Authors Response regarding: "Accuracy of snow depth estimation in mountain and prairie environments by an unmanned aerial vehicle"

By: Phillip Harder, Michael Schirmer, John Pomeroy, and Warren Helgason

The author response is comprised a summary of the major changes made in the manuscript, the point by point responses to the reviewers and lastly a version of the revised manuscript with all of the changes tracked.

SUMMARY OF MAJOR CHANGES

- Portions of the manuscript have been reduced, removed or modified to make it more concise.
- References have been updated to reflect current state of current complementary publications
- More information has been provided on the flights characteristics of the Ebee platform, camera specifications, vegetation characteristics at the alpine site and snow surface and depth measurement protocols at the various sites.
- More information on the problematic flights, how they were identified and why the problems occurred is now included.
- The extent of the snow covered areas remove from analysis with the identification of erroneous points is now included.
- The results are now more explicitly discussed with respect to platform. Specifically, we contrast our fixed wing results with multirotor results of other recent studies.
- The discussion of the use of the complementary orthomosaics to observe snow processes now has a more complete discussion. Additional figures highlight the quantification of snow covered area and visualize the problems of snowcover estimation in the presence of exposed stubble.

Reponses to the reviewers:

Author responses in red

Response to Reviewer 1

Regarding General Comments Paragraph 1:

"More information, context, and discussion regarding the UAV system would help frame the results and conclusions presented in the study. The world of UAVs and their payloads is broad and quickly expanding. Given the diversity of aircraft, cameras, and processing techniques available, the authors should refrain from representing the results from one UAV system (theirs) as indicative of UAV / SfM snow estimation techniques as a whole. Quantitative results may be particular to the UAV system of choice. More discussion of how the choice of aircraft, payload, and processing software may have influenced results is needed."

We agree that articulating our results more clearly in terms of the platform that we use (fixed wing Ebee RTK) will help us to frame our results in the context of recent work that have used multirotor platforms and differentiate more clearly the results that come from unmanned and manned platforms. The revised manuscript now reflects this context more clearly. More information is also included on the UAV platform and camera.

Regarding General Comments Paragraph 2:
"For example, the Sensefly Ebee Real Time Kinematic aircraft was shown to be sensitive to wind speeds greater than 6 ms-1. While this conclusion may be useful to future surveyors (i.e. it may not be worth their time to collect data on windy days), other platforms, such as rotary aircraft or even delta wings with more sophisticated autopilots, may be able to compensate and collect consistent data at higher wind speeds."

A minor edit has been made to the wind sensitivity value. Initially a wind speed greater than 6 $m\ s^{-1}$ was reported to lead to an increase in DSM errors. Re-examination shows that any differences in DSM error with respect to wind speed were not larger for wind speeds up to 10 $m\ s^{-1}$ and this value is now used in the paper. This value is obviously platform specific.

"Do the authors recommend future surveys use rotary platforms?"

It has been suggested that multirotor UAV's may be more stable and return better data products in windy conditions (Bühler, et al., 2016). However, there have not been any direct comparison studies that the authors are aware of that validate such assertions. A general statement regarding the use of fixed wing vs. multirotor is challenged by the broad range of UAV designs and capabilities on the market. We see that the only clear benefit of using a multirotor platform is that larger, heavier, potentially more sophisticated, sensors can be carried (which may improve DSM accuracy as our camera's exposure settings were found to generate erroneous points) and landing accuracy is higher. Disadvantages of multirotor UAVs are that flight speeds and areal coverage are more limited than for fixed wing UAVs. We now note in the manuscript that the Ebee RTK returns data at resolutions that are more than sufficient for our purposes (3cm $pixel^{-1}$), can cover much larger areas and has a higher wind resistance (>14 m/s) than many multirotors – this seems to be a clear overall advantage. Landing accuracy (+/- 5 m) was also sufficient to locate a landing location in the complex topography of the alpine site. The more important issue relative to any comparison between platform types is that all UAVs will have limited flight times and results will be compromised if conditions are windy. A direct comparison between fixed wing and multirotor platforms is necessary to determine exactly how snow depth errors of various platforms may respond to variations in wind speed and lighting conditions. Until then, based on this experience and results of other recent studies (Vander Jagt et al., 2015; Bühler et al., 2016; De Michele et al., 2016), the sufficient image quality, reasonably good high-wind stability, suitable launching and landing procedures for alpine and prairie environments that are noted in the revised manuscript, in conjunction with the clear advantages in fixed wing range, may make fixed wing platforms preferable to the multi-rotor UAVs that have been described in the snow literature to date.

"Or does the decreased flight range / endurance of rotary aircraft compared to fixed wings outweigh the increased stability?"

This is platform specific but comparing this experience and results of other recent studies (Vander Jagt et al., 2015; Bühler et al., 2016; De Michele et al., 2016) would suggest that if the reported errors are similar than the increased range/ endurance of fixed wing platforms hold an advantage. That being said one cannot say anything with certainty without a direct side-by-side comparison. The manuscript has been amended with this discussion as noted above.

"How much of an operational concern is wind sensitivity, given that snow precipitation events and wind events frequently coincide?"

The reviewer does raise the concern that snow precipitation and wind events do sometimes coincide but those events should not be of concern as any UAV should not be flying in a snow event and certainly not in a blowing snow storm because limited visible range (Pomeroy and Male, 1988) would make such operations illegal. Regulatory constraints (in Canada and other regions) restrict operations to visual line of sight, which is significantly hampered by snow in the atmosphere. Practically, airborne snow would significantly obscure surface features as seen from the UAV, reducing its ability to resolve the surface with SfM – there is no point in flying.

The most important consideration when planning to map snow depth with any UAV should be whether the anticipated signal to noise ratio will allow for direct estimates of snow depth or snow depth change. A discussion of platform type and its role in data quality that reflects these points is now in the revised manuscript.

Regarding General Comments Paragraph 3:
"A similar discussion of the camera payload would be useful to readers as well. What is the specific model of the Canon IXUS used in this study? A quick Google search yields at least a dozen different models. What are the specifications of the camera? In particular, what is the bit depth? The point about the camera automatically adjusting exposure based on center-weighted values and overexposing some scenes, causing erroneous points, is important. More discussion of this type is useful – for example, that those planning a UAV snow survey should avoid cameras with automatic light metering. Also, the authors mention their system is not equipped with a stabilizing gimbal, which clearly increased wind sensitivity and decreased vertical accuracy. A 3-axis gimbal capable of maintaining an ideal camera orientation is a common feature of many consumer or "prosumer" level UAVs. A gimbal would certainly increase the quality of the SfM inputs, and therefore perhaps the snow depth resolution. Readers interested in snow, but perhaps UAV/SfM novices, would benefit from a more detailed discussion of the camera system used in the study."

We concur that more details on the camera system would be beneficial. The camera a Canon PowerShot ELPH 110 HS, (which is the same as a Canon IXUS 125 HS) is used to capture red, green and blue band imagery and is modified to be triggered by the autopilot. Exposure settings are automatically adjusted based on a centre-weighted light metering and results may be improved in the future if one could manually adjust exposure settings (not possible with Canon ELPH). Most small fixed wing UAV's do not employ a gimbal due to the space and weight requirements for such arrangements and in the case of the Ebee RTK the camera is fixed in the UAV body. To stabilize the camera when taking photos the UAV cuts power to the motor to minimize vibrations and levels the entire UAV resulting in consistent nadir image orientation. The camera has a 16.1 Mp 1/2.3-inch CMOS sensor and stores images as JPEGs, resulting in images with 8-bit depth for the three color channels. These details are now in the revised manuscript and addressed in the discussion of errors.

Regarding General Comments Paragraph 5:
"Whether or not the erroneous points caused by overexposure are included in the authors' results is unclear upon first reading. For example, section 3.1 (256 - 257) reads "These results exclude areas affected by erroneous points, as described in section 3.3.2, which was small compared to the total snow-covered area." Which results are the authors referencing? Are the authors speaking generally about every single treatment? Or just the alpine-bare? For example, the authors should consider replacing "These results" with "The alpine-bare results" or "All results." In general, an instance of the word "this" or "these" which lacks a referent can be confusing to the reader because they are unsure as to what precisely the writer is referring. After reading section 3.1 it seems the authors did not include the erroneous points for some or all of the results - but upon referencing section 3.3.2 (322 - 324) the reader finds conflicting information: "Erroneous points could be eliminated with the removal of overexposed images. However, reducing the number of images in such a large amount caused a larger bias and gaps in the point cloud, which made this method inappropriate." Are the overexposed erroneous points included in the results or not? If the erroneous points are included, specify which results are impacted."

We appreciate the reviewer's identification of a confusing discussion on the identification and removal (or not) of erroneous points. This discussion has been simplified and limited to section 3.3.2. Some of the erroneous points encountered in early processing, only on alpine snow, coincide with snow surface measurement locations. On certain days, these errors limited the number of useful surface measurements. Incidentally, the erroneous points are located several metres above the surrounding surface, and thus are obvious and simple to exclude and so it does not make sense to include these in the error statistics.

The areas removed for each flight (as a percentage of the total snow covered area (SCA)) varied between 2% at the beginning of melt when the surface was predominantly snow-covered and 22% near the end of melt when a small number of snow patches persisted. The values of the removed SCA are now noted in the revised manuscript. The point of this discussion was to note how we approached the errors in the hope of helping others who may encounter this issue in the future.

Regarding General Comments Paragraph 6:
”Although the literature is sparse regarding SfM estimates of snow, the authors must be wary of comparing results derived from much different methods. For example, in the discussion section (286 - 292) the results are contrasted against the findings of Nolan et al. 2015, despite their methods using a manned aircraft. Similarly, Buhler et al. 2015 is a reference to a manned aircraft experiment. Given the topic sentence of this section begins “Differencing of *UAV* derived DSMs…” (emphasis mine) some readers may find the contrast of the authors’ results with that of a manned aircraft campaign misleading. Also, the 30 cm mean error reported by Nolan et al. is a geolocation error rather than a snow depth error. Snow depth errors were reported as 10 cm, and rigorously documented. Mean snow depths are not reported by Nolan et al. and thus as readers we cannot calculate or assess the SNR of his results, but it does seem like this study is suggesting a higher snow depth threshold for measurements than Nolan et al. That needs to be addressed.”

We agree it is important to differentiate that the imagery in this study was collected with a small fixed wing UAV rather than a multirotor or manned aircraft. The main difference between these studies is the collection platform, as application of SfM is fundamentally the same. Different processing software: Agisoft versus Pix4D Mapper versus Postflight Terra (and even between versions of Postflight Terra as we noticed) will give different results but the SfM principles are all the same. The differences in platform will lead to differences in the accuracy of image geotags and orientation, image resolution, bit depth and image overlaps. Regardless, very similar errors are being reported from the many recent studies applying SfM to snow despite the range of platforms and software being employed- this suggests to us that the greatest sources of uncertainty is the SfM procedure, followed by the differences in platform characteristics. The revised manuscript differentiates more clearly between the sources of uncertainty and the platforms used in the referenced studies. The 30cm mean error that we attribute to snow depth error from Nolan et al. (2015) was an error and are grateful that the reviewer brought it to our attention. It is corrected in the revised manuscript.

Regarding Technical Corrections:
Both errors were typos, we thank the reviewer for noticing them, and they are corrected in the revised manuscript.

Regarding Overall recommendation:
”I recommend the authors make revisions to the paper based on the comments above. In general, the authors need to discuss the results appropriately with respect to the referenced work and use more precise language. Also, given the limited scope of this study, readers will prefer a considerably shorter paper. Striving for concision may improve the clarity of the paper as well. The paper could easily be shortened by up about 30%.”

We thank the reviewer for these detailed comments. More precise language will be implemented in the revised manuscript and we will adjust how we reference similar studies to more appropriately reflect their results and the platforms they used in contrast to this study. Efforts were made to be more concise and the revised manuscript is shorter than before and includes recommended details on the UAV platform, camera and more explicit discussion of this work with respect to platform type.

Response to Reviewer 2:

Regarding General Comment 1:

"There appear to be a low number (or potentially a low number) of snow depth data used to evaluate depths retrieved from SfM. In some areas of the manuscript this is clear (e.g. observations range between 3 to 19 in the Alpine), but in the prairie, measurements 'between and at 34 snow stakes' is ambiguous. In addition, the reader is left unaware of the spatial coverage of these measurements (within each airborne measurement area) nor how representative they are. At the very least I would expect the n-value to be included in tables 1 and 2. Currently in the literature the amount of in-situ evaluation data for airborne SfM studies are highly variable, e.g. De Michele et al. (2015) tens of depths, Bühler et al. (2015) hundreds of depth, Nolan et al. (2015) thousands of depths. So while this comment should not be seen as an impediment to publication, where very low numbers of in-situ data exist, this needs strong justification or perhaps judicious exclusion from analyses."

