

# Analysis of Image-Based Classification of Foraminiferal Tests

Kamal Ranaweera<sup>a</sup>, Santo Bains<sup>b</sup>, and Dileepan Joseph<sup>a,\*</sup>

<sup>a</sup>*Electrical and Computer Engineering, University of Alberta, Canada*

<sup>b</sup>*Research and Technology Group, BP Plc, UK*

---

## Abstract

Two factors govern the performance of computerized systems for classifying foraminiferal tests (forams) from optical microscopy images: (1) the amount of information captured by the images; and (2) the ability of computer algorithms to exploit the information. This paper provides an analysis of the former, to understand the efficacy and limitations of digital images in foram representation. To this end, 500 images were captured to show default (unmanipulated) and alternative views of 250 variable-grade specimens taken from 25 core samples. The specimens (mainly planktic forams) were classified by a specialist three times: first, from the default views alone; second, from the default and alternative views together; and third, from the archived particles. Image and particle-based classifications were compared using statistical methods to establish significance. Default views alone were sufficient to realize low incorrect genus and species rates and high correct genus rates, irrespective of specimen quality, view side, and image quality. Availability of alternative views had only a minimal impact on performance, at least for the statistical sample. However, image quality, which was primarily limited by depth-of-field issues, had a significant impact on performance. To facilitate this work, an online database called the Microfossil Wiki was developed. All images and classifications referred to here are accessible via the wiki.

## **1 Introduction**

With recent advancements in digital cameras, computer vision, and robotics, one should be able, in theory, to build a computerized system to automate the physical separation, or *picking*, and sorting of foraminiferal tests (forams) from sediment samples. Such a system should essentially: (a) capture images of sieved particles through an optical microscope, (b) identify the forams and foraminiferal taxa based on visual features, and (c) pick the identified forams and sort them by genus or species for further analysis, such as laboratory work to determine the chemical composition of the tests. One could also envision an automatic foram picker where identification is done *after* picking and sorting, i.e. where particles are first grouped according to image similarity alone (Ranaweera et al., 2009).

If micropaleontologists agree that automatic foram picking is desirable, a series of steps are required to engineer such a complex system. The first step is to analyze the efficacy and limitations of image-based classification of forams, in contrast to particle-based classification. Such an analysis has not been done before. Instead, the literature reports on a few systems geared towards fully-automatic identification of forams, mainly for faunal counts, and comments indirectly on image-based classification itself. Furthermore, the state of the art may not represent the best way towards an automatic foram picker as the latter does not require fully-automatic identification (Ranaweera et al., 2009). Thus, it is very difficult to extract from the literature the information essential to guide work on an automatic foram picker.

---

\* Please address correspondence to this author.

*Email address:* dil.joseph@ualberta.ca (Dileepan Joseph).

For example, the research group led by Dr. Thonnat in the French National Institute for Research in Computer Science and Control analyzed the automatic identification of planktic forams (Liu et al., 1994; Yu et al., 1996). The heart of their system consisted of a rule-base that associated visual features of specimens with their taxons. Given three images that show selected sides of an unclassified specimen, the system automatically extracted visual features using computerized tools and inferred the taxon with the help of its rule base. Although the idea was sound, the associated analysis is of limited use for several reasons. Firstly, images were obtained by digitizing scanning electron microscopy (SEM) photographs that were published in the *Atlas of Late Cretaceous Globotruncanids* (Robaszynski et al., 1996). These images were of carefully-selected high-quality specimens, i.e. holotypes, which are not representative of sediment samples. Secondly, the use of SEM as opposed to optical microscopy is time consuming and expensive, especially for foram picking. Thirdly, no evidence was given to justify the need for three views of each specimen, a task that is particularly complex with SEM. Finally, the researchers did not publish any performance measures of their system to date.

The Micropaleontology/Geophysiology group led by Prof. Thierstein in ETH Zurich has also been working on automated foram identification. The group has developed three fully-automatic systems based on artificial neural networks (ANNs) (Bollmann et al., 2004): (1) ALFA, which uses reflected-light microscopy; (2) COGNIS-Light, which uses transmitted-light microscopy; and (3) COGNIS-SEM, which uses SEM. In an experiment that involved 14 species, the SEM system correctly classified 82% of specimens in the data set, but 37% of specimens in the classified groups were incorrectly labelled. The performance of the transmitted-light system was tested on a binary problem of identifying a rare species (*F. profunda*). While this system correctly classified 93% of *F. profunda*, 80% of the classified group

consisted of non-*F. profunda* specimens. One cannot determine from the analysis whether these false positives were due to limitations of the images or of the ANN. Finally, the performance of the reflected-light system (ALFA) was not reported.

