Application of index number theory to the construction of a water quality index: Aggregated nutrient loadings related to the areal extent of hypoxia in the northern Gulf of Mexico

Citation	Whittaker, G., Barnhart, B., Färe, R., & Grosskopf, S. (2015). Application of index number theory to the construction of a water quality index: Aggregated nutrient loadings related to the areal extent of hypoxia in the northern Gulf of Mexico. Ecological Indicators, 49, 162-168. doi:10.1016/j.ecolind.2014.10.003
DOI	10.1016/j.ecolind.2014.10.003
Publisher	Elsevier
Version	Version of Record
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

Ecological Indicators 49 (2015) 162–168

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Application of index number theory to the construction of a water quality index: Aggregated nutrient loadings related to the areal extent of hypoxia in the northern Gulf of Mexico

Gerald Whittaker [a], Bradley Barnhart [a,*], Rolf Färe [b], Shawna Grosskopf [b]

[a] National Forage Seed Production Research Center, Agricultural Research Service, USDA, 3450 Campus Way, Corvallis, OR 97330, USA
[b] Department of Economics, Oregon State University, 319D Ballard Extension Hall, Corvallis, OR 97331, USA

ARTICLE INFO

Article history:
Received 28 January 2014
Received in revised form 3 October 2014
Accepted 7 October 2014

Keywords:
Eutrophication
Surface water quality
Gulf hypoxia
Economic productivity
Distance functions

ABSTRACT

Numerous studies have linked individual nutrient loadings from the Mississippi and Atchafalaya Rivers to the growth of the hypoxic, or oxygen depleted, zone in the northern Gulf of Mexico. However, in the discussion of policy to remediate Gulf hypoxia, it is beneficial for stakeholders and policymakers to obtain a single measure for water quality that characterizes information from multiple water pollutants. This study aggregates loadings from six nutrients measured at the entrance to the Gulf of Mexico into a single time-varying index of water quality. The index is constructed using traditional index number theory originating from economic production theory, mainly, Shephard's distance functions calculated using data envelopment analysis (DEA). The methodology is an advance over other index construction schemes because the determined metric weights are endogenous, calculated from the data itself, and do not require external user input. To validate the index, May values of the index are used within a statistical regression model to model the areal extent of Gulf hypoxia using mid-July cruise measurements from 1985 to 2013, excluding 1989 when no cruise data were available. Regression results ($R^2_{adj} = 0.81$) suggest the index is successful at aggregating multiple pollutants into a single measure of water quality and may be useful for tracking their aggregated effect on the growth of the hypoxia area in the northern Gulf of Mexico. Calculation of the water quality index described here is automatic in the sense that no human intervention is required for variable selection, statistical analysis or assignment of weights. This is very useful for specifying a water quality objective in a multiple objective optimization for watershed management.

1. Introduction

Water quality indices have been developed and used to characterize a wide variety of phenomena including drinking water (Beamonte Córdoba et al., 2010), bioassessment (Aguiar et al. 2014; Blanchet et al., 2008; Kanno et al., 2010; Stoddard et al., 2008), fresh water habitat (Pinto et al., 2009; Simaika and Samways, 2011), effect of agriculture on stream water (Justus et al., 2010; Shiels, 2010), river water quality (Feio et al., 2009; Navarro-Llácer et al., 2010), ecological condition (Jordan et al., 2010; Marchini et al., 2009; Seilheimer et al., 2009; Tran et al., 2008) and variable reduction for selection of variables to monitor (Kantoussan et al., 2010). A recent survey of approaches for constructing indices (Bierman et al., 2011) cite the techniques of cluster analysis (Hargiss et al., 2008; Khalil et al., 2010; Styers et al., 2010), principal components analysis, factor analysis (Blocksom and Johnson, 2009; Leunda et al., 2009; Liou et al., 2004; Ou et al., 2010; Tran et al., 2010), discriminant analysis (Feio et al., 2009; Kane et al., 2009), 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 (Beamonte Córdoba et al., 2010; Castoldi et al., 2009; Ghosh and Mujumdar, 2010; Sin et al., 2009; Taheriyoun et al., 2010).

In all of the studies cited above, the index is most commonly validated by a statistical comparison with observations of variables that characterize the purpose of the index. The underlying assumption of these water quality indices is that the calculated index is representative of the included components. How such an index should be constructed and what properties the index should have to ensure that it is representative of its included constituents have been the subject of extensive research over the past several hundred years. Original work on index theory began in 1707, where

* Corresponding author. Tel.: +1 541 738 4182.
E-mail address: bradleybarnhart@gmail.com (B. Barnhart).

http://dx.doi.org/10.1016/j.ecolind.2014.10.003
1470-160X/© 2014 Published by Elsevier Ltd.

a market basket of goods was utilized to compare prices among regions (see Diewert and Nakamura, 1993). Development of the consumer price index dates to 1823 (Lowe, 1823), and some of the properties necessary for the construction of a representative index were known by the late nineteenth century (see for example Pierson, 1895). For the theory of index numbers, see Diewert (1987). Nonetheless, few studies today utilize foundational index theory in order to construct water quality indices.

In this paper, we develop a water quality index that is based on the foundation of economic index theory. Our approach follows the work of Malmquist (1953) for constructing a consumer quantity index. In particular we apply Shephard's (1953, 1970) input distance functions to aggregate multiple water quality constituents into a single indicator of water quality. Caves et al. (1982) demonstrated that a Malmquist-type quantity index can be formulated using ratios of Shephard's (1953, 1970) distance functions. In practice, the distance functions are calculated using a linear programming method known as data envelopment analysis (DEA) (Charnes et al., 1978) or activity analysis (von Neumann, 1937). The method ensures that resulting indexes satisfy several mathematical properties including homogeneity, time-reversibility, transitivity, and dimensionality. Also, the weights applied to each metric in the construction of the water quality index are endogenous in our approach. This means they are not externally selected or subject to user bias, but are rather calculated from the data themselves. The technique provides a solution to common discussion on environmental performance indicators that often declare the need for index construction techniques that endogenously calculate metric weights and remove user bias (Bellenger and Herlihy, 2009, 2010; Tran et al., 2008; Tyteca, 1996; Zhou et al., 2007).