We agree that the number of verification points in this analysis is quite variable. Manual snow depth observation protocols were different at the alpine and prairie sites due to the dynamics of the melt processes, and logistics. The locations of the manual snow observations were fixed throughout time at the prairie site. Each stubble treatment zone had 17 observation points identified by a physical stake for a total of 34 points at the prairie site. In contrast, the alpine site did not have a fixed snow course and snow depth measurements were limited by logistics and thus ranged between 3 and 19 sites. While the number of snow measurements is limited and variable at the alpine site, there were 100 surface measurements that were continually snow free which that had very similar errors over the course of the campaign to those of the snow surfaces. Considering the snow covered and non-snow covered surface errors together one can see that despite the limited n of error measurements specific to snow, these were not different from the large sample over bare ground. In contrast to other studies which are limited to assessing accuracy over a single or small number of flights we assessed accuracy over a large number of flights over a season. Therefore, the total number of surface observations available to assess accuracy was high. At the alpine site, absolute snow surface accuracy was assessed at 101 points and snow depth accuracy was assessed at 83 (five probe average at each point corresponds to 415 individually probed depths) points. At the prairie site, absolute snow surface and snow depth accuracy was assessed at the same 646 points. This information is now included in the tables. The locations of the points used to assess snow depth and the alpine bare surfaces are plotted in the site figure (Fig 1ab). The prairie site is very homogenous so evaluation points are quite representative of the study area. The alpine evaluation points are not as representative of the areal variation in snowpacks due to steep and inaccessible slopes but do reflect the variabilities in snow depth observed. These points are clarified in the manuscript.

Regarding General Comment 2:
"Quantification of SCA is demonstrated in Fig 8, and only briefly mentioned in section 3.4. The authors mention this is not discussed in this paper. This leads the reader to ask why not? If data are available to do this in a more thorough manner than currently presented, then this analysis would make an exceptionally valuable contribution to the literature, increase the scientific value of this paper and should definitely be included."

This is a good comment. The quantification of SCA has been added as an objective of the paper and the manuscript section on quantification of SCA has been expanded. The discussion of orthomosaic accuracy is complementary to that for the DSM so not much text is needed to include this. The additional step needed to assess SCA from orthomosaics is to implement a classification scheme and some options such as traditional supervised/unsupervised classification as well as object-oriented classification are now discussed with a clearer example. Compared to estimating snow depth from DSMs, calculating SCA from an orthomosaic is relatively simple and so is discussed concisely.

Specific Edits:

While NIR imagery was attempted, as it is not used in any of the results or discussion I suggest excluding it from this paper.

- For the sake of brevity and lack of results all references to NIR will be removed.

While written in a very readable style, the manuscript in its current form could be shortened in many areas, losing extraneous text that is not relevant to the main thrust of the argument. This will provide room for select expansion of sections in greater detail that are currently vague. Some suggestions for sections to delete or shorten considerably are: Ln 11-14; Ln 29-32; Ln 93-97; Ln 98-104; Ln 115-118; Ln 146-149; Ln 152- 155; Ln 266-269; Ln 342-345; Ln 408-412. Could much of the information in Ln 168- 181 be put in a table, making this section much more concise?

- Many of the identified sections have been edited to reduce redundancy and/or make more concise.

Ln 137: Could the size of the areas measured be explicitly mentioned?

- The prairie site was 65 hectares but the UAV consistently mapped ~100 hectares (to ensure the area of interest was captured). The alpine site was 24 hectares in size. These areas are listed in the revised manuscript.

Ln 205: Why was vegetation negligible? I'd like more information about the nature of the vegetation here to justify this claim for the creation of DSMs.

- Alpine site vegetation was sparse and where it did exist was limited to short grasses on the ridgetop (<10cm) and shrubs and coniferous trees in deep gullies on the shoulders of the ridge. To avoid potential errors in detecting change associated with vegetation obscuring the snow, springing up as snowpack ablated or growing, accuracy assessment points (the 100 points surveyed) with no vegetation (bare ground or exposed rock) were selected. Other errors, such as offsets or tilts, which are minimized through inclusion of GCPs, had a greater impact on DSM accuracy than vegetation. This is clarified in the revised manuscript.

Ln 205 – 'most of the flights' – this is vague. How many flights? Did this affect the analyses?

- Not all flights throughout the measurement campaign had concurrent snow measurements. Only 8 flights did and this is clarified in the revised manuscript

Ln 219 – (linked to previous vegetation comment) While vegetation is said to be negligible I need more convincing that grasses, particularly on 24 July at the Alpine site after 'spring up' once the snow has cleared, would not have any impact on the on the ability to pick the ground surface from photos. I expect this concern can be allayed through local knowledge, but it needs to be made explicitly and clearly here as it has been a big issue in the past at other sites.

- See answer to previous comment regarding vegetation. These grasses were very sparse.

Ln 240: Please give more details describing what 'dynamic conditions' and 'surface characteristics' are.

- Dynamic conditions reflect changes in lighting due to variability in cloud cover and wind over the course of the flight and surface characteristics reflect changes in vegetation exposure and their shadows. This is clarified in the revised manuscript.

Ln 242: Please define either here or very clearly in 3.3.1 how 'problematic flights' are defined. Currently this is, at best, vague.

- Agreed and fixed. Problematic flights were identified upon on examination of the DSMs - we could easily see that the generated surfaces clearly did not represent the snow surface (rough, with gaps in point clouds). For four of these flights this was due to high wind conditions (> 10 ms-1) and challenging light conditions that were also reflected in quite high RMSE values. One flight at the alpine site had a bias much larger than the other flights. To date we have not been able to come up with a reasonable explanation for this situation beyond the fact that it increases with the inclusion of GCPs. Diagnosis of this error is hampered by the "black box" nature of the software, we cannot examine intermediate steps to determine where the error originates. The identification of these 'problematic flights' is more rigorously defined in section 3.3.1 of the revised manuscript.

Ln 255: Give more explanation on what is meant by 'limited observations' and why this doesn't affect the detection of differences.

- That sentence was poorly constructed and did not convey what was intended. It is changed in the revised manuscript

Ln 283: No correlation is presented. Do you mean 'related'? If so please change the terminology? If not, please add the statistical correlations.

- For the sake of brevity, the brief discussion of bias correction and the associated figures is now removed.

Ln 325-340: Uncertain that this section on SGM is that useful. Proprietary software (last sentences of this paragraph) is always problematic for scientific understanding, but somewhat unavoidable for much SfM processing. Also, please explain what '2.5D' means.

- The section of SGM is very specific to the processing software that we did use and while important to replicate/understand how we dealt with the erroneous points it is now shortened to be more concise. 2.5D refers to the type of point cloud that is used in the DSM generation. 2.5D point clouds are point clouds that do not have overlapping elements. The best way of conceptualizing this is to consider the figure at the following link: https://support.pix4d.com/hc/en-us/articles/202556289-Difference-between-a-3D-and-a-2-5D-Model#gsc.tab=0. This is clarified in the revised manuscript.

Ln 376-381: I consider this just speculation. Suggest removal.
- Removed in the revised manuscript.

Ln 335: 'were' rather than 'where'.
- Corrected in the revised manuscript.

Ln 373-375: Repetitive use of 'This'. Hard to understand what 'this' is referring to. Please re-write this section with increased clarity.
- Agreed. Section is rewritten.

Ln 472: De Michele et al. 2015 is now in TC rather than TCD.
- Reference is now updated

Ln 597 & 601: Is the mean of the absolute values not the same as RMSE? If so, then stick with RMSE as terminology.

- This is the mean of the bias values from the various flights. Since bias can be negative the absolute of bias values is used to ensure that the magnitudes of the biases are preserved. This should read (is updated in revised manuscript) “mean of absolute bias values”. This is different from RMSE, which is the root of the mean squared error.

Fig 1 c) – Is this short or tall stubble – please specify.

- Tall stubble and is now specified in the caption.

Fig 5 – Opening sentence of caption - introduce ‘Alpine’ as well as the prairie sites.

- Corrected in the revised manuscript.

Fig 7 – Add ‘100’ on the y-axis of both plots.

- Corrected in the revised manuscript.

Manuscript with tracked changes

Abstract

~~The quantification of~~Quantifying the spatial distribution of snow is crucial to predict and assess ~~snow as a~~its water resource potential and understand land-atmosphere interactions~~ in cold regions. Typical~~. High-resolution remote sensing ~~approaches to quantify~~of snow depth ~~have focused on~~has been limited to terrestrial and airborne laser scanning and more recently with application of Structure from Motion (SfM) techniques to airborne (manned and unmanned) ~~photogrammetry~~imagery. In this study photography from a small unmanned aerial vehicle (UAV) was used to generate digital surface models (DSMs) and orthomosaics for snowcovers at a cultivated agricultural Canadian Prairie and a sparsely-vegetated Rocky Mountain alpine ridgetop site using ~~Structure from Motion (~~SfM~~).~~. The ~~ability~~accuracy and repeatability of this method to quantify snow depth, changes in depth and its spatial variability was assessed for different terrain types over time. Root mean square errors in snow depth estimation from ~~the~~differencing snow covered and non-snow covered DSMs were 8.8 cm for a short prairie grain stubble surface, 13.7 cm for a tall prairie grain stubble surface and 8.5 cm for an alpine mountain surface. This technique provided ~~meaningful~~useful information on maximum snow accumulation and snow-covered area depletion at all sites, while temporal changes in snow depth could also be quantified at the alpine site due to the deeper snowpack and consequent higher signal-to~~-~~-noise~~-~~ ratio. The application of SfM to UAV photographs ~~can estimate snow depth~~returns meaningful information in areas with mean snow depth > 30 cm~~ – this restricts its utility for studies~~, however the direct observation of ~~the ablation~~snow depth depletion of shallow~~, windblown~~ snowpacks with this method is not feasible. Accuracy varied with surface characteristics, sunlight and wind speed during the flight, with the most consistent performance found for wind speeds < ~~6 m s^{-1}, clear skies, high sun angles and surfaces with negligible vegetation cover. Relative to surfaces having greater contrast and more identifiable features, snow surfaces present unique challenges when applying SfM to imagery collected by a small UAV for the generation of DSMs. Regardless, the low cost, deployment mobility and the capability of repeat-on-demand flights that generate DSMs and orthomosaics of unprecedented spatial resolution provide exciting opportunities to quantify previously unobservable small-scale variability in snow depth and its dynamics~~10 m s^{-1}, clear skies, high sun angles and surfaces with negligible vegetation cover.

1. Introduction

Accumulation, redistribution, sublimation and melt of seasonal or perennial snowcovers are defining features of cold region environments. The dynamics of snow have incredibly important impacts on land-atmosphere interactions and can constitute significant proportions of the water resources necessary for socioeconomic and ecological functions (Armstrong and Brun, 2008; Gray and Male, 1981; Jones et al., 2001). Snow is generally quantified in terms of its snow water equivalent (SWE) through measurements of its depth and density. Since density varies less than depth (López-Moreno et al., 2013; Shook and Gray, 1996) much of the spatial variability of SWE can be described by the spatial variability of snow depth. Thus, the ability to measure snow depth, and its spatial distribution, is crucial to assess and predict how the snow water resource responds to meteorological variability and landscape heterogeneity. Observation and prediction of snow depth spatial distribution is even more relevant with the anticipated and observed changes occurring due to a changing climate and land use (Dumanski et al., 2015; Harder et al., 2015; Milly et al., 2008; Mote et al., 2005; Stewart et al., 2004).

The many techniques and sampling strategies employed to quantify snow depth all have strengths and limitations (Pomeroy and Gray, 1995). Traditionally, manual snow surveys have been used to quantify snow depth and density along a transect. The main benefit of manual snow surveying is that the observations are a direct measurement of the snow water equivalent; however, it requires significant labour, is a destructive sampling method and can be impractical in complex, remote or hazardous terrain (DeBeer and Pomeroy, 2009; Dingman, 2002). Many sensors exist that can measure detailed snow properties non-destructively, with a comprehensive review found in Kinar and Pomeroy (2015), but non-destructive automated sensors, such as acoustic snow depth rangers (Campbell Scientific SR50) or SWE analyzers (Campbell Scientific CS275 Snow Water Equivalent Sensor), typically only provide point scale information and may require significant additional infrastructure or maintenance to operate properly. Remote sensing of snow from satellite and aerial platforms quantify snow extent at large scales. Satellite platforms can successfully estimate snow-covered area but problems remain in quantifying snow depth, largely due to the heterogeneity of terrain complexity and vegetation cover. To date, Light Detection And Ranging (LiDAR) techniques have provided the highest resolution estimates of snow depth spatial distribution from both terrestrial (Grünewald et al., 2010) and airborne platforms (Hopkinson et al., 2012). The main limitations encountered are available areas of observation (sensor viewshed) for the terrestrial scanner and the prohibitive expense and long lead time needed for planning repeat flights for the aerial scanner (Deems et al., 2013). Typically, airborne LiDAR provides data with a ground sampling of nearly 1 m and a vertical accuracy of 15 cm (Deems and Painter, 2006; Deems et al., 2013). While detailed, this resolution still does not provide observations of the spatial variability of snow distributions that can address microscale processes such as snow-vegetation interactions or wind redistribution in areas of shallow snowcover, and the frequency of airborne LiDAR observations are typically low, except for NASA's Airborne Snow Observatory applications in California (Mattmann et al., 2014).

