Dear Authors,
I carefully read your responses to the reviewer’s comments and your revised manuscript. You addressed most of the reviewer’s comments and you improved the manuscript. Thank you.
However, there are a few points that I would like to you to work on. Most of them are related to open questions of the reviewers in their comments. While you answer very well to these questions in your point to point response, too often, you do not include that information in the text of your manuscript. I think you should include such information because, if the reviewer asked that question, it is most probably because the information was missing or should be clarified in the manuscript. Here is the list of things I would like you to consider in a revised version of your manuscript.

Response: Thank you for your comment. We integrated the requested changes in the manuscript, and provide a second markup version.

• I suggest that you add the explanation provided to Lazarus comment’s about the glue somewhere in the text, maybe in section 2.3

Response: A paragraph about this was added in second 2.2 when discussing about the choice of the glue, on lines 166 to 169.

• I also suggest to add on line 286 : “(...) and to send any suggestion to improve the taxonomical framework of the database” or something similar.

Response: This was added as suggested on lines 301 to 302.

• I suggest you add something in the discussion about the usefulness of improving the identification to a level that is useful for paleoenvironmental research.

Response: A sentence was added on lines 317 to 322.

• While your answer to Lazarus comment at bottom of your page 5 (your point by point response to reviewers), I still think you should add a comment about the usability of this system and his limitation in the text of the MS.

Response: This is now discussed in section 3.2, on lines 317 to 326 and in the conclusions, on lines 399 to 409.

• Again, it might be interesting to add a comment in the text about the appropriate sample size for radiolarian in terms of number of specimens to be counted.

Response: A small description was added to lines 162 to 165

• Sample coverage: I suggest that you add a disclaimer about how geographic variation in morphology, or variation over time in lineages might affect the system’s performance by blurring between species distinctions. I understand that you are going to detail this in an upcoming study, but I also acknowledge that the reviewer has a good point here.

Response: A paragraph was added at the end of section 3.2, on lines 322 to 326.

• Reviewer Lazarus suggested to move some of the preparation method to SOM. Why do you think it is useful to keep it in the main text body? Maybe you have very good reasons to keep here, but please provide an answer.

Response: Even though Lazarus suggest that part of the preparation method could be moved to supplementary material, as it is relatively standard, some part of the preparation are specific to the new workflow, and moreover, as the paper was change to a technical note, we would like to keep the preparation procedure and technical description in the main text, instead of segmenting it and put a part of it in the supplementary material.
• It would be nice to mention in the text of the MS that autostacking does not affect correct identification, even for delicate specimens such as the one mentioned by reviewer Friesenhagen (something more than point 3 of your conclusion).

Response: This was added to section 2.4, on lines 191 to 192.

• Once again, you answer the question of reviewer Friesenhagen about the ontogenetic stages, but you do not add anything about it in the MS. I invite you to explain somewhere that ontogenetic stages are visible in numerous classes and, when they are sufficiently imaged, can be distinguished, for instance in section 2.3

Response: This section was added in section 3.1, on line 264 to 266.

• Friesenhagen asks: “What is the procedure for (intact) specimens which extend over the borders of the 324 FOV and are parted/bisected? (…)”. You should include your answer in the manuscript.

Response: This sentence was added in the text lines 194 to 196.

• Fig.1 Please add a scale.

Response: Scale bars were added regarding important measurements on each of the designs.

• Fig. 2 There is some shading on the legend of each panel that looks unnecessary to me, making the letters less clear.

Response: The shadings were removed from Figure 2.

• L 213: What is a ResNet50? please add a reference to resnet50 topology.

Response: References were added to the text and a small description, on lines 223 to 224.

• L 277 and following: I guess it should read “(…) then trained to be recognised with a current overall precision accuracy of just above 90 % (90.1 %) over every class. The average precision is above 85 % (85.6 %) and the (…)”.

Response: The sentence was corrected in the text.

• Fig 4. Caption: add that the figure is also available as an Excel spreadsheet in supplement.

Response: A sentence was added to the Figure 4.

• L 305 and following: Fig. 6 should be changed to Fig. 5

Response: Figure 6 was corrected to Figure 5 everywhere in the text.

• L 485. Marchant at al should be accepted, or in press, but please make it consistent throughout the MS.

Response: Marchant et al. is currently in press, this was corrected in the text.
• Section Biostratigraphy: I’m puzzled that no data or figure supports this section. Where is the Excel sheet that you are referring too?

Response: As we are afraid that the excel sheet is not directly usable for biostratigraphic reconstruction, we decided to remove it. We also emended the Section “Biostratigraphy” to only discuss about the interest of this method for biostratigraphic studies.

• Figure A1 is not called for in the text. The caption of this figure should contain some information about the samples

Response: Figure A1 was referred to Appendix A in the text, this was corrected. Caption was emended, and Appendix B and Appendix C were renamed Table A1 and Table A2 in the text.

• I suggest you also refer to your web site and make more publicity about its content. It is quite nice!

Response: More references to the website were added in the text (from 1 to 4 references to it now). Let us know if it’s too much.

• L 395 (competing interest) : I think you should declare that one co-author, Luc Beaufort, is an associate editor of Climate of the Past.

Response: A sentence was added to the competing interests section.

I think these points can be very easily addressed, and I’m therefore looking forward to your revised manuscript.
Best Regards
Pierre Francus, associate editor

Response: Thank you very much for your comments, we hope that the manuscript it now worth publishing,

Martin Tetard and co-authors.
A NEW AUTOMATED RADIOLARIAN IMAGE ACQUISITION, STACKING, PROCESSING, SEGMENTATION, AND IDENTIFICATION WORKFLOW

Martin Tetard¹, Ross Marchant¹,², Giuseppe Cortese³, Yves Gally¹, Thibault de Garidel-Thoron¹, and Luc Beaufort¹

¹Aix Marseille Univ, CNRS, IRD, Coll France, INRAE, CEREGE, Aix-en-Provence, France.
²Present address: School of Electrical Engineering and Robotics, Queensland University of Technology, Brisbane, Australia.
³GNS Science, Lower Hutt, New Zealand.

