

A review of

**Super-high-resolution aerial imagery datasets of permafrost landscapes
in Alaska and northwestern Canada**

by Tabea Rettelbach et al.

Introduction

This paper presents airborne imagery data and associated products processed from them, acquired over a several year period in the Arctic. It is clear that a tremendous amount of work and expense went into the collection of these data and that they will be useful in a wide variety of studies. However, the paper itself falls short of the mark for ESSD's requirements and I recommend publication only after substantial revisions. That being said, I do not think it will take much work to revise the paper and my comments here are suggestions to the authors to create a paper that will cast the widest net possible to convince others to use and get the most out of their data.

In broad brush strokes what needs to be greatly improved is:

- 1) The description of the photogrammetric system
- 2) The description of the acquisition flight planning choices
- 3) The description of the data's accuracy and precision

Some other sections are perhaps over-described, but these comments are not as critical. For instance, there is a comprehensive literature review of permafrost topics which seems to have little bearing on the rest of the paper – either this should be reduced or later the paper should elaborate in more detail how this literature review affected their SPECIFIC flight planning and future science questions. For example, were there acquisitions specifically designed to look at lake drainages, ice wedge melt, beaver ponds, etc., and what questions will these data help answer? If so, which PARTICULAR flight blocks align with which topic? Similarly, there was a tremendous amount of detail on the image processing steps such as vignetting – is there a reason these steps can't be reduced to a single sentence? That is, was there something unique about this processing or will the information provided be important to someone using the data? Etc.

The term 'super-high-resolution' is used in the title and throughout the paper and this needs to be changed. What is 'super-high' to you may be coarse to someone else. Especially when it comes to modern airborne photogrammetry, there is nothing 'super-high' about 10 cm GSD. Also, the term resolution is not the best choice in most of these cases, though commonly used. A better choice is GSD, which is used elsewhere in the paper, when talking about the area covered by a single pixel and reserving 'resolution' to discuss whether the shape of an ice wedge or a tussock is resolved or not, though that's a little nitpicky (though not in the title). Similarly, I feel the rest of the title is a disservice to the data the authors are presenting – what is described in this paper are blocks of images processed into orthomosaics and DEMs, the imagery itself is just an intermediate step in this case. You want people to read the paper and use the processed data products, right? If so pick a title that will draw savvy users into doing so.

So overall I think there is too little detail where detail matters to future users of these data and too much detail on topics that won't be of much to them, and at least the parts with too little details need to be

addressed to give these data the longest legs possible. The paper also needs some reorganization, as important information on methods or results is sprinkled into somewhat random locations throughout the text.

My review focuses on these broad brush strokes and some science questions/comments I have, as I think it's premature to discuss word choices or section structure though in general the writing is clear and well written so those comments would be few in number any way. I would be happy to re-review this paper or answer any questions that I could in the meantime.

Major Revisions

1) System Description.

The paper essentially lacks a section on the photogrammetric system itself and this is unacceptable for a data paper on photogrammetric products in ESSD. While its fine to reference other papers that contain various details, the broad brush strokes **MUST** be included here if this is the first time this system has been used for this purpose, as seems to be the case. There is a brief section that describes the camera, but even the camera description is insufficient. Here is what **MUST** be addressed at **MINIMUM** in my opinion:

- Was there a GNSS system installed? If so, give some basics about it.
- No mention was made that you recognize that the GNSS antenna is not at the camera and that you dealt with the lever arms appropriately in flight direction and crabbing.
- Exactly how was the camera triggered and how was the time of photo capture recorded relative to the GNSS data stream? What is the timing accuracy?
- What is the resulting spatial accuracy of the photo centers? This **CONTROLS** the precision and accuracy of a final gridded products so must be stated or referenced.
- In what ways is this custom camera superior to a Nikon D800 or D850, which were available at the time of these acquisitions and have far superior megapixels and a huge dynamic range?
- What is the dynamic range of the MACS sensor in EV?
- What is the focal length of the lens for each sensor?
- What are the pixel dimensions of the sensor and what are the swath widths for each at a typical GSD?
- What camera parameters are fixed and which are set in the air? For example, focus, aperture, shutter speed, iso. What values were used here (in general) and how were they determined or changed in flight? What minimum in shutter speed were used in particular and how does this relate to pixel blur caused by the aircraft's motion?
- What is a Polar 5 aircraft? Later it is described as a modified DC3 (which I find really cool) but why is such a huge plane required here compared to a more maneuverable aircraft which burns less fuel and more easily makes tight turns for grids or following irregular features like rivers? Were there other sensors installed that required the room? Were any of these sensors turned on at the time of the photogrammetric acquisitions and thus have some utility to users of the photogrammetric data? Was the primary mission of these flights to do photogrammetry or something else?