Only a small number of studies have used Shephard's (1953, 1970) input distance functions or similar approaches in order to aggregate multiple constituents into an indicator, or to produce environmental quality indexes using production theory (Bellenger and Herlihy, 2009, 2010; Färe et al., 2004; Zhou et al., 2008). Prompted by an environmental indicator review by Tyteca (1996), Färe et al. (2004) demonstrated a DEA approach to calculate an environmental performance index that simultaneously accounted for resources used, good outputs produced and bad outputs emitted. Most recently, Bellenger and Herlihy (2009, 2010) used a distance function approach to construct environmental performance indicators using six macroinvertebrate metrics and compared them to an existing Environmental Protection Agency index of biotic integrity in the Appalachian mountains.

This paper reiterates the utility of distance function techniques to aggregate multiple constituents into a single indicator. We apply these approaches to construct a water quality index utilizing six nutrient loadings supplied to the Gulf of Mexico from the Mississippi and Atchafalaya Rivers. The time-varying index, representing the aggregated effect of nutrient loadings, will then be validated by using it to model the areal extent of Gulf hypoxia.

The area of the summer hypoxic (oxygen $\leq 2\,\mathrm{mg\,l^{-1}}$) region in the northern Gulf of Mexico has been widely studied and publicized for nearly two and a half decades (Forrest et al., 2011; Obenour et al., 2012; Rabalais et al., 2001; Scavia et al., 2013). Since 1997, numerous sources have demonstrated high correlations of the Gulf hypoxic area with nutrient loadings delivered from the Mississippi and Atchafalaya Rivers (Forrest et al., 2011; Greene et al., 2009; Scavia et al., 2003; Turner et al., 2006; Wiseman et al., 1997). To track the variability of Gulf hypoxia, and perhaps the progress in reducing its size, shelfwide measurement cruises have been conducted annually in late July since 1985 (Rabalais et al., 2007). This data set has enabled researchers to model the response of Gulf hypoxia to a variety of forcing variables, including nitrogen and phosphorus concentrations, discharge, and wind speed (Forrest et al., 2011; Scavia and Donnelly, 2007; Scavia et al., 2003, 2004). May nitrogen loadings have been found to have the highest correlation with July hypoxia areas (Forrest et al. (2011) state $R^2 = 0.24$ using 24 measurements). Therefore, the May nitrogen loadings have been used to predict hypoxia areas in the months before each annual July cruise (Forrest et al., 2011; Scavia and Donnelly, 2007; Scavia et al., 2003, 2004, 2013). However, no single model for Gulf hypoxia incorporates more than two water pollutants (total N and total P) simultaneously, despite research that other nutrients may greatly affect eutrophication and hypoxia area, including silica and individual types of N and P (Correll, 1998; Howarth et al., 2011). Also, we are not aware of any previous study that has aggregated multiple nutrient loadings into a composite index of water quality for this region. This paper will construct a monthly water quality index from observations of six nutrient loadings measured at the mouth of the Gulf of Mexico, including dissolved nitrite plus nitrate, total organic nitrogen plus ammonia nitrogen (total Kjeldahl nitrogen), dissolved ammonia, total phosphorous, dissolved orthophosphate, and dissolved silica. The index will be validated by use in a statistical regression in order to model Gulf hypoxia area. Model results will be compared to results from the latest biophysical scenario and forecast model for Gulf hypoxia (Scavia et al., 2013).

In Section 2, the theory and methods necessary to construct the index will be discussed. Section 3 will discuss the data used as well as motivation for the formulation of the statistical regression model. Section 4 will display the resulting index and regression model results while Section 5 will include discussion and conclusions.

2. Theory

In general, an index attempts to compare two vectors x^0 and x^1, where each vector may contain N inputs so that $x^0, x^1 \in \Re^N_+$. The idea of Malmquist (1953) was to introduce a benchmark curve (an indifference curve, in his case) and measure the distance of an input vector from the curve. In this study, we assume that the water characteristics are undesirable, and should be minimized. A two-dimensional case is illustrated in Fig. 1, where we wish to compare vector x^0 of distance OC with vector x^1 of distance OB. The benchmark that we use for comparison is the best practice benchmark (II), where we are "indifferent" as to the choice among points that lie on the curve. Each distance DC and AB is the distance that can be proportionally reduced for each vector to reach the benchmark (II). We conclude for this example that x^1 has better water quality than x^0 because it requires a smaller contraction to reach II than x^0 does.

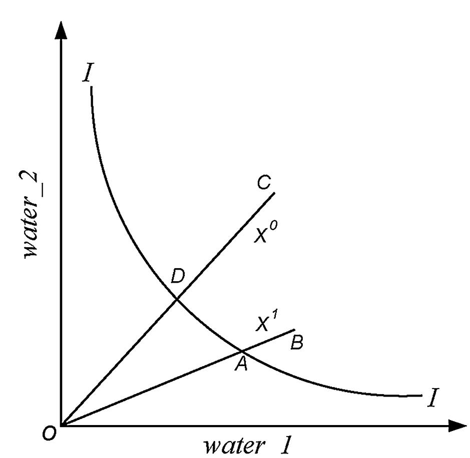

Fig. 1. Demonstration of index construction with two constituents of water quality.

