Evaluation of long term Northern Hemisphere snow water equivalent products

Colleen Mortimer[1], Lawrence Mudryk[1], Chris Derksen[1], Kari Luojus[2], Ross Brown[1], Richard Kelly[3], and Marco Tedesco[4]

5 [1]Climate Research Division, Environment and Climate Change Canada, Toronto, Canada
[2]Finnish Meteorological Institute, Helsinki, Finland
[3]Department of Geography and Environmental Management, University of Waterloo, Canada
[4]Lamont Doherty Earth Observatory, Columbia University, Palisades NY; NASA Goddard Institute for Space Studies, New York, USA

10 *Correspondence to*: Colleen Mortimer (colleen.mortimer@canada.ca)

Abstract. Nine gridded northern hemisphere snow water equivalent (SWE) products were evaluated as part of the European Space Agency (ESA) Satellite Snow Product Intercomparison and Evaluation Exercise (SnowPEx). Three categories of datasets were assessed: (1) those utilizing some form of reanalysis (the NASA Global Land Data Assimilation System version 2 – GLDAS-2; the European Centre for Medium-Range Weather Forecasts (ECMWF) interim land surface reanalysis – ERA-Interim/Land and ERA5; the NASA Modern-Era Retrospective Analysis for Research and Applications version 1 (MERRA) and version 2 (MERRA2); the Crocus snow model driven by ERA-Interim meteorology – Crocus); (2) passive microwave remote sensing combined with daily surface snow depth observations (ESA GlobSnow v2.0); and (3) standalone passive microwave retrievals (NASA AMSR-E SWE versions 1.0 and 2.0) which do not utilize surface snow observations. Evaluation included validation against independent snow course measurements from Russia, Finland, and Canada, and product intercomparison through the calculation of spatial and temporal correlations in SWE anomalies. The standalone passive microwave SWE products (AMSR-E v1.0 and v2.0 SWE) exhibit low spatial and temporal correlations to other products, and RMSE nearly double the best performing product. Constraining passive microwave retrievals with surface observations (GlobSnow) provides comparable performance to the reanalysis-based products; RMSE over Finland and Russia for all but the AMSR-E products is ~50 mm or less, with the exception of ERA-Interim/Land over Russia. Using a seven-dataset ensemble that excluded the standalone passive microwave products reduced the RMSE by 10 mm (20%) and increased the correlation from 0.67 to 0.78 compared to any individual product. The overall performance of the best multi-product combinations is still at the margins of acceptable uncertainty for scientific and operational requirements; only through combined and integrated improvements in remote sensing, modeling, and observations will real progress in SWE product development be achieved.

30

1 Introduction

Temporally (~20–30 years) and spatially (~10–20 km) consistent estimates of daily SWE over seasonal snow covered land are required for many applications including climate model evaluation (Mudryk et al., 2018a), verification of seasonal forecasts (Sospedra-Alfonso et al., 2016), annual updates to climate assessments (e.g. Mudryk et al., 2018b; 2019), and determination of freshwater availability (Barnett et al. 2005; Clark et al. 2011). There is a growing number of gridded SWE datasets available to the snow community, but these are typically affected by one or more critical shortcomings related to:

1. Challenges in using point measurements: Meaningful spatially continuous information can be derived from surface observations for regions and time periods with a sufficiently dense observing network (Dyer and Mote, 2006; Brown and Derksen, 2013); as an alternative to snow depth, snowfall measurements can also be integrated (Broxton et al., 2016). However, both snow depth and snowfall measurements from single point locations are intrinsically limited by a lack of confidence in how they capture the landscape mean across coarse grid cells (Meromy et al., 2013), which is particularly problematic in areas of mixed forest vegetation, open areas prone to wind redistribution, and complex topography (most snow covered regions fall into at least one of these categories). Furthermore, there remain expansive alpine and northern regions with insufficient coverage by conventional observing networks (Brown et al., 2019).

2. Reliance on models driven by atmospheric reanalysis: Most modern reanalysis products include output of land surface variables such as SWE (Balsamo et al., 2015; Gelaro et al., 2017); alternatively the meteorology from these datasets can be used to force snow models (Brown et al., 2003; Brun et al., 2013). While these snow schemes are of varying complexity, they typically do not account for important processes such as snow-vegetation interactions and redistribution by blowing snow. In addition, the spread in SWE estimates among differing reanalyses is large: not only do differences between snow models introduce uncertainties (Mudryk et al., 2015), but model-based approaches are also sensitive to the precipitation forcing, which itself is challenging to validate in complex terrain and observation sparse regions (Lundquist et al., 2015; Henn et al., 2018). There may also be temporal inconsistencies in the forcing data related to changes in the observational streams assimilated in the reanalyses (Robertson et al., 2011).

3. Coarse spatial resolution: Whether derived from passive microwave satellite measurements or some form of model reanalysis, the typical resolution of existing gridded SWE datasets is 25 to 100 km. While synoptic scale patterns can be resolved at this resolution, spatial variability in SWE due to topographic and land cover heterogeneity is not adequately captured. Coarse resolution is a particularly critical limitation in alpine regions, which are masked out completely in some products (e.g. Takala et al., 2011). While this is a reasonable decision for some coarse resolution products, it nevertheless is a source of frustration for users. Coarse resolution also makes validation of SWE products challenging: the validation of large grid cells with single point measurements is conceptually unsatisfying and statistically non-robust. Regional climate models can provide higher resolution SWE information, but the computational cost related to complex atmospheric physics schemes

2

is, at least at present, a limiting factor in producing long time series (Wrzesien et al., 2018). There may be potential for cross-polarized C-band SAR to provide high spatial resolution snow depth information in mountain areas (Lievens et al., 2019), but these estimates currently lack a physical explanation. Since cross-polarized C-band SAR data are only available since the launch of Sentinel-1A in 2014, there is limited potential to provide climate-relevant time series.

4. Inability of remote sensing data to constrain uncertainty: The number of purely satellite derived SWE datasets is limited, and uncertainty in standalone passive microwave retrievals can be high (Kelly et al., 2003). The combination of passive microwave and surface snow depth measurements (within the GlobSnow product; Takala et al., 2011) was shown to yield performance similar to snow models driven by atmospheric reanalysis (Mudryk et al., 2015), but it relies heavily on background fields and constraints generated from re-gridded surface snow depth observations (Pulliainen, 2006). The microwave remote sensing community has made great progress in understanding and quantifying error sources (snow microstructure, deep snow, wet snow, vegetation, lake ice), all of which are exacerbated by the coarse resolution of passive microwave measurements (Foster et al., 2005; Durand et al., 2011; Lemmetyinen et al., 2011; Durand and Liu, 2012).

Previous studies have demonstrated the potential for using multi-product SWE ensembles in order to improve estimates of observed snow-related quantities (e.g. SWE and snow cover fraction, integrated snow mass, snow cover extent and trends in these quantities) and to constrain uncertainty (Mudryk et al., 2015, 2017, 2018a; Krinner et al., 2018). The intent in such a strategy is that uncorrelated errors between products of the same type average out, so the limitations and shortcomings of a given class of products offset one another. Ideally, such ensembles would draw from as many types of products as possible and use multiple versions of each type of product. To date, these ensembles have relied heavily on models driven by atmospheric reanalyses and include only a single dataset (GlobSnow) that utilizes remote sensing. While SWE or snow depth products can be derived using InSAR techniques (Deeb et al., 2011) and airborne LiDAR data (Painter et al., 2016), such products are only available for regionally and temporal limited domains. Hence, the long time series of passive microwave measurements provide the most straightforward pathway to increase the use of satellite data within observational SWE ensembles. Before existing passive microwave derived SWE products can be included, however, an assessment is needed because of markedly different climatological patterns (Fig. 1; discussed further in Sect. 3.1). The specific objectives of this study are to evaluate gridded Northern Hemisphere SWE products by (1) *validation* with independent surface observations, and (2) *intercomparison* through calculation of the spatial and temporal correlations in SWE anomalies.

2 Datasets and methods

2.1 Gridded SWE products

We evaluate three categories of northern hemisphere gridded SWE products: (1) standalone passive microwave retrievals (AMSR-E SWE v1.0 and v2.0); (2) passive microwave estimates combined with surface snow depth observations (GlobSnow

v2.0), and (3); products which utilize some form of reanalysis (Crocus, GLDAS-2, ERA-Interim/Land, ERA5, MERRA, MERRA2). A summary of these nine SWE datasets is provided in Table 1. All the products provide SWE directly and are available at daily or sub-daily frequency. For the four products available at sub-daily frequency, we either obtained daily mean

95 versions directly from the product's distribution site (MERRA, MERRA2) or sampled a consistent sub-daily snapshot for each calendar day (ERA-Interim/Land, ERA-5) which we consider to be representative of the daily mean value. The analyses described subsequently in Section 2.3 were conducted for the period 2002 –2010 to maximize temporal overlap between products.

1. *Standalone passive microwave*: The NASA AMSR-E SWE v1.0 product
100 (https://nsidc.org/data/AE_DySno/versions/2, Tedesco et al., 2004) is described in Kelly (2009) and evaluated in Tedesco and Narvekar (2010). Brightness temperature thresholds are utilized to identify shallow and non-shallow dry snow areas, with the depth of shallow snow set to 5 cm (Kelly et al., 2003). SWE is retrieved based on a brightness temperature difference approach (37-19 GHz; based on the original formulation of Chang et al. (1990)) with enhancements to account for the influence of vegetation, to address deeper snowpacks (through the use of 10 GHz
105 measurements), and to consider the dynamic influence of snow grain size (based on the assumption that as snow depth increases, the depth average grain size increases). Snow depth is converted to SWE using the snow climate classification of Sturm et al. (1995) and snow density climatologies from Brown and Braaten (1998) and Krenke (2004). Building on the v1.0 AMSR-E SWE product, NASA's current v2.0 AMSR-E SWE algorithm utilizes an artificial neural network, snow emission modelling, and climatological snow depth data for the estimation of snow
110 depth, and the detection of dry versus wet snow conditions (Tedesco and Jeyaratnam 2016). Snow density maps based on Sturm et al. (2010) are employed for conversion of retrieved snow depth to SWE. Unlike the GlobSnow approach described next, both NASA AMSR-E SWE algorithms are self-contained and do not rely on any external temporally variable snow measurements.

2. *Synergistic passive microwave + in situ:* The European Space Agency GlobSnow v2.0 SWE product (data available
115 at www.globsnow.info) is based on a retrieval method first described in Pulliainen (2006). The approach evolved from standalone passive microwave algorithms (so it also relies on 19 and 37 GHz measurements), but the retrieval also integrates daily surface snow depth measurements. First, daily climate station snow depth observations are kriged to form a continuous background field independent of passive microwave retrievals. This first guess snow depth field is used as input to two iterations of forward microwave emission model simulations, one to estimate grain size, the
120 second to estimate snow depth (Takala et al., 2011). A temporally and spatially fixed snow density value of 0.24 g cm^{-3} is applied to convert snow depth to SWE. Alpine areas are excluded due to the known limitations of this technique in regions with complex sub-grid topographical heterogeneity (Takala et al., 2011).

