

Biological taxonomic problem solving using fuzzy decision-making analytical tools

Janice L. Pappas*

Museum of Zoology, University of Michigan, 1109 Geddes Avenue, Ann Arbor, MI 48109-1079, USA

Received 9 December 2004; received in revised form 19 December 2005; accepted 10 January 2006

Available online 20 February 2006

Abstract

Biological taxonomy is at the heart of species identifications. Such identifications are instrumental in biodiversity studies, ecological assessment, and phylogenetic analysis, among other studies. Fuzzy measures and classification integration was used to analyze shape groups of the diatom *Asterionella* using fuzzy Fourier shape coefficients and fuzzy morphometric measures. Based on this analysis, six shape groups were determined with specimen membership assignments at or exceeding the crossover point (0.5). Fuzzy average overlap values were approximately at or just over the crossover point, indicating similarity in developmental stages. In further analysis, spatial and temporal data from specimen samples were used in conjunction with fuzzy membership assignment values. Spatial and temporal variables were ranked and fuzzified based on the mode. The modes were then weighted by degree of importance as determined by an expert in diatom research. The weighted fuzzy modes for each specimen in each shape group were aggregated as a weighted sum. The normalized relative cardinality for each specimen defined the degree of suitability that a specimen belonged to a shape group, and the expert evaluated the result. While morphological data specifies inheritance (shape) and development (morphometry), spatial and temporal data were proxies for reproductive isolation. These biological principles constrained and defined the direction of analysis and defined each shape group as a species to the degree specified by each specimen. This fuzzy decision-making process provided a simple way to aggregate scant available data and a linguistic solution in a taxonomic study understandable to a biologist.

© 2006 Elsevier B.V. All rights reserved.

Keywords: *Asterionella*; Biology; Classification integration; Diatoms; Expert systems; Information fusion; Multiple criteria evaluation; Systematics and taxonomy

1. Introduction

In biological classification systems, prior to determining phylogenetic relationships, it is necessary to identify each organism as having a sufficient number of differences for comparison purposes to other organisms that are similar on a species or genus level. Being confident of the identity of an organism has important implications for researchers in a wide variety of studies. One group of such organisms is the diatoms, and they are used as bioindicators of environmental conditions such as lake acidification (e.g., [1]), eutrophication [11], hydrologic and climatic change in lakes (e.g., [8]), and in oceans (e.g., [23]). These studies and others are of interest not only to ecologists, but also to those studying biodiversity.

* Tel.: +1 734 764 0456; fax: +1 734 763 4080.

E-mail address: jpgappas@umich.edu (J.L. Pappas).

One of the more difficult tasks is identification of such organisms that have few distinguishing morphological features. Diatoms have siliceous frustules that fit together like a Petri dish, and they have many shapes and patterns. They range in size from 3 to 300 μm and are found worldwide. They reproduce sexually, but the meiotic life cycle is unknown for all but a few diatoms [22]. To complicate matters, mitotic division in diatoms results in size diminution and slight shape distortions occur [16]. Sometimes, the identification of an ontogenetic (developmental) life cycle stage of one diatom is easily mistaken to be a different species. Identification of a given diatom is sometimes difficult and encompasses some degree of uncertainty.

Traditionally, diatom classification has been largely based on frustule morphology. Most often, diatoms are viewed in fixed mounts on glass slides using a compound light microscope. Other forms of microscopy have been used to augment morphological studies through studies of ultrastructure. Empirical data and measurement constitutes a large portion of the evidence gathered when trying to determine the taxonomic status of a given diatom. As a result, a scientific name of a diatom is a hypothesis to be tested. The species-naming process by a taxonomist occurs with reference to previous work. Vague or incomplete historical documentation, scant or missing data, and qualitative or descriptive attributes are some of the kinds of information that may be available to the taxonomist. Decision-making necessarily includes uncertainties in relation to the available information at a given time.

At the outset, initial sorting and binning of diatom specimens is largely based on shape. Diatom shape is an inherited property [15,16] and is considered to be a standard starting point in morphological studies. Shape descriptors may be qualitative, but quantitative shape analysis has been used as a tool to determine the range of shape variation in a given species or for comparisons of a taxon with type specimens.

2. Background on *Asterionella*, shape groups, and taxonomic sorting

What do we know from the historical record of species identification of *Asterionella*? The record is scant at best [17,20]. If it exists, the biologist usually refers to the type specimen for each named species. In this case, only the neotype of *Asterionella formosa* exists, was named by Körner [14], and is housed at the Botanical Gardens and Museum, University of Berlin, and the taxon was originally described by Hassall [12]. No photographs exist, and contemporaries of Hassall as well as some 20th century researchers relied on Hassall's description without examining specimens themselves. A few early drawings exist of specimen outlines only. Because of the lack of scrutiny of early work and few subsequent studies detailing morphology of specimens, researchers in aquatic ecology, limnology, and paleolimnology assign the neotype's name to all specimens that resemble *Asterionella*. However, some researchers have questioned the validity of species identifications of *Asterionella*, including Stoermer and Yang [25] and Stoermer and Ladewski [24]. Fuzzy biological taxonomic studies are a recent occurrence [17,21], although fuzzy multivariate analysis has been used with regard to morphological variation and ecological gradients [19].

In previous studies [17,21], Fourier shape analysis of specimen outline was used on taxa from the araphid pennate diatom genus *Asterionella* [18]. This genus has few morphological features, including striae, a labiate process, and a sternum without a raphe. All species in this diatom genus have a bilobate shape (Fig. 1). For photographic images of all the specimens used, see [17].

The multivariate statistical technique, principal components analysis (PCA), was used to ordinate fuzzified Fourier coefficients defining each *Asterionella* shape outline. Fuzzification was accomplished by mean-correcting the coefficients. From the first PCA of shape variance for each specimen, the resultant shape gradient in fuzzy PCA shape space was subsectioned into seven groups. Circumscribed specimens were designated as shape groups, and overlap of adjacent shape groups occurred to some degree. Since diatom shape is largely inherited, each shape group is, to some degree, an indication of a distinct species. This initial sorting technique is used to reduce some of the arbitrariness in taxonomic decision-making.

A second sorting of specimens within shape groups is then necessary. Sometimes, there is overlap between shape groups, as there was in this case. Morphometric measures other than shape were used for this next sorting. Measurable quantities common to all *Asterionella* are length, head pole width, mid-valve width, and foot pole width (Fig. 1). (see [17] for more details). Combinations of these measures as ratios can be used as a numerical way to distinguish specimens that look approximately alike. Two fuzzy measures were created using classification integration [26,30]. They were used to determine degree of overlap and degree that specimens belonged to their assigned shape group [17].

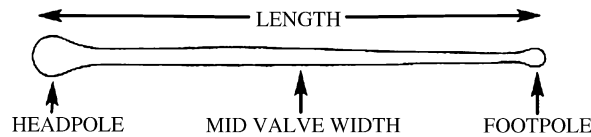


Fig. 1. Generalized *Asterionella* bilobate shape with identifiers used in morphometric analysis.

Degree of overlap was quantified by a fuzzy measure. However, does the overlapped region represent similarity between two separate species (two separate shape groups), or as a lower level taxonomic designation, such as similar subspecies from the same species? Despite our results, we still cannot say to what degree each specimen is which species. This is in large part because of the way in which diatoms vegetatively reproduce, viz., they undergo size diminution, complicating the morphological differences in shape among some species.

