Response to Topical Editor Dr. Xin Li:

Comments to the author:

The manuscript has been greatly improved, however, it still needs to revise some parts, please refer to the reviewer's comments for further revision.

We thank Topical Editor Dr. Xin Li for handling this manuscript. We have carefully revised the manuscript to address the issues and comments raised by the reviewer during this round. In particular, we have revised the figures and texts related to the error analysis. We have also corrected the style and language issues raised by the reviewer. Please see detailed information in the following point-to-point responses.

Response to the Reviewer:

(1) The authors addressed all reviewer comments and improved the manuscript quality consistently. However, in my opinion, before the manuscript can be accepted for publication, the part of the manuscript regarding the error analysis must be improved. After these issues are solved, the manuscript should be accepted for publication.

We thank this reviewer for her/his valuable time in reviewing this manuscript again and for providing thorough and insightful comments. We have carefully revised the manuscript to address the issues and comments raised during this round. In particular, we have revised the figures and texts related to the error analysis. We have also corrected the style and language issues raised by the reviewer.

Point-to-point responses are listed below. Comments are in **black**, the authors' responses are in **blue**, and the revisions in the manuscript are in **red**.

- Section 3.5 is unclear in two points:

(2) a) I don't understand why the authors choose to compute the rRMSE relative to the variation in X, (max(X)-min(X)). For the case min(x)=0 this would be the maximum value (which is the case for this dataset), and if the variation in X is small the relative values diverge. I would use either the mean value as for the STE, or normalize each error (Y-X) by X before computing the root mean. Please change it or give a valid reason for using this type of normalization.

We have changed the relative RMSD (rRMSD) to normalized RMSD (nRMSD), and the relative STE (rSTE) to normalized STE (nSTE) for simplification. We have revised to use the mean value to normalize the RMSD. Please see below for the revised texts:

The Root Mean Square Difference (RMSD), normalized RMSD (nRMSD), STE, and normalized STE (nSTE) are four error indicators used in this study. The RMSD of two data (X and Y) are given by $RMSD = \sqrt{\sum(X_i - Y_i)^2/N}$, where N is the number of elements in the sample. The nRMSD is given by $nRMSD = RMSD/(\mathrm{mean}(X))$. The STE of one data (Z) is given by $STE = \sigma_Z/\sqrt{N_Z}$, where σ_Z is the standard deviation of the data Z and N_Z is the number of elements in Z. The nSTE is given by $nSTE = STE/\bar{Z}$, where $\bar{Z}$ is the mean of the sample.

The nRMSD normalized by the mean value is larger than by the (Max-Min). We have added several sentences in the revised Section 4.1 to explain this issue:

"… We show the correlation coefficient (r), RMSD, and nRMSD values for each comparison. It should be noted that the nRMSD (snow depth > 5 cm) is significantly lower than the nRMSD (all), which is because the reference value (i.e., the mean snow depth) was used to normalize the RMSD. A large portion of snow depths in the study area is lower than 5 cm, yielding a lower mean value when involving all the data than only using the > 5 cm data. The same principle applies to the following Figures 9, 10, and 11. Nevertheless, the metrics only represent the comparison during the intermediate process of the data set production. Users can define their own rules to use the data according to the quality flags in the published data set."

(3) b) It is also not clear from the text which dataset Z is used for determining the STE. I assume this are the different traces used for one datapoint. Please explain more it in detail. Since the STE is given for each sample should it not be more correct to use the standard deviation? Moreover, if Z is the sample of traces used for one datapoint then N must be the number of traces, which is different of the N used for RMSE which is the number of datapoints. Use different symbols if this is the case.

We apologize for the unclear description of the STE. Yes, the different traces used for one data point are used to determine the STE. We have added explanations in the titles of Figures 9, 16, and 19 to clarify this issue:

"… The error bar of each point is the standard error (STE) of the snow depths for all the available tracks of this point. …"

We have revised the expression of the STE to replace N with N_Z to distinguish these two numbers. Please see below:

"…The STE of one data (Z) is given by $STE = \sigma_Z/\sqrt{N_Z}$, where σ_Z is the standard deviation of the data Z and N_Z is the number of elements in Z. …"

We agree with the reviewer that the standard deviation (STD) is also a helpful indicator. In this study, we use the STE instead of STD, the same as that used in the published west U.S. PBO H_2O GNSS-IR snow depth data set on the National Snow and Ice Data Center (NSIDC, https://nsidc.org).