Correspondence: M. Tetard (tetard@cerege.fr)

Abstract. Identification of microfossils is usually done by expert taxonomists and requires time and a significant amount of systematic knowledge developed over many years. These studies require manual identification of numerous specimens in many samples under a microscope, which is very tedious and time consuming. Furthermore, identification may differ between operators, biasing reproducibility. Recent technological advances in image acquisition, processing, and recognition now enable automated procedures for this process, from microscope image acquisition to taxonomic identification.

A new workflow was developed for automated radiolarian image acquisition, stacking, processing, segmentation, and identification. The protocol includes a newly proposed methodology for preparing radiolarian microscopic slides. We mount 8 samples per slide, using a recently developed 3D-printed decanter that enable the random and uniform settling of particles, and minimise the loss of material. Once ready, slides are automatically imaged using a transmitted light microscope. About 4000 specimens per slide (500 per sample) are captured in digital images which include stacking techniques to improve their focus and sharpness. Automated image processing and segmentation is then performed using a custom plugin developed for the ImageJ software. Each individual radiolarian image is automatically classified by a convolutional neural network (CNN) trained on a Neogene to Quaternary radiolarian database (currently 21,746 images, corresponding to 132 classes) using the software, ParticleTrieur.

The trained CNN has an overall accuracy of about 90%. The whole procedure, including the image acquisition, stacking, processing, segmentation and recognition, is entirely automated via a LabVIEW interface, and takes approximately 1 hour per sample. Census data count and classified radiolarian images are then automatically exported and saved. This new workflow paves the way for the analysis of long-term, radiolarian-based palaeoclimatic records from siliceous remains-bearing samples.

Copyright statement. TEXT
1 Introduction

The term radiolarians currently refers to the polycystine radiolarian orders Spumellaria and Nassellaria, whose shell is made of opaline silica, relatively well preserved in the fossil record by comparison with the Acantharia and Phaeodaria groups. They are marine micro-organisms whose siliceous shells are found in the sedimentary record since their appearance during the Cambrian period (Boltovskoy, 1999; Lazarus et al., 2005; Suzuki and Not., 2015). While they have been originally neglected for a long time for biostratigraphical studies due to several documented cases of recurrent evolution in the overall morphology of some taxa (e.g. Schrock and Twenhofel, 1953; Campbell, 1954; Bjørklund and Goll, 1979), radiolarian taxonomy and stratigraphy have significantly progressed due to Deep Sea Drilling Project (DSDP) studies since 1968 (Sanfilippo et al., 1985) and are currently of major interest. Radiolarians are commonly used in biostratigraphy by documenting the presence / absence of key marker species, as well as in palaeoceanographic reconstructions of past productivity, temperature, and variability of water masses, wherein these approaches rely instead on relative species abundances. For both these approaches, radiolarians are particularly useful in high latitude settings (e.g. the Southern Ocean) where both the preservation and species diversity of calcareous microfossils are very low.

Indeed, radiolarian’s delicate siliceous remains have been proved important for decades in micropalaeontological studies focussing on palaeoenvironmental reconstructions from various oceanic areas to investigate primary and export productivity (e.g. Welling et al., 1992; Lazarus, 2002; Abelmann and Nimmergut, 2005; Lazarus et al., 2006; Hernández-Almeida et al., 2013; Matsuzaki et al., 2019), sea surface temperature (e.g. Abelmann et al., 1999; Lazarus, 2002; Cortese and Abelmann, 2002; Lüer et al., 2008; Panitz et al., 2015; Kamikuri, 2017; Hernández-Almeida et al., 2017; Matsuzaki et al., 2019), water masses (e.g. Welling et al., 1992; Kamikuri et al., 2009; Kamikuri, 2017; Hernández-Almeida et al., 2017; Matsuzaki et al., 2019) and oxygenation (e.g. Matsuzaki et al., 2019) across the Cenozoic. At present, radiolarians assemblages are considered to be consistent and valuable micropalaeontological bio-indicators as they are largely distributed in all oceans since their appearance and can be very abundant in sediments (e.g. Sanfilippo et al., 1985; Boltovskoy, 1998; Hernández-Almeida et al., 2017).

However, despite their usefulness for such investigations, radiolarians are not as used as other microfossil groups such as benthic and planktic foraminifera, or nannofossils such as coccolithophorids. Experts on living and fossil radiolarians are relatively few, and some radiolarian species still lack a satisfactory taxonomy, especially for taxa within the order Spumellaria (Riedel, 1967; Sanfilippo et al., 1985). Identification of a substantial and sufficient number of specimens per sample (usually about 300 for reliable assemblage composition estimations Fatela and Taborda, 2002) is very time-consuming and requires a consistent and detailed taxonomic knowledge. Moreover, as it is common for all microfossil groups, and especially true for radiolarians, determination and taxonomy of recovered specimens can be different between studies as it can be biased by the subjective appreciation of the operator, influencing reproducibility of the census counts.

Recent technological advances in image acquisition, processing, and recognition now enable automated procedures, from microscopic slide field-of-view acquisition to taxonomic identification, that can ease radiolarian studies. In the early 1980’s, some authors had already proposed to automatically analyse the size and shape of a large number of digitised images of
assemblages of microfossils (Budai et al., 1980), in order to investigate the variability of their morphology and use it as a palaeoenvironmental descriptor. For more than 20 years now, the CEREGE laboratory has been a pioneer in automated image acquisition and recognition for several microfossil groups. Dollfus and Beaufort (1999) developed a structured multi-layer fat Neural network for coccolith recognition, that was first applied in 2001 to Late Pleistocene primary productivity reconstructions (Beaufort et al., 2001). This formed the base for the following Système de Reconnaissance Automatique de Coccolithes (SYRACO) workflow, that used dynamic neural networks (Beaufort and Dollfus, 2004) and is still operating today. For the past few years, the field of computer vision has seen the emergence and development of convolutional neural networks (CNNs), a deep-learning approach that enables the automated classification of large sets of images. Convolutional neural networks are a class of deep neural networks that consist of an input and an output layer, and multiple convolutional layers. This architecture is similar to the organisation of the visual cortex in the human brain.