Morphological data is usually collected with regard to dead, fixed specimens. Such data are compared to ranges of values that are dictated by the fact that diatoms diminish in size as they vegetatively reproduce. During sexual reproduction, diatoms may undergo auxosporulation and produce an initial cell that is morphologically somewhat unlike its subsequently asexually produced cell. Specimens are obtained from samples collected at a particular depth in the water column and along a transect delineated over the surface of the body of water. Locally, specimens from the same sample spatially occur within a narrow range of water volume. Globally, specimens in different samples are quite distant from each other, given their size on the scale of micrometers, occurring in a body of water such as Lake Michigan, at a volume of approximately 5000 km³. What makes the taxonomic problem suitable for further fuzzy analysis involves the availability of data about the samples from which the specimens were obtained.

Although there are different definitions as to what is meant by a “species”, one way to differentiate species is to determine how spatial proximity plays a role with regard to reproductive isolation. The implication is that if two specimens are “far enough” away from each other, they have less chance of combining their genetic material, thereby signifying the potential of being (or becoming) separate species. Spatial dimension of reproduction is a dynamic process, and in the long run, an everchanging one that is a part of natural selection. From an environmental perspective, if two specimens were collected at different times of the year, then there is the potential that each specimen has a different tolerance to say, temperature, and are therefore, potentially, different species. This data will comprise the basis of the current study.

3. Synopsis on previous studies using classification integration and fuzzy measures

To reiterate, in previous studies [17,21], classification integration and fuzzy measures were used to determine two types of membership for *Asterionella* specimens. The first way classification integration and fuzzy measures were used was in measuring degree of overlap between shape groups, or to what degree members from each of two adjacent shape groups contribute to overlap. The second way classification integration and fuzzy measures were used was in measuring degree of membership for each specimen in its assigned shape group. Fuzzy measures were based on ratios using head pole width, foot pole width, and mid-valve width or a combination to form membership functions on the interval [0, 1] that were ordered. Widths of head pole, foot pole, and mid-valve region were interdependent with respect to shape of *Asterionella*.

The crossover rule of 0.5 was used to determine degree of acceptability. That is, unless degree of overlap equaled or exceeded 0.5, *Asterionella* shape groups were considered to be separate species to the degree the fuzzy classification indicated. In addition, if each specimen’s fuzzy classification indicated a value of 0.5 or greater, then that specimen remained in its assigned shape group.

3.1. Fuzzy measure overlap

Classification integration [30] was accomplished using Sugeno’s integral [26] to evaluate degree of agreement between partial evidence and the level of importance of that evidence to determine fuzzy average overlap between pairs of shape groups. For (X, \mathcal{F}) , a measurable space, where $A = X \in \mathcal{F}$, $g : \mathcal{F} \rightarrow [0, \infty]$ is a fuzzy measure, let

$h : X \rightarrow [0, 1]$ be a membership function where X is finite, then

$$E = \int_A h(x) \circ g(\bullet), \tag{1}$$

where $h(x)$, a measurable function of partial evidence and $g(\bullet)$, an importance measure. To obtain the solution, E the evaluation of a composition of functions [5,26,30] is calculated as

$$E = \sup_{a \in [0,1]} [a \wedge g(A \cap F_a)] \tag{2}$$

where $\{x_i, \dots, x_n\}$ defines the i th to n th attributes head pole width, foot pole width, mid-valve width or some combination [17], \vee and \wedge are the maximum and minimum operators, respectively [5,30], $F_{a^x} = \{x | h(x) \geq a\}$, where A is a fuzzy set of $h(x)$ with respect to g for all $a \in [0, 1]$, $A = X$, \cap is the intersection, and sup is the supremum or least upper bound [26]. When g is a possibility measure, Eq. (2) (see [5])

$$E = \bigvee_{i=1}^n [h(x_i)g \wedge (x_i)]. \tag{3}$$

Let μ_t be a membership function equal to the ratio of mid-valve width to head pole width. For each shape group, S_i , partial evidence about x_j is

$$h_t(x_j) = \frac{\sum_{j=1}^m \mu_t(x_j)}{m}, \tag{4}$$

where m is the total number of elements in a shape group, and $h(x_j)$ is a fuzzy average [27]. For combinations of shape groups, S_i and S_l , with membership functions $\mu_t(x_j)$ and $\mu_t(x_k)$, respectively, for all $x_j \in S_i$, $x_k \in S_l$:

$$h_t(x_{j,k}) = \frac{1}{m+q} \left\{ \sum_{j=1}^m \mu_t(x_j) + \sum_{k=1}^q \mu_t(x_k) \right\}, \tag{5}$$

where m is the total number of x_j elements in the i th shape group, S_i , q is the total number of x_k elements in the l th shape group, S_l , and the result was a fuzzy average overlap [7,27] as partial evidence, $h(x_{j,k})$, for a combination of shape groups with commutativity [30].

The importance measure, $g(\bullet)$, is calculated as a possibility measure [30]. For each shape group, this measure is based on the ratio of foot pole width to head pole width. For finite X , let $S : X \rightarrow [0, 1]$ be a membership function on measurable space $(X, P(X))$ where $P(X)$ is the power set of subsets (shape groups) and \prod is a possibility measure. Then, $\prod(\emptyset) = 0$ and $\prod(X) = 1$. For any S_i of X ,

$$\prod \left(\bigcup_{i=1} S_i \right) = \sup_i \prod (S_i), \tag{6}$$

where sup is the supremum or least upper bound [5,13,30,34], and $g(\bullet) = \sup_i \prod(S_i)$. In addition, for combinations of shape groups, $g(x)$ for all $x_j \in S_i$, $x_k \in S_l$

$$g(x_{j,k}) = \pi(x_{j,k}) = \max_{i,l} \min(\pi_{S_i}(x_j), \pi_{S_l}(x_k)), \tag{7}$$

where S_i and S_l represent different shape groups, and the fuzzy set theoretic operators supremum and minimum are utilized on possibility distributions $\pi_{S_i}(x_j)$ and $\pi_{S_l}(x_k)$ [5]. Properties of fuzzy measures are boundedness, monotonicity [26,27], and subadditivity or lower semicontinuity [30].

Aggregation of partial evidence and importance measures provides the overall evaluation of a given shape group. The value 0.5 was used as a crossover point to judge $h(x)$, $g(x)$, and E [30].

3.2. Fuzzy membership assignment

Classification integration was also used to determine the degree that a given specimen belonged to its assigned group [17] as fuzzy membership assignment. In this case, the importance measure is calculated as Sugeno’s measure,

g_λ -fuzzy measure. For two attributes for each shape group in which each specimen was considered, two measures were calculated. One measure, g_1 , represents ratio of foot pole width to head pole width. The other measure, g_2 , represents ratio of mid-valve width to head pole width.

Sugeno's measure, g_λ as $\mu(x_i)$ [30] is

$$g_\lambda(X) = \mu(X) = \frac{1}{\lambda} \left[\prod_{i=1}^n (1 + \lambda \cdot \mu(\{x_i\})) - 1 \right] \quad (8)$$

for any λ where

$$\lambda \in \left(-\frac{1}{\sup \mu}, \infty \right) \cup \{0\}. \quad (9)$$

If $\lambda > 0$, then g is a belief measure; if $-1 < \lambda < 0$, then g is a plausibility measure; if $\lambda = 0$, then g is a probability measure [5,30].

As a defining factor for each group, a ratio of normalized mid-valve width to foot pole width represents $h(x)$ for each specimen. Partial evidence represents the degree of consistency that a particular specimen belongs to a particular group. Values equal to one indicate complete consistency that a specimen is assigned to the correct group. Values equal to zero indicated that the specimen did not belong to that group. The crossover point was 0.5. For each specimen, fuzzy membership assignment was 0.5 or greater, and g was calculated to be a plausibility measure.