In terms of reflected-light microscopy, Dr. Beaufort et al. of the European Centre for Research and Teaching of the Geosciences and the Environment also developed an ANN-based system, called SYRACO 2 (Dollfus and Beaufort, 1999). This system operates on position-normalized 64-by-64 pixel images. The challenge of recognizing high level patterns from raw pixel values was delegated to a five-layer *fat* ANN. To improve performance, the system was extended with an iterative sequence of image transformations, which were governed by feedbacks obtained from a bank of six other ANNs (Beaufort and Dollfus, 2004). This version of SYRACO 2 achieved good performance in classifying 13 types of coccoliths. But the authors themselves commented on the difficulty in understanding its operation.

While the published analyses of COGNIS-Light and SYRACO 2 provide evidence that optical microscopy images contain sufficient information to identify calcareous microfossils, it is impossible to separate the information content of the digital representation from the capabilities of the ANN classifiers. Also, despite the excellent progress, computerized identification of forams remains an open problem, so much so that automatic foram picking has hardly, if at all, been considered in the literature. It is worthwhile, therefore, to fill in important gaps of previous studies, especially those that are relevant to non-ANN classifiers in development (Ranaweera et al., 2009), ones that may indicate a good way to automate foram picking.

This paper analyzes the suitability of digital images in foram representation by comparing image and particle-based classifications made by the same specialist, using specimens sampled from a site rich in forams. Image-based classification

was done twice: first, using one image of each specimen obtained without particle manipulation; and second, using the first image as well as another image obtained with particle manipulation. Hence, the paper also analyzes the impact of particle manipulation on classification. Finally, the paper investigates the impact of non-idealities in: specimen quality, as some particles were broken or had overgrowths; view side, as some particles were photographed from a side with few features; and image quality, as some images suffered from blurring or shadowing.

## 2 Materials and Methods

Specimens were obtained from a sediment core acquired at the tropical Pacific site 865 of the *Ocean Drilling Program*.<sup>1</sup> Samples were taken over 25 intervals along various depths of the B core from 104.15 to 135.85 meters below the sea floor (mbsf), and are of late Palaeocene age (Kelly et al., 1996).

Ten particles were randomly picked from the size fraction 250–300  $\mu\text{m}$  of each core sample, yielding a total of 250 specimens. Each particle was dropped onto an opaque glass slide from approximately 5 cm above the surface to acquire a naturally-stable position. An image was captured using a Pixelink PL-A622C microscopy camera mounted on a Zeiss Stemi-2000C microscope. This image is called the *default view* of the specimen. The particle was then manipulated and an *alternative view* was captured. The alternative view shows the most detailed side of the specimen if the default view did not show it. Otherwise, the alternative view shows the approximate opposite side of the default view. Particles were illuminated using a Zeiss KL-1500 fibre-optic light source, operating at a color temperature of 3,050 K, that was mounted at a fixed location.

---

<sup>1</sup> <http://www.iodp.org/>

A computer program was developed not only to capture these images but also to upload them to the *Microfossil Wiki*, a digital archive summarized below. All known information about each specimen (size fraction, depth below sea floor, core section, etc.) was stored in the wiki. After photography, the particles were physically archived on cardboard slides using water-based glue and suitable labels.

Thus, the 250 specimens were represented by 500 images and 250 particles. Each specimen was classified three times by a specialist—Dr. Bains, who has worked on such samples before (Bains et al., 1999)—in the following order: first, by inspecting the default view only; second, by inspecting the default and alternative views together; and third, by inspecting the particle with a binocular microscope.

For the image-based classifications, which were done online, the genus, species, state (unbroken or broken), overgrowths (without or with), view side (dorsal, edge, or ventral), and image quality (good, fair, or poor) were labelled for each specimen. The particle-based classifications were accomplished by the standard practice of manipulating the specimens and adjusting microscope focus and illumination. Noted attributes were the same, except for view side and image quality.

## 2.1 *Image-Based Classifications*

The Microfossil Wiki was developed to facilitate image-based classification and to disseminate the data used in this work.<sup>2</sup> The wiki is a dynamic web site built upon a relational database. It displays single or multiple images of each specimen along with a caption that describes each image. Sample images are given in Fig. 1.

