Characterization of Northern Hemisphere snow water equivalent datasets, 1981–2010

L. R. Mudryk*

Department of Physics, University of Toronto, Toronto, Ontario, Canada.

C. Derksen

Climate Research Division, Environment Canada, Toronto, Ontario, Canada.

P. J. Kushner

Department of Physics, University of Toronto, Toronto, Ontario, Canada.

R. Brown

Climate Research Division, Environment Canada, Ouranos, Montréal, Québec, Canada.

Corresponding author address: Department of Physics, University of Toronto, 60 St. George Street, Toronto, Ontario, Canada, M5S1A7.

E-mail: mudryk@atmosp.physics.utoronto.ca

ABSTRACT

Five daily, gridded Northern Hemisphere snow water equivalent (SWE) datasets are analyzed over the 1981–2010 period in order to quantify the spatial and temporal consistency of satellite retrievals, land surface assimilation systems, physical snow models, and reanalyses. While the climatologies of total Northern Hemisphere snow water mass (SWM) vary among the datasets by as much as 50%, their interannual variability and daily anomalies are comparable, showing moderate to good temporal correlations (between 60% and 85%) on both interannual and intraseasonal time scales. Wintertime trends of total Northern Hemisphere SWM are consistently negative over the 1981–2010 period among the five datasets but vary in strength by a factor of 2–3. Examining spatial patterns of SWE indicates that the datasets are most consistent with one another over boreal forest regions compared to Arctic and alpine regions. Additionally, the datasets derived using relatively recent reanalyses are strongly correlated with one another and show better correlations with the satellite product (GlobSnow) than do those using older reanalyses. Finally, a comparison of eight reanalysis datasets over the 2001–2010 period shows that land surface model differences control the majority of spread in the climatological value of SWM, while meteorological forcing differences control the majority of the spread in temporal correlations of SWM anomalies.

1. Introduction

The seasonal cycle of terrestrial snow cover and snow mass has a notable influence on the Northern Hemisphere energy budget, water balance and geochemical cycles. Snow water equivalent (SWE) is expected to respond in a complex way to projected temperature and precipitation changes with the magnitude and sign of the response varying with climate regime and elevation (Brown and Mote 2009). Verification of such responses in climate models and the initialization of snow in seasonal to decadal prediction systems requires a gridded, observational SWE dataset with well-characterized uncertainty (De Lannoy et al. 2010). For snow cover extent, intercomparison of existing data has led to estimation of uncertainties in SCE anomalies and trends (Brown et al. 2010; Derksen and Brown 2012) as well as improved documentation and understanding of systematic differences and inhomogeneities (Brown and Derksen 2013; Mudryk et al. 2014).

A similar quantitative understanding of uncertainties in the Northern Hemisphere is lacking for SWE datasets apart from some more limited comparisons cited below. To address this gap, we compare an ensemble of daily, gridded datasets in order to fully characterize inter-dataset spread and produce a multi-dataset mean. Out intent is to make available the mean and spread of the SWE datasets analyzed here on the National Center for Atmospheric Research Climate Data Guide portal. All of the datasets include observations (e.g. satellite measurements, observed inputs to reanalysis) as at least a component of the data generation, but otherwise draw from a variety of sources including remote sensing, station data, land surface assimilation systems, and reanalysis-driven snow models of varying complexity. In particular we use: (1) the GlobSnow (version 2) analysis, combining satellite-based passive microwave retrievals and ground-based weather station data (Takala et al. 2011); (2) the Global Land Data Assimilation System Version 2 (GLDAS-2) product (Rodell et al. 2004); (3) the European Centre for Medium-Range Forecasts Interim Land

Reanalysis (ERA-I-Land) which uses a simple snow scheme (Balsamo et al. 2013); (4) the Modern Era Retrospective Analysis for Research and Applications (MERRA) which uses an intermediate complexity snow scheme (Rienecker et al. 2011); and, (5) SWE from the Crocus snow scheme, a detailed physical snowpack model driven by meteorology from ERA-Interim (Brun et al. 2013).

Evaluations of the above datasets or of the land surface models and snow schemes used to produce them have been conducted at specific locations (e. g., Wang et al. 2010; Dutra et al. 2012; Brun et al. 1992; Langlois et al. 2009; Stieglitz et al. 2001); however, such local validations do not necessarily represent the datasets' hemisphere-wide fidelity. Indeed, for coarsely gridded datasets such as these, a meaningful hemisphere-scale evaluation with ground measurements is not a trivial undertaking. Single point climate station measurements are inappropriate for validation of coarse grid cell SWE datasets, and datasets with detailed sub-grid measurements over multiple seasons and locations are very limited. Alpine areas, with large sub-grid gradients in elevation and associated snow properties are particularly challenging. Efforts to address such scaling challenges are ongoing, but are outside the scope of this study.

Instead, the objective of this analysis is to exploit the use of multiple datasets to robustly characterize the spatial and temporal agreement in SWE climatologies and anomalies at the hemispheric scale. While the climate modeling community has long recognized the strength in using data from a large number of climate models, such an ensemble-based approach has been less readily adopted by the observational community. Data assessments and intercomparisons have typically focused on identifying the best product. This approach can produce the potentially misleading impression that a single dataset is capable of characterizing the observational truth for all regions and seasons. In reality, variables like SWE are particularly challenging to characterize with coarse resolution gridded datasets due to significant subgrid heterogeneity in horizontal (i.e. snow depth) and vertical (i.e. snow stratigraphy) properties. Recent intercomparisons limited only to snow models

(Rutter et al. 2009) reinforce this perspective, demonstrating that most models are good at simulating certain aspects of observed snow conditions, but less good at simulating others. Therefore, we argue that the analysis of multiple datasets is necessary to understand current uncertainties and inconsistencies within SWE datasets. In turn, this understanding of uncertainty can enhance the use of observational snow analyses. For instance, dataset spread can inform the assignment of uncertainty to observations necessary for land surface data assimilation in numerical weather prediction and seasonal forecast applications (Orsolini et al. 2013; De Lannoy et al. 2010; Koster et al. 2010; Jeong et al. 2013; Drewitt et al. 2012). Consideration of spread also reduces the sensitivity of climate model evaluation to the selection of a single dataset for evaluation (where disagreement among data sets is large enough to affect the significance of SWE biases depending on the choice of observational data).

The remainder of this paper is organized as follows. Details of the SWE datasets and how they are combined into a multi-dataset mean are provided in Section 2. In Section 3, several metrics are used to compare individual datasets with one another as well as the multi-dataset average. Section 4 contains a summary and discussion of key results.

2. Data and Methods

a. Datasets

The SWE datasets used in this study are chosen based on two main criteria: complete Northern Hemisphere spatial coverage (with the exception of an alpine mask applied to GlobSnow) and continuous availability through the satellite era (we use 1981–2010 as our analysis period). We also require relatively homogenous SWE time series for the entire analysis period, which we

diagnose from trends in the time series of global snow mass (see subsec. 2c). The component datasets analyzed in this study are described below and summarized in Table 1.

The GlobSnow (version 2) product Takala et al. (2011, www.globsnow.info) is the only satellite-based product in our analysis; however, it also uses ground-based weather station data in the SWE retrieval. Estimates of snow grain size are first derived for grid cells containing weather station snow depth measurements by optimizing agreement between microwave snow model simulations and observed satellite passive microwave brightness temperatures at 19 and 37 GHz. These local estimates of grain size are interpolated via kriging across the Northern Hemisphere, and used in a second round of emission model simulations for which grain size is fixed and SWE is optimized. Resolution of the product is 25 km. GlobSnow retrievals over complex terrain (defined in subsection 2b) are masked from the standard product due to well known uncertainties related to sub-grid heterogeneity in snow properties and microwave signatures (Tong et al. 2010).

ERA-Interim-Land is a reanalysis product (Balsamo et al. 2013) that diagnoses SWE using the Hydrology Tiled ECMWF Scheme for surface Exchanges over Land (HTESSEL, Balsamo et al. 2009) driven by meteorological forcing from the ERA-Interim atmospheric reanalysis. The snow scheme used is a simple single-layer scheme for dry snow (no liquid water content). Snow density and albedo changes follow closely the formulation proposed by Douville et al. (1995). The precipitation values used to force the land model are corrected using the Global Precipitation Climatology Product version 2.1. Product resolution is $3/4^\circ \times 3/4^\circ$.

MERRA (Rienecker et al. 2011) is a National Aeronautics and Space Administration (NASA) atmospheric reanalysis product generated with version 5.2.0 of the Goddard Earth Observing System (GEOS-5) Atmospheric General Circulation Model and Atmospheric Data Assimilation System (ADAS). SWE is diagnosed from a hydrological catchment-based land surface model (called Catchment, Koster et al. 2000). Catchment uses an intermediate complexity snow scheme

with up to three snow layers describing snow accumulation, melting, refreezing and compaction in response to surface meteorological conditions (Stieglitz et al. 2001). Product resolution is $1/2^\circ \times 2/3^\circ$. In section 2c we also analyze the related product MERRA-Land. This second product is produced by re-running only the land surface component of MERRA forced by atmospheric data from the standard MERRA product except for precipitation which is forced by the gauge-based National Oceanic and Atmospheric Adminstration's (NOAA) Climate Prediction Center "Unified" (CPCU) precipitation dataset. The canonical reference for the MERRA-Land product is Reichle et al. (2011); however, the current working product differs in several ways from the preliminary version described therein, including in the choice of precipitation forcing which Reichle et al. (2011) lists as the Global Precipitation Climatology Project (GPCU) dataset. Henceforth, unless explicitly referred to as MERRA-Land, all references to MERRA data refer to the standard Rienecker et al. (2011) product.

The Crocus SWE dataset is from the Interactions between Soil Biosphere-Atmosphere (ISBA) land surface model driven by ERA-Interim meteorology. The Crocus snow scheme (Brun et al. 2013) is embedded in ISBA in place of the usual snow scheme. Crocus is a detailed snowpack model with multiple historical snow layers possible, each representing a separate snowfall event. Each layer is described by the thickness, temperature, dry density, liquid water content and grain type (dendricity, spericity, size, and age). Resolution of the data is $1^\circ \times 1^\circ$.

GLDAS version 2 (Rodell et al. 2004) is another NASA product that estimates SWE based on the National Centers for Environmental Prediction, Oregon State University, Air Force, and Hydrologic Research Lab (Noah) land surface model version 3.3 (Chen et al. 1996; Koren et al. 1999) constrained by assimilated observations and land surface parameters. This version is forced using the Princeton meteorological forcing dataset (Sheffield et al. 2006). We also use four different GLDAS version 1 products that have been forced using the Global Data Assimilation System

(GDAS) with precipitation adjustments from the Climate Prediction Centers Merged Analysis of Precipitation (CMAP). Because of changes in the forcing data in GLDAS version 1 between 1979 to 2001, the products are unsuitable for analysis over the full 1981–2010 period. However there are no changes to the forcing meteorology after 2001, and we analyze the products over the 2001–2010 period in Section 3c. The four GLDAS-1 products each use different land surface models to diagnose SWE: the Variable Infiltration Capacity (VIC) land surface model (Liang et al. 1994), the Community Land Model (CLM) version 2.0 (Bonan et al. 2002), the Mosaic land surface model (Koster and Suarez 1994), and the Noah land surface model version 2.7. The corresponding snow schemes implemented in these models range from simple single layer schemes in the Noah and Mosaic models to intermediate complexity schemes in CLM and VIC. Resolution of all GLDAS products used is $1^\circ \times 1^\circ$.

b. Methods

For each dataset we acquired daily SWE at the native resolution for the 1981–2010 period. Between 1981 and 1987, GlobSnow is only available approximately every second day with occasional gaps of longer duration. We linearly interpolated any temporal gaps in the data using the two nearest dates with available SWE, assuming SWE is uniformly zero between June 30 and September 7.

Each dataset was interpolated to a regular $1^\circ \times 1^\circ$ longitude-latitude grid. Before interpolating we excluded snow from land ice/glaciers and large lakes based on the MERRA land fraction mask, which specifies the fractional area occupied by land ice or lakes in a given grid cell. We upscaled the MERRA land fraction mask to the native resolution of the other datasets and removed snow from any grid cell containing a nonzero fraction of land ice. Snow over lakes in MERRA was

removed in proportion to the percentage of the grid cell area occupied by lakes. The remaining datasets already mask snow over large lakes, which we considered sufficient for our purposes.

In order to treat alpine and non-alpine regions separately, we also upscaled the 25km-resolution binary alpine mask applied in the GlobSnow data processing chain to the 1° grid. Topographic information for this mask was derived from ETOPO5 (National Geophysical Data, 1988) data which contains global elevation information at a resolution of 5 arc minutes, an appropriate resolution given the 25-km scale of the GlobSnow product. Grid cells are considered mountainous if the standard deviation of the elevation within is larger than 200m. Once interpolated to the native resolution of the remaining SWE datasets, the mask represents the fraction of a given grid cell which is of non-alpine land type. Multiplying the regridded SWE data by the fractional alpine mask provides a second SWE field consisting only of non-alpine SWE; the difference between the two fields represents alpine only SWE. Separating the alpine and non-alpine SWE in this simple manner is equivalent to evenly partitioning SWE based on the proportion of a given land type over grid cells which are not 100% alpine or 100% non-alpine. Separating alpine and non-alpine SWE was necessary firstly because the alpine mask was applied to GlobSnow, but also in order to isolate the relative uncertainty/spread in the datasets over complex terrain (which poses unique challenges due to topographic variability) compared to non-alpine regions. Because of the alpine mask applied to GlobSnow, there is only non-alpine SWE available for this product.

We construct the dataset mean ($\overline{h}$) from the first five datasets listed in Table 1 by averaging the datasets available over a given grid cell accounting for land type as follows:

$$\overline{h} = \frac{1}{N}\sum_{i \neq GS} h_i + \frac{1}{N} h_{GS} + \frac{1}{N(N-1)}\sum_{i \neq GS} h_i f, \tag{1}$$

where h_i represents the SWE value for a particular product at the grid cell, f is the alpine fraction of the grid cell, N is the total number of data products, and h_{GS} refers to the GlobSnow product.

This procedure reduces to an unweighted average of four products (all but GlobSnow) over strictly alpine grid cells and an unweighted average of all five products over strictly non-alpine cells.

We also perform a time series analysis over three land masses: the entire Northern Hemisphere land mass (referred to as NH), the North American continent (NA), and the Eurasian continent (EU). We exclude Greenland from all three definitions. Where possible we further decompose the time series resulting from each of the land masses into three mutually exclusive land-types: mid-latitudes (non-alpine land regions below 60N), Arctic (non-alpine land regions above 60N) and alpine. We determine the total snow water mass (SWM) by summing the equivalent volume of snow water per grid cell over the appropriate spatial domain and converting to mass using the density of pure water. We use these daily time series to calculate climatologies, anomalies, standard deviations and linear trends. Henceforce, we reserve the use of SWM to describe SWE which has been spatially aggregated over a given region to form a time series. When discussing the spatially varying field we will use the term SWE.

c. Diagnosis of temporal discontinuities

For the inclusion of a reanalysis-based dataset in this comparison over the 1981–2010 period, we require a relatively homogenous time series of NH SWM. An acceptable level of homogeneity is empirically diagnosed by comparing the sign and magnitude of trends of NH SWM. On this basis we reject three established datasets (MERRA-Land, CMC and GLDAS-1) for which we find spurious trends. Figure 1 shows how NH SWM trends from these three datasets compare with those from GlobSnow, GLDAS-2, MERRA, ERA-I-Land and Crocus. While the five latter datasets exhibit a range of negative SWM trends (solid lines in Fig. 1), the trends of the three remaining datasets are exceedingly different (broken lines in Fig. 1). To illustrate how trends are indicative of inhomogeneities in global SWM, we show a time series of NH SWM for MERRA-

Land in Figure 2a. Except for the first two snow seasons (which have exceedingly large positive anomalies compared to the rest of the time series), anomalies are almost exclusively negative before 1998 with a negative trend, and positive after 1999 with a weak, positive trend. Examining separate time series for EU and NA (not shown) shows some covariability (12%) but that Eurasian variability dominates the total (64%) compared to NA (12%). Figure 2b shows a spatial map of the difference in climatologies between 1983-1998 and 1999-2010. The latter period has additional SWE over most of Eurasia. Comparing the 1981-1982 climatology to the 1983-2010 climatology also shows additional SWE uniformly across Eurasia (not shown). Further analysis indicates the additional SWE is a result of discontinuities present in the precipitation data used to force the MERRA-Land model. Similar evidence of temporal discontinuities can be found in the other two datasets with spurious trends. In the case of CMC, changes in spatial resolution of the precipitation forcing that was used to drive the analysis (specifically, increased resolution after 2006) result in a noticeable discontinuity in SWM anomalies at that time (not shown). The discontinuity produces the positive trend in SWM seen in Figure 1 because the higher resolution precipitation forcing better resolves deep alpine snow. In GLDAS version 1 (trend shown in Fig. 1 for the Noah LSM only) a discontinuity occurs at the end of 1998 and results in a strongly negative winter season trend roughly 3–4 times larger than that found in any of the other datasets.

Comparing the five principal datasets of our analysis, Crocus, ERA-I-Land and MERRA have more strongly negative NH SWM trends than GlobSnow stemming from more strongly negative trends over Eurasia with comparable or more weakly negative trends over North America (continental breakdown not shown). GLDAS-2 shows much weaker trends through the entire snow season than the other four datasets. There is some consensus for decreasing NH winter SWE over the 1981–2010 period that is consistent with *in situ* trends to less snow cover in many regions of the NH (Vaughan et al. 2013).

3. Results

a. Climatological snow water equivalent

Total NH SWM climatologies are shown in Figure 3a for the five SWE datasets and the multi-dataset mean. A key finding is the large amount of spread in climatological SWM among the products — during the seasonal peak they vary by as much as 50%. In order to display total NH SWM for GlobSnow alongside the other datasets we use the average of the non-Globsnow datasets in alpine regions so that the additional SWM seen in the GlobSnow climatology stems from differences over non-alpine regions.

To better determine which datasets and regions account for the spread among the individual climatologies, we decompose the differences between the individual datasets and the multi-dataset mean by continental domain and land type (Fig. 3b). Three seasonal periods are presented (calculated as two-month averages) corresponding to before-peak (December, January), near-peak (February, March), and after-peak (April, May) SWM. GLDAS-2 shows the largest difference from the multi-dataset mean with less SWM over both continents during all month-pairs and for all land types. The ERA-I-Land product has the largest positive bias relative to the multi-dataset mean, stemming primarily from differences with the other datasets over Eurasia (especially in the Arctic). SWE in GlobSnow peaks earlier in the season than the other datasets. The MERRA and Crocus datasets show the smallest differences from the mean although for Crocus the similarity is in part because of opposite-signed differences in Arctic and alpine regions: Crocus has the most SWM of all the datasets in the alpine regions of both continents, but it has less SWM in the Eurasian Arctic than any of the other datasets except for GLDAS-2.

A close examination of Figure 3b also shows that alpine and Arctic regions generally contribute the most to the differences in climatological SWM, which is consistent with our a priori assump-

tions that SWE is most poorly constrained in these regions. These regions have poor agreement among precipitation data, sparse observing networks, and complex snow processes that occur, including snow redistribution in the Arctic and complex elevational gradients in alpine regions, all of which will increase the uncertainty for the datasets. We show this result more explicitly in Figure 4 which shows the total spread (range of dataset climatologies) in NH SWM according to land type. The spread over Arctic and alpine regions is comparable to one another but roughly 2–3 times larger than that over mid-latitude regions.

We also examine the spatial distribution of the multi-dataset mean and its spread in Figure 5. The climatology compares well with known features of the observed SWE distribution (Brown and Mote 2009). Examining the ratio of climatological SWE to the spread among the component datasets provides a measure of the signal to noise ratio. Regions with a ratio greater than one coincide approximately with the boreal forest regions of North America and Eurasia. That we have the best agreement in the datasets over these regions is reasonable for several reasons. These are regions where snow cover has well-defined start and end points to the season and where midlatitude winter cyclones are well-represented in numerical weather prediction models. Further south of this zone, differences in air temperature will have increasingly important impacts on accumulation through differences in snow-on dates and rain/snow separation. Further north of this zone uncertainty in the precipitation forcing increases. Furthermore, because the taiga snowpack is generally spatially homogeneous (Sturm et al. 1995), wind redistribution and topographic effects are minimal. The relative consistency of the datasets across the NH boreal forest is further supported by other measures of agreement among the datasets (discussed later, see Figs. 8 and 11).

b. Snow water equivalent variability

We present time series of anomalous total NH non-alpine SWM in Figure 6 (anomalies calculated separately for each dataset using its own climatology and individually detrended). Some years show good agreement in the sign and evolution of SWM anomalies over the winter (e.g. 1986/87, 1990-1992, 2001/02) while other years show larger spreads in the anomalies (1988/89, 1996/97, 2009/10). Note the datasets evident as an outlier for a given time period varies among all 5 component datasets (i.e. GLDAS-2 in 1989; Crocus in 1993; GlobSnow in 1994; MERRA in 1999; ERA-I-Land in 2003).

In order to determine the strength of agreement between datasets, we present pair-wise correlations between individually detrended SWM time series in Figure 7. Each value is determined by correlating the detrended SWM anomaly time series for a given pair of datasets for all winter days (NDJFMA) over the entire 1981–2010 period (such that each time series for a given pair contains 181×30 days). Calculated as such, correlations reflect both intraseasonal and interannual covariance. Slightly different approaches were also evaluated: (1) determine the intraseasonal correlation of winter season days (NDJFMA) for each year separately and average the thirty resulting correlations; and (2) calculate the interannual correlation for each calendar day and average the correlations over the winter season (NDJFMA), but the results are largely insensitive to the method (within 5%) and the ranking of which datasets correlate best with one another shifts only slightly among pairs involving GLDAS-2.

Because correlations involving GlobSnow only consider non-alpine snow, the datasets have been ranked by their correlations over non-alpine regions only, which differ slightly from those that consider all land types. Except for two pairs, using non-detrended SWM time series affects the correlations by less than 5%. The correlations that are most strongly affected by detrending involve pairs of datasets with the smallest and largest trends — GLDAS-2 with either MERRA or Crocus — and these differences in correlation stem from differences over Eurasia rather than

North America (not shown). Examining separately the time series for each continent, we note that all datasets correlate better over NA than over EU. GLDAS-2 shows the weakest correlation with other datasets. GlobSnow has slightly higher correlations, and the remaining three reanalysis-based datasets are correlated with one another at $\sim 80\%$. An increased correlation among these three datasets is expected since MERRA and Crocus both use the same forcing meteorology and the ERA-Interim meteorology is itself well correlated with that of MERRA. This point is made in Rienecker et al. (2011) and further investigated in Section 3c. Finally we also show the mean SWE pattern correlation for each dataset pair in Figure 7 (calculated daily and averaged over all winter-time days of the thirty year period). The mean pattern correlation is lower than the corresponding temporal correlation of total snow mass. This result may be due to the presence of opposite-signed spatial biases that cancel when spatially aggregated into a SWM time series.

Figure 8 presents a spatial map of the temporal correlation. For this figure, rather than spatially aggregating snow water before correlating, we calculate the temporal correlation between pairs of SWE datasets for each grid cell individually. Values plotted are the average among all possible pairs of datasets (for alpine regions, the correlations shown are the average of the six dataset pairs that exclude GlobSnow, and for non-alpine regions they are average of all ten pairs). Interestingly, alpine regions don't have noticeably weaker correlations than neighboring regions despite the comparatively large climatological spread indicated in Figure 4. Correlations are lower in Arctic regions and the marginal snow zones and peak over the taiga, consistent with the high signal-to-noise present in the climatology (Fig. 5b).

Despite the reasonable spatial and temporal correlations shown in Figures 7 and 8, the inter-dataset spread of SWM anomalies is comparable to the interannual variability. This result can be approximately assessed from an examination of Figure 6 but is shown explicitly in Figure 9a, where for each day over the 1981–2010 period, anomaly spread was calculated as the difference

between the datasets with the maximum and minimum SWM anomaly. This comparability in magnitude means that on average at least one of the datasets fails to agree on the sign of the anomaly. When the same analysis is applied to non-detrended data, it is apparent that differences in trends among the datasets are not responsible for the majority of the spread.