Several workflows inspired by SYRACO and now using CNNs were successively developed at CEREGE and applied to microfossil taxa (e.g. Marchant et al., in press; Bourel et al., 2020). Regarding radiolarians, previous attempts mainly focussed on the identification step. Apostol et al. (2016) used morphometrical measurements and support vector machine on four radiolarians species recovered from Triassic sediments. In 2017, Keçeli et al. investigated SEM images of 27 selected triassic species. Renaudie et al. (2018) recently achieved promising results focussing on the automated identification of species from the same genus with transmitted light microscope images. They obtain an overall identification accuracy of 73 %, achieved over 16 species from 2 genera, where the morphological difference between species can be very tricky.

In this paper, we also propose a workflow for the automated identification of radiolarians. Our approach differs in that we wanted to generate a neural network that could recognise most of the common radiolarian species, rather than those of a specific genus, in order to investigate their abundance (relative and absolute) and diversity, and thus use them as bio-indicators to reconstruct palaeoenvironmental parameters. It is necessary to obtain a large database of images covering the common species in order to train the network. Out of the modern living 400 to 500 polycystine species, about 100 are relatively common (Boltovskoy, 1998), however, they have yet to be imaged to create a database for automated recognition purposes. Some online Cenozoic radiolarian databases have already existed for a few years (e.g. WoRaD, Boltovskoy et al., 2010; radiolaria.org, Dolven and Skjerpen, 2006; Radworld, Caulet et al., 2006; see Lazarus et al., 2015 for an extensive review of the existing databases), however these are more directed to creating a catalogue for taxonomic purposes. As such, we created an exhaustive and participative database specifically for CNN training and automated recognition purposes, called AutoRadio (Automated Radiolarian database, visible at: https://autoradio.cerege.fr). To achieve this goal, a new protocol to obtain standard images for inclusion in the database was required (square images of individual white specimens on a black background, using a stacking technique if possible) which was also developed in this study.
2 Material and methods

2.1 Material

Radiolarian microfossils to be used in this study were extracted from several sediment cores. Core MD97-2140 was retrieved from the centre of the West Pacific Warm Pool (WPWP; latitude: 2°02’N; longitude: 141°46’E) at a water depth of 2547 m during the Marion Dufresne IMAGES III-IPHIS cruise in 1997 (Beaufort et al., 1997). This core is currently stored at the CEREGE laboratory, France. The sediments consist of a greyish and compact calcareous nannofossil ooze, also containing abundant radiolarian and foraminiferal faunas (de Garidel-Thoron et al., 2005).

Several samples were chosen to extract siliceous microfossils and thus construct a radiolarian images database. Their depths within the recovered core are: 3-4 cm, (6.3 ka, de Garidel-Thoron et al., 2005), 48-49 cm (11.8 ka), 82-83 cm (16.4 ka), 98-99 cm (18.8 ka), 245-246 cm (38.0 ka), 363-364 cm (53.3 ka), 405-406 cm (63.0 ka), 417-418 cm (67.8 ka), 487-488 cm (77.7 ka), 648-650 cm (120.4 ka), 727-728 cm (141.4 ka). For detail on sample processing and slide preparation, the reader is referred to section 2.2.

Middle Miocene to Quaternary samples retrieved from the WPWP were subsequently used to increase the number of rare and absent species in the database. These cores were also taken during the Marion Dufresne IMAGES III-IPHIS cruise: Core MD97-2138 (1°25’S, 146°24’E; 1960 mbsl; samples 1760-1761 cm, 2670-2671 cm, 3151-3152 cm); and from IODP Expedition 363 (Rosenthal et al., 2018): Holes U1483A (13°05.24’S, 121°48.25’E; samples from sections 9H-4W, 14H-5W), U1483B (samples 6H-6W, 18H-2W), U1486B (2°22.34’S, 144°36.08’E; samples 3H-3W, 6H-4W, 13H-4W), and U1486C (21H-4W) as well as about 150 samples from Hole U1488A (2°02.59’N, 141°45.29’S; samples E6H-3W to 35F-2W).

2.2 Random Settling Protocol

A new protocol was developed as a proposed standard methodology for preparing radiolarian microscopic slides. It places 8 samples per standard 76x28 mm slide using 12x12 mm cover slides on which radiolarians are randomly and uniformly decanted using a new 3D-printed decanter (Fig. 1a-b). The 3D file for this new decanter was designed online, using the Autodesk, Inc. 3D design platform Tinkercad (https://www.tinkercad.com/), and is available for free at https://github.com/microfossil/Decanter. Two other versions of the decanter were designed for standard 32x24 mm and 40x22 mm cover slides, also commonly used in micropalaeontology, and are also available online. Custom-sized decanters can also be designed on demand. Our decanters were printed on a Raise3D fused deposition layer type printer using 1.75 mm R3D Premium PLA filament for a material cost of about 1 euro. Approximately 30 g of filament was used, and 4.5 hours were needed to print the model using a standard resolution layer height of at least 0.20 mm.

A random settling technique was preferred to a standard smear slide preparation as the objective of this study is a detailed quantitative faunal analysis with investigation of the relative abundances of each taxon (Sanfilippo et al., 1985). Indeed, a random settling technique provides a more uniform distribution of the residue resulting in less clumped particles, which are also easier to capture digital images of each specimen. The new decanter minimises the loss of material, and a slide guide (Fig. 1c) can also be used to align cover slides. During development, various shapes and sizes of tank were tested, and the one
presented herein was the best compromise between the quantity of sample material required, loss of residue that would not settle on the slide, and quantity of microfossil residue recovered. This method is an improved version of the original random settling method developed by (Moore, 1973), adapted to radiolarian studies (Boltovskoy, 1998) that provides an even and random distribution of the shells on a slide, and modified by (Beaufort et al., 2014) to mount up to 8 samples on a single micropalaeontological glass slide. The use of this new device is simple: a 12x12 mm cover slide is placed in the middle of each tank and maintained centred by the fins. A solution containing radiolarians in suspension for each sample is then poured onto each tank and after a few minutes of settling on the cover slides, water is vacuumed out from each hole.