3. *Land surface models and reanalysis*: Six SWE datasets derived from combinations of models driven by reanalysis meteorology were used for comparison with the passive microwave products: the NASA Global Land Data

4

125　　　Assimilation System version 2 – GLDAS-2; the European Centre for Medium-Range Weather Forecasts (ECMWF) interim land surface reanalysis – ERA-Interim/Land; and ECMWF ReAnalysis version 5 - ERA5; the NASA Modern-Era Retrospective Analysis for Research and Applications, version1 – MERRA, and version2 - MERRA2; the Crocus snow model driven by ERA-Interim meteorology – Crocus. We refer to these datasets as *snow analyses*. It is important to note that spread among the snow analyses does not only depend on differences in the forcing data; in fact, a

130　　　substantial portion of the spread stems from differences in the complexity and parametrizations of their respective snow schemes (see Mudryk et al., 2015). For example, both Crocus and ERA-Interim/Land use the same forcing data but employ different land models with different snow schemes which yield significantly different validation results (Section 3.2). The impact of snow depth observations also differs between reanalysis products. Snow depth observations are directly assimilated into ERA5. For ERA-Interim/Land, however, only the forcing meteorology

135　　　includes explicit assimilation of point snow depth measurements (the SWE produced by ERA-Land does not). Therefore, for ERA-Interim/Land, the use of snow depth information is one step removed from the final SWE estimates compared to ERA5, although the assimilation of snow information impacts variables such as lower tropospheric temperatures which obviously have an indirect impact on snow.

140　**2.2 Snow course data**

The suite of gridded SWE products described in Section 2.1 are validated with a network of in situ snow course measurements from multiple national and regional agencies. These data consist of manual gravimetric snow measurements made at multiple locations along a pre-defined transect that are averaged to obtain a single SWE value for a given snow course on a given day. Measurements are collected along the same transect multiple times each snow season. By averaging multiple samples along a

145　transect, the resulting SWE measurement provides better representation of sub-grid scale variability than a single point measurement and so is more suitable for evaluation of SWE at the scale of the gridded products. These snow course data are fully independent of the point snow depth measurements assimilated into GlobSnow and ERA5. Transect length, number of samples collected along each transect, and sample aggregation methods differ among reporting agencies as described below.

Russia has a long-term snow course network located near 517 meteorological stations (Bulygina et al., 2011). The snow survey

150　transects extend for 1 to 2 km in open areas, and 500 m at forested sites. Measurements are made every ten days when at least half of the visible area around a station is snow-covered, except at forested sites where measurements are made once per month prior to 20 January. Sampling frequency is increased to five days during the spring snow melt season. The Finnish snow course network, maintained by the Finnish Environment Institute (SYKE), consists of approximately 200 transects distributed across the country. Measurements are conducted monthly around the 15th of each month, with a subset of snow courses also measured

155　at the end of each month. Each snow course is 2 to 4 km long and extends through variable land cover consistent with the surrounding landscape (Haberkorn, 2019).

The Canadian snow course data are a recently updated collection pooled from a series of national and regional networks described in Brown et al. (2019). There is no comprehensive national strategy in Canada to obtain a spatially representative collection of snow course measurements. Snow courses are maintained by various jurisdictions resulting in a spatially heterogeneous sample distribution heavily biased towards population centres. For 2002 through 2010, there were more than 1000 unique snow course locations across Canada with varying sampling frequency. Measurements are typically made around the 1st and 15th of each month during the snow season (November to April). Snow courses are roughly 150 to 300 m long consisting of five to ten sampling locations (Brown et al., 2019). While the network density is sparse across Canada and the transects are shorter in length than the Russian and Finnish data, previous analysis suggests the measurements still capture reasonable landscape mean values (Neumann et al., 2006).

Because snow course measurements are only acquired during the snow season and zero SWE values are not reported in a consistent manner across all jurisdictions, zero SWE is not a reliable measure of snow-free conditions. All zero snow course observations were therefore removed prior to spatiotemporal aggregation (Sect. 2.3); SWE product zero values were also excluded. Finally, it is difficult to attach specific uncertainty values to the snow survey measurements because non-standard sampling tools are used between snow courses (e.g. no consistent snow corer diameter). Full discussion of snow course measurement protocols and instruments is available elsewhere (Goodison et al., 1981; Brown et al., 2019; Haberkorn et al., 2019), but there is no doubt that uncertainty associated with the individual measurements (± approximately 5%; Brown et al., 2019) is overwhelmed by uncertainty in how the snow course measurements represent the landscape mean at the scale of the gridded SWE products.

2.3 Validation and intercomparison methods

We assess the gridded products in two separate analyses conducted for the snow season (defined here as November – April (NDJFMA)) from November 2002 – April 2010. The first assessment is termed a *validation* because it evaluates each gridded product using the snow course data as a measure of ground truth. While the relative sparseness of the snow course measurements limits the assessment's spatial and temporal completeness, it nonetheless considers a broad range of snow classes covering both Northern Hemisphere continents (Fig. 2) and considers seasonal variability from November through April over eight years of interannual variability. We are unaware of any other studies that have evaluated the breadth of products examined here with similarly representative data and with comparable spatial and temporal coverage. The second assessment is termed an *intercomparison* and is similar to the analysis performed in Mudryk et al. (2015). This second type of analysis is spatially and temporally complete (across the seasons and period considered). We use this analysis as it provides a more complete measure of differences among the products and is able to more readily discern differences and discontinuities between products than the validation analysis (see results regarding ERA5 in Sect. 3.3).

For the *validation* analysis, SWE product grid cells must be matched in both space and time with the snow course measurements. To achieve this, snow course observations from Canada and Finland were first grouped into bi-weekly periods using a 16 day window centred on the 1st or 15th of each month. Likewise, over Russia, observations were grouped into ten-day periods centred on the typical measurement dates (10th, 20th, 30th of each month). For each temporal grouping, snow course measurements falling within a given 25 x 25 km EASE grid cell (Brodzik et al., 2012) were averaged together, thereby forming a gridded snow course field (Fig. 2). Roughly 30% of these snow course grid cells had two or more separate snow courses which were averaged together while the remaining 70% had only one snow course observation. Grouping the snow course data had the largest impact over Canada and Russia where 35% and 20% of grid cells, respectively, had multiple snow courses. Although Finland's snow course network is representative of the landscape's different snow-climate classes(Sturm et al., 1995), in Canada, and to a lesser extent over Russia, tundra environments which are often remote, are under-sampled while maritime and alpine snow types are oversampled (Fig. 2).

For the validation analysis, we included all nine products in Table 1, to consider the range of available products and show the difference in performance between subsequent product generations (e.g. MERRA to MERRA2). For a given measurement date, each EASE grid cell with snow course data was paired with corresponding SWE values from each of the nine gridded products. The paired SWE values correspond to the grid cell at each product's native resolution that intersects with the centroid of the snow course EASE grid cell. In order to fairly compare how the gridded products perform against one another, only snow course data from EASE grid cells with corresponding paired values from all nine of the SWE products were analysed. This means that regions of complex topography are implicitly excluded from the validation analysis because they are masked in GlobSnow. Analyses were conducted for the snow season only (November – April). Bias and root mean squared error (RMSE) were calculated for each product-snow course pair and then averaged over the full November 2002 – April 2010 time period; correlation was calculated from all data pairs for the November 2002 – April 2010 period. To understand the influence of seasonality on product performance, bias, RMSE and correlation were also computed across all years for each biweekly period (10 day period for Russia). Validation statistics were calculated separately for each national snow course dataset in order to separate any sensitivity to differences in snow course measurement protocol and sample distribution. Finally, to determine the influence of SWE magnitude on product performance, all snow course-product SWE pairs were binned into 10 mm increments according to the snow course SWE. For each 10 mm increment the average product SWE was plotted against the bin midpoint.

The *intercomparison* analysis does not consider the snow course measurements, only the nine gridded SWE products. For this analysis, daily SWE from each product was interpolated to a regular 1° x 1° longitude–latitude grid. SWE values over glaciers and large lakes were excluded based on the MERRA land fraction mask (consistent with Mudryk et al., 2015). To determine the strength of agreement among datasets we use three metrics, all applied to SWE or snow mass anomalies (i.e. with the seasonal cycle removed). We only consider anomalies due to the results from Mudryk et al. (2015), which demonstrated that while different snow products can have substantial spread in their climatological snow estimates, one can and should expect a

reasonable degree of agreement in their interannual and intraseasonal variability. First, we considered the correlation between each product's time series of daily Northern Hemisphere snow mass anomalies (SWE integrated over the entire Northern Hemisphere land area). Each product's time series was calculated using its respective climatology (determined for the snow season over the November 2002–April 2010 period). A correlation coefficient was calculated for each pair of datasets by correlating the two snow mass anomaly time series cropped to the snow season (November–April) over the April 2002 – November 2010 period. Secondly, we considered correlations between the patterns of anomalous SWE fields. Daily SWE anomalies were calculated for each product using its respective climatology. For each dataset pair, we calculated the daily pattern correlation between the two anomalous SWE fields and averaged the sequence of correlation values over the snow season for the 2002–2010 period. These first two metrics are bulk measures of agreement, specifically in their estimates of Northern Hemisphere snow mass anomalies and the average agreement of their pattern correlations. Finally, we also considered 'local' correlation maps of anomalous SWE. As above, we calculated daily anomalous SWE fields. Then for each dataset pair, we calculated the correlation coefficient between the daily time series of anomalous SWE at each location on the 1° x 1° grid. The correlation calculation only considers the snow season (November – April) over the November 2002 – April 2010 period. This third metric allows us to consider which regions agree more and less among the various products.

3 Results

3.1 Climatology

There is notable disagreement in the climatological SWE distribution over the Northern Hemisphere land area between standalone passive microwave products and the other data sources (Fig. 1). The pattern of high and low SWE between western and eastern Siberia is reversed for the snow analyses and GlobSnow versus the two AMSR-E algorithms. This inconsistency across Eurasia was also identified in analysis of older versions of passive microwave derived SWE data (e.g. Rawlins et al., 2007), reanalysis, and climate model simulations (see Fig. 2 in Clifford et al., 2010). The AMSR-E products also fail to capture a pronounced region of high SWE in eastern Canada present in the other datasets. The GlobSnow climatology is in close agreement with the snow analyses, particularly over Eurasia. The snow analyses and GlobSnow also agree with other SWE climatologies derived from other sources covering different time periods and thus not included in this study (see Brown and Mote, 2009; Liston and Hiemstra, 2011).

The difference in climatological SWE patterns is not solely due to the well-documented systematic underestimation in passive microwave retrievals when SWE exceeds 150 mm (Markus et al., 2006). Eastern Siberia is a low winter-season precipitation environment with very cold surface temperatures. These are ideal conditions for a thin, low density snowpack (see Liston and Hiemstra, 2011), likely composed primarily of faceted snow grains due to kinetic metamorphism, as seen in the Canadian Arctic (Derksen et al., 2014) and Alaskan North Slope (Hall, 1987). Thin snow composed of large faceted grains results in

exaggerated scattering relative to the amount of SWE (Hall et al., 1991), hence the comparatively large SWE estimates for the standalone passive microwave products.

The reason the standalone passive microwave products fail to capture higher SWE in western Siberia, Russia, northern Europe, and eastern Canada is less clear, but may be related to weaker scattering signatures from smaller grained and deeper snow, which is further masked by microwave emission from forest cover. The ability of GlobSnow to retain sensitivity to deeper snow than the AMSR-E products is due to the assimilation of daily surface snow depth observations which work to 'nudge' the retrievals to higher values (Pulliainen, 2006). In observation-sparse regions such as northern Quebec, the GlobSnow estimates are more heavily weighted to the passive microwave retrievals, which increases uncertainty in these areas (Larue et al., 2017; Brown et al., 2018) compared to forested, deep snow regions with a dense observation network such as Finland (Takala et al., 2011).

3.2 Comparison with surface measurements

The nine gridded SWE datasets were compared to Canadian, Finnish, and Russian snow course measurements for all snow seasons (November–April) over the November 2002 – April 2010 period. A summary of the validation results is provided in Fig. 3.

All products exhibit weaker skill over Canada, where the RMSE for all products is roughly twice that of Finland and Russia. Larger absolute bias and RMSE over Canada may be attributed, in part, to a higher average SWE since the mean SWE of all snow course grid cells used for validation (Sect. 2.3) is 143 mm in Canada compared to 96 mm in Finland and 76 mm in Russia. However, the RMSE, expressed as a percentage of the mean observed SWE (of grid cells used in the analysis) is still higher over Canada for almost all products, indicative of poorer relative performance. The exception is ERA-Interim/Land which has poorer relative performance over Russia than over either Finland or Canada, consistent with the product intercomparison from Mudryk et al. (2015).

