Comment on essd-2022-63
Anonymous Referee #2


Review for Hu et al. (2022)

Summary

The authors have produced a daily gridded snow-depth dataset for the northern hemisphere for the period 1980 to 2019 using machine learning, specifically using a random forest method. The dataset incorporates remote sensing data from multiple products derived from various sensors, reanalysis data, and in situ measurements. Different combinations of datasets are chosen for different periods based on data availability. Additional datasets such as land surface type and topographic information are incorporated into the scheme as additional input variables to improve the estimated depth values. The authors indicate that the scheme improves snow depth estimates substantially relative to the best available current products using various metrics including the coefficient of determination, root mean squared error, and mean absolute error. The fused dataset is less accurate at high elevations, and would benefit in terms of accuracy from further validation and input datasets.

General comments
The dataset appears to be well thought out and such a dataset is indeed important and potentially useful for many applications from climate studies to water resource management. However I feel the manuscript requires major changes with regard to presentation as well as further evidence and descriptions to support the authors’ claims. I do not feel it is suitable for final publication in the present form. In general:

- The methods used to produce the dataset are not adequately discussed. It appears that the authors have discussed some of the methods and their reasons for choices e.g. which machine learning algorithm to use, in previous studies, but these decisions and choices should also be summarized in this description paper for the benefit of users of this dataset.
- I am concerned that the in situ locations used for validation are the same locations as the training data, and therefore the method may artificially appear to be successful. I believe the authors mention performing some sensitivity experiments at the end of the paper but I think a section should be devoted to exploring some of the choices made and their impact on the dataset. Additionally is any uncertainty information for the training in situ datasets incorporated into the analysis? How might this affect the results?
- It would be greatly beneficial to users of the dataset to have an estimate of uncertainty for all points and times in the final dataset, or at least to provide some flags for low/high/medium quality data, derived from the data used here. This should be provided if possible.
- The discussion section is very brief and now more or less summarizes the paper. The purpose of the discussion section should be: e.g. discussing limitations of the methods and data used. I believe the discussion section needs more elaboration on this.
- I’m missing important information on the timing of the snow depth values in the different snow depth products and in situ observations. Only for the meteorological station observations from China you mention the measurement time of 8am, but not for the other products and observations. This potential mismatch in timing may induce a larger bias than is actually the case. Please elaborate on these timings or, if the timings are not available, elaborate on the potential effect of different timings between each product/observation.
- The writing in the manuscript needs considerable improvement. Spelling and grammar should be reviewed by a fluent English speaker. Sometimes paragraphs are too short and should be combined. In particular I suggest using present rather than past tense in most cases which would improve readability. The text is sometimes redundant and should be revised to avoid repetition of information.
- Please add citations for the remote sensing and in situ datasets in the data description section.

Specific comments
L16: The abstract is a bit long and could be shortened somewhat.

L16: Please briefly elaborate on this statement. Why is it important for these disciplines?

L25-26: I suggest starting a new sentence here, e.g. “and topographic data. Here we incorporated these datasets as independent input variables to a random forest regressor to generate a gridded northern hemisphere snow depth dataset for the 1980 to 2019 period.”

L32: Change “was distributed…” to “was in the range of -5 to 5 cm.”

L39-42: Does this sentence belong here or should it be in the data availability section?

L54: What are manual observations? Do you mean in situ observations?

L58: Can you quantify the “highest elevations”? 
L68: Please briefly explain what it means when a snow depth data set saturates.

L69-70: Please elaborate on the structural limitations.

L74: Suggest changing to read ”Additionally, some reanalysis datasets...” unless you mean that some reanalysis datasets overestimate snow depth at high latitudes.

L105: Please elaborate on ”combined and integrated improvements”.

L107-108: Why is this a new approach? It is more precise compared to what? What is the auxiliary information here?

L109-110: What does it mean that the fusion is improved? Do you mean that errors were reduced?

L110-111: Please note which ”machine learning methods” were used.
L112: Please quantify the increased estimation accuracy if an estimate was provided.

