Review of Bevington et al. "Satellite telemetry of surface ablation observations to inform spatial melt modelling, Place Glacier, British Columbia, Canada", submitted to the Cryosphere.

The submitted manuscript presents a novel setup of automated 'smart stakes' that transmit temperature, humidity and ultrasonic surface height data via Iridium satellite telemetry from the Place Glacier in Canada. The manuscript presents results from a summer campaign during 2024 finding a range of melt factors for snow and ice, which is variable based upon the unique temperature records stemming from each of the four smart stakes. The study finds that such networks can be highly valuable to improve attribution of melt to specific warm events with a near-realtime transmission of data. The authors provide a very detailed and thorough description of the logistical and technical setup of the smart stakes which is well written and beneficial to the community and to other scientists looking for ideas/inspiration for similar monitoring on glaciers. The study is generally interesting and the structure, writing and figures are very well put together. The authors should also be commended for the hard work that has gone into establishing and maintaining such a nice network. While the quality of the work is mostly good, I am left questioning what is learned from some of the scientific findings in several places and I think that some of the data could and should be much better leveraged to inform a more detailed model. I believe that my comments compose a major revision to the manuscript before it should be published in the journal.

Major Comments

- The authors describe their framework as "comprehensive for understanding glacier melt dynamics", but their modelling approach is incredibly simplified, to the author's own admission. What is unclear is why the authors have not attempted to include a more detailed modelling approach based upon energy balance (if such data are available from the Ridge or Forefield AWS is not made clear), or even an enhanced temperature index model (e.g. Pellicciotti et al., 2005) that includes solar radiation (again data dependent from the off-glacier data). At the very least, the exclusion of accumulation (L323) when some level of height change is interpretable from the sonic rangers (with uncertainties) is perplexing. Even for a summer balance, such storm events, which can be critical for the surface albedo and energy budget, are important to include. Without that, I feel that the overall findings from the model are not overly informative or valuable, which does a disservice to the incredible data collection made by these 'smart stakes'. This is especially the case as the authors use high spatio-temporal data about snow cover from PlanetScope satellites. I believe the authors should test some form of model that accounts for summer accumulation and ideally at least shortwave radiation. Even if the modelling framework cannot be completely revised due to data constraints, some testing of this with nearby station or reanalysis data (for example) should be used to identify shortcomings of the model in the context of this new sensor network and build a more robust discussion surrounding its value to this "understanding of melt dynamics", when such dynamics are informed by a single DDF per site.

- The authors combine a lot of information from both their smart stake network and satellite (PlanetScope) and airborne (LiDAR) information. Some of this information is used to support or understand the absolute elevation changes seen from the sonic rangers over the summer. However, the authors do not provide much perspective on such a setup where no detailed and contemporaneous airborne LiDAR surveys are available (which is almost always the case in glaciological research). For example, can the authors simply report height change relative to the start of the stake setup, without absolute numbers constrained by the LiDAR? The use of dGPS is also introduced here, but not discussed or presented any further. Is the data from those dGPS surveys sufficient without the LiDAR information to gain the same understanding from the telemetered information? This needs some more discussion and mention when presenting this as a more generalisable approach (which one feels is trying to be generalised when reading the very nice set up details for the Arduino etc). Also related to this discussion is the comparison to the 'smart stake' set up of Cremona/Landmann et al, which are cited. How and why is this approach superior or worth the extra effort/challenges/cost?

- I am left a little confused about the implementation of the lapse rate to distribute the air temperatures across the glacier, as the details are a little vague and the stations utilised include an airport station some distance away and a mix of on- and off-glacier weather stations/smart stakes. The authors state (L418) ". This spatial variability in temperature regimes underscores the importance of distributed temperature measurements across glaciers for accurate melt modeling" referring to the differences of temperature off- and on-glacier, but provides an unclear reasoning for this with respect to modelling, given the fact that they do not explore the relevance of using point-measured temperatures (smart stakes) vs. interpolation / lapsed temperatures and its impact on model accuracy.

  Therein lies another main issue with the manuscript for me: Do the observed air temperature records at 4 stakes significantly benefit our overall modelling ability (given the issues with model, as above) compared to, say, 3 stakes, 2 or even 1, or only the off-glacier AWS? Would other studies likely benefit from just one station in the ablation zone and one in the accumulation zone, halving their costs? The authors need to i) establish a clear method and rationale for their lapse rate choice (just an hourly variable, linear lapse rate with all off- and on-glacier stations - Fig. S3?) - but why also the airport station included?, and ii) leverage this interesting and novel information to discuss more the value it brings to understand physical processes (e.g. boundary layer temperatures - Ayala et al., 2015 which includes Place Glacier) and melt dynamics and improve modelling.

