Response to Anonymous Referee #2

We greatly appreciate your valuable suggestions and comments. We studied your comments carefully and made substantial revisions accordingly. We hope that our revisions meet the requirements. Our modifications and responses to the reviewer's comments are listed as follows. The reviewer's comments are in blue text, while the authors' responses are in black text. All revisions in the revised manuscript are in red text.

A very interesting paper to develop an improved SWE dataset based on existing multi-source SWE products using machine learning. However, there are some points which need to be well addressed. Please see following comments:

Reply: We greatly appreciate your careful review of our manuscript.

1. The ridge regression machine learning algorithm in Section 2.3 needs to be better described.

- Why ridge regression is selected among so many machine learning algorithms? The advantages of using ridge regression over others for this application need to be highlighted.

Reply: The text has been modified to improve clarity.

The advantages of the ridge regression model are the following:

1) The ridge regression algorithm can effectively solve the multicollinearity problem, i.e., the independence of training products and models. In this study, the reanalysis data based on SWE products cannot make the products and models independent of each other, i.e., they are prone to the multicollinearity problem, which leads to distorted model estimation or difficulty in accurate estimation. It is, therefore, a good way to use the ridge regression algorithm to address this problem.
2) The ridge regression model has high accuracy and stability for "ill-conditioned" data. In this study, we used various forms of gridded SWE datasets, and the accuracy of each gridded SWE dataset varied. A small change in one of the SWE products involved in the training will cause a large error in the final calculation results. Therefore, using the ridge regression algorithm to address "ill-conditioned" SWE data is an effective method.

In addition, we provide an additional explanation of the advantages of the ridge regression model in Section 2.3.

"Generally, since the reanalysis data based on SWE products cannot make the products and models independent of each other, i.e., they are prone to the multicollinearity problem, which leads to distorted model estimation or difficulty in performing accurate estimations. In contrast, the ridge regression model can successfully solve the multicollinearity problem, i.e., the independence of training products and models. In addition, when integrating multiple SWE products, the accuracy of each SWE dataset is likely to differ. A small change in one of the SWE products involved in the training will cause a significant error in the final calculation results, while the ridge regression model has high accuracy and stability for this "ill-conditioned" SWE data".

- What is the requirement of ridge regression in terms of variable dependence? Does it require predictor variables and target variable independent? Are the station data from GHCN or Russian snow survey used (or partially used) for generating the gridded SWE products used in this paper? If so, the target variable and predictor variables are intrinsically related. How does this impact this algorithm? Please provide more information to explain and justify.

Reply: The text has been modified to improve clarity.

Regarding the question "What is the requirement of ridge regression in terms of variable dependence? Does it require predictor variables and target variable independent?, the ridge regression model is flexible in the choice of predictor variables and does not require the predictor and target variable to be independent of each other. It is effective in solving machine learning problems with correlations between predictor and target variables (Duzan and Shariff, 2015; Saleh et al., 2019).

Regarding the question "Are the station data from GHCN or Russian snow survey used (or partially used) for generating the gridded SWE products used in this paper?", a few of the gridded SWE products used in this paper use hemispheric-scale snow course (HSSC) observations (in this revised version, the GHCN data have been replaced by HSSC data) data and Russian snow survey data (RSSD), and the predictor and target variables are intrinsically related.

Regarding the question "How does this impact this algorithm?**"**, this "intrinsically related" does reduce the accuracy of SWE estimates and leads to a reduction in the stability of SWE estimates. However, most of the current SWE observations are inevitably subject to such a multicollinearity problem. The ridge regression model is a biased estimation method specifically designed for the problem of multicollinearity data (Duzan and Shariff, 2015; Saleh et al., 2019). It has good tolerance to "ill-conditioned" data (Hoerl and Kennard, 1970b; Guilkey and Murphy, 1975). The ridge regression model can solve the multicollinearity problem of predictor and target variables and effectively reduce the impact of this problem on the preparation accuracy of the RRM SWE product.

We have provided detailed supplementary explanations in the Introduction and Section 2.3.

Introduction:

"The ridge regression model is a biased estimation method specifically designed to address the problem of multicollinear data (Duzan and Shariff, 2015; Saleh et al., 2019). It has good tolerance to "ill-conditioned" data and has a good effect in using SWE data to address the multicollinearity problem (Hoerl and Kennard, 1970b; Guilkey and Murphy, 1975)."

