Summary of Revisions

RC = Reviewer comment

We thank all three reviewers for their positive and constructive feedback. In order to provide a quick overview of the changes to the to-be-revised manuscript, we give a summary here:

- The title has been changed to: “Exploring the ocean and atmosphere coupled system with a data science approach applied to observations from the Antarctic Circumnavigation Expedition” (following RC3.3).
- We have added research questions in the introduction for a framework that better structures the manuscript as a whole (following RC1.6).
- The methods description has been revised substantially to make the language more accessible to non-data scientists (following the general and several targeted comments of Reviewer #1).
- Section 5 (description of individual LVs) will be moved to a new appendix A to substantially shorten the manuscript. We now summarize the outcome of all LVs briefly in a revised section 4.1, and highlight the novel aspects we found there as well. We give give one condensed description of LV9 as example in a revised section 4.2. (following RC1.7, 3.1, 3.4)

RC = Reviewer Comment, AC = Author Comment, new suggested text in blue

Answer to Reviewer 3

Anonymous Referee #3, 02 Jun 2021

General comments:

RC3.1: I find this paper uses an interesting approach that has a potentially high value and high impact for the ocean-atmosphere interdisciplinary research community. The paper takes the observations from the Antarctic Circumnavigation Expedition (ACE, austral summer 2016/2017) cruise and combines them with a sparse Principal Component Analysis (sPCA) to understand how different observed variables are linked together and to the general context (e.g. distance from land, cyclone activity, etc.). The paper is also very long, which makes reading and understanding the entire content of the paper and really getting into the new conclusions that result from this study extremely difficult.

I support this paper as a proof of concept for this approach, but I find the science questions posed (or hypotheses) and conclusions in the study are very weak. This paper should be published after the comments from the other reviewers and the comments below are addressed.

AC3.1: We thank the reviewer for the positive and constructive feedback. We appreciate the reviewer's remark that the approach presented in our paper might be a valuable addition to the way that our community analyses large, heterogeneous data sets. We also agree that the paper is very long, which might be a disadvantage to clearly communicate our message. Therefore, we have taken multiple measures (see summary of revisions and detailed responses) to shorten the paper (e.g. moved the individual presentation of all 14 LVs to the appendix A). In addition, we have taken an effort to more clearly state the science questions (see AC3.2 and AC1.6) and made a dedicated effort to clearly state the novel results (in the new section 4.1, see AC3.4 and AC1.7) and the advantages of the sPCA (in the new section 3.5, see AC1.2). We hope that these changes will help to better bring across our key messages. Our detailed responses are provided below.

Major comments:

RC3.2: Most of the conclusions made using this very complex analysis are simplified statements of well known phenomena. So, I'm not sure what is the added value of this approach compared to what is already known. This is seen in the various "In summary" statements that come at the end of each section that focuses on the latent variables (LVs). This is seen most clearly in the summary for LV7 and LV10, which mostly put things into a seasonal and diurnal cycle context. I do not see what we have learned by using this "data science" approach. One way to address this would be to acknowledge in the abstract and very early in the study that there are no main scientific conclusions using data science in this study, but that this sets up the methodology that can be used in the future for this purpose.

AC3.2: We agree that the value of this manuscript lies first and foremost in setting up the method for future studies, which were designed a priori around interdisciplinary research questions. The sPCA fills an important gap in this regard, because it is more powerful than simple correlation analysis, and it allows to relate a large number of variables, which reflect processes of different time scales and at a level of detail that comprehensive Earth System Models cannot address.

One of the key aspects of this analysis is that it provides the possibility for an untargeted, and therefore more objective and unbiased, analysis, whereas traditional methods are often biased by a certain method that is tailored to a specific question. We should have pointed this aspect out more clearly and also more clearly state which of the results are novel and which are well known aspects.

