I thank Leseurre et al for providing the revised manuscript. While clearly improved, there are still some issues that needs to be dealt with before I can recommend that the manuscript is accepted for publication.

Comments on responses to major issues 1-10 in initial review.

1. *Testing the scheme to reconstruct February and July data.*

   I proposed to test this scheme using the high frequency SST and SSS data collected by the TSG. This has not been done. While a figure showing the trends in higher frequency TSG SST and SSS data have been added to the manuscript (Fig 3), this does not constitute an actual test of the reconstruction scheme. Please add and discuss a side-by-side comparison of trends in SST and SSS as measured by the TSG in February and July, vs. those shown in Fig 4 based on a mixture of measured and reconstructed data. This could be in the form of a figure or a table.

   I see that there are clear discrepancies between the SST and SSS trends in Fig 3 and 4 for some of the time periods. The 1993-1997 period, the data in Fig 3 shows increasing SST and SSS, while for the corresponding period in Fig 4, SST has no trend and SSS is decreasing. The winter SST trends for the second period (2001-2007) are also different (increasing in Fig 3 decreasing in Fig 4).

2. *Comparison with trends in mapped products*

   I asked for a comparison with SST trends from the NOAA objectively analysed product; this would shed light on how representative the data (in Fig. 4) are of large-scale interannual phenomena. In their response the authors states this will be included: "We will also use SST reanalysis to confirm the trends for the selected periods". This has not been included (i.e. comparison to regional SST trends from NOAA OI SST product (https://climatedataguide.ucar.edu/climate-data/sst-data-noaa-optimal-interpolation-oi-sst-analysis-version-2-oisstv2-1x1)

Overall, for comment 1 & 2. I would recommend to add a subsection in the Results section, between current 3.1 and 3.2, that compares the SST and SSS trends as presented in Fig 4, with those from higher frequency TSG data and also regional SST trends from the NOAA OI product. This would, before anything else, inform readers about how accurate the trends shown in Fig 4 are, and the extent to which they represent larger scale interannual variations.

3. *pCO_2 driver decomposition*

   It is good to see that a decomposition of the drivers of pCO_2 and pH trends have now been included. However, please include citations for the equation (e.g. Keeling et al (2004) and Fröb et al. (2019)). Also, it is not clear what parts the columns for DIC/sDIC and TA/sTA in Fig 5 show.

Specifically, is DIC the total DIC driver while sDIC is the part of the DIC driver not related to salinity? Please specify this in the caption.

4. *Trends in salinity normalised DIC and TA*

   I am happy with the response and actions

5. *Inclusion of figure with trends in air-sea pCO2 difference*

   This is fine

6. *Uncertainties*

   Please explain in the methods section how the uncertainty of the trends presented in Fig 4 was determined.

7. *Trend 2001-2008 and reconstructed data*

   OK, would be worthwhile to mention this

8. *Anthropogenic vs natural carbon*

   This has been addressed nicely

9. *Discussion*

   This is now very exhaustive

10. *Long term trend*

    OK

New comments

Page 1 line 18: 'the trend of …CO2….is slightly less than the atmospheric….and the pH decrease.' Odd sentence, please revise.

Page 1 line 28: Consider using 'multiannual' instead of 'pluriannual'

Page 2 line 3: replace 'indices' with 'values'

Page 2 line 33-43: The paper by Goris et al: Constraining projection-based estimates of the future North Atlantic carbon uptake, *Journal of Climate, 31*, 3959-3978, 2018, shows clearly that many ESM struggle to represent mixing correctly in the NASPG, which render them unable to reproduce the actual trends.

Page 3 line 7-9: It is not so much the shoaling of the lysocline that is the problem, but that these northern surface waters have initially low $CaCO_3$ saturation because of high natural (preindustrial) DIC concentrations because of low temperatures giving high CO_2 solubility. Please revise.

Page 3 line 35: Replace 'conducts' with 'conduits'

Page 5, line 5: State the pressure used, i.e. 1013.25 hPa

Page 5: Please add a subsection on how uncertainties were determined: the uncertainties of the trend slopes, the uncertainties of the driver decomposition (i.e. the 1000 random perturbations) and any other uncertainty calculation. For the uncertainties in the trend regression, how did you deal with the uncertainty of each point in the y direction (e.g. the SST values for each year has a standard deviation associated with them shown as error bars in Fig 4(a), how was this accounted for when calculating the regression and its uncertainty? As far as I know this is not straightforward).

Page 6 line 5: You used the seasonal climatology to adjust data from January/March and June/August to February and July, not 'to complement winter and summer'. Please state here the typical magnitude of these adjustments.

