Major concerns:

1. *Lack of evidence for sub-surface ice layers: The manuscript often refers to ice cover on melt lakes as well as sub-surface ice layers without showing evidence that such features are indeed present. In fact, most of the features that are described as "sub-surface ice layer" very well match the appearance of "second return" artefacts that result from the ATLAS sensor being saturated due to specular returns from flat surfaces, and many photons returning to the instrument during its dead-time without being detected. This issue is also briefly described in Martino et al. (2020), and further detailed in the "Specular Returns" section of the known issues document for the ATL03 product. I would highly recommend that the authors consult these two sources and decide whether they are still convinced that these second returns are signals from sub-surface ice layers rather than just artefacts in the data. If so, I would like to see convincing evidence for the claimed widespread existence of such sub-surface ice layers in a revised manuscript.*

We thank the reviewer for this insight and have examined both texts suggested. We had, earlier, discussed the specific case of Lake Ayse with Dr. Martino when this work was being developed and proceeded with the code because this did *not* seem to be a case of specular returns or instrument echo. However, the case for Lake Julian is a bit more ambiguous. The subsurface ice layer detection is intended to locate ice layers, but will definitely also capture both of phenomena described. We have added text (as follows) to make a less bold claim than in the original and also underline that the detection of ice layers is inclusive of both cases we deem to be real as well as specular returns/instrument echo. Additionally, we have included imagery (which is not covered by labels) in Fig. 10 which will hopefully make apparent why we are interpreting these physical structures the way that we are.

<u>In Section 3.1, introducing the issue and relating it to Lake Ayse:</u>
Two potential sources of ambiguity with the subsurface ice classification are: (a) the possibility for specular returns and (g) apparent multiple surface returns which resulting from instrument echo. Specular returns over flat water (implying high energy return), return a strong surface as well as multiple layers below the surface spaced according to the ATLAS deadtime (1m below the surface and a potential tertiary return below that). Echoes produced by electronic noise in the instrument, which also frequently occur very smooth water surfaces, can similarly produce a strong return at the surface with double echoes at ~2.3m and ~4.2m below the surface. (Martino et al., 2020). The categorization of subsurface ice (as in with Lake Ayşe in Fig. 3) are reliant on visual inspection. In this case, we assume subsurface ice because the layer is less than 1m from the surface and shows trailing photons towards a weakly-resolved lake bottom rather than a distinctive sharp horizontal layer with no curved bottom return. If this were a specular return, we would expect a high energy surface return to obfuscate the lake bottom entirely.

<u>In Section 5.2, discussing drainage points of Lake Julian:</u>
The *Watta* designation of "ice surface" here is likely, but not unambiguous, as this method will capture both real ice surface and false dual returns as detailed in Section 3.1. The main attributes typical of the false dual returns are a strong top surface over surface water in a flat region, followed by weaker returns at predictable intervals (~1m and ~1m for the specular return, ~2.3 and ~4.2m for the instrument echo). In this case, a specular return would be the most likely cause for a false dual-return due the spacing. However, we first note that the surface in this region is not flat and we do not see the predicted strong surface return followed by a weaker echo ("surface" and "ice" layers are of equal thickness). Additionally (a) in the case of Pt. B, the top layer contains no dual return (b) in the case of C, a distinct gap occurs in the surface. Both of these correspond to ice/water in imagery. For Pt A and Pt D, imagery suggests that these points occur at a convergence of streams. These could, however be either water or ice as no distinctive bottom return is detected. With the current available information, the "ice surface" detection

will still require manual inspection; future improvements to the code may account for the known issue with false dual returns as knowledge in this area develops.

2. *Insufficient information to make methods repeatable: One of the main objectives of this manuscript is to present the new Watta algorithm. Yet, the algorithm is not described in much detail, and the information provided is certainly not enough to replicate the results in this study. I think the easiest way to fix this would be to archive the already existing matlab code on a platform such as Zenodo (which was used by the authors to archive the lake depth data set). If the authors do not want to share their code, I think that they should provide some more detailed pseudo code (including a list of parameters and chosen values), or a more detailed text description at the very least. Furthermore, from the information provided in the manuscript, it does not seem possible for readers to check the underlying data themselves. This means that it is very hard to verify any claims made based on "visual inspection". No locations are given for the lakes under consideration, and ICESat-2 data is only referred to by its track number (not the spot, so this could refer to any of the six beams) and plotted against along-track distance with the zero point seemingly arbitrarily chosen. At the very least, I think that the authors need to provide latitude and longitude coordinates for each of the lakes considered in this study and to specify the ICESat-2 spot (GT1L, GT1R, GT2L, GT2R, GT3L, GT3R) for each lake section used with Watta. This information could be included in table S1 in the supplement. Since each of of the lakes has an associated ICESat-2 overpass, this information could be automatically extracted directly from the corresponding ATL03 or ATL06 data by, for example, using the median latitude for each of the segments shown and then querying for the corresponding longitude along the ground track. (Other information useful to readers (spot, beam type, acquisition time, etc.) could easily be printed out at the same time.)*