Crocus had the smallest bias over both Canada (-22 mm) and Russia (-2.3 mm); Crocus and ERA5 had the strongest correlations over Canada (~0.7). ERA5 had the lowest RMSE and strongest correlation over Finland (33 mm, 0.8; tied with ERA-Interim/Land) and Russia (38 mm, 0.8), and the lowest bias over Finland (0.8 mm). Performance of the standalone passive microwave products (AMSR-E) is noticeably weaker for all regions and validation statistics (with the exception of bias over Russia). RMSE for the standalone passive microwave products is nearly double that of the best performing product for both Finland and Canada, with slightly better results over Russia. For Finland and Russia, bias ranged between ±15 mm for all datasets except ERA-Interim/Land (> +20 mm) and the standalone passive microwave products and MERRA2 over Russia (+17 mm). Over Canada, bias ranged from -23 to -51 mm for all but the AMSR-E products (bias of -78 to -90 mm). For all regions, correlation coefficients for all but the standalone passive microwave products were ~0.5 and greater. The

9

AMSR-E products exhibited lower or even negative correlations with snow course measurements for all three reference datasets.

We find that among GlobSnow, Crocus, ERA-Interim/Land, ERA5, GLDAS-2, MERRA and MERRA2 no individual product consistently performs best with respect to the RMSE, bias, and correlation statistics *across all regions*. This is an important

285 finding, as it shows no clear advantage to using a single type of snow analysis, whether it is remote sensing combined with surface observations, an external snow model driven by reanalysis meteorology, or the land surface schemes within reanalyses. With higher RMSEs, greater bias, and weaker correlations relative to the other seven datasets, this assessment shows the standalone passive microwave algorithms do not perform in a comparable fashion to the other products.

To determine the influence of SWE magnitude on product performance, all three reference snow course datasets were binned

290 into 10 mm increments for comparison with the gridded SWE estimates (Sect. 2.3, Fig. 4). Crocus and MERRA perform similarly, with reasonable agreement up to about 150 mm of SWE and a tendency to underestimate SWE for deeper snow and overestimate SWE for shallow snow. MERRA2 behaves in a similar fashion, but slightly overestimates SWE below ~150 mm. The performance of GLDAS-2 and ERA5 is similar to Crocus and MERRA except that they both underestimate SWE across a larger range of reference values (>100 mm), consistent with the negative bias in Fig. 3b (with the exception of ERA5 over

295 Finland). GlobSnow overestimates SWE up to ~100 mm and underestimates above ~130 mm while ERA-Interim/Land overestimates SWE up to ~180 mm, consistent with the positive bias over Russia and Finland (Fig. 3b) The AMSR-E v1.0 product exhibits low sensitivity to SWE, especially for values >70 mm and overestimates low SWE values. Better results were found for the newer AMSR-E v2.0 product, although the retrievals plateau at about 100 mm, and show no sensitivity to further SWE increases.

300 To quantify the influence of seasonality on product performance, validation statistics (RMSE, bias, correlation) were computed at a bi-weekly time step (10 days for Russia) for 2002 through 2010 (Sect. 2.3). Figure 5 shows the monthly evolution from November through April over Russia and provides insight into both the seasonal evolution of product-specific uncertainty and the spread in uncertainty between products. In general, RMSE and bias magnitude both increase over the course of the snow season. Early in the snow season, the RMSE and bias magnitudes are low because snow is shallow, although even small errors

305 can produce high relative RMSE. As SWE increases through the snow accumulation season, the RMSE and the spread in RMSE between products increases. While not true for every product, bias also tends to become increasingly negative over the course of the snow season. By the end of the snow season, inter-product spread in RMSE and bias are at a maximum. Peak uncertainty late in the season is driven by cumulative errors over the entire season, differences in the timing of snow melt onset, and different melt rates. Whereas the RMSE and bias evolve over the course of the snow season, the magnitude of

310 correlation for all but the AMSR-E products is stable. This is an encouraging result as it indicates that SWE anomalies should be reasonably realistic throughout the season, even if climatological amounts of SWE differ strongly between analyses. A similar seasonal evolution of product specific uncertainties is observed for both Finland and Canada (not shown).

The analyses summarized in Figures 2–4 indicate that Crocus, MERRA and ERA5 perform slightly better than the other reanalysis-based products and GlobSnow, while the two AMSR-E products perform substantially worse. To what extent do these conclusions suggest that one should choose a single gridded SWE product as the 'best' dataset? We address this question by analyzing how the error statistics (RMSE and correlation) of multiple-product combinations compare to those of individual products. Such multi-product SWE ensembles have previously been employed to charaterize uncertainty (e.g. Mudryk et al., 2015, 2017, 2018a; Krinner et al., 2018). Here we demonstrate that such ensembles also tend to improve overall accuracy. The two AMSR-E products were excluded from this analysis because of the low correlation with snow course measurements as illustrated in Figs. 3c, 4h, 4i, and 5c. Further, for this analysis, we did not separate error statistics by country (Russia, Finland and Canada are considered on aggreate). Figure 6a confirms the conclusion that Crocus, MERRA and ERA5 perform slightly better than the other products since the average of all product combinations that involve those particular snow analyses have lower RMSE and higher correlation than averages involving the remaining products. However, we find that combinations of products often have a lower RMSE and higher correlation than individual products. For example, *any* possible combination of two or more products has improved RMSE and correlation compared to GLDAS-2, GlobSnow or ERA-Interim/Land considered individually (not shown explicitly). For MERRA and MERRA2, more than 90% of all possible combinations of two or more products have improved RMSE and correlation compared to the single product. For Crocus, approximately 40% of product combinations have improved RMSE and correlation [than the single product] while for ERA5, 70% of all possible product combinations have lower RMSE and 35% have higher correlation. This tendency for multi-product combinations to have improved accuracy is demonstrated generally in Figure 6b. As the number of products included in a multi-product combination increases, the correlation improves and the RMSE decreases with the lowest RMSE and highest correlation attained when all seven products are combined. This improvement in accuracy suggests that, to some extent, each product has randomized errors which are averaged out by considering multiple products. Because the RMSE of even the best performing products is at the margins of acceptable uncertainty for operational (<15%; Rott et al., 2010; Larue et al. 2017) and scientific (10-25%; Derksen and Nagler, 2019) requirements, the increase in accuracy represents a simple method to yield performance gains.

3.3 Correlation Analysis

To determine the strength of agreement among datasets, temporal and spatial correlation analysis was performed as described in Section 2.3. In preparing the datasets for intercomparison, a very strong negative trend since 1980 was found for ERA5 snow mass. This is driven by a stepwise discontinuity introduced by the assimilation of satellite derived binary snow/no-snow estimates starting in 2004 (Patricia de Rosnay, personal communication; Fig. 7). While this change addressed a positive snow extent bias during the melt season (e.g. Orsolini et al., 2019), it renders the raw ERA5 snow mass time series unsuitable for climate analysis. We therefore considered ERA5 separately from the other snow analyses (Crocus, GLDAS-2, MERRA2, ERA-Interim/Land) for the intercomparison analysis. Results obtained substituting MERRA with MERRA2 were similar so

only those including MERRA2 are presented. In the subsequent analysis R4 refers to a suite of four products (Crocus, GLDAS-2, MERRA2 and ERA-Interim/Land) that rely on reanalysis in some way.

Each of the products in the R4 suite exhibit moderately strong spatial and temporal correlations with each other (Fig. 8). The correlations, ranging between 0.5 and 0.7, represent the average of the six pairwise combinations of these four products. The agreement among these four datasets is consistent with the expected coherence of their forcing meteorologies and the relative influence of land model and meteorological forcing on hemispheric scale snow mass previously established by Mudryk et al. (2015). While Figure 8 illustrates that the spatial patterns of ERA5 snow mass anomalies are comparable to those of GlobSnow and the R4 products, the stepwise discontinuity in its climatology lowers the correlation of its snow mass time series. It is possible to correct for this discontinuity in an *ad hoc* manner by adjusting the snow mass starting in the fall of 2004 by the difference in the climatology before and after the discontinuity. Applying this correction yields correlation values more in line with those seen among the R4 products and GlobSnow (dashed symbol in Fig. 8). For the snow analyses and GlobSnow, the mean pattern correlation is lower than the corresponding temporal correlation of total snow mass (Fig. 8). This may be due to the presence of opposite-signed spatial biases that cancel when spatially aggregated into a snow mass time series. In contrast to the snow analyses and Globsnow, there is a lack of temporal and spatial correlation between the AMSR-E products and the R4 datasets. Spatially, this is an expected result given the differences in climatological SWE patterns shown in Fig. 1. The weak temporal correlation means the snow mass anomalies do not evolve in-phase with the other products as the snow season evolves.

Further insight is gained through the calculation of correlation maps among groups of datasets (Fig. 9), where temporal correlations of daily SWE are calculated analogous to Northern Hemisphere snow mass but for each grid cell. As expected, the reanalysis datasets are strongly correlated to each other (R4-R4 and R4-E5 in Fig. 9). Correlations between GlobSnow and the R4 products are strong across most snow covered regions of the Northern Hemisphere (GS-R4), with the exception of parts of Arctic Canada and the ephemeral snow zones of both North America and Eurasia (note that alpine areas are masked in the GlobSnow product). As noted earlier, the performance of GlobSnow is closely tied to the density of snow depth observations used as inputs to the retrievals (Larue et al., 2017; Brown et al., 2018) which likely contributes to the low correlations in parts of Arctic Canada where there are relatively few observations. The NASA AMSR-E v1.0 dataset exhibits very weak anomaly correlations with the R4 datasets (N1-R4), and even negative correlations over the boreal forest of North America and parts of central and eastern Siberia. The AMSR-E v2.0 algorithm shows improved anomaly correlations over eastern Siberia (N2-R4; likely by better accounting for the combination of shallow snow and large snow grains found in this region (Tedesco and Jeyaratnam, 2016)) and the boreal forest of North America, although correlations remain weak over the remainder of the snow-covered northern hemisphere.

4. Conclusions and Discussion

In this study, we compared three types of northern hemisphere gridded SWE products: (1) those utilizing some form of reanalysis (Crocus, ERA-Interim/Land, ERA5, GLDAS-2, MERRA, MERRA2); (2) passive microwave remote sensing combined with surface observations (GlobSnow v2.0); and (3) standalone passive microwave retrievals (AMSR-E v1.0 and v2.0). There is past evidence of acceptable algorithm performance for standalone passive microwave products, particularly in open environments with relatively shallow snow (Derksen et al., 2004; Vuyovich et al., 2014), or when SWE retrievals are converted to snow cover extent (Brown et al., 2010). At the continental scale, however, the standalone AMSR-E SWE products have stark differences in climatological SWE patterns compared to other available products (see Fig. 1).

Evaluation against snow course measurements from Russia, Finland, and Canada show higher RMSE and bias, and lower correlation for standalone passive microwave products compared to the seven other datasets (Fig. 3). While uncertainty for all products tends to increase with deeper snow, this is a critical issue for the AMSR-E products because of pronounced negative bias even at relatively low SWE values (<100 mm; Fig. 4 and 5). Although there is no single product that consistently performs best over all regions with respect to bias, RMSE and correlation, Crocus and ERA5 do perform best across the range of snow conditions captured by the validation dataset. However, while a particular product may outperform others over some regions, this is no guarantee that it will do so everywhere, so we are not recommending any one product. Furthermore, we have demonstrated that averaging multiple products together tends to lead to additional accuracy improvements (Fig. 6), while as exemplified by ERA5, a single product may have properties which lend itself to one type of analysis but make it unsuitable for others.

