



OPEN

DATA DESCRIPTOR

CASCADE: Dataset of extant coccolithophore size, carbon content and global distribution

Joost de Vries¹✉, Alex J. Poulton², Jeremy R. Young³, Fanny M. Monteiro¹, Rosie M. Sheward⁴, Roberta Johnson⁵, Kyoko Hagino⁶, Patrizia Ziveri^{5,7} & Levi J. Wolf¹

Coccolithophores are marine calcifying phytoplankton important to the carbon cycle and a model organism for studying diversity. Here, we present **CASCADE** (Coccolithophore Abundance, Size, Carbon And Distribution Estimates), a new global dataset for 139 extant coccolithophore taxonomic units. **CASCADE** includes a trait database (size and cellular organic and inorganic carbon contents) and taxonomic-unit-specific global spatiotemporal distributions (Latitude/Longitude/Depth/Month/Year) of coccolithophore abundance and organic and inorganic carbon stocks. **CASCADE** covers all ocean basins over the upper 275 meters, spans the years 1964-2019 and includes 33,119 gridded taxonomic-unit-specific abundance observations. Within **CASCADE**, we characterise the underlying uncertainties due to measurement errors by propagating error estimates between the different studies. This error propagation pipeline is statistically robust and could be applied to other plankton groups. **CASCADE** can contribute to (observational or modelling) studies that focus on coccolithophore distribution and diversity and the impacts of anthropogenic pressures on historical populations. Additionally, our new taxonomic-unit-specific cellular carbon content estimates provide essential conversions to quantify the role of coccolithophores on ecosystem functioning and global biogeochemistry.

Background & Summary

Coccolithophores are marine phytoplankton and important calcifiers, impacting the ocean carbon cycle through the production of calcite (inorganic carbon)¹⁻³ and primary production (organic carbon, ca. 2 to 10% of global PP⁴; and up to 40% regionally⁵). However, it is still unclear how climate change will impact coccolithophores due to many uncertainties regarding their ecology and global carbon stocks. Several issues predicate this problem: (1) *in situ* observations of coccolithophores are in cellular concentrations and need to be converted into carbon content to infer their climatic impacts; (2) conversion of abundance into carbon content generally relies on laboratory measurements, which are sparse for coccolithophores because most species are hard to culture; (3) there are uncertainties associated with measurement error and natural variability in cell size and cellular organic and inorganic contents; and (4) *in situ* observations consist of different studies which have to be extracted from multiple sources, and might use different synonyms to refer to the same taxonomic unit.

Here, we provide new and improved estimates of cell size and cellular Particulate Organic Carbon (POC) and Particulate Inorganic Carbon (PIC) contents for 139 coccolithophore taxonomic units. For this, we collated direct measurements from *in situ* water samples from the literature and developed novel allometric functions of cellular POC and PIC based on cell size. We also compiled an extensive global abundance dataset, which includes 33,119 gridded observations and covers all ocean regions and seasons over the 1965-2019 period for the same 139 coccolithophore taxonomic units. Combining cell size and carbon content estimates with the abundance dataset, we calculated the global spatial-temporal distribution (Latitude, Longitude, Depth, Month and Year) of water-column taxonomic-unit-specific PIC and POC stocks for the 139 coccolithophore taxonomic units. To ensure data quality, we thoroughly checked for errors associated with taxonomic unit synonyms

¹BRIDGE, School of Geographical Sciences, University of Bristol, Bristol, BS8 1HB, UK. ²The Lyell Centre for Earth and Marine Science, Heriot-Watt University, Edinburgh, EH14 4BA, UK. ³Department of Earth Sciences, University College London, London, WC1E 6BS, UK. ⁴Institute for Geosciences, Goethe-University Frankfurt, Frankfurt am Main, 60438, Germany. ⁵ICTA-UAB, Universitat Autònoma de Barcelona, Barcelona, 08193, Spain. ⁶Marine Core Research Institute, Kochi University, Nankoku, 783-8502, Japan. ⁷Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, 08010, Spain. ✉e-mail: joost.devries@bristol.ac.uk

and misspellings for all datasets. We only included methods that resolve individual coccolithophore taxonomic units for the abundance dataset and excluded methods that only measured total coccolithophore abundances. Besides, the dataset includes water-column abundance data previously not readily available because they were only included in hard-to-access formats such as supplemental PDF tables, notebooks and floppy disks. Finally, we carefully quantified uncertainties in the cellular POC and PIC content estimates through sample reconstruction and allometric regression error propagation.

This approach significantly improves on previous carbon content and stock estimates by (1) expanding and improving coccolithophore cell size estimates provided in MAREDAT⁶, (2) re-evaluating the allometric scaling of coccolithophore POC content with cell size using a much larger dataset (42 vs 9 observations included in a previous coccolithophore-specific allometric scaling estimate⁷), (3) estimating a novel allometric scaling function for PIC based on 961 measurements from 56 taxonomic units, and (4) providing the first global spatial-temporal distribution estimate of taxonomic-unit-specific coccolithophore PIC distributions.

The overall dataset is valuable for training species distribution models and validating mechanistic ecosystem models, which are needed to estimate coccolithophore contributions to the global carbon cycle and the impact of climate change. The dataset can also help study coccolithophore diversity and drivers of their distributions. The long period (1965–2019) covered by our dataset also makes it an excellent tool for investigating historical anthropogenic impacts on coccolithophore ecology. Finally, our dataset can help identify gaps in the spatial and temporal coverage of coccolithophore abundance, size, and carbon content, which can guide future sampling efforts.

Methods

The dataset of global coccolithophore abundance and associated carbon stocks was created by combining (1) new estimates of taxonomic-unit-specific cell size and cellular POC and PIC contents for different coccolithophore taxonomic units from field samples and (2) a new compilation of global coccolithophore abundance observations from water-column samples. The sources of these data and data processing used to create the final dataset are summarised in Fig. 1 and detailed below.

We compiled global coccolithophore abundance observations from water-column samples and estimated taxonomic-unit-specific cell size, POC and PIC contents for taxonomic units with at least 20 abundance observations (139 out of 226 taxonomic units). We did not compile abundances for rarer taxonomic units (<20 observations) because their impact on the carbon cycle is likely small, and there is high uncertainty in estimating their cell sizes with so few measurements.

Cell size estimates were compiled from the literature^{6–16} and measured using Light Microscopy (LM)¹⁷ and Scanning Electron Microscopy (SEM)¹⁸ for all 139 taxonomic units. Cellular POC estimates were compiled from laboratory observations of POC content for 11 taxonomic units from 7 studies^{7,8,10,12,15,19,20}. Finally, we compiled cellular PIC content estimates for 3 taxonomic units from 2 studies^{21,22} and provide morphometric *in situ* measurements for 56 taxonomic units from the 14th cruise of the Atlantic Meridional Transect (AMT) programme (April–June 2004^{18,23}).

We used a model of allometric scaling for coccolithophore taxonomic units without direct measurements of cellular POC and PIC contents, converting their cell size estimates into cellular carbon content. Specifically, we modelled allometric scaling using Generalized Linear Models (GLMs), a class of linear regression models that can explicitly characterise the distribution of the measurement data even when the outcome is not normally distributed.

We also accounted for any underlying uncertainties at each step in our analysis by propagating errors. We achieved this for measurements such as cell size and cellular POC and PIC contents by sampling the data based on the standard error of the mean and bootstrapping the resulting distributions. We propagated the error using simulations obtained from the GLMs for allometric estimates. More details of these error propagation methods are provided below.

Taxonomy, misspellings, and synonyms. We followed the latest coccolithophore taxonomic definitions provided by NannoTax3 (www.mikrotax.org/ - last accessed January 2024)²⁴. We considered life cycle phases (e.g. HET and HOL) but did not consider finer taxonomic resolution than species level (i.e., no classification of sub-species, morphospecies, morphotypes, or varieties in the dataset), as they are poorly resolved in our *in situ* observation compilation and can be relatively subjective. All records in our dataset were thoroughly checked for taxonomic unit misspellings and synonyms to ensure consistent, accurate and up-to-date taxonomy. These synonyms and misspellings are provided as a YAML file in the data archive²⁵.

We excluded *Reticulofenestra sessilis* HET from our dataset, as this species occurs exclusively in association with a centric diatom of the genus *Thalassiosira*²⁶, which makes estimating POC and PIC stock for this organism challenging.

Additionally, we excluded rare taxonomic units (88 taxonomic units²⁵, <20 global occurrences) because they are highly unlikely to form a significant contribution to the carbon cycle or diversity calculations (even regionally). Furthermore, we lack cellular PIC and POC estimates for rare taxonomic units due to their infrequent occurrence.

Life cycle associations and coccolith morphology. Coccolithophores have a distinct ‘haplo-diplontic’ life cycle, which allows them to grow and divide in two different life cycle phases (‘haploid’ and ‘diploid’)^{27–29}. These two life cycle phases are morphologically distinct, with more heavily calcified diploid life cycle phases generally utilizing a heterococcolith (‘HET’) morphologies and more lightly calcified haploid cells utilizing naked, holococcolith (‘HOL’), ceratolith (‘CER’), or polycrater (‘POL’) morphologies^{8,30–32}. Furthermore, haploid and diploid life cycle phases for a given species tend to have similar cell sizes^{8,31,33}.

