Study of fragmented fossil diatoms using an invariant correlation method

C. Elizabeth Villalobos-Flores\textsuperscript{a}, J. Alvarez-Borrego\textsuperscript{b}, J. L. Pech-Pacheco\textsuperscript{c}, G. Cristóbal\textsuperscript{c} and E. Castro-Longoria\textsuperscript{b}

\textsuperscript{a} Facultad de Ciencias de la UNAM, Circuito Exterior Universitario, C.P. 04510 D.F., Delegación Coyoacán, México

\textsuperscript{b} Centro de Investigación Científica y de Educación Superior de Ensenada, División de Física Aplicada, Depto. de Óptica, Km. 107 carretera Tijuana-Ensenada, Ensenada, B.C., México

\textsuperscript{c} Instituto de Óptica, Daza de Valdés, CSIC, Serrano 121, Madrid España

ABSTRACT

The taxonomic identification of diatom species that constituted phytoplankton communities in remote times is determining in several research fields like ecology, evolution, paleoecology and biostratigraphy. In the last 30 years the use of fossil diatoms like environmental indicators has become of prime importance. However the use of these organisms is limited since they are found in sediment samples mostly fragmented or pulverized. This may lead to confusion and loss of information. In this work we used invariant correlation to identify 12 species of fossil diatoms. With this method we were able to identify the diatom species from only a small fragment of the organisms. This methodology can be used for the development of an automated system of plankton identification. An automatized identification of diatoms would be able to guarantee a faster identification and also would reduce the time necessary for accomplishing analysis of samples highly fragmented.

Keywords: invariant correlation, diatoms, marine phytoplankton, automatic systems

1. INTRODUCTION

Diatoms are one of the basic sources for the formation of organic matter in the ocean, and actively participate in sedimentation, not only during recent periods of time but throughout the remote past. The presence of diatom valves in marine paleoenvironments has been used for the study of climatic changes as well as geomorphological processes\textsuperscript{1,2}. One of the most important features for the recognition of diatom species as paleoenvironment and paleoclimatic indicators is the high correlation between the groups of diatoms and their ecophysiological behaviour, this permit the species association with diverse geological and hydrodynamic features.

The identification of diatom fossils requires the analysis of a great number of valves per sample. Generally, to obtain relative abundances and diversity indexes, diatom counts must go from 400 to $10^7$ structures per gram\textsuperscript{3}. The analysis of these samples requires a great amount of time and experience and, on the other hand, the samples analysed frequently contain material with different fragmentation degrees and this can lead to confusion and loss of information. Therefore, it is necessary the development of new techniques to facilitate the species recognition, even with fragmented organisms.
It is important to note that there is an increasing interest in the development of new techniques that simplify the analysis of biological samples, specially those groups of organisms that provide information for the study of spatial and seasonal changes in biological composition and ecosystem diversity as well as the climatic change and global environment problems. In this respect there are some optical techniques and photo-optic systems that are being calibrated with in situ measurements and this needs the development of automated systems for identification of biological material. In this way it would be possible the generation of databases that would serve to facilitate studies of biological and ecological properties.

The presentation is organized as follows. Section 2 provides a description of the invariant correlation method showing the 12 species involved. Section 3, the results obtained with the digital algorithm in the identification of the species presented are discussed. Section 4 summarizes our conclusions.

2. INVARIANT CORRELATION METHOD

Images of 21 fossil diatom species from Pacific Ocean were selected. These images have been identified, selected and published. The images were digitized with a resolution of 600 dpi using a Mustek® digitizer. Each image was recorded in a 256x256 matrix, f(x,y). Each image was randomly fragmented, 49 fragments were obtained from each image. We worked with 21 species, but in this work we show the results of 12 of them only (Fig. 1). Each fragmented image was compared with its original image and with the other fragments in order to find the minimum information required for the identification of the species. The comparison was realized using an invariant correlation method. A numerical simulation was performed in order to correlate diatom fossils species with phase-only filters (POF). POF is invariant to rotation and scale. All steps were developed digitally. In figure 2 we present our method step by step. Figure 2 shows the image, f(x,y), Step 1. In Step 2 the scale factor, \( r^{1/2} \), is applied to the image f(x,y). This process is what differentiates the scale transform from the Mellin-transform. After these steps, we mapped the cartesian coordinates to polar coordinates, Step 3. We introduced a bilinear interpolation of the first data of coordinates conversion. This is done to avoid the aliasing which in one way or another affects the identification of the objects, Step 4. The image obtained in Step 4 is the one that is going to be correlated with the phase-only filters (filters which have information about species to be recognized). In Step 5, we obtain the bidimensional scale transform via a FFT. Phase-only filters \( S_{POF}(u_p, v_0) \) is defined as