An early deployment of a high resolution digital camera on a remote controlled gasoline powered model helicopter in 2004 permitted unmanned digital aerial photography to support studies of shrub emergence and snowcovered area depletion in a Yukon mountain shrub tundra environment (Bewley et al., 2007). Since then, Unmanned Aerial Vehicles (UAVs) have become increasingly popular for small-scale high-resolution remote sensing applications in the earth sciences. The current state of the technology is due to advances in the capabilities and miniaturization of the hardware comprising UAV platforms (avionics/autopilots, Global-positioning systems (GPS), Inertial Momentum Units (IMUs) and cameras) and the increases in available computational power to end users for processing imagery. The conversion of raw images to orthomosaics and digital surface models takes advantage of Structure from Motion (SfM) algorithms (Westoby et al., 2012). These computationally intensive algorithms simultaneously resolve camera pose and scene geometry through automatic identification and matching of common features in multiple images. With the addition of information on the respective camera location, or if feature locations are known, then georeferenced point clouds, orthomosaics and Digital Surface Models (DSMs) can be generated (Westoby et al., 2012). Snow is a challenging surface for SfM techniques due to its relatively uniform surface and high reflectance relative to snow-free areas, which limit identifiable features (Nolan et al., 2015). The resolution of the data products produced by UAVs depends largely on flight elevation and sensor characteristics but can promise accuracies down to 2.6 cm in the horizontal and 3.1 cm in the vertical (Roze et al., 2014). ~~The vertical accuracy of the (DSM) is generally 1 - 3 times the ground sample distance (GSD) (Strecha, 2011).~~ The unprecedented spatial resolution of these products may be less important than the fact these platforms are deployable at a high, user-defined, frequency below cloud cover, which can be problematic for airborne or satellite platforms. Manned aerial platforms have the advantage of covering much larger areas (Nolan et al., 2015) with a more mature and clear regulatory framework (Marris, 2013; Rango and Laliberte, 2010) than small UAVs. However, the greater expenses associated with acquisition, maintenance, operation and training ~~of~~required for manned platforms (Marris, 2013), relative to small UAVs, are significant (Westoby et al., 2012). ~~Small UAVs overcome the limitation of terrestrial LiDAR viewshed constraints and in principle can generate DSMs equally well for complex and flat terrain.~~ Many snow scientists have expressed great enthusiasm in the opportunities UAVs present and speculate that ~~the data~~ they ~~produce~~ may drastically change the quantification of snow accumulation and ablation (Sturm, 2015).

The roots of SfM are found in stereoscopic photogrammetry, which has a long history in topographic mapping (Collier, 2002). Major advances in the 1990's in computer vision (Boufama et al., 1993; Spetsakis and Aloimonost, 1991; Szeliski and Kang, 1994) ~~building upon the development of automated feature matching algorithms (Förstner, 1986; Harris and Step, 1988) has led to the removal of certain data inputs, such as camera location, orientation or sensor characteristics, which simplifies the application of this technique.~~has automated and simplified the data requirements to go from a collection of overlapping 2D images to 3D points clouds, relative to traditional photogrammetry. Significant work by the geomorphology community has pushed the relevance, application and further development of this technique into the earth sciences (Westoby et al., 2012). Recent application of this technique to snow depth estimation has used imagery captured by manned aerial platforms (Bühler et al., 2015; Nolan et al., 2015) and increasingly with small UAVs (~~De Michele et al., 2015;~~ Vander Jagt et al., 2015; Bühler et al., 2016~~). These~~; De Michele et al., 2016). The manned aircraft examples have reported vertical accuracies ~~(root mean square errors) from the manned platforms of~~ of 10cm (Nolan et al., 2015) and 30 cm (Bühler et al., 2015) with horizontal ~~resolution between~~ resolutions of 5-20 cm (Nolan et al., 2015) and 2 m (Bühler et al., ~~2015) and from the UAV~~ 2015). Unmanned aircraft examples have shown similar accuracies and resolution with vertical errors of reported to be ~10 cm with a horizontal ~~of resolution~~resolutions between 50 cm (Vander Jagt et al., 2015) and 10 cm (Bühler et al., 2016). The accuracy of ~~assessment~~assessments of the De Michele et al. (~~2015~~2016), Vander Jagt et al. (2015), and Bühler et al. (2016) studies were limited to a small number of snow depth maps, Bühler et al. (2016) had the most with four maps, ~~and~~but more are needed to get a complete perspective on the performance of this technique and its repeatability under variable conditions.

~~The advent of UAVs and their promise to generate orthomosaics and DSMs of the earth surface at the centimeter scale at a high observational frequency is exciting. Testing of this technology applied to snow has been limited, thus a careful assessment is required of the accuracy achievable with varying weather, terrain, and vegetation, and also of its temporal repeatability.~~ The overall objective of this paper is to assess the accuracy of snow depth as estimated by imagery collected by small UAVs and processed with SfM techniques. Specifically, this paper will; 1) assess the accuracy of UAV-derived snow depths with respect to the deployment conditions and heterogeneity of the earth surface; specifically variability in terrain relief, vegetation characteristics and snow depth, and 2) identify and assess opportunities for UAV generated data to advance understanding and prediction of snowcover and snow depth dynamics.

2. Sites and Methodology

2.1 Sites

The prairie field site (Fig. 1a) is representative of agricultural regions on the cold, windswept Canadian prairies, where agriculture management practices control vegetation physical characteristics which, in turn, influence snow accumulation (Pomeroy and Gray, 1995). There is little elevation relief and the landscape is interspersed with wooded bluffs and wetlands. Snowcover is typically shallow (maximum depth < 50 cm) with development of a patchy and dynamic snow-covered area during melt. Data collection occurred at a field site near Rosthern, Saskatchewan, Canada in spring 2015 as part of a larger project studying the influence of grain stubble exposure on snowmelt processes. The 65-hectare0.65km^2 study site was divided into areas of tall stubble (35 cm) and shorter stubble (15 cm). Wheat stubble, clumped in rows ~30 cm apart, remained erect throughout the snow season, which has implications for blowing snow accumulation, melt energetics and snow cover depletion (Fig. 1c). Pomeroy et al. Snow(1993, 1998) describes the snow accumulation dynamics and snowmelt energetics inof similar environments have been described by .Pomeroy et al. (1993, 1998).

The alpine site, located in Fortress Mountain Snow Laboratory in the Canadian Rocky Mountains, is characterized by a ridge oriented in SW-NE direction (Fig. 1b, d) at an elevation of approximately 2300 m. The average slope at the alpine site is ~15 degrees with some slopes > 35 degrees. Large areas of the ridge were kept bare by wind erosion during the winter of 2014/2015 and wind redistribution caused the formation of deep snowdrifts on the leeward (SE) side of the ridge, in surface depressions and downwind of krummholz. Vegetation is limited to short grasses on the ridgetop while shrubs and coniferous trees become more prevalent on gullies on the shoulders of the ridge. Mean snow depth of the snow-covered area at the start of the observation period (May 13, 2015) was 2 m (excluding snow-free areas) with maximum depths over 5 m. The snow albedo differed between clean snow and that which had dust deposition from localized sources. The0.32 km^2 study area was divided between a North and a South area (red polygons in Fig. 1b) due to UAV battery and hence flight area limitations. Snow accumulation dynamics and snowmelt energetics in in the same environment have been described by DeBeer and Pomeroy (2010, 2009),) and MacDonald et al. (2010) and Musselman et al. (2015) and in similar environments by Egli et al. (2012), Grünewald et al. (2010), Mittaz et al. (2015) and Reba et al. (2011).describe the snow accumulation dynamics and snowmelt energetics of the area.

2.2 Methodology

2.2.1 Unmanned Aerial Vehicle - flight planning – operation - data processing

A Sensefly Ebee Real Time Kinematic (RTK) UAV (Fig. 2a) was used to collect imagery over both sites. It is marketed as a complete system, including the UAV The platform andis bundled with flight control and image processing software, to provide a complete system capable of survey grade accuracy without the use of ground control points (GCPs) (Roze et al., 2014). The Ebee RTK is a hand launched, fully autonomous, battery powered delta wing UAV with a wingspan of 96 cm and a weight of ~0.73 kg including payload. Maximum flight time is up to 45 minutes with cruising speeds between 40-90 km h^{-1}. A consumer grade camera, a Canon IXUS, captured imagery that was taggedA modified consumer grade camera, a Canon PowerShot ELPH 110 HS, is captured red, green and blue band imagery and is triggered by the autopilot. The camera is fixed in the UAV body, there is no stabilizing gimbal as often seen on multirotor UAVs, but when taking a photo the UAV cuts power to the motor to minimize vibrations and levels the entire UAV resulting in consistent nadir image orientation. The camera has a 16.1 Mp 1/2.3-inch CMOS sensor and stores images as JPEGs, resulting in images with 8-bit depth for the three color channels. Exposure settings are automatically adjusted based on a center weighted light metering. Images are geotagged with location and camera orientation information supplied by RTK corrected Global Navigation Satellite System (GNSS) positioning and IMU, respectively. A Leica GS15 base station supplied the RTK corrections to the UAV that resolve image locations to an accuracy of ± 2.5 cm. Bühler et al. (2015) found that snow depth mapping improved with the use of near-infrared (NIR) imagery as the NIR spectrum is sensitive to variations in snow grain size and water content (Dozier and Painter, 2004), which increases the contrast and complexity of the snow surface. A NIR camera, a customized Canon S110, was also flown repeatedly during this campaign

~~(three times at alpine site and 16 times at prairie site) and captured imagery in three bands; green, red and NIR (850 nm) bands.~~Ebee to resolve image locations to an accuracy of ± 2.5 cm. The Ebee was able to fly in all wind conditions attempted but image quality, location and orientation became inconsistent ~~and/or was missed~~ when wind speed at the flight altitude (as observe by an on-board pitot tube) approached ~~or exceeded~~ 14 m s^{-1}.

At the prairie site, ~~flight altitudes were ~100 m with 60% lateral and 75% longitudinal photo overlaps, which translated into mapping of up to 100 hectares per flight at a resolution of ~3 cm pixel^{-1}. Figure 2b provides a typical flight plan generated by~~ the ~~eMotion flight control software that was used on the prairie site. The~~ UAV was flown 22 times ~~during~~over the course of the melt period (6 to 30 March 2015) with three ~~more~~ flights over ~~a~~the snow free surface between 2 and 9 April 2015. A loaner Ebee, from Spatial Technologies, the Ebee distributor, performed the first 11 flights at the prairie site due to technical issues with the Ebee RTK. The geotag errors of the non-RTK loaner Ebee were ±5 m (error of GPS Standard Positioning Service) and therefore required GCPs to generate georeferenced data products. At the Alpine site, to reduce variations in the height of the UAV above the surface in complex terrain, flight plans were adjusted using a 1 m resolution DEM, derived from a LiDAR DEM. The UAV was flown 18 times over melt from 15 May to 24 June 2015 with four flights over bare ground on 24 July 2015. Table 1 summarises flight plan attributives of the respective sites. Figure 2b provides a typical flight plan generated by the eMotion flight control software for the prairie site.

~~Default settings for difficult terrain were chosen for the alpine site, these include a lateral overlap of 85% and a longitudinal overlap of 75%, with a flight altitude of 100 m. Two flights with perpendicular flight paths covered the south and north part of the alpine study area. To reduce variations in flight altitudes, flight plans were adjusted to ensure a more consistent flight altitude using a 1 m resolution DEM, derived from an available airborne LiDAR scan. The UAV was flown 18 times from 15 May to 24 June 2015 with four flights over bare ground on 24 July 2015.~~

Postflight Terra 3D 3 (version 3.4.46) ~~was used to process~~ processed the imagery to generate DSMs and orthomosaics. Though the manufacturer suggested that they are unnecessary with RTK corrected geotags (error of ±2.5 cm), all processing included GCPs ~~(locations highlighted in Fig. 1).~~. At the prairie site, 10 GCPs comprised of five tarps and five utility poles were distributed throughout the study area~~.~~ (blue points in Fig. 1a). At the alpine site, the north and south areas had five and six GCPs~~,~~ (blue points in Fig. 1b), respectively comprised of tarps (Fig. 3a) and easily identifiable rocks (Fig. 3b) spread over the study area.

