Summary:

A data-paper serves usually a dual purpose and is somewhere between a technical report describing the measurements/ dataset and a research paper presenting detailed analyses. The manuscript addresses both, and describes datasets from Werenskioldbreen, a glacier in the Arctic archipelago of Svalbard, that cover both glacier mass balance measurements and operation of an automated weather station. These data have been made available and the manuscript describes the meteorological and glaciological measurements, presents some characteristics of each dataset and puts them in relation to each other.

I applaud the authors for collecting and publishing these datasets from a generally data-sparse environment such as the Arctic and I will recommend publication of the manuscript once some of the identified shortcomings have been rectified.

- Description and presentation of data: although nominal sensor uncertainties are stated in Tab 1, I miss a more thorough discussion of data quality, clearly stated quality thresholds used for discarding data, procedures for gap filling etc. The problem of varying sensor height above the surface is not at all addressed. Not much is learned from e.g. Fig 2 that presents 7 years of data in hourly resolution. Instead, values could be presented aggregated in monthly boxes to enhance figure readability (for instance Schuler et al. 2014). Adding supporting information about data completeness/quality would give the reader a quicker overview than a listing of all gaps in the main text. Some figures present uncertainties, but it is never explained what these uncertainties represent (standard deviation, RMS error or other measures?).
- Datasets in the zenodo repository come in plain CSV format without explanatory readme-file or other metadata. Even with the manuscript at hand, it needs some intuition to identify the individual variables and which units the values are affiliated
with. This must clearly improved to make the dataset useful. At the other repository https://ppdb.us.edu.pl, the meteorological data are available in netCDF format and contain all required metadata. This should be the case also for the data on zenodo. Especially given the difficulties finding the data at PPDB, that the other reviewer already commented on.

- Throughout the manuscript, the authors repeat statements about the significance of the dataset, but after all these statements appear more persuasive rather than convincing and would benefit from better support (or more modest wording).
- Related to the point above, the analyses to show the significance of the data should be improved; 10 years too short for climate trend analysis. Instead the data could be discussed in light of existing longterm series from Hornsund. Do trends agree? How do differences vary (seasonally, decade)? Such a discussion would provide good arguments for the necessity of these measurements, why should we need additional series if they are only linearly related to other operational series?

Details:

L12: unique...in what sense?

L15: 10 minutes, hourly and daily resolution

L15: snow data is not from 49 individual sampling locations but the dataset consists of 49 datapoints, reword the sentence.

L16: spell out SWE (undefined abbreviation).

L16: since you refer to Cogley et al 2011, you should stick with its terminology: ...consists of seasonal and annual surface mass balance measurements... (“surface annual winter, summer and net balance” is at least confusing if not inappropriate).

L18: this is an example for the unsupported, persuasive wording mentioned above: “high-quality and long-term datasets”. the quality is not well assessed in the manuscripts and 10 years is not really long-term.

L18: ...to serve as local forcing data...

L24: “...and significantly expand our knowledge” not per se, but they may contribute to that. Such a bold statement would require a reference.
L26: “...does not adequately address...” why not? needs a reference

L27: Both, atmosphere and ocean... (both are parts of the climate system).


L31: define JJA

L33: ...the highest air temperature since beginning of the measurements... (add a reference)

L37: minimum extent in the history...(add reference)

L40: add a reference

L40/41: what disparities? I do not understand.

L43: must be...developed in Svalbard. Why there?

L47: ...located in South Spitsbergen (Fig. 1)...

L49 ...covered ...in 2008...

L51: what is "the internal marginal zone"?

L53: ...to be treated as a well-defined ...
L55: please mark the location of the Baranowski station on the map (Fig 1)?

L66: “...placed in the ice (drilled ca 6 m deep)” why not using a floating station in the ablation area?

L80: between the glacier surface and sensor level

L80: ...were systematically replaced... how often? Every year? More often? Or only if needed?

Tab 1: I understand that sensor level above surface will continuously vary, but at least you could indicate original height or the target heights at which you wish the sensors to be...

L83: is this the measurement interval or the recoding interval?

L87: The locations have been chosen to cover the elevation range... (combine with next sentence)

L90: did you really use an ice core drill or the Kovacs stake auger? (same brand, different tool)

L92: the properties of the snowpack

L101: ...renovated during maintenance visits.

L102: ...pandemic travel restrictions inhibiting access to field area. (please state the years when this happened)

Tab2: please identify coordinates with Northing and Easting. The coordinates of stake #9 apparently have been swapped.
What is the value of labeling GPS and SR50 measurements if these are not presented?

L110: ...annual and seasonal point mass balance. (you do not measure ablation or accumulation but mass balance!)

