The Cryosphere Discuss.,
https://doi.org/10.5194/tc-2020-248-AC1, 2020

Interactive comment on "Snow Ensemble Uncertainty Project (SEUP): Quantification of snow water equivalent uncertainty across North America via ensemble land surface modeling" *by* Rhae Sung Kim et al.

Rhae Sung Kim et al.

rhaesung.kim@nasa.gov

Received and published: 10 December 2020

Dear Dr. Essery,

Thank you for your constructive comments. Based on the feedback, we have made several changes and improvements to the manuscript. Below we have provided a response to each of the comments provided.

Richard L.H. Essery (Referee1)'s comments:

Kim et al. quantify uncertainty in estimates of snow storage by a set of operational models and forcing datasets. This is a topic of considerable interest, given the importance of snow storage for water resources and the difficulty of estimating it by remote sensing. The paper is interesting and generally well written; my comments focus on clarifying the methodology.

119 Why was the three-layer snow scheme in JULES, described by Best et al. (2011) and operational at the Met Office, not used?

The reviewer is correct that currently, the Met Office uses a three-layer snow scheme which was deployed in 2018, post the modeling setups were devised for this study.

We have revised the text to clarify in lines 124-125 on page 4, "Note that the UKMO currently uses a three-layer scheme in JULES, which wasn't available in NASA LIS at the time this study was devised."

122 "not tuned in this study to assess current configurations" is ambiguous. I assume that "to assess current configurations, parameters were not tuned" is what is intended.

We have revised this sentence in lines 125-126 on page 4 as: "In order to assess current configurations, initial model conditions and model parameters used in the operational set-up were not tuned in this study."

130 State the original spatial and temporal resolutions of the forcing datasets.

We have added the following sentences in the main manuscript and S2.

Main manuscript (in lines 133-134 on page 5): "Original spatial and temporal resolutions for these datasets are described in Section S2".

Section S2.1 (in lines 107-108 on page 4 in supporting information): "MERRA2 has a native spatial resolution of 0.5° latitude by 0.625° longitude (roughly 50 km) and hourly temporal resolution."

Section S2.2 (in lines 113-115 on page 4 in supporting information): "The GDAS model grids have been upgraded from roughly 80 km (since 2000), ∼60 km (Oct. 2002), ∼38 km (Jun. 2005), 27 km (Jul. 2010), to ∼13 km (January 2015). The temporal resolution is 3-hour."

Section S2.3 (in lines 119-122, on page 4 in supporting information): "In this study, the operational real-time data from the ECMWF-Integrated Forecast System (IFS) are used; the meteorological fields are provided on a 0.25-degree grid (roughly 25 km) at a 3-hour interval, and generated by assimilating available atmospheric observations every 12 hours into a forecast model with surface meteorological fields (e.g., precipitation and radiation), which are diagnosed from the model output (Dee et al., 2011; Flemming et al., 2015)."

132 Doesn't SRTM extend only to 60N? How is the downscaling achieved up to 71.875N?

The reviewer is correct that SRTM extends only to 60N. We use the GTOPO30 dataset for regions north of 60N. We have modified the text in lines 136-139 on pages 5 as follows:

"Meteorological inputs of near surface air temperature, relative humidity, surface pressure, and downward longwave radiation are downscaled by applying a lapse-rate and hypsometric adjustments using the 5 km Shuttle Radar Topography Mission (SRTM; between 60N and 60S) and the USGS Global 30 arc second elevation (GTOPO30; north of 60N) elevation datasets."

134 How are relative humidity, wind and longwave radiation downscaled? I don't think that this is described in Cosgrove et al. (2003).

The wind fields are not adjusted for topographic differences. This was erroneously stated in the manuscript and we have removed it. The lapse-rate correction of near surface air temperature, relative humidity, surface pressure, and downward longwave radiation is described in Cosgrove et al. (2003). The manuscript also had the incorrect

"Cosgrove et al." reference. It has been corrected in line 140 on page 5 to the following: Cosgrove, B. A., et al. (2003), RealâĂŘtime and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project, J. Geophys. Res., 108, 8842, doi:10.1029/2002JD003118, D22.

137 It is worth noting that Kumar et al. (2013) concluded that topographic adjustments to radiation should be included in models with resolutions finer than 16 km, but the adjustments are likely to be small at 5 km resolution.

Thanks for pointing this out. We have added this sentence in lines 143-146 on page 5 as follows:

"Kumar et al. (2013) demonstrated that these adjustments are particularly important for improving snow simulations over midlatitude domains in regions of complex topography and concluded that these adjustments should be included in models with resolutions finer than 16 km, but the adjustments are likely to be small at 5 km resolution."

154 Should this be "If the observation is more than 10% higher than the highest ensemble member, then the rank is set to 13"?

Thanks for this correction. We've changed this sentence in lines 160-161 on page 5, "If the observation is more than 10% higher than the highest ensemble member, then the rank is set to 13."

161 Conventionally, an integral would not have the "X=" in its limits. "P0 represents the observations occurrence" is not clear – it is a step function at the observed value.

We have replaced the equation 1 and its description as follows (in lines 168-170 on page 6), please find the attached file to see this response):

184 To be clear, CMC includes an estimate of SWE, but only snow depth observations are used in the CMC analysis.

We have replaced this sentence in lines 191-192 on page 6 as follows:

"Despite providing an estimate of SWE, in this analysis, we evaluate the CMC modeled snow depth fields since the CMC only uses snow depth observations in its analysis."

222 Give some context for what can be regarded as a "low" value of CRPS.

As noted in the text, CRPS can be considered analogous to mean absolute error, but for ensembles. It is hard to define a single value to define a "low" value of CRPS. We have modified the text in lines 227-228 on page 8 as follows to quantify what we mean by "low":

"Over most of the domain, including the northeast/Midwest U.S. and high plains, the CRPS values are low (0-100 mm), where a low (good) score indicates a small ensemble spread that agrees with SNODAS and UA data."

243 Some information should be given (earlier, or refer to S1) on how rain/snow partitioning schemes differ between the LSMs. Three of the four are identical.

We have added this information in lines 248-251 on page 8 as follows:

"These highly complex terrains have relatively high snowfall precipitation, and the large spread is partially due to different rain/snow partitioning schemes in each LSM. While Noah2.7.1, JULES, and CLSMF-2.5 use a simple temperature threshold of 0oC to distinguish rainfall and snowfall precipitation, Noah-MP3.6 includes a transition temperature range described in Jordan (1991) (see Table S1)."

We've also added Jordan's (1991) scheme for the Noah-MP3.6's precipitation partitioning method in the Table S1 (please find the attached file to see this response).

284 Raleigh et al. (2016) and Guenther et al. (2019) added uncertainty to single forcing datasets, rather than using multiple forcing datasets as here.

We have revised this sentence in lines 297-299 on page 10 as follows:

"For example, Raleigh et al. (2016) and Günther et al. (2019) showed the forcing data to be the primary driver of SWE uncertainty in their study, which used a single forcing

dataset with added uncertainty and focused on a limited number of relatively small sites mostly in mountainous terrains."

381 Is SWS averaged over the entire time period meaningful for comparing different snow classes, when it conflates the amount and duration of snow cover? How about looking at average annual maximum SWS?

Thanks for this comment. We agree that looking at the average annual maximum SWS is another interesting way. However, using annual maximum SWS in this study is not ideal because we're evaluating over a large spatial extent. Choosing the date of the annual maximum SWS is also challenging. The date of the annual maximum SWS would be different between snow classes and these are also different from the date of domain maximum SWS.

456 For "current operational capabilities", note that several countries now have operational limited area numerical weather prediction models with spatial resolutions on the order of 1 km.

We have revised this sentence in lines 469-471 on page 15 as follows:

"A primary goal of this study is to establish a baseline assessment of current global- or continental-scale operational capabilities and identify potential opportunities where improvements or SWE observations could inform both science and application needs."

Figure 2 caption

Average CRPS. Move the information "13 is more snow than all ensembles and 0 is less snow than all ensembles" from the figure to the caption

We have corrected this caption.