Image captions have editable fields, which are highlighted when an authorized ed-

---

<sup>2</sup> <http://www.ece.ualberta.ca/~imagesci/microfossil/wiki>

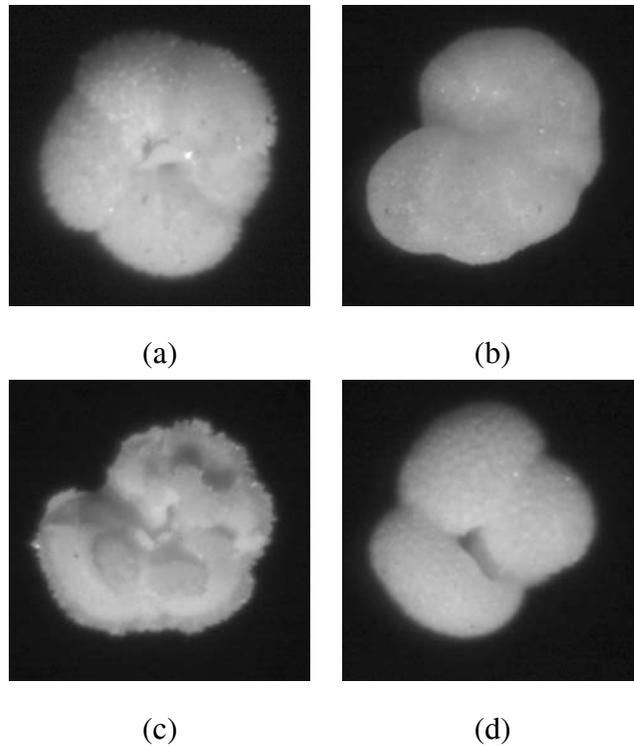


Fig. 1. Sample images from the Microfossil Wiki, showing default views of four genera: (a) *Acarinina*, (b) *Anomalinoidea*, (c) *Morozovella*, and (d) *Subbotina*.

itor logs in. For example, an editor may see the following caption for an image:

This is ?, in ? state ? overgrowths, of ?. It was obtained at 132.6 mbsf from site ODP 865 B in the size fraction 250 to 300  $\mu\text{m}$ . You are looking at its ? side (image quality: ?).

By choosing keywords from menus that pop up when the cursor touches the fields, i.e. the question marks above, the editor may revise the caption to say:

This is a benthic foram, in an unbroken state without (or with minor) overgrowths, of *Anomalinoidea*. It was obtained at 132.6 mbsf from site ODP 865 B in the size fraction 250 to 300  $\mu\text{m}$ . You are looking at its dorsal side (image quality: good).

The editor may leave any field unedited or revert an edited field to an unedited state, i.e. a question mark. Such fields are considered *unknown*.

Table 1

Genus and species counts of the specimens used in this analysis, according to particle-based classification. Planktic forams comprised 97% of the specimens.

Genus	Count	%
<i>Acarinina</i>	37	15.2
<i>Anomalinoides</i>	2	0.8
<i>Morozovella</i>	177	72.5
<i>Subbotina</i>	22	9.0
<i>Unknown</i>	6	2.5
<b>Total</b>	<b>244</b>	<b>100</b>

Species	Count	%
<i>Ac. mckannai</i>	14	5.7
<i>Ac. soldadoensis</i>	19	7.8
<i>M. acuta</i>	4	1.6
<i>M. subbotinae</i>	8	3.3
<i>M. velascoensis</i>	124	50.9
<i>Unknown</i>	75	30.7
<b>Total</b>	<b>244</b>	<b>100</b>

Data in the Microfossil Wiki may be searched using a similar natural language interface. Search results are displayed as images and captions that can be navigated. However, search results may also be downloaded in several formats, which may be imported into data analysis tools like Microsoft Excel and MATLAB.

## 2.2 Particle-Based Classifications

After the image-based classifications were completed, the physical archive was opened and particles were unglued and classified with a binocular microscope. Somewhere during this process, one particle was lost. Also, there were non-unique classifications for five other particles. Excluding these six, 244 particles were used in the subsequent analysis. Table 1 gives the genus and species counts of these specimens, and Table 2 breaks down the specimen quality.

Although the Microfossil Wiki was only used to make image-based classifications, it has been programmed to offer particle-based classifications as well. Therefore, all images and classifications described here are publically available.

Table 2

Breakdown of specimen quality, as determined by particle-based classification. A specimen may be in an unbroken or broken state and may be without or with overgrowths.

Quality	Unbroken	Broken	Unknown	Subtotal
W/o overgrowths	124	4	0	128
With overgrowths	79	16	0	95
Unknown	15	5	1	21
Subtotal	218	25	1	244

### 3 Results

Essentially, an image-based classification was taken to be correct if it agreed with the corresponding particle-based classification, assumed to be correct.

