Review:
Measuring snow specific surface area with 1.30 and 1.55 um bidirectional reflectance factors
A. Schneider, M. Flanner, and R. De Roo

Summary:

The authors present a novel empirical measurement method to derive snow specific surface area (SSA) from reflectance measurements that may be used to develop a tool for snow SSA measurements in the field, called the NERD. The impact of some snow physical characteristics on measured bidirectional reflectance factors (BRFs) is evaluated, but eventually the SSA measurement method is formulated by resorting to an empirical relationship between measured snow BRF and snow SSA determined by computed tomography (CT) for a small number of samples, without providing a quantitative uncertainty assessment or validation of the SSA measurement method.
While the manuscript also contains some interesting mostly qualitative results beyond the presentation of the measurement method, e.g., the influence of the 'snow grain shape' used for modeling BRFs and the preliminary results given in Fig. 7, several major points should be addressed before publication is considered. The most critical ones are: overly optimistic and premature claims about the presented measurement principle that are based on very limited analysis results; overly simplified and misleading generalizations about previously presented measurement methods that can also be used to derive snow SSA; there is no validation of the presented SSA measurement method against an independent measurement method and no quantitative discussion of SSA measurement uncertainties. The manuscript could also benefit from a more streamlined structure and a clearer writing style.

General comments:

1)
Validation of the NERD measurement principle would entail comparing snow SSA values measured with the NERD to independent SSA measurements, e.g., SSA measured by other optical measurement methods or by methane gas absorption (or at the very least a vast number of CT measurements that are not used for developing the NERD measurement method).
The authors have merely shown the consistency of BRF measurements (not SSA measurements) performed with the NERD on highly uniform, diffusely scattering reflectance standard surfaces. Neither the effects of extensive volume-scattering in snow nor possible directional/specular effects at the snow surface are included here. For example, is there light leakage into or out of the NERD on uneven surfaces, and what effect do oriented surface structures (small dents or ripples on the surface) have on measured BRF (I would expect this to be of greater concern here than for previously presented methods that are based on diffuse instead of directional reflectance)?
Furthermore, no systematic and quantitative uncertainty assessment of snow SSA measurements with the NERD is presented (although the impact of various sources of uncertainty on measured BRF is discussed), which is an integral part of any presentation of a novel measurement method, as described in the 'Guide to the expression of uncertainty in measurement', for example. Particularly an intercomparison of multiple measurement methods or comparisons of in situ and remote sensing observations, as alluded to in the conclusions of the manuscript, require a thorough assessment of the uncertainties affecting each measurement method to allow drawing reliable conclusions from such comparisons. In this manuscript, only a rough estimate of the variability of an empirically derived exponential relationship between BRF and SSA is interpreted as SSA measurement uncertainty, based on a low number of sample measurements. For empirical measurement methods,

i.e., measurement methods that are not based on physical measurement models but on statistical correlation as presented in this manuscript, a large number of samples is required to guarantee a high statistical significance. A low number of samples can easily lead to highly underestimated (or overestimated) measurement uncertainties over the entire or parts of the measurement range, especially when not validated against independent measurement methods.
A detailed NERD SSA measurement uncertainty assessment and validation would also include how the mismatch between CT measurement volume and NERD measurement volume affects the analysis. While suggesting to use the NERD to determine the large-scale spatial variability of snow SSA, the authors fail to discuss this very effect at a small scale of ~10 cm, relevant for the NERD measurement principle, e.g., how do CT samples and NERD measurements match on the probed snow blocks, did they use enough CT samples to provide a reasonable estimate of the spatial variability within the snow volume probed by the NERD, did they only use the very top of the snow samples for their CT analysis because only this part is probed by the NERD due to the long NIR wavelengths of its illumination sources?