Accordingly, we will be releasing the current version of *Watta* matlab code to the public. Additionally, we will be adding details about (a) ICESat-2 data, specifically the spot as well as the maximum latitude and longitude of the segment within Supplemental information (b) concurrent imagery IDs from all sources. Together, these should allow users to reproduce these segments with minimal effort.

Minor concerns / suggestions:

1. *I think some similar ICESat-2 shallow bathymetry literature should be cited. There are a few articles out there with similar methodology, just applied to satellite- derived bathymetry outside the polar regions. Examples would be Albright and Craig (2020) or Thomas et al. (2020).*

These have been included in the text.

2. *It is unclear to me how the numbered lakes relate to the named lakes. Some sort of explanation for choosing to refer to some lakes by numbers and some by names should be included. A map with the locations of all lakes would be great.*

To clarify, we have added lake names and numbers in Supplementary material. A map with all locations is difficult to make meaningful as the lakes are too small for the total area of each pass. However, we are including specific latitude/longitude boundaries for the lake segments from ICESat-2 (organized by RGT, and therefore able to be mapped to the AOIs present in Figure 1) in order to enable readers to locate and reproduce lakes more easily.

3. *It is not clear to me from the text how geolocation and co-registration relate to each other here. Can you explain how matching the ATL03 point cloud with the GIMP-2 DEM to reduce square error will improve the co-registration between ICESat-2 and Landsat 8 data? To my understanding, GIMP-2 DEM elevations are mostly derived from WorldView stereo imagery, and if there is a significant difference in acquisition time between the image underlying the DEM and the Landsat 8 / ICESat-2 lake observations then surface topography could have changed significantly in the meantime due to ice flow or surface processes. With a geolocation accuracy of roughly 5 meters for both ICESat-2 and Landsat 8, I would assume that simply mapping both datasets to the same CRS would give better results than the intermediate use of DEM elevations. Admittedly, this might be me not fully understanding co-registration, yet it would be nice to be provided with some more detail/explanation, or to see evidence that this intermediate step using DEM elevations actually improves the method in a meaningful way*.

In fact, there was not really any improvement in coregistration using this process (and none that impacted the segment used for the empirical depth estimate). This step was performed nevertheless as a check as Landsat 8 and ICESat-2 geolocation was based on the GIMP-2 DEM (but this process was not transparent to us), and to monitor for anything that looked like large-scale deviation from the DEM due to changes in ice flow (although this was minimal as lakes largely conformed to bed topography. We decided to keep this step in the workflow to potentially use imagery sources in the future which were geolocated to another DEM.

Regarding temporal changes in surface topography, we note that the portion of the ICESat-2 beams that were used for coregistration were long enough to incorporate large-scale relief (which would not be affected by ice flow).

Admittedly, perfect geolocation of high-resolution imagery given ice flow is a challenge with this work, and to our knowledge, this is ongoing research for other groups.

4. The results shown on the figures could be made somewhat more accessible to readers: I would suggest to plot any ATL03 data with latitude on the horizontal axis while also including a scale bar for along-track distance. This, along with the information about which ICESat-2 track and spot is shown on which date, would already be enough for readers to figure out where to find all the underlying data. Plotting ICESat-2 ground tracks on top of imagery or image-based depth estimated wherever applicable would help readers with visual verification of some of the claims made in the text. A graticule on some maps/imagery would help as well.

To address the issue of repeatability, we are detailing the specific location of lakes (latitude/longitude boundaries of ICESat-2 segments used) in Supplemental Table 1 in addition specific concurrent imagery identifiers used for each calculation. We are avoiding including a full catalogue of imagery and overlying ICESat-2 tracks for two reasons (1) the overlying track tends to obfuscate the image substantially (2) the resulting file size (for sufficient resolution of imagery) is very large given all of the lakes in question.

The choice of axes was actually a specific request made by audience members within a previous presentation of this material. As much of the focus of this paper is related to small-scale features, we chose this axis to allow the reader to easily understand the length of lakes or ice cover.