Processing involved three steps. First, initial processing extracted features common to multiple images, optimized external and internal camera parameters for each image, and generated a sparse point cloud. The second step densified the point cloud and the third step generated a georeferenced orthomosaic and a DSM. Preferred processing options varied between the sites, with the semi global matching algorithm in the point densification used to minimize erroneous points ~~that were~~ encountered at the alpine site (see Sect 3.3). Generated orthomosaics and DSMs had a horizontal resolution of 3.5 cm at the prairie site and between 3.5 cm and 4.2 cm at the alpine site.

2.2.2 Ground truth and snow depth data collection

To assess the accuracy of the generated DSMs and their ability to measure snow depth, detailed observations of the land surface elevation and snow depth ~~over the course of snowcover ablation were made.~~were collected. At the prairie site a GNSS survey, utilizing a Leica GS15 as a base station and another GS15 acting as a RTK corrected rover, measured the location (x, y and z) of ~~34~~17 snow stakes on each stubble treatment to an accuracy of < ± 2.5 cm. This gives 34 observation points at the prairie site (locations identified as red dots in Fig. 1a). Over ~~the~~ melt period, the snow depth was measured with a ruler at each point (error of ± 1 cm~~) along snow surveys between and at each of~~). Adding the ~~34 snow survey stakes. Combining the~~ manually measured snow depths ~~measured by the snow surveys and their~~to the corresponding land surface elevations from the GNSS survey gives snow surface ~~elevation points that can be~~elevations at each observation point directly ~~compared~~comparable to the UAV derived DSM.

At the alpine site, 100 land surface elevations were measured at points with negligible vegetation (bare soil or rock outcrops) with a GNSS survey to determine the general quality of the DSMs. ~~Vegetation was negligible at these locations. For most of the~~For eight flights a GNSS survey was also performed on the snowcover~~.~~ (all measurement locations over the course of campaign are highlighted in Fig. 1b). To account for the substantial terrain roughness and to avoid measurement errors in deep alpine snowpacks, ~~the snowcover~~ snow surface elevation was ~~directly determined by the~~measured via GNSS survey and snow depth ~~was measured with~~estimated from the average of five snow depth measurements in a 0.4 m x 0.4 m square at ~~these locations. The average snow depth of these five values was then compared to the snow depth determined by the UAV~~that point. Time constraints and inaccessible steep snow patches limited the number of snow depth measurements to between three and ~~20 measurements per flight.~~ 19 measurements per flight. While the number of accuracy assessment points over snow is limited for each flight the cumulative number of points over the course of the campaigns used to assess accuracy over all flights is not; at the alpine site there were 101 GNSS surface measurements and 83 averaged snow depth measurements available, and at the prairie site 323 measurements on each stubble treatment.

At both the prairie and alpine site, ~~GCP location measurement employed~~ the same GNSS RTK surveying method established GCP locations. Snow surveys (maximum one per day) and DSMs (multiple per day) are only compared if from the same days.

2.2.3 Snow depth estimation

~~Snow depth was estimated by subtracting~~Subtracting a DSM ~~representing~~of a snow~~-~~ free ~~period~~surface from a DSM ~~representing a period with snowcover. This assumes that~~of a snow covered surface results estimate snow depth if snow ablation is the only ~~cause of change in the~~things changing surface elevations between the ~~dates of image capture.~~observation periods. Vegetation is limited over the areas of interest at the alpine site and any spring up of grasses or shrubs is insignificant, based upon local observations, with respect to the large snow depths observed (upto 5m). The wheat stubble at the prairie site is unaffected by snow accumulation or ablation. The snow-free DSMs corresponded to imagery collected on 2 April for the prairie site and 24 July for the ~~prairie and~~ alpine ~~sites, respectively~~site.

2.2.4 Accuracy assessment

The accuracy of the UAV-derived DSM ~~or~~and snow depth was estimated by calculating the root mean square error (~~RSME~~RMSE), mean error (bias) and standard deviation of the error (SD) with respect to the manual measurements. The ~~RSME~~RMSE quantifies the overall difference between manually measured and UAV derived values~~. Bias~~, bias quantifies the mean magnitude of the over (positive values) or under (negative values) prediction of the DSM with respect to manual measurements~~. The~~, and SD quantifies the variability of the error.

2.2.5 Signal-to-Noise Calculation

The signal-to-noise ratio (SNR) compares the level of the snow depth signal with respect to the measurement error to inform when meaningful information is available. The SNR is calculated as the mean measured snow depth value divided by the standard deviation of the error between the observed and estimated snow depths. The Rose criterion, commonly applied in image processing literature, is used to define the threshold SNR where the UAV returns meaningful snow depth information; this is further described in Rose (1973). The Rose criterion proposes a SNR ≥ 4 for the condition at which the signal is sufficiently large to avoid mistaking it for a fluctuation in noise. ~~-~~Ultimately, the acceptable signal to noise ratio depends upon the user's error tolerance (Rose, 1973).

3. Results and Discussion

3.1 Absolute surface accuracy

~~The accuracy of the DSMs is summarized in Figure 4 and Table 1 by presenting the errors for the individual flights and a summary of all the flights, respectively.~~ The accuracy of the DSMs relative to the measured surface points ~~are variable due~~varies with respect to ~~dynamic~~light conditions at time of photography and ~~the~~differences in snow surface characteristics and extent. This is seen in the RMSE for individual flights varying from 4 cm to 19 cm~~.~~ (Fig. 4). Only a few problematic flights, which will be discussed in section 3.3.1, showed larger ~~RMSE of up to 32 cm~~RMSEs, which are marked in blue in Figure 4. In general, the accuracy of the DSMs as represented by the mean RMSEs in Table ~~1~~2, were comparable between the prairie short stubble (8.1 cm), alpine-bare (8.~~1~~7 cm~~),~~) and alpine-snow (7.5 cm) sites and were greater over the prairie tall stubble (11.5 cm). Besides the five (out of 43~~ flights~~) problematic flights~~, which will be discussed in section 3.3.1~~, accuracy was relatively consistent over time at all sites. To clarify, the prairie flights simultaneously sampled the short and tall stubble areas, thus there were only three problematic flights at the prairie site in addition to the two at the alpine site (~~Figure~~Fig. 4). The larger error at the tall stubble is due to snow and vegetation surface interactions. Over the course of melt, the DSM gradually became more representative of the stubble surface rather than the snow surface~~, as the snow surface dropped below the stubble height. This highlights a problem in applying SfM to estimate snowcover, as the most prominent features, in this case exposed stubble, are preferentially weighted to represent the surface. The bias, especially for tall stubble, becomes positive resulting in over prediction of the surface, as the snow surface drops beneath the stubble height. The number of observations on alpine-snow is limited (Fig. 4) but no obvious differences were detected with respect to the alpine-bare soil (determined by 100 observations). These results exclude areas affected by erroneous points, as described in section 3.3.2, which was small compared to the total snow-covered area.~~. More points are matched on the high contrast stubble than the low contrast snow leading to the DSM being biased to reflect the stubble surface. This is apparent in the increasing tall stubble bias as the snow surface drops below the stubble height. By comparing the many alpine-bare points to the limited number of alpine-snow points (3 to 19) the relative difference in errors between the snow and non-snow surfaces was assessed. The benefit of the large amount of alpine-bare points (100) reveals the general errors, offsets and tilts in the DSM. It is concluded that the snow surface errors are not appreciably different from the non-snow surface errors.

The ~~manufacturer suggests that~~ RTK level accuracy ~~on~~of the camera geotags ~~without the use of GCPs can~~ is supposed to produce products with similar accuracy~~ to~~, without the use of GCPs, as those generated with standard GPS positioning and the use of GCPs (Roze et al., 2014).~~ This was assessed with~~ DSMs created with and without GCPs for flights where the Ebee's camera geotags had RTK-corrected positions with an accuracy of ± 2.5 cm~~. This amounted to nine~~ tested this claim. Nine flights ~~at~~from the prairie site and 22 flights ~~at~~from the alpine site met the requirements for this test. Inclusion of GCPs had little effect on the standard deviation of error with respect to surface observations, but resulted in a reduction of the mean absolute error of the bias from 27 cm to 10 cm and from 14 cm to 6 cm at the prairie and alpine site, respectively.

~~The generated NIR DSMs had rough surfaces, large biases and gaps due to SfM not being able to resolve the surface features. Despite possible advantages over visible imagery due to greater snow contrast, it was not possible to generate reliable results using the images from this customized Canon S110 NIR camera.~~

3.2 Snow depth accuracy

The snow depth errors were similar to that of the surface errors with the alpine and short stubble sites having very similar errors, with mean RMSEs of 8.5 cm and 8.8 cm, but much larger errors over tall stubble, with mean RMSE of 13.7 cm (Fig. 5 and Table ~~2~~3). Snow depth errors were larger than the surface errors as the errors from the snow-free and snow-covered DSMs are additive in the DSM differencing. The usability of snow depth determined from DSM differencing requires comparison of signal-to-noise. Signal-to-noise, SNR in Fig. 5, clearly demonstrates that the deep alpine snowpacks have a large signal relative to noise and provide very useable information on snow depth both at maximum accumulation and during most of the snowmelt period (SNR >7). In contrast, the shallow snowpack at the prairie site, despite a similar absolute error to the alpine site, demonstrates decreased ability to retrieve meaningful snow depth information over the course of snowmelt; the signal became smaller than the noise. Applying the Rose criterion of a SNR ~4, it is apparent that only the first flight at the short stubble and the first two flights at the tall stubble provided useful information on the snow depth signal.

~~The error of the estimated snow depth~~This is ~~correlated to the bias;~~relevant when applying this ~~is most apparent at the prairie site where the estimated, shallow, snow depth varies~~technique to other areas with ~~the bias. With bias correction, the mean snow depth, as demonstrated in Fig. 6, shows a relatively coherent time evolution for a shallow snow cover.~~

~~Differencing of UAV derived DSMs provides meaningful but limited information about snow depth. Reliable information is limited to the peak accumulation period at the prairie site, which is typical of~~ shallow, wind redistributed seasonal snowcovers such as those that cover prairie, steppe and tundra in North and South America, Europe and Asia. This is in contrast to other studies which ~~suggest~~do not limit where this technique can be ~~universally adopted for snow depth mapping despite reporting a RMSE of up to 30 cm~~reasonably applied (Bühler et al., ~~2015~~2016; Nolan et al., ~~2015). Errors of such a magnitude are inappropriate for estimating the depth of shallow snowcovers.~~2015).

3.3 Challenges

3.3.1 UAV Deployment Challenges

An attractive attribute of UAVs, relative to manned aerial or satellite platforms, is that they allow "on-demand" responsive data collection. While deployable under most conditions encountered, the ~~significant~~ variability in the DSM RMSEs is likely due to the environmental factors at time of flight including wind conditions, sun angle, flight duration, cloud cover and cloud cover variability. In high wind conditions (>14 m s^{-1}) the UAV struggled to maintain its preprogrammed flight path~~.~~ as it is blown off course when cutting power to take photos. This resulted in missed photos and inconsistent density in the generated point clouds. ~~This UAV does not employ~~Without a ~~gimbal to stabilize~~gimballed camera ~~orientation and thus~~ windy conditions also resulted in blurry images ~~from the unstable platform~~ that deviate from the ideal vertical orientation. The flights for the DSMs with the greatest RMSEs had the highest wind speeds as measured by the UAV. Four of the five problematic flights were due to high winds (>10 m s^{-1}) and were identified by relatively low-density point clouds with significant gaps which rendered DSMs that did not reflect the snow surface characterises.

As the system relies on a single camera traversing the areas of interest, anything that may cause a change in the reflectance properties of the surface will complicate post-processing and influence the overall accuracy. Consistent lightning is important with a preference for clear, high sun conditions to minimize ~~shadow dynamics.~~changes in shadows. Diffuse lighting during cloudy conditions ~~resulted~~results in little contrast over the snow surface and large gaps in the point cloud over snow, especially when the snow cover was homogeneous. Three flights under these conditions could not be used and were not included in the previously shown statistics. Clear conditions and patchy snowcover led to large numbers of overexposed pixels (see Sect 3.3.2). Low sun angles should be avoided as orthomosaics from these times are difficult to classify with respect to the large and dynamic surface shadows present and the relatively limited reflectance range.

