Anonymous Referee #1

As a companion to the lead author's paper published (with some different co-authors) in Nature late last year (DOI:10.1038/s41598-017-15486-3) , Stranne and colleagues present what I would term a technical-demonstration paper showing how the depth of the ocean surface mixed layer may be sensed acoustically. The demonstration is set in the Arctic using data from two icebreaker cruises that sampled in both ice-covered and open-ocean regions at various vessel speeds. While the spin is largely positive, several limitations to the technique are discussed that will constrain where and when the approach will yield scientifically useful information.

Apologies for the cumbersome terminology, but I believe it is important to distinguish between the ocean surface "mixed layer" and the "mixing layer." I consider the latter to be the span that is actively being stirred vertically at the time of observation; it can be (and is frequently) thinner than the mixed layer whose base might mark the maximum depth of turbulent stirring in the past.

We agree with the reviewer on the terminology. Here we are following the definition of De Boyer, where the threshold value is chosen as to avoid the shallower mixing layer, caused by diurnal variability. This is now explicitly stated in the revised ms.

My sense is that the acoustic technique presented by Stranne and colleagues preferentially identifies the base of this deeper, possibly remnant surface mixing layer (in part due to the typically larger vertical gradient at the mixed layer base and its greater depth - at least in the data sets presented that were acquired from a large-draft vessel). Either way, I believe it is important to recognize this distinction and discuss if/how each class of "mixed" layer might be observed acoustically.

We agree with the reviewer. We have added a brief discussion on the mixed layer definition and how we seem to be imaging the "same" MLD with the acoustic method as we (and De Boyer) derive from CTD data using the threshold method.

One reason for worrying about (weak) stratification within the surface mixed layer is its possible manifestation of restratification processes including submesoscale instabilities (see Timmermans et al., 2011, doi: 10.1175/JPO-D-11-0125.1). Indeed, restratification processes are just as important as the surface stress and buoyancy forcing cited in the paper's introduction (page 1, line 41) in controlling mixed layer depth.

This is a good point. In the revised ms we mention "lateral advection" in relation to buoyancy fluxes, with reference to the Timmermans paper, as suggested by the reviewer.

The authors employ the de Boyer Montegut et al. (2004) protocol (at times modified to use a smaller temperature difference criterion) to estimate the depth of the mixed layer depth in CTD data used as ground truth for their acoustic scheme. Adoption of a technique based on temperature is a bit odd for Arctic data since at cold temperatures, density is so strongly controlled by salinity.

The density threshold approach presented in the De Boyer paper was tested with close to identical results. We opted to use and display the results from the temperature threshold method, as it is simpler plus there are more temperature data available (in e.g. WOD) than there are salinity data, thus rendering this method more useful in a general sense. Note that the same problems we had with the

temperature threshold (we had to adjust it for the central Arctic Ocean) also showed up for the density threshold. This is now stated in the manuscript.

While I doubt it would change the main conclusions of the paper, it might be worth trying the Holte and Talley (JAOTech, DOI: 10.1175/2009JTECHO543.1) algorithm, particularly for those cases where there was disagreement between the methods

As is pointed out by the reviewer, the focal point of this paper is on the fact that we can observe the MLD acoustically with a high success rate. As this is mainly a methods paper, the comparison with different protocols for deriving MLD from CTD data is, in our opinion, of secondary importance. We hope that a larger acoustic data set (or a compilation of several acoustic data sets) will be used in the future to study these differences more systematically. Currently, the number of groundtruth CTD stations are far too few for any conclusions to be drawn in this regard.

I would also quibble with the exclusion of very shallow mixed layers in the analysis, though I certainly understand the constraints deriving from vessel draft and acoustic blanking period.

We determine the presence of an MLD from CTD data by visually inspecting the profiles (as explicitly stated in the original ms). Note that the visually determined MLD can be shallower than 10 m and for all these, the acoustic method is interpreted as failing (this caused more than 50% of the failures during the SWERUS cruise). This is explained in the second comment to Table 1 (double asterisk). If we were to consider only the failures where the MLD is deeper than 10 m, our statistics would look much more convincing.