Reference:

Larson, K. M. and E. E. Small. 2017. *Daily Snow Depth and SWE from GPS Signal-to-Noise Ratios, Version 1*. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: https://doi.org/10.5067/Z02Y1HGNFXCH.

(4) - I don't agree with the argument of the authors that it is not necessary to indicate the RMSE and rRMSE of the data in figure 11. It is true that comparison between in-situ and GNSS data is tricky due to the different footprint. However, some information is better than no information. The footprint point can still be explained in the text. Please indicate RMSE and rRMSE for figure 11 a1,b1 and c1. The same is true for sentence at line 510.

We have added the RMSD and nRMSD in Figure 11 a1, b1, and c1. Please see the revised figure in the following Comment # (5). We have also added a sentence to explain the footprint issue:

“Due to the inconsistent footprint between the GNSS and in-situ measurements, the error metrics presented in Figure 11 are for reference only and do not represent factual accuracies.”

(5) Moreover, the second part of figure 11 is visually very appealing but unfortunately the higher bars cover the points in the back. I suggest using a simple color plot since the z-axis is already indicated by the color. (At least I think since the z-axis label is missing). I also find valuable to use a color plot similar to Figurea1,b1 also for fig. 11a1,b,c1. It would indicate the density of point relative to the 1:1 line.

We have remade Figure 11 a2, b2, and c2 using the display style of “tile”. We have also changed to use density plot for (a1), (b1), and (c1). To keep consistence with the previous Figures 9 and 10, we changed to use the > 5 cm data to plot (a1), (b1), and (c1). The revised figure 11 is shown below:

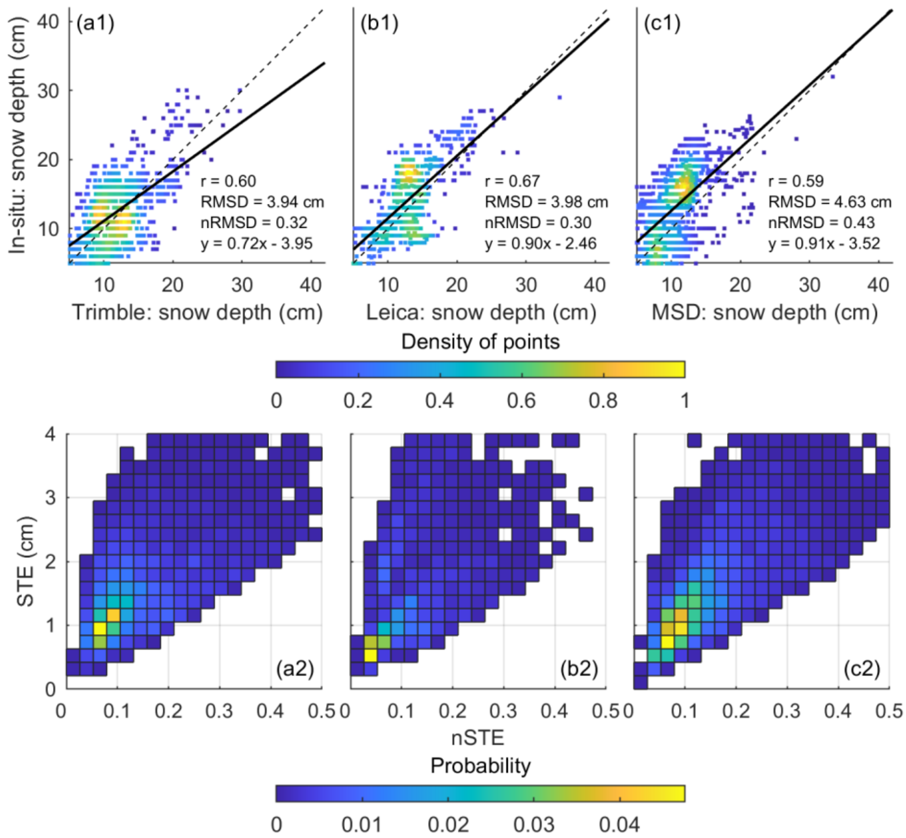


Figure 11. Comparisons of the GNSS-derived snow depth and the in-situ measurements from different types of GNSS receivers: (a1) Trimble; (b1) Leica; (c1) Minshida (MSD), and the histogram of the standard error (STE) and nSTE of snow depths for different types of GNSS receivers: (a2) Trimble; (b2) Leica; (c2) MSD. The number of sites representing Trimble, Leica, and MSD is 20, 5, and 24. The GNSS snow depths values are greater than 5 cm in this figure. To prevent other possible effects besides the receiver type, the STE of snow depths is less than 1 cm (63% of the entire data) in (a1), (b1), and (c1). RMSD: Root Mean Square Difference; nRMSD: normalized RMSD.