It is suggested that multirotor UAV's may be more stable and return better data products in windy conditions (Bühler, et al., 2016). There have not been any direct comparison studies that the authors are aware of that validate such assertions. A general statement regarding the use of fixed wing vs. multirotor is also impossible with the broad spectrum of UAVs and their respective capabilities on the market. The only clear benefit of using a multirotor platform is that larger, potentially more sophisticated, sensors can be carried and landing accuracy is higher. That being said the Ebee RTK returns data at resolutions that are more than sufficient for our purposes (3cm pixel^{-1}), can cover much larger areas and has a higher wind resistance (>14 m/s) than many multirotors. Landing accuracy (+/- 5 m) was also sufficient to locate a landing location in the complex topography of the alpine site. The more important issue relative to any comparison between platform types is that all UAVs will have limited flight times and results are compromised if conditions are windy and light is inconsistent. Until a direct platform comparison study is conducted this experience, and results of other recent studies (Vander Jagt et al., 2015; Bühler et al., 2016; De Michele et al., 2016), suggests that fixed wing platforms, relative to multi-rotor platforms, have similar accuracy and deployment constraints but a clear range advantage.

3.3.2 Challenges applying Structure from Motion over snow

Erroneous points over snow were generated ~~by~~in post-processing with the default settings at the alpine ~~sites~~site. These points were up to several metres above the actual snow surface and were mainly located at the edge of snow patches, but also on irregular and steep snow surfaces in the middle of a snow patch. The worst cases occurred during clear sunny days over south-facing snow patches, ~~where the whole snow patch was~~which were interspersed with these erroneous points. These points are related to the overexposure of snow pixels in the ~~raw~~ images~~,~~ which ~~typically occurred during direct sunlight over a small snow-covered area. A typical image with overexposed snow pixels~~ had bare ground in the centre and small snow patches on the edges. ~~The Canon IXUS camera~~ This is a consequence of the automatically ~~adjusts~~adjusted exposure based on centre-weighted light metering ~~and is not adjustable. Erroneous~~of the Canon ELPH camera. It is recommended that erroneous points could be ~~eliminated~~minimized with the removal of overexposed images~~. However, reducing the number of images in such a large amount caused a larger~~; however this increased the bias and led to gaps in the point cloud, which made this ~~method~~approach inappropriate.

The semi-global matching (SGM) option with optimization for 2.5D point clouds (point clouds with no over lapping points) proved to be the best parameter setting within the post-processing software Postflight Terra 3D. Semi-global matching was employed to improve results on projects with low or uniform texture images, while the optimization for 2.5D removes points from the densified point cloud (SenseFly, 2015).

The SGM option removed most of the erroneous points with best results if processing was limited to individual flights. Including images from additional ~~perpendicular~~ flights ~~or merging subareas with overlapping images~~ resulted in a rougher surface with more erroneous points. This ~~is likely due to~~may be caused by changes in the surface lighting conditions between flights, which challenges SfM. ~~However, there was no additional bias introduced by the use of~~Bias did not change when using SGM ~~and~~though some linear artefacts were visible when compared to default settings. These linear artefacts caused the ~~standard deviation of the error~~sd to increase from 1 cm to 3 cm on bare ground. Areas with remaining erroneous points ~~where~~were identified and excluded from the presented analysis. ~~The ability to reduce these erroneous points~~Table 3 summaries the extent of the areas removed with ~~SGM depended on the version~~respect to the snow covered area at the alpine site. The fifth problematic flight identified (1 June flight over north area of ~~Postflight Terra 3D. Results achieved~~alpine site) had a much larger bias with ~~version 3.4.46 were much better than results from the later version 4.0.81. This suggests that future users should test different versions to achieve optimal results~~the inclusion of GCPs and the reason for this cannot be determined. The "black box" nature of this proprietary software and small number of adjustable parameters clearly limits the ~~applications~~application of this post-processing tool for scientific ~~applications~~purposes.

3.4 Applications

The distributed snow depth maps generated from UAV imagery are of great utility for understanding snow processes at previously unrealized resolutions, spatial coverages and frequencies. ~~These products may directly lead to a greater understanding of snow phenomena and/or inform, initialize and validate distributed models at a high resolution. Figure 7~~Figure 6 provides examples of UAV derived distributed snow depth maps. The identification of snow dune structures, which correspond to in-field observations, is a qualitative validation that UAV derived DSM differencing does indeed provide reasonable information on the spatial variability of snow depth. Actual applications will depend upon the surface, snow depth and other deployment considerations as discussed.

~~In the prairies, as discussed earlier, it is reasonable to use this technique to measure peak snow accumulation. Besides providing an estimate of the total snow volume, this technique can also inform snow cover depletion curve estimation and description (Pomeroy et al., 1998). Simple snow cover depletion models can be parameterized with estimates of the mean and standard deviation of the snow depth (Essery and Pomeroy, 2004), which otherwise are obtained from snow surveying. For 2015, the bias corrected peak snow accumulation at the short stubble site had a mean of 28.2 cm and a standard deviation of 7.2 cm while the tall stubble site had a mean of 38 cm and standard deviation of 6.2 cm. These values correspond to coefficients of variation of 0.255 and 0.173, at the short and tall stubble sites respectively, which are similar to previous observations from corresponding landforms/surfaces (Pomeroy et al., 1998). While not discussed in this paper, the classification of the orthomosaics can quantify snow-covered area (SCA), providing a validation tool for depletion prediction (Fig. 8a). Orthomosaics have the same horizontal accuracy and resolution as the DSMs; the vertical errors are irrelevant as orthomosaics lack a vertical component. Interpretation of snow processes from orthomosaics is therefore possible regardless of surface characteristics or snow depth.~~

Applications at the alpine site also include the ability to estimate the spatial distribution of snow depth change due to ablation (Fig. ~~8b~~7). To obtain ablation rates, the spatial distribution of snow density is still needed but it may be estimated with a few point measurements or with parameterizations dependent upon snow depth (Jonas et al., 2009; Pomeroy and Gray, 1995). In Fig. ~~8b~~7 the mean difference in snow depth between the two flights was 0.9 m; this gives a SNR of ~11 which is more than sufficient to confidently assess the spatial variability of melt.

Despite the limitations and deployment considerations discussed, ~~UAVs are~~the Ebee RTK was capable of providing accurate data at ~~unprecedented~~very high spatial and temporal resolutions~~ that can advance understanding~~. A direct comparison between fixed wing and multirotor platforms is necessary to determine how snow depth errors may respond to variations in wind speed and lighting conditions. Until then, based on this experience and results of ~~snow processes. The~~other recent studies (Vander Jagt et al., 2015; Bühler et al., 2016; De Michele et al., 2016), we do not expect there to be large differences in errors between platform type. Rather, the most important consideration ~~is~~ when planning to map snow depth with a UAV should be whether the anticipated ~~signal-to-noise ratio~~SNR will allow for direct estimates of snow depth or snow depth change. ~~This~~The SNR issue limits the use of this technique to areas with snow depths or observable changes sufficiently larger than the SD of the error. ~~This analysis established this threshold, at a minimum, to be ~30 cm.~~We propose a mean snow depth threshold of ~30 cm is necessary to obtain meaningful information on snow depth distribution with current technology. This threshold is equal to four times the mean observed SD (Rose criterion), but will vary with the application, site and user's error tolerance. ~~Regardless of the accuracy of the absolute surface values, the relative variability within the DSM may offer fresh insights into the spatial variability of snow depth and snow surface roughness. Previous work on the statistical properties of snow depth (Deems et al., 2006; Shook and Gray, 1996) and snow surface roughness (Fassnacht et al., 2009; Manes et al., 2008) could be extended to consider even finer, centimetre-scale, variability over large areas.~~

The use of SfM in shallow snow environments, such as on the Canadian Prairies, is therefore limited to measuring near-maximum snow depths. Besides providing an estimate of the total snow volume, this information can also inform snow cover depletion curve estimation and description (Pomeroy et al., 1998). Simple snow cover depletion models can be parameterized with estimates of snow depth mean and coefficient of variation (Essery and Pomeroy, 2004), which otherwise need to be obtained from snow surveying. For 2015 coefficients of variation from the peak snow depth maps were 0.255 and 0.173, at the short and tall stubble sites respectively, which are similar to previous observations from corresponding landforms/surfaces (Pomeroy et al., 1998).

In addition to parameterising snow cover depletion models, UAV data could also be used to test their performance as Structure from Motion processing of UAV images produces orthomosaics in addition to DSMs. Sequences of orthomosaics are especially useful to quantify the spatio-temporal dynamics of snow covered area (SCA) depletion processes. Orthomosaics are complementary products to DSMs and their quality is subject to the same deployment conditions as DSMs. Orthomosaics have the same horizontal accuracy and resolution as the DSMs but without a vertical component any DSM vertical errors are irrelevant. Interpretation of SCA from orthomosaics is therefore possible regardless of surface characteristics or snow depth. The classification of orthomosaics to quantify surface properties will introduce error, and can be challenging in changing light conditions, which changes the spectral response of snow or non-snow covered areas across the surface. Typical supervised and unsupervised pixel based classification procedures can be readily applied. Since UAV imagery is at a much higher resolution than satellite or airborne imagery classification differences in spectral response due to varying light conditions can be compensated for by using object oriented classification which also takes into account shape, size, texture, pattern and context (Harayama and Jaquet, 2004).

An example of a snow-covered depletion curve for the prairie site is presented in Fig. 8. A simple unsupervised classification of the orthomosaic into snow and non-snow classes quantifies the earlier exposure of the tall wheat stubble relative to the short wheat stubble. The tall stubble surface is an illustrative example of the advantages UAVs offer for SCA quantification. Tall stubble is a challenging surface to quantify SCA on as snow is prevalent for a time below the exposed stubble surface rendering other remote sensing approaches inappropriate. From an oblique perspective, the exposed stubble obscures the underlying snow and prevents the classification of SCA from georectification of terrestrial photography (Fig. 9). Due to the surface heterogeneity on small scales (stubble, soil and snow all regularly occurring within 30 cm) satellite, and most aerial, imagery struggles with clearly identifying SCA. To identify features accurately, in this case exposed stubble versus snow, multiple pixels are needed per feature (Horning and DuBroff, 2004). The 3.5 cm resolution of the orthomosaic corresponds to approximately three pixels to span the 10 cm stubble row which is sufficient for accurate SCA mapping over a tall stubble surface. The advantages of high-resolution UAV orthomosaics are obviously not limited to SCA mapping of snow between wheat stubble and can be readily applied to other challenging heterogeneous surfaces where SCA quantification was previously problematic. Snow cover data at this resolution can quantifying the role of vegetation on melt processes at a micro-scale, which can in turn inform and validate snowmelt process understanding.

4. Conclusions

~~A new tool, a small UAV that took photographs from which~~ The accuracy of DSMs and orthomosaics ~~were~~, generated through application of SfM techniques to imagery captured by a small UAV, was evaluated in two different environments, mountain and prairie, to verify its ability to quantify snow depth and its spatial variability ~~for varying weather conditions~~ over the ablation period. The introduction of functional UAVs to the scientific community requires a critical assessment of what can reasonably be expected from these devices over ~~the~~ seasonal snowcover. Snow represents one of the more challenging surfaces for UAVs and SfM techniques to resolve due to the lack of contrast and high surface reflectance. Field campaigns assessed the accuracy of the Ebee RTK system over flat prairie and complex terrain alpine sites subject to wind redistribution and spatially variable ablation associated with varying surface vegetation and terrain characteristics. The mean accuracies of the DSMs were 8.1 cm for the short stubble surface, 11.5 cm for the tall surface and 8.7 cm for the alpine site. These DSM errors translate into mean snow depth errors of 8.8 cm, 13.7 cm and 8.5 cm over the short, tall and alpine sites respectively. Ground control points were needed to achieve this level of accuracy. ~~Error varied with bias, which allowed application of a bias correction to improve the accuracy of the snow depth estimates, but this required additional surface observations.~~ The SfM technique provided meaningful information on maximum snow depth at all sites, and snow depth depletion could also be quantified at the alpine site due to the deeper snowpack and consequent higher signal-to-noise ratio. These findings demonstrate that SfM can be applied to accurately estimate snow depth and its spatial variability only in areas with snow depth > 30 cm. This restricts its application for shallow, windblown snowcovers. Snow depth estimation accuracy varied with wind speed, surface characteristics and sunlight; the most consistent performance was found for wind speeds < ~~6m~~10 m s^{-1}, surfaces with insignificant vegetation cover, clear skies and high sun angles. The ability to generate good results declined over especially homogenous snow surfaces and southerly aspects in mountain terrain. Clear sky conditions were favourable for high snow-covered fractions with limited snow surface brightness contrast. During snowmelt with reduced snow-covered fraction, clear sky conditions caused overexposure of snow pixels and erroneous points in the point clouds.

The challenges of applying SfM to imagery collected by a small UAV over snow complicate the generation of DSMs and orthomosaics relative to other surfaces with greater contrast and identifiable features. Regardless, the unprecedented spatial resolution of the DSMs and orthomosaics, low costs and "on-demand" deployment provide exciting opportunities to quantify previously unobservable small-scale variability in snow depth that will only improve the ability to quantify snow properties and processes.