3.3. Fuzzy morphological taxonomic decision-making

Overlap values were calculated between groups within a region of similar total variance from PCA. Two regions were defined—one with shape groups I, II and III, and the other one with shape groups IV, V and VI. Generally, but not exclusively speaking, each region defined two sets of specimens differentiated by geographic location. The first one (shape groups I, II and III) included most of the specimens from Lakes Superior and Huron, Green Bay and other areas of northern Lake Michigan. There were a few specimens from central and southern Lake Michigan as well. In the second region (shape groups IV, V and VI), most of the specimens were from all of Lake Michigan (excluding Green Bay), but also included a few specimens from Lakes Superior and Huron [17].

Results from fuzzy average overlap calculations were in the range of 0.53–0.60. The degree of overlap suggests that paired comparisons of shape groups are as much similar as dissimilar (0.53–0.56), or are slightly more similar to each other (0.55–0.60). That is, all fuzzy overlap values are similar. This measure somewhat indirectly embodies the developmental stages of this diatom, in that shape of one species at one developmental stage may look like the shape of another species at a different developmental stage (overlap). Fuzzy average overlap indicates to what degree specimens from adjacent groups are similar and that biological sorting at the species level is not a simple sorting problem [17].

Each specimen's shape group assignment was calculated to be from 0.49 (one case) to 0.73, with a degree of plausibility (Sugeno's measure) from 0.6 to 0.9. A check of each specimen's plausible fuzzy assignment value at approximately the crossover point indicates that specimen shape group assignment is correct to some degree [17].

At any point in a taxonomic analysis, an expert in the field can weigh in on outcomes for species designations. Moreover, with additional evidence, a new fuzzy analysis may be useful. This additional evidence must be gleaned from available historical sources. Such data includes information about sample collection location and when the sample was collected. More specifically, data on latitude and longitude of the sample location, time of year when the sample was collected, and general information about circulation patterns in the bodies of water from which the samples were collected are available with reference to the diatoms used in fuzzy morphological analyses.

4. Purposes of this study

Previous morphometric studies were useful in finding fuzzy group boundaries and their properties (i.e., fuzzy average overlap) and specimen fuzzy membership in groups. The current study focuses on adding new information to what has already been done using techniques from fuzzy decision-making, including information fusion, expert systems, and multiple criteria evaluation. I want to test the fuzzy biological taxonomic shape groups devised with respect to spatial and temporal data and reassess the degree to which specimen assignment in groups is true and also in agreement with

an expert in diatom research. That is, given the nature of the evidence and the uncertainties therein, I want to determine to what degree that each specimen in its assigned shape group may be called a “species”. Results from this study will provide a simple, understandable way to decrease arbitrariness in the taxonomic decision making process for biologists by providing a useful, analytical tool for taxonomists when data is scarce, absent, or indirectly applicable.

5. Methods

Data on sample location as latitude and longitude, long-term circulation patterns in the Great Lakes [2], and the time of year when specimen samples were collected are used. These data constitute four spatial and/or temporal variables. Latitude and longitude are numerical values relating to the position of sample collection and are local variables. Long-term circulation patterns are a composite of data with regard to time of year and position in the Great Lakes. The premise of this data is that circulation patterns dictate whether regions of a given lake will incur water mixing, and thereby a more likely chance that specimens would come in close proximity and interbreed. This is a regional or global variable. Time of year is a monthly designation and considered to be a temporal variable on a local scale. The variables latitude, longitude and time of year pertain to samples from which the specimens were collected have definite numerical or categorical specificity. Long-term circulation patterns represent a composite of many years of data applied to regions of the Great Lakes and are indirectly applicable to the regions where samples were collected. The data are more general and are labeled as cyclonic, coastal, annual, or not applicable and are designated as 0, 1, 2, and 3, respectively. Sample location and time of year collected are the factors that determine which categorical variable is given to that sample.

Each of these variables is characterized as a quintuple, (x, T, U, J, M) , where x is the variable with the name defined as the primary term, T is the set of linguistic values of x as terms, U is the universe of discourse, J is a rule of syntax for generating terms of linguistic values of x , and M is a semantic rule indicating meaning of the linguistic values by assigning a fuzzy set to the terms.

5.1. Ranking variables within categories

For each shape group, spatial and temporal data are ranked based on how frequently particular values appear in each category. That is, the mode or relative maximum is used to represent the value that approximately occurs most often. For example, if most of the specimens occur at approximately the same latitude, that value is assigned “most frequent” and ranked as number one. The next most frequently occurring is assigned “more frequent” and ranked as number two. The least frequently occurring is assigned “least frequent” and ranked as number three. If there is more than one relative maximum within each category (bimodal), the expert decides which relative maximum to use. The premise is that specimens that occur most closely together and/or at the same time of year are more likely to be the same species. This premise places a biological meaning and constraint on the spatial and temporal data being used as ranked modal values. The same process is used with regard to all variables and is the rule of syntax for generating linguistic values, J . For the universe of discourse, U , that is spatial and temporal variables, where x is the primary term, for each variable, the term sets for T are

$$T(\text{frequent}) = \{\text{least frequent}, \text{more frequent}, \text{most frequent}\} \quad (10)$$

and are recursively determined as

$$\begin{aligned} T^0 &= \emptyset, \\ T^1 &= \text{least frequent}, \\ T^2 &= \text{least frequent}, \text{more frequent}, \\ T^3 &= \text{least frequent}, \text{more frequent}, \text{most frequent}. \end{aligned} \quad (11)$$

That is,

$$T^{i+1} = \{\text{frequent}\} \cup \text{more } T^i \quad (12)$$

and

$$\text{more } T^i = \bigcup_{R \in T^i} \{\text{more } R\}. \tag{13}$$

Next, each specimen in each shape group is ranked per variable according to the rankings for the whole group. The semantic rule, $M(x)$, is that each variable rank is fuzzified as

$$v(R) = 1 - \frac{R}{\sum R}, \tag{14}$$

where $v(R)$ is the variable rank of x which defines the membership function $\mu(x)$ of x , and R is rank of variable with a value of 1, 2, 3, or 4. Ranking and fuzzification of each variable and each specimen is accomplished in the same way for each shape group. This fuzzy ranking acts as a fuzzy restriction on x given the semantic rule, M .

5.2. Weighting of variables between categories

Values from classification integration for each specimen assignment to a shape group are used with the fuzzified spatial and temporal data in a new fuzzy attribute matrix. The values for each category must be weighted to represent their degree of importance in taxonomic decision-making. To determine the weights, the expert identifies the weights using linguistic terms. The categories are ranked and a function is applied such that the calculated fuzzy values reflect the expert’s opinion. The effect is to “weight” the categories with respect to each specimen per shape group.

The primary term is “important”, and linguistic modifiers or hedges used are “very”, “somewhat”, “slightly”, and “not very”. Although morphological data is more important than spatial or temporal data, fuzzy specimen shape group assignment values, which represent morphological data in this matrix are “not very important” since the linguistic modifiers refer to spatial and temporal data only. However, morphological data still takes precedence since its fuzzy value in the matrix must exceed those for spatial and temporal data (Stoermer, personal communication). The weightings for values in this fuzzy matrix as the universe of discourse are determined by the term set

$$T(\text{important}) = \{\text{very}^*, \text{somewhat}^*, \text{slightly}^*, \text{not very}^*\}, \tag{15}$$

where $*$ represents the primary term. The rule of syntax for generating linguistic hedges is defined as

$$\begin{aligned} \text{Very} &= x^2, \\ \text{Somewhat} &= 1 - (1 - x^2), \\ \text{Slightly} &= 1 - (1 - x^3), \\ \text{Not very} &= \sqrt{x}, \end{aligned} \tag{16}$$

where $\tilde{A} = \{a_i\}$, $i = 1, \dots, n$, is the fuzzy set for each row (specimen) of the matrix generated from the semantic rule that orders the importance of each variable category with respect to the others as $\mu_{\tilde{A}}(x) : X \rightarrow \{0, 1\}$, $\forall x \in X$.

