Review of “Pacific Southern Ocean coccolithophore-derived particulate inorganic carbon (PIC): A novel comparative analysis of in-situ and satellite-derived measurements”. By Saavedra-Pellitero et al. Submitted to EGUsphere

Synopsis: This paper evaluates the coccolithophores in 13 water samples from a transect south of New Zealand to Antarctica (Dec. 2004 to Jan. 2005) plus19 more water samples from a transect at the western end of Drake Passage (Feb.-March 2016). They made SEM images of the water samples, then calculated the PIC concentration of each sample using the morphometric properties of the coccoliths. Then they used level-3 satellite data to estimate PIC concentrations. These satellite data were binned to 4km-resolution and a mix of temporal binning (daily, 8-day average and one-month time-averaged satellite images but mostly one-month binning). They then concluded that Subantarctic waters and waters north of the Polar Front, the NASA PIC algorithm worked reasonably well. South of the Polar Front, however, there were few coccolithophores yet the satellite observations still suggested elevated PIC. The authors suspect that other reflective particles than PIC were causing the PIC algorithm to be in error.

General comments

This paper deals with an interesting topic, the coccolithophores of the Southern Ocean and what appear to be false, high-concentrations of coccolithophores as derived by satellite when the coccolithophore concentrations were observed to be low in their 32 water samples. Unfortunately, the paper is lacking in some fundamental key aspects and this reviewer cannot recommend publication it its current form. The key shortfalls are: (1) the NASA PIC algorithm measures the concentration of suspended PIC. Nowhere in the paper did the authors ever actually measure PIC analytically in order to judge the errors in the satellite approach, (2) the authors “measured” PIC using an approach based on the morphology of coccoliths and the numbers of coccoliths observed on coccolithophore cells under the SEM, calculating PIC volume. Nowhere did they use analytically-derived PIC estimates to examine the error of this SEM-based approach, (3) the remote sensing approach used here was inadequate for estimating PIC in these Southern Ocean waters, (4) their conclusion that the PIC algorithm does not perform well south of the Polar Front leads them to hypothesize that something else is causing the error. They hypothesized in the abstract that the reflectance signal from these waters was “due to the prevalence of small opal particles or unknown highly reflective particles (such as *Phaeocystis* aggregations) or suspended sediment “ and later in the paper, they mention other possible sources of errors such as microbubbles, floating loose ice or high concentrations of other particulate matter such as glacial flour”, which could cause errors. Unfortunately, barely any evidence was provided for these possibilities. I will elaborate on each of these points below.

Analytical PIC measurements

The gold-standard for measuring PIC in the ocean uses analytical measurement of either particulate Calcium (using any of several methods: particulate calcium analyses using flame photometry or optical emission spectrometry, or carbonate carbon measurements using coulometry). Any of these techniques would have been ideal to use to absolutely validate the PIC concentrations with quantifiable accuracy and precision.

PIC volume measurements using SEM

The technique that these authors used (Young & Ziveri, 2000), involves SEM imaging of filtered coccolithophores for the multiple morphotypes of *E. huxleyi* (the most abundant coccolithophore species in these waters). Using this laborious technique, they quantified the following

morphotypes for coccolithophores: Type A, Type A over-calcified, Type B, Type B/C, Type C, Type O single layered and Type O double-layered) as well as a few other coccolithophore species (Young & Ziveri, 2000). These measurements involve (for hundreds of cells) quantifying the length of the major and minor axes of the coccoliths, number of t-elements, ray width, tube width, and a shape factor. Each of these measurements will have an error, which would compound to the overall error of the coccolith volume, hence the PIC per cell. No such error bars were ever provided for the PIC measurements of coccolith morphometrics from their 32 samples. Moreover, it was critical to provide the accuracy and precision of the PIC morphometric measurements, measured against calibrated, analytical PIC measurements. As for the influence of diatoms in the Southern waters causing error in the satellite measurements of PIC, the only evidence was shown is four SEM images that contained diatoms. They never presented the abundance or biomass data of diatoms present in the samples, (nor an estimate of the biogenic silica associated with them, for that matter). It would have been nice to know some of the other features of these waters, too, such as chlorophyll concentrations, POC concentrations, TSS (total suspended sediments) to tease out what the mystery particles were causing the discrepancy.