- Image-based 1 (IB-1) classifications were made using the default views only, which were the images captured without any manipulation of particles. They serve to assess how much useful information was available in the default views alone, which were relatively easy to obtain.
- Image-based 2 (IB-2) classifications were made using the default and alternative views together, which provided more information than the default view alone with minimal particle manipulation. Comparing IB-1 and IB-2 classifications helps to assess the value of particle manipulation.
- Random a priori (RAP) classifications were made statistically, according to the distributions of known classes in the particle-based (PB) classifications. RAP classifications establish a useful benchmark for IB classifications. To be effective, the latter must be better than the former.

All metrics that were computed are binomial proportions  $\hat{p}$  between 0% and 100% over a statistical sample of size  $n$ . Consequently, 95% confidence intervals were

also computed using the robust *Wilson score* method (Agresti and Coull, 1998). The endpoints of the interval are the  $p_0$  solutions of the quadratic equation:

$$(p_0 - \hat{p})^2 - z_{.05/2}^2 p_0(1 - p_0)/n = 0, \quad (1)$$

where 95% of the probability mass in the standard normal distribution lies within  $\pm z_{.05/2}$  of the mean, i.e.  $z_{.05/2} = 2$ . Confidence intervals are used to generalize an observation from the statistical sample, i.e. the 244 specimens, to the statistical population, i.e. all specimens of size 250–300  $\mu\text{m}$  in the 25 core samples.

An underlying assumption of the Wilson score method is that the outcome of each trial, e.g. the classification of a specimen, is statistically independent of other trials. While this is a reasonable assumption for IB classifications, it poses a problem for RAP classifications. With the latter, the same data is being used to define a classifier and to assess its performance. An approximate solution, which has been employed, is to substitute  $n/2$  for  $n$  in the above formula.

### 3.1 *Correct and Incorrect Classifications*

Four quantities were defined to measure overall success and failure in the IB classifications of genus and species. Although such classifications involve more than two *known* possibilities in the PB classifications, the correctness or incorrectness of an IB classification represents a binomial trial with a true or false outcome.

- The correct genus rate (CGR) was the proportion of specimens, with known genus by PB classification, that were correctly identified by IB classification:

$$CGR = \frac{\text{Number correctly identified}}{\text{Number with known genus}}. \quad (2)$$

- The incorrect genus rate (IGR) was the proportion of specimens, with known genus by PB classification, that were incorrectly identified by IB classification:

$$IGR = \frac{\text{Number incorrectly identified}}{\text{Number with known genus}}. \quad (3)$$

- The correct and incorrect species rates (CSR and ISR) were similarly defined.

The CGR and IGR (or CSR and ISR) may not sum to one. This is because the wiki allowed the genus (or species) to be left unclassified, i.e. classified as *unknown*, if it was impossible to make a judgement based on the information in the image(s).

For the RAP classifications, the CGR and IGR were computed as follows:

$$CGR = \sum_{i=1}^N p_i^2, \quad (4)$$

$$IGR = 1 - CGR. \quad (5)$$

Here,  $N$  represents the number of known genera, by PB classification, and  $p_i$  represents the proportion of the  $i^{th}$  genus amongst specimens with known genera, also by PB classification. RAP classifications are analogous to casting an  $N$ -sided die, where each side is weighted by  $p_i$ , to decide the genus of a specimen. Unlike with IB classifications, the CGR and IGR of RAP classifications always sum to one. This is because all sides of the die represent known genera. The CSR and ISR of RAP classifications were similarly computed.

### 3.1.1 Genus and Species Rates

Fig. 2 gives the overall correct and incorrect rates of RAP and IB classifications. With respect to the RAP benchmarks, images were sufficient to realize low in-

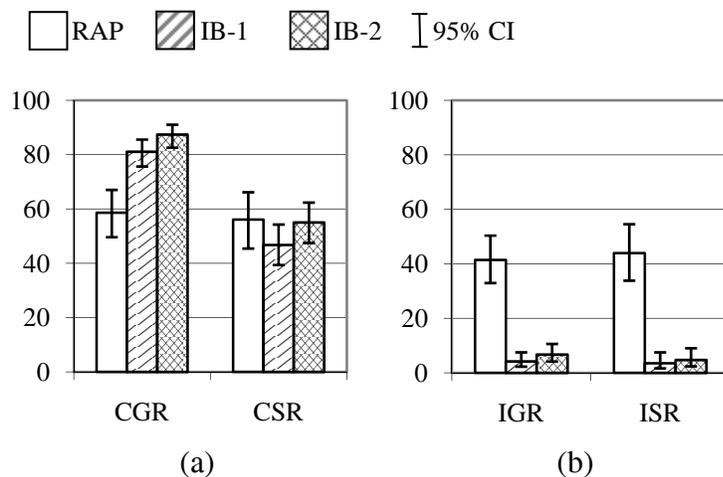


Fig. 2. Overall performance: (a) correct genus and species rates of the RAP, IB-1, and IB-2 classifications; and (b) incorrect genus and species rates of the same. Information in the default view was sufficient to identify the genus of 81% of the specimens but the species of only 47%. Incorrect classification rates were 4% in both cases.

correct genus and species rates and high correct genus rates. Essentially, it was impossible to classify species from one or two images more often than it was impossible to classify genus. These conclusions may be generalized to the statistical population, as they are supported by non-overlapping confidence intervals.