The normalized weighted fuzzy modal values, $h(x)$, are calculated as

$$h(x) = \frac{h_c(x)}{\max_{a_i \in C} h_c(a_i)} \quad \forall x \in X, \tag{17}$$

where $h_c(x)$ is an approximate fuzzy histogram (see e.g., [29]) for the c th variable category of all C categories. That is, the fuzzy membership values calculated from Eq. (16) as a_i are normalized and the result is $h_c(x)$ for each category for each specimen.

5.3. Fuzzy weighted attribute matrix and aggregation

Using the normalized weighted fuzzy attribute matrix, aggregation of the information is necessary to determine the degree to which each specimen belongs to its assigned shape group. The method of aggregation chosen is the weighted (ranking) sum. This aggregation operator is chosen because the categorical variables are independent of each other. Morphological data is biologically based. Latitude, longitude, long-term circulation patterns, and time of year are

measured independent of one another and on different scales. Linguistic modifier values are based on the mode for each category per column independent of other categories, and the weighted sum for each row is independent of other rows.

The rows in the weighted fuzzy attribute matrix are also suitability ratings (or weighted fuzzified rankings) [28] for each specimen with respect to the variable categories for each shape group. A sum of suitability ratings for each specimen would represent the total suitability [28] of that specimen in that shape group. Alternatively, the sum of the values for each specimen represents a fitness value with respect to that shape group. This fitness value may indicate the approximate optimum value that the specimen can attain with respect to that shape group.

For each shape group, the approximate optimum is defined as the mode for each category and is the fuzzy set $G(x) = \{g_j\}$, $j = 1, \dots, m$. A weighted sum of these values is the threshold against which each specimen is judged. The weighted fuzzy mode threshold or threshold weighted sum is

$$|G(x)| = \sum_{j=1}^m g_j, \quad (18)$$

where x is the category modal value of latitude, longitude, long-term circulation pattern, and time of year, and by the extension principle [32], the sum is a fuzzy addition.

For each specimen, the weighted sum is the relative cardinality [35] given as

$$\|A\| = \frac{|A|}{|G|}, \quad (19)$$

where $|G|$ is the threshold weighted sum and $|A| = \sum_{x \in X} \mu_A(x)$ is the cardinality of each specimen. The relative cardinalities are then normalized, and these values represent the degree that the specimen belongs to the shape group it was assigned to given combined spatial and temporal factors and fuzzy morphological analytical results. The membership is defined for the i th specimen in each shape group as

$$\mu_{\text{weighted}}(x_i) = \begin{cases} 1 & \text{if } x_i = x_{\max}, \\ 0.5 < x_i < 1 & \text{otherwise.} \end{cases} \quad (20)$$

5.4. Expert evaluation

Expert evaluation is crucial in biological taxonomic decision-making. Dr. Eugene F. Stoermer from University of Michigan is an expert in diatom research and has studied diatoms for over 40 years. The specimens used for this study were viewed from strewn mounts on microscope slides and were collected from samples from the Great Lakes in the 1960s and 1970s and are a part of Dr. Stoermer's research collection.

There are two types of data available to analyze the taxonomic problem. Initially, morphological analyses are the accepted way to perform a taxonomic study. This type of analysis takes precedence over spatial and temporal data analysis, but does not entirely supercede it (Stoermer, personal communication). Spatial and temporal data are useful proxies with regard to reproductive isolation. In this study, Dr. Stoermer was consulted and made decisions on the relative importance of morphological vs. spatial and temporal variables (Section 5.2), the relative importance among spatial and temporal variables (Section 5.2), the modal value for a category when more than one relative maximum exists (Section 5.2), linguistically express the degree of truth about the normalized fuzzy aggregation values for each specimen (Section 5.3), and evaluate the final results.

5.5. Fuzzy decision set and linguistic truth-value approximation

Fuzzy weighted membership values for each specimen in each shape group are evaluated with respect to the degree their membership is true. Since membership is already true to the degree 0.49 based on morphology alone, re-evaluation with spatial and temporal data implies that membership values will be $a_i = [0.5, 1.0]$ for the i th specimen in each shape group. A fuzzy truth set is devised from the membership values interval that is common for all shape groups. A linguistic decision is generated by devising meaningful labels for the fuzzy truth set, and these labels are determined by linguistic approximation using a semantic equivalence [32,33]. That is, I want to find a linguistic approximation that is close in

meaning to the degree that each specimen's assignment to a shape group is true and the degree of agreement expressed by the expert.

For a fuzzy truth set, τ , an equivalent set L is defined by

$$\mu_L(a) = \mu_\tau(\mu_B(a)), \quad (21)$$

where $\mu_B(a)$ is the degree of agreement of the results by the expert. The equivalent set is inclusively $L(\text{Truth}) = \{\text{true, mostly true, more or less true, not as true}\}$. Additional linguistic terms may be used (e.g., see [3]). A fuzzy gradient assessment table of intervals of suitability of specimen assignment, a , equivalent set of linguistic truth-values, and expert's degree of agreement is devised. A linguistic solution is presented of the results.

6. Results

Relative cardinality reflects the situation of each specimen per shape group with respect to what is actually known about the specimens biologically and taxonomically. None of the specimens are actually completely excluded or included in each shape group. The values are maximized in order of importance. By normalizing the values, a rescaling is useful to indicate the degree to which each specimen belongs to its assigned shape group. The values normalized by the fuzzy mode exhibit the "typicality" of each specimen in each shape group, and those specimens with values of 1.0 characterize each shape group.

For each shape group, fuzzy matrices were devised of specimen by category values, including spatial, temporal, and morphological values. Fuzzy values based on the mode (as calculated from Eq. (14)) for each of the six shape groups are presented in Table 1. Weighted fuzzy modal values (as calculated from Eq. (16)) are presented in Table 2. Normalized weighted fuzzy modal values (as calculated from Eq. (17)), their sum (as calculated from Eq. (19)), and the normalized sum (as calculated from Eq. (20)) are presented in Table 3.

From the fuzzy modal values in each category, there is a widespread departure between the spatial and temporal values and the morphological values. From Table 1, fuzzy modes for spatial and temporal variables indicate a difference in importance for each specimen in contrast to fuzzy morphological values. In some cases, fuzzy morphological values are identical among all specimens in a given shape group (Table 1—shape groups 1, 3, and 5), whereas spatial and temporal fuzzy modes are not. An interesting result is that since spatial and temporal variables are independent, latitude and longitude fuzzy modes are mostly different within shape groups (Table 1—shape groups 2, 4, 5, and 6).

From Table 2, the relative level of importance between spatial and temporal variables is reflected in the weighting applied. As previously stated, morphological values (local) take precedence over spatial and temporal values (global). Subsequently, the order of importance across spatial and temporal values range from latitude and longitude to time of year to long-term circulation pattern.

Normalized weighted modes in Table 3 reflect the relative position each specimen has to every other specimen in each shape group. Those specimens with a value of 1.0 define their assigned shape groups. The sum, which is normalized for each specimen, gives the value of suitability that the specimen has with respect to belonging to its shape group. A value of 1.0 for a specimen characterizes the shape group morphologically, spatially, and temporally. Other specimens relative to those members with a value of approximately 1.0 indicate the degree of suitability that these specimens have in their assigned shape group.

For each shape group, the normalized weighted sum for each specimen was evaluated by the expert, and his judgment was rendered as to the degree he agreed with the result. His evaluation was compiled as a solution to be viewed for all shape groups. In this regard, and given the results from Eq. (20), the equivalence set of linguistic approximations was assigned to four subsets on the interval [0.5, 1.0], and the results of the evaluation are presented in Table 4. The expert's degree of agreement was 1.0 (Stoermer, personal communication).