We also quantify the relative contribution of each dataset to the total spread, defined as the *attributable spread*. To calculate this quantity for each grid cell on a given day and year, we partition the total spread between just two of the five datasets: those with the maximum and minimum anomaly on that day. The magnitude of spread attributed to each is the difference between each dataset's anomaly and the multi-dataset mean. This definition allows a particular dataset to accrue more of the total spread for the given day if it is further from the mean value. The remaining datasets are attributed no spread for the given day, but may be attributed spread on other days. This definition may be extended to a time series comprising any number of datasets as long as only two component datasets are attributed the full spread on a given day (as expected in the case with only two datasets, each would always receive half of the total spread). By definition, the sum of attributed spread across the total number of component datasets is equal to the total spread. Figure 9b shows the amount of spread attributed to each of the five component datasets averaged over the 1981–2010 period. In order to account for alpine regions in GlobSnow, the mean of the remaining four datasets was used to fill in alpine SWM in GlobSnow. This accounting means that spread is attributed to GlobSnow only when it stems from SWE differences in non-alpine regions.

Ranked in sequence, MERRA is attributed the least spread, followed by Crocus, ERA-I-Land and GLDAS-2 while GlobSnow is attributed the most. While the first three datasets constitute roughly half of the total spread, GLDAS-2 and GlobSnow together are responsible for the remaining half, each contributing approximately one quarter of the total. Note that ERA-I-Land's attributable spread peaks during the spring and that GlobSnow has relatively low attributable spread

during the spring, but a large amount during the fall and early winter. We may understand these two results by considering that attributable spread convolves the frequency that a given dataset is an outlier with its distance from the mean anomaly. Figure 10a demonstrates that ERA-I-Land's standard deviation (which will be related to its distance from the mean dataset anomaly) peaks during the spring while that of GlobSnow peaks during the fall and early winter. Each of these two datasets are also more frequently outliers during these seasons (as opposed to Crocus which shows a similar spring peak in interannual variability but is not a frequent outlier). The combination of these two traits means each dataset accrues more spread during the time of year that it shows increased variability. It is possible to account for these separate effects by standardizing the time series and examining the spread in the standardized time series (Fig. 10b). The spread now attributed to each dataset is more constant throughout the snow season (total spread of all datasets is around 2 standard deviations) and more closely reflects the proportion of days and years that the particular dataset is an outlier. Calculated as such, GlobSnow, Crocus, MERRA and ERA-I-Land all contribute approximately equally to the total spread; by contrast, GLDAS-2 contributes approximately twice as much to the total spread.

The spatial distribution of signal to noise for SWE anomalies is shown in Figure 11a. Consistent with the relative magnitudes of SWM anoamlies and spread shown in Figure 9a, the signal-to-noise ratio of the local SWE distribution is less than one nearly everywhere. It is larger in the boreal forest region than elsewhere consistent with results for climatological SWE and for temporal correlations of daily SWE. The signal-to-noise ratio among only the MERRA, ERA-I-Land and Crocus datasets is substantially larger indicating better agreement among these three datasets (Fig. 11b).

c. Relative influence of land model and meteorological forcing on reanalysis-based datasets

Finally, we examine the relative influence of differences in the precipitation forcing versus differences in the treatment of snow processes in the land model on the resulting correlation and spread of SWM time series. For this analysis we use two separate groups of reanalysis-based datasets. The first of these two groups (Group 1) contains the MERRA, ERA-I-Land and Crocus datasets. We consider these datasets to form a single group because they use meteorological forcing from either the MERRA or ERA-Interim reanalyses, which have been shown to be well-correlated with one another (Rienecker et al. 2011). The second group (Group 2) contains four GLDAS-1 products. Each of these datasets has been forced using the same GDAS forcing data, however each one uses a different land model as described in Section 2 and Table 1. Because the forcing data for GLDAS-1 contains changes over the 1981–2010 period, we restrict our analysis to the 2001–2010 period over which the forcing data is consistent.

Figure 12a shows that the spread in climatologies within each of the two groups is affected by the particular land models that are included. The spread shown in the grey shading (Group 1 plus GLDAS-2) is comparable to that found among the Group 2 products and remains comparable with the exclusion of the Noah-associated products from both groupings (but is decreased by about half). Note that both versions of GLDAS that use the Noah land surface model are outliers with similar (but especially low) SWM climatologies, despite using different meteorological forcing. This shows it is possible to obtain similar climatologies from the same land surface model while using different meteorological forcing. We also see that using the same meteorological forcing but different land models (Group 2) results in a large climatological spread. These results imply that differences among the land models generate the majority of the climatological spread. We examine the effect on the correlation of the SWM time series in Figure 12b. Datasets within each of the two

groups correlate very well with one another over all continental domains and over all land types. However the correlations of SWM time series between the two groups are substantially lower, especially over alpine and Arctic regions. This result suggests that differences in the meteorological forcing exert a larger influence on the resulting SWM correlation than differences in the details of the land-model used to produce the data. We also note that agreement with the GlobSnow product (an independent estimate of SWM) is higher for Group 1 datasets (circles, right column) than for Group 2 datasets (diamonds, right column). Since the meteorological forcing data for Group 1 is more recent, this may represent an improvement in the accuracy of the more recent reanalyses.

4. Conclusion

We have presented a comparison of five daily, gridded Northern Hemisphere SWE datasets over the 1981–2010 period. Our intent is to make available the multi-dataset mean and corresponding spread of both the climatology and anomalies as part of the National Center for Atmospheric Research (NCAR) Climate Data Guide (https://climatedataguide.ucar.edu/climate-data). Our analysis has shown that the individual datasets exhibit a large amount of spread in their total snow water mass (SWM) climatologies (Figs. 3 and 4) as well as their anomalies (Fig 9a). Despite the large spread, the SWM time series show moderate to good correlations with one another, approximately 85% for the three datasets using modern reanalysis meteorological forcings; these correlations are higher over North America than over Eurasia (Fig. 7). Boreal regions not only have the lowest relative amount of spread (highest signal-to-noise) for both climatological SWE (Fig. 5) and SWE anomalies (Fig. 11), but these regions also have the highest temporal correlation among the SWE datasets (Fig. 8). We have also examined the relative influence of the particular land surface model compared to the choice of meteorological forcing using a suite of reanalysis-derived datasets. The former accounts for the majority of the spread in the climatologies while the latter exert a larger

influence on the resulting SWM correlation (Fig. 12). More modern reanalysis-derived datasets also show improved correlation with the satellite product, GlobSnow, compared to the previous generation (Fig. 12).

This analysis of gridded SWE data has yielded important insights on the amount of spread, and the strength of spatial and temporal correlation between independent SWE datasets. At this point, the relatively high agreement between some groups of datasets (i.e. modern reanalysis-derived datasets such as MERRA, ERA-I-Land, and Crocus) does not imply lower bias compared to the ground truth, only that these datasets are generally consistent with each other. The agreement between these three datasets is not surprising given the commonalities in how SWE is derived: modern reanalysis meteorology and high quality precipitation forcing coupled with state of the art land surface models. A product like GlobSnow adopts an entirely different approach by blending *in situ* snow depth observations and satellite passive microwave measurements through the use of a microwave snow emission model. Going forward the challenge to the community is how to combine these unique perspectives with their respective strengths and sources of error.

Given the spread in climatology, these results highlight the sensitivity of climate analysis to the selection of an individual SWE dataset for model evaluation from the available pool of data. For example, evaluation of simulated SWE with ERA-I-Land (highest pre-melt SWE) versus GLDAS (lowest pre-melt SWE) could lead to entirely different interpretations of model performance and bias. There are also seasonal and regional differences which are also important for users to consider. For example, ERA-I-Land has very different biases over NA compared to EU, Crocus has different biases for Arctic versus alpine snow, while GlobSnow exhibits different seasonality than the multi-dataset mean.

For applications like climate model evaluation, it is straightforward to see the value in using an ensemble of SWE datasets for evaluation of a multi-model ensemble of simulations. This ap-

proach would illustrate the overlap (or lack thereof) in simulated climate realizations against the uncertainties in characterizing our current climate, which is evident when a single dataset is used to characterize reality (for instance Derksen and Brown 2012). For land surface data assimilation applications, an ensemble approach could be used to statistically characterize observational uncertainty, an important requirement for the assimilation of the model first guess with observations.

