Morphometrics of Southern Ocean diatoms using high throughput imaging and semi-automated image analysis

Dissertation

zur Erlangung des Akademischen Grades

eines Doktors der Naturwissenschaften

– Dr. rer. nat. –

im Fachbereich 2 (Biologie/Chemie)

der Universität Bremen

vorgelegt von

Michael Kloster

Bremerhaven, Januar 2018

Erster Gutachter: Dr. Andres Salvador Rigual-Hernández

Zweiter Gutachter: Dr. Gastón Osvaldo Almandoz

Dritter Gutachter: Dr. Giuseppe Cortese

Tag und Ort des öffentlichen Kolloquiums: 17. April 2018, Universität Bremen

Table of contents

1 General Introduction

This thesis introduces methods I developed for automated morphometric analysis of diatom slides, and their application in two research settings in the Southern Ocean. The following sections describe the organisms and their habitat, as well as introduce established and the newly developed methods.

1.1 Diatoms

Diatoms are unicellular, eukaryotic photosynthetic microorganisms (Round et al. 1990), so-called microalgae, which sometimes form chains or other types of colonies. Their taxonomic classification still is under debate, and so they are referred to as either class Bacillariophyceae (Kociolek et al. 2017) or division Bacillariophyta, with Bacillariophyceae being one of three classes (Round et al. 1990). More than 74,000 names of diatom taxa have been described (Kociolek et al. 2017), and estimations suggest that up to 100,000 or even 10,000,000 diatom species might exist (Norton et al. 1996, Mann and Vanormelingen 2013). A major criterion for differentiating them is their morphology, which comes in very diverse designs, with their symmetry being either bilateral (pennates) or rotational with bi- or multipolar features (centrics).

Habitats of diatoms are planktonic or benthic, where the latter refers to moist or submerged surfaces (Round et al. 1990). Diatoms dwell in any aquatic habitat (ocean, lake, stream, brackish water), and some even in terrestrial environments like moist soil, tree bark or within wet mosses (Pfister et al. 2017), but also attached to or even within sea ice (Thomas and Dieckmann 2002); some species live epiphytic (Winter and Duthie 2000) or epizoic (Frankovich et al. 2015).

Diatoms are one of the most important primary producers. They contribute substantially to both, phytoplankton (Round et al. 1990) and global biogenic carbon fixation, with their contribution to the latter estimated to be ca. 20 - 25 % (Nelson et al. 1995, Mann 1999). Thus they play an important role in food webs and the ocean carbon cycle (Smetacek 1999). Exporting organic carbon to the deep sea, they can sequestrate CO_2 from the atmosphere (Smetacek et al. 2012), thus affecting global climate as well. Accordingly, boosting diatom growth by fertilization on large scale has been discussed as a possible geoengineering technique for counteracting global warming (Denman 2008, Shepherd 2009, Assmy et al. 2013).

A unique feature of diatoms is their cell wall being heavily impregnated with silica ($SiO_2 \cdot nH_2O$) (Round et al. 1990), enclosing the cell material within a robust shell. These shells display a broad range of shapes and fine structures, depending on the species, and are referred to as frustules. The frustule is bipartite, consisting of two usually similar parts fitting into each other, with their arrangement often described as resembling a hat box or pill box. The two subunits are called thecae, where the larger epitheca overlaps the smaller hypotheca. The overlapping area is comprised by cinguli or girdle bands, whilst the thecae's terminal surfaces are termed valves. Because of their heavy utilization of silica, diatoms play an important role in the global silica cycle (Tréguer and De La Rocha 2013).

The diatom life cycle is determined by their unique reproduction scheme, which is closely linked to cell size changes. Most of the time, they reproduce vegetatively by mitotic cell division. In doing so, each daughter cell inherits one of the two thecae, which now becomes the new epitheca regardless of its former position, and is complemented by a newly formed smaller hypotheca (Round et al. 1990). This results in one daughter cell being the same size as its parent – whilst the other daughter cell is smaller. Over time, as the population grows due to ongoing cell division, the proportion of smaller cells increases, dragging down the average cell size of the population. This ongoing shrinking

is referred to as the Macdonald-Pfitzer rule (Macdonald 1869, Pfitzer 1869), and can continue only until cells reach their physiologically possible lower size limit. For most diatom species, the only way out of this impasse is by auxospore formation. This is mostly preceded by meiosis to produce haploid gametes, which then fuse (sexual reproduction). The details of these processes differ between diatom species: iso- and anisogamy were observed, as well as flagellated, amoeboid, or immotile gametes, and heteromixis, automixis or parthenogenesis (Round et al. 1990). The resulting auxospore does not bear a typical bipartite frustule, and thus can expand to produce the so-called "initial cell". This is the largest cell of the whole size cycle, and it is enclosed afresh in a silicified frustule. Subsequently, this initial cell can propagate for some time by cell division again, producing increasingly smaller offspring, until another sexual event becomes necessary to restitute cell size.

Sexual reproduction in diatoms can be induced only if cell size falls below ca. 30 – 75 % of its maximum (Davidovich 2001), but a broad variety of possible external triggers has been observed and discussed (Chepurnov et al. 2004). These depend on taxon and include, for example, changes in light regime, temperature and nutrient availability (Drebes 1977), salinity (Schultz and Trainor 1968), ammonia released by grazing events (Moore et al. 2017), or the presence of specific pheromones (Sato et al. 2011, Moeys et al. 2016). However, for most species the actual triggers are unknown.

The double-track reproduction scheme of diatoms allows for fast population growth under favorable conditions, by conducting vegetative reproduction, whilst, at the same time, it allows for adaptation to unfavorable conditions by sexual recombination of genes. This might help to explain the ecological success of diatoms, which becomes apparent in their broad geographic distribution and their high contribution to global phytoplankton abundance.

Changes in cell size during the diatom life cycle are often accompanied by more or less subtle alterations in the frustule's shape and structure (Woodard et al. 2016). Other external factors also affect these, like fluctuations in the environment and changes in the climate regime. Consequently, not only species composition (Smol and Stoermer 2010, Almeida et al. 2014, Tan et al. 2017), but also variations in cell morphometry might be applied as indicators of water properties, as well as for reconstructing past conditions like for example location of water-masses, nutrient availability and water temperature (Potapova and Snoeijs 1997, Cortese and Gersonde 2007, Shukla et al. 2009, Shukla et al. 2013).

1.2 Southern Ocean

The Southern Ocean surrounds Antarctica, connects the major oceanic basins and is an important site for the global overturning thermohaline circulation (Rintoul et al. 2001). Thus it is important to the global ocean-climate-system, as the largest upwelling area, as well as the main regions for formation of cold, high density water masses sinking from the ocean surface to depth, are located within it (Marshall and Speer 2012). It is characterized by extreme environmental conditions, with severe storms and strong intra-seasonal changes in sea ice coverage and radiation regime. One of its major features is the Antarctic Circumpolar Current (ACC), which is the largest of all ocean currents (Rintoul et al. 2001), driven eastwards by strong winds, with the Subantarctic and the Polar Front comprising its main fronts (Orsi et al. 1995). The Southern Ocean is the largest high-nutrient low-chlorophyll region in the world ocean (Martin et al. 1990), with restricted phytoplankton growth due to limitation in iron (Boyd et al. 2007). In some areas and occasionally, this iron limitation is alleviated by input from upwelling (Tagliabue et al. 2014), islands (Planquette et al. 2007), icebergs (Smetacek et al. 2004) or aeolian dust (Gao et al. 2001, Erickson et al. 2003). The dominant phytoplankton group of the Southern Ocean are diatoms, contributing 89 % of regional primary production, i.e.

4.0 Pg C · y^{-1} (Rousseaux and Gregg 2014). Since those surface water masses of the ACC, which submerge at the northern boundary of the Southern Ocean, fuel lower latitude upwelling (Sarmiento et al. 2004), diatom silica consumption in the ACC influences the global potential for diatom growth and biological carbon sequestration (Kienast et al. 2006, Assmy et al. 2013).

The growth of strongly silicified diatoms has led to an accumulation of siliceous sediment below the ACC, the so called "diatom ooze belt" (Kozlova 1966, Defelice and Wise 1981, Zielinski and Gersonde 1997, Crosta et al. 2005, Shukla et al. 2016). This accounts for approximately one third of the global biogenic silica accumulation (Tréguer and De La Rocha 2013, Tréguer 2014), and represents a unique climate archive (Combes et al. 2007, Cortese et al. 2007, Shukla et al. 2013, Esper and Gersonde 2014a).

Due to the often harsh environmental conditions, sampling in the Southern Ocean is mostly restricted to spring and summer, in limited regions at any one time point. Year-round sampling can be accomplished only by sediment traps.

1.3 Fragilariopsis kerguelensis

I focused mostly on the pelagic pennate diatom *Fragilariopsis kerguelensis* (O'Meara) Hustedt (Hasle 1965), because it plays an important role in ecological, biogeochemical and paleoclimate research (Zielinski and Gersonde 1997, Smetacek et al. 2004, Assmy et al. 2013). It is endemic to the Southern Ocean (Hasle 1965), being frequently present, and sometimes dominant (Hart 1942), in the ice-free open waters of the Antarctic Circumpolar Current, at water temperatures roughly between 0 and 10 °C (Zielinski and Gersonde 1997, Pinkernell and Beszteri 2013). Here it is a main contributor to depletion of dissolved silica from surface waters (Assmy et al. 2013) and its export to the sediments of the diatom ooze belt. This makes *F. kerguelensis* an important species in the global silica cycle, and since it thrives well in high-nutrient low-chlorophyll waters, high abundance of *F. kerguelensis* in sediments is interpreted as an indicator for low-carbon high-silica exporting regimes (Smetacek et al. 2004, Geibert et al. 2005, Abelmann et al. 2006). Besides this species' abundance being used in transfer functions to determine past sea surface temperature (Zielinski et al. 1998, Esper and Gersonde 2014b), morphometric features of its valves have been proposed as proxies for paleoclimate research, linked to productivity, iron availability, (summer) sea surface temperature, the position of the Antarctic Polar Front, and to glacial / interglacial conditions (Fenner et al. 1976, Cortese and Gersonde 2007, Cortese et al. 2012, Shukla et al. 2013).

1.4 Traditional methods for diatom analyses

Scanning electron microscopy and modern biomolecular analyses have become widespread in diatom research. Nevertheless, conventional light microscopy still is the most common method for investigating diatoms, applied almost since microscopy became a scientific research tool. Samples containing living cells, or dead cells fixed with Lugol's iodine solution or formol, are usually investigated using inverted light microscopes. However, since fine structural details, that are also important for taxonomic identification, mostly cannot be observed from such material, often the organic cell compounds are removed by oxidizing agents and acids, and only the siliceous frustules are investigated, at resolutions up to the optical limit. For light microscopic imaging, diatom frustules are embedded in high-refractive mounting media like Naphrax or special industrial glues (e.g. Norland Optical Adhesive NOA 61, Norland Products Inc., Cranburry, New Jersey, USA), to increase the otherwise very poor contrast of the silica structures. Unfortunately, conventional light microscopy comes with major drawbacks: Processing speed is slow, it requires highly trained experts, and results are scarcely documented by images. As a consequence, the traditional method often has

issues regarding data archival, quality and reproducibility, related among others to taxonomic inconsistencies among different investigators.

A possible remedy to this might be newer methods of automated light microscopic imaging and image analysis. High throughput imaging methods have become more and more available in the last years, also in marine research. Systems like FlowCam (Fluid Imaging Technologies, Inc., Scarborough, Maine, USA) or Imaging FlowCytoBot (Olson and Sosik 2007) allow for high-throughput in situ or in-line analyses of living or fixed micro-phytoplankton and even of oxidized diatom samples (Spaulding et al. 2012, Anglès et al. 2015). However, these systems neither provide the highest optical resolution, which mostly is essential for detailed taxonomic diatom analyses, nor get the complete cell into focus. Furthermore, they often cannot retrieve views from the usually indispensable pervalvar angle, especially for chain-forming species. What is more, these methods cannot cover sediment and benthic samples. And finally, these systems are not suitable for processing samples already embedded on slides, which is the common form of storing diatom material for in-house documentation as well as for long term archival in publically accessible collections.

1.5 Collections

Diatom collections play an important role in scientific research, as public type repositories, providers of material, documenting archive or contact point, especially since the processed siliceous material, usually mounted on microscopy slides, has a shelf life in the order of hundreds of years. One of the largest diatom collections in the world is hosted at the Hustedt diatom study centre (Simonsen 1987), which is today integrated into the Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research in Bremerhaven, Germany. Present work in this collection, including my thesis work, focuses on taxonomy, biogeography and morphometrics (Pinkernell and Beszteri 2013, Beszteri et al. submitted) of mainly Southern Ocean diatoms. Its herbarium originates from the private collection of Friedrich Hustedt, a diatomist of high renown, who, during his life from 1886 to 1968, described ca. one third of the 6,000 diatom species known at his time. Over the years, the collection was extended by the complete collections of Reimer Simonsen, Grete Hasle and Kurt Krammer, as well as by contributions of many more, and currently comprises ca. 100,000 microscopy slides and 50,000 samples of diatom material. These cover a broad range of taxa and sampling methods, as well as geographical and geological origins. Some specimens were sampled during historic expeditions like Challenger or Gauss (Heiden and Kolbe 1928, Hinz et al. 2015), but also contemporary expeditions contribute to the diversity and importance of this collection.

1.6 History of automated diatom analysis

Diatom morphometry already dates back several decades (Pappas et al. 2014). Traditionally, it utilized multivariate analyses of measurements and countings (Genkal and Kuz'min 1979, Theriot and Stoermer 1984). Valve outline was mostly analyzed by application of Legendre polynomials (Stoermer and Ladewski 1982, Pappas and Stoermer 2003), Fourier analysis (Mou and Stoermer 1992, Pappas et al. 2001), or based on landmarks (Potapova and Hamilton 2007, Woodard et al. 2016).

Attempts at automating diatom morphometry already started more than 40 years ago (Cairns et al. 1974, Johnston and Stoermer 1976). Some 20 years ago, the first project attempting automatic large-scale diatom imaging and image analysis was conducted, the ADIAC (Automated Diatom Identification And Classification) project (du Buf and Bayer 2002), which I will introduce in detail below. Girault (2010) developed automated microscopy and imaging techniques employing a computer-controlled microscope and the open-source image processing software ImageJ (Rasband

1997-2016) to investigate size characteristics of centric diatoms. Recently, a FlowCam system was used for automated size measurement of pennate diatoms (Spaulding et al. 2012), and automated diatom taxa identification was addressed by means of classical machine learning or deep learning (Bueno et al. 2017, Pedraza et al. 2017).

Software implementing automated diatom morphometry is rare. ImageJ (Rasband 1997-2016) offers a broad range of image processing and analysis functions, but often solid background knowledge as well as some programming experience is required to combine these into a usable workflow. One attempt to overcome these requirements is the automated phytoplankton identification system PlanktoVision (Schulze et al. 2013). This was realized as a plugin for ImageJ, offering a complete image processing and classification pipeline, classifying object outline by elliptic Fourier descriptors (Claude 2008) as well as texture by a neural network. During the work of my master's thesis (Kloster 2013) I developed DiatoMorphoTo, which implements a workflow for finding diatom valves of certain shapes, reviewing them and exporting their outline as well as basic morphometric descriptors. Development of this program continued and finally culminated in the software SHERPA, which I will introduce below. Automatic characterization, classification and description of biological images are facilitated by the VisionBioShape package (Stela and Monleón-Getino 2016) for the R programming environment (R Core Team 2015). This application focuses on object outline only, as described by elliptic Fourier descriptors; the resulting feature vectors can be classified by a broad range of machine learning methods. DiaCurv (Wishkerman and Hamilton 2017) is another software for investigating diatom outlines, calculating their curvature by identifying the chord and the sagitta from the mid-point.

1.7 The ADIAC project

The ADIAC (Automated Diatom Identification And Classification) project (du Buf and Bayer 2002), was the first extensive pilot study regarding automated large-scale diatom identification by means of computer-based image processing. The combined efforts of a European consortium, with working groups from the Royal Botanic Garden Edinburgh and universities of Newcastle, Strasbourg, Bern, Groningen, Madrid and Faro participating, brought together experts in image processing and pattern recognition as well as diatom taxonomists. They investigated methods of automating slide scanning, image acquisition, reference image database construction, image processing, feature extraction and feature classification, for use in diatom identification. With the project running from 1998 to 2001, nowadays some of its aspects have been outrun by commercially available developments, like automated slide scanners. Also diatom image databases, augmented with taxonomic and ecological metadata, have become publically available, like for example DiatomBase (Kociolek et al. 2017), AlgaTerra (Jahn and Kusber), the Hustedt database (Beszteri et al.) or the Antarctic Freshwater Diatoms database (Spaulding et al.). However, these databases address the taxonomist's need for information, but do not cover the demand for extensive annotated training sets, which are essential for automated classification. These require libraries depicting each diatom species by hundreds of images, to cover the morphological variability within the species, including plastic variance as well as ontogenetic size variability. However, even the original ADIAC image database (Bayer et al. 1999) covers only 470 taxa by 2300 images, with a maximum of ca. 20 images each for just 50 taxa. Other major aspects covered by ADIAC regarded automatic diatom contour extraction, morphometry-based classification, compilation of training data, and quality assessment of classification results. Even though their approaches were very promising, the methods they developed did not become widespread in the diatomist community, probably because the technical developments remained unavailable to a broad group of users, and "*shape analysis of diatoms was something vaguely known*

but only used by very few diatomists" at that time, as the final ADIAC report stated (du Buf and Bayer 2002). As of nowadays, utilization of shape analysis has drawn much more attention in diatom research (Mann et al. 2004b, Beszteri 2005, Beszteri et al. 2005, Veselá et al. 2012, Olenici et al. 2017, Wishkerman and Hamilton 2017), but nevertheless, methods for automating the accompanying procedures did not mature into routine application. I tried to address this shortcoming, substantially inspired by the ADIAC project, but focusing on developing a complete diatom light microscopic image acquisition, processing and analysis pipeline, which is easily operated without deep knowledge of the technical aspects, and thus suitable for everyday use by diatomists. This was achieved by combining a commercially available slide scanning microscope with my self-developed image processing and diatom morphometry software SHERPA.

1.8 Objectives of the thesis

The aim of my thesis work was to develop a high throughput, high optical resolution light microscopy-based diatom analysis workflow, and to apply it for investigating morphometrics in Southern Ocean diatoms. In the first part of the thesis, I report on results of the workflow development, and first tests of its applicability in routine use. Subsequently, two major applications of the system were executed, investigating different material types, and focusing on *F. kerguelensis* as an exemplary Southern Ocean diatom species. The first study investigated life cycle events of this species by morphometric signals from a two-year sediment trap time series. In the second study, fossil sediment core material was examined for morphometric changes related to changes in environmental conditions during glacial-interglacial transitions.

1.9 Thesis outline

This work is structured in four major subjects:

1.9.1 Semi-automated image analysis

The first step in this study regarded investigation and implementation of image processing methods for automated diatom morphometry. To do so, I revised the software DiatoMorphoTo, which originated from my master's project (Kloster 2013), and integrated new methods, resulting in the novel diatom morphometry software SHERPA. The programming part included, among other tasks, implementation of new features like template matching via elliptic Fourier descriptors (Claude 2008, Claude 2010), calculation of various morphometric descriptors, and an enhanced ranking system for internal assessment of result quality. Also, I tested the software extensively, extended the template library massively, and provided detailed documentation, including a user guide, a technical manual and instructions for optimizing settings. SHERPA and the corresponding documentation were published in the frame of **Publication I**, which describes the software, the concepts behind it, the methods employed and the general usage. Two short practical applications assessed the applicability of SHERPA on material from different sources: 1) Sediment core material was used to test the software's aptitude to analyze material containing debris, broken valves and overlapping objects, and compared different segmentation methods; 2) parts of the work done to differentiate six *Sellaphora pupula* demes by minute differences in their valve outline (Mann et al. 2004b) were successfully reproduced. SHERPA was made available for all to download and use under www.awi.de/sherpa.

1.9.2 High throughput imaging and morphometry workflow

The second step in this study implemented a complete high-throughput diatom morphometry workflow, from image acquisition to extraction and analysis of feature descriptors. For automated microscopic imaging, I used the Metafer slide scanning microscope (MetaSystems Hard & Software GmbH, Altlussheim, Germany). This commercially available system was designed especially for

research purposes, with a broad range of possible applications, though primarily targeting medical applications, that is, digital histology and pathology. Its imaging capabilities comply with the needs of diatom microscopy, especially since it can make use of oil-immersion objectives and apply focus-stacking. Nevertheless, its image processing and classification functions turned out to be insufficient for finding and measuring diatom valves. To overcome this limitation, we developed and tested several scenarios of combining the Metafer slide scanner with SHERPA. The most promising workflow replicated ideas which had been supposed in the ADIAC project, but were not be implemented at that time due to problems related to re-locating positions, in which diatom valves were found, at sufficient precision. My approach employed a low-resolution scan of the slide area, followed by locating diatom valves of interest with SHERPA. In a second scanning cycle, only these valves of interest were imaged, with a high-resolution oil immersion objective and applying focus stacking. The resulting focus-enhanced images were subjected to high-precision morphometric measurements by SHERPA. To render this two-step scanning procedure possible, I extended SHERPA with a function for exporting valve outline coordinates into the Metafer VSAI format. Furthermore, I increased the processing speed of SHERPA, by code optimization, improved linking of external functions and parallelization of image processing routines, and I developed and implemented a novel routine for measuring costae distance or striae density for the taxon targeted in the next two chapters of this thesis. After I honed the workflow by extensive testing, it was applied in a showcase scenario, which investigated basic valve morphometrics of *F. kerguelensis* from sediment core material. During this test, I found and documented valves of this species much larger than described before (Hasle 1965, Fenner et al. 1976, Cefarelli et al. 2010), extending the valve length range (apical axis) reported for this species by ca. 45 %.

The workflow and the findings, along with an updated version of SHERPA, were published in **Publication II**, as part of a special issue of Applied Sciences regarding "Automated Analysis and Identification of Phytoplankton Images". This special issue comprised three other publications (Bueno et al. 2017, Pedraza et al. 2017, Rojas Camacho et al. 2017), showing the growing interest in the topic.

1.9.3 Life cycle analysis from sediment trap samples

For testing the potential of this high-throughput diatom morphometry workflow, I applied it to measure features of ca. 29,000 *F. kerguelensis* valves from a two year sediment trap time series. This enabled me to investigate life cycle events of the species, as they are represented by morphometric changes. Here I mainly analyzed changes in distributions of valve length, although I also looked at other features. In **Publication III**, which is a manuscript ready to be submitted for publication, I present this data set and discuss possible causes of changes in valve size distributions, such as vegetative proliferation, grazing, adaptation to environmental conditions, an overwintering strategy and sexual activity.

1.9.4 Morphometric changes related to glacial-interglacial transitions

As a final showcase, I investigated further morphometric features from ca. 12,000 *F. kerguelensis* valves I had measured from the Southern Ocean sediment core PS1768-8 (Gersonde and Hempel 1990, Gersonde 2003). This sediment core had been studied intensively before (Zielinski 1993, Zielinski et al. 1998, Gersonde and Zielinski 2000, Gersonde et al. 2003, Gersonde et al. 2005, Abelmann et al. 2006, Esper and Gersonde 2014a, b). Nevertheless, focusing on sampling depths covering glacial-interglacial transitions, I found two previously overlooked *F. kerguelensis* morphotypes, which differed by the shape of their valve tips. These differences were so subtle that they did not become apparent by visual inspection, but only by the value of the morphometric

descriptor rectangularity. In **Publication IV**, which is ready to be submitted for publication, I describe this data set, changes in relative abundances of both morphotypes throughout two glacial-interglacial transitions, and suggest possibilities for improving paleo-proxies based on *F. kerguelensis* valve morphometrics which have been proposed for summer sea surface temperature, iron availability, position of the Antarctic Polar Front or glacial / interglacial regimes (Cortese and Gersonde 2007, Cortese et al. 2012, Shukla et al. 2013, Shukla and Crosta 2017). In this publication, I also underline the taxonomic and paleoceanographic improvement gained by precisely measuring morphometric features like valve area (as opposed to roughly estimating it, as has previously routinely been done in *F. kerguelensis* morphometric proxy studies).

2 Publications

2.1 List of publications

This thesis is based on the following publications:

I. M. Kloster, G. Kauer, B. Beszteri (2014)

SHERPA: an image segmentation and outline feature extraction tool for diatoms and other objects

BMC Bioinformatics 2014, 15:218, doi:10.1186/1471-2105-15-218
Received: 15 May 2014, accepted: 27 May 2014, published: 25 June 2014

II. M. Kloster, G. Kauer, B. Beszteri (2017)

Large-Scale Permanent Slide Imaging and Image Analysis for Diatom Morphometrics

Applied Sciences "Automated Analysis and Identification of Phytoplankton Images" 2017, 7, 330, doi:10.3390/app7040330
Received: 1 February 2017, accepted: 23 March 2017, published: 28 March 2017

III. M. Kloster, A. S. Rigual-Hernández, L. Armand, G. Kauer, T. W. Trull, B. Beszteri

Insights into the life cycle of the Southern Ocean diatom *Fragilariopsis kerguelensis* through high-throughput microscopy of sediment trap samples

To be submitted to Journal of Phycology

IV. M. Kloster, O. Esper, G. Kauer, B. Bestzeri

Morphometry of the diatom *Fragilariopsis kerguelensis* from Southern Ocean sediment: High-throughput measurements show second morphotype occurring during glacials

Submitted to Marine Micropaleontology

Other publications prepared with contributions of this candidate from the period of this work:

V. B. Beszteri, C. Allen, G. O. Almandoz, L. Armand, M. Á. Barcena, H. Cantzler, X. Crosta, O. Esper, R. W. Jordan, G. Kauer, C. Klaas, M. Kloster, A. Leventer, J. Pike, A. S. Rigual-Hernández

Quantitative comparison of taxa and taxon concepts in the diatom genus *Fragilariopsis*

Submitted to Journal of Phycology

2.2 Statement of the candidate´s contribution to the publications included in this dissertation

Publication I

The candidate investigated image analysis methods for their potential applicability to diatoms, developed the software SHERPA, compiled the template library and wrote the accompanying documentation. Some basic concepts of it originate from the software DiatoMorphoTo, which the candidate developed during his master's project (Kloster 2013). B. Beszteri provided help concerning diatom morphometry and recreated the analysis Mann et al. (2004b) did on differentiating six *Sellaphora pupula* demes by minute differences in their valve outline. G. Kauer provided help with programming questions. The manuscript was written by the candidate in discussion with the co-authors.

Publication II

Bank Beszteri proposed the two-cycle workflow. The candidate developed this workflow and programmed the export of object coordinates into the Metafer VSAI format, along with the other extensions and refinements of SHERPA, as well as the R script for removing doublets. The workflow was honed by the candidate, Gerhard Kauer and Bank Beszteri. Sample material was contributed by Oliver Esper and Rainer Gersonde. Friedel Hinz, Sean Seegert, Sarah Olischläger and Fenina Buttler prepared the samples and microscopy slides. The candidate scanned the samples, did SHERPA analyses including manual check of the taxonomic identity and segmentation quality for each valve outline found, and performed all downstream morphometric analyses. The manuscript was written by the candidate in discussion with the co-authors.

Publication III

Andres S. Rigual-Hernández proposed to investigate *F. kerguelensis* life cycle from sediment trap material, which was provided by Tom W. Trull. Andres S. Rigual-Hernández prepared the samples and microscopy slides. The candidate performed microscopic scanning, morphometric measurements using SHERPA including manual check of the taxonomic identity and segmentation quality for each valve outline found, and data analysis. Life cycle signals in the observed size distributions were interpreted by the candidate in discussion with Leanne Armand, Andres S. Rigual-Hernández and Bank Beszteri. The manuscript was written by the candidate, supported by the co-authors; Andres S. Rigual-Hernández provided the text and map of the oceanographic setting (Figure 1).

Publication IV

Sample material was contributed by Oliver Esper and Rainer Gersonde. Friedel Hinz, Sean Seegert, Sarah Olischläger, Fenina Buttler and Fabian Altvater prepared the samples and microscopy slides. The candidate scanned the samples, analyzed *F. kerguelensis* valves using SHERPA including manual check of the taxonomic identity and segmentation quality for each valve outline found, programmed the R scripts for the morphometric analyses and conducted them. The manuscript was written by the candidate in discussion with the co-authors.

2.3 Publication I – SHERPA: An image segmentation and outline feature extraction tool for diatoms and other objects

Michael Kloster: Corresponding author, AWI and HSEL, eMail: sherpa@michael-kloster.de

Gerhard Kauer, HSEL, eMail: gerhard.kauer@hs-emden-leer.de

Bánk Beszteri, AWI, eMail: bank.beszteri@awi.de

AWI: Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research,
Am Handelshafen 12, 27570 Bremerhaven, Germany

HSEL: University of Applied Sciences Emden / Leer, Constantiaplatz 4, 26723 Emden, Germany

2.3.1 Abstract

Background

Light microscopic analysis of diatom frustules is widely used both in basic and applied research, notably taxonomy, morphometrics, water quality monitoring and paleo-environmental studies. In these applications, usually large numbers of frustules need to be identified and / or measured. Although there is a need for automation in these applications, and image processing and analysis methods supporting these tasks have previously been developed, they did not become widespread in diatom analysis. While methodological reports for a wide variety of methods for image segmentation, diatom identification and feature extraction are available, no single implementation combining a subset of these into a readily applicable workflow accessible to diatomists exists.

Results

The newly developed tool SHERPA offers a versatile image processing workflow focused on the identification and measurement of object outlines, handling all steps from image segmentation over object identification to feature extraction, and providing interactive functions for reviewing and revising results. Special attention was given to ease of use, applicability to a broad range of data and problems, and supporting high throughput analyses with minimal manual intervention.

Conclusions

Tested with several diatom datasets from different sources and of various compositions, SHERPA proved its ability to successfully analyze large amounts of diatom micrographs depicting a broad range of species. SHERPA is unique in combining the following features: application of multiple segmentation methods and selection of the one giving the best result for each individual object; identification of shapes of interest based on outline matching against a template library; quality scoring and ranking of resulting outlines supporting quick quality checking; extraction of a wide range of outline shape descriptors widely used in diatom studies and elsewhere; minimizing the need for, but enabling manual quality control and corrections. Although primarily developed for analyzing images of diatom valves originating from automated microscopy, SHERPA can also be useful for other object detection, segmentation and outline-based identification problems.

Keywords: Diatom, Segmentation, Outline, Elliptic Fourier analysis, Shape descriptors, Morphometrics, Automated slide scanning

Abbreviations: CDF, convexity defection factor; CDMs, convexity defect measures; CHMDF, convex hull maximum distance factor; CSV, comma separated values; EFDs, elliptic Fourier descriptors; EFDIs, elliptic Fourier descriptor invariants; PCAF, percent concave area fraction; RATS, robust automated threshold selector; SHERPA, tool for SHapE Recognition, Processing and Analysis; TIFF, tagged image file format

2.3.2 Background

Diatoms are a group of photosynthetic protists producing uniquely ornamented and diversely shaped silicate shells (Round et al. 1990). They are present in all aquatic and wet habitats and, with an estimated 10^5 species, they represent the most species rich algal group (Mann and Vanormelingen 2013). Diatom assemblage composition reflects the abiotic and biotic features of their respective habitats, and is widely used for making inferences about environmental conditions in water quality monitoring and paleontology (Smol and Stoermer 2010). Due to a combination of traditional and practical reasons, the most widely applied method for diatom investigations is based on light microscopic analysis of so called permanent slides, prepared using the silicate frustules after cleaning them of organic material (Round et al. 1990).

Size and shape distributions of diatom populations are measured and analyzed in a number of different fields, including taxonomy (Droop et al. 2000, Kingston and Pappas 2009, Poulíčková et al. 2010, Abarca et al. 2014), ecology (Crawford et al. 1997, Jewson et al. 2008, Shimada et al. 2009, Jewson et al. 2010), and paleontology (Mann et al. 2004b, Cortese and Gersonde 2007, Marchetti and Cassar 2009, Cortese et al. 2012, Shukla et al. 2013). In such studies, dozens to hundreds of specimens are routinely investigated from each of several slides, and measurements are usually performed by one of the following methods: 1) through an ocular micrometer directly on images seen in the microscope by the investigator (Kumar et al. 2009); 2) as manual (mostly, length) measurements on digital live images presented on a computer screen (Shukla et al. 2013, Abarca et al. 2014); 3) as manual (again mostly, length) measurements on saved digital images using general purpose image analysis software (Shimada et al. 2009); 4) combination of manual measurements and measurements obtained by custom-developed macros or extensions of general purpose image analysis software like ImageJ (Shukla et al. 2013) or Optimas (Droop et al. 2000, Mann et al. 2004b).

There is a considerable methodological gap between these approaches and the sometimes rather sophisticated methods which have been applied to diatoms in the image analysis literature for instance in the project ADIAC (du Buf and Bayer 2002), or by others including (Grima et al. 2003, Álvarez-Borrego and Solorza 2010, Luo et al. 2011). Much of the experience gained in diatom image analysis studies should in principle be transferable to diatom morphometrics and would have the potential to speed up the latter and make it more accurate and reproducible. However, these methods have remained practically inaccessible to diatomists due to a lack of publicly available and user friendly implementations of image processing and analysis methods suitable for diatom analyses. Most of the diatom image analysis literature does not explicitly state which software tool or framework was used for implementing the applied methodology. Although this practice reflects a focus upon algorithms and methods, as opposed to software, and is probably well suited for readers with their main area of expertise lying in computer science and image analysis, translating these methodological experiences into routinely practicable workflows has remained a challenge beyond the qualification of most, if not all, diatomists, as illustrated by the almost complete lack of reports on re-use of these methods beyond the groups which developed them. The only case known to us where implementations of individual algorithms have been made available publicly is represented by the small collection of MATLAB and C source code files available under (Marshall et al.). However,

even these only represent fragments of a practically applicable analysis workflow and are virtually inaccessible to most diatomists (at least to the overwhelming subset lacking familiarity with MATLAB / C programming).

Several of the individual algorithms tested and applied in diatom image analyses in the above cited works represent standard image analysis methods, with widely available implementations in general purpose image analysis software like ImageJ (Rasband 1997-2016). Thus, it could be argued that such software should also be perfectly suited for the needs of diatomists. However, in our experience, whereas for instance ImageJ can be useful for processing and analyzing individual diatom images or small collections thereof, building a workflow for high throughput work with it requires serious programming capabilities, a reason probably hindering the use of such software in diatom studies. For instance, a number of segmentation algorithms can successfully be applied to diatom valves, but it is often found that a different method works best for different objects, depending not only on valve structure (and thus, also taxonomy) but also upon minor details of how the object lies relative to the focal plane and to neighboring objects (du Buf and Bayer 2002). Whereas one can easily apply a handful different segmentation algorithms to an image in for instance ImageJ, deciding which one gives best results in a case-by-case manner can be challenging. Doing so programmatically to enable batch processing of large numbers of images with minimal manual interaction would go beyond the capabilities of most non-image-analysis-expert users of ImageJ. Since diatom images are notoriously difficult to segment due to the optical properties of the silicate shells (low contrast, strong halo around outline, huge structural and shape diversity), chaining together individual analysis steps to an automated workflow also requires some kind of quality control. Differentiating objects of interest (diatom frustules, or, in particular cases, frustules of a particular group of diatoms) from other objects found by segmentation methods (sediment particles, debris, non-target species) would also require considerable programming skills to implement in ImageJ.

The outline represents a rather information rich aspect of the morphological variability of diatom frustules, and its shape and size contains substantial taxonomic and life cycle related information especially in the case of pennate diatoms (even if it has to be noted that diatom identification at the species level is mostly impossible based on outline shape alone). The main approaches for quantitative characterization of outline shapes in diatom morphometrics have included the use of simple heuristic shape descriptors like rectangularity (Droop et al. 2000), ellipticity, compactness (du Buf and Bayer 2002, Rosin 2003); Legendre-polynomials ((Kingston and Pappas 2009) and the large body of literature cited therein); Fourier descriptors (Mou and Stoermer 1992, Pappas and Stoermer 2001, du Buf and Bayer 2002); and landmarks and semi-landmarks (Falasco et al. 2009, Frankova et al. 2009, Vesela et al. 2009, Poulickova et al. 2010, Veselá et al. 2012, Kermarrec et al. 2013). Although further methods have been developed, some specifically for diatoms, notably the segment shape analysis approach (Loke and Du Buf 2002) successfully applied in (Mann et al. 2004b), these have not become widely used. General purpose morphometrics software (SUNY Stony Brook 2004, Klingenberg 2011) is available for landmark and semi-landmark digitization and analysis, but using such software, landmark points need to be digitized individually and manually, hindering high throughput analyses. For other types of outline descriptors, some software support is available (see e.g. examples for software tools capable of calculating elliptic Fourier coefficients under (SUNY Stony Brook 2004)), but again not as part of routinely applicable workflows supporting the analysis of large numbers of images.

With SHERPA presented in the present paper, we address these gaps and introduce an easy-to-use tool for segmenting and analyzing light microscopic images of diatom frustules, and for extracting a

number of outline features useful for diatom morphometrics (but potentially in other fields as well). Our goals were to develop a tool that implements 1) a full image analysis workflow from image segmentation to outline feature extraction, specifically adapted to diatom images, but potentially useful for other objects where outline shape is informative; 2) multiple segmentation methods and an automated selection of the best result for each segmented object; 3) matching of object outlines against a set of template outlines to enable both taxonomically selective as well as broader analyses; 4) object scoring and ranking to support quality checking; 5) extraction of a wide range of outline shape descriptors for further analyses; 6) supporting processing of large batches of images by minimizing the need for manual interaction, but leaving the possibility for it in case it should be required, e.g., to correct outlines for diatom valves with minor overlaps with neighboring objects. Software implementing statistical and / or machine learning methods for exploration, analysis, and classification of large multivariate data sets is widely available both commercially and free of charge for users at a wide range of levels of computer fluency (ranging for instance, from the easy-to-use PAST (Hammer 2013) or JMP (SAS 2013) to the more challenging, but also more versatile statistical analyses systems like R (R Core Team 2015) or SPSS (IBM Analytics 2013)). Accordingly, we decided to not include this functionality in our tool but rather generate output that can be loaded for downstream analyses into the user's statistical tool of choice.

2.3.3 Implementation

SHERPA, the tool for "**SH**ap**E** **R**ecognition, **P**rocessing and **A**nalysis", offers an image processing workflow focused on the identification and measurement of object outlines (see Figure 1). Though it was developed focusing on analyzing diatom valves, SHERPA can also handle other object classes. Starting point are micrographs, obtained by optical microscopy, or similar images. For each depicted object, the respective outline is detected and compared to a set of templates which characterize representative shapes of interest. Detected objects receive quality scores and are ranked accordingly, reflecting the chance of representing a relevant object. The aim of this step is to reduce the effort required for sorting out unwanted objects. Suboptimal results can be revised manually to improve yield if necessary, and selected results can be exported along with a set of descriptors for further morphometric scrutiny.

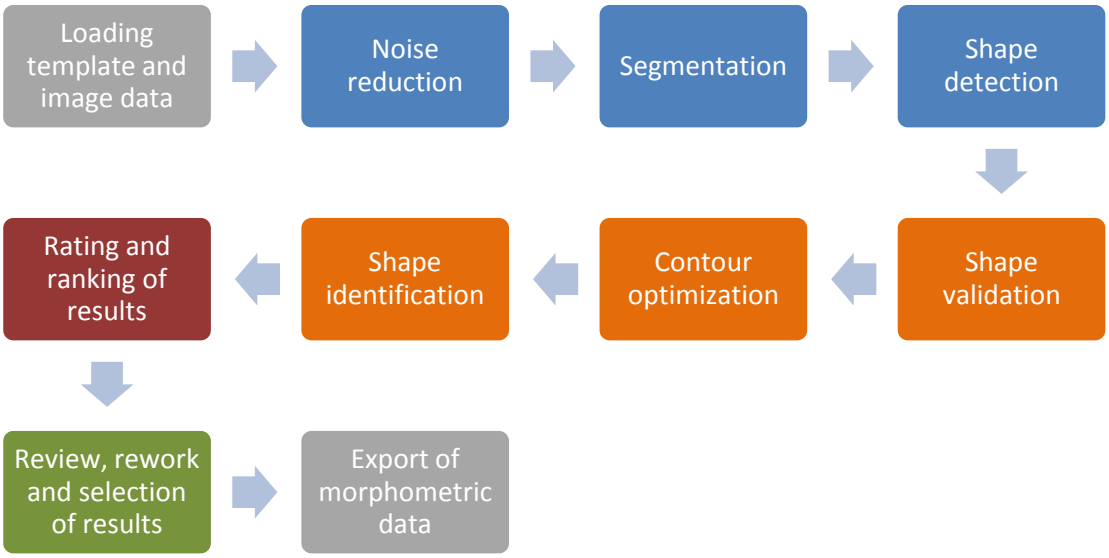


Figure 1: Structure of SHERPA's image processing pipeline / workflow.

This way, extensive image collections can be processed in a fully automated manner or with minimal manual intervention. Irrelevant data, originating from debris, damaged or unwanted objects, can be sorted out with little or no user intervention at all, while relevant objects are identified and

measured. The exported morphometric descriptors allow for a detailed and specific analysis based on tools like R (R Core Team 2015), and questions about variation in outline shape and size can easily be investigated.

One of the main strengths of SHERPA is its easily to follow workflow and plain user interface, which combine different techniques into a simple to use, yet powerful tool, which does not demand deeper expertise in image processing and programming. This distinguishes SHERPA from general purpose image analysis solutions like ImageJ (Rasband 1997-2016), which usually require experience in image processing and a lot of manual intervention or skills in scripting (Table 1 lists the main features of SHERPA which go beyond those supported by ImageJ).

Table 1: Comparison of features of SHERPA and ImageJ

Feature	SHERPA	ImageJ
Integrated workflow for segmentation, identification and measurement of objects	Yes	No
Automatic combination of multiple segmentation methods	Yes	No
Automatic combination of multiple contour optimization methods	Yes	No
Convexity defect measures	Yes	No
Ranking of segmentation results	Yes	No
Quick interactive review of results	Yes	No

In order to create a low level entry point for novice users, extensive documentation is provided along with the software, including a comprehensive manual, a quick-start guide, a tutorial on how to achieve suitable settings in a straightforward way, and a technical description of the analysis process and extracted morphometric features.

SHERPA was developed for Windows7 64 Bit using C# / .NET 4.0. Most image processing functions are realized based on OpenCV 2.4.2 (itseez 2012), whose DLLs are wrapped for .NET by Emgu CV 2.4.2 (Emgu 2012), and on ITK 4.2 (Kitware and al. 2012) called via external executables. “Microsoft .NET Framework 4” (Microsoft Corporation 2013a) and the “Microsoft Visual C++ 2010 SP1 Redistributable Package (x64)” (Microsoft Corporation 2013b) have to be installed prior to running SHERPA. A 32 Bit version of SHERPA is available, but its usage is not recommended because it might run out of memory resources when analyzing large amounts of data.

2.3.3.1 Input data

Image data to be analyzed can depict objects either as dark structures on bright background (like obtained e.g. using bright field microscopy) or as bright structures on dark background (like obtained e.g. using dark field microscopy). Objects are identified by shape information. For proper results, object outlines should be focused as precisely as possible. Minor blurring will affect the accuracy of outline detection, while extensive fuzziness might impede usable results. For an optimal identification yield the sample density should be sparse without overlapping objects.

Templates provide prototypes of relevant shapes, containing silhouettes of each suitable object type (see some example diatom templates in Figure 2). A broad collection of templates depicting diatom valves is provided along with SHERPA (see under “Results and discussion”). However, for good results, a set of templates depicting the morphological variability of the objects under investigation must be generated. Depending on the object of interest, several templates might be needed to cover the range of shapes corresponding to one type (species). In the case of our objects of primary focus,

diatom valves, templates should cover the range of shape variation occurring during size reduction for each taxon concerned (see some examples in Figure 2 e-g).

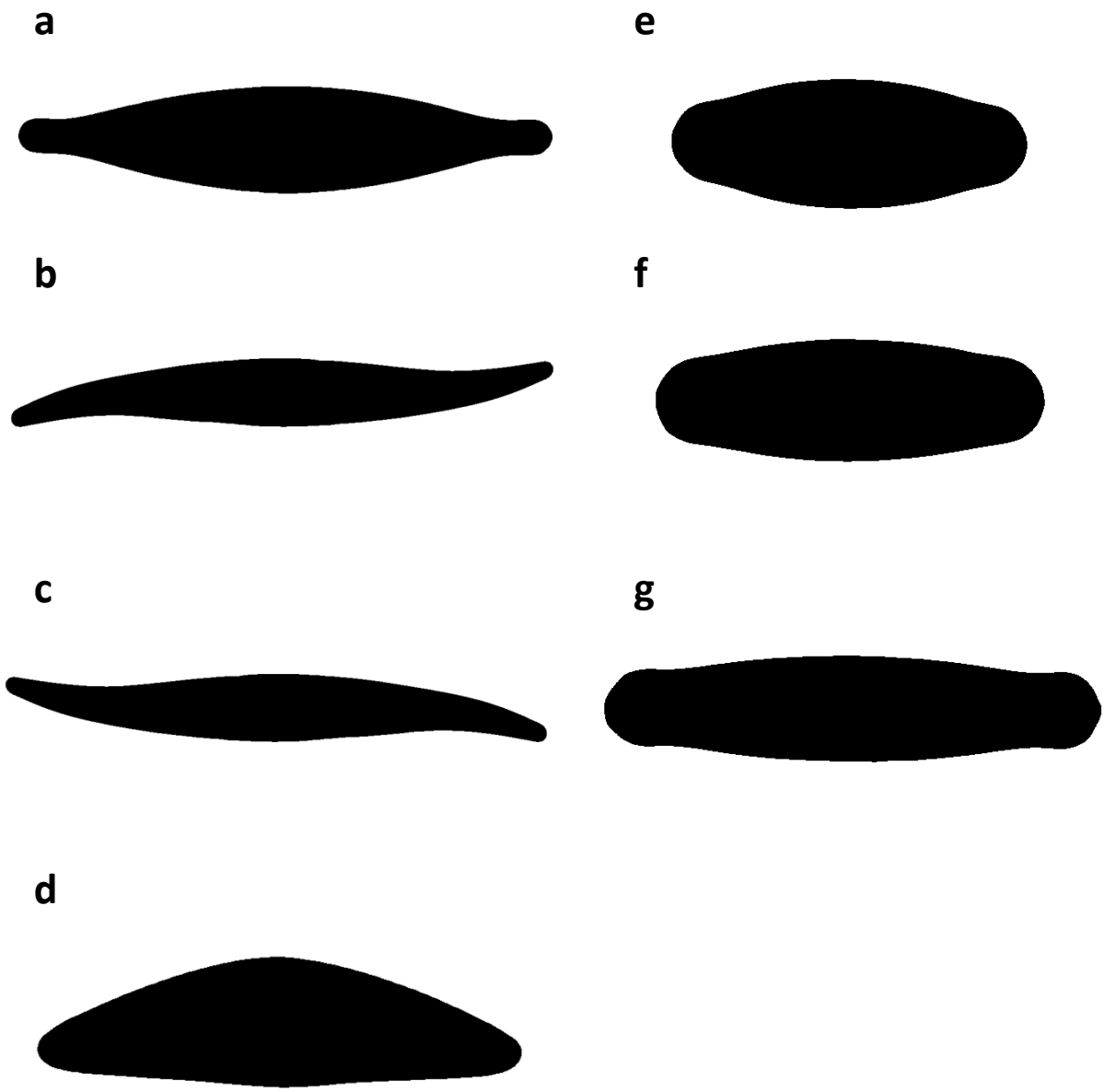


Figure 2: Seven exemplary templates used for shape detection. a) a typical *Navicula*, b) *Gyrosigma*, c) the same Gyrosigma mirrored, d) *Cymbella helvetica*, e-g) different variations of *Sellaphora pupula*. All shapes were derived from ADIAC data (du Buf 2003)

Since templates are matched to object shapes by using elliptic Fourier analysis (see below under "Shape identification"), the identification process is insensitive to size, rotation and position. However, it is not invariant to mirroring, so for objects which do not have symmetry with respect to an axis, two templates need to be used (see Figure 2 b-c).

2.3.3.2 Image processing

Image data is converted into shape information by applying a consecutive set of image processing functions:

Noise reduction can be performed by applying Gaussian or median filtering.

Image segmentation separates objects from image background by using up to five different procedures (see Figure 3). Segmentation algorithms implemented are Otsu's thresholding (Otsu 1979), Canny edge detector (Canny 1986), robust automated threshold selector (RATS) (Lehmann 2006) and adaptive thresholding (Bradski and Kaehler 2008), p. 138 ff., where Otsu's thresholding can additionally be combined with histogram equalization (Bradski and Kaehler 2008), p. 186 ff. for analyzing images with poor contrast. Whilst for most segmentation procedures a single set of parameters is provided, RATS can be applied running a whole range of sigma values as a kind of "brute force" approach for trying to successfully segment even difficult data. Since only the outer contour of each object is analyzed, segmentation errors within the object's interior are negligible.

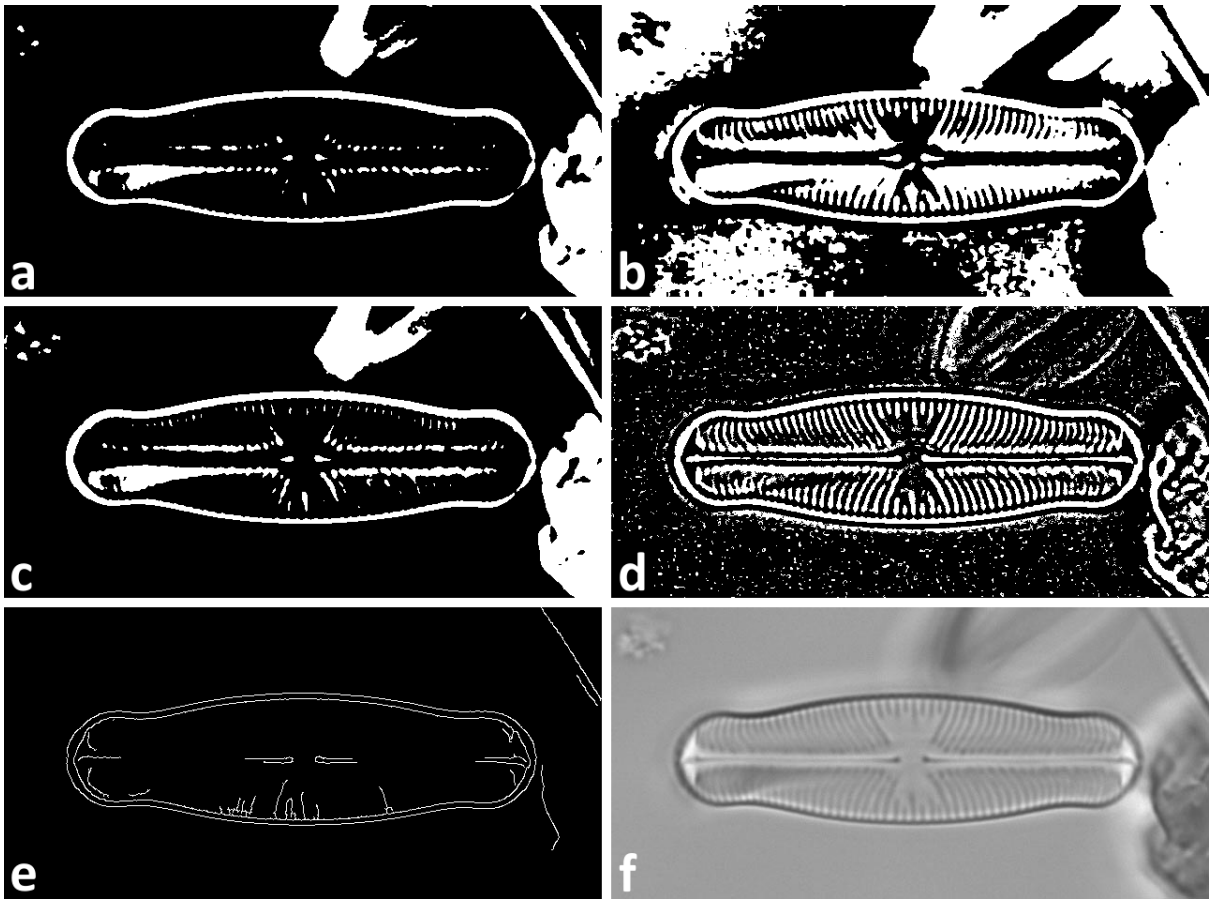


Figure 3: Results of different segmentation procedures. a) Otsu's thresholding, b) Otsu's thresholding combined with histogram equalization, c) robust automated threshold selector (RATS), d) adaptive thresholding, e) Canny edge detector, f) original image data. For each object (white) only the outer contours are analyzed subsequently.

All segmentation procedures can be applied simultaneously. This allows for an increased yield of detected objects, since each procedure presents its own advantages and disadvantages, depending on the image data quality, but this approach can generate manifold results for a single object (see Figure 4). To prevent multiple detection, for each object only the one result will be taken into consideration, which produces the best matching value for any template (according to elliptic Fourier analysis, see below under "Shape identification"). Two shapes are considered as belonging to the same object if the centroid of one shape lies within the area of the other.

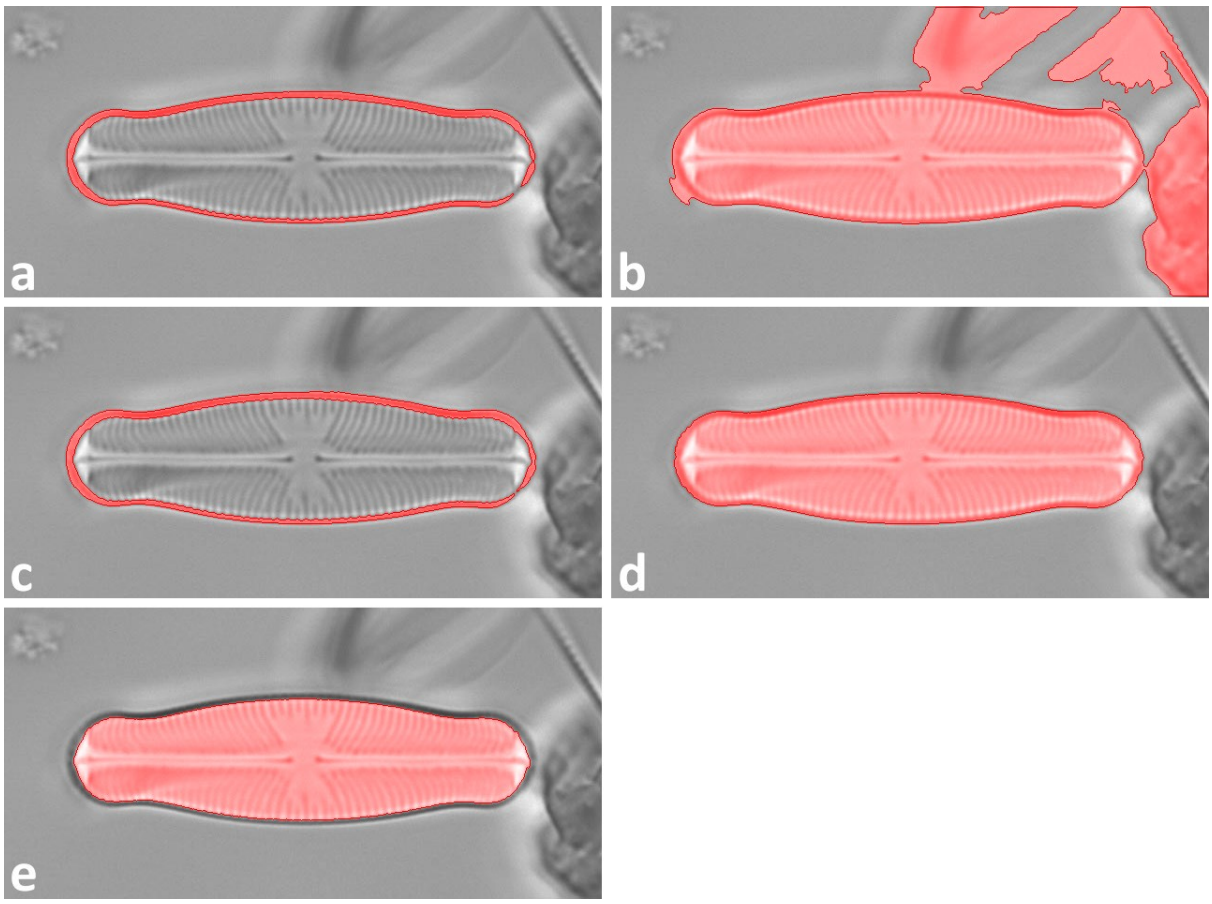


Figure 4: Multiple shapes (highlighted red) detected for a diatom valve according to different segmentation procedures (compare to Figure 3). a) Otsu's thresholding, b) Otsu's thresholding combined with histogram equalization, c) robust automated threshold selector (RATS), d) adaptive thresholding, e) Canny edge detector. Only the result matching best to one of the templates (according to elliptic Fourier analysis, see below under "Shape identification") is taken for analysis.

Shape detection is accomplished by following each object outline using an algorithm by (Sklansky 1982). The outer object contour is the starting point for subsequent analysis steps.