L114: How do you know that the positive aspects of each product are incorporated into the fused data set?

L115-116: Please describe what the candidate independent variables are and why they are candidates.

L117: Please elaborate on the different bins.

L118: The data fusion framework has already been proposed in Hu et al. 2021. Please clarify that this paper isn’t proposing the framework but is presenting the dataset and validation by comparing it with observational snow depth data.

L123: Please elaborate on what validation works are.
L143-144: Can you elaborate on the calibration?

L144-145: From this sentence it seems that the snow depth can only have a depth of 5 cm. Is that what you mean to say? Is this a minimum snow depth value?

L145-146: Please elaborate briefly on the spatio-temporal interpolation.

L153-154: Is this less than a 30% deviation with respect to observations? Please clarify.

L155-157: Please provide a citation for this statement.

L163: In L67 you say this product excludes the area above 35N. Please check and adjust accordingly.

L164-165: Precision and accuracy are two different concepts. Do you mean you use both of them?
L166-167: If the attempt was successful, please elaborate. If not, this sentence is not necessary.

L172: Please elaborate on how you get daily data from the 6-hourly data. Do you take the mean?

L174: Do you mean in the process of “making” the MERRA-2 data set? If so, please rewrite to make that clear.

L175: Please elaborate on improving the quality of the data.

L178: Is this nearest neighbor interpolation? Please clarify.

L181-182: As noted in the general comments, please provide citations for these datasets.

L182: What is the spatial distribution of the GHCN data set? Is it distributed sufficiently across the NH to be able to draw conclusions about regions outside of China and Russia?
L187-188: Please explain the meaning of “per five days data.”

L189: Please elaborate on the rigorous data standards.

L191-192: What is a quality checked field? Please elaborate.

L192: What was the method of removing the anomalous snow depth fields? Is this the quality checking procedure?

L197-198: Please elaborate on the inter-annual consistency and climatological outlier check.

L198: The amount of station sites used in the two Russian data sets are missing.

L200: Are the seven data sets mentioned here different from the four data sets described in section 2.2? If not, please elaborate.
L200: How/why did you choose these data sets? Could sites not used in the training also be chosen for this purpose?

L202-203: It might be good to mention the specific years that are covered by the other data sets as well.

L203-204: What is meant by “snow depth retrieved model” are these simulations of snow in earth system models? Please clarify.

L207: What else does the auxiliary data include?

L211-212: Can you justify this assumption? How might this impact the results?

L212-213: Please elaborate on what you mean with “snow depth data as a whole”.

L215: Which dataset is being referred to here, GMTED2010 and/or GTOPO30? Please elaborate.

L220-221: I believe RFR is the abbreviation for random forest fusion framework. Please clarify. Although this is discussed in another study, it would be helpful to include details as to why the RFR method showed the best performance. Also a brief description of each of these methods should be provided.

L232-233: Please elaborate on what is meant by “different models were established”, also what is meant by “15 models can be employed to train and verify the model.”

L236-237: Additional details are needed here. Why is the random forest model the best?

L244: Please elaborate on the “leave-one-year-out” cross-validation.

L259-260: This is confusing. Suggest revising to read: “As noted above the fused dataset provides continuous daily data from 1980 to 2019, with several gaps.” Then the gaps can be mentioned. It is not clear whether the gaps occur every year or whether only certain years have gaps.
L264-265: Please elaborate on this.

L266-267: Briefly explain why these areas are excluded.

L271: Please elaborate on why the NHSD and GlobSnow inevitably have a large amount of data missing.

L272: Why do data gaps in 2 of the 7 snow depth products lead to data gaps in the fused data set? Shouldn't the other datasets be able to fill the gap? Please elaborate.

L275-276: Do the data gaps arise because of the striping you mention in L274-275? Please elaborate and clarify.

L276-277: This seems an important limitation of the machine learning fused framework. Please elaborate on why this happens and what it means for your results.