We now highlight these new aspects that the analysis was able to depict in section 4.1, the abstract, and conclusions. They include:

- New insights into the Southern Ocean water cycle, where surprisingly, our large-scale assessment of concurrent precipitation and salinity measurements does not yield a direct response of the surface ocean salinity to precipitation events. Instead, we here show that variations in surface ocean salinity are driven by the climatological (long-term) patterns set by surface freshwater fluxes integrated over time-scales longer than synoptic events (LV1) and seasonal melting on sea ice (LV9).
- We also find a latitudinal distribution of the nutrient availability and its effect on the productivity, which is highlighted in LV11, LV6 and LV8. This shows, at the largest scale ever reported, nutrient limitation regimes for the subantarctic front, south of the polar front and associated with the island mass effect as previously reported.
- The sPCA produced unexpected results for some of the reactive trace gases, notably isoprene (LV7). This result points towards a complex interplay between the seasonality of emissions (sources) and seasonality of oxidation pathways (sinks), which, coupled with the potential effect of transport from terrestrial sources, paint a very complex picture for atmospheric isoprene in the Southern Ocean.

We also think that this contribution has provided a valuable overview of Southern Ocean processes on different time and spatial scales. In addition, the published datasets are a benchmark for the current state of the Southern Ocean, against which data in several years or decades time can be compared. We have added the following to the abstract in l. 11ff:

"Our results provide a proof of concept that sPCA with uncertainty analysis is able to identify temporal patterns from diurnal to seasonal cycles, as well as geographical gradients and "hotspots" of interaction between environmental compartments. While confirming many well known processes, our analysis provides novel insights into the Southern Ocean water cycle (freshwater fluxes), trace gases (interplay between seasonality, sources and sinks), and microbial community (nutrient limitation and mass island effects at the largest scale ever reported). Our results establish…"

And in l. 24:

"It thereby fills an important gap between simple correlation analyses and complex Earth System Models. The former would not be able to relate such a large number of variables, while the latter is less constrained by observations and comes with analytical challenges to depict single processes."

And in the introduction in l.42 (This addition in l. 42 is also a response to reviewer comment RC1.6):

"To explore interactions between the Southern Ocean system components, we apply an unsupervised learning method, sparse principal component analysis (sPCA). Application of the sPCA has two objectives: i) conducting an untargeted and therefore more objective analysis of data, where the method is less tailored to the science question as compared to more traditional regression analysis, and ii) to target a set of specific research questions (RQ):

RQ1: Is sparse principal component analysis an adequate tool to extract interaction processes inherent to a heterogeneous and short data set, which describes environmental variability?

RQ2: Is it possible to identify geographic locations ("hotspots") that are common to several interaction processes?

RQ3: Which are the key observed environmental variables that strongly contribute to several interaction processes?

Specific answers to RQ1 are given in section 3.5, with respect to model limitations and advantages, and 6.2, with respect to interaction processes. RQ2 is answered in section 6.1 and RQ3 in section 6.3. Note that we focus on the proof of concept of the sparse principal component method by basing the interpretation primarily on the known processes of the Southern Ocean climate system. New scientific insights from this novel approach are described in section 4.1.

Just as a point of clarification, the summaries at the end of each section 5.x are meant for the quick reader to grasp the essence. There are more interesting and potentially novel details in the descriptions, which can inspire researchers with an interest in the specific processes to explore those further. And of course, in a way the temporal patterns in LV7 and LV10 are trivial, but thinking this the other way around, it would not be a good sign if LV7 and LV10 did not feature, because this is an obvious performance check.

RC3.3: The paper should be re-titled to more clearly reflect the paper content. The paper focuses on all of the aspects of the ACE cruise, not just biogeochemistry and physics. I would recommend something more general like "Understanding processes observed in the southern ocean-atmosphere system using ACE observations combined with data science".

AC3.3: Thank you for the suggestion. We have retitled the paper:

Exploring the ocean and atmosphere coupled system with a data science approach applied to observations from the Antarctic Circumnavigation Expedition

We spell out ACE, because there was another cruise a couple of decades ago in the Southern Ocean called the Aerosol Characterization Experiment (ACE).

RC3.4: I recommend that the authors work on shortening the paper by moving some of the very lengthy discussion into supplementary materials or into an annex to make this paper more readable. I would like the authors to get to the point of what was learned in addition to what is already known more quickly.