Acknowledgements

The authors wish to acknowledge the reliable assistance of Spatial Geomatics, Ltd of Calgary, Alberta who provided strong technical support and access to a dGPS unit, courtesy Dr. Cherie Westbook that made this research possible. Funding was provided by NSERC Research Tools and Instruments and Discovery grants, the NSERC Changing Cold Regions Network, the NSERC Postgraduate Scholarships-Doctoral Program, the Global Institute for Water Security and the Canada Research Chairs programme. Logistical support from Fortress Mountain Ski Resort, the University of Calgary Biogeoscience Institute and field assistance from May Guan, Angus Duncan, Kevin Shook, Sebastian Krogh and Chris Marsh of the Centre for Hydrology and post-processing support from Chris Marsh are gratefully noted.

References

Armstrong, R. and Brun, E.: Snow and Climate: Physical Processes, Surface Energy Exchange and Modeling, Cambridge University Press, Cambridge, UK., 222 pp., 2008.

Bewley, D., Pomeroy, J. W. and Essery, R.: Solar Radiation Transfer Through a Subarctic Shrub Canopy, Arctic, Antarct. Alp. Res., 39(3), 365–374, 2007.

Boufama, B., Mohr, R. and Veillon, F.: Euclidean Constraints for Uncalibrated Reconstruction, in 4th International Conference on Computer Vision (ICCV ’93), 466–470, IEEE Computer Society, Berlin Germany., 1993.

Bühler, Y., Marty, M., Egli, L., Veitinger, J., Jonas, T., Thee, P. and Ginzler, C.: Snow depth mapping in high-alpine catchments using digital photogrammetry, Cryosph., 9, 229–243, doi:10.5194/tc-9-229-2015, 2015.

Bühler, Y., Adams, M. S., Bösch, R., and Stoffel, A.: Mapping snow depth in alpine terrain with unmanned aerial systems (UASs): potential and limitations, Cryosph., 10, 1075-1088, doi:10.5194/tc-10-1075-2016, 2016.

Collier, O.: The Impact on Topographic Mapping of Developments in Land and Air Survey: 1900-1939, Cartogr. Geogr. Inf. Sci., 29, 155–174, 2002.

DeBeer, C. M. and Pomeroy, J. W.: Modelling snow melt and snowcover depletion in a small alpine cirque, Canadian Rocky Mountains, Hydrol. Process., 23(18), 2584– 2599, 2009.

DeBeer, C. M. and and Pomeroy, J. W.: Simulation of the snowmelt runoff contributing area in a small alpine basin, Hydrol. Earth Syst. Sci., 14(7), 1205–1219, 2010.

~~Fassnacht, S. and Elder, K.: Fractal Distribution of Snow Depth from Lidar Data, J. Hydrometeorol., 7, 285–297, 2006.~~

Deems, J., Painter, T. and Finnegan, D.: Lidar measurement of snow depth: a review, J. Glaciol., 59(215), 467–479, doi:10.3189/2013JoG12J154, 2013.

Deems, J. S. and Painter, T. H.: LiDAR measurement of snow depth: accuracy and error sources, in Proceedings of the 2006 International Snow Science Workshop, 330-338, Telluride, Colorado., 2006.

De Michele, C., Avanzi, F., Passoni, D., Barzaghi, R., Pinto, L., Dosso, P., Ghezzi, A., Gianatti, R., and Della Vedova, G.: Microscale variability of snow depth using U.A.S. technology, The Cryosphere~~ Discuss., 9, 1047–1075~~, 10, 511-522, doi:10.5194/~~tcd-9-1047-2015, 2015.~~tc-10-511-2016, 2016

Dingman, S. L.: Snow and Snowmelt, in:Physical Hydrology, 2nd ed., Prentice Hall, Upper Saddle River, New Jersey., 166-219, 2002.

~~Dozier, J. and Painter, T. H.: Multispectral and Hyperspectral Remote Sensing of Alpine Snow Properties, Annu. Rev. Earth Planet. Sci., 32(1), 465–494, doi:10.1146/annurev.earth.32.101802.120404, 2004.~~

Dumanski, S., Pomeroy, J. W. and Westbrook, C. J.: Hydrological regime changes in a Canadian Prairie basin, Hydrol. Process., 2015.

~~Egli, L., Jonas, T., Grünewald, T., Schirmer, M. and Burlando, P.: Dynamics of snow ablation in a small Alpine catchment observed by repeated terrestrial laser scans, Hydrol. Process., 26(10), 1574–1585, doi:10.1002/hyp.8244, 2012.~~

Essery, R. and Pomeroy, J. W.: Implications of spatial distributions of snow mass and melt rate for snow-cover depletion: theoretical considerations, Ann. Glaciol., 38(1), 261–265, doi:10.3189/172756404781815275, 2004.

~~Fassnacht, S. R., Williams, M. W. and Corrao, M. V: Changes in the surface roughness of snow from millimetre to metre scales, Ecological Complexity, 6, 221–229, doi:10.1016/j.ecocom.2009.05.003, 2009.~~

~~Förstner, W.: A feature-based correspondence algorithm for image matching, Int. Arch. Photogramm. Remote Sens., 26, 150–166, 1986.~~

Gray, D. M. and Male, D. H.: Handbook of Snow: Principles, Processes, Management, and Use, Pergamon Press, Toronto, Canada., 776 pp., 1981.

Grünewald, T., Schirmer, M., Mott, R. and Lehning, M.: Spatial and temporal variability of snow depth and ablation rates in a small mountain catchment, Cryosph., 4(2), 215–225, doi:10.5194/tc-4-215-2010, 2010.

Harayama, A., Jaquet, J.: Multi-source object-oriented classification of landcover using very high resolution imagery and digital elevation model. UNEP, Switzerland, 2004.

Harder, P., Pomeroy, J. W. and Westbrook, C. J.: Hydrological resilience of a Canadian Rockies headwaters basin subject to changing climate, extreme weather, and forest management, Hydrol. Process., 20, doi:10.1002/hyp.10596, 2015.

~~Harris, C. and Step, M.: A combined corner and edge detector, in Proceedings of the Fourth Alvey Vision Conference, 147–151, Manchester., 1988.~~

Hopkinson, C., Pomeroy, J. W., Debeer, C., Ellis, C. and Anderson, A.: Relationships between snowpack depth and primary LiDAR point cloud derivatives in a mountainous environment, Remote Sens. Hydrol., 352, 1–5, 2012.

Horning, N. DuBroff, N.: Myths and misconceptions about remote sensing. Version 1.0. American Museum of Natural History, Center for Biodiversity and Conservation. 4/29/2016 [online] Available from: http://biodiversityinformatics.amnh.org, 2004.

Vander Jagt, B., A, L., L, W. and M, T. D. D.: Snow Depth Retrieval with UAS Using Photogrammetric Techniques, Geosciences, 5, 264–285, doi:10.3390/geosciences5030264, 2015.

Jonas, T., Marty, C. and Magnusson, J.: Estimating the snow water equivalent from snow depth measurements in the Swiss Alps, J. Hydrol., 378, 161–167, doi:10.1016/j.jhydrol.2009.09.021, 2009.

Jones, H., Pomeroy, J. W., Walker, D. and Hoham, R.: Snow Ecology: An Interdisciplinary Examination of Snow-Covered Ecosystems, Cambridge University Press, Cambridge, UK., 384 pp., 2001.

Kinar, N. and Pomeroy, J. W.: Measurement of the physical properties of the snowpack, Rev. Geophys., 53, doi:10.1002/2015RG000481, 2015.

López-Moreno, J. I., Fassnacht, S. R., Heath, J. T., Musselman, K. N., Revuelto, J., Latron, J., Morán-tejeda, E. and Jonas, T.: Small scale spatial variability of snow density and depth over complex alpine terrain : Implications for estimating snow water equivalent, Adv. Water Resour., 55, 40–52, doi:10.1016/j.advwatres.2012.08.010, 2013.

MacDonald, M. K., Pomeroy, J. W. and Pietroniro, A.: On the importance of sublimation to an alpine snow mass balance in the Canadian Rocky Mountains, Hydrol. Earth Syst. Sci., 14(7), 1401–1415, 2010.

~~Manes, C., Guala, M., Lo, H., Bartlett, S., Egli, L. and Lehning, M.: Statistical properties of fresh snow roughness, Water Resour. Res., 44, 1–9, doi:10.1029/2007WR006689, 2008.~~

Marris, E.: Fly, and bring me data, Nature, 498, 156–158, 2013.

Mattmann, C. A., Painter, T., Ramirez, P. M., Goodale, C., Hart, A. F., Zimdars, P., Boustani, M., Khudikyan, S., Verma, R., Caprez, F. S., Deems, J., Trangsrud, A., Boardman, J. and Ave, A.: 24 hour near real time processing and computation for the JPL Airborne Snow Observatory, in Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, 5222–5225., 2014.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundezewicz, Z. W., Lettenmaier, D. P. and Stouffer, R. J.: Stationarity Is Dead: Whither Water Management?, Science, 319, 573–574, 2008.

~~Mittaz, C., Imhof, M. and Hoelzle, M.: Snowmelt Evolution Mapping Using an Energy Balance Approach over an Alpine Terrain, Arctic, Antarct. Alp. Res., 34, 274–281, 2015.~~

Mote, P. W., Hamlet, A. F., Clark, M. P. and Lettenmaier, D. P.: Declining mountain snowpack in western North America, Bull. Am. Meteorol. Soc., 86(1), 39–49, doi:10.1175/BAMS-86-1-39, 2005.

~~Musselman, K. N., Pomeroy, J. W. and Link, T. E.: Variability in shortwave irradiance caused by forest gaps: Measurements, modelling, and implications for snow energetics, Agric. For. Meteorol., 207, 69–82, doi:10.1016/j.agrformet.2015.03.014, 2015.~~

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

Pomeroy, J. W. and Gray, D. M.: Snow accumulation, relocation and management, National H., Environment Canada, Saskatoon, SK., 144 pp., 1995.

Pomeroy, J. W., Gray, D. M. and Landine, P. G.: The Prairie Blowing Snow Model: characteristics, validation, operation, J. Hydrol., 144(1-4), 165–192, 1993.

Pomeroy, J. W., Gray, D. M., Shook, K., Toth, B., Essery, R. L. H., Pietroniro, A. and Hedstrom, N.: An evaluation of snow accumulation and ablation processes for land surface modelling, Hydrol. Process., 12, 2339–2367, 1998.

Rango, A. and Laliberte, A.: Impact of flight regulations on effective use of unmanned aircraft systems for natural resources applications, J. Appl. Remote Sens., 4, doi:10.1117/1.3474649, 2010.

~~Reba, M. L., Marks, D., Winstral, A., Link, T. E. and Kumar, M.: Sensitivity of the snowcover energetics in a mountain basin to variations in climate, Hydrol. Process., 3321, 3312–3321, doi:10.1002/hyp.8155, 2011.~~

Rose, A.: The Visual Process, in: Vision: Human and Electronic, Plenum Press, New York., 1-28, 1973.

Roze, A., Zufferey, J.-C., Beyeler, A. and Mcclellan, A.: eBee RTK Accuracy Assessment, Lausanne, Switzerland., 7 pp., 2014.

SenseFly: Menu Process Options Point Cloud Densification, 1/10/2015 [online] Available from: https://sensefly.zendesk.com/hc/en-us/articles/204542828-Menu-Process-Options-Point-Cloud-Densification, 2015.

Shook, K. and Gray, D. M.: Small-Scale Spatial Structure of Shallow Snowcovers, Hydrol. Process., 10, 1283–1292, 1996.

Spetsakis, M. and Aloimonost, J.: A Multi-frame Approach to Visual Motion Perception, Int. J. Comput. Vis., 6, 245–255, 1991.

Stewart, I. T., Cayan, D. R. and Dettinger, M. D.: Changes in Snowmelt Runoff Timing in Western North America under a `Business as Usual' Climate Change Scenario, Clim. Change, 62(1-3), 217–232, doi:10.1023/B:CLIM.0000013702.22656.e8, 2004.

~~Strecha, C.: The Accuracy of Automatic Photogrammetric Techniques on Ultra-Light UAV Imagery, in Photogrammetric Week Series, edited by D. Fritsch, 289–294, Stuttgart., 2011.~~

Sturm, M.: White water: Fifty years of snow research in WRR and the outlook for the future, Water Resour. Res., 51, 4948–4965, doi:10.1002/2015WR017242, 2015.

Szeliski, R. and Kang, S.: Recovering 3D Shape and Motion from Image Streams Using Nonlinear Least Squares, J. Vis. Commun. Image Represent., 5, 10–28, 1994.

Westoby, M., Brasington, J., Glasser, N., Hambrey, M. and Reynolds, J.: "Structure-from-Motion" photogrammetry: A low-cost, effective tool for geoscience applications, Geomorphology, 179, 300–314, doi:10.1016/j.geomorph.2012.08.021, 2012.