A more formal approach to index construction is presented in economic production theory (Bellenger and Herlihy, 2009, 2010; Färe et al., 2004; Shephard, 1970). Consider a set of producers that form the technology for a given production process defined as

$$T = \{(x, y) : x \text{can produce} y\} \tag{1}$$

where x is a vector of inputs, and y is a vector of outputs. If outputs are held fixed the input requirement set is defined as

$$L(y) = \{x : (x, y) \in T\}. \tag{2}$$

Again, we assume water pollutants are inputs to be minimized, and set all outputs equal to one. The benchmark isoquant (II) in Fig. 1 represents the lower bound of the input requirement set. That is, it is the minimum observed input combinations required to achieve a fixed output $y = 1$ (Bellenger and Herlihy, 2009). The distances of each input vector, x^0 and x^1, to the benchmark can be calculated using Shephard's distance functions (Shephard, 1970),

$$D_i(1, x^k) = \sup\left\{\lambda : \frac{x^k}{\lambda} \in L(1)\right\}, \tag{3}$$

where λ is the necessary proportional contraction to reduce the input vector to the benchmark curve. In our Fig. 1 example, $D_i(1, x^0) = DC/OC$ and $D_i(1,x^1) = AB/OB$.

In the following we assume that the technology is homothetic so that the distance function takes the form

$$D_i(1, x) = \frac{\hat{D}_i(x)}{F(1)} \tag{4}$$

Then, following Caves et al. (1982), a quantity index (named Malmquist after the extensive work of Malmquist (1953)) can be calculated as

$$Q(x^j, x^k) = \frac{D_i(1, x^j)}{D_i(1, x^k)} = \frac{\hat{D}_i(x^j)}{\hat{D}_i(x^k)} \tag{5}$$

where x^j and x^k are two input vectors to be compared. This index satisfies the following properties due to Fisher (1922):

1. Homogeneity $Q(\alpha x^j, x^k) = \alpha Q(x^j, x^k), \alpha > 0$
2. Time-reversibility $Q(x^j, x^k) \times Q(x^k, x^j) = 1$
3. Transitivity (or circularity test) $Q(x^j, x^k) \times Q(x^k, x^l) = Q(x^j, x^l)$
4. Dimensionality $Q(\alpha x^j, \alpha x^k) = Q(x^j, x^k)$.

Homogeneity or the proportionality condition implies a proportional increase to the overall index when an increase occurs to one of the input vectors. Time reversibility guarantees that 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". The transitivity property means that whether a fixed base or a chain of observations is used to calculate the index, the result will be the same. Dimensionality means that if we change the units of measurement for each metric by the same positive number α, then the index remains unchanged. The statistical methods referred to in Section 1 (cluster, principle component, and factor analysis) do not ensure that comparison of observations is based on a best practice curve (benchmark curve or isoquant), or that the resulting index has properties 1 through 4.

In the context of economics, the Malmquist quantity index in Eq. (5) is especially useful to compare productivity changes of production units between two spatial or temporal periods (Caves et al., 1982; Färe et al., 2004). For the purpose of this study, however, we are interested in viewing the time-variability of water quality as compared only to the overall benchmark, or best quality.

In practice, the distances in Eq. (3) are calculated 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 $x \in \Re^N_+$ observed at $t = 1, \ldots, T$ time periods. As discussed further in the next section, we utilize six nutrient loadings as water quality characteristics that are measured at the input of the Mississippi and Atchafalaya Rivers to the mouth of the Gulf of Mexico. We used monthly data from 1981 to 2013. Our data is represented generally as

$$x^t, t = 1, \ldots, T, x^t \in \Re^N_+. \tag{6}$$

We specify the benchmark using all observations as

$$\left\{\begin{array}{ll} x = (x_1, \ldots, x_N) : & \sum_{t=1}^{T} z^t x^t_n \le x_n n = 1, \ldots, N, \\ & \sum_{t=1}^{T} z^t = 1 \\ & z^t \ge 0, t = 1, \ldots, T \end{array}\right\} \tag{7}$$

where $z^t \ge 0$ 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), as shown in Fig. 1. In our context, the benchmark is the best observed water quality as defined from the lower boundary of the input data set. Note that the benchmark is computed from the data itself, and does not require user input regarding the specification of metric weights. In contrast to statistical methods, this approach requires no assumptions about distributions of data or the functional form of the relationships among variables. However, we do assume that there is no measurement error. The primary disadvantage of this assumption is that an outlying observation will shift the position of the whole benchmark surface. It is easy to find observations that cause this problem, since the outlier will be the only observation that determines the benchmark in its neighboring space. To calculate the water quality index (WQI), equivalent to Shephard's distance function in Eq. (3), we compute

$$\begin{array}{ll} & (\mathrm{WQI}(t'))^{-1} = (D(x^{t'}))^{-1} = \min\lambda \\ \text{s.t.} & \sum_{t=1}^{T} z^t x^t_n \le \lambda x^{t'}_n n = 1, \ldots, N \\ & \sum_{t=1}^{T} z^t = 1 \\ & z^t \ge 0, t = 1, \ldots, T \end{array} \tag{8}$$

for each observation. The WQI has a lower limit of one representing the best observed quality. Larger values of WQI represent poorer water quality. For some applications, Malmquist-type quantity indexes can be calculated using Eqs. (8) and (5) to compare two indicator values. However, for this study we will analyze the time variability of the WQI. To validate that the index is representative of its underlying constituents and their variations, it will be used in a statistical regression to describe the variation of the hypoxic region in the northern Gulf of Mexico.

3. Data

To construct the WQI, monthly nutrient loadings as measured from the Mississippi and Atchafalaya Rivers at the mouth of the Gulf of Mexico were obtained from the USGS (Aulenbach et al., 2007), available on line at http://toxics.usgs.gov/hypoxia/mississippi/nutrient_flux_yield_est.html (accessed August 8, 2013). The six constituents, measured monthly from 1981 to 2013 and reported by the USGS, included dissolved nitrite plus nitrate, total organic nitrogen plus ammonia nitrogen (total Kjeldahl nitrogen), dissolved ammonia, total phosphorous, dissolved orthophosphate, and dissolved silica loadings. The loadings were estimated from measured data using the adjusted maximum likelihood estimation (AMLE) method and the LOADEST program. Monthly data from October 2012 to May 2013 were designated preliminary data because the rating curves to convert stream stage to stream discharge are not finalized until after the watershed year. For more information regarding the nutrient loading measurements or estimation methods used, see Aulenbach et al. (2007) or http://toxics.usgs.gov/hypoxia/mississippi/nutrient_flux_yield_est.html (accessed August 8, 2013).