AC3.4: We appreciate the reviewer's point of view and suggest the following: Section 4 has been renamed "Sparse PCA results", Section 4.1 is now "Short summary of all latent variables and new insights". and contains the text here below, which is merged from the original section 4.1 first paragraph and section 5.8 “Short summary of all latent variables”, and contains new additions to highlight the new insights. We also provide a condensed description of LV9 in a new section 4.2 to give one prominent example with new insights. The remainder of section 4 stays in the main manuscript. The manuscript then continues with the former section 6 “Discussion”. We highlight the new text in blue.

This is the new section 4.1:

“Figure 3 shows the time series of the 14 LVs, where the blue dots indicate the average of the principal components of the bootstrap runs and the shading indicates the 95% confidence interval (±2 standard deviations). The 14 LVs can be related to physical, biological and/or chemical processes, or changes in the environment that influence the variance of OVs within each LV. We name each LV according to the process or environmental condition, which they reflect (Table 2). Overall, the sPCA solution describes 55% of the variability of the 111 OVs. Here we provide a short summary for all LVs, and in section 4.2 an example description of LV9. Detailed interpretations for each LV are provided in Appendix A.

The largest signal by far originates from the large-scale horizontal temperature and pressure gradients that exist between the low and high latitudes. The effect of these gradients on physical properties of the surface ocean and its activity are mostly captured in the two climatic zone signals (LV1 and LV14). The latitudinal temperature and pressure gradients give rise to the meridional advection of cold and warm air (LV3) with implications on cyclone activity (LV13) and the freshwater cycle with the intermittent character of precipitation events (LV4).

The sPCA led to some new insights into the Southern Ocean water cycle. We were able to systematically identify the different modes of variability in the isotopic signal of marine boundary layer water vapour. δ18Ovap and δ2Hvap show significant contributions to climatological signals (LV1) and the RH environment (LV3), while d_{excvap} mainly reflects the contrasting air-sea moisture fluxes in different RH environments. While an excess of precipitation over evaporation is generally thought to cause a relatively fresh Southern Ocean surface (Dong et al., 2007; Ren et al., 2011), surprisingly, our large-scale assessment of concurrent precipitation and salinity measurements does not yield a direct response of the surface ocean salinity to precipitation events. Instead, we here show that variations in surface ocean salinity are driven by the climatological (long-term) patterns set by surface freshwater fluxes integrated over time-scales longer than synoptic events (LV1) and seasonal melting on sea ice (LV9).

We also find a latitudinal distribution of the nutrient availability and its effect on the productivity, which is highlighted in LV11, LV6 and LV8. This confirms, at the largest scale ever reported, nutrient limitation regimes for the subantarctic front, south of the polar front and associated with the island mass effect as previously reported (Pollard et al. 2002; Blain et al. 2007; Cassar et al 2007; Weber and Deutsch 2010). Moreover, the sPCA successfully decouples the high spatial and temporal variability of iron-limited (LV8) and iron-fertilized blooms (LV6) and their dependence on nutrient availability (LV11), helping to identify the macro- and micro-nutrients responsible for changes to the biogeochemistry and microbial community structure and the source of those nutrients e.g. upwelling, aeolian deposition, sea-ice melt.

The method further highlights the effects of diurnal variability of solar forcing on phytoplankton photosynthetic efficiency and trace gas oxidation (LV10) as well as that of the seasonal variation of the solar forcing on dissolved as well as atmospheric trace gas concentrations and seasonal cycle in microbial productivity (LV7). While the sPCA confirmed known seasonal trends for a number of relatively long-lived key atmospheric trace gases (methane, CO and ozone), it produced unexpected results for some of the reactive trace gases, notably isoprene (LV7). This result points towards a complex interplay between the seasonality of emissions (sources) and seasonality of oxidation pathways (sinks), which, coupled with the potential effect of transport from terrestrial sources, paint a very complex picture for atmospheric isoprene in the Southern Ocean. Further future analysis is required to better understand these processes.