L299-300: The in situ observations are at the point-scale, while the fused data set is at 0.25 deg resolution. Can you comment on errors introduced from this comparison?
L302-306: Where do you get these conclusions from? If from figure A2, please refer to that figure.

L304: What does “its overestimation and underestimation were obvious” mean? Please elaborate.

L305-306: What does “and there were many points of underestimating and overestimating disorderly distribution” mean? Please rewrite.

L309: Suggest changing “BIAS” to “bias” throughout. The statement here is unclear. Suggest revising to: “The fused data bias fell mostly between -5 and +5 cm, with 88.31% of the bias falling within that range.”

L317-318: What does “percentage of each interval” mean? The percentage of the total amount of data?

L320-323: In section 4.2 you say you use 90% of the in situ observations for model training while you retain the other 10% for model verification. I believe the 10% of measurements are taken from the same locations as the other 90% while these are separate locations. But this is unclear. How do the locations here relate to the other in situ locations mentioned earlier? Would it be possible to also exclude some of those locations to improve the analysis?
L324: Please briefly elaborate on which regions you mean.

L325: I suggest extending the analysis shown in Table 1 to also be performed on the other snow depth datasets. This will reveal the success of the various methods assessed against the independent in situ measurements. As it stands the analysis only describes the strengths and weaknesses of the fused dataset without showing its performance against other datasets.

L329: It would be good to mention the countries or regions these sites are in.

L329-330: Not necessary to explain abbreviations of R2, RMSE, and MAE, you already did this.

L330: Does this mean it is impossible to calculate the R2? I'm not sure why a large error would impede you from calculating the R2.

L330: Suggest changing the column descriptions in Table 2 from “RMSE / cm” to ‘RMSE [cm]”. Same for the other column descriptions. It now reads as “RMSE per cm”.
L332-333: “... their accuracies were still relatively high compared to those of other gridded snow depth products”. Please elaborate on which snow products you mean.

L333-334: Not sure what this means, please rewrite. I also do not see any inflection points, which is where the direction of the curvature changes. The curves in Figure 4 are all in the same direction.

L338-339: Here it appears that a comparison is made with the performance of the original gridded datasets at this site. However, the data is not provided. As noted above it would be best to also include that analysis, perhaps as a set of tables in the appendix.

L340: Please elaborate briefly on what a relative low elevation is.

L340: Please elaborate briefly on the “better performance”. The performance is better than what?

L341: Please describe what is in the file, rather than the file type.
L344: Please describe why this site is better suited for measuring precipitation.

L346: In L342 you say that SBBSA is located in a basin. Is the basin above 3700?

L348-349: Are two sites sufficient to characterize an entire basin? How large is the basin? Please elaborate briefly. And what does 'the large area snow depth' mean? The area of the 0.25 deg pixel or of a larger regional scale?

L350: Please elaborate on what you mean with “but this site has a higher altitude”.

L352-354: How do you get to these conclusions? Please elaborate briefly. I believe that the authors are discussing changes in snow depth with elevation. There is a rapid change in depth with elevation that cannot be captured in the fused dataset. This is consistent with a larger bias for the highest snow depths. Please clarify.

L369-381: This paragraph needs rewriting. Please clarify the meaning of “relative frequency of BIAS”, “slightly overestimated trend”, and “distribution charts of relative frequency”.

L390-391: Accuracy cannot have poor precision. Data can have accuracy and precision.
Please rewrite.

L391: Use either elevation or altitude consistently throughout the manuscript.

L392: Consistency is probably not what you mean here. Please rewrite.

L393: What do you mean with “both snow depth and error of the fused dataset were greater”? Please clarify.

L394-395: Move this sentence to earlier in the paragraph when you talk about these elevation ranges.

L400: Discuss when these changes (decrease followed by increase) occur in the timeseries.

L407-408: What do you mean with this sentence? Figure 6b shows high snow depth values in the west, as well as the east. Please adjust. Also remove the word 'distribution'.

L408: From the spatial pattern in Canada? That is probably not what you mean, but this sentence does make it look like that. Please adjust.