Remote sensing approach

The authors used questionable remote sensing techniques to derive the PIC concentration. It is not clear exactly what they did here and more information is needed to assess this. They did say that their satellite PIC estimates were based on Level 3 data, highly binned satellite data in space and time. For example, in creating Level 3 data, daily, one kilometer resolution measurements would have been aggregated and averaged into 4km pixels. In addition, using data at 4km resolution could likely have included multiple overpasses over a day, especially given the high latitude. Multiple overpasses mean the radiances are associated with multiple azimuthal and nadir-viewing geometries. Why is this a problem? For algorithm development, one should always use Level 2 data. These data are recovered from a single overpass, first calculated from unprocessed instrument data (Level 0), to Level 1A data (reconstructed, unprocessed instrument counts), converted to calibrated water-leaving radiance data (Level 1B) at a given solar zenith and nadir viewing angle and those radiances are finally processed to geophysical variables as Level 2 data (e.g., PIC concentrations using the PIC algorithms at hand). Using already-binned, level 3 data can lead to spurious results for relating radiances to PIC due to the already-mentioned bin-averaged data, likely collected from multiple overpasses, and potentially applying an arithmetic average when the variables are actually log normal in distribution (requiring a geometric average) or vice versa. Algorithm validation should always be done with Level 2 data as per Bailey and Werdell (2006). There's too much binning that goes on to get from L2 to L3 to do an accurate comparison. It looks like, due to abundant cloudy skies in the Southern Ocean, these authors had to use monthly composites for the most part, to get the satellite PIC data to match with their ship samples. Bailey and Werdell (2006) recommend overpasses within ±3 hours of field observations (not even ±24 hours), in order to properly validate any ocean colour-derived variable. Here, the monthly level-3 composites would have been good to a temporal resolution of ±15d for the most part, hardly a “coincident” matchup! This is a long temporal difference to validate against. One can go from a phytoplankton bloom to background phytoplankton concentrations in just a few days, especially when grazers and viruses are devouring them! An error propagation calculation is essential, for all the measurements described in this paper (satellite, PIC volume), in order to assess the true errors.

Causes for PIC error in the satellite approach south of the Polar Front

On both transects, the authors saw a drop-off in the PIC volume calculations south of the Polar Front. However, the possible reasons are numerous, such as dense diatom populations, dense

Phaeocystis aggregations, floating ice crystals, bubbles or other suspended sediments, to name a few. Yes, they saw some diatoms in these waters south of the Polar Front (see their Fig. 9c and d) but upon close inspection, there were diatoms in the waters from north of the Polar Front, too (see their Fig. 9a and 9b). They also saw coccolithophores in their SEM from south of the Polar Front (see Fig. 9c). Nonetheless, they ultimately concluded that diatoms must be responsible for this error. Note, the SEM technique would never have visualized soft-bodied *Phaeocystis* in an identifiable form, ice crystals or bubbles, even if they were present. Suspended lithogenic sediments would have been visualized under SEM, however. Finally, foramifera can be abundant PIC sources in these Southern Ocean waters, too, contributing ~50% of the PIC in waters at 55-60^{o}S (Trull *et al.*, 2018) but they aptly point out that due to the optical properties of large PIC particles (they don't scatter light much per unit PIC), the satellite might not see their reflectance. But this still could be a source of error in their PIC volume calculation (which included only coccolithophores, not foraminifera or fragments of foraminfera). This paper desperately needs an error analysis so that the authors can put error bars on the PIC estimates that appear in the figures and tables.