Four suitability intervals were devised, and each interval had a linguistic approximation from the equivalent set. Each specimen was then assigned a linguistic approximation based on Table 4. The linguistic solution for all specimens in all shape groups is given in Table 5.

By using 0.75 as a "crossover" point from the interval [0.5, 1.0], most of the shape groups have the majority of their members in the "true" and "mostly true" categories. Shape groups 1, 3, and 4 have all but one member at or greater than 0.75. Shape group 2 has two members and shape groups 5 and 6 have four members at less than 0.75.

Table 1
Fuzzy modal values of spatial and temporal data and fuzzy morphological values for each specimen in its assigned shape group

Shape group	Specimen	Latitude	Longitude	Long term circulation	Time of year	Fuzzy morphological values
1	111	0.67	0.67	0.83	0.83	0.68
	116	0.67	0.67	0.83	0.83	0.68
	192	0.83	0.83	0.83	0.83	0.68
	193	0.83	0.83	0.83	0.83	0.68
	195	0.83	0.83	0.83	0.83	0.68
	198	0.83	0.83	0.83	0.83	0.62
	19F	0.83	0.83	0.83	0.83	0.68
	19H	0.83	0.83	0.83	0.83	0.68
	19T	0.83	0.83	0.83	0.83	0.68
	391	0.50	0.50	0.67	0.67	0.68
2	O51	0.83	0.83	0.83	0.83	0.67
	O52	0.83	0.83	0.83	0.83	0.67
	115	0.67	0.67	0.67	0.83	0.67
	11H	0.67	0.67	0.67	0.83	0.61
	127	0.67	0.83	0.83	0.67	0.57
	129	0.67	0.83	0.83	0.67	0.67
	197	0.50	0.50	0.67	0.83	0.60
	Aa4	0.83	0.83	0.50	0.50	0.66
3	118	0.67	0.67	0.83	0.67	0.60
	142	0.67	0.67	0.67	0.67	0.60
	764	0.83	0.83	0.83	0.83	0.60
	767	0.83	0.83	0.83	0.83	0.60
	768	0.83	0.83	0.83	0.83	0.60
	76E	0.83	0.83	0.83	0.83	0.60
4	123	0.50	0.83	0.90	0.60	0.73
	12E	0.50	0.83	0.90	0.60	0.72
	135	0.67	0.83	0.80	0.90	0.73
	136	0.67	0.83	0.80	0.90	0.68
	137	0.67	0.83	0.80	0.90	0.73
	148	0.67	0.67	0.90	0.70	0.73
	14F	0.67	0.67	0.90	0.70	0.73
	14H	0.67	0.67	0.90	0.70	0.73
	191	0.50	0.50	0.60	0.60	0.61
	A5	0.67	0.83	0.80	0.90	0.73
	AT	0.83	0.83	0.70	0.80	0.73
	Aa3	0.83	0.83	0.70	0.80	0.73
	Aa5	0.83	0.83	0.70	0.80	0.73
	Aa6	0.83	0.83	0.70	0.80	0.73
AaE	0.67	0.83	0.80	0.90	0.73	
5	112	0.67	0.50	0.67	0.67	0.61
	126	0.67	0.67	0.83	0.50	0.61
	141	0.83	0.83	0.83	0.83	0.61
	143	0.83	0.83	0.83	0.83	0.61
	149	0.83	0.83	0.83	0.83	0.61
	14W	0.83	0.83	0.83	0.83	0.61
	199	0.67	0.50	0.67	0.67	0.61
	Aa2	0.83	0.67	0.50	0.50	0.61
6	12T	0.50	0.83	0.67	0.50	0.63
	131	0.83	0.83	0.83	0.83	0.63
	132	0.83	0.83	0.83	0.83	0.63
	133	0.83	0.83	0.83	0.83	0.63
	134	0.83	0.83	0.83	0.83	0.63
	144	0.50	0.50	0.67	0.50	0.63
	194	0.50	0.67	0.50	0.67	0.63
	196	0.50	0.67	0.50	0.67	0.63
	A1	0.67	0.83	0.83	0.83	0.49
	A4	0.67	0.83	0.83	0.83	0.63
	A9	0.67	0.83	0.83	0.83	0.63
	AW	0.67	0.83	0.83	0.83	0.63

Table 2
 Weighted fuzzy modal values of spatial and temporal data and fuzzy morphological values for each specimen in its assigned shape group

Shape group	Specimen	Weighted latitude	Weighted longitude	Weighted long-term circulation	Weighted time of year	Weighted fuzzy morphological values
1	111	0.44	0.44	0.58	0.69	0.82
	116	0.44	0.44	0.58	0.69	0.82
	192	0.69	0.69	0.58	0.69	0.82
	193	0.69	0.69	0.58	0.69	0.82
	195	0.69	0.69	0.58	0.69	0.82
	198	0.69	0.69	0.58	0.69	0.79
	19F	0.69	0.69	0.58	0.69	0.82
	19H	0.69	0.69	0.58	0.69	0.82
	19T	0.69	0.69	0.58	0.69	0.82
	391	0.25	0.25	0.30	0.44	0.82
2	O51	0.69	0.69	0.57	0.69	0.82
	O52	0.69	0.69	0.57	0.69	0.82
	115	0.45	0.45	0.30	0.69	0.82
	11H	0.45	0.45	0.30	0.69	0.78
	127	0.45	0.69	0.57	0.45	0.75
	129	0.45	0.69	0.57	0.45	0.82
	197	0.25	0.25	0.30	0.69	0.77
	Aa4	0.69	0.69	0.13	0.25	0.81
3	118	0.45	0.45	0.57	0.45	0.77
	142	0.45	0.45	0.30	0.45	0.77
	764	0.69	0.69	0.57	0.69	0.77
	767	0.69	0.69	0.57	0.69	0.77
	768	0.69	0.69	0.57	0.69	0.77
	76E	0.69	0.69	0.57	0.69	0.77
	4	123	0.25	0.69	0.73	0.36
12E		0.25	0.69	0.73	0.36	0.85
135		0.44	0.69	0.51	0.81	0.85
136		0.44	0.69	0.51	0.81	0.82
137		0.44	0.69	0.51	0.81	0.85
148		0.44	0.44	0.73	0.49	0.85
14F		0.44	0.44	0.73	0.49	0.85
14H		0.44	0.44	0.73	0.49	0.85
191		0.25	0.25	0.22	0.36	0.78
A5		0.44	0.69	0.51	0.81	0.85
AT		0.69	0.69	0.34	0.64	0.85
Aa3		0.69	0.69	0.34	0.64	0.85
Aa5		0.69	0.69	0.34	0.64	0.85
Aa6		0.69	0.69	0.34	0.64	0.85
AaE		0.44	0.69	0.51	0.81	0.85
5	112	0.44	0.25	0.30	0.44	0.78
	126	0.44	0.44	0.58	0.25	0.78
	141	0.69	0.69	0.58	0.69	0.78
	143	0.69	0.69	0.58	0.69	0.78
	149	0.69	0.69	0.58	0.69	0.78
	14W	0.69	0.69	0.58	0.69	0.78
	199	0.44	0.25	0.30	0.44	0.78
	Aa2	0.69	0.44	0.13	0.25	0.78
6	12T	0.25	0.69	0.30	0.25	0.79
	131	0.69	0.69	0.58	0.69	0.79
	132	0.69	0.69	0.58	0.69	0.79
	133	0.69	0.69	0.58	0.69	0.79
	134	0.69	0.69	0.58	0.69	0.79
	144	0.25	0.25	0.30	0.25	0.79
	194	0.25	0.44	0.13	0.44	0.79
	196	0.25	0.44	0.13	0.44	0.79
	A1	0.44	0.69	0.58	0.69	0.70
	A4	0.44	0.69	0.58	0.69	0.79
	A9	0.44	0.69	0.58	0.69	0.79
	AW	0.44	0.69	0.58	0.69	0.79