Line-by-line comments:

Line 14: From just reading the abstract it is unclear what is meant by "corrected" depth. Line 17: Landsat 8
Altered to depth corrected for refraction

Line 18: You are stating 46 lakes, but I am counting 45 lakes in table S1, plus the five named lakes in Figure 1 for a total of 50 lakes? Line 22: please spell out CAMBOT the first time you use it (Continuous Airborne Mapping By Optical Translator)
The additional lake did not have coinciding imagery. In short: the 46th lake was used in Watta development, but only 45 had overlaying imagery with which to extract lake volumes.
We spelled out CAMBOT on its first occurance.

Line 40: You mention "both ice sheets" here, yet surface melt in Antarctica has not been discussed. While this paper only uses data over Greenland, I think it would be beneficial to briefly mention surface melt in Antarctica and how it is believed to be connected to ice shelf disintegration via hydrofracture.
A sentence in the following paragraph now discusses Antarctic ice sheets

Line 54: Since you specify for Sentinel-2, can you specify what the "higher spatial and temporal resolution" is?

Sentence has been altered from:
Commercial satellite imagery, which is poised to expand
substantially in the future, can help fill the gap in coverage of small-scale melt and melt-induced features at a higher spatial and temporal resolution, complementing estimates resolved from Sentinel-2.
To:
Commercial satellite imagery, which is poised to expand substantially in the future, can help fill the gap in coverage of small-scale melt and melt-induced features at a higher spatial (<3m) and temporal (multiple daily passes) resolution, complementing estimates resolved from Sentinel-2.

Line 56: This sounds very wordy. Could simply say "ICEsat-2 now makes it possible to replace…"
Altered accordingly

Line 60: typo (bathymmetry → bathymetry)
Corrected (here and in other locations)

Line 64: I am sceptical about the presence of ice layers under the water surface. See above in the major concerns section.
Addressed separately (see above)

Line 64: Maybe here I would specify that by "the native resolution of the ATL03 photon cloud" you mean the 0.7m laser pulse frequency in along-track distance
added

Line 65: typo (wen → when)
fixed

Line 85 / Figure 1: Can you add latitude and longitude labels, or preferably a graticule in the right panel? Please also specify in the caption that RGT = ICESat-2 "Reference Ground Track", not "repeat ground track". The ground tracks that should be repeated (in the polar regions) are the six spots GT1L,

GT1R, GT2L, GT2R, GT3L and GT3R for each numbered track. The RGT should be the point directly at the nadir, so unless ATLAS is pointing off-nadir it should fall right between GT2L and GT2R.
Corrected as requested, but with graticule added to the bottom left panel (as the addition in the main panel created too much visual noise). However, in order to better identify the locations of specific lakes, we will be adding max/min lat/lon values to Supplemental Table 1 (in addition to imagery identifiers associated with each lake). This should also address concerns about repeatability.

Line 92: This paper is largely about ICESat-2 so you might want to spell it out: "Ice, Cloud, and Land Elevation Satellite" and possibly ATLAS = Advanced Topographic Laser Altimeter System
Altered

Line 98: It is unclear to me what you mean by "using x signal photons per shot". Are you referring to the expected number of signal photons that ATLAS will detect per pulse? Are these values over land ice? Is there a citation for these values?
These are from the ICESat-2 science specs https://icesat-2.gsfc.nasa.gov/science/specs, also in Neumann et al., 2019 (cited)

Line 101: typo (MacGruder → Magruder)
Altered

Line 114: TOA has not been defined before → top of atmosphere
Altered

Line 121: It is unclear to me here what the role of the DEM is in geolocation. Line 124: "each area was approximately on average" makes no sense?
Altered (deleted "on average") and a direct reference to the Imagery Processing section is included (as this is a bit difficult to summarize)

Line 137: There are only 45 lakes in the supplemental table?
This is because only 45 lakes had both coincident imagery and ICESat-2. One lake was used for the development of Watta, but did not have coincident imagery that was usable (the lake drained too quickly afterwards). We have corrected the abstract accordingly.