For validation, the calculated WQI will be analyzed and used in a statistical regression to model annual hypoxic areas in the northern Gulf of Mexico. Measured hypoxic area data were obtained from mid-summer shelf-wide cruises between 1985 and 2013. This excludes the year 1989 when a complete cruise measurement was not obtained. Details of cruise data collection methods as well as the data itself can be found at http://www.gulfhypoxia.net/ (accessed August 8, 2013). The model fit will be compared to the latest biophysical model for hypoxia growth from Scavia et al. (2013). The regression model is fitted using *R* and utilizes the method of ordinary least squares (OLS) to estimate the unknown coefficient in the linear regression.

4. Results

4.1. Nutrient aggregation

Distance functions from economic production theory were used to aggregate multiple constituents of nutrient loadings into a monthly time series index of water quality. The individual nutrient time series and the calculated WQI values are shown in Fig. 2. WQI values of one indicate the best observed quality, and values greater than one represent poorer water quality. Fig. 2 shows that WQI values are proportional to the nutrient loadings, where higher nutrient loadings represent poorer water quality. The WQI exhibits a distinct seasonal variability as emphasized in Fig. 3. The mean monthly values along with single standard deviations are shown. It is clear that the WQI values are lowest, or of best possible quality, in September and highest in April and May. The former is typically when crops are near harvest and the latter when planting and fertilizer use is most frequent. In addition to the seasonal variability, there exists a long-term trend of increased water quality (decreased WQI) from 1981 to 2013, as shown in Fig. 2.

4.2. Index validation

We could find no study where a water quality index had been calculated for the Mississippi and Atchafalya Rivers at their outlets into the Gulf of Mexico for comparison with the calculated WQI. Therefore, our approach to validation was to compare the WQI with measured areas of Gulf hypoxia. The May values of the monthly WQI were correlated with available cruise measurements of hypoxia area from 1985 to 2013 and achieved a Pearson correlation coefficient of 0.353 (p-value = 0.066). The only year removed from the time series was 1989 when measured cruise data were not available. For comparison, a Pearson correlation coefficient of 0.534 (p-value = 0.003) was obtained between hypoxia area and the total sum of all nitrogen components, including dissolved nitrite plus nitrate, total organic nitrogen plus ammonia nitrogen (total Kjeldahl nitrogen), and dissolved ammonia. The low, and perhaps statistically insignificant, correlation between the WQI and the measurements of hypoxia area does not invalidate the index; the WQI is still able to describe the aggregated variability of its underlying constituents. But it is clear the combined influence of the multiple constituents is less correlated with hypoxia area than nitrogen loadings alone.

To continue our validation of the index, May values of the WQI were used within a statistical linear regression to model the mid-summer area of Gulf hypoxia:

$$\text{Area} = \alpha(\text{WQI} - 1) \tag{9}$$

where Area is the hypoxia area measured in July, WQI is the May water quality index, and α is a determined coefficient. The model essentially sets the area equal to zero when the index is at its minimum, or best quality—equal to 1.

The results of the model fit with y-intercept set to zero are displayed in Table 1. The model achieves a significant $R^2_{\text{adj}} = 0.81$ with p-value = $1.51e-11$. The model residual standard error is 6432 km^2, and the F-statistic is 122.7 on 27° of freedom. The t-value represents the coefficient's estimate divided by its standard errors. Pr(>|t|) shows the probability that the null hypothesis would produce a result greater than the achieved t-value. The estimated coefficient demonstrates a statistically significant t-value of 11.08 with Pr(>|t|) = $1.51e-11$. Therefore, these results validate the fact that the WQI successfully describes the aggregated effect of multiple pollutants in a single, time-varying quantity. These results were compared with a linear regression model of total N loadings alone where the y-intercept was set to zero, and are also shown in Table 1. The N loadings alone model achieves a significant $R^2_{\text{adj}} = 0.89$ with p-value = $1.08e-14$ and the model residual standard error is 4925 km^2. Using the Wilcoxon rank sum test, we found that the residuals between the N loadings model and the WQI model were not significantly different (W = 360, p-value = 0.6084). An analysis of the residuals did not reveal any outliers, which pose a particular problem for construction of the benchmark.

5. Discussion

The WQI presented here provides a unique application of economic production theory to aggregate multiple pollutants into a single WQI that represents their overall characteristics. The relative weights of the index are formulated endogenously, from the data itself. While this is advantageous for index construction, the method is limited by the quality of the data used because the formulation of the benchmark and the resulting index values are based on observational data. If inaccurate data measurements are used in the formulation of the index, the resulting benchmark and index values will surely be misleading. Also, for this study no measurement errors were considered for nutrient loadings or for cruise measurements of Gulf hypoxia. Significant measurement errors could lead to a benchmark that is not characteristic of the surrounding environment, and would cause index values that are not representative of the actual water quality of the system. Note that if the WQI is to be compared with other water quality measurements, perhaps at a different location or time not included in the analysis, the benchmark will need to be recalculated before an appropriate comparison can be made. This need for recalculation is no different than that is required by other indexes, for example, the Consumer Price Index (CPI) that is updated monthly (http://www.bls.gov/cpi/; accessed August 7, 2014). Nonetheless, the approach for constructing the WQI from the data itself as opposed to assigning subjective weights to various components is preferable.

Validation for a WQI requires comparison with phenomena that the index is meant to quantify or represent. We believe the rather low correlation (R = 0.353, p-value = 0.066) between the WQI values and the measured Gulf hypoxia area does not invalidate its ability to aggregate multiple constituents into a single value.