Figure 6 Does "without outliers" mean that outliers are omitted from the diagram (in which case, how is an outlier defined?) or that the ends of the whiskers are the SWE range?

The outlier is defined as a value that is more than 1.5 times the interquartile range away from the top or bottom of the box.

"without outliers" has been corrected in Figures 6 and 7 to "with outliers (defined as more than 1.5 times the interquartile range (between 25% and 75%)) omitted".

There is room to put the LSM and forcing dataset labels on the figure axis, removing the need for the reader to relate the figure legend to the boxes and removing the need for colour.

The plot has been modified.

If I understand correctly, five statistics (minimum, 25th percentile, median, 75th percentile and maximum mean annual average of SWE) have been calculated from just three data points for each LSM (mean annual average SWE for three different forcing datasets). Why not just show the points?

These statistics were calculated using the annual average of SWE for each year (i.e., interannual variability). For the LSM group, we used 8 annual averages of SWE (from 2009 to 2017) for three different forcing datasets (8*3). For the forcing dataset group analysis, 8 annual averages of SWE for four different (8*4) LSMs were used.

We have added these details in Figure 6 caption as follows: "Distribution of North America mean annual average of SWE (i.e., interannual variability), grouped by the LSMs and forcing datasets (e.g., the box of Noah-MP3.6 represents the distribution of mean SWE, averaged from Noah-MP3.6 runs with all forcing datasets; the box of MERRA2 represents the distribution of mean SWE, averaged from all LSM runs with MERRA2 forcing data). For the LSM group, we used 8 annual averages of SWE (from 2009 to 2017) for three different forcing datasets (total of 8*3). For the forcing dataset group, 8 annual averages of SWE for four different LSMs (total of 8*4) were used. The red line indicates SWE median; top and bottom of box are the 75th and 25th percentiles, and top and bottom of whiskers represent the maximum and minimum

SWE with outliers (defined as more than 1.5 times the interquartile range (between 25% and 75%)) omitted."

Figures 1, 3, 4 and 11 Rainbow colour scales are widely deprecated. If latitudes and longitudes are not going to be marked, remove the redundant grid lines. For the figures that use a divergent colour scale centered on white, the coastline would be a nice addition.

Thank you for the suggestion. Figures 1,3,4 and 11 have been revised.

Figures 7, 9 and 10 All of the colours in these figures are redundant.

Figures 7,9 and 10 have been revised.

References

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., Mcnally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N. and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, doi:10.1002/qj.828, 2011.

Essery, R., Rutter, N., Pomeroy, J., Baxter, R., Stahli, M., Gustafsson, D., Barr, A., Bartlett, P. and Elder, K.: SnowMIP2: An evalution of forest snow process simulation, Bulletin of the American Meteorological Society, doi:10.1175/2009BAMS2629.1, 2009.

Flemming, J., Huijnen, V., Arteta, J., Bechtold, P., Beljaars, A., Blechschmidt, A. M., Diamantakis, M., Engelen, R. J., Gaudel, A., Inness, A., Jones, L., Josse, B., Katragkou, E., Marecal, V., Peuch, V. H., Richter, A., Schultz, M. G., Stein, O. and Tsikerdekis, A.: Tropospheric chemistry in the integrated forecasting system of ECMWF, Geoscientific Model Development, doi:10.5194/gmd-8-975-2015, 2015.

Günther, D., Marke, T., Essery, R. and Strasser, U.: Uncertainties in snowpack simulations assessing the impact of model structure, parameter choice, and forcing data error on point-scale energy balance snow model performance, Water Resources Research, doi:10.1029/2018WR023403, 2019.

Jordan, R.: A one-dimensional temperature model for a snow cover: Technical documentation for SNTHERM.89., 1991.

Raleigh, M. S., Livneh, B., Lapo, K. and Lundquist, J. D.: How does availability of meteorological forcing data impact physically based snowpack simulations?, Journal of Hydrometeorology, 17(1), 99–120, doi:10.1175/JHM-D-14-0235.1, 2016.