Section 2.3.:

"The ridge regression model is flexible in the choice of predictor variables and does not require the predictor and target variable to be independent of each other. It can effectively solve the multicollinearity problem of predictor and target variables as well as reduce the impact of this problem on the training model (Duzan and Shariff, 2015; Saleh et al., 2019). Generally, since the reanalysis data based on SWE products cannot make the products and models independent of each other, i.e., they are prone to the multicollinearity problem, which leads to distorted model estimation or difficulty in performing accurate estimations. In contrast, the ridge regression model can successfully solve the multicollinearity problem, i.e., the independence of training products and models. In addition, when integrating multiple SWE products, the accuracy of each SWE dataset is likely to differ. A small change in one of the SWE products involved in the training will cause a significant error in the final calculation results, while the ridge regression model has high accuracy and stability for this "ill-conditioned" SWE data".

References

Hoerl, A. E. and Kennard, R. W.: Ridge regression: Biased estimation for nonorthogonal problems, Technometrics, 12, 55-67,1970

Guilkey, David K., and James L. Murphy.: Directed ridge regression techniques in cases of multicollinearity, Journal of the American Statistical Association, 70, 769-775, 1975

Duzan, H. and Shariff, N. S. B. M.: Ridge regression for solving the multicollinearity problem: review of methods and models, Journal of Applied Science, 2015.

Saleh, A. M. E., Arashi, M., and Kibria, B. G.: Theory of ridge regression estimation with applications, John Wiley & Sons2019.

- It is unclear how DEM is used in the ridge regression. Is it used as a predictor variable in addition to AMSR-E/AMSR2, ERA-Interim, GLDAS, GlobSnow, ERA5-land SWE or just used to evaluate different model performances related to elevation? There are many important factors impacting the SWE estimation in addition to DEM. Have you also considered adding more variables to further improve the model? Please provide more information to explain.

Reply: The text has been modified to improve clarity.

In this study, the DEM was used as an important environmental feature information input to the ridge regression model and was included in the model training. The DEM is an auxiliary terrain feature variable in addition to the five SWE prediction feature variables AMSR-E/AMSR2 SWE, ERA-Interim SWE, GLDAS SWE, GlobSnow SWE, and ERA5-land SWE.

In addition to the DEM, meteorological elements, NDVI, land type, and other factors will affect the SWE estimation. Unfortunately, our current training model does not consider these factors in detail, which is a limitation of the current RRM SWE product. Currently, there are two main reasons why we do not consider these impact factors. The uncertainty of the model will be greater after adding these impact factors, and the accuracy of the RRM SWE product may not be improved. However, an investigation considering these impact factors in detail is a significant workload and cannot currently be performed. In future studies, we will carefully consider these factors in detail to further improve the accuracy of the RRM SWE product.

We added a description of the role of DEM data in Section 2.3.

"The digital elevation model (DEM) was used as an important environmental feature input to the ridge regression model and was included in the model training. The DEM is an auxiliary terrain feature variable in addition to the five SWE prediction feature variables, AMSR-E/AMSR2 SWE, ERA-Interim SWE, GLDAS SWE, GlobSnow SWE, and ERA5-land SWE."

$$\hat{\beta}^{ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}, \quad (1)$$

where $\hat{\beta}^{ridge}$ is the extremum solution function of ridge regression and p is the number of gridded SWE product variables involved in training. x_i are the prediction feature variables, which contain two parts, one set contains the main feature variables of the gridded SWE products, and the other part consists of the DEM auxiliary feature variables. y_i is the observed snow water equivalent, and λ, β, β_j and β_0 are the parameters to be solved. $1,\cdots,N$ is the sample of the training dataset. $\lambda\sum_{j=1}^{p}\beta_j^2$ is the penalty function terms.

We added limitations of the RRM SWE product and directions for future improvement in the Results and Discussion sections.

"Then, in addition to the DEM, meteorological elements, NDVI, land type, and other factors will affect the SWE estimation. Unfortunately, our current training model does not consider these factors in detail, which is a limitation of the current RRM SWE product."

- The authors mentioned in the model training process, "... reduced the training data appropriately for the regions with denser training data, and make it close to the amount of training data in the sparse region" Please be more specific how this is done. Is it through randomly selecting training data in the regions with denser data? If so, what is the amount of the training samples used? Please provide more information.

Reply: This suggestion has been accepted, and the text has been modified accordingly.