What remains to be determined is how to select the SWE datasets for inclusion in an observational ensemble. Thresholds based on attributable spread or detrended anomaly correlations could be the basis for selecting a sub-group of datasets, but this must be done with caution. In the absence of an evaluation with representative ground measurements, agreement between a subset of datasets does not necessarily represent better overall accuracy in representing reality. Comparisons with independent reference measurements in order to determine the dataset bias relative to ground truth are ultimately necessary and currently underway within the European Space Agency's Satellite Snow Product Intercomparison and Evaluation Experiment (SnowPEX). In advance of a comparison with high quality *in situ* reference datasets quantifying the spread between different available products, as was accomplished in this study, is an important step in informing users of the level of spatial and temporal agreement between products and the relationship of individual datasets to the multi-dataset mean.

Acknowledgments. We acknowledge funding from the Natural Sciences and Engineering Research Council of Canada's Climate Change and Atmospheric Research initiative via the Canadian Sea Ice and Snow Evolution Network. We also thank Eric Brun for providing data from the Crocus snowpack model for this analysis.

References

Balsamo, G., A. Beljaars, and K. Scipal, 2009: A revised hydrology for the ECMWF Model: Ver-

ification from field site to terrestrial water storage and impact in the integrated forecast system. *J. Hydrometeor.*, **10**, 623–643, doi:10.1175/2008JHM1068.1.

Balsamo, G., and Coauthors, 2013: ERA-Interim/Land: a global land water resources dataset. *Hydrol. Earth Syst. Sci. Discuss.*, **10**, 14 705–14 745, doi:10.5194/hessd-10-14705-2013.

Bonan, G. B., K. W. Oleson, M. Vertenstein, S. Levis, X. Zeng, Y. Dai, R. E. Dickinson, and Z.-L. Yang, 2002: The Land Surface Climatology of the Community Land Model Coupled to the NCAR Community Climate Model. *J. Climate*, **15**, 3123–3149, doi:10.1175/1520-0442(2002)015⟨3123:TLSCOT⟩2.0.CO;2.

Bransnett, B., 1999: A global analysis of snow depth for numerical weather prediction. *J. Appl. Meteorol.*, **38**, 726–740, doi:10.1175/1520-0450(1999)038⟨0726:AGAOSD⟩2.0.CO;2.

Brown, R., C. Derksen, and L. Wang, 2010: A multi-data set analysis of variability and change in Arctic spring snow cover extent, 1967-2008. *J. Geophys. Res.*, **115 (D14)**, D16111, doi:10.1029/2010JD013975.

Brown, R. D., and C. Derksen, 2013: Is Eurasian October snow cover extent increasing? *Environ. Res. Lett.*, **8**, 024 006, doi:10.1088/1748-9326/8/2/024006.

Brown, R. D., and P. W. Mote, 2009: The Response of Northern Hemisphere Snow Cover to a Changing Climate. *J. Climate*, **22**, 2124, doi:10.1175/2008JCLI2665.1.

Brun, E., P. David, M. SUDUL, and G. Brunot, 1992: A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting. *J. Glaciol.*, **38**, 13–22.

Brun, E., V. Vionnet, A. Boone, B. Decharme, Y. Peings, R. Valette, F. Karbou, and S. Morin, 2013: Simulation of Northern Eurasian Local Snow Depth, Mass, and Density Using a Detailed

Snowpack Model and Meteorological Reanalyses. *J. Hydrometeor.*, **14**, 203–219, doi:10.1175/JHM-D-12-012.1.

Chen, F., and Coauthors, 1996: Modeling of land surface evaporation by four schemes and comparison with fife observations. *J. Geophys. Res.*, **101 (D3)**, 7251–7268, doi:10.1029/95JD02165.

De Lannoy, G. J. M., R. H. Reichle, P. R. Houser, A. K. R., N. E. Verhoest, and R. N. Pauwels, 2010: Satellite-scale snow water equivalent assimilation into a high-resolution land surface Model. *J. Hydrometeor.*, **11**, 352–369, doi:10.1175/2009JHM1192.1.

Derksen, C., and R. Brown, 2012: Spring snow cover extent reductions in the 2008–2012 period exceeding climate model projections. *Geophys. Res. Lett.*, **39**, L19504, doi:10.1029/2012GL053387.

Douville, H., J.-F. Royer, and J.-F. Mahfouf, 1995: A new snow parameterization for the météo-france climate model. *Climate Dyn.*, **12**, 21–35, doi:10.1007/BF00208760.

Drewitt, G., A. A. Berg, W. J. Merryfield, and W. Lee, 2012: Effect of realistic soil moisture initialization on the Canadian CanCM3 seasona forecast model. *Atmos.–Ocean*, **50**, 466–474, doi:10.1080/07055900.2012.722910.

Dutra, E., P. Viterbo, M. A. M. Pedro, and G. Balsamo, 2012: Complexity of snow schemes in a climate model and its impact on surface energy and hydrology. *J. Hydrometeor.*, **13**, 521–538, doi:10.1175/JHM-D-11-072.1.

Jeong, J., H. W. Linderholm, S. Woo, C. Folland, B. Kim, S. Kim, and D. Chen, 2013: Impacts of snow initialization on subseasonal forecasts of surface air temperature for the cold season. *J. Climate*, **26**, 1956–1972, doi:10.1175/JCLI-D-12-00159.1.

514 Koren, V., J. Schaake, K. Mitchell, Q.-Y. Duan, F. Chen, and J. M. Baker, 1999: A parameteriza-
515 tion of snowpack and frozen ground intended for ncep weather and climate models. *J. Geophys.*
516 *Res.*, **104**, 19 569–19 585, doi:10.1029/1999JD900232.

517 Koster, R. D., and M. J. Suarez, 1994: The components of a 'SVAT' scheme and their effects
518 on a GCM's hydrological cycle. *Adv. in Water Resour.*, **17**, 61–78, doi:10.1016/0309-1708(94)
519 90024-8.

520 Koster, R. D., and Coauthors, 2010: Contribution of land surface initialization to subseasonal
521 forecast skill: First results from a multi-model experiment. *Geophys. Res. Lett.*, **37**, L02402,
522 doi:10.1029/2009GL041677.

523 Langlois, A., and Coauthors, 2009: Simulation of snow water equivalent (swe) using ther-
524 modynamic snow models in québec, canada. *J. Hydrometeor.*, **10**, 1447–1463, doi:10.1175/
525 2009JHM1154.1.

526 Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based
527 model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.*,
528 **99 (D7)**, 14 415–14 428, doi:10.1029/94JD00483.

529 Mudryk, L. R., P. J. Kushner, and C. Derksen, 2014: Interpreting observed northern hemisphere
530 snow trends with large ensembles of climate simulations. *Climate Dyn.*, **43**, 345–359, doi:10.
531 1007/s00382-013-1954-y.

532 Orsolini, Y. J., R. Senan, G. Balsamo, F. J. Doblas-Reyes, F. Vitart, A. Weisheimer, A. Carrasco,
533 and R. E. Benestad, 2013: Impact of snow initialization on sub-seasonal forecasts. *Climate*
534 *Dyn.*, **41**, 1969–1982, doi:10.1007/s00382-013-1782-0.

Reichle, R. H., R. D. Koster, G. J. M. de Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Touré, 2011: Assessment and Enhancement of MERRA Land Surface Hydrology Estimates. *J. Climate*, **24**, 6322–6338, doi:10.1175/JCLI-D-10-05033.1.

Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, doi:10.1175/JCLI-D-11-00015.1.

Rodell, M., and Coauthors, 2004: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, doi:10.1175/BAMS-85-3-381.

Rutter, N., and Coauthors, 2009: Evaluation of forest snow processes models (snowmip2). *J. Geophys. Res.*, **114 (D6)**, n/a–n/a, doi:10.1029/2008JD011063, d06111.

Sheffield, J., G. Goteti, and E. F. Wood, 2006: Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling. *J. Climate*, **19**, 3088, doi:10.1175/JCLI3790.1.

Stieglitz, M., A. Ducharne, K. Koster, and M. Suarez, 2001: The impact of detailed snow physics on the simulation of snow cover and subsurface thermodynamics at continental scales. *J. Hydrometeor.*, **2**, 228–242.

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

Takala, M., K. Luojus, J. Pulliainen, C. Derksen, L. Lemmetyinen, J.-P. Kärnä, and J. Koskinen, 2011: Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements. *Remote Sens. Environ.*, **115**, 35173529, doi:10.1016/j.rse.2011.08.014.

Tong, J., S. J. Déry, P. L. Jackson, and C. Derksen, 2010: Testing snow water equivalent retrieval algorithms for passive microwave remote sensing in an alpine watershed of western Canada. *Can. J. Remote Sens.*, **36**, S74–S86, doi:10.5589/m10-009.

Vaughan, D. G., and Coauthors, 2013: Observations: Cryosphere Supplementary Material. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley, Ed., Available from www.climatechange2013.org and www.ipcc.ch.

Wang, Z., X. Zeng, and M. Decker, 2010: Improving snow processes in the noah land model. *J. Geophys. Res.*, **115 (D20)**, doi:10.1029/2009JD013761.

LIST OF TABLES

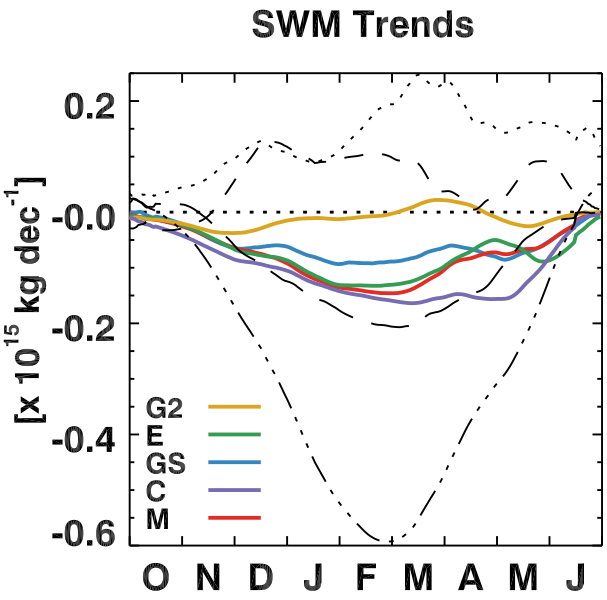
Dataset	Abbreviation	Snow Scheme	Land Model	Forcing Data	Resolution	Reference
GlobSnow	GS	satellite passive microwave + *in situ*[1]			25 km	Takala et al. (2011)
ERA-I-Land	E	simple	HTESSEL	ERA-Interim	$3/4^\circ \times 3/4^\circ$	Balsamo et al. (2013)
MERRA	M	intermediate	Catchment	MERRA	$1/2^\circ \times 2/3^\circ$	Rienecker et al. (2011)
Crocus	C	complex	ISBA	ERA-Interim	$1^\circ \times 1^\circ$	Brun et al. (2013)
GLDAS-2	G2	simple	Noah 3.3	Princeton Met.	$1^\circ \times 1^\circ$	Rodell et al. (2004)
GLDAS-1*	G1n	simple	Noah 2.7	GDAS+CMAP	$1^\circ \times 1^\circ$	Rodell et al. (2004)
	G1m	simple	Mosaic			
	G1v	intermediate	VIC			
	G1c	intermediate	CLM			
Can. Met. Centre*	CMC	simple + *in situ*[2]		GEM	35km	Bransnett (1999)
MERRA-Land*	ML	intermediate	Catchment	MERRA	$1^\circ \times 1^\circ$	Reichle et al. (2011)

TABLE 1. Summary of products referred to in this study. The first five datasets are used in developing the multi-dataset mean over the 1981–2010 period. The remaining datasets (marked with $*$) are established SWE datasets that meet the NH domain criteria but which contain temporal discontinuities as analyzed in Section 2c that compromise their use in the multi-dataset mean. The four GLDAS-1 datasets are suitable for analysis over the restricted, 2001–2010 period (see Sec. 3c).

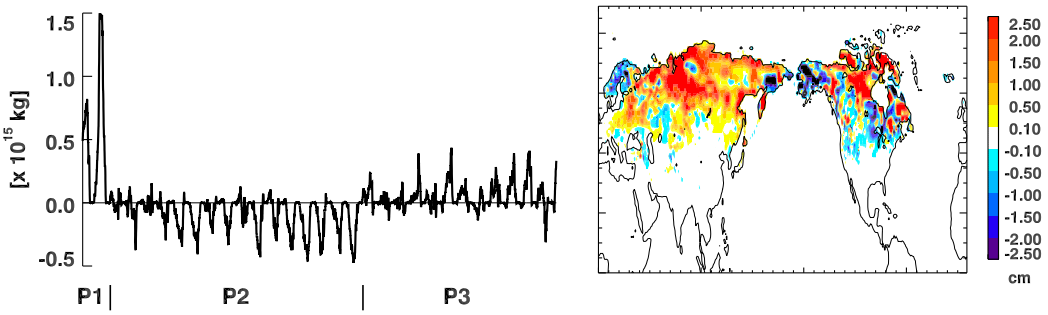
[1] GlobSnow is based on combined information from satellite passive microwave retrievals and *in situ* observations from weather stations. See text for details.

[2] CMC computes snow depths based on combined information from *in situ* observations and a simple snow scheme, driven by temperature and precipitation from the Global Environmental Multiscale Model (GEM). Depths are converted to SWE using climatological snow density information.

LIST OF FIGURES



649 FIG. 1. Trends in NH SWM for the five principal datasets over 1981–2010 (colors as marked). The aver-
650 age trend of the four reanalysis-derived datasets over alpine regions has been added to the trend of the Glob-
651 Snow product in order to compare trend magnitudes. Also shown are trends for CMC (1999–2010, dotted),
652 GLDAS(Noah) version 1 (1981–2010, dot-dash), MERRA-Land (1983–1998, long-dash) and MERRA-Land
653 (1999–2010, short dash).



654 FIG. 2. a) Time series of anomalous NH SWM from MERRA-Land. b) Difference in climatological SWE
655 between 1999-2010 and 1983-1998 periods for MERRA-Land.

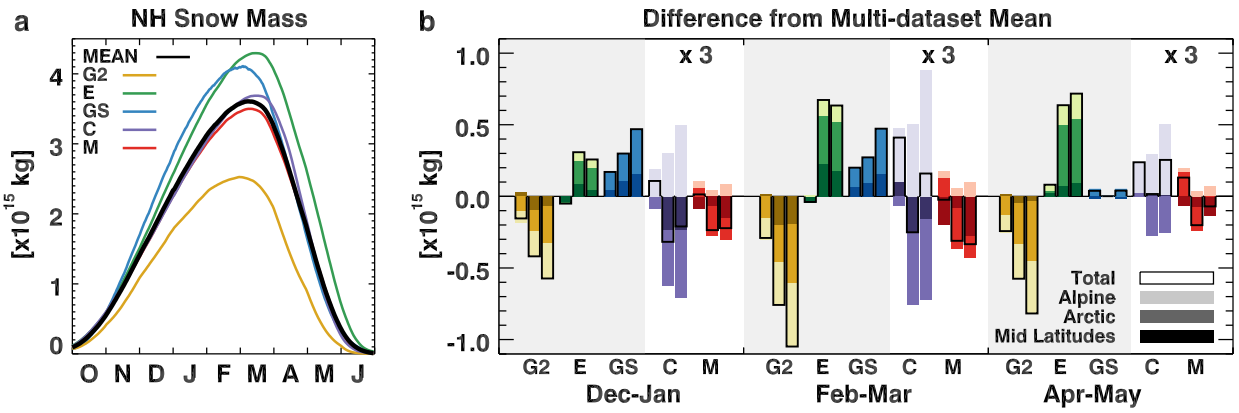


FIG. 3. a) Climatological total NH SWM for individual datasets and the multi-dataset mean for the 1981–2010 period. The average of the NH alpine SWM time series from the four reanalysis datasets is added to the GlobSnow time series in order to appropriately compare total NH SWM. b) Stacked bar chart of SWM differences between individual datasets and the multi-dataset mean over pairs of months for various regions. Each individual dataset is colored as in panel a) and for a given season shows a sequence of three bars corresponding to NA, EU and NH regions. Separate contributions from the three land-types are stacked vertically using different shades for mid-latitudes, Arctic and alpine regions; total difference summed over all land types on a given continent is shown by the bar with the black outline. Crocus and MERRA differences have been multiplied by three; GLDAS-2, ERA-I-Land and GlobSnow differences are unaltered.