Table 3

Normalized weighted fuzzy modal values of spatial and temporal data and fuzzy morphological values for each specimen in its assigned shape group

Specimen	Normalized weighted latitude	Normalized weighted longitude	Normalized weighted long-term circulation	Normalized weighted time of year	Normalized weighted fuzzy morphological values	SUM	Normalized sum
111-1	0.64	0.64	1.00	1.01	1.01	4.30	0.86
116-1	0.64	0.64	1.00	1.01	1.01	4.30	0.86
192-1	1.01	1.01	1.00	1.01	1.01	5.02	1.00
193-1	1.01	1.01	1.00	1.01	1.01	5.02	1.00
195-1	1.01	1.01	1.00	1.01	1.01	5.02	1.00
198-1	1.01	1.01	1.00	1.01	0.96	4.98	0.99
19F-1	1.01	1.01	1.00	1.01	1.01	5.02	1.00
19H-1	1.01	1.01	1.00	1.01	1.01	5.02	1.00
19T-1	1.01	1.01	1.00	1.01	1.01	5.02	1.00
391-1	0.36	0.36	0.51	0.64	1.01	2.89	0.57
O51-2	1.00	1.00	1.00	1.00	1.00	5.00	1.00
O52-2	1.00	1.00	1.00	1.00	1.00	5.00	1.00
115-2	0.65	0.65	0.53	1.00	1.00	3.83	0.77
11H-2	0.65	0.65	0.53	1.00	0.95	3.78	0.76
127-2	0.65	1.00	1.00	0.65	0.92	4.22	0.84
129-2	0.65	1.00	1.00	0.65	1.00	4.30	0.86
197-2	0.36	0.36	0.53	1.00	0.94	3.20	0.64
Aa4-2	1.00	1.00	0.22	0.36	0.99	3.57	0.71
118-3	0.65	0.65	1.00	0.65	1.01	3.96	0.79
142-3	0.65	0.65	0.53	0.65	1.01	3.49	0.70
764-3	1.00	1.00	1.00	1.00	1.01	5.00	1.00
767-3	1.00	1.00	1.00	1.00	1.01	5.00	1.00
768-3	1.00	1.00	1.00	1.00	1.01	5.00	1.00
76E-3	1.00	1.00	1.00	1.00	1.01	5.00	1.00
123-4	0.36	1.01	1.00	0.44	1.01	3.82	0.88
12E-4	0.36	1.01	1.00	0.44	1.00	3.81	0.87
135-4	0.64	1.01	0.70	1.00	1.01	4.36	1.00
136-4	0.64	1.01	0.70	1.00	0.97	4.32	0.99
137-4	0.64	1.01	0.70	1.00	1.01	4.36	1.00
148-4	0.64	0.64	1.00	0.60	1.01	3.90	0.89
14F-4	0.64	0.64	1.00	0.60	1.01	3.90	0.89
14H-4	0.64	0.64	1.00	0.60	1.01	3.90	0.89
191-4	0.36	0.36	0.30	0.44	0.92	2.38	0.55
A5-4	0.64	1.01	0.70	1.00	1.01	4.36	1.00
AT-4	1.01	1.01	0.47	0.79	1.01	4.28	0.98
Aa3-4	1.01	1.01	0.47	0.79	1.01	4.28	0.98
Aa5-4	1.01	1.01	0.47	0.79	1.01	4.28	0.98
Aa6-4	1.01	1.01	0.47	0.79	1.01	4.28	0.98
AaE-4	0.64	1.01	0.70	1.00	1.01	4.36	1.00
112-5	0.64	0.36	0.51	0.64	1.00	3.16	0.63
126-5	0.64	0.64	1.00	0.36	1.00	3.65	0.73
141-5	1.01	1.01	1.00	1.01	1.00	5.02	1.00
143-5	1.01	1.01	1.00	1.01	1.00	5.02	1.00
149-5	1.01	1.01	1.00	1.01	1.00	5.02	1.00
14W-5	1.01	1.01	1.00	1.01	1.00	5.02	1.00
199-5	0.64	0.36	0.51	0.64	1.00	3.16	0.63
Aa2-5	1.01	0.64	0.22	0.36	1.00	3.23	0.64
12T-6	0.36	1.01	0.51	0.36	1.00	3.25	0.65
131-6	1.01	1.01	1.00	1.01	1.00	5.02	1.00
132-6	1.01	1.01	1.00	1.01	1.00	5.02	1.00
33-6	1.01	1.01	1.00	1.01	1.00	5.02	1.00
134-6	1.01	1.01	1.00	1.01	1.00	5.02	1.00
144-6	0.36	0.36	0.51	0.36	1.00	2.60	0.52
194-6	0.36	0.64	0.22	0.64	1.00	2.87	0.57
196-6	0.36	0.64	0.22	0.64	1.00	2.87	0.57
1-6	0.64	1.01	1.00	1.01	0.89	4.54	0.90
A4-6	0.64	1.01	1.00	1.01	1.00	4.66	0.93
A9-6	0.64	1.01	1.00	1.01	1.00	4.66	0.93
W-6	0.64	1.01	1.00	1.01	1.00	4.66	0.93

Specimen number followed by a hyphenated shape group number. The sum and normalized sum are also indicated.

Table 4
Fuzzy gradient assessment table of suitability intervals, equivalent set of linguistic truth-values, and expert's degree of agreement

Suitability interval	Truth-value	Degree of agreement
$0.500 \leq a < 0.625$	Not as true	1
$0.625 \leq a < 0.750$	More or less true	1
$0.750 \leq a < 0.875$	Mostly true	1
$0.875 \leq a < 1.00$	True	1

7. Discussion

Morphological studies are incomplete in many ways, and morphology is not the only criterion for judging species identification. Classification integration was useful in defining shape groups and their similarity with respect to fuzzy average overlap and degree of membership per shape group. However, this is not entirely sufficient to make taxonomic decisions.

Classification integration of morphological data determined that in six shape groups, specimen assignment was plausible. The assignments were not necessarily unique or optimal, but plausibility was a sufficient basis as a beginning with the intention of further analyses. To begin with, the number of shape groups was unknown and information about the specimens was limited. From classification integration, the expected result of fuzzy average overlap values at approximately the crossover value or slightly larger indicated similarity among pairs of shape groups. This is an indication of developmental stages that are difficult to distinguish by morphological analysis alone.

The additional analyses using sample data was helpful in finding the degree of truth in specimen assignment to each shape group. While morphology was indicative of inheritance (shape) and development (morphometry), sample spatial and temporal data were useful as proxies for reproductive isolation.

From the results, most of the specimens are suitable in the shape group to which they were assigned, and characterized the shape group. Those specimens in the “not as true” category might be candidates for re-evaluation. Specimens in the “more or less true” category may or may not be re-evaluated. This largely depends on the purpose to which the current results might be used in future research, or how the current results are received by the diatom taxonomy research community. In any case, the results indicate that linguistic solution to the suitability of specimens in their assigned shape groups is useful in taking the next step in taxonomic studies, which is to identify shape groups as species groups.

The entire analysis was directed toward determining the degree to which each shape group is a separate species. The results from this analysis showed that six shape groups are the basis for six hypotheses, namely, each shape group can be tested further with respect to species identification. That is, if a species is named for each shape group, each species designation is a hypothesis subject to further testing. Such testing using fuzzy logic must be defined and constrained by biological principles and any other possible types of biological information acquired.