The new radiolarian slide preparation protocol follows the following steps (#2 to 7 have been adapted from a similar procedure used to process limestone and calcareous sediments):

1. Weigh the sediment.

2. Put about 1 g of sediment (depending on the abundance of radiolarians) in a 200 mL beaker and add a few drops of distilled water to disaggregate it.

3. Add a few mL of 37 % Hydrochloric acid (HCl) until the end of the effervescence.

4. Further add a few drops of % HCl to ensure the end of the effervescence.
5. Pour the solution and rinse the beaker over a 50 µm sieve.

6. Clean the residues in the sieve using a pressure sprayer until they appear whitish.

7. Rinse the residues using distilled water.

8. Weigh a clean glass storage vial.

9. Pour the residues from the sieve to the vial using distilled water.

10. Once the residues have decanted, remove the excess water using a pipette.

11. Place the vial into the oven (about 50 °C) until the residues are dry.

12. Weigh again the vial to calculate the weight of the recovered residues.

13. Gently tap the vial to unstick the residues from the bottom of the vial.

14. Put a 12x12 mm licked or flame burned cover slide into one tank of the decanter.

15. Take about 0.6 to 1 mg of siliceous residue and drop it onto 3.5 mL of distilled water.

16. Shake this solution to suspend the residue and quickly pour it into the corresponding tank.

17. Wait until the residues have decanted (few seconds to minutes) and slowly vacuum out the water from the hole (Fig. 1b).

18. Place the decanter in the oven (about 50 °C) to dry the cover slide.

19. When dry, remove the cover slide from the tank using plastic tweezers and glue it to a standard glass slide (76x28 mm) using optical glue (e.g., NOA81, refractive index of 1.56).

Regarding step #2, the reader should take into consideration the fact that the absolute abundance of radiolarians varies massively in sediment samples from various parts of the ocean. The amount of sediment dissolved into HCl should thus be customised according to the expected abundance.

Regarding step #15, 0.6 mg of residue corresponds to the best compromise between having a sufficient number of radiolarian specimens and them touching or overlapping too much, for a 12x12 mm cover slide (see Appendix A Figure A1). This is desirable as often touching specimens cannot be individually segmented from images, leading to "double" images containing two or more specimens, which cannot be easily classified or assigned to a species count. Distilled water was preferred to ethanol as it leads to less clustering of specimens.

The volume above the cover slide in the tank corresponds to about 45 % of the total volume of the tank. According to the average weight of a radiolarian specimen (about 0.5 µg Takahashi and Honjo, 1983), 0.6 mg of siliceous residues after chemical treatment should then contain about 1,200 radiolarians, if not "contaminated" by other siliceous particles, of which about 600 should fall on the cover slide, thus resulting in at least 300 specimens that should be available for identification (minimum
required to characterise an assemblage by most of the statistical studies, e.g. Fatela and Taborda, 2002). This was confirmed by our tests showing an average of 473 complete identifiable radiolarian specimens per sample, or at least exhibiting more than 50% of their shell, including at least the medullary shells for spumellarians, and cephalis and thorax for nassellarians (excluding specimens touching each other and broken specimens). Depending on the goal and accuracy of your study, this issue can be easily addressed in the sample preparation by pouring a solution of the same sample in the 8 tanks of the decanter (or more or less tanks according to the abundance of radiolarian in the sediment). This way, no changes are required for the image acquisition part of the workflow. Other testing found that Norland Optical Adhesives NOA81 glue was preferred to other mounting media such as NOA74 or Naphrax due to its refractive index, consistency and long-term preservation. Although due to the glue’s viscosity, air bubbles can be trapped in some perforated-type shells, common in collosphaeridae for example. Although not ideal, images of specimens containing bubbles, and that were still recognisable were kept into the database to integrate this variability into the neural network. Although time consuming, metal coating (using C or Au/Pd for example) is also a very efficient way of increasing contrast prior to mounting specimens on the slides. The darkfield illumination technique was too inconsistent in the produced images that further tests were not carried out.

2.3 Automated Image acquisition

Particular emphasis was placed on acquiring high-quality slide images, as being able to recognise different radiolarian species depends on having clearly visible features. However, it has to be noted that, no matter the image quality, very small features that can be taxonomically important (for example bladed vs cylindrical spines) are likely to be difficultly learned by the network as the morphological variability between every pictures of each class is likely to play a most important role, before the network can focus on such small features. For each radiolarian microscopic slide, the 8 cover slides (corresponding to 8 samples) are automatically and consecutively imaged using a Leica DMR 6000 B automated transmitted light microscope (200x magnification using a HCX PL FLUOTAR 20 x magnification Leica lens) and a Hamamatsu ORCA Flash4.0 LT camera, controlled via a LabVIEW (National Instruments) interface. The microscope parameters were set as: Intensity: 10; Depth of field: 38; Aperture: 33; and the condenser was lowered by 9 mm from the glass slide. The LabVIEW acquisition software parameters were set as: Exposure: 9 ms; Gain: 1. These settings provided the maximum contrast between the glass shells and their mounting medium.

For each sample, 324 fields of view (18 x 18 FOVs of 660 x 660 μm each within each 12x12 mm cover glass) were imaged using a multi-focal technique (Fig. 2). For each FOV, 15 images were acquired by incrementally stepping the Z focus position through the microscopic slide (step size: 10 μm) to cover a total focal distance of 150 μm, which corresponds to the thickness of most radiolarian species. This acquisition step takes exactly 1 hour per sample, and thus 8 hours per slide.