Comparing IB-1 and IB-2 classifications, the availability of an alternative view led to an increase in the correct rates, which is desirable, but also an increase in the incorrect rates, which is undesirable. At the species level, the alternative view was needed to reach the correct rate of RAP classifications. At the genus level, the relative improvement afforded by the alternative view was less significant. Overall, the default view alone provided almost as much taxonomic information as the default and alternative views together, at least for the statistical sample.

### 3.1.2 Impact of Specimen Quality

Fig. 3 shows the impact of specimen quality, as determined by PB classifications, on the correct and incorrect rates of IB-1 classifications. Rates were calculated sep-

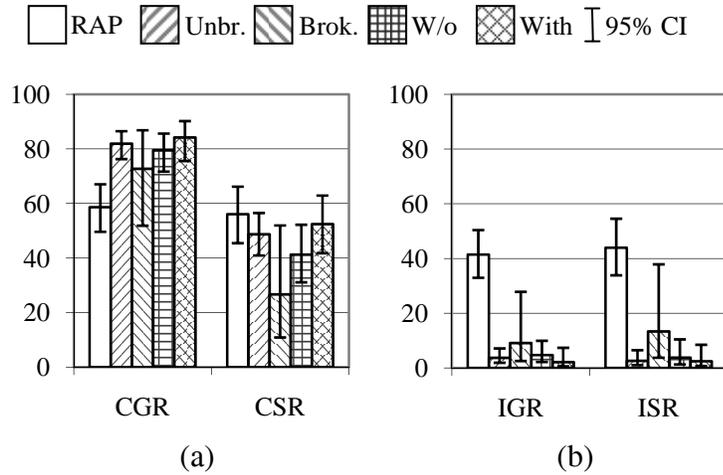


Fig. 3. Impact of specimen quality on performance: (a) correct genus and species rates of the RAP and IB-1 classifications, with the latter conditional on particle state (unbroken or broken) and overgrowths (without or with); and (b) incorrect genus and species rates of the same. Unbroken specimens, and those with overgrowths, provided more information to identify species. Specimen quality had less impact on genus identification.

arately for particles in an unbroken or broken state, as well as for those without and with overgrowths. Overall rates of the RAP classifications are given for reference. Wide confidence intervals occur for the broken-particle rates because few specimens had that quality (Table 2). As specimens were sieved into the 250–300  $\mu\text{m}$  size fraction before picking, small broken particles were excluded.

Broken particles correlated with a decrease and increase in correct and incorrect rates, respectively, both of which are undesirable. To a lesser degree, particles with overgrowths correlated with an increase and decrease in correct and incorrect rates, respectively, both of which are desirable. These trends are less pronounced at the genus level. While it is not surprising that breakages complicate IB-1 classification, it is surprising that overgrowths facilitate it. However, these observations cannot be generalized to the statistical population with high confidence.

What may be generalized to the statistical population with at least 90%, i.e. 95% squared, confidence is the following. At the genus level, with respect to the RAP benchmarks, default views were sufficient to realize low incorrect rates irrespective

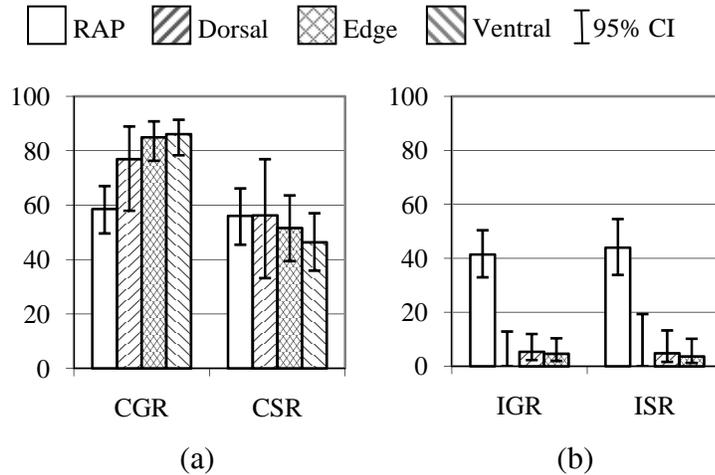


Fig. 4. Impact of view side on performance: (a) correct genus and species rates of the RAP and IB-1 classifications, with the latter conditional on dorsal, edge, and ventral sides; and (b) incorrect genus and species rates of the same. Edge and ventral sides provided more information to identify genus than the dorsal side. However, details in the edge and ventral sides were often blurred, which reduced the capacity to identify species.

of particle state and overgrowths. They were also sufficient to realize high correct rates irrespective of overgrowths. At the species level, the same images were sufficient to realize low incorrect rates irrespective of overgrowths.