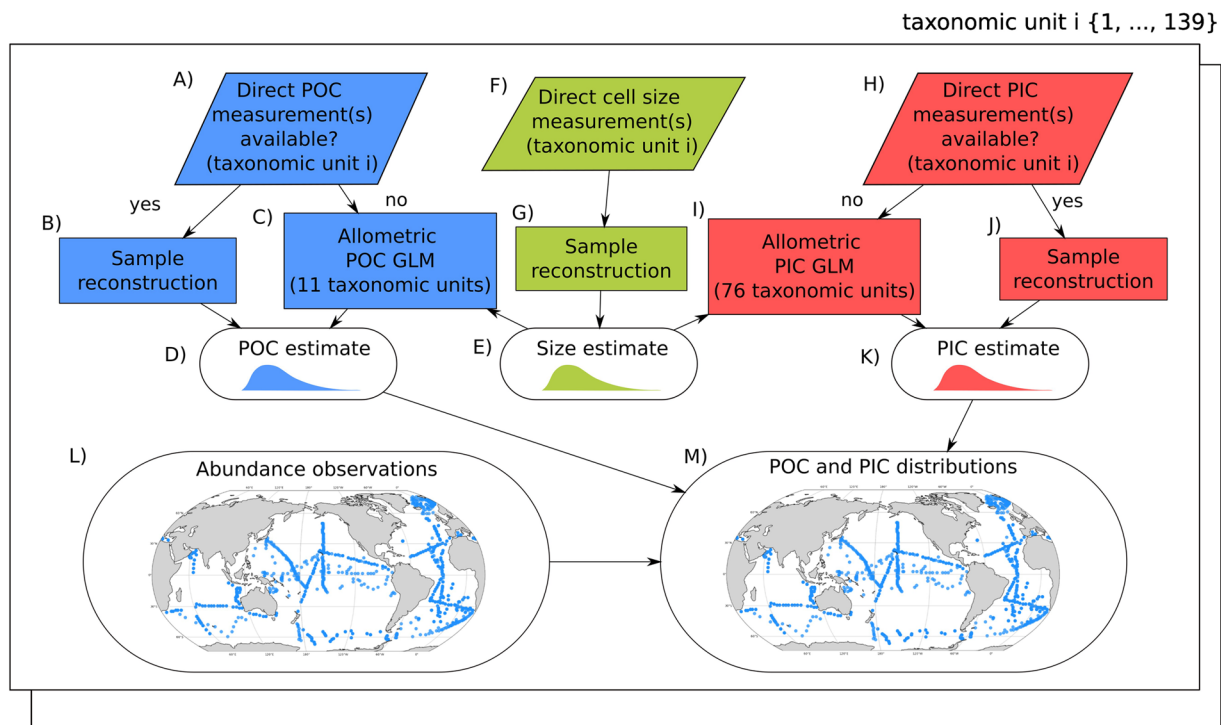


Fig. 1 Graphical Abstract: An overview of CASCADE dataset's pipeline. We compiled taxonomic-unit-specific cell size and cellular organic carbon (POC) and inorganic carbon (PIC) contents for 139 coccolithophore taxonomic units. We then combined them with a global abundance observational dataset to create a gridded data product (Latitude/Longitude/Depth/Month/Year) of the global distributions of taxonomic-unit-specific coccolithophore carbon stocks. The pipeline consists of the following steps for a given taxonomic unit (i). (A) Direct POC laboratory measurements of taxonomic unit i are compiled from the literature. (B) If POC measurements are available for taxonomic unit i , measurements are statistically reconstructed to create a merged estimate of POC content of taxonomic unit i . (C) If direct measurements are unavailable for taxonomic unit i , an allometric Generalised Linear Model (GLM) based on cell size and POC contents of all taxonomic units with observations is used to estimate taxonomic unit i 's cellular POC content from its merged cell-size distribution (E). The merged cell-size estimate of taxonomic unit i is created by compiling cell-size estimates of taxonomic unit i (F), which are then statistically reconstructed to propagate the error distributions of measurements from different studies (G). For cellular PIC content estimates, a similar process is applied. (H) PIC measurements are conducted and compiled from the literature. (J) If direct measurements are available for taxonomic unit i , its PIC measurements are statistically reconstructed and used to create a merged estimate of its PIC content (K). If no direct estimates are available for taxonomic unit i , an allometric Generalised Linear Model (GLM) based on cell size and PIC contents of all taxonomic units with observations is used to estimate taxonomic unit i 's cellular PIC content from its merged cell-size estimate (E). Finally, global spatial-temporal abundances of taxonomic unit i are compiled (L) and then converted into organic carbon (D) and inorganic (K) stocks, providing a global distribution of cellular carbon stocks for taxonomic unit i (M). This pipeline is then applied for each taxonomic unit (139 in total).

While there are some notable exceptions to these trends, as in some cases naked morphologies have been observed to be diploid³⁴, and some diploid morphotypes can be lightly calcified⁹, we can nonetheless use this information in our analysis to make better estimates of cell sizes and PIC contents for taxonomic units for which we do not have direct measurements. We thus compiled coccolith type and known life cycle associations from NannoTax3²⁴ for each taxonomic unit in our dataset (Supplementary Table 1).

Propagating uncertainties. All data in our compilation have underlying uncertainties due to instrument measurement error and natural variations of cell size and cellular carbon content, which can vary between studies. Accurately capturing these uncertainties and sources of variability is important, as these are needed to determine confidence in standing stock estimates or when making inferences about ecological processes such as drivers of diversity or impacts of anthropogenic climate change. However, propagating and integrating sources of uncertainty across studies is particularly challenging when combining multiple studies with studies using different measuring conditions, populations, or instruments. As a result, each study generates different estimates with unique error distributions. Another challenge is when we rely on multiple variables to estimate the cellular carbon contents (e.g., estimating POC content based on size), we also need to propagate the uncertainties across all variable estimates.

Here, we implemented various forms of sample reconstruction to propagate uncertainties across the different methods and estimates from our multiple studies. Conceptually, the idea is to generate large sets of random

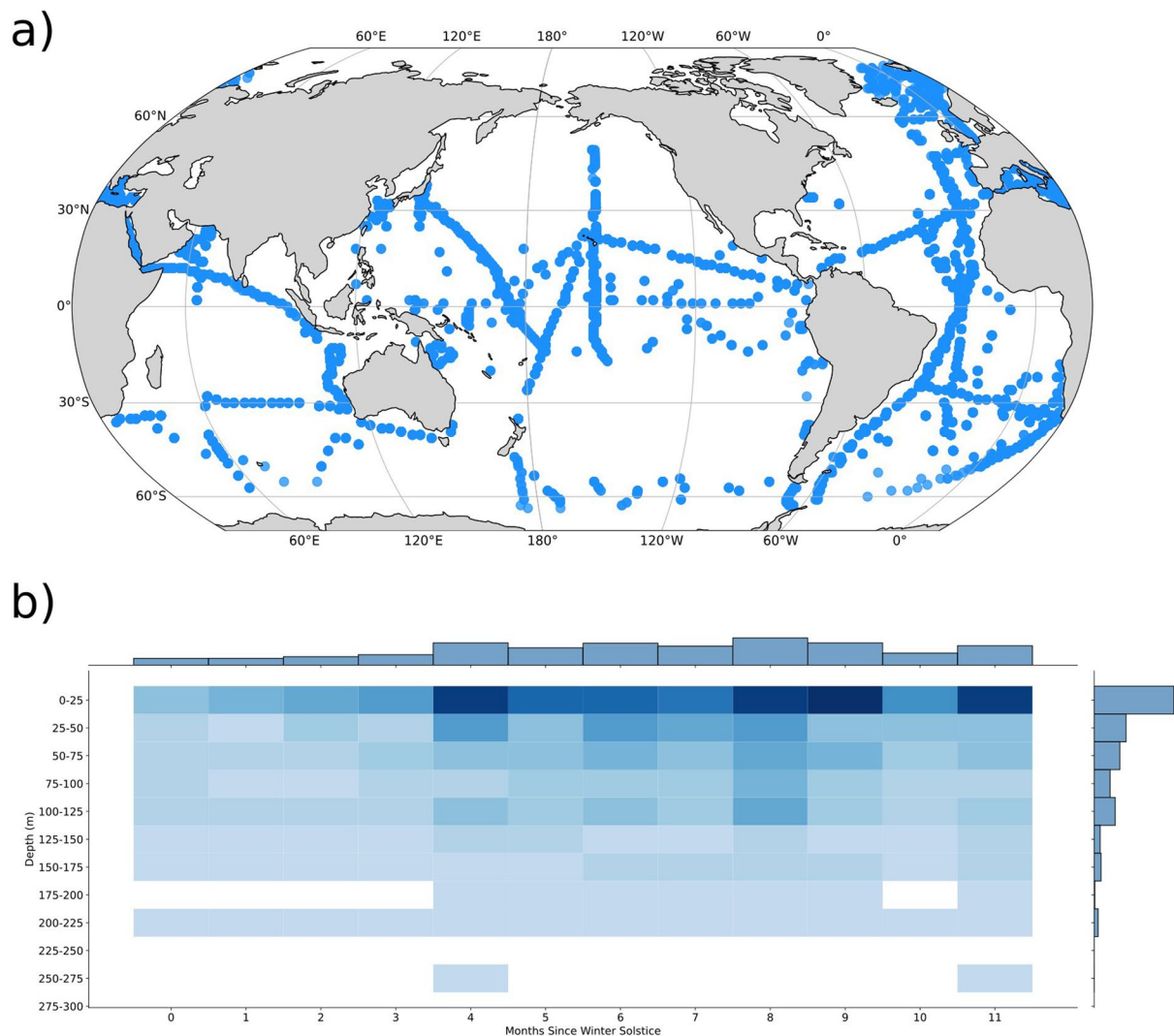


Fig. 2 Abundance sample distribution. (a) Latitude and Longitude; (b) Depth and time. Note that most samples are in summer months and in the top 25m of the water column.

values of the estimated variables (e.g., cell size). These replicates represent the potential outcomes of each study given their error distribution and allow us to model the uncertainty distribution of each study. We then combined each study's replicated distribution set to produce merged distribution of possible measurements, representing the combined uncertainties of all different studies.

For example, imagine a given species i is measured in two observational studies, each with estimates of mean cell size and standard error of the mean (SE) in cell sizes. Study A's mean size is $10\ \mu\text{m}$ with a SE of $2\ \mu\text{m}$, while study B's mean size is $12\ \mu\text{m}$ with a SE of $3\ \mu\text{m}$. Assuming the distribution of cell sizes is normally distributed, we generate a set of cell size values for each study based on its mean and standard error of the mean. Study A's set might be 11.6, 6.0, 9.7, while study B's set might be 7.6, 18.0, 16.9. We can then append these two sets of numbers to create a new set 11.6, 6.0, 9.7, 7.6, 18.0, 16.9. Finally, this merged distribution can be smoothed using Bayesian bootstrapping and then used to calculate a newly merged mean and confidence interval based on percentile estimates.

In our study, we used different methods to simulate these sets of values depending on the context and type of data available. The method above (assuming a zero-truncated normal distribution) was used for direct measurements with only a mean and SE. For direct measurements where the original data was available, we used Bayesian bootstrapping for sample reconstruction instead. Bootstrapping is a type of resampling that uses random sampling by replacement to create new distributions while maintaining the original uncertainty distribution³⁵. Bayesian bootstrapping is a generalised version of bootstrapping, which further allows weighting of the sub-samples to create smoother distributions³⁶. We used Bayesian bootstrapping if possible because the method preserves the original data's underlying distribution without making any normality assumptions while creating smoother reconstructed sample distributions. In some cases, studies only reported maximum and minimum values instead of SE; for such studies, we estimated SD using a modified version of the 'range rule'.

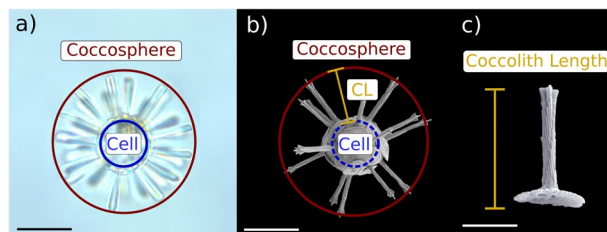


Fig. 3 Comparison of cell size and coccospere size for the taxonomic unit *Rhabdosphaera clavigera* HET. Coccolithophore cells (blue circles) are covered in coccoliths, forming a larger coccospere (red circles). **(a)** By measuring the difference between coccospere and cell size using light microscopy, cell sizes can be estimated from SEM images by measuring coccospere size and applying a conversion factor **(b)**. **(c)** Alternatively, if LM images are unavailable, SEM measurements of coccolith length (CL, orange line) can be used to estimate cell size by subtracting twice the coccolith length from the coccospere diameter. SEM and LM images of *R. clavigera* HET coccospere and coccolith by Jeremy Young.

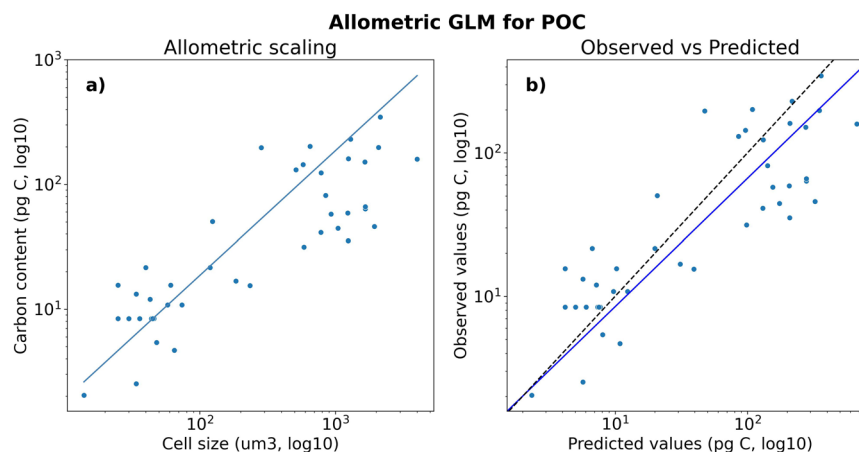


Fig. 4 GLM allometric scaling of coccolithophore cellular POC content **(a)** Fitted GLM allometric scaling (line) compared to observations (points). **(b)** Comparison between the GLM predicted and observed cellular POC content. The black line represents a 1:1 prediction between observed and fitted values. Note that x- and y-axis are on a log10 scale.

$$SD = (\text{Maximum} - \text{Minimum})/2 \quad (1)$$

This first-order approximation assumes a normal distribution and that the reported maximum and minimum values represent one standard deviation from the mean. Based on simulation (Fig. 6a), this tends to overestimate the standard deviation even when the original sample size is small and thus is a conservative estimate. This is critical since all studies reporting only min/max values measured only a few cells^{37–44}. We thus use division by two instead of the standard range rule, which assumes min and max represent two deviations from the mean (i.e. division by four)⁴⁵, and underestimates the standard deviation if the sample size is small (Fig. 6b). SE was then estimated from the SD by assuming a sample size of three.