2)
Without validation of NERD SSA measurements, and thus without an essential part of any presentation of a novel measurement method, some of the made statements seem rather unfounded. For example, the authors seem to repeatedly suggest that the NERD can accurately measure BRFs for snow, but how do they know that? Only BRF measurements on reflectance standards are somewhat validated as their nominal reflectance values are known and as they are roughly compared in this manuscript to reflectance measurements of reflectance standards that were performed with previously presented measurement methods. Going on to claim accurate snow SSA measurements based on these basic results and without any quantitative validation seems to be even more unfounded than claiming accurate snow BRF measurements.
The authors also stress the high cost of other snow SSA measurement methods and the low cost of their anticipated NERD instrument. Yet, I fail to see how the development of a useful measurement tool in the field based on their measurement principle will lead to a price of the instrument of less than thousands of USD, similar to the cost of some of the other optical measurement methods that can be used to derive snow SSA. Can the authors give a realistic cost estimate to justify such claims, factoring in development, prototyping, weather sealing, installation of permanent and sturdy components, … of a portable NERD measurement tool? If they mean the NERD will be cheaper than a CT or a high-resolution spectrometer, then they should not overly generalize by including all previously presented measurement methods in this statement, while they actually only talk about some of those (see also point 3 below).

3)
Particularly bothersome are some generalizations and omissions when discussing previously presented measurement methods and the motivation behind and the potential benefits of the NERD. Fast and nondestructive snow SSA measurements can be obtained with the InfraSnow (introduced by Gergely et al. 2014) or by contact spectroscopy (introduced by Painter et al. 2007), for example. No sampling is required, no samples are destroyed. In fact, the InfraSnow was developed for some of the same reasons and applications as the NERD, as stated by Gergely et al. (2014), and its presentation additionally included a quantitative uncertainty analysis and measurement validation, yet none of this is mentioned in this manuscript. Instead, all previously presented measurement methods for deriving snow SSA are falsely lumped together, and the manuscript gives the impression that the anticipated NERD method will be the first and only fast, nondestructive snow SSA measurement method, for example. This is poor 'scientific' work.
The authors should either summarize and discuss the various measurement methods separately and in much more detail without overly generalizing and thus without making misleading and false claims, or they should simply state that they are attempting to develop a novel measurement tool that allows fast, nondestructive SSA measurements without suggesting that such measurements are

not possible with (any of the) previously developed measurement methods. Also, NIR photography should be included as a reference in the list of currently available optical SSA measurement methods:

```
@ARTICLE{Matzl+06,
 author = {M. Matzl and M. Schneebeli},
 title = {Measuring specific surface area of snow by near-infrared photography},
 journal = jg,
 year = {2006},
 volume = {52},
 pages = {558--564},
 number = {179},
 doi = {10.3189/172756506781828412}
}
```

4)
The main results in the context of the presentation of the NERD measurement method are summarized in Fig. 6. The effect of relevant sources of uncertainty should be discussed quantitatively instead of only stating the variability of the fitted exponential function for the extremely limited number of samples used to derive the fit (see point 1).
Here, another effect is of interest: The small penetration depth of just a few mm or less at long NIR wavelengths is clearly much shallower than the penetration depth of visible light which forms the main contribution to overall solar irradiation. So, if snow surface SSA is measured at long NIR wavelengths, how realistic is it to analyze overall snow albedo based on the derived SSA value?
It may be that the top few mm of the snow cover that determine long-wavelength NIR reflectance do not represent the full near-surface snow that determines overall snow albedo, e.g., a thin layer of surface hoar or some very small windblown snow fragments deposited at the very surface. Such a discussion would also add further scientific value to the study beyond presenting a novel SSA measurement method.

5)
Terminology: Throughout the manuscript, 'error' should be changed to 'uncertainty' or 'difference', depending on the context. Usage of 'error' when actually talking about 'uncertainty' or simple 'differences' is deprecated in measurement science and avoided to guarantee a more precise and meaningful terminology (see the 'Guide to the expression of uncertainty in measurement'), especially as a realistic assessment of uncertainties (and not errors) becomes increasingly important in remote sensing and climate modeling applications.