Taxonomic decision-making using fuzzy analytical methods has many advantages. First, the aggregation method chosen is simple. Spatial and temporal data used are independent and additive therefore, the usual pitfalls of using a weighted sum were not realized. There are no interactions to be concerned with since the mode was used as the criterion for evaluation. Second, the data used were constrained by biological information and expert opinion. This provides a narrow range of options for meaningful fuzzy data analysis. Third, spatial and temporal data do not depend on being ordered. A quasi-ordering occurs once the values are ranked in each category variable. Fourth, this analysis uses the combination of local and global variables, although taxonomic solutions are local but become global. All the variables are constrained by biological principles, viz., shape inheritance, morphological changes with respect to development, and reproductive isolation that implies species differentiation. Fifth, linguistic solutions are readily recognizable and accepted by biological taxonomists. In addition, more than four linguistic terms may be employed in the equivalent set, and this provides for flexibility in arriving at a linguistic solution.

This study provides a way to do taxonomy that is recognizable and understandable to taxonomists, but reduces the arbitrariness in species identification problems. Taxonomic decision-making is not an exercise in codifying an individual's preferences, since such decisions must rest on some sort of evidence no matter how little of it is available. This study makes fuzzy decision making in taxonomy accessible to taxonomists, and taxonomists could assess one another about their decisions. Linguistic solutions, such as those presented in Table 5, are highly acceptable in a field such as biological taxonomy where numerical solutions are not typical or well understood. This study shows how fuzzy

Table 5
Linguistic solution for each specimen in each shape group based on Table 4

Shape group	Specimen	Normalized SUM	Linguistic solution
1	111	0.86	mostly true
	116	0.86	mostly true
	192	1.00	true
	193	1.00	true
	195	1.00	true
	198	0.99	true
	19F	1.00	true
	19H	1.00	true
	19T	1.00	true
	391	0.57	not as true
2	O51	1.00	true
	O52	1.00	true
	115	0.77	mostly true
	11H	0.76	mostly true
	127	0.84	mostly true
	129	0.86	mostly true
	197	0.64	more or less true
	Aa4	0.71	more or less true
3	118	0.79	mostly true
	142	0.70	more or less true
	764	1.00	true
	767	1.00	true
	768	1.00	true
	76E	1.00	true
4	123	0.88	true
	12E	0.87	mostly true
	135	1.00	true
	136	0.99	true
	137	1.00	true
	148	0.89	true
	14F	0.89	true
	14H	0.89	true
	191	0.55	not as true
	A5	1.00	true
	AT	0.98	true
	Aa3	0.98	true
	Aa5	0.98	true
	Aa6	0.98	true
AaE	1.00	true	
5	112	0.63	more or less true
	126	0.73	more or less true
	141	1.00	true
	143	1.00	true
	149	1.00	true
	14W	1.00	true
	199	0.63	more or less true
	Aa2	0.64	more or less true
6	12T	0.65	more or less true
	131	1.00	true
	132	1.00	true
	133	1.00	true
	134	1.00	true
	144	0.52	not as true
	194	0.57	not as true
	196	0.57	not as true
	A1	0.90	true
	A4	0.93	true
	A9	0.93	true
	AW	0.93	true

logic, approximate reasoning, and information fusion are what actually happen in solving a biological problem, viz., taxonomy and species identification.

From this study, there are issues that need to be explored further. One problem was how to determine which fuzzy mode to use for each category. In this study, consultation with the expert was used. Solving this problem in a more analytical way might be accomplished by creating a fuzzy measure to find the minimum distance between the “ideal” for the category and each of the relative maxima. There is still the potential for ties among distance minima. Perhaps a resampling method would be an alternative to use in choosing the relative minimum to use. Alternatively, a resemblance measure based on a fuzzy histogram [29] might be devised to compare a specimen with what is “typical” for a given shape group.

Another issue is that of independence of variables and method of aggregation [6]. In this study, the variables to be aggregated were independent, and therefore additive. If variables used are not independent, then non-additive aggregation is warranted. In this regard, the Choquet integral, among others, might be a suitable option [4,9,10]. This study provided an initial solution to the problem of aggregating spatial and temporal data with morphological data. Although the methods used here do not necessarily provide for a unique solution, given the scant evidence and analysis of specimens relative to each other, the solution is a worthy starting point. In addition, there is the potential that another expert would reassign specimens to different groups, or even weight the spatial and temporal variables differently. This study is flexible in that amendments can be made to accommodate other types of taxonomic studies, or studies that have a similar compilation of a variety of inputs. It can be adapted to include more than one expert or more than one decision-making group. More than one expert may be consulted to contribute to the decision-making process. In this case, the process would become more complex, but would be accommodated by some form of fuzzy group decision-making [34]. In essence, the methods used here are simple, adaptable, and understandable to biological taxonomists, especially with the incorporation of linguistic solutions.

It may be that reassigning suboptimal specimens to other shape groups would produce a more optimal solution. By suboptimal I mean those specimens with a value of less than 1.00. Another scenario might be that some of the specimens with suboptimal assignment do not belong to any of the shape groups. That is, some specimens may belong to a yet-to-be described shape group. A fuzzy permutation program reassigning suboptimal specimens to other shape groups and re-evaluating results until the more optimal solution is obtained might be useful. The permutation can be of any type and crosschecked with other permutation methods (e.g., bootstrapping, jackknifing, Monte Carlo permutation). Alternatively, fuzzy decision trees may be used [31].

Perhaps one way to refine results further would be to find relative densities for each shape group with respect to fuzzy overlap values. If it could be determined which specimens contributed the largest degree to fuzzy average overlap, then those particular specimens would be candidates for reassignment and testing, based on a newly devised fuzzy measure for developmental stage.

Biological taxonomy is not a static enterprise since a scientific name is a hypothesis to be tested, so an optimal or unique solution is not necessarily a realistic expectation. The current results are constrained by biological information, starting with predefined shape groups from an earlier morphological study and incorporating expert opinion in the decision making process. This provides a sense of validity to the enterprise of biological taxonomy that is not merely opinion, albeit authoritative opinion.

Taxonomic studies are necessary precursors to phylogenetic analyses and reconstruction of phylogenetic relations among taxa. This study has importance for other biological studies, in general, in that a formal way of assessing species identification is a very important indicator of the species concept in biology and how that concept is applied. Another adjunct study to this one includes exploring fuzzy position with respect to an organism and climate or habitat quality since latitude and longitude are independent variables. This study is only the beginning of possibilities for applications of fuzzy logic and decision-making to biological studies.

8. Conclusions

My study was conducted to show that biological taxonomic decision-making may be augmented by the use of fuzzy sets and fuzzy decision-making. Information about specimens from the genus *Asterionella* was scant, and some of the sample data was used as proxies for inferred biological information. Degree of suitability in specimen assignment to shape groups was determined after initial sorting using classification integration of morphometric data. Degree of truth

in specimen assignment, as given in Table 5, reflects acceptability of the results to an expert in diatom research. That is, the degree to which each specimen is a species was determined using the methodology outlined in my study. Throughout my study, biological principles of developmental stage or ontogeny (morphology and morphometrics), reproductive isolation, and species concepts constrained the analysis to reflect a realistic way to conduct biological taxonomy. Moreover, not only was a numerical result per specimen achieved, but also a linguistic solution was determined that makes this study accessible to biologists.

In conducting biological taxonomic studies, researchers will never have all the morphotypes or “species” available for empirical assessment. My study shows that given a limited amount of information and a limited number of specimens, biological taxonomy using fuzzy decision-making can be accomplished to reflect a realistic portrayal of the state of species designations at a given time.

Acknowledgements

I would like to thank Dr. E. F. Stoermer for participating in this study as an expert in diatom research and three anonymous reviewers for helpful comments on this manuscript.