Correlation analysis performed with respect to both space and time shows consistent behaviour with strong statistical agreement among the six reanalysis-based products and GlobSnow (consistent with Mudryk et al., 2015), which clearly benefits from the ingestion of daily surface snow depth data into the retrievals compared to the standalone passive microwave datasets. ERA5 also assimilates point snow depth observations into a state-of-the-art assimilation system and yields excellent validation results. The slightly stronger validation for ERA5 compared to GlobSnow suggests the impact of the ERA5 assimilation system, which ingests multiple data streams, improves the SWE estimates more than the impact of passive microwave remote sensing on the GlobSnow retrievals (which also assimilates point snow depth observations). However, it is important to highlight that the validation results do not convey that the raw ERA5 snow mass time series contains a significant discontinuity in 2004, caused by an abrupt change to assimilate satellite derived snow extent information. So, while ERA5 may provide one of the better SWE estimates for instantaneous applications like numerical weather prediction, the data are unsuitable (at least in an uncorrected form) for climate analysis.

As with any continental-scale evaluation, our results may (or may not) apply to small regions or local domains, and the validation results do not apply to alpine areas which contribute a large proportion (~30%; Wrzesien et al., 2019) to the total

northern hemispheric SWE. In areas of complex terrain, uncertainty in meteorological forcing within reanalyses, particularly precipitation amount and phase (Lundquist et al., 2019) must also be considered. Further, in alpine regions the coarse resolution of the gridded SWE products (25 km or more) does not lend itself to comparison with snow course observations because of limited representativeness of surface observations in complex terrain and across elevation gradients; a different validation approach is likely needed for mountain areas.

The AMSR-E products exhibit weak spatial agreement and negative temporal anomaly correlations with the other datasets (Fig. 8 and 9). The retrieval of SWE solely from passive microwave measurements is a difficult challenge, and despite the best efforts of many research groups over many decades, passive microwave-based standalone algorithms do not perform as well as other methods that make use of ancillary snow depth measurements or snow models. Although there are many attractive attributes (wide swath, all-weather imaging, long legacy time series, and theoretical sensitivity to SWE under simplified assumptions) passive microwave data has always been a measurement of opportunity for snow applications, not an ideal measurement system. This introduces intrinsic biases and errors into the standalone retrieval scheme because of the non 'optimal' nature of these measurements for snow applications.

Despite these challenges, there are opportunities to utilize satellite passive microwave measurement as a component of SWE product development moving forward. Machine learning operators show potential for the radiance-based assimilation of brightness temperatures (e.g. Forman and Reichle 2014) analogous to how L-band brightness temperatures are assimilated for improved soil moisture analyses. Assimilation approaches also show potential for addressing challenges posed by stratigraphy (Durand et al., 2011; Andreadis and Lettenmaier, 2012) and deep snow (Li et al., 2012). While coarse resolution is an inherent challenge with satellite passive microwave measurements, enhanced resolution products spanning multiple decades are now available (Long and Brodzik, 2016; Takala et al., 2017).

The combination of brightness temperature measurements, surface snow depth observations, and forward radiometric modeling are able to produce skillful SWE products. This approach was already used successfully within the ESA GlobSnow project, and will be further enhanced within the ESA Climate Change Initiative (CCI) snow project. It is important to note that the brightness temperature component of the GlobSnow/Snow CCI retrieval has direct heritage to standalone passive microwave retrieval approaches which date back to the first generation of passive microwave imagers launched in the 1970s. This also suggests that research focusing on passive microwave interactions with snow parameters should not be neglected as, ultimately, better understanding of the underlying physics is a positive step for algorithm improvement.

While the continued development of a remote sensing capabilities for SWE represents an important observational capability, it is necessary to also appreciate the quality of the large scale model derived SWE products. The combination of reanalysis meteorology and snow models yields very useful snow information, which can be refined as forcing data (particularly

precipitation) and snow models continue to improve. Only through combined and integrated improvements in remote sensing, modeling, and observations will real progress in SWE product development be achieved and sustained.

References

455 Andreadis, K., and Lettenmaier, D.: Implications of representing snowpack stratigraphy for the assimilation of passive microwave satellite observations, J. Hydrometeorol., 13, 1493–1506, https://doi.org/10.1175/JHM-D-11-056.1, 2012.

Balsamo, G., Albergel, C., Beljaars, A. , Boussetta, S. , Brun, E. , Cloke, H. , Dee, D. , Dutra, E., Muñoz-Sabater, J., Pappenberger, F. , de Rosnay, P., Stockdale, T., and Vitart, F. : ERA-Interim/Land: a global land surface reanalysis data set,
460 Hydrology and Earth System Science, 19, 389–407, https://www.hydrol-earth-syst-sci.net/19/389/2015, 2015.

Barnett, T. P., Adam, J.C., and D.P. Lettenmaier, D. P.: Potential impacts of a warming climate on water availability in snow-dominated regions, Nature, 438, 303–309. https://doi.org/10.1038/nature04141, 2005.

465 Baseman, J. and Fernandez-Prieto D. (eds.). ESA-CliC Earth Observation and Arctic Science Priorities. Tromso, Norway, 20 Jan 2015. http://www.climate-cryosphere.org/media-gallery/1533-esa-clic-arctic-2015, 2015.

Brodzik, M. J., Billingsley, B., Haran, T., Raup, B., & Savoie, M. H.: EASE-Grid 2.0: Incremental but significant improvements for Earth-gridded data sets, ISPRS International Journal of Geo-Information, 1, 32-45, 470 https://doi.org/10.3390/ijgi1010032, 2012.

Brown, R., and Braaten, R.: Spatial and temporal variability of Canadian monthly snow depths, 1946-1995, Atmos.-Ocean, 36, 37–54. https://doi.org/10.1080/07055900.1998.9649605, 1998.

475 Brown, R., Brasnett, B., and Robinson, D.: Gridded North American monthly snow depth and snow water equivalent for GCM evaluation. Atmos.-Ocean, 41, 1–14. https://doi.org/10.3137/ao.410101, 2003.

Brown, R., and Mote, P.: The response of Northern Hemisphere snow cover to a changing climate. J. Climate, 22, 2124–2145, https://doi.org/10.1175/2008JCLI2665.1, 2009.
480
Brown, R., Derksen, C., and Wang, L.: A multi-dataset analysis of variability and change in Arctic spring snow cover extent, 1967–2008. J. Geophys. Res., 115, D16111, https://doi.org/10.1029/2010JD013975, 2010.

Brown, R., and Derksen, C.: Is Eurasian October snow cover extent increasing?, Env. Res. Lett., 8, 024006, 485 https://doi.org/10.1088/1748-9326/8/2/024006, 2013.

Brown, R., Tapsoba, D. and Derksen, C.: Evaluation of snow water equivalent datasets over the Saint-Maurice river basin region of southern Québec. Hydrol. Process., 32, 2748–2764, https://doi.org/10.1002/hyp.13221, 2018.

490 Brown, R. D., Fang, B. and Mudryk, L. Update of Canadian historical snow survey data and analysis of snow water equivalent trends, 1967-2016. Atmos.-Ocean, 57, 1–8, https://doi.org/10.1080/07055900.2019.1598843, 2019.

Broxton, P. D., Dawson, N., and Zeng, X.: Linking snowfall and snow accumulation to generate spatial maps of SWE and snow depth, Earth and Space Science, 3, 246–256, https://doi.org/10.1002/2016EA000174, 2016.
495
Brun, E., Vionnet, V., Boone, A., Decharme, B., Peings, Y. , Vallette, R. , Karbou, F., and Morin, S.: Simulation of northern Eurasian local snow depth, mass, and density using a detailed snowpack model and meteorological reanalyses, J. Hydrometeorol., 14, 203–219, https://doi.org/10.1175/JHM-D-12-012.1, 2013.

16

500   Bulygina, O., Groisman, P Ya, Razuvaev, V., and Korshunova, N.: Changes in snow cover characteristics over Northern Eurasia since 1966, Env. Res. Lett., 6, 045204, https://doi.org/10.1088/1748-9326/6/4/045204, 2011.

Copernicus Climate Change Service (C3S): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate, Copernicus Climate Change Service Climate Data Store (CDS), 2019-02-19, https://cds.climate.copernicus.eu/cdsapp#!/home,
505   2017.

Chang, A., Foster, J., and Hall, D.: Satellite sensor estimates of northern hemisphere snow volume. Int. J. Remote Sens., 11, 167–171. https://doi.org/10.1080/01431169008955009, 1990.

510   Clark, M. P., Hendrix, J., Slater, A. G., Kavetski, D., Anderson, B. , Cullen, N. J. , Kerr, T., Hreinsson, E. O., and Woods, R. A.: Representing spatial variability of snow water equivalent in hydrologic and land-surface models: a review, Water Resour. Res., 47, W07539, https://doi.org/10.1029/2011WR010745, 2011.

Clifford, D.: Global estimates of snow water equivalent from passive microwave instruments: history, challenges and future
515   developments. Int. J. Remote Sens., 31, 3707–3726, https://doi.org/10.1080/01431161.2010.483482, 2010.

Deeb, E., Forster, R., and Kane, D.: Monitoring snowpack evolution using interferometric synthetic aperture radar on the North Slope of Alaska, USA, Int. J. Remote Sens., 32, 3985–4003, https://doi.org/10.1080/01431161003801351, 2011.

520   Derksen, C., Brown, R., and Walker, A.: Merging conventional (1915-92) and passive microwave (1978–2002) estimates of snow extent and water equivalent over central North America, J. Hydrometeorol., 5, 850–861, https://doi.org/10.1175/1525-7541(2004)005<0850:MCAPME>2.0.CO;2, 2004.

Derksen, C., Lemmetyinen, J., Toose, P., Silis, A., Pulliainen, J., and Sturm, M.: Physical properties of Arctic versus subarctic
525   snow: Implications for high latitude passive microwave snow water equivalent retrievals, J. Geophys. Res. - Atmospheres, 119, 7254–7270, https://doi.org/10.1002/2013JD021264, 2014.

Derksen, C. and Nagler, T.: ESA CCI+ Snow ECV: User Requirements Document, version 1.0, January 2019.

530   Durand, M., Kim, E., Margulis, S., and Molotch, N.: A first-order characterization of errors from neglecting stratigraphy in forward and inverse passive microwave modeling of snow, IEEE Geosci. Remote S. Lett., 8, 730–734, https://doi.org/10.1109/LGRS.2011.2105243, 2011.

17

Durand, M., and Liu, D.: The need for prior information in characterizing snow water equivalent from microwave brightness temperatures, Remote Sens. Environ., 126, 248–257, https://doi.org/10.1016/j.rse.2011.10.015, 2012.

Dyer, J., and Mote, T.: Spatial variability and trends in observed snow depth over North America. Geophys. Res. Lett., 33, https://doi.org/10.1029/2006GL027258, 2006.

Forman, B. A., and Reichle, R. H.: Using a support vector machine and a land surface model to estimate large-scale passive microwave brightness temperatures over snow-covered land in North America, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8, 4431-4441, https://doi.org/10.1109/JSTARS.2014.2325780, 2014.

Foster, J. L., Sun, C., Walker, J. P., Kelly, R., Chang, A., Dong, J., and Powell, H.: Quantifying the uncertainty in passive microwave snow water equivalent observations, Remote Sens. Environ., 94, 187–203, https://doi.org/10.1016/j.rse.2004.09.012, 2005.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella,S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Climate, 30, 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

GMAO (Global Modeling and Assimilation Office), tavg1_2d_lnd_Nx: MERRA 2D IAU Diagnostic, Land Only States and Diagnostics, Time Average 1-hourly V5.2.0, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [22 November, 2013], 10.5067/YL8Z7MICQZF9, 2008.

GMAO (Global Modeling and Assimilation Office), MERRA-2 tavg1_2d_lnd_Nx: 2d,1-Hourly,Time-Averaged,Single-Level, Assimilation,Land Surface Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [13 July, 2016], 10.5067/RKPHT8KC1Y1T, 2015.