### 3.1.3 Impact of View Side

One may analyze the impact on performance of dorsal, ventral, or edge sides when limited to the default view. Fig. 4 shows the correct and incorrect rates of RAP and IB-1 classifications, where rates of the latter were calculated separately for each side. At the genus level, images of edge or ventral sides were sufficient to realize high correct rates. At the genus or species level, images of any side were sufficient to realize low incorrect rates. These observations may be generalized to the statistical population with over 90% confidence.

In the statistical sample, correct genus rates of IB-1 classifications were very close for the edge and ventral sides but somewhat lower for the dorsal side. This is because the dorsal side of many specimens was relatively uniform and contained

fewer features. Nonetheless, as indicated by the broad confidence interval, few specimens (11%) presented the dorsal side in their default view. This is likely because a specimen has to settle on a rather unstable ventral side in order to show the dorsal side to the camera. Consequently, the majority (83%) landed showing either the edge or ventral side—the sides that gave better correct genus rates.

Correct species rates of IB-1 classifications had the opposite trend in the statistical sample—the less informative dorsal side facilitated classification. This slight trend may be understood by considering image quality. Whereas a small proportion (3.7%) of the dorsal images were rated poor, much higher proportions (37.2% and 18.5% respectively) of the edge and ventral images were rated poor. The latter were more likely to be blurred by the depth-of-field limits of optical microscopy, due to the non-planar geometry of edge and ventral sides. It was evidently more difficult to classify species than genus when images were not clear.

### *3.1.4 Impact of Image Quality*

Image quality was classified based on the ability to see clearly distinguishable specimen features. This was primarily a measure of image sharpness but was also a measure of illumination suitability. Fig. 5 breaks down image quality of the default view in terms of specimen quality and view side.

Specimen quality had little bearing on image quality. Evidently, they are different measures. Irrespective of particle state and overgrowths, about 25%, 50%, and 25% of default-view images were rated good, fair, and poor, respectively. However, view side and image quality were correlated. For example, the relative decrease in good ratings and increase in poor ratings for highly non-planar edge sides may be generalized to the statistical population with 90% confidence.

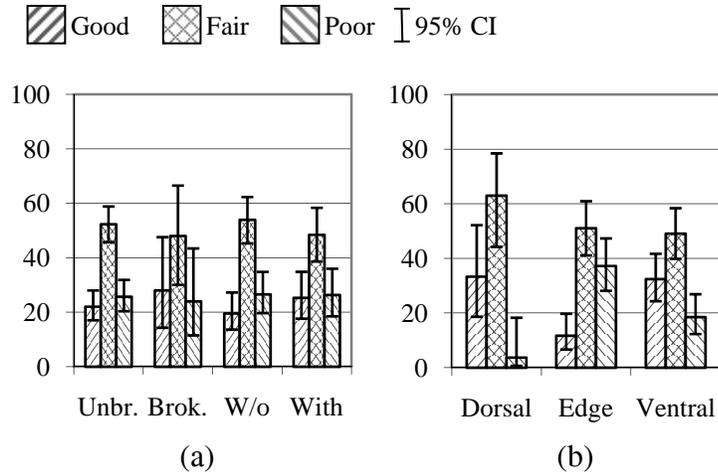


Fig. 5. Image quality of the default view, broken down in terms of: (a) specimen quality; and (b) view side. Ventral sides were less often rated fair and more often rated poor, when compared to dorsal sides. Edge sides were less often rated good and more often rated poor, when compared to dorsal or ventral sides. Particle state (unbroken or broken) and overgrowths (without or with) had much less influence on image quality.