References

- [1] R.W. Battarbee, D.F. Charles, S.S. Dixit, I. Renberg, Diatoms as indicators of surface water acidity, in: E.F. Stoermer, J.P. Smol (Eds.), *The Diatoms: Applications for the Environmental and Earth Sciences*, Cambridge University Press, Cambridge, 1999, pp. 85–127.
- [2] D. Beletsky, J.H. Saylor, D.J. Schwab, Mean circulation in the Great Lakes, *J. Great Lakes Res.* 25 (1999) 78–93.
- [3] R.E. Bellman, L. Zadeh, Local and fuzzy logics, in: J.M. Dunn, G. Epstein (Eds.), *Modern Uses of Multiple-Valued Logic*, D. Reidel Publishing Company, Dordrecht, 1977, pp. 105–165.
- [4] M. Ceberio, F. Modave, An interval-valued, 2-additive Choquet integral for multicriteria decision making, 10th Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU), July 2004, Perugia, Italy.
- [5] D. Dubois, H. Prade, *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, 1980 393 pp.
- [6] D. Dubois, H. Prade, R. Yager, Merging fuzzy information, in: J.C. Bezdek, D. Dubois, H. Prade (Eds.), *Fuzzy Sets in Approximate Reasoning and Information Systems*, Kluwer Academic Publishers, Boston, 1999, pp. 337–401.
- [7] J.M. Dunn, Indices of partition fuzziness and detection of clusters in large data sets, in: M.M. Gupta, G.N. Saridis, B.R. Gaines (Eds.), *Fuzzy Automata and Decision Process*, North-Holland Publishers, Amsterdam, 1977, pp. 271–283.
- [8] S.C. Fritz, B.F. Cumming, F. Gasse, K.R. Laird, Diatoms as indicators of hydrologic and climatic change in saline lakes, in: E.F. Stoermer, J.P. Smol (Eds.), *The Diatoms: Applications for the Environmental and Earth Sciences*, Cambridge University Press, Cambridge, 1999, pp. 41–72.
- [9] M. Grabisch, C. Labreuche, Fuzzy measures and integrals in mcda, in: J. Figueira, S. Greco, M. Ehrgott (Eds.), *Multiple Criteria Decision Analysis*, Kluwer Academic Publishers, Boston, 2004, pp. 1–50.
- [10] M. Grabisch, M. Roubens, Application of the Choquet integral in multicriteria decision making, in: M. Grabisch, T. Murofushi, M. Sugeno (Eds.), *Fuzzy Measures and Integrals—Theory and Applications*, Physica Verlag, Germany, 2000, pp. 415–434.
- [11] R.I. Hall, J.P. Smol, Diatoms as indicators of lake eutrophication, in: E.F. Stoermer, J.P. Smol (Eds.), *The Diatoms: Applications for the Environmental and Earth Sciences*, Cambridge University Press, Cambridge, 1999, pp. 85–127.
- [12] A.H. Hassall, Memoir on the organic analysis or microscopic examination of the water supplied to the inhabitants of London and the suburban districts, *The Lancet* 1 (1850) 230–235.
- [13] A. Kandel, *Fuzzy Techniques in Pattern Recognition*, Wiley, New York, 1982 356 pp.
- [14] H. Körner, Morphologie und taxonomie der diatomeengattung *Asterionella*, *Nova Hedwigia* 20 (1970) 557–724.
- [15] D.G. Mann, Why didn't Lund see sex in *Asterionella*? A discussion of the diatom life cycle in nature, in: F.E. Round (Ed.), *Algae and the Aquatic Environment*, Biopress, Bristol, 1988, pp. 384–412.
- [16] D.G. Mann, The species concept in diatoms, *Phycologia* 38 (1999) 437–495.
- [17] J.L. Pappas, Fourier shape analysis and shape group determination by principal component analysis and fuzzy measure theory of *Asterionella* Hassall (Heterokontophyta, Bacillariophyceae) from the Great Lakes, Dissertation, University of Michigan, 2000.
- [18] J.L. Pappas, G.W. Fowler, E.F. Stoermer, Calculating shape descriptors from Fourier analysis: shape analysis of *Asterionella* (Heterokontophyta, Bacillariophyceae), *Phycologia* 40 (2001) 440–456.
- [19] J.L. Pappas, E.F. Stoermer, Multidimensional analysis of diatom morphologic and morphometric phenotypic variation and relation to niche, *Ecoscience* 2 (1995) 357–367.
- [20] J.L. Pappas, E.F. Stoermer, *Asterionella* Hassall (Heterokontophyta, Bacillariophyceae): taxonomic history and quantitative methods as an aid to valve shape differentiation, *Diatom* 17 (2001) 47–58.
- [21] J.L. Pappas, E.F. Stoermer, Fourier shape analysis and fuzzy measure shape group differentiation of Great Lakes *Asterionella* (Heterokontophyta, Bacillariophyceae), in: A. Economou-Amilli (Ed.), *Proc. 16th Internat. Diatom Symp.*, Amvrosiou Press, Athens, 2001, pp. 485–501.
- [22] F.E. Round, R.M. Crawford, D.G. Mann, *The Diatoms—Biology & Morphology of the Genera*, Cambridge University Press, Cambridge, 1990 747 pp.

- [23] C. Sancetta, Diatoms and marine paleoceanography, in: E.F. Stoermer, J.P. Smol (Eds.), *The Diatoms: Applications for the Environmental and Earth Sciences*, Cambridge University Press, Cambridge, 1999, pp. 374–386.
- [24] E.F. Stoermer, T.B. Ladewski, Apparent optimal temperature for the occurrence of some common phytoplankton species in southern Lake Michigan, *Great Lakes Research Division Publication* 18, University of Michigan, Ann Arbor, Mi, 1976 pp. 50.
- [25] E.F. Stoermer, J.J. Yang, Plankton diatom assemblages in Lake Michigan, *Great Lakes Research Division Special Report No. 47*, University of Michigan, Ann Arbor, Mi, 1969 168pp.
- [26] M. Sugeno, Fuzzy measures and fuzzy integrals: a survey, in: M.M. Gupta, G.N. Saridis, B.R. Gaines (Eds.), *Fuzzy Automata and Decision Processes*, North-Holland Publishers, New York, 1977, pp. 89–102.
- [27] T. Terano, K. Asai, M. Sugeno, *Fuzzy Systems Theory and its Applications*, Academic Press, San Diego, 1992 268pp.
- [28] R.M. Tong, P.P. Bonissone, Linguistic solutions to fuzzy decision problems, in: H.-J. Zimmermann, L.A. Zadeh, B.R. Gaines (Eds.), *Fuzzy Sets and Decision Analysis*, North-Holland, Amsterdam, 1984, pp. 323–334.
- [29] C. Vertan, N. Boujemaa, Embedding fuzzy logic in content based image retrieval, in: *19th Internat. Meeting of North American Fuzzy Information Processing Society, NAFIPS*, Atlanta, 2000.
- [30] Z. Wang, G.J. Klir, *Fuzzy Measure Theory*, Plenum Press, New York, 1992 354 pp.
- [31] Y. Yuan, M.J. Shaw, Induction of fuzzy decision trees, *Fuzzy Sets and Systems* 69 (1995) 125–139.
- [32] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, Parts 1, 2, and 3, *Inform. Sci.* 8 (1975) 199–249, 8 (1975) 301–357, 9 (1975) 43–80.
- [33] L.A. Zadeh, PRUF—a meaning representation language for natural language, *Internat. J. Man–Machine Stud.* 10 (1978) 395–460.
- [34] H.-J. Zimmerman, *Fuzzy Sets, Decision Making and Expert Systems*, Kluwer Academic Publishers, Boston, 1987 335pp.
- [35] H.-J. Zimmerman, *Fuzzy Set Theory—and its Applications*, second revised ed, Kluwer Academic Publishers, Boston, 1991 391pp.