Clim. Past Discuss.,
https://doi.org/10.5194/cp-2020-76-RC1, 2020

Climate of the Past Discussions

Open Access

Interactive comment on "A new automated radiolarian image acquisition, stacking, processing, segmentation, and identification workflow" *by* Martin Tetard et al.

David Lazarus (Referee)

david.lazarus@mfn.berlin

Received and published: 21 July 2020

General comments

As the authors note, human collection of routine occurrence data for radiolarians or other organisms is time consuming, requires rare, expensive expert workers, and suffers from inconsistencies between data collectors. Automated collection of occurrence data is likely to prove revolutionary to fields such as micropaleontology, where the vast numbers of specimens and species preserved in the fossil record means that the quantity and quality of the data can be expanded by orders of magnitude. This is a rapidly developing subject and papers are appearing in quick succession. Most work in mi-

Printer-friendly version

Discussion paper

cropaleontology however so far has been on the most intensely studied groups - pollen, planktonic foraminifera and coccolithophores. Only a very few studies have been done on radiolarians, despite their importance to carbon cycling, polar biogeochronology and evolution research. This ms is, so far as I know, the first attempt to essentially throw an AI system at an entire assemblage without a human pre-selection process of the input images. It is a very valuable contribution to see how well this works, and where bottlenecks or barriers arise.

One very important advantage of this as a fully automated system, including primary image acquisition, is that, having identified common categories and trained the system to identify them, future image acquisition can concentrate on the less common categories, resulting in a great improvement int the efficiency of the taxonomic specialist who should be examining and tagging only unknown/poorly documented-rare forms. They might wish to point this out more clearly in the ms.

They post their data and code, or pre-prints very openly - kudos. As Marchant et al. submitted gives the technical component of the AI system I will not comment further here on it.

The ms, and the study itself, are well done and the ms is worthy of publication with only moderate revisions. Nonetheless there are several things that should be improved prior to acceptance.

Critique

Citation

The ms does not cite some other prior attempts to automate radiolarian identification, should try to cite some of these since there are so few of them: e.g. Apostol, L. A., Ma ÌĄrquez, E., Gasmen, P., and Solano, G. (2016); Keçeli, A. S., Kaya, A., and Keçeli, S. U. (2017).

Images and Taxonomy

Stacked glassy images are used in this project as the basis for identification. These are arguably the best single image to use if a study is done with just one image per specimen (not actually a requirement for this type of work). However, stacked images often do not show important interior characters (for example in actinommids, pylonids and plagiacanthids, and in many instances in other families as well). These interior characters are important for species identification in many taxa and I found the images sometimes to be frustrating to interpret as an expert in the taxonomy of these forms. The standard method of imaging and presentation in research for this material is to show a manually focussed set of a few unstacked images in transmitted light. Possibly there is a way to use the raw images better? Also, precisely because such stacked images are uncommon in the field it is not clear how easy it will be for community members to add to their open image database, despite the laudable call for contributions. Very few researchers for that matter have access to the, for micropaleontologic standards, complex, expensive equipment used by the authors in this study, tho perhaps only manually generated images are needed as input.

The images are only ca 300 pixels in size. This is a bit marginal. While sufficent for most features, taxonomically important small features can get lost (bladed vs cylindrical spines, etc). Computation costs increase with image size but if possible I would suggest nearly doubling the image resolution for future work, particularly if the database used is going to be promoted as a standard for future contributions.

The cited 17K image database is significantly smaller in numbers of species images. There are non radiolarian broad categories like 'spicule' - 8 categories, ca 8K images, plus supraspecific spp group categories - nearly 20 named,ca 4K images. The number of tagged species images is thus well under 10K. This distinction should be noted clearly in the ms.

For many taxa images have multiple image types, e.g. Acrosphaera spinosa specimens largely covered with some sort of milky bubble (preparation or image stacking artefact?) - 20 of 60 images. This problem is fairly common, seen in many folders.

This particular image problem does not appear in routine sample preparation of similar materials and should be explained, as also the effect on identification accuracy.

There are many, mostly minor issues with the taxonomic image database, tho some of these do not appear in the ms itself due to dropping rare image categories in the analysis.