For direct measurements with no SE, original data or min and max values, SE was simulated by bootstrapping from normalized SEs from other measurements of the same taxonomic unit. If no other SEs were available for the same taxonomic unit, cross-study-and-cross-taxonomic-unit normalized SEs were used.

Uncertainty in Generalised Linear Modelling estimates. When direct measurements were unavailable, carbon content was estimated based on cell size ('allometric estimates'). For these estimates, errors were propagated by accounting for model uncertainty of the slope used to convert cell size into cellular carbon content. We achieved this by bootstrapping of training data before fitting multiple GLMs (one for each bootstrapped sample).

Implementation in Python. All error propagation was conducted in Python using either Numpy⁴⁶ (np) or Statsmodels⁴⁷ (sm). If only mean and SD were known, np.random.normal was used, which was zero-truncated by removing values below zero using np.clip. If the original data was available, np.random.choice.dirichlet was used for Bayesian bootstrapping. To simulate GLM distributions, sm.GLM.get_distribution was used.

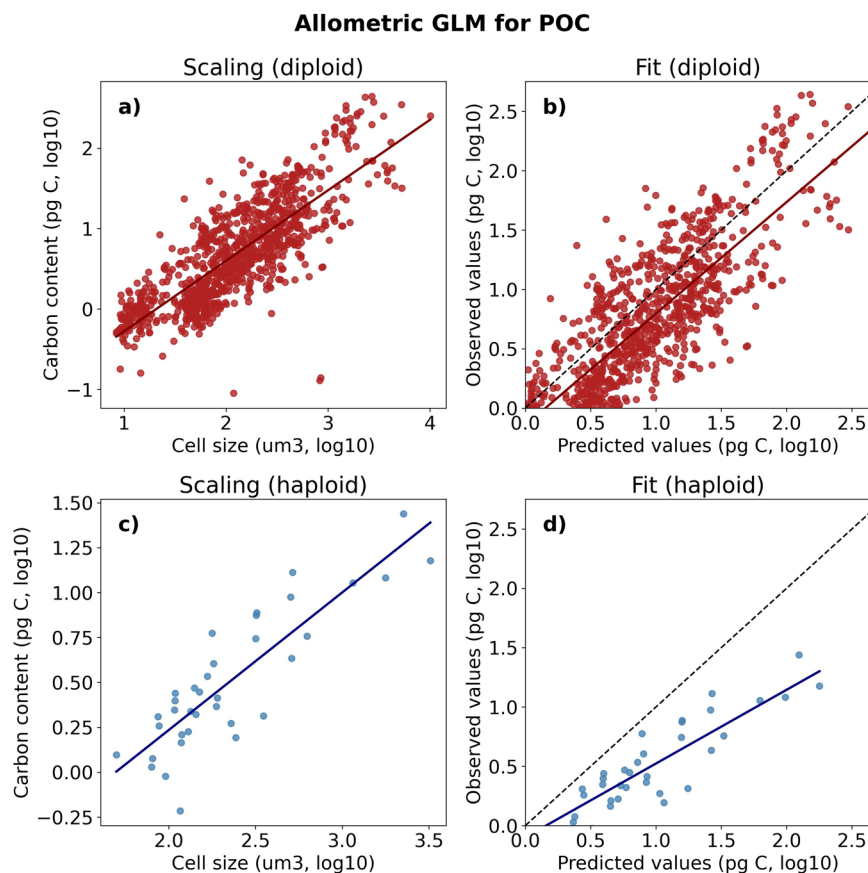


Fig. 5 GLM allometric scaling of coccolithophore cellular PIC content per life stages. **(a)** Fitted GLM allometric scaling (red line) over trained data (red points) for the coccolithophore diploid morphotypes. **(b)** Comparison between the GLM fitted and observed cellular PIC content for diploid morphotypes **(c)** Fitted GLM allometric scaling (blue line) over trained data (blue points) for the coccolithophore haploid morphotypes. **(d)** Comparison between the GLM fitted and observed cellular PIC content for haploid morphotypes. Note that x and y-axis are on a log₁₀ scale.

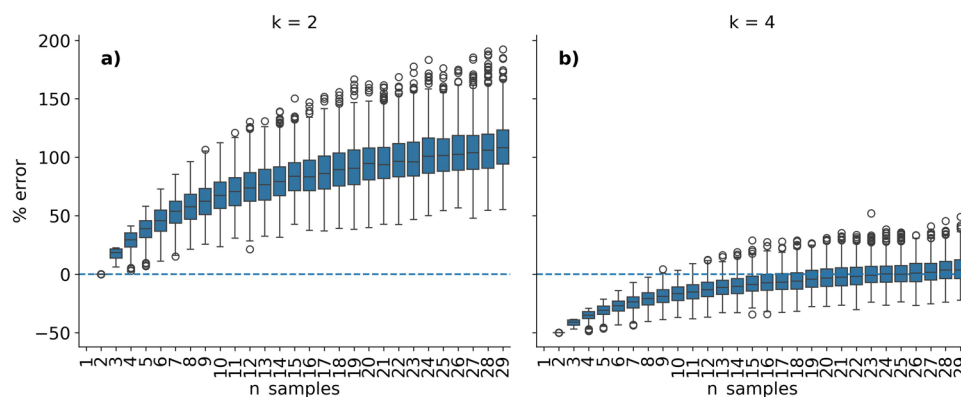


Fig. 6 The percentage error of SD when estimated based on min-max values. **(a)** Percentage error of estimated SD based on min-max when values are divided by two. **(b)** Percentage error of estimated SD based on min-max when values are divided by four. The horizontal line is 0 percentage error; values above this line represent over-estimates of SD (conservative estimates), while values below this line represent under-estimates of SD. Note that when min-max values are divided by two **(a)**, SD is never underestimated, even when the sample size is 1.

Coccolithophore cell size. We compiled coccolithophore cell size estimates from the literature. The cell size component of our dataset is sourced from a range of observation and measurement types. It includes new scanning electron microscopy measurements¹⁸ and light microscopy measurements¹⁷ from plankton samples, size

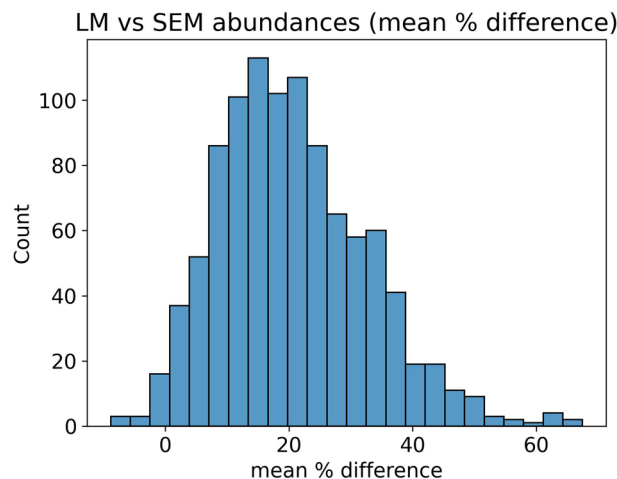


Fig. 7 Comparison of abundance estimates from *in situ* observations counts of taxonomic units collected from the same water sample using SEM or LM. Data was acquired from Bollman *et al.*⁷⁰ and re-analysed using Bayesian bootstrapping of relative % difference values for each sample and taxonomic unit. Note that we merged sub-species and varieties and excluded non-taxonomic-unit-specific counts (e.g. undefined *Syracosphaera* sp.).

measurements from laboratory cultures^{6–8,10–13,15,16,48,49}, and literature morphometric estimates of cell size⁷. Each measurement type has its advantages and disadvantages, which are discussed below.

Regarding observation types, laboratory cell size measurements generally benefit from more samples but are usually limited to single genetic strains of a taxonomic unit grown under a controlled range of environmental conditions. Therefore, cell size measurements from laboratory cultures do not capture the full natural variation of cell sizes that occur *in situ*, where the size distribution of taxonomic unit populations will be influenced by genetic and phenotypic diversity and a wider range of environmental conditions⁵⁰. However, a smaller quantity of cell size measurements from plankton samples are available, and the observations are often spatially biased (similarly to the availability of abundance data). Here, we overcome some of these limitations by combining *in situ* (plankton) and laboratory measurements of cell size to estimate the overall expected distribution of cell size of individual coccolithophore taxonomic units.

Regarding measurement types, LM provides more accurate cell size measurements than SEM measurements. This is because only the coccosphere surface (i.e. including the inorganic carbon exoskeleton, the coccosphere) is observable through SEM, whereas, under LM, fine adjustments to the focal depth of the microscope allow the observer to image a cross-section of the cell that reveals the internal size of the coccosphere (approximately the diameter of the cell cytoplasm)⁴⁸. Therefore, cell size measurements from SEM must be derived from coccosphere size measurements. In addition, both LM and SEM can be spatially and temporally biased and might thus fail to capture the full range of natural intra-specific variations. Therefore, to maximise the use of available data, we combined both SEM and LM measurements for each cell size estimate, except if there was a disagreement in median cell size 5-fold times greater between LM and SEM measurements for a given taxonomic unit, in which case only LM (as the more accurate cell size measurement) was used. Generally, there was a good agreement between LM and SEM (Fig. 8).

We also used resampling to propagate any uncertainties in the observed measurements to maintain the uncertainty distributions resulting from natural intraspecific variations in cell size and measurement errors of each input size dataset. For a detailed description, please refer to the section Uncertainty Propagation.

In situ scanning electron estimates of cell size. Extant coccolithophore cell size estimates from plankton samples were derived from a new dataset of taxonomic unit-level morphometric traits measured on legacy SEM images from 16 samples (10,665 images) from AMT-14 ([zenodo.11483788](https://zenodo.org/record/11483788))¹⁸. From the SEM images ([zenodo.10571820](https://zenodo.org/record/10571820))⁵¹, the coccosphere diameter (long axis and short axis measurement) of each intact coccosphere encountered was measured using the freeware Fiji (v1.53a)⁵². The SEM morphometric trait database includes 961 coccosphere size measurements (as equivalent spherical diameter) for 56 taxonomic units. For each coccosphere, cell size was estimated from measured coccosphere size using a taxonomic-unit-specific conversion factor that defines the percentage of coccosphere volume that is represented by cell size (Fig. 3). The conversion factor was derived from LM measurements of cell size and coccosphere size on the same coccospheres¹⁷. The methodology for estimated cell size from coccosphere size measurements on SEM images and related uncertainties is described in detail in the associated publication²³.