2.3.3.3 Shape processing and analysis

Shapes derived from image processing might be flawed due to segmentation problems or overlapping objects, and they can depict anything from objects of interest to debris and foreign particles. To increase the yield of usable results and to sort out irrelevant data, shapes can be optimized and are evaluated according to their chance of depicting a relevant object.

Shape validation reduces the amount of data to be analyzed to speed up the analysis processes. Each image's segmentation can result in hundreds or even thousands of separate objects, with most of them usually not depicting relevant ones (see Figure 5). Objects will be rejected if their size is outside a user defined range, or if they are within close proximity to the image border, where the chance is high that they were truncated by the camera's field of view.

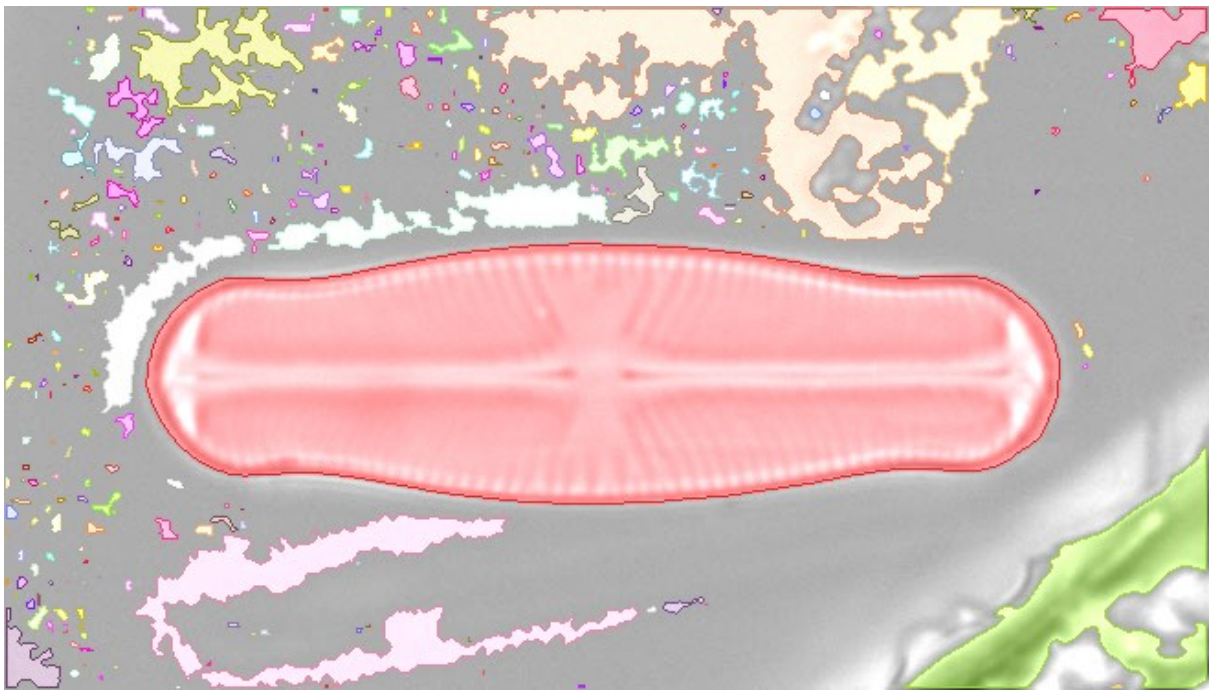


Figure 5: Shapes detected after segmentation (highlighted in different colors). Most of them do not depict relevant objects. Only the shape of the diatom valve will pass validation, other objects are too small or too close to the image border and hence are excluded from further analysis.

Contour optimization can optionally be applied to increase the yield of usable results. Due to debris, overlapping structures, damages or segmentation flaws, not all objects can be segmented successfully. However, some contours can be "repaired" by applying morphological operators (Gonzalez and Woods 2007) "Opening", "Closing" and combinations of these two (see Figure 6). Small indentations and bulges are removed this way and the yield of usable results can increase significantly, but at the expense of accuracy of the derived outlines, reliability of the convexity defect measures (see below), and processing time. For each object, only the result matching best to one of the templates (see "Shape identification" below) is taken for further analysis.

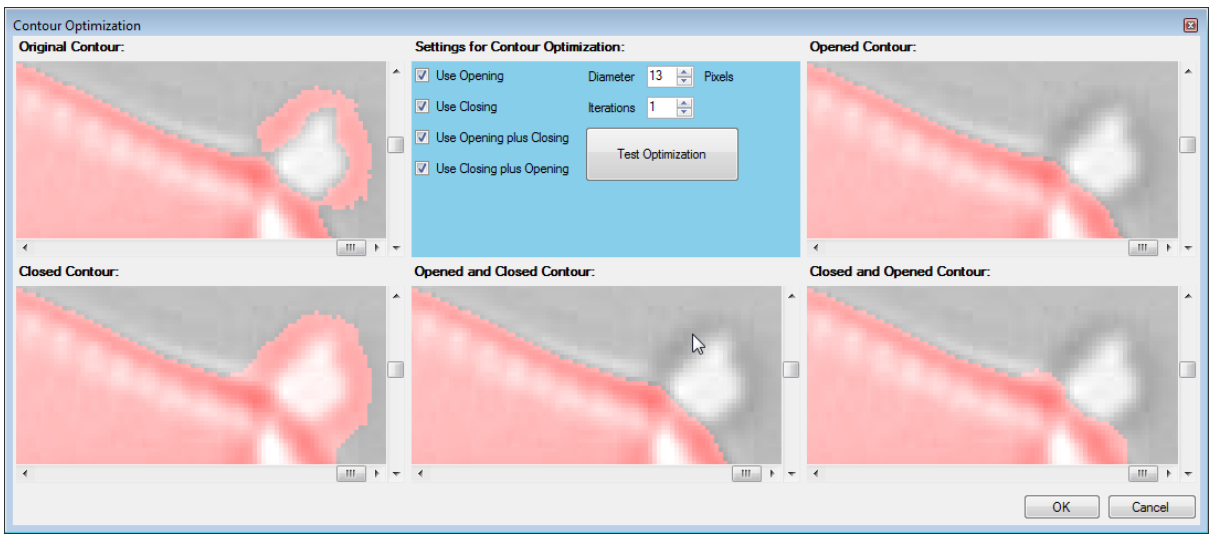


Figure 6: Effects of contour optimization, shapes are highlighted in red. The bulge of the original contour (see top left) can be eliminated successfully by applying morphological opening (see top right) or opening followed by closing (see bottom center).

Manual rework is an option if a shape is distorted due to segmentation flaws, but the corresponding object is essential as a valid result. SHERPA offers functions for redrawing a contour like in a painting

program, for smoothing it and for applying morphological operators (see above) with individual settings to it, as well as to expand the outline to its convex hull.

Shape identification identifies objects by comparing their shapes with templates via elliptic Fourier analysis (Claude 2008, Claude 2010). Matching is accomplished by summing up the squared differences of the normalized elliptic Fourier descriptors of object and template outline; the template having the lowest matching value is assigned to the object. The number of harmonics to be used for Fourier analysis is configurable, appropriate base points are assigned along the object perimeter at steady intervals, with the starting point being the leftmost point with respect to the major axis (see Figure 7).

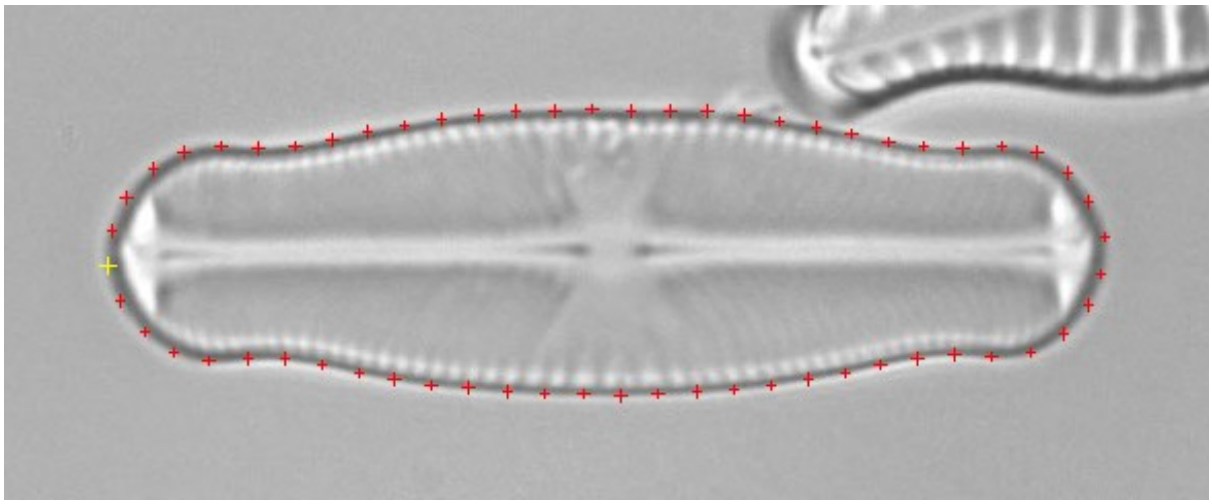


Figure 7: Base points (colored crosses) used for elliptic Fourier analysis, spaced equally along the object outline. The starting point is highlighted yellow.

2.3.3.4 Rating and ranking

The assignment of template and object can be incorrect either because no matching template is available, or because the object shape is distorted due to imperfect segmentation. To estimate the chance of a shape to represent a relevant object, two groups of criteria are evaluated. The first type of criteria judges the quality of shape identification plus some object features (see “Matching and quality indicators” below and Table 2), whereas the second type provides information about contour convexity (see “Convexity defect measures” below and Table 3). The user can define cut-off values for each criterion. Results are ranked by the number of criteria they fulfill. Appropriate cut-off values will depend on a number of factors, including types of objects of interest and representativeness of the template set. A guide on how to achieve appropriate settings is provided along with SHERPA’s documentation.

Matching and quality indicators rate the matching between shape and template and some properties which help to distinguish objects of interest from irrelevant ones, like e.g. width / height-ratio and standard deviation of the texture gray levels within the central part of the object (see Table 2).

Table 2: Matching and quality indicators used for ranking

EFDIs Match with Template	Matching between elliptic Fourier descriptor invariants (EFDIs) of object and template shape (Claude 2008, Claude 2010)**.**
Hu Match for EFDIs Template	Matching between the Hu invariants(Hu 1962) of the object and the template which matches best according to EFDIs.
Optimization Method	Morphological Operator used to improve the object contour. If an optimization was applied to derive a shape, its ranking is degraded, because the resulting outline might be inaccurate.
Standard Deviation of inner 50 %	Standard deviation of the gray level distribution within the object boundaries. Only the inner 50 % of the area are analyzed. This way, diatom valves, normally containing striae / costae / areolae, can be distinguished from empty girdle bands which can produce good outline matching but have a homogenous interior.
Width / Height Ratio	Ratio between object width and height. Usually objects of a certain type have a ratio within a certain range.
Contour Smoothness	Estimation of the object contour smoothness. The actual object outline usually is quite smooth, especially for diatom valves, whilst contours distorted by segmentation inaccuracies or failures usually are rough. The ratio between the outline perimeter and that of the outline smoothed by a Gaussian filter provides information about the contour smoothness.
Formfactor	Heuristic descriptor “formfactor” (Russ 2011)

Convexity defect measures (CDMs) are calculated based on differences of area and / or perimeter between a contour and its convex hull, the latter being the smallest area which encloses the contour without containing any concave parts.

Table 3: Convexity defect measures used for ranking

Absolute measures	
CDF	“Convexity Defection Factor”, depicts the percentaged difference between area resp. perimeter of contour and convex hull (Kloster 2013)
PCAF	The “Percent Concave Area Fraction” compares the areas of contour and convex hull (Nafe and Schlote 2002).
CHMDF	For the “Convex Hull Maximum Distance Factor” each convexity defect’s maximum distance between contour and convex hull is calculated. For distances larger than $\sqrt{2} \cdot$ pixelwidth the squares of the distances are summed up to the CHMDF (Kloster 2013).
Relative measures	
CDF-Match	Ratio of CDF of object and template
PCAF-Match	Ratio of PCAF of object and template
Compactness-Match	Ratio of heuristic descriptor “compactness” between object and template shape

Absolute measures result from the object and are judged directly by their values, relative measures result from comparing values between object and best matching template.

If only convex shapes are of interest, these measures (see Table 3, “Absolute measures”) are excellent features to decide about segmentation quality. This is because for convex shapes, even small indentations or bulges caused by erroneous segmentation will produce noticeable concave

parts within the outline (see Figure 8), which significantly increase the CDMs. When enabling the setting “Force Convexity” in SHERPA, only absolute values of the object’s CDMs are evaluated, and only convex templates are taken into consideration. When doing so, most segmentation problems are detected clearly, and segmentation quality can be judged quite precisely based on absolute values of the convexity defect measures.

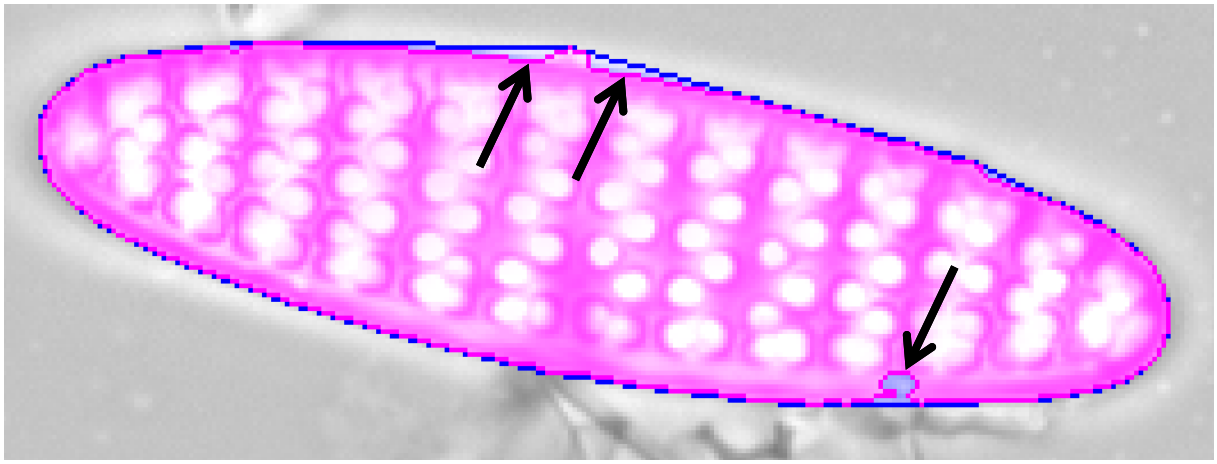


Figure 8: Typical convexity defects. The object area is highlighted in purple, its convex hull in blue. Black arrows depict significant convexity defects caused by segmentation faults, resulting in indentions resp. bulges of the contour outline.

This approach will not work for objects which naturally contain concave parts. If the data contains convex as well as concave objects, SHERPA’s feature “Use Convexity” can be activated. In this case, only if the best matching template is convex, CDMs are evaluated by their absolute values derived from the respective object shape (like when using “Force Convexity”). If the best matching template is concave, some CDMs plus the heuristic descriptor “compactness” (Russ 2011) of the object will be compared to those derived from the best matching template (see Table 3, “Relative measures”).

When the set of objects to be detected contains both convex and concave outlines and convexity analysis is employed (i.e. “Use Convexity” or “Force Convexity” is enabled), the template set should be composed with special care. The situation to be avoided is that the best match of a concave object becomes a convex template, which can happen if no proper concave template is provided. In this case, the object convexity will be judged by absolute values even though it is concave, which will result in a failure of convexity defect measures and hence in a poor ranking.

If neither “Use Convexity” nor “Force Convexity” are activated, only a relative comparison of some CDMs between object and template plus an evaluation of the form factor takes place, regardless if the best matching template is convex or concave. The object’s CDMs are not judged directly. This is usually a good choice if it is not known in advance if all relevant objects are convex and / or there is no extensive library of templates yet.

It should be noted that detection of segmentation flaws is much less accurate when an object’s convexity defect measures are compared to those of the template instead of being judged by their absolute values. So if only convex objects are of interest, choosing “Force Convexity” will provide a more precise ranking and might save some manual reviewing.

Heuristic descriptors rectangularity (du Buf and Bayer 2002), ellipticity (Rosin 2003), triangularity (Rosin 2003), roundness (Russ 2011) and convexity (Zunic and Rosin 2002, Russ 2011) are calculated for exporting but not evaluated by SHERPA.

2.3.3.5 Review, rework and selection of results

Analysis results can be reviewed for verification and for selecting data to be exported in a comfortable manner (see Figure 9). For each object passing validation (see above under “Shape processing and analysis”), the path to the original image file the object was found in, the name of the segmentation method, the path to the best matching template file, values of basic morphometric variables (e.g. width, height), values of quality and convexity defect measures, and ranking are displayed. Objects detected can be displayed, along with their detected outlines, their enclosing convex hull, the points used for elliptic Fourier analysis as well as their best matching templates. Shapes containing segmentation errors can be reworked manually to increase the yield of usable results. Quality indicators, rankings and morphometric variables are updated after manual reworking.

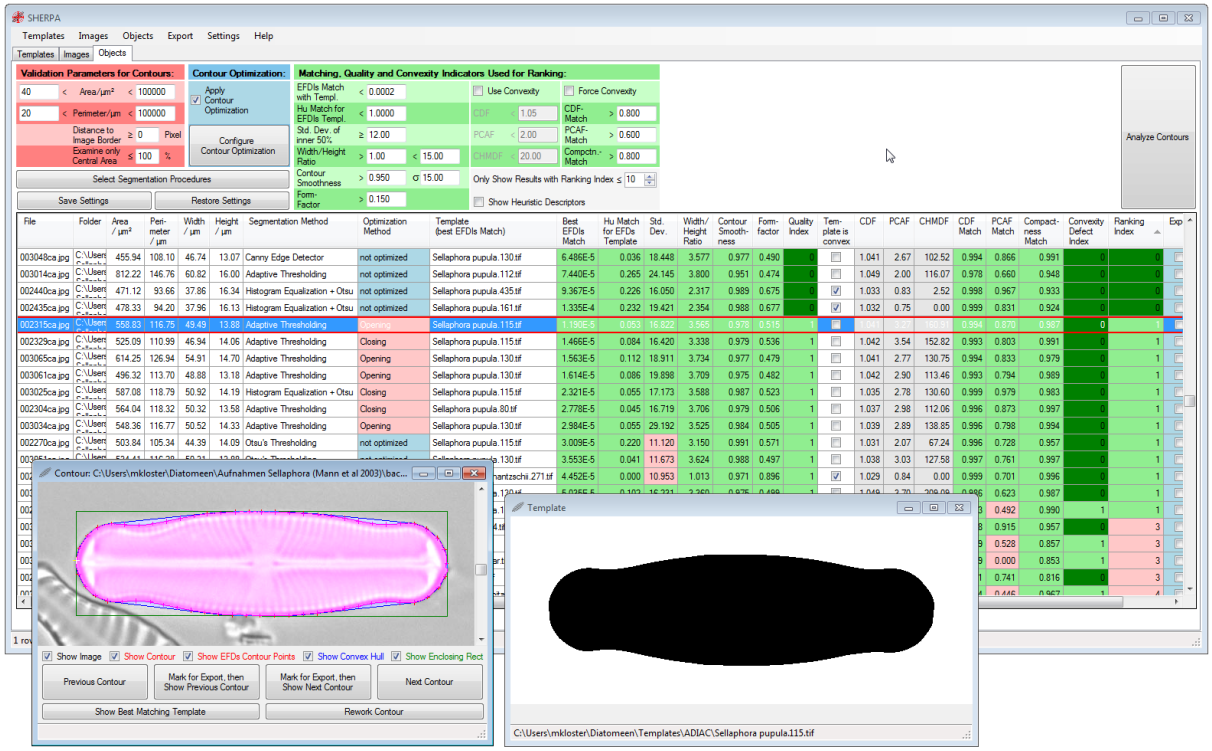


Figure 9: Screenshot of SHERPA. Analysis settings and results (background), a single result (bottom left, detected object highlighted in purple) and its best matching template (bottom center) are displayed.

2.3.3.6 Data export

Selected results can be exported to a set of CSV and TIFF files for further morphometric analysis using tools like e.g. “R” (R Core Team 2015). Results can be exported to a table containing all the information displayed by SHERPA, plus some additional morphometric values (see Table 4). All relevant settings of SHERPA used to create these results are stored into a separate file. Optionally, the image data cropped to the object region, the coordinates of the object outline, the coordinates of the outline points used for elliptic Fourier analysis, and the resulting descriptors can be exported to separate files for each result. Detailed information on all features is included in the manual and the “Technical Details” document linked within SHERPA’s help menu.

Table 4: Exportable features

Name of feature	Description
Source Image	Path to raw image data file
Area	Object area
Perimeter	Object perimeter
Width	Object width (along major axis)
Height	Object height (perpendicular to major axis)
Rotation Angle	Rotation angle of the major axis
Segmentation Method	Segmentation method used to derive the object shape
Optimization Method	Optimization method applied to the object shape
Best Template (EFDIs)	Path to the best matching template (according to matching of elliptic Fourier descriptor invariants)
Template Difference (EFDIs)	Value for matching of elliptic Fourier descriptor invariants between object and best matching template
Hu-match for best EFDIs-Template	Value of matching of Hu invariants between object shape and best matching template
Standard Deviation	Standard deviation of texture gray levels within the inner 50 % of the object boundaries
Width/Height-Ratio	Aspect ratio of the object shape
Smoothed Perimeter Ratio	Ratio between the perimeters of the smoothed and the original contour; smoothing is performed by Gaussian filtering of the contour coordinates.
Quality Index	Number of fulfilled quality indicators
Template is convex	Indicator showing if the best matching template is convex
Convexity is used	Indicator showing if convexity was judged directly to calculate convexity indicators (use of absolute convexity measures)
Rectangularity	Heuristic descriptor
Compactness	Heuristic descriptor
Ellipticity	Heuristic descriptor
Triangularity	Heuristic descriptor
Roundness	Heuristic descriptor
Convexity by perimeter	Heuristic descriptor
Convexity by area	Heuristic descriptor
Formfactor	Heuristic descriptor
CDF	Convexity defect measure
PCAF	Convexity defect measure
CHMDF	Convexity defect measure
CDF-Match	Ratio of CDF between object and template
PCAF-Match	Ratio of PCAF between object and template
Compactness-Match	Ratio of heuristic descriptor “formfactor” between object and template
Convexity Defect Index	Number of fulfilled absolute or relative convexity indicators
Ranking Index	Ranking for object shape, i.e. estimation of quality and relevance of result
Contour Image	Name of the file containing the image data cropped to the object area

Contour Image top left Corner	Coordinates of the top left corner of the cropped object image with respect to the raw data
Image Moments (mu)	Image moments of the object shape
Hu Invariants (Hu)	Hu-Invariants of the object shape

2.3.4 Results and discussion

For the following analyses, bright field micrographs of valves of different diatom species and from different sources were analyzed. All results were produced without manually reworking or resorting detected shapes, relying solely on SHERPA's automated functions for segmentation, contour optimization and result ranking.

2.3.4.1 Templates

To facilitate use of SHERPA for generic diatom recognition and analysis, we prepared a library covering a wide range of diatom outline shapes, containing about 450 templates. This compilation is mainly based on the outline shape classification scheme and accompanying diagrams from (Barber and Haworth 1981), *Fragilariopsis* data sets from a surface sediment sample (Zielinski et al. 1998), and upon the extensive ADIAC diatom image database available online (du Buf 2003), although the ADIAC data is not fully covered by the current template library. For the latter two, SHERPA was used for image segmentation to detect shapes previously not represented in the template set: Shapes with a poor template matching value were screened manually. If they were depicting relevant valves and segmentation quality was satisfactory, they were converted into additional templates employing the built-in functions of SHERPA. Because diatom shapes vary widely among taxa, as well as during the life cycle of even a single taxon, it is crucial to check the presence of a representative set of templates for taxa of interest when using SHERPA for analyzing a particular type of diatom samples.

2.3.4.2 **Sellaphora** *data as example for identification accuracy*

To demonstrate the usability of SHERPA, we analyzed a set of images from one of the classical model taxa of diatom microdiversity, the *Sellaphora pupula* (Kützing) Mereschkowsky complex s.l. *S. pupula* has been known as a morphologically highly variable diatom species during most of the 20th century. However, Mann and colleagues demonstrated in a series of papers (cumulating in (Mann et al. 2004b)) that sympatric demes of this diatom "species" formed reproductively isolated groups, that could also be diagnosed using molecular markers and also differed in minute morphological / morphometric features, including (but not limited to) minor differences in their valve outlines. In their 2004 investigation (Mann et al. 2004b), Mann et al. used Legendre-polynomials and contour segment analysis for comparing outline morphology of six *S. pupula* demes (since that study, also formally recognized as distinct species). They made the images upon which the analyses were based publicly available (Mann et al. 2004a), which we used in this analysis.

All five segmentation methods plus contour optimization were applied to analyze a total of 383 micrographs focused on the outlines of *Sellaphora* valves (see Table 5). Most of the valves were clearly isolated, without overlapping structures and only little amount of debris, so this might not be a typical data set, but serves as an example on how specific the identification process works. Since contours of *S. pupula* contain concave parts, convexity was not taken into account for judging segmentation quality directly (i.e. neither "Use convexity" nor "Force convexity" were activated in SHERPA).

Table 5: Results analyzing 383 images (Mann et al. 2004a) depicting *Sellaphora* valves (plus one centric diatom)

	Identified as *Sellaphora pupula*	**Identified as other [1)]**
Ranking 0	318	4
Ranking 1	25	7
Ranking 2	2	1

[1)] One centric diatom was present in the data set, the other valves identified as being not *Sellaphora* have a similar shape and therefore cannot be distinguished when using the large template set.
All five segmentation methods were used (RATS with σ range 1 to 11) and contour optimization was applied.

Considering only results of ranking 0 to 2, which usually is the range for objects without significant segmentation flaws and good coverage by templates, 357 (93 %) of the valves contained in the data set were successfully segmented (see Figure 10). When using the comprehensive template library, most of the results were assigned correctly to one of the 18 *Sellaphora pupula* templates derived from the ADIAC dataset (no template was created from the *Sellaphora* data set itself). Only about 3 % of the results were assigned to templates of other species, which had shapes very similar to *S. pupula*. One centric diatom was actually present in the data and correctly identified as a disc-shaped type, clearly distinct from the others. When using only the 18 *Sellaphora pupula* templates instead of the whole template library, the yield was identical (apart from the single centric diatom), with all valves correctly identified.

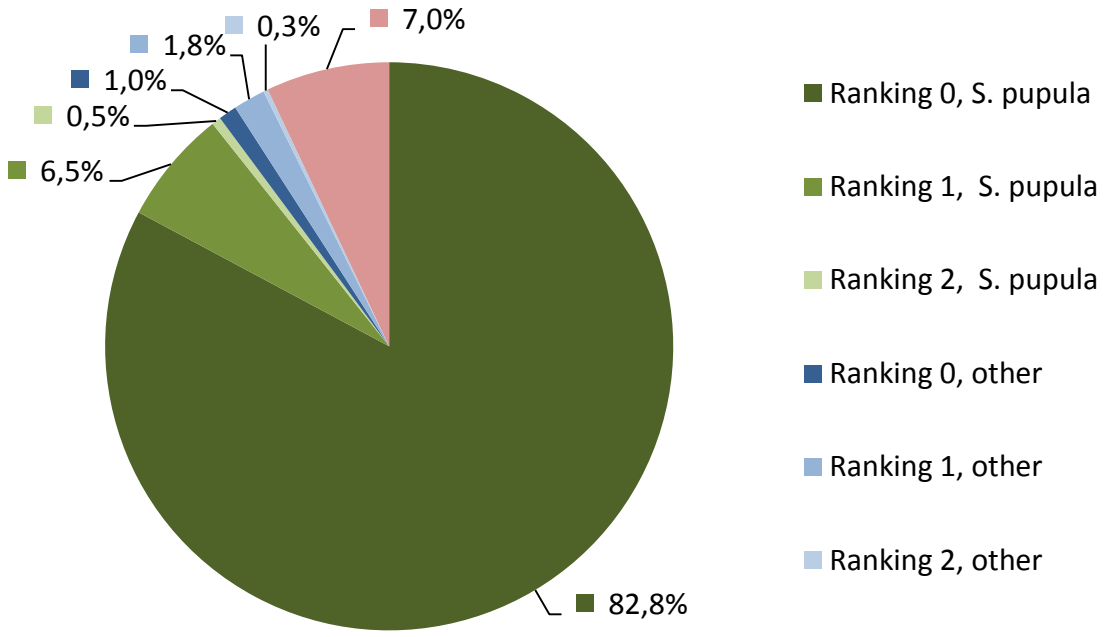


Figure 10: Percentage of different rankings and identifications for the *Sellaphora* data set (compare Table 5). About 93 % of the valves were segmented successfully (green and blue), about 90 % were identified correctly as *S. pupula* (green), about 7 % were not segmented successfully (red).

Results having a ranking above 2 are not listed, because they were caused by partly unfocused outlines, overlapping objects or debris and would have needed manual inspection and reworking.

2.3.4.3 Fragilariopsis *data as example for segmentation quality*

As a typical data set, 773 micrographs originating from sediment core PS1768-8 (Zielinski et al. 1998) and mainly showing *Fragilariopsis kerguelensis,* plus broken valves, debris and overlapping objects, were analyzed. The data was obtained using a Metafer slide scanning system (Metasystems,

Altlussheim, Germany), applying the implemented autofocus and stacking functions. Because not all valves were lying parallel to the focal plane, outlines were partly out of focus or blurred despite of stacking. Since the outline of *F. kerguelensis* is completely convex, SHERPA's "Force Convexity" feature was used to improve judging of segmentation quality.

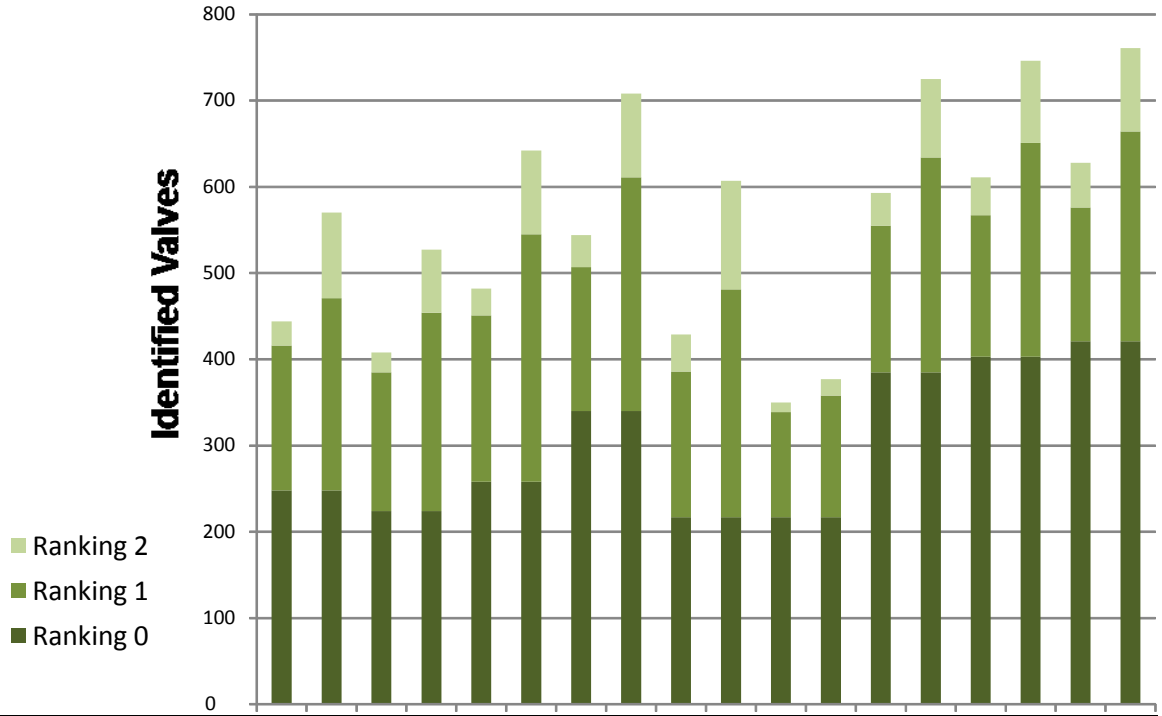
Again, the full template set covering a broad range of diatom species was used. Although *Fragilariopsis* valves were mostly identified correctly, some were assigned to templates of other similarly shaped species, and some correctly identified valves of other species were present. Undamaged valves could successfully be distinguished from artifacts like broken ones or debris. In some cases, objects like girdle bands or spherical structures were identified as relevant valves (usually at a ranking index 2 or worse), because of their shape similar to those of other diatom species in the template library. This problem can be overcome by using only *Fragilariopsis* templates.

All segmentation methods available in SHERPA were applied separately, as well as in combination, to compare the yield of usable results (see Table 6 and Figure 11). As expected, the best yield is achieved when using all segmentation methods, employing RATS with a wide range of σ, and applying contour optimization. When combining the individual strengths of the different methods plus contour optimization, even objects which are difficult to segment can be handled successfully; although not always without contour inaccuracies (see Figure 12). Since applying the whole range of methods drastically increases the time needed for analysis, using only Otsu's thresholding, Canny edge detector, adaptive thresholding and Otsu's thresholding plus histogram equalization might be a practicable choice for preliminary or quick analyses.

Table 6: Results for *Fragilariopsis* data for different combinations of segmentations methods and contour optimization

Otsu's thresholding	Histogram equalization	RATS (σ = 3)	RATS (σ = 1-11)	Adaptive thresholding	Canny edge detector	Contour optimization	Ranking 0 total	Ranking 1 total	Ranking 2 total[2)]	Total ranking 0 to 2
✓							248	168	28	444
✓						✓	248	223	99	570
	✓						224	161	23	408
	✓					✓	224	230	73	527
		✓					258	193	31	482
		✓				✓	258	287	97	642
			✓				340	167	37	544
			✓			✓	340	271	97	708
				✓			217	169	43	429
				✓		✓	217	264	126	607
					✓		217	122	11	350
					✓	✓	217	141	19	377
✓	✓			✓	✓		385	170	38	593
✓	✓			✓	✓	✓	385	249	91	725
✓	✓	✓		✓	✓		403	164	44	611
✓	✓	✓		✓	✓	✓	403	248	95	746
✓	✓		✓	✓	✓		421	155	52	628
✓	✓		✓	✓	✓	✓	421	243	97	761

[2)] Whilst results of ranking 0 and 1 contain nearly only correctly segmented valves of *Fragilariopsis* and a few of other species, ranking 2 also contains few results of girdle bands incorrectly identified as valves.
The more methods are combined, the higher is the yield.



Otsu's thresholding	✓	✓										✓	✓	✓	✓	✓	✓
Histogram eq. + Otsu			✓	✓								✓	✓	✓	✓	✓	✓
RATS (σ = 3)					✓	✓								✓	✓		
RATS (σ = 1-11)							✓	✓								✓	✓
Adaptive thresholding									✓	✓			✓	✓	✓	✓	✓
Canny edge detector											✓	✓	✓	✓	✓	✓	✓
Contour optimization		✓		✓		✓		✓		✓		✓		✓		✓	

Figure 11: Results for *Fragilariopsis* data for different combinations of segmentations methods and contour optimization (compare Table 6). The more methods are combined, the higher is the yield.

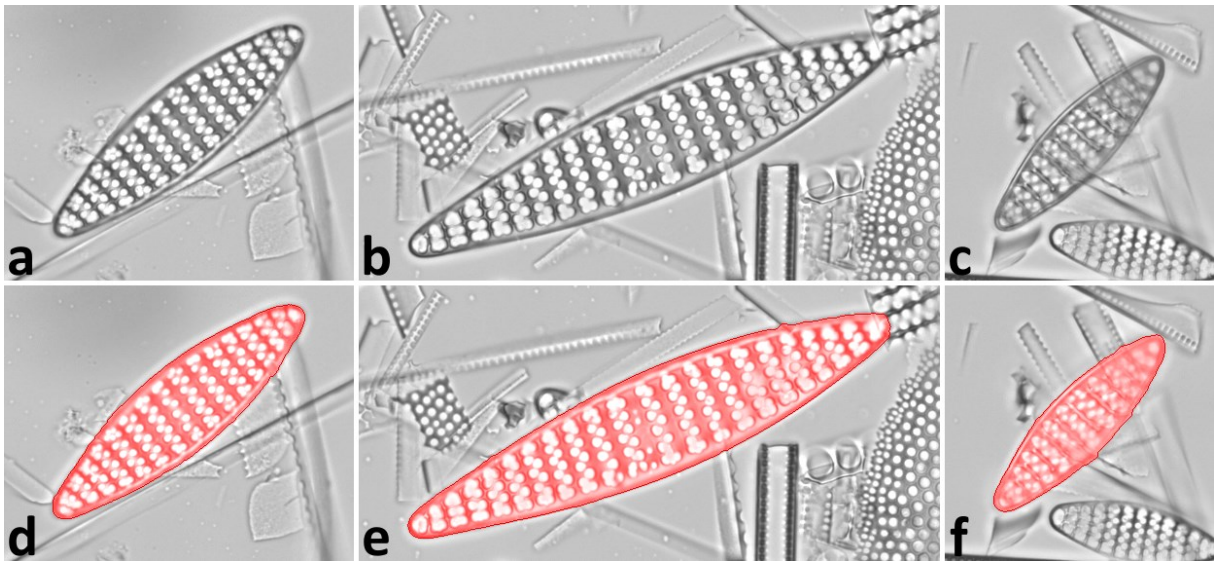


Figure 12: Successful segmentation in the presence of debris and overlapping objects. a) – c) Micrographs of *Fragilariopsis* valves, d) – f) segmented shapes (highlighted red) after contour optimization, using d) adaptive thresholding, e) Otsu's thresholding, f) RATS (σ = 3.0). By application of multiple segmentation methods and contour optimization even problematic objects could be extracted, since often at least one of the methods succeeded, but partly at the expense of contour accuracy (see the small bulges in the object contours).

2.3.4.4 Comparison of segmentation methods

88 valves of the *Fragilariopsis* data were successfully segmented by each of the five segmentation methods (RATS with σ = 3.0) without applying contour optimization. Area, perimeter, width and height obtained by the different segmentation methods were compared by calculating their percentage deviation for each of these valves. The deviations for all valves were compared (see Equation 1). This illustrates the variation of the object contours produced by the different segmentation methods, which is about ±1 % around the center value between the minimum / maximum values (see Figure 13).

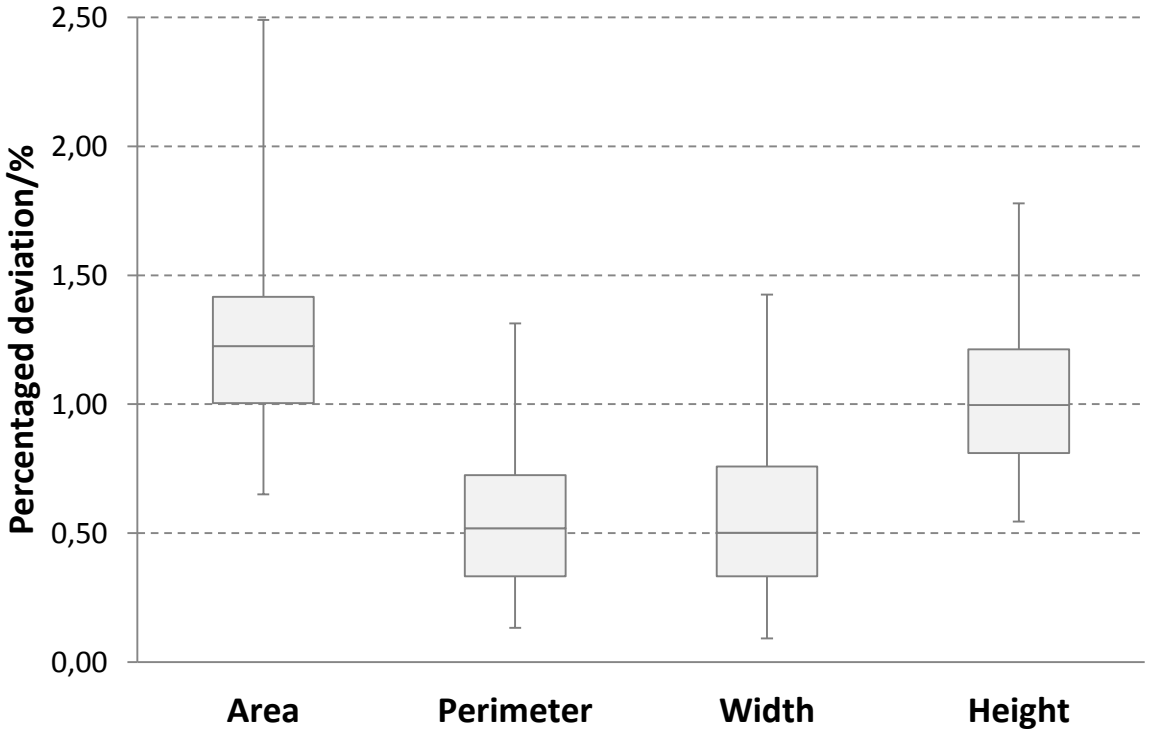


Figure 13: Percentaged deviation of features around the minimum / maximum center when using all five segmentation methods. The deviation is about ± 1 % around the center value.

$$Percentaged\ deviation = \frac{MAX - MIN}{MAX + MIN} \cdot 100\ \% \qquad \textbf{(Equation 1)}$$

With *MAX* = maximum, *MIN* = minimum value for a feature (area, perimeter, etc.) when using multiple segmentation methods.

2.3.4.5 Further analysis using R

As a benchmark experiment, and to illustrate how data exported by SHERPA can be used in further analyses, we imported both the classical morphometric features and the elliptic Fourier descriptors (EFDs) calculated by SHERPA for the 356 *Sellaphora* valves from the first above described experiment into the open source statistical data analysis environment R (R Core Team 2015). In R, we reproduced those plots from (Mann et al. 2004b) for which features used were captured by SHERPA (see Figure 14; besides outline features, Mann et al. also measured a number of features characterizing striae density, orientation and the terminal bars which are not captured by SHERPA).

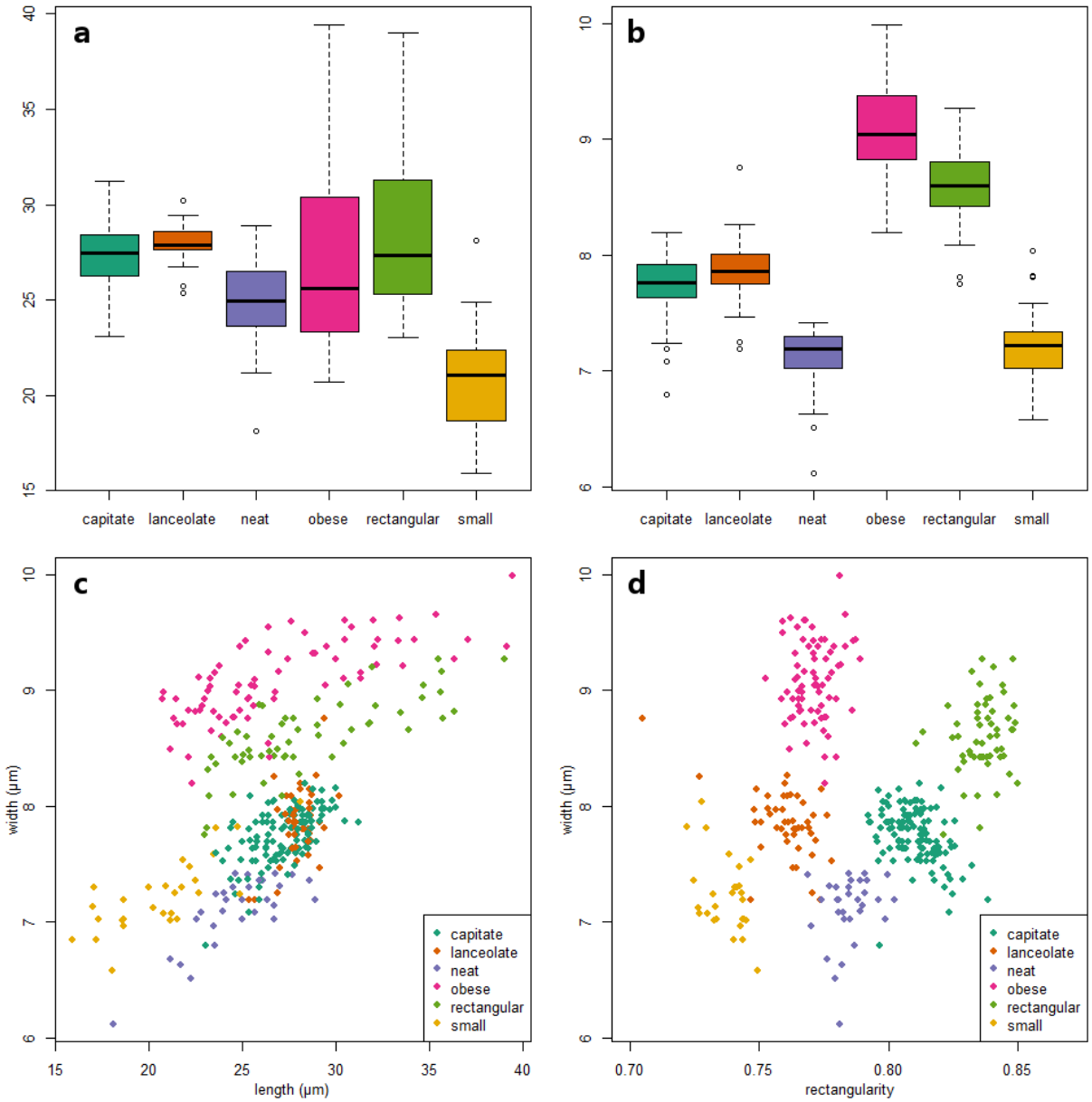


Figure 14: Reproduction of plots from (Mann et al. 2004b) using the same variables. a) valve length, b) valve width, c) valve width vs. length, d) valve width vs. rectangularity, corresponding to Figures 5, 6, 10 and 14 from (Mann et al. 2004b). In the box plots in a) and b), the thick horizontal lines represent the medians; the boxes the range from the first to the third quartile; and whiskers +/− 1.58 times the interquartile range. Individual values outside these ranges are displayed as circles.

The plots correspond to Figures 5, 6, 10 and 14 from Mann et al., based on valve length, width and rectangularity. These figures rather accurately correspond to those in the original publication, with the exception of a single “lanceolate” valve with an extremely low rectangularity value of 0.705: such a low value does not appear in the original publication and it is also extremely low when compared with the other values exported from SHERPA. This outlier reflects a segmentation problem caused by a shadow overlapping the valve outline which can easily be fixed using the “Manual rework” feature of SHERPA, resulting in a rectangularity value of 0.757 which hardly differs from the value given for the same valve by Mann et al. (0.760). In order to illustrate the accuracy of the methods when applied in a fully unsupervised manner, we opted to keep the original value for Figure 14 a) and for the following classification exercise. When applying a cross-validation linear discriminant analysis based on classical morphometric features extracted by SHERPA (randomly selected 50 % of objects used to train the model, the remaining 50 % is then classified against it, in 100 iterations), classification accuracies of the six demes (species) range from 98.9 % to 100 % (median: 100 %).

EFDIs performed less well in linear discriminant analysis (77.5 - 92.7 % accuracy, median: 88.2 %, in an identical cross-validation, see Figure 15), but the classical morphometric features still demonstrate that the set of features extracted by SHERPA provides a robust basis for downstream outline-based classification, especially when considering the small differences in outline shapes among the *Sellaphora* groups.

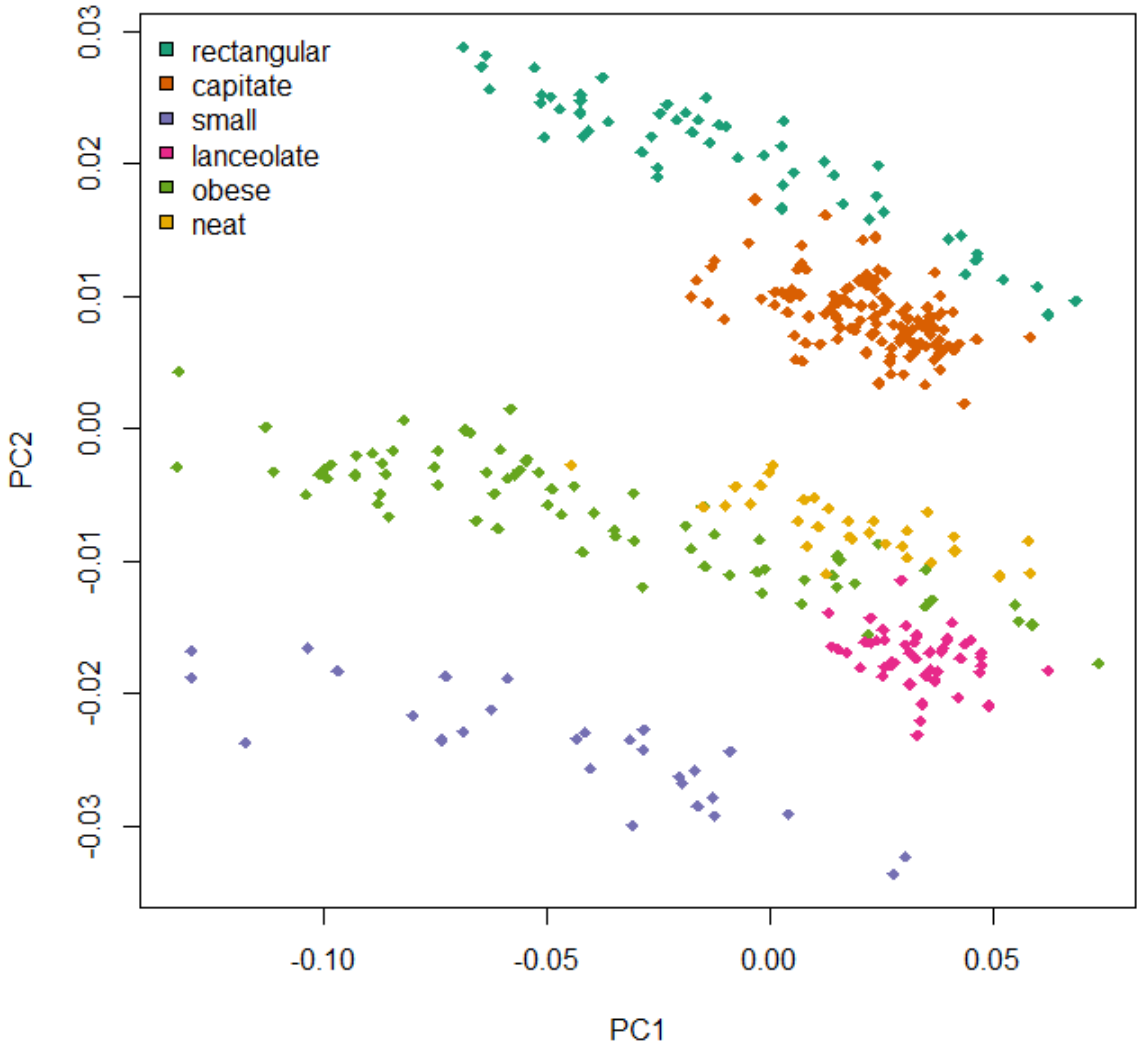


Figure 15: Principal component analysis of elliptic Fourier descriptor invariants for the *Sellaphora* data set. EFDIs have a comparable discriminatory power to the Legendre polynomials used by (Mann et al. 2004b), differentiating the three main shape groups but not the individual demes / species within each shape group.

2.3.4.6 Future development

Besides improving performance, the next steps in SHERPA's development will concern the analysis of texture und structural features to improve versatility and identification specificity.

2.3.5 Conclusions

SHERPA provides a useful tool for diatom identification and morphometrics, enabling mass screenings, since it greatly reduces the amount of work needed to be performed by human interaction. Manual revision required for best results can be accomplished in a quick and effective manner, supported by a ranking based on matching and quality indicators.

The degree of identification reliability reflects both the range of templates used and the diversity present in the analyzed samples. In spite of depending solely on outline shape, good identification accuracy can be reached using customized template sets. Combining multiple segmentation methods

improves the identification rate without significantly impairing result accuracy, and, combined with contour optimization, even objects showing segmentation artifacts can be analyzed successfully. For convex shapes, convexity defect measures provide an effective way to judge segmentation quality, hence allowing identification of flawed object outlines.

The approach of restricting SHERPA to the identification of relevant objects and the calculation of their morphometric features enables an adaptation to specific problems / target taxa. Downstream analyzes or classification can be performed using widely available commercial or free statistical software tools, e.g. “R”.

Availability and requirements

Project name: SHERPA.
Project home page: http://www.awi.de/sherpa
Operating system(s): Windows7 64 Bit (32 Bit version available)
Programming language: C#
Other requirements: .NET 4.0
License: Freeware, royalty-free, non-exclusive
Any restrictions to use by non-academics: none

Abbreviations

CDF, “Convexity Defection Factor”, a convexity defect measure; CDMs, Convexity defect measures; CHMDF, “Convex Hull Maximum Distance Factor”, a convexity defect measure; EFDs, Elliptic Fourier descriptors; EFDIs, Elliptic Fourier descriptor invariants; PCAF, “Percent Concave Area Fraction”, a convexity defect measure; SHERPA, Tool for “**SH**ap**e R**ecognition, **P**rocessing and **A**nalysis”

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

MK developed SHERPA and its image processing workflow, performed the data analyses, and is the main author of this paper. GK mentored the beginning steps of SHERPA (at this point called “DiatoMorphoTo” and part of MK’s master thesis) and revised the manuscript. BB strongly contributed to the morphometric aspects of SHERPA as well as this paper, took care of the “R” part and was the main information source on diatom taxonomy. All authors read and approved the final manuscript.

Authors’ information

MK started developing SHERPA as part of his master thesis (at that time called “DiatoMorphoTo”) at the HSEL, supervised by GK and in collaboration with BB. Since graduation he works at the Friedrich Hustedt Diatom Study Centre, AWI, under supervision of BB to develop SHERPA. He mainly works at the interface between biology and informatics, focusing on image processing, data visualization and automation.

GK is a professor in bioinformatics and has been working for more than 15 years in the areas of genome analysis, microscopy, image processing, image interpretation and development of bioinformatics methods for genome and proteome analysis. His recent works regard high performance reconstruction of structures from high resolution image stacks of extensive microscopic objects, and limited three-dimensional reconstruction from stereoscopic images of biological tissues and organisms. GK is also the author of E.L.M.I. (Expert System for Light Microscopy).

BB is a diatomist / bioinformaticist, curator of the Hustedt Diatom Study Centre. His research currently focuses on taxonomy, biogeography and morphometrics of Antarctic diatoms.

Acknowledgements
Thanks to Rainer Gersonde for providing the slides from sediment core PS1768-8 used for the *Fragilariopsis* test, and to Nike Fuchs and Fabian Altvater for scanning and sorting the *Fragilariopsis* images used.

2.3.6 References

Abarca, N., Jahn, R., Zimmermann, J. and Enke, N. (2014). "Does the cosmopolitan diatom *Gomphonema parvulum* (Kützing) Kützing have a biogeography?" Plos One **9**(1): e86885.

Álvarez-Borrego, J. and Solorza, S. (2010). "Comparative analysis of several digital methods to recognize diatoms." Hidrobiológica **20**: 158-170.

Barber, H. G. and Haworth, E. Y. (1981). A guide to the morphology of the diatom frustule. Ambleside, Cumbria, UK, Freshwater Biological Association.

Bradski, G. and Kaehler, A. (2008). Learning OpenCV : Computer Vision with the OpenCV Library. Sebastopol, O'Reilly.

Canny, J. (1986). "A Computational Approach to Edge Detection." IEEE Transactions on Pattern Analysis and Machine Intelligence **30**: 125-147.

Claude, J. (2008). Morphometrics with R. New York, Springer Science+Business Media, LLC.

Claude, J. (2010). "Morphometrics with R - Errata 1.81" from
http://www.isem.univ-montp2.fr/recherche/files/2012/01/Morphometrics_errata1.81.pdf

Cortese, G. and Gersonde, R. (2007). "Morphometric variability in the diatom *Fragilariopsis kerguelensis*: Implications for Southern Ocean paleoceanography." Earth and Planetary Science Letters **257**(3-4): 526-544.

Cortese, G., Gersonde, R., Maschner, K. and Medley, P. (2012). "Glacial-interglacial size variability in the diatom *Fragilariopsis kerguelensis*: Possible iron/dust controls?" Paleoceanography **27**.

Crawford, R. M., Hinz, F. and Rynearson, T. (1997). "Spatial and temporal distribution of assemblages of the diatom *Corethron criophilum* in the Polar Frontal region of the South Atlantic." Deep Sea Research Part II: Topical Studies in Oceanography **44**(1): 479-496.

Droop, S. J., Mann, D. G. and Lokhorst, G. M. (2000). "Spatial and temporal stability of demes in *Diploneis smithii*/*D. fusca* (Bacillariophyta) supports a narrow species concept." Phycologia **39**(6): 527-546.

du Buf, H. (2003). "ADIAC public image files" Retrieved 07.02.2013, from
http://rbg-web2.rbge.org.uk/ADIAC/pubdat/downloads/public_images.htm

du Buf, H. and Bayer, M. M. (2002). Automatic Diatom Identification. New Jersey, London, Singapore, Hong Kong, World Scientific Publishing Co. Pte. Ltd.