Goodison, B. E., Ferguson, H. L., and McKay, G.A.: Measurement and data analysis. In: Handbook of Snow (D. M. Gray and D. H. Male, eds.), pp. 191–274. Reprint. Caldwell, NJ, USA, The Blackburn Press, 1981.

Haberkorn, A. (Ed.): European Snow Booklet, 363 pp., doi:10.16904/envidat.59, 2019.

18

Hall, D.: Influence of depth hoar on microwave emission from snow in northern Alaska, Cold Reg. Sci. Technol., 13, 225-231, https://doi.org/10.1016/0165-232X(87)90003-6, 1987.

Hall, D., Sturm, M. , Benson, C. , Chang, A., Foster, J. , Garbeil, H., and Chacho, E.: Passive microwave remote and in situ measurements of Arctic and Subarctic snow covers in Alaska, Remote Sens. Environ., 38, 161–172, https://doi.org/10.1016/0034-4257(91)90086-L, 1991.

Henn, B., Newman, A., Livneh, B., Daly, C., and Lundquist, J.: An assessment of differences in gridded precipitation datasets in complex terrain, J. Hydrol., 556, 1205–1219, https://doi.org/10.1016/j.jhydrol.2017.03.008, 2018.

Hersbach, H., Bell, W., Berrisford, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Simmons, A., Soci, C., Dee, D.: Global reanalysis: goodbye ERA-Interim, hello ERA5, ECMWF Newsletter., 159, 17-24, https://doi.org/10.21957/vf291hehd7, 2019.

Kelley, R. E., Change, A. T., Tsang, L., and Foster, J. L.: A prototype AMSR-E global snow area and snow depth algorithm, IEEE Trans. Geoci. Remote S., 41, 230-242, https://doi.org/10.1109/TGRS.2003.809118, 2003.

Kelly, R.E.J.: The AMSR-E Snow Depth Algorithm: Description and Initial Results, Journal of the Remote Sensing Society of Japan, 29, 307–317, https://doi.org/10.11440/rssj.29.307, 2009.

Krenke, A. Edited by National Snow and Ice Data Center. Former Soviet Union Hydrological Snow Surveys, 1966-1996, Version 1. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. https://10.7265/N58C9T60, 1998, updated 2004.

Krinner, G., Derksen, C., Essery, R., Flanner, M., Hagemann, S., Clark, M., Hall, A., Rott, H., Brutel-Vuilmet, C., Kim, H., Ménard, C. B., Kim, H., Ménard, C. B., Mudryk, L., Thackeray, C., Wang, L., Arduini, G., Balsamo, G., Bartlett, P., Boike, J., Boone, A., Chéruy, F., Colin, J., Cuntz, M., Dai, Y., Decharme, B., Derry, J., Ducharne, A., Dutra, E., Fang, X., Fierz, C., Ghattas, J., Gusev,Y., Haverd, V., Kontu, A., Lafaysse, M., Law, R., Lawrence, D., Li,W., Marke, T., Marks, D., Ménégoz, M., Nasonova, O., Nitta, T., Niwano, M., Pomeroy, J., Raleigh, M. S., Schaedler, G., Semenov, V., Smirnova, T. G., Stacke, T., Strasser, U., Svenson, S., Turkov, D., Wang, T., Wever, N., Yuan, H., Zhou, W., and Dan Zhu, D.: ESM-SnowMIP: assessing snow models and quantifying snow-related climate feedbacks, Geosci. Model Dev., 11, 5027-5049, https://doi.org/10.5194/gmd-11-5027-2018, 2018.

Larue, F., Royer, A., De Sève, D., Langlois, A., Roy, A. and Brucker, L.: Validation of GlobSnow-2 snow water equivalent over Eastern Canada, Remote Sens. Environ., 194, 264–277, https://doi.org/10.1016/j.rse.2017.03.027, 2017.

Lemmetyinen, J., Kontu, A., Kärnä, J.-P. , Vehviläinen, J. , Takala, M.,and Pulliainen, J.: Correcting for the influence of frozen lakes in satellite microwave radiometer observations through application of a microwave emission model, Remote Sens. Environ., 115, 3695–3706, https://doi.org/10.1016/j.rse.2011.09.008, 2011.

Li, D., Durand, M., and Margulis, S.: Potential for hydrologic characterization of deep mountain snowpack via passive microwave remote sensing in the Kern River basin, Sierra Nevada, USA, Remote Sens. Environ., 125, 34–48, https://doi.org/10.1016/j.rse.2012.06.027, 2012.

Liston, G., and Hiemstra, C.: The changing cryosphere: pan-Arctic snow trends (1979–2009), J. Climate, 24, 5691–5712, https://doi.org/10.1175/JCLI-D-11-00081.1, 2011.

Long, D., and Brodzik, M.-J.: Optimum image formation for spaceborne microwave radiometer products. IEEE Geosci. Remote S., 54, 2763–2779, https://doi.org/10.1109/TGRS.2015.2505677, 2016.

Lundquist, J.D., Hughes, M., Henn, B., Gutmann, E.D., Livneh, B., Dozier, J., and Neiman, P.: High-elevation precipitation patterns: using snow measurements to assess daily gridded datasets across the Sierra Nevada, California, J. Hydrometeorol., 16, 1773–1792, https://doi.org/10.1175/JHM-D-15-0019.1, 2015.

Markus, T., Powell, D., and Wang, J.: Sensitivity of passive microwave snow depth retrievals to weather effects and snow evolution, IEEE Geosci. Remote S., 44, 68–77, https://doi.org/10.1109/TGRS.2005.860208, 2006.

Meromy, L., Molotch, N. P., Link, T. E., Fassnacht, S. R., and Rice R.: Subgrid variability of snow water equivalent at operational snow stations in the western USA, Hydrological Processes, 27, 2383–2400, https://doi.org/10.1002/hyp.9355, 2012.

Mudryk, L., Derksen, C., Kushner, P., and Brown, R.: Characterization of Northern Hemisphere snow water equivalent datasets, 1981–2010, J. Climate, 28, 8037–8051, https://doi.org/10.1175/JCLI-D-15-0229.1, 2015.

Mudryk, L., Kushner, P., Derksen, C., and Thackeray, C.: Snow cover response to temperature in observational and climate model ensembles, Geophys. Res. Lett., 44, 919–926, https://doi.org/10.1002/2016GL071789, 2017.

Mudryk, L., Derksen, C., Howell, S., Laliberté, F., Thackeray, C., Sospedra-Alfonso, R., Vionnet, V., Kushner, P., and Brown, R.: Canadian snow and sea ice: historical trends and projections, The Cryosphere, 12, 1157–1176, https://doi.org/10.5194/tc-635   12-1157-2018, 2018a.

Mudryk, L., Brown, R., Derksen, C., Luojus, K., Decharme, B., and Helfrich, S.: Terrestrial Snow Cover [in Arctic Report Card], 28 November 2018, https://www.arctic.noaa.gov/Report-Card, 2018b.

640   Mudryk, L., Brown, R., Derksen, C., Luojus, K., and Dechame, B.: Terrestrial Snow Cover [in: "State of the Climate 2018"], Am. Meteorol. Soc., 100, S181–S185, https://doi.org/10.1175/2019BAMSStateoftheClimate.1, 2019.

Neumann, N., Smith, C., Derksen, C., and Goodison, B.: Characterizing local scale snow cover using point measurements during the winter season, Atmos.-Ocean, 44, 257–269, https://doi.org/10.3137/ao.440304, 2006.
645
Orsolini, Y., Wegmann, M., Dutra, E., Liu, B., Balsamo, G., Yang, K., de Rosnay, P., Zhu, C., Wang, W., Senan, R., and Arduini, G.: Evaluation of snow depth and snow cover over the Tibetan Plateau in global reanalyses using in situ and satellite remote sensing observations, The Cryosphere, 13, 2221–2239, https://doi.org/10.5194/tc-13-2221-2019, 2019.

650   Painter, T., Berisford, D., Boardman, J., Bormann, K., Deems, J., Gehrke, F., Hedrick, A., Joyce, M., Laidlaw, R., Marks, D., Mattmann, C., McGurk, B., Ramirez, P., Richardson, M., Skiles, S. M., Seidel, F., and Winstral, A.: The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo, Remote Sens. Environ., 184, 139–152, https://doi.org/10.1016/j.rse.2016.06.018, 2016.

655   Pulliainen, J.: Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations, Remote Sens. Environ., 101, 257–269, https://doi.org/10.1016/j.rse.2006.01.002, 2006.

Rawlins, M.A., Fahnestock, M., Frolking, S. and Vörösmarty, C.J.: On the evaluation of snow water equivalent estimates over 660   the terrestrial Arctic drainage basin, Hydrol. Process., 21, 1616–1623, https://doi.org/10.1002/hyp.6724, 2007.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and 665   Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, J. Climate, 24, 3624–3648, https://doi.org/10.1175/JCLI-D-11-00015.1, 2011.

Robertson, F. R., Bosilovich, M. G., Chen, J., and Miller, T. L.: The effect of satellite observing system changes on MERRA water and energy fluxes, J. Clim., 24, 5197–5217, doi:10.1175/2011JCLI4227.1, 2011.

670

Rodell, M., Houser, P. R., Jambor, U. E. A., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The global land data assimilation system, B. Am. Meteorol. Soc., 85, 381–394, https://doi.org/10.1175/BAMS-85-3-381, 2004.

675 Rott, H., Yueh, S. H., Cline, D. W., Duguay, C., Essery, R., Haas, C., Hélière, F., Kern, M. G., Malnes, E., Nagler, T., Pulliainen, J., Rebhan, H., and Thompson, A.: Cold regions hydrology high-resolution observatory for Snow and Cold Land Processes, Proc. IEEE, 98, 752–765, https://doi.org/10.1109/JPROC.2009.2038947, 2010.

Sospedra-Alfonso, R., Mudryk, L., Merryfield, W., and Derksen, C.: Representation of snow in the Canadian seasonal to
680 interannual prediction system. Part I: Initialization, J. Hydrometerol., 17, 1467-1488, https://doi.org/10.1175/JHM-D-14-0223.1, 2016.

Sturm, M., Holmgren, J., Liston, G.: Global Seasonal Snow Classification System. Version 1.0. UCAR/NCAR - Earth Observing Laboratory. https://doi.org/10.5065/D69G5JX5, 2009. Accessed 14 Feb 2020.

685

Sturm, M., Holmgren, J., and Liston, G.: A seasonal snow cover classification system for local to global applications, J. Climate, 8, 1261–1283, https://doi.org/10.1175/1520-0442(1995)008<1261:ASSCCS>2.0.CO;2, 1995.

Sturm, M., Taras, B. , Liston, G. , Derksen, C. , Jonas, T. , and Lea, J.: Estimating snow water equivalent using snow depth
690 data and climate classes, J. Hydrometeorol., 11, 6, 1380–1394, https://doi.org/10.1175/2010JHM1202.1, 2010.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J.-P. ,and Koskinen, J.: Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements, Remote Sens. Environ., 115, 3517–3529, https://doi.org/10.1016/j.rse.2011.08.014, 2011.

695

Takala, M., Ikonen, J., Luojus, K., Lemmetyinen, J., Metsämäki, S., Cohen, J., Arslan, A. N., and Pulliainen, J.: New snow water equivalent processing system with improved resolution over Europe and its applications in hydrology, IEEE Journal of Selected Topics on Applied Remote Sensing, 10, 428–436, https://doi.org/10.1109/JSTARS.2016.2586179, 2017.

22

700 Tedesco, M., Kelly, R., Foster, J. L., and Change, A. T.: AMSR-E/Aqua Daily L3 Global Snow Water Equivalent EASE-Grids, Version 2. Boulder, Colorado USA, NASA Snow and Ice Data Center Distributed Active Archive Center, https://doi.org/10.5067/AMSR-E/AE_DYSNO.002, 2004.