L111: what you refer to as “surface mass balance” seems to correspond to the “glacier wide mass balance”. The first considers only mass changes at the surface of the glacier, the second is the integral over the surface area of the glacier, see Cogley 2011.

L155ff: which DEM did you use for which period? How did you deal with transitions between them?

L120: travel restrictions started in Mar 2020, so they cannot be blamed for data gaps in 2017-2020...

Fig2: how much are the values influenced by changing sensor levels as the surface melts down/ sensors are releveled? Add some information about completeness of the records (e.g. percentage of time). The shortwave radiation records could be analyzed in terms of albedo and its seasonal changes; adding potential clear sky radiation would give an impression about cloud effects.

For a melting snow/ glacier surface, the surface temperature is capped at the melting point and so is the outgoing longwave radiation (316 W m-2). Discuss why you observe upwelling radiation in excess of this value!

L130 ...the difference ...between...and...

L131: how does the temperature lapse rate vary with time?

L131: deriving a temperature trend from a 7 years series is questionable, especially without further remarks on statistical significance. How does this compare to the long-term series from nearby Hornsund?

L133/134: how much is the trend in summer influenced by the increasing distance between the sensor and the cooling surface as the snow/ ice melts down?
In 2009-2020, the AWS measured... this is not continuously measured at all: rel humidity was recorded in 2011 and then again 2016-2019, wind direction only since 2016 (according to information given in Tab1).

If the sensors are not robust enough for polar conditions why did you not use others that are better suited?

...are of great value for solving specific scientific problems. Could you give some examples to support this claim?

No, you cannot measure these by only 1-2 visits, you do measure the balance!

...on the physical properties...

Better description of error estimation is required.

Fig 3: y-label should be "Water equivalent"

Why are the stake values not sorted according to stake number? This is a confusing sequence...

The observation that the variability of annual balances is dominated by the variability of summer balances is not new, add references. This is actually much better seen in Fig 6.

...was generally low.

...for calculations of the glacier-wide surface mass balance.

...was then averaged weighted by layer thickness.

Mention more about the method of Sturm (2010)
Tab3: add the SWE values as well.

How did the snow density vary in space/ with elevation?

Fig 5: explain which error measure is indicated by the whiskers.

Present the coefficients of the regressions shown in both plots. The slope of the line in Fig 5a would correspond to the bulk density. How much of the variability in Fig5b is due to variability in snow thickness versus variations in density? Better to present the measurements instead of the inferred quantity.

L199: $R^2 = 0.62$ is not “very high correlation”!!

L206: as mentioned above, your “surface mass balance” corresponds to “glacier-wide mass balance” (Cogley 2011)

Fig 6: typo in label on y-axis: Surface mass balance

State the trends and add a measure for their significance.

Comment on the records winter balance in 2014. Has this also been observed on other glaciers?

L 216 fff: again: summer and winter mass balances

L222: summer balance decreases by -0.14 m (negative value marks a decrease!!)

L222: …acceleration of mass loss.

L226: …according to the model...
L228: the summer balance decreased from -1.23...to -1.68

L230: observations (1999-2019)...this is confusing, you present data for 2009-2019

In the following lines you fuse the model results and your measurements to infer a long-term evolution. This is questionable as long as you do not show the level of agreement between model and observations. However, this is not possible to derive since the two series do not overlap in time (model 1959-2002, measurements 2009-2019)

L247: better explain how the error estimate has been derived

L248: better explain the gap filling: has the Hornsund data been adjusted using a regression or similar?

Fig 7, caption is misleading, the figure shows cumulative ablation, not daily ablation.

What value for DDF has been used? How has it been determined/ selected?

How are eqs 2 and 3 used in the calculations? For the entire glacier? Or per stake location (and according adjustment of temperature records)?

L254: ....determines meltwater runoff volumes.

L265: ...could not be maintained with high frequency.

L268: Similarly, malfunctioning of sensors...

L272: These data spikes...

L273: ...such spikes are believed to be artefacts.
L274: ...exceeding a permitted variability... (state the value of this important threshold!)

L276: ...compared with records from other weather stations...

Could you show or discuss the results of these comparisons? Show the agreement or at least state a quality measure.

are replaced/ corrected values in the published datasets marked as such? (distinct from valid measurements)

L284-306 do not belong to a quality-control section but rather to a separate section “dataset structure”

L307: the first sentence can be removed without losses.

L310-317: consider moving this to the “methods” section

L318: what are the results of this comparison?

L321-326: consider moving this to the “methods” section

L348: “high-quality and long-term” ...see comments above.

References:
