analysis (Khalil et al., 2010; Styers et al., 2010), principal components analysis and factor analysis (Ou et al., 2010; Tran et al., 2010), discriminant analysis (Feio et al., 2009; Kane et al., 2009), spatial weighting using multivariate statistical analysis, and fuzzy logic (Ghosh and Mujumdar, 2010). This work has been extended to include errors in measurement of the constituent variables used to construct the water quality index (Ghosh and Mujumdar, 2010; Taheriyoun et al., 2010).

The fundamental idea for construction of an index is that two or more quantities of interest can be represented by a single number. The defining issue for calculation of an index is how the quantities one is interested in can be weighted and combined. The underlying assumption of these water quality indices is that the calculated index is representative of the measured components in some sense. How such an index should be constructed and the necessary properties to ensure that it is representative have been the subject of research since the late nineteenth century (see for example Pierson, 1895). The interpretation of how an index is “representative” of the underlying data has to led to several approaches, including the economic, test (or axiomatic), and the Divisia approaches. We have been unable to find any reference to this large, mature subject literature in any study of water quality indices.

We accept that the axiomatic approach to construction of index is necessary to answer the question “what is a good index?”. Indices have to meet the requirements of homogeneity, time reversibility, transitivity, and dimensionality (Fisher, 1922):

1. homogeneity  $Q(\lambda x^t, x^{t+1}) = \lambda Q(x^t, x^{t+1})$ ,  $\lambda > 0$ ,
2. time-reversibility  $Q(x^t, x^{t+1}) \bullet Q(x^{t+1}, x^t) = 1$ , that is, when the quantity for one time period is exchanged with another, the resulting index is the reciprocal of the original index. Diewert (1992) p. 218 remarks that Pierson (1895) was so upset with the “fact that many of the commonly used index number formulae did not satisfy this test that he proposed that the entire concept of an index number should be abandoned.”
3. transitivity (or circularity test)  $Q(x^t, x^{t+1}) \bullet Q(x^{t+1}, x^{t+2}) = Q(x^t, x^{t+2})$ . This property means that whether a fixed base or a chain of observations is used to calculate the index, the result will be the same.
4. dimensionality  $Q(\lambda x^t, \lambda x^{t+1}) = Q(x^t, x^{t+1})$ , that is, if we change the units of measurement for each commodity by the same positive number  $\lambda$ , then the index remains unchanged.

Our motivation for the study of water quality indices lies in the specification of objectives in a multiple objective optimization approach to agri-environmental policy analysis. In our integrated modeling system, we require that a WQI is can be calculated without human intervention. This additional requirement limits acceptable approaches to those where the weights used in WQI calculation are endogenous. That is, the decision to assign importance to a variable is based solely on the data available for the calculation, and is carried out in a computational algorithm. This limits our choices to statistical methods or Malmquist index related methods.

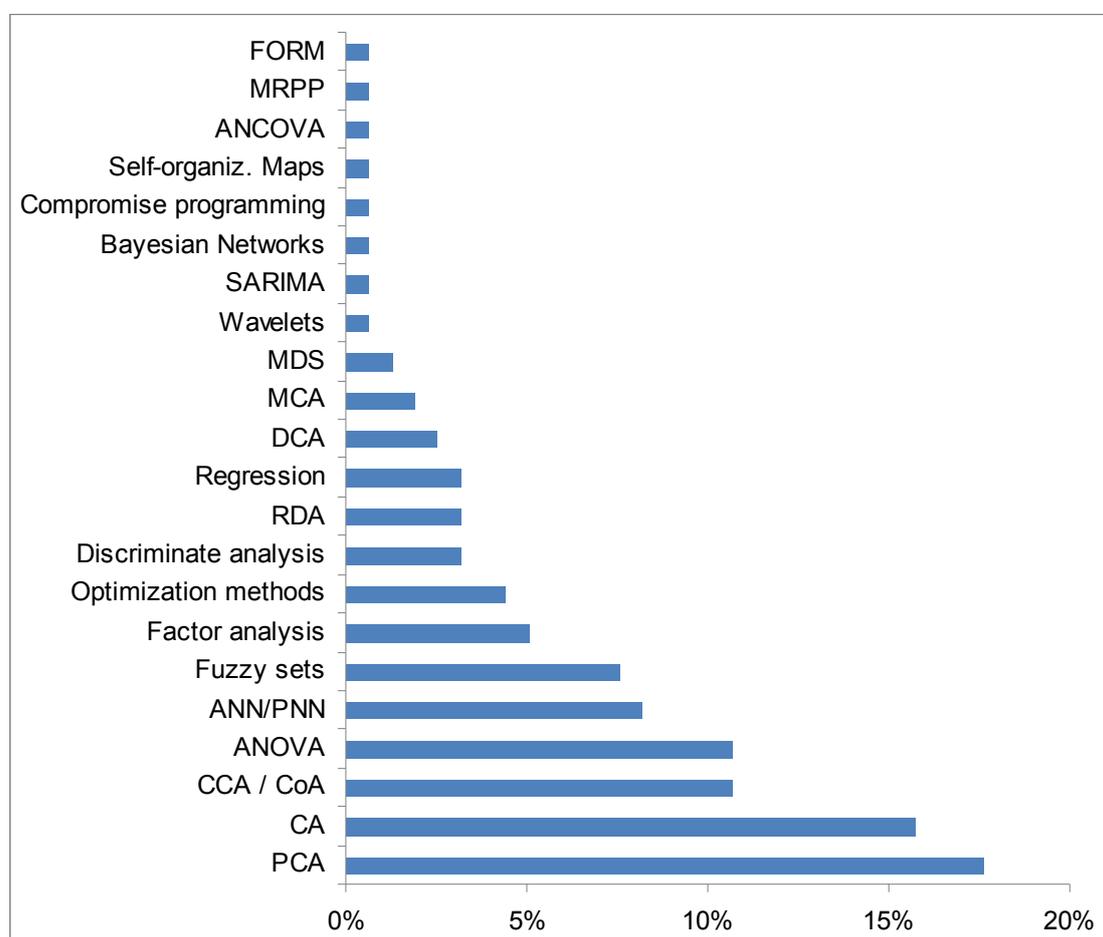
We carried out an exhaustive review of papers dealing with the development or application of indices in water quality related studies to get an overview over the most applied statistical methods in this field. In the next step, we tested the indices calculated by these methods for time reversibility. As own contribution to index development, we test a water quality index that is based on the foundation of index theory. Our approach is based on the Malmquist (1953) consumer quantity index. In particular

we apply Shephard's (1953) input distance functions and following Malmquist, use ratios of them to define the index. To calculate the water quality index, we apply a method alternatively known as Data Envelopment Analysis (DEA, (Charnes et al., 1978) or Activity Analysis (von Neumann, 1937). For the theory of index numbers, see Diewert (1987). The weights applied to each metric in the construction of the water quality index are endogenous using this approach. In the second section, we discuss the theoretical basis for the construction of a water quality index.

## 2. Methods

### 2.1 Review of water quality index methods

We found 824 papers in ISI web of science (state November 2011) by using "environment index management" for Topic and "Environmental Sciences, Ecology, Water Resources, Agriculture Multidisciplinary" as refining terms. 189 of the papers used statistical methods – the others relied on externally defined weighting schemes or weights assigned by experts. Figure 1 shows the results of our analysis.



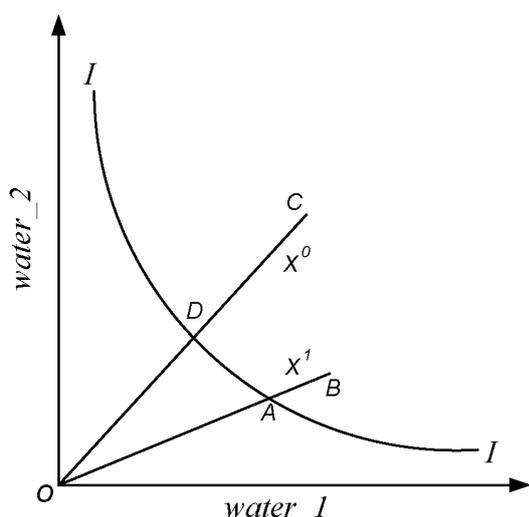
**Figure 1.** Methods used in the literature reviewed.

Most studies used principal component analysis, followed by cluster analysis and canonical correspondence analysis / correspondence analysis. Analysis of Variance (17), Artificial Neural Networks (15), Fuzzy sets (13) as well Factor Analysis (11) are next in the ranking. Other methods include discriminate analysis, SARIMA (multiplicative seasonal autoregressive integrated moving average model), multiple regression, MRPP (Multiresponse Permutation Procedure), MDS (Multi-dimensional scaling), FORM (first-order reliability method) and MCA (Multicriteria analysis).

## 2.2 The Malmquist index

The idea of Malmquist(1953) was to introduce a benchmark curve (an indifference curve in this case) and

measure the distance of each vector from the curve. The ratio of these distances defines the index. The benchmark curve can be calculated from the data or can be set, e.g. by legislative water quality thresholds or other regulative thresholds.



**Figure 2.** Construction of the Malmquist WQI with two constituents of water quality.

The two dimensional case is illustrated in Figure 2, where we wish to compare vector  $x^0$  of length  $OC$  with vector  $x^1$  of length  $OB$ . The benchmark that we use for comparison is the best practice benchmark  $II$ , where we are indifferent as to choice among points that lie on the curve. Each distance  $AB$  and  $DC$ , known as Shephard's input distance function, is the distance that each vector can be proportionally reduced to reach the benchmark ( $II$ ). The ratio of the input distance functions, i.e.  $(OB / OA) / (OC / OD)$  defines the Malmquist input quantity index, and we conclude for this example that  $x^1$  has better water quality than  $x^0$ , it requires a smaller contraction to reach  $II$  than  $x^0$  does.

In this paper we calculate the distances using a linear programming method known as Data Envelopment Analysis (DEA) (Charnes et al., 1978) or Activity Analysis (von Neumann, 1937). We assume there

are  $N$  characteristics of water quality  $x \in \mathfrak{R}^N$ , and  $k=1, \dots, K$  measurements (observations) of water characteristics at  $t=1, \dots, T$  time periods. Our data is represented as:

$$x_k^t, \quad k=1, \dots, K, \quad t=1, \dots, T, \quad x_k^t \in \mathfrak{R}_+^N. \quad (1)$$

The  $z_k^t \geq 0$ , are intensity variables. The intensity variables form the convex hull of all the data points.

The lower boundary of this set is the best practice benchmark isoquant  $II$  in Figure 2. For each observation  $(k', t)$  we compute:

$$\begin{aligned} (D(x_k^t))^{-1} &= \min \theta \\ \text{s.t.} \quad &\sum_{k=1}^K \sum_{t=1}^T z_k^t x_{kn}^t \leq \theta x_{k'n}^t \quad n=1, \dots, N, \\ &\sum_{k=1}^K \sum_{t=1}^T z_k^t = 1 \\ &z_k^t \geq 0, \quad t=1, \dots, T, \quad k=1, \dots, K. \end{aligned} \quad (2)$$

Suppose we compare  $(k', t)$  with  $(k', t+1)$ , then the Malmquist water quality index (WQI) is

$$WQI_{k'} = \frac{D(x_k^{t+1})}{D(x_k^t)}. \quad (3)$$

In this application, we compare all measurements through all time periods. If  $WQI = 1$ , water quality is equal to the best practice.  $WQI$  values greater than 1 indicate a poorer water quality. To implement

the calculation, we use a linear program to solve (2) for each observation, then calculate (3), the water quality index.

### 3. Comparison of alternative calculation of indices

The fundamental difficulty in comparing methods for calculation of index lies in the fact that no “observed” index exists to compare with alternative approaches to calculation. In all of these studies, the index is most commonly validated by a statistical comparison with observations of variables that characterize the purpose of the index. Correlation is often used as the measure in comparison of indices to observations (see Schletterer et al. (2010), for example). Lopez y Royo et al. (2010) combine validation with index construction by individually relating the constituent variables to the condition of seagrass, then combining the resulting scaled categories to calculate an index. Comparison of indices with simulation results have also been used for validation (Feio et al., 2009).

In this study, we compare the results of a Malmquist index number calculation based on a synthetic data set with the most common statistical approach, principle components analysis (PCA). PCA is representative of the statistical approach in that it establishes the weights for calculation of an index based a statistical property of the data, in this case, variance. The use of a statistical property of the data endogenizes the weights, an important characteristic of the index to be created. Unfortunately, this is only useful in the special case there the statistical property, here the variance, is also the property of interest in comparing observations. In any other situation, there is no reason to expect that variance has any relation to the weights in calculation of a useful index.

We constructed a synthetic data set with 2 iid variables: variable 1, mean = 50, variance =1; variable 2, mean = 50, variance = 2, 100 observations each. Principal components were estimated using the Princomp package in R (cf. Table 1). Since variable 1 has a low variance compared to variable 2, variable 1 will dominate a WQI calculated using a method that depends on explanation of variance.

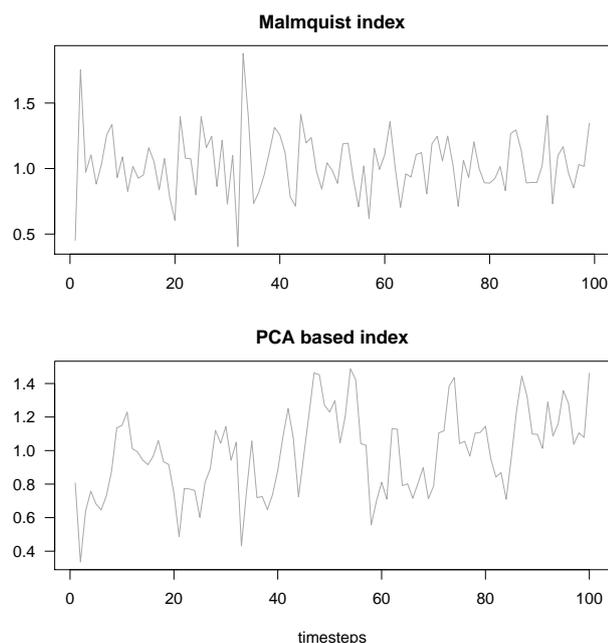
A WQI is commonly calculated from the PCA results using the product of the proportion of variation and the loadings (eigenvalues) for each constituent (variable) (see Qian, 2007 for an example). To demonstrate where PCA fails in construction of an index we generated an output variable from  $0.9 \times$  variable 1 plus a random, normally distributed error. This synthetic output variable represents information that may be used to assign importance such as fish population. The correlation between the synthetic WQI based on the PCA and the synthetic output is very low, 0.06. There is no reason to expect a better correlation because there is no connection between any statistical property of the PCA based index and a comparison of variables. For heteroscedastic data, one can expect even more problems.

**Table 1.**Results of the PCA approach

	Importance of components			Loadings	
	Comp.1	Comp.2		Comp.1	Comp.2
Standard deviation	1.90	1.02	Variable 1	0.14	0.99
Proportion of Variance	0.78	0.22	Variable 2	0.99	0.14
Cumulative Proportion	0.78	1.0			

The Malmquist approach is in contrast to the PCA nonparametric, and requires no assumptions about functional form or distribution. When we analyzed the same synthetic data set using the Malmquist index approach, the WQI correlation with the synthetic output increased to 0.21. Furthermore, the output can be added to the linear programming problem so that the Malmquist index takes account of the relation of the output to the inputs in calculation of the WQI. This is an important feature lacking in almost all other methods.

Data on constituents of water quality are commonly collected through time, and the observed values are often correlated between time periods. Termed autocorrelation for a single time series, this common property of water quality measurements can cause calculation of a WQI to fail because the property of time reversibility is not met. The Malmquist WQI has the property of time reversibility. We demonstrate this problem by again comparing PCA for calculation of a WQI with a Malmquist WQI. Two time series were generated:  $x_t = 0.9 x_{t-1} + \alpha_t$  where  $\alpha$  is iid with mean 3, variance 1 for times series one and iid with mean 9, variance 3 for time series two. We calculated a WQI using PCA and the Malmquist WQI following the equations given in section 2.2 (cf. Figure 3 for the time series of the index values), and tested for time reversibility using the consistent metric entropy test of asymmetry as described by Maasoumi and Racine (2009). The chosen test is nonparametric, and takes account of autocorrelation. The test for time reversibility (symmetry) was rejected (p-value: 0.04) for the WQI constructed using PCA, but could not be rejected (p-value: 0.40) for the Malmquist WQI.