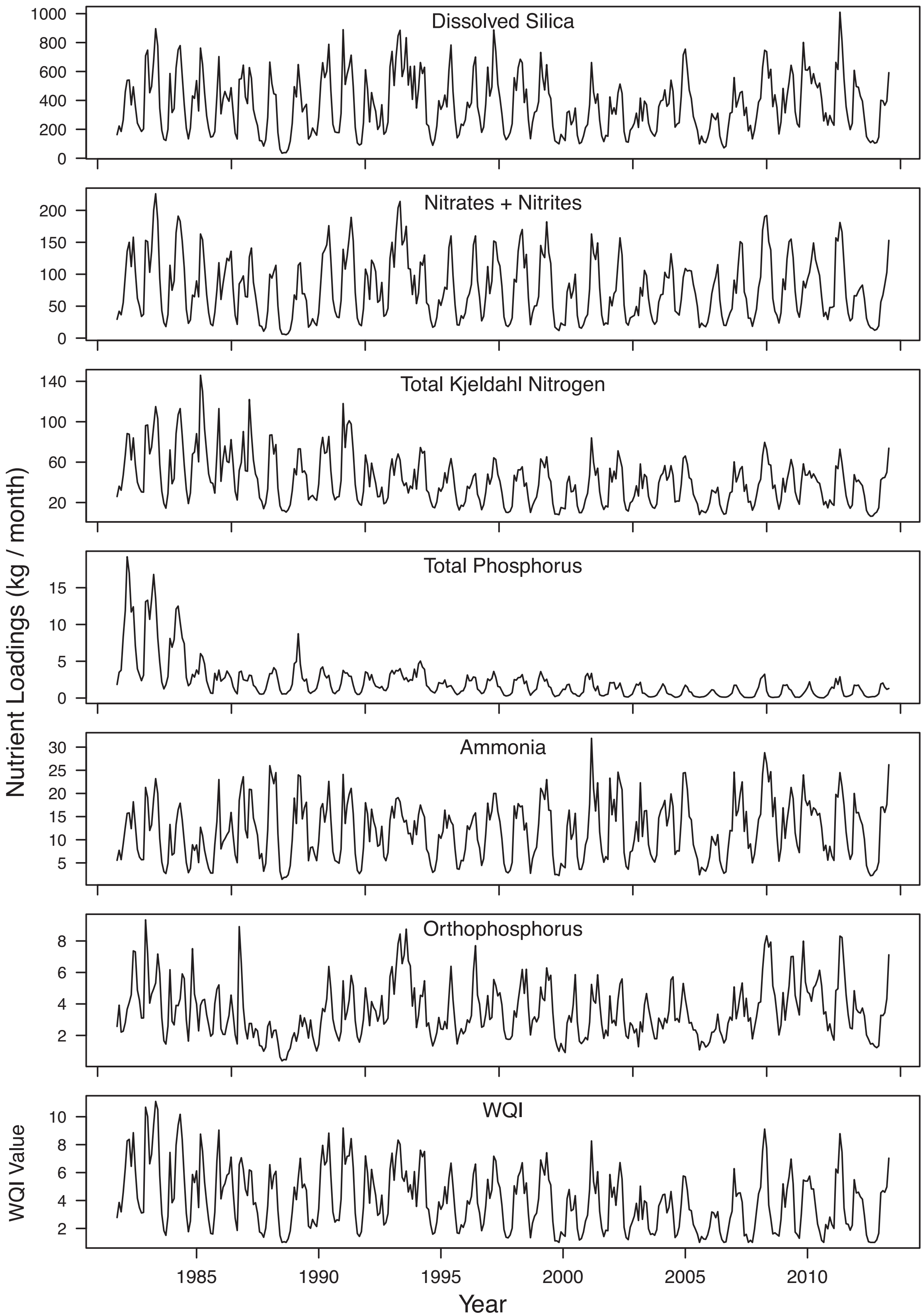

Fig. 2. Total monthly nutrient flux from Mississippi and Atchafalaya Rivers entering the Gulf of Mexico from 1981 to 2013. Also shown is the water quality index that was constructed using DEA and all six nutrient loadings. Note that larger values of WQI indicate poorer quality and correspond to increases in nutrient levels.

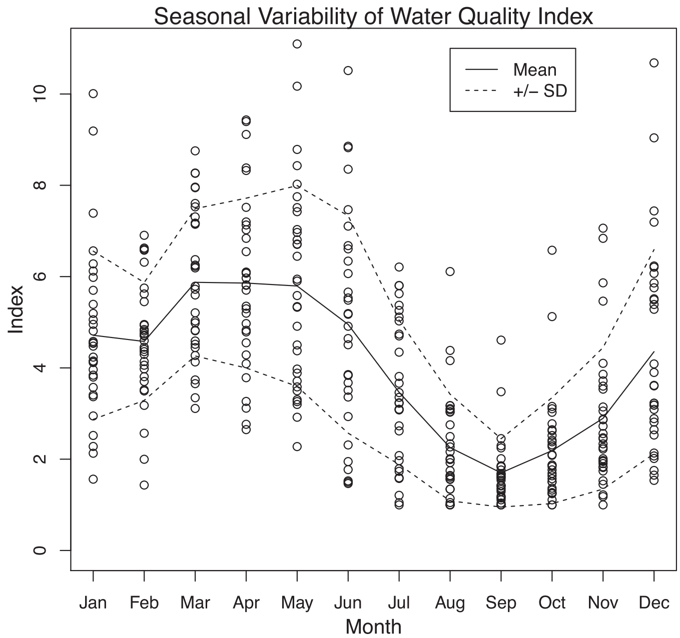

Fig. 3. Seasonality of the WQI. Lines shown are the mean and standard deviations for each month. Notice the lowest values (best water quality) occur in September while the highest values (poorest quality) are in April and May.

Table 1
Linear regression model results.

Model results	Multiple R^2	Adjusted R^2	Resid. std. error	p-value
WQI model	0.820	0.813	6432	1.51e – 11
Total N model	0.894	0.890	4925	1.08e – 14

Coefficient	Estimate	Std. error	t-value	Pr(>|t|)
WQI model	2785	251.4	11.08	1.51e – 11
Total N model	0.072	0.005	15.11	1.08e – 14

Rather, it suggests that these components do not solely describe the variation of the annual area of Gulf hypoxia.