2.4 Automated Image Processing and Segmentation

Image processing and segmentation is performed via a second LabVIEW interface. For each FOV, the batch of 15 images is automatically stacked using Helicon Focus 7 (Helicon Soft) and saved following a CoreName-SampleName-FOVNumber.jpg pattern (Fig. 2). As the shells are outlined by the different refractive indices between the shell and the mounting medium,
we didn’t experienced any identification issue with regard to the stacking step, even with small and / or delicate shells. Every stacked FOV image is then processed and segmented into individual specimen images using a custom plugin (AutoRadio_Segmenter.ijm) developed for the ImageJ / Fiji software (V1.52n Schneider et al., 2012). Regarding specimens that are cut between two FOVs, if a part of the shell contains the first chambers (usually for nasselaria) it would be identified as the correct class, and the second part would be identified as "broken", to prevent a double identification in the correct class. The processing steps are:

1. Open a stacked FOV image.
2. Subtract its background.
3. Adjust the minimum and maximum greyscale value to increase its contrast.
4. Invert the image and create a mask.
5. Threshold it in order to binarise it.
6. Blur it and threshold it again to obtain the overall shape of each particle.
7. Separate particles that are in contact with each other (require the configurable Biovoxxel "Water Irregular Features” plugin, available at: https://github.com/biovoxxel/BioVoxel_Toolbox).
8. Define regions of interest (ROIs) for each particle.
9. Restore ROIs corresponding to every particle on the original FOV image.
10. Create a square vignette for each particle.
11. Save it into the corresponding ”Core” folder and ”Sample” subfolder.

Each sample results in approximately 1,000 to 3,000 individual segmented vignettes after the automated image processing and segmentation step.

2.5 Database building and CNN training

ParticleTrieur is a dedicated software program developed at CEREGE (Marchant et al., in press), that enables the operator to visualise and assign vignettes to manually defined classes, and uses the k-NN (k-nearest neighbours) algorithm to aid in identification by self-learning and progressively suggesting identification once enough radiolarian pictures are identified (the reader is referred to Marchant et al. (in press) for more information). Using this software, a large dataset of radiolarian taxa images (called the AutoRadio Database) was progressively built (the current version of the database used in this study can be downloaded at: http://microautomate.cerege.fr/database/ and is freely accessible online as a catalogue on https://autoradio.cerege.fr/database/). It
**Figure 2.** Automated radiolarian image acquisition, processing and identification workflow. 1, 2 (red rectangle): Automated acquisition steps. 3 (Orange rectangle): Automated FOV images stacking step. 4, 5, 6 (Purple rectangle): Automated FOV images processing and segmentation steps. 7 (Blue rectangle): Automated recognition step. 8 (Green rectangle): Automated export of classified images, census counts, and morphometric measurements.

is currently composed of 21,746 images, corresponding to 132 classes/taxa. Each class contains between 1 and about 1,000 images.
Once labelled, this database was used to train a CNN (convolutional neural network) for the automated taxonomical identification of radiolarian vignettes resulting from the automated microscope image acquisition, processing, and segmentation steps. The best results were obtained using a resnet50 topology (residual nets with a depth of 50 layers; He et al., 2015) with added cyclic (Dieleman, 2016) and gain layers (resnet50_cyclic_gain_tl) (resnet50_cyclic_gain_tl; see Marchant et al., in press, for a detailed description of this training process). Greyscale images resized to 256x256 px, a batch size (number of images presented per training iteration) of 64, 30 epochs and four drops for the adaptive learning rate system (ALR), and augmentation (Marchant et al., in press). This training process last about 30 min and generates two files that can then be used for automated recognition (network_info.xml and frozen_model.pb files).

2.6 Automated Taxonomic Identification

Once individual vignettes of radiolarian specimens are generated and saved during the ImageJ processing and segmentation step, they are automatically opened in ParticleTrieur using its server mode, controlled by the second LabVIEW interface. These vignettes are then automatically assigned to a class using the trained CNN. Individual vignettes are then automatically moved into folders corresponding to their core and sample, and subfolders corresponding to their assigned class. Using one microscope, about 8,000 individuals from two slides (16 cover slides corresponding to 16 samples) can be imaged per day (about 500 specimens per sample, from the original 1000 to 3000 vignettes per sample). This fully automated stacking, processing, segmentation and identification step takes about 50 min per sample and operates in parallel to the image acquisition step.

Two types of data are then automatically exported (Fig. 2):

(1) For each sample, a "sample results" file is generated which assembles metadata and morphometric measurements. Each taxonomic ID is then returned to the LabVIEW interface and indexed with its corresponding vignette name (also containing the core, sample, FOV and vignette numbers in column) into a .txt file for each vignette (in row). For each specimen, morphometric measurements, such as "Area", "Diameter", "Major Axis", "Minor Axis", "Circularity", "Roundness", "Solidity", and "Eccentricity" are also automatically appended to the .txt file.

(2) For each core, census data counts of each sample are automatically compiled. A "core results" file is generated during this process where the abundance of each taxon (in column) for each sample (in row) is automatically incremented.

3 Results and discussion

3.1 Description of the database

Of the 21,746 images used to construct the database, 132 morphoclasses were created. Of all these classes, 124 belong to Neogene to Quaternary radiolarian taxa (116 classes corresponding to species or groups of two to three species and containing 11,126 images, 7 classes corresponding to genera and containing 1,932 images, and 1 corresponding to family and containing 677 images) and are part of the Spumellaria families Actinommidae, Coccodiscidae, Heliodiscidae, Litheliidae, Pyloniidae, Spongodiscidae, and Tholoniidae; and of the Nassellaria families Artostrobiidae, Cannobotryidae, Carpocaniidae, Collozoidae,
Plagiacanthidae, Pterocorythidae, Theoperidae, and Trissocyclidae (see Fig. 3, which includes some example images). Eight non-radiolarian classes (corresponding to "background", "broken" specimens, air "bubble", "diatom", "double", "porous fragments", siliceous "particles", and "spicule" and containing 8,011 images) were also defined to train the network to recognise these non-radiolarian images that usually represent 1/2 to 4/5 of the total vignettes.