**Figure 3.** The Malmquist and the PCA based index for the simulated time series data.

#### 4. Summary and conclusions

The mathematical properties required for the construction of an index have been studied for more than century, and are well known. The Malmquist index approach to construction of a water quality index is derived from this knowledge base, and provides a method that is easily calculated, objective, and requires no information beyond the raw data set. The index has been shown to have the required property of time reversibility. A survey of methods used for water quality index construction revealed that statistical approaches are commonly used, where the water constituents are weighted based on statistical properties of the data. We demonstrated that in two common situations, PCA as the most commonly used statistical approach failed to produce a meaningful WQI, and the Malmquist index succeeded. Generally speaking, it can be expected that the calculation of a WQI using a statistical approach will provide a poor WQI if the statistical properties of water quality constituents do not happen to coincide with a knowledgeable evaluation of importance. The properties of the Malmquist WQI will prevent such problems.

#### REFERENCES

- Charnes, A., Cooper, W. W. & Rhodes, E., Measuring the efficiency of decision making units. *European Journal of Operational Research* 2, 429-444, 1978.
- Diewert, W. E. Index Numbers. In *The New Palgrave: A Dictionary of Economics*, 767-780. J. Eatwell, Milgate, M., Newman, P., eds. London: The Macmillan Press, 1987.
- Feio, M. J., Almeida, S. F. P., Craveiro, S. C. & Calado, A. J., A comparison between biotic indices and predictive models in stream water quality assessment based on benthic diatom communities. *Ecological Indicators* 9, 497-507, 2009.
- Fisher, I., *The Making of Index Numbers*. Houghton Mifflin, Boston, 1922.
- Ghosh, S. & Mujumdar, P. P., Fuzzy waste load allocation model: A multiobjective approach. *Journal of Hydroinformatics* 12, 83-96, 2010.
- Jarvie, H.P., Neal, C. & Tappin, A.D., European land-based pollutant loads to the North Sea: an analysis of the Paris Commission data and review of monitoring strategies. *The Science of the Total Environment* 194-195, 39-58, 1997.
- Jordan, S. J., Lewis, M. A., Harwell, L. M. & Goodman, L. R., Summer fish communities in northern Gulf of Mexico estuaries: Indices of ecological condition. *Ecological Indicators* 10, 504-515, 2010.
- Justus, B. G., Petersen, J. C., Femmer, S. R., Davis, J. V. & Wallace, J. E., A comparison of algal, macroinvertebrate, and fish assemblage indices for assessing low-level nutrient enrichment in wadeable Ozark streams. *Ecological Indicators* 10, 627-638, 2010.
- Kane, D. D., Gordon, S. I., Munawar, M., Charlton, M. N. & Culver, D. A., The Planktonic Index of Biotic Integrity (P-IBI): An approach for assessing lake ecosystem health. *Ecological Indicators* 9, 1234-1247, 2009.
- Kanno, Y., Vokoun, J. C. & Beauchene, M., Development of dual fish multi-metric indices of biological condition for streams with characteristic thermal gradients and low species richness. *Ecological Indicators* 10, 565-571, 2010.
- Khalil, B., Ouarda, T. B. M. J., St-Hilaire, A. & Chebana, F., A statistical approach for the rationalization of water quality indicators in surface water quality monitoring networks. *Journal of Hydrology* 386, 173-185, 2010.
- Lopez y Royo, C., Casazza, G., Pergent-Martini & C., Pergent, G., A biotic index using the seagrass *Posidonia oceanica* (BiPo), to evaluate ecological status of coastal waters. *Ecological Indicators* 10, 380-389, 2010.
- Malmquist, S., Index numbers and indifference surfaces. *Trabajos de Estadística* 4, 209-242, 1953.
- Marchini, A., Facchinetti, T. & Mistri, M., F-IND: A framework to design fuzzy indices of environmental conditions. *Ecological Indicators* 9, 485-496, 2009.
- Maasoumi, E. and Racine, J.S., "A robust entropy-based test of asymmetry for discrete and continuous processes," *Econometric Reviews*, 28, 246-261, 2009.
- Navarro-Llácer, C., Baeza, D. & de las Heras, J., Assessment of regulated rivers with indices based on macroinvertebrates, fish and riparian forest in the southeast of Spain. *Ecological Indicators* 10, 935-942, 2010.