Line 150: I think it might be good to point out somewhere that if the empirical estimates are "time, location and sensor specific", then your method is currently limited to producing valid depth estimates for imagery scenes that overlap with an ICESat-2 overpass over a melt lake within that scene and a three-day window. This is a limitation that the physical models don't have.
This is a fair point and has been explicitly noted around Line 77

Line 166: How are "outliers" detected?
We have added the following text to clarify: "whereby the number of standard deviations used to detect an outlier and the number of photons used to calculate a mean (window) increase with over several steps"

Line 191: What you describe here sounds exactly like artefacts in the data that come from ATLAS's dead-time when the sensor is oversaturated by a specular return. If you really believe that this is sub-surface ice in some cases, then I would need to see evidence for that to be convinced. (see major concerns section)
Addressed in the main response.

Line 201: missing full stop after "lake edges"
Fixed

Lines 205-208: It is not clear from the text how matching the ATL03 point cloud with the GIMP-2 DEM to reduce square error will improve the co-registration between ICESat-2 and Landsat 8 data. GIMP-2 DEM elevations are mostly derived from WorldView stereo imagery, and if there is a significant difference in acquisition time between the image underlying the DEM and the Landsat 8 / ICESat-2 lake observations then surface topography could have changed significantly in the meantime due to ice flow or surface processes. With a geolocation accuracy of roughly 5 meters for both ICESat-2 and Landsat 8, I would assume that simply mapping both datasets to the same CRS would give a better co-registration than the intermediate use of DEM elevations. If this is not the case, it would be nice to see some sort of proof that this intermediate step using DEM elevations actually improves coregistration.
In fact, there was not really any improvement in coregistration using this process (and none that impacted the segment used for the empirical depth estimate). This step was performed nevertheless as a check as Landsat 8 and ICESat-2 geolocation was based on the GIMP-2 DEM (but this process was not transparent to us), and to monitor for anything that looked like large-scale deviation from the DEM due to changes in ice flow (although this was minimal as lakes largely conformed to bed topography. We decided to keep this step in the workflow to potentially use imagery sources in the future which were geolocated to another DEM.

Regarding temporal changes in surface topography, we note that the portion of the ICESat-2 beams that were used for coregistration were long enough to incorporate large-scale relief (which would not be affected by ice flow).

Admittedly, perfect geolocation of high-resolution imagery given ice flow is a challenge with this work, and to our knowledge, this is ongoing research for other groups.

Line 208: Do you mean a margin of 0.2 degrees in latitude and/or longitude?
The reference was to latitude, and text has been added accordingly

Line 217: two commas after NDWI_ice, missing full stop before "Boundaries".
Altered

Line 224: typo (MacGruder → Magruder)
Altered

LIne 225-226: "a line 6m in each direction perpendicular to the ICESat-2 beam" seems like a rather confusing way to describe a circle of 6 m radius around the location of the photon.
Indeed! Altered accordingly.

Line 230 / Figure 3: Please plot the ground track of the ATL03 data shown in the top left panel on top of the depth estimates shown in the bottom left panel. Please spell out "Elevation" and "Along-track distance" in the top left panel. Also, why is along-track distance going from roughly -50 m to 800 m? I think the ICESat-2 convention is that along-track distance is measured from the last equator crossing? It would probably be more helpful for the reader if elevation was plotted against latitude, with a scale bar indicating along-track distance.
The choice of axes was actually a specific request made by audience members within a previous presentation of this material. As much of the focus of this paper is related to small-scale features, we chose this axis to allow the reader to easily understand the length of lakes or ice cover. However, with regard to the location of the lakes, we will be including the latitude/longitude extents of each lake within supplementary material, which should allow for the specific identification of features if desired.

Line 245: spell out "2" → two
Altered

Line 259-260: If performance evaluation is done by "visual inspection", it would be nice if the reader could also get to see a few examples of imagery with precise ICESat-2 ground tracks plotted on top, for their own visual inspection.
Actually, "visual inspection" here refers to Watta-calculated depths from ATL03 alone, which is largely the objective of Supplemental Figure S3. We have made this more explicit by addition additional depths.

Line 260: correlation coefficient between what? NDWI and Watta-derived depth?
Text added to make this explicit

Line 264-265: "reference ground track (RGT) 1222, Lake 3 in Fig. S4": Should be referring to Fig. S3.
Altered

Line 265-266: "the presence of subsurface ice did not always preclude the presence of a strong bottom return" → This suggests to me that it's even more likely that this "subsurface" ice layer might not exist, and that it's actually the sensor saturation and dead-time effect. (see major concerns section)
Addressed in the main response, although we note that a bottom return was present, it was just somewhat weaker.