The linear regression model using WQI values described 81% of the interannual variability of Gulf hypoxia. Also, the residual standard errors for both the N loadings alone and the WQI models were near or above 5000 km^2. This result suggests that both models cannot reliably predict that the task force goal of reducing Gulf hypoxia area to less than 5000 km^2 can be reached even if all nutrients are removed from the Gulf of Mexico. This does not imply that the hypoxia goal can never be met; instead, it reiterates the views of present literature that mechanistic models are needed to more accurately simulate the fate and transport of nutrients from a variety of sources to the Gulf in order to determine their overall effect on hypoxia area levels near the hypoxia goal (Rabotyagov et al., 2014a; Whittaker et al., 2015).

For comparison, Scavia et al. (2013) used a biophysical model to model and forecast annual Gulf hypoxia area and volume. Their biophysical model used an adapted Streeter-Phelps river model that simulates the effect of organic matter point sources on oxygen concentrations downstream (Scavia et al., 2013). They specified three models, utilizing only nitrogen loadings, where model calibration coefficients were either treated as annually varying, constant across all years, or varying only for 'storm years' and 'wind years' (Scavia et al., 2013). These models gave fits of $R^2 = 1.00$, $R^2 = 0.50$, and $R^2 = 0.69$, respectively, where the first model was noted as overparameterized (Scavia et al., 2013). Model comparisons with null results using the Akaike information criterion (AIC), where smaller values indicates a better fit, gives values of 6.68 and 22.29 for the WQI regression and third biophysical model, respectively. The large difference is most likely because the AIC penalizes models with larger number of parameters.

Note that the WQI linear regression model used all available cruise data and did not account for wind or storm years. This is contrary to all current hypoxia models available today that utilize some mechanism in order to account for wind velocities, whether by specific inclusion in models (Forrest et al., 2011; Scavia et al., 2013) or by removing particular years from analysis (Evans and Scavia, 2011; Greene et al., 2009; Scavia and Donnelly, 2007; Scavia et al., 2003; Turner et al., 2006). We purposefully refrained from utilizing only recent measurements or excluding annual measurements based on conditions including high wind velocities, proximity to hurricane occurrences, or atypical area measurements. Even without these modifications, the main conclusion remains clear that the WQI captures the intrinsic characteristics of its included constituents, and describes a large percentage of hypoxia interannual variability.

This work does not conclude that the Malmquist WQI is a superior method for hindcasting or forecasting the annual variation of Gulf hypoxia. Instead, we simply state that the index accurately represents the aggregated characteristics and variations of its underlying constituents. Further research should be conducted to use the WQI (perhaps with different combinations of constituents) to improve current hypoxia models. Temperature and wind velocity could be included in multivariable regressions along with WQI, or perhaps included within the WQI itself. The WQI values could also be used to compare the spatial and temporal variations of water quality between different agricultural fields or land management practices.

It is an important characteristic of the Malmquist index approach to constructing a water quality index that no human intervention is required. Some new tools for watershed management use genetic algorithms, where nutrient loading or water quality is evaluated at each step in the search for optima (Arabi et al., 2006; Rabotyagov et al., 2014b). Our particular interest is optimization of agri-environmental policy, where optimal parameters for targeting environmental policy are found using a hybrid genetic algorithm (Whittaker et al., 2009). The Malmquist WQI is ideal for this approach, since it has to be evaluated at each generation of the genetic algorithm. Therefore, future research should also use the Malmquist WQI to specify maximization of water quality as one of multiple objectives.

References

Aguiar, F.C., Segurado, P., Urbanič, G., Cambra, J., Chauvin, C., Ciadamidaro, S., Dörflinger, G., 2014. Comparability of river quality assessment using macrophytes: a multi-step procedure to overcome biogeographical differences. Sci. Total Environ. 476, 757–767.
Arabi, M., Govindaraju, R.S., Hantush, M.M., 2006. Cost-effective allocation of watershed management practices using a genetic algorithm. Water Resour. Res. 42, W10429.
Aulenbach, B.T., Buxton, H.T., Battaglin, W.A., and Coupe, R.H, 2007. Streamflow and nutrient fluxes of the Mississippi–Atchafalaya River Basin and subbasins for the period of record through 2005. U.S. Geological Survey Open-File Report 2007-1080. http://toxics.usgs.gov/pubs/of-2007-1080/index.html.
Beamonte Córdoba, E., Casino Martínez, A., Veres Ferrer, E., 2010. Water quality indicators: comparison of a probabilistic index and a general quality index. The case of the Confederación Hidrográfica del Jácar (Spain). Ecol. Indic. 10, 1049–1054.
Bellenger, M.J., Herlihy, A.T., 2009. An economic approach to environmental indices. Ecol. Econ. 68, 2216–2223.
Bellenger, M.J., Herlihy, A.T., 2010. Performance-based environmental index weights: are all metrics created equal? Ecol. Econ. 69, 1043–1050.
Bierman, P., Lewis, M., Ostendorf, B., Tanner, J., 2011. A review of methods for analysing spatial and temporal patterns in coastal water quality. Ecol. Indic. 11, 103–114.
Blanchet, H., Lavesque, N., Ruellet, T., Dauvin, J.C., Sauriau, P.G., Desroy, N., Desclaux, C., Leconte, M., Bachelet, G., Janson, A.L., Bessineton, C., Duhamel, S., Jourde, J., Mayot, S., Simon, S., de Montaudouin, X., 2008. Use of biotic indices in semi-enclosed coastal ecosystems and transitional waters habitats—implications for