In the summer Arctic sea ice zone, the upper ocean can be stratified all the way up to the ice-ocean interface. (Drainage of melt ponds is an important summertime stratification mechanism.) The authors show one such example in figure S1 but I worry such stratification is common throughout much of the Arctic in summer, and that observations from a deep-draft vessel will give biased results. Speaking of vessel draft, there is of course the strong possibility that ships disturb the near-surface stratification, introducing yet another sourceof error.

In terms of the acoustic method, the placement of the echosounder transponder on the vessel's hull defines an absolute limit in terms of how close to the sea surface we can make observations. As discussed in the ms, the pulse length puts an additional constraint. This is inevitable.

Regarding CTD data – the method of making CTD casts from vessels have been around for many decades, and the accuracy and problems involved with such data acquisition can be found in the literature. In the high Arctic Ocean, with engines turned off, and the vessel drifting with the ice at a typical speed of less than 1 m/s, the risk of the vessel itself interfering with the shallow stratification is limited. CTD operations in high waves can be problematic due to the mixing induced by the CTD rosette (moving up and down through the water column with the waves). This is rarely a problem in the Arctic, however, and most of our CTD data seem to be reliable up to one meter or so from the surface.

My final general comment concerns the interpretation of the acoustic observations, exemplified by the sentence on page 2 line 20: "The increased SNR of wideband echosounders have made it possible to map density stratification in the ocean." The authors don't actually invert their acoustic data to estimate the ocean density profile. Rather, it is my understanding that they equate regions of enhanced acoustic backscatter with regions of enhanced vertical density gradients, the one discussed here being the mixed layer base.

This is really a question of subtle semantics. It is true that we are not mapping the actual properties of the stratification, but we are mapping density stratification in the sense that we are observing the location of more or less sharp transitions between water masses of different density, in time and space.

I continue with more specific comments/suggestions:

Page 1 line 32-34: I note that light is also a significant factor impacting phytoplankton growth, which can be impacted by MLD and residence time for phytoplankton near the air-sea interface.

Light, oxygen and nutrients place important constraints on the primary production, but these are often indirectly controlled by the MLD, as noted by the reviewer. Here we state that the MLD is one of the main factors controlling the primary production, and we hope that the interested reader will go on to read the two papers that we cite (where other aspects on primary production are discussed in detail).

Page 1 line 47: the term "temporal sampling frequency" could be confusing - I initially thought of the sampling rate of the CTD instruments, not the time between vertical profiles.

We agree with the reviewer and we have now deleted the words "temporal sampling".

Page 2 top: I note that the remote sensing observations from the GRACE satellite mission are indicative of more than surface ocean properties.

Agreed. We have changed the sentence to "..essentially restricted to near sea-surface properties".

Page 2 paragraph starting with line 3: I found it curious that this brief history doesn't begin with echosounding to determine water depth. See http://oceanexplorer.noaa.gov/history/electronic/electronic.html

Here we present a brief summary of acoustic water column mapping specifically.

Figure 1 (and others): I found the quality of the figures in this pdf to be not as crisp as I like. I'm hoping this is just a consequence of the review copy that was made available to me and that the published document will be better (i.e., quality more like the similar figures in the lead author's recent Nature paper).

Yes, this is a PDF issue. We guarantee that figure quality will be acceptable in the final version.

Page 3 line 21: The authors might wish to temper this phrase: "Together, the SWERUSC3 and AO2016 expeditions spanned the breadth and depth of the Arctic Ocean..." No observations were obtained in the Canada Basin for example.

Actually, 3 of the total 21 CTD stations from the Arctic Ocean 2016 cruise were from within the Canada Basin (see Fig 1). We do agree with the reviewer, however, and have reformulated the sentence to "..spanned much of the breadth and depth…"

Page 4 line 38: "A CTD [profile] was collected..."

Fixed

Page 5 line 35: this sentence has no real content. Much better to make a technical statement and cite a figure in support.

We are not sure what the reviewer means by this comment. What kind of technical statement are we supposed to make? The figure represents our main result and different aspects of the figure are cited further down in the text.

Figure 2 and those similar: Please give the location and date that these data were collected. In this caption and those similar, panel B should, in my opinion, say CTD profile, not profiles, or CTD-derived temperature and salinity profiles.

Fixed

Page 6 line 11: "EK80 data [are] available "

Fixed

Page 8 line 3: might be good to note the different criteria for MLD used by these previous authors.

Fixed

In summary, I believe that after revision, this work will be suitable for publication in Ocean Science.