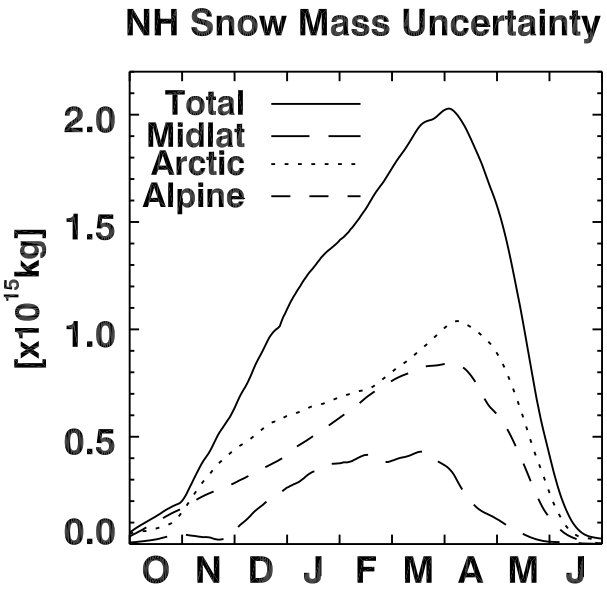
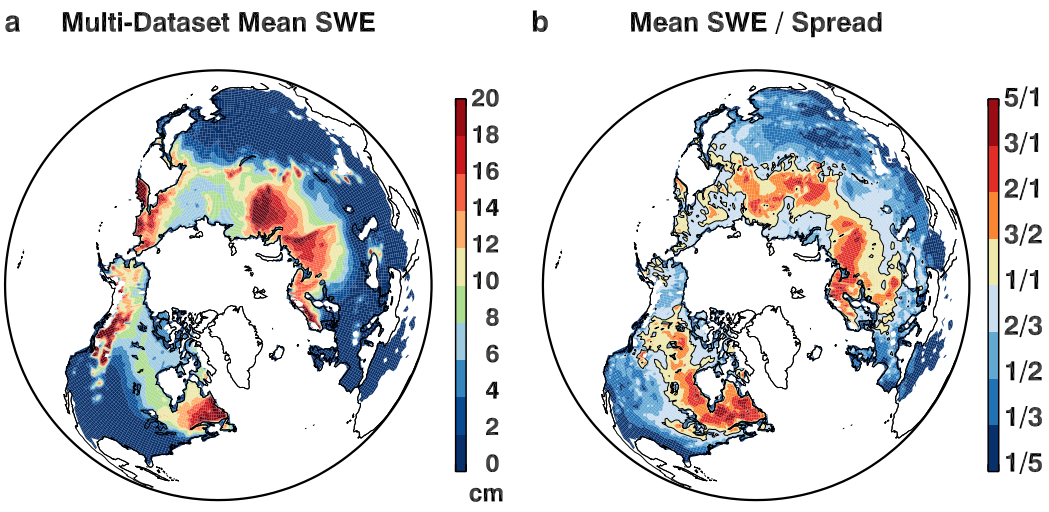
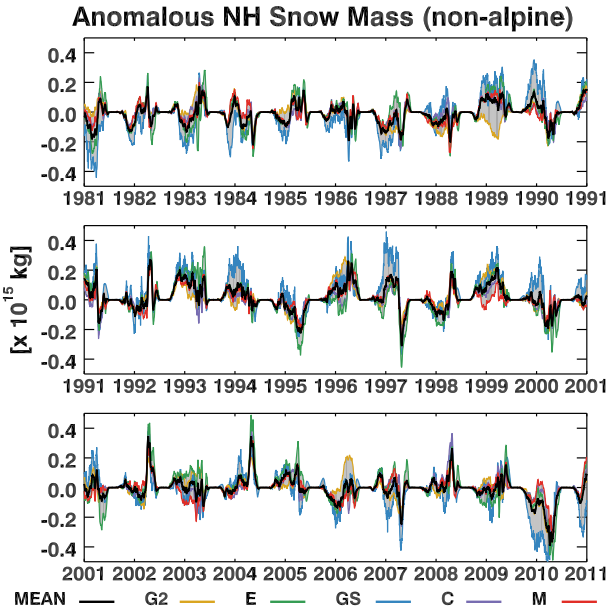


FIG. 4. Spread among climatologies for NH SWM by region over the 1981–2010 period.



665 FIG. 5. a) Climatology of multi-dataset mean SWE for February-March over 1981–2010 period. b) Ratio of
666 climatological SWE to spread among the five component datasets calculated for February-March over 1981–
667 2010 period. The black contour delineates the 1:1 ratio.



668 FIG. 6. a) Anomalous total NH non-alpine SWM for the five individual datasets (individually detrended).

669 Grey shading marks the range of anomalies (spread) among the datasets on a given day.

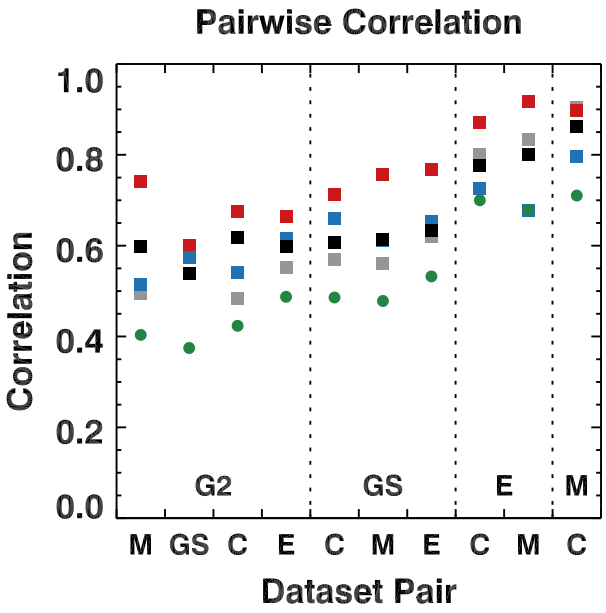


FIG. 7. a) Correlations of daily (NDJFMA) SWM time series (1981–2010) between pairs of datasets. Correlations are shown for detrended time series of NH (black), NA (red) and EU (blue) SWM as well as non-detrended NH (grey) SWM. Also shown are mean values of the SWE pattern correlation between dataset pairs (green circles), calculated daily and averaged over the NDJFMA season and 1981–2010 period.

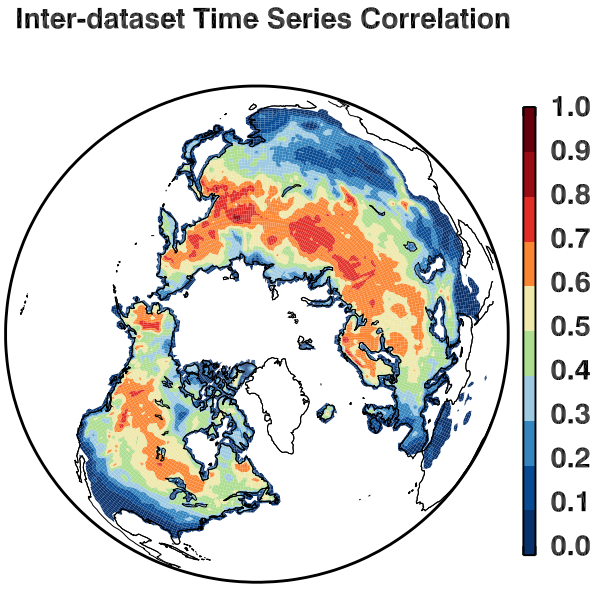


FIG. 8. a) Spatial map of the mean correlation between pairs of daily (NDJFMA) SWE time series (1981–2010). Correlations of detrended time series are performed for a given grid cell location between all available pairs of datasets and averaged together.