L409: The snow depth of the Tibetan Plateau was also less than what? Please clarify.

L418: What is a "distribution area". Please clarify.

L418-419: What about Scandinavia, Svalbard, eastern Siberia, and Alaska?

L419: What do you mean with "eastern European plain"? The little area slightly east of the European Alps? That hardly seems like an important area to mention given all the other large areas with high snow depth values.

L421: What does "relatively smooth" mean? Please clarify.
L421-422: Are the authors referring to the machine learning methods with regard to dividing into seasons, and the method of validation when referring to dividing snow depth into different intervals?

L422: What does “more reasonable and precise” mean?

L430: What is a “changing trend”? Suggest replacing this with simply “trend”.

L431: What does a test value of -3.28 mean?

L431-432: What shows a significantly decreasing trend? Can you quantify this?

L433: Please quantify the trends.

L446-451: Redundant and does not belong in the discussion section.
L454: Please elaborate on “based on experience”.

L454: Add citations to “previous studies”.

L462-465: This should be in the results section.

L463: What do you mean with “different spatial positions in the training sample (same time), different times of training samples”? Please clarify.

L465-467: What do you mean with this?

L476-468: This is not a proper way to train and verify the ML model. These years may differ significantly in climate, and thus in snow depth. You need more years of training data to train the ML model.

L468-470: Please clarify what this means.
L470-471: What do you mean with this? In L465 you say that you use all the NH data because of the generalization ability of ML. Also, please cite these claims.

L471-472: Here you say again that the ML model is able to generalize. Please clarify.

L472: Please elaborate on “new training is advisable”.

L472: Not clear what you mean with eliminating “one variable”. What variables are these?

L474-475: Here you say again that the ML model cannot generalize across different spatial locations. This argument is inconsistent and needs to be revised.

L477: Add citation to "as found in previous studies".
L492: Do you mean accuracy instead of precision?

L493: If you've validated this, your conclusion cannot be “likely more accurate...”. You should be able to have a firmer conclusion. Also missing citations.

Technical comments

L18: Replace “product” with “products”.
L25: Change “incorporated” to “incorporating”.

L27: Replace “different time period” with “a different time period” or “different time periods”.

L34: Replace “under” with “for”.

L46: Replace “is measured” with “are measured”.

L48: Replace “spatial-temporal” with “spatio-temporal”.

L59: Remove “retrieved”.

L59: Replace “spatiotemporal” with “spatio-temporal”.
L69: Change “susceptive” to “susceptible”.

L73: Remove comma after “latitudes”.

L89: Change “showed” to “have exhibited”.

L92: Replace “Mudrky” with “Mudryk”.

L99: Replace “plain” with “plains” and “forest” with “forested”.

L100: Replace “satisfying” with “satisfactory”. Can you quantify this statement?

L104: Not sure the word “even” is necessary here.
L106: Remove "real".

L108: Replace "the ANN model" with "their ANN model".

L108-109: Replace "to have a lower ... than" with "to have a reduced MAE of 40% compared to an MAE of 60% of".

L112: Replace "compared with" with "compared to".

L113: Remove "products".

L124: Replace semicolon with period.

L124: Replace "summarized" with "discussed".
L161: Replace “included some in situ” with “includes a number of in situ”.

L162: Replace “mountain” with “mountainous”.

L166: Replace “mountain” with “mountainous”.

L166: Remove comma after “study”.

L169: Remove “from the fourth generation of reanalysis”.

L170: Replace “from” with “by”.

L173: Combine into one paragraph.
L193: Remove “also”.

L195: Not sure what this sentence means. Please rewrite.

L209: Remove “covers”.

L210: Remove “land” after Hemisphere.

L218: In L30 you use indexes as the plural for index. Here you use indices. Both are correct but it's best to be consistent throughout the manuscript.

L220: Replace “try fuse” with “generate fused”.

L220: Replace “datasets at” with “datasets of”.
L222-223: Replace “was referenced from” with “can be found in”.

L233-235: Replace “existing accuracy assessment” with “an existing accuracy assessment” or “existing accuracy assessments”.