Tedesco, M., and Narvekar, P.: Assessment of the NASA AMSR-E SWE product, IEEE Journal of Selected Topics in Applied
705 Earth Observations and Remote Sensing, 3, 141–159, https://doi.org/10.1109/JSTARS.2010.2040462, 2010.

Tedesco, M., and Jeyaratnam, J.: A new operational snow retrieval algorithm applied to historical AMSR-E brightness temperatures, Remote Sensing, 8, 1037, https://doi.org/10.3390/rs8121037, 2016.

710 Vuyovich, C. M., Jacobs, J. M., and Daly, S. F.: Comparison of passive microwave and modeled estimates of total watershed SWE in the continental United States, Water Resour. Res., 50, 9088–9102. https://doi.org/10.1002/2013WR014734, 2014.

Wrzesien, M. L., Durand, M. T., Pavelsky, T. M., Kapnick, S. B., Zhang, Y., Guo, J., and Shum, C. K..: A new estimate of North American mountain snow accumulation from regional climate model simulations, Geophys. Res. Lett., 45, 1423–1432,
715 https://doi.org/10.1002/2017GL076664, 2018.

Wrzesien, M. L., Pavelsky, T. M., Durand, M. T., Dozier, J., and Lundquist, J. D.: Characterizing biases in mountain snow accumulation from global data sets, Water Res. Res., 55, 9873–9891, https://doi.org/10.1029.2019WR025350, 2019.
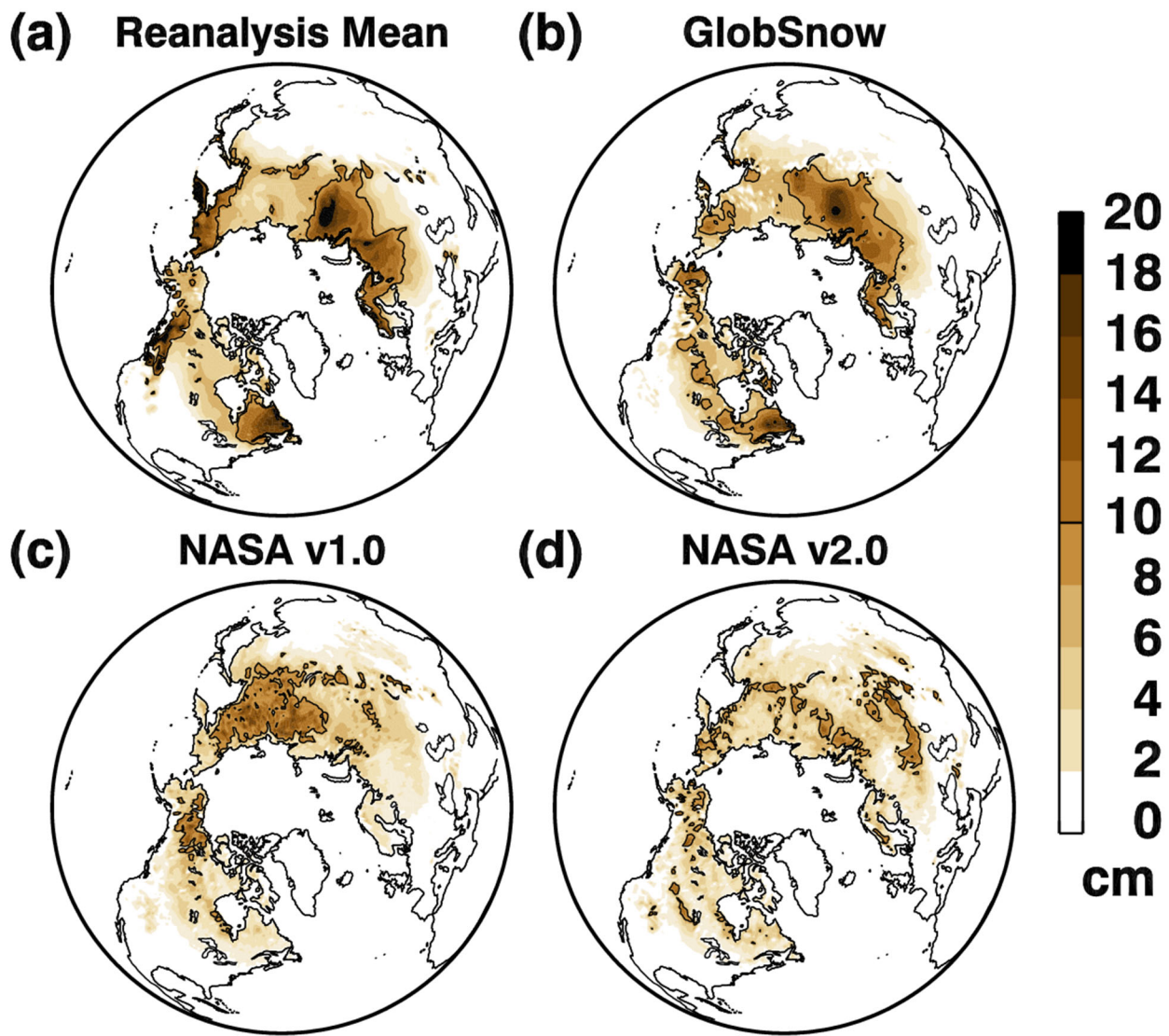
Table 1. Summary of SWE products evaluated in this study.

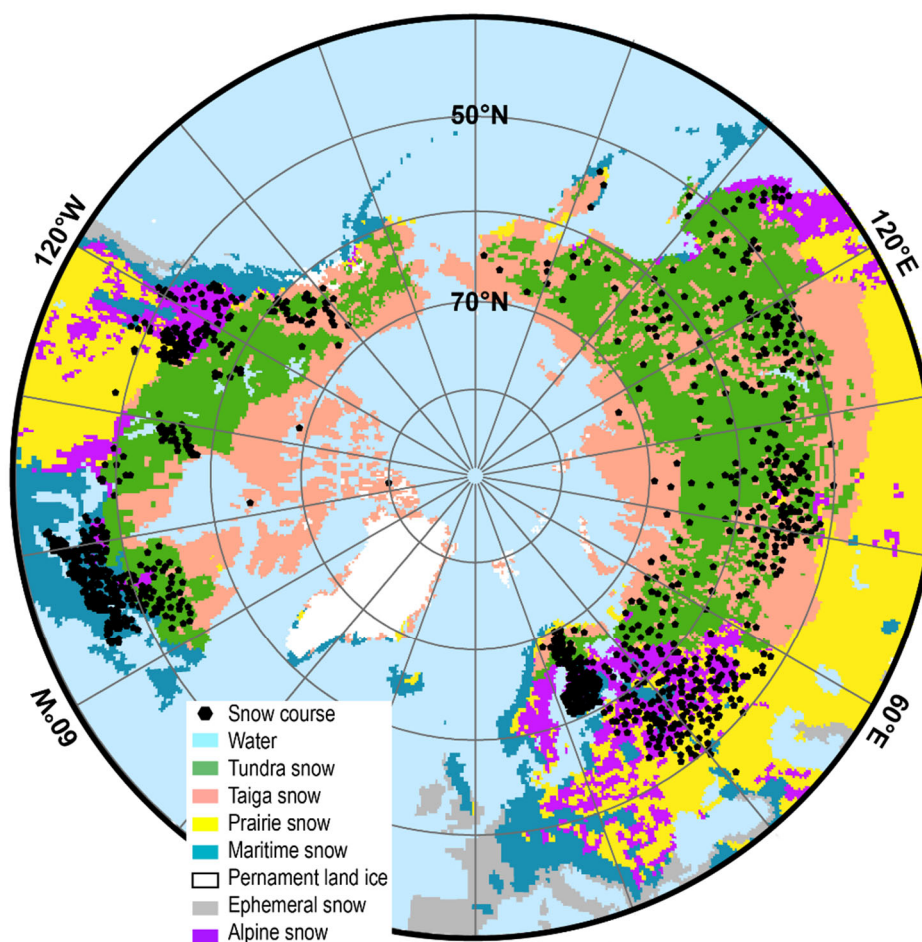| Data Product	Method	Ancillary/ Forcing Data	Resolution	Reference/ Availability
GlobSnow v2.0	Passive microwave + in situ	Weather station snow depth measurements	25 km	Takala et al., 2011 www.globsnow.info
NASA AMSR-E v1.0	Standalone passive microwave		25 km	Kelly (2009) nsidc.org
NASA AMSR-E v2.0	Microwave + ground station climatology	Weather station snow depth climatology	25 km	Tedesco and Jeyaratnam (2016) nsidc.org†
ERA-Interim/Land	HTESSEL land surface model	ERA-interim	0.75° x 0.75°	Balsamo et al (2015) www.ecmwf.int
ERA5	HTESSEL land surface model	ERA5	0.25° x 0.25°	Hersbach et al. (2019) C3S (2017)
MERRA	Catchment land surface model	MERRA	0.5° x 0.67°	Rienecker et al (2011) GMAO (2017a)
MERRA2	Catchment land surface model	MERRA2	0.5° x 0.625°	Gelaro et al. (2017) GMAO (2017b)
Crocus	ISBA land surface + Crocus snow model	ERA-interim	1° x 1°	Brun et al (2013) ‡
GLDAS-2	Noah 3.3 land surface model	Princeton Met.	1° x 1°	Rodell et al (2004) disc.gsfc.nasa.gov