Emgu (2012). "Emgu CV, a cross platform .NET wrapper for the OpenCV image processing library, Version 2.4.2" from
http://www.emgu.com

Falasco, E., Blanco, S., Bona, F., Goma, J., Hlubikova, D., Novais, M. H., Hoffmann, L. and Ector, L. (2009). "Taxonomy, morphology and distribution of the *Sellaphora stroemii* complex (Bacillariophyceae)." Fottea **9**.

Frankova, M., Poulickova, A., Neustupa, J., Pichrtova, M. and Marvan, P. (2009). "Geometric morphometrics - a sensitive method to distinguish diatom morphospecies: a case study on the sympatric populations of *Reimeria sinuata* and *Gomphonema tergestinum* (Bacillariophyceae) from the River Becva Czech Republic." Nova Hedwigia **88**.

Gonzalez, R. C. and Woods, R. E. (2007). Digital Image Processing, 3rd Ed. Upper Saddle River, New Jersey, Prentice Hall.

Grima, C., Tadeo, F., Álvarez, T. and Arribas, J. L. (2003). Diatoms classification using frequency domain techniques. Jornadas de Automática. León.

Hammer, Ø. (2013). "PAST" from
http://folk.uio.no/ohammer/past/

Hu, M. K. (1962). "Visual Pattern Recognition by Moment Invariants." In IRE Transactions on Information Theory **8**.

IBM Analytics (2013). "SPSS" from
http://www-01.ibm.com/software/analytics/spss/

itseez (2012). "OpenCV (Open Source Computer Vision Library) Version 2.4.2" from
http://opencv.org/

Jewson, D. H., Granin, N. G., Zhdanov, A. A., Gorbunova, L. A., Bondarenko, N. A. and Gnatovsky, R. Y. (2008). "Resting stages and ecology of the planktonic diatom *Aulacoseira skvortzowii* in Lake Baikal." Limnology and Oceanography **53**(3): 1125-1136.

Jewson, D. H., Granin, N. G., Zhdarnov, A. A., Gorbunova, L. A. and Gnatovsky, R. Y. (2010). "Vertical mixing, size change and resting stage formation of the planktonic diatom *Aulacoseira baicalensis*." Eur. J. Phycol. **45**(4): 354-364.

Kermarrec, L., Bouchez, A., Rimet, F. and Humbert, J. F. (2013). "First evidence of the existence of semi-cryptic species and of a phylogeographic structure in the *Gomphonema parvulum* (Kützing) Kützing Complex (Bacillariophyta)." Protist **164**.

Kingston, J. and Pappas, J. L. (2009). "Quantitative shape analysis as a diagnostic and prescriptive tool in determining *Fragilariforma* (Bacillariophyta) taxon status."

Kitware and al., e. (2012). " Insight Segmentation and Registration Toolkit (ITK), Version 4.20" from
http://www.itk.org/

Klingenberg, C. P. (2011). "MorphoJ: an integrated software package for geometric morphometrics." Mol Ecol Resour **11**.

Kloster, M. (2013). Digitale Bildsignalverarbeitung in der Bioinformatik: Methoden zur Segmentierung und Klassifizierung biologischer Merkmale am Beispiel ausgewählter Diatomeen. Emden, University of Applied Sciences Emden/Leer. **Master of Engineering**.

Kumar, M. S. R., Ramaiah, N. and Tang, D. (2009). "Morphometry and cell volumes of diatoms from a tropical estuary of India." Indian Journal of Marine Sciences **38**.

Lehmann, G. (2006). "Robust Automatic Threshold Selection." The Insight Journal **2006** (July - December).

Loke, R. E. and Du Buf, H. (2002). Identification by curvature of convex and concave segments. Automatic diatom identification. H. du Buf and M. M. Bayer. Singapore, World Scientific Publishing.

Luo, Q., Gao, Y., Luo, J., Chen, C., Liang, J. and Yang, C. (2011). "Automatic Identification of Diatoms with Circular Shape using Texture Analysis." Journal of Software **6**(3): 428-435.

Mann, D. G., McDonald, S. M., Bayer, M. M., Droop, S. J. M., Chepurnov, V. A., Loke, R. E., Ciobanu, A. and Buf, J. M. H. d. (2004a). "Algae World: Mann et al. 2004: images and morphometric data" from http://rbg-web2.rbge.org.uk/algae/research/mann_etal_2004_data.html

Mann, D. G., McDonald, S. M., Bayer, M. M., Droop, S. J. M., Chepurnov, V. A., Loke, R. E., Ciobanu, A. and du Buf, J. M. H. (2004b). "The *Sellaphora pupula* species complex (Bacillariophyceae): morphometric analysis, ultrastructure and mating data provide evidence for five new species." Phycologia **43**(4): 459-482.

Mann, D. G. and Vanormelingen, P. (2013). "An inordinate fondness? The number, distributions, and origins of diatom species." Journal of Eukaryotic Microbiology **60**(4): 414-420.

Marchetti, A. and Cassar, N. (2009). "Diatom elemental and morphological changes in response to iron limitation: a brief review with potential paleoceanographic applications." Geobiology **7**(4): 419-431.

Marshall, D., Martin, R., Rosin, P., Hicks, J., Mann, D. G., Droop, S. J. M. and Bayer, M. M. "DIADIST: Diatom and desmid identification by shape and texture." from http://www.cs.cf.ac.uk/diadist/code.htm

Microsoft Corporation (2013a). "Download Microsoft .NET Framework 4 (Web Installer) from Official Microsoft Download Center." from http://www.microsoft.com/en-us/download/details.aspx?id=17851

Microsoft Corporation (2013b). "Download Microsoft Visual C++ 2010 SP1 Redistributable Package (x64) from Official Microsoft Download Center." from http://www.microsoft.com/en-us/download/details.aspx?id=13523

Mou, D. and Stoermer, E. F. (1992). "Separating *Tabellaria* (Bacillariophyceae) shape groups based on Fourier descriptors." Journal of Phycology **28**(3): 386-395.

Nafe, R. and Schlote, W. (2002). "Methods of Shape Analysis of two-dimensional closed Contours - A biologically important, but widely neglected Field in Histopathology." Electronic Journal of Pathology and Histology **8**(2): 1-18.

Otsu, N. (1979). "A Threshold Selection Method from Gray-Level Histograms." IEEE Transactions on Systems, Man and Cybernetics **9**(1): 62-66.

Pappas, J. and Stoermer, E. (2001). Fourier shape analysis and fuzzy measure shape group differentiation of Great Lakes *Asterionella* Hassall (Heterokontophyta, Bacillariophyceae). Proceedings of the 16th International Diatom Symposium.

Poulickova, A., Vesela, J., Neustupa, J. and Skaloud, P. (2010). "Pseudocryptic diversity versus cosmopolitanism in diatoms: a case study on *Navicula cryptocephala* Kütz. (Bacillariophyceae) and morphologically similar taxa." Protist **161**.

Poulíčková, A., Veselá, J., Neustupa, J. and Škaloud, P. (2010). "Pseudocryptic diversity versus cosmopolitanism in diatoms: a case study on *Navicula cryptocephala* Kütz.(Bacillariophyceae) and morphologically similar taxa." Protist **161**(3): 353-369.

R Core Team (2015). "R: a language and environment for statistical computing." from https://www.R-project.org

Rasband, W. S. (1997-2016). "ImageJ" from https://imagej.nih.gov/ij/

Rosin, P. L. (2003). "Measuring shape: ellipticity, rectangularity, and triangularity." Machine Vision and Applications **14**(3): 172-184.

Round, F. E., Crawford, R. M. and Mann, D. G. (1990). The diatoms: the biology and morphology of the genera. Cambridge, Cambridge University Press.

Russ, J. C. (2011). The Image Processing Handbook, Sixth Edition. Boca Raton, London, New York, CRC Press.

SAS (2013). "JMP" from http://www.jmp.com/

Shimada, C., Nakamachi, M., Tanaka, Y., Yamasaki, M. and Kuwata, A. (2009). "Effects of nutrients on diatom skeletal silicification: evidence from *Neodenticula seminae* culture experiments and morphometric analysis." Marine Micropaleontology **73**(3): 164-177.

Shukla, S. K., Crosta, X., Cortese, G. and Nayak, G. N. (2013). "Climate mediated size variability of diatom *Fragilariopsis kerguelensis* in the Southern Ocean." Quaternary Science Reviews **69**: 49-58.

Sklansky, J. (1982). "Finding the convex hull of a simple polygon." Pattern Recognition Letters **1**(2): 79-83.

Smol, J. P. and Stoermer, E. F. (2010). The diatoms: applications for the environmental and earth sciences, Cambridge University Press.

SUNY Stony Brook (2004). "TPS series" from http://life.bio.sunysb.edu/morph/

Vesela, J., Neustupa, J., Pichrtova, M. and Poulickova, A. (2009). "Morphometric study of *Navicula* morphospecies (Bacillariophyta) with respect to diatom life cycle." Fottea **9**.

Veselá, J., Pavla Urbánková, Kateřina Černá and Jiří Neustupa (2012). "Ecological variation within traditional diatom morphospecies: diversity of *Frustulia rhomboides* sensu lato (Bacillariophyceae) in European freshwater habitats." Phycologia **51**(5): 552-561.

Zielinski, U., Gersonde, R., Sieger, R. and Fütterer, D. (1998). "Quaternary surface water temperature estimations: Calibration of a diatom transfer function for the Southern Ocean." Paleoceanography **13**(4): 365-383.

Zunic, J. and Rosin, P. L. (2002). A Convexity Measurement for Polygons. Proceedings of the British Machine Vision Conference 2002. Cardiff, UK.

2.4 Publication II – Large-Scale Permanent Slide Imaging and Image Analysis for Diatom Morphometrics

Michael Kloster [1,2], Oliver Esper [1], Gerhard Kauer [2] and Bánk Beszteri [1,*]

[1] Hustedt Diatom Study Centre, Section Polar Biological Oceanography, Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Am Handelshafen 12, 27570 Bremerhaven, Germany; michael.kloster@awi.de (M.K.); Oliver.Esper@awi.de (O.E.)

[2] Hochschule Emden/Leer, Constantiaplatz 4, 26723 Emden, Germany; gerhard.kauer@hs-emden-leer.de

* Correspondence: bank.beszteri@awi.de; Tel.: +49-471-4831-1530

Academic Editors: Gabriel Cristobal, Saúl Blanco Lanza, Gloria Bueno and Jan Kybic

2.4.1 Abstract

Light microscopy analysis of diatom frustules is widely used in basic and applied research, notably taxonomy, morphometrics, water quality monitoring and paleo-environmental studies. Although there is a need for automation in these applications, various developments in image processing and analysis methodology supporting these tasks have not become widespread in diatom-based analyses. We have addressed this issue by combining our automated diatom image analysis software SHERPA with a commercial slide-scanning microscope. The resulting workflow enables mass-analyses of a broad range of morphometric features from individual frustules mounted on permanent slides. Extensive automation and internal quality control of the results helps to minimize user intervention, but care was taken to allow the user to stay in control of the most critical steps (exact segmentation of valve outlines and selection of objects of interest) using interactive functions for reviewing and revising results. In this contribution, we describe our workflow and give an overview of factors critical for success, ranging from preparation and mounting through slide scanning and autofocus finding to final morphometric data extraction. To demonstrate the usability of our methods we finally provide an example application by analysing *Fragilariopsis kerguelensis* valves originating from a sediment core, which substantially extends the size range reported in the literature.

Keywords: slide scanning; virtual slide; SHERPA; *Fragilariopsis kerguelensis*; microscopy; Metafer

Abbreviations: ADIAC, Automated Diatom Identification And Classification; AIC, Akaike information criterion; BIC, Bayesian information criterion; CSV, comma separated values; LED, light emitting diode; MIS, marine isotope stadium; SHERPA, tool for SHapE Recognition, Processing and Analysis; XML, extensible markup language

2.4.2 Introduction

Large-scale automated image acquisition and analysis methods are spreading in both biomedical and environmental research. In the aquatic realm, examples include sea floor imaging as well as in and ex situ imaging of pelagic particles, phyto- and zooplankton (Olson and Sosik 2007, Álvarez et al. 2011, Álvarez et al. 2012). Similar methods, meant to enable highly automated imaging and image analysis workflows, were previously developed for diatom permanent slides during the ADIAC (Automated Diatom Identification And Classification) project (du Buf and Bayer 2002). In spite of the highly promising results (several algorithms tested reached better-than-human identification success), the methods developed have not achieved widespread application. The main reason for this is that both

the hard- and software developed in that project are highly customized innovations that were only prototyped, but did not become available to a wider group of users.

Development of digitally controlled light microscopes, slide scanning and virtual slide systems, widely available programming libraries for computer vision, machine learning, as well as, more recently, deep convolutional neural networks, are currently making workflows similar to or even beyond those drafted by ADIAC more readily available to a wider user community, including diatomists. However, the analysis of large image sets, as they can easily be produced by slide scanning microscopes, has remained challenging for most diatomists not trained in image analysis. To fill this gap, we developed SHERPA, a user-friendly software tool conducting segmentation of such images and morphometric characterization of objects of interest (Kloster et al. 2014). This tool incorporates some of the ideas and experiences reported previously (du Buf and Bayer 2002, Claude 2008), and contributes a number of novel ideas to support quick but highly precise morphometric characterization of diatom outlines. Matching segmented objects against a library of shape templates representing objects of interest, and a refined quality scoring and ranking system, represent the major innovations of this software, whereas its workflow allows for an automated as well as a manual, but massively computer-assisted, analysis.

To illustrate the applicability of the described methods, here we investigate valve size distributions of *Fragilariopsis kerguelensis* from sediment core PS1768-8 using our workflow. This sediment core was retrieved at 52°35.58’ S, 4°28.56’ E close to the Antarctic Polar Front from a depth of 3299 m by a gravity corer (Gersonde and Hempel 1990). Its 9.03 m cover ca. 140,000 years, with *F. kerguelensis* being the dominant species (Zielinski et al. 1998). This diatom is the most prominent endemic diatom species of the Southern Ocean as well as the main opal contributor to the Southern Ocean diatom ooze belt (Zielinski and Gersonde 1997). Its abundance is an indicator of low carbon, high silica-exporting regimes in the Southern Ocean (Smetacek et al. 2004, Abelmann et al. 2006, Sachs et al. 2009), and its morphology has been hypothesized to be related to oceanic currents, nutrient (esp. iron) availability and summer sea surface temperature (Cortese and Gersonde 2007).

In this contribution, we show that a light microscopic imaging and image analysis workflow for diatom permanent slides can now be implemented combining commercially available components and the freely available SHERPA software. After introducing some options of potential implementations and highlighting some methodological issues, we draft a workflow resembling that initially recommended by ADIAC, which we are successfully applying in morphometric work. Finally, we provide the example application mentioned above to demonstrate the usability of our methods.

2.4.3 Materials and Methods

We implemented a complete workflow for semi-automated light-microscopic imaging, measurement and analysis of diatom valves (Figure 1, the numbering refers to the workflow steps as they are described below). Content to be analysed consists of cleaned frustules mounted on permanent slides (step 1); the according raw material can originate from planktonic or benthic samples, sediment traps, sediment cores or cultures. Data collection and processing consists of two cycles of slide scanning and image analysis via SHERPA: The first imaging cycle (steps 2–4) serves the purpose of locating valves of interest at low magnification. During the second cycle (steps 5–7), only these valves of interest are imaged at high magnification and measured at high precision. This method provides a good trade-off between manual effort, time consumption and data economy. Final analysis of the morphometric features measured by SHERPA is realized via R scripts customized to the particular problem (step 8).

For image acquisition, we use a Metafer slide scanning system (MetaSystems, Altlussheim, Germany) operating with a ZEISS AxioImager.Z2 using ZEISS objectives Plan-APOCHROMAT 10×/0.45, Plan-NEOFLUAR 20×/0.5 and Plan-APOCHROMAT 63×/1.4 with oil immersion. Images are obtained with a CoolCube 1m monochrome camera (MetaSystems).
In combination with the system's three-colour (red, green, blue) LED illumination, colour images can be recorded by taking three images of each position and combining them in silico. Since permanent slides prepared of oxidized diatom frustules contain no relevant colour information, we solely use blue LED illumination in all imaging procedures to achieve the highest possible optical resolution.

Regarding diatom analysis, the Metafer slide scanning system is capable of two main modes of operation. In area scan mode, a user defined rectangular or circular region on a slide is imaged in overlapping fields-of-view; these images can be combined (stitched) into large panoramic overview images (virtual slides) using the VSlide software (version 1.1.101, MetaSystems). In position list mode, a list of user-defined positions, spread arbitrarily over the slide, can be imaged. The vertical position of objects can be determined by automatic creation of a focus map, i.e., based on interpolation between reference points, or by autofocusing at each position captured, or using a combination of both. Each object/field-of-view can optionally be photographed at multiple focus depths, and the resulting images can, again optionally, be combined (stacked) to extended focus depth images.

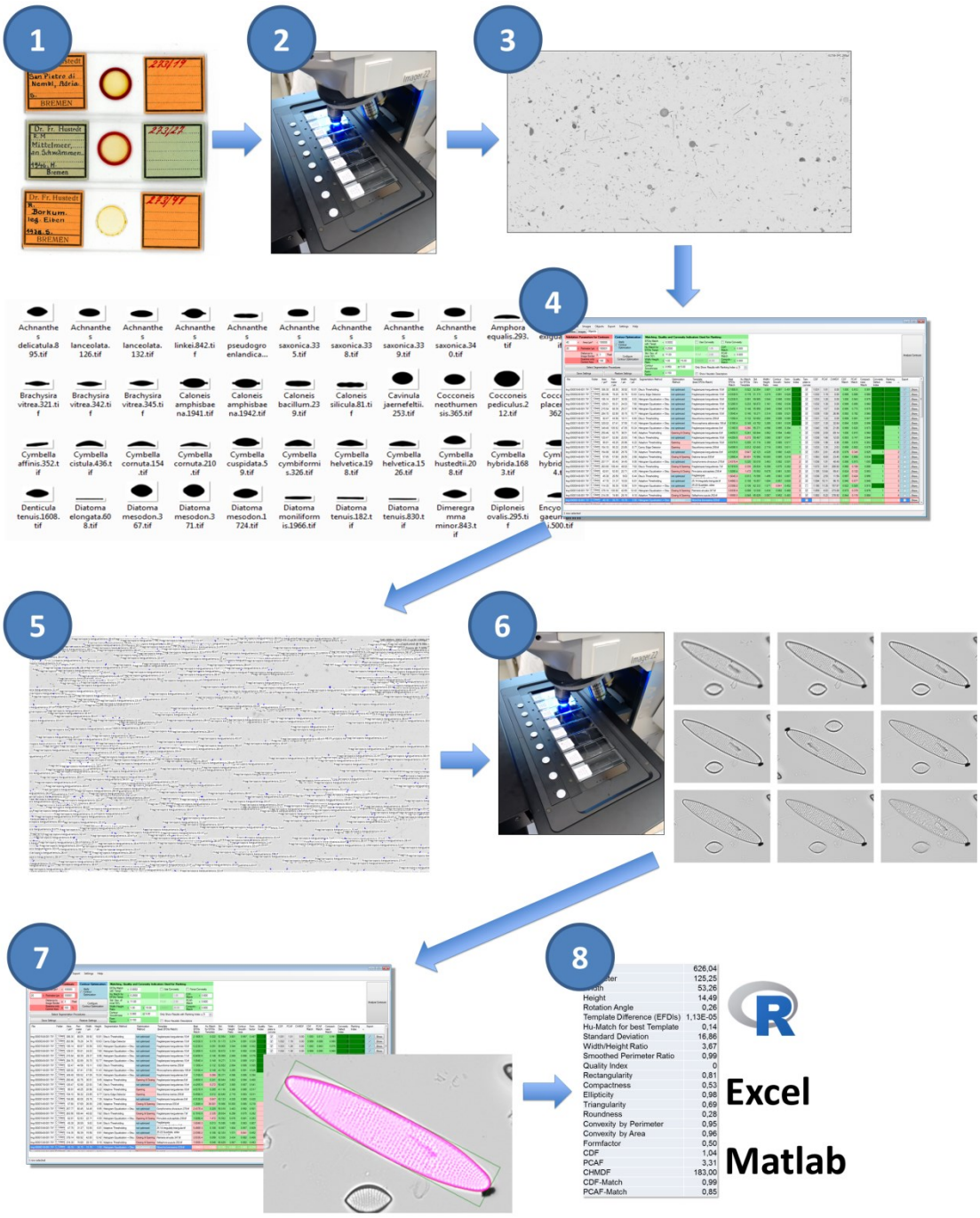


Figure 1: Schematic overview of our imaging/image analysis workflow, composed of the following steps: (1) Sample preparation, (2) low-resolution scan for locating valves of interest, (3) stitching of the virtual slide, (4) selecting objects of interest for high-resolution scan, (5) setting up the position list for high-resolution scan, (6) high-resolution scan of selected objects, (7) high-precision measurement of morphometric features, and (8) data export and post-processing.

All parameters regarding the operation of the slide scanner are recorded in so-called classifiers. These include settings for illumination (intensity, colour) and imaging (exposition, number and distances of focal planes to image) as well as parameters related to stacking focal plane images. Further settings are related to image post-processing and object detection, the latter of which is unfortunately too simplistic to be useful in diatom analyses; these are not applied in the workflows presented in this contribution. The actual classifier settings for each run are saved alongside the images produced, enabling the reproduction of an imaging run or the transparent application of a standardized set of settings to multiple slides. The only important component of an image acquisition workflow that is not explicitly documented in the classifier archive files is the autofocus settings. Settings for finding the focus are, similarly to classifiers, recorded as named parameter combinations,

and the classifier files only record this name. Metafer settings and configuration files, as used in this project, are provided as Supplementary Materials.

Image analysis and object measurement is realized by our software SHERPA (version 1.1c) (Kloster et al. 2014), which is a freely available Microsoft Windows software developed specifically for diatom morphometry. SHERPA can apply a set of different image processing methods for detecting objects in light microscopic images, and compares their contour outline to a set of shape templates to identify objects of interest. An internal rating and ranking system assesses the relevance and quality of results to reduce efforts for reviewing them manually, but nevertheless enables for manual review of objects found and for correction of faulty segmentations which lead to flawed object outlines. All parameter settings can be saved and restored to reproduce the settings. For the selected objects of interest, coordinates with respect to the images they were found in, as well as a wide variety of morphometric feature descriptors, can be exported. The coordinates (in our workflow originating from the low-resolution area scans) can be saved into an XML annotation file which can be read by the VSViewer software (version 2.1.103, Metasystems, Altlussheim, Germany) and subsequently converted into a position list for high-resolution imaging in the Metafer imaging software (version 3.10.4, Metasystems, Altlussheim, Germany). For morphometric analysis, measurements obtained by SHERPA can be exported as CSV files along with cut-outs of the original images and intermediate image processing results. We process this CSV output by R scripts that allow for a quick, easy and flexible analysis of various morphometric descriptors, but, being simple comma-delimited text tables, they can be viewed and analysed just as well with any other spreadsheet/data analysis software.

The number of parameters influencing the results of an image analysis procedure with SHERPA is substantial. Settings regard the parametrization of pre-processing applied for each of one or more segmentation methods and the optional subsequent contour optimization, the identification of valves of interest, and the filtering, scoring and ranking of results. To ensure transparency and reproducibility, SHERPA supports saving a combination of settings and re-applying it to several image data sets. The according configuration files and templates are provided as Supplementary Materials.

In the following, we describe the workflow illustrated in Figure 1 in detail (the numbering in that figure refers to the workflow steps as they are described below):

2.4.3.1 Step (1): Preparation

Throughout this contribution, microscopic imaging is performed from so-called permanent slides, prepared by embedding oxidized frustule samples into a mounting medium. Since strong object contrast is vital for successfully automating image processing, we recommend using a high refractive index mountant like Naphrax. Another critical issue is sample density. On the one hand, the density has to be sparse and uniform enough to preferably show objects clearly separated from each other, because overlapping impedes automatic valve detection by SHERPA and thus could increase manual efforts required for selecting and reworking valve shapes. On the other hand, the density has to be high enough for focus map creation to find objects at random positions on the slide. We qualitatively found that optimal sample density for the automated procedure is somewhat lower than that for manual counting.

2.4.3.2 Step (2): Low-resolution area scan for locating valves of interest

Our two-cycle microscopy and image analysis procedure starts with a low-resolution area scan (see Figure 2), creating a slide overview with a magnification appropriate to the investigated diatom. Usually we employ a 20× magnification objective, with a lamp intensity of 150, in combination with a maximal integration time of 0.011 s. Individual fields-of-view overlap by 64 µm in both *X* and *Y*

directions. Finding the focus for each image position is performed in two steps: First, at the start of each run, a focus map of the slide is prepared. For this, the optimal focus position is located over a 2 mm grid spanned over the target area by determining the value of a focus function at 25 positions in 5 µm steps and finding the *Z* position giving the maximum absolute local contrast (focusing mode 0: absolute local contrast, contrast mode 0). Subsequently, the focus map is created by linear interpolation of determined optimal focus *Z* values over (*X*, *Y*), covering the target area. After the focus map is created, the area scan itself is performed by fine focusing at each position, followed by final image capture. For fine focusing, the value of the focus function is determined (focusing mode 2: horizontal quadratic normalized gray level distances, pixel distance 2, contrast mode 0), this time in 40 planes in *Z*-steps of 2 µm. This second, fine focusing step can sometimes be omitted, but it often leads to substantial improvement in the sharpness of resulting images (depending on density and contrast of diatom valves). Finally, images at 10 focus positions in *Z*-distances of 2 µm are captured and stacked together to extended focus depth images which are saved for further processing. These stacked images represent the input to SHERPA for locating diatom valves, as well as to the stitching procedure to create a virtual slide.

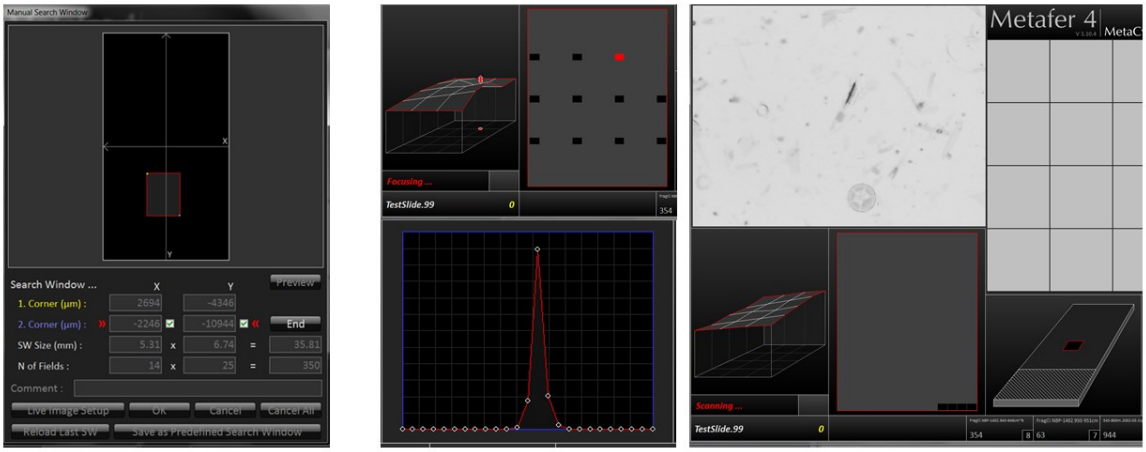


Figure 2: Overview of the low-resolution slide area scan procedure (step 2 of the workflow). Left: manual selection of rectangular target area. Centre: focus map creation over a 2 mm × 2 mm grid. Right: focus stack capture in overlapping fields-of-view over target area.

2.4.3.3 Step (3): Stitching of virtual slides

The overlapping low-resolution images are stitched together to a virtual slide using the VSlide software (Metasystems, Altlussheim, Germany). These virtual slides enable for visualizing the valves of interest as identified and selected via SHERPA in the context of the whole slide (see steps 4 & 5 below) and also for manually adding further positions for the later high-resolution scan (see step 6 below). Although in principle, it is not strictly required to stitch images to a virtual slide at this stage, this is necessary with the imaging system we are using for two practical reasons: First, the object coordinates relative to individual field-of-view images (as recorded by SHERPA) are transformed to coordinates relative to the whole slide scanned at this step. Second, import of positions into a virtual slide is necessary for feeding a list of positions back into the imaging software in the specific Metafer imaging system we are using.

2.4.3.4 Step (4): Selecting objects of interest for high-resolution scan via SHERPA

For locating and selecting valves of interest, the low-resolution images (step 2) are processed by SHERPA. Object outlines are compared to a user defined set of shape templates, in case of the example analysis explicated below to the *F. kerguelensis* templates provided with SHERPA. The template set selected defines the taxonomic scope (more precisely, the scope in terms of outline

shapes) of an analysis. Hereby analyses focusing on a narrow range of outline shapes are possible, like in our example application, as are analyses targeting a broad diversity of diatom shapes. For the latter purpose, SHERPA provides a set of ca. 500 outline templates covering a broad range of diatom shapes, which can be further extended by the user to match the specific set of taxa of interest from a particular habitat. Shape optimization and manual corrections are usually not applied since this analysis step is used only for localizing valves, not for measuring them. Convexity analysis often fails on low-resolution images, even for species with strictly convex outlines, especially if the image contrast or sharpness is not optimal. Even though we applied it in our example investigation (see below), we recommend not to do so at this workflow step, at the optional expense of increased efforts for sorting out potential false positives manually. To ensure that settings for validation and ranking are comparable to those of the high-resolution scan (like minimum required object area/perimeter), it is important to set the correct magnification ratio (parameter "Micrometer Factor"; also saved in the settings file). Valves of interest are either picked automatically by selecting all results up to (including) ranking index 2, or by manually reviewing and selecting results. Manual selection is generally conducted at least up to ranking index (including) 2, and can be extended up to higher indices if the amount of highest quality results is not sufficient. The position list for the subsequent high-resolution scan (step 6) is exported as a Metafer VSAI file and saved into the same directory and with the same name as the according virtual slide files.

2.4.3.5 Step (5): Setting up the position list for high-resolution scan

We use the VSViewer software (Metasystems, Altlussheim, Germany) to import the position list provided by SHERPA (step 4) into the virtual slide (step 3), using the "import annotations" function; valve positions are displayed by their annotated outline (see blue markings in Figure 3). The annotation function of the viewer enables to manually add positions of interest if some objects deemed relevant were missed by SHERPA (see red box in Figure 3). This allows for combining (semi-)automated and manual selection of valves of interest.

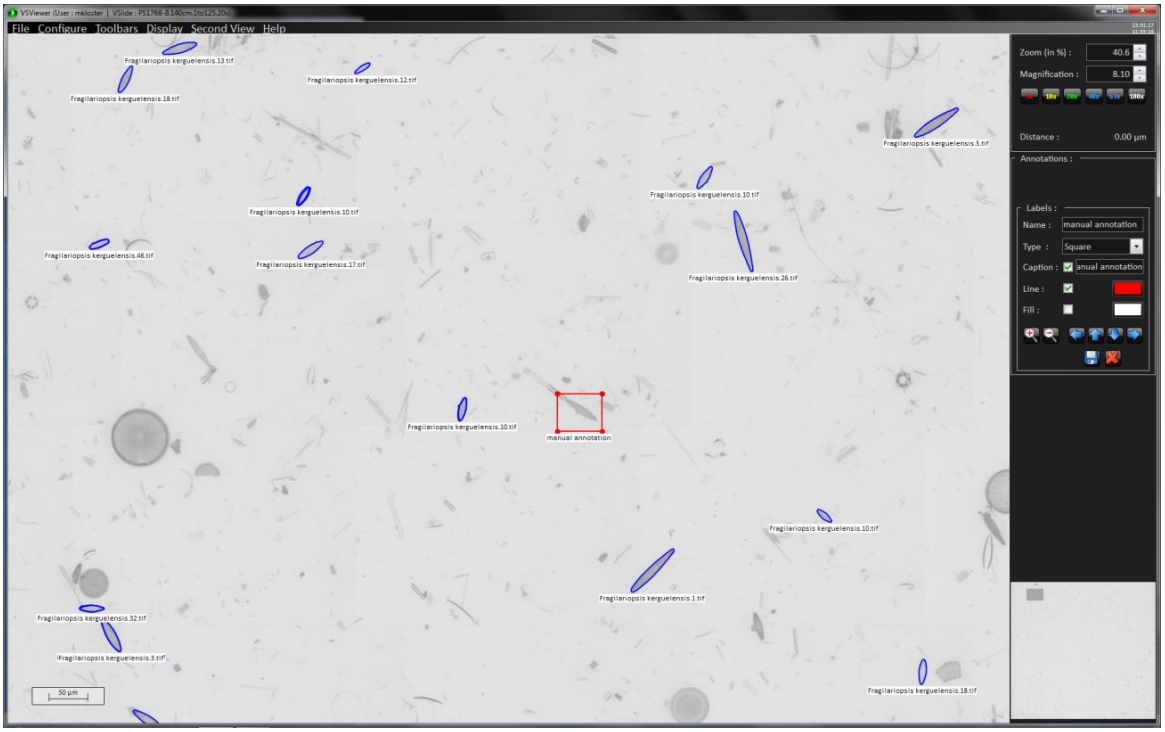


Figure 3: Screenshot of VSViewer showing a subarea of the virtual slide of a 20× scan. *F. kerguelensis* valve outlines imported from SHERPA are highlighted in blue; additional manual annotations are in red. The positions annotated in this low magnification image are used for the selective high-resolution scans in step 6.

2.4.3.6 Step (6): High-resolution position scan of selected objects of interest for precise morphometric measurements

The annotated virtual slide from step 5 contains the (*X*, *Y*) position data required for selectively imaging the valves of interest. This high-resolution scan uses the 63× objective with oil-immersion and captures 20 focus levels in 0.2 µm *Z*-distances after two focusing steps, a first, coarse autofocusing followed by a fine focusing. This is necessary because object positions imported from a virtual slide do not preserve information on the vertical (*Z*) location of individual objects, so these have to be determined from scratch again. In the first autofocus run, the value of the focus function is determined at 40 *Z*-positions in 2 µm distances for a coarse determination of the optimal focal plane (focusing mode 2: horizontal quadratic normalized grey level distances, pixel distance 2, contrast mode 0). The focus is narrowed down by a second scan measuring at 60 focus levels in 0.5 µm distances (focusing mode 0: absolute local contrast, contrast mode 0). Subsequently, 20 focal planes around the optimal one are captured in 0.2 µm distances. An extended focus image is produced by stacking the individual focal planes, using the method "CombPlanesFF" with focusing mode 0 and a submatrix size of 128. The individual focal plane images, as well as the extended focus image, can be saved for further processing. In general, for analysis with SHERPA, we only use the extended focus image, since this clearly depicts the valve's outline as well as its interior ornamentation, even if not all structures can be visualized completely within a single focal plane because of their three-dimensional extent, or because the valve surface is slightly tilted with respect to the focal plane (see Figure 4). Keeping the individual focal plane images leads to a significantly increased demand for storage space, but allows for better manual evaluation of ambiguous species classifications, since in uncertain cases the three-dimensional structure of features, as well as the influence of stacking artefacts, can be judged. The individual focal plane images also allow one to see if a valve lies obliquely, leading to a perspective distortion of its outline shape and size. So far no software is available to make use of this information, but in theory it seems possible to develop automated procedures to correct morphometric measurements for the bias resulting from oblique valve orientation (Kloster 2013). The maximum possible bias can be estimated according to the Pythagorean theorem (see Figure 5): $s = \sqrt{m^2 + t^2}$ with s being the actual valve size, m the measured size, and t the tilt. Our setup limits the tilt to a maximum of 4 µm, which results in a maximum error of about 8 % for measurements of 10 µm. For larger measurements the error decreases, e.g., for 36 µm, which is the average valve length in our investigation, the maximum error lessens to ca. 0.6 %.

2.4.3.7 Step (7): High-precision measurement of morphometric features from high-resolution images via SHERPA

The high-resolution extended focus (i.e., stacked) images are then analysed with SHERPA to measure the valves of interest. Since the low-resolution position list only sets the central position of the field-of-view, images captured in the second, high-resolution run can also depict objects that are not relevant for analysis, but might be detected by SHERPA. To avoid false positives, manual selection of the final set of valves of interest can be conducted a second time, a process which is massively supported by SHERPA. Usually the same settings and shape templates are applied as for the low-resolution scan, the only difference being the magnification ratio set according to the high-resolution objective, and enabling convexity analysis if this is suitable for the species under investigation (i.e., its valve outline does not contain concave/indented parts). For the example analysis explicated below, contour optimization was disabled and convexity analysis enabled. Like for the low-resolution analysis, manual review and selection of results should be conducted at least until ranking index (including) 2, and may proceed to higher indices. Manual rework for correcting inaccurately

segmented valve outlines can be applied if deemed necessary. Selected results can be saved to CVS files along with cut-outs of the original images depicting the valve and intermediate stages of image processing to enable for a later (re-)analysis with other tools than SHERPA.

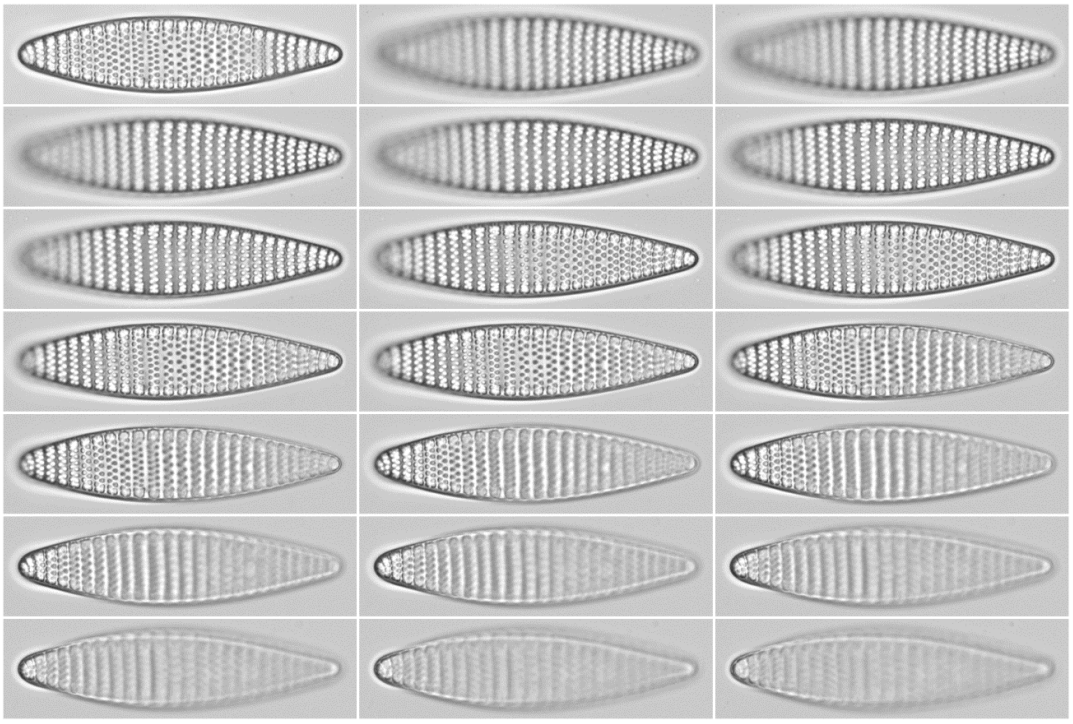


Figure 4: The extended focus image (top left) clearly depicts valve outline and ornamentation, in contrast to the individual focal planes (anything but top left) it is stacked from. On the other hand, the focal plane images preserve information about the obliquity of the valve that is lost in the extended focus image.

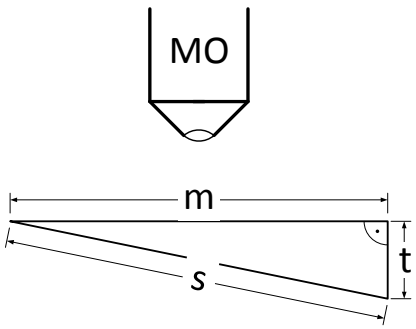


Figure 5: Measured (m) compared to actual object size (s) for a given tilt (t); MO = microscope objective.

2.4.3.8 Step (8): Data export and post-processing

Our workflow can produce duplicate results for objects of interest if some of them are captured on more than one field-of-view images. This happens if a valve of interest lies so near to another one that it is also captured in the image centred on the latter. Removing these doublets is the first important step of downstream analyses. Duplicate objects are identified based on a threshold applied to the summed up distances of basic geometric feature measures like apical and transapical axes length, area, circumference and orientation angle. We implemented this approach to be independent of object coordinates supplied by the Metafer slide scanner, enabling to process data originating also from manual microscopy or other slide scanning systems. The according R function (based on R version 3.2.2 (R Core Team 2015)) is supplied in the Supplementary Materials.

2.4.4 Results

Applying our automated approach, we analysed about 12,000 valves from 72 slides of samples taken from 21 different depths within sediment core PS1768-8 covering marine isotope stadiums (MIS) 1, 2, 5 and 6 (see Table A1). The according CSV file is supplied in the Supplementary Materials. Because this core originates from the central area of the opal belt it consists mainly of diatoms, which are generally well to moderately preserved and do not show significant effects of dissolution on the preserved diatom valves (Zielinski et al. 1998). Our observations substantially expand the size ranges known for this taxon compared to previous reports (see Table 1 and Figure 6).

Table 1: Summary on *F. kerguelensis* valve sizes from our investigation. Extreme values are highlighted in bold type.

		Apical Axis Length (µm)				Transapical Axis Length (µm)			
MIS	***n***	**min**	**max**	**mean**	**median**	**min**	**max**	**mean**	**median**
1	2853	11.3	101.1	36.8	34.1	**3.7**	13.5	9.0	9.1
2	5691	**11.1**	95.5	35.2	33.1	4.6	**14.2**	9.2	9.2
5	2610	12.8	**118.9**	37.0	33.6	4.1	14.1	9.0	9.1
6	705	14.8	94.4	36.1	33.5	5.6	13.8	10.0	10.0
all	11859	11.1	118.9	36.0	33.4	3.7	14.2	9.2	9.2

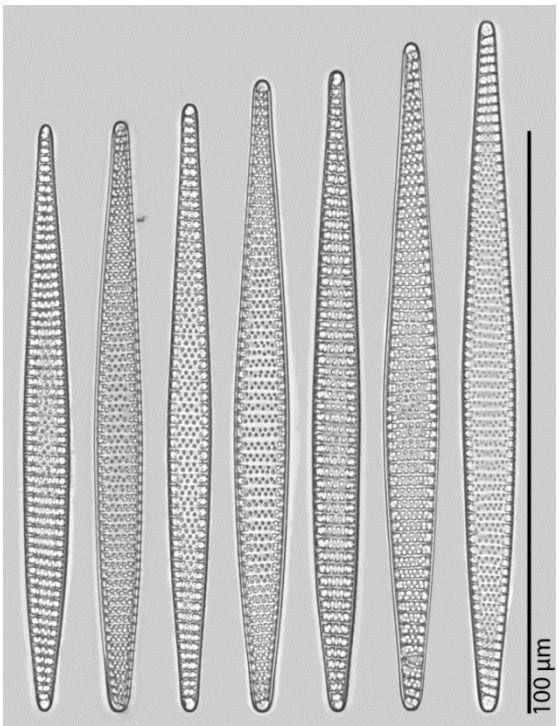


Figure 6: *F. kerguelensis* valves of extreme apical axis lengths (101.0 to 119.9 µm). Left: MIS 1, all others come from MIS 5.

The valves with the longest apical axis were found in the MIS 5 samples. To test whether the species shows a different distribution of apical axis lengths between this and the other MISs, Kolmogorov–Smirnov tests were used (see Tables A2 and A3). MIS 1 represents the most recent material we analysed and thus the best to be compared to the investigations of (Hasle 1965, Fenner et al. 1976, Cefarelli et al. 2010). Between this and MIS 5 the two-sided test did not support a significant difference. This indicates that the occurrence of very large valves is not caused by a general shift in the valve length distribution of the species during MIS 5. The overall distribution of apical axis length is right-skewed (see Figure 7 top). To the contrary, distribution of the transapical axis length is symmetric (see Figure 7 bottom) but platykurtic (test for non-normality; see Appendix C). For individual analysis of each MIS see Figures A1 and A2 in the Supplementary Materials.

Ca. 5.4 % of all valves (505 in total) are sized outside the dimensions reported in the literature of (Hasle 1965, Fenner et al. 1976, Cefarelli et al. 2010), depicted by the dashed rectangle in Figure 8.

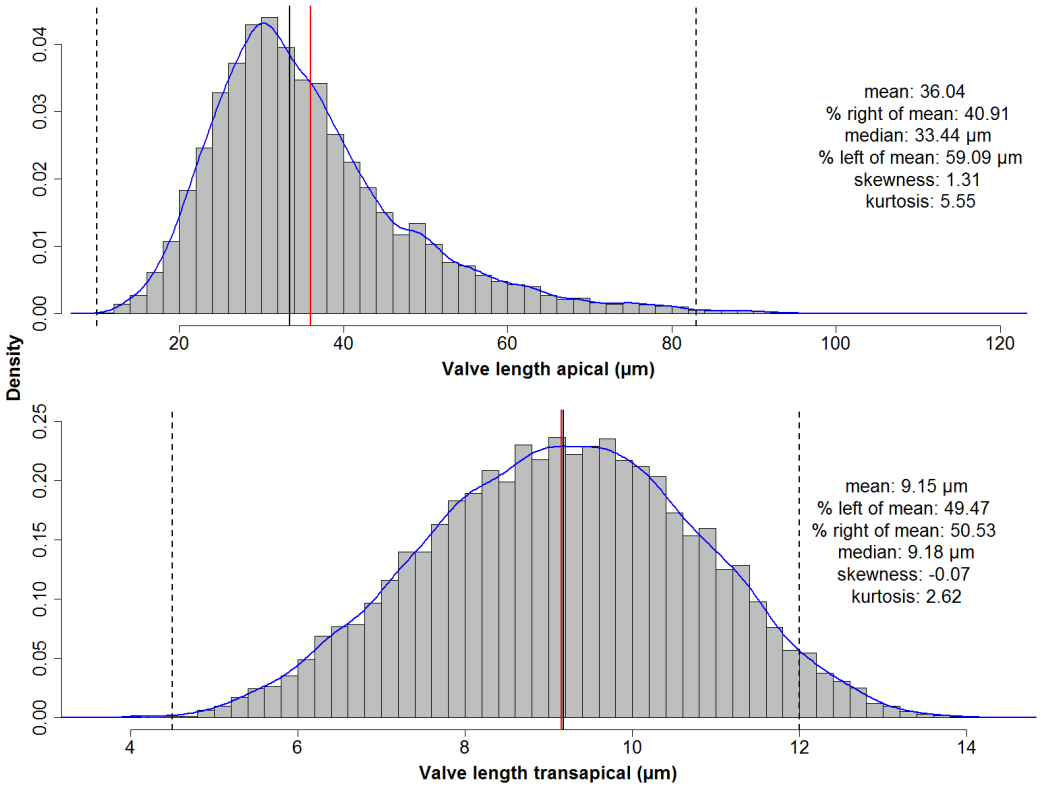


Figure 7: Distribution of basic valve dimensions along the apical and transapical axes for all MISs. Red lines depict the mean, black lines the median values, and blue lines the estimated density curves. Dashed lines depict the valve size ranges reported in the literature. Whilst for the apical axis the distribution is right-skewed, the distribution of lengths of the transapical axis is symmetrical but not normal (platykurtic, i.e., has thinner tails than a normal distribution).

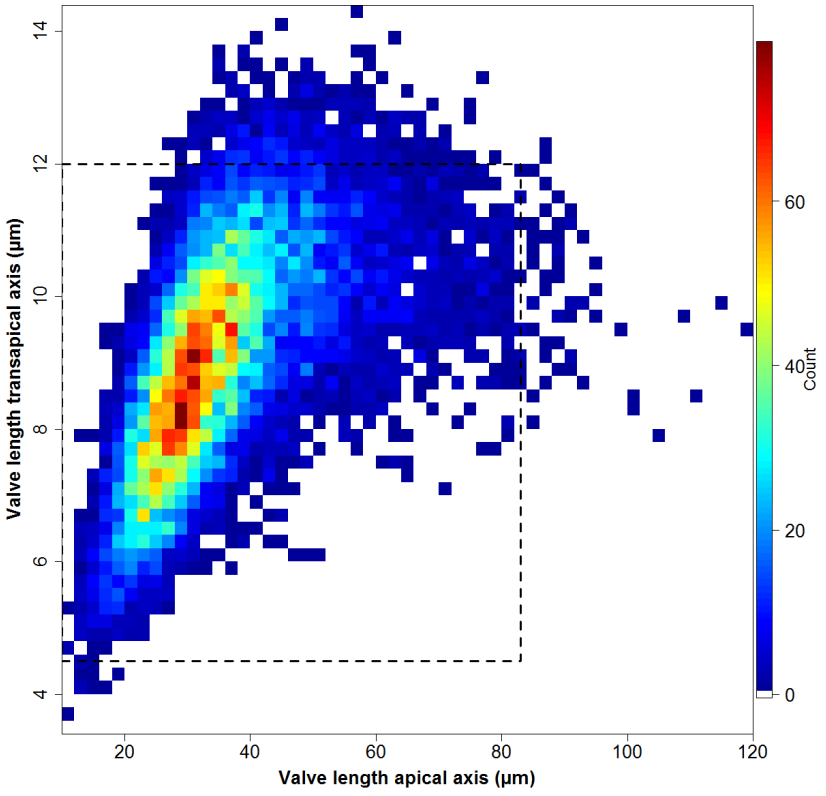


Figure 8: Counts of valves of different apical versus transapical lengths for all MISs. The dashed rectangle depicts the valve size ranges reported in the literature.

We modelled the relationship between transapical and apical axis length using linear models with an increasing order of polynomial terms (see Figure A3). A third-order polynomial gave a better fit (with a R^2 of 0.496) than a first- or second-order regression, and lowest values of the Akaike resp. Bayesian information criteria. Adding a term of order 4 did not significantly improve the model fit. This means that, in spite of what might be suggested by the symmetric distribution of transapical valve measurements (see Figure 7, bottom), the transapical tends to increase with apical length only up to about 50 µm valve length. Above this value, the transapical length stays roughly constant or even decreases slightly with increasing apical length.

2.4.5 Discussion

Imaging and image processing methods have been developed and spread rapidly for marine research in the recent years (Schulz et al. 2010, Briseño-Avena et al. 2015, Schulz et al. 2015, Mullen et al. 2016). In contrast to many of these techniques, our approach is conducted ex situ on oxidized material, as has been the common practice in diatom research for nearly 200 years. This enables us to combine traditional preparation methodology with recent advances in imaging and automation, and the application of the latter to existing sets of samples. During the ADIAC project a prototyped automated slide scanning procedure was developed, the main limitation of which consisted of the mechanical parts. Since then, commercial slide scanning systems have become available off the shelf, and advances in camera technology improved the image quality significantly.

Optical resolution: In the last decade, a variety of slide scanning microscopes have become available, which are capable of producing virtual slides by stitching together overlapping field-of-view images. In many biomedical applications, however, a much lower optical resolution is sufficient, and 40× or lower objective magnification is the highest available in many slide scanning systems on the market. For applying to diatoms, however, it is important to select a slide scanning system that can also robustly operate with high-resolution oil immersion objectives. A different high-throughput imaging and measurement method, using unmounted, but still acid-cleaned diatom frustule material, has been proposed previously by (Spaulding et al. 2012). In special settings, like for monospecific cultures, their FlowCam-based method can probably provide a superior throughput compared to our workflow. However, the range of morphometric descriptors seems limited to cell length in the case of the FlowCam. Also, the accuracy of measurements is expected to be higher in the case of images obtained using high-resolution oil immersion objectives from permanent slides, and probably even higher, but definitely more precisely reproducible, than data obtained by time-consuming manual measurement (whether using an ocular micrometer or software-assisted linear measurement between manually selected points in a digital image). The variation of valve measurements induced by using multiple segmentation methods simultaneously in SHERPA is usually around 1 % (Kloster et al. 2014).

Focus: Autofocusing at a specified slide position (*X*, *Y*) is realized by capturing images at different focus depths (*Z*), determining the value of a focus function at each *Z* position and finding the location of its maximum. The focus function used, the *Z*-range captured, as well as the steps size, can be controlled by the user. Finding the best focus position is a low-level component of the imaging workflow that is applied at a number of different steps throughout, and autofocus settings for different steps (focus map creation versus final imaging) and for different objectives need to be optimized separately.

An area scan consists of a number of steps (see Figure 2 for a graphical overview): After selection of the target area, the *Z* position providing the maximum of a focus function is determined over a grid covering the target area. Using linear or spline interpolation, a focus map is created estimating the height (*Z* coordinate) of the optimal focus position across the whole target area. The way the focus grid is created can be controlled by the user in fine detail. Respective parameters regard the resolution of the grid for focus map creation, the autofocus settings applied in this step, as well as the interpolation method. The target area is divided into overlapping fields-of-view that are captured one by one. For each (*X*, *Y*) position, the *Z* position can either be determined from the interpolated focus map, or from a renewed autofocusing around that *Z* position (fine focus).

There are two main reasons why images captured in a single focus position determined from a grid interpolation are often not optimally sharp: first, minute local variation in frustule positions in the Z direction on a preparation often exceeds the depth of focus of even relatively low-resolution (10×, 20×) objectives. Second, in the case of several taxa, relevant morphological information is located in different focal planes across a single valve (for instance, the outline is often best discernible in a different *Z* position than valve face structures like striation). One of the possible measures to improve this situation is conducting a renewed autofocusing at each image position, which enhances the quality of the resulting images (more homogeneous sharpness), but comes at the cost of increased processing time. Another possible measure to address local variation in object *Z* positions is to capture and combine an image stack at each (*X*, *Y*) position. If this is performed over a range that covers the heterogeneity in *Z* positions across the slide, the quality of the stacked images can be comparable to or even better than those captured after renewed autofocusing even if a second autofocusing step is omitted. The trade-off for this measure is a substantially increased processing time. In our experience, however, it is difficult to estimate in advance the degree of *Z* position heterogeneity before a scan. We found, therefore, that the most generally applicable procedure is to combine a renewed autofocusing with image stacking (as opposed to capturing a single image only at each position). Renewed autofocusing ensures that the images are taken in a *Z*-range that fits the object height distribution in the specific field-of-view. On the other hand, capturing and combining images at multiple focus depths ensures that differences in object height within a single field-of-view do not lead to only those objects being pictured sharply which dominate the autofocus function in that field of view (which can be the largest or the most contrasting objects and are not necessarily the object of interest). Focus stacking also improves single-specimen images in the sense that features of the valve lying in different focus depths (for instance, valve outline as well as valve face) are visualised sharply, although in the case of diatoms with complex three-dimensional valve structures, this might create disturbing artefacts.

We note that it is not necessary to combine (stack) images taken at different Z but identical (*X*, *Y*) positions into an extended depth-of-focus image. It is also possible to keep the individual *Z*-position images and analyse them separately or in a combined manner in downstream processing. However, we currently do not see that this could lead to an improvement in the quality of the low-resolution area scan analysis, and so decided to only keep the stacked images for further processing in this case, which helps saving disk storage space. In the case of high-resolution images, we see a strong potential in analysing image stacks as opposed to stacked extended depth-of-focus images, but development of methods specifically suited to the former is substantially more challenging than analyses of stacked images and currently is not part of our workflow.

Investigated Features: During the ADIAC project, a large variety of morphometric features (e.g., Gabor features or Legendre polynomials) were measured and classified for taxonomic identification

of diatoms, utilizing a compilation of mostly prototyped methods. In contrast, SHERPA investigates less and partly different features (e.g., elliptic Fourier descriptors (Claude 2008)) and focuses on finding, imaging and measuring diatoms based on their shape as part of a routine workflow for morphometry purposes.

Data economy: An area scan, as described above, can in principle be performed with any user-selected objective, including oil immersion objectives. Accordingly, it would be thinkable to design an imaging/image analysis workflow where diatom slides are scanned over their whole area using a high-resolution oil immersion objective directly. This is indeed possible (and we have tested the procedure) but causes two difficulties: One is an enormous increase in storage capacity required to save the images. To capture the same physical surface area of a slide, roughly 10 times as many images captured by a high-resolution 63× objective are necessary when compared to a 20× objective. The difference is almost 40-fold when comparing the 63× with a 10× objective. In addition, the probability of individual diatom frustules not being captured fully on any single field-of-view image increases, hampering downstream analyses. This can in principle be helped by increasing the overlap between neighbouring fields-of-view, but to ensure that most frustules are captured fully in at least one field-of-view, the overlap between neighbouring images should be in the range of the size of the largest frustules in a sample, which is substantial. Taking 64 µm, a not uncommon diatom size, as an example (this matches the overlap between neighbouring fields-of-view in our 20× pre-scan): this corresponds to around 627 pixels in images captured at 63× objective magnification. A full field-of-view image has 1360 × 1024 pixels, i.e., the necessary overlap would correspond to almost half the image width and over 60 % of image height. This leads to a further substantial increase of storage capacity, due to the fact that more images are required to capture the same slide area. Besides an increase in storage space requirements by more than an order of magnitude, processing time for downstream steps is also increased in a similar manner.