L240: Change “second period include” to “second period includes”.

L245: It is 2022 now, so the data set you’re using does not cover the last 40 years. Please rewrite.

L247-248: Change to read “We evaluated the accuracy of the fused snow depth and the original gridded snow depth products against the in situ observations.”

L249: Change to “snow depth products as follows:”

L252: Make sure the variables in the text are aligned with the rest of the text. They are elevated right now.
L252: Both variables are now called $S_i$. Please change one of them and adjust accordingly.

L255: Combine this with the previous paragraph.

L255: Replace “variation trend” with “trend” and “We” with “we”.

L270-271: Replace “large data missing exist” with “large amounts of data are missing”.

L271: Remove “were”.

L272: Change “resulting in the similar missing in the” to “resulting in similar data gaps in the”

L278: Please make the projection part of the sentence more clear.
L278: Replace “spatio” with “The spatial”.

L279: You already mentioned the GeoTiff file type. This sentence can be removed.

L281-282: The first part of this sentence can be removed; you elaborate on the filename format in the next sentence.

L285: This sentence is not necessary.

L288: Remove comma after “2019”.

L289: Do you mean “machine learning model training”?

L305: Remove “as a reanalysis snow depth product”.
L308: Replace “snow depth” with “fused snow depth”.

L310-311: Remove “This also indicated that the consistency between the fused snow depth and ground station observations was very good over the entire Northern Hemisphere”.

L333: Replace “The fused snow depth can accurately estimate deeper snow” with “The fused snow depth product contains accurate estimates of deeper snow”.

L340: Replace “an” with “a”.

L344: Replace semicolon with period.

L345: Replace “shallower” with “smaller”.

L346: Replace “land cover type of this pixel” with “land cover type of this site”.
L347-348: Replace “range was varied” with “varies”

L351: Remove “During the winter ... at this altitude”.

L363: (Fig. 3) I suggest compressing this figure so that it fits on one page. This could be done by removing some of the locations and moving them to the appendix, compressing the y-axis and reducing space between figures. Please adjust the x-ticks to improve readability. Suggestion: fewer x-ticks and mention just the year, not the month/day. This will also reduce the size of each sub-figure.

L368: Remove “levels of”.

L369: Replace “levels” with “depths”.

L370: Combine paragraphs.
L371-372: The wording is strange here. I would suggest noting that for 90% of the data, the bias falls between -5 and 5 cm. Similar wording can be applied throughout.

L378-379: Replace “In the last ... than 50 cm;” with “For snow depths larger than 50 cm;”.

L379-380: Replace ”Although the ... were underestimated” with “Although the estimates for large snow depths are underestimated”.

L380: Figure 5 needs to be referenced in the beginning of this paragraph, not at the end.

L381: What is the difference between “small error” and “high accuracy” in this sentence? They seem to mean the same. Please clarify.

L386: (Fig. 5) Is the legend item "Frequent Count" meant to represent the bias between fused snow depth and in situ observations? If so, please clarify that in the legend. Also, please explain the legend item "Gauss" in the legend. This must be a gaussian distribution fit to the data. ?

L397: Remove abbreviation explanations. They have already been explained.
L400: Replace “slowly” with “slow”.

L406: Replace “shallower” with “less”.

L409: Replace “shallower” with “less”.

L419: Replace “the farthest east of Canada” with “eastern Canada”.

L419: Replace “Alps” with “European Alps”.

L420: Remove capitalization of the seasons.

L424: Please make the x-ticks in plots a, b, and c consistent. Either one tick every 20 degs, or every 40 degs.
L427: Remove “of change”.

L433: Replace “area” with “of the area”.

L437: Remove “very”.

L439: Replace “changed response times” with “change rate”.

L446: Replace “spatio-temporal” with “spatio-temporally”.

L476: Remove “Regarding the limitations of this study,”.

L483: Replace “theses” with “these”.
L490: Replace “consistency” with “agreement”.

L498: Replace “leaning” with “learning”.