2) Flight Planning.

The section labeled Survey Design does not adequately describe why flight parameter decisions were made and these are important given the unusual choices that were described. As I understand the data described here (which I have not attempted to download), the authors are only treating the data that were acquired in blocks. Yet, little information is provided about these blocks. Figure 1 sort of shows their general location but this is largely obscured by the black lines which are apparently irrelevant to this paper and by the large spatial scale. Figure 1 needs to be revised to show only the blocks (in a, b, c) with enough scale to see exactly where they are and how many lines are within each block (or annotate that). As it appears at this scale, most of these blocks are only two passes? If true, this is important to know. Further, the text (in this section and in 2.1) describes a variety of great reasons that drove flight-planning decision-making, but there is nothing I found that relates back to specific blocks presented here – the blocks should be color coded or otherwise annotated to refer to their relevance according to scientific driver and the text limited and focused to only those scientific topics actually covered by the data here, if you want to entice others to make the most use of them. And better yet, the references should relate to the blocks too – if you mapped an area specifically because some paper noted something of scientific significance occurring there, this should be made clear to the reader who may be interested in that topic or area so that they are motivated to find your data. For example, did you map any beaver dams? Or fire scars? Etc And which blocks were those? And any villages mapped as blocks need to be identified on the figures.

In terms of flight planning, no information was given on the choice of side lap. Why were these sidelaps chosen? Do you believe there was some photogrammetric advantage to using 28% rather than 60%? If this paper and the products described here were essentially opportunistic (that is, the flights were flown for other reasons than creating these data products) that's fine, but this needs to be stated clearly to make clear you are not proposing something non-standard as being superior. No information that I could find indicates how many flight lines composed each block or how accurately you believe they were flown. It's also unclear what the relevance of 'viewing angles' is, what we really need to know is how many image pairs cover each pixel. For 60% sidelap and 80% overlap, this should be 8-10. This places strong controls on precision. But we also don't know the focal length, and this controls the base-height ratio and thus also controls accuracy. When flying grids, did you attempt to maintain a constant AGL? Was this averaged? How did you determine the flying height AGL in mission planning and how did you maintain it while flying?

There seems to be a variety of information related to flight planning sprinkled throughout the remaining text – this needs to be consolidated here so that a savvy reader has all the information they need in a single spot.

3) Data quality.

In my opinion, there is simply no useful data quality information here at all and this MUST be addressed. The authors state that many of their locations were selected due to the availability of prior data at these locations, yet there are no comparisons to these data for data quality purposes. Why not? There is not even a reference I could find to prior studies of data quality using this system. Especially given their poor choice of side laps (apparently chosen for lidar purpose?), these photogrammetric DEMs need a rigorous accuracy and precision assessment for each side lap. From section 5.3 I'm surmising that their aircraft was equipped and was using lidar on every flight (???) – if this is true, they have the opportunity to compare EVERY photogrammetrically-derived DEM to their lidar and this should be done if not on all of them then a large subset capturing both flight planning differences and terrain differences. For such small areas, this should only take a day or two total, if that. I mean what's the point of writing this paper

and archiving these data if not to be used by others? And how can they be used by others for anything useful without SOME understanding of topographic accuracy and precision?

Your Figure 3 flowchart does not indicate anything about photo-center geolocation or GNSS interaction – this needs to be updated to make clear how you selected your initial positions for photo centers and in the text stated what you believe the accuracy of those positions are. The accuracy of these photo centers CONTROLS the precision of your DEMs so it needs to be clearly stated and rigorously examined. Please understand too that the Pix4D processing report gives no useful information on actual errors – it merely gives the MISFIT between the values you fed it and the values it determined in the bundle adjustment. You also must specify within Pix4D what you believe the accuracy of your photo centers is so that it won't go too crazy with adjusting them, and you should make clear in the paper (given all of the other uncertainties and problems using an opportunistic data set) what that value is given the novelty of using MACS for this purpose.

Section 5.3 indicates that there are serious data quality issues here and I do not believe they are attributable to the causes given. Horizontal accuracy should be within 1-2 pixels, perhaps 50 cm at most, if this work is done to modern standards. I was mapping thousands of square kilometers at 10 cm ten years ago at 1-2 pixel accuracy and that is what scientists expect of data acquired since then (especially in 2021), so if you are getting 2-4 m horizontal mismatches then you need to make this very clear up front and determine how typical this is of the data you are providing. Vertical accuracy stated for this single project is poor but given the lack of ground control that's fine, the data are easily shifted vertically to match the lidar and in a sense for change detection the data could have no vertical reference and still be just as useful as long as common zero-change reference points are found and detrended in the comparison. What seems completely missing and ESSENTIAL is any discussion of vertical precision – these data were nominally collected and published for the purpose of change detection and the accuracy of change detection is described ONLY by the vertical precision of the individual data sets being compared. A rigorous assessment of vertical precision is required here and is done by DEM-differencing with a reference data set and examining the standard deviation or 95% RMSE of difference. The authors mention somewhere that several blocks were acquired several times (perhaps at different AGL?) – these DEMs should be assessed for horizontal and vertical accuracy and precision too. Why would you not? If you want people to use these data in the future, you need to indicate what sorts of questions can be addressed by them! For example, can you use these data to detect permafrost thaw slumps before they occur or is it only the gross failures that can be assessed? Can you use these data to assess ice wedge melt? Etc. Provide examples of this, like in Figure 13 but for cool stuff that actually worked well to excite and motivate readers to use your data.

Here are some of my papers and blogs which give a sense of what I mean by a rigorous accuracy and precision assessment for reference, each slightly different based on prior research and current topic. I'm not saying you need to do things my way (I chose my papers for my convenience), but you do need to leave the reader with a clear sense of the scientific questions that can be assessed with your data. You'll also notice that there are overlaps between some of our data sets that can be used for your data quality comparison.

Nolan, M., Larsen, C., and Sturm, M.: Mapping snow depth from manned aircraft on landscape scales at centimeter resolution using structure-from-motion photogrammetry, *Cryosphere*, 9, 1445–1463, <https://doi.org/10.5194/tc-9-1445-2015>, 2015.

Nolan, M. and DesLauriers, K., 2016. Which are the highest peaks in the US Arctic? Fodar settles the debate. *The Cryosphere*, 10(3), pp.1245-1257.

Swanson, D. K. and Nolan, M.: Growth of Retrogressive Thaw Slumps in the Noatak Valley, Alaska, 2010–2016, Measured by Airborne Photogrammetry, *Remote Sens-basel*, 10, 983, <https://doi.org/10.3390/rs10070983>, 2018.

Gibbs, A. E., Nolan, M., Richmond, B. M., Snyder, A. G., and Erikson, L. H.: Assessing patterns of annual change to permafrost bluffs along the North Slope coast of Alaska using high-resolution imagery and elevation models, *Geomorphology*, 336, 152–164, <https://doi.org/10.1016/j.geomorph.2019.03.029>, 2019.

<https://fairbanksfodar.com/science-in-the-1002-area/>

<https://fairbanksfodar.com/fodar-makes-50-billion-measurements-of-snow-depth-in-arctic-alaska/>

<https://fairbanksfodar.com/the-first-fodar-map-of-denali-alaska/>

<https://fairbanksfodar.com/west-coast-village-data-delivered/>

In section 4, labelled as describing data and metadata structure, there is a paragraph on GNSS accuracy (?!). This is the only mention that you had on board GNSS (which should be in methods) and the accuracies given here are exceptionally crude – 2 m vertically? How can this be? Using modern PPP processing, exclusive of blunders or poor system design, you should be achieving < 10 cm positioning and more like 1-2 cm. More detail needs to be provided on this in the method and processing sections (there is no information on GNSS processing or photo center geolocation methods that I could find). Mention is made here that the photogrammetric data are going to be reprocessed once the GNSS data are reprocessed – why then are you publishing this paper and these data now? Don't you think this will simply add confusion by publishing multiple versions? GNSS processing, even when tightly coupled to IMU, takes only a few hours and it seems these are small blocks that only take a few hours each to process photogrammetric, so I think this should be done before this paper is published, along with a rigorous accuracy assessment.

