Why Do Global Reanalyses and Land Data Assimilation Products Underestimate Snow Water Equivalent?

PATRICK D. BROXTON, XUBIN ZENG, AND NICHOLAS DAWSON

Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, Arizona

(Manuscript received 2 March 2016, in final form 10 August 2016)

ABSTRACT

There is a large uncertainty of snow water equivalent (SWE) in reanalyses and the Global Land Data Assimilation System (GLDAS), but the primary reason for this uncertainty remains unclear. Here several reanalysis products and GLDAS with different land models are evaluated and the primary reason for their deficiencies are identified using two high-resolution SWE datasets, including the Snow Data Assimilation System product and a new dataset for SWE and snowfall for the conterminous United States (CONUS) that is based on PRISM precipitation and temperature data and constrained with thousands of point snow observations of snowfall and snow thickness. The reanalyses and GLDAS products substantially underestimate SWE in the CONUS compared to the high-resolution SWE data. This occurs irrespective of biases in atmospheric forcing information or differences in model resolution. Furthermore, reanalysis and GLDAS products that predict more snow ablation at near-freezing temperatures have larger underestimates of SWE. Since many of the products do not assimilate information about SWE and snow thickness, this indicates a problem with the implementation of land models and pinpoints the need to improve the treatment of snow ablation in these systems, especially at near-freezing temperatures.

1. Introduction

Accurate representation of the amount of seasonal snow on the land surface is important because seasonal snow cover has a large influence on land–atmosphere interactions (Robinson and Kukla 1985). For example, because of its high albedo, snow influences how much solar energy is absorbed by the land surface, and further, it affects the turbulent exchange of water and energy between the land surface and the atmosphere, as well as the transfer of heat to the subsurface (due to its low thermal conductivity). Hydrologic simulation also depends critically on correct representation of snow because the hydrology of cold regions is heavily influenced by annual snowmelt (Bowling et al. 2003; Bales et al. 2006). Incorrect representation of snow water

equivalent (SWE) and subsequent snowmelt may also have lasting impacts after the snow has melted. For example, the timing and magnitude of snowmelt influences streamflow, soil moisture, and vegetation health well into the summer. In addition, there can be a connection between snow and summertime rainfall (Hahn and Shukla 1976; Gutzler and Preston 1997).

There remains substantial uncertainty about the representation of snow on the ground in many reanalysis and Global Land Data Assimilation System (GLDAS; Rodell et al. 2004) products, as evidenced by the wide spread of SWE and snow depth simulations in such systems (e.g., Mudryk et al. 2015). Incorrect representation of SWE might lead to incorrect representation of climate variables, even subsequent to snowmelt. For example, an incorrect amount of snowmelt may influence simulation of runoff and soil moisture well after the snow has melted. In some cases, these variables are used for applications outside of the specific reanalysis or GLDAS model. For example, soil moisture from GLDAS has been used to initialize global models for subseasonal forecasts (Koster et al. 2004; De Goncalves et al. 2006) and to supplement observational soil moisture datasets for hydrologic land–atmosphere coupling studies (e.g., Zhang et al. 2008; Syed et al. 2008; Rodell

et al. 2007). Reanalyses are routinely used to study and monitor variables that may be heavily influenced by the amount of snow on the ground (e.g., near-surface air temperature).

One of the challenges of snow evaluation in large-scale reanalyses and GLDAS products is determining the quantity of snow that is actually on the ground. Scale differences make it difficult to use in situ point data to determine how much snow there is over larger areas, as points may not be representative of larger grid boxes (Brasnett 1999; Brown et al. 2003). Point data can be upscaled to a gridded surface using a variety of methods (e.g., Molotch et al. 2005; Erxleben et al. 2002; López-Moreno and Nogués-Bravo 2006; Dixon et al. 2014; Dawson et al. 2016). Often, this involves a fusion of modeling with in situ observations (Carroll et al. 2001; Brown et al. 2003; Drusch et al. 2004) or satellite observations (Takala et al. 2011). Even though such combinations generate snow fields that are consistent with the available observations, there is still uncertainty about how well these products can be treated as "ground truth."

An even more critical factor that complicates the evaluation of snow in large-scale products is the difficulty to disentangle various factors that might lead to simulation errors. For one thing, the forcing data can be uncertain, causing ambiguity about whether simulation errors are the result of forcing data errors, land model deficiencies, or both. There are specific instances where simulation errors can clearly be alleviated by improving either the land models themselves or the input data to those models. For example, Pan et al. (2003) found that differences between observed and model-generated SWE in a North American Land Data Assimilation System (NLDAS) model could be largely alleviated by adjusting precipitation data to match observations. Numerous other studies have also shown that there are differences of snow simulation with different models with the same forcing (e.g., Schlosser et al. 2000; Etchevers et al. 2002; Bowling et al. 2003; Essery et al. 2009; Rutter et al. 2009; Essery et al. 2013; Chen et al. 2014; Magnusson et al. 2015; Mudryk et al. 2015) and that new model formulations (e.g., for snow albedo) can lead to greater agreement between observed and modeled SWE (e.g., Barlage et al. 2010; Livneh et al. 2010; Wang et al. 2010). It is uncertain, however, whether deficiencies in snow simulation in reanalyses and GLDAS products are more commonly related to the land model components of these systems, to data assimilation, or to the input forcing data.

In this study, we seek to disentangle these various factors by evaluating