For the region with dense training data, the training data were not simply selected randomly, but sample points that were spatially uniformly distributed were selected for the training as much as possible based on the latitude and longitude information of the observational points. The sample sizes of the training, validation and test datasets were divided according to the ratio of 7:2:1, where the numbers of training, validation and test set samples were 271651, 77614 and 38807, respectively.

We have provided detailed supplementary explanations in Section 2.3.

"The sample points that were spatially uniformly distributed were selected for training as much as possible based on the latitude and longitude information of the observational points."

"The total number of samples N in the training dataset is 271651. The sample sizes of the training data set, validation data set and test data set are divided according to the ratio of 7:2:1, where the numbers of training set, validation set and test set samples are 271651, 77614 and 38807, respectively."

2. What is the range of the generated SWE product? Would be nice to see the range of generated SWE, and also the performances of SWE at different ranges. In Figure 4, it shows the data all below 400mm. What is the capability of this RRM product for capturing deeper snow? Please add more descriptions or discussions on this.

Reply: This suggestion has been accepted, and the text has been modified.

Regarding the question of "What is the range of the generated SWE product?**"**, the range of the RRM SWE product is all land regions north of 45° N. The spatial range of the RRM SWE product is consistent with that of the DEM. In Figure 1, we labeled the range of the RRM SWE product in detail.

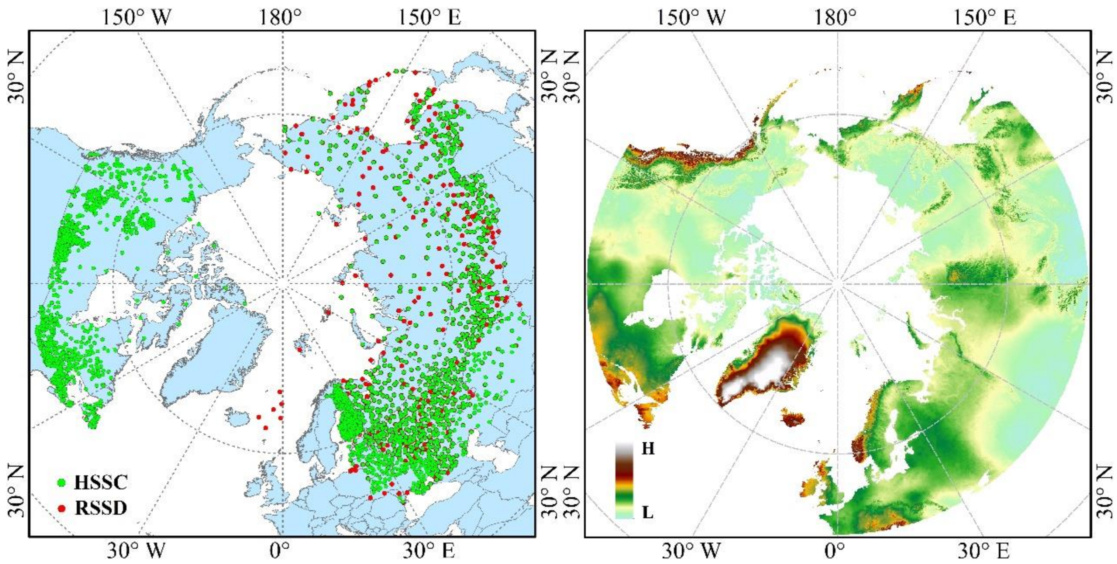

Figure 1: The DEM and snow survey stations of the research region. The right subgraph shows the DEM, and the left subgraph shows the SWE observational stations. HSSC, hemispheric-scale snow course; RSSD, the Russian snow survey station. The spatial range of the RRM SWE product is consistent with that of DEM.).

Regarding the question of "Would be nice to see the range of generated SWE, and the performances of SWE at different ranges,**"** we evaluated the performances of the RRM SWE product in three representative regions: Russia, Canada, and Finland.

We have made the following changes in the Abstract:

"The average MAE and RMSE of the RRM SWE products are 0.22 and 19.92 mm at different altitude intervals and 0.21 and 27.00 mm at different regions, respectively."

We have made the following changes in Section 2.4.2:

"In addition, we also evaluated the performances of the RRM SWE product in three representative regions: Russia, Canada, and Finland."

We have made the following changes in Section 3.2:

Table 3: Error list for the station data and RRM SWE product in different regions.

