

Line (initial)	Comment	Reviewer	Response (revision or comment)	Revised Line
Section 2.6	In section 2.6 it is unclear how the trends are calculated. A linear regression of the imputed and observed time series of monthly means? If so, the results may not be applicable to the most common approach currently used in the ocean carbon community, which is to apply a linear regression to deseasoned monthly means, where seasonal variability is removed (Bates 2001; Bates et al. 2014; Takahashi et al. 2009). In this approach, gaps are filled if months are missing in the climatological monthly means used to remove the seasonal signal (see Figure 2 in Takahashi et al. 2009), but data gaps within the resulting time series of deseasoned monthly means are not (see Figure 3 in Takahashi et al. 2009). If the trends presented by Vance et al. are on data that include additional noise from seasonal variability, would the impact of different gap-filling techniques on resulting trends be the same for the deseasoned approach where seasonal variability is removed?		We have applied the method of Takahashi et al. 2009 to seasonally detrend the time series before applying the linear regression to establish the long term trends for observations and comparison to imputed time series. With this we have updated Section 2.6, adding Equation 10, and updated associated figures and results.	225
NA	Many of the figure titles and labels in the supplemental need larger fonts.		Supplemental figures have been revised with larger font and sizing that improves visibility.	
301	It may be useful if the authors could describe why these measurement uncertainties were chosen. Also, is μ_{matm} supposed to be μ_{atm} ?		Our reasoning for these uncertainties was based on typical performance for field and lab measurements. We have added clarification to this section for these selections, with some references and fixed any typos.	255
340-345	How these uncertainty assessments were done are not clear. For example, the beginning of this paragraph refers to measurement uncertainty, but line 340 refers to uncertainty of monthly means. In the next sentence, the authors say that annual data from the WHOTS mooring are used to estimate uncertainty for HOT data, but it's unclear what "annual data" means when mooring data are 3-hourly and why moored pCO ₂ data would be used to determine uncertainty of HOT measurements of DIC.		We agree that this was not clearly communicated as written. The word annual was misleading here as it was referring to using a year of data. The point here though was to evaluate the uncertainty associated with averaging mooring data to monthly values as well as estimating the uncertainty associated with treating individual samples as monthly averages. WHOTS pCO ₂ data was used to estimate the daily variability in DIC at HOT and served as a proxy to estimate the uncertainty associated with monthly averaging. This combined with KE0 and Papa provided a narrow range, from which we took the upper limit and applied to all sites. We have revised this section for clarity.	Section 2.7
655	"lease" should likely be "least"		1 done	
Fig 11	One of the labels for the y axes on the right is cut off.		1 done	Fig 11
Fig 14	It's hard to tell the difference between the grays. Why not use the same color scheme as kernel density curves?		1 This figure was updated to match the color scheme used in other figures for visual consistency as suggested.	Fig 14
NA	Overall I think the authors have done a very good job at responding to the issues raised by the reviewers. I'm happy to accept it in its present form, given the small issues highlighted below are amended. One preference I would still have though is for Figures 10 and 13 to present the anomalies, rather than leaving the reader to try and visualise them themselves. This is particularly as there are so many panels on the figures with so many lines on each panel - using the majority of the real estate for showing the observational variation doesn't appear to me to be the most efficient use of space.		2 Figures 10 and 13 were split into A and B parts and revised changed to residual plots as suggested.	
16	...annual budgets [and] interannual and climatic variability"		2 done	16
35	"...over varied durations and may [be] trained with either in-situ..."		2 done	35
152 and thro	but also throughout. When citing previous studies inline, I think the citation style should be "...from Lueker et al., (2000) rather than (Lueker, 2000)". Similarly "...Kf from (Dickson, 1979)..." should be replaced by "...Kf from Dickson (1979)..."		2 done	NA
158	should be (O'Reilly et al., 1998)		2 Fixed - this was an EndNote MS Word format typo (correctly referenced in library)	
158-160	- need citations for MODIS and VIIRS here		2 Citations and acknowledgements were added for MODIS and VIIRS data.	Section 2.2
189	should be either 'measurement is' or 'measurements are'		2 done	
257	typo for absolute		2 done	
296	should be equation 10?		2 Equation numbering was fixed	
337	should be equation 11?		2 Equation numbering was fixed	
339	measurements rather than measurement		2 done	
348	should be equation 9 instead of equation 8? In fact, check all references to equations as these seem to be going awry up to this point.		2 Equation numbering was fixed	
510	Figure 3, could you explicitly put in the legend which colour is which? You can derive it from the differences in the trends of course, but it wouldn't hurt to include it.		2 Added legend here and additional visualization to indicate time series were truncated to Sept. 1997 coincident with remotely sensed chlorophyll records.	Fig. 3
567	Kernel density curves. For readers that haven't come across these curves before, I'd recommend adding a line describing what they show, and what the optimal should be. (You have this in Fig 7 & 11 captions, just not in the text)		2 We added language here that was consistent with the captions for Figs. 7 & 11.	361-362

616	Fig 10. There is a lot of information on this plot. Maybe have it over two pages, with three locations on each page? I still believe that anomalies from the observed of the timeseries would be more powerful and easier to interpret, showing the strengths of the different methods over different data gaps more clearly (this applies to the seasonal cycles shown in Fig 13 too, these are already shown in Fig 2 for observations, so showing anomalies from the observed for each imputation method would be easier to interpret).		Figures 10 and 13 were split into A and B parts and revised changed to residual plots as suggested. The composite time series versions from this revision were retained but moved into split 3-sided figures in the Supplemental Materials.	
697	typo of althought	2	done	
881	... limited evaluation [of] errors...	2	done	
899	change 'less than' to 'less that'	2	done	
965	Should it be 'These were less than..'	2	done	
968	equating instead of equate	2	done	
880	It might be worth here stating what is thought to be the physical cause of the difference in performance given the different levels of missing data. Is it that the extremes of temperature and their DIC concentrations need to be captured so as to best enable the different imputation techniques (as temperature has the greatest correlation with DIC)? Or is it something else?	2	It is not clear if this comment / question is in reference to only the MLR performance across sites or the performance of each imputation model across sites, nor the distinction between physical causes for performance and missing levels of data. In either case, our interpretation of this comment would require additional analysis to appropriately answer without conjecture.	NA
1133	'with acceptable accuracy' rather than 'with acceptable accurately'	2	done	