Response to Anonymous Referee #2:

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):

Using photon-counting lidar is a new method for mapping ice sheets and glaciers. The paper presents the first results obtained by the MABEL system over complex glacier surfaces, such as heavily crevassed glaciers and lakes during melt conditions. These results are vital for assessing the performance of the ATLAS system to be flown on ICESat-2. They also highlight some of the advantages of the dense spatial sampling of photon-counting laser altimetry, for example for estimating crevasse geometry or melt pond shape and depth.

The paper could benefit from reorganizing the results and discussion sections by organizing them according to the surface feature types and measurement goals, instead of cycling through the different sites and goals several times. For example, the description of the altimetry profiles, photon histograms, and drop in the number of background photons over melt ponds are presented in three different sections.
In reading through the manuscript and incorporating the edits based on review, we have reorganized the manuscript a bit, but kept it categorized in a series of topics (photon density, bias/precision, surface characterization, and slope). While we realize some things will still be repeated in this simplified structure, we have worked to minimize that repetition. Other structural changes include: 1) we moved some MABEL beam detail out of the Intro and combined it with other beam detail in the Data section; 2) we merged the Lower Taku WorldView and GPS data sections; 3) we merged the 'Slope' results, independent of flat or steep terrain; and, most importantly, 4) we have shortened and reworked the Lower Taku Results section.

Also, some of the figures are too minuscule to appreciate the details of the surface as depicted by the photon-counting system. For example, parts of the Bagley Icefield MABEL transect and corresponding photographs (Figure 6) covering the melt ponds and crevasses could be enlarged to the surface features (photograph) and structure (photons) motor clearly. The drop in the number of background photons over the melt ponds is very evident in Fig. 6.b – a beautiful example. However, the importance of this observation is almost lost in the details.
We agree. We note that many of the figures have new numbers, based on the reorganization of the manuscript. Most of the figures were simply made larger (*e.g.,* Figs. 2, 3, 4, 7, 9, and 10). But 3 others had to be reworked such that their details became more apparent (*e.g.,* Figs. 5, 6, and 8). And we added a new figure to address one of the reviewer's detailed comments. We believe we have addressed the reviewers concerns.

Less convincing is the explanation about the melt pond depth determination. Showing the major elements of the melt pond in Figure 7 would help. Where is the elevation corresponding to the surface of the melt pond? And the bottom? Comparison of Figures 3 and 7 suggest that either the surface of the melt ponds is rough, or the maximum return is from below the surface. Can the authors distinguish between these cases? Also, I assume that the ~0-2 photons per bin between 1397.5 and 1399.8 meters is due to returns from the ponds, i.e., volume scattering. Is this correct? Does the histogram depict the bottom of the pond?

The authors agree that the analysis of the photon data over the melt ponds allows for debate. We have added elements to the figure, including surface and the after-pulse. We agree that there isn't a distinct feature on the histograms that suggests a return from the bottom of the pond; we have added this text to the manuscript. We also assume that the spread of the 532-nm histogram is associated with volume scattering throughout the pond; we have added that language as well. Some of the differences between the 2 histogram figures (formerly 3 and 7) are the result of variable length scales and, thus, very different photon counts associated with each analysis.

The connection between the results obtained by the MABEL measurements and those expected from ATLAS is not well articulated. In particular, the paper should explain better how slope accuracy and spatial scales impact the accuracy of change detection with ATLAS. Also, the errors of the MABEL measurements should be better quantified to assess its performance and the implications for ATLAS/ICESat-2.

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."* However, we cannot state further conclusions (e.g., accuracy and spatial scale comparisons), as these would require the instruments to have more similar radiometry.

For example, on page 4, line 29- page 5, line 3, the authors mention a different range bias of each of the MABEL beams. As the local slope is determined from elevations measured by different beams, this bias is expected to have an impact on the accuracy of slope determination. How stable are the range biases? How were the beams calibrated to one another? Was there any pointing bias, which would translate to additional elevation and slope errors? Page 7, lines 7-9 describe a calibration over ocean surface but does not mention how stable the offset was and if there was a pointing bias or not.

The reviewer is correct: MABEL's overall geolocation, and thus range issues, are an area of ongoing engineering discussions and limit this lidar from being utilized for a broader range of altimetry applications. However, the range biases are suitable for the overall goals of MABEL as they relate to ICESat-2 (algorithm development and error analysis).