The sPCA solution also clearly highlights aerosol sources (especially for INP and fluorescent aerosol) on or in the proximity of islands and continents (LV5), which was previously not as evident (Moallemi et al., 2021). We observe a clear link between wind speed and sea state and the concentration of large sea spray aerosol (LV12), tying them to the most wind-driven regions of the Southern Ocean. In contrast to that, the smaller accumulation mode particles (LV2) are ubiquitous, because of their long lifetime and various source processes contributing to their abundance.

4.2 Marginal sea ice zone and snowfall (LV9)

LV9 has a very distinct regional signal that is mostly active during Leg 2 of the cruise, with a clear peak between 27 January and 2 February 2017 when the ship was going through sea ice while approaching and leaving the Mertz region ([ref to old] Figure 12a and b), explaining about 3.4(±0.6)% of the variance of all 111 variables (Table 2). The largest contribution to this LV comes from the sea ice concentration (C_i), i.e. fraction of surface area covered by sea ice ([ref to old] Figure 12c), which was unusually low during the austral summer season 2016/2017 (Schlosser et al., 2018).

The sPCA highlights four interesting characteristics of the coupled ocean, ice, and atmosphere system in the melting sea ice region. Firstly, positive LV9 periods are associated with a low surface ocean salinity (S_{sw}) and density (sigma0,sw; ([ref to old] Figure 12c). These relatively fresh and light surface waters suggest a stable surface ocean stratification associated with recently melted sea ice, confirming previous observations (Haumann et al., 2016). While other surface freshwater fluxes such as snow and glacial melt could have been responsible for the low salinity surface ocean, the absence of a low delta18Osw in LV9 suggests no significant contribution of these fluxes. A second interesting observation is the large contribution of the wave period (Tm−1,1) to LV9 (([ref to old] Figure 12c), with a significantly longer wave period in the partially ice covered region when LV9 is positive. Therefore, the sPCA confirms that ice floes in the marginal ice zone dissipate wave energy (Squire, 2020; Ardhuin et al., 2020) with a faster rate for short-wave components of the spectrum (Meylan et al., 2018). Thirdly, net community production (NCP) and phytoplankton biomass (Chlafluo) are both positively correlated with LV9. Therefore, the sea ice melt appears to increase the water column productivity most likely through iron fertilization (Lannuzel et al., 2008, 2016), and/or enhanced water column stratification, relieving light limitation (Vernet et al., 2008; Cassar et al., 2011; Eveleth et al., 2017). A fourth aspect of LV9 is the large contribution of snowfall (SR). While a higher snowfall compared to rainfall is expected near the Antarctic coast in summer, it is unclear if there is a link between snowfall and the presence of sea ice in LV9 - an aspect that requires further investigation. However, the sPCA suggests an atmospheric boundary layer over sea ice that is dominated by Antarctic continental air masses near the surface with moist and warm advection aloft (see back trajectories in supplementary information section S4) producing snowfall at times. Antarctic air masses near the surface in LV9 are indicated by the very low abundances of heavy water molecules (delta2Hvap and delta18Ovap) in the atmospheric water vapour (w), and a low carbon monoxide (CO) concentration.

The presence of sea ice thus helps to maintain Antarctic air masses properties over the ocean by forming a barrier between the ocean and the atmosphere, limiting the influence of surface fluxes on the air mass before it reaches the open ocean (see e.g. Renfrew and Moore, 1999). Therefore, the sea ice influences the vertical atmospheric boundary layer structure, possibly favoring snowfall.

References:

Dong, S., Gille, S. T., and Sprintall, J. (2007). An Assessment of the Southern Ocean Mixed Layer Heat Budget. *Journal of Climate*, 20(17), 4425-4442. https://doi.org/10.1175/JCLI4259.1.

Moallemi, A., Landwehr, S., Robinson, C., Simó, R., Zamanillo, M., Chen, G., et al. (2021). Sources, Occurrence and Characteristics of Fluorescent Biological Aerosol Particles Measured over the Pristine Southern Ocean. *Journal of Geophysical Research: Atmospheres*, 126, e2021JD034811. https://doi.org/10.1029/2021JD034811

Ren, L., Speer, K., and Chassignet, E. P. (2011). The mixed layer salinity budget and sea ice in the Southern Ocean. *Journal of Geophysical Research*, 116, C08031. https://doi.org/10.1029/2010JC006634.