There are really too many incorrect names being used in the ms and supporting image database. While not a problem for testing (the name tags as such have no effect on the system) these should be corrected. I have not gone through all 100 folders but based on a sampling of these I note the following. 'Strichocorys spp' - no such genus, content appears to be Phormostichoartus pitomorphus, Theocalyptra davisiana - correct name (since 30 years) is Cycladophora davisiana, Calyptra cervus instead of intended? name Corocalyptra cervus, tho recommended name is Eucecryphalus cervus - and see also below, Pseudodictyophimus gracilipes, kilmari, not Dictyophimus (again, since decades) - which importantly changes the family assignment.

There are also some typos such as Stylotractus instead of Stylatractus universus. Lastly there is at least one instance of possible oversplitting - i. e. Dictyocoryne truncatum vs Euchitonia triangulum. What is the difference?

There are also a certain number of specimens in the individual taxon folders that are not con-specific. A brief sampling yields:

image name given correct name total images in category

14875 Cyrtopera langucula Artopilium undulatum 6

11233 Anthocyrtidium ophirense unknown but pores much too large to be conspecific 20

01235 Zygocircus productus Zygocircus piscicaudatus 67

13329 Amphisphaera sp. B appears conspecific w. some specimens labeled as Druppatractus irregularis 21/28

mult. Calyptra cervus multiple Cycladophora species including [u1486-..] Cycladophora cabrilloensis

The large majority of the identifications (so far as the image quality allows) appear to be correct as monospecific classes, even if sometimes the name for the class are not. An attempt to provide a standard set of names (provisional of course as taxonomy is always being revised) was given by a group of taxonomists in Lazarus et al. 2015. I sugggest for the sake of data standardization that they use these, or at least provide a taxonomic appendix where they explain any variant usage.

Performance measurements

The time has come, the Walrus said, To talk of many things: Of shoes and ships and sealing-wax, Of cabbages and kings (Lewis Carroll, Through the Looking Glass, 1871)

Radiolarians are to a highly unusual degree morphologically diverse. There are at least a dozen topologically highly distinct Baupläne alone in living/late Neogene assemblages. Distinctions between these broad morphologic groups are trivially easy. This is quite in contrast to most other clades of organisms, and in particular to planktonic foraminifera, which were used by this research group to initially develop their algorithms and work flows. Planktonic foraminifera are morphologically very conservative (once described to me -by a foram specialist- as 'basically just popcorn'), and can be considered for an imaging system as a single broader category for analysis purposes. For radiolarians, it is really not very informative to know that the system can distinguish between forms as radically different as, well, cabbages and kings. The true test of performance is its capacity to accurately identify and distinguish between species within topologic/taxonomically similar groupings, such as within radiolarian families/subfamilies. This is not only for purely taxonomic reasons, but also for the utitily of a system in applied research. Radiolarians encode ecologic-environmental information almost entirely at the level of individual species. Attempts to use genus or family level taxa as proxies in applied paleoenvironmental research have yielded almost no useful signals. Geologic age in the Cenozoic is partially recoverable at the genus level, but at a resolution so poor as to make it uninteresting for actual use.

The ms therefore needs to clearly separate the performance of the system in distinguishing between morphologically/taxonomically similar forms and the ability to distinguish extremely dissimilar forms. I suggest in both the statistical analyses of output and in the creation/ organisation of the figures, and the image database for download, that a clustering to higher categories is done, e.g. radiolarian orders and families, plus 'other' for non radiolarian categories such as particles, diatoms or background. The results for pairings of related taxa such as Lophophaena hispida and Peromelissa phalacra, both within the lophophaenid subfamily of Plagiacanthidae (performance values ca 70%) suggest that the accuracy for applied uses may be significantly lower than the current bulk statistics suggest.

It is also important to report separately performance in identifying radiolarian species vs identifying broad categories such as 'spp.', diatoms, particles etc.

Lastly, when errors Are made, the nature of these is significant. It is hard to understand how this system could mis-identify a Pseudodictyophimus gracilipes [incorrectly named Dictyophimus gracilipes in the ms] with Hexacontium spp. - these taxa are in different orders, and have [to a taxonomic expert] fully different morphologies. Some sort of statistic giving not just the error rate, but the type of error - misclassified as to species in same family, different family in order, or different order should be given.