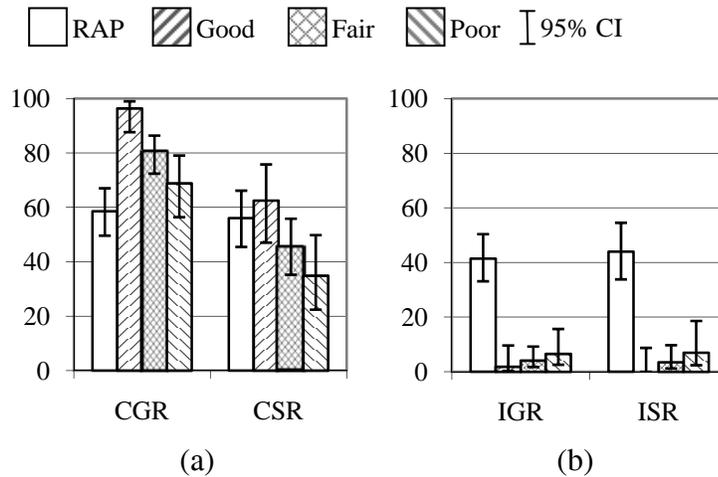


Fig. 6. Impact of image quality on performance: (a) correct genus and species rates of the RAP and IB-1 classifications, with the latter conditional on good, fair, and poor images; and (b) incorrect genus and species rates of the same. Good images provided sufficient information to classify genus and species correctly 96% and 63% of the time, respectively, with incorrect classification rates of 2% and 0%, respectively.

While it was correlated with view side, and could also be specified from images alone, image quality was still a different measure. Fig. 6 shows the correct and incorrect rates of RAP and IB-1 classifications, where rates of the latter were calculated separately for good, fair, and poor images. View side (Fig. 4) had less of an impact on correct genus and species rates than did image quality (Fig. 6).

Moreover, the following observations may be generalized to the statistical population with about 90% confidence or better. The correct genus rate was higher with good images than with either fair or poor images. The correct species rate was higher with good images than with poor images. Incorrect genus and species rates of IB-1 classifications were low, with respect to the RAP benchmarks, irrespective of image quality. In other words, specimens with better images were classified more often and such classifications turned out to be correct, not incorrect.

#### **4 Discussion**

The problems of computerized identification and digital representation of forams were decoupled in this paper. Specimens have a physical representation, which micropaleontologists trust, but can also have a digital representation, which may facilitate various micropaleontological studies. Analyses like the one introduced here provide a means to build trust in the digital representation, and to examine how the digital representation may be revised to deserve more trust.

One can envision foram studies that rely on the suitability of digital representation but do not require the fully-automatic identification found in the literature. COGNIS-Light and SYRACO 2, the state of the art for calcareous microfossil identification with optical microscopy, are complex ANN systems that operate on simple digital images. As ANN systems require training on a labelled data set that contains many examples of each class expected, it is hard to see how the approach could be applied to benthic forams, where the number of extant species is large.

A computer-aided system in development, which follows from this work, does not require training (Ranaweera et al., 2009). It is a relatively simple system that op-

erates on default-view images. While preliminary results are promising, the main limitation seems to be the digital representation. The same system indicates a way towards automatic foram picking, one that does not require automatic foram identification but that does benefit from a good digital representation.

The Microfossil Wiki was used here to test two types of digital representations: (1) default-view images alone; and (2) default and alternative-view images together. Other representations are in development that also employ optical microscopy. To further build trust in a digital representation of forams, expensive kinds of microscopy may prove essential. But an analysis like the one introduced here would still be needed to test the suitability of these future representations.

The digital representations studied here were limited by a low depth of field of the optical system. The depth of field may be improved by using a microscope with a smaller aperture, and compensating for the dimmer signal with a longer exposure. The camera must have low noise under these conditions. Alternately, one may use techniques such as *blur estimation* to automatically identify poor-quality images (Tong et al., 2004). Specimens in these images could then be photographed under better conditions or may be excluded if immaterial to the end application.

## **5 Conclusion**

This paper analyzed the efficacy and limitations of optical microscopy images in foram representation. The analysis involved 250 specimens from the B core acquired at site 865 (tropical Pacific) of the Ocean Drilling Program. Ten specimens were picked randomly from each of 25 depths between 104.15 and 135.85 mbsf. Each specimen was photographed twice to obtain default and alternative views,

where only the latter required particle manipulation. The specimens were classified three times by the same specialist: first, by inspecting only default-view images; second, by inspecting both default and alternative-view images; and third, by inspecting archived particles with a binocular microscope.

Image-based (IB) and particle-based (PB) classifications were compared. Correct and incorrect genus and species rates were defined to analyze overall performance. Statistical benchmarks were defined, using taxon proportions from PB classifications, to establish success or failure of IB classifications. Moreover, 95% confidence intervals were computed so that observations could be made not only for the statistical sample but also for the statistical population.