(6) Below a few minor comments. Mostly style and language. The line numbers refer to the document with track changes.

Line 40: delete good. “the potential” is sufficient.

Fixed.

Line 47: delete "years of"

Fixed.

Lines 68-69: ''remote sensing has a long revisiting period (>20 days) and high cost …"

Fixed.

Line 99: " … for typical GNSS-IR sites…"

Fixed.

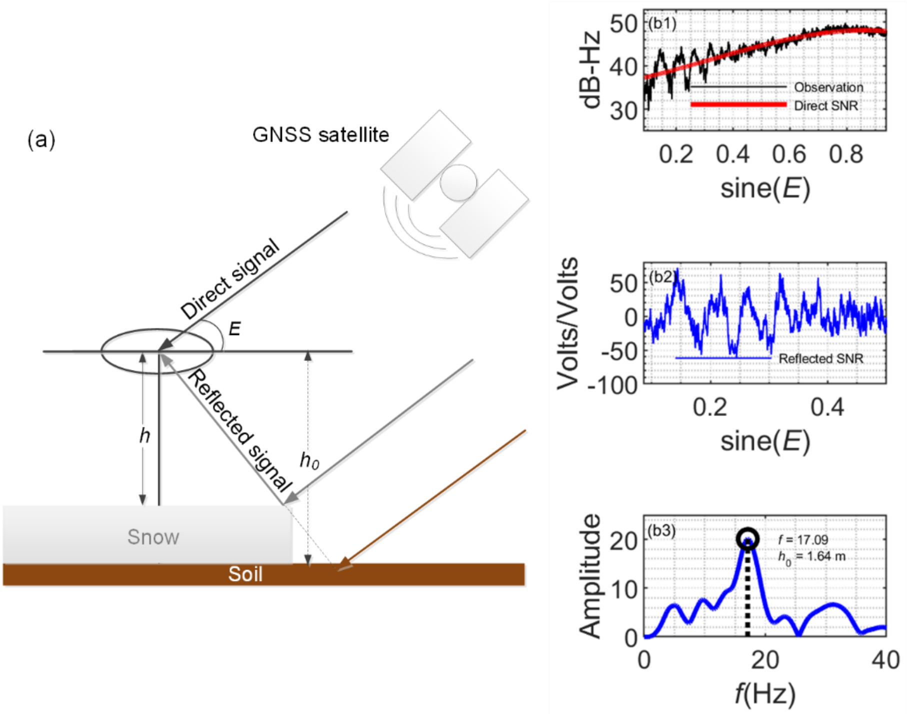
Line 128: It is not clear to me in which sense snow is a natural disaster for pastural area if is providing fresh water. Do you mean the lack of snow?

Sorry for the unclear description. We have revised this sentence as below:

Northern China lies between latitudes 25°N and 55°N and longitudes 70°E and 140°E and includes humid, semi-humid, semi-arid, and arid zones. Snow is the primary freshwater resource ~~over~~ in this area ~~and the main natural disaster in winter~~. Sudden snowstorms or long-lasting deep snow is one of the major natural disasters for pastoral areas because it affects livestock grazing.

Figure 4: Please indicate E in (a).

We have added E in (a). Please see below:



Line 285: Indicate the return period. Do the tracks repeat every day?

We have clarified this issue. Please see below:

It is worth mentioning that GPS ground tracks have sidereal repeatability and reappear at the same azimuth every day.

Line 303: housing instead of small house.

Fixed.

Line 336: It is not clear to which site you are referring to. Figure 5 has 3 sites.

We have clarified this issue. Please see below:

At the same time, it is impossible to derive correct h_0 values for “bgfc” in other orientations that have buildings or trees; This phenomenon can be verified from the photo of the site in Figure 5.

Line 411: “…site over from October…”. Delete ‘’over’’.

Fixed.

Line 421: included instead of involved.

Fixed.

Line: 424: Please change to “… possible error due to vegetation…”

Fixed.

Line 474: See comment above about STE. Do you indicate the same STE for each point?

For one specific point, the different traces are used to determine the STE of this point. Therefore, each point has its own STE value. Please see detailed responses in the previous Comment # (3).

Lines 475-476: It is not clear to me what the authors mean with this sentence. Did you filter all data with large STE. In this case why? Specify how many outliers (percentage) were deleted and precise which other effects besides GNSS frequency you expect.