Page 6 line 16-17: Salinity is also low in the green box, but DIC is not correspondingly lower. I.e. the low DIC in the southern (red) box are not only a result of the low salinities, please explain more.

Page 6 line 30: 'Section (3.3)' should be 'Section 2.3'

Page 6: It would be good to include the uncertainty of the trends that are stated.

Page 7: Use of TrOCA method. The TrOCA method has issues and it is not very accurate. See for example https://www.biogeosciences.net/7/723/2010/bg-7-723-2010.pdf. Some cautionary remarks would be appropriate, for example that you believe it is robust enough to determine Cant trends even if the absolute values are questionable.

Page 7 line 7: The DIC trend of 0.9 from GLODAPv2 – was this something you calculated. What were the lat/lon bounds

Page 7 line 7: GLODAPv2, not GLODAP-V2

Page 7 line 16: replace 'very homogeneous' with 'a'

Page 7 lines 11-18. I don't understand some of these numbers. From Gruber et al., 2019 you find a Cant increase between 1994 and 2007 of 8.5 μmol/kg. This gives a trend of 0.65 μmol/kg yr. You assume this applies for the surface as well, to give a total increase of 10.1 μmol/kg for 1994-2007. But why the change?, if the subsurface trend from Gruber applies in the surface layer, this would give 8.5 μmol/kg there as well, not 10.1 μmol/kg.

My sense is that the Cant increase from Gruber et al (0.65 μmol/kg yr) is similar to the one you estimate using TrOCA (0.6 μmol/kg yr), but both of these are lower than the observed DIC increase in SURATLANT (0.7-0.9 μmol/kg yr) and estimate from GLODAPv2 (0.9 μmol/kg yr). So altogether, the DIC is increasing slightly faster than explained by Cant alone. Please revise and simplify.

Page 8 line 11: Please simplify to: Because interannual variability is more pronounced in summer than in winter because of the added influence of biological activity.

Page 8 line 14-15: Please include citations to these studies that have evaluated trends in different seasons.

Page 9 line 16: Mention the source of these independent pCO2 observations (ship, principal investigators, cite any papers by PIs that describe the data).

Page 10 line 14: 'last decade , 2001-2017…', better with 'last period, 2008-2017 for summer and 2008-2015 for winter'.

Page 10 line 29-30: The case for low primary production in 1996, is very weak. 1996 was characterised by very negative NAO index. Another year with very low NAO index was 2010, and in that year primary production was exceptionally high. So if anything – by analogy – I would expect strong primary production in 1996.

Page 11 line 2-4: Attribution of summer 2005 - 2007 high fCO2 values. Low productivity and/or deep (winter) mixing suggested. This can and should be checked with satellite Chl and Argo MLD data. For the record, Fröb et al., 2016 (Irminger Sea deep convection injects oxygen and anthropogenic carbon to the ocean interior, *Nature Communications*, 7:13244, 2016) shows that winter mixing was *not* deep in these years in the central-west Irminger Sea. Therefore primary production is the more likely candidate.

Page 11, line 18-20. This is cherry picking, what about the earlier time periods when the discrete data does not reflect the trends in higher frequency TSG data (such as winter 1993-1998 noted earlier in this review)? The extent to which these discrete samples reflect larger scale variations should be investigated using a systematic approach – not an ad hoc as is currently done.

Page 11 line 31. Write as GLODAPv2.2019

Page 12 line 32-33. This is also related to the way models do/don't represent winter mixing, see earlier cited Goris et al paper.

Page 13 line 10. Last sentence needs revision.

Page 14 lines 1-10. The 2010 primary production event is nicely discussed in Henson et al, Unusual subpolar North Atlantic phytoplankton bloom in 2010, JGR, 2013

Page 14 line 23-24. Please name ship (Skogafoss) and principal investigator (Wanninkhof NOAA/AOML) of these data (it is always nice to give specific and concrete credit to the correct people)

Page 14-15, 'Such large changes in pH trends … would not have been resolved when using fCO2 data and TA/S relation instead of TA measurements'. I am not convinced this is actually the case. pH determined from fCO2 and TA is not very sensitive to TA, but depends most on fCO2, which in itself reflects underlying DIC/TA changes; pH and fCO2 are strongly correlated. I think such large changes could be visible in pH determined from fCO2 and TA (from TA-S relationship). You may want to check with a simple calculation before making such a statement. (i.e. combine the calculated fCO2 with TA from TA-S relationship and calculate pH trends.)

Page 15, line 11 You state 'indirect methods' but refer to only one, Denvil-Sommer et al., 2019. This generalisation does not seem correct unless you compare with other mapped pCO2/flux products.