\[
S_{POF}(u_p, v_0) = \exp \left[-i\phi(u_p, v_0) \right] \quad \text{where} \quad |S(u_p, v_0)| = 1
\]

is equal to one, and \( u_p, v_0 \) are variables in frequency domain, Step 6. Correlation of digital filter, Step 7 with scale transform of the image generate a very low correlation value for geometrically dissimilar organisms and a high correlation value for geometrically similar organisms.

![Image](image.png)

Figure 1. a) *Actinocyclus ingens*, b) *Coscinodiscus nodulifer var1*, c) *Coscinodiscus nodulifer Schmidt*, d) *Actinocyclus ellipticus*, e) *Actinocyclus ellipticus moronensis*, f) *Nitzchia paereinholdii*, g) *Nitzchia reinholdii*, h) *Thalassiosira oestruppii var1*, i) *Thalassiosira oestruppii var2*, j) *Thalassiosira domifacta*, k) *Asteromphalus imbricatus*, l) *Pseudotriceratium cinamomeum*. 
3. RESULTS AND DISCUSSION

Figure 3 shows the result for *Actinocyclus ingens* when is correlated with all the entire organism and also with fragments of the 21 species, between them 12 of the species showed above. The x-axis shows the entire and fragmented diatoms and the y-axis shows the correlation values. The plot shows two lines, one of them corresponds to the correlation of the entire diatom versus the 21 species and the other one shows the correlation between a fragment of the specie, 13.5% (shown by the

Figure 3. Invariant correlation of the entire and fragmented *Actinocyclus ingens* diatom. The arrow indicates the plot of the fragmented diatom result.
arrow in the plot) versus the 21 species. We can see in this figure that even with this percentage of fragmentation it is still possible to identify the species. The next table shows the grade of fragmentation of the 12 species presented, where still is possible to identify the species. Therefore this method works perfectly because with so few information is possible to recuperate the information about the entire specie.

### Table 1

Minimum fragment identified by invariant correlation of 12 fossil diatom species

<table>
<thead>
<tr>
<th>Identification number of each specie</th>
<th>Scientific name</th>
<th>Fragment range</th>
<th>Percentage minimum identified. Fragment # (percent area)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td><em>Actinocyclus ingens</em></td>
<td>1-50</td>
<td>36 (13.5%)</td>
</tr>
<tr>
<td>B</td>
<td><em>Coscinodiscusnodulifer var</em></td>
<td>51-100</td>
<td>82 (12.0%)</td>
</tr>
<tr>
<td>C</td>
<td><em>Coscinodiscusnodulifer Schmidt</em></td>
<td>101-150</td>
<td>140 (4.0%)</td>
</tr>
<tr>
<td>D</td>
<td><em>Actinocyclus ellipticus</em></td>
<td>151-200</td>
<td>192 (5.0%)</td>
</tr>
<tr>
<td>E</td>
<td><em>Actinocyclus ellipticus morenensis</em></td>
<td>201-250</td>
<td>240 (4.5%)</td>
</tr>
<tr>
<td>F</td>
<td><em>Nitzchia paereinholdii</em></td>
<td>301-350</td>
<td>341 (7.0%)</td>
</tr>
<tr>
<td>G</td>
<td><em>Nitzchia reinholdii</em></td>
<td>1001-1050</td>
<td>1038 (12.5%)</td>
</tr>
<tr>
<td>H</td>
<td><em>Thalassiosira oestruppii var 1</em></td>
<td>401-450</td>
<td>438 (13.5%)</td>
</tr>
<tr>
<td>I</td>
<td><em>Thalassiosira oestruppii var 2</em></td>
<td>451-500</td>
<td>491 (6.0%)</td>
</tr>
<tr>
<td>J</td>
<td><em>Thalassiosira domifacta</em></td>
<td>501-550</td>
<td>535 (10.0%)</td>
</tr>
<tr>
<td>K</td>
<td><em>Asteromphalus imbricatus</em></td>
<td>551-600</td>
<td>588 (12.0%)</td>
</tr>
<tr>
<td>L</td>
<td><em>Pseudotriceratium cinamomeum</em></td>
<td>601-650</td>
<td>637 (8.0%)</td>
</tr>
</tbody>
</table>

### REFERENCES