Line 266: "(e.g. Lake 7, RGT 1169, Fig. S4)": Should be referring to Fig. S3. Also, I don't really see anything indicative of subsurface ice in Lake 7, RGT 1169, Fig. S3.
Addressed in the main response

Line 279: It sounds like you are using the R^2 for performance evaluation of the empirical model, but this would mean to evaluate the model on the data that was used to generate the model in the first place. So it should be made clear that the R^2 cannot be considered a performance metric for a model across an entire lake basin, and rather that it merely indicates how well you were able to fit the empirical model to the data along the given ICESat-2 ground track. However, the underlying model is rather simple and based on physics, so overfitting is probably not much of an issue here.
Additional text has been added to make this more explicit

Line 292: typo (there're were → there were)
Altered

Line 293-294: "future users would be able to select bands or combinations [...] that provide the greatest fidelity to ICESat-2 based observations": I know what you mean by that, but the way it is phrased it sounds like a bulletproof recipe for overfitting the data...
We have dropped the clause "that provide the greatest fidelity to ICESat-2 observations" to avoid this

Line 304: Technically GT3L describes a "spot", not a beam. Two beams (one strong, one weak) will alternate in pointing at that particular spot, switching off whenever ICESat-2 performs a yaw flip.
We have made this clearer by rephrasing as "Over this spot, covered by the 3l beam"

Line 309: typo (lake → lakes)
Fixed

Line 313 / Figure 5: "Sentinel-2 (l,m) and Planet SkySat (n,o)": should be "Sentinel-2 (k,l) and Planet SkySat (m,n)". Also, what does the red box in panel c indicate?
Altered. Added "Red box in (c)highlights region where underlying crevassing is captured

Line 328 / Figure 6: Can you show the ICESat-2 ground track on the right panels? It is very hard to see what is going on without that information. Also, it is pretty clear from context what the abbreviations Sent/LSat/SSat/PS/R/G mean here, but at least somewhere you should specify this for clarity. Also please try to stay consistent across all figures. I have noticed images with labels "Sentinel-2", "Sentinel", "Sent" and "S" across the figures in the paper, and they all refer to the same thing.
We have added an explanation of the abbreviations in the figure caption and altered the designation for "Sent" in other figures to make this a bit more consistent when possible. The ground track over Lake Ayse is shown in Figure 1, which we have made explicit in the figure caption for Figure 6.

Line 344: You want to refer to Supplemental Fig. S4 here, not S5
Altered accordingly (here and elsewhere)

Line 356 / Figure 7: Can you label lake Niels and lake Julian on the left panel?
Altered accordingly

Line 407: Ice motion should not be adjusted for in geolocation?
Because geolocation will fit to Landsat imagery (itself geolocated using the GIMP-2 DEM), we remain reliant on how well Landsat captures ice motion following from the GIMP-2 DEM upon which it's based. It remains possible that minor ice motion will not be perfectly captured (This would require feature-tracking which is outside the scope of this study).

Lines 409-419: What's shown in cyan in fig 9d does not look like ice cover to me. Also none of the satellite imagery seems to show the presence of ice cover. Can you corroborate your claims about ice layers? From looking at the figure, I would guess those are specular returns from water surfaces. (see also major concerns section)
Addressed in the Main Concerns above

Line 425: This is not the reference ground track for track 727. This must be GT1L (based on looking at the data myself), which is roughly 3.3 km offset from the RGT! The big stars used to show the locations A-D very much cover the actual features, which makes it hard to see any of the things discussed in the text. Can you plot the actual precise ground track 727 GT1L for this overpass on panel e as a (very) fine line, and indicate locations A-D with arrows pointing at the features without covering them.
What is shown is, in fact, the track for gt1l. We have altered the text to clarify this. We are reluctant to add a line to the image in panel (e) as even a fine line obfuscates much of the image. However, we have altered this figure to include two panels (on the two days surrounding the ICESat-2 pass) where the precise location of the apparent dual return is indicated with arrows (thus not obfuscating the imagery at the location).

Lines 458-463: This paragraph about Antarctica does not fit into the conclusions section. The information about ice shelf stability considerations, etc. would fit nicely into the introduction/background information about surface melt, where "both ice sheets" are already mentioned. Then, in the conclusion section you could just briefly mention that Watta could be used in Antarctica as well.
As suggested, the text applying to Antarctica has been moved into the Introduction.

Line 469-470: The goal of implementing Watta in an open-source framework is commendable and would certainly be beneficial to the scientific community. Yet, it would also be helpful to publish the already existing matlab code along with the manuscript. (also likely the easiest way to address my major concern about methods repeatability)
Addressed in Main Concerns.