- The value of the independent validation stakes (the manual stake measurements - Fig. 10) is not made overly clear and raises questions about the model. My understanding from the text and figure is that the period of manual measurements is longer (by a month) than the model period (limited due to the smart stake temperature operation). However, the information from Fig 10 shows that the model is still largely over-estimating the summer mass balances (is more negative). It is not so clear why that is (I suspect it is an albedo issue on the lower glacier which is not considered by the degree day modelling), but it does not build confidence in the value provided by the smart stakes if one wants to understand the mass balance of the whole glacier. For example, is it *just* the model, or also the distribution of temperatures? Given the reasonable performance of the model at the stakes themselves (Fig.8), it is possibly the latter. But how sensitive is the model to that choice of data usage and does it undermine the value of the stakes if we still cannot confidently interpolate/distribute the temperature data between them, even on a small glacier? The authors need to build a clearer argument for this and provide more discussion to support how much is model uncertainty and how much is forcing uncertainty. I appreciate that this is not necessarily a study about modelling, so an exhaustive investigation of this is not necessary, but reasonable discussion and reflection on this for the reader is certainly needed.

Specific Comments

L69: Please revise the second objective of the manuscript here to be more precise. How exactly are they being compared to the PlanetScope imagery for snow cover and how does that help the overall message of the work?

L94/95: Can the authors provide the final date ranges here that they use in the model study?

L98-101: The authors introduce these weather stations, but it's not made clear if the forefield and ridge stations are permanent, what they measure and with what sensors/intervals/uncertainties. The authors should briefly report this information and particularly in light of the information that they could provide to a more detailed model (major comments above).

L127: The reported wake up times are naturally very short, due to battery life considerations, as is well reported by the authors. The equilibrium response times of the sensor according to the manufacturer are up to 30 seconds for the 63% range, but did the authors make any tests of these values compared to a continuously logging temperature sensor at the other weather stations, for example? Are the sensors also comparing well when placed together? Some short report of this would be beneficial.

L109-175: This section is very detailed and informative. The authors have done a great job to provide all of the necessary logistical and technical considerations. The authors should, however, provide a short table to summarize some of the information about chips/boards and instruments used for their final setup.

L153: Can the authors show some comparison of these temperature/RH observations compared to the AWS sensors or another reference? Are the accuracy records only taken from the manufacturer? What about a comparison between smart stake sensors before installation?

L196-198: As mentioned in my major comment above, it is unclear why exactly it is necessary to provide the absolute height changes via comparison with the airborne LiDAR, especially if the sites were dGPS located at the time of installation and stake re-setting (L215). Perhaps I have missed something clear here, but the authors should state their reasoning for this, and, as I mention above, provide discussion for the cases where this data is not available.

L218-226: Did the authors test other available datasets that did not require a user licence (even if researcher access is available for Planet when applying)? What about using Landsat and deriving albedo (e.g. Naegeli et al., 2019) to aid the development of an enhanced temperature index model? The authors run a distributed model using a 5 m LiDAR and 3m snow cover map, but it is unclear how much additional benefit that very high resolution has.

L222: How many of the 73 scenes remain after removal of those with haze/partial coverage etc?

Section 2.4.1: As per the major comment, please consider revising the modelling strategy or adapting it to test against cases with a more detailed model.

L241: if spring densities are not considered, why mention them here?

L269: The authors claim that accumulations are false when the air temperature is above 0°C, but snowfalls are possible at temperatures a few degrees above zero (e.g. Jennings et al., 2018), particularly in humid environments, like that of the study site. The authors should elaborate on their confidence of these false values and at what temperatures they are typically occurring.

L306-309: This short section requires some additional information. As per my major comment above, it is not clear how the lapse rates are applied in the glacier-wide model (only monthly means as in Fig. 6 or hourly variable?) and why the authors also include the airport station in this calculation. What is the goal of using those stations and in such a way? Do the authors wish only to make use of off-glacier data (that is typical in mountain regions) to see the value of smart stakes equipped with T/RH sensors? Do they only want to use on-glacier data to say that air temperature relates to melting of the glacier? How often is the lapse rate non-linear? How well does 1-3 stations represent the air temperature variability of another station in a leave-one-out analysis? Does it matter to have 4 stations? What is meant by a “near-normal” lapse rate (L309)? Please expand this section to explain the approach better and justify it within the context of conducting glaciological research with these very nice smart stakes.

L306: In this section it should also be clarified if the authors make any adjustments given the variable height of the temperature sensors above the surface. As the stakes melt out, the sensors will increasingly become independent of the boundary layer temperature variability due to the density driven katabatic winds (under warm weather conditions - Ayala et al., 2015). They will also become less affected by sources of uncertainty due to high albedo and heating errors of the radiation shield. These factors should also be considered when evaluating sub-period variability in calculated snow and ice melt factors.

L314: Are these melt factors useful when comparing the smart stake stations? Station 4’s ice melt factor is clearly much higher due to the exposure of ice during a warmer period of observation. How high are the melt factors for the other stations if considering the same period as the snow-free period at station 4? The authors mention the variable nature of these factors, which is well established, but by how much are the differences?

L316: The R2 of what? The authors refer to Fig. 7 again here? Please say it explicitly.

L322-323: It is unclear how the ‘accumulation’ is added into the model results and what this approach really tells us about the response of the glacier to warming (the melt dynamics that the authors mention). So the glacier is still melting, but the height of the surface is superimposed on that, and the melt factor (if seen by Planet images) is set to snow? Again, I find this a key limitation that does a disservice to the great work in creating and maintaining this network.