† The v2 product is not available via NSIDC over the 2002–2010 period, however data using the same algorithm is available from July 2012–present. Contact authors for availability.
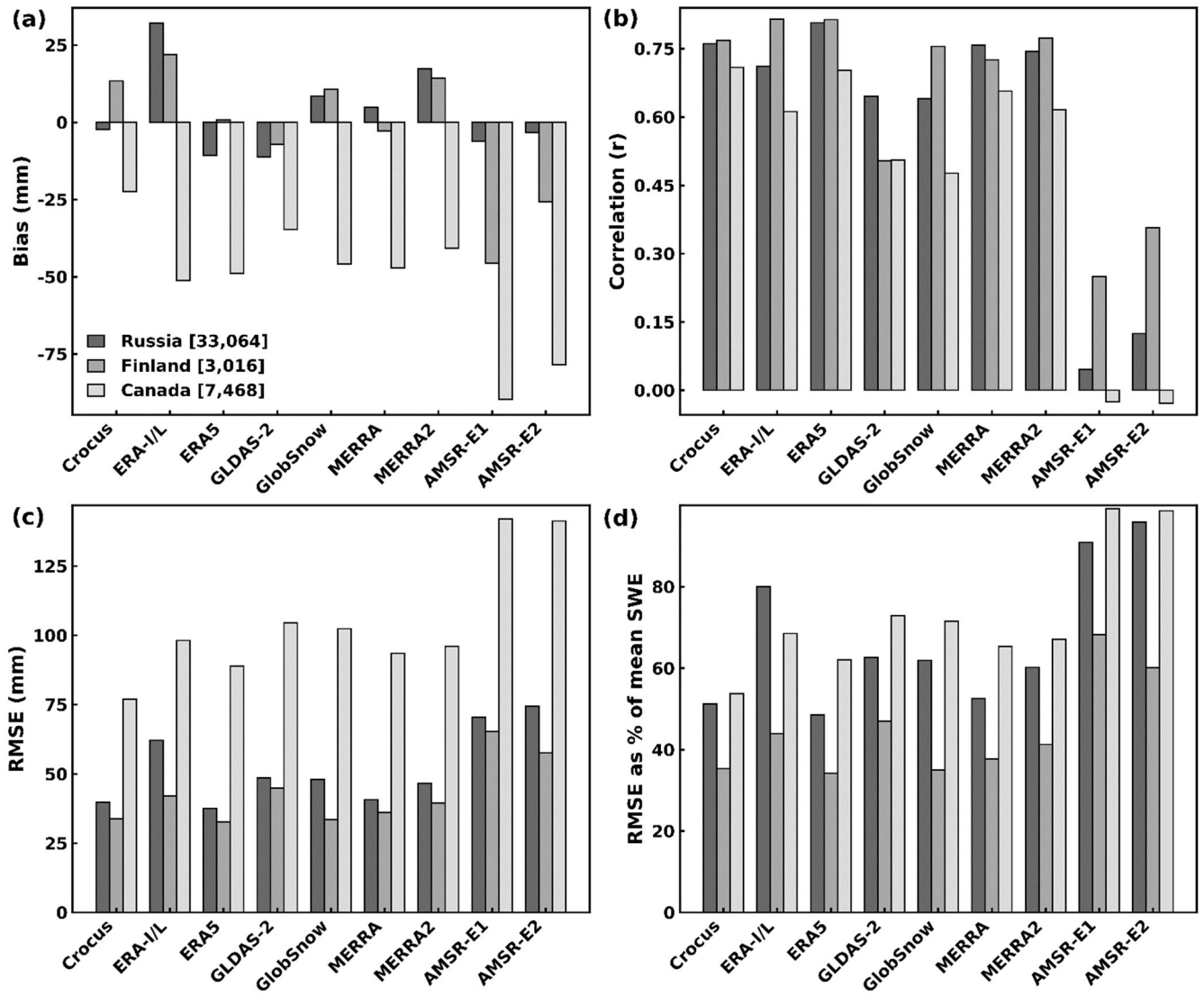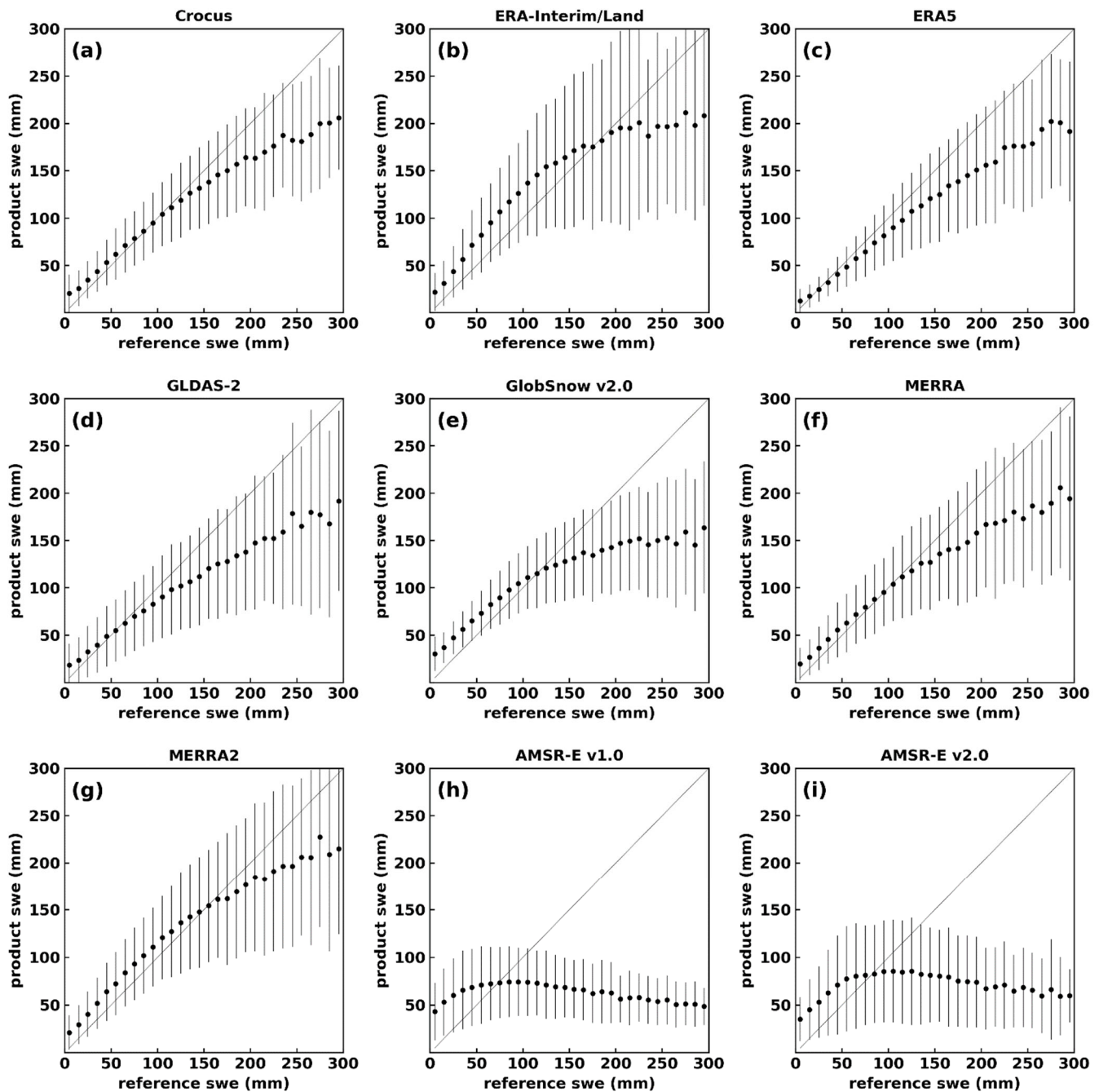‡ Contact authors for availability.

735 **Figure 1.** Mean January, February, and March (JFM) SWE over the 2003 – 2010 period for (a) four reanalysis driven products (GLDAS-2, ERA-Interim/Land, Crocus, and MERRA2); (b) GlobSnow v2.0; (c) NASA AMSR-E SWE v1.0; (d) NASA AMSR-E SWE v2.0.
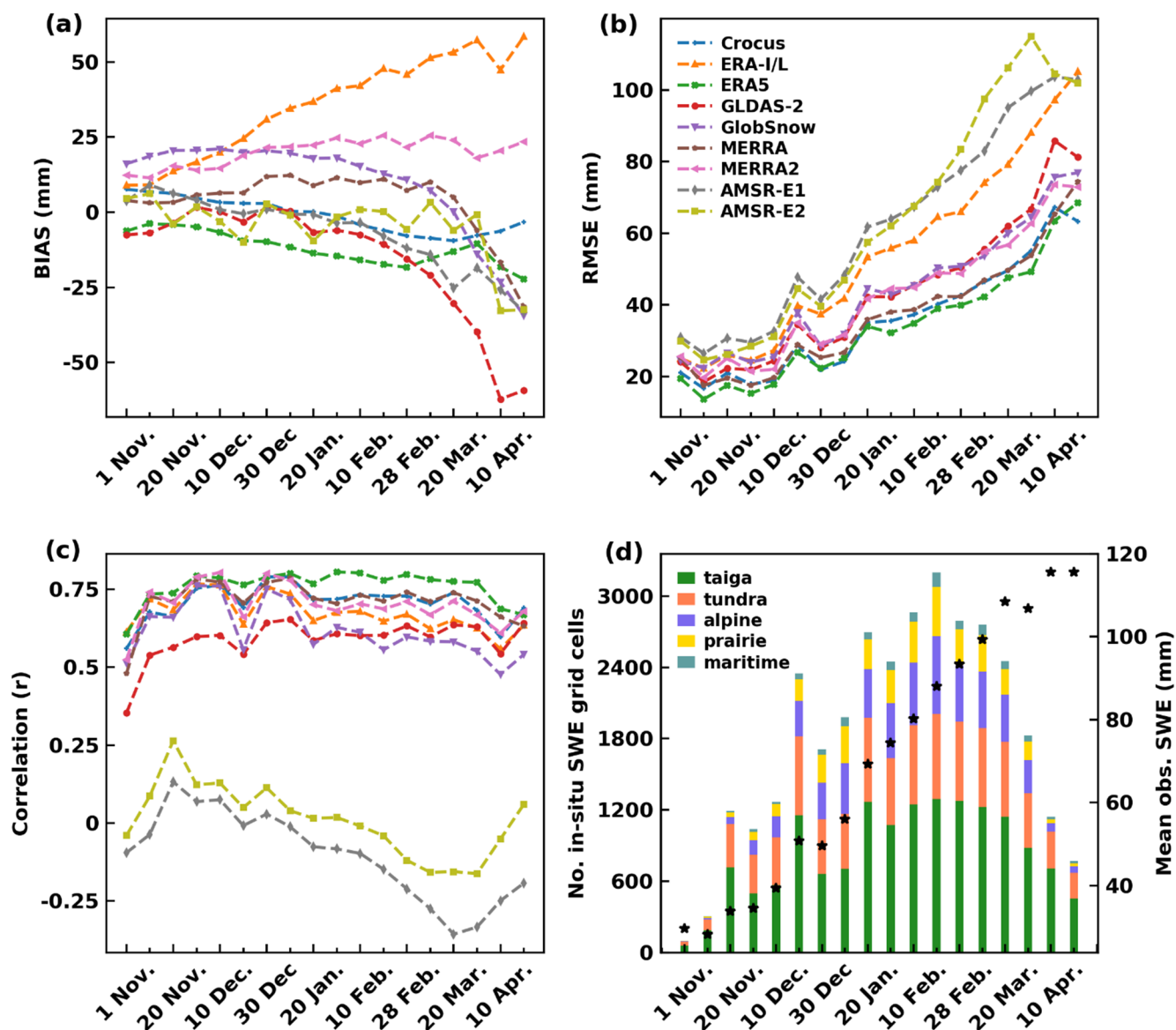
25

Figure 2. Centroid of 25 km EASE grid cells with snow course observations used in the analysis (Sect. 2.3) overlaid on snow-climate classes (Sturm et al., 2009).

Figure 3. Validation statistics (a: bias; b: correlation; c: RMSE; d: RMSE as percentage of mean SWE) for the nine SWE products for November through April, 2002–2010 [ERA-I/L = ERA-Interim/Land; GlobSnow = GlobSnow v2.0]. Total number of grid cells with snow course measurements in square brackets in panel a.
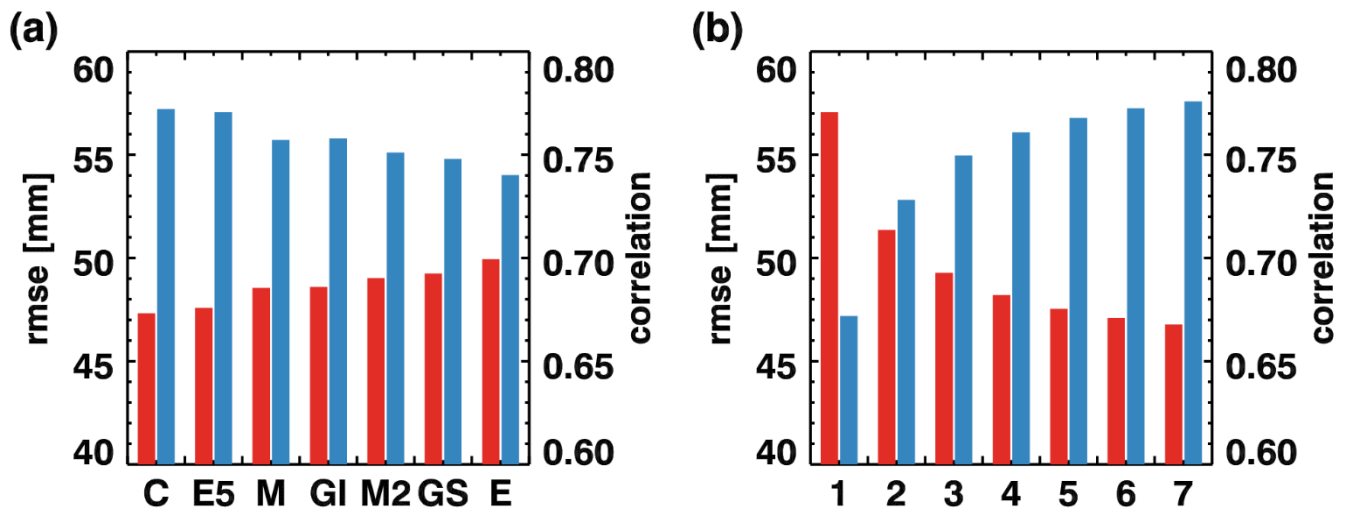
745