- Ou, Y., Wang, X. & Li, M., Relating watershed characteristics to non-point source pollution in miyun reservoir watershed using principal components analysis (PCA). Chengdu, 2010.
- Pierson, N. G., 1895. Further considerations on index-numbers. *Economic Journal* 6, 127-131.
- Navarro-Llácer et al., 2010.
- Pinto, R., Patricio, J., Baeta, A., Fath, B. D., Neto, J. M. & Marques, J. C., Review and evaluation of estuarine biotic indices to assess benthic condition. *Ecological Indicators* 9, 1-25, 2009.
- Qian, Y., Migliaccio, K.W., Wan, Y. and Li, Y., Surface water quality evaluation using multivariate methods and a new water quality index in the Indian River Lagoon, Florida. *Water Resources Research*, 43: W08405, 2007.
- Schletterer, M., Füreder, L., Kuzovlev, V. V. & Beketov, M. A., Testing the coherence of several macroinvertebrate indices and environmental factors in a large lowland river system (Volga River, Russia). *Ecological Indicators* 10, 1083-1092, 2010.
- Seilheimer, T. S., Mahoney, T. P. & Chow-Fraser, P., Comparative study of ecological indices for assessing human-induced disturbance in coastal wetlands of the Laurentian Great Lakes. *Ecological Indicators* 9, 81-91, 2009.
- Sharifdoost, M., Mahmoodi, S. & Pasha, E., A Statistical Test for Time Reversibility of Stationary Finite State Markov Chains. *Applied Mathematical Sciences* 3(52), 2563-2574, 2009.
- Shephard, R. W., Cost and Production Functions. Princeton University Press, Princeton, New Jersey, 1953.
- Shiels, D. R., Implementing landscape indices to predict stream water quality in an agricultural setting: An assessment of the Lake and River Enhancement (LARE) protocol in the Mississinewa River watershed, East-Central Indiana. *Ecological Indicators* 10, 1102-1110, 2010.
- Simaika, J. P. & Samways, M. J., Comparative assessment of indices of freshwater habitat conditions using different invertebrate taxon sets. *Ecological Indicators*, In Press, Corrected Proof.
- Stoddard, J. L., Herlihy, A. T., Peck, D. V., Hughes, R. M., Whittier, T. R. & Tarquinio, E., A process for creating multimetric indices for large-scale aquatic surveys. *Journal of the North American Benthological Society* 27, 878-891, 2008.
- Styers, D. M., Chappelka, A. H., Marzen, L. J. & Somers, G. L., Scale matters: Indicators of ecological health along the urban-rural interface near Columbus, Georgia. *Ecological Indicators* 10, 224-233, 2010.
- Taheriyoun, M., Karamouz, M. & Baghvand, A., Development of an entropy-based Fuzzy eutrophication index for reservoir water quality evaluation. *Iranian Journal of Environmental Health Science and Engineering* 7, 1-14, 2010.
- Tran, C. P., Bode, R. W., Smith, A. J. & Kleppel, G. S., Land-use proximity as a basis for assessing stream water quality in New York State (USA). *Ecological Indicators* 10, 727-733, 2010.
- Ullrich, A. & Volk, M., Influence of different nitrate-N monitoring strategies on load estimation as a base for model calibration and evaluation. *Environmental Monitoring and Assessment* 171: 513-527, 2010.
- von Neumann, J., A Model of General Economic Equilibrium. In Ergebnisse eines mathematischen Kolloquiums, 1935-36. K. Menger, ed: Translated and reprinted in 1945 in *Review of Economic Studies* 13 (1), 1-9, 1937.