In situ light microscopy estimates of cell size. With light microscopy, we can directly observe the internal space defined by the coccosphere, which, for most taxonomic units, can be assumed to correspond closely to the cell volume. For this study, slides from previous work that had common, well-preserved coccospheres were selected. The slides had been prepared during various research cruises, but in all cases, by filtration of seawater onto cellulose filters of 0.2 to 0.4 μm pore size. After filtration, the filters were oven-dried, and then a portion of them was

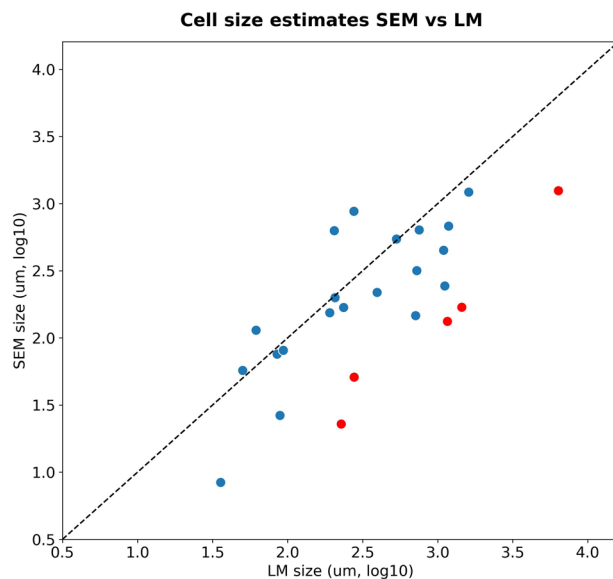


Fig. 8 Agreement between light microscopy (LM) and scanning electron microscopy (SEM) cell size estimates. For LM cell size was measured directly, while for SEM cell size was estimated based on coccosphere size and coccolith thickness. For taxonomic units where the disagreement was greater than 5-fold (red points), only LM estimates were used. Such taxonomic units included: *Algirosphaera robusta* HET, *Helicosphaera pavementum* HET, *Rhabdosphaera clavigera* HET, *Syracosphaera mediterranea* HOL *wettsteinii* type, and *Umbilicosphaera sibogae* HET.

immediately mounted on a glass slide using low-viscosity optical adhesive (Norland Optical Adhesive NOA 74). Coccospheres were identified using polarising light microscopy for cell volume measurement and then imaged in cross-polars and bright fields. The saved images were measured using the image analysis program Fuji⁵². This work focused on the most common/numerically important taxonomic units but excluded *Emiliana*, *Gephyrocapsa* and *Coccolithus*, for which extensive data is available from the literature. This dataset can be accessed from Zenodo ([zenodo.10572754](https://zenodo.org/record/10572754))¹⁷.

Laboratory light microscopy measurements of cell size. Our compilation of published laboratory cell size estimates consists of 11 different studies^{6,7,9–13,15,16,49,53}. We checked the method used for each study and converted diameter values to volume where appropriate. We did not include studies that only measured coccosphere size because converting such estimates to cell size adds significant uncertainty due to variations in coccosphere thickness. A concatenated dataset of the 11 studies can be found on Zenodo ([zenodo.12794780](https://zenodo.org/record/12794780))²⁵.

Literature morphometric estimates of cell size. For taxonomic units with no direct cell size measurements but where coccosphere size measurements were available in the literature, we similarly converted coccosphere size to an estimate of cell size by subtracting a taxonomic-unit-specific estimate of coccolith thickness or coccosphere thickness from coccosphere diameter (Fig. 3). These size conversion factors were taken from previous literature⁷ or newly estimated for this study for taxonomic units not included in other datasets²⁴.

We have not used the literature morphometric cell size estimates from MAREDAT⁶, which estimated cell dimensions for all extant coccolithophore taxonomic units by assuming a fixed ratio of 0.6 between coccosphere and cell volume (i.e. assuming that cell volume is consistently 60% of coccosphere volume). This assumption introduces significant uncertainties, as the ratio between coccosphere and cell size can vary from 0.3 to 0.9 between taxonomic units⁶. In extreme cases, this leads to 3 to 8-fold differences between estimates of cell volume and, subsequently, POC cellular content.

Therefore, we followed the approach described in Villiot *et al.*⁷ for our cell size estimates. Briefly, we compiled coccosphere diameter (or long axis and short axis diameter measurements for non-spherical taxonomic units) and coccolith thickness from the literature. Cell dimensions were then estimated by subtracting twice the coccolith thickness from the coccosphere diameter (Fig. 3b–c). Finally, cell volume was calculated by assuming a spherical or prolate spheroid coccosphere shape, depending on the taxonomic unit (Table 1). All taxonomic units were assumed to have a spherical coccosphere except for *Syracosphaera aurisinae* HET and *Calciopappus caudatus* HET, which were assumed to have a prolate spheroid coccosphere.

HOL life cycle cell size. Since haploid and diploid life cycle phases for a given taxonomic unit tend to have similar cell sizes^{8,31,33}, if available, the associated HET life cycle phase(s) values were used to estimate the cell size of the HOL phase lacking size measurements. A mean of associated HET life cycle phases was used for HOL morphotypes with multiple HET life cycle phases. For example, for the HOL life cycle phase *Helladosphaera*

Species	Vol (mean)	Vol (SD)	Cell shape	Ref.	obs.
<i>Alisphaera gaudii</i> POL	134.2	79.7	Spherical	Cros and Fortuno ³⁷	26
<i>Calciopappus caudatus</i> HET	23.4	3.9	Prolate spheroid	Gaarder and Ramsfjell ³⁸	177
<i>Calicasphaera blokii</i> HOL	33.5	—	Spherical	Cros and Fortuno ³⁷	32
<i>Calicasphaera concava</i> HOL	20.6	—	Spherical	Cros and Fortuno ³⁷	27
<i>Calyptosphaera sphaeroidea</i> HOL	346.4	186.2	Spherical	Cros and Fortuno ³⁷	81
<i>Helicosphaera HOL confusus type</i>	472.3	181.0	Spherical	Cros and Fortuno ³⁷	27
<i>Helladosphaera pienaarii</i> HOL	190.0	99.5	Spherical	Norris ³⁹	25
<i>Ophiaster formosus</i> HET	51.5	25.0	Spherical	Cros and Fortuno ³⁷	187
<i>Ophiaster hydroideus</i> HET	44.7	3.0	Spherical	Cros and Fortuno ³⁷	490
<i>Pappomonas sp. type 5</i> HET	144.3	16.5	Spherical	Cros and Fortuno ³⁷	31
<i>Sphaerocalyptra adenensis</i> HOL	70.0	31.0	Spherical	Cros and Fortuno ³⁷	40
<i>Syracosphaera aurisinae</i> HET	942.7	150.0	Prolate spheroid	Lecal-Schlauder ⁴⁰	91
<i>Syracosphaera borealis</i> HET	81.7	19.3	Spherical	Okada and McIntyre ⁴¹	45
<i>Syracosphaera reniformis</i> HET	59.8	19.0	Spherical	Kleijne and Cros ⁴²	39
<i>Syracosphaera squamosa</i> HET	11.7	2.5	Spherical	Kleijne and Cros ⁴²	48
<i>Syracosphaera strigilis</i> HET	91.7	37.1	Spherical	Cros and Fortuno ³⁷	24
<i>Wigwamma antarctica</i> HET	65.4	—	Spherical	Thomsen <i>et al.</i> ⁴³	41
<i>Zygospaera marsilii</i> HOL	174.4	37.6	Spherical	Borsetti <i>et al.</i> ⁴⁴ Cros and Fortuno ³⁷	45

Table 1. Species for which cell size was estimated based on literature morphometric estimate of coccosphere size and coccolith thickness. Diameter (or length and width for prolate spheroid morphologies) was estimated by subtracting twice the coccolith thickness from coccosphere length. After this, cell volume (Vol) was estimated based on the cell shape. For these species' estimates, SDs were unavailable and, if possible, estimated by assuming the min and max values represented the 68% quantiles: $SD = (max - min) / 2$.

cornifera HOL estimates from *Syracosphaera nodosa* HET³⁰ and *Syracosphaera noroitica* HET⁵⁴ were used, while for the HOL life cycle phase *Sphaerocalyptra quadridentata* HOL estimates from *Algirosphaera robusta* HET⁵⁵⁻⁵⁷ and *Rhabdosphaera clavigera* HET³⁷ were used.

The full list of life cycle associations utilized in this study can be found in Supplementary Table 1 and is provided as a YAML file in the Zenodo data archive ([zenodo.12794780](https://zenodo.org/record/12794780))²⁵.

Cellular POC content. *Laboratory combustion estimates of POC.* Direct measurements of cellular POC were compiled from previously published laboratory data. We compiled all studies with direct measurements of coccolithophore cellular POC content (11 studies)^{6-8,10-13,15,16,48,49}. This included combustion estimates, which were treated with an acid solution to remove PIC. We excluded studies that utilised combustion but did not specify an acid treatment, as it was unclear if such studies measured POC or a net cellular carbon content (PIC and POC) instead¹⁶. We also excluded studies that utilised wet oxidation treatments¹³.

Overall, our cellular POC data compilation included 11 taxonomic units and 42 observations. If available, for each study, we also compiled cell size estimates which were used for our allometric GLM scaling (7 studies, 11 taxonomic units, and 42 observations). The POC dataset is included as part of the Zenodo data archive ([zenodo.12794780](https://zenodo.org/record/12794780))²⁵.

Where available, direct measurements of cellular POC were resampled to create a merged estimate of cellular POC content. For a detailed description, please refer to the section Uncertainty Propagation.

Allometric scaling and cellular POC content estimation. For coccolithophore taxonomic units which did not have direct cellular POC content measurements, we applied an allometric scaling function to estimate the cellular POC content from cell size. For this, we created a new allometric scaling function by fitting a Generalised Linear Model (GLM) to a significantly larger number of coccolithophore observations from previous studies (42 vs 4 included in Menden-Deuer and Lessard⁵⁸ and 9 in Villiot *et al.*⁷). GLMs improve data fit compared to traditional linear regressions because GLMs can 1) constrain the predicted POC to be strictly positive by defining a Gamma distribution and 2) retain the training data's error distributions⁵⁹.

We used all available data points from our compilation (42 data points) representing 11 taxonomic units to fit a GLM allometric scaling for cellular POC content. We applied this GLM scaling to estimate the cellular POC content of taxonomic units with no direct estimates. The model has a mean prediction error of 77 pg C per cell and a root mean squared prediction error of 137 pg C per cell. Relative to the observed mean value of 67 pg C per cell, this constitutes relative MAE values of 113% and relative RMSE values of 200% (Fig. 4).

We compared the performance of the Gamma distributed GLM model with the more commonly used Gaussian distributed linear regression^{7,58}. The Gaussian distributed linear regression models did not perform as well with lower Cox and Snell pseudo-R-squared values (0.52 vs 0.69) and higher Akaike Information Criterion (AIC) information loss (539 vs 399) and was thus not used.

The allometric scaling slope of our GLM compares well with previous literature, with a slope of 0.72 when cell size and POC are plotted on a log10 scale (Fig. 4). This estimated slope falls between the previously reported scaling slope of 0.7 reported for coccolithophores⁷ and 0.95 previously reported for protists⁵⁸.

Cellular PIC content. We compiled estimates of cellular PIC from the literature. For taxonomic units with direct measurements, we resampled these observations to determine the mean and standard deviation in cellular PIC content for each measured taxonomic unit. To estimate PIC of taxonomic units with no direct measurements, we used an GLM allometric scaling of cellular PIC content, which we fitted based on the observed values.

In situ scanning electron microscope estimates of PIC. 56 of our taxonomic units have direct cellular PIC estimates based on Sheward *et al.*²³. Underlying assumptions and methodological descriptions of these estimates are provided in full detail in Sheward *et al.*²³. In brief, Sheward *et al.*²³ estimated the cellular PIC content of 56 taxonomic units using morphometrics. This method combines estimates of coccolith PIC content based on coccolith morphology and size⁶⁰ and then multiplies this estimate by the number of coccoliths observed per coccosphere to estimate the cellular PIC content. Coccolith size measurements and cellular counts were made using SEM images collected during the Atlantic Meridional Transect 14 cruise (28 April to 1 June 2004), which covered 47.03° S to 49.25° N⁶¹. The images ([zenodo.10571820](https://zenodo.org/record/10571820))⁵¹ and resulting dataset ([zenodo.11483788](https://zenodo.org/record/11483788))¹⁸ can be found on Zenodo.