Pollard, R., Lucas, M., and Read, J.: Physical controls on biogeochemical zonation in the Southern Ocean, Deep Sea Research Part II: Topical Studies in Oceanography, 49, 3289–3305, 2002.

Blain, S., Quéguiner, B., Armand, L., Belviso, S., Bombled, B., Bopp, L., Bowie, A., Brunet, C., Brussaard, C., Carlotti, F., Christaki, U., Corbière, A., Durand, I., Ebersbach, F., Fuda, J.-L., Garcia, N., Gerringa, L., Griffiths, B., Guigue, C., Guillerm, C., Jacquet, S., Jeandel, C., Laan, P., Lefèvre, D., Lo Monaco, C., Malits, A., Mosseri, J., Obernosterer, I., Park, Y.-H., Picheral, M., Pondaven, P., Remenyi, T., Sandroni, V., Sarthou, G., Savoye, N., Scouarnec, L., Souhaut, M., Thuiller, D., Timmermans, K., Trull, T., Uitz, J., van Beek, P., Veldhuis, M., Vincent, D., Viollier, E., Vong, L., and Wagener, T.: Effect of Natural Iron Fertilization on Carbon Sequestration in the Southern Ocean, Nature, 446, 1070–1074, https://doi.org/10.1038/nature05700, 2007

Cassar, N., Bender, M. L., Barnett, B. A., Fan, S., Moxim, W. J., Levy, H., and Tilbrook, B.: The Southern Ocean biological response to aeolian iron deposition, Science, 317, 1067–1070, 2007.

Weber, T. S. and Deutsch, C.: Ocean nutrient ratios governed by plankton biogeography, Nature, 467, 550–554, 2010.

Haumann, F. A., Gruber, N., Münnich, M., Frenger, I., and Kern, S.: Sea-Ice Transport Driving Southern Ocean Salinity and Its Recent Trends, Nature, 537, 89–92, https://doi.org/10.1038/nature19101, 2016.

Squire, V. A.: Ocean Wave Interactions with Sea Ice: A Reappraisal, Annual Review of Fluid Mechanics, 52, 37–60, https://doi.org/10.1146/annurev-fluid-010719-060301, 2020.

Ardhuin, F., Otero, M., Merrifield, S., Grouazel, A., and Terrill, E.: Ice break-up controls dissipation of wind-waves across Southern Ocean sea ice, Geophysical Research Letters, 47, e2020GL087 699, https://doi.org/10.1029/2020gl087699, 2020.

Meylan, M. H., Bennetts, L. G., Mosig, J. E. M., Rogers, W. E., Doble, M. J., and Peter, M. A.: Dispersion Relations, Power Laws, and Energy Loss for Waves in the Marginal Ice Zone, Journal of Geophysical Research: Oceans, 123, 3322–3335, https://doi.org/10.1002/2018JC013776, 2018.

Lannuzel, D., Schoemann, V., De Jong, J., Chou, L., Delille, B., Becquevort, S., and Tison, J.-L.: Iron study during a time series in the western Weddell pack ice, Marine Chemistry, 108, 85–95, 2008.

Lannuzel, D., Vancoppenolle, M., van der Merwe, P., de Jong, J., Meiners, K., Grotti, M., Nishioka, J., and Schoemann, V.: Iron in Sea Ice: Review and New Insights, Elementa: Science of the Anthropocene, 4, https://doi.org/10.12952/journal.elementa.000130, 2016.

Vernet, M., Martinson, D., Iannuzzi, R., Stammerjohn, S., Kozlowski, W., Sines, K., Smith, R., and Garibotti, I.: Primary production within the sea-ice zone west of the Antarctic Peninsula: I-Sea ice, summer mixed layer, and irradiance, Deep-Sea Research Part II: Topical Studies in Oceanography, 55, 2068–2085, https://doi.org/10.1016/j.dsr2.2008.05.021, 2008.