Rutter, N., Essery, R., Pomeroy, J., Altimir, N., Andreadis, K., Baker, I., Barr, A., Bartlett, P., Boone, A., Deng, H., Douville, H., Dutra, E., Elder, K., Ellis, C., Feng, X., Gelfan, A., Goodbody, A., Gusev, Y., Gustafsson, D., HellströM, R., Hirabayashi, Y., Hirota, T., Jonas, T., Koren, V., Kuragina, A., Lettenmaier, D., Li, W.-P., Luce, C., Martin, E., Nasonova, O., Pumpanen, J., Pyles, R. D., Samuelsson, P., Sandells, M., SchäDler, G., Shmakin, A., Smirnova, T. G., StäHli, M., StöCkli, R., Strasser, U., Su, H., Suzuki, K., Takata, K., Tanaka, K., Thompson, E., Vesala, T., Viterbo, P., Wiltshire, A., Xia, K., Xue, Y. and Yamazaki, T.: Evaluation of forest snow processes models (SnowMIP2), Journal of Geophysical Research: Atmospheres, 114(D), D06111, doi:10.1029/2008JD011063, 2009.

Please also note the supplement to this comment:
https://tc.copernicus.org/preprints/tc-2020-248/tc-2020-248-AC1-supplement.pdf

Interactive comment on The Cryosphere Discuss., https://doi.org/10.5194/tc-2020-248, 2020.

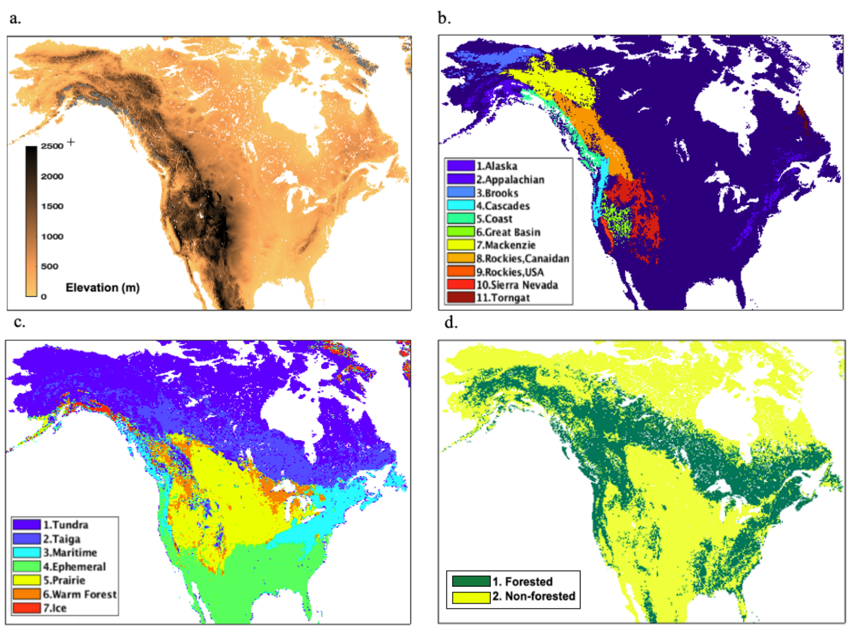

Figure 1: Snow Ensemble Uncertainty Project (SEUP) domain: (a) domain with terrain elevation. Grey areas indicate the excluded glacier regions, (b) individual mountain domains, (c) individual snow class domains, and (d) land cover classification used in this study.

Fig. 1.