As such, more rigorous assessments of bias and bias stability have not been part of prior MABEL deployment plans (*e.g.,* targeting many calibration targets throughout an individual flight). The previous deployments have included 'pitch-and-roll' maneuvers over open ocean (which supports a calibration that should minimize pointing biases) and flights over stable surfaces that have been surveyed (*e.g.,* Brunt et al., 2014 describes results from flying over well-surveyed airport departure aprons); language that addresses this has been added to the text:

"Prior to Level 2A data processing, MABEL ranges are corrected for these channel-specific optical path lengths using a calibration derived from data recorded during aircraft pitch and roll maneuvers performed over stretches of open ocean. We assume that this calibration mitigates the larger channel biases, including those associated with errors in pointing. However, other smaller-scale channel biases may still exist; these smaller-scale channel bias corrections were on the order of decimeters."

But for some of the slope assessments done here, we had to make assumptions on bias and bias stability. In this manuscript, we describe, and have expanded upon, our calibration technique in section 3.1, where we present the histograms over the open ocean. We choose a calibration beam and then subtract the mean differences between the other beams and the calibration beam to calibrate them to one another. For the histograms, we also detrend the data to remove long-wavelength ocean effects. We have modified the text to include these details; and we added detrending summary text:

"The detrending of each beam takes into account all of these effects; this correction ranged from 0.11 to 0.29 m over the 3000 m of along-track data used for this analysis."

For the slope assessments made on the Juneau Icefield, we make the assumption that the calibration is valid over the 75 km between the ocean site and the survey site (we have added this caveat to the language).

The explanation of DEM "migration" over the Lower Taku Glacier needs an overhaul. It is very hard to follow the details. Maybe a map with velocity vectors could help? What is the expected accuracy of the WV-2 DEM? Page 13, line 1-6: what is "the difference between the DEM and the true elevation"?. Lines 4-6: was the MABEL range bias determined from ocean measurements? How was the surface melt estimated, any reference?

We agree. We have substantially shortened and reworked the Lower Taku Results section. We believe this makes this section read much better. And the reviewer makes an excellent suggestion: we have added the velocities (and scaled vectors) to the Figure. The expected elevation accuracy of the WV DEM is on the order of meters (which is described in the section mentioned by the reviewer; Page 13, line 1-6; we have added a new reference, Shean et al., 2016, for this as well); this is effectively the difference between the *WorldView-2* DEM *elevation* and the true elevation. MABEL range bias was determined using the open ocean data in Figure 4. All of these details have been added to the text of this section. The surface melt was assessed at the GPS sites to be 2.3 m using ablation wires, by one of the co-authors; we have added this text to the manuscript.

Detailed comments:

Mention the wavelength domains for 532 nm – green, and 1064 nm – near infrared and briefly summarize the current knowledge of penetration, surface and volume scattering with references.

This is an excellent addition to the manuscript. We edited existing text and added the following:

"MABEL (discussed in detail in McGill et al., 2013) is a multibeam, photon-counting lidar, sampling at both 532 (green) and 1064 (near infrared) nm wavelengths using short (~1.5 ns) laser pulses. The dual wavelength instrument design was intended to assess green-wavelength light penetration in water or snow (McGill et al., 2013). Deems et al. (2013) provides a review of lidar use for snow studies and describes how light at 532 and 1064 nm wavelengths interacts with snow surfaces. Light penetration into a snow surface is a function of both grain size (with larger snow-grain size resulting in increased volumetric scattering, and therefore increased light penetration) and wavelength (with 532 nm light having lower absorption than 1064 nm light, which ultimately produces increased light penetration at the shorter wavelength). Deems et al. (2013) also note that light penetration into snow surfaces is extremely difficult to accurately measure."

Page 2, line 29: "ATLAS model"? Do you refer to the sensor model of ATLAS?

We were referring to the ATLAS instrument performance model. However, in light of other edits associated with the MABEL/ATLAS comparisons, this has been deleted.

Page 2-3: A figure comparing ATLAS and MABEL geometries would be helpful

Agreed. We adapted a figure from Brunt et al. (2014) to only include ICESat-2 and MABEL. Additions to the figure include across-track sampling length-scales.

Page 3, 9-12: A brief summary about the expected accuracy and potential interpretation difficulties of the photon-counting altimetry in winter and summer conditions would be useful, e.g., penetration depth of green laser beam or impact of high surface reflectance on range bias (dead time issue)

This is a great comment and makes for a great addition to the manuscript. We have added language that directly addresses this comment. The text includes language associated with NIR and green studies, which are still ongoing:

"In winter, increased albedo, reduced ice-sheet surface roughness, and reduced solar background and backscatter in the atmosphere all lead to an increased signal-to-noise ratio and an increase in photon-retrieval density (i.e., the number of, and temporal distribution of photons transmitted and recorded by the lidar). In general, with increased photon-retrieval density, we expect better surface measurement precision. In the extreme case, the photon-retrieval density may be sufficiently high that the instrument receiver does not have the time required to process the incoming photon information before receiving more. This effect is referred to as 'instrument dead time' and can produce a positive surface elevation bias. In summer, reduced albedo, increased ice-sheet surface roughness, and increased solar background leads to a decrease in photon-retrieval density and signal-to-noise ratios, compromising measurement precision. The Alaska 2014 campaign also aimed to investigate how light at 532 and 1064 nm wavelengths interacts with the surface in melting conditions, and how this may affect the statistics of the 532 nm signal photons and overall elevation accuracy."