Specific comments:

page 1 line 1: Is SSA an important physical property because it directly affects solar radiation, or is it important for another reason and it also affects solar radiation? The authors should specify this.

page 1 line 2f: This is a misleading generalization, if no further details are given about the various different SSA measurement methods (see point 3 above). The authors should remove this sentence and instead focus on the analysis and description of the NERD measurement principle in the abstract, and, e.g., on its possible future use to track snow SSA evolution on a time scale of hours.

page 1 line 17f: I do not understand this sentence. What is 'positive snow internal albedo feedback'? Is this the same as 'positive albedo feedback' in the previous sentence? This sentence could probably be rewritten to clarify.

p.2 l.4: What 'particles'? Snow is not a granular material of individual particles or grains but a material characterized by a complex 3D microstructure with continuous ice and air phases. Probably this should be the 'snow microstructure' or are the authors talking specifically about modeling snow here as a matrix of suspended ice particles?

p.2. l.22: What is 'grain growth'? Snow is not a granular material. 'grain growth' could simply be left out. … where solar heating induces a further decrease in SSA, …

p. 3 l. 1 – 6: This is a misleading generalization (see point 3 above). Discuss different methods separately and in more detail, or simply state (1) that different methods to measure snow SSA for different applications have been presented previously (including the corresponding references) and (2) that this study describes a novel measurement method for fast and nondestructive SSA measurements (without trying to motivate it by making misleading and false claims about other measurement methods). Readers can then go back to the cited studies for details and see for themselves how different measurement methods compare to each other and what may be an advantage or disadvantage for different applications.
Because a short and still adequate, i.e., not misleading or false, description of all relevant previously presented SSA measurement methods may be difficult to achieve within a few of lines of text, the following approach could be used:
keep lines 1 – 3 and add reference Matzl and Schneebeli 2006 (see point 3 above), delete lines 4 – 6, keep line 6f and add: … is widely sought after, which not all (or which only few) previously presented measurement methods allow. Here, we introduce ...

p. 4 l. 18 – 25: This is not a validation for determining BRFs of snow with the NERD due to the very different nature of snow (uneven surface, extensive volume-scattering). So, l. 24f would be more correctly rephrased as: … to obtain BRFs on smooth reflectance standards with a measurement uncertainty of …, or simply delete this statement.

p. 4 l. 32: pixel (2D) or voxel (3D equivalent of pixel)?

p.4 l. 33f: Can the authors give the temperature of the CT and CT sample more accurately than below 0°C, if snow can't survive much more than 15 min? How does this affect snow SSA evolution during the duration of CT measurements? Has this CT been used for snow SSA measurements previously, is the CT resolution high enough to yield reliable snow SSA values, have previous CT measurements of snow with this CT been validated against other measurement methods (or other CT setups)? This information should be included in the text.

p. 5 l. 10: Delete 'relatively'.

p. 5 l. 20: What is 'highly sintered' snow. Old hard snow?

p. 6 l. 9: What is 'visibly apparent snow metamorphosis'? Is this temperature-gradient metamorphism or equal-temperature metamorphism or both, resulting in what type of snow (e.g., depth hoar or melt-freeze or other)? The authors should specify this in the text or simply list the physical properties of the snow sample and refer the reader to the CT images for further information.

p. 6 l. 13ff: Has this Monte Carlo model been validated or at least used for snow previously? Is there any indication of what the expected uncertainty is for applying the model to snow (and not only to Lambertian surface scattering as described on p.7)? Such information should be included in the text, if it is available.

p. 6 l. 23: Why are at least 100 thousand photons used per simulation, while commonly millions of photons are needed for complex Monte Carlo raytracing simulations. Have the authors checked that an increase in photons beyond the photon numbers that they have chosen does not lead to significant changes in the Monte Carlo modeling results for snow? If this is the case both for the tested Lambertian surfaces and for Monte Carlo simulations for volume-scattering snow, they should state this in the text. Or they should state how much a further increase in the number of photons may change the modeling results for snow.