An extensive overview of the existing Neogene to Quaternary literature was used for the taxonomy and identification of each class, and to define as accurately as possible our assemblages and the observed taxa (including Ling and Anikouchine, 1967; Nigrini and Moore, 1979; Nigrini and Lombari, 1984; Boltovskoy and Jankilevich, 1985; Cault and Nigrini, 1988; Takahashi, 1991; Abelmann, 1992; Boltovskoy, 1998, 1999; Sharma et al., 1999; Nigrini and Sanfilippo, 2001; Itaki et al., 2003; Kamikuri et al., 2009; Zhang et al., 2009; Lazarus et al., 2015; Matsuzaki et al., 2015; Motoyama et al., 2016; Boltovskoy et al., 2017; Matsuoka, 2017; Zhang and Suzuki, 2017; Sandoval, 2018). Synonymies were also taken into account, especially regarding the work of Boltovskoy (1998, 1999). This means that a few species were regrouped into a single class when a significant morphological gradation was observed and when the limit between the considered species was blurry (e.g., *Eucyrtidium acuminatum* and *E. hexagonatum*; *Sithocampe arachnea* and *S. lineata*; *Actinomma henningsmoeni* and *A. leptodermum*). Conversely, ontogenetic stages are clearly visible in numerous species level classes and, once sufficiently imaged, could be distinguished into separated classes. All manual taxonomic IDs during the building of the database were reviewed by a radiolarian taxonomy expert (G. Cortese) to ensure consistent and accurate identifications.

### 3.2 Results of the CNN training

One of the best ways to assess the efficiency of a trained CNN is to look at its confusion matrix (Fig. 4; the original excel spreadsheet is available as supplementary material, see Appendix BTable A1). Right before the training step, the dataset is automatically split into two subsets: one being the training set, and the second one the test set. The data split chosen for this study is 1/5. This means that 4/5 of the original images are used for training (training set) while the remaining 1/5 (test set) of the original images are used for testing the CNN efficiency by calculating several indices. The efficiency results are then represented by the overall accuracy Eq. (1), precision Eq. (2) recall Eq. (3), and individual recall for each class, with these terms defined as:

\[
\text{Accuracy} = \frac{\text{Number of images correctly classified}}{\text{total number of images}}
\]  

(1)

It is the overall performance of the system regardless of class. If you select a random image from the dataset and classify it, the overall accuracy is the probability (in %) that the returned classification is correct.

\[
\text{Precision} = \frac{\text{Number of images that were classified as class } N \text{ and actually belong to class } N}{\text{total number of images classified as class } N}
\]  

(2)
Figure 3. Examples of radiolarian vignettes generated by the automated acquisition, processing and recognition workflow. (a) *Lamprocyclus maritalis*. (b) *Lampropyrtis hannai*. (c) *Theocorythium trachelium*. (d) *Pterocanium trilobum*. (e) *Pterocanium praetextum*. (f) *Eucryphalus sestrodiscus*. (g) *Eucrytidium acuminatum / hexagonatum*. (h) *Acrosphaera spinosa*. (i) *Solenosphaera chierchiae*. (j) *Collosphaera tuberosa*. (k) *Didymocytis tetrahalamus tetrahalamus*. (l) *Hexactinum* spp. (m) *Stylatractus neptunus*. (n) *Heliodiscus asteriscus*. (o) *Tetrapyle octacantha* group. Scale bar 100 µm.

Precision is a metric for a specific class: it is the probability (in %) that an image classified as class N is actually from class N, divided by the total number of images classified as class N.

\[
Recall = \frac{\text{Number of images in class } N \text{ that were correctly classified}}{\text{total number of images in class } N}
\] 

(3)
Recall is, for a specific class, the probability (in %) that a random image from class N is correctly classified, divided by the number of images belonging to class N. Recall is basically the accuracy of a single class. Individual recall scores for each class are visible in the confusion matrix (Fig. 4) as the % of class N (in row) that was identified as various classes (in column). For example, for the first row "Acanthodesmia vinculata", 95 % of the images belonging to this class were correctly identified, while 5 % were classified as "Lophospyris pentagona pentagona". If the CNN training was perfect, the diagonal should only exhibit "100" values. The single overall recall and precision scores are the respective values averaged across all the classes.

During the CNN training, all classes containing less than 10 images (corresponding to rare species, currently lacking images) were automatically fused into a single "other" class. Of the original 132 classes, 109 classes (including 101 radiolarian classes from Middle Miocene to Quaternary, and 8 non-radiolarian classes) were then trained to be recognised with a current overall precision accuracy of just above 90 % (90.1 %) over every class. The average precision is above 85 % (85.6 %) and the average recall is about 81 % (80.7 %). A closer look at the matrix shows that classes with a low recall score usually correspond to classes containing an insufficient number of images (rare species, difficult to get on slide, that would require a significant amount of samples to be processed before enough individual images are generated), usually less than 30 images (e.g., Axoprunum acuilonium: 20 %, contains only 25 images; Clathrocanium coercatum: 50 %, contains only 12 images), while several hundreds of images per class are usually recommended. More images, about at least 150 (ideally 300) in total for each class, as defined above, are then likely to increase the recall and accuracy of these under-represented classes. To this aim, the database will be updated and populated gradually through the automated processing of new samples. As the aim objective of this database is to be free-access and participative online, people are encouraged to send and / or add pictures of these under-represented classes—and send any suggestion to improve the taxonomical framework of the database (see the online contact form at https://autoradio.cerege.fr/contact/).