In terms of the statistical sample, default-view images sufficed to correctly identify genus and species 81% and 47% of the time, respectively. With respect to statistical benchmarks, the former was a success but the latter was a failure. However, incorrect classifications were about 4% for both genus and species, which were both successes. The primary difference between genus and species classification is that it was often impossible to classify species from the information visible in images. While the availability of alternative-view images led to improvements, they were small. Manipulation did not make a worthwhile difference.

The impact on IB classification of specimen quality, view side, and image quality was also analyzed, when limited to the default view. Of these, image quality had the most significant impact. The subset of specimens whose images were rated good, at the time of taxon classification, had correct genus and species rates of 96% and 63%, respectively. Image quality did not correlate with particle state (unbroken or broken) or overgrowths (without or with) but did correlate with view side, although they were not the same. In general, planar dorsal sides had the best ratings, highly

non-planar edge sides had the worst ratings, and non-planar ventral sides were in between. Image quality was primarily a measure of sharpness.

In terms of the statistical population, default-view images were sufficient with well over 90% confidence to obtain low incorrect genus and species rates, irrespective of particle overgrowths, view side, and image quality. The same may be said with about 90% confidence for particle state (unbroken or broken). Default-view images were also sufficient with over 90% confidence to obtain high correct genus rates, irrespective of particle overgrowths, for unbroken particles, for edge or ventral sides, and for good or fair images. In summary, default-view images provided trustworthy representations of variable-grade forams, primarily for not misidentifying genus or species and secondarily for identifying genus.

To facilitate this work, an online database called the Microfossil Wiki, which is searchable and editable, was developed. Data referred to in this paper, called the *Nov. 2006 (a)* batch, may be accessed freely via the wiki, in both human and machine-readable formats. This entails 500 images and 750 classifications of 250 specimens. The data is made available to stimulate further research in computerized identification and digital representation of forams.

## **Acknowledgments**

The authors are grateful to Alberta Ingenuity for support. They also thank: Adam Harrison, Anna Pienkowski-Furze, and Mark Furze of the University of Alberta, Canada; Richard Hesketh and Haydon Bailey of Network Stratigraphic Consulting Ltd, UK; Richard Norris of the Scripps Institution of Oceanography, USA; Øyvind Hammer of the Oslo Geological Museum, Norway; and anonymous reviewers.

## References

- Alan Agresti and Brent A. Coull. Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician*, 52(2):119–126, May 1998.
- Santo Bains, Richard M. Corfeld, and Richard D. Norris. Mechanisms of climate warming at the end of the paleocene. *Science*, 285:724–727, 1999.
- L. Beaufort and D. Dollfus. Automatic recognition of coccoliths by dynamical neural networks. *Marine Micropaleontology*, 51(1):57–73, 2004.
- J. Bollmann, P.S. Quinn, M. Vela, B. Brabec, S. Brechner, M.Y. Corts, H. Hilbrecht, D.N. Schmidt, R. Schiebel, and H.R. Thierstein. Automated particle analysis: Calcareous microfossils. *Image Analysis, Sediments and Paleoenvironments*, 7: 229–252, 2004.
- D. Dollfus and L. Beaufort. Fat neural network for recognition of position-normalised objects. *Neural Networks*, 12(3):553–560, 1999.
- D.C. Kelly, T.J. Bralower, J.C. Zachos, I.P. Silva, and E. Thomas. Rapid diversification of planktonic foraminifera in the tropical pacific (odp site 865b) during the late paleocene thermal maximum. *Geology*, 24:423–429, 1996.
- S. Liu, M. Thonnat, and M. Berthod. Automatic classification of planktonic foraminifera by a knowledge-based system. In *Artificial Intelligence for Applications, Proceedings of the Tenth Conference on*, pages 358–364, March 1994.
- Kamal Ranaweera, Adam P. Harrison, Santo Bains, and Dileepan Joseph. Feasibility of computer-aided identification of foraminiferal tests. *Submitted to Marine Micropaleontology*, March 2009.
- F. Robaszynski, M. Caron, J.M. Gonzalez Donoso, and A.A.H. Wonders. *Atlas of Late Cretaceous Globotruncanids*, volume 26, pages 113–123. 1996.
- Hanghang Tong, Mingjing Li, Hongjiang Zhang, and Changshui Zhang. Blur de-

tection for digital images using wavelet transform. In *IEEE International Conference on Multimedia and Expo*, volume 1, pages 27–30. IEEE, June 2004.

Shan Yu, Pierre Saint-Marc, Monique Thonnat, and Mark Berthod. Feasibility study of automatic identification of planktic foraminifera by computer vision. *Journal of Foraminiferal Research*, 26(2):113–123, April 1996.