the implementation of the European Water Framework Directive. Ecol. Indic. 8, 360–372.
Blocksom, K.A., Johnson, B.R., 2009. Development of a regional macroinvertebrate index for large river bioassessment. Ecol. Indic. 9, 313–328.
Castoldi, N., Bechini, L., Stein, A., 2009. Evaluation of the spatial uncertainty of agro-ecological assessments at the regional scale: the phosphorus indicator in northern Italy. Ecol. Indic. 9, 902–912.
Caves, D.W., Christensen, L.R., Diewert, W.E., 1982. The economic-theory of index numbers and the measurement of input, output, and productivity. Econometrica 50, 1393–1414.
Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. Eur. J. Oper. Res. 2, 429–444.
Correll, D.L., 1998. The role of phosphorus in the eutrophication of receiving waters: a review. J. Environ. Qual. 27 (2), 261–266.
Diewert, W.E., Nakamura, A., 1993. Essays in Index Number Theory: Volume 1. Emerald Group Publishing, Bingley, UK.
Diewert, W.E., 1987. Index numbers. In: Eatwell, J., Milgate, M., Newman, P. (Eds.), The New Palgrave: A Dictionary of Economics. The Macmillan Press, London, pp. 767–780.
Diewert, W.E., 1992. Fisher ideal output, input, and productivity indexes revisited. J. Prod. Anal. 3, 211.
Evans, M.A., Scavia, D., 2011. Forecasting hypoxia in the Chesapeake Bay and Gulf of Mexico: model accuracy, precision, and sensitivity to ecosystem change. Environ. Res. Lett. 6.
Färe, R., Grosskopf, S., Hernandez-Sancho, F., 2004. Environmental performance: an index number approach. Resour. Energy Econ. 26, 343–352.
Feio, M.J., Almeida, S.F.P., Craveiro, S.C., Calado, A.J., 2009. A comparison between biotic indices and predictive models in stream water quality assessment based on benthic diatom communities. Ecol. Indic. 9, 497–507.
Fisher, I., 1922. The Making of Index Numbers. Houghton Mifflin, Boston.
Forrest, D.R., Hetland, R.D., DiMarco, S.F., 2011. Multivariable statistical regression models of the areal extent of hypoxia over the Texas-Louisiana continental shelf. Environ. Res. Lett. 6.
Ghosh, S., Mujumdar, P.P., 2010. Fuzzy waste load allocation model: a multiobjective approach. J. Hydroinformatics 12, 83–96.
Greene, R.M., Lehrter, J.C., Hagy, J.D., 2009. Multiple regression models for hindcasting and forecasting midsummer hypoxia in the Gulf of Mexico. Ecol. Appl. 19, 1161–1175.
Hargiss, C.L.M., DeKeyser, E.S., Kirby, D.R., Ell, M.J., 2008. Regional assessment of wetland plant communities using the index of plant community integrity. Ecol. Indic. 8, 303–307.
Howarth, R., Chan, F., Conley, D.J., Garnier, J., Doney, S.C., Marino, R., Billen, G., 2011. Coupled biogeochemical cycles: eutrophication and hypoxia in temperate estuaries and coastal marine ecosystems. Front. Ecol. Environ. 9 (1), 18–26.
Jordan, S.J., Lewis, M.A., Harwell, L.M., Goodman, L.R., 2010. Summer fish communities in northern Gulf of Mexico estuaries: indices of ecological condition. Ecol. Indic. 10, 504–515.
Justus, B.G., Petersen, J.C., Femmer, S.R., Davis, J.V., Wallace, J.E., 2010. A comparison of algal, macroinvertebrate, and fish assemblage indices for assessing low-level nutrient enrichment in wadeable Ozark streams. Ecol. Indic. 10, 627–638.
Kane, D.D., Gordon, S.I., Munawar, M., Charlton, M.N., Culver, D.A., 2009. The Planktonic Index of Biotic Integrity (P-IBI): an approach for assessing lake ecosystem health. Ecol. Indic. 9, 1234–1247.
Kanno, Y., Vokoun, J.C., Beauchene, M., 2010. Development of dual fish multi-metric indices of biological condition for streams with characteristic thermal gradients and low species richness. Ecol. Indic. 10, 565–571.
Kantoussan, J., Marc Ecoutin, J., Fontenelle, G., Morais, L.T.d., Laë, R., 2010. Ecological indicators based on trophic spectrum as a tool to assess ecosystems fishing impacts. Ecol. Indic. 10, 927–934.
Khalil, B., Ouarda, T.B.M.J., St-Hilaire, A., Chebana, F., 2010. A statistical approach for the rationalization of water quality indicators in surface water quality monitoring networks. J. Hydrol. 386, 173–185.
Leunda, P.M., Oscoz, J., Miranda, R., Ariño, A.H., 2009. Longitudinal and seasonal variation of the benthic macroinvertebrate community and biotic indices in an undisturbed Pyrenean river. Ecol. Indic. 9, 52–63.
Liou, S.M., Lo, S.L., Wang, S.H., 2004. A generalized water quality index for Taiwan. Environ. Monit. Assess. 96, 35–52.
Lowe, J., 1823. The Present State of England in Regard to Agriculture, Trade and Finance, 2nd ed. Longman, Hurst, Rees, Orme and Brown, London.
Malmquist, S., 1953. Index numbers and indifference surfaces. Trabajos Estadistica 4, 209–242.
Marchini, A., Facchinetti, T., Mistri, M., 2009. F-IND: a framework to design fuzzy indices of environmental conditions. Ecol. Indic. 9, 485–496.
Navarro-Llácer, C., Baeza, D., de las Heras, J., 2010. Assessment of regulated rivers with indices based on macroinvertebrates, fish and riparian forest in the southeast of Spain. Ecol. Indic. 10, 935–942.
Obenour, D.R., Michalak, A.M., Zhou, Y.T., Scavia, D., 2012. Quantifying the impacts of stratification and nutrient loading on hypoxia in the northern Gulf of Mexico. Environ. Sci. Technol. 46, 5489–5496.
