Response to Anonymous Referee #1:

We thank the anonymous reviewer for their very detailed and highly constructive comments. Edits based on your input (and that of the other reviewer) have substantially improved this manuscript.

Here are our responses (in red) to your specific comments (in black):

This paper presents an evaluation of MABEL data over selected regions with the motivation of proving the performance of a single-photon counting lidar system in support of the upcoming ICESat-2 mission. In particular, the authors are investigating if the planned collection strategy for ICESat-2 will support accurate assessment of local slope in order to determine the difference between surface slope and true elevation change over the ice sheets. The study is also intended to determine the resolution of the measurements in terms of identifying small melt ponds and crevasses within the surveyed area. The paper is written well and provides a thorough analysis of the data collected by MABEL. However, the true connection of MABEL with ATLAS, the ICESat-2 onboard instrument, isn't quite clearly presented. As such, the manuscript provides knowledge about MABEL and allows for a generalized confidence in the ICESat-2 mission but certainly could expand on comparison of the two systems particularly with the radiometric differences, variations in processing schemes and length scales beyond 70 cm sampling rates. It would also be beneficial to specifically relate the MABEL statistics with the quantitative performance goals (science requirements) of the satellite with respect to ice sheet elevation change derivation. That said, this is still an important piece of work as the community looks toward another space-based laser altimetry mission and it is definitely relevant to this journal.
We received two excellent reviews of this manuscript. Both Reviewer 1 and 2 make note of the connection between MABEL and ATLAS. We have added a bit of text throughout to address this main criticism (we have captured much of this in our responses to specific comments below). Most notably, there is a large paragraph at the start of the Discussion section that addresses this in detail. In summary, given that the measured MABEL signal-photon density is generally less than that predicted for ATLAS, and given our lack of quantitative knowledge of the throughput of MABEL, it is hard for us to scale MABEL results to ATLAS. However, what we can say is that *"if the ATLAS signal-photon density and signal-to-noise ratios are within 30% of its measurement requirements (and thus mimics the MABEL performance documented in this study), ATLAS can be used to measure surface slopes over both relatively flat ice-sheet interior conditions and steeper glaciers such as the Lower Taku Glacier, and identify melt ponds."* Please see our other comments below that directly address the ATLAS/MABEL concerns.

Specific comments.

· In the abstract the purpose of MABEL is to support geophysical algorithm development, simulate key elements of the sampling strategy and assess elements of the resulting data that may vary seasonally. The last programmatic goal of MABEL seems out of place in terms of truly providing relevant information in terms of trying to separate out the errors in the MABEL measurements from seasonal variations. Later in the Introduction this goal isn't mentioned and other goals are added in; the text isn't consistent.

The reviewer makes a great point. The 'seasonal' language stems from a comparison between the timing of this AK campaign (July/August) and an earlier campaign based in Iceland (April/May). We have changed the text to separate overall MABEL goals with the goals of this particular deployment (which makes the abstract stronger). Further, we have made the language between the Abstract and the third and fourth paragraphs of the Introduction consistent. The Abstract (which summarizes the overall change) now reads:

"Given the new technology of ATLAS, an airborne instrument, the Multiple Altimeter Beam Experimental Lidar (MABEL), was developed to provide data needed for satellite-algorithm development and ICESat-2 error analysis. MABEL was deployed out of Fairbanks, Alaska in July 2014 to provide a test data set for algorithm development in summer conditions with water saturated snow and ice surfaces."

· This publication submission and the one from the author in 2014 (similar topic) cite the same reference with respect to the details of ICESat-2 but the specifications are not consistent (e.g. 10 m footprint versus 14 m footprint). Assumedly the specifications of the instrument have changed over the last 6 years since the reference was published? What else has changed? Isn't there another citation more current?