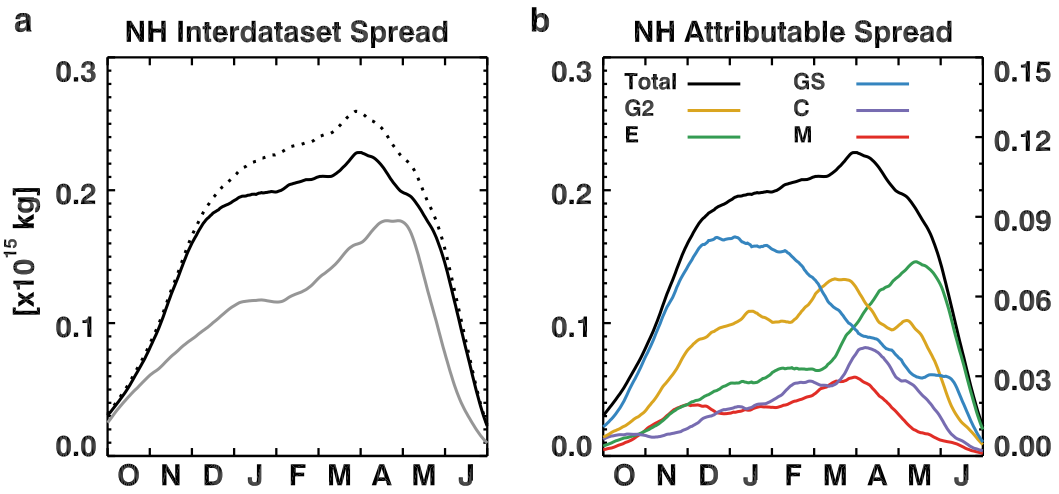
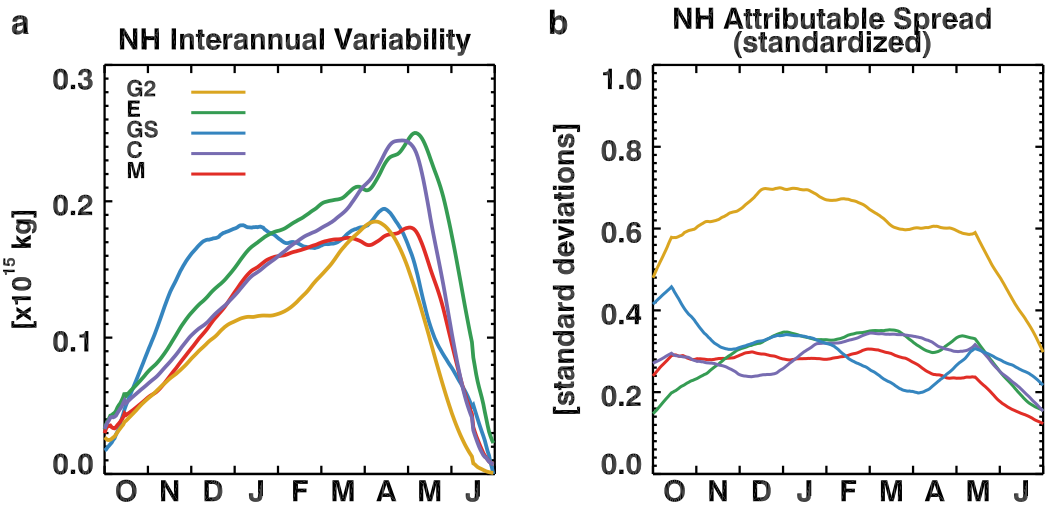


FIG. 9. a) Spread in anomalous SWM among detrended datasets (black, solid), non-detrended datasets (black, dotted) and interannual variability of multi-dataset mean (grey, solid) over the 1981–2010 period. b) Spread in anomalous SWM attributable to each of the five component dataset time series (1981–2010). Total spread (black) is measured on left axis. Spread attributable to individual datasets measured on right axis (2× smaller). A smoothing window of 30 days has been applied to all curves for clarity. See text for definitions of spread and attributable spread.



683 FIG. 10. a) NH interannual variability of SWM for each dataset. b) Attributable spread calculated from
684 standardized SWM for each dataset. A smoothing window of 30 days has been applied to all curves for clarity.

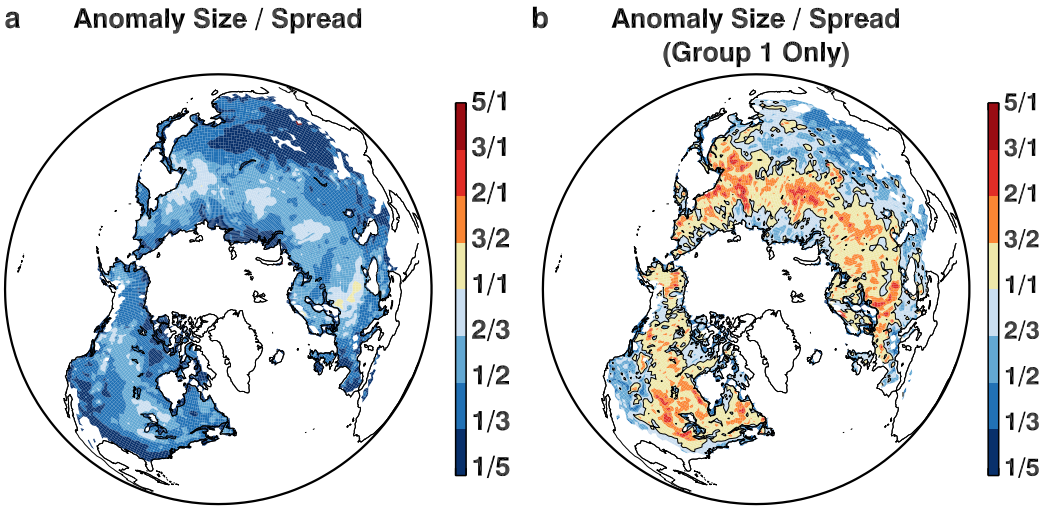


FIG. 11. a) Ratio of absolute anomaly size for multi-dataset mean to dataset spread, calculated daily and averaged over February-March and 1981–2010 period. b) Same as a) but only for the MERRA, ERA-I-Land and Crocus datasets (Group 1, see Sec. 3c). The black contour delineates the 1:1 ratio.

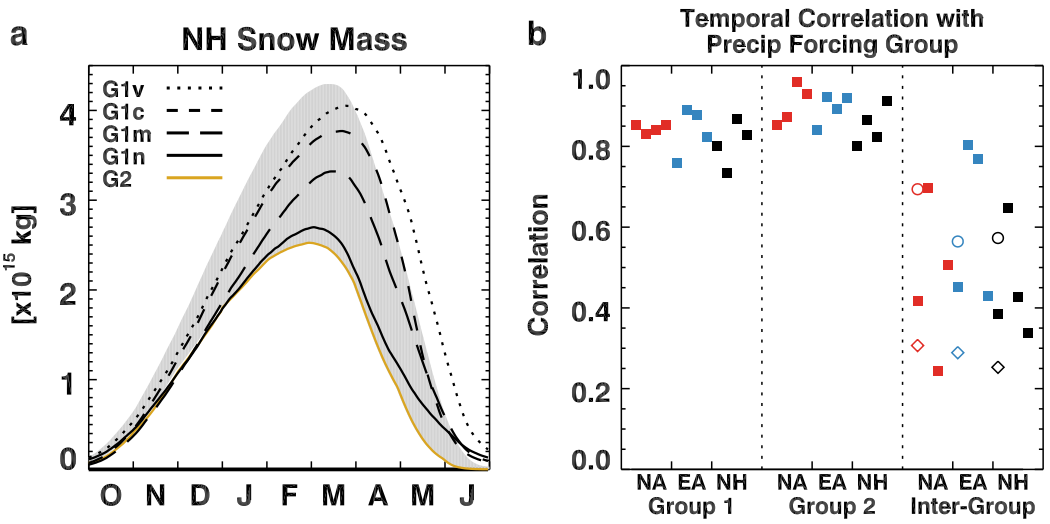


FIG. 12. a) Spread in total NH SWM for Group 1 plus GLDAS-2 (grey shading). Climatologies for the four GLDAS-1 products (Group 2) are shown in the labeled curves. The single GLDAS-2 product is also shown for comparison. b) Mean pairwise correlation among different groups of SWM time series (2001–2010). For each of the three groupings the average correlation is calculated for NA (red), EU (blue) and NH (black) SWM time series. For each land mass and group, four averages are shown in sequence for SWM summed over all land types, only mid-latitude, only the Arctic and only alpine regions. Mean correlations of group 1 (circles) and group 2 (diamonds) total NH SWM time series (all land types) with GlobSnow are also shown. Correlations with GlobSnow over individual land types are omitted for clarity.