Ou, Y., Wang, X., Li, M., 2010. Relating watershed characteristics to non-point source pollution in miyun reservoir watershed using principal components analysis (PCA), Chengdu.
Pierson, N.G., 1895. Further considerations on index-numbers. Econ. J. 6, 127–131.
Pinto, R., Patrício, J., Baeta, A., Fath, B.D., Neto, J.M., Marques, J.C., 2009. Review and evaluation of estuarine biotic indices to assess benthic condition. Ecol. Indic. 9, 1–25.
Rabalais, N.N., Turner, R.E., Wiseman, W.J., 2001. Hypoxia in the Gulf of Mexico. J. Environ. Qual. 30, 320–329.
Rabalais, N.N., Turner, R.E., Sen Gupta, B.K., Boesch, D.F., Chapman, P., Murrell, M.C., 2007. Hypoxia in the northern Gulf of Mexico: does the science support the plan to reduce, mitigate, and control hypoxia? Estuar. Coast 30, 753–772.
Rabotyagov, S.S., Kling, C.L., Gassman, P.W., Rabalais, N.N., Turner, R.E., 2014a. The economics of dead zones: causes, impacts, policy challenges, and a model of the Gulf of Mexico hypoxic zone. Rev. Environ. Econ. Policy 8 (1), 58–79.
Rabotyagov, S.S., Valcu, A.M., Kling, C.L., 2014b. Reversing property rights: practice-based approaches for controlling agricultural nonpoint-source water pollution when emissions aggregate nonlinearly. Am. J. Agric. Econ. 96, 397–419.
Scavia, D., Donnelly, K.A., 2007. Reassessing hypoxia forecasts for the Gulf of Mexico. Environ. Sci. Technol. 41, 8111–8117.
Scavia, D., Rabalais, N.N., Turner, R.E., Justic, D., Wiseman, W.J., 2003. Predicting the response of Gulf of Mexico hypoxia to variations in Mississippi River nitrogen load. Limnol. Oceanogr. 48, 951–956.
Scavia, D., Justic, D., Bierman, V.J., 2004. Reducing hypoxia in the Gulf of Mexico: advice from three models. Estuaries 27, 419–425.
Scavia, D., Evans, M.A., Obenour, D.R., 2013. A scenario and forecast model for Gulf of Mexico hypoxic area and volume. Environ. Sci. Technol. 47, 10423–10428.
Seilheimer, T.S., Mahoney, T.P., Chow-Fraser, P., 2009. Comparative study of ecological indices for assessing human-induced disturbance in coastal wetlands of the Laurentian Great Lakes. Ecol. Indic. 9, 81–91.
Shephard, R.W., 1953. Cost and Production Functions. Princeton University Press, Princeton, New Jersey.
Shephard, R.W., 1970. Theory of Cost and Production Functions. Princeton University Press, Princeton.
Shiels, D.R., 2010. 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. Ecol. Indic. 10, 1102–1110.
Simaika, J.P., Samways, M.J., 2011. Comparative assessment of indices of freshwater habitat conditions using different invertebrate taxon sets. Ecol. Indic. 11 (2), 370–378.
Sin, G., Gernaey, K.V., Neumann, M.B., van Loosdrecht, M.C.M., Gujer, W., 2009. Uncertainty analysis in WWTP model applications: a critical discussion using an example from design. Water Res. 43, 2894–2906.
Stoddard, J.L., Herlihy, A.T., Peck, D.V., Hughes, R.M., Whittier, T.R., Tarquinio, E., 2008. A process for creating multimetric indices for large-scale aquatic surveys. J. North Am. Benthol. Soc. 27, 878–891.
Styers, D.M., Chappelka, A.H., Marzen, L.J., Somers, G.L., 2010. Scale matters: indicators of ecological health along the urban-rural interface near Columbus, Georgia. Ecol. Indic. 10, 224–233.
Taheriyoun, M., Karamouz, M., Baghvand, A., 2010. Development of an entropy-based fuzzy eutrophication index for reservoir water quality evaluation. Iran. J. Environ. Health Sci. Eng. 7, 1–14.
Tran, L.T., O'Neill, R.V., Smith, E.R., Wagner, P.F., Mehaffey, M., 2008. Watershed-based self- and peer-appraisal indices for integrated environmental assessment with a case study of the Mid-Atlantic region. Ecol. Indic. 8, 308–315.
Tran, C.P., Bode, R.W., Smith, A.J., Kleppel, G.S., 2010. Land-use proximity as a basis for assessing stream water quality in New York State (USA). Ecol. Indic. 10, 727–733.
Turner, R.E., Rabalais, N.N., Justic, D., 2006. Predicting summer hypoxia in the northern Gulf of Mexico: riverine N, P, and Si loading. Mar. Pollut. Bull. 52, 139–148.
Tyteca, D., 1996. On the measurement of the environmental performance of firms – a literature review and a productive efficiency perspective. J. Environ. Manag. 46, 281–308.
von Neumann, J., 1937. A Model of General Economic Equilibrium. In: Menger, K. (Ed.), Ergebnisse eines mathematischen Kolloquiums, 1935–36. Translated and reprinted in 1945 in Review of Economic Studies, vol. 13 (1), pp. 1–9.
Whittaker, G., Confesor Jr., R., Griffith, S.M., Färe, R., Grosskopf, S., Steiner, J.J., Mueller-Warrant, G.W., Banowetz, G.W., 2009. A hybrid genetic algorithm for multiobjective problems with activity analysis-based local search. Eur. J. Oper. Res. 193, 195–203.
Whittaker, G., Barnhart, B.L., Srinivasan, R., Arnold, J.G., 2015. Cost of areal reduction of Gulf hypoxia through agricultural practice. Sci. Total Environ. 505, 149–153.
Wiseman, W.J., Rabalais, N.N., Turner, R.E., Dinnel, S.P., MacNaughton, A., 1997. Seasonal and interannual variability within the Louisiana coastal current: stratification and hypoxia. J. Mar. Syst. 12, 237–248.
Zhou, P., Ang, B.W., Poh, K.L., 2007. A mathematical programming approach to constructing composite indicators. Ecol. Econ. 62, 291–297.
Zhou, P., Ang, B.W., Poh, K.L., 2008. Measuring environmental performance under different environmental DEA technologies. Energy Econ. 30, 1–14.