Due to these complications, we finally adopted a two-step imaging procedure which was previously suggested and tested by the ADIAC project (du Buf and Bayer 2002). This consists of a low magnification area scan, followed by an image analysis step locating diatom frustules in the images, and by a high-resolution multiple focal planes imaging limited only to slide positions where objects of interest (frustules) were located (see Figure 1 steps 2–6). This procedure keeps storage space requirements relatively low while still allowing high precision of morphometric characterization. In order to be able to implement this workflow, software functionality for exporting outline point coordinates of objects found from SHERPA and importing these into virtual slides in the VSViewer software was developed in communication with Metasystems GmbH.

User efforts: By all efforts at automation and supporting high throughput analyses, the quality of final results from this workflow can be substantially improved by human intervention at multiple steps of the workflow. Although in some situations it can be satisfactory to receive a subsample of diatom valves of interest in a defined slide area, there are also use cases when it is important to not miss (m)any objects of interest (quantitative counts). Also, diatom valves representing taxa of interest cannot be differentiated from other species (in the case of taxon-specific analyses), and from other types of objects commonly occurring on diatom slides (like sediment particles or unidentifiable fragments of broken diatoms) at a 100 % accuracy and specificity. Furthermore, segmentation can be misled by touching or overlapping objects, minor fractures in valves or other anomalies. The workflow using slide scanning and SHERPA allows a fine scale control to address these issues. After the first slide scan generating an overview at low objective magnification, the target objects located by SHERPA can be manually checked and remaining non-diatom objects sorted out before further

processing. In SHERPA, one can either use stricter quality settings to speed up or completely skip manual processing and to reduce the number of non-diatom objects detected at the price of perhaps missing more valves of interest; or use more permissive quality criteria potentially allowing a higher yield at the price of more false positives (requiring more processing time to remove these manually).

If it is critical not to miss any valves lying in the slide area of interest, there is a possibility for manual interaction with the virtual slide once object positions and outlines detected by SHERPA have been imported into it. At this stage, the virtual slide can be scanned visually and diatom objects missed by SHERPA can be marked manually in addition so that they are also imaged at high resolution in the next step.

Finally, to improve segmentation results in the case of individual problematic valves (e.g., ones overlapping with other objects), a manual correction of outlines is possible at the stage of SHERPA analysis of high-resolution images. The morphometric features, as well as quality scores and ranking of the object, are updated immediately after such manual corrections, and the fact that a manual correction was applied to the object is recorded in the results table exported by SHERPA for further processing. Again, a decision as to whether a manual processing at this step is sensible and how much time to invest in it can be decided case by case based on the purpose of the investigation.

Practical example: To give a simple demonstration of the workflow introduced, we presented a basic morphometric characterization of *Fragilariopsis kerguelensis*, one of the most common diatom species in Southern Ocean sediments from the late Pleistocene on to present days (Cortese and Gersonde 2008). The fact that we can substantially extend on the hitherto reported size range of the species (see Figure 9) seems to simply reflect a substantial increase in sample size. Although qualitatively, we see many of the longest valves in samples from MIS5, a Kolmogorov–Smirnov test did not support a significant difference in apical length distributions of this sample from more recent sediment layers. Whereas taxonomic descriptions circulating in the literature might be based on measuring dozens, perhaps up to a few hundred valves (Hasle 1965, Cefarelli et al. 2010), using the partially automated workflow presented, we could measure about 12,000 valves. Ca. 0.54 % (in total 64) of the *F. kerguelensis* valves analysed had an apical axis length longer than the maximum of 83 µm reported before by (Hasle 1965, Fenner et al. 1976, Cefarelli et al. 2010). The longest valve with 118.9 µm apical axis length (see Table 1 and Figure 6) is ca. 45 % longer than the maximum reported before. The shortest valve with 11.1 µm apical axis length is smaller than the minimum sizes reported by (Cefarelli et al. 2010) and (Fenner et al. 1976), but larger than the 10 µm reported by (Hasle 1965). The transapical axis length ranges from 3.7 to 14.2 µm, compared to 4.5 to 12 µm reported by (Fenner et al. 1976), where 3.66 % (in total 434) valves were larger, and 0.08 % (in total nine) valves were smaller with regard to this axis than reported before.

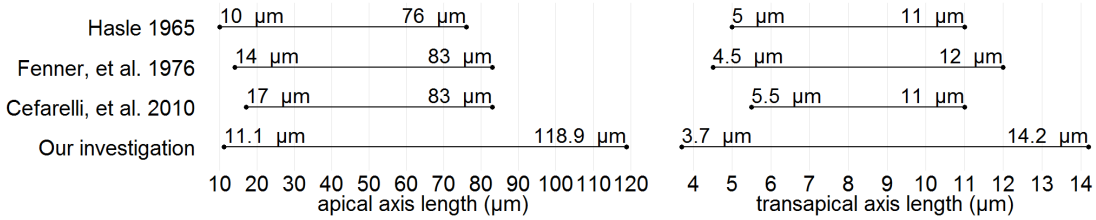


Figure 9: Valve size ranges reported in the literature compared to our findings.

Whilst the distribution of apical valve axis lengths is right-skewed (see Figure 7 top), as can be expected based on size reduction accompanying vegetative divisions, the symmetric distribution of transapical valve axis lengths (see Figure 7 bottom) shows that the size reduction somehow is relaxed

along this axis for *F. kerguelensis.* Whereas it could be suspected, based on this symmetric distribution, that transapical valve width might remain constant with size decrease, this is not the case. The observed pattern is rather a roughly linear relationship of transapical width with apical length up to length of around 50 µm, and a constant or slightly decreasing transapical width with increasing length above that. This relation between apical and transapical valve length can be described by a third-order polynomial that explains ca. 50 % of the observed variance (see Figure A3). More details of the morphometrics of the species will be presented in a follow-up publication.

2.4.6 Conclusions

We describe a partially automated diatom slide imaging and image analysis workflow consisting of commercially and freely available components, making such a procedure widely applicable for routine research. Extending the size range reported for *F. kerguelensis* (quite substantially, by nearly 45 % in the case of apical length) nicely illustrates the advantages of a (semi-)automated approach for microscopic morphometry of diatom valves, primarily that it makes documenting and measuring large numbers of specimens feasible with moderate effort. The method also enables one to flexibly balance throughput/manual effort required versus quality (specificity/sensitivity of detecting objects of interest). This workflow is still far from a fully automated optical diatom analysis system also capable of taxonomic identification of imaged valves, as envisioned earlier (du Buf and Bayer 2002); nevertheless, it implements substantial components and is usable for performing diatom morphometric research already in its current form.

Supplementary Materials: The following are available online at http://awi.de/sherpa: Settings for Metafer4 software, the SHERPA software, SHERPA settings und shape templates, R functions for reading SHERPA output and removing duplicates from each dataset, morphometric data from PS1768-8. The latter one also is available at https://doi.org/10.1594/PANGAEA.873993.

Acknowledgments: This work was supported by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the priority programme 1158 “Antarctic Research with comparative investigations in Arctic ice areas” by grant nr. BE4316/4-1, KA1655/3-1.

Author Contributions: Michael Kloster, Gerhard Kauer and Bánk Beszteri developed the microscopy workflow. Michael Kloster developed SHERPA, and performed the measurements and analysed the data under the supervision of Bánk Beszteri and Gerhard Kauer. Oliver Esper contributed the sediment material. Michael Kloster and Bánk Beszteri wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

2.4.7 Supplements

2.4.7.1 Appendix A

Samples from sediment core PS1768-8.

Table A1: Information on the analysed samples from sediment core PS1768-8.

Depth (cm)	Age (years)	MIS	*n*
60	10,195	1	85
80	11,182	1	698
100	12,071	1	857
110	12,516	1	553
120	12,960	1	322
130	13,405	1	83
140	13,849	1	255
150	14,912	2	1659
160	16,130	2	1458
170	17,347	2	765
180	18,565	2	441
190	19,782	2	650
200	21,000	2	718
780	119,500	5	471
800	123,700	5	437
810	125,800	5	550
820	127,900	5	420
830	130,000	5	732
840	132,100	6	166
850	134,200	6	265
870	138,400	6	274

2.4.7.2 Appendix B

Comparing apical valve axis length distributions from each MIS for significant differences.

Table A2: The *p*-values of the two-sided Kolmogorov–Smirnov test show a significant difference between valve apical axis length distributions of MIS 1 & 2, 1 & 6 and 2 & 5 (see bold highlights).

	MIS 1	MIS 2	MIS 5	MIS 6
MIS 1	1.0000	NA	NA	NA
MIS 2	**0.0000**	1.0000	NA	NA
MIS 5	0.2808	**0.0000**	1.0000	NA
MIS 6	**0.0429**	0.1680	0.4442	1.0000

Table A3: The *p*-values of the one-sided Kolmogorov–Smirnov test (Alternative hypothesis "greater") show that the central tendency is greater in MIS 1 compared to MIS 2 and MIS 6, as well as for MIS 5 compared to MIS 2 (see bold highlights).

	MIS 1	MIS 2	MIS 5	MIS 6
MIS 1	1.0000	**0.0000**	0.1408	**0.0214**
MIS 2	0.7994	1.0000	0.9193	0.4559
MIS 5	0.3137	**0.0000**	1.0000	0.2246
MIS 6	0.8622	0.0841	0.7043	1.0000

2.4.7.3 Appendix C

The Shapiro–Wilk and the Lilliefors (Kolmogorov–Smirnov) normality tests do not support the hypothesis that the symmetrical distribution of transapical axis length is normal (see R output below).

R console output:

```
> shapiro.test(fkergdata$Height, 5000)) # -> very low p-value, not normally distributed

    Shapiro-Wilk normality test

data: fkergdata$Height, 5000)

W = 0.99782, p-value = 1.589e-06


> lillie.test(fkergdata$Height) # very low p-value, -> not normally distributed

    Lilliefors (Kolmogorov-Smirnov) normality test

data: fkergdata$Height

D = 0.015253, p-value = 1.72e-06
```

2.4.7.4 Appendix D

For each MIS we found valves exceeding the size ranges reported before (see Figures A1 and A2).

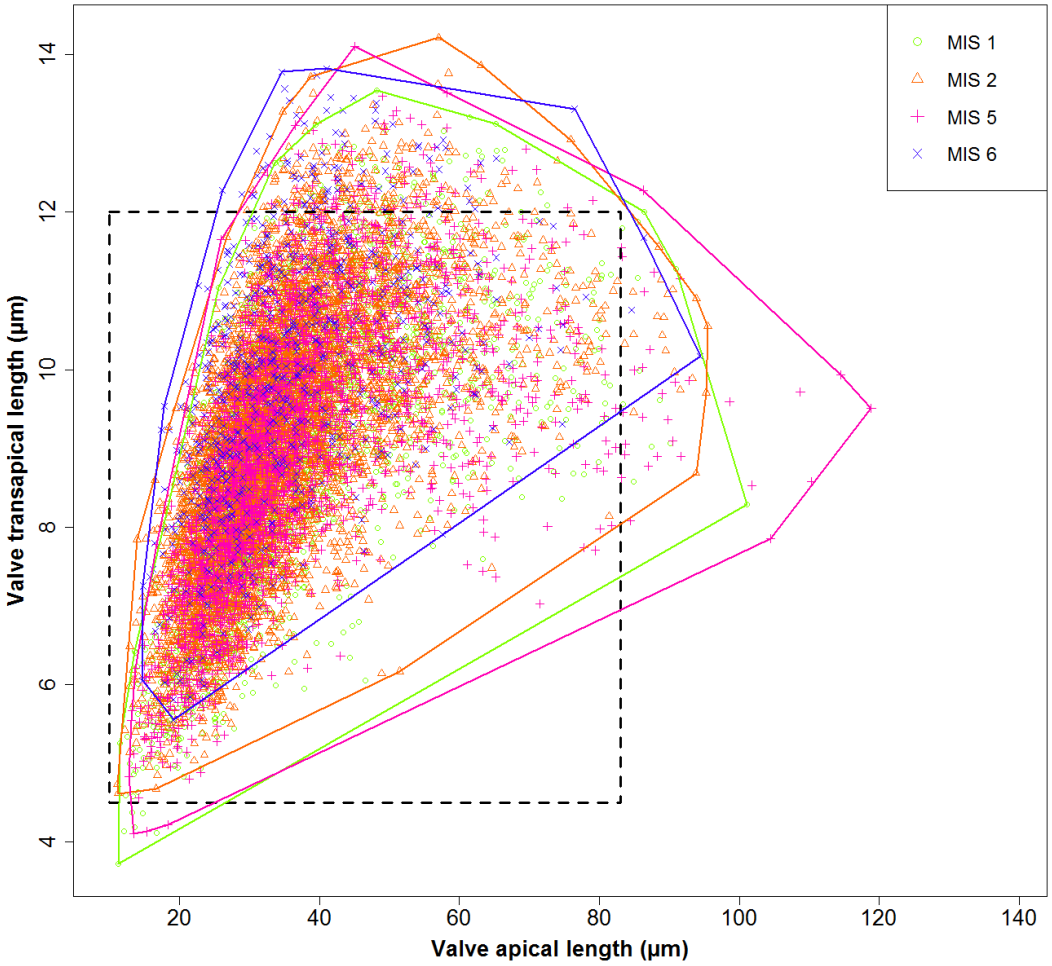


Figure A1: Valve dimensions for different sampling depths; coloured lines depict the convex hull around the corresponding data points. Especially during MIS 5, extremely large valves were found. The dashed rectangle depicts the range of valve sizes previously reported in the literature.

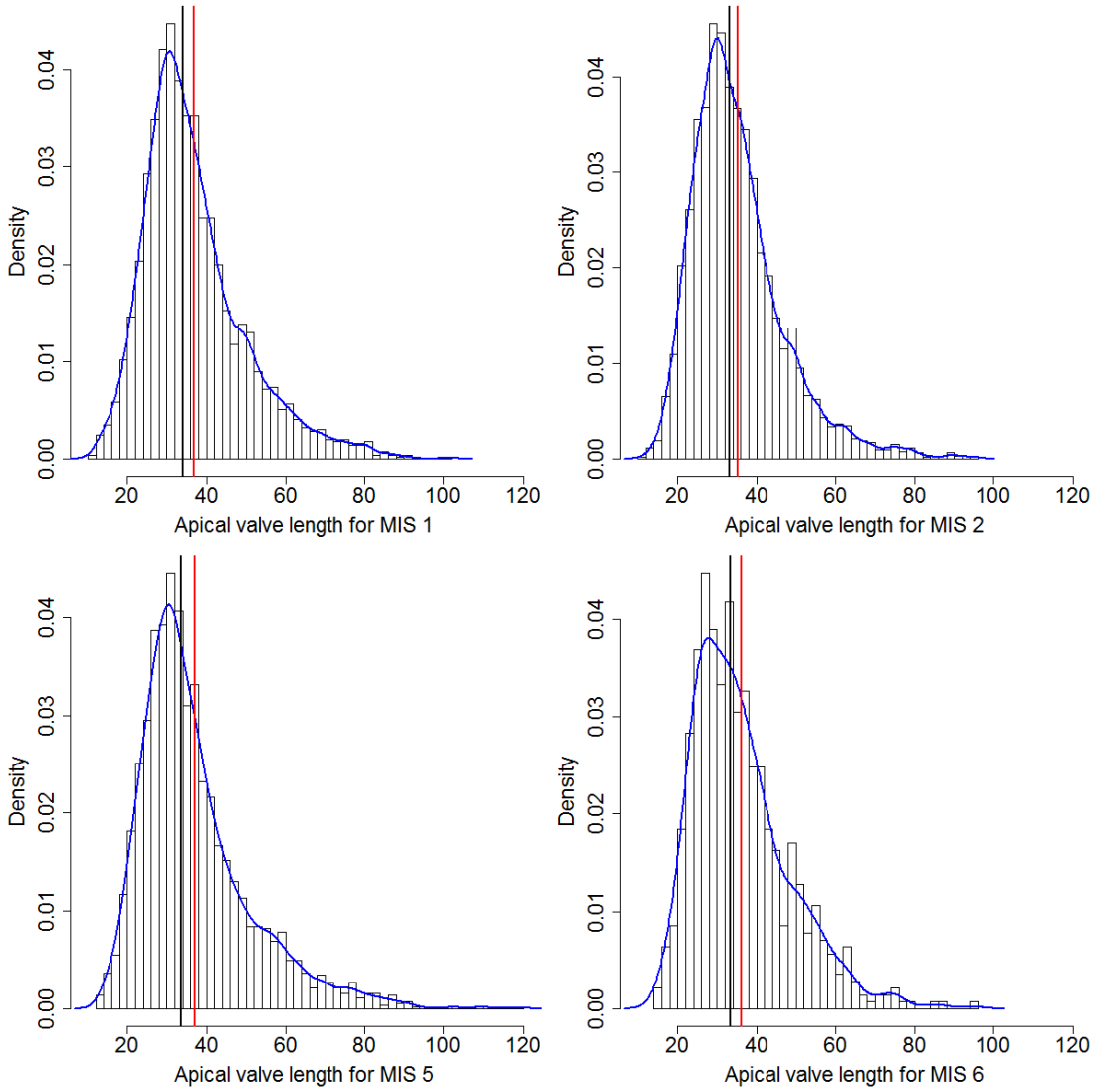


Figure A2: Distribution of apical axis length for each MIS. Red lines depict the mean, black lines the median values, and blue lines the estimated density curves.

2.4.7.5 Appendix E

Modelling apical versus transapical valve axis length. A third-order polynomial gave the best result regarding R^2, Akaike and Bayesian information criteria (see the green curve in Figure A3).

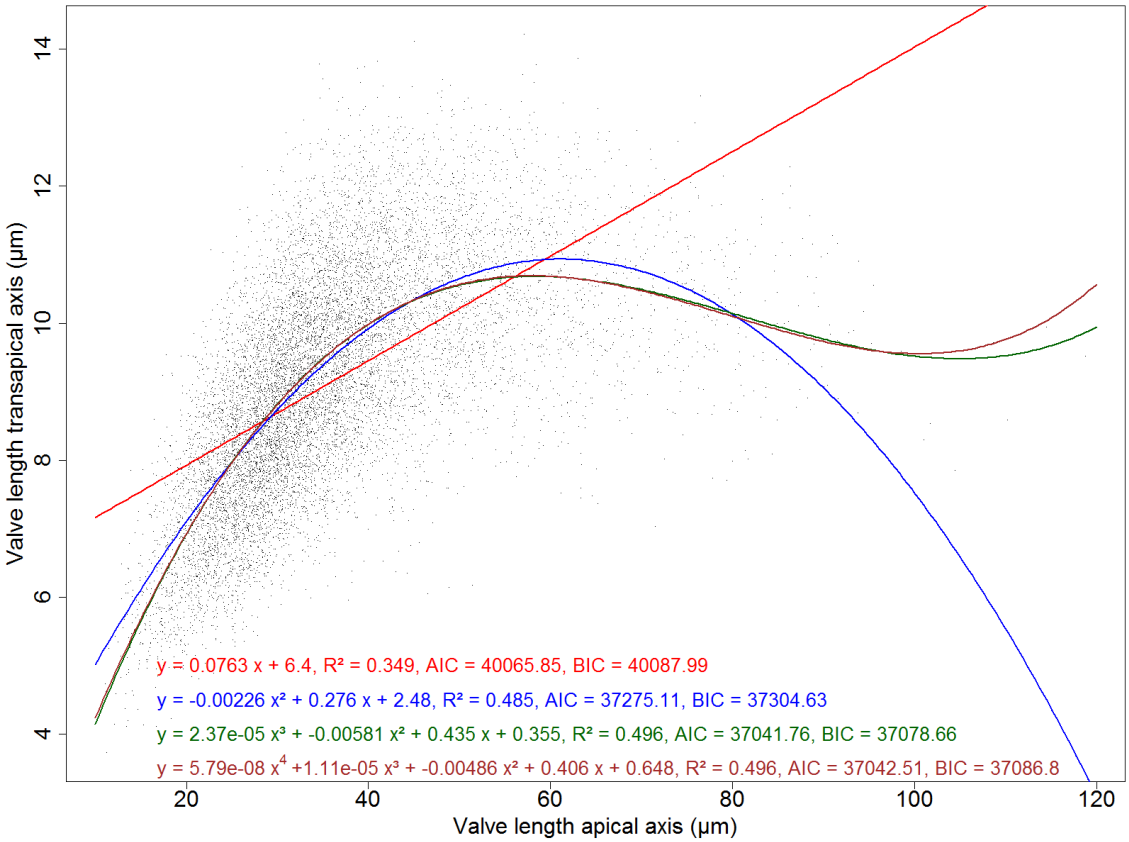


Figure A3: Models of apical versus transapical valve axis length. The polynomial model of order three (green) gives the lowest value of the Akaike (AIC) and of the Bayesian information criterion (BIC). The polynomial model of order four (brown) does not provide a significant improvement.

Statistical assessment of polynomial model order one:

R console output:

```
> ## modelling valve transapical vs. apical length
> x <- fkergdata$Width # apical axis lengths
> y <- fkergdata$Height # transapical axis lengths
> # poly(1) model
> lmWidthHeightPoly1 = lm(y~x)
> summary(lmWidthHeightPoly1)
Call:
lm(formula = y ~ x)
Residuals:
    Min      1Q  Median      3Q     Max
-6.5225 -0.8975 -0.0182  0.8838  4.7325

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.4029050  0.0365284  175.29   <2e-16 ***
```

```
x           0.0762691  0.0009571   79.69   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 1.31 on 11857 degrees of freedom
Multiple R-squared:  0.3488,   Adjusted R-squared:  0.3487
F-statistic:  6350 on 1 and 11857 DF, p-value:  <2.2e-16
```

```
> anova(lmWidthHeightPoly1)
Analysis of Variance Table
Response: y
             Df Sum Sq Mean Sq F value    Pr(>F)
x             1  10900 10899.5  6349.7  <2.2e-16 ***
Residuals 11857  20353     1.7
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Statistical assessment of polynomial model order two:
R console output:

```
> # poly(2) model
> lmWidthHeightPoly2 = lm(y~ x + I(x^2))
> summary(lmWidthHeightPoly2)
Call:
lm(formula = y ~ x + I(x^2))
Residuals:
    Min      1Q  Median      3Q     Max
-4.5560 -0.8131 -0.0312  0.7789  6.0982
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.481e+00  7.707e-02   32.19   <2e-16 ***
x            2.765e-01  3.668e-03   75.37   <2e-16 ***
I(x^2)      -2.260e-03  4.028e-05  -56.11   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 1.165 on 11856 degrees of freedom
Multiple R-squared:  0.4854,   Adjusted R-squared:  0.4853
F-statistic:  5592 on 2 and 11856 DF,  p-value: <2.2e-16
```

```
> anova(lmWidthHeightPoly2)
Analysis of Variance Table
Response: y
```

```
             Df  Sum Sq Mean Sq F value    Pr(>F)
x             1 10899.5 10899.5  8035.2  <2.2e-16 ***
I(x^2)        1  4270.5  4270.5  3148.2  <2.2e-16 ***
Residuals 11856 16082.4     1.4
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Statistical assessment of polynomial model order three:

R console output:

```
> # poly(3) model
> lmWidthHeightPoly3 = lm(y~ x + I(x^2) + I(x^3))
> summary(lmWidthHeightPoly3)
Call:
lm(formula = y ~ x + I(x^2) + I(x^3))
Residuals:
    Min      1Q  Median      3Q     Max
-4.4314 -0.7984 -0.0278  0.7729  4.3329
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.553e-01  1.576e-01   2.254   0.0242 *
x            4.350e-01  1.091e-02  39.879   <2e-16 ***
I(x^2)      -5.806e-03  2.335e-04 -24.869   <2e-16 ***
I(x^3)       2.372e-05  1.539e-06  15.415   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 1.153 on 11855 degrees of freedom
Multiple R-squared:  0.4955,   Adjusted R-squared:  0.4954
F-statistic:  3881 on 3 and 11855 DF,  p-value: <2.2e-16


> anova(lmWidthHeightPoly3)
Analysis of Variance Table

Response: y
             Df  Sum Sq Mean Sq F value    Pr(>F)
x             1 10899.5 10899.5 8195.52  <2.2e-16 ***
I(x^2)        1  4270.5  4270.5 3211.02  <2.2e-16 ***
I(x^3)        1   316.0   316.0  237.62  <2.2e-16 ***
Residuals 11855 15766.4     1.3
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Statistical assessment of polynomial model order four:

R console output:

```
> # poly(4) model
> lmWidthHeightPoly4=lm(y~ x + I(x^2) + I(x^3) + I(x^4))
> summary(lmWidthHeightPoly4)
Call:
lm(formula = y ~ x + I(x^2) + I(x^3) + I(x^4))
Residuals:
    Min      1Q  Median      3Q     Max
-4.4436 -0.7984 -0.0282  0.7705  4.3359
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.478e-01  3.061e-01   2.117   0.0343 *
x            4.064e-01  2.787e-02  14.584  <2e-16 ***
I(x^2)      -4.858e-03  8.820e-04  -5.507 3.72e-08 ***
I(x^3)       1.105e-05  1.146e-05   0.964   0.3350
I(x^4)       5.787e-08  5.190e-08   1.115   0.2649
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 1.153 on 11854 degrees of freedom
Multiple R-squared:  0.4956,   Adjusted R-squared:  0.4954
F-statistic:  2911 on 4 and 11854 DF,  p-value: < 2.2e-16


> anova(lmWidthHeightPoly4)
Analysis of Variance Table
Response: y
             Df  Sum Sq Mean Sq   F value Pr(>F)
x             1 10899.5 10899.5 8195.6842 <2e-16 ***
I(x^2)        1  4270.5  4270.5 3211.0849 <2e-16 ***
I(x^3)        1   316.0   316.0  237.6264 <2e-16 ***
I(x^4)        1     1.7     1.7    1.2432 0.2649
Residuals 11854 15764.7     1.3
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

2.4.8 References

Abelmann, A., Gersonde, R., Cortese, G., Kuhn, G. and Smetacek, V. (2006). "Extensive phytoplankton blooms in the Atlantic sector of the glacial Southern Ocean." Paleoceanography **21**(1).

Álvarez, E., López-Urrutia, Á. and Nogueira, E. (2012). "Improvement of plankton biovolume estimates derived from image-based automatic sampling devices: application to FlowCAM." Journal of Plankton Research **34**(6): 454-469.

Álvarez, E., López-Urrutia, Á., Nogueira, E. and Fraga, S. (2011). "How to effectively sample the plankton size spectrum? A case study using FlowCAM." Journal of Plankton Research **33**(7): 1119-1133.

Briseño-Avena, C., Roberts, P. L. D., Franks, P. J. S. and Jaffe, J. S. (2015). "ZOOPS- O2: A broadband echosounder with coordinated stereo optical imaging for observing plankton in situ." Methods in Oceanography **12**: 36-54.

Cefarelli, A. O., Ferrario, M. E., Almandoz, G. O., Atencio, A. G., Akselman, R. and Vernet, M. (2010). "Diversity of the diatom genus *Fragilariopsis* in the Argentine Sea and Antarctic waters: morphology, distribution and abundance." Polar biology **33**(11): 1463-1484.

Claude, J. (2008). Morphometrics with R. New York, Springer Science+Business Media, LLC.

Cortese, G. and Gersonde, R. (2007). "Morphometric variability in the diatom *Fragilariopsis kerguelensis*: Implications for Southern Ocean paleoceanography." Earth and Planetary Science Letters **257**(3-4): 526-544.

Cortese, G. and Gersonde, R. (2008). "Plio/Pleistocene changes in the main biogenic silica carrier in the Southern Ocean, Atlantic Sector." Marine Geology **252**(3-4): 100-110.

du Buf, H. and Bayer, M. M. (2002). Automatic Diatom Identification. New Jersey, London, Singapore, Hong Kong, World Scientific Publishing Co. Pte. Ltd.

Fenner, J., Schrader, H. and Wienigk, H. (1976). "Diatom phytoplankton studies in the southern Pacific Ocean, composition and correlation to the Antarctic Convergence and its paleoecological significance." Initial Reports of the Deep Sea Drilling Project **35**: 757-813.

Gersonde, R. and Hempel, G. (1990). Die Expeditionen ANTARKTIS-VIII/3 und VIII/4 mit FS "Polarstern" 1989 = The expeditions ANTARKTIS VIII/3 and VIII/4 of RV "Polarstern" in 1989. Bremerhaven, Alfred Wegener Institute for Polar and Marine Research. **74**.

Hasle, G. R. (1965). "*Nitzschia* and *Fragilariopsis* species studied in the light and electron microscopes. III. The genus *Fragilariopsis*." Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo, I. Matematisk-Naturvidenskapelig Klasse. Ny Serie **21**: 1-49.

Kloster, M. (2013). Digitale Bildsignalverarbeitung in der Bioinformatik: Methoden zur Segmentierung und Klassifizierung biologischer Merkmale am Beispiel ausgewählter Diatomeen. Emden, University of Applied Sciences Emden/Leer. **Master of Engineering**.

Kloster, M., Kauer, G. and Beszteri, B. (2014). "SHERPA: an image segmentation and outline feature extraction tool for diatoms and other objects." BMC bioinformatics **15**: 218.

Mullen, A. D., Treibitz, T., Roberts, P. L. D., Kelly, E. L. A., Horwitz, R., Smith, J. E. and Jaffe, J. S. (2016). "Underwater microscopy for in situ studies of benthic ecosystems." Nature Communications **7**: 12093.

Olson, R. J. and Sosik, H. M. (2007). "A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging FlowCytobot." Limnology and Oceanography: Methods **5**(6): 195-203.

R Core Team (2015). "R: a language and environment for statistical computing." from https://www.R-project.org

Sachs, O., Sauter, E. J., Schlüter, M., Rutgers van der Loeff, M. M., Jerosch, K. and Holby, O. (2009). "Benthic organic carbon flux and oxygen penetration reflect different plankton provinces in the Southern Ocean." Deep Sea Research Part I: Oceanographic Research Papers **56**(8): 1319-1335.

Schulz, J., Barz, K., Ayon, P., Luedtke, A., Zielinski, O., Mengedoht, D. and Hirche, H.-J. (2010). "Imaging of plankton specimens with the lightframe on-sight keyspecies investigation (LOKI) system." Journal of the European optical society-rapid publications **5**.

Schulz, J., Möller, K. O., Bracher, A., Hieronymi, M., Cisewski, B., Zielinski, O., Voss, D., Gutzeit, E., Dolereit, T. and Niedzwiedz, G. (2015). "Aquatische Optische Technologien in Deutschland." Meereswissenschaftliche Berichte.

Smetacek, V., Assmy, P. and Henjes, J. (2004). "The role of grazing in structuring Southern Ocean pelagic ecosystems and biogeochemical cycles." Antarctic Science **16**(4): 541-558.

Spaulding, S. A., Jewson, D. H., Bixby, R. J., Nelson, H. and McKnight, D. M. (2012). "Automated measurement of diatom size." Limnology and Oceanography: Methods **10**(11): 882-890.

Zielinski, U. and Gersonde, R. (1997). "Diatom distribution in Southern Ocean surface sediments (Atlantic sector): Implications for paleoenvironmental reconstructions." Palaeogeography Palaeoclimatology Palaeoecology **129**(3-4): 213-250.

Zielinski, U., Gersonde, R., Sieger, R. and Fütterer, D. (1998). "Quaternary surface water temperature estimations: Calibration of a diatom transfer function for the Southern Ocean." Paleoceanography **13**(4): 365-383.

2.5 Publication III – Insights into the life cycle of the Southern Ocean diatom *Fragilariopsis kerguelensis* through high-throughput microscopy of sediment trap samples

Michael Kloster[1]
Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Am Handelshafen 12, 27570 Bremerhaven, Germany
Hochschule Emden/Leer, Constantiaplatz 4, 26723 Emden, Germany

Andres S. Rigual-Hernández
Área de Paleontología, Departamento de Geología, Universidad de Salamanca, 37008 Salamanca, Spain

Leanne Armand
The Australian National University, 142 Mills Road Acton ACT 2601, Australia

Gerhard Kauer
University of Applied Sciences Emden/Leer, Constantiaplatz 4, 26723 Emden, Germany

Thomas W. Trull
Antarctic Climate and Ecosystems Cooperative Research Centre, Commonwealth Scientific and Industrial Research Organisation - Oceans and Atmosphere, Hobart, 7001, Australia

Bánk Beszteri
Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Am Handelshafen 12, 27570 Bremerhaven, Germany

[1]Author for correspondence: michael.kloster@awi.de, Fon: +49(471)4831-1532

2.5.1 Abstract

The life cycle of the Southern Ocean diatom *Fragilariopsis kerguelensis* is known roughly, but many of its details have not been surveyed in the species' natural habitat so far. We investigated material from a two year sediment trap time series by high-throughput imaging and image analysis, looking for morphometric signals which indicate life cycle stages. Based on changes in valve length distributions, we found three general seasonal phases corresponding to proliferation, ceasing and stagnation of the population. Furthermore we observed several transient signals, which might give further clues on population development and life cycle events. We discuss possible sources of these signals, and their relevance as possible indications for proliferation, grazing, adaption to environmental conditions, for an overwintering strategy and sexual activity.

Key index words
Morphometry, phytoplankton, SHERPA, Slide Scanner, valve size distribution

Abbreviations
ACC, Antarctic Circumpolar Current; APF, Antarctic Polar Front; $BSiO_2$, biogenic silica; EIFEX, European Iron Fertilization Experiment; HLNC, high-nutrient low-chlorophyll; n, sampling size; N, population size; *p*, p-value; PCC, Pearson correlation coefficient; PFZ, Polar Frontal Zone; POC, particulate organic carbon; RI, ranking index; RFI, relative frequency of valves within the initial cell size range; SST, sea surface temperature

2.5.2 Introduction

The pelagic pennate *Fragilariopsis kerguelensis* (O'Meara) Hustedt is endemic to the Southern Ocean, where it is one of the dominant diatom species (Hart 1942). The cells were described by (Hasle 1965) as 10 - 76 µm long and 5 - 11 µm wide, which we were able to extend up to 119.9 µm length (Kloster et al. 2017), and build colonies in form of chains, comprised of up to a dozen cells connected to each other. Thriving in the ice-free open waters of the Antarctic Circumpolar Current (ACC), it prefers surface water temperatures roughly between 0 and 10 °C (Zielinski and Gersonde 1997, Pinkernell and Beszteri 2013). A major fraction of the silicic acid brought to the Southern Ocean surface waters by the global overturning thermohaline circulation is taken up by *F. kerguelensis*, making this species the most important silica sinker of the Southern Ocean (Assmy et al. 2013). Indeed, *F. kerguelensis* is the main component of a band of silica-rich sediments that encircle Antarctica, known as the "diatom ooze belt" (Kozlova 1966, Defelice and Wise 1981, Zielinski and Gersonde 1997, Crosta et al. 2005, Shukla et al. 2016). This diatom ooze belt accounts for approximately one third of the global biogenic Si accumulation (Tréguer and De La Rocha 2013, Tréguer 2014), representing the largest sink for Si in the global ocean. Moreover, owing to the capacity of *F. kerguelensis* to thrive well in high-nutrient, low-chlorophyll waters, high abundances of *F. kerguelensis* in Southern Ocean sediments are regarded as a proxy for low-carbon high-silica exporting regimes (Smetacek et al. 2004, Abelmann et al. 2006).

F. kerguelensis has a bipartite cell wall, which is strongly silicified (Hasle 1965) and encases the cytoplasm in two pill-box like frustules. These are exceptionally robust, which might be an adaptation against grazing (Verity and Smetacek 1996, Hamm et al. 2003). The frustules are comprised of two thecae, and their terminal faces are denominated as valves. Valve area and shape were proposed as proxies for reconstructing past positions and nutrient characteristics of the Antarctic Polar Front (APF) and the diatom ooze belt (Cortese and Gersonde 2007), as well as iron availability during glacial-interglacial periods (Cortese et al. 2012, Shukla et al. 2013, Nair et al. 2015, Shukla and Crosta 2017). Furthermore, transfer functions, based on species compositions which were mostly dominated by *F. kerguelensis* (Zielinski et al. 1998, Gersonde et al. 2003, Gersonde et al. 2005), as well as other morphometric features of this species, have been related to summer sea surface temperature (Fenner et al. 1976, Cortese and Gersonde 2007).

F. kerguelensis' life cycle is characterized by a unique reproduction scheme (Round et al. 1990), which is tightly connected to their morphology. One of the two thecae is slightly smaller than the other, where the larger epitheca overlaps the smaller hypotheca. During cell division, each of the two resulting daughter cells inherits one of these thecae, but in both cases the inherited theca becomes the new epitheca regardless of its size, whilst the newly built theca is a smaller hypotheca. Because of this, one of the daughter cells has the same size as its parent – and the other is smaller. Accordingly, over time the size distribution of cells has an increasing amount of smaller cells. During this proceeding asexual reproduction, cells only become sexually inducible if their size falls below a certain limit, which usually is between 30 - 75 % of their maximum size (Davidovich 2001). Actual triggers of sexual reproduction are still unclear but considered to be diverse (Chepurnov et al. 2004). Among them, changes in the light regime, temperature and nutrient availability (Drebes 1977), as well as the emission of specific pheromones (Sato et al. 2011, Moeys et al. 2016), have been reported as possible candidates to trigger sexual reproduction. *F. kerguelensis* exhibits a dioecious mating system where opposite mating types, originating from different strains, need to be present to start gametogenesis (Chepurnov et al. 2005, Fuchs et al. 2013). The naked gametes are formed within gametangia, and, after successful mating, the resulting zygote expands to a large-sized auxospore. Inside the auxospore, the so called initial cell is formed. The initial cell has a newly

formed frustule that constitutes the largest member of the subsequent vegetative size reduction cycle. Hence, sexuality plays a double role in the diatom life cycle: it ensures genetic diversity, and restores maximum cell size. Based on recent work with *Seminavis robusta* (to be published, personal talk with G. Pohnert) it appears that for cells that have just entered the size range where sexual reproduction is possible, gametogenesis will only be triggered if all of the other factors are optimal. In contrast, at the lower end of the size range, where size-restitution is a matter of survival, also sub-optimal conditions will trigger sexual events. Accordingly, cell size distributions provide information on the controls as well as the details of diatom life cycles.

Assessments of cell size distributions require large sampling sizes, and thus diatom life cycle investigations based on them are conducted only rarely. For example (Crawford et al. 1997) determined the size of 1,000 valves per sample of *Corethron criophilum*, and (D'Alelio et al. 2010) investigated *Pseudo-nitzschia multistriata* size temporal variations analyzing 200 cells per sample. In both cases, cells were sized manually by optical light microscopy. Newer technical developments provide automation of such measurements; for example the FlowCAM® (Fluid Imaging Technologies) system was utilized by (Spaulding et al. 2012) to assess cell length distributions for three species of diatoms, with sampling sizes between 320 and 672.

(Fuchs et al. 2013) shed light on several aspects of the mating system and life cycle traits of *F. kerguelensis* using laboratory cultures. However, laboratory experiments are unrealistic in many aspects, particularly because they are only capable of reproducing natural conditions in limited ways. Therefore, field observations in the natural habitat of *F. kerguelensis* are essential to complement experimental observations in cultures, yet annual and multi-annual records of diatom populations in the pelagic waters of the Southern Ocean are rare due to logistical limitations. In this paper we present results of such study. We document the valve length distributions of *F. kerguelensis* populations collected during two years (2002 - 2004) by a sediment trap deployed in the Australian sector of the Polar Frontal Zone, in the middle of the Antarctic Circumpolar Current. The samples were analyzed using a Metafer automated slide scanner (MetaSystems Hard & Software GmbH, Altlussheim, Germany) and the diatom morphometry software SHERPA (Kloster et al. 2014). Our results represent the first observations of seasonal life cycle of *F. kerguelensis* in its natural habitat, clarifying influences of reproduction and grazing on the population, and hypothesizing an overwintering strategy.

2.5.3 Material and Methods

We analyzed material collected by a bottom-tethered sediment trap located at 54° S 140° E in 800 m depth, covering the period from November 2002 to October 2004. Details on the sediment trap, sampling and material processing are described in (Rigual-Hernández et al. 2015).

2.5.3.1 Oceanographic setting

The sediment trap site is located within the central area of the Polar Frontal Zone (PFZ) along the 140° E meridian. The PFZ is delimited to the north by the Subantarctic Front (SAF) and to the south by the Antarctic Polar Front (APF). These two deep-reaching fronts account for a significant part of the eastward transport of the Antarctic Circumpolar Current (Nowlin and Clifford 1982, Rintoul and Sokolov 2001) while the PFZ waters are characterized by weak currents and relatively homogeneous physical and chemical properties. The PFZ waters south of Tasmania can be defined as a high-nutrient low-chlorophyll (HNLC) regime, most likely due to the low dissolved iron concentrations in the region (Lannuzel et al. 2011). Indeed, it has been suggested that primary production is co-limited

by iron availability and light in the PFZ waters (Boyd et al. 2001, Lannuzel et al. 2011). Mixed-layer depths in this region are often shallower than 200 m (Rintoul and Trull 2001).

Phytoplankton communities in the PFZ surface waters south of Tasmania exhibit high abundances of diatoms, as well as coccolithophores, flagellates and relatively low concentrations of cyanobacteria (Wright et al. 1996, Popp et al. 1999, Kopczynska et al. 2001, de Salas et al. 2011).

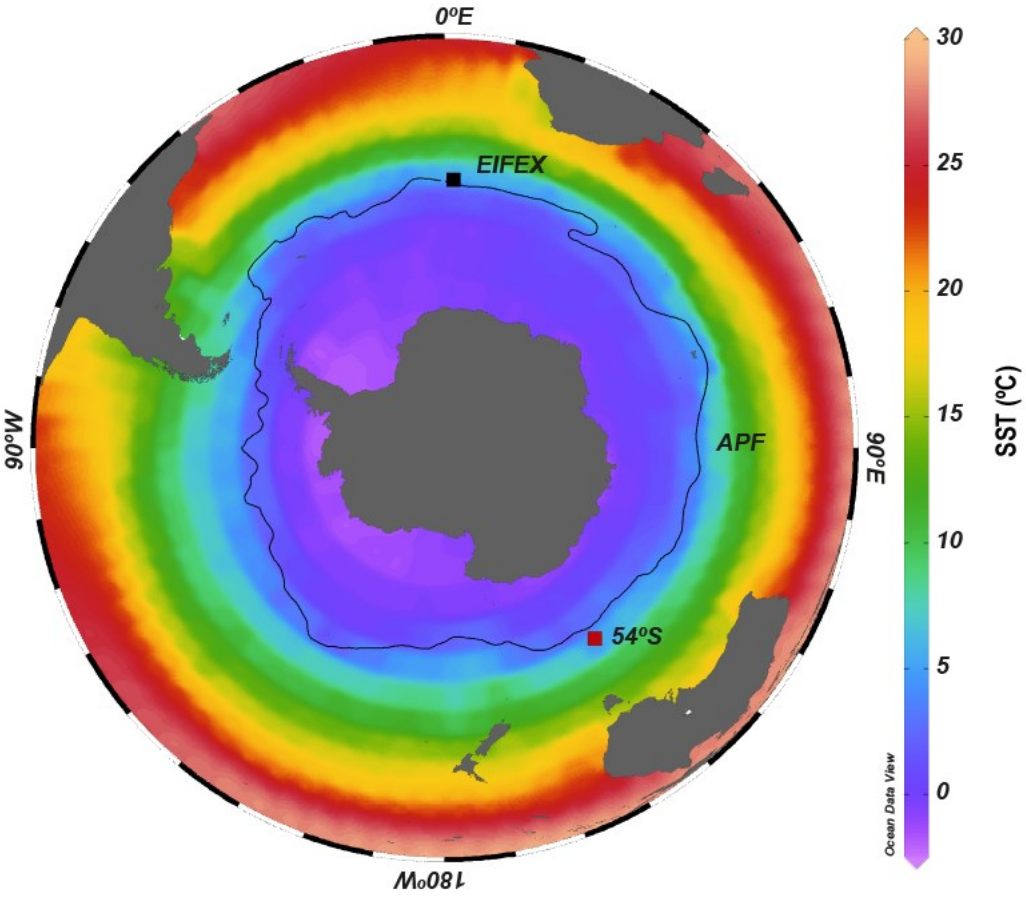


Figure 1: Map of the Southern Ocean with annual sea surface temperatures (SST; World Ocean Atlas 2013) showing the location of the 54° S station and the site of the European Iron Fertilization Experiment (EIFEX). Black line represents the position of the Antarctic Polar front (Orsi et al. 1995).

2.5.3.2 Microscopy

Microscope slides were prepared by embedding the processed siliceous material into Norland Optical Adhesive NOA 61 (Norland Products Inc., Cranburry, New Jersey, USA). Microscopy and morphometric measurements were conducted using a Metafer automated slide scanner (MetaSystems Hard & Software GmbH, Altlussheim, Germany) and our diatom morphometry software SHERPA version 1.1c (Kloster et al. 2014) according to our previously described workflow (Kloster et al. 2017). Briefly, a rectangular area on each slide was scanned, in overlapping fields of view, with a Zeiss Plan-NEOFLUAR 20× objective. These low resolution images were compared by SHERPA to a template library of diverse outline shapes representing the genus *Fragilariopsis* (see supplement ”SHERPA settings”), to locate valves of interest in the captured area. Positions where valves were found were then revisited in the Metafer system with a high resolution oil immersion objective (Zeiss Plan-APOCHROMAT 63×/1.4), each captured in 20 different focal planes, and the resulting images combined to extended depth of focus images. The latter were then analyzed with SHERPA for morphometric measurements. Settings files used for SHERPA analyses of both, the low

and high resolution images as well as the according shape templates, are included in the supplementary material, see “SHERPA settings”.

Because of the relatively low refractive index (1.56) of the NOA 61 mountant, imaging contrast was poorer than for instance in Naphrax, causing a decreased accuracy in object segmentation and outline detection. To compensate this shortcoming, increased manual intervention was necessary in SHERPA, for manual post-editing as well as reviewing segmentation results (see supplement “Modified SHERPA procedure”).

2.5.3.3 Data analysis

For tracing auxosporulation / size restitution events, the relative abundance of very large valves was determined. *F. kerguelensis* average valve length peaks in the Southern Ocean sediments beneath the APF (Cortese and Gersonde 2007). Close to the APF, in the Atlantic sector of the Southern Ocean (49° S 2° E), the only observation of *F. kerguelensis* sexual activity in its natural habitat was made during the EIFEX experiment (Assmy et al. 2006). Since our sampling site also is close to the APF, we assume that data of valve length measured at both sites should be comparable. For the EIFEX experiment, auxospores of *F. kerguelensis* containing initial thecae were reported to be at least 76 µm long. We assumed that valves longer than 75.5 µm originate from either initial cells or from cells which descended from them by only few divisions, and refer to them as within the initial cell size range. These large cells were very rare, and their counts are affected significantly by the “ranking index” (RI) up to which results were taken into account (please see supplement “Modified SHERPA procedure” and (Kloster et al. 2014, Kloster et al. 2017) for further information). Accordingly, relative abundance calculated from counts utilizing different highest RIs might be biased (respective results are listed in supplement “Valves within the initial cell size range”). To compensate for that, they were searched for in an additional reviewing run, by examining high-resolution results of RI 0 to 6 for each slide, regardless of the maximum RI used for retrieving the valve length distribution. The corresponding counts were normalized to the amount of material suspension applied to the slide, the scanned slide area (approximated by the number of fields of view scanned at low magnification), the length of the sampling period for the corresponding cup of the sediment trap, and the *F. kerguelensis* flux rates from (Rigual-Hernández et al. 2015) to characterize the relative frequency of valves within the initial cell size range (RFI) in the population (see Equation 1). These values are using an arbitrary scale and are only comparable between each other.

$$RFI = \frac{(number\ of\ valves\ \geq 75.5\ \mu m) \cdot 10^{12}}{amount\ of\ suspension \cdot nFOVs \cdot spl \cdot valve\ flux} \quad \textbf{(Equation 1)}$$

with nFOVs = number of scanned fields of view, spl = sampling period length

A rough estimation of the growth rate (see Equation 2) is possible for periods with an ongoing decrease of mean valve length, based on an average size reduction of 0.28 µm / division observed by (Fuchs et al. 2013), and assuming that the observed size decrease was produced by vegetative proliferation only.

$$est.\,growth\ rate\ (divisions\ per\ day) = \frac{l_2 - l_1}{(t_2 - t_1) \cdot (-0.28\ \mu m/division)} \quad \textbf{(Equation 2)}$$

with l_1, l_2 = mean valve length at time t_1, t_2

Of the morphometric parameters determined by SHERPA, for this publication we present the apical valve lengths and the variation in their distribution over time. Downstream analysis of counts and valve sizes was performed using scripts programmed in R (R Core Team 2015) (see supplement "R-scripts"). Measurements of additional features like transapical length, valve area, costae distance and F (Fenner et al. 1976) or F* (Cortese and Gersonde 2007) are included in supplement "Additional measurements". The complete data set is provided in supplement "R" of the annexed data medium.

2.5.4 Results

We observed distinct seasonal trends for the two sampling periods 2002/03 and 2003/04, mostly with similarity regarding seasonal trends, but differences in detail. The according trend-lines in (Figure 2 & 3 a, c – h) were derived by LOESS smoothing, with the exception of chlorophyll-*a* data, where linear interpolation was used to keep the graph consistent with the original publication.

2.5.4.1 Chlorophyll-a concentration

The chlorophyll-*a* data were plotted to provide ecological background to the ensuing morphometric measurements (Figure 2 & 3 a and supplement "Chlorophyll-*a* data", derived from NASA's Giovanni online data system) as they present an estimation of algal biomass accumulation in the area between 55.5–52.5° S and 130–150° E, aiming to reflect the chlorophyll-*a* concentration of the region sampled by the 54ºS site traps (Rigual-Hernández et al. 2015). The concentration was measured monthly, with the exceptions of July and August, when no data is available.

2.5.4.2 Sediment trap material

Signals originating from the sediment trap material are postponed by an estimated 2 to 6 weeks compared to the chlorophyll-*a* data. This is due to the time required for material to reach the sediment trap in 800 m depth (discussed below). Below, dates are referred to by the middle day of the sampling period. Sampling periods and intervals differed for both seasons: In 2002/03, a period of 290 days was covered, from November 18th 2002 to September 4th 2003, with sampling intervals mostly covering 10 days. For season 2003/04, a period of 378 days was covered, from September 27th 2003 to October 9th 2004, with sampling intervals for the most part integrating over 14 days.

Valve flux rates (Figure 2 & 3 b) for *F. kerguelensis* were estimated previously (Rigual-Hernández et al. 2015) and show different patterns for both seasons. In 2003/04, the flux was significantly lower compared to the previous season, but the period of high flux lasted nearly twice as long, from mid-spring / November until the middle of autumn / April.

Valve length distributions were measured from nearly 29,000 *F. kerguelensis* valves, and were mostly asymmetrical (Figure 2 & 3 c); because of this we use the median instead of the mean as measure of central tendency, even though both exhibit the same patterns. Their variability displayed the following commonalities between both seasons: Median valve size decreased during late spring and first half of summer, then increased until ca. mid-autumn, and, after a decrease, stayed stable during the rest of the year, with a slight increase around late winter / early spring. The same applies to valve width (transapical axis length), valve area, costae distance and F or F*; the corresponding data are presented in supplement "Additional measurements".

In the season 2002/03, median valve length varied pronouncedly over the year, ranging from 20.8 to 43.9 µm (ca. 24 to 38 µm for the smoothed median). The minimum was reached in mid-January 2003 / mid-summer, the maximum middle of April / mid-autumn. After a decrease until May / late autumn, median lengths remained stable until mid-winter at ca. 33 µm, and increased to 38 µm

towards late August / late winter.
During sampling season 2003/04, this increasing trend continued, and peaked at 41.1 µm in early October 2003. Within the timeframe comparable to the preceding year (mid-November 2003 until early September 2004), median valve lengths varied less, from 28.5 to 38.5 µm (ca. 30 to 38 µm for the smoothed median). Shifts were less pronounced compared to the previous season, but inflexion points of trend lines, laid on skewness as well as minimal and maximal sizes, occurred at about similar points of time. The minimum was reached in late January 2004 / mid-summer, an intermediate maximum of 37.4 µm in middle of April 2004 / mid-autumn, and the absolute maximum in early August 2004 / mid-winter. After median valve lengths decreased in late autumn, they remained stable at ca. 36 µm, and increased to 40.9 µm for the last sample of the season in October 2004.

Skewness of valve length distributions (Figure 2 & 3 d) was elevated throughout, that is measurements of valve lengths were right-skewed, with prolonged tails at large sizes, with the exception of cup 1.16. Changes in skewness were similar during both seasons, although less pronounced in the second one. Skewness followed the patterns of valve length medians in an inversed manner, with a Pearson correlation coefficient (PCC) of -0.82 (N = 41, $p = 3.384 \cdot 10^{-11}$). This reveals that valve length distributions with higher medians were more symmetric.

Minimal and maximal valve length (Figure 2 & 3 e, f), depicted by the first and 99th percentile, showed a similar general pattern as the median valve lengths (N = 41, first percentile: PCC = 0.85, $p = 3.274 \cdot 10^{-12}$; 99th percentile PCC = 0.78, $p = 1.759 \cdot 10^{-09}$).

Ratio of reproduction size classes (Figure 2 & 3 g): We defined here sexually inducible size classes operationally as below 31.5 µm apical valve length, after (Assmy et al. 2006) who observed gametangia of maximal 31 µm length. The percentage of sexually inducible cells varied significantly, following median valve length with a PCC of -0.98 (N = 41, $p < 2.2 \cdot 10^{-16}$).
In season 2002/03, the fraction of the size class capable of sexual reproduction ranged from 86.6 to 17.9 % of the whole population, with the maximum reached in late January 2003 / mid-summer, and the minimum in late April 2003 / mid-autumn. Subsequently, the proportion of cells within the sexually inducible size range rose up to 48 % and decreased throughout the winter 2003.
During sampling season 2003/04, this decreasing trend continued, resulting in only 7.5 % of the population being capable of sexual reproduction by October 2003. Within the time frame comparable to the previous year, a similar but less pronounced pattern was visible, where the maximum of 58.7 % occurred at the end of January 2004 / mid-summer, an intermediate minimum of 26.2 % in the middle of April 2004 / mid-autumn, and the absolute minimum in early August 2004 / mid-winter. For the last sample of the season, in early October 2004, the sexually inducible size fraction dropped down to 14.8 %.

The relative frequency of valves within the initial cell size (RFI, see Figure 2 & 3 h) refers to valves of at least 75.5 µm apical valve length. These were quantified separately, and their relative abundance is expressed on an arbitrary scale, not as percentage of total measured population (see Equation 1). This was necessary because valves in this size range represented only a very minor part of the populations and occurred in very low numbers in the overall samples. RFI varied in the range between 0 and 0.4 and showed distinct patterns for both seasons investigated (see Figure 2 & 3 h and supplement “Valves within the initial cell size range”): It exhibited transient peaks in mid-summer for both seasons, and elevated periods for winter 2003 and throughout mid-summer to mid-winter 2004.

Season 2002/03: A low relative frequency between 0 and 0.07 persisted from late November 2002 / late spring until mid-April 2003 / mid-autumn, with the exception of 0.14 in late January 2003 / mid-summer. Subsequently, the frequency rose up to a maximum of 0.28 in late August 2003 / end of winter.
Season 2003/04: During early October and mid-December 2003, the relative frequency of valves within the initial cell size range dropped to 0.01. After an increase to a maximum of 0.4 by January 2004 / mid-summer, it fluctuated at an elevated level of ca. 0.17 until end of June / mid-winter, and decreased subsequently.

Explicit values of the above mentioned measurements are listed in supplement "Summary statistics of *F. kerguelensis* valve size distributions".

2.5.4.3 Seasonal characteristics

The general patterns of valve length distribution shifts were similar between both seasons, despite varying extent of the shifts and substantial differences regarding Chlorophyll-*a* concentration as well as *F. kerguelensis* valve flux. These patterns were strongly influenced by the ratio between smaller and larger cells, which becomes apparent in the distributions' shape, as well as by their mean or median (Figure 2 & 3 c, d). The median valve length correlated with the percentage of sexually inducible cells (Figure 2 & 3 g turquoise bars) with a PCC of -0.98 (N = 41, $p < 2.2 \cdot 10^{-16}$). Because of the variable and mostly pronounced asymmetry of valve length distributions, we prefer regarding the median over their mean as a major signal, but both of them showed similar trends. The common seasonal patterns are summarized below by dividing the year into three different phases (see colored areas in Figure 2 & 3 b-h), according to the inflexion points of trend curves laid onto valve length median (Figure 2 & 3 c), distribution skewness (Figure 2 & 3 d) and extremes (Figure 2 & 3 e, f).

- Phase 1 (highlighted in orange color in Figure 2 & 3): During late spring and early summer, average valve lengths decreased. Percentage of smaller valves (below the sexual cut-off), as well as minimal and maximal valve lengths, increased. By the end of this phase, valve length distributions were highly right-skewed and dominated by smaller cells.
- Phase 2 (highlighted in green color): From late summer to mid-autumn, average valve lengths increased. Percentage of smaller valves dropped, whereas minimal and maximal valve lengths increased, skewness of valve length distributions tapered off.
- Phase 3 (highlighted in blue color): During late autumn and winter, average, minimal and maximal valve lengths remained stagnant at high values, with a slight increase towards the end of the phase / early spring.