Table 1: Flight plan specifications

Variable	Prairie Site	Alpine Site
Flight altitude	100m	100m
Lateral overlap	70%	85%
Longitudinal overlap	70%	75%
Ground resolution	3 cm pixel^{-1}	3 cm pixel^{-1}
Number of flights (over snow/over non-snow)	22/3	18/4
Approximate area surveyed per flight	1 km^2	0.32 km^2

Table ~~2~~1: Absolute ~~snow depth~~surface accuracy ~~summary*~~summary[a]

Area	Variable	Mean~~*~~ (cm)	~~Max~~Maximum (cm)	~~Min~~Minimum (cm)	Total Points [c]
alpine-bare	RMSE	8.7	15	4	1120
alpine-bare	Bias [b]	5.6	11	1	1120
alpine-bare	SD	6.2	12	3	1120
alpine-snow~~Alpine~~	RMSE	~~8~~7.5	14~~.0~~	3	101
alpine-snow~~Alpine~~	Bias~~**~~ [b]	4.~~1~~4	~~11.0~~13	~~0~~1	101
~~Alpine~~alpine-snow	SD	~~7.1~~5.4	~~12.0~~13	3	101
Short	RMSE	8.~~8~~1	12.5~~15.8~~	~~0~~4.4	357
Short	Bias~~**~~ [b]	~~5~~4.4	~~15~~11.2	0	357
Short	SD	6.~~1~~3	9.5	~~10.~~3.2	~~0~~357
Tall	RMSE	11.5~~13.7~~	~~27.2~~18.4	~~0~~4.9	357
Tall	Bias~~**~~ [b]	~~9.8~~6.6	~~26.4~~17.5	0.3	357
Tall	SD	8.~~3~~4	14.2~~13.9~~	~~0~~3.1	357

~~*~~[a] summary excludes ~~four~~five flights identified to be problematic ~~due to windy conditions~~

~~**~~[b] mean of absolute bias values

[c] cumulative points used to assess accuracy over all assessed flights

Table 2: Absolute snow depth accuracy summary [a]

Area	Variable	Mean (cm)	Maximum (cm)	Minimum (cm)	Total Points [c]
Alpine	RMSE	8.5	14.0	3	83
Alpine	Bias [b]	4.1	11.0	0	83
Alpine	SD	7.1	12.0	3	83
Short	RMSE	8.8	15.8	0	323
Short	Bias [b]	5.4	15.2	0	323
Short	SD	6.1	10.3	0	323
Tall	RMSE	13.7	27.2	0	323
Tall	Bias [b]	9.8	26.4	0	323
Tall	SD	8.3	13.9	0	323

[a] summary excludes two flights identified to be problematic

[b] mean of absolute bias values

[c] cumulative points used to assess accuracy over all assessed flights

Table 3: Summary of areas excluded due to erroneous points with respect to snow covered area at Alpine site.

Flight [a]	Snow covered area (%)	Percentage of snow covered area excluded (%)
5-19_N	45.9	0.0
5-20_S	32.6	2.0
5-22_N	39.8	0.0
6-01_N	24.0	0.0
6-08_N	12.5	3.2
6-18_N	5.3	19.3
6-24_N	3.1	21.9
6-24_S	3.7	18.9

[a] month-day_portion of study area

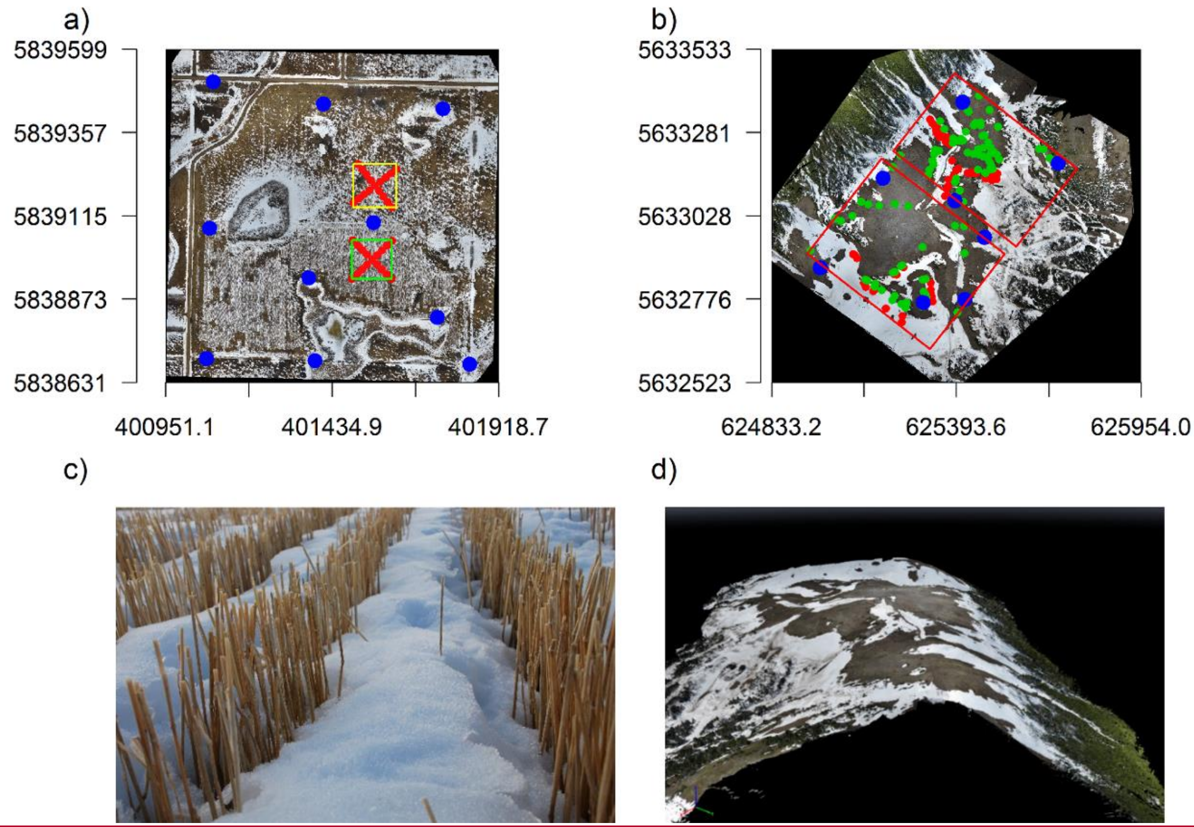

Figure 1: Orthomosaics of a) the prairie site located near Rosthern, Saskatchewan and b) the alpine site at Fortress Mountain Snow Laboratory, Kananaskis, Alberta . The prairie site image (March 19, 2015) has polygons depicting areas used for peak snow depth estimation over short (yellow) and tall (green) stubble. The alpine site image (May 22, 2015) was split into two separately processed subareas (red polygons). Red points in a) and b) are locations of manual snow depth measurements while green points at the alpine site b) were used to test the accuracy of the DSM over the bare surface. Ground control point (GCP) locations are identified as blue points. Axes are UTM coordinates for the prairie site (UTM zone 13N) and alpine site (UTM zone 11N). The defining feature of the prairie site was the c) wheat stubble (tall) exposed above the snow surface and at the alpine site was the d) complex terrain as depicted by the generated point cloud (view from NE to SW).

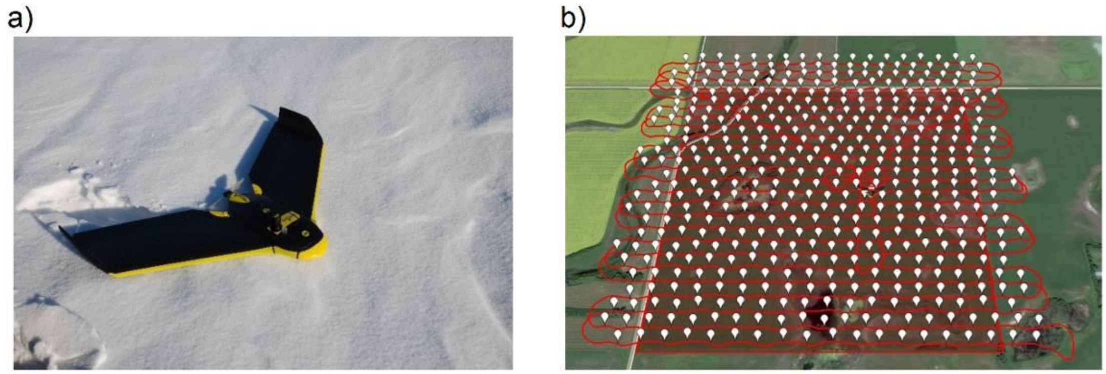

Figure 2: a) Sensefly Ebee RTK, b) a typical flight over the prairie site where red lines represent the flight path of UAV and the white placemarks represent photo locations.

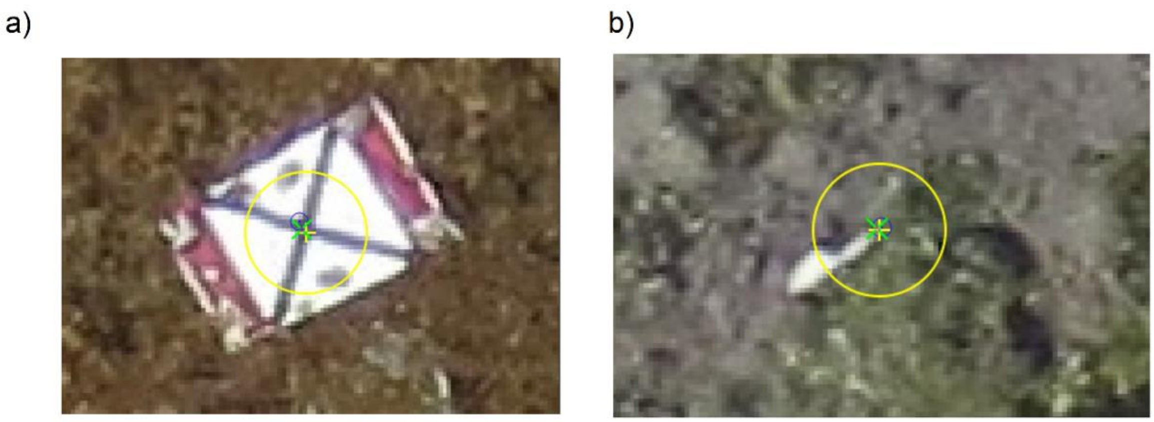

Figure 3: Examples of ground control points that included a) tarps (2.2 m x 1.3 m) and b) identifiable rocks at the same magnification as the tarp.

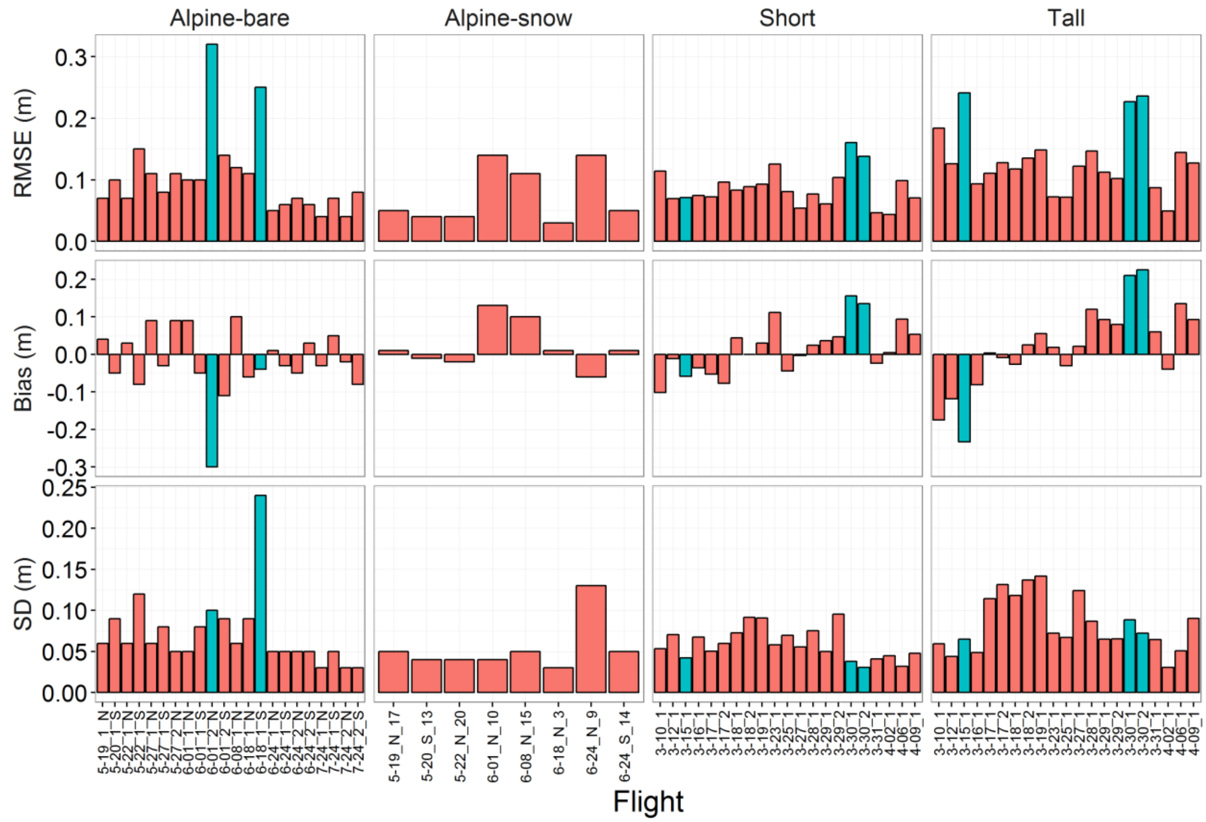

Figure 4: Root mean square error (RMSE, top row), Bias (middle row) and standard deviation (SD) of DSMs with respect to surface over alpine-bare, alpine-snow, and short and tall stubble at prairie site, respectively. Blue bars highlight problematic flights and are excluded from summarization in Table ~~1~~2. X-axis labels represent month-date-flight number of the day (to separate flights that occurred on the same day). Alpine-bare accuracies are separated into north or south areas, reflected as _N or _S at the end. The last number in the alpine-snow x-axis label is the number of observations used to assess accuracy as they vary between 3 and 20.