p. 6 l. 24ff: Are the ice particle scattering properties obtained for randomly oriented or preferentially oriented ice particles (horizontally, vertically, something else?) within the snow matrix? How realistic is this assumption for the analyzed snow types? I would intuitively expect that random orientation should be the most realistic assumption in general, but some snow does show strong anisotropy. This information should be included in the manuscript.
Similarly, are all ice particle types equally realistic representations of natural snow? Particularly hexagonal plates seem rather extreme when intuitively compared to the 3D microstructure of natural snow, which is also confirmed in Figs. 4 and 5. Maybe it would be more realistic not to include hexagonal plates in the analysis, which could also streamline the discussion?

p. 6 l. 29f: Delete this sentence, it is a duplicate of l. 22f.

p. 7 l. 14ff: How are the stark differences in CT and NERD measurement volumes included in the analysis? How does this affect the analysis results (see also point 1 above)? This should be discussed in the text.

p. 7 l. 24: How can this be claimed without validating snow BRF measurements with the NERD against any other independent snow BRF measurements, e.g., a gonioreflectometer, and without providing any quantification of the term 'relatively accurate' when comparing NERD BRF measurements and modeled BRFs? The authors should either delete this sentence or provide a detailed quantification instead of a vague statement.

p. 8 l. 10f: Better: With the InfraSnow, Gergely et al. (2014) were able to determine the reflectance values of nominal 25 %, 50 %, and 99 % reflectance standards to within an accuracy of better than 1 %.

p.8 l. 11: 'directional-hemispherical reflectance' is not correct here due to the diffusing cone in the InfraSnow that prevents direct illumination of the snow surface and instead guarantees predominantly diffuse illumination. 'directional-' should be removed.

p. 8 l. 12 – 16: Best to remove these two sentences. Lambertian reflectance standards are only part of the testing performed for the InfraSnow. Additionally, various other surfaces, including snow, are tested. This should be included if l.12,13 are kept in the text. The second sentence does not add important information and is highly speculative, especially when trying to translate the results found for reflectance standards to reflectance measurements on snow, due to the very different nature of NIR light scattering in snow and the differences in the applied measurement techniques (see also point 1 above).

p. 8 l. 28ff: How does this BRF uncertainty affect the derived snow SSA?
And how could light leakage to and from the outside of the NERD due to an uneven snow surface or due to specular reflections also add to the uncertainty in BRF measurements (see also point 1 above)?

p. 8 l. 32ff: The authors should indicate which Figure or Table they are referring to.

p.9 l. 1f: What is an estimate of this uncertainty then? Can the authors provide a quantification? Otherwise, it is better not to include such statements.

p.10 l.3 – 18: I do not see the immediate relevance of this discussion. This could be mostly removed to streamline the manuscript and focus on more crucial results, which are given in the next paragraph. I would prefer to see the authors try to include an actual quantitative uncertainty assessment of their measurement method, including the effect of grain shape on their modeling results instead of this ancillary discussion.

p. 10 l. 22: I do not see a compelling reason in the presented data why this relationship has to be an exponential function. Could it be something else (linear relationship), and would this significantly change any of the results? If so, this effect should be included in the SSA measurement uncertainty.
Also, what does a constant SSA measurement uncertainty across the entire SSA range mean for using the NERD to measure different snow types? High relative SSA measurement uncertainty for snow characterized by low SSA (what snow types are those?) and low relative SSA measurement uncertainty for snow characterized by high SSA (what snow types are those?). This should be added to the text to illustrate the uncertainty in the derived BRF-to-SSA relationship beyond the mere presentation of the values for the root mean square differences.

p. 11 l. 9: Delete 'accurate'. There is no quantitative validation of the SSA measurements in the manuscript. 'quick, reliable, and repeatable' convey the full picture. Even 'repeatable' could be removed because it is implied by 'reliable'.

p. 11 l. 16,17,18: The authors should replace 'will' with 'can' or 'could' or 'may', or preface each of these statements by 'We believe that ….' or 'We intend to use the NERD for ...' or 'The analysis indicates the potential for ….' or similar. Selling such statements as foregone conclusions seems like a far stretch given the limited analysis in the manuscript and the inherent uncertainty of future developments (see also point 1 above).

Caption Figure 3: Snow is not a granular material. Better: … 'as the snow microstructure gets coarser' or 'is characterized by more rounded shapes'.