Specific comments:
P1 L23: Spelling "Coccolith"
P1 L26: Which satellite product was used? What sensor?
P2 L47: They should also cite the earlier references who were really the first ones to show this (Holligan *et al.*, 1993; Robertson *et al.*, 1994).
P3 L89: change word "concentrations" to "measurements". Next sentence, change "Most" to "many" since there are a significant number of subpolar studies that enumerated free coccoliths.
P3 L94: See also (Young *et al.*, 2014)
P4 L99: The Oliver et al paper is now published (Oliver *et al.*, 2023) and it included non-bloom areas, too, I believe.
P4 L104: Does aim #1 include both plated coccolithophores as well as detached coccoliths?
P4 L117 and occurrences henceforth in the manuscript: In the field of physical oceanography, salinity data should not be reported with units of ‰.
P5 L135: Cite also (Holligan *et al.*, 2010)
P5 L143: R/V Italica should be italicized.
P5 L144: Is XX referring to the roman numeral or is this some sort of place-holder that needs to be filled in?
P5 L147-154: The authors are critical of how few stations are available in NASA's remote sensing validation of the PIC algorithm (n=42; noting that these are clear-sky matchups, using L2 satellite data within +/- 3 hours of an overpass (see general comments above). It should be noted, too, there is very strong under-sampling in this study, with a total of 32 water samples from which the paper is based, and these samples are based on Level 3, binned data which should not be used in satellite validation.
P6 L168: There are some details the authors might elaborate on. A SEM only views one side of a coccosphere. What about the coccoliths one can't see? Is this taken into consideration with the shape factors? It looks like they approximate the coccolith volume in equation 1 as a cube (L^3), why not a sphere?
P6 L186: For those not familiar with the Young and Ziveri method, could you say what "tube" you are talking about?
P7 L192: What fraction of your samples used daily, 8-daily or monthly PIC values? Was there a difference on how well the validation agreed when you used the shorter time scale validations? You should state this. I would suspect that since you show monthly images in your Figure 1, then most of your comparisons were done with monthly images. Is this correct? Your comment at line 204-205 would suggest that the number of daily match-ups was extremely low. Your table 2 lists these out but for the actual validation, if you had a daily, an 8-d and a monthly

measurement, which did you use? Please explicitly state this! Using 8-d composites or monthly composites means that your validations were hardly +/- 3h (the standard for true satellite validation (Bailey & Werdell, 2006)…see general comments above)! Also, which MODIS instrument did you use? Both were aloft during these cruises.
P7 L207-239- It is not clear what exactly the PCA analysis was for (as related to the validation of the PIC volume method). This should be stated in the lead sentence for that paragraph.
P8 L230-"Almost all" doesn't equal "solely"
P8 L233- What is the accuracy of the standard PIC algorithm? How does this compare to the accuracy of the morphometric approach? What is the absolute calibration?
standard for PIC?
P8 L234- Samples probably too small (volume-wise) to adequately sample forams and certainly not pteropods, other sources of Southern Ocean PIC . See Trull et al. (Trull *et al.*, 2018) for data on this. These can add significant PIC to the water.
P9 L261 I hope that in the Discussion, you compare this carbon per coccolith with other estimates (for which there are many).
P10 L301- And temporal scales, too!
P10 L307- Yes, but those are 1:1 clear-sky ship versus satellite comparisons using analytically-derived PIC concentrations. Here, the authors are using a technique with unquantified errors (coccolith-morphometrics) which are not calibrated to analytically-derived PIC. See general comments above.
P10 L310- Of course there are other particles! You are only seeing the diatoms and coccolithophores and maybe armored dinos in the SEM since they would preserve well on the filter. But there are huge abundances of purely organic cels like *Phaeocystis*, picoplankton and nanoplankton (without any sort of mineral coverings to make them visible in the SEM).
P11 L335-336- The attribution of light intensity and iron limitation are meaningless here since no such data were taken in this study (to know if it is even remotely feasible).
P11 L351-356- You desperately need an error analysis. You are presenting data from some 32 water samples (all surface, or some at depth?) The level-3 satellite data (which originate over the top two optical depths in a weighted fashion) are averaging over large pixel sizes. Did you pick just one pixel or the center pixel and surrounding 8 pixels? Most of the matchups were done with monthly averages (which may indeed have only resulted from a single overpass given the cloudiness in this region. A lot can happen to a population in one month! The potential errors on both sides are enormous. And you have no standard (analytical PIC measurements) on which to judge any of these. See general comments.
P15 L454-455- I'm sorry, but this is not a very definitive comparison, with no absolute standard analytical measurement.
P15 L458- Define "good". Nowhere do you ever plot one versus the other, nor present statistics.
P15 L460- How do you know that the morphometric algorithm is also not biased, especially when there is no comparison to gold-standard analytical PIC measurements.
P15 L467- But no where did you present the actual abundance of diatoms. What about the organic particles that you can't see with the SEM?
P15 L472- See Trull (Trull *et al.*, 2018) for elsewhere in the Southern Ocean. They show varying influence of other larger calcifiers!
P15 L485- Regarding future research, you never defined the accuracy of the morphometric method, nor discussed the accuracy of the satellite method.
Fig. 1 legend- What do big dots and little dots mean. Are the big dots the actual samples that you estimated coccolith morphometrics?
Fig. 3 legend- Are all these measurements from the surface? 3m depth? I don't recall seeing this. Please add error bars to the data when possible. I'm not sure these colors will necessarily be visible to a color-blind person. Please use correct color pallet. I'm not sure "electric blue" is