We filtered out the data with large STE (i.e., STE > 1 cm). We have specified the percentage of valid data in the titles of Figures 9, 10, and 11.

The 1 cm is a rigorous threshold. The manuscript states that GNSS constellation, frequency, and receiver type are key factors affecting snow depth accuracy. If we compare one of the three factors, we should prevent the other two and other random errors from cross-influence. In other words, we should ensure a snow depth value is “accurate” under the defined condition. Therefore, we use the STE = 1 cm threshold to prevent this issue.

We have added explanations in the revised manuscript Section 4.1:

“The intra-comparisons of the snow depths are executed from three aspects, i.e., comparison of different GNSS constellations, frequency bands, and receivers. If we compare one of the three factors, we should prevent the other two and other random errors from cross-influence. In other words, we should ensure a snow depth value is “accurate” under the defined condition. Therefore, in this section, we use a rigorous threshold of STE = 1 cm to filter out the outliers. …”

We have also remade Figures 9 and 10 considering the reviewer’s comments comprehensively, i.e., we revised (or added) the nRMSD, revised the style of the density plot, and used a consistent 1 cm threshold for all the figures. For the convenience of the reviewer, the revised Figures 9 and 10 are shown below:

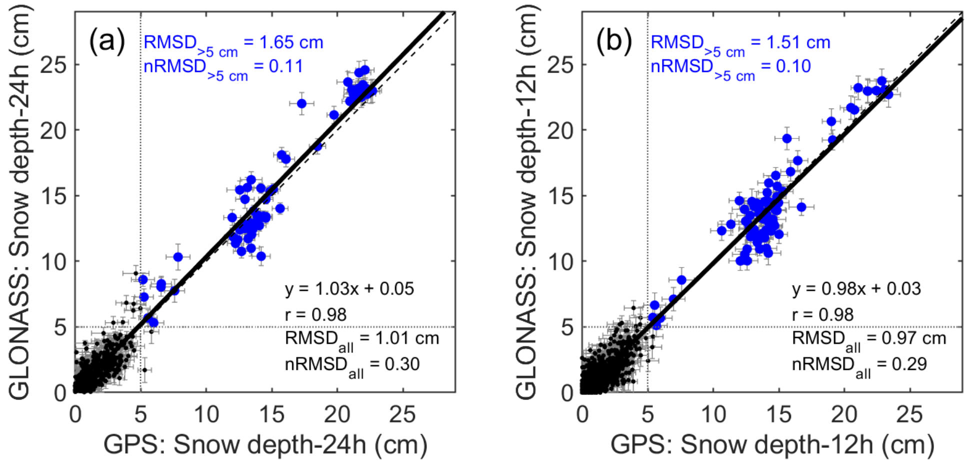


Figure 9. Correlations of 24 h/12 h snow depths from GPS and GLONASS observations. (a) 24 h; (b) 12 h. The error bar of each point is the standard error (STE) of the snow depths for all the available tracks of this point. Four available sites, i.e., hltl, hlhl, bfqe, and bttl, during the GPS/GLONASS overlapped periods (i.e., the year 2014 and 2015) are used to plot this figure. For each point in the figure, the number of valid observations is more than five. To prevent other possible effects besides the GNSS constellation, the STE of snow depths is less than 1 cm (90% for the 24 h data and 76% for the 12 h data). Blue points are with the retrieved GPS and GLONASS snow depths greater than 5 cm. RMSD: Root Mean Square Difference; nRMSD: normalized RMSD.