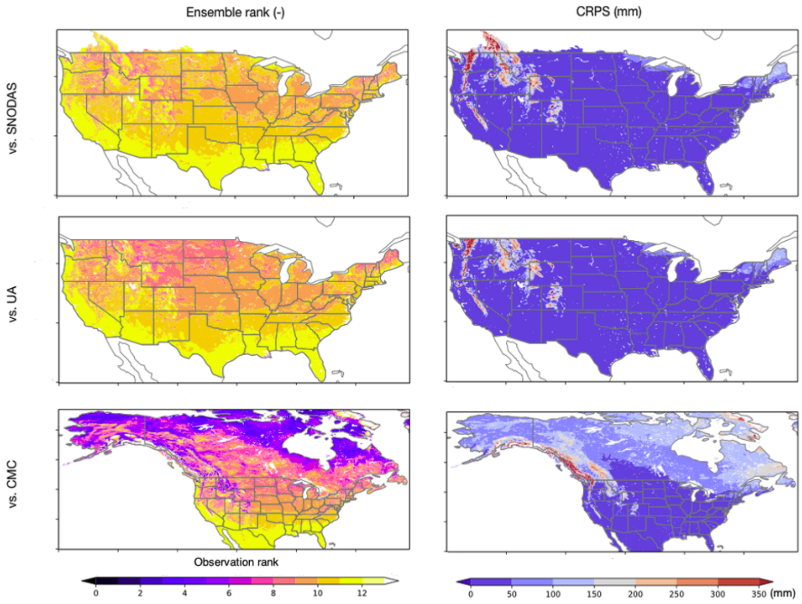

Figure 2: Maps of average ensemble rank (left column) and average Continuous Rank Probability Score (CRPS (mm); right column) from the SEUP ensemble compared to SNODAS (top row), UA (middle row), and CMC (bottom row). SWE is used for SNODAS and UA comparisons, whereas snow depth is used for CMC comparison. Ensemble rank represents the rank of the reference data within the SEUP ensemble. Rank 13 represents more snow than all ensembles and rank 0 is less snow than all ensembles. CRPS, which is the extension of mean absolute error to ensemble evaluation, provides a measure of the degree of agreement between the SEUP ensemble and the reference data.

Fig. 2.

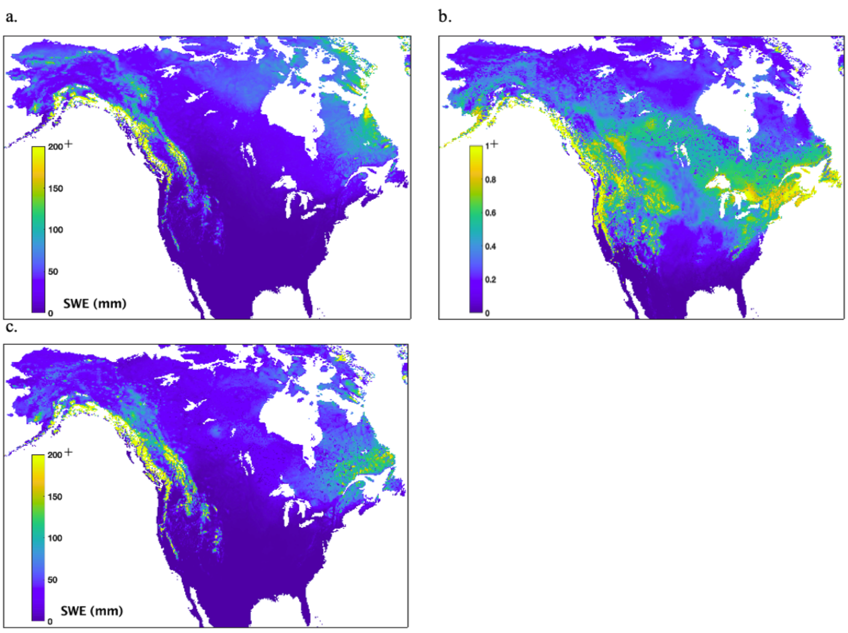

Figure 3: (a) Spatial distributions of ensemble mean SWE, (b) the coefficient of variation of ensemble mean SWE, and (c) the range of ensemble mean SWE. The ensemble mean SWE is computed by taking an average of 3 hourly SWE from 12 ensembles over the entire study time period (from 2009 to 2017).

Fig. 3.