Laboratory morphometric estimates of PIC. We also compiled morphometric estimates of cellular PIC content from the literature, which were available for three different taxonomic units: *Coccolithus pelagicus* HET, *Calcidiscus leptoporus* HET and *Helicosphaera carteri* HET^{21,22}. For *Calcidiscus leptoporus* HET, two sub-species of *Calcidiscus leptoporus* HET were measured: *Calcidiscus leptoporus* subsp. *quadriperforatus* HET and *Calcidiscus leptoporus* subsp. *leptoporus* HET. While some authors consider these subspecies to be a separate species^{62–64}, *C. leptoporus* HET morphotypes are hard to distinguish morphologically and are not distinguished in our *in situ* abundance compilation. We thus considered both *Calcidiscus leptoporus* subsp. *quadriperforatus* HET and *Calcidiscus leptoporus* subsp. *leptoporus* HET as a single taxonomic unit here. The original datasets ([PANGAEA.836841](https://pangaea.de/dataset/PANGAEA.836841))²¹ and ([PANGAEA.865403](https://pangaea.de/dataset/PANGAEA.865403))²² can be found on PANGAEA.

Laboratory combustion estimates of PIC. We did not include laboratory combustion estimates of PIC. Such measurements include discarded coccoliths and thus overestimate cellular PIC. While the contribution of discarded coccoliths depends on the taxonomic unit, since, for example, for the taxonomic unit *Calcidiscus leptoporus* HET the number of discarded coccoliths is low⁶⁵, the ratio between attached and discarded coccoliths can be over 6x for *Emiliania huxleyi* HET if the culture is not maintained at optimum conditions⁶⁶.

PIC allometric scaling. For taxonomic units which did not have direct cellular PIC content measurements, we followed a similar approach taken for POC, where we used a GLM allometric scaling function to estimate cellular PIC content from cell size. However, we also include the life cycle phase as a variable to estimate PIC. Life cycle phases strongly influence cellular PIC contents, with haploid morphologies generally containing much lower cellular PIC contents^{8,30–32}. To define the life cycle phase in the model, heterococcolith (HET), and Nannolith (NANO) life cycle phases were labelled as ‘diploid’ while ceratolith (CER), holococcolith (HOL) and polycrater (POL) were labelled as ‘haploid’. While HET cells might not be strictly diploid and HOL, CER and POL cells might not be strictly haploid³⁴, for simplicity, and since it will not influence PIC estimates, we refer to them as such here. Life cycle phases were included by one-hot-encoding the variables using the Pandas⁶⁷ `pd.get_dummies` function.

We fitted the GLM allometric scaling for the cellular PIC content based on 961 data points (925 diploid and 36 haploid) (Fig. 5). The models were used to estimate the cellular PIC content of 76 taxonomic units. The GLM PIC model has mean prediction error of the model is 13 pg C per cell, with a root mean squared prediction error of 34 pg C per cell. Relative to the observed mean value of 16 pg C per cell, this constitutes relative MAE values of 77% and relative RMSE values of 211%.

***In situ* abundance observations.** We compiled water-column observations of coccolithophore abundance from the literature (6,166 samples from 62 studies, Table 2 and Table 3, some of which have previously been included in coccolithophore abundance compilations^{6,28}, which were extracted from their respective PANGAEA data archives ([PANGAEA.785092](https://pangaea.de/dataset/PANGAEA.785092))⁶⁸ and ([PANGAEA.922933](https://pangaea.de/dataset/PANGAEA.922933))⁶⁹. The observations included in our coccolithophore abundance compilation are restricted to methods that can resolve taxonomic unit identity: Scanning Electron Microscopy (SEM) and light microscopy (brightfield and polarised, LM). We excluded observations from methods that cannot resolve taxonomic units, such as flow cytometric measurements and data records, where methods used for taxonomic identification were unspecified Table 1.

Previous literature suggests that LM underestimates total coccolithophore abundances relative to SEM counts^{70–72}. This is due to difficulty identifying lightly calcified taxonomic units with LM relative to SEM observations. To test if this holds for observations of the same taxonomic unit (instead of total coccolithophore abundance counts), we re-analysed data presented in Bollman *et al.*⁷⁰ who thoroughly compared taxonomic-unit-specific coccolithophore abundance estimates using both SEM and LM measurements from the same water column-samples. However, rather than comparing total coccolithophore abundances, we compared taxonomic-unit-specific observations only. Furthermore, we grouped abundances of sub-species and varieties. We estimated the percentage difference between LM and SEM abundance estimates for each sample and then conducted Bayesian bootstrapping to find the mean percentage difference. We found that abundance estimates

Reference	Method	Samples	Survey Period
Andruleit <i>et al.</i> ⁸⁶	SEM	71	1993–1993
Andruleit <i>et al.</i> ⁸⁷	SEM	21	2000–2000
Andruleit <i>et al.</i> ⁸⁸	SEM	45	1999–1999
Andruleit <i>et al.</i> ^{6†}	SEM	48	2001–2002
Baumann <i>et al.</i> ⁷⁸	SEM	98	1987–1995
Baumann <i>et al.</i> ⁸⁹	SEM	34	1993–1993
Boeckel <i>et al.</i> ⁹⁰	SEM	56	1998–2000
Bonomo <i>et al.</i> ⁹¹	SEM	92	2008–2008
Bonomo <i>et al.</i> ⁹²	SEM	40	2012–2012
Bonomo <i>et al.</i> ⁹³	SEM	120	2009–2009
Bonomo <i>et al.</i> ⁹⁴	SEM	301	2010–2014
Bonomo <i>et al.</i> ⁹⁵	SEM	133	2016–2016
Cepek <i>et al.</i> ⁹⁶	SEM	33	1993–1993
Cerino <i>et al.</i> ⁷²	SEM	77	2011–2013
Charalampopoulou <i>et al.</i> ⁹⁷	SEM	59	2008–2008
Charalampopoulou <i>et al.</i> ⁹⁸	SEM	89	2009–2009
Cortes <i>et al.</i> ⁸¹	SEM	171	1994–1996
Cros <i>et al.</i> ⁹⁹	SEM	113	1996–1996
D'Amario <i>et al.</i> ¹⁰⁰	SEM	44	2013–2013
Daniels <i>et al.</i> ³²	SEM	19	2012–2012
Dimiza <i>et al.</i> ⁵⁷	SEM	190	2001–2004
Dimiza <i>et al.</i> ¹⁰¹	SEM	99	2001–2013
Dimiza <i>et al.</i> ¹⁰²	SEM	12	2016–2016
Eynaud <i>et al.</i> ¹⁰³	SEM	44	1995–1995
Giraudeau <i>et al.</i> ¹⁰⁴	SEM	146	2014–2014
Godrijan <i>et al.</i> ⁷¹	SEM	23	2008–2009
Guerreiro <i>et al.</i> ¹⁰⁵	SEM	108	2010–2010
Guptha <i>et al.</i> ¹⁰⁶	SEM	12	1992–1992
Hagino <i>et al.</i> ¹⁰⁷	SEM	225	1964–1992
Haidar <i>et al.</i> ⁸⁰	SEM	193	1991–1994
Karatsolis <i>et al.</i> ¹⁰⁸	SEM	68	2013–2014
Keuter <i>et al.</i> ¹⁰⁹	SEM	38	2018–2019
Keuter <i>et al.</i> ¹¹⁰	SEM	196	2017–2019
Kinkel <i>et al.</i> ¹¹¹	SEM	40	1994–1996
Kleijne 1985 ^{112–114‡}	SEM	100	1985–1985
Luan <i>et al.</i> ¹¹⁵	SEM	52	2013–2013
Malinverno <i>et al.</i> ¹¹⁶	SEM	72	1997–1997
Malinverno <i>et al.</i> ¹¹⁷	SEM	13	2005–2005
Okada <i>et al.</i> ⁷⁵	SEM	226	1968–1969
Oviedo <i>et al.</i> ¹⁴	SEM	81	2011–2012
Patil <i>et al.</i> ¹¹⁸	SEM	48	2010–2010
Poulton <i>et al.</i> ⁶¹	SEM	137	2003–2005
Saavedra-Pellitero <i>et al.</i> ¹¹⁹	SEM	109	2010–2010
Schiebel <i>et al.</i> ¹²⁰	SEM	47	1995–1997
Schiebel <i>et al.</i> ¹²¹	SEM	47	2004–2004
Silver ^{6*}	SEM	13	2004–2004
Smith <i>et al.</i> ¹²²	SEM	25	2011–2012
Supraha <i>et al.</i> ¹²³	SEM	63	2013–2013
Takahashi <i>et al.</i> ¹²⁴	SEM	118	1996–1996

Table 2. Scanning Electron Microscopy (SEM) studies compiled as part of the abundance compilation. †Dataset published as part of O'Brien *et al.*⁶. ‡Dataset resulting from Snellius-II Cruise Gx (1985) which was published in part in several manuscripts.

from LM differ by $\approx 19\%$ (95% CI [1, 47%], Fig. 7). Thus, we did not consider the sampling method when converting our compilation of global abundances into POC and PIC stocks. Abundances were converted into POC and PIC by multiplying cellular carbon content distributions for each taxonomic unit by abundance values at each sample location.

Reference	Method	Samples	Survey Period
Acri <i>et al.</i> ¹²⁵	LM	39	2015–2016
Assmy <i>et al.</i> ¹²⁶	LM	28	2004–2004
Dimiza <i>et al.</i> ¹⁰¹	LM	9	2011–2011
Estrada <i>et al.</i> ¹²⁷	LM	229	1990–1990
Estrada <i>et al.</i> ⁷⁶	LM	400	2010–2011
Guerreiro <i>et al.</i> ⁷⁴	LM	150	2018–2018
Sal <i>et al.</i> ⁷³	LM	474	1992–2001
Valencia-Vila <i>et al.</i> ¹²⁸	LM	75	2000–2001
Estrada <i>et al.</i> ¹²⁹	LM	40	1985–1985
Grados <i>et al.</i> ¹³⁰	LM	9	1995–1995
Karentz ^{6†}	LM	1	2000–2000
Marshall <i>et al.</i> ¹³¹	LM	32	1965–1965
Widdicombe <i>et al.</i> ⁷⁹	LM	482	1992–2008

Table 3. Light Microscopy (LM) studies compiled as part of the abundance compilation. † dataset published as part of O'Brien *et al.*⁶.

The dataset includes data from 11 Atlantic Meridional Transect (AMT) cruises that occurred between 1995–2018. This data comes from three different publications: AMT 1–6⁷³, AMT 12, 14, 15 and 17⁶¹, and AMT 28⁷⁴. Additionally, we included all the sampling efforts conducted by Okada, Honjo, and McIntyre in the Pacific and Atlantic Oceans between 1968–1972^{41,75}, which were previously unavailable to the community. Other noteworthy datasets are data from Snellius-II Expedition⁶³, the Malaspina-2010 expedition⁷⁶, and a series of expeditions in the North Atlantic from 1987 to 1995^{77,78}. The dataset also contains data from several time series, such as the English Channel ‘L4’ time series⁷⁹, the Bermuda ‘BATS’ time series (1991–1994)⁸⁰, the Hawaii ‘HOTS’ time series (1994–1996, 2004)^{81,82}, and two a mesotrophic coastal ecosystem time series in the Adriatic Sea: ‘RV-001’ (2008–2009)⁷¹, and the ‘C1-LTER1’ (2011–2013)⁷².

Data grid and number of observations. The abundance compilation is provided in a non-gridded concatenated form and a gridded version with a 1-degree resolution with 5-m depth levels and a monthly temporal resolution. Gridding was conducted using Pandas in Python⁶⁷. Mean values were utilised for grid cells with more than one measurement.

The concatenated dataset contains 6,166 samples for 139 taxonomic units and 42,975 abundance observations. The gridded dataset comprises 4,307 samples and 33,119 abundance observations.

Spatial-temporal bias. Within our dataset, most of the samples were collected within the top 50 m of the water column and are biased towards the northern hemisphere and Atlantic Ocean (Fig. 2). Nonetheless, the Southern Ocean is well represented in our compilation, covering every major Ocean basin. The fewest samples are observed in the winter and below 100 m.