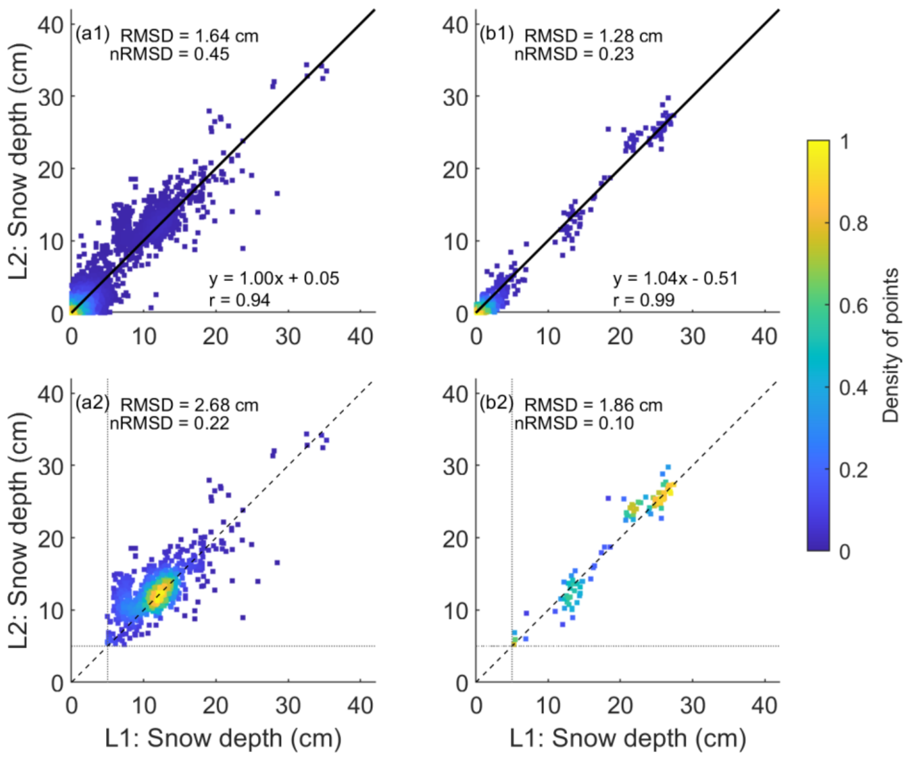


Figure 10. Correlations of snow depth from different GNSS frequencies. (a1) GPS L1 vs. GPS L2; (b1) GLONASS L1 vs. GLONASS L2. The color bar represents the density of points; (a2) Same as (a1) but with snow depths greater than 5 cm; (b2) Same as (b1) but with snow depths greater than 5 cm. Fifty-one high-quality GPS sites of CMA and four GPS/GLONASS compatible sites are respectively used to plot (a1, a2) and (b1, b2). For each point in the figure, the number of valid observations is more than five. To prevent other possible effects besides the GNSS frequency, the STE of each snow depth is less than

1 cm in all the subfigures (61% for the GPS data and 70% for the GLONASS data). RMSD: Root Mean Square Difference; nRMSD: normalized RMSD.

Lines 496-500: It is not cleat to me relative to which data the RMSE and rRMSE indicated in this paragraph were computed. Please specify.

We apologize for the unclear description. The r, RMSD, and nRMSD values are the average values from Figures 9 ~ 11. For example, in Figure 9, the RMSD of the 24-hour and 12-hour GPS/GLONASS results is 1.01 cm and 0.97 cm respectively, then the mean RMSD is 0.99 cm. We have revised the corresponding texts to clarify this issue:

"From the comprehensive intra-comparisons shown in Figures 9 ~ 11, we conclude that the snow depths derived from different GNSS constellations, frequency bands, and receivers have overall good agreement. The average values of the metrics shown in Figures 9 ~ 11 are summarized as follows: mean r = 0.98, mean RMSD = 0.99 cm, and mean nRMSD (snow depth >5 cm) = 0.11 for different GNSS constellations, mean r = 0.97, mean RMSD = 1.46 cm, and mean nRMSD (snow depth >5 cm) =0.16 for different frequency bands, and mean r = 0.62 for different GNSS receivers. Therefore, it is feasible to combine all these results to produce the snow depth data set in this study."

Lines 518-521: All snow depth values? I see just 16 sites but the dataset contains much more sites. Please correct or specify.

We have clarified this issue. Please see below:

Figure 12 presents all the GNSS snow depth values ~~derived in the produced GSnow-CHINA data set~~ of the 16 GNSS sites, regardless of its quality, to give a comprehensive illustration of the data.

Figure 14 and corresponding text: Can you give an explanation for the underestimation of mean snow depth by PMW? It has a larger footprint but why is the mean snow depth over a larger footprint smaller?

We have added explanations for this issue. Please see below:

For the mean values shown in (b2), the GNSS and in-situ have a better agreement than the PMW because of the significant difference in their spatial footprint. Most sites are located in the region with evergreen coniferous forest, which prevents the PMW data from acquiring reliable snow depth values due to its wider observation extent of 25 km.

Line 578: Change to "PMW data were available only for the period 2016-2020."

Fixed.

Line 689: Provide a reference of the precious study.

We have added the reference:

Wan, W., Larson, K. M., Small, E. E., Chew, C. C., and Braun, J. J.: Using geodetic GPS receivers to measure vegetation water content, GPS Solutions, 19, 237-248, 10.1007/s10291-014-0383-7, 2015.

Line 693: change to ''30 seconds for which is impossible …"

Fixed.