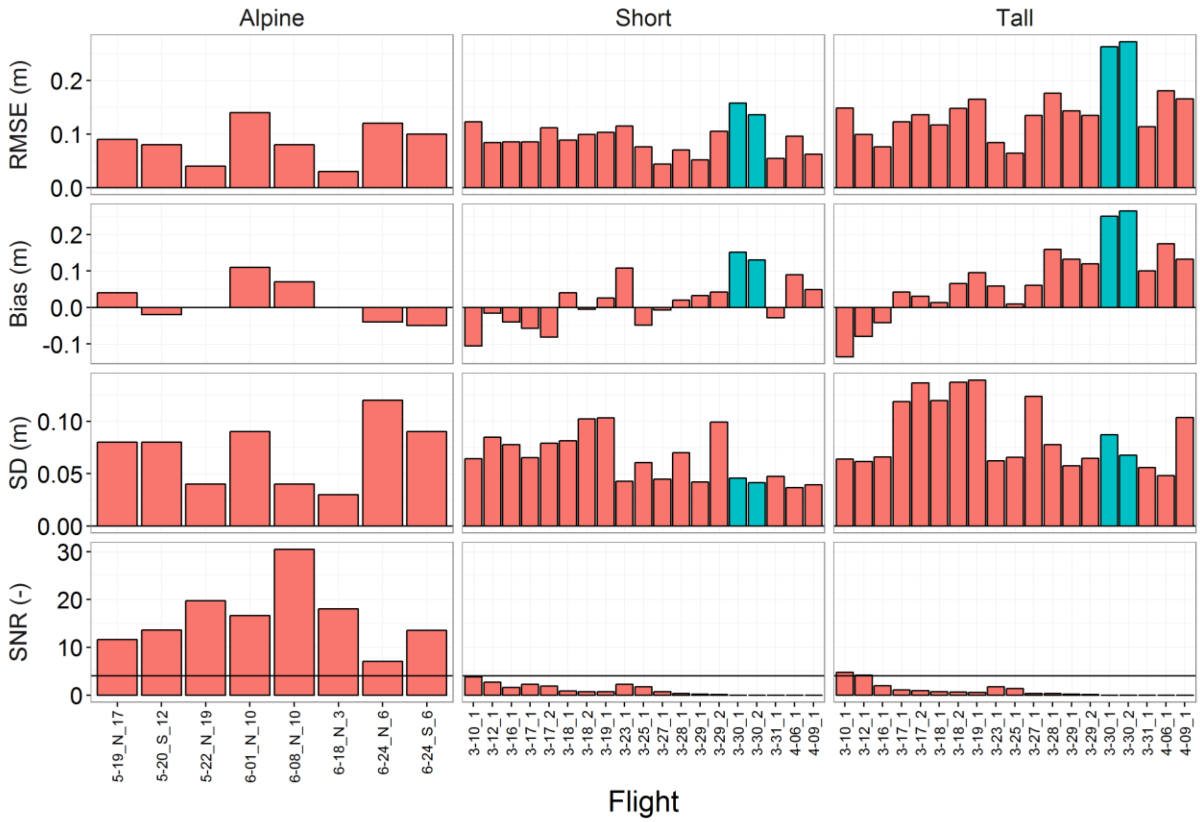

Figure 5: Estimated UAV snow depth error with respect to observed snow depth for the alpine site and the short and tall stubble treatments at prairie site. Blue bars highlight problematic flights and are excluded from summarization in Table ~~1~~3. X-axis labels represent month-date. The last value in prairie labels is the flight of the day (to separate flights that occurred on the same day). Alpine labels separate the north or south flight areas, reflected as _N or _S respectively, and the last value is the number of observations used to assess accuracy as they vary between 3 and 19. Horizontal line in the SNR plots is the Rose criterion (SNR=4) that is used to identify flights with a meaningful snow depth signal.

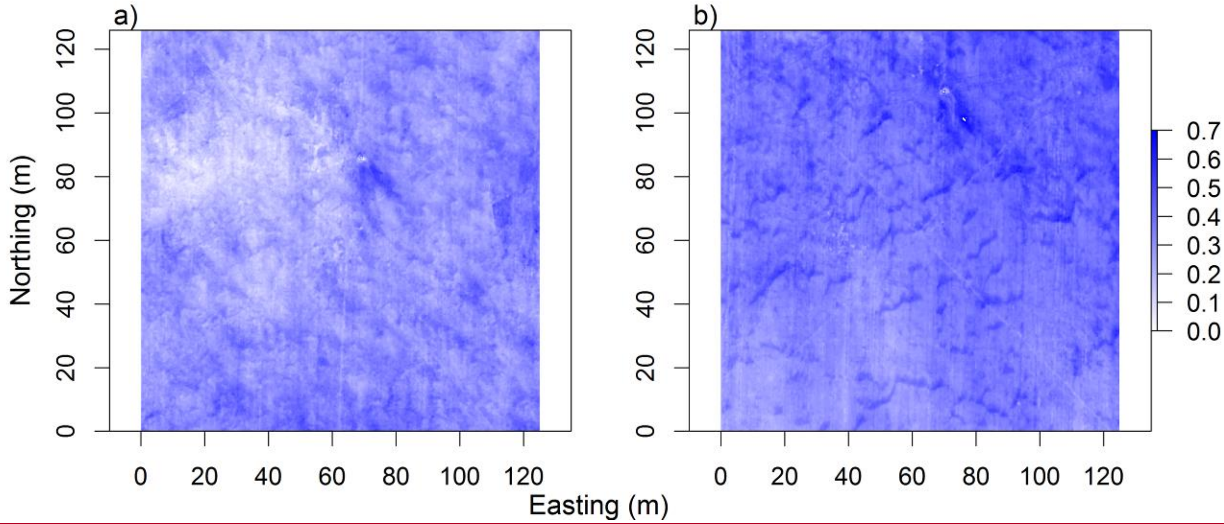

Figure 6: Bias corrected distributed snow depth (meters) for a) short and b) tall stubble treatments at peak snow depth (March 10, 2015) at the prairie site.

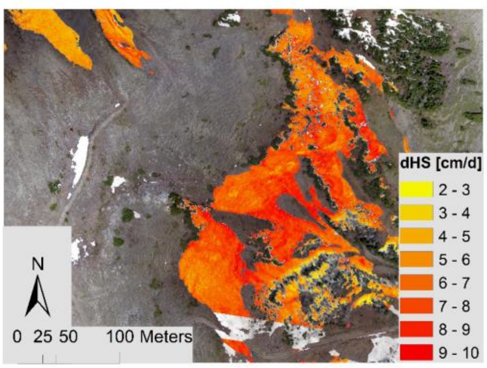

Figure 7: Snow depth change per day (dHS d^{-1}) between May 19 and June 1 in the northern portion of the alpine site.

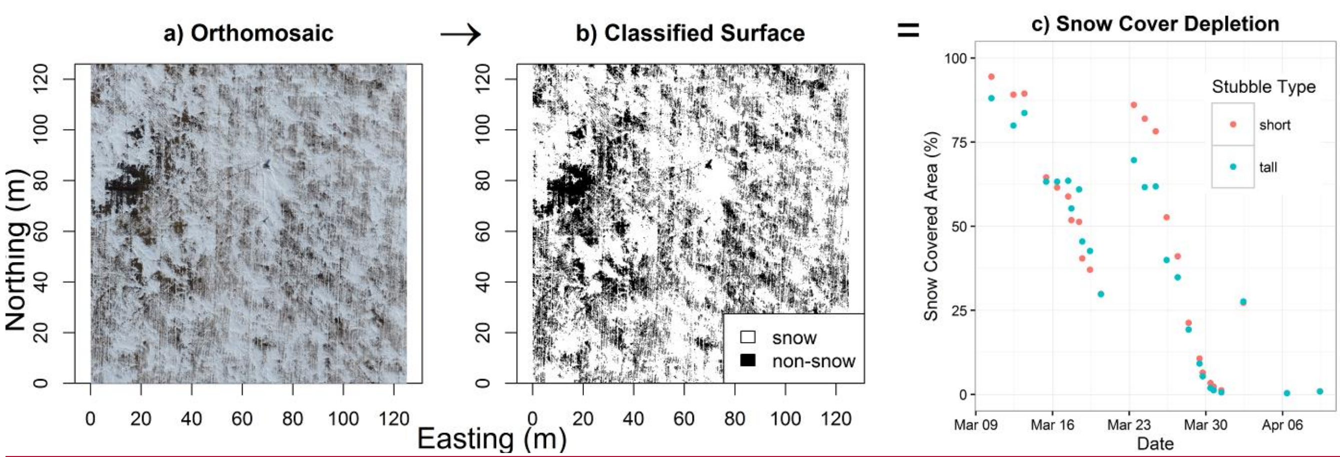

Figure 8: Estimation of snow covered area requires an a) orthomosaic which is then b) classified into snow and non-snow covered area. W This produces a c) snow cover depletion curve when a sequence of orthomosaics are available. The short and tall stubble surface snow covered areas at the prairie site are contrasted, with a snowfall event evident on March 23.

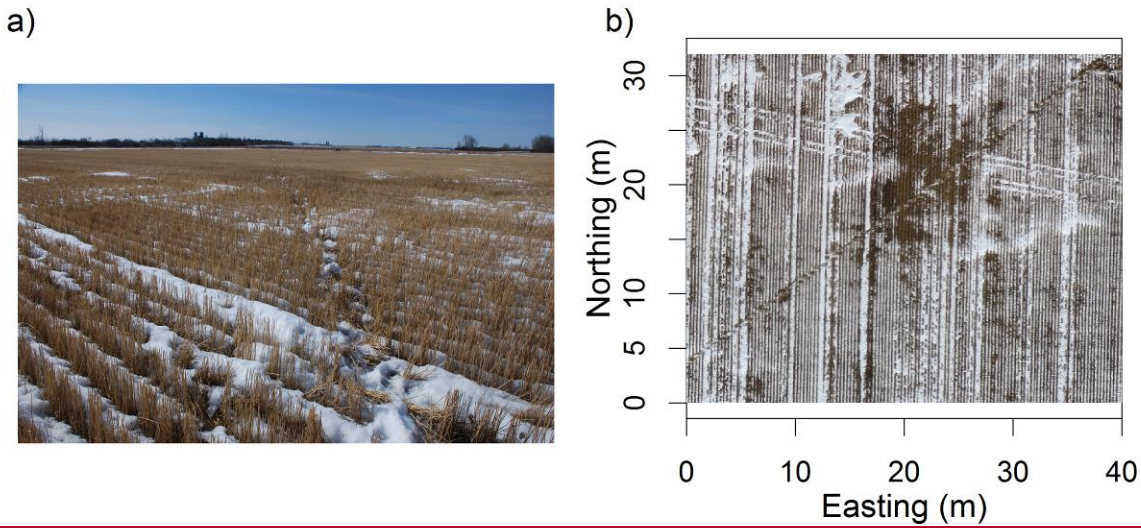

Figure 9: a) An oblique photograph demonstrates the issue of tall stubble obscuring underlying snowcover when considered in contrast to b) a UAV orthomosaic of the same area on the same date that clearly shows widespread snowcover.