The reviewer is correct. Some aspects of ICESat-2 have changed since the Abdalati et al. (2010) reference. We have noted this in the first paragraph of the Introduction. An update (Markus et al., *submitted*) has recently been submitted to Remote Sensing of the Environment. We have added that reference. Text changes include:

"Abdalati et al. (2010) provides an early overview of the ATLAS concept and overall design. While the measurement goals of ATLAS remain as described in Abdalati et al. (2010), some of the details have evolved (Markus et al., submitted)."

· Shouldn't there be some discussion on how the MABEL data compares to ATLAS beyond just data density (e.g. photons interpreted as signal for a given length scale)? Does the density impact the performance statistics of slope determination? It seems that data gaps hinder characterization of the surface (elevation change), as would environmental impacts associated with topography and radiometry/reflectance.

Density and data gaps definitely have an impact on the characterization of surface slope; Figs. 8 and 10 indicate how clouds lead to gaps in the assessment. We have added a fair bit of text to the MABEL instrument and data descriptions that address instrumentation differences. Further, we have added a paragraph at the start of the Discussion section that addresses MABEL/ATLAS comparison issues, and thus directly addresses some of this comment. However, since MABEL measured signal-photon densities were not as high as the predicted ATLAS values, and given our lack of quantitative knowledge of the throughput of MABEL, it was hard for us to scale MABEL results to ATLAS and provide quantitative ATLAS performance statistics.

· The value of MABEL is undeniable as a test bed instrument for ATLAS but it isn't clear in this publication how it directly simulates ICESat-2 performance in a quantitative way. Many of the conclusions are vague analysis language such as " the analysis proves that ICESat-2 will be able to provide a robust assessment of across-track slope". What does robust mean in this context? Valid? Precise? Consistent performance?

This is similar to the comment above, and to the overall critique from this reviewer. In the new Discussion paragraph, we acknowledge that due to the fact that MABEL measured signal-photon densities were not as high as the predicted ATLAS values, and given our lack of quantitative knowledge of the throughput of MABEL, it was hard for us to scale MABEL results to ATLAS, and thus make direct quantitative assessments. However, we agree with the reviewer that 'robust' was not a meaningful word; we have replaced it with *'valid'* in one instance and *'non-aliased'* in another.

· Why isn't the m-atlas data used in this situation?

M-ATLAS is a dataset produced by the ICESat-2 Project Science Office. It is primarily intended for ATLAS algorithm testing. This dataset uses MABEL data to simulate expected ATLAS photon point clouds to more closely match our model predictions of ATLAS instrument performance. M-ATLAS is intended to assess photon point clouds over homogeneous surfaces, where combined beams are looking at a similar surface. Much of the analysis in this study focuses on heterogeneous surfaces, which include small-scale features that are not common from beam to beam (small melt ponds and crevasses). Thus, we did not look at the M-ATLAS data products. Since this is a very preliminary dataset, used primarily for internal purposes, and since readers will not be familiar of this dataset, we prefer to refrain from incorporating it into this manuscript.

· Will the same surface interpretation algorithm be used on ICESat-2 data as MABEL? Does the processing scheme implemented have any impact on the comparable performance? How does the surface interpretation algorithm change with changing data density (i.e. laser power degradation over time)?

The NASA GSFC surface finder is a generic surface-finding algorithm used for a wide range of surface types, and therefore a conservative estimate of surface signal (it flags more photons as signal); ATLAS algorithms will be specific to the surface type (*e.g.,* glacier, sea ice, ocean, vegetation, etc.) and more rigorous with respect to returns identified as surface signal (these will flag fewer photons as signal, relative to the GSFC algorithm). We have added language to address this. All of the algorithms have some form of along-track interpretation length scales. Thus, if data density changes, either as a result of environmental (*e.g.,* cloud cover) or instrumental (*e.g.,* degradation over time) effects, the surface finders are compromised. Since the along-track length-scales will be unique to the surface type, we prefer to leave a more formalized discussion of data density impact on algorithm surface detection for the ICESat-2 algorithm documents (in process).

· If MABEL has a horizontal error of 2 m, how does this impact the slope comparison with the GPS surveys?