Completeness of taxonomic/morphologic coverage

The authors have clearly made an effort to look at the entire assemblage of radiolarians, which is perhaps the most distinctive, laudable and novel aspect of this study. It must however be noted that the number of actual examined species is less than 80, while radiolarian diversity in tropics-subtropics is ca 500 (just in the sediments, not counting those, relatively few, species in the plankton that do not preserve). Thus only about 15% of the diversity in these assemblages has been incorporated into the study. Adding a larger percentage of the species is a stated goal of their project and quite correct. However, as more species are added, the number of closely related species pairings will increase, in relation to comparisons between distantly related forms. This is likely to have a negative effect on the performance of the system, as accuracy in similar pairings appears to be fairly low at present. This may not affect using the system for classic assemblage based proxies of paleoenvironmental conditions as these can be based on relatively few, selected species or even species groups. There is not much demand however at the for this type of work, as it has been largely displaced by geochemical methods. Possibly having a cheap system to generate the data will revive it, but I am somewhat skeptical. Biostratigraphy however remains important, along with a variety of emerging themes related to evolution and biodiversity. These studies though need to use a larger fraction of the assemblage, distinguish closely related forms, and/or include rarer species. (Indeed it seems to be an instance of Murphy's Law that important biostratigraphic markers are so hard to find in many slides...). The usability of this system will only become apparent when it has grown to include more taxa, including many closely related and rare forms. The ms should make these limitations clear. There is no general answer to what level of accuracy is 'adequate', but I would suggest that for many biostratigraphic and biodiversity studies error rates should be closer to 1% than 10%, which may prove quite challenging for AI systems to achieve. Alternatively analytic methods in these fields will have to be revised to handle much noisier data.

There is an additional issue in the adequacy of counting only a few hundred individuals to represent an assemblage of radiolarians. This is adequate only for the rather small fraction of species (usually <10% of the species richness) that are present at several percent abundance in the assemblage, as a closer reading of the short paper the authors cite (Fatela and Taborda) would reveal. While a cut-off of several percent is indeed frequently used in paleoenvironmental proxy studies, this is a seriously inadequate degree of coverage for either biostratigraphy or evolution/biodiversity research.

Printer-friendly version

Discussion paper

Indeed, this problem is indirectly illustrated in the inadequate numbers of specimens for many species in their training-test image sets. There is in fact a very large body of sophisticated literature in ecology on determining the adequacy of sample sizes for different degrees of assemblage diversity and desired completeness of the resulting sample. See as starting points Chao et al. (2020) Ecol. Res.; Dornelas et al. (2012) Proc. R.Soc. B, and the brief discussion related to radiolarian assemblages in Lazarus et al. (2018) PeerJ. For radiolarians, in many cases the appropriate sample size is several thousand specimens.

Sample coverage

Detailed information is given for this in the SOM, but essentially all material is from a single location in the western equatorial Pacific. I miss a discussion, or at least a disclaimer, of how geographic variation in morphology, or variation over time in lineages might affect the system's performance by blurring between species distinctions. It would also be nice to know in more detail the ages of the samples and the sources of the age information.

New sample preparation method

In this section a new variant of a coverslip holder is described. Although the goal of making slides with very few individuals seems to me to somewhat quixotic given the sample adequacy issues mentioned above, the idea of a custom designed holder that can be manufactured by 3-D printing is novel and is, adapted perhaps to full size cover slips, a useful addition to the literature. The chemical and other preparation steps are fairly standard and, though important to include, could be moved to the SOM.

Figures and Tables

The confusion matrix (fig. 4) is useful but a much more readable table listing each species, numbers/percents correctly and incorrectly classified and the top 3 error categories they were assigned to would be very helpful. I spent too much time scrolling figure 4 around my screen.

Tiddles

I think the citation to Lazarus et al. 2015 line 23 should be Lazarus 2005 while line 235 should be to Lazarus et al. 2015 - they have inverted these.

Interactive comment on Clim. Past Discuss., https://doi.org/10.5194/cp-2020-76, 2020.