Cassar, N., DiFiore, P. J., Barnett, B. A., Bender, M. L., Bowie, A. R., Tilbrook, B., Petrou, K., Westwood, K. J., Wright, S. W., and Lefevre, D.: The influence of iron and light on net community production in the Subantarctic and Polar Frontal Zones, Biogeosciences, 8, 227–237, https://doi.org/10.5194/bg-8-227-2011, 2011.

Eveleth, R., Cassar, N., Sherrell, R., Ducklow, H., Meredith, M., Venables, H., Lin, Y., and Li, Z.: Ice melt influence on summertime net community production along theWestern Antarctic Peninsula, Deep Sea Research Part II: Topical Studies in Oceanography, 139, 89–102, https://doi.org/10.1016/j.dsr2.2016.07.016, 2017.

Renfrew, I. A. and Moore, G. W. K.: An Extreme Cold-Air Outbreak over the Labrador Sea: Roll Vortices and Air–Sea Interaction, Monthly Weather Review, 127, 2379–2394, https://doi.org/10.1175/1520-0493(1999)127<2379:AECAOO>2.0.CO;2, 1999.

RC3.5: The authors should discuss how different timescales of processes that occur in nature that control the observed variables that were seen as a snapshot in space and time on the ship. Is it fair to group things into a data science approach variables that are observed in the atmosphere, ice, and ocean that have very different lifetimes and controlling factors that may not be co-located (i.e. relating them in the same space and time may give the wrong correlations/dependencies compared to what happens in nature)?

AC3.5: This is an important question and has been addressed in section 3.5 “Model limitations and advantages”, and section 6.2 “Atmosphere-ocean interactions”. The two main limitations we highlight are:

a) There is no underlying temporal model, meaning two observations sampled within a short period of time are more related than two observations sampled within a longer period of time - one example is the lack of observation of the relation between dissolved DMS and aerosol MSA. The sPCA does not model time, and therefore lags and nonlinear temporal effects between measurements are not taken into account. However, note that we perform an independent temporal resampling prior to sPCA (as preprocessing) in order to homogenize temporal resolution. This comes with the drawback of potentially increasing relatedness between measurements acquired within the resampling time window, but also has the benefit of increasing the temporal correlation of each OV.

b) The strict linearity means non-linear process cannot be considered. We highlight the key observations on time scale relations from section 6.2 here below. In essence, we find if processes happen on sufficiently different time scales, sPCA succeeds in not relating them, as their covariance is usually low. To prevent solutions driven by noise and spurious correlations also along different spatio-temporal scales, we introduce the use of bootstrapping, which allows us to focus only on significant relationships.

Here below follow excerpts from the manuscript that address these points:

L. 1023: *In most LVs, we find a coinciding activation of variables in the Atmospheric dynamics and thermodynamics and in the Oceanic dynamics and thermodynamics category, which are related to local coupling of wind and waves, larger-scale variations of air and water temperature, and characteristics of the ocean currents. These LVs only activate OVs from the Atmospheric dynamics and thermodynamics category, but not from the Oceanic dynamics and thermodynamics. One possible explanation for the absence of a clear influence on the ocean is that the precipitation (LV4) and the diurnal cycle (LV10) represent strong variation of atmospheric OVs on time scales of less than a day, which might be too short to trigger considerable oceanic variability of detectable strength.*

L. 1033: Links between ocean and atmosphere are visible for LVs with a strong low-frequency (> 1month) component like the climatic zones (LV1; Figure 22a), the seasonal signal (LV7; Figure 22g), and intermediate frequencies (in the order of days) such as sea ice cover (LV9; Figure 22i), and cyclone activity (LV13; Figure 22m). LVs which happen *on short time scales, for example strong precipitation related variations of LV4, trigger only a weak (w < 1) marine reaction…*

L. 1049: The above observations show that our analysis targets processes that manifest themselves in rather local correlations, such as the established link between wind speed and sea state or correlations based on smooth variations over time and space, such as the large-scale horizontal gradients in the air and sea water temperature and the hydrological cycle. To include processes occurring with a time lag or those affected by transport across larger scales, the coupling with air mass back trajectory analysis

provides a valuable extension to infer potential relations of the observed signals with up-wind conditions and air mass history, for example the advection of cold or warm air (see section 5.2.1),...