While the ability of the network to distinguishing between morphologically / taxonomically very dissimilar taxa is very strong (Fig. 4: almost every values out of the squares corresponding to family groupings in the confusion matrix are 0; most of the specimens are usually assigned to their correct family: actinommidae: 93 %; coccodiscidae: 93 %; heliodiscidae: 100 %; lithelididae: 88 %; pyloniidae: 95 %; spongodiscidae: 96 %; artostrobiidae: 97 %; cannobotryidae: 95 %; carpocaniidae: 93 %; collozoidae: 99 %; plagiacanthidae: 87 %; pterocorythidae: 98 %; theoperidae: 96 %; trissocyclidae: 99 %), we also tested its ability to distinguish between morphologically very similar forms, usually corresponding to closely related taxa (species or genera) by computing the accuracy of each radiolarian family present in the database (using the number of specimens of the test set and recall score for each class). Overall, the intra-family accuracy for each family is very high (actinommidae: 80 %; coccodiscidae: 89 %; heliodiscidae: 100 %; lithelididae: 84 %; pyloniidae: 90 %; spongodiscidae: 89 %; artostrobiidae: 93 %; cannobotryidae: 95 %; carpocaniidae: 93 %; collozoidae: 91 %; plagiacanthidae: 85 %; pterocorythidae: 92 %; theoperidae: 93 %; trissocyclidae: 94 %). For each class, the three classes where most of the misclassification occurs are summarised in Appendix C Table A2. Most of the misclassification usually happens with classes of the same family, or with the "broken" class, where a specific part of the investigated class might be recognised (e.g. part of a cephalis or part of a thorax, but always incomplete).
Figure 4. Confusion matrix showing the overall and individual accuracy, precision and recall for the 109 trained classes. Squared groupings correspond to radiolarian families. This figure is also available as a scorable Excel spreadsheet in supplement (Table A1).

As attempts to use genus or higher level taxa as radiolarian proxies in paleoenvironmental research have yielded almost no useful signals, we tried to integrate in our network as many classes correspond to species level taxa as possible. However, the whole workflow is a compromise between distinguishing as many species as possible, and trying to keep a good accuracy for each class, which mostly depends on the growing number of images in each of them. The more images will be progressively added to the database, the more accurate will be the identification, and the more we will be able to go to down to the species level for each class. Eventually, it has to be noted that geographic variation in morphology that might affect the system’s
performance was not yet taken into account, although it is planned to add samples from other locations to cover the species specific variation in morphology linked with spatial distribution. The morphological variation over time (within lineages for example) should not affect the system’s performance as each class contains images of specimens usually covering the species lifespan.

3.3 Accuracy of the trained CNN on a random set of samples

In order to test the reliability and reproducibility of our trained CNN on actual samples, a slide on which 8 cover slides containing siliceous particles from 8 random samples with variable radiolarian abundances was selected. Four Quaternary samples (400 to 6,400 years BP) from Core U1488A and 4 Miocene sample (10.116 to 10.694 ma) from Core U1483 (both from IODP Expedition 363) were then prepared, and their identification scores computed. This slide was automatically imaged, FOV pictures were automatically segmented and individual vignettes were automatically identified using the trained CNN. After a manual verification of every automated identification, 6 indices were computed: (1) the % of radiolarian images recognised as radiolarians (Fig. 6e5a); (2) the % of radiolarian images recognised as the correct radiolarian taxa (Fig. 6e5c); (3) the % of non-radiolarian images recognised as non-radiolarian particles (Fig. 6b5b); (4) the % of non-radiolarian images recognised as the correct particle class (Fig. 6d5d); (5) the % of non-radiolarian images recognised as radiolarian (non-radiolarian false positive; Fig. 6e5e); and (6) the % of radiolarian recognised as non-radiolarian (radiolarian false positive; Fig. 6f5f).

Overall, 7,800 vignettes were identified and manually checked among the 8 samples containing between 444 and 1,502 images each. The abundance of radiolarians ranges from 176 to 697 specimens per sample. The results show that the 6 indices exhibit very close values between the 8 samples. In average, the proportion of radiolarians actually recognised as radiolarian is very high, about 100 % (Fig. 6a5a) and the proportion of radiolarians identified as the correct radiolarian taxa is about 93 % (Fig. 6b5b). Almost all radiolarian images are thus recognised as radiolarian with a 7 % error regarding their species identification. Regarding the non-radiolarian images, more than 95 % are recognised as non-radiolarian (Fig. 6e5c) and again about 95 % are assigned to the correct class (Fig. 6d5d).

False positive identifications were also investigated and are relatively low. Among all the images identified as non-radiolarians, only 0.08 % should be assigned to radiolarians, and among all the images automatically recognised as radiolarians, about 6 % are non-radiolarian images. Within these 6 %, most of the non-radiolarian images confused with radiolarians exhibit radiolarian features and correspond to the non-radiolarian classes ”broken” and ”double” that either contain incomplete radiolarians, or radiolarians touching each other and cannot be assigned to a single species. These false positives are then usually assigned, in the ”broken” class case, to the species partially present in the image, or in the ”double” class case, to one of the species that can be distinguished.

3.4 BiostratigraphyInterest for biostratigraphic studies

To explore the applicability of the automated radiolarian identification workflow is also of interest for biostratigraphic studies, as radiolarian faunal events, such as first occurrences (FOs) and last occurrences (LOs) of radiolarian taxa (about 30 zones were defined for the Cenozoic (Sanfilippo et al., 1985)) were compiled into an Excel spreadsheet. Here, we decided to
Figure 5. Identification indices evaluated on 8 Quaternary and Miocene random samples recovered from cores U1483 and U1488A (IODP Exp. 363).
focus on the biostratigraphy are commonly used for biostratigraphic studies of the Neogene to Quaternary interval using the existing zonation (Nigrini, 1971; Lazarus et al., 1985; Johnson et al., 1989; Moore, 1995; Sanfilippo and Nigrini, 1998; Nigrini and Sanfilippo, 2001; Vigour, 2002). The known stratigraphic ranges of species (101 Middle Miocene to Quaternary radiolarian classes) included in our database were then compiled into an Excel spreadsheet that automatically suggests the age of any sample, according to the composition of its radiolarian assemblage. As this spreadsheet follows the architecture of the automatically generated census data file for each core (core results file), it can be easily filled by copy-pasting this content. This operative workflow, that is automated from the image acquisition to the census counts and can suggest an age for the processed sample could thus significantly contribute to the field of biostratigraphy.

3.5 Application to other datasets and other studies

To test the potential application and limits of our trained CNN on existing sets of images, we compiled various individual images of radiolarians from the literature including unstacked optical microscope images and scanning electron microscope (SEM) images. We then performed a simple colour inversion of the optical microscope image to obtain white specimens on a dark background. Of the hundred images tested, about half were correctly recognised while the other ones were mostly assigned to the "background" class, likely due to the blurry shell edges of the unstacked images, and to the "broken" class, as only part of the shell were probably recognised. While this 50% accuracy on a random set of unstacked optical microscope and SEM images may seem relatively low and arbitrary, it is very encouraging and promising for the development of future and extensive neural networks for automated radiolarian recognition regardless of the imaging method.