Science/technical questions

- Reference is made to the total area covered by your data. How were these areas calculated? This data purports to present only the blocks that were processed photogrammetrically. I have mapped blocks that are 6000 km² and there is no way that your small blocks add up to anything near this amount as is stated in your conclusions, so I believe this is misleading and potentially disingenuous. In this paper you should limit your discussion only to the blocks you are presenting and I think that will help focus the paper in many ways overall, though a single panel figure (like 1d) is fine to set the context.
- Great words are used to describe the MACS camera, but the results don't seem that impressive to me. These words should be toned down and a discussion made comparing to modern prosumer cameras which seem far superior to me based on your results. I understand completely that this project may be stuck with the data it has and that this is perhaps an opportunistic project based on those data – that is all fine, but be clear about this. If you are not proposing that everyone should use a MACS camera, then be clear about that. Just because it was a great thing 10 years ago and now is outdated doesn't mean that's bad, just be clear and honest about it.

- Why did you use Pix4D rather than other options? It's fine that you did, but there are other (probably better) options like Metashape – why did you not use that? If Pix4D was your only options for whatever reason, that's fine – just be clear. Also be clear that everything needed for someone to reprocess the data on their own is provided, if that is indeed the case.
- The data are described as multispectral and some discussion occurs on radiometric scaling, but I didn't understand it and my gut says that it is a bit unfair to describe these data as multispectral if that word is to retain any useful meaning. I mean RGB is technically multispectral but we don't refer to it as such. I didn't understand section 3.1.2 at all so this section should be cleaned up. And without radiometric calibration on the ground or some other means, again I'm not sure you're making a good case or instilling confidence in your readers for calling the system multispectral. By 'shutter timing' did you mean 'shutter speed'? If so, why are your RGB cameras not using the same shutter speed? And how are you ensuring that they were acquired simultaneously?
- Did you really provide Pix4D with O,P,K or was it actually yaw, pitch, roll? Just double checking.
- What is the value of combining the point clouds for RGB and nIR in making gridded elevation models? Clearly they are measuring slightly different things and different contrast features – are you making an argument that this will lead to improved results? What analyses can you provide that back that up? You mention in Section 3.3 that it yielded the “best” results but give no indication of how you determined this.
- In Figures 7-9 you show data examples, but the location map seems to indicate enormous areas covered in these blocks (presumably that's what the red area is on the location map?) which is not what Figure 1 shows. Could you clarify?
- In Section 5.2 you mention cloud cover requiring 'longer sensor exposure' – do you mean shutter speed here? Is the MACS system not capable of adjusting ISO? Can you clarify this? Also can you specify what range of shutter speeds you used and the speed of the airplane and the associated percentage of pixel blur while the shutter was open?
- Here you also mention HDR techniques but I did not understand it. Can you clarify? Are you attempting to merge several photos together? That's what HDR normally means. Are you taking two photos at each intended location but with different shutter speeds? How exactly were these multiple photos used and how does this affect DEM accuracy and precision compared to using a single photo? Or did you just use a single photo (which is then not HDR)? Does this mean that you had no ability to change shutter speed in flight?
- In section 5.3 you describe acquired the TVC in race track format rather than flying adjacent flight lines in grid sequence. Having tried this myself occasionally, I can tell you that my conclusion is not that changing illumination (that is clouds or something) but rather the sun angle causes the increased errors. Even though there is not much vegetation here, the primary contrast features picked by the photogrammetric software are shadows, and over a 3 hour acquisition the shadow direction is changing 45 degrees in the Arctic. So it's always best, from what I found, to minimize the time between adjacent flight lines for this reason and only use the race track approach when logistics call for it. For example, if you are mapping a road or field site and it looks like the weather wont hold for the entire time you need, map the highest priority location in the center first so you're sure you get it then expand in a racetrack format until the weather finally calls the show. Otherwise if you start at one side of a block and fly in a normal grid sequence, you may not reach the most important area before the weather shuts you down. Same thing but worse if you spiral in on your highest priority from the outside.
- Section 5.4 on water areas does not match my experiences. The claim is made here, I think, that white caps are usable photogrammetric features. If the goal is just to get any topographic result so that an orthoimage can be made that may be true. But the photogrammetric bundle block

adjustment depends on the observed parallax in contrast features to be solely due to topography – if the contrast features are moving (like shadows, waves, cars, etc) then the topographic measurement will be thrown off. It seems that this is recognized here, but it is not clear why the topic is addressed and additional clarity would be useful if I am missing something.

- Reference is made in several places that these data will be useful as training data for machine learning use in satellite-based studies but no mention I could find was given as to how or for what scientific purposes. These comments should either be removed or described in more detail, especially in reference to specific blocks in this dataset and presumably especially those that repeated prior mapping.