Region	MAE	RMSE (mm)	R	R^2
Russia	0.20	26.39	0.89	0.79
Canada	0.23	29.31	0.87	0.76
Finland	0.21	25.29	0.89	0.79

"The RRM SWE product has good performance in different regions, and its RMSE in Russia, Canada, and Finland are 26.39 mm, 29.31 mm, and 25.29 mm, respectively; additionally, the performance of the RRM SWE product in different regions is basically similar (Table 3). The RRM SWE product not only performs well at different altitudes but also in different regions, and it has good stability."

Regarding "In Figure 4, it shows the data all below 400mm. What is the capability of this RRM product for capturing deeper snow? Please add more descriptions or discussions on this,**"** we retrained the model using hemispheric-scale snow course observational data to generate the new RRM SWE product. In Figure 4, the evaluation results for the SWE above 400 mm were added. The results show that the RRM SWE product still has a higher ability to capture the SWE above 400 mm than other products.

We modified Figure 4 and added a discussion of the related content in the revised manuscript.

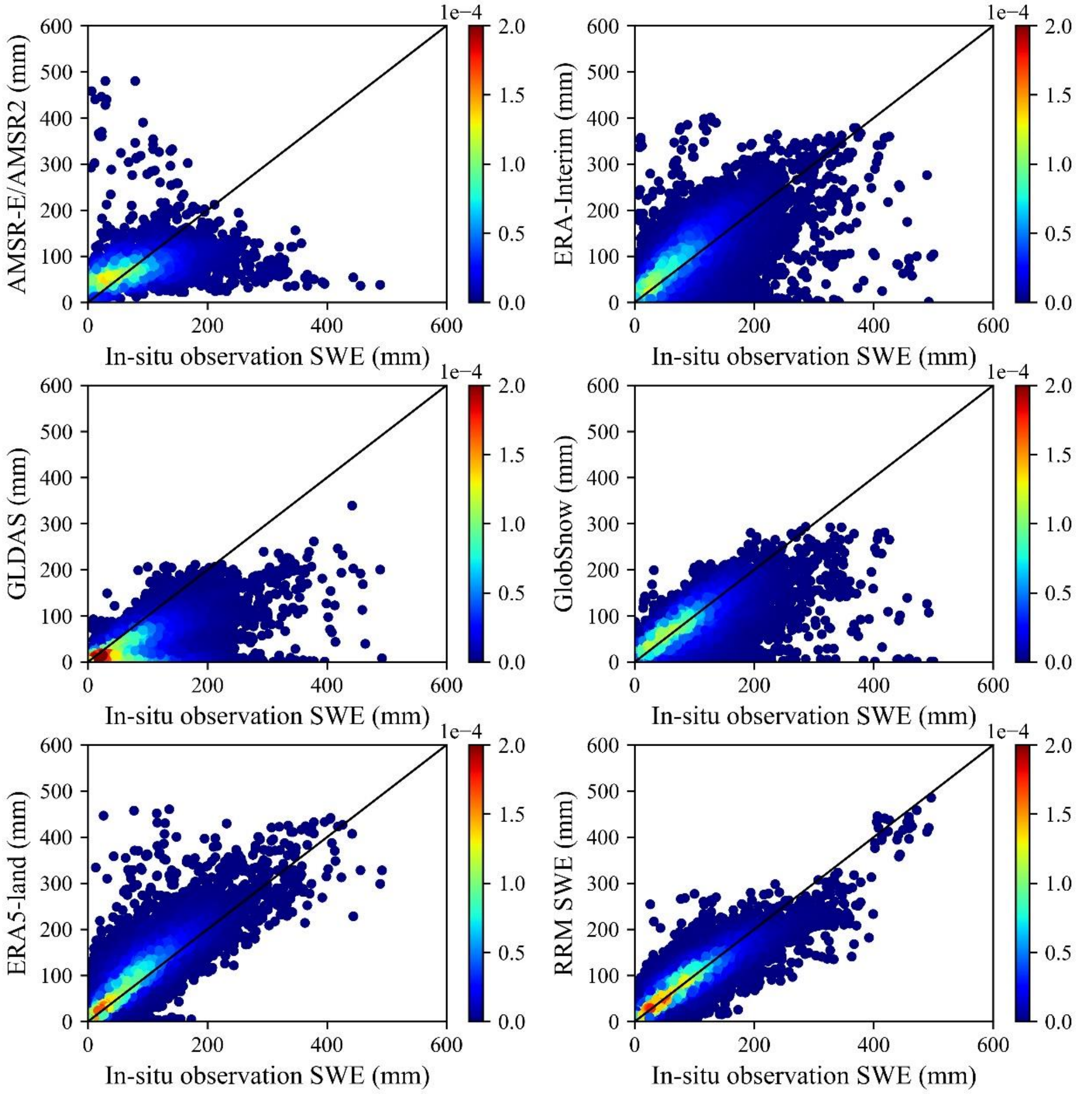

Figure 4: Error verification density diagram (a total of 38807 sample points were used for verification). The color bar represents the value of kernel density estimation. The closer the high-density area is to the 1:1 line, the higher the verification accuracy of the dataset is at most of the measuring stations.

"For an SWE above 400 mm, the MAE and RMSE of the RRM SWE product and the measured SWE are 0.35 and 43.57 mm, respectively. Although the RRM SWE product is better than other products in capturing the SWE above 400 mm, it is still not as good at capturing the SWE below 400 mm relative to itself."

3. In Figure 1, the stations block the DEM information, making it difficult for the audiences. I would suggest to change figure 1 into a two-panel figure, with one for stations and the other for DEM.

Reply: Thank you for your suggestion. Figure 1 has been modified.

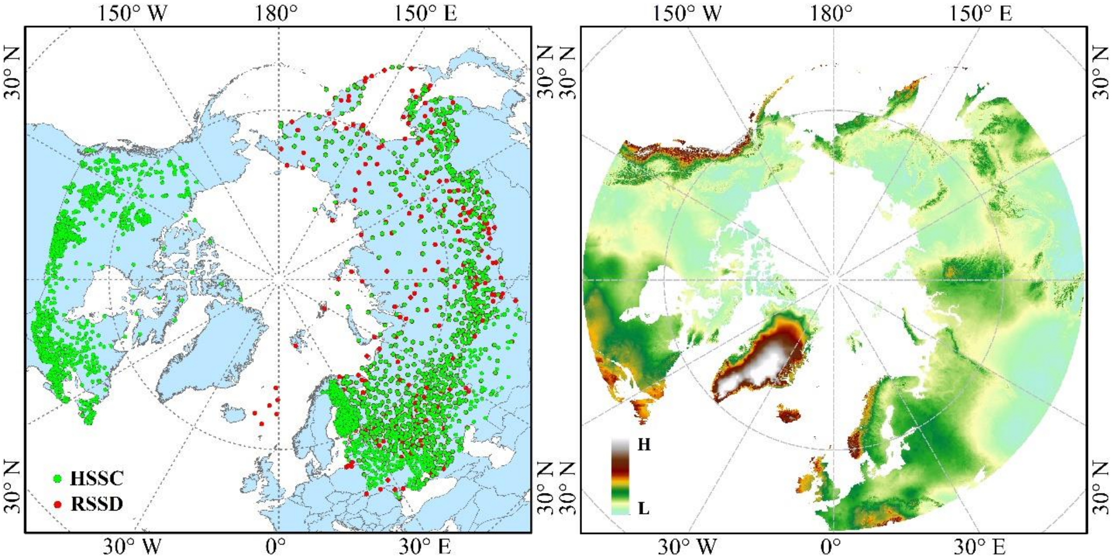

Figure 1: The DEM and snow survey stations of the research region. The right subgraph shows the DEM, and the left subgraph shows the SWE observational stations. HSSC, hemispheric-scale snow course; RSSD, the Russian snow survey station. The spatial range of the RRM SWE product is consistent with that of the DEM.

4. From Table 2, the improvement of the RRM product is not so big as compared to ERA5-land and GlobSnow SWE alone. It would also be nice to see the comparisons with other fusion methods, such as the simple multi-source data average or regular linear regression. Also, it shows that GLDAS SWE has poor performance as compared to other products. Have you considered to leave it out and further improve the machine learning model?

Reply: Thank you for your suggestion; the text has been modified accordingly.

Because of systematic errors in the NOAA GHCN SWE data, we retrained, tested, and validated the model using the new hemispheric-scale snow course (HSSC) observational data and Russian snow survey data (RSSD) to generate a new version of the RRM SWE product and re-evaluated the product.

In addition, we compared the RRM SWE product with the SWE dataset obtained by the multisource data average method. The evaluation results are shown in Table 2 and Figure 3.