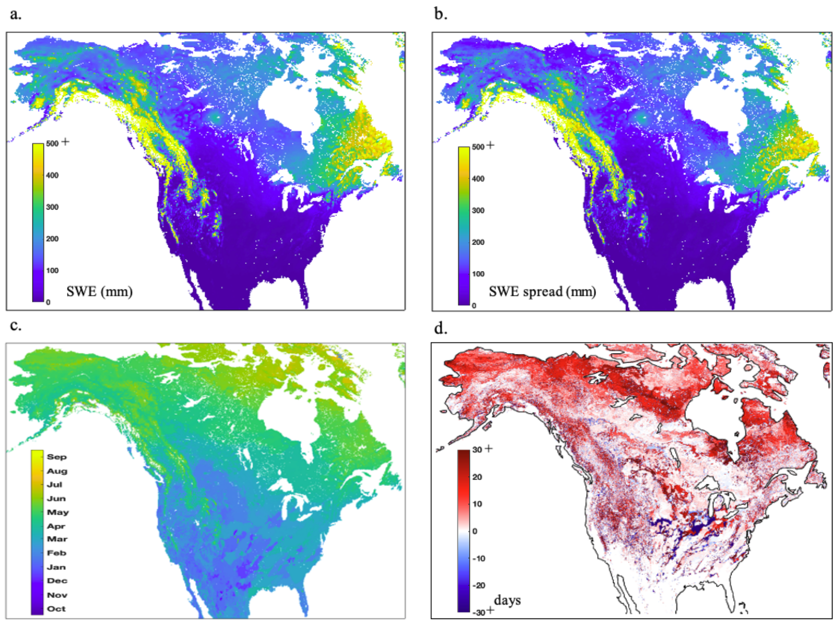

Figure 4: Spatial distributions of (a) the peak SWE amount, (b) the highest SWE spread amount, (c) the average day of year (DOY) with the highest ensemble SWE spread, and (d) the difference of average DOY between the highest ensemble SWE spread and the peak SWE (we are only showing/examining places where the DOY difference exist).

Fig. 4.

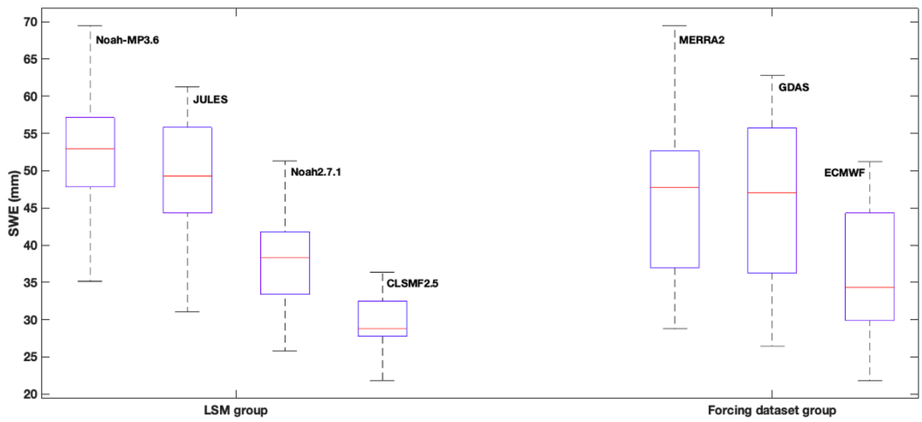

Figure 6: Distribution of North America mean annual average of SWE (i.e., interannual variability), grouped by the LSMs and forcing datasets (e.g., the box of Noah-MP3.6 represents the distribution of mean SWE, averaged from Noah-MP3.6 runs with all forcing datasets; the box of MERRA2 represents the distribution of mean SWE, averaged from all LSM runs with MERRA2 forcing data). For the LSM group, we used 8 annual averages of SWE (from 2009 to 2017) for three different forcing datasets (total of 8*3). For the forcing dataset group, 8 annual averages of SWE for four different LSMs (total of 8*4) were used. The red line indicates SWE median; top and bottom of box are the 75th and 25th percentiles, and top and bottom of whiskers represent the maximum and minimum SWE with outliers (defined as more than 1.5 times the interquartile range (between 25% and 75%)) omitted.

Fig. 5.