Data Records

The CASCADE dataset is available at Zenodo ([zenodo.12794780](https://zenodo.org/record/12794780))²⁵. The repository contains five main folders: 1) “Concatenated literature”, which contains the merged datasets of size, PIC and POC and which were corrected for taxonomic unit synonyms; 2) “Resampled cellular datasets”, which contains the resampled datasets of size, PIC and POC in long format as well as a summary table; 3) “Gridded datasets”, which contains gridded datasets of abundance, PIC and POC; 4) “Classification”, which contains YAML files with synonyms, family-level classifications, and life cycle phase associations and definitions; 5) “Species list”, which contains spreadsheets of the “common” (>20 obs) and “rare” (<20 obs) species and their number of observations. Within the data archive a README file is provided describing the directory structure in detail.

Technical Validation

Species misspellings and synonyms. Taxonomic units were checked for synonyms and misspellings following taxonomy as reported on the NannoTax3 website²⁴. During compilation, taxonomic units were checked to be in our database, and if flagged to be absent, they were added as synonyms or as new taxonomic units in the database.

Spatial-temporal location. Latitude, longitude, date, and depth were checked to be of the correct datatype and within the expected range.

Abundance observations. Cell concentrations were compared to the original publication and converted to cells per litre where appropriate.

Cell size. Where multiple size measurements were available for a single taxonomic unit, we compared cell size estimates between studies and flagged any taxonomic units which showed greater than 5-fold disagreement

between studies. Estimates for such taxonomic units were compared against the original literature, and dropped were appropriate. Dropped values included estimates of *Syracosphaera pulchra* HET and *Helicosphaera wallichii* HET from Villiot *et al.*⁷. Considerable disagreements in volume estimates were also observed for *Syracosphaera prolongata*; however, due to the unusual nature of the coccosphere morphology and the extensive range of size estimates reported in the literature, all estimates were kept in the compilation.

POC and PIC regression. We checked the performance of our POC GLM and PIC GLM for Mean Average Error (MAE), Root Mean Squared Error (RMSE), and Cox and Snell pseudo-R-squared. Size, POC and PIC values were checked to be strictly positive and finite before fitting the GLMs.

Code availability

All code used to generate and validate the dataset is publicly available on GitHub (<https://github.com/nanophyto/CASCADE>)⁸³. All pipelines were written in Python (3.11.4) and can be reproduced by running the Jupyter notebooks⁸⁴ included in the repository. The Python package dependencies required to reproduce CASCADE, are provided in a YAML file, and can be installed to an Anaconda environment⁸⁵ by following the instructions provided in the CASCADE GitHub README.

Received: 15 February 2024; Accepted: 30 July 2024;

Published online: 24 August 2024

References

- Ziveri, P., de Bernardi, B., Baumann, K. H., Stoll, H. M. & Mortyn, P. G. Sinking of coccolith carbonate and potential contribution to organic carbon ballasting in the deep ocean. *Deep-Sea Research Part II: Topical Studies in Oceanography* **54**, 659–675, <https://doi.org/10.1016/j.dsr2.2007.01.006> (2007).
- Engel, A., Szlosek, J., Abramson, L., Liu, Z. & Lee, C. Investigating the effect of ballasting by *caco3* in *emiliania huxleyi*: I. formation, settling velocities and physical properties of aggregates. *Deep-Sea Research Part II: Topical Studies in Oceanography* **56**, 1396–1407, <https://doi.org/10.1016/j.dsr2.2008.11.027> (2009).
- Zeebe, R. E. Loscar: Long-term ocean-atmosphere-sediment carbon cycle reservoir model v2.0.4. *Geoscientific Model Development* **5**, 149–166, <https://doi.org/10.5194/gmd-5-149-2012> (2012).
- Poulton, A. J., Adey, T. R., Balch, W. M. & Holligan, P. M. Relating coccolithophore calcification rates to phytoplankton community dynamics: Regional differences and implications for carbon export. *Deep-Sea Research Part II: Topical Studies in Oceanography* **54**, 538–557, <https://doi.org/10.1016/j.dsr2.2006.12.003> (2007).
- Poulton, A. J. *et al.* The 2008 *emiliania huxleyi* bloom along the patagonian shelf: Ecology, biogeochemistry, and cellular calcification. *Global Biogeochemical Cycles* **27**, 1023–1033, <https://doi.org/10.1002/2013GB004641> (2013).
- O'Brien, C. J. *et al.* Global marine plankton functional type biomass distributions: Coccolithophores. *Earth System Science Data* **5**, 259–276, <https://doi.org/10.5194/essd-5-259-2013> (2013).
- Villiot, N., Poulton, A. J., Butcher, E. T., Daniels, L. R. & Coggins, A. Allometry of carbon and nitrogen content and growth rate in a diverse range of coccolithophores. *Journal of Plankton Research* **43**, 511–526 (2021).
- Fiorini, S., Middelburg, J. J. & Gattuso, J. P. Testing the effects of elevated *pco2* on coccolithophores (prymnesiophyceae): Comparison between haploid and diploid life stages. *Journal of Phycology* **47**, 1281–1291, <https://doi.org/10.1111/j.1529-8817.2011.01080.x> (2011).
- Gafar, N., Eyre, B. & Schulz, K. Particulate inorganic to organic carbon production as a predictor for coccolithophorid sensitivity to ongoing ocean acidification. *Limnology and Oceanography Letters* **4**, 62–70 (2019).
- Gerecht, A. C., Šupraha, L., Langer, G. & Henderiks, J. Phosphorus limitation and heat stress decrease calcification in *emiliania huxleyi*. *Biogeosciences* **15**, 833–845 (2018).
- Langer, G. *et al.* Species-specific responses of calcifying algae to changing seawater carbonate chemistry. *Geochemistry, geophysics, geosystems*, **7** (2006).
- Llewellyn, C. & Gibb, S. Intra-class variability in the carbon, pigment and biomineral content of prymnesiophytes and diatoms. *Marine Ecology Progress Series* **193**, 33–44 (2000).
- Mullin, M., Sloan, P. & Eppley, R. Relationship between carbon content, cell volume, and area in phytoplankton. *Limnology and oceanography* **11**, 307–311 (1966).
- Oviedo, A., Ziveri, P., Álvarez, M. & Tanhua, T. Is coccolithophore distribution in the mediterranean sea related to seawater carbonate chemistry? *Ocean Science* **11**, 13–32, <https://doi.org/10.5194/os-11-13-2015> (2015).
- Šupraha, L., Gerecht, A. C., Probert, I. & Henderiks, J. Eco-physiological adaptation shapes the response of calcifying algae to nutrient limitation. *Scientific Reports* **5**, 16499 (2015).
- Verity, P. G. *et al.* Relationships between cell volume and the carbon and nitrogen content of marine photosynthetic nanoplankton. *Limnology and Oceanography* **37**, 1434–1446 (1992).
- Young, J. R. Extant coccolithophore cell size measurements. *Zenodo* <https://zenodo.org/doi/10.5281/zenodo.10572753> (2024).
- Sheward, R. M. & Poulton, A. J. Cellular morphological trait dataset for extant coccolithophores from the atlantic ocean. *Zenodo* <https://doi.org/10.5281/zenodo.11483788> (2024).
- Langer, G., Nehrke, G., Probert, I., Ly, J. & Ziveri, P. Strain-specific responses of *emiliania huxleyi* to changing seawater carbonate chemistry. *Biogeosciences* **6**, 2637–2646, <https://doi.org/10.5194/bg-6-2637-2009> (2009).
- Sett, S. *et al.* Temperature modulates coccolithophorid sensitivity of growth, photosynthesis and calcification to increasing seawater *pco2*. *PLoS ONE*, **9**, <https://doi.org/10.1371/journal.pone.0088308> (2014).
- Sheward, R. M., Daniels, C. J. & Gibbs, S. J. Growth rates and biometric measurements of coccolithophores (*Coccolithus pelagicus*, *Coccolithus braarudii*, *Emiliania huxleyi*) during experiments. *PANGAEA* <https://doi.org/10.1594/PANGAEA.836841> (2014).
- Sheward, R. M., Poulton, A. J., Gibbs, S. J., Daniels, C. J. & Bown, P. R. Coccosphere geometry measurements from culture experiments on the coccolithophore species *Calcidiscus leptoporus*, *Calcidiscus quadriporus* and *Helicosphaera carteri*. <https://doi.org/10.1594/PANGAEA.865403> (2016). Supplement to: Sheward, R. M. *et al.* Physiology regulates the relationship between coccosphere geometry and growth phase in coccolithophores. *Biogeosciences*, **14**(6), 1493–1509, <https://doi.org/10.5194/bg-14-1493-2017> (2017).
- Sheward, R. M. *et al.* Cellular morphological trait dataset for extant coccolithophores from the atlantic ocean. *Scientific Data* **11**, 720 (2024).
- Young, J. R., Bown, P., Howe, R. & Lees, J. Nanotax3 website. <https://www.mikrotax.org/> (2024).
- de Vries, J. *et al.* Coccolithophore abundance, size, carbon and distribution estimates (cascade). *Zenodo* <https://doi.org/10.5281/zenodo.12794779> (2024).
- Young, J. R. *et al.* A guide to extant coccolithophore taxonomy. *Journal of Nanoplankton Research*, 125 (2003).