L337: Define ‘stack’ in this context.

Section 4.7: Please see my major comment regarding the mis-match of the validation stakes and the reasoning/discussion about this.

L371: Can the authors specify what % if the total observation period these melt % occur under? This might help to give a more clear context.

L387-389: Can the authors elaborate on what is learned from the use of these smart stakes compared to normal stake measurements, or smart stakes based upon cameras (i.e. Landmann et al. 2021)? Again, given the over-simplified model approach and my concerns about the mismatch of the model (Fig. 10), the true value of the author's setup is not so clear.

L406: What is a less pronounced lapse rate? Do the authors refer to a shallower lapse rate (where the rate of change in temperature with elevation is less?). If so, state this clearly and provide an average value for context. It is still not clear from my reading what causes the melt rates to be so high for S4. Is this because of an ice melt factor that is derived from a short period of ice exposure and warm temperatures?

L416-419: Given that Place Glacier was used in the creation/testing of two models of air temperature distribution (Shea and Moore, 2010; Ayala et al., 2015), it is surprising that there was no comparison of this in the manuscript or discussion of these approaches using the airport or ridge/forefield stations as forcing. I understand that it is not the main aim of the work to look at temperature distribution, but, as mentioned before, the main value of the observed temperatures, linked to these such statements, are not so clearly demonstrated.

L421: Why did the authors not test the variable melt factors? My above point again related to the very high melt factor for S4.

L423-424: As mentioned, the authors should also highlight the value of their smart stakes for cases where airborne LiDAR data are not available, and where visible imagery might be increasingly limited by cloud during the melt season (e.g. parts of the Himalaya / S-E Tibet).

L442: The authors made dGPS measurements when setting up and resetting the SmartStakes. Is there a reason that ice velocity measurements are still unavailable? Perhaps I have missed something here.

Figures / Tables

Table 1: The authors should add what data are measured at each station

Fig. 3: Please add the station numbers to these plots, just for clarity.

Fig. 4: Change the legend to read 'start of ice *exposure*'.

Fig. 6: I think that this figure would benefit more from excluding the ECCC (maybe the current version could stay in the SI?), and zooming into the glacier area. It would be more valuable to see what the lapse rates are just by fitting to the on-glacier smart stakes and also perhaps when using only the two off-glacier stations (assuming no on-glacier data).

Fig. 7: Specify in the caption whether these are derived from the model or from the observations.

Fig. 9: I would invert the colour scales in all subplots, as the PDD, snow and melt are more intuitive if the oranges and yellows represent larger melt values/temperatures.

Fig. 10: Please add a colour to the circle to show the mean snow duration or something that might help to interpret the reason for the mis-match and its cause (see major comment).

Fig. 11: I think it would be clearer to indicate each site with a different colour, rather than the temperatures, as this mostly replicates the y-axis melt ranges.

Cited Works

Ayala, A., Pellicciotti, F., & Shea, J. (2015). Modeling 2m air temperatures over mountain glaciers: Exploring the influence of katabatic cooling and external warming. *Journal of Geophysical Research: Atmospheres*, *120*, 1–19. https://doi.org/10.1002/2015JD023137.Received

Cremona, A., Huss, M., Landmann, J. M., Borner, J., & Farinotti, D. (2023). European heat waves 2022: contribution to extreme glacier melt in Switzerland inferred from automated ablation readings. *Cryosphere*, *17*(5), 1895–1912. https://doi.org/10.5194/tc-17-1895-2023

Jennings, K. S., Winchell, T. S., Livneh, B., & Molotch, N. P. (2018). Spatial variation of the rain-snow temperature threshold across the Northern Hemisphere. *Nature Communications*, *9*(1), 1–9. https://doi.org/10.1038/s41467-018-03629-7

Landmann, J. M., Künsch, H. R., Huss, M., Ogier, C., Kalisch, M., and Farinotti, D.: Assimilating near-real-time mass balance stake readings into a model ensemble using a particle filter, The Cryosphere, 15, 5017–5040, https://doi.org/10.5194/tc-15-5017-2021, 2021.

Naegeli, K., Huss, M., & Hoelzle, M. (2019). Change detection of bare-ice albedo in the Swiss Alps. *The Cryosphere*, *13*, 397–412. https://doi.org/https://doi.org/10.5194/tc-13-397-2019

Pellicciotti, Francesca., Brock, Ben. W., Strasser, Ulrich., Burlando, Paolo., Funk, Martin., & Corripio, Javier. G. (2005). An enhanced temperature-index glacier melt model including the shortwave radiation balance : development and testing for Haut Glacier d ' Arolla , Switzerland. *Journal of Glaciology*, *51*(175), 573–587.

Shea, J. M., & Moore, R. D. (2010). Prediction of spatially distributed regional-scale fields of air temperature and vapor pressure over mountain glaciers. *Journal of Geophysical Research*, *115*(D23), D23107. https://doi.org/10.1029/2010JD014351