The reviewer makes a great point here. We believe that the averaging that is done to determine an elevation for MABEL at given points should sufficiently rectify the geolocation concern. The MABEL elevations that defined the MABEL surface for slope determination were based on the mean of signal photons around the points of intersection of the MABEL ground track and the trends of the GPS survey lines. However, based on

this comment, we slightly modified our approach. Originally, the MABEL mean elevations that defined the MABEL surface were based on the closest 100 signal photons to the point of intersection with the GPS trend; we have changed this to be signal photons within a 5-m radius, to address the reviewers concerns and to base the analysis on the length scale of geolocation error. This led to a slight change to Figure 5, but not in the overall conclusions. We added the following text to address the reviewer's concern:
"*Elevations at those nodes were determined by taking an average of the elevations of the signal photons within a 5 m radius of those points, to take into account MABEL horizontal geolocation uncertainty."*

· What is the gridding resolution used on the MABEL data for in situ slope comparison? Does this cause any aliasing?
The MABEL grid is 1 x 1 m. This is sufficiently small for length scales appropriate to ICESat-2 slope determination (90 m). We have added this scale to the text.

· Were there any other conclusions to the goal of is there sub-surface sampling happening other than '532 nm appears to be sub-surface sampling but there is also an after pulse". It was hard to determine the results of this investigation thread.
This is a great question and still an area of active research. Geolocation limitations compromised our ability to further interrogate this topic with this dataset. We have added the following text:
"*Penetration of 532 nm wavelength light into the surface, be it a melt pond or snow, is an ongoing area of research for ICESat-2 algorithm development. MABEL geolocation uncertainty, and the fact that the 1064 and 532 nm beams do not have coincident footprints for more direct comparison (as the 1064 nm beams lead the 532 nm beams by ~60 m), compromised our ability to further interrogate this topic with this dataset, as the data could not be precisely co-registered spatially. Due to these limitations, a separate campaign with a different photon-counting laser altimeter (with both a more accurate geopositioning system and coincident 1064 and 532 nm footprints) was deployed to Thule, Greenland, in July and August 2015 (Brunt et al., 2015). Processing and analysis of that dataset are still ongoing.*
Analysis of MABEL data over small melt ponds on the Bagley Icefield in Alaska provided a preliminary assessment of how green-wavelength photon-counting systems will interact with water on an ice surface..."

· MABEL beams are said to have non-uniform transmit energy? Is this average pulse energy relative to other beams or are there spatial differences of the energy distribution (Gaussian distribution), or both? The comment "generally not the same pulse shape" is ambiguous. Does the changing or different pulse shapes have an impact on this study?
Our original statement was intended to relate the average pulse energy of the beams relative to each other. We have added this language to the text:
"Relative to one another, the MABEL beams have non-uniform average transmit energy."
But the reviewer made us also reconsider additional text in this paragraph. In reading the next couple of sentences, which contained the statement "transmit-pulse shapes are generally not the same", we realized that this was not our intent; the paragraph is really

discussing transmit energy for beams with unique optical paths. We have changed the statement to:

"…transmit-pulse energies are generally not equal."

We then removed the next sentence that continued the discussion about pulse shape. Based on previous MABEL results, pulse shapes have been fairly symmetric (from Brunt et al., 2014: "…analysis of MABEL data from the ice-sheet interior indicates that the returned pulse shape is largely symmetric."). Results here over open ocean (Figure 3) suggest the same. Changes to the pulse shape would undoubtedly have an effect on this. Broadening of the pulse would most likely lead to larger values for both accuracy and precision. Our estimates of accuracy and precision are consistent with previous results. Thus, we do not believe that MABEL pulse shapes are changing.

· Are the unique beam range biases on MABEL only due to the optical path of each beam? It seems like there are other influences on the ranges than just optical path but the text implies this is the only reason.

There are certainly environmental conditions that could contribute to an overall bias (*e.g.,* temperature of the laser). Our expectation is that environmental influences would have a uniform effect on all beams. We have added language that addresses this. However, the 'unique' nature of each beam bias stems from the unique optical path lengths (due to unique optical fiber lengths) of each beam through the instrument.

