We thank Professor Roy Rasmussen for commenting on the manuscript.

1) Area incorrect: I agree that the area in my original paper is incorrect. We had a number of prior designs of the hotplate and this is probably left over from one of those sensor heads. Boudala et al. (2014) also found this.

We referenced Boudala et al. (2014). In our opinion, explaining how the conversion factor is handled in Rasmussen et al. (2011) (R11) requires consideration of the area effect, and the 20% lowering described here in R11:

"In practice, this conversion factor was 20% lower because of the imperfect heat transfer from the precipitation to the hot plate (losses to the air, e.g.). The actual conversion rates were determined by comparing the predicted precipitation rate from the hotplate to the measured rates from a GEONOR precipitation gauge in the Double-Fence Intercomparison Reference (DFIR) shield (see discussion below)."

Because the area effect and the 20% lowering are linked, we discuss them together, and thus we repeat the finding of Boudala et al. (2014).

Not sure how I missed it in proof reading. However, I never expected the factor to be perfect, so I fully expected it to need calibration, so at the time I was mainly worried about the hotplate working and not the exact value of the theoretical factor. I think the algorithm worked well for the data we collected, I never thought that the Yankee algorithm would have the same factor (in fact, I don't think it does).

As we stated L268-L274, we **assumed** that the YES algorithm has incorporated R11's surface area and R11's distinction between the theoretical and actual conversion factors.

The reasoning used on page 12 is correct, however, I expected the calibration factor not to be exactly equal to the terms which is why I called it a "calibration" factor and not a constant.

Agreed, it's not a constant. To be consistent with R11, we used "energy conversion factor."

2) The snow particle collection efficiency is the most uncertain part of the original hotplate algorithm. It worked well for the data I had a Marshall, but I was not sure it would work as well elsewhere. The recent WMO solid precipitation experiment evaluated three hotplates, and they performed well, in fact, they had the lowest RMSE of all the snow gauges tested. I attribute this to the aerodynamic profile, making the dependence of snow particle type less than weighing gauges. This report will be coming out before January 1, 2018. We did spend a lot of time worrying about the level that the wind speed was taken from, so your discussion on this is useful. However, the Yankee algorithm uses the wind from the hotplate itself, so the raw data is from 2 meters.

As we explain on L295-L298, we used information conveyed in a personal communication from YES to formulate the snow particle catch efficiency function. In the manuscript, this is indicated "Y12." From that personal communication, the Y12 efficiency function is based on the hotplate-derived wind speed. It is our understanding that there is **no** adjustment of the hotplate-derived wind speed that is input to the Y12 snow particle catch efficiency function.

3) Do you have photos of the sites?

Yes, photographs are available. These are in the references provided in the first column of Table 2.

Are they flat?

No, however, our calculation of the sensible energy output term (the dominant term in the power budget) is based on site-specific calibrations. This is discussed L418-L438.

The hotplate will be biased if the wind is not horizontal or there is upstream blocking. What I would also like to see is an uncertainty analysis of the truth gauge (ETI) as compared to the hotplate data.

Figure 12b shows that when we apply the catch efficiency function recommended by R11 (E R11) our result compares well with weighing gauge measurements that were also corrected for inefficient catch. The Figure 12b is a scatterplot of the 19 analyzed OWL snow events in our data set. Development of an uncertainty analysis for the NOAH-II gauge (aka, ETI gauge) would require significantly more information. For example, as shown in Kochendorfer et al. (2017), this can involve hundreds of concurrent measurements from a DFIR and the gauge under test.

Kochendorfer, J., Rasmussen, R., Wolff, M., Baker, B., Hall, M. E., Meyers, T., Landolt, S., Jachcik, A., Isaksen, K., Brækkan, R., and Leeper, R.: The quantification and correction of wind-induced precipitation measurement errors, Hydrol. Earth Syst. Sci., 21, 1973-1989, https://doi.org/10.5194/hess-21-1973-2017, 2017

4) The paper is quite detailed. If possible, I would try to focus on less of the details of the testing and more on the results. How well is the hotplate performing given the current algorithm?

As stated in the abstract, and in the conclusion, the manuscript has two objectives: 1) incorporation of the radiation measurements into the power budget, and thus into what we refer to as the Wyoming Algorithm, and 2) reconciling R11, and possibly what YES is doing in their real-time processor, with the Wyoming Algorithm. In our opinion, the detail provided is necessary for meeting these objectives.

Results presented in Figure 11 (rain) and in Figure 12b (snow) address your second point.

How much accuracy is gained by including radiation (%)?

The effect of radiation is addressed L476 to L484.

5) I also tried to calculate the Nusselt number but never got a satisfactory comparison to a flat plat. I assumed this was due to the addition of the ridges. As a result I left this out of the original paper.