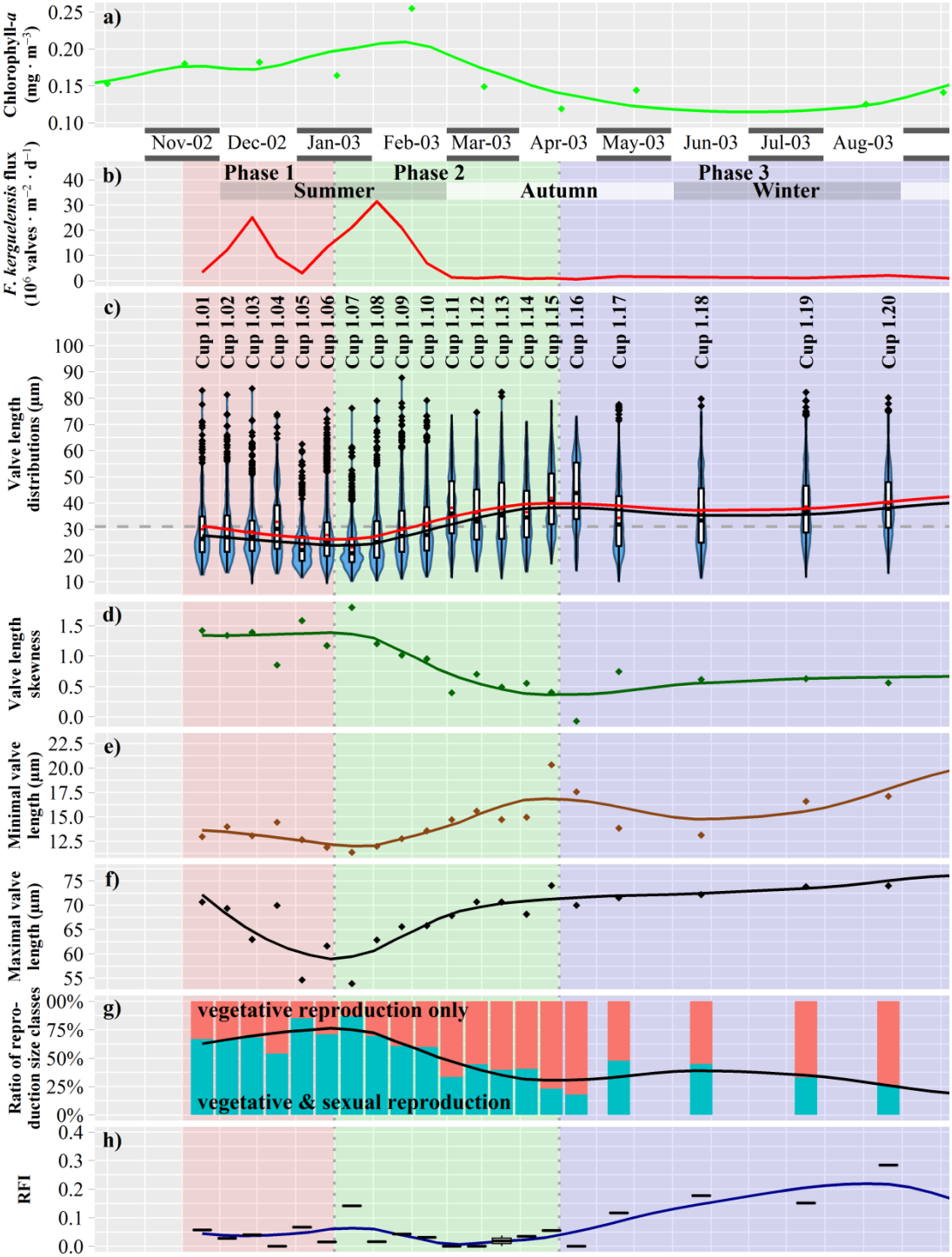


Figure 2: Temporal variability during sampling season 2002/03, x-axis refers to the middle day of sampling periods. a) Chlorophyll-a concentration in the upper water column. b-h) Sediment trap data from 800 m depth, regarding *F. kerguelensis*; depending on the sinking speed, the timeline is delayed by ca. 2 to 6 weeks compared to the chlorophyll data. Trend lines are derived by LOESS smoothing: b) Valve flux. c) Valve length. Blue violin plots depict the value distributions, box plots represent quartiles, red dots indicate the mean, black dots inside the boxplots the median. The dashed line depicts 31 µm, the size limit below which sexual reproduction can occur. d) Seasonal changes of the valve length skewness. e) Minimal valve length by 0.01 quantile. f) Maximal valve length by 0.99 quantile. g) Ratio of size fractions which are sexually inducible (that are capable of vegetative as well as sexual reproduction, with a valve length below 31 µm, highlighted in turquoise), or capable of vegetative reproduction only (valve length above 31 µm, highlighted in orange). h) Relative frequency of valves within the initial cell size range (RFI), as an indicator for initial cells.

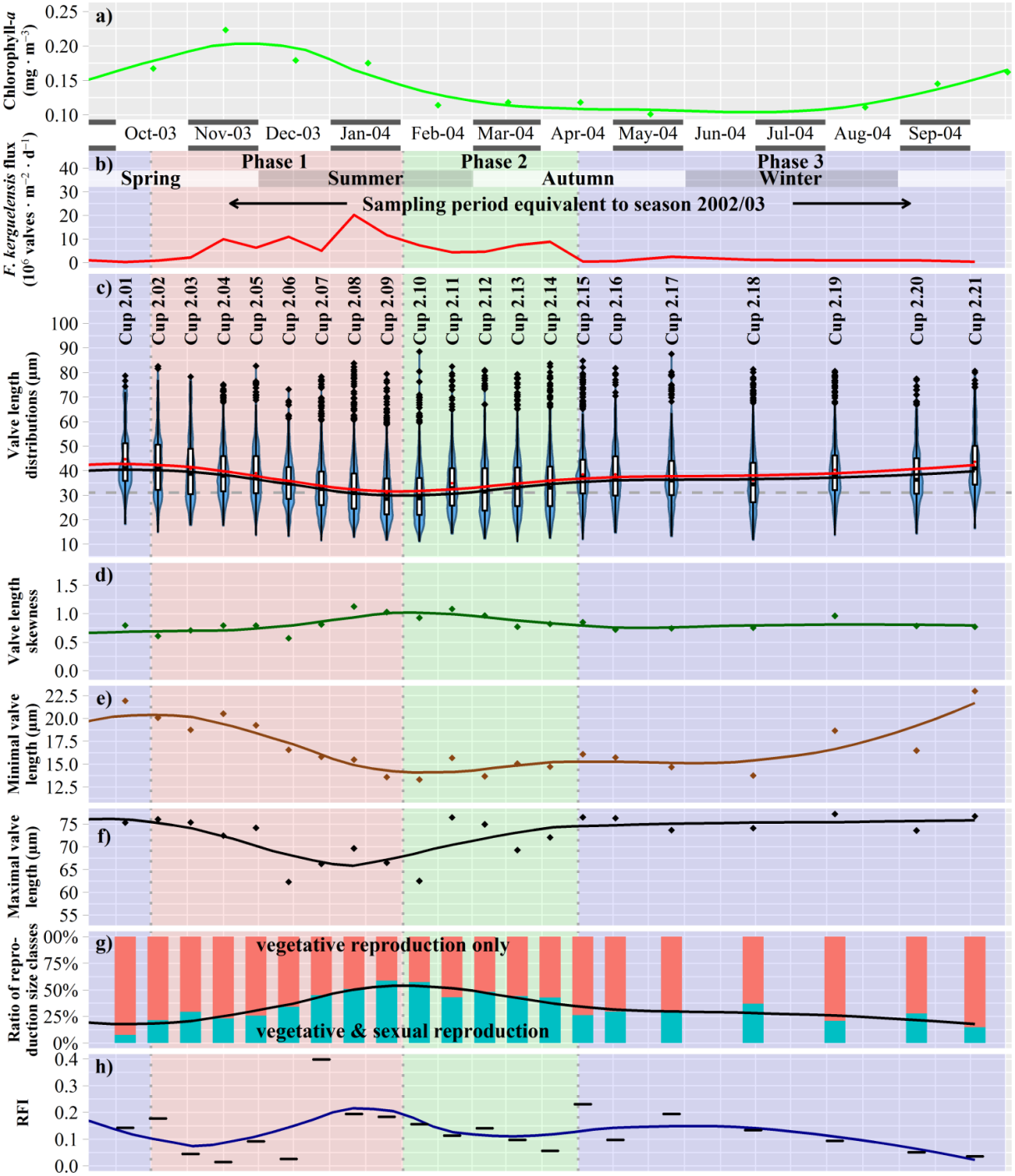


Figure 3: Temporal variability during sampling season 2003/04, x-axis refers to the middle day of sampling periods. a) Chlorophyll-a concentration in the upper water column. b-h) Sediment trap data from 800 m depth, regarding *F. kerguelensis*; depending on the sinking speed, the timeline is delayed by ca. 2 to 6 weeks compared to the chlorophyll data. Trend lines are derived by LOESS smoothing: b) Valve flux. c) Valve length. Blue violin plots depict the value distributions, box plots represent quartiles, red dots indicate the mean, black dots inside the boxplots the median. The dashed line depicts 31 µm, the size limit below which sexual reproduction can occur. d) Seasonal changes of the valve length skewness. e) Minimal valve length by 0.01 quantile. f) Maximal valve length by 0.99 quantile. g) Ratio of size fractions which are sexually inducible (that are capable of vegetative as well as sexual reproduction, with a valve length below 31 µm, highlighted in turquoise), or capable of vegetative reproduction only (valve length above 31 µm, highlighted in orange). h) Relative frequency of valves within the initial cell size range (RFI), as an indicator for initial cells.

2.5.5 Discussion

Because of the unique reproduction scheme of diatoms, events in their life cycle can be traced from cell size distributions. Continuous long-term monitoring of phytoplankton is, however, not feasible in the Southern Ocean due to the costs and efforts of year-round on-site sampling. To overcome this problem, and as a test case for our diatom morphometry workflow (Kloster et al. 2017), we

investigated sediment trap material collected at 800 m depth at sampling intervals between 10 and 45 days, covering two complete years. *F. kerguelensis* was selected as target species due to its importance in regulating ocean nutrient stoichiometry, key role at the base of Southern Ocean ecosystems, and to its usefulness as proxy for reconstructing past sea-surface conditions (Zielinski and Gersonde 1997, Smetacek et al. 2004, Assmy et al. 2013). By looking at year-round variations in valve length distributions, we provide new insights into the life cycle of *F. kerguelensis* in its natural habitat by using morphometrics based on high throughput microscopy and image analysis.

2.5.5.1 Particles collected by the sediment trap

It is important to be aware of the varying and differing sampling intervals. To keep figures and discussion congruent throughout the discussion, we refer dates to the time line of the sediment trap signals. This is postponed by several weeks compared to the situation within the upper water column, depending on size, shape and buoyancy of the sinking particles, perturbation of the mixed layer and processing by zooplankton. (Ebersbach et al. 2011) investigated particle fluxes in a similar environmental setting in Polar Frontal Zone waters at 54° S 147° E. Their observations were made during late January / early February 2007, in a situation where the phytoplankton bloom was strongly declining. They used an indented rotating sphere trap at 320 m depth to determine particle sinking speed. About 30 % of the biogenic silica was observed to be sinking slower than 10 m $\cdot$ d^{-1}, about 30 % sinking between 10 and 100 m $\cdot$ d^{-1}, and about 40 % sinking faster than 100 m $\cdot$ d^{-1}. In a gel trap at 290 m depth, small diatom chains or individual diatom cells comprised 74 % of the particle counts, whilst fecal pellets and aggregates comprised only 26 % thereof, but, due to their size, nearly 100 % of the flux volume. (Rigual-Hernández et al. 2016) estimated sinking rates ranging between 20 and 50 m $\cdot$ d^{-1} from November to January, for covering the distance between our sediment trap (800 m) and a second trap installed at 1,500 m. For the same site, (Closset et al. 2015) reported a sinking rate estimation of maximal 35 m $\cdot$ d^{-1} between July 1999 and August 2000, with a strong seasonal variability and the highest rates achieved during spring and summer. (Waite and Nodder 2001), using different methods, found sinking rates between 0.47 and 2.4 m $\cdot$ d^{-1} in February 1999 at 61° S, 141° E between 5 m and 40 - 70 m depth during a bloom induced by iron fertilization, where *F. kerguelensis* contributed substantially to the population. Based on these numbers, we estimate the time lag for signals from actively living populations in the surface mixed layer compared to signals from material collected by the sediment trap at 800 m depth, to be roughly between two and six weeks, but there is a large and not quantifiable uncertainty even in this broad estimate.

Diatoms reaching the sediment trap depth are mainly dead, with few exceptions of cells possibly surviving grazing (Jansen and Bathmann 2007), but nevertheless ending up as part of a fecal pellet or aggregate sinking down. Possible reasons for diatom death are natural mortality, starving, pathogens or grazing. Close to our study area (54° S 147° E), grazing clearance rates from deckboard incubations suggest that 42 % of the primary production per day were consumed by grazing (Pearce et al. 2011). Although the latter observations were limited to a short observational period and to a single site, the results of (Pearce et al. 2011) suggest that grazing plays an important role in the control of particle export at out sampling site. Furthermore, (Assmy et al. 2013) described *F. kerguelensis* life cycle strategy as involving *"controlled, low-level, quasiconstant mortality in this species with sinking out of empty frustules and recycling of cytoplasm in the surface layer"*, which would provide for a perpetual trickling of *F. kerguelensis* valves down into the sediment trap.

2.5.5.2 Population dynamics

We can draw indirect inferences about population dynamics from species specific valve flux rates and surface chlorophyll-*a* concentration (Figure 2 & 3 b, a). These were highly seasonal, with considerable

differences between the investigated seasons:
In summer 2002/03, the productive season of *F. kerguelensis* comprised two periods of enhanced valve export, indicated by peaking valve flux rates in combination with elevated chlorophyll-*a* concentrations. During the first flux maximum, *F. kerguelensis* provided the largest fraction of the valve flux with ca. 40 % (see supplement "Flux data"). The flux from the second bloom was even more strongly dominated by *F. kerguelensis*, by ca. 70 %. After termination of this bloom, the chlorophyll-*a* concentration remained high, while flux of biogenic silica stayed low, whereas fluxes of $CaCO_3$ and particulate organic carbon were within roughly the same range as during the bloom (see (Rigual-Hernández et al. 2015) and supplement "Flux data"). This can probably be explained by proliferation of phytoplankton which did not reach the sediment trap, succeeding the productive phase of strongly silicified diatoms, for example *Chaetoceros* subg. Hyalochaetae or *Phaeocystis*.
In 2003/04, the *F. kerguelensis* productive season lasted much longer, and compared to the previous year, it both started earlier and ended later. Changes in flux rates were less pronounced, with only a single, less prominent peak at about the time of year corresponding to the second bloom of the previous season. This flux event was dominated by *F. kerguelensis* by ca. 65 % (see supplement "Flux data") and corresponded, taking into account the shifted time lines, to elevated chlorophyll-*a* concentration.

2.5.5.3 Changes in valve lengths distributions

Shifts in valve length distributions toward shorter cells indicate vegetative reproduction, which adds progressively smaller cells to the population, according to the Macdonald-Pfitzer rule (Macdonald 1869, Pfitzer 1869). Shifts towards larger cells can be caused either by loss of smaller cells, or by sexual reproduction (combination of a loss of small gametangial frustules and addition of initial cells). Stagnant distributions are a signal for low reproductive (incl. vegetative and sexual) activity. We observed three phases of valve length distribution shift, which are interpreted in the following way in this framework:

- During phase 1 (highlighted in orange color in Figure 2 & 3) the median valve lengths were dragged down by an accumulating proportion of short valves. This was probably caused by prevailing vegetative reproduction, adding an increasing quantity of progressively smaller cells to the population. We assume that the reason for decreasing minimal valve length is that high division rates forced cells to converge towards their lower size limit. We also hypothesize that the reason for decreasing maximal valve length is that their absolute abundance was not replenished significantly by sexual reproduction (discussed below), and their relative abundance was reduced by a "dilution effect" caused by the massive increase in the abundance of smaller cells.
- During phase 2 (highlighted in green color) median valve lengths increased by a loss of short valves (see Figure 4 for a detailed view). This seems to reflect a combined effect of ceasing vegetative reproduction and selective removal of smaller cells from the population. The latter might have been caused by a substantial gametogenesis event, the incapacity of smaller cells to resist declining environmental conditions, like reduced light availability, or by grazing selective to smaller cells. The latter two could also explain the increase of minimal valve length, whilst maximal valve length might rise due to an inversion of the "dilution effect" described above.
- During phase 3 (highlighted in blue color) valve length distributions were mostly stagnant and contained the lowest proportion of short valves. This indicates low (probably light limited) reproductive activity during the winter.

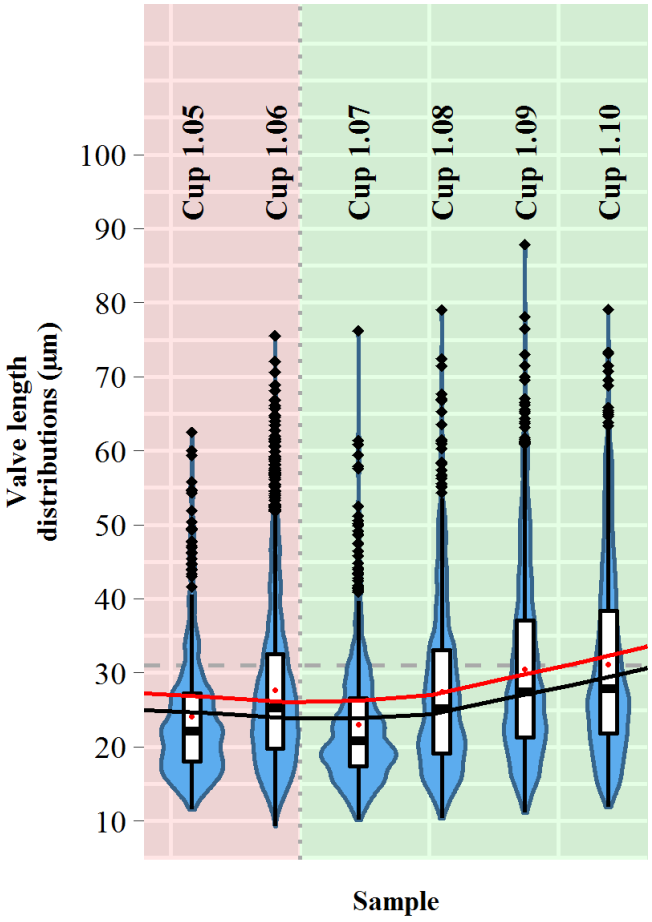


Figure 4: Temporal variability of valve length distributions showing the ongoing loss of short valves during phase 2 of sampling seasons 2002/03 (detail from Figure 2). Blue violin plots depict the valve length distributions, box plots represent quartiles, red dots indicate the mean, black dots inside the boxplots the median, red and black lines the respective trend lines derived by LOESS smoothing. The dashed line depicts 31 µm, the size limit below which sexual reproduction can occur. Phase 1 is highlighted by red, phase 2 by green background.

2.5.5.4 Sexual reproduction

Sexual stages of diatom life cycles are directly related to cell size. Only cells at the lower range of a species' size spectrum are sexually inducible and can produce gametangia. In contrast, the largest diatom cells (initial cells) originate from auxospores usually generated by sexual reproduction. These subsequently propagate by vegetative reproduction for prolonged time periods, and add size cohorts to the population by producing an exponentially increasing number of progressively shrinking cells, according to the MacDonald-Pfitzer rule (Macdonald 1869, Pfitzer 1869). In the case of *Pseudo-nitzschia multistriata,* (D'Alelio et al. 2010) observed distinctly multimodal size distributions, which allowed the detection of both size fractions mentioned above, as well as of explicit cohorts in a population.

Gametangia of *F. kerguelensis* have been reported to be between 6.9 µm (Fuchs et al. 2013) and 31 µm (Assmy et al. 2006) or 36 µm long (Fuchs et al. 2013). We relate valves within that size range to cells which are sexually inducible, taking the lower value as a conservative estimate. Accordingly, elevated abundance of these valves (see Figure 2 & 3 c - portions below the dashed lines, and g - turquoise bars) indicates a population composition beneficial for sexual propagation. Cells within this size range were present continuously throughout both sampling seasons, but dominated the population most strongly in mid-summer, at the transition between phases 1 and 2 (see Figure 4 for a detailed view). This also coincides with environmental conditions favorable for vegetative growth, which can also be interpreted as favoring sexual reproduction by increasing the chance of encounter of gametes.

Auxospores of *F. kerguelensis* containing initial thecae have been reported to be at least 76 µm long (Assmy et al. 2006). Taking into account the precision of our measurements, we relate valves of at

least 75.5 µm length to either initial cells or close descendants of them. Elevated abundance of these valves is a clear indicator that a sexual event has occurred, even though its explanatory power is restricted for our data, because these valves are rare and account for not more than ca. 2 %, and often much less, of all results (see supplement "Valves within the initial cell size range"). To overcome the problem that valves of this size might be underrepresented by the method we used for measuring valve size distributions, we use the relative frequency of initial cells (RFI, see Equation 1 and Figure 2 & 3 h) as a comparable relative measure for the abundance of cells within the initial cell size range.

Clearest indication of a synchronized sexual reproduction event in a size distribution time series would be a period with abundant cells in the sexually inducible size range, followed by the decrease thereof and the appearance of valves in the initial size range, and further followed by the emergence of a distinct large size cohort moving towards smaller sizes with time. In our data, we did not see such a clear-cut signal, but we observed a number of cases where the relative abundance of small cells decreased rather suddenly.

This happened for instance around January 2003, at the transition between phases 1 and 2 (see Figure 4 for a detailed view). Here, within the short time between cups 1.06 – 1.08, we saw a strong increase, followed by a strong decrease in the portion of valves below ca. 20 µm length. At the same time (cup 1.07), RFI transiently peaked at ca. 0.14, but was mostly below 0.05 throughout the rest of the summer. Because this transient peak was not followed by similarly increased RFI values in subsequent samples, it does not clearly signal a sexual event. A similar transient decrease in the proportion of small cells was observed between cups 1.04 – 1.06, but its significance is restricted by the low quantity of valves sampled from cup 1.04. A further similar, but less pronounced, event took place in February 2004, again at the transition between phases 1 and 2 (cups 2.09 – 2.11). Like in the other listed cases, these signals can be produced by massive and synchronized sexual reproduction, which would produce increased amounts of short valves shed from gametangia, but we do not think this probable, because we saw no indication of massive sexual reproduction in the RFI (see below) shortly afterwards. During late autumn and winter, from May to October 2003, the situation was inversed. Here cells within the sexually inducible size range comprised only 26 to 48 % of the population, whilst RFI persisted at an elevated level of ca. 0.18. This might signal sexual activity of a detectable, although still low, proportion of the population.

Subsequently, RFI appeared low throughout late spring and early summer / November until middle of December 2003. However, the smoothed trend line indicated a slight increase, which can be explained by the high value for late December; to keep the amount of smoothing consistent throughout Figure 2 c – h, we nevertheless decided to stick to the same smoothing procedure. Throughout second half of summer, autumn and early winter, from late December 2003 until end of June 2004, after a transient maximum of ca. 0.4, RFI fluctuated at values around 0.17, comparable to the previous winter, with the portion of sexually inducible cells ranging between 26 and 59 %. This possibly indicates a low level sexual activity going on during a longer period compared to the previous season, with a decline starting in July 2004 / mid-winter.

Comparing both seasons, the signals showed similarities as well as discrepancies: For both seasons investigated, a substantially elevated RFI was observed in mid-summer, where conditions for sexual reproduction were favorable in terms of the percentage of cells within the sexually inducible size range. Also environmental conditions were good enough to enable blooms. But for 2002/03 it is unclear if the high RFI signals a sexual event, due to its transient appearance. In contrast, for

2003/04, the transient maximum initiated a prolonged period of elevated values, which lasted for 8 months, from late December 2003 until August 2004 (mid-summer until late winter). A similar period was observed for season 2002/03, where RFI was elevated for ca. 6 months, from May until October 2003 (late autumn and winter). Throughout both of these time frames, the proportion of cells within the sexually inducible size range was below 50 %.

Size distributions of individual cohorts are defined by the size of initial cells, the size decrease per cell division and the number of cell divisions before entering the next sexual phase. For *F. kerguelensis* ca. 0.28 µm average length decrease per division was estimated by (Fuchs et al. 2013). The shortest auxospores that contained initial thecae, as a proxy for the initial cell, was reported to be 76 µm, and the longest gametangium, as a proxy for the largest cell possibly entering the subsequent sexual phase, 31 µm long (Assmy et al. 2006). From our data, a rough estimation of *F. kerguelensis* growth rate (see Equation 2) is ca. 0.73 divisions $\cdot$ d^{-1} for the period between cups 1.02 and 1.05, and ca. 0.48 divisions $\cdot$ d^{-1} for the period between cups 2.01 and 2.09. This is within the range we observed for culture experiments at 6 °C and 100 µE $\cdot$ s^{-1} $\cdot$ m^{-2} radiation under nutrient-replete conditions (unpublished data). Within this size range between initial and sexually inducible cell size, at least ca. 160 cell divisions can occur according to the MacDonald-Pfitzer rule, which would take not less than 220 days with the maximum growth rate of maximal 0.73 divisions $\cdot$ d^{-1} we estimated. Considering the short productive season of Southern Ocean, we speculate that the vegetative phase of the *F. kerguelensis* life cycle spans several years. Indeed, this idea is consistent with the duration of the life cycle of phylogenetically closely related species, such as *Pseudo-nitzschia multistriata* (Lundholm et al. 2002), that has been reported to be of two years (D'Alelio et al. 2010).

However, from our *F. kerguelensis* data, according size cohorts cannot be observed in a similarly clear manner. Shifts in the size distributions depended mostly on the percentage of the lower size fraction, without substantial augmentation of very large cells. Furthermore, our data do not contain obvious size cohorts, although distributions often show long tails towards large sizes. Which means that the presence of multiple cohorts in these populations is probable, but they are not as clearly distinct as in *P. multistriata*.

The prolonged periods of elevated RFI, which mostly outlast late autumn and winter, and the lack of clearly distinct size cohorts, might indicate that *F. kerguelensis* performs asynchronous sexuality during periods of weak vegetative growth, according to (Edlund and Stoermer 1997). Nevertheless, as noted above, alternative explanations for the selective removal of smaller size classes from the population in autumn are also conceivable (discussed below), and the percentage of cells within the initial cell size range is generally low. Based on this, it cannot be excluded that a low proportion of the population would engage in sexual reproduction throughout the year. Before the present study, the only reported observation of *F. kerguelensis* sexual reproduction in its natural habitat was made by (Assmy et al. 2006) during the European Iron Fertilization Experiment EIFEX (Smetacek et al. 2005), undertaken in the Atlantic sector of the Southern Ocean in the vicinity of the APF (49° S 2° E; see Figure 1) between mid-February and mid-March 2004. (Assmy et al. 2006) documented a significant increase in auxospore concentration of up to 11.4 $\cdot$ 10^6 auxospores $\cdot$ m^{-2} approximately 10 days after iron release, whilst outside the iron-fertilized fertilized patch the auxospore concentration was low (0 to 3.5 $\cdot$ 10^6 auxospores $\cdot$ m^{-2}) and persistent throughout the experiment. In total, auxospores accounted for 0.03 to 0.4 % of the *F. kerguelensis* population. Assuming a delay of 2 to 6 weeks between the production of *F. kerguelensis* cells and their collection by the sediment trap, sinking auxospores would have reached the depth of the sediment trap between late February and early May. The later part of this timeframe is overlapping with the early period of increased RFI

for season 2002/03, and for 2003/04 the complete timeframe is overlapping with the period of increased RFI. That auxosporulation was observed only one single time in nature, contradicts the periods of elevated RFI lasting 5 to 8 months per year in our data. This discrepancy might result from auxospores being overlooked; or the auxospores witnessed by (Assmy et al. 2006) originated from rarer and stronger sexual activity than we observed; or that such a strong signal might have been blurred due to the highly variable sinking times in our investigation.

Based on these observations, we propose that *F. kerguelensis* propagates sexually at very low rates by asynchronous sexuality, possibly throughout the whole year, with some periods of elevated sexual activity, which however appear at different times in different years. This might be indicative of an opportunistic sexual behavior, where the probability of encounter between sexually inducible cells is rather low, and is the main limiting factor of sexual reproduction. This scenario would perhaps not be unrealistic for an immotile species with no flagellated gametes, although we cannot support it with a clear correlation between population density (as reflected by *F. kerguelensis* valve flux) with RFI (PCC = -0.11, N = 43, p = 0.478). Furthermore, with the exception of cup 1.07, which might not represent a sexual event, we do not see increased RFI in summer 2002/03, even though the population was massively dominated by cells within the sexually inducible size range at this time. In contrast, throughout summer 2003/04 RFI was elevated, whilst the population was much more heterogeneous regarding its size fractions, and the percentage of sexually inducible cells generally was lower, compared to the previous year. The share of valves originating from initial cells or their close descendants always stayed at low rates, and only one single observation of auxospores in the species' natural habitat was reported until now (Assmy et al. 2006). Therefore, it is probable that sexual reproduction is not exclusively controlled by population density, but also by environmental factors, especially since we saw an elevated RFI during different periods between both years investigated. However, the lack of in situ measurements of physical and chemical parameters of the water column during the sediment trap experiment make it difficult to establish clear links between environmental drivers and *F. kerguelensis* life cycle. Moreover, during periods where vegetative propagation is low due to unfavorable conditions, the population would benefit from increased sexual activity, mixing up the genes for better adaptation. In contrast, in situations where the population already is adapted well to the environmental conditions, ongoing vegetative proliferation would avoid the need for (costly) sexual recombination.

2.5.5.5 Ecological-physiological effects of cell size

Size distributions in diatom populations are often interpreted as almost mechanical consequences of vegetative (stepwise size decrease) versus sexual (sudden large size increase) reproduction, in the framework we also used above. Nevertheless, cell size also has important physiological and ecological effects which - either by slight physiological regulation of cell size changes or by differential selection across different size classes - might also influence the size composition of diatom populations.

Reduced cell size is beneficial for coping with depleted nutrient conditions, because of a higher surface to volume ratio, allowing for more efficient nutrient uptake. For the High Nutrient – Low Chlorophyll (HNLC) regions as Southern Ocean, not macro-nutrients, but iron is known to be a major factor limiting phytoplankton growth (Martin et al. 1990). For *F. kerguelensis* under prolonged iron limitation, (Timmermans and Van Der Wagt 2010) reported a decrease of 50 % in surface area and volume, and thus an increase in surface-volume ratio of 25 %, whilst (Hoffmann et al. 2007, Wilken et al. 2011) found stronger silicification under iron-limited conditions. These findings indicate that iron limitation would result in the production of small and heavily silicified *F. kerguelensis* frustules. This

physiological response could potentially reduce the risk of being grazed. This applies especially during the productive season, when usually iron availability is low due to consumption by proliferating diatoms, whilst grazing pressure might be high due to proliferating grazers. Furthermore, light absorption efficiency and diatom growth rate increase with decreasing cell size (Geider et al. 1986). Also, (Laney et al. 2012) documented higher division rates on the daughter cells of *Ditylum brightwellii* that inherited the hypotheca, compared to the ones inheriting the epitheca. These differences in division rates of the daughter cells would give an additional boost to shifting a population's size distribution towards smaller cells.

Larger cells, as (Roselli and Basset 2015) stated for the phytoplankton found in marine and transitional zones, *"[…] have nutrient storage capacity in nutrient-rich and fluctuating environments, (Grover 1991, Stolte and Riegman 1996, Litchman et al. 2009), motility and ability to control buoyancy in high physical mixing conditions (Naselli-Flores and Barone 2003), […] anti-predator strategies against grazing pressure (Thingstad 1998)."*. Even though this statement refers to general phytoplankton inter-species competition, it might also give clues regarding intra-species competition, explaining why larger *F. kerguelensis* cells might be fitter to endure the winter conditions. Furthermore, larger cells might be able to sustain enough supplies until next spring, to start the first bloom as soon as conditions become favorable again, without needing to restock first. And they can support this bloom with as many cell divisions as possible before running into their lower size limit, which, for restoring large cell size, would require sexual reproduction. Since this reduces the growth rate (Waite and Harrison 1992), it would impede proliferation, even with environmental conditions being favorable enough to enable the bloom.

Grazing, if size selective, can also impact diatom cell size distributions substantially. For example euphausiids or copepods are known to selectively graze distinct size classes. However, the lack of accompanying observations of the zooplankton communities in the surface layer during our experiment precludes the assessment of the grazing impact on *F. kerguelensis* populations. *Euphausia superba* was reported to show size-dependent filtering efficiency, but its aptitude for grazing especially on small *F. kerguelensis* cells is not clear, since ambiguous ranges for the preferred size range are given by (Assmy 2004): *"mainly […] 20 - 25 µm"*, (Meyer and El-Sayed 1983): *"solitary cells smaller than 20 µm (maximum diameter) being filtered less rapidly than larger diatoms and chain forming species"*, and (Quetin and Ross 1985) investigating clearance rates on phytoplankton from 4 to 40 µm cell size: *"does not feed very effectively on small round cells"*. Nevertheless, grazing might affect sexual reproduction and thus give a possible explanation on the variable timing of elevated RFI we found: Prior to sexual reproduction, (Fuchs et al. 2013) observed breaking of long *F. kerguelensis* chains into individual cells. Since this offers more points of attack, individual cells become more prone to grazing. Asynchronous sexual propagation, or a timing avoiding high grazer densities, might be a strategy to overcome this. In contrast, (Moore et al. 2017) show that the ammonia released by grazing events can trigger sexual reproduction in centric diatoms such as *Thalassiosira pseudonana*.

Considering these effects of cell size shows that changes in *F. kerguelensis* size distributions, as we observed them over the course of the year, might have adaptive relevance, besides being the effects of reproduction: During phase 1, the ongoing shift towards an increased portion of smaller cells is probably beneficial for coping with the progressively depleted nutrient conditions, whilst, as a side-effect of iron depletion, stronger silicification might improve protection from grazing. In phase 2, smaller cells were removed from the population, which might have been due to nutrient depletion or declining light availability, or because of size-selective grazing. The latter one can also explain the

strong sudden drop in the portion of valves below ca. 20 µm length, observed for cups 1.07 – 1.08 (and also cups 1.05 – 1.06 and 2.10 – 2.11). Throughout phase 3, mostly larger cells persisted. We hypothesize that the loss of smallest cells during the transition to winter relates to starving due to size-constrained energy storage capacity, which is supported by the pronounced increase of the minimal valve length from mid-winter to early spring. In contrast, larger cells with larger storage capacities and better active buoyancy regulation might be advantaged in surviving winter, when reduced light availability limits photosynthesis and growth.

2.5.5.6 Possible limitations and benefits of our analysis

Because of the remoteness of the study area, our data is limited to information available via remote-sensing (chlorophyll-*a* concentration) and from sediment trap material (fluxes and morphometric measurements). Sediment trap samples do not necessarily give an unbiased picture of processes in the surface mixed layer. The collected material is influenced by numerous and varying factors like mortality (intrinsic and extrinsic, that is by grazing, parasites (Kühn et al. 2004) or pathogens), sinking speed, aggregation, mixed layer depth, varying sampling resolution and many more. Material collected by a sediment trap is unavoidably integrated over time, and the possibly large variability in sinking speed between cells / valves sinking out individually versus as part of larger aggregates might confound temporal signals to an unknown degree. Furthermore, water masses passing the trap were driven by the ACC, which goes on continuously at a speed of ca. 1 to 2 km · h^{-1} (Hofmann 1985) and meanders (Moore et al. 1999) over the fixed position of the trap. Thus, samples also were integrated over an area, the so-called statistical funnel (Siegel and Deuser 1997, Siegel et al. 2008), and most certainly over different population patches. (Smetacek et al. 2002) observed a distinct patchiness of diatom populations along the Antarctic Polar Front on the mesoscale, within the magnitude of some tens of kilometers. These patches can contain different populations, which might exhibit different life cycle stages and propagation behavior. Accordingly, signals most probably are blurred, possibly biased and definitely postponed, compared to the surface situation. Nevertheless, until continuous long-term monitoring of the upper water column phytoplankton becomes more wide spread in the Southern Ocean (two recent efforts have been successful using moored water samplers; (Eriksen et al. in press; Blain et al., in preparation), sediment traps represent one of the few available tools that allow us for investigating *F. kerguelensis* life cycle in its natural habitat.

By applying our workflow for large scale permanent slide imaging and image analysis, we were able to investigate precise morphometric information of nearly 29,000 *F. kerguelensis* valves, covering a period of 2 years by 41 samples. This enabled us to assess valve length distributions instead of average values only, from which we hoped better insights into life cycle processes. From our data, we see cell size changes which do not only reflect the species life cycle, but might also be beneficial in respect to growth and survival. Even though size cohorts depicting sexual propagation were not as clearly visible as for *Pseudo-nitzschia multistriata*, we found indication of low sexual activity lasting for prolonged periods throughout varying times of the year, which can be interpreted according to (Edlund and Stoermer 1997) as asynchronous sexuality during periods of weak vegetative growth, with the vegetative phases lasting at least several years.

To tease apart the relative role of population density versus environmental factors in controlling the frequency of sex and changes in valve size in this species, more observations will be necessary, especially regarding the surface environmental conditions. Here we lack data on nutrient availability and plankton composition, which would be invaluable for interpreting our results. Further research will also be necessary to address possible biases affecting sediment trap material during its way from the surface mixed layer. Besides direct observation of events, also the influence of immigrating

populations could be assessed by DNA analyses, and a quantification of auxosporulation could be conducted by RNA analyses comparing the activity of meiosis- versus housekeeping genes, although these would also rely on long-term sampling which is currently hardly possible apart from sediment traps.

2.5.6 Conclusions

Our data suggests how an *F. kerguelensis* life cycle could be envisioned as follows: During the productive phase in late spring and summer, the population proliferates by vegetative reproduction. The size reduction introduced by this is possibly advantageous in coping with the progressive shortage of nutrients, with the additional advantage of stronger silicification caused by iron-depletion, resulting in cells being less prone to grazing. Subsequently, a major share of smaller cells is removed from the population due to grazing or declining environmental conditions. During the light-limited period of late autumn and winter, mostly larger cell persist, presumably due to their better buoyancy control and increased storage capacity, which might also be beneficial for starting a bloom in the following spring. Low sexual activity lasts for prolonged periods at different times of the year, which can be interpreted according to (Edlund and Stoermer 1997) as asynchronous sexuality during periods of weak vegetative growth, which might be beneficial for avoiding predation or for generating novel genetic variants possibly better adapted to unfavorable environmental conditions.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the priority programme 1158 “Antarctic Research with comparative investigations in Arctic ice areas” by grant nr. BE4316/4-1, KA1655/3-1, and by an outgoing scholarship, as well as travel expenses, granted by the Helmholtz Graduate School for Polar and Marine Research (POLMAR). Part of this work was made possible by the Australian Government’s Australian Antarctic Science Grant Program (project number 4078) and Macquarie University (A. Rigual-Hernández and L. Armand). The sediment trap samples were collected by the ACE CRC with logistic support from the Australian Antarctic Division (under grants AAS1156 and AAS2256 to T.W. Trull).

We would like to express our gratitude to Fenina Butler for scanning most slides from season 2002/03.

Author contributions

Andres S. Rigual-Hernández initiated the project idea, prepared the microscopy slides, provided data and background information on the sediment trap and wrote the paragraphs about the oceanographic setting. Leanne Armand, Bánk Beszteri and Gerhard Kauer supervised the project. Thomas W. Trull provided the sediment trap material. All of the aforementioned persons contributed to the paper’s content. Slide scanning for season 2003/04 and partly 2002/03, image analysis and measurements via SHERPA, data analysis and paper writing were performed by Michael Kloster.

2.5.7 Supplements

2.5.7.1 Data

The raw data of morphometric measurements are provided in supplement “R” on the annexed data medium.

2.5.7.2 SHERPA settings

The configuration files and shape templates are provided in supplement “SHERPA” on the annexed data medium. Settings for the low-resolution scan can be found in “54S-800m.20x.dmini”, for the high-resolution scan in “54S-800m.63x.dmini”

2.5.7.3 Modified SHERPA procedure

We adapted our established analysis workflow to compensate for the low imaging contrast of the mounted valves: Increased manual intervention was exerted for manual post-editing, as well as reviewing segmentation results by also addressing results of lower quality than described in (Kloster et al. 2017). "Quality" in this context relates to the question if a) an object is of interest, and b) the image processing and segmentation worked out well. SHERPA supports the user in tackling this problem by calculating the "ranking index" (RI), which increases with declining result quality.

Accordingly, for each slide, when selecting valves of interest in the low-resolution (20× magnification) scans, results from RIs of 0 to 6, and for sparse samples up to RI 10, were reviewed. If ca. 100 consecutive results of one particular RI did not represent suitable *F. kerguelensis* valves, the remaining results of this RI were skipped, because the probability that they contained further useful *F. kerguelensis* valve images was very low, and the procedure was resumed with results of the following RI. From the high-resolution scans (63× magnification), for each slide results were reviewed, reworked as needed, and *F. kerguelensis* valves selected for RIs 0 to at least 2. If the sample size was not sufficient (mostly below 300), results of higher RIs were added. The same applied if results of the highest RI changed the valve length distribution significantly, which was checked by a Kolmogorov-Smirnov test (see supplement "KS-test results regarding RI"). In some cases, duplicate slides were scanned, to obtain higher sampling sizes.

Costae distance measurement is an experimental SHERPA feature and, until now, was tested on *Fragilariopsis* valves only (Beszteri et al. submitted). This frequency based method is mostly robust towards segmentation flaws, nevertheless results of costae distance ≥ 5 µm were assumed to be incorrect and thus rejected, which applied to ca. 0.5 % of all data. The valve image was segmented by Adaptive Thresholding (Bradski and Kaehler 2008), resulting in costae and valve border colored white, whilst striae and image background were colored black. The segmented image was sampled along the central 80 % of the valve's major (apical) axis by the Bresenham algorithm (Bresenham 1965). Along this sampled line, each white section was replaced by a 5 pixel wide binomial-smoothed peak, resulting in an artificial 2d-image of costae positions. This was transformed into the frequency domain by discrete Fourier transform (DFT). Average costae distance was calculated from the frequency of the maximum magnitude of the Fourier transformed.

2.5.7.4 Valves within the initial cell size range

Table S1: Abundance, percentage and relative frequency of valves within the initial cell size range.

Cup	Sample size [*1]	n of valves within the initial cell size range (≥ 75.5 µm valve length) [*1]	% of valves within the initial cell size range (≥ 75.5 µm valve length) [*1]	Rel. frequency of valves within the initial cell size range (RFI) [*2]
1.01	300	2	0.67	0.06
1.02	569	1	0.18	0.03
1.03	525	1	0.19	0.04
1.04	178	0	0.00	0.00

1.05	455	0	0.00	0.07
1.06	1630	1	0.06	0.02
1.07	574	1	0.17	0.14
1.08	736	1	0.14	0.02
1.09	1010	3	0.30	0.04
1.10	907	1	0.11	0.03
1.11	338	0	0.00	0.00
1.11	338	0	0.00	0.00
1.12	409	0	0.00	0.00
1.13	478	2	0.42	0.02
1.13	478	2	0.42	0.00
1.14	420	0	0.00	0.03
1.15	379	2	0.53	0.05
1.16	145	0	0.00	0.00
1.17	793	5	0.63	0.12
1.18	592	3	0.51	0.18
1.19	955	4	0.42	0.15
1.20	811	4	0.49	0.28
2.01	160	2	1.25	0.14
2.02	327	5	1.53	0.18
2.03	430	5	1.16	0.04
2.04	600	0	0.00	0.01
2.05	467	4	0.86	0.09
2.06	718	0	0.00	0.03
2.07	1639	5	0.31	0.40
2.08	1360	9	0.66	0.19
2.09	1228	4	0.33	0.18
2.10	1345	3	0.22	0.16
2.11	553	7	1.27	0.11
2.12	689	7	1.02	0.14
2.13	1377	7	0.51	0.10
2.14	1525	8	0.52	0.06
2.15	673	9	1.34	0.23
2.16	473	6	1.27	0.10
2.17	840	6	0.71	0.19
2.18	998	7	0.70	0.13
2.19	559	13	2.33	0.09
2.20	460	2	0.43	0.05
2.21	210	3	1.43	0.04

*[1]) These values originate from the count performed during the morphometric analysis.

*[2]) These values originate from the additional analysis for finding cells within the initial cell size range only.

RFI is calculated according to Equation 1.

2.5.7.5 R scripts

The R scripts used for analyzing the data are provided in supplement “R” of the annexed data medium.

2.5.7.6 *Additional measurements*

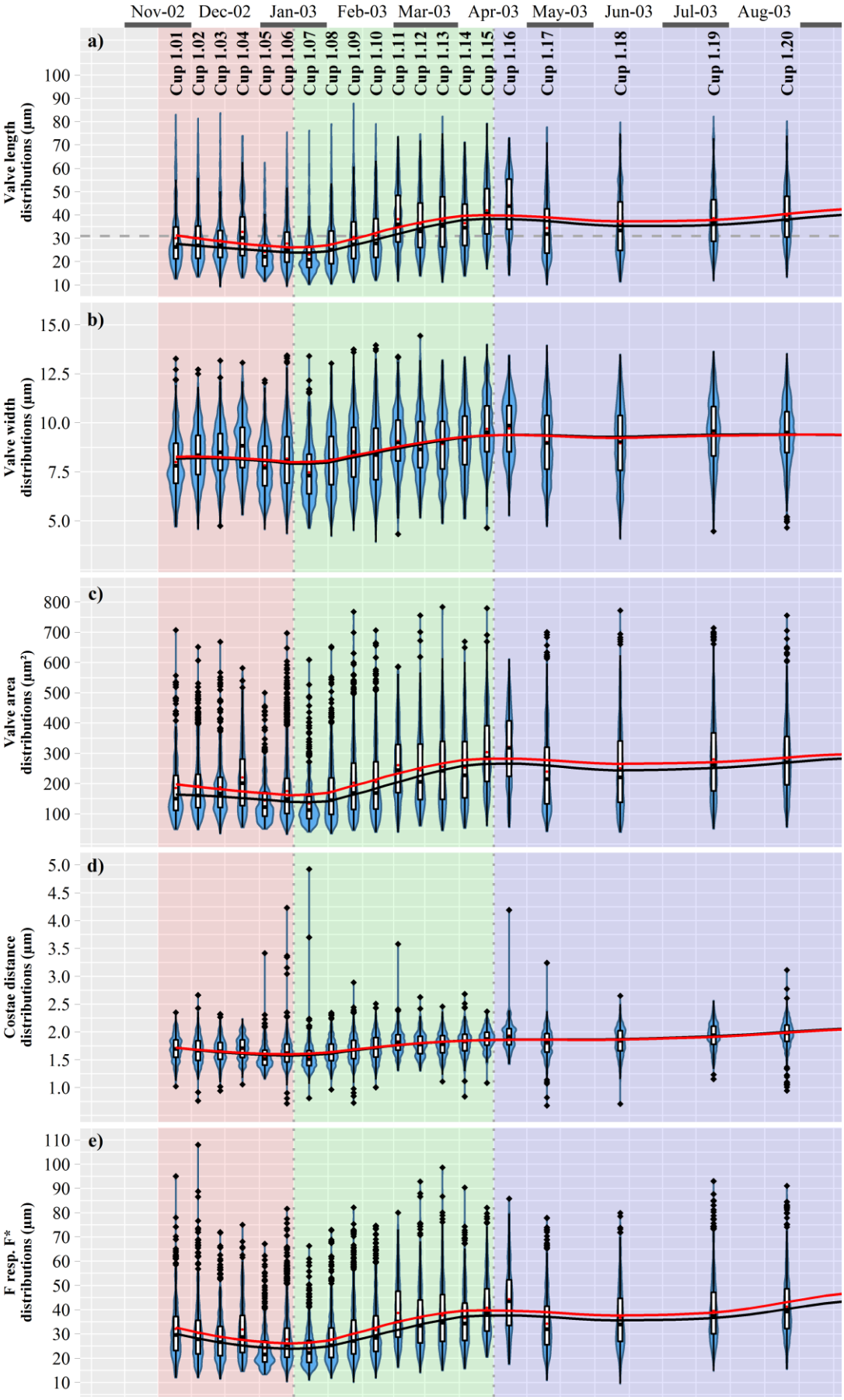


Figure S1: Season 2002/03, distribution of a) valve length (cmp. Fig 2 & 3 c), b) valve width (transapical), c) valve area, d) costae distance, e) F (Fenner et al. 1976) or F* (Cortese and Gersonde 2007). To compensate for outliers, results for d) and e) with costae distances ≥ 5 µm were excluded.

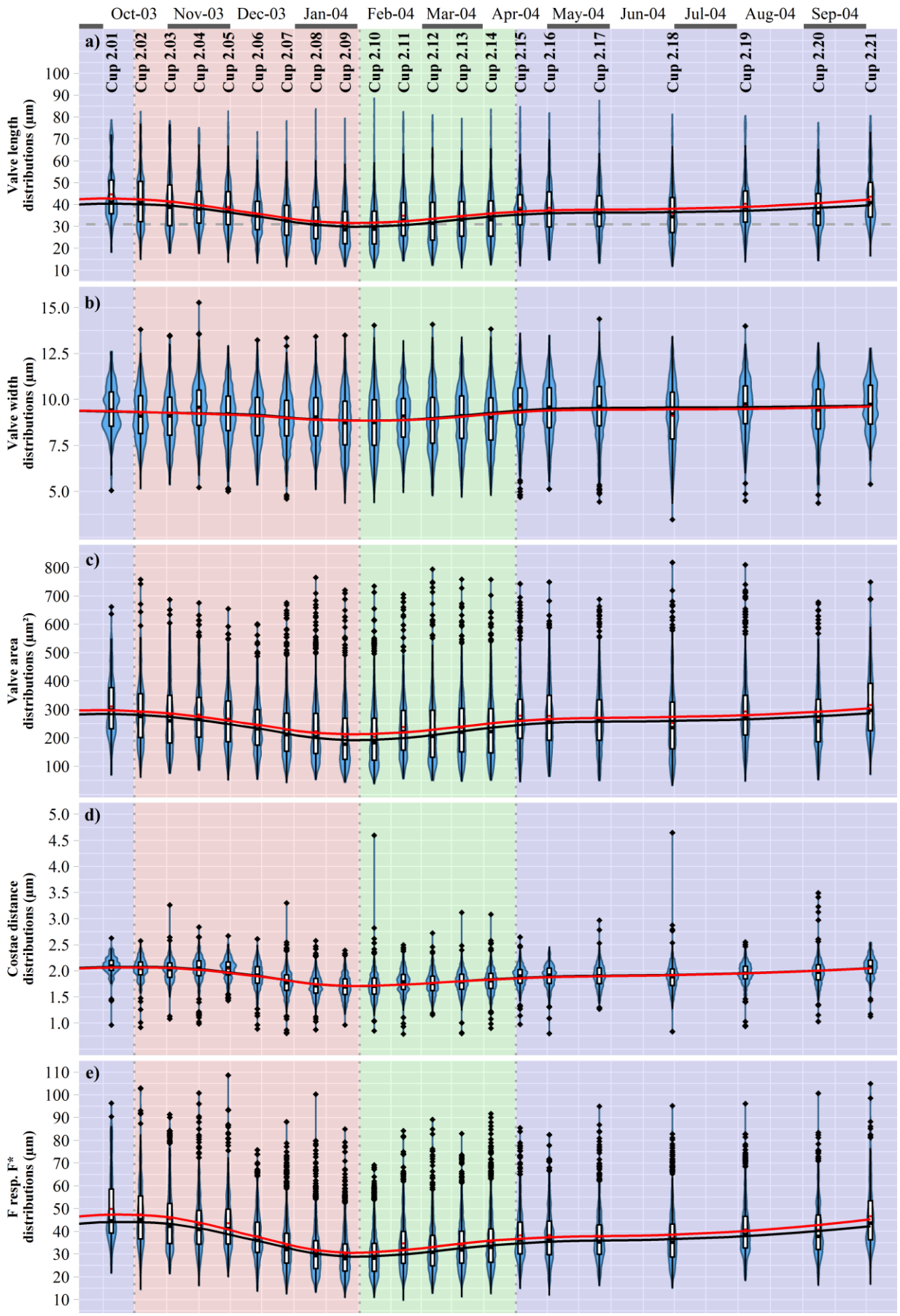


Figure S2: Season 2003/04, distribution of a) valve length (cmp. Fig 2 & 3 c), b) valve width (transapical), c) valve area, d) costae distance, e) F (Fenner et al. 1976) or F* (Cortese and Gersonde 2007). To compensate for outliers, results for d) and e) with costae distances ≥ 5 µm were excluded.

2.5.7.7 Chlorophyll-a data

Table S2: Chlorophyll-*a* concentration derived from NASA's Giovanni online data system. NA = not available.

Date	**Chlorophyll-*a* concentration ($mg \cdot m^{-3}$)**
2002-08-16	0.138
2002-09-16	0.142
2002-10-16	0.153
2002-11-16	0.18
2002-12-16	0.182
2003-01-16	0.164
2003-02-15	0.255
2003-03-16	0.149
2003-04-16	0.119
2003-05-16	0.144
2003-06-16	NA
2003-07-16	NA
2003-08-16	0.125
2003-09-16	0.141
2003-10-16	0.167
2003-11-16	0.223
2003-12-16	0.179
2004-01-16	0.175
2004-02-15	0.114
2004-03-16	0.118
2004-04-16	0.118
2004-05-16	0.101
2004-06-16	NA
2004-07-16	NA
2004-08-16	0.111
2004-09-16	0.145
2004-10-16	0.162

2.5.7.8 Summary statistics of F. kerguelensis valve size distributions.

Table S3: Statistics of valve length measurements (compare Figure 2 & 3).

Cup	Middle day of sampling period	Length of sampling period (d)	Sample size	Valve length					% within sexually inducible size range [*1]	Rel. frequency of valves within the initial cell size range (RFI) [*2]
				Mean (µm)	Median (µm)	Skewness (µm)	Min (µm)	Max (µm)		
1.01	2002-11-23	10	300	30.0	26.3	1.4	13.0	70.7	66.7	0.06
1.02	2002-12-03	10	569	30.2	27.0	1.3	14.0	69.3	66.1	0.03
1.03	2002-12-13	10	525	29.1	27.0	1.4	13.1	63.0	68.0	0.04
1.04	2002-12-23	10	178	32.8	30.2	0.9	14.4	70.0	53.9	0.00
1.05	2003-01-02	10	455	24.1	22.1	1.6	12.7	54.6	85.1	0.07
1.06	2003-01-12	10	1630	27.7	25.3	1.2	11.9	61.6	70.9	0.02
1.07	2003-01-22	10	574	23.1	20.8	1.8	11.4	53.9	86.6	0.14
1.08	2003-02-01	10	736	27.5	25.1	1.2	12.0	62.8	69.4	0.02
1.09	2003-02-11	10	1010	30.5	27.4	1.0	12.8	65.6	60.6	0.04
1.10	2003-02-21	10	907	31.1	27.8	1.0	13.6	65.8	59.6	0.03
1.11	2003-03-03	10	338	38.1	35.9	0.4	14.7	67.8	33.4	0.00
1.12	2003-03-13	10	409	36.1	33.0	0.7	15.6	70.7	44.5	0.00
1.13	2003-03-23	10	478	37.5	35.3	0.5	14.7	70.7	39.5	0.02
1.14	2003-04-02	10	420	36.4	34.5	0.6	15.0	68.1	40.5	0.03
1.15	2003-04-12	10	379	41.9	40.6	0.4	20.3	74.0	23.0	0.05
1.16	2003-04-22	10	145	44.0	43.9	-0.1	17.5	70.0	17.9	0.00
1.17	2003-05-09	25	793	34.3	31.8	0.7	13.8	71.5	47.7	0.12
1.18	2003-06-11	40	592	35.9	33.3	0.6	13.1	72.2	44.8	0.18
1.19	2003-07-23	45	955	38.6	36.8	0.6	16.6	73.8	33.0	0.15
1.20	2003-08-25	20	811	39.7	37.8	0.6	17.1	74.0	26.4	0.28
2.01	2003-10-04	14	160	44.7	41.1	0.8	21.9	75.3	7.5	0.14
2.02	2003-10-18	14	327	42.4	40.9	0.6	20.1	76.0	21.4	0.18
2.03	2003-11-01	14	430	40.8	38.7	0.7	18.7	75.4	29.3	0.04
2.04	2003-11-15	14	600	39.6	38.0	0.8	20.5	72.5	23.0	0.01
2.05	2003-11-29	14	467	38.9	37.6	0.8	19.3	74.2	25.7	0.09
2.06	2003-12-13	14	718	35.6	34.7	0.6	16.6	62.3	34.0	0.03
2.07	2003-12-27	14	1639	33.6	32.5	0.8	15.8	66.2	44.9	0.40
2.08	2004-01-10	14	1360	32.5	30.8	1.1	15.5	69.6	50.9	0.19
2.09	2004-01-24	14	1228	30.5	28.5	1.0	13.6	66.5	58.7	0.18
2.10	2004-02-07	14	1345	30.3	28.8	0.9	13.3	62.5	57.3	0.16
2.11	2004-02-21	14	553	34.9	32.9	1.1	15.7	76.5	43.0	0.11
2.12	2004-03-06	14	689	33.3	31.5	1.0	13.7	74.9	48.2	0.14
2.13	2004-03-20	14	1377	34.1	33.0	0.8	15.1	69.3	44.0	0.10

2.14	2004-04-03	14	1525	34.7	33.1	0.8	14.7	72.1	42.8	0.06
2.15	2004-04-17	14	673	38.4	37.4	0.9	16.1	76.5	26.2	0.23
2.16	2004-05-01	14	473	38.5	36.9	0.7	15.7	76.3	29.6	0.10
2.17	2004-05-25	35	840	37.4	36.0	0.7	14.7	73.7	28.8	0.19
2.18	2004-06-29	35	998	36.0	34.4	0.8	13.8	74.1	37.0	0.13
2.19	2004-08-03	35	559	40.3	38.5	1.0	18.7	77.2	20.8	0.09
2.20	2004-09-07	35	460	38.4	36.3	0.8	16.5	73.6	27.8	0.05
2.21	2004-10-02	14	210	43.5	40.9	0.8	23.0	76.7	14.8	0.04

*1) refers to valves < 31.5 µm apical length

*2) see Equation 1

For cup 1.04, sampling size was low (n=178), even though this sample was taken within the productive period in late December 2002. Compared to the adjacent samples, valve density was unexpectedly sparse on the according slide. Also for cups 1.16, 2.01 and 2.21 sampling size was low (< 300), but here the slide density approximately accorded with the adjacent samples. We nevertheless decided to keep these measurements, since for the discussion of our results we mostly relate to smoothed general trends, which are not significantly changed by these samples.