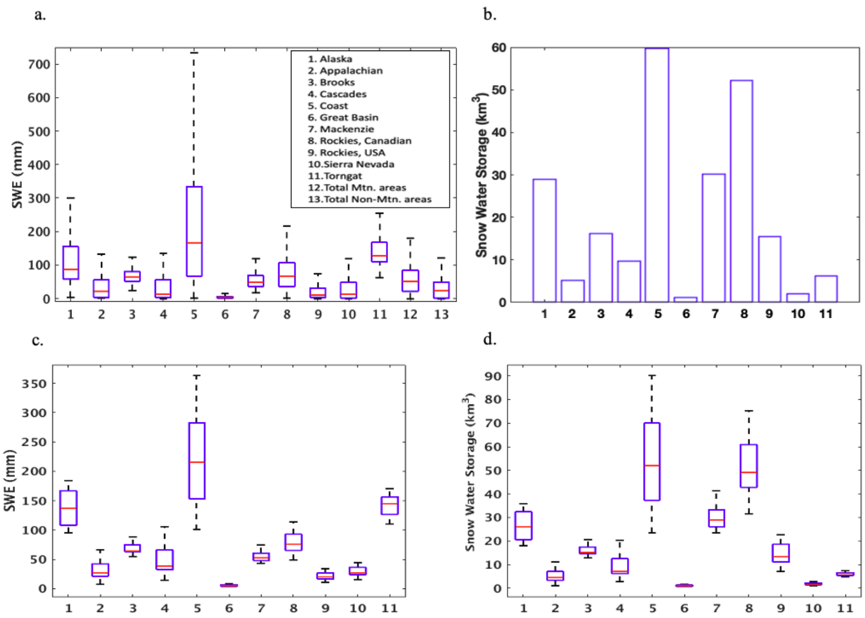

Figure 7: (a) Spatial variability of ensemble mean SWE (in millimeters) within each mountain range. Red line indicates SWE median; top and bottom of box are the 75th and 25th percentiles and top and bottom of whiskers represent max and min SWE with outliers (defined as more than 1.5 times the interquartile range (between 25% and 75%)) omitted. (b) Total snow water storage (SWS; in cubic kilometers) within each mountain range, computed from average of ensemble mean SWE over entire time period. The spread of ensembles for (c) domain and time averaged SWE and (d) time averaged SWS for different mountain range.

Fig. 6.

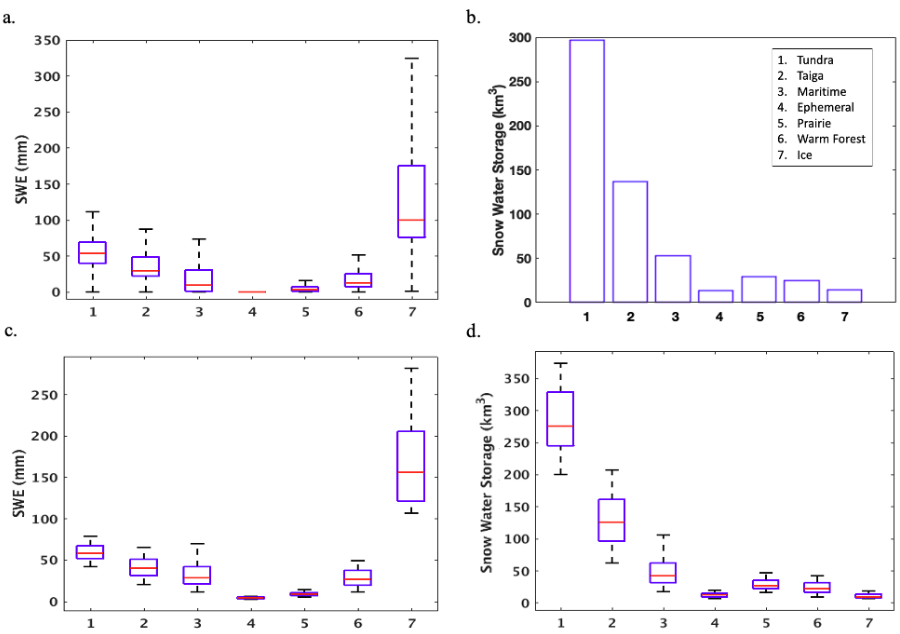

Figure 9: Same as Figure 7, but for each snow class.