a good color descriptor to a color-blind person. What are the units of "tube width"? The triangle and diamond symbols are so small that one can't discern them.
Fig. 4 legend- Same comments as for Fig. 3
Fig. 5 - I'm confused for panel (C) since it shows 38 samples just for panel C (New Zealand transect) and 16 samples for the Drake Passage transect. These don't match your original numbers?
Fig. 6- panel b- confusion again over the numbers of samples in the New Zealand transect (here showing 13 samples) and Drake Passage transect (here showing 15 samples). Symbols in these figures are tiny! Make them bigger.
Fig. 7- I do not understand how this PCA figure advances the goal of the paper. Eliminate?
Fig. 8- Not sure a color-blind person will be able to interpret the colors in this. Can you add error bars to this given a thorough error analysis? What is "C-Calcita" referring to? PIC?
Fig. 9- I see diatoms in all four SEMs. Did you quantify them in any way? I see plated coccolithophores and coccoliths in panel c (from south of the Polar Front). What does this mean for the hypothesis?

Summary
In summary, there are so many unquantified errors in these measurements (for both coccolith morphometrics and satellite PIC) that this reviewer cannot recommend publication of this paper in its current form. If it is to be published, it will require a major revision.

References

Bailey, S. W., & Werdell, P. J. (2006), A multi-sensor approach for the on-orbit validation of ocean color satellite data products, *Rem. Sens. Environ.*, *102*, 12-23.

Holligan, P. M., Charalampopoulou, A., & Hutson, R. (2010), Seasonal distributions of the coccolithophore, *Emiliania huxleyi*, and of particulate inorganic carbon in surface waters of the Scotia Sea, *Journal of Marine Systems*, *82*(4), 195-205.

Holligan, P. M., et al. (1993), A biogeochemical study of the coccolithophore, *Emiliania huxleyi*, in the north Atlantic, *Global Biogeochem. Cycles*, *7*(4), 879-900.

Oliver, H., McGillicuddy, J., D. .J. , Krumhardt, K. M., Long, M. C., Bates, N. R., Bowler, B. C., Drapeau, D. T., & Balch, W. M. (2023), Environmental drivers of coccolithophore growth in the Pacific sector of the Southern Ocean *Global Biogeochem. Cycles*, *37*(11), https://doi.org/10.1029/2023GB007751.

Robertson, J. E., Robinson, C., Turner, D. R., Holligan, P., Watson, A. J., Boyd, P., Fernandez, E., & Finch, M. (1994), The impact of a coccolithophore bloom on oceanic carbon uptake in the northeast Atlantic during summer 1991, *Deep-Sea Research Part I*, *41*(2), 297-314, doi:10.1016/0967-0637(94)90005-1.

Trull, T. W., Passmore, A., Davies, D. M., Smit, T., Berry, K., & Tilbrook, B. (2018), Distribution of planktonic biogenic carbonate organisms in the Southern Ocean south of Australia: A baseline for ocean acidification impact assessment, *Biogeosciences*, *15*(1), 31-49, doi:10.5194/bg-15-31-2018.

Young, J. R., Poulton, A. J., & Tyrrell, T. (2014), Morphology of Emiliania huxleyi coccoliths on the northwestern European shelf - Is there an influence of carbonate chemistry?, *Biogeosciences*, *11*(17), 4771-4782, doi:10.5194/bg-11-4771-2014.

Young, J. R., & Ziveri, P. (2000), Calculation of coccolith volume and its use in calibration of carbonate flux estimates, *Deep-Sea Research Part II: Topical Studies in Oceanography*, *47*(9-11), 1679-1700, doi:10.1016/S0967-0645(00)00003-5.