2.5.7.9 Flux data

Table S4: Flux data from (Rigual-Hernández et al. 2015).

Cup	Middle day of sampling period	Length of sampling period (d)	Total Mass Flux ($mg \cdot m^{-2} \cdot d^{-1}$)	$BSiO_2$ ($mg \cdot m^{-2} \cdot d^{-1}$)	$BSiO_2$ (%)	$CaCO_3$ ($mg \cdot m^{-2} \cdot d^{-1}$)	$CaCO_3$ (%)	POC ($mg \cdot m^{-2} \cdot d^{-1}$)	POC (%)	*F. kerguelsnsis* flux (10^6 valves $\cdot m^{-2} \cdot d^{-1}$)	*F. kerguelensis* percentage of all diatoms
1.01	2002-11-23	10.0	96.7	46.8	48	18.3	19	3.6	4	3.2	56.0
1.02	2002-12-03	10.0	190.6	110.6	58	37.5	20	4.6	2	12.1	43.6
1.03	2002-12-13	10.0	363.9	203.6	56	81.0	22	12.3	3	25.1	38.4
1.04	2002-12-23	10.0	240.7	134.0	56	47.2	20	11.4	5	9.5	26.3
1.05	2003-01-02	10.0	125.6	54.4	43	45.4	36	4.8	4	3.0	44.3
1.06	2003-01-12	10.0	158.9	93.7	59	35.0	22	5.4	3	13.2	71.7
1.07	2003-01-22	10.0	360.2	219.1	61	68.9	19	8.6	2	21.1	73.3
1.08	2003-02-01	10.0	213.8	113.7	53	48.3	23	5.5	3	31.4	69.7
1.09	2003-02-11	10.0	396.5	253.8	64	56.0	14	5.1	1	20.9	74.3
1.10	2003-02-21	10.0	60.8	26.6	44	23.7	39	2.2	4	6.9	40.5
1.11	2003-03-03	10.0	58.5	16.8	29	23.8	41	5.5	9	1.3	67.3
1.12	2003-03-13	10.0	89.3	26.6	30	37.0	41	7.0	8	0.9	63.1
1.13	2003-03-23	10.0	49.5	13.6	28	25.1	51	3.4	7	1.5	64.0
1.14	2003-04-02	10.0	33.5	7.6	23	18.2	54	2.7	8	0.8	78.0
1.15	2003-04-12	10.0	31.9	6.2	19	20.1	63	1.6	5	1.0	84.5
1.16	2003-04-22	10.0	21.6	5.8	27	14.0	65	1.0	4	0.6	78.7
1.17	2003-05-09	25.0	23.8	7.3	31	12.8	54	0.9	4	1.7	85.8
1.18	2003-06-11	40.0	23.7	10.8	46	7.8	33	1.0	4	1.4	86.7

1.19	2003-07-23	45.0	16.9	8.0	47	5.5	33	0.7	4	1.0	83.7
1.20	2003-08-25	20.0	33.6	13.5	40	11.9	35	1.8	5	2.1	81.0
2.01	2003-10-04	14.0	15.1	8.1	54	4.5	30	0.5	3	0.3	76.2
2.02	2003-10-18	14.0	23.4	12.6	54	7.0	30	0.8	3	0.8	73.9
2.03	2003-11-01	14.0	34.8	18.6	54	10.8	31	1.1	3	2.2	77.0
2.04	2003-11-15	14.0	121.0	54.6	45	19.9	16	2.0	2	9.9	85.8
2.05	2003-11-29	14.0	75.5	35.9	48	14.6	19	1.6	2	6.3	80.8
2.06	2003-12-13	14.0	178.6	111.6	62	40.2	23	3.6	2	10.9	55.0
2.07	2003-12-27	14.0	232.6	134.8	58	43.6	19	7.2	3	5.0	48.7
2.08	2004-01-10	14.0	182.0	114.5	63	33.7	19	6.2	3	20.2	65.2
2.09	2004-01-24	14.0	121.9	81.7	67	21.0	17	3.3	3	11.7	69.6
2.10	2004-02-07	14.0	63.9	37.0	58	10.8	17	2.7	4	7.3	65.7
2.11	2004-02-21	14.0	28.5	16.6	58	7.0	24	1.0	4	4.4	74.1
2.12	2004-03-06	14.0	31.0	18.4	59	7.7	25	0.9	3	4.6	73.7
2.13	2004-03-20	14.0	66.2	45.3	68	10.2	15	1.0	2	7.4	75.2
2.14	2004-04-03	14.0	84.6	52.6	62	16.9	20	0.8	1	8.9	76.9
2.15	2004-04-17	14.0	15.6	10.1	65	4.9	31	0.8	5	0.5	80.8
2.16	2004-05-01	14.0	10.6	6.9	65	4.0	38	0.5	4	0.6	87.2
2.17	2004-05-25	35.0	19.6	13.7	70	3.5	18	0.4	2	2.5	73.0
2.18	2004-06-29	35.0	19.3	11.3	59	4.5	23	1.1	6	1.2	75.8
2.19	2004-08-03	35.0	14.5	6.0	41	2.8	20	2.0	14	1.0	80.3
2.20	2004-09-07	35.0	10.3	5.0	49	3.8	37	0.5	5	1.0	90.3
2.21	2004-10-02	14.0	7.3	3.6	49	2.5	34	0.5	7	0.4	87.6

The flux data also is available at:
https://data.aad.gov.au/metadata/records/AAS_4078_diatoms_biogenic_flux_subantarctic/

2.5.7.10 KS-test results regarding RI

The results of Kolmogorov-Smirnov tests, assessing the influence of adding results from subsequent ranking indices, can be found on the annexed data medium in the file “Supplement KS-test results regarding RI.pdf”.

2.5.8 References

Abelmann, A., Gersonde, R., Cortese, G., Kuhn, G. and Smetacek, V. (2006). "Extensive phytoplankton blooms in the Atlantic sector of the glacial Southern Ocean." Paleoceanography **21**(1).

Assmy, P. (2004). Temporal development and vertical distribution of major components of the plankton assemblage during an iron fertilization experiment in the Antarctic Polar Frontal Zone, Universität Bremen.

Assmy, P., Henjes, J., Smetacek, V. and Montresor, M. (2006). "Auxospore formation by the silica-sinking, oceanic diatom *Fragilariopsis kerguelensis* (Bacillariophyceae)." Journal of Phycology **42**(5): 1002-1006.

Assmy, P., Smetacek, V., Montresor, M., Klaas, C., Henjes, J., Strass, V. H., Arrieta, J. M., Bathmann, U., Berg, G. M., Breitbarth, E., Cisewski, B., Friedrichs, L., Fuchs, N., Herndl, G. J., Jansen, S., Krägefsky, S., Latasa, M., Peeken, I., Röttgers, R., Scharek, R., Schüller, S. E., Steigenberger, S., Webb, A. and Wolf-Gladrow, D. (2013). "Thick-shelled, grazer-protected diatoms decouple ocean carbon and silicon cycles in the iron-limited Antarctic Circumpolar Current." Proceedings of the National Academy of Sciences **110**(51): 20633-20638.

Beszteri, B., Allen, C., Almandoz, G. O., Armand, L., Barcena, M. Á., Cantzler, H., Crosta, X., Esper, O., Jordan, R. W., Kauer, G., Klaas, C., Kloster, M., Leventer, A., Pike, J. and Rigual-Hernández, A. S. (submitted). "Quantitative comparison of taxa and taxon concepts in the diatom genus *Fragilariopsis*." Journal of Phycology.

Boyd, P. W., Crossley, A. C., DiTullio, G. R., Griffiths, F. B., Hutchins, D. A., Queguiner, B., Sedwick, P. N. and Trull, T. W. (2001). "Control of phytoplankton growth by iron supply and irradiance in the subantarctic Southern Ocean: Experimental results from the SAZ Project." Journal of Geophysical Research: Oceans **106**(C12): 31573-31583.

Bradski, G. and Kaehler, A. (2008). Learning OpenCV : Computer Vision with the OpenCV Library. Sebastopol, O'Reilly.

Bresenham, J. E. (1965). "Algorithm for computer control of a digital plotter." IBM Systems journal **4**(1): 25-30.

Chepurnov, V. A., Mann, D. G., Sabbe, K., Vannerum, K., Casteleyn, G., Verleyen, E., Peperzak, L. and Vyverman, W. (2005). "Sexual reproduction, mating system, chloroplast dynamics and abrupt cell size reduction in *Pseudo-nitzschia pungens* from the North Sea (Bacillariophyta)." European Journal of Phycology **40**(4): 379-395.

Chepurnov, V. A., Mann, D. G., Sabbe, K. and Vyverman, W. (2004). "Experimental studies on sexual reproduction in diatoms." International review of cytology **237**: 91-154.

Closset, I., Cardinal, D., Bray, S. G., Thil, F., Djouraev, I., Rigual-Hernández, A. S. and Trull, T. W. (2015). "Seasonal variations, origin, and fate of settling diatoms in the Southern Ocean tracked by silicon isotope records in deep sediment traps." Global Biogeochemical Cycles **29**(9): 1495-1510.

Cortese, G. and Gersonde, R. (2007). "Morphometric variability in the diatom *Fragilariopsis kerguelensis*: Implications for Southern Ocean paleoceanography." Earth and Planetary Science Letters **257**(3-4): 526-544.

Cortese, G., Gersonde, R., Maschner, K. and Medley, P. (2012). "Glacial-interglacial size variability in the diatom *Fragilariopsis kerguelensis*: Possible iron/dust controls?" Paleoceanography **27**.

Crawford, R. M., Hinz, F. and Rynearson, T. (1997). "Spatial and temporal distribution of assemblages of the diatom *Corethron criophilum* in the Polar Frontal region of the South Atlantic." Deep Sea Research Part II: Topical Studies in Oceanography **44**(1): 479-496.

Crosta, X., Romero, O., Armand, L. K. and Pichon, J.-J. (2005). "The biogeography of major diatom taxa in Southern Ocean sediments: 2. Open ocean related species." Palaeogeography, Palaeoclimatology, Palaeoecology **223**(1–2): 66-92.

D'Alelio, D., d'Alcala, M. R., Dubroca, L., Sarn, D., Zingone, A. and Montresor, M. (2010). "The time for sex: A biennial life cycle in a marine planktonic diatom." Limnology and Oceanography **55**(1): 106-114.

Davidovich, N. (2001). Species specific sizes and size range of sexual reproduction in diatoms. Proceedings of the 16th international diatom symposium, University of Athens, Greece.

de Salas, M. F., Eriksen, R., Davidson, A. T. and Wright, S. W. (2011). "Protistan communities in the Australian sector of the Sub-Antarctic Zone during SAZ-Sense." Deep Sea Research Part II: Topical Studies in Oceanography **58**(21–22): 2135-2149.

Defelice, D. R. and Wise, S. W. (1981). "Surface lithofacies, biofacies, and diatom diversity patterns as models for delineation of climatic change in the southeast Atlantic Ocean." Marine Micropaleontology **6**(1): 29-70.

Drebes, G. (1977). Sexuality [The biology of diatoms: chapter 9]. Oxford, Blackwell. **Botanical Monographs; vol. 13:** 498.

Ebersbach, F., Trull, T. W., Davies, D. M. and Bray, S. G. (2011). "Controls on mesopelagic particle fluxes in the Sub-Antarctic and Polar Frontal Zones in the Southern Ocean south of Australia in summer—Perspectives from free-drifting sediment traps." Deep Sea Research Part II: Topical Studies in Oceanography **58**(21): 2260-2276.

Edlund, M. B. and Stoermer, E. F. (1997). "Ecological, evolutionary, and systematic significance of diatom life histories." Journal of Phycology **33**(6): 897-918.

Eriksen, R., Trull, T., Davies, D., Jansen, P., Davidson, A., Westwood, K. and Van den Enden, R. (in press). "Seasonal succession of phytoplankton community structure from autonomous sampling at the Australian Southern Ocean Time Series site." Marine Ecology Progress Series.

Fenner, J., Schrader, H. and Wienigk, H. (1976). "Diatom phytoplankton studies in the southern Pacific Ocean, composition and correlation to the Antarctic Convergence and its paleoecological significance." Initial Reports of the Deep Sea Drilling Project **35**: 757-813.

Fuchs, N., Scalco, E., Kooistra, W. H. C. F., Assmy, P. and Montresor, M. (2013). "Genetic characterization and life cycle of the diatom *Fragilariopsis kerguelensis*." European Journal of Phycology **48**(4): 411-426.

Geider, R. J., Platt, T. and Raven, J. A. (1986). "Size dependence of growth and photosynthesis in diatoms: a synthesis." Marine Ecology Progress Series: 93-104.

Gersonde, R., Abelmann, A., Brathauer, U., Becquey, S., Bianchi, C., Cortese, G., Grobe, H., Kuhn, G., Niebler, H. S., Segl, M., Sieger, R., Zielinski, U. and Futterer, D. K. (2003). "Last glacial sea surface temperatures and sea-ice extent in the Southern Ocean (Atlantic-Indian sector): A multiproxy approach." Paleoceanography **18**(3).

Gersonde, R., Crosta, X., Abelmann, A. and Armand, L. (2005). "Sea-surface temperature and sea ice distribution of the Southern Ocean at the EPILOG Last Glacial Maximum—a circum-Antarctic view based on siliceous microfossil records." Quaternary Science Reviews **24**(7-9): 869-896.

Grover, J. P. (1991). "Resource competition in a variable environment: phytoplankton growing according to the variable-internal-stores model." The American Naturalist **138**(4): 811-835.

Hamm, C. E., Merkel, R., Springer, O., Jurkojc, P., Maier, C., Prechtel, K. and Smetacek, V. (2003). "Architecture and material properties of diatom shells provide effective mechanical protection." Nature **421**(6925): 841-843.

Hart, T. J. (1942). "Phytoplankton periodicity in Antarctic surface waters." Discovery reports **8**: 261-356.

Hasle, G. R. (1965). "*Nitzschia* and *Fragilariopsis* species studied in the light and electron microscopes. III. The genus *Fragilariopsis*." Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo, I. Matematisk-Naturvidenskapelig Klasse. Ny Serie **21**: 1-49.

Hoffmann, L., Peeken, I. and Lochte, K. (2007). "Effects of iron on the elemental stoichiometry during EIFEX and in the diatoms *Fragilariopsis kerguelensis* and *Chaetoceros dichaeta*." Biogeosciences **4**(4): 569-579.

Hofmann, E. E. (1985). "The large-scale horizontal structure of the Antarctic Circumpolar Current from FGGE drifters." Journal of Geophysical Research: Oceans **90**(C4): 7087-7097.

Jansen, S. and Bathmann, U. (2007). "Algae viability within copepod faecal pellets: evidence from microscopic examinations." Marine Ecology Progress Series **337**: 145-153.

Kloster, M., Esper, O., Kauer, G. and Beszteri, B. (2017). "Large-Scale Permanent Slide Imaging and Image Analysis for Diatom Morphometrics." Applied Sciences **7**(4): 330.

Kloster, M., Kauer, G. and Beszteri, B. (2014). "SHERPA: an image segmentation and outline feature extraction tool for diatoms and other objects." BMC bioinformatics **15**: 218.

Kopczynska, E. E., Dehairs, F., Elskens, M. and Wright, S. (2001). "Phytoplankton and microzooplankton variability between the Subtropical and Polar Fronts south of Australia: Thriving under regenerative and new production in late summer." Journal of Geophysical Research: Oceans **106**(C12): 31597-31609.

Kozlova, O. g. G. (1966). Diatoms of the Indian and Pacific Sectors of the Antarctic, Israel Program for Scientific Translations;[available from the US Clearinghouse for Federal Scientific and Technical Information, Springfield, Va.].

Kühn, S., Medlin, L. and Eller, G. (2004). "Phylogenetic position of the parasitoid nanoflagellate *Pirsonia* inferred from nuclear-encoded small subunit ribosomal DNA and a description of *Pseudopirsonia* n. gen. and *Pseudopirsonia mucosa* (Drebes) comb. nov." Protist **155**(2): 143-156.

Laney, S. R., Olson, R. J. and Sosik, H. M. (2012). "Diatoms favor their younger daughters." Limnology and Oceanography **57**(5): 1572-1578.

Lannuzel, D., Bowie, A. R., Remenyi, T., Lam, P., Townsend, A., Ibisanmi, E., Butler, E., Wagener, T. and Schoemann, V. (2011). "Distributions of dissolved and particulate iron in the sub-Antarctic and Polar Frontal Southern Ocean (Australian sector)." Deep Sea Research Part II: Topical Studies in Oceanography **58**(21–22): 2094-2112.

Litchman, E., Klausmeier, C. and Yoshiyama, K. (2009). "Contrasting size evolution in marine and freshwater diatoms." Proceedings of the National Academy of Sciences **106**(8): 2665-2670.

Lundholm, N., Daugbjerg, N. and Moestrup, Ø. (2002). "Phylogeny of the Bacillariaceae with emphasis on the genus *Pseudo-nitzschia* (Bacillariophyceae) based on partial LSU rDNA." European Journal of Phycology **37**(1): 115-134.

Macdonald, J. D. (1869). "On the structure of the Diatomaceous frustule, and its genetic cycle." Journal of Natural History **3**(13): 1-8.

Martin, J. H., Fitzwater, S. E. and Gordon, R. M. (1990). "Iron deficiency limits phytoplankton growth in Antarctic waters." Global Biogeochemical Cycles **4**(1): 5-12.

Meyer, M. A. and El-Sayed, S. Z. (1983). "Grazing of *Euphausia superba* Dana on natural phytoplankton populations." Polar biology **1**(4): 193-197.

Moeys, S., Frenkel, J., Lembke, C., Gillard, J. T., Devos, V., Van den Berge, K., Bouillon, B., Huysman, M. J., De Decker, S. and Scharf, J. (2016). "A sex-inducing pheromone triggers cell cycle arrest and mate attraction in the diatom *Seminavis robusta*." Scientific reports **6**.

Moore, E. R., Bullington, B. S., Weisberg, A. J., Jiang, Y., Chang, J. and Halsey, K. H. (2017). "Morphological and transcriptomic evidence for ammonium induction of sexual reproduction in *Thalassiosira pseudonana* and other centric diatoms." Plos One **12**(7): e0181098.

Moore, J. K., Abbott, M. R. and Richman, J. G. (1999). "Location and dynamics of the Antarctic Polar Front from satellite sea surface temperature data." Journal of Geophysical Research: Oceans **104**(C2): 3059-3073.

Nair, A., Mohan, R., Manoj, M. and Thamban, M. (2015). "Glacial-interglacial variability in diatom abundance and valve size: Implications for Southern Ocean paleoceanography." Paleoceanography **30**(10): 1245-1260.

Naselli-Flores, L. and Barone, R. (2003). Steady-state assemblages in a Mediterranean hypertrophic reservoir. The role of *Microcystis* ecomorphological variability in maintaining an apparent equilibrium. Phytoplankton and Equilibrium Concept: The Ecology of Steady-State Assemblages, Springer**:** 133-143.

Nowlin, W. D. J. and Clifford, M. (1982). "The kinematic and thermohaline zonation of the Antarctic Circumpolar Current at Drake Passage." Journal of Marine Research **40**: 481-507.

Orsi, A. H., Whitworth, T. and Nowlin, W. D. (1995). "On the meridional extent and fronts of the Antarctic Circumpolar Current." Deep Sea Research Part I: Oceanographic Research Papers **42**(5): 641-673.

Pearce, I., Davidson, A. T., Thomson, P. G., Wright, S. and van den Enden, R. (2011). "Marine microbial ecology in the Sub-Antarctic Zone: rates of bacterial and phytoplankton growth and grazing by heterotrophic protists." Deep Sea Research Part II: Topical Studies in Oceanography **58**(21): 2248-2259.

Pfitzer, E. (1869). "Über den Bau und die Zellteilung der Diatomeen." Botanische Zeitung **27**: 774-776.

Pinkernell, S. and Beszteri, B. (2013). A species distribution model of the marine diatom *Fragilariopsis kerguelensis*. EurOceans Hot Topics Conference - A Changing Ocean, Las Palmas, Gran Canaria.

Popp, B. N., Trull, T., Kenig, F., Wakeham, S. G., Rust, T. M., Tilbrook, B., Griffiths, B., Wright, S. W., Marchant, H. J., Bidigare, R. R. and Laws, E. A. (1999). "Controls on the carbon isotopic composition of southern ocean phytoplankton." Global Biogeochemical Cycles **13**(4): 827-843.

Quetin, L. B. and Ross, R. M. (1985). Feeding by Antarctic Krill, *Euphausia superba*: Does Size Matter? Antarctic Nutrient Cycles and Food Webs. W. R. Siegfried, P. R. Condy and R. M. Laws. Berlin, Heidelberg, Springer Berlin Heidelberg**:** 372-377.

R Core Team (2015). "R: a language and environment for statistical computing." from https://www.R-project.org

Rigual-Hernández, A. S., Trull, T. W., Bray, S. G. and Armand, L. K. (2016). "The fate of diatom valves in the Subantarctic and Polar Frontal Zones of the Southern Ocean: Sediment trap versus surface sediment assemblages." Palaeogeography, Palaeoclimatology, Palaeoecology **457**: 129-143.

Rigual-Hernández, A. S., Trull, T. W., Bray, S. G., Cortina, A. and Armand, L. K. (2015). "Latitudinal and temporal distributions of diatom populations in the pelagic waters of the Subantarctic and Polar Frontal Zones of the Southern Ocean and their role in the biological pump." Biogeosciences Discuss. **12**(11): 8615-8690.

Rintoul, S. R. and Sokolov, S. (2001). "Baroclinic transport variability of the Antarctic Circumpolar Current south of Australia." Journal of Geophysical Research: Oceans **106**: 2815-2832.

Rintoul, S. R. and Trull, T. W. (2001). "Seasonal evolution of the mixed layer in the Subantarctic zone south of Australia." Journal of Geophysical Research: Oceans **106**(C12): 31447-31462.

Roselli, L. and Basset, A. (2015). "Decoding size distribution patterns in marine and transitional water phytoplankton: from community to species level." Plos One **10**(5): e0127193.

Round, F. E., Crawford, R. M. and Mann, D. G. (1990). The diatoms: the biology and morphology of the genera. Cambridge, Cambridge University Press.

Sato, S., Beakes, G., Idei, M., Nagumo, T. and Mann, D. G. (2011). "Novel Sex Cells and Evidence for Sex Pheromones in Diatoms." Plos One **6**(10): e26923.

Shukla, S. K., Crespin, J. and Crosta, X. (2016). "*Thalassiosira lentiginosa* size variation and associated biogenic silica burial in the Southern Ocean over the last 42kyrs." Marine Micropaleontology **127**: 74-85.

Shukla, S. K. and Crosta, X. (2017). "*Fragilariopsis kerguelensis* size variability from the Indian subtropical Southern Ocean over the last 42 000 years." Antarctic Science **29**(2): 139-146.

Shukla, S. K., Crosta, X., Cortese, G. and Nayak, G. N. (2013). "Climate mediated size variability of diatom *Fragilariopsis kerguelensis* in the Southern Ocean." Quaternary Science Reviews **69**: 49-58.

Siegel, D. and Deuser, W. (1997). "Trajectories of sinking particles in the Sargasso Sea: modeling of statistical funnels above deep-ocean sediment traps." Deep Sea Research Part I: Oceanographic Research Papers **44**(9): 1519-1541.

Siegel, D. A., Fields, E. and Buesseler, K. O. (2008). "A bottom-up view of the biological pump: Modeling source funnels above ocean sediment traps." Deep Sea Research Part I: Oceanographic Research Papers **55**(1): 108-127.

Smetacek, V., Assmy, P. and Henjes, J. (2004). "The role of grazing in structuring Southern Ocean pelagic ecosystems and biogeochemical cycles." Antarctic Science **16**(4): 541-558.

Smetacek, V., Bathmann, U. and Helmke, E. (2005). "The expeditions ANTARKTIS XXI/3-4-5 of the research vessel Polarstern" in 2004 = Die Expeditionen ANTARKTIS XXI/3-4-5 des Forschungsschiffes "Polarstern" 2004." Berichte zur Polar-und Meeresforschung (Reports on Polar and Marine Research) **500**.

Smetacek, V., Klaas, C., Menden-Deuer, S. and Rynearson, T. A. (2002). "Mesoscale distribution of dominant diatom species relative to the hydrographical field along the Antarctic Polar Front." Deep Sea Research Part II: Topical Studies in Oceanography **49**(18): 3835-3848.

Spaulding, S. A., Jewson, D. H., Bixby, R. J., Nelson, H. and McKnight, D. M. (2012). "Automated measurement of diatom size." Limnology and Oceanography: Methods **10**(11): 882-890.

Stolte, W. and Riegman, R. (1996). "A model approach for size-selective competition of marine phytoplankton for fluctuating nitrate and ammonium." Journal of Phycology **32**(5): 732-740.

Thingstad, T. F. (1998). A theoretical approach to structuring mechanisms in the pelagic food web. Eutrophication in Planktonic Ecosystems: Food Web Dynamics and Elemental Cycling, Springer**:** 59-72.

Timmermans, K. R. and Van Der Wagt, B. (2010). "Variability in Cell Size, Nutrient Depletion, and Growth Rates of the Southern Ocean Diatom *Fragilariopsis kerguelensis* (Bacillariophyceae) after Prolonged Iron Limitation1." Journal of Phycology **46**(3): 497-506.

Tréguer, P. J. (2014). "The southern ocean silica cycle." Comptes Rendus Geoscience **346**(11): 279-286.

Tréguer, P. J. and De La Rocha, C. L. (2013). "The world ocean silica cycle." Annual review of marine science **5**: 477-501.

Verity, P. G. and Smetacek, V. (1996). "Organism life cycles, predation, and the structure of marine pelagic ecosystems." Marine Ecology Progress Series: 277-293.

Waite, A. and Harrison, P. J. (1992). "Role of sinking and ascent during sexual reproduction in the marine diatom *Ditylum brightwellii*." Marine Ecology Progress Series: 113-122.

Waite, A. M. and Nodder, S. D. (2001). "The effect of in situ iron addition on the sinking rates and export flux of Southern Ocean diatoms." Deep Sea Research Part II: Topical Studies in Oceanography **48**(11): 2635-2654.

Wilken, S., Hoffmann, B., Hersch, N., Kirchgessner, N., Dieluweit, S., Rubner, W., Hoffmann, L. J., Merkel, R. and Peeken, I. (2011). "Diatom frustules show increased mechanical strength and altered valve morphology under iron limitation." Limnology and Oceanography **56**(4): 1399-1410.

Wright, S. W., Thomas, D. P., Marchant, H. J., Higgins, H. W., Mackey, M. D. and Mackey, D. J. (1996). "Analysis of phytoplankton of the Australian sector of the Southern Ocean: comparisons of microscopy and size frequency data with interpretations of pigment HPLC data using the 'CHEMTAX' matrix factorisation program." Marine Ecology Progress Series **144**: 285-298.

Zielinski, U. and Gersonde, R. (1997). "Diatom distribution in Southern Ocean surface sediments (Atlantic sector): Implications for paleoenvironmental reconstructions." Palaeogeography Palaeoclimatology Palaeoecology **129**(3-4): 213-250.

Zielinski, U., Gersonde, R., Sieger, R. and Fütterer, D. (1998). "Quaternary surface water temperature estimations: Calibration of a diatom transfer function for the Southern Ocean." Paleoceanography **13**(4): 365-383.

2.6 Publication IV – Morphometry of the diatom *Fragilariopsis kerguelensis* from Southern Ocean sediment: High-throughput measurements show second morphotype occurring during glacials

Michael Kloster[1]
Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Am Handelshafen 12, 27570 Bremerhaven, Germany
Hochschule Emden/Leer, Constantiaplatz 4, 26723 Emden, Germany

Gerhard Kauer
Hochschule Emden/Leer, Constantiaplatz 4, 26723 Emden, Germany

Oliver Esper
Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Am Alten Hafen 26, 27568 Bremerhaven, Germany

Nike Fuchs
Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Am Handelshafen 12, 27570 Bremerhaven, Germany

Bánk Beszteri
Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Am Handelshafen 12, 27570 Bremerhaven, Germany

[1]Author for correspondence: michael.kloster@awi.de, Fon: +49(471)4831-1532

Key index words: paleoceanography, SHERPA, proxy, valve area, rectangularity, Holocene

Abbreviations: λ, mixing proportion; MIS, marine isotope stage; μ, mean; n, sampling size; N, population size; *p*, p-value; PCC, Pearson Correlation Coefficient; R^2, square of PCC; σ, standard deviation; SSST, summer sea surface temperature

2.6.1 Abstract

Fragilariopsis kerguelensis is one of the most abundant diatoms in the sediments of the Southern Ocean. Its morphometric features have been proposed as proxies, based on links with, for example, iron availability, sea surface temperature and glacial / interglacial conditions. We investigated morphometric changes in *F. kerguelensis* valves in a well-studied sediment core record, focusing on transition periods between glacials and interglacials. By applying a high-throughput diatom imaging and morphometry workflow, we found two clearly distinct shape classes, differentiated by their rectangularity. We relate these classes to two different morphotypes, where one of them occurred prominently only in glacial samples, whilst the other persisted throughout. Paleo-proxies based on *F. kerguelensis* valve morphometric features might benefit from differentiating these two morphotypes, since both contribute to a different extent to the feature means under different environmental conditions. Furthermore, the ratio of their abundance also appears as a promising candidate for a paleo-proxy for summer sea surface temperature. Distinguishing between the two morphotypes only became possible by image analysis for precisely measuring diatom valve outlines and area, highlighting the potential of such methods for diatom analyses.

2.6.2 Introduction

The siliceous sediment below large areas of the Southern Ocean contains a diatom record which has extensively been used for paleoceanographic reconstructions (Gersonde et al. 2005, Abelmann et al. 2006, Crosta 2009, Shukla et al. 2009). Reconstructions of paleo-temperatures and of sea ice cover are mostly performed using transfer functions linking taxonomic composition of sediment assemblages to environmental parameters (Zielinski et al. 1998, Gersonde et al. 2003, Crosta et al. 2004, Esper and Gersonde 2014a, Esper and Gersonde 2014b). But observations of intraspecific morphometric variation correlated with environmental conditions have also led to ideas to use such intraspecific variability to learn about past oceanographic conditions (Cortese and Gersonde 2007, Nair et al. 2015, Shukla et al. 2016, Shukla and Crosta 2017). Most attention in this regard was directed at marine diatom species *Fragilariopsis kerguelensis* which is, at least in terms of valve counts, the most abundant diatom species in Southern Ocean sediments (Abelmann and Gersonde 1991, Warnock et al. 2015, Shukla et al. 2016). Variation in quantitative characters like valve length, width and area, as well as striation density, and the composite descriptor F or F* (Equation 4), were proposed to be related to productivity, iron availability, (summer) sea surface temperature, the position of the Antarctic Polar Front and to glacial / interglacial regimes (Fenner et al. 1976, Cortese and Gersonde 2007, Cortese et al. 2012, Shukla et al. 2013). The reliability of these proxies is still under discussion, because *F. kerguelensis* has a highly variable valve morphology (Hasle 1965), which can be affected by multiple, partly intercorrelated causes (Crosta 2009, Cortese et al. 2012, Shukla et al. 2013, Shukla and Crosta 2017). Further drawbacks of these studies are that a) characters beyond simple length measurements, in particular valve area, could only be estimated, not measured directly, b) the data was taken from very limited sample sizes (average 38 (Cortese and Gersonde 2007), 40 (Cortese et al. 2012) or 100 (Shukla et al. 2013) valves per sample), c) the descriptors were only assessed by their means or ranges of means, and d) they only considered a rather limited set of morphometric features.

As a test case for a high throughput diatom morphometrics workflow combining automated microscopic imaging by a slide scanning system and image analyses using the custom developed software tool SHERPA (Kloster et al. 2014, Kloster et al. 2017), we were interested in whether any novel insights can be gained into patterns of *F. kerguelensis* valve morphometric variability in a paleoceanographic context. Thus, we analyzed material from Southern Ocean sediment core PS1768-8 (Gersonde and Hempel 1990, Gersonde 2003), which was taken at 52°35.58´ S 4°28.56´ E, close to the average position of the Antarctic Polar Front, from a depth of 3299 m. This 9.03 m long gravity core covers the past ca. 150,000 years and has already been investigated extensively (Zielinski 1993, Zielinski et al. 1998, Gersonde and Zielinski 2000, Gersonde et al. 2003, Gersonde et al. 2005, Abelmann et al. 2006, Esper and Gersonde 2014b). The use of a partially automated workflow (Kloster et al. 2017) enabled us to process large sample sizes (mostly between ca. 300 and 1,000 valves per sample), and to assess a broad range of morphometric features, from which we were able to analyze their distributions and not only means. Furthermore, valve area was measured directly and at a considerable precision, instead of being estimated based on length and width measurements. Observing distributions of diverse morphometric descriptors, we found that the *F. kerguelensis* “populations” in some layers of this sediment core consisted of two morphotypes, separated by the morphometric descriptor rectangularity (Droop 1995). Below we describe these findings and the occurrence of these morphotypes in two glacial to interglacial transitions in the core, and provide an initial exploration of whether their appearance provides possibilities for developing novel paleo-proxies, in particular related to summer sea surface temperature (SSST).

2.6.3 Methods

We measured valve features with the diatom morphometry software SHERPA (Kloster et al. 2014), applying the workflow and methods described in (Kloster et al. 2017). The sediment core was obtained and processed for diatom analyses as described in (Zielinski 1993). Valve densities on the original slides used by (Zielinski 1993) were rather high. Since this caused difficulties for automated analysis, we re-mounted oxidized material at lower densities. This was done by drying ca. 1 ml diluted valve suspension onto cover slips in 12-well cell culture plates and embedding in Naphrax. Sampling depths were chosen with emphasis on transitions between glacial and interglacial periods (Table 1). Results from sampling depths 800 – 820 cm were obtained by employing a low-magnification scan using a 10× objective, for all other sampling depths a 20× objective was used; settings for SHERPA and employed shape templates are supplied on the annexed data medium in supplement "SHERPA". Data analysis was performed in R (R Core Team 2015), the respective scripts can be found on the annexed data medium in supplement "R". We modeled rectangularity distributions by univariate normal mixtures, fitted by an EM algorithm (Benaglia et al. 2009).

To compare our valve area measurements, as performed by SHERPA by counting pixels lying within the segmented valve area (we will refer to this value as measured valve area), with literature values (which we will refer to as estimated valve area below), we applied the equation given in (Shukla et al. 2013):

$$Estimated\ valve\ area = apical\ length * transapical\ length * 0.8 \quad \textbf{(Equation 1)}$$

We calculated the morphometric descriptor rectangularity (in image processing often referred to as heuristic descriptor) as ratio between the area of the valve and the area of the enclosing rectangle, oriented along the major valve axes (Droop 1995):

$$Rectangularity = \frac{area\ of\ valve}{area\ of\ enclosing\ rectangle} \quad \textbf{(Equation 2)}$$

Ellipticity was calculated based on moment invariants according to (Rosin 2003) as:

$$Ellipticity = \begin{cases} 16\,\pi^2\,I_1 & if\ I_1 \leq \frac{1}{16\,\pi^2} \\ \frac{1}{16\,\pi^2\,I_1} & otherwise \end{cases} \quad \textbf{(Equation 3)}$$

$$with\ I_1 = \frac{\mu_{20}\mu_{02} - \mu_{11}^2}{\mu_{00}^4}$$

and μ_{qp} being the central moments

F or F* was calculated according to (Fenner et al. 1976, Cortese and Gersonde 2007):

$$F = \frac{apical\ valve\ length}{transapical\ valve\ length} * width\ of\ 5\ costae \quad \textbf{(Equation 4)}$$

Costae distance measurement is an experimental SHERPA feature and, until now, was tested on *Fragilariopsis* valves only (Beszteri et al. submitted). This frequency based method is mostly robust towards segmentation flaws, nevertheless results of costae distance ≥ 5 µm were assumed to be incorrect and thus rejected, which applied to ca. 0.5 % of all data. The valve image was segmented by Adaptive Thresholding (Bradski and Kaehler 2008), resulting in costae and valve border colored white, whilst striae and image background were colored black. The segmented image was sampled along the central 80 % of the valve's major (apical) axis by the Bresenham algorithm (Bresenham 1965). Along this sampled line, each white section was replaced by a 5 pixel wide binomial-smoothed peak, resulting in an artificial 2d-image of costae positions. This was transformed into the frequency domain by discrete Fourier transform (DFT). Average costae distance was calculated from the frequency of the maximum magnitude of the Fourier transformed.

The strength of correlations, calculated by the Pearson Correlation Coefficient (PCC), is reported below, following the scheme (Evans 1996) suggested (0.00 - 0.19 = very weak; 0.20 - 0.39 = weak; 0.40 - 0.59 = moderate; 0.60 - 0.79 = strong; 0.80 - 1.00 = very strong).

2.6.4 Results

We analyzed sediment core PS1768-8 by sampling 21 different depths, covering the previous ca. 140,000 years, marine isotope stages (MIS) 1, 2, 5 and 6, and reconstructed SSST between 0.5 and 4.4 °C (Esper and Gersonde 2014b, Abelmann et al. 2015), with emphasis on the transitions between glacial and interglacial periods. Morphometric features of ca. 12,000 valves were measured; the according data is supplied in supplement "R" on the annexed data medium. Table 1 shows summarized data on the different sampling depths.

Table 1: Paleological data on sediment core PS1768-8 (Esper and Gersonde 2014b) and summarized results. Reconstructed SSST ≤ 0.9 °C is highlighted in bold letters.

Sampling depth (cm)	Index referring to Figure 2	Age (yrs)	SSST (°C)	MIS	Sampling size
60	a	10,195	2.7	1	85
80	b	11,182	4.4	1	698
100	c	12,071	2.3	1	857
110	d	12,516	2.2	1	553
120	e	12,960	1.7	1	322
130	f	13,405	1.7	1	83
140	g	13,849	1.8	1	255
150	h	14,912	**0.9**	2	1,659
160	i	16,130	**0.8**	2	1,458
170	j	17,347	**0.6**	2	765
180	k	18,565	**0.8**	2	441
190	l	19,782	**0.5**	2	650
200	m	21,000	**0.8**	2	716
780	n	119,500	2.2	5	471
800	o	123,700	3.5	5	437
810	p	125,800	3.1	5	550
820	q	127,900	1.9	5	420
830	r	130,000	**0.7**	5	732
840	s	132,100	**0.7**	6	166
850	t	134,200	**0.5**	6	265
870	u	138,400	**0.6**	6	274
All depths combined					11,857

2.6.4.1 Valve rectangularity

Combining measurements from all sampling depths, rectangularity showed a clearly bimodal distribution (Figure 1), which we approximated by two overlapping normal distributions, calculated by an EM algorithm for mixtures of univariate normals (Benaglia et al. 2009). Means of these distributions differed by ca. 0.055 (rectangularity is unitless and is constrained within the range 0 - 1) or 7 %. In the following, we refer to them as classes of low and high rectangularity, marked by a green and a maroon curve in Figure 1, respectively. The low rectangularity class was defined by mean (μ) of ca. 0.72 and standard deviation (σ) of ca. 0.014, contributing ca. 70 % (λ) of the valves. The high rectangularity class had a μ of ca. 0.78, a σ of ca. 0.016, and contributed ca. 30 % of all valves investigated. For simplifying separation of these two classes, we used a threshold of 0.753 to distinguish between them, corresponding to the point where the posterior probability of an observation belonging to either class is equal in the normal mixture model presented in Figure 1. A more precise estimation of the contribution of both morphotypes to any single sample / population is possible by fitting the normal mixture with fixed μ and σ and optimizing the value of λ to the

individual distribution, as illustrated in Figure 3, but to avoid confusion, we only report the values from the simpler, threshold-based estimates in the following text unless stated otherwise.

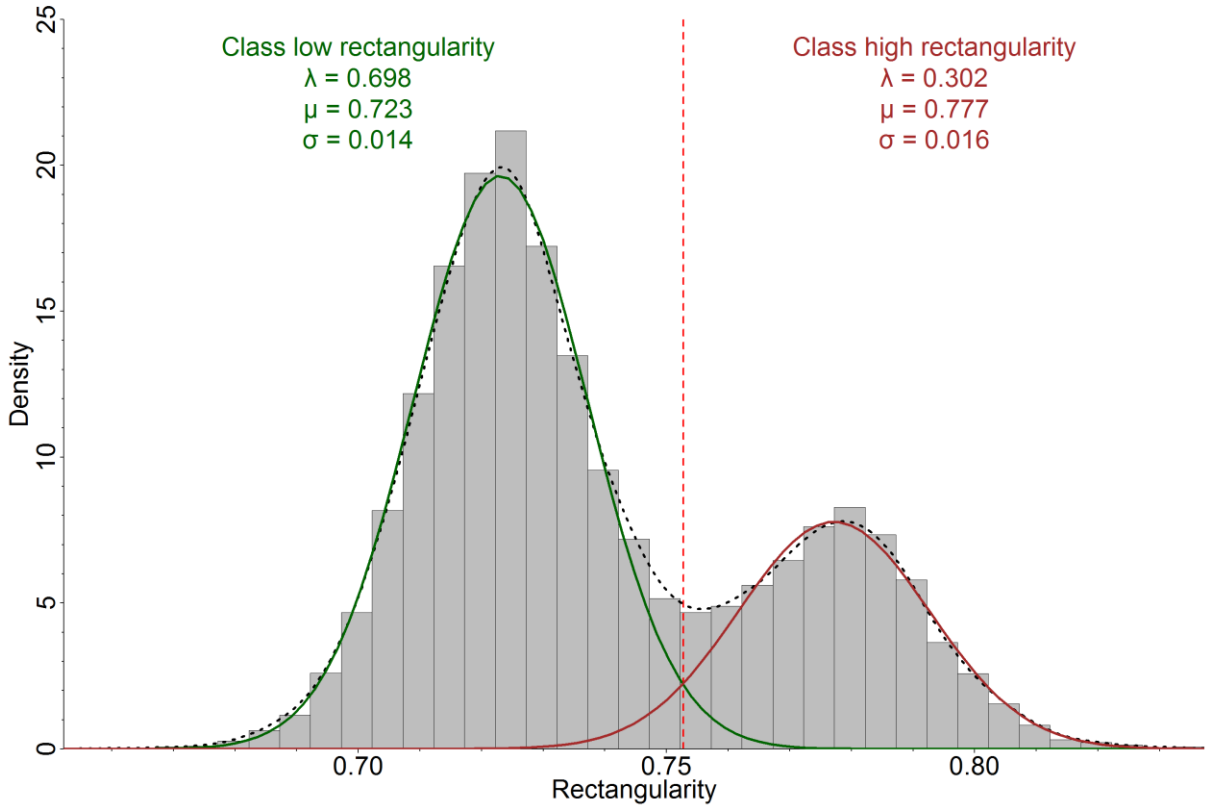


Figure 1: Bimodal distribution of the morphometric descriptor rectangularity over all sampling depths. Green and maroon lines depict the low and high rectangularity classes, based on normal distributions defined by the values stated in the graphic. The dashed red line marks the threshold of 0.753 used to separate the two classes; the dotted black line depicts the density estimate. λ, μ and σ refer to mixing proportion, mean and standard deviation.

As Figure 2 shows, rectangularity distributions from individual sampling depths displayed different patterns, appearing either more uni- or more bimodal; two typical examples are given in Figure 3, and all histograms in Figure 2. Valves from the low rectangularity class comprised between ca. 50 % and nearly 100 % of the *F. kerguelensis* populations (Figure 4). The high rectangularity class preferentially occurred in core depths corresponding to reconstructed SSST values ≤ 0.9 °C (Figure 2 blue color), and contributed ca. 20 – 50 % of valves measured. For layers with the reconstructed value of SSST lying ≥ 1.7 °C (Figure 2 red color), the high rectangularity class contributed only ca. 5 – 15 %, with the single exception of 33 % for 110 cm sampling depth (Figure 4). SSST between 0.9 °C and 1.7 °C was not covered by our samples. Two examples, which represent the samples corresponding to the highest and the lowest reconstructed SSST we investigated, are given in Figure 3: For 80 cm sampling depth (4.4 °C SSST), the population was dominated by the low rectangularity class by 95 %; the high rectangularity class was represented only marginally (Figure 3 a). In contrast, for 190 cm depth (0.5 °C SSST), both classes contributed nearly equally (Figure 3 b).

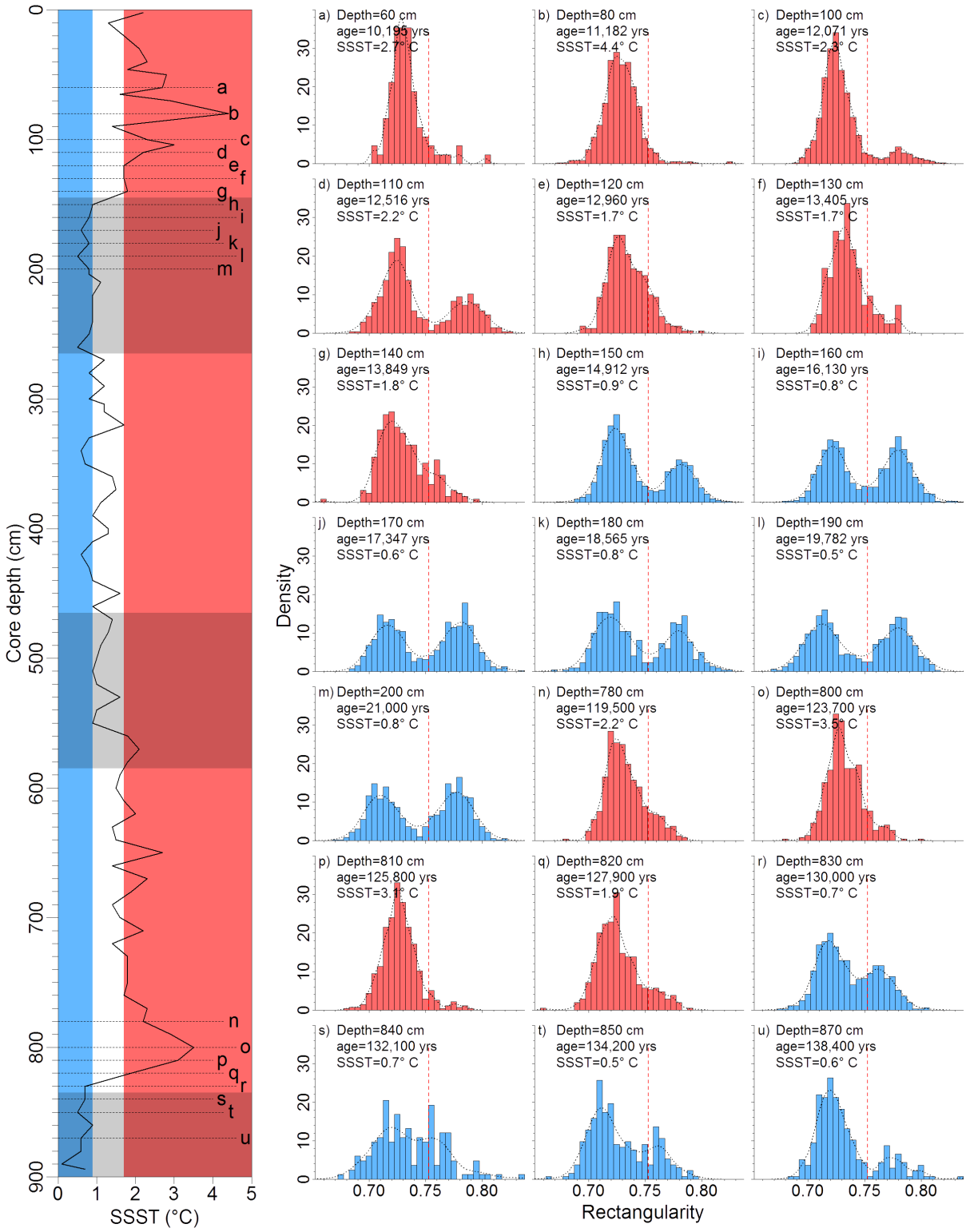


Figure 2: SSST reconstructed using a diatom transfer function and rectangularity histograms for samples from different depths. Left: depth of core related to SSST, after (Esper and Gersonde 2014b). Horizontal dotted lines mark the sampling depths we investigated, referred to by letters a to u. Shaded areas indicate glacial periods according to MIS. Right: rectangularity histograms, a) – u) refer to the different sampling depths as indicated in the left hand panel and in Table 1. SSST ≤ 0.9 °C is marked in blue, ≥ 1.7 °C in red color. Vertical dashed red lines in the histograms mark the threshold of 0.753 we used for separating the two rectangularity classes; the dotted black curves depict the density estimates.

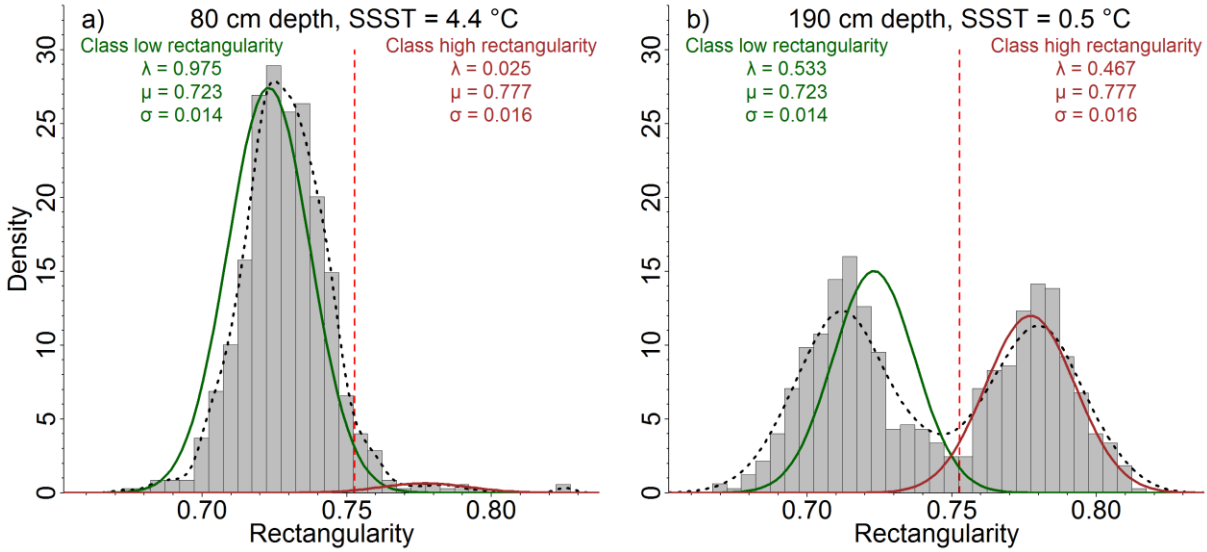


Figure 3: Exemplary rectangularity histograms and contribution of rectangularity classes for a) 80 cm and b) 190 cm sampling depth. Green and maroon lines depict the low and high rectangularity classes as shown in Figure 1. The dashed red lines mark the threshold of 0.753 we used for separating the two rectangularity classes; the dotted black lines depict the density estimates. λ, μ and σ refer to mixing proportion, mean and standard deviation.

The relative contribution of both rectangularity classes across all investigated core depths is summarized in Figure 4. Samples from MIS 6 showed a much lower contribution of the high rectangularity class than samples with similar reconstructed SSST values from MIS 2, whilst for 110 cm sampling depth, it contributed exceptionally strong compared to nearby SSSTs. Apart from these cases, a general trend of increased contribution of the high rectangularity class at low reconstructed SSST values could be observed.

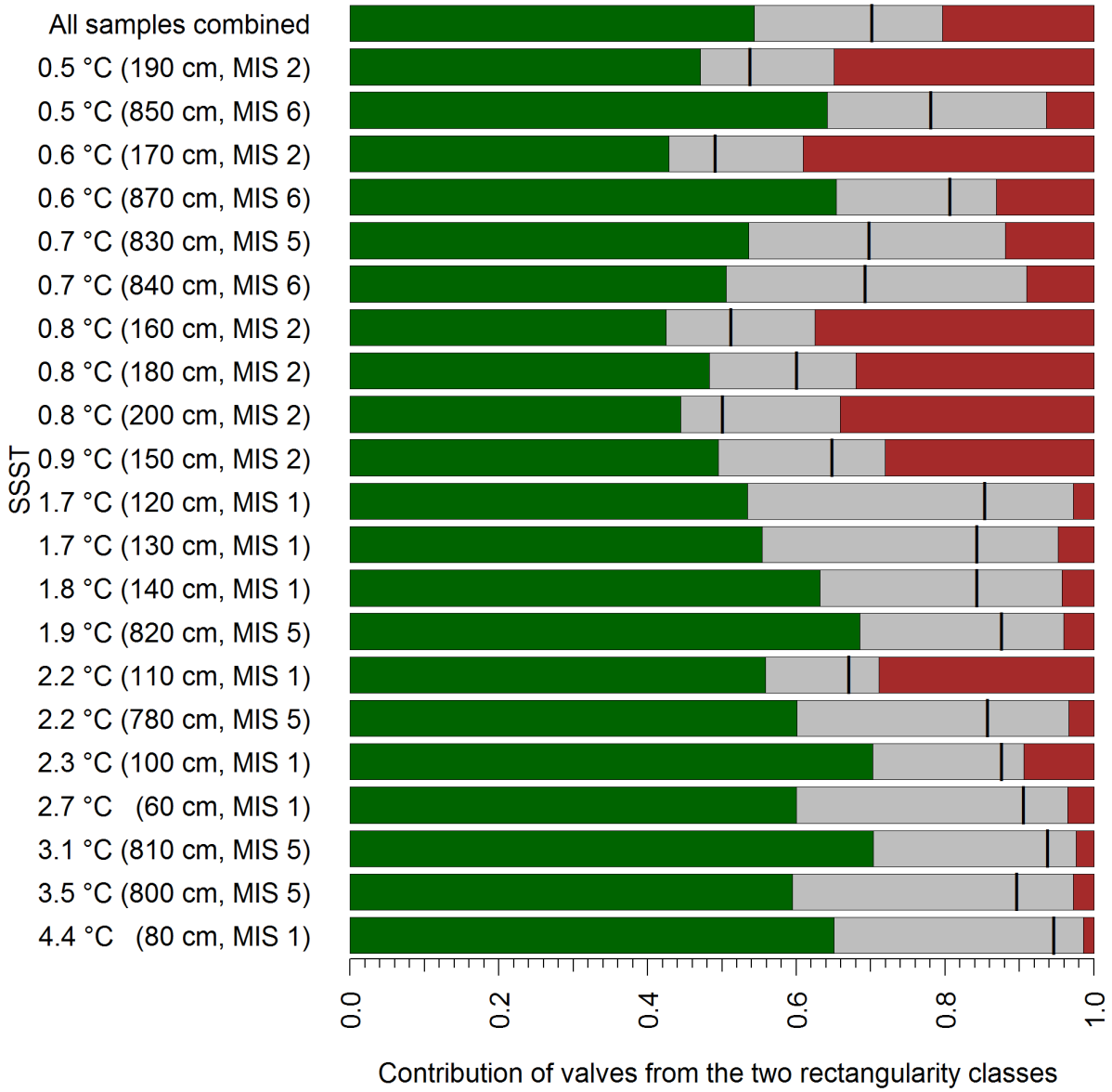


Figure 4: Contribution of valves from the two rectangularity classes from different reconstructed SSSTs and sampling depths. The proportion of valves belonging to the low rectangularity class with a high confidence is shown in green, for the high rectangularity class in maroon. Gray color marks the proportion of valves for which class assignment was uncertain, i.e. the posterior probability of belonging to one class differed from that of the other class by less than a factor of 100, according to the EM algorithm. Black marks indicate the class border by a rectangularity threshold of 0.753. MIS 2 and 6 depict glacial, 1 and 5 interglacial periods.