Fig. 7.

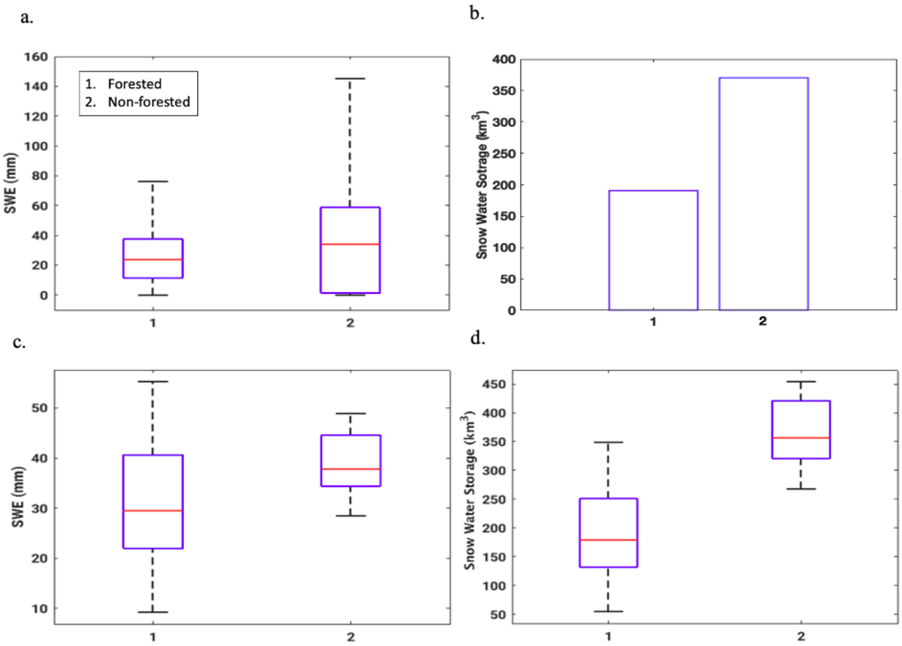

Figure 10: Same as Figure 7, but for forested areas vs non-forested areas.

Fig. 8.

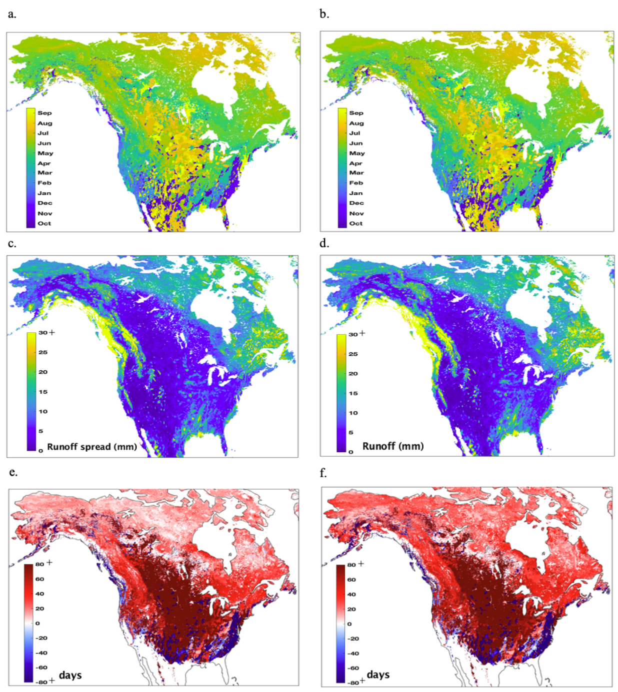

Figure 11: Spatial distributions of (a) the average day of year (DOY) with the highest ensemble *R* spread, (b) the average DOY with the peak *R*, (c) the highest *R* spread amount, (d) the peak *R* amount, (e) the difference of average DOY between the highest *R* spread and the highest SWE spread, and (f) the difference of average DOY between the peak *R* and the peak SWE.

Fig. 9.