The new evaluation results show that the accuracy of the RRM SWE product is significantly improved compared to that of the ERA5-land SWE and GlobSnow SWE products. In addition, although the multisource data average method can improve the accuracy of SWE products to some extent (better than AMSR-E/AMSR2 SWE and GLDAS SWE), the improvement of this method is still very limited. The RRM SWE product has a significant advantage over the multisource data average method, and its accuracy is much higher than that of the simple multisource data average method.

Regarding "Additionally, it shows that GLDAS SWE has poor performance compared to other products. Have you considered to leave it out and further improve the machine learning model?**",** in preparing the RRM SWE product, the machine learning algorithm we constructed can automatically select the better-performing SWE training product based on the validation dataset. Therefore, although the accuracy of the GLDAS SWE product is poor, it does not affect the training results of machine learning algorithms or the accuracy of the RRM SWE product.

We have added specific content to the revised manuscript.

"In addition, we compared the RRM SWE product with the SWE dataset obtained by the multisource data average method."

Table 2: Error list for the station data and grid snow water equivalent products.

Error type	MAE	RMSE (mm)	R	R^2
ERA-Interim	0.43	46.81	0.69	0.48
AMSR-E/AMSR2	0.49	52.39	0.47	0.22
GLDAS	0.58	65.25	0.52	0.27
GlobSnow	0.32	40.99	0.70	0.49
ERA5-land	0.32	37.02	0.84	0.71
Multisource data average	0.44	52.00	0.51	0.26
RRM SWE	0.21	25.37	0.89	0.79

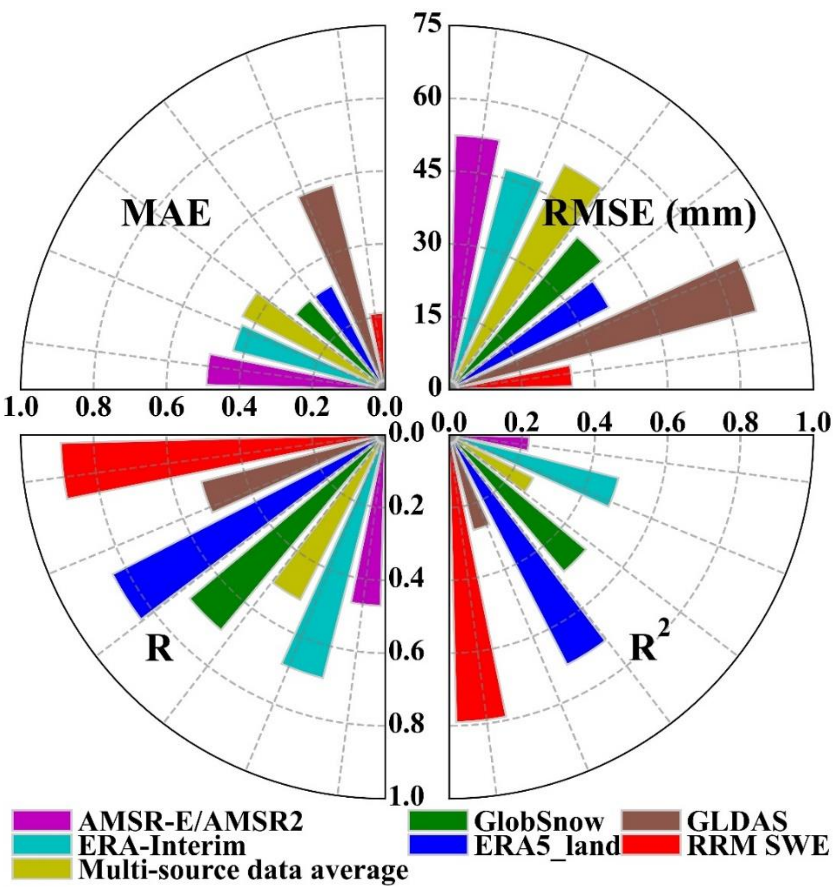

Figure 3: Accuracy comparison of various snow water equivalent products. The upper left sector represents the MAE, the upper right sector represents the RMSE, the lower-left sector represents R, and the lower right sector represents R^2. The sector axis represents the size of the error, and the color represents different SWE datasets.

"Although the multisource data average method can improve the accuracy of SWE products to some extent (better than AMSR-E/AMSR2 SWE and GLDAS SWE), the improvement of this method is still very limited. The RRM SWE product has a significant advantage over the multisource data average method, and its accuracy is much higher than that of the simple multisource data average method."