2.6.4.2 Rectangularity compared to valve length distribution

Since in *F. kerguelensis*, like in most other diatoms with elongate valves, valve outline shape changes accompany size decrease, one might expect rectangularity to show such size dependence, and the presence of two somewhat distinct size classes might then even explain the bimodal rectangularity distributions observed. That this was not the case is illustrated by Figure 5. Two peaks appeared in the joint distribution of rectangularity and valve length, but they appeared at similar valve lengths (around 25 - 30 µm), and were separated clearly on the rectangularity axis. Thus, we can exclude a bimodal size distribution as an explanation of the bimodal rectangularity distributions.

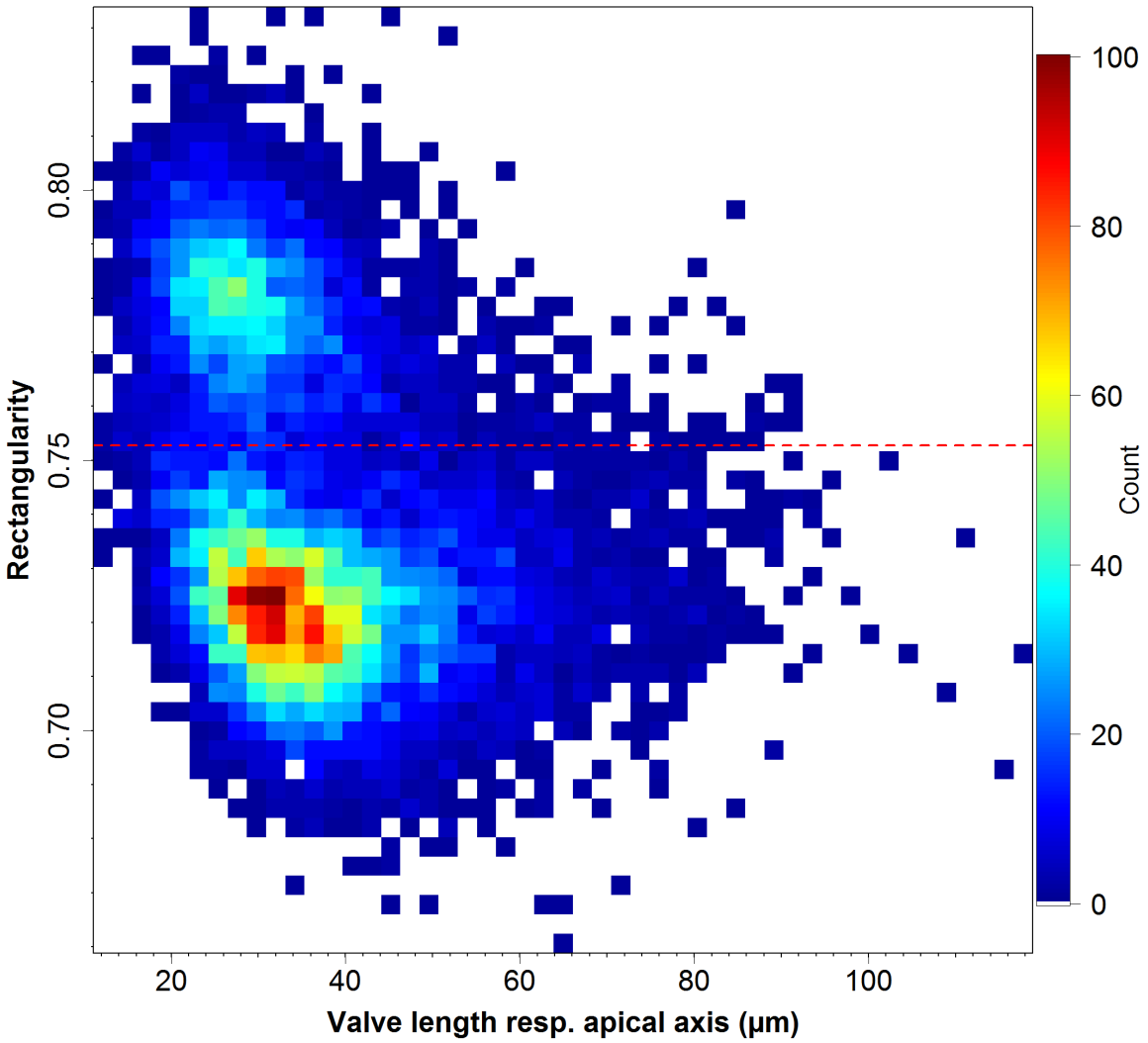


Figure 5: Counts of valves of different rectangularity, compared to valve length, for measurements from all depths combined. The dashed red lines mark the threshold of 0.753 we used for separating the two rectangularity classes.

2.6.4.3 Effect of separating two rectangularity classes on means of further morphometric features

We compared mean values of feature from the individual rectangularity classed with their overall means (i.e. both rectangularity classes combined), to see how large a difference it makes to separate both classes. Separated by the two distinct classes of rectangularity, means of some morphometric features showed substantial differences from the overall means (Table 2), in the case of valve length, width and area by ca. 10 – 20 %. In contrast, costae distance remained nearly constant between both classes, and the combined feature F or F* (Equation 4) showed only little difference (ca. 5 %).

Table 2: Morphometric features, separated according to the two rectangularity classes; differences compared to values from both rectangularity classes combined are in brackets; differences ≥ 10 % absolute are marked in bold letters.

	Both rectangularity classes combined	Low rectangularity class	High rectangularity class
Rectangularity mean (µm)	0.74	0.72 (−2.2 %)	0.78 (+5.24 %)
Length mean (µm)	36.04	38.33 (+6.4 %)	**30.63 (-14.99 %)**
Length median (µm)	33.44	35.67 (+6.6 %)	**28.77 (-13.98 %)**
Width mean (µm)	9.15	9.56 (+4.5 %)	**8.18 (-10.61 %)**
Area mean (µm²)	250.96	272.17 (+8.4 %)	**200.96 (-19.92 %)**
Costae distance mean (µm)	1.91	1.92 (+0.1 %)	1.91 (-0.24 %)
F or F* mean (µm)	37.79	38.59 (+2.1 %)	35.90 (-4.99 %)

2.6.4.4 Correlations between SSST and means of morphometric features

SSST values reconstructed by (Esper and Gersonde 2014b) correlated stronger with means of morphometric features from the low rectangularity class (Table 3), with the exception of mean valve length and F or F*, which correlated stronger for the high rectangularity class. Correlation was substantially stronger for one of the classes than for both combined only in the case of mean valve area and length.

Table 3: Pearson correlation coefficient (PCC) for correlation of SSST to means of morphometric features per sampling depth. Coefficients with highest absolute values in each row are marked in bold letters. N = 21; *** marks PCC with $p < 0.001$, ** marks PCC with $p < 0.01$, * marks PCC with $p < 0.05$, · marks PCC with $p < 0.1$, unmarked PCC have $p \geq 0.1$.

Feature	PCC for both rectangularity classes combined	PCC for low rectangularity class	PCC for high rectangularity class
Mean rectangularity	-0.544 *	**0.662** **	-0.444 *
Mean valve area	-0.211	**-0.590** **	0.276
Mean valve length	0.183	-0.313	**0.567** **
Median valve length	0.044	**-0.382** ·	-0.132
Mean valve width	-0.497 *	**-0.702** ***	-0.610 **
Mean costae distance	0.256	**0.341**	-0.222
Mean F or F*	0.540 *	0.449 *	**0.698** ***

2.6.4.5 Valve area

The formula (Equation 1) used previously by (Shukla et al. 2013) overestimated valve areas, as measured by SHERPA, on average by ca. 9 % (Table 4).

Table 4: Comparison of estimated versus measured valve area

Sampling depth (cm)	Mean measured area (µm²)	Mean estimated area (µm²)	Difference between means of measured and estimated area (µm²)
60	256.28	280.74	24.45 (+9.54 %)
80	255.59	280.70	25.10 (+9.82 %)
100	271.57	298.49	26.92 (+9.91 %)
110	243.33	264.52	21.19 (+8.71 %)
120	191.71	209.05	17.35 (+9.05 %)
130	269.82	294.33	24.50 (+9.08 %)
140	267.04	293.29	26.25 (+9.83 %)
150	246.58	267.50	20.92 (+8.48 %)
160	232.57	250.35	17.78 (+7.64 %)
170	244.12	262.96	18.84 (+7.72 %)
180	258.83	280.68	21.86 (+8.44 %)
190	263.64	286.41	22.77 (+8.64 %)
200	258.43	279.98	21.56 (+8.34 %)
780	270.81	295.87	25.06 (+9.25 %)
800	206.07	225.25	19.18 (+9.31 %)
810	251.56	277.07	25.51 (+10.14 %)
820	268.40	295.90	27.50 (+10.25 %)
830	263.58	288.43	24.84 (+9.42 %)
840	252.08	276.04	23.97 (+9.51 %)
850	266.11	294.75	28.64 (+10.76 %)
870	279.22	306.83	27.61 (+9.89 %)
All depths	**250.96**	**273.46**	**22.49 (+8.96 %)**

For the low rectangularity class, we found a moderate linear correlation of mean valve area to SSST with $R^2 = 0.348$ (Figure 6; PCC = -0.59, N = 21, $p = 0.0049$ for the slope and $p < 2.2 \cdot 10^{-16}$ for the intercept). For the high rectangularity class, R^2 was below 0.08, giving no indication of a correlation to SSST.

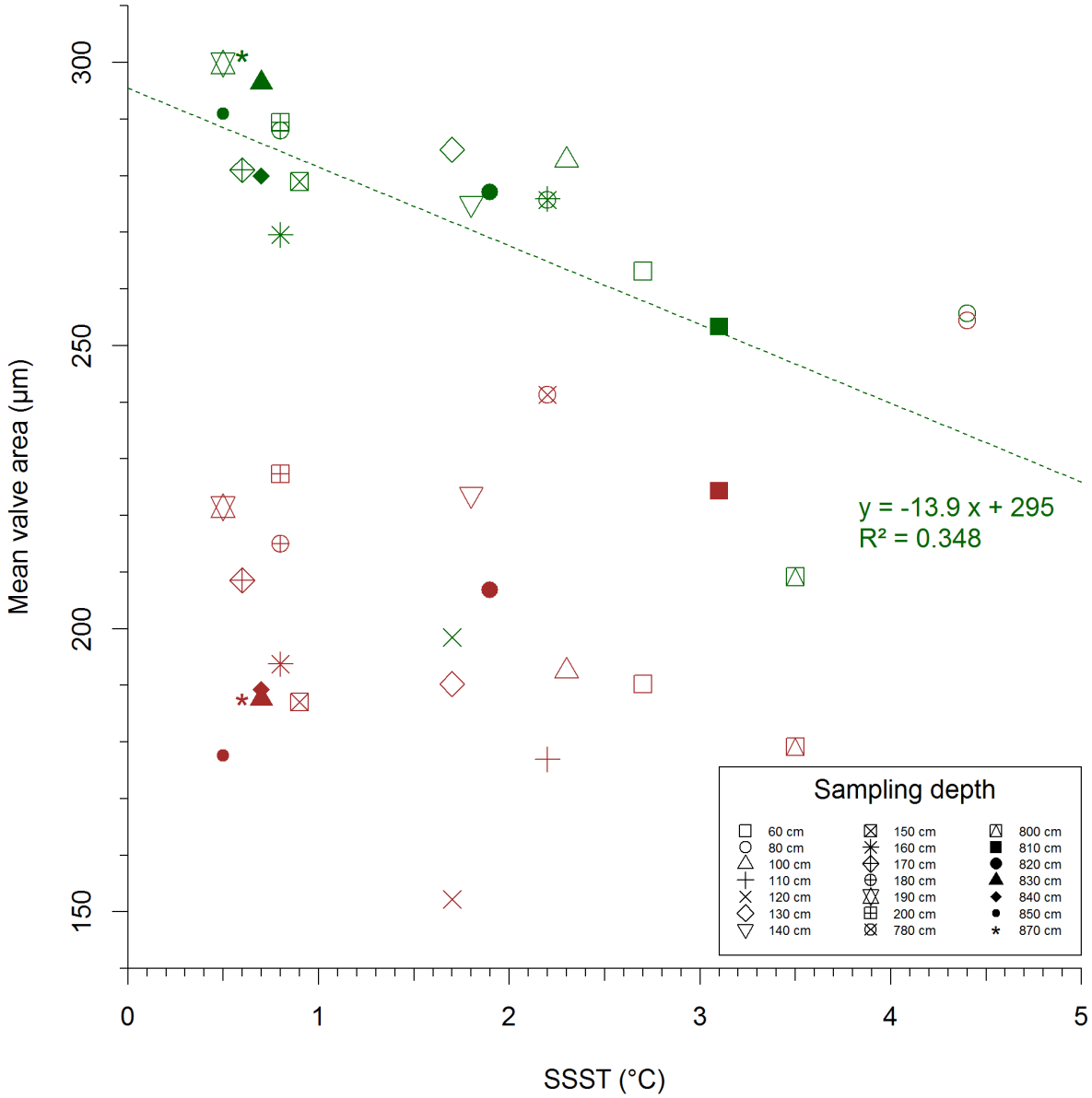


Figure 6: Comparison of mean valve area and SSST, for low rectangularity class (green) and high rectangularity class (maroon); sampling depths are depicted by the indicated symbols. Ca. 35 % of the variability in mean valve area within the low rectangularity class can be explained by a linear correlation to SSST (dashed green line), and less than 8 % for the high rectangularity class mean area (not shown).

2.6.5 Discussion

We investigated morphometric changes in *F. kerguelensis* valves in a well-studied sediment core record. Unexpectedly, we found two clearly distinct classes, differentiated by their rectangularity. Means of these classes differed by only ca. 7 %, yet they were clearly different, as visible in histograms shown in Figures 1 – 3. Means of several morphometric features were substantially different for both classes (Table 2), and proportion of valves matching these two classes varied over sampling depths of sediment core PS1768-8 (Figures 2 & 4). We found a substantial offset in valve area estimated by established methods, compared to our measurements (Table 4). Furthermore, we observed that differentiating *F. kerguelensis* low rectangularity morphotype leads to an improved correlation between valve area and reconstructed summer sea surface temperature (Figure 6 and Table 3).

2.6.5.1 Morphotypes distinct by rectangularity

F. kerguelensis morphology, as well as material from sediment core PS1768-8, has already been investigated thoroughly. Nevertheless, we observed two classes of different rectangularity (Figure 1)

which have not been recognized before. A few considerations on valve ellipticity (in a qualitative sense, not to be confused with the morphometric descriptor ellipticity) have, however, been made in the literature (see below). These might be related to the aspect of morphological variability captured by rectangularity.

We defined a low rectangularity class with $\mu = 0.723$ and $\sigma = 0.014$, and a high rectangularity class with $\mu = 0.777$ and $\sigma = 0.016$, by fitting a normal mixture model to the combined data of all sampling depths (Figures 1 & 2). For some depths, the distributions were slightly shifted, compared to the overall average position (Figure 3 and supplement "Two rectangularity classes for each sampling depth"), but still nevertheless separable. Whenever both classes were clearly visible, also transitional forms, covering the whole range between them, were present; yet rectangularity distributions appeared strongly bimodal in most of these cases. The two classes overlapped only slightly; for separating them we applied a threshold of 0.753 for rectangularity, as derived from the normal mixture model fit to the full data set (i.e., all core depths pooled; see Figures 1 & 4).

Table 5: Correlations between several valve properties and morphometric descriptors. N = 11587; $p < 0.001$ is marked with ***.

Correlated properties	**PCC**
a) Valve length & ellipticity	-0.608 ***
b) Valve length & rectangularity	-0.313 ***
c) Rectangularity & ellipticity	0.769 ***

We refer to these rectangularity classes as to two morphotypes; exemplary valves are depicted in Figure 7. Morphotype I (Figure 7 a – d) exhibits a lower rectangularity (mean 0.72), due to its rather pointed tips, resulting in a more lanceolate shape. In contrast, morphotype II (Figure 7 e – h) has a higher rectangularity (mean 0.78) with rounded tips, resulting in a more elliptical shape. A shape transition from lanceolate towards elliptic valve outlines can often be observed as a result of size reduction in pennate diatoms (Woodard et al. 2016), and might be reflected in the morphometric descriptors ellipticity (Equation 3) as well as rectangularity (Equation 2). In line with this idea, valve length in our data correlated strongly with ellipticity (Table 5 a), but only weakly with rectangularity (Table 5 b), even though both descriptors correlate strongly with each other (Table 5 c). Furthermore, the distribution of ellipticity is not as clearly bimodal as that of rectangularity. Therefore differences between the two morphotypes seem to be better captured by rectangularity, even though visually they appear to relate to a less, or to a more elliptical valve shape, respectively. We showed in Figure 5 that these differences do not seem to be caused by mitotic size reduction in our case.

The proportion of these morphotypes in individual core depths is related to different SSST regimes, like e.g. distinct oceanic water masses or environmental changes during glacial / interglacial periods (Figures 2 – 4). Morphotype I represents the majority of *F. kerguelensis* valves in all our samples, contributing between ca. 50 and nearly 100 % (Figure 3), with its relative abundance being highest during interglacial periods, for SSST ≥ 1.7 °C (Figure 4). In contrast, proportion of morphotype II was highest during glacial periods, for SSST ≤ 0.9 °C; here it comprised more than ca. 20 % of *F. kerguelensis* valves. For SSST of 1.7 °C and above, it contributed less than ca. 15 %, with the exception of 33 % for sampling depth 110 cm, corresponding to 2.2 °C SSST.

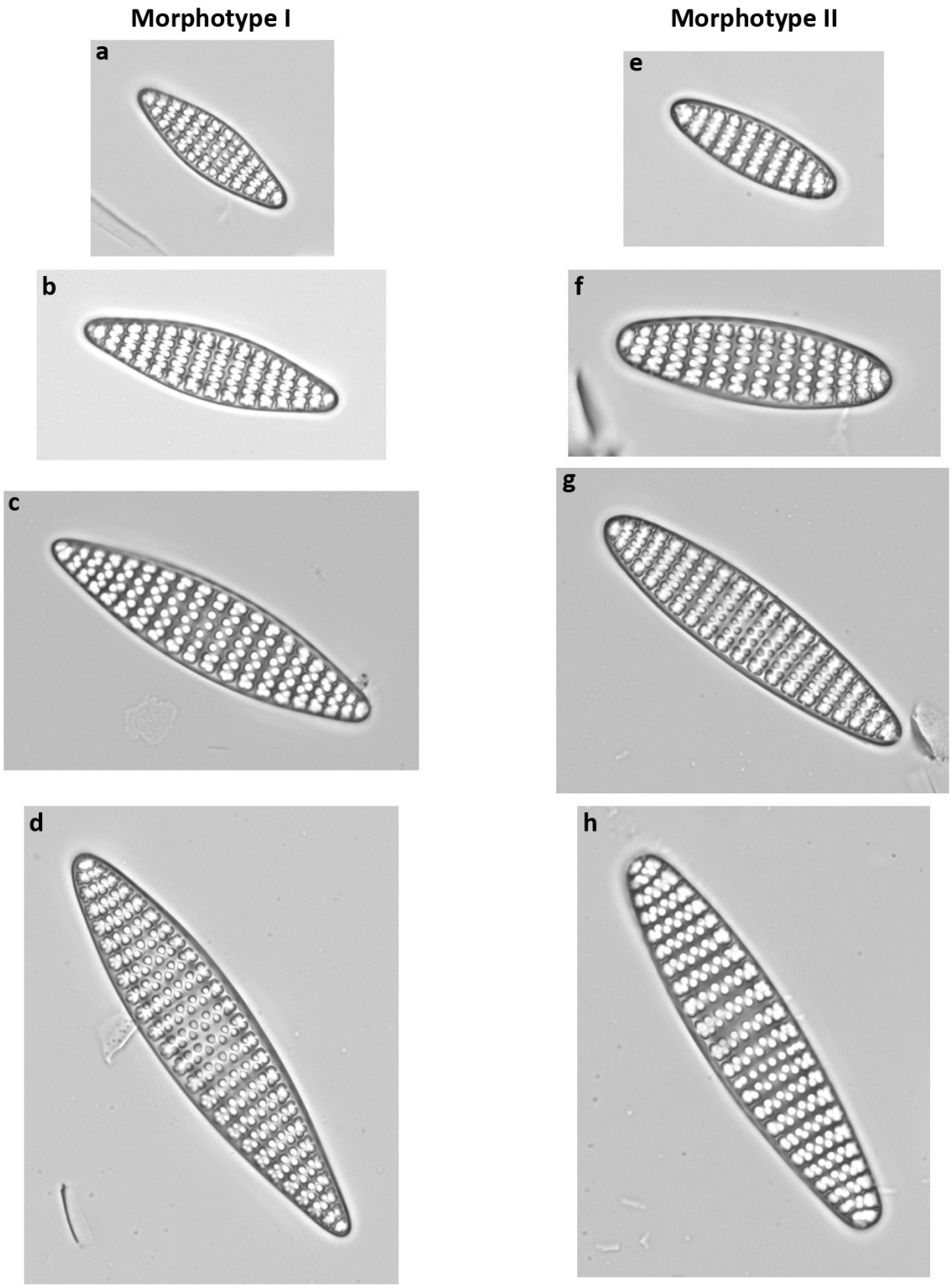


Figure 7: Examples of the two *F. kerguelensis* morphotypes. a – d) morphotype I with rather pointed tips and more lanceolate shape. e – h) morphotype II with roundish tips and more elliptical shape. Valve lengths are ca. 20, 30, 40, 50 µm from top to bottom.

Whether these morphotypes correspond to morphological variants reported previously in the species is in part difficult to decide. (Donahue 1970) proposed *Nitzschia kerguelensis* var. *ovalis* (*Nitzschia kerguelensis* being synonymous with *F. kerguelensis*), documented by six valve images. Judging by its name, one could hypothesize that this variety corresponds to the more elliptical morphotype II we observed. We digitized and analyzed with SHERPA Donahue's published images of

her var. *ovalis*, resulting in rectangularity ranging from 0.70 to 0.72. These values are at the lower side of the distributions for our low rectangularity class, depicting the more lanceolate morphotype I. She further reported a higher abundance of the proposed var. *ovalis* north of the Antarctic Convergence. This is also in contrast with our findings that morphotype II is abundant mostly at temperatures at the lower end of *F. kerguelensis* physiological temperature range. Accordingly, we conclude that the *Nitzschia kerguelensis* var. *ovalis*, as proposed by (Donahue 1970), is not congruent with the *F. kerguelensis* morphotype II we found.

(Frenguelli 1960) also described an elliptically shaped variant in this species as *F. antarctica* var. *elliptica* (*F. antarctica* being synonymous with *F. kerguelensis*). However, this variety was rejected by (Hasle 1965) because she observed that more elliptical valves usually occurred at lower sizes, and was of the opinion that this resulted from allometric change in valve outline shape. (Cefarelli et al. 2010) agreed, since for Frenguelli's material they "*found a close correlation between form and valve size*". Our study is the first where the question of correlation between size and valve outline shape in *F. kerguelensis* is addressed quantitatively. Rectangularity was only weakly correlated with valve length (PCC = -0.313, N = 11857, $p < 2.2 \cdot 10^{-16}$, Figure 5), and thus probably not a result of mitotic size reduction. From the single valve drawing which (Frenguelli 1960) used to illustrate *F. antarctica* var. *elliptica*, we measured a rectangularity value of 0.78, which conforms to the more elliptical morphotype II we found. (Cefarelli et al. 2010) presented further light microscopic images from Frenguelli's original material, from which we also measured rectangularity with SHERPA. One of two var. *elliptica* examples had a rectangularity of 0.769, which matches the high rectangularity class well, but this measurement might be imprecise due to partly defocused valve edges. However, the other example of var. *elliptica*, as well as the three *F. kerguelensis* var. *kerguelensis* examples, were within the overlapping range of the two rectangularity classes (0.746 – 0.758), so their class assignment is ambiguous. Nevertheless, at Frenguelli's sampling site (under ice near Port Martin winter station, Adélie Land at 66°49´ S), a SSST < 0.9 °C can be expected, which conforms with increased abundance of morphotype II in our sediment core samples. Thus it seems possible that the var. *elliptica* described by (Frenguelli 1960) matches the morphotype II we found. In this case, our analyses can be taken as evidence against the arguments put forward by (Hasle 1965) and (Cefarelli et al. 2010) against the biological validity of Frenguelli's proposed variety.

Based on only fossil material, it is difficult to decide what the biological background of the observed morphotypic differentiation might be. One possibility would be that the two morphotypes correspond to two different species or subspecies, that is, groups of populations which are either reproductively isolated, or in reproductive contact, but under strong divergent selection upon their valve morphologies. Another possibility is phenotypic plasticity, that is, it is possible that *F. kerguelensis* cells can (possibly depending on environmental conditions, for example, temperature) switch their valve outline shapes from lanceolate to more elliptic. When looking at morphometric characters besides rectangularity in our light microscopic data set, we see very little other differences between both morphotypes. In spite of this, the presence of two distinct taxa might still be a more probable explanation for our findings, since we know of no examples from other diatoms of a morphological switch which is similar to that between our two rectangularity classes.

2.6.5.2 F. kerguelensis valve properties as paleo-proxy

For *F. kerguelensis,* valve area, valve length and the combined factor F or F* have been proposed as potential paleoclimate proxies (Cortese and Gersonde 2007, Cortese et al. 2012, Shukla et al. 2013, Nair et al. 2015, Shukla and Crosta 2017). Our methods and findings might have implications for this application of *F. kerguelensis* morphometrics in two aspects. First, using slide scanning combined

with image analyses allows increasing sample sizes, and, probably even more important, a more precise measurement of morphometric characters apart from simple length measurements. The most relevant case in point is valve area: estimations according to Equation 1 (Shukla et al. 2013) overestimated values, compared to our direct measurements by ca. 9 % on average (Table 4). This is evident by considering that the rectangularity from our measurements, having an overall mean of 0.74, corresponds to their correction-factor of 0.8. If this was a consistent bias across different samples or core depths, this would make little difference; however, due to the varying contributions of two morphotypes (Figure 4) and their distinct rectangularity, the bias affects valve area estimates from different communities to different extents. Accordingly we propose that, whenever feasible, actual measurement of valve area should be preferred over estimation based on measuring valve length and width only.

The second aspect with potential relevance for *F. kerguelensis* morphometric proxies is our finding of two different morphotypes. Whether both morphotypes correspond to distinct (sub-)species or to phenotypic variants of the same species, their relative occurrence shows a strong correspondence with paleo-environmental conditions, since the high rectangularity morphotype II occurred preferentially during glacial periods. Based on this, it is both possible that the relative contribution of both morphotypes might become useful as (part of) a paleoceanographic proxy, and/or that explicitly distinguishing both morphotypes might improve the precision or resolution of any one of the previously proposed *F. kerguelensis* morphometric proxies. We address these points shortly below.

$$Ratio\ of\ rectangularity\ classes = \frac{number\ of\ valves\ with\ rectangularity \leq 0.753}{number\ of\ valves\ with\ rectangularity > 0.753} \quad \textbf{(Equation 5)}$$

As for the relative contribution of both morphotypes, we found that SSST showed a strong linear relationship with the ratio of contribution (defined in Equation 5; Figure 8), explaining ca. 80 % of the SSST variability (PCC = 0.89, N = 21, $p = 6 \cdot 10^{-8}$ for the slope and $p = 0.008$ for the intercept).

As for possible improvement of morphometric proxies by differentiating both morphotypes, we showed in Table 2 that several morphometric characters differed slightly between both morphotypes, e.g. by up to ca. 20 % for mean valve area. Since the proportion of the two morphotypes varied significantly between different samples (Figure 4), this might have a confounding effect on morphometric trends as observed for *F. kerguelensis* as a whole (i.e., not differentiating both morphotypes). We showed in Figure 6 and Table 3 that differentiating both morphotypes, and using valve area from only one of them, leads to an increased correlation with reconstructed SSST values. To decide whether this character would be a more precise proxy is not possible based on sediment core material only; to address these aspects, we are currently investigating surface sediment material and plankton net catches.

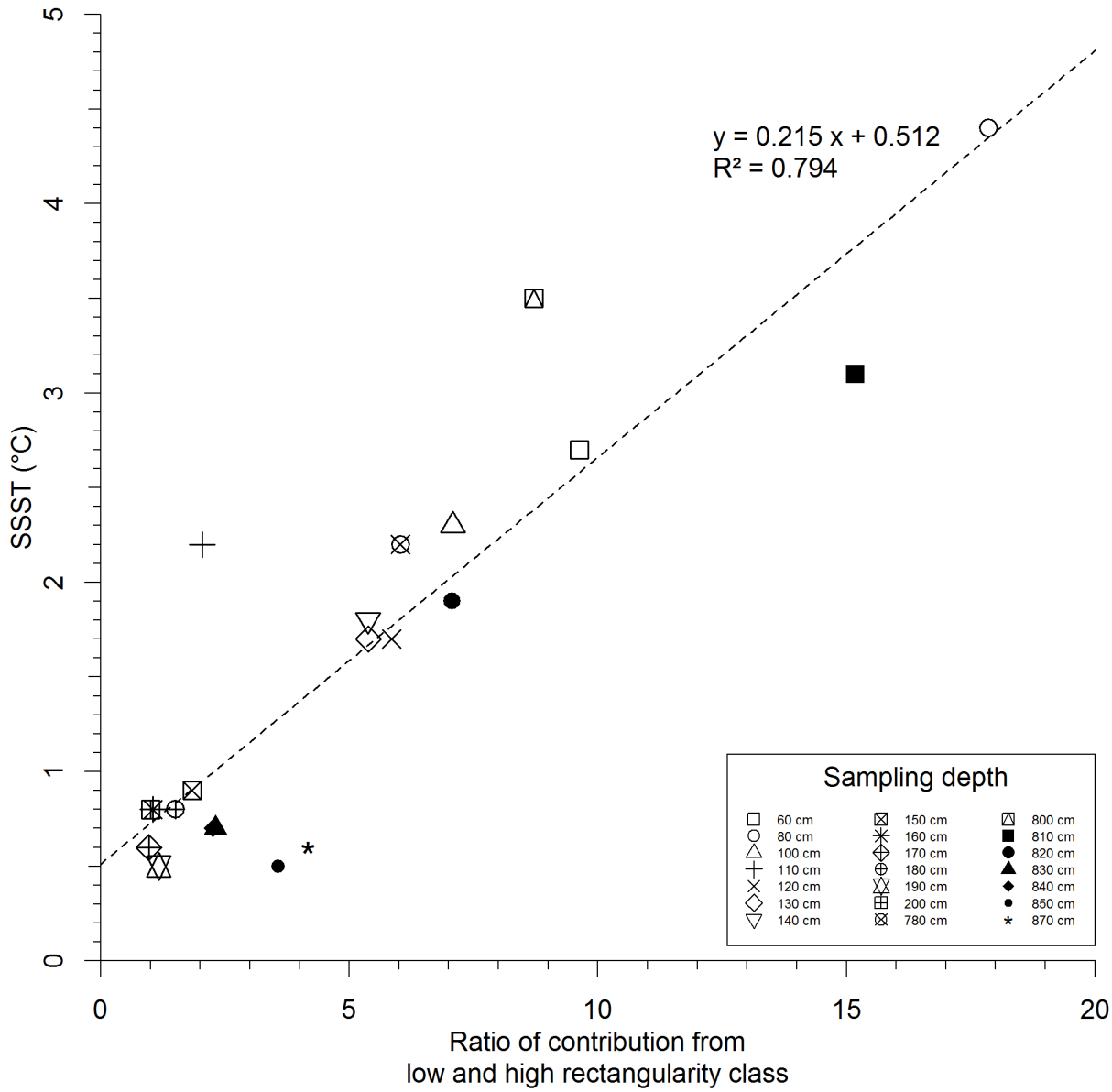


Figure 8: Correlation between ratio of contribution from the two rectangularity classes and SSST. Ca. 80 % of the SSST variability can be explained by a linear correlation with the ratio (dashed line).

2.6.6 Conclusion

We found two morphotypes of *F. kerguelensis*, distinguished by characteristics of the valve tips, which appear to be either rather pointed, resulting in a more lanceolate valve shape, or rather rounded, resulting in a more elliptical valve shape. This disparity becomes apparent in the morphometric descriptor rectangularity, which indicates the presence of two clearly distinct classes for the material we investigated. These classes can be separated easily by a rectangularity threshold, and increased occurrence of morphotype II might be related to lower SSST.

Paleo-proxies based on *F. kerguelensis* valve morphometric features might benefit from differentiating these two morphotypes, since both contribute to a different extent to the feature means under different environmental conditions. Furthermore, the ratio of abundance of both morphotypes also appears as a promising candidate for a paleo SSST proxy, which however needs to be substantiated by further examination of recent material from known environmental conditions.

Acknowledgements
This work was supported by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the priority programme 1158 “Antarctic Research with comparative investigations in Arctic ice areas” by grant nr. BE4316/4-1, KA1655/3-1.

We would like to express our gratitude to Rainer Gersonde for supplying sediment core PS1768-8 material, as well as to Friedel Hinz, Sean Seegert, Sarah Olischläger, Fenina Buttler and Fabian Altvater for preparing material used in this study and in previous preliminary experiments.

Author contributions
Bánk Beszteri and Gerhard Kauer supervised the project. Oliver Esper contributed material samples and paleoceanographic background. Nike Fuchs conducted preliminary measurements which were instrumental for designing this study. All of the aforementioned persons contributed to the paper’s content. Slide scanning, image analysis and measurements via SHERPA, data analysis and paper writing were performed by Michael Kloster.

2.6.7 Supplements

2.6.7.1 Data
The raw data of morphometric measurements are provided in supplement “R” on the annexed data medium.

2.6.7.2 SHERPA settings
The configuration files and shape templates are provided in supplement “SHERPA” on the annexed data medium. Settings for the low-resolution scan can be found in “54S-800m.20x.dmini”, for the high-resolution scan in “54S-800m.63x.dmini”

2.6.7.3 Two rectangularity classes for each sampling depth

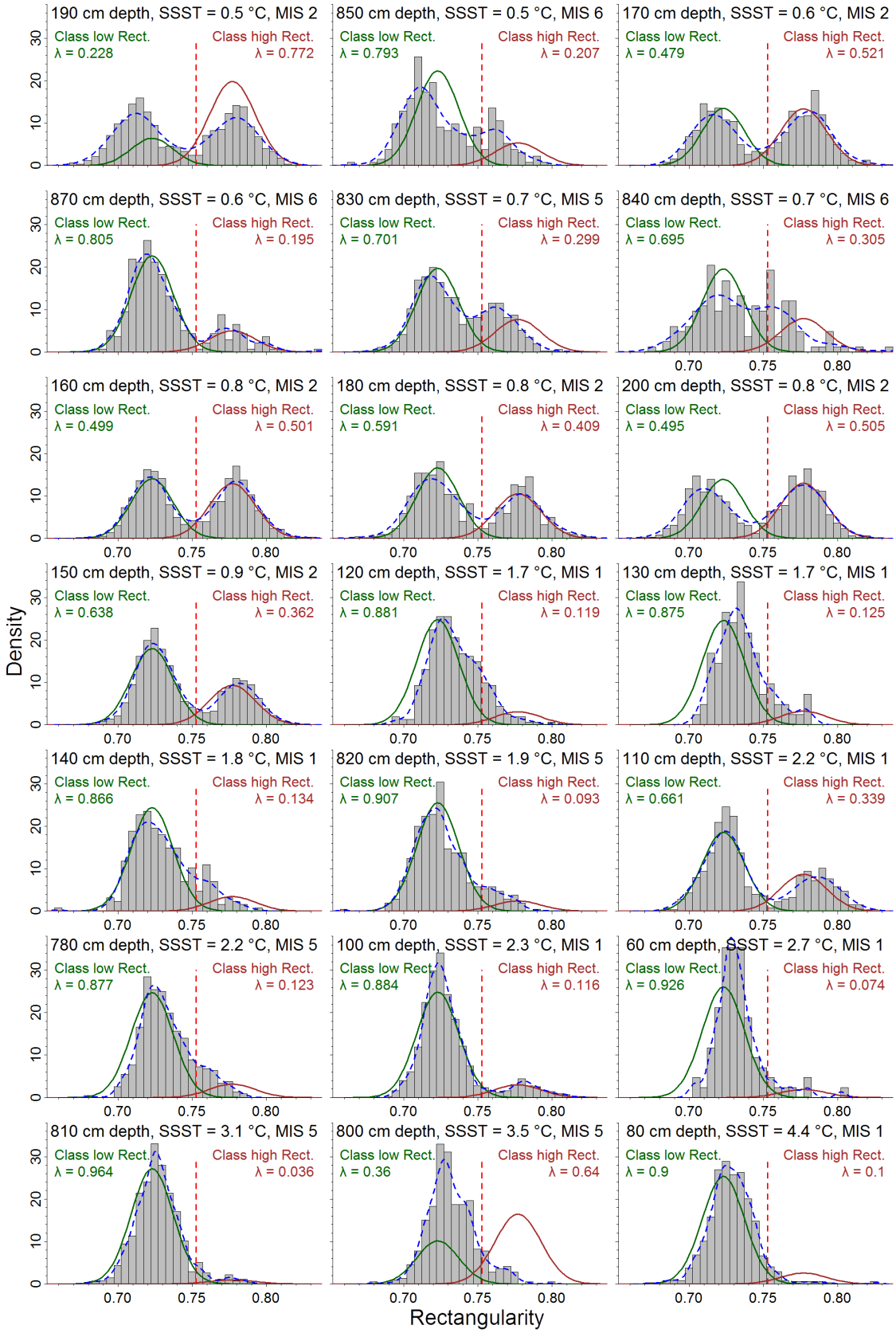


Figure A1: Rectangularity histograms and contribution of rectangularity classes for each sampling depth. Green and maroon lines depict the low and high rectangularity classes as shown in Figure 1. Normal mixtures were fitted by an EM algorithm (Benaglia et al. 2009) with fixed μ and σ (according to the values stated in Figure 1) and optimizing the value of λ to the individual distribution. The dashed red lines mark the threshold of 0.753 we used for separating the two rectangularity classes; the dotted black lines depict the density estimates. λ, μ and σ refer to mixing proportion, mean and standard deviation.

2.6.8 References

Abelmann, A. and Gersonde, R. (1991). "Biosiliceous Particle-Flux in the Southern-Ocean." Marine Chemistry **35**(1-4): 503-536.

Abelmann, A., Gersonde, R., Cortese, G., Kuhn, G. and Smetacek, V. (2006). "Extensive phytoplankton blooms in the Atlantic sector of the glacial Southern Ocean." Paleoceanography **21**(1).

Abelmann, A., Gersonde, R., Knorr, G., Zhang, X., Chapligin, B., Maier, E., Esper, O., Friedrichsen, H., Lohmann, G. and Meyer, H. (2015). "The seasonal sea-ice zone in the glacial Southern Ocean as a carbon sink." Nature Communications **6**.

Benaglia, T., Chauveau, D., Hunter, D. R. and Young, D. S. (2009). "mixtools: An R Package for Analyzing Finite Mixture Models." Journal of Statistical Software **32**(6): 1-29.

Beszteri, B., Allen, C., Almandoz, G. O., Armand, L., Barcena, M. Á., Cantzler, H., Crosta, X., Esper, O., Jordan, R. W., Kauer, G., Klaas, C., Kloster, M., Leventer, A., Pike, J. and Rigual-Hernández, A. S. (submitted). "Quantitative comparison of taxa and taxon concepts in the diatom genus *Fragilariopsis*." Journal of Phycology.

Bradski, G. and Kaehler, A. (2008). Learning OpenCV : Computer Vision with the OpenCV Library. Sebastopol, O'Reilly.

Bresenham, J. E. (1965). "Algorithm for computer control of a digital plotter." IBM Systems journal **4**(1): 25-30.

Cefarelli, A. O., Ferrario, M. E., Almandoz, G. O., Atencio, A. G., Akselman, R. and Vernet, M. (2010). "Diversity of the diatom genus *Fragilariopsis* in the Argentine Sea and Antarctic waters: morphology, distribution and abundance." Polar biology **33**(11): 1463-1484.

Cortese, G. and Gersonde, R. (2007). "Morphometric variability in the diatom *Fragilariopsis kerguelensis*: Implications for Southern Ocean paleoceanography." Earth and Planetary Science Letters **257**(3-4): 526-544.

Cortese, G., Gersonde, R., Maschner, K. and Medley, P. (2012). "Glacial-interglacial size variability in the diatom *Fragilariopsis kerguelensis*: Possible iron/dust controls?" Paleoceanography **27**.

Crosta, X. (2009). "Holocene size variations in two diatom species off East Antarctica: Productivity vs environmental conditions." Deep Sea Research Part I: Oceanographic Research Papers **56**(11): 1983-1993.

Crosta, X., Sturm, A., Armand, L. and Pichon, J.-J. (2004). "Late Quaternary sea ice history in the Indian sector of the Southern Ocean as recorded by diatom assemblages." Marine Micropaleontology **50**(3): 209-223.

Donahue, J. G. (1970). Diatoms as Quaternary Biostratigraphic and Paleoclimatic Indicators in High Latitudes of the Pacific Ocean, Columbia University. **Ph. D.**

Droop, S. (1995). A morphometric and geographical analysis of two races of *Diploneis smithii*/*D. fusca* (Bacillariophyceae) in Britain. Proceedings of the 13th International Diatom Symposium. Biopress Ltd, Bristol.

Esper, O. and Gersonde, R. (2014a). "New tools for the reconstruction of Pleistocene Antarctic sea ice." Palaeogeography, Palaeoclimatology, Palaeoecology **399**: 260-283.

Esper, O. and Gersonde, R. (2014b). "Quaternary surface water temperature estimations: New diatom transfer functions for the Southern Ocean." Palaeogeography, Palaeoclimatology, Palaeoecology **414**: 1-19.

Evans, J. D. (1996). Straightforward statistics for the behavioral sciences, Brooks/Cole.

Fenner, J., Schrader, H. and Wienigk, H. (1976). "Diatom phytoplankton studies in the southern Pacific Ocean, composition and correlation to the Antarctic Convergence and its paleoecological significance." Initial Reports of the Deep Sea Drilling Project **35**: 757-813.

Frenguelli, J. (1960). "Diatomeas y silicoflagelados recogidas en Tierra Adelia durante las Expediciones Polares Francesas de Paul-Emile Victor (1950-1952)." Revue Algologique **1**: 3-48.

Gersonde, R. (2003). "Documentation of sediment core PS1768-8." Alfred Wegener Institute - Polarstern core repository.

Gersonde, R., Abelmann, A., Brathauer, U., Becquey, S., Bianchi, C., Cortese, G., Grobe, H., Kuhn, G., Niebler, H. S., Segl, M., Sieger, R., Zielinski, U. and Futterer, D. K. (2003). "Last glacial sea surface temperatures and sea-ice extent in the Southern Ocean (Atlantic-Indian sector): A multiproxy approach." Paleoceanography **18**(3).

Gersonde, R., Crosta, X., Abelmann, A. and Armand, L. (2005). "Sea-surface temperature and sea ice distribution of the Southern Ocean at the EPILOG Last Glacial Maximum—a circum-Antarctic view based on siliceous microfossil records." Quaternary Science Reviews **24**(7-9): 869-896.

Gersonde, R. and Hempel, G. (1990). Die Expeditionen ANTARKTIS-VIII/3 und VIII/4 mit FS "Polarstern" 1989 = The expeditions ANTARKTIS VIII/3 and VIII/4 of RV "Polarstern" in 1989. Bremerhaven, Alfred Wegener Institute for Polar and Marine Research. **74**.

Gersonde, R. and Zielinski, U. (2000). "The reconstruction of late Quaternary Antarctic sea-ice distribution—the use of diatoms as a proxy for sea-ice." Palaeogeography, Palaeoclimatology, Palaeoecology **162**(3-4): 263-286.

Hasle, G. R. (1965). "*Nitzschia* and *Fragilariopsis* species studied in the light and electron microscopes. III. The genus *Fragilariopsis*." Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo, I. Matematisk-Naturvidenskapelig Klasse. Ny Serie **21**: 1-49.

Kloster, M., Esper, O., Kauer, G. and Beszteri, B. (2017). "Large-Scale Permanent Slide Imaging and Image Analysis for Diatom Morphometrics." Applied Sciences **7**(4): 330.

Kloster, M., Kauer, G. and Beszteri, B. (2014). "SHERPA: an image segmentation and outline feature extraction tool for diatoms and other objects." BMC bioinformatics **15**: 218.

Nair, A., Mohan, R., Manoj, M. and Thamban, M. (2015). "Glacial-interglacial variability in diatom abundance and valve size: Implications for Southern Ocean paleoceanography." Paleoceanography **30**(10): 1245-1260.

R Core Team (2015). "R: a language and environment for statistical computing." from https://www.R-project.org

Rosin, P. L. (2003). "Measuring shape: ellipticity, rectangularity, and triangularity." Machine Vision and Applications **14**(3): 172-184.

Shukla, S. K., Crespin, J. and Crosta, X. (2016). "*Thalassiosira lentiginosa* size variation and associated biogenic silica burial in the Southern Ocean over the last 42kyrs." Marine Micropaleontology **127**: 74-85.

Shukla, S. K. and Crosta, X. (2017). "*Fragilariopsis kerguelensis* size variability from the Indian subtropical Southern Ocean over the last 42 000 years." Antarctic Science **29**(2): 139-146.

Shukla, S. K., Crosta, X., Cortese, G. and Nayak, G. N. (2013). "Climate mediated size variability of diatom *Fragilariopsis kerguelensis* in the Southern Ocean." Quaternary Science Reviews **69**: 49-58.

Shukla, S. K., Mohan, R. and Sudhakar, M. (2009). "Diatoms: a potential tool to understand past oceanographic settings." Current Science: 1726-1734.

Warnock, J. P., Scherer, R. P. and Konfirst, M. A. (2015). "A record of Pleistocene diatom preservation from the Amundsen Sea, West Antarctica with possible implications on silica leakage." Marine Micropaleontology **117**: 40-46.

Woodard, K., Kulichová, J., Poláčková, T. and Neustupa, J. (2016). "Morphometric allometry of representatives of three naviculoid genera throughout their life cycle." Diatom Research **31**(3): 231-242.

Zielinski, U. (1993). Quantitative Bestimmung von Paläoumweltparametern des Antarktischen Oberflächenwassers im Spätquartär anhand von Transferfunktionen mit Diatomeen = Quantitative estimation of palaeoenvironmental parameters of the Antarctic surface water in the Late Quaternary using transfer functions with diatoms, Alfred Wegener Institute for Polar and Marine Research.

Zielinski, U., Gersonde, R., Sieger, R. and Fütterer, D. (1998). "Quaternary surface water temperature estimations: Calibration of a diatom transfer function for the Southern Ocean." Paleoceanography **13**(4): 365-383.

3 Synthesis

Methods for automated identification and classification of light microscopic diatom images have been developed, tested and proposed already more than 15 years ago by the ADIAC project (du Buf and Bayer 2002). However, these methods did not become widespread, owing to a lack of ready-to-use equipment, as well as to deficits in usability of some of the methods, which possibly generated some reluctance in their application. As of nowadays, automated image processing has become a routine tool in many areas of research and industry. Nevertheless, before this work, no user friendly software was available which covered the broad range of methods required for diatom morphometry. I filled this gap with SHERPA, a tool for SHapE Recognition, Processing and Analysis, as described in **Publication I**. SHERPA was developed particularly for processing light microscopic images of oxidized slide-mounted diatom material, tackling limitations of this type of specimen (like poor contrast and contamination with debris) as well as making use of its peculiarities (like limited range of geometric shapes to investigate, and their convexity properties). These are, besides several image processing knacks, addressed by SHERPA's internal quality control system, which allows sorting out unwanted or flawed results to reduce user efforts. For downstream analysis, SHERPA calculates descriptors for ca. 30 morphometric features, which are exported, along with cut-outs of the valve images, into standardized image and data file formats. I successfully tested the applicability of SHERPA on two datasets, originating from different sources and thus manifesting distinct image properties. As a test case for material containing debris, broken valves and overlapping objects, I investigated material from a sediment core, imaged by the Metafer slide scanner by a high-resolution scan of a continuous slide area. By competitive application of multiple segmentation methods, SHERPA was mostly able to successfully segment even this complicated material. Comparing the different segmentation methods SHERPA offers, I found that differences they introduced in measurements of valve area, perimeter, width and height were mostly in the range of 1.5 % or less. For the second test case, from original image data (Mann et al. 2004a) showing complete and isolated valves without debris, we successfully recreated parts of the work (Mann et al. 2004b) did on differentiating six *Sellaphora pupula* demes by minute differences in their valve outline. Since SHERPA is not limited to processing diatom images, but can handle light microscopic images depicting structures with a more or less stable outline, it also has been proposed for analyzing phytoliths (Evett and Cuthrell 2016). These results demonstrate SHERPA's applicability for diatom morphometry, by massively reducing user efforts, by measuring a broad range of morphometric features, and by doing so at high precision, which, in combination with high optical resolution of input images, enables a high taxonomic resolution for downstream taxonomic classification.

The software SHERPA, which I developed during this work, enables for precise automated morphometric measurement of diatom valves from slide-mounted material.

For facilitating large-scale morphometric investigations, not only measurement, but also microscopic imaging needs to be automated. A two-cycle workflow, with the first cycle using a low-magnification objective for locating valves, and a second cycle for revisiting theses vales for imaging them with a high-magnification objective, has already been suggested by the ADIAC project more than 15 years ago (du Buf and Bayer 2002). However, due to technical limitations, they were not able to implement this procedure at that time. Nowadays, these limitations have been overcome by commercially

available slide scanning systems. Nevertheless, these systems do not offer functionality tailored to the need of diatom morphometry, and they are not designed to enable integration with self-developed software. For the Metafer slide scanner I established such an integration, which enables annotating object coordinates retrieved by SHERPA into so called “virtual slides”, to use those for setting up a list of positions to be revisited during the high-resolution imaging step. Even though this connection requires slight manual user interaction, it enables for imaging and measuring up to a few thousand valves per day, depending on specimen and preparation quality. Making possible the integration of these originally independent components, and adapting them to the specific question, required a lot of optimization for each of the workflow steps and their harmonization. With the multitude of possible settings pertaining to a large number of individual workflow steps, documentation of analyses became another issue to be automated. Our approach enables for automatically storing the imaging setup of the slide scanner as well as the image processing settings used in SHERPA, along with all the measurements and image data. This enables for later reanalysis of the images, and even re-imaging of the diatom valves, ensuring reproducibility and reusability of data, whilst the two-cycle workflow provides a good trade-off between manual effort, time consumption and data economy. I put the applicability of this workflow to the test by measuring the length of ca. 12,000 *F. kerguelensis* valves from sediment core material. This sampling size, which was orders of magnitude more extensive compared to conventional morphometric investigations, enabled for expanding the known size range for the species’ valve length substantially, by ca. 45 %, which shows the benefit of such a large-scale approach. The workflow, as well as the test case, has been published in **Publication II**.

The workflow I developed by combining a commercially available slide scanning microscope with SHERPA, facilitates measurement of ca. 30 morphometric features from up to a few thousand diatom valves per day. This approach provides a good trade-off between manual effort, time consumption and data economy, and it ensures reproducibility and reusability of data by automated documentation of imaging and analyses settings.

After this workflow was established, I applied it in a large-scale investigation, measuring ca. 29,000 *F. kerguelensis* valves, and looking for morphometric signals related to the diatom life cycle. Direct observations of the species in its natural habitat are not possible year-round in the pelagic Southern Ocean, thus only little is known about its seasonal development. One of very few possible remedies is sampling by moored or drifting sediment traps, which automatically collect material trickling down from the upper water column in a time-controlled manner. From such material, I traced signals represented by changes in valve length distributions, occurring on different time scales. The ratio of shorter and larger cells indicated three seasonal stages, with the population proliferating (mid-spring to mid-summer), ceasing (by declining environmental conditions or grazing, from mid-summer to mid-autumn) and stagnating (with a higher proportion of very large cells persisting over late autumn and winter). These phases occurred at similar, though not identical points in time during both observed years, and the pace and degree of changes between neighboring time points also differed between both years. The increased abundance of smaller cells during summer indicated conditions beneficial for sexual propagation. However, we did not observe clear size cohorts, which would indicate strong sexual events affecting a major part of the population. Instead, I found only a slightly elevated relative abundance of valves within the size range of initial cells, which I interpreted as

signal of low sexual activity. Since this occurred for prolonged periods at different times of both years, it is not clear if this signal was caused by events of *F. kerguelensis* life cycle. The signal could also result from the nature of sediment trap material, which usually does not give a very clear image of the processes in the upper water column. I nevertheless interpret these data as indication of asynchronous sexuality during periods of weak vegetative growth. Whilst changes in diatom cell size are usually interpreted exclusively in the frame of the life cycle, I also pointed out their interlinkage with physiological and ecological effects, which might contribute to an explanation of the patterns I observed in valve length distributions. This investigation, which will be published in **Publication III,** became feasible only by means of my automated high-throughput workflow, and helped elucidating the life cycle of *F. kerguelensis* in its natural habitat.

Unlike for my other investigations, oxidized diatom frustules were embedded not in Naphrax, but in Norland Optical Adhesive NOA 61. The lower refractive index of this mountant resulted in weaker contrast of siliceous structures, complicating automated segmentation and necessitating increased efforts in manual review and reworking.

As a first larger scale research project using the diatom imaging and morphometry workflow I developed, I analyzed 29,000 *F. kerguelensis* valves from a two-year sediment trap time series. I interpreted the observed changes in valve length distributions by relating them to life cycle events as well as to physiological and ecological effects of cell size.

The last extensive showcase of my workflow regarded *F. kerguelensis* morphometric changes on a geological time scale from a sediment record. Features like valve length, valve area, and costae distance in relation to the valve's aspect ratio (as the combined factor F or F*), have previously been proposed as paleoceanograpic proxies for summer sea surface temperature (SSST), iron availability and glacial / interglacial conditions. Whilst these proxies were based on the quantification of basic geometric valve properties, we scrutinized also more complex morphometric descriptors, measured from ca. 12,000 *F. kerguelensis* valves originating from sediment core samples. Doing so, we found that the formula used by (Shukla et al. 2013) overestimated the valve area by ca. 9 % on average, compared to actual measurements by SHERPA. For the morphometric descriptor rectangularity, I found two clearly distinct classes, which I termed different morphotypes. Whilst one of the morphotypes persisted throughout the samples, the other one occurred prominently only during glacials. If the second morphotype represents a cryptic taxon, or if it is a result of phenotypic plasticity, is not clear at the moment. However, both morphotypes contributed differently to means of morphometric features used or proposed as paleo-proxies. Accordingly, differentiating between the morphotypes might improve the accuracy of these proxies. I also found that the abundance ratio of both morphotypes correlated very strongly with the reconstructed SSST, making it a promising candidate for a paleo SSST proxy. These findings are ready to be published in **Publication IV**. The fact that the two morphotypes have not been recognized before, even though the species, as well as the material I used, has been investigated thoroughly before by highly skilled diatomists, illustrates the potential of my automated diatom morphometry workflow. To answer the question if the two morphotypes relate to cryptic taxa or to phenotypic plasticity needs further investigation of recent (for example phytoplankton net) samples and of *F. kerguelensis* cultures.

From sediment core material, I identified two previously unrecognized different morphotypes of *F. kerguelensis.* This finding not only has implications for existing paleo-proxies, but might also provide a new proxy for summer sea surface temperature. This study clearly underlines that large-scale imaging based methods not only have the potential to speed up tedious analyses, but can also lead to substantially novel scientific insights.

4 Summary

Since the ADIAC project, which ended more than 15 years ago, not much progress in automating morphometric analysis of diatoms from slide-mounted material has been published, and no ready-to-use system has become available. This thesis work is the first to implement such a system completely, covering all aspects of the underlying imaging and image processing pipeline, by combining a commercially available slide scanning microscope with my diatom morphometry software SHERPA. I was able to show the applicability as well as the potential of this approach by executing a series of smaller and two large-scale morphometry projects. The extensive sampling sizes, which were made possible only by the new workflow, enabled the first observations of life cycle related size distribution changes of *Fragilariopsis kerguelensis* in its natural habitat, leading to hypotheses on influences of reproduction, grazing and environmental changes in one of the most important diatom species of the Southern Ocean. In a second large-scale investigation, SHERPA's precise morphometric measurements revealed a second *F. kerguelensis* morphotype, which has not been recognized before, even though the species, as well as the very material I analyzed, have been investigated intensely before by experienced diatomists; a result not disqualifying their work, but rather underlining that explicit and precise quantification of morphological information has a strong potential to generate novel scientific insights. This new morphotype has implications on the utilization of paleo-proxies which are based on geometrical valve features of *F. kerguelensis*. Differentiating both morphotypes might improve established methods and possibly provides a new proxy for summer sea surface temperature.

As image processing and classification nowadays develop faster than ever before, further development will allow application of my workflow in automatic diatom identification, water quality monitoring, paleoceanography, paleolimnology and many other fields of diatom research.