RC3.6: How do non-local processes get integrated into this approach? This is not currently clear for me.

AC3.6: This is a good point. The method succeeds by itself in including several large-scale and longer temporal features: climatic zones and large-scale horizontal gradients are represented by LV1 and LV14, large-scale weather systems feature in LV13, LV7 highlights seasonal patterns. In addition to that, we have included back trajectory analyses to understand how in situ observations carry signatures of air mass history (i.e. larger spatial and temporal extent). This was somewhat addressed in the discussion section in l. 1051. Following the reviewer's comment, we have made the formulation more explicit (l. 1051):

To better understand the ability of the sPCA to capture non-local processes occurring with a time lag or those affected by transport across larger scales, we analyse air mass back trajectories. This analysis provides a valuable extension to infer potential relations of the observed signals with up-wind conditions and air mass history. Two examples are the advection of cold or warm air (see LV3 - Meridional cold and warm air advection, Appendix A), and the removal of accumulation mode aerosols during successive precipitation events (see LV2 - Drivers of the cloud condensation nuclei population, Appendix A)."

RC3.7: The authors should expand their discussion of missing data and the influence this has on their analysis (as noted by reviewer 1).

AC3.7: To answer this question, we have to distinguish between two types of *missing data:* Data measured by sensors but filtered or dropped due to quality control or sensor failures, and data that is missing, because their temporal resolution is too low. For the former, we deal with them explicitly by using imputation strategies and temporal averaging to cope with uneven sampling and spurious missing data. The temporal interval of imputation has been selected based on the overall OV temporal resolution, and selected by comparing different strategies. We describe this in detail in Sections 3.3 and 3.4.

For the second category, i.e. data that has been sampled less frequently than our temporal resampling interval of 3 hours, we perform iterative imputation by sPCA model inversion. We employ this strategy in order to provide continuous LVs along the temporal dimension, but we cannot verify the quality of this imputation strategy or identify variations on time scales shorter than the actual sampling frequency. For this reason, OVs with very low temporal resolution have a lower number of datapoints, which affects the strength of the correlations between different OVs. In addition, such missing values generally reduce the significance of the results after bootstrapping, which tends to assign larger standard deviations and lower median weights to sparser OVs. It results that OVs with lower correlations are generally discarded by the sPCA. The lower significance and correlations, result in the tendency of assigning lower importance of sparsely measured OVs for the corresponding LVs. For example, the mixed-layer depth is only derived from the relatively sparse CTD and XBT profile locations and therefore has a much lower temporal resolution compared to the other OVs in our data set. As a consequence, it appears to be less important for air-sea exchange processes and biological production in our results as one might expect. This issue is a clear limitation of our study that is important to consider when interpreting results.

Future work can be devoted to the inclusion of OVs temporal models within a sPCA like strategy, in order to better estimate the contribution of missing data of the two types described above. Ideally, imputation will not only be made based on linear dependencies between the input OVs, but also accounting temporal co-variations, potentially providing more robust decomposition solutions with respect to gaps in measurements.

Proposed manuscript changes:

- New sentences, L266: "The data filling performed at the preprocessing step is complementary to the data imputation performed by sPCA. While the former is an independent data filling based on temporal averages, the latter can be seen as an estimation based on inverting the sPCA model on missing data, corresponding to a regression from non-missing OVs. The more correlated the OVs to the one containing a missing data point to be estimated, the better the estimation. The lower significance and correlations, result in the tendency of assigning lower importance of sparsely measured OVs for the corresponding LVs. For example, the mixed-layer depth is only derived from the relatively sparse water column profile locations and therefore has a much lower temporal resolution compared to the other OVs in our data set. As a consequence, it appears to be less important for air-sea exchange processes and biological production in our results as one might expect. This issue is a clear limitation of our study that is important to consider when interpreting results. "

Minor comments:

RC3.8: There are a few small typos as noted by reviewer 2. I suggest a careful re-reading before publication.

AC3.8: Thank you for pointing this out. We have corrected all typos we found.