4 Conclusions

A new automated radiolarian workflow was developed and consists of a sequence of six steps:

1. A new microscopic slide preparation protocol to enable an efficient automated image acquisition on transmitted light microscopes and decrease the loss of material, as this can limit the investigation of samples where radiolarians are scarce.

2. Automated microscope image acquisition that can automatically image microscopic slides bearing up to 8 samples (324 FOV images per sample) at different focal depths (15 images per FOV, every 10 µm in depth).

3. Automated stacking of each batch of FOV images (using depth maps) to generate a single clear FOV with clearly distinct radiolarian specimens.

4. Automated FOV image processing (contrast enhancement, B&W inversion) and segmentation to generate individual images for every radiolarian specimen.

5. Automated radiolarian recognition using a CNN, as well as calculating morphometric measurements.
6: Automated export of census data per sample (usually about 500 radiolarian images per sample), and storage of radiolarian images in folders corresponding to their taxonomic identification for every sample.

The whole procedure is then entirely automated from the image acquisition to the census counts, and only required from the operator to prepare the micropalaeontological slides and put them under the microscope. The operative workflow described in this study can thus perform complex, tedious, time-consuming tasks such as taxonomic identification and census counts by producing reliable, reproducible, and accurate results. Moreover, as the system can identify most of the common Miocene to Quaternary species, taxonomic specialists can focus on unknown and poorly documented forms. This workflow is achieved using a polyvalent and extensive radiolarian image database (currently 21,746 images) and a ResNet CNN trained using transfer learning for modern and Neogene radiolarian identification. The CNN is currently able to recognise 109 classes with an average precision of about 90%, an overall score that was also obtained on a test performed on 8 random samples containing about 7,800 images.

Although the database already incorporates 124 radiolarian taxa, the main limitation of our system is that it does not cover yet the entire radiolarian diversity that can be relatively high in tropical and subtropical areas (about 500 species), but only equatorial sediments which however exhibit numerous worldwide species. In order to continue to increase its efficiency, more images are required, particularly so for rare species—and from other oceanic regions. The more samples will be processed using the automated workflow, the more images will be progressively added to the database. To this aim, the database was also made free-access and participative to rapidly increase the number of images (see online at https://autoradio.cerege.fr), especially for rare species where the recall score is relatively low, most likely due to low numbers of training images for these taxa.

This new workflow and associated CNN has the potential to make paleoclimate studies more approachable and feasible, along with biostratigraphy for very long sequences. It can be easily installed in other laboratories equipped with an automated microscope as most of our developments are made free-access. The radiolarian census data can then be used to investigate the radiolarian assemblages variability for biostratigraphical purposes, and to develop, apply and improve existing assemblage-based palaeoenvironmental proxies such as SSTs (e.g., radiolarian-based palaeotemperatures for the late Quaternary, Cortese and Abelmann, 2002; Subtropical (ST) Index, Lüer et al., 2008; Radiolarian Temperature Index (RTI), applied to Miocene samples, Kamikuri, 2017) and paleoproductivity (e.g., Upwelling Radiolarian Index (URI), Caulet et al., 1992; Water Depth Ecology index (WADE), Lazarus et al., 2006). It also enables the investigation of evolutionary trends, appearance of new species, and rate of evolutionary change, a fascinating topic regarding radiolarians and other microfossil groups.

This dataset and following studies also enable the fast and accurate measurement of numerous morphometric parameters for each vignette that was assigned a class in the automated recognition step. In addition to the previous research applications, the morphometry aspect provides the possibility to investigate the link between the morphological variability of a species or an assemblage through time along a sedimentary record and elaborate/test scenarios to explain such variability. This new workflow will now be used on two Neogene to Recent sedimentary records from IODP Expedition 363 (Hole U1483A, Hole U1488A), recovered in the West Pacific Warm Pool.
Data availability. The original version of the AutoRadio database used in this study can be downloaded at: http://microautomate.cerege.fr/dat. It is currently composed of 21,746 images, corresponding to 132 classes/taxa.

Code and data availability. A semi-automated version of the AutoRadio_Segmenter.ijm plugin (automated image processing performed on ImageJ / Fiji), developed to process a root folder (“Core”), containing subfolders (“Samples”) of images (“FOVs”) is available online for free at https://github.com/microfossil/ImageJ-LabView-Scripts. To use it, download the .ijm file and save it into the ImageJ/plugins folder, and it will be available to use after restarting ImageJ / Fiji.

Author contributions. MT designed the experiment, performed its technical aspects, including the image preprocessing and write the first draft of the manuscript. RM developed ParticuleTrieur. MT and GC established the taxonomy of the training set. YG developed the automation of the microscope. LB and TdGT were at the origin of the project. All authors participated in the writing of the manuscript.

Competing interests. The authors declare that co-author, Luc Beaufort, is an associate editor of Climate of the Past.

Disclaimer. TEXT

Acknowledgements. We thanks IODP-France for financial support for this project. This work was also supported by the French National Research Agency (ANR) as part of the French platform called Nano-ID (EQUIPEX project ANR-10-EQPX-39-01) and the ANR project FIRST (ANR-15-CE4-0006-01). We also thanks the program Ocean Acidification from the french Foundation for Research on Biodiversity (FRB), and the Ministry for the Ecological and Inclusive Transition (MTES) in supporting the project COCCACE.
References

Beaufort, L., de Garidel-Thoron, T., Mix, A. C., Pisias, N. G.: ENSO-like forcing on oceanic primary production during the Late Pleistocene. Science 293, 2440–2444.


21


Matsuzaki, K. M., Itaki, T. and Tada, R.: Paleoceanographic changes in the Northern East China Sea during the last 400 kyr as inferred from radiolarian assemblages (IODP Site U1429), Prog. Earth Planet. Sci., 6, 1–21, 2019.