· The sentence on page 6, lines 1-2 doesn't seem to make sense or is incomplete.

Agreed. We removed this sentence and added the necessary detail to other sentences in this paragraph.

· The authors rely quite a bit on the along-track signal density. As such, it might be helpful to present a table with MABEL performance (density statistics) comparison to what is expected with ICESat-2 design cases under certain conditions (radiometric conditions, topography, weather). This could be presented as an augmented Table 1.

Great idea. We have added a row to Table 1 that provides ATLAS information, appropriate to the surface type and environmental conditions. For the open ocean, we have provided a range, as we do not know the wind state during MABEL data acquisition.

· Does the noise in the process (background and detector noise) affect the interpretation of the surface and the subsequent determination of local slope?

We do not believe so. The MABEL data used for this particular analysis were identified as surface returns by the GSFC algorithm. While that algorithm is conservative and may misidentify background as signal, we believe that our averaging technique mitigates any adverse effect on the result. The MABEL elevations that defined the MABEL surface for slope determination were based on the mean of the signal photons within a 5-m radius of those points of intersection (generally 20 – 50 signal photons, which should be sufficient to average out the small number of background outliers).

· Does this MABEL analysis provide confidence that ICESat-2 will satisfy its science requirements? There is extensive elevation bias and precision discussion for MABEL for this study area but none of the results are projected to a quantitative ICESat-2 performance. Is that projection relevant here? Will a user get similar precision from ICESat-2 measurements for a single pass over this area and see the same detail of the surface topography?

This comment addresses the key critique identified by both reviewers of this manuscript. As we stated above, we have added a paragraph at the start of the Discussion section that addresses this comment. In that paragraph, we acknowledge that due to the fact that MABEL measured signal-photon densities were not as high as the predicted ATLAS values, and given our lack of quantitative knowledge of the throughput of MABEL, it was hard for us to scale MABEL results to ATLAS, and thus make direct quantitative assessments. However, what we can quantitatively say that:

"… if the ATLAS signal-photon density and signal-to-noise ratios are within 30% of its measurement requirements (and thus mimics the MABEL performance documented in this study), ATLAS can be used to measure surface slopes over both relatively flat ice-sheet interior conditions and steeper glaciers such as the Lower Taku Glacier, and identify melt ponds."

· Page 11, line 27 seems to indicate that the 1064 nm beam would penetrate the water surface, which is not the case.

The reviewer is correct; this language was intended to compare how the different wavelengths interact with water, not to suggest that 1064 penetrates the surface. We have changed this language:

"This figure depicts how light at 532 and 1064 nm wavelengths interacts with the surface of the melt pond, and how the melt pond affects the statistics of the 532 nm return signal."

· Can you address the relationship to length scale along-track and the derived performance associated with accuracy and precision of the MABEL measurements? How do the length scales translate to ICESat-2?

For many of the performance metrics presented here, we used a 0.7 m along-track length-scale for analysis, to match that of ICESat-2. Since absolute accuracy was not required for this analysis, we did not determine an overall bias for MABEL. For precision, we used a 3-km stretch of open water to assess the individual beams. Results here were comparable to those determined over a 1.4-km departure apron, in a previous study (Brunt et al., 2014). Thus, we are confident that these values represent overall MABEL surface measurement precision. However, as stated previously, we cannot directly, or quantitatively, scale MABEL performance or precision results to ATLAS.

· How do your conclusions of MABEL performance metrics allow for accurate change detection as related to ICESat-2 expected performance?

This comment also addresses the reviewer's key critique of this manuscript. As we stated above, we have added a paragraph at the start of the Discussion section that addresses this comment. In that paragraph, we make only relative assessments of MABEL measured performance relative to ATLAS predicted performance. Ultimately, we summarize that if ATLAS performance is within 30% of its measurement requirements, ATLAS can be used to measure flat and steep surface slopes, and identify melt ponds. But we cannot make further conclusions, as these would require the instruments to have more similar radiometry.