27. Frada, M. J., Bendif, E. M., Keuter, S. & Probert, I. The private life of coccolithophores. *Perspectives in Phycology* **6**, 11–30, <https://doi.org/10.1127/pip/2018/0083> (2019).
28. de Vries, J. *et al.* Haplo-diplontic life cycle expands coccolithophore niche. *Biogeosciences* **18**, 1161–1184, <https://doi.org/10.5194/bg-18-1161-2021> (2021).
29. Dassow, P. V. & Montresor, M. Unveiling the mysteries of phytoplankton life cycles: Patterns and opportunities behind complexity. *Journal of Plankton Research* **33**, 3–12, <https://doi.org/10.1093/plankt/fbq137> (2011).
30. Cros, L., Kleijne, A., Zeltner, A., Billard, C. & Young, J. New examples of holococcolith–heterococcolith combination coccospheres and their implications for coccolithophorid biology. *Marine Micropaleontology* **39**, 1–34 (2000).
31. Fiorini, S., Middelburg, J. J. & Gattuso, J. P. Effects of elevated CO₂ partial pressure and temperature on the coccolithophore *Syracosphaera pulchra*. *Aquatic Microbial Ecology* **64**, 221–232, <https://doi.org/10.3354/ame01520> (2011).
32. Daniels, C. J. *et al.* Species-specific calcite production reveals coccolithus pelagicus as the key calcifier in the arctic ocean. *Marine Ecology Progress Series* **555**, 29–47, <https://doi.org/10.3354/meps11820> (2016).
33. de Vries, J., Monteiro, F., Langer, G., Brownlee, C. & Wheeler, G. A critical trade-off between nitrogen quota and growth allows *Coccolithus braarudii* life cycle phases' to exploit varying environment. *EGU Sphere* **2023**, 1–28, <https://doi.org/10.5194/egusphere-2023-880> (2023).
34. Frada, M. J. *et al.* Morphological switch to a resistant subpopulation in response to viral infection in the bloom-forming coccolithophore *emiliana huxleyi*. *PLoS Pathogens* **13**, 1–17, <https://doi.org/10.1371/journal.ppat.1006775> (2017).
35. Efron, B. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics: Methodology and distribution*, 569–593 (Springer, 1992).
36. Rubin, D. B. The Bayesian bootstrap. *The Annals of Statistics*, 130–134 (1981).
37. Cros, L. & Fortuño, J. M. Atlas of northwestern mediterranean coccolithophores. *Scientia Marina* **66**, 1–182 (2002).
38. Gaarder, K. R. & Ramsfjell, E. A new coccolithophorid from northern waters, *Calciopappus caudatus* n. gen., n. sp. *Nytt Magasin for Botanikk* **2**, 155–156 (1954).
39. Norris, R. E. Indian ocean nannoplankton. ii. holococcolithophorids (calyptrosphaeraceae, prymnesiophyceae) with a review of extant genera 1. *Journal of Phycology* **21**, 619–641 (1985).
40. Lecal-Schlauder, J. *Recherches morphologiques et biologiques sur les Coccolithophorides Nord-Africains* (Masson, 1951).
41. Okada, H. & McIntyre, A. Modern coccolithophores of the Pacific and North Atlantic oceans. *Micropaleontology*, 1–55 (1977).
42. Kleijne, A. & Cros, L. Ten new extant species of the coccolithophore *Syracosphaera* and a revised classification scheme for the genus. *Micropaleontology*, 425–462 (2009).
43. Thomsen, H. A., Buck, K. R., Coale, S. L., Garrison, D. L. & Gowing, M. M. Nanoplanktonic coccolithophorids (prymnesiophyceae, haptophyceae) from the Weddell Sea, Antarctica. *Nordic Journal of Botany* **8**, 419–436 (1988).
44. Borsetti, A. & Cati, F. Il nannoplankton calcareo vivente nel Tirreno centro-meridionale. *Giornale di Geologia* **43**, 157–174 (1978).
45. Triola, M. *Elementary Statistics 11th edition* (Addison Wesley, Boston, MA, 2010).
46. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362, <https://doi.org/10.1038/s41586-020-2649-2> (2020).
47. Seabold, S. & Perktold, J. statsmodels: Econometric and statistical modeling with Python. In *9th Python in Science Conference* (2010).
48. Sheward, R. M., Poulton, A. J., Gibbs, S. J., Daniels, C. J. & Bown, P. R. Physiology regulates the relationship between coccosphere geometry and growth phase in coccolithophores. *Biogeosciences* **14**, 1493–1509, <https://doi.org/10.5194/bg-14-1493-2017> (2017).
49. Oviedo, A. M., Langer, G. & Ziveri, P. Effect of phosphorus limitation on coccolith morphology and element ratios in mediterranean strains of the coccolithophore *emiliana huxleyi*. *Journal of Experimental Marine Biology and Ecology* **459**, 105–113 (2014).
50. Gibbs, S. J. *et al.* Species-specific growth response of coccolithophores to palaeocene - eocene environmental change. *Nature Geoscience* **6**, 10–14, <https://doi.org/10.1038/ngeo1719> (2013).
51. Poulton, A. J. & Sheward, R. M. Scanning electron microscopy images of the coccolithophore community during Atlantic meridional transect (AMT) 14. <https://doi.org/10.5281/zenodo.10571820> (2024).
52. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682 (2012).
53. Fiorini, S., Middelburg, J. J. & Gattuso, J.-P. Testing the effects of elevated pCO₂ on coccolithophores (prymnesiophyceae): Comparison between haploid and diploid life stages 1. *Journal of Phycology* **47**, 1281–1291 (2011).
54. Young, J. R. & Geisen, M. Xenospheres—associations of coccoliths resembling coccospheres. *Journal of Nannoplankton Research* **24**, 27–35 (2002).
55. Kamptner, E. Die Coccolithineen der Südküste von Istrien. *Annalen des Naturhistorischen Museums in Wien*, 54–149 (1940).
56. Triantaphyllou, M. & Dimiza, M. Verification of the algirosphaera robusta–sphaerocalyptra quadridentata (coccolithophores) life-cycle association. *Journal of Micropaleontology* **22**, 107–111 (2003).
57. Dimiza, M. D., Triantaphyllou, M. V. & Dermitzakis, M. D. Seasonality and ecology of living coccolithophores in eastern mediterranean coastal environment (Andros Island, Middle Aegean Sea). *Micropaleontology*, 159–175 (2008).
58. Menden-Deuer, S. & Lessard, E. J. Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton. *Limnology and Oceanography* **45**, 569–579, <https://doi.org/10.4319/lo.2000.45.3.0569> (2000).
59. Gelman, A. & Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research (Cambridge University Press, 2006).
60. Ziveri, P. *et al.* Stable isotope 'vital effects' in coccolith calcite. *Earth and Planetary Science Letters* **210**, 137–149, [https://doi.org/10.1016/S0012-821X\(03\)00101-8](https://doi.org/10.1016/S0012-821X(03)00101-8) (2003).
61. Poulton, A. J., Holligan, P. M., Charalampopoulou, A. & Adey, T. R. Coccolithophore ecology in the tropical and subtropical Atlantic Ocean: New perspectives from the Atlantic meridional transect (AMT) programme. *Progress in Oceanography* **158**, 150–170, <https://doi.org/10.1016/j.pocean.2017.01.003> (2017).
62. Sáez, A. G. *et al.* Pseudo-cryptic speciation in coccolithophores. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 7163–7168, <https://doi.org/10.1073/pnas.1132069100> (2003).
63. Kleijne, A. *Morphology, taxonomy and distribution of extant coccolithophorids (calcareous nannoplankton)* (Drukkerij FEBO BV, 1993).
64. Geisen, M. *et al.* Species level variation in coccolithophores. *Coccolithophores: from molecular processes to global impact*, 327–366 (2004).
65. Langer, G. *et al.* Calcium isotope fractionation during coccolith formation in *emiliana huxleyi*: Independence of growth and calcification rate. *Geochemistry, Geophysics, Geosystems* **8**, 1–11, <https://doi.org/10.1029/2006GC001422> (2007).
66. Rosas-Navarro, A., Langer, G. & Ziveri, P. Temperature effects on sinking velocity of different *emiliana huxleyi* strains. *Plos one* **13**, e0194386 (2018).
67. pandas development team, T. pandas-dev/pandas: Pandas, Zenodo <https://doi.org/10.5281/zenodo.10537285> (2024).
68. O'Brien, C. J. Global distributions of coccolithophores abundance and biomass - Gridded data product (NetCDF) - Contribution to the MAREDAT World Ocean Atlas of Plankton Functional Types. PANGAEA <https://doi.org/10.1594/PANGAEA.785092> (2012).
69. de Vries, J. C. *et al.* Global SEM coccolithophore abundance compilation. PANGAEA <https://doi.org/10.1594/PANGAEA.922933> (2020).

70. Bollmann, J. *et al.* Techniques for quantitative analyses of calcareous marine phytoplankton. *Marine Micropaleontology* **44**, 163–185, [https://doi.org/10.1016/S0377-8398\(01\)00040-8](https://doi.org/10.1016/S0377-8398(01)00040-8) (2002).
71. Godrijan, J., Young, J. R., Pfannkuchen, D. M., Precali, R. & Pfannkuchen, M. Coastal zones as important habitats of coccolithophores: A study of species diversity, succession, and life-cycle phases. *Limnology and Oceanography* **63**, 1692–1710, <https://doi.org/10.1002/lno.10801> (2018).
72. Cerino, F., Malinverno, E., Fornasaro, D., Kralj, M. & Cabrini, M. Coccolithophore diversity and dynamics at a coastal site in the gulf of trieste (northern adriatic sea). *Estuarine, Coastal and Shelf Science* **196**, 331–345, <https://doi.org/10.1016/j.ecss.2017.07.013> (2017).
73. Sal, S., López-Urrutia, Á., Irigoien, X., Harbour, D. S. & Harris, R. P. Marine microplankton diversity database. *Ecology* **94**, 1658–1658 (2013).
74. Guerreiro, C. V. *et al.* Response of coccolithophore communities to oceanographic and atmospheric processes across the north-and equatorial atlantic. *Frontiers in Marine Science* **10**, 1119488 (2023).
75. Okada, H. & Honjo, S. The distribution of oceanic coccolithophorids in the pacific. In *Deep Sea Research and Oceanographic Abstracts*, **20**, 355–374 (Elsevier, 1973).
76. Estrada, M. *et al.* Phytoplankton across tropical and subtropical regions of the atlantic, indian and pacific oceans. *PLoS One* **11**, e0151699 (2016).
77. Baumann, K. H., Andruleit, H., Schröder-Ritzrau, A. & Samtleben, C. Spatial and temporal dynamics of coccolithophore communities during low production phases in the norwegian-greenland sea. *Grzybowski Foundation Special Publication* **5**, 227–243 (1997).
78. Baumann, K. H., Andruleit, H. & Samtleben, C. Coccolithophores in the nordic seas: Comparison of living communities with surface sediment assemblages. *Deep-Sea Research Part II: Topical Studies in Oceanography* **47**, 1743–1772, [https://doi.org/10.1016/S0967-0645\(00\)00005-9](https://doi.org/10.1016/S0967-0645(00)00005-9) (2000).
79. Widdicombe, C. E., Eloire, D., Harbour, D., Harris, R. P. & Somerfield, P. J. Long-term phytoplankton community dynamics in the western english channel. *Journal of Plankton Research* **32**, 643–655, <https://doi.org/10.1016/j.plankt/fbp127> (2010).
80. Haidar, A. T. & Thierstein, H. R. Coccolithophore dynamics off bermuda (n. atlantic). *Deep-Sea Research Part II: Topical Studies in Oceanography* **48**, 1925–1956, [https://doi.org/10.1016/S0967-0645\(00\)00169-7](https://doi.org/10.1016/S0967-0645(00)00169-7) (2001).
81. Cortés, M. Y., Bollmann, J. & Thierstein, H. R. Coccolithophore ecology at the hot station. *Deep-Sea Research Part II: Topical Studies in Oceanography* **48**, 1957–1981, [https://doi.org/10.1016/S0967-0645\(00\)00165-X](https://doi.org/10.1016/S0967-0645(00)00165-X) (2001).
82. Silver, M. Vertigo km0414 phytoplankton species data and biomass data: abundance and fluxes from ctds (2009).
83. de Vries, J. & Wolf, L. J. nanophyto/cascade: v0.1.1, <https://doi.org/10.5281/zenodo.12797198> (2024).
84. Kluyver, T. *et al.* Jupyter notebooks - a publishing format for reproducible computational workflows. In Loizides, F. & Schmidt, B. (eds.) *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, 87–90 (IOS Press, Netherlands, 2016).
85. Anaconda, I. Anaconda (2024). Version 2024.02-1 (2024).
86. Andruleit, H., Stäger, S., Rogalla, U. & Čepeck, P. Living coccolithophores in the northern arabian sea: Ecological tolerances and environmental control. *Marine Micropaleontology* **49**, 157–181, [https://doi.org/10.1016/S0377-8398\(03\)00049-5](https://doi.org/10.1016/S0377-8398(03)00049-5) (2003).
87. Andruleit, H. Living coccolithophores recorded during the onset of upwelling conditions off oman in the western arabian sea. *Journal of Nannoplankton Research* **27**, 1–14 (2005).
88. Andruleit, H. Status of the java upwelling area (indian ocean) during the oligotrophic northern hemisphere winter monsoon season as revealed by coccolithophores. *Marine Micropaleontology* **64**, 36–51, <https://doi.org/10.1016/j.marmicro.2007.02.001> (2007).
89. Baumann, K., Boeckel, B. & Čepeck, M. Spatial distribution of living coccolithophores along an east- west transect in the subtropical south atlantic. *Journal of Nannoplankton Research* **30**, 9–21 (2008).
90. Boeckel, B. & Baumann, K. H. Vertical and lateral variations in coccolithophore community structure across the subtropical frontal zone in the south atlantic ocean. *Marine Micropaleontology* **67**, 255–273, <https://doi.org/10.1016/j.marmicro.2008.01.014> (2008).
91. Bonomo, S. *et al.* Living coccolithophores from the gulf of sirte (southern mediterranean sea) during the summer of 2008. *Micropaleontology* **58**, 487–503 (2012).
92. Coccolithophores from near the voltorno estuary (central tyrrhenian sea). *Marine Micropaleontology*, **111**, 26–37, <https://doi.org/10.1016/j.marmicro.2014.06.001> (2014).
93. Bonomo, S. *et al.* Living coccolithophores community from southern tyrrhenian sea (central mediterranean — summer 2009). *Marine Micropaleontology* **131**, 10–24, <https://doi.org/10.1016/j.marmicro.2017.02.002> (2017).
94. Bonomo, S. *et al.* Living and thanatocoenosis coccolithophore communities in a neritic area of the central tyrrhenian sea. *Marine Micropaleontology* **142**, 67–91, <https://doi.org/10.1016/j.marmicro.2018.06.003> (2018).
95. Bonomo, S., Schroeder, K., Cascella, A., Alberico, I. & Lirer, F. Living coccolithophore communities in the central mediterranean sea (summer 2016): Relations between ecology and oceanography. *Marine Micropaleontology* **165**, 101995 (2021).
96. Čepeck, M. Zeitliche und räumliche variationen von coccolithophoriden-gemeinschaften im subtropischen ost-atlantik: Untersuchungen an plankton, sinkstoffen und sedimenten (1996).
97. Charalampopoulou, A., Poulton, A. J., Tyrrell, T. & Lucas, M. I. Irradiance and ph affect coccolithophore community composition on a transect between the north sea and the arctic ocean. *Marine Ecology Progress Series* **431**, 25–43, <https://doi.org/10.3354/meps09140> (2011).
98. Charalampopoulou, A. *et al.* Environmental drivers of coccolithophore abundance and calcification across drake passage (southern ocean). *Biogeosciences* **13**, 5917–5935, <https://doi.org/10.5194/bg-13-5917-2016> (2016).
99. Cros, L. & Estrada, M. Holo-heterococcolithophore life cycles: Ecological implications. *Marine Ecology Progress Series* **492**, 57–68, <https://doi.org/10.3354/meps10473> (2013).
100. D’Amario, B., Ziveri, P., Grelaud, M., Oviedo, A. & Kralj, M. Coccolithophore haploid and diploid distribution patterns in the mediterranean sea: Can a haplo-diploid life cycle be advantageous under climate change? *Journal of Plankton Research* **39**, 781–794, <https://doi.org/10.1093/plankt/fbx044> (2017).
101. Dimiza, M. D. *et al.* The composition and distribution of living coccolithophores in the aegean sea (ne mediterranean). *Micropaleontology* **61**, 521–540 (2015).
102. Dimiza, M. *et al.* Seasonal living coccolithophore distribution in the enclosed coastal environments of the thessaloniki bay (thermaikos gulf, nw aegean sea). *Revue de micropaléontologie* **69**, 100449 (2020).
103. Eynaud, F., Giraudeau, J., Pichon, J. J. & Pudsey, C. J. Sea-surface distribution of coccolithophores, diatoms, silicoflagellates and dinoflagellates in the south atlantic ocean during the late austral summer 1995. *Deep-Sea Research Part I: Oceanographic Research Papers* **46**, 451–482, [https://doi.org/10.1016/S0967-0637\(98\)00079-X](https://doi.org/10.1016/S0967-0637(98)00079-X) (1999).
104. Giraudeau, J. *et al.* A survey of the summer coccolithophore community in the western barents sea. *Journal of Marine Systems* **158**, 93–105, <https://doi.org/10.1016/j.jmarsys.2016.02.012> (2016).
105. Guerreiro, C. *et al.* Late winter coccolithophore bloom off central portugal in response to river discharge and upwelling. *Continental Shelf Research* **59**, 65–83, <https://doi.org/10.1016/j.csr.2013.04.016> (2013).
106. Guptha, M. V., Mohan, R. & Muralinath, A. S. Living coccolithophorids from the arabian sea. *Rivista Italiana di Paleontologia e Stratigrafia* **100**, 551–573 (1995).
107. Hagino, K. & Okada, H. Intra- and infra-specific morphological variation in selected coccolithophore species in the equatorial and subequatorial pacific ocean. *Marine Micropaleontology* **58**, 184–206, <https://doi.org/10.1016/j.marmicro.2005.11.001> (2006).