Figure 4. Performance of SWE datasets versus reference snow course SWE ± 1 standard deviation for (a) Crocus; (b) ERA-Interim/Land; (c) ERA5; (d) GLDAS-2; (e) GlobSnow v2.0; (f) MERRA; (g) MERRA2; (h) AMSR-E v1.0; (i) AMSR-E v2.0. SWE values above 300 mm are not shown.
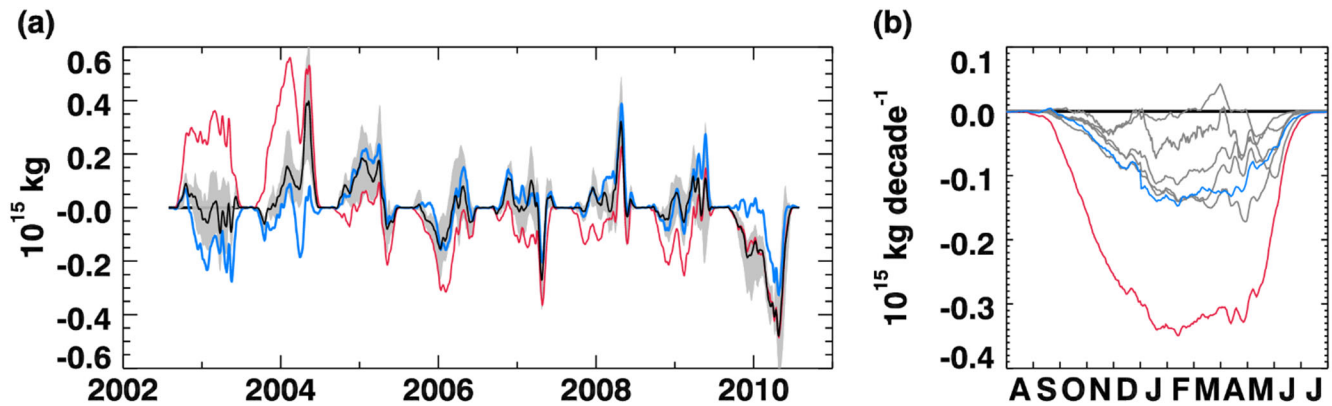
28

Figure 5. (a) Bias (b) RMSE and (c) correlation coefficient relative to the Russia snow course dataset (Sect. 2.1) for each ten day time step over the 2002–2010 period. (d) Number of grid cells with snow course observations by Sturm et al. (2009) snow class (bars, left-hand axis), mean observed SWE (stars, right-hand axis) [ERA-I/L = ERA-Interim/Land, GlobSnow = GlobSnow v2.0].
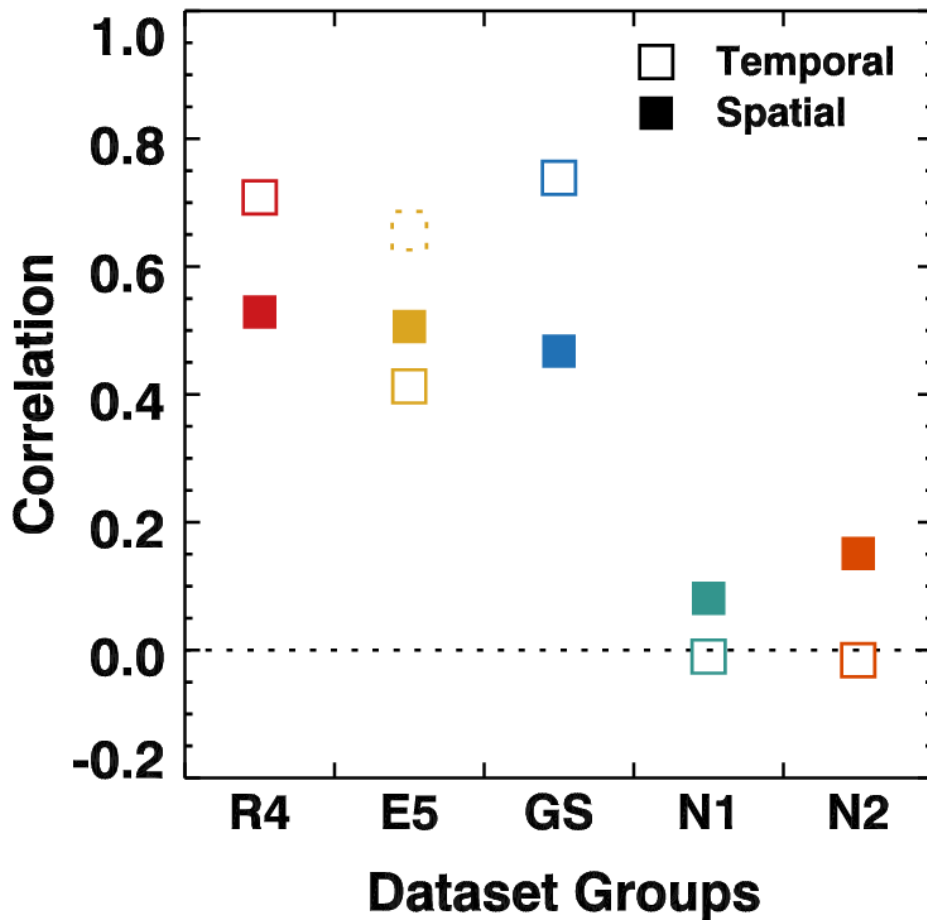
Figure 6. RMSE (red) and correlation (blue) of snow course measurements with various combinations of SWE products. (a) Average of all combinations that contain the specified individual product [C = Crocus, E5 = ERA5, M = MERRA, Gl = GLDAS-2, M2 = MERRA2, GS = GlobSnow v2.0, E = ERA-Interim/Land] and (b) average of all combinations of N products as specified on x-axis.
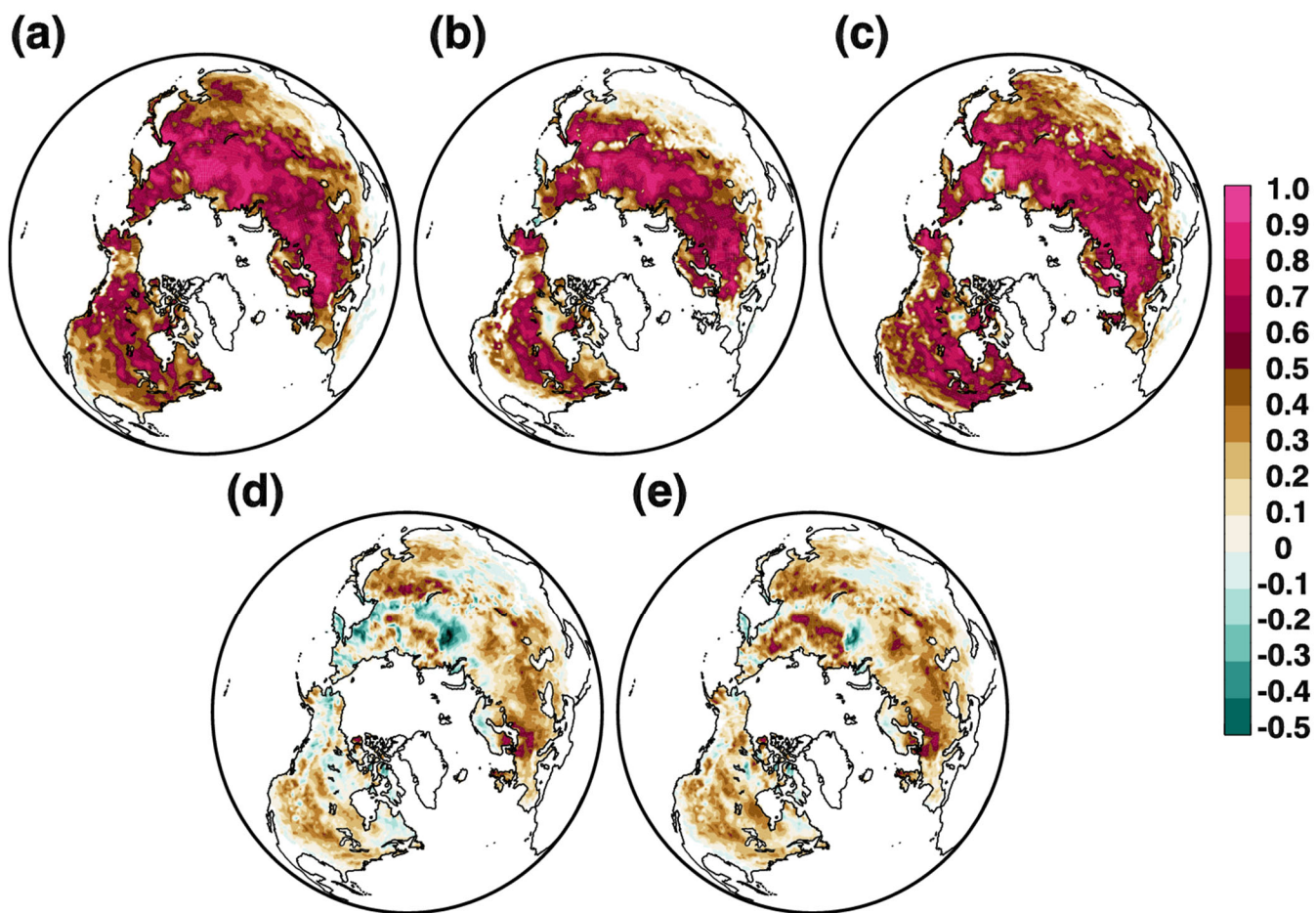

Figure 7. (a) Average Northern Hemisphere snow mass anomalies (black) and spread (shading) calculated from five component products: MERRA2, Crocus, GlobSnow v2.0, GLDAS-2 and ERA-Interim/Land along with snow mass anomalies from raw (red) and corrected (blue) ERA5 values. (b) Trends (1981–2010) from the five component time series used for the average in panel a (grey) along with trends from the raw (red) and corrected (blue) ERA5 time series. The ERA5 discontinuity occurs in January 2004.

30

Figure 8. Temporal and spatial correlations among groups of products over the 2002–2010 time period. Temporal correlations assess the extent to which anomalous northern hemispheric snow mass jointly evolves between pairs of datasets while spatial correlations assess the pattern correlation of SWE fields for pairs of datasets; see text (Sect. 2.3) for details. R4 = the average of six pairwise correlations between Crocus, GLDAS-2, ERA-Interim/Land, and MERRA2. E5 = the average of 4 pairwise correlations between ERA5 and each R4 product. GS = the average of 4 pairwise correlations between GlobSnow v2.0 and each R4 product. N1 = the average of four pairwise correlations between AMSR-E v1.0 and each R4 product. N2 = the average of four pairwise correlations between AMSR-E v2.0 and each R4 product. The dotted square shows the impact of correcting the E5 snow mass anomalies for a discontinuity introduced in 2004.

Figure 9. Correlation maps (2002–2010) for four reanalysis driven products (Crocus, GLDAS-2, ERA-Interim/Land, and MERRA2) relative to: (a) each other (mean correlation between the four reanalysis driven products); (b) GlobSnow v2.0; (c) ERA5; (d) NASA AMSR-E SWE v1.0; (e) NASA AMSR-E SWE v2.0.