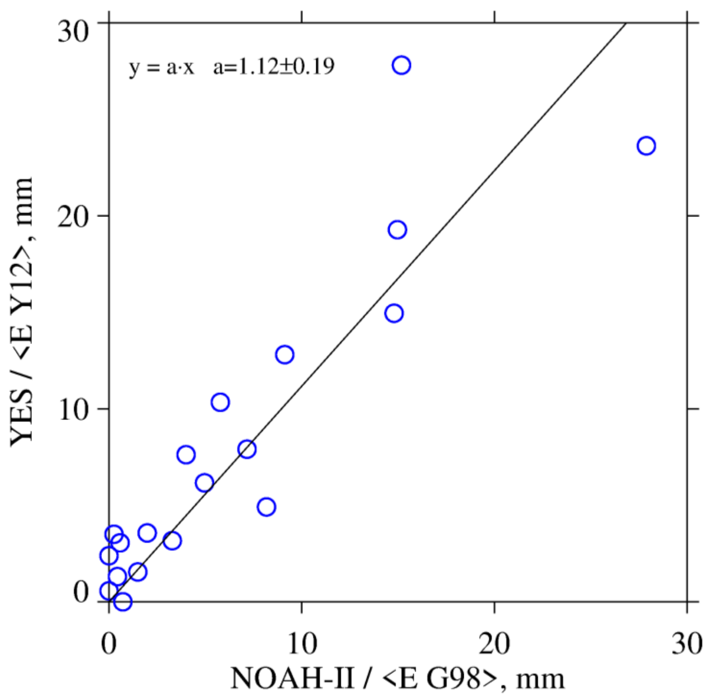
This is discussed in response to your point #3. We feel that the approach we have developed is useful because it allows incorporation of the sensible energy term into power budget (Equation 3, manuscript).

6) One of the major findings in running a hotplate that I found is that its performance depends on whether it is outdoors or indoors. Thus, the outdoor turbulence, in my experience, makes the hotplate perform a lot different than in a wind tunnel. In the wind tunnel, the data are very clean and everything works as you would expect. Outdoors, the turbulence on the bottom plate and upper plates impact the cooling differently, causing me to take a 5 minute average before initiating accumulation.

We have not used a wind tunnel. Just to be clear, our indoor testing was in the hangar or in the laboratory. There was ventilation, but the observed wind speeds were below the detection threshold of the hotplate (0.1 m/s).

7) The final comparison to outdoor data (pages21-23) is confusing to me (and I expect an independent reader). The SPICE evaluation of the YES hotplate suggests that it is 10% high, yet you find it 10% or more low. I am not sure if this is due to using the wrong level for the wind or a different wind correction algorithm. If YES has a different catch efficiency algorithm than I do I don't know where they got it from as they did not do any outdoor testing in comparison to a truth gauge. I think it might be useful to have a discussion on these data next time you are in Boulder or I am in Laramie. We could also do a conference call, but it might take some time to figure this out.

I presume, what you are comparing, in SPICE, are the values produced by the internal YES algorithm. In the manuscript, these are referred to as hotplate-derived accumulations. The figure here shows the 19 analyzed OWL snow events with hotplate-derived accumulations (ordinate) vs NOAH-II values (abscissa). The collection efficiencies applied here are event averages and these were divided into accumulations that were not corrected for inefficient catch. Data presented in the figure are available in our Table 5. The slope of the fit line is 1.1, consistent with the "10% high" you mention, but there is large statistical uncertainty due to the relatively small number of points and scatter in the measurements.



8) Conclusions: I think you can state that the area is incorrect, but, again, the factor was not stated as a constant but a calibration factor. I think this point is over-stated and was already made by Boudala et al. (2014).

Please see our related comments above. We feel our repeating of R11's distinction between a theoretical and an actual energy conversion factor is appropriate.

9) I would state the main conclusions in the final section. What I want to hear about is:

a) Is a radiation correction important to the hotplate?

This is addressed in the body of the manuscript. We agree that the information should be restated in the conclusions.

b) How well does the Yankee hotplate do compared to field observations? The wind speed in the Yankee hotplate is from the hotplate itself, so it self consistent. The fact that the unit has performed well for me and SPICE over the past 5 years suggests that the algorithm if fundamentally sound.

At OWL, the YES-derived wind speeds are biased low relative to the anemometer-derived wind speeds. This is discussed L491 to L516. We speculate that the hotplate-derived accumulations in Figure 12a are statistically smaller than NOAH-II values because 1) the Y12 catch efficiency function does not apply a height-adjusted U, or 2) because of the hotplate underestimate of U, compared to the anemometer. We conclude that untangling these two possibilities is beyond the scope of the study.

c) Any suggestions on how to improve the catch efficiency? How variable is it from storm to storm?

Event-averaged catch efficiency values are in Table 5.

I think a histogram plot of the delta precipitation as compared to a truth gauge as I did in my 2011 paper would be very useful. Are the differences biased one direction or another? Do we need to take into account snow type, for instance.

There are 19 analyzed OWL snow events in our data set. In our view, the best way to present these is in a scatterplot. This is done in Figure 12a – c. Meaningful consideration of the additional effects (e.g., wind direction) will require more data.

Are there are observations of snow type? The largest discrepancy I saw was for graupel. Was there graupel particles observed during OWLS?

In our data set there are three OWL snow events (out of 19) with either riming or graupel mentioned in the notes taken by Leah Campbell and Jim Steenburgh. In two of these (OWL-05 and OWL-20; Table 5) the UW hotplate accumulation (not corrected for inefficient catch) divided by <E R11>, is larger that the NOAH-II accumulation (not corrected for inefficient catch) divided by <E G98>. This can be verified using values in our Table 5.