Page 4, line 3-5: What assumptions are made to classify the photons into signal and noise/background?

While the details are discussed in Brunt et al. (2014), we agree with the reviewer that his section needed a few more details, especially with respect to the assumptions. We have added the text:

"The algorithm is based on histograms of photon arrival times in 25 m along-track segments and 10 m vertical bins and assumes a random distribution of background photons and a symmetric return pulse. Further details of this surface-finding algorithm are described in Brunt et al. (2014). The GSFC algorithm is applicable to a wide range of surface types, while most ICESat-2 standard data product algorithms are surface-type specific (e.g., glacier, sea ice, ocean, vegetation, etc.) and more rigorous with respect to returns identified as surface signal."

Page 5, line 13: what is the resolution of the camera? Number of pixels, rows/columns? Is it a B/W or color camera? How often were images taken? Was there an overlap between consecutive images at the nominal flying height?

Again, we agree with the reviewer that more detail is needed. We have added the following detail about the camera: 1) 6000x4000 pixels per image; 2) at 65,000 feet, this is approximately <3 m/pixel; 3) color images; and 4) 3-second intervals for ~30% overlap.

Page 5, line 25: Was the "standard" Landsat 8 spectral reflectance product used or did the authors derive their own reflectance? What wavelengths were used for the model? Was the panchromatic band used for Bagley Field because of the better spatial resolution?

We edited and added the following text to the Landsat discussion:

"We assessed the performance of OLI's coastal blue, blue, green, red, and panchromatic channels in retrieving supraglacial lake depths. Ultimately, the models establish a relationship between Landsat 8 top-of-atmosphere (TOA) comparing pre-drainage spectral reflectance values over the lakes with a post-drainage digital elevation model (DEM), derived from WorldView-2 imagery acquired from the Polar Geospatial Center at the University of Minnesota, using image-processing software (ERDAS). Our analysis indicated that for shallow lakes (depth < 5 m), red and panchromatic band data are most suitable for supraglacial bathymetry. Because of the relatively small size of the lakes in our study area, we chose the panchromatic channel for the better spatial resolution."

Page 8, line 17: how large was the slope caused by wind stress or dynamic ocean topography? How large is the geoid undulation?

The reviewer brings up an excellent point. Our goal was to use a stretch of open ocean to calibrate the beam elevations to one another in a relative sense. We used a 3-km stretch of data to do this. This should be significantly smaller than the length-scale of variations in dynamic topography and geoid undulation (changing ~1 to 10 m over length scales of ~100 to 1000 km). Wind stress is probably a bigger term on our 3-km length scale. While we can't separate these terms, detrending the data should account for all of these terms simultaneously. We have therefore provided the numbers associated with the detrending:

"We calibrated the beam elevations to one another to remove the unique beam elevation biases; relative bias corrections ranged from 0.03 to 0.73 m. We then detrended the surface elevations based on a linear fit to the signal photons to remove any elevation differences associated with wind stress or the relatively small effects of ocean dynamic topography and geoid undulation. The detrending of each beam takes into account all of these effects; this correction ranged from 0.11 to 0.29 m over the 3000 m of along-track data used for this analysis."

Page 8, line 27: under some operational conditions, such as??

The causes of the after-pulse are still not fully understood. We added the following language that explains this and provides more constraints on what we do know, especially with respect to the instrument/aircraft configuration of MABEL:

"The exact conditions for after-pulsing are not completely understood, but are most likely the result of temperature drifts in the fundamental laser system. These occur due to changing environmental conditions within the instrument pod in the aircraft, and/or changes in efficiency of the coolant system. The cooling system relies upon passive external fins exposed to ambient cold conditions at altitude and these conditions (temperature, airflow) change during flight. The secondary laser pulses are primarily seen in the 1064 nm returns, and are minimized when the 1064 nm source is frequency-doubled to generate 532 nm beams."

Page 9, line 1: Was the second pulse removed by visual inspection and manual editing?

We manually removed these when doing statistical analysis. We have added this to the text.

Page 10, line 20: flat, HORIZONTAL surface?

Yes, that is correct. However, based on other edits, and for continuity in this section, we have replace 'flat, HORIZONTAL' surface with *'detrended'* surface.