108. Karatsolis, B. T. *et al.* Coccolithophore assemblage response to black sea water inflow into the north aegean sea (ne mediterranean). *Continental Shelf Research* **149**, 138–150, <https://doi.org/10.1016/j.csr.2016.12.005> (2017).
109. Keuter, S. *et al.* Seasonal patterns of coccolithophores in the ultra-oligotrophic south-east levantine basin, eastern mediterranean sea. *Marine Micropaleontology* **175**, 102153 (2022).
110. Keuter, S., Koplovitz, G., Torfstein, A. & Frada, M. J. Two-year seasonality (2017, 2018), export and long-term changes in coccolithophore communities in the subtropical ecosystem of the gulf of aqaba, red sea. *Deep Sea Research Part I: Oceanographic Research Papers* **191**, 103919 (2023).
111. Kinkel, H., Baumann, K. H. & Cepek, M. Coccolithophores in the equatorial atlantic ocean: Response to seasonal and late quaternary surface water variability. *Marine Micropaleontology* **39**, 87–112, [https://doi.org/10.1016/S0377-8398\(00\)00016-5](https://doi.org/10.1016/S0377-8398(00)00016-5) (2000).
112. Kleijne, A., Kroon, D. & Zevenboom, W. Phytoplankton and foraminiferal frequencies in northern indian ocean and red sea surface waters. *Netherlands Journal of Sea Research* **24**, 531–539 (1989).
113. Kleijne, A. Holococcolithophorids from the indian ocean, red sea, mediterranean sea and north atlantic ocean. *Marine micropaleontology* **17**, 1–76 (1991).
114. Kleijne, A. Extant rhabdosphaeraeaceae (coccolithophorids, class prymnesiophyceae) from the indian ocean, red sea, mediterranean sea and north atlantic ocean. *Scripta Geologica* **100**, 1–63 (1992).
115. Luan, Q., Liu, S., Zhou, F. & Wang, J. Living coccolithophore assemblages in the yellow and east china seas in response to physical processes during fall 2013. *Marine Micropaleontology* **123**, 29–40, <https://doi.org/10.1016/j.marmicro.2015.12.004> (2016).
116. Malinverno, E. Coccolithophorid distribution in the ionian sea and its relationship to eastern mediterranean circulation during late fall to early winter 1997. *Journal of Geophysical Research* **108**, 8115, <https://doi.org/10.1029/2002JC001346> (2003).
117. Malinverno, E., Triantaphyllou, M. V. & Dimiza, M. D. Coccolithophore assemblage distribution along a temperate to polar gradient in the west pacific sector of the southern ocean (january 2005). *Micropaleontology* **61**, 489–506, <https://doi.org/10.1007/BF01874407> (2015).
118. Patil, S. M. *et al.* Biogeographic distribution of extant coccolithophores in the indian sector of the southern ocean. *Marine Micropaleontology* **137**, 16–30, <https://doi.org/10.1016/j.marmicro.2017.08.002> (2017).
119. Saavedra-Pellitero, M., Baumann, K. H., Flores, J. A. & Gersonde, R. Biogeographic distribution of living coccolithophores in the pacific sector of the southern ocean. *Marine Micropaleontology* **109**, 1–20, <https://doi.org/10.1016/j.marmicro.2014.03.003> (2014).
120. Schiebel, R. *et al.* Distribution of diatoms, coccolithophores and planktic foraminifers along a trophic gradient during sw monsoon in the arabian sea. *Marine Micropaleontology* **51**, 345–371, <https://doi.org/10.1016/j.marmicro.2004.02.001> (2004).
121. Schiebel, R. *et al.* Spring coccolithophore production and dispersion in the temperate eastern north atlantic ocean. *Journal of Geophysical Research* **116**, 1–12, <https://doi.org/10.1029/2010JC006841> (2011).
122. Smith, H. E. *et al.* The influence of environmental variability on the biogeography of coccolithophores and diatoms in the great calcite belt. *Biogeosciences* **14**, 4905–4925, <https://doi.org/10.5194/bg-14-4905-2017> (2017).
123. Šupraha, L., Ljubešić, Z., Mišanović, H. & Henderiks, J. Coccolithophore life-cycle dynamics in a coastal mediterranean ecosystem: Seasonality and species-specific patterns. *Journal of Plankton Research* **38**, 1178–1193 (2016).
124. Takahashi, K. & Okada, H. Environmental control on the biogeography of modern coccolithophores in the southeastern indian ocean offshore of western australia. *Marine Micropaleontology* **39**, 73–86, [https://doi.org/10.1016/S0377-8398\(00\)00015-3](https://doi.org/10.1016/S0377-8398(00)00015-3) (2000).
125. Acri, F. *et al.* Lter northern adriatic sea (italy) marine data from 1965 to 2015 (version 3). *Zenodo* <https://doi.org/10.5281/ZENODO.3465097> (2019).
126. Assmy, P. Phytoplankton abundance measured on water bottle samples at stations ps65/424-3, 514-2, 570-4 and 587-1, <https://doi.org/10.1594/PANGAEA.603388>, <https://doi.org/10.1594/PANGAEA.603393>, <https://doi.org/10.1594/PANGAEA.603398> and <https://doi.org/10.1594/PANGAEA.603400> (2007).
127. Estrada, M., Varela, R. A., Salat, J., Cruzado, A. & Arias, E. Spatio-temporal variability of the winter phytoplankton distribution across the catalan and north balearic fronts (nw mediterranean). *Journal of Plankton Research*, **21** (1999).
128. Valencia-Vila, J., De Puellas, M. F., Jansá, J. & Varela, M. Phytoplankton composition in a neritic area of the balearic sea (western mediterranean). *Journal of the Marine Biological Association of the United Kingdom* **96**, 749–759 (2016).
129. Estrada, M. Phytoplankton assemblages across a nw mediterranean front: Changes from winter mixing to spring stratification. *Oecologia aquatica* **10**, 157–185 (1991).
130. Grados, C., Flores, G., Villanueva, P., Chang, F. & Ayón, P. Phytoplankton abundance at stations off Paita in August 1995, Piura, Peru, <https://doi.org/10.1594/PANGAEA.603265> and <https://doi.org/10.1594/PANGAEA.603267> (2007).
131. Marshall, H. G. Phytoplankton distribution off the north carolina coast. *American Midland Naturalist*, 241–257 (1969).

Acknowledgements

J.d.V., F.M.M., and L.J.W. were supported by UKRI NERC funding (Cocco Trait, NE/X001261/1). A.J.P. was supported by UKRI NERC funding (Cocco Trait, NE/X001261/1) and co-funding from the Horizon Europe Funding programme and UK Research and Innovation (OceanICU, agreement number 10054454). R.M.S. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project number 447581699. R.J. and P.Z. acknowledge the contribution to the Spanish MINECO project Global biodiversity of marine planktonic calcifiers (BIOCAL) (PID2020-113526RB-I00). Data derived from the Atlantic Meridional Transect was funded by the UK Natural Environment Research Council through its National Capability Long-term Single Centre Science Programme, Climate Linked Atlantic Sector Science (grant number NE/R015953/1). This study contributes to the international IMBeR project and is contribution number 405 of the AMT programme.

Author contributions

J.d.V. conceived the manuscript. J.d.V. and L.J.W. conceived the error propagation methods. J.d.V. and L.J.W. wrote, tested and validated the code. J.d.V. compiled the data. J.d.V. conducted the analysis. J.d.V. and A.J.P. compiled size data. A.J.P. compiled POC data. J.d.V. compiled PIC data. J.R.Y. and R.M.S. contributed size measurement data. R.M.S. contributed PIC measurement data. P.Z., and R.J. contributed abundance data from the Mediterranean Sea. K.H. contributed abundance data from the Okada cruises. J.d.V., F.M., A.J.P., R.M.S., and L.J.W. contributed to manuscript writing. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-024-03724-z>.

Correspondence and requests for materials should be addressed to J.d.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024