5 Zusammenfassung

Seit das ADIAC Projekt vor mehr als 15 Jahren abgeschlossen wurde, sind nur wenige signifikante Fortschritte vorgestellt worden bezüglich der Automatisierung morphometrischer Analysen von Diatomeen, welche auf Objektträgern montiert sind, und gebrauchsfertige Systeme sind bisher nicht verfügbar. Diese Arbeit ist die erste, welche ein umfassendes Gesamtsystem implementiert das sämtliche Aspekte einer entsprechenden Bildaufnahme und -verarbeitungspipeline abdeckt; dies wurde erreicht durch die Verknüpfung eines kommerziellen Slide-Scanning-Mikroskops mit meiner Diatomeen-Morphometriesoftware SHERPA. Durch eine Reihe kleinerer und zwei großmaßstäbliche Morphometrieprojekte konnte ich sowohl die Eignung, als auch das Potential dieser Lösung zeigen. Die großen Stichprobenumfänge, welche erst durch dieses System ermöglicht wurden, gestatteten die ersten Beobachtungen von Änderungen der Größenverteilungen von *Fragilariopsis kerguelensis* in Abhängigkeit von ihrem Lebenszyklus, in ihrem natürlichen Lebensraum. Diese mündeten in Hypothesen hinsichtlich der Einflüsse von Reproduktion, Fressfeinden und Umweltveränderungen auf eine der wichtigsten Diatomeen-Arten des Südozeans. In einer zweiten großangelegten

Untersuchung führten die durch SHERPA ermöglichten präzisen morphometrischen Messungen zur Entdeckung eines zweiten *F. kerguelensis* Morphotypen, welcher bisher übersehen wurde, obwohl sowohl die Art, als auch eben jenes Material, das ich analysiert habe, bereits von erfahrenden Diatomologen ausgiebig untersucht wurde. Dieser Fund soll nicht die bisherigen Arbeiten disqualifizieren, sondern deutlich machen, dass die eindeutige und präzise Quantifizierung morphologischer Informationen ein großes Potential hinsichtlich der Erlangung neuer wissenschaftlicher Erkenntnisse birgt. Die Entdeckung des zweiten Morphotypen hat Auswirkungen auf die Anwendung von Paläo-Proxies, denen geometrische Eigenschaften der Valvenform von *F. kerguelensis* zugrunde liegen. Beide Morphotypen zu differenzieren könnte die etablierten Methoden verbessern, und möglicherweise sogar einen neuen Proxy für die Oberflächentemperatur des Meereswassers im Sommer bieten.

Die immer schneller fortschreitende Entwicklung von Methoden zur Bildverarbeitung und -klassifizierung wird die Verwendung meines Workflows im Bereich der automatischen Identifizierung von Diatomeen, der Wasserqualitätsüberwachung, der Paläo-Ozeanographie und -Limnologie, sowie vielen weiteren Feldern der Diatomeenforschung ermöglichen.

6 Abbreviations

ACC, Antarctic Circumpolar Current; ADIAC, Automated Diatom Identification And Classification; AIC, Akaike information criterion; APF, Antarctic Polar Front; BIC, Bayesian information criterion; $BSiO_2$, biogenic silica; CDF, convexity defection factor; CDMs, convexity defect measures; CHMDF, convex hull maximum distance factor; CSV, comma separated values; EFDs, elliptic Fourier descriptors; EFDIs, elliptic Fourier descriptor invariants; EIFEX, European Iron Fertilization Experiment; HLNC, high-nutrient low-chlorophyll; λ, mixing proportion; LED, light emitting diode; μ, mean; MIS, marine isotope stadium; n, sampling size; N, population size; *p*, p-value; PCAF, percent concave area fraction; PCC, Pearson correlation coefficient; PFZ, Polar Frontal Zone; POC, particulate organic carbon; R^2, square of PCC; RATS, robust automated threshold selector; RI, ranking index; RFI, relative frequency of valves within the initial cell size range; σ, standard deviation; SHERPA, tool for SHapE Recognition, Processing and Analysis; SSST, summer sea surface temperature; SST, sea surface temperature; TIFF, tagged image file format; XML, extensible markup language

7 References

Abarca, N., Jahn, R., Zimmermann, J. and Enke, N. (2014). "Does the cosmopolitan diatom *Gomphonema parvulum* (Kützing) Kützing have a biogeography?" Plos One **9**(1): e86885.

Abelmann, A. and Gersonde, R. (1991). "Biosiliceous Particle-Flux in the Southern-Ocean." Marine Chemistry **35**(1-4): 503-536.

Abelmann, A., Gersonde, R., Cortese, G., Kuhn, G. and Smetacek, V. (2006). "Extensive phytoplankton blooms in the Atlantic sector of the glacial Southern Ocean." Paleoceanography **21**(1).

Abelmann, A., Gersonde, R., Knorr, G., Zhang, X., Chapligin, B., Maier, E., Esper, O., Friedrichsen, H., Lohmann, G. and Meyer, H. (2015). "The seasonal sea-ice zone in the glacial Southern Ocean as a carbon sink." Nature Communications **6**.

Almeida, S. F., Elias, C., Ferreira, J., Tornes, E., Puccinelli, C., Delmas, F., Dorflinger, G., Urbanic, G., Marcheggiani, S., Rosebery, J., Mancini, L. and Sabater, S. (2014). "Water quality assessment of rivers using diatom metrics across Mediterranean Europe: a methods intercalibration exercise." Science of the Total Environment **476-477**: 768-776.

Álvarez-Borrego, J. and Solorza, S. (2010). "Comparative analysis of several digital methods to recognize diatoms." Hidrobiológica **20**: 158-170.

Álvarez, E., López-Urrutia, Á. and Nogueira, E. (2012). "Improvement of plankton biovolume estimates derived from image-based automatic sampling devices: application to FlowCAM." Journal of Plankton Research **34**(6): 454-469.

Álvarez, E., López-Urrutia, Á., Nogueira, E. and Fraga, S. (2011). "How to effectively sample the plankton size spectrum? A case study using FlowCAM." Journal of Plankton Research **33**(7): 1119-1133.

Anglès, S., Jordi, A. and Campbell, L. (2015). "Responses of the coastal phytoplankton community to tropical cyclones revealed by high-frequency imaging flow cytometry." Limnology and Oceanography **60**(5): 1562-1576.

Assmy, P. (2004). Temporal development and vertical distribution of major components of the plankton assemblage during an iron fertilization experiment in the Antarctic Polar Frontal Zone, Universität Bremen.

Assmy, P., Henjes, J., Smetacek, V. and Montresor, M. (2006). "Auxospore formation by the silica-sinking, oceanic diatom *Fragilariopsis kerguelensis* (Bacillariophyceae)." Journal of Phycology **42**(5): 1002-1006.

Assmy, P., Smetacek, V., Montresor, M., Klaas, C., Henjes, J., Strass, V. H., Arrieta, J. M., Bathmann, U., Berg, G. M., Breitbarth, E., Cisewski, B., Friedrichs, L., Fuchs, N., Herndl, G. J., Jansen, S., Krägefsky, S., Latasa, M., Peeken, I., Röttgers, R., Scharek, R., Schüller, S. E., Steigenberger, S., Webb, A. and Wolf-Gladrow, D. (2013). "Thick-shelled, grazer-protected diatoms decouple ocean carbon and silicon cycles in the iron-limited Antarctic Circumpolar Current." Proceedings of the National Academy of Sciences **110**(51): 20633-20638.

Barber, H. G. and Haworth, E. Y. (1981). A guide to the morphology of the diatom frustule. Ambleside, Cumbria, UK, Freshwater Biological Association.

Bayer, M., Pullan, M., Droop, S., Mann, D., Juggins, S., Lewis, M., Clarke, A., Head, R. and Weckström, K. (1999). "ADIAC Diatom Image Database." from http://rbg-web2.rbge.org.uk/ADIAC/db/adiacdb.htm

Benaglia, T., Chauveau, D., Hunter, D. R. and Young, D. S. (2009). "mixtools: An R Package for Analyzing Finite Mixture Models." Journal of Statistical Software **32**(6): 1-29.

Beszteri, B. (2005). Morphometric and molecular investigations of species limits in *Cyclotella meneghiniana* (Bacillariophyceae) and closely related species, Universität Bremen. **Dr. rer. nat.**

Beszteri, B., Ács, É. and Medlin, L. (2005). "Conventional and geometric morphometric studies of valve ultrastructural variation in two closely related *Cyclotella* species (Bacillariophyta)." European Journal of Phycology **40**(1): 89-103.

Beszteri, B., Allen, C., Almandoz, G. O., Armand, L., Barcena, M. Á., Cantzler, H., Crosta, X., Esper, O., Jordan, R. W., Kauer, G., Klaas, C., Kloster, M., Leventer, A., Pike, J. and Rigual-Hernández, A. S. (submitted). "Quantitative comparison of taxa and taxon concepts in the diatom genus *Fragilariopsis*." Journal of Phycology.

Beszteri, B., Hinz, F. and Crawford, R. M. "Hustedt Diatom Collection Database." Retrieved 18 Dec 2017, from http://hustedt.awi.de/

Boyd, P. W., Crossley, A. C., DiTullio, G. R., Griffiths, F. B., Hutchins, D. A., Queguiner, B., Sedwick, P. N. and Trull, T. W. (2001). "Control of phytoplankton growth by iron supply and irradiance in the subantarctic Southern Ocean: Experimental results from the SAZ Project." Journal of Geophysical Research: Oceans **106**(C12): 31573-31583.

Boyd, P. W., Jickells, T., Law, C., Blain, S., Boyle, E., Buesseler, K., Coale, K., Cullen, J., De Baar, H. J. and Follows, M. (2007). "Mesoscale iron enrichment experiments 1993-2005: Synthesis and future directions." Science **315**(5812): 612-617.

Bradski, G. and Kaehler, A. (2008). Learning OpenCV : Computer Vision with the OpenCV Library. Sebastopol, O'Reilly.

Bresenham, J. E. (1965). "Algorithm for computer control of a digital plotter." IBM Systems journal **4**(1): 25-30.

Briseño-Avena, C., Roberts, P. L. D., Franks, P. J. S. and Jaffe, J. S. (2015). "ZOOPS- O2: A broadband echosounder with coordinated stereo optical imaging for observing plankton in situ." Methods in Oceanography **12**: 36-54.

Bueno, G., Deniz, O., Pedraza, A., Ruiz-Santaquiteria, J., Salido, J., Cristóbal, G., Borrego-Ramos, M. and Blanco, S. (2017). "Automated Diatom Classification (Part A): Handcrafted Feature Approaches." Applied Sciences **7**(8): 753.

Cairns, J. J., Dickson, K. L., Slocomb, J. P., Almeida, S. P., Eu, J. K. T., Liu, C. Y. C. and Smith, H. F. (1974). Microcosm pollution monitoring, ; Virginia Polytechnic Inst. and State Univ., Blacksburg.

Canny, J. (1986). "A Computational Approach to Edge Detection." IEEE Transactions on Pattern Analysis and Machine Intelligence **30**: 125-147.

Cefarelli, A. O., Ferrario, M. E., Almandoz, G. O., Atencio, A. G., Akselman, R. and Vernet, M. (2010). "Diversity of the diatom genus *Fragilariopsis* in the Argentine Sea and Antarctic waters: morphology, distribution and abundance." Polar biology **33**(11): 1463-1484.

Chepurnov, V. A., Mann, D. G., Sabbe, K., Vannerum, K., Casteleyn, G., Verleyen, E., Peperzak, L. and Vyverman, W. (2005). "Sexual reproduction, mating system, chloroplast dynamics and abrupt cell size reduction in *Pseudo-nitzschia pungens* from the North Sea (Bacillariophyta)." European Journal of Phycology **40**(4): 379-395.

Chepurnov, V. A., Mann, D. G., Sabbe, K. and Vyverman, W. (2004). "Experimental studies on sexual reproduction in diatoms." International review of cytology **237**: 91-154.

Claude, J. (2008). Morphometrics with R. New York, Springer Science+Business Media, LLC.

Claude, J. (2010). "Morphometrics with R - Errata 1.81" from http://www.isem.univ-montp2.fr/recherche/files/2012/01/Morphometrics_errata1.81.pdf

Closset, I., Cardinal, D., Bray, S. G., Thil, F., Djouraev, I., Rigual-Hernández, A. S. and Trull, T. W. (2015). "Seasonal variations, origin, and fate of settling diatoms in the Southern Ocean tracked by silicon isotope records in deep sediment traps." Global Biogeochemical Cycles **29**(9): 1495-1510.

Combes, H. J., de la Rocha, C. L., Esper, O., Abelmann, A., Gersonde, R. and Shemesh, A. (2007). "Diatom delta C-13, delta N-15, and C/N since the Last Glacial maximum in the Southern Ocean: Evidence for regional and ecological influences." Geochimica Et Cosmochimica Acta **71**(15): A436-A436.

Cortese, G., Abelmann, A. and Gersonde, R. (2007). "The last five glacial-interglacial transitions: A high-resolution 450,000-year record from the subantarctic Atlantic." Paleoceanography **22**(4).

Cortese, G. and Gersonde, R. (2007). "Morphometric variability in the diatom *Fragilariopsis kerguelensis*: Implications for Southern Ocean paleoceanography." Earth and Planetary Science Letters **257**(3-4): 526-544.

Cortese, G. and Gersonde, R. (2008). "Plio/Pleistocene changes in the main biogenic silica carrier in the Southern Ocean, Atlantic Sector." Marine Geology **252**(3-4): 100-110.

Cortese, G., Gersonde, R., Maschner, K. and Medley, P. (2012). "Glacial-interglacial size variability in the diatom *Fragilariopsis kerguelensis*: Possible iron/dust controls?" Paleoceanography **27**.

Crawford, R. M., Hinz, F. and Rynearson, T. (1997). "Spatial and temporal distribution of assemblages of the diatom *Corethron criophilum* in the Polar Frontal region of the South Atlantic." Deep Sea Research Part II: Topical Studies in Oceanography **44**(1): 479-496.

Crosta, X. (2009). "Holocene size variations in two diatom species off East Antarctica: Productivity vs environmental conditions." Deep Sea Research Part I: Oceanographic Research Papers **56**(11): 1983-1993.

Crosta, X., Romero, O., Armand, L. K. and Pichon, J.-J. (2005). "The biogeography of major diatom taxa in Southern Ocean sediments: 2. Open ocean related species." Palaeogeography, Palaeoclimatology, Palaeoecology **223**(1–2): 66-92.

Crosta, X., Sturm, A., Armand, L. and Pichon, J.-J. (2004). "Late Quaternary sea ice history in the Indian sector of the Southern Ocean as recorded by diatom assemblages." Marine Micropaleontology **50**(3): 209-223.

D'Alelio, D., d'Alcala, M. R., Dubroca, L., Sarn, D., Zingone, A. and Montresor, M. (2010). "The time for sex: A biennial life cycle in a marine planktonic diatom." Limnology and Oceanography **55**(1): 106-114.

Davidovich, N. (2001). Species specific sizes and size range of sexual reproduction in diatoms. Proceedings of the 16th international diatom symposium, University of Athens, Greece.

de Salas, M. F., Eriksen, R., Davidson, A. T. and Wright, S. W. (2011). "Protistan communities in the Australian sector of the Sub-Antarctic Zone during SAZ-Sense." Deep Sea Research Part II: Topical Studies in Oceanography **58**(21–22): 2135-2149.

Defelice, D. R. and Wise, S. W. (1981). "Surface lithofacies, biofacies, and diatom diversity patterns as models for delineation of climatic change in the southeast Atlantic Ocean." Marine Micropaleontology **6**(1): 29-70.

Denman, K. L. (2008). "Climate change, ocean processes and ocean iron fertilization." Marine Ecology Progress Series **364**: 219-225.

Donahue, J. G. (1970). Diatoms as Quaternary Biostratigraphic and Paleoclimatic Indicators in High Latitudes of the Pacific Ocean, Columbia University. **Ph. D.**

Drebes, G. (1977). Sexuality [The biology of diatoms: chapter 9]. Oxford, Blackwell. **Botanical Monographs; vol. 13:** 498.

Droop, S. (1995). A morphometric and geographical analysis of two races of *Diploneis smithii*/*D. fusca* (Bacillariophyceae) in Britain. Proceedings of the 13th International Diatom Symposium. Biopress Ltd, Bristol.

Droop, S. J., Mann, D. G. and Lokhorst, G. M. (2000). "Spatial and temporal stability of demes in *Diploneis smithii*/*D*. *fusca* (Bacillariophyta) supports a narrow species concept." Phycologia **39**(6): 527-546.

du Buf, H. (2003). "ADIAC public image files" Retrieved 07.02.2013, from http://rbg-web2.rbge.org.uk/ADIAC/pubdat/downloads/public_images.htm

du Buf, H. and Bayer, M. M. (2002). Automatic Diatom Identification. New Jersey, London, Singapore, Hong Kong, World Scientific Publishing Co. Pte. Ltd.

Ebersbach, F., Trull, T. W., Davies, D. M. and Bray, S. G. (2011). "Controls on mesopelagic particle fluxes in the Sub-Antarctic and Polar Frontal Zones in the Southern Ocean south of Australia in summer—Perspectives from free-drifting sediment traps." Deep Sea Research Part II: Topical Studies in Oceanography **58**(21): 2260-2276.

Edlund, M. B. and Stoermer, E. F. (1997). "Ecological, evolutionary, and systematic significance of diatom life histories." Journal of Phycology **33**(6): 897-918.

Emgu (2012). "Emgu CV, a cross platform .NET wrapper for the OpenCV image processing library, Version 2.4.2" from
http://www.emgu.com

Erickson, D. J., Hernandez, J. L., Ginoux, P., Gregg, W., McClain, C. and Christian, J. (2003). "Atmospheric iron delivery and surface ocean biological activity in the Southern Ocean and Patagonian region." Geophysical Research Letters **30**(12).

Eriksen, R., Trull, T., Davies, D., Jansen, P., Davidson, A., Westwood, K. and Van den Enden, R. (in press). "Seasonal succession of phytoplankton community structure from autonomous sampling at the Australian Southern Ocean Time Series site." Marine Ecology Progress Series.

Esper, O. and Gersonde, R. (2014a). "New tools for the reconstruction of Pleistocene Antarctic sea ice." Palaeogeography, Palaeoclimatology, Palaeoecology **399**: 260-283.

Esper, O. and Gersonde, R. (2014b). "Quaternary surface water temperature estimations: New diatom transfer functions for the Southern Ocean." Palaeogeography, Palaeoclimatology, Palaeoecology **414**: 1-19.

Evans, J. D. (1996). Straightforward statistics for the behavioral sciences, Brooks/Cole.

Evett, R. R. and Cuthrell, R. Q. (2016). "A conceptual framework for a computer-assisted, morphometric-based phytolith analysis and classification system." Journal of Archaeological Science **68**(Supplement C): 70-78.

Falasco, E., Blanco, S., Bona, F., Goma, J., Hlubikova, D., Novais, M. H., Hoffmann, L. and Ector, L. (2009). "Taxonomy, morphology and distribution of the *Sellaphora stroemii* complex (Bacillariophyceae)." Fottea **9**.

Fenner, J., Schrader, H. and Wienigk, H. (1976). "Diatom phytoplankton studies in the southern Pacific Ocean, composition and correlation to the Antarctic Convergence and its paleoecological significance." Initial Reports of the Deep Sea Drilling Project **35**: 757-813.

Frankova, M., Poulickova, A., Neustupa, J., Pichrtova, M. and Marvan, P. (2009). "Geometric morphometrics - a sensitive method to distinguish diatom morphospecies: a case study on the sympatric populations of *Reimeria sinuata* and *Gomphonema tergestinum* (Bacillariophyceae) from the River Becva Czech Republic." Nova Hedwigia **88**.

Frankovich, T. A., Sullivan, M. J. and Stacy, N. (2015). "Three new species of *Tursiocola* (Bacillariophyta) from the skin of the West Indian manatee (*Trichechus manatus*)." Phytotaxa **204**(1): 33-48.

Frenguelli, J. (1960). "Diatomeas y silicoflagelados recogidas en Tierra Adelia durante las Expediciones Polares Francesas de Paul-Emile Victor (1950-1952)." Revue Algologique **1**: 3-48.

Fuchs, N., Scalco, E., Kooistra, W. H. C. F., Assmy, P. and Montresor, M. (2013). "Genetic characterization and life cycle of the diatom *Fragilariopsis kerguelensis*." European Journal of Phycology **48**(4): 411-426.

Gao, Y., Kaufman, Y., Tanre, D., Kolber, D. and Falkowski, P. (2001). "Seasonal distributions of aeolian iron fluxes to the global ocean." Geophysical Research Letters **28**(1): 29-32.

Geibert, W., Rutgers van der Loeff, M. M., Usbeck, R., Gersonde, R., Kuhn, G. and Seeberg-Elverfeldt, J. (2005). "Quantifying the opal belt in the Atlantic and southeast Pacific sector of the Southern Ocean by means of 230Th normalization." Global Biogeochemical Cycles **19**(4): n/a-n/a.

Geider, R. J., Platt, T. and Raven, J. A. (1986). "Size dependence of growth and photosynthesis in diatoms: a synthesis." Marine Ecology Progress Series: 93-104.

Genkal, S. and Kuz'min, G. (1979). "Biometric analysis of main structural elements of valves in the species of genus *Stephanodiscus* Ehr.(Bacillariophyta)." Botanicheskii zhurnal.

Gersonde, R. (2003). "Documentation of sediment core PS1768-8." Alfred Wegener Institute - Polarstern core repository.

Gersonde, R., Abelmann, A., Brathauer, U., Becquey, S., Bianchi, C., Cortese, G., Grobe, H., Kuhn, G., Niebler, H. S., Segl, M., Sieger, R., Zielinski, U. and Futterer, D. K. (2003). "Last glacial sea surface temperatures and sea-ice extent in the Southern Ocean (Atlantic-Indian sector): A multiproxy approach." Paleoceanography **18**(3).

Gersonde, R., Crosta, X., Abelmann, A. and Armand, L. (2005). "Sea-surface temperature and sea ice distribution of the Southern Ocean at the EPILOG Last Glacial Maximum—a circum-Antarctic view based on siliceous microfossil records." Quaternary Science Reviews **24**(7-9): 869-896.

Gersonde, R. and Hempel, G. (1990). Die Expeditionen ANTARKTIS-VIII/3 und VIII/4 mit FS "Polarstern" 1989 = The expeditions ANTARKTIS VIII/3 and VIII/4 of RV "Polarstern" in 1989. Bremerhaven, Alfred Wegener Institute for Polar and Marine Research. **74**.

Gersonde, R. and Zielinski, U. (2000). "The reconstruction of late Quaternary Antarctic sea-ice distribution—the use of diatoms as a proxy for sea-ice." Palaeogeography, Palaeoclimatology, Palaeoecology **162**(3-4): 263-286.

Girault, F. E. (2010). Size and textural patterns in centric diatoms since the Miocene. ETH ZURICH, Doctor of Sciences.

Gonzalez, R. C. and Woods, R. E. (2007). Digital Image Processing, 3rd Ed. Upper Saddle River, New Jersey, Prentice Hall.

Grima, C., Tadeo, F., Álvarez, T. and Arribas, J. L. (2003). Diatoms classification using frequency domain techniques. Jornadas de Automática. León.

Grover, J. P. (1991). "Resource competition in a variable environment: phytoplankton growing according to the variable-internal-stores model." The American Naturalist **138**(4): 811-835.

Hamm, C. E., Merkel, R., Springer, O., Jurkojc, P., Maier, C., Prechtel, K. and Smetacek, V. (2003). "Architecture and material properties of diatom shells provide effective mechanical protection." Nature **421**(6925): 841-843.

Hammer, Ø. (2013). "PAST" from
http://folk.uio.no/ohammer/past/

Hart, T. J. (1942). "Phytoplankton periodicity in Antarctic surface waters." Discovery reports **8**: 261-356.

Hasle, G. R. (1965). "*Nitzschia* and *Fragilariopsis* species studied in the light and electron microscopes. III. The genus *Fragilariopsis*." Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo, I. Matematisk-Naturvidenskapelig Klasse. Ny Serie **21**: 1-49.

Heiden, H. and Kolbe, R. W. (1928). Die marinen Diatomeen der Deutschen Südpolar-Expedition 1901–1903. Berlin, W. de Gruyter.

Hinz, F., Crawford, R. M., Fenner, J. and Beszteri, B. (2015). "Scanning electron microscopic investigation of *Vanhoeffenus antarcticus* Heiden from type material." Diatom Research **30**(1): 55-64.

Hoffmann, L., Peeken, I. and Lochte, K. (2007). "Effects of iron on the elemental stoichiometry during EIFEX and in the diatoms *Fragilariopsis kerguelensis* and *Chaetoceros dichaeta*." Biogeosciences **4**(4): 569-579.

Hofmann, E. E. (1985). "The large-scale horizontal structure of the Antarctic Circumpolar Current from FGGE drifters." Journal of Geophysical Research: Oceans **90**(C4): 7087-7097.

Hu, M. K. (1962). "Visual Pattern Recognition by Moment Invariants." In IRE Transactions on Information Theory **8**.

IBM Analytics (2013). "SPSS" from
http://www-01.ibm.com/software/analytics/spss/

itseez (2012). "OpenCV (Open Source Computer Vision Library) Version 2.4.2" from
http://opencv.org/

Jahn, R. and Kusber, W.-H. "AlgaTerra Information System" Retrieved 18 Dec 2017, from
http://www.algaterra.org

Jansen, S. and Bathmann, U. (2007). "Algae viability within copepod faecal pellets: evidence from microscopic examinations." Marine Ecology Progress Series **337**: 145-153.

Jewson, D. H., Granin, N. G., Zhdanov, A. A., Gorbunova, L. A., Bondarenko, N. A. and Gnatovsky, R. Y. (2008). "Resting stages and ecology of the planktonic diatom *Aulacoseira skvortzowii* in Lake Baikal." Limnology and Oceanography **53**(3): 1125-1136.

Jewson, D. H., Granin, N. G., Zhdarnov, A. A., Gorbunova, L. A. and Gnatovsky, R. Y. (2010). "Vertical mixing, size change and resting stage formation of the planktonic diatom *Aulacoseira baicalensis*." Eur. J. Phycol. **45**(4): 354-364.

Johnston, E. and Stoermer, E. (1976). "Computer analysis of phytoplankton cell images." Microscope **v. Second Quarter 1976, 24**.

Kermarrec, L., Bouchez, A., Rimet, F. and Humbert, J. F. (2013). "First evidence of the existence of semi-cryptic species and of a phylogeographic structure in the *Gomphonema parvulum* (Kützing) Kützing Complex (Bacillariophyta)." Protist **164**.

Kienast, S. S., Kienast, M., Jaccard, S., Calvert, S. E. and François, R. (2006). "Testing the silica leakage hypothesis with sedimentary opal records from the eastern equatorial Pacific over the last 150 kyrs." Geophysical Research Letters **33**(15): n/a-n/a.

Kingston, J. and Pappas, J. L. (2009). "Quantitative shape analysis as a diagnostic and prescriptive tool in determining *Fragilariforma* (Bacillariophyta) taxon status."

Kitware and al., e. (2012). " Insight Segmentation and Registration Toolkit (ITK), Version 4.20" from
http://www.itk.org/

Klingenberg, C. P. (2011). "MorphoJ: an integrated software package for geometric morphometrics." Mol Ecol Resour **11**.

Kloster, M. (2013). Digitale Bildsignalverarbeitung in der Bioinformatik: Methoden zur Segmentierung und Klassifizierung biologischer Merkmale am Beispiel ausgewählter Diatomeen. Emden, University of Applied Sciences Emden/Leer. **Master of Engineering**.

Kloster, M., Esper, O., Kauer, G. and Beszteri, B. (2017). "Large-Scale Permanent Slide Imaging and Image Analysis for Diatom Morphometrics." Applied Sciences **7**(4): 330.

Kloster, M., Kauer, G. and Beszteri, B. (2014). "SHERPA: an image segmentation and outline feature extraction tool for diatoms and other objects." BMC bioinformatics **15**: 218.

Kociolek, J. P., Balasubramanian, K., Blanco, S., Coste, M., Ector, L., Liu, Y., Kulikovskiy, M., Lundholm, N., Ludwig, T., Potapova, M., Rimet, F., Sabbe, K., Sala, S., Sar, E., Taylor, J., Van de Vijver, B., Wetzel, C. E., Williams, D. M., Witkowski, A. and Witkowski, J. (2017). "DiatomBase" Retrieved 12/11/2017, from
http://www.diatombase.org/

Kopczynska, E. E., Dehairs, F., Elskens, M. and Wright, S. (2001). "Phytoplankton and microzooplankton variability between the Subtropical and Polar Fronts south of Australia: Thriving under regenerative and new production in late summer." Journal of Geophysical Research: Oceans **106**(C12): 31597-31609.

Kozlova, O. g. G. (1966). Diatoms of the Indian and Pacific Sectors of the Antarctic, Israel Program for Scientific Translations;[available from the US Clearinghouse for Federal Scientific and Technical Information, Springfield, Va.].

Kühn, S., Medlin, L. and Eller, G. (2004). "Phylogenetic position of the parasitoid nanoflagellate *Pirsonia* inferred from nuclear-encoded small subunit ribosomal DNA and a description of *Pseudopirsonia* n. gen. and *Pseudopirsonia mucosa* (Drebes) comb. nov." Protist **155**(2): 143-156.

Kumar, M. S. R., Ramaiah, N. and Tang, D. (2009). "Morphometry and cell volumes of diatoms from a tropical estuary of India." Indian Journal of Marine Sciences **38**.

Laney, S. R., Olson, R. J. and Sosik, H. M. (2012). "Diatoms favor their younger daughters." Limnology and Oceanography **57**(5): 1572-1578.

Lannuzel, D., Bowie, A. R., Remenyi, T., Lam, P., Townsend, A., Ibisanmi, E., Butler, E., Wagener, T. and Schoemann, V. (2011). "Distributions of dissolved and particulate iron in the sub-Antarctic and Polar Frontal Southern Ocean (Australian sector)." Deep Sea Research Part II: Topical Studies in Oceanography **58**(21–22): 2094-2112.

Lehmann, G. (2006). "Robust Automatic Threshold Selection." The Insight Journal **2006** (July - December).

Litchman, E., Klausmeier, C. and Yoshiyama, K. (2009). "Contrasting size evolution in marine and freshwater diatoms." Proceedings of the National Academy of Sciences **106**(8): 2665-2670.

Loke, R. E. and Du Buf, H. (2002). Identification by curvature of convex and concave segments. Automatic diatom identification. H. du Buf and M. M. Bayer. Singapore, World Scientific Publishing.

Lundholm, N., Daugbjerg, N. and Moestrup, Ø. (2002). "Phylogeny of the Bacillariaceae with emphasis on the genus *Pseudo-nitzschia* (Bacillariophyceae) based on partial LSU rDNA." European Journal of Phycology **37**(1): 115-134.

Luo, Q., Gao, Y., Luo, J., Chen, C., Liang, J. and Yang, C. (2011). "Automatic Identification of Diatoms with Circular Shape using Texture Analysis." Journal of Software **6**(3): 428-435.

Macdonald, J. D. (1869). "On the structure of the Diatomaceous frustule, and its genetic cycle." Journal of Natural History **3**(13): 1-8.

Mann, D. G. (1999). "The species concept in diatoms." Phycologia **38**(6): 437-495.

Mann, D. G., McDonald, S. M., Bayer, M. M., Droop, S. J. M., Chepurnov, V. A., Loke, R. E., Ciobanu, A. and Buf, J. M. H. d. (2004a). "Algae World: Mann et al. 2004: images and morphometric data" from http://rbg-web2.rbge.org.uk/algae/research/mann_etal_2004_data.html

Mann, D. G., McDonald, S. M., Bayer, M. M., Droop, S. J. M., Chepurnov, V. A., Loke, R. E., Ciobanu, A. and du Buf, J. M. H. (2004b). "The *Sellaphora pupula* species complex (Bacillariophyceae): morphometric analysis, ultrastructure and mating data provide evidence for five new species." Phycologia **43**(4): 459-482.

Mann, D. G. and Vanormelingen, P. (2013). "An inordinate fondness? The number, distributions, and origins of diatom species." Journal of Eukaryotic Microbiology **60**(4): 414-420.

Marchetti, A. and Cassar, N. (2009). "Diatom elemental and morphological changes in response to iron limitation: a brief review with potential paleoceanographic applications." Geobiology **7**(4): 419-431.

Marshall, D., Martin, R., Rosin, P., Hicks, J., Mann, D. G., Droop, S. J. M. and Bayer, M. M. "DIADIST: Diatom and desmid identification by shape and texture." from http://www.cs.cf.ac.uk/diadist/code.htm

Marshall, J. and Speer, K. (2012). "Closure of the meridional overturning circulation through Southern Ocean upwelling." Nature Geoscience **5**(3): 171-180.

Martin, J. H., Fitzwater, S. E. and Gordon, R. M. (1990). "Iron deficiency limits phytoplankton growth in Antarctic waters." Global Biogeochemical Cycles **4**(1): 5-12.

Meyer, M. A. and El-Sayed, S. Z. (1983). "Grazing of *Euphausia superba* Dana on natural phytoplankton populations." Polar biology **1**(4): 193-197.

Microsoft Corporation (2013a). "Download Microsoft .NET Framework 4 (Web Installer) from Official Microsoft Download Center." from http://www.microsoft.com/en-us/download/details.aspx?id=17851

Microsoft Corporation (2013b). "Download Microsoft Visual C++ 2010 SP1 Redistributable Package (x64) from Official Microsoft Download Center." from http://www.microsoft.com/en-us/download/details.aspx?id=13523

Moeys, S., Frenkel, J., Lembke, C., Gillard, J. T., Devos, V., Van den Berge, K., Bouillon, B., Huysman, M. J., De Decker, S. and Scharf, J. (2016). "A sex-inducing pheromone triggers cell cycle arrest and mate attraction in the diatom *Seminavis robusta*." Scientific reports **6**.

Moore, E. R., Bullington, B. S., Weisberg, A. J., Jiang, Y., Chang, J. and Halsey, K. H. (2017). "Morphological and transcriptomic evidence for ammonium induction of sexual reproduction in *Thalassiosira pseudonana* and other centric diatoms." Plos One **12**(7): e0181098.

Moore, J. K., Abbott, M. R. and Richman, J. G. (1999). "Location and dynamics of the Antarctic Polar Front from satellite sea surface temperature data." Journal of Geophysical Research: Oceans **104**(C2): 3059-3073.

Mou, D. and Stoermer, E. F. (1992). "Separating *Tabellaria* (Bacillariophyceae) shape groups based on Fourier descriptors." Journal of Phycology **28**(3): 386-395.

Mullen, A. D., Treibitz, T., Roberts, P. L. D., Kelly, E. L. A., Horwitz, R., Smith, J. E. and Jaffe, J. S. (2016). "Underwater microscopy for in situ studies of benthic ecosystems." Nature Communications **7**: 12093.

Nafe, R. and Schlote, W. (2002). "Methods of Shape Analysis of two-dimensional closed Contours - A biologically important, but widely neglected Field in Histopathology." Electronic Journal of Pathology and Histology **8**(2): 1-18.

Nair, A., Mohan, R., Manoj, M. and Thamban, M. (2015). "Glacial-interglacial variability in diatom abundance and valve size: Implications for Southern Ocean paleoceanography." Paleoceanography **30**(10): 1245-1260.

Naselli-Flores, L. and Barone, R. (2003). Steady-state assemblages in a Mediterranean hypertrophic reservoir. The role of *Microcystis* ecomorphological variability in maintaining an apparent equilibrium. Phytoplankton and Equilibrium Concept: The Ecology of Steady-State Assemblages, Springer**:** 133-143.

Nelson, D. M., Tréguer, P., Brzezinski, M. A., Leynaert, A. and Quéguiner, B. (1995). "Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation." Global Biogeochemical Cycles **9**(3): 359-372.

Norton, T. A., Melkonian, M. and Andersen, R. A. (1996). "Algal biodiversity." Phycologia **35**(4): 308-326.

Nowlin, W. D. J. and Clifford, M. (1982). "The kinematic and thermohaline zonation of the Antarctic Circumpolar Current at Drake Passage." Journal of Marine Research **40**: 481-507.

Olenici, A., Blanco, S., Borrego-Ramos, M., Momeu, L. and Baciu, C. (2017). "Exploring the effects of acid mine drainage on diatom teratology using geometric morphometry." Ecotoxicology **26**(8): 1018-1030.

Olson, R. J. and Sosik, H. M. (2007). "A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging FlowCytobot." Limnology and Oceanography: Methods **5**(6): 195-203.

Orsi, A. H., Whitworth, T. and Nowlin, W. D. (1995). "On the meridional extent and fronts of the Antarctic Circumpolar Current." Deep Sea Research Part I: Oceanographic Research Papers **42**(5): 641-673.

Otsu, N. (1979). "A Threshold Selection Method from Gray-Level Histograms." IEEE Transactions on Systems, Man and Cybernetics **9**(1): 62-66.

Pappas, J., Kociolek, P. and Stoermer, E. (2014). "Quantitative morphometric methods in diatom research."

Pappas, J. and Stoermer, E. (2001). Fourier shape analysis and fuzzy measure shape group differentiation of Great Lakes *Asterionella* Hassall (Heterokontophyta, Bacillariophyceae). Proceedings of the 16th International Diatom Symposium.

Pappas, J. L., Fowler, G. and Stoermer, E. F. (2001). "Calculating shape descriptors from Fourier analysis: shape analysis of *Asterionella* (Heterokontophyta, Bacillariophyceae)." Phycologia **40**(5): 440-456.

Pappas, J. L. and Stoermer, E. F. (2003). "Legendre shape descriptors and shape group determination of specimens in the *Cymbella cistula* species complex." Phycologia **42**(1): 90-97.

Pearce, I., Davidson, A. T., Thomson, P. G., Wright, S. and van den Enden, R. (2011). "Marine microbial ecology in the Sub-Antarctic Zone: rates of bacterial and phytoplankton growth and grazing by heterotrophic protists." Deep Sea Research Part II: Topical Studies in Oceanography **58**(21): 2248-2259.

Pedraza, A., Bueno, G., Deniz, O., Cristóbal, G., Blanco, S. and Borrego-Ramos, M. (2017). "Automated Diatom Classification (Part B): A Deep Learning Approach." Applied Sciences **7**(5): 460.

Pfister, L., Wetzel, C. E., Klaus, J., Martínez-Carreras, N., Antonelli, M., Teuling, A. J. and McDonnell, J. J. (2017). "Terrestrial diatoms as tracers in catchment hydrology: a review." Wiley Interdisciplinary Reviews: Water **4**(6): e1241-n/a.

Pfitzer, E. (1869). "Über den Bau und die Zellteilung der Diatomeen." Botanische Zeitung **27**: 774-776.

Pinkernell, S. and Beszteri, B. (2013). A species distribution model of the marine diatom *Fragilariopsis kerguelensis*. EurOceans Hot Topics Conference - A Changing Ocean, Las Palmas, Gran Canaria.

Planquette, H., Statham, P. J., Fones, G. R., Charette, M. A., Moore, C. M., Salter, I., Nédélec, F. H., Taylor, S. L., French, M., Baker, A. R., Mahowald, N. and Jickells, T. D. (2007). "Dissolved iron in the vicinity of the Crozet Islands, Southern Ocean." Deep Sea Research Part II: Topical Studies in Oceanography **54**(18): 1999-2019.

Popp, B. N., Trull, T., Kenig, F., Wakeham, S. G., Rust, T. M., Tilbrook, B., Griffiths, B., Wright, S. W., Marchant, H. J., Bidigare, R. R. and Laws, E. A. (1999). "Controls on the carbon isotopic composition of southern ocean phytoplankton." Global Biogeochemical Cycles **13**(4): 827-843.

Potapova, M. and Hamilton, P. B. (2007). "Morphological and ecological variation within the *Achnanthidium minutissimum* (Bacillariophyceae) species complex." Journal of Phycology **43**(3): 561-575.

Potapova, M. and Snoeijs, P. (1997). "The natural life cycle in wild populations of *Diatoma moniliformis* (Bacillariophyceae) and its disruption in an aberrant environment." Journal of Phycology **33**(6): 924-937.

Poulickova, A., Vesela, J., Neustupa, J. and Skaloud, P. (2010). "Pseudocryptic diversity versus cosmopolitanism in diatoms: a case study on *Navicula cryptocephala* Kütz. (Bacillariophyceae) and morphologically similar taxa." Protist **161**.

Poulíčková, A., Veselá, J., Neustupa, J. and Škaloud, P. (2010). "Pseudocryptic diversity versus cosmopolitanism in diatoms: a case study on *Navicula cryptocephala* Kütz.(Bacillariophyceae) and morphologically similar taxa." Protist **161**(3): 353-369.

Quetin, L. B. and Ross, R. M. (1985). Feeding by Antarctic Krill, *Euphausia superba*: Does Size Matter? Antarctic Nutrient Cycles and Food Webs. W. R. Siegfried, P. R. Condy and R. M. Laws. Berlin, Heidelberg, Springer Berlin Heidelberg**:** 372-377.

R Core Team (2015). "R: a language and environment for statistical computing." from https://www.R-project.org

Rasband, W. S. (1997-2016). "ImageJ" from https://imagej.nih.gov/ij/

Rigual-Hernández, A. S., Trull, T. W., Bray, S. G. and Armand, L. K. (2016). "The fate of diatom valves in the Subantarctic and Polar Frontal Zones of the Southern Ocean: Sediment trap versus surface sediment assemblages." Palaeogeography, Palaeoclimatology, Palaeoecology **457**: 129-143.

Rigual-Hernández, A. S., Trull, T. W., Bray, S. G., Cortina, A. and Armand, L. K. (2015). "Latitudinal and temporal distributions of diatom populations in the pelagic waters of the Subantarctic and Polar Frontal Zones of the Southern Ocean and their role in the biological pump." Biogeosciences Discuss. **12**(11): 8615-8690.

Rintoul, S. R., Hughes, C. W. and Olbers, D. (2001). Chapter 4.6 The antarctic circumpolar current system. International Geophysics. G. Siedler, J. Church and J. Gould, Academic Press. **77:** 271-XXXVI.

Rintoul, S. R. and Sokolov, S. (2001). "Baroclinic transport variability of the Antarctic Circumpolar Current south of Australia." Journal of Geophysical Research: Oceans **106**: 2815-2832.

Rintoul, S. R. and Trull, T. W. (2001). "Seasonal evolution of the mixed layer in the Subantarctic zone south of Australia." Journal of Geophysical Research: Oceans **106**(C12): 31447-31462.

Rojas Camacho, O., Forero, M. and Menéndez, J. (2017). "A Tuning Method for Diatom Segmentation Techniques." Applied Sciences **7**(8): 762.

Roselli, L. and Basset, A. (2015). "Decoding size distribution patterns in marine and transitional water phytoplankton: from community to species level." Plos One **10**(5): e0127193.

Rosin, P. L. (2003). "Measuring shape: ellipticity, rectangularity, and triangularity." Machine Vision and Applications **14**(3): 172-184.

Round, F. E., Crawford, R. M. and Mann, D. G. (1990). The diatoms: the biology and morphology of the genera. Cambridge, Cambridge University Press.

Rousseaux, C. and Gregg, W. (2014). "Interannual Variation in Phytoplankton Primary Production at A Global Scale." Remote Sensing **6**(1): 1.

Russ, J. C. (2011). The Image Processing Handbook, Sixth Edition. Boca Raton, London, New York, CRC Press.

Sachs, O., Sauter, E. J., Schlüter, M., Rutgers van der Loeff, M. M., Jerosch, K. and Holby, O. (2009). "Benthic organic carbon flux and oxygen penetration reflect different plankton provinces in the Southern Ocean." Deep Sea Research Part I: Oceanographic Research Papers **56**(8): 1319-1335.

Sarmiento, J. á., Gruber, N., Brzezinski, M. and Dunne, J. (2004). "High-latitude controls of thermocline nutrients and low latitude biological productivity." Nature **427**(6969): 56-60.

SAS (2013). "JMP" from
http://www.jmp.com/

Sato, S., Beakes, G., Idei, M., Nagumo, T. and Mann, D. G. (2011). "Novel Sex Cells and Evidence for Sex Pheromones in Diatoms." Plos One **6**(10): e26923.

Schultz, M. E. and Trainor, F. R. (1968). "Production of male gametes and auxospores in the centric diatoms *Cyclotella Meneghiniana* and *C. cryptica*." Journal of Phycology **4**(2): 85-88.

Schulz, J., Barz, K., Ayon, P., Luedtke, A., Zielinski, O., Mengedoht, D. and Hirche, H.-J. (2010). "Imaging of plankton specimens with the lightframe on-sight keyspecies investigation (LOKI) system." Journal of the European optical society-rapid publications **5**.

Schulz, J., Möller, K. O., Bracher, A., Hieronymi, M., Cisewski, B., Zielinski, O., Voss, D., Gutzeit, E., Dolereit, T. and Niedzwiedz, G. (2015). "Aquatische Optische Technologien in Deutschland." Meereswissenschaftliche Berichte.

Schulze, K., Tillich, U. M., Dandekar, T. and Frohme, M. (2013). "PlanktoVision-an automated analysis system for the identification of phytoplankton." BMC bioinformatics **14**: 115.

Shepherd, J. G. (2009). Geoengineering the climate: science, governance and uncertainty, Royal Society.

Shimada, C., Nakamachi, M., Tanaka, Y., Yamasaki, M. and Kuwata, A. (2009). "Effects of nutrients on diatom skeletal silicification: evidence from *Neodenticula seminae* culture experiments and morphometric analysis." Marine Micropaleontology **73**(3): 164-177.

Shukla, S. K., Crespin, J. and Crosta, X. (2016). "*Thalassiosira lentiginosa* size variation and associated biogenic silica burial in the Southern Ocean over the last 42kyrs." Marine Micropaleontology **127**: 74-85.

Shukla, S. K. and Crosta, X. (2017). "*Fragilariopsis kerguelensis* size variability from the Indian subtropical Southern Ocean over the last 42 000 years." Antarctic Science **29**(2): 139-146.

Shukla, S. K., Crosta, X., Cortese, G. and Nayak, G. N. (2013). "Climate mediated size variability of diatom *Fragilariopsis kerguelensis* in the Southern Ocean." Quaternary Science Reviews **69**: 49-58.

Shukla, S. K., Mohan, R. and Sudhakar, M. (2009). "Diatoms: a potential tool to understand past oceanographic settings." Current Science: 1726-1734.

Siegel, D. and Deuser, W. (1997). "Trajectories of sinking particles in the Sargasso Sea: modeling of statistical funnels above deep-ocean sediment traps." Deep Sea Research Part I: Oceanographic Research Papers **44**(9): 1519-1541.

Siegel, D. A., Fields, E. and Buesseler, K. O. (2008). "A bottom-up view of the biological pump: Modeling source funnels above ocean sediment traps." Deep Sea Research Part I: Oceanographic Research Papers **55**(1): 108-127.

Simonsen, R. (1987). "Atlas and catalogue of the diatom types of Friedrich Hustedt." Vol. 1 Catalogue X, 525 Vol. 2 Atlas, plates 1-395, Vol. 3 Atlas, plates 396-772 , Cramer bei Borntraeger, Berlin und Stuttgart .

Sklansky, J. (1982). "Finding the convex hull of a simple polygon." Pattern Recognition Letters **1**(2): 79-83.

Smetacek, V. (1999). "Diatoms and the ocean carbon cycle." Protist **150**(1): 25-32.

Smetacek, V., Assmy, P. and Henjes, J. (2004). "The role of grazing in structuring Southern Ocean pelagic ecosystems and biogeochemical cycles." Antarctic Science **16**(4): 541-558.

Smetacek, V., Bathmann, U. and Helmke, E. (2005). "The expeditions ANTARKTIS XXI/3-4-5 of the research vessel Polarstern" in 2004 = Die Expeditionen ANTARKTIS XXI/3-4-5 des Forschungsschiffes

"Polarstern" 2004." Berichte zur Polar-und Meeresforschung (Reports on Polar and Marine Research) **500**.

Smetacek, V., Klaas, C., Menden-Deuer, S. and Rynearson, T. A. (2002). "Mesoscale distribution of dominant diatom species relative to the hydrographical field along the Antarctic Polar Front." Deep Sea Research Part II: Topical Studies in Oceanography **49**(18): 3835-3848.

Smetacek, V., Klaas, C., Strass, V. H., Assmy, P., Montresor, M., Cisewski, B., Savoye, N., Webb, A., d'Ovidio, F. and Arrieta, J. M. (2012). "Deep carbon export from a Southern Ocean iron-fertilized diatom bloom." Nature **487**(7407): 313-319.

Smol, J. P. and Stoermer, E. F. (2010). The diatoms: applications for the environmental and earth sciences, Cambridge University Press.

Spaulding, S., Esposito, R., Lubinski, D., Horn, S., Cox, M., D., M., Alger, A., Hall, B., Mayernick, M., Whittaker, T. and Yang, C. "Antarctic Freshwater Diatoms." Retrieved 18 Dec 2017, from http://huey.colorado.edu/diatoms/

Spaulding, S. A., Jewson, D. H., Bixby, R. J., Nelson, H. and McKnight, D. M. (2012). "Automated measurement of diatom size." Limnology and Oceanography: Methods **10**(11): 882-890.

Stela, B. and Monleón-Getino, A. (2016). "Facilitating the Automatic Characterisation, Classification and Description of Biological Images with the VisionBioShape Package for R." Open Access Library Journal, 2016, vol. 3, p. e3108.

Stoermer, E. and Ladewski, T. (1982). "Quantitative analysis of shape variation in type and modern populations of Gomphoneis herculeana." Nova hedwigia. Beihefte.

Stolte, W. and Riegman, R. (1996). "A model approach for size-selective competition of marine phytoplankton for fluctuating nitrate and ammonium." Journal of Phycology **32**(5): 732-740.

SUNY Stony Brook (2004). "TPS series" from http://life.bio.sunysb.edu/morph/

Tagliabue, A., Sallée, J.-B., Bowie, A. R., Lévy, M., Swart, S. and Boyd, P. W. (2014). "Surface-water iron supplies in the Southern Ocean sustained by deep winter mixing." Nature Geoscience **7**(4): 314-320.

Tan, X., Zhang, Q., Burford, M. A., Sheldon, F. and Bunn, S. E. (2017). "Benthic Diatom Based Indices for Water Quality Assessment in Two Subtropical Streams." Frontiers in microbiology **8**: 601.

Theriot, E. and Stoermer, E. F. (1984). "Principal Component Analysis of Variation in *Stephanodiscus rotula* and *S. niagarae* (Bacillariophyceae)." Systematic Botany **9**(1): 53-59.

Thingstad, T. F. (1998). A theoretical approach to structuring mechanisms in the pelagic food web. Eutrophication in Planktonic Ecosystems: Food Web Dynamics and Elemental Cycling, Springer**:** 59-72.

Thomas, D. N. and Dieckmann, G. S. (2002). "Antarctic Sea Ice - a Habitat for Extremophiles." Science **295**(5555): 641-644.

Timmermans, K. R. and Van Der Wagt, B. (2010). "Variability in Cell Size, Nutrient Depletion, and Growth Rates of the Southern Ocean Diatom *Fragilariopsis kerguelensis* (Bacillariophyceae) after Prolonged Iron Limitation1." Journal of Phycology **46**(3): 497-506.

Tréguer, P. J. (2014). "The southern ocean silica cycle." Comptes Rendus Geoscience **346**(11): 279-286.

Tréguer, P. J. and De La Rocha, C. L. (2013). "The world ocean silica cycle." Annual review of marine science **5**: 477-501.

Verity, P. G. and Smetacek, V. (1996). "Organism life cycles, predation, and the structure of marine pelagic ecosystems." Marine Ecology Progress Series: 277-293.

Vesela, J., Neustupa, J., Pichrtova, M. and Poulickova, A. (2009). "Morphometric study of *Navicula* morphospecies (Bacillariophyta) with respect to diatom life cycle." Fottea **9**.

Veselá, J., Pavla Urbánková, Kateřina Černá and Jiří Neustupa (2012). "Ecological variation within traditional diatom morphospecies: diversity of *Frustulia rhomboides* sensu lato (Bacillariophyceae) in European freshwater habitats." Phycologia **51**(5): 552-561.

Waite, A. and Harrison, P. J. (1992). "Role of sinking and ascent during sexual reproduction in the marine diatom *Ditylum brightwellii*." Marine Ecology Progress Series: 113-122.

Waite, A. M. and Nodder, S. D. (2001). "The effect of in situ iron addition on the sinking rates and export flux of Southern Ocean diatoms." Deep Sea Research Part II: Topical Studies in Oceanography **48**(11): 2635-2654.

Warnock, J. P., Scherer, R. P. and Konfirst, M. A. (2015). "A record of Pleistocene diatom preservation from the Amundsen Sea, West Antarctica with possible implications on silica leakage." Marine Micropaleontology **117**: 40-46.

Wilken, S., Hoffmann, B., Hersch, N., Kirchgessner, N., Dieluweit, S., Rubner, W., Hoffmann, L. J., Merkel, R. and Peeken, I. (2011). "Diatom frustules show increased mechanical strength and altered valve morphology under iron limitation." Limnology and Oceanography **56**(4): 1399-1410.

Winter, J. G. and Duthie, H. C. (2000). "Stream epilithic, epipelic and epiphytic diatoms: habitat fidelity and use in biomonitoring." Aquatic Ecology **34**(4): 345-353.

Wishkerman, A. and Hamilton, P. B. (2017). "DiaCurv: a value-based curvature analysis application in diatom taxonomy." Diatom Research **32**(3): 351-358.

Woodard, K., Kulichová, J., Poláčková, T. and Neustupa, J. (2016). "Morphometric allometry of representatives of three naviculoid genera throughout their life cycle." Diatom Research **31**(3): 231-242.

Wright, S. W., Thomas, D. P., Marchant, H. J., Higgins, H. W., Mackey, M. D. and Mackey, D. J. (1996). "Analysis of phytoplankton of the Australian sector of the Southern Ocean: comparisons of microscopy and size frequency data with interpretations of pigment HPLC data using the 'CHEMTAX' matrix factorisation program." Marine Ecology Progress Series **144**: 285-298.

Zielinski, U. (1993). Quantitative Bestimmung von Paläoumweltparametern des Antarktischen Oberflächenwassers im Spätquartär anhand von Transferfunktionen mit Diatomeen = Quantitative estimation of palaeoenvironmental parameters of the Antarctic surface water in the Late Quaternary using transfer functions with diatoms, Alfred Wegener Institute for Polar and Marine Research.

Zielinski, U. and Gersonde, R. (1997). "Diatom distribution in Southern Ocean surface sediments (Atlantic sector): Implications for paleoenvironmental reconstructions." Palaeogeography Palaeoclimatology Palaeoecology **129**(3-4): 213-250.

Zielinski, U., Gersonde, R., Sieger, R. and Fütterer, D. (1998). "Quaternary surface water temperature estimations: Calibration of a diatom transfer function for the Southern Ocean." Paleoceanography **13**(4): 365-383.

Zunic, J. and Rosin, P. L. (2002). A Convexity Measurement for Polygons. Proceedings of the British Machine Vision Conference 2002. Cardiff, UK.

8 Acknowledgements

First I would like to thank Bank Beszteri and Gerhard Kauer, without whom I most probably never would have undertaken the adventure of writing a PhD thesis. I owe very special thanks to Bank, for supervising my PhD projects, and for enabling me to work at AWI directly after finishing my master's studies, where I started developing SHERPA. Another special thanks is owed to Gerhard, for supervising, supporting and motivating me throughout. These two persons are the bright minds and bright lights guiding my way through the sometimes rough waters of my academic journey.

I would also like to express my gratitude to Andres Salvador Rigual-Hernández, Gastón Osvaldo Almandoz and Giuseppe Cortese, who reviewed this thesis, for investing time and efforts to support a, how I feel, very late start into the realm of academic research.

I also thank everybody else from the Husted diatom study centre, especially Friedel, Fenina, Sarah, Lena, Ute and Stefan, for their support and friendship, and also Nike for supporting my first steps in automated diatom image analysis.

A very heartfelt thank-you to Andres Salvador Rigual-Hernández and Leanne Armand. Andres came up with the idea of investigating *F. kerguelensis* life cycle from sediment trap material and supported me throughout this project. Leanne provided indispensable help by being a great source of knowledge as well as a treasurer of precious old material, and she supervised my first steps in this project in Sydney, Australia. I was granted to get there by an outgoing scholarship founded by the Helmholtz Graduate School for Polar and Marine Research (POLMAR). This enabled me to have one of the best times of my life and to re-adjust my perspective; thank you so much, Claudia and Claudia and everybody else from POLMAR.

Also a very warm thank-you to Oliver Esper, Christine Klaas, Anya Waite, Tanja Glawatty and everybody else at AWI who supported me by words and deeds; any supporter not finding her / his name listed here, please forgive me, I never before found such a supportive and motivating environment as at AWI.

Another thank is owed to everybody using SHERPA, for showing me that my work might be of relevance not only for myself, and to *F. kerguelensis* for being such a patient and fruitful object of investigation ;-)

Finally, a very big and warm thank-you to my biggest fan, Andrea, for supporting me throughout the final phase of writing my thesis, for being my anchor to reality and the light of my life. Do you realize that without SHERPA, we probably never would have met each other?

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the priority programme 1158 "Antarctic Research with comparative investigations in Arctic ice areas" by grant nr. BE4316/4-1, KA1655/3-1, and by an outgoing scholarship, as well as travel expenses, granted by the Helmholtz Graduate School for Polar and Marine Research (POLMAR).

9 Annex: Data medium

The data medium contains the data accompanying the papers, as well as R scripts used for analysis and visualization and other supplementary material. It can be downloaded from PANGAEA: https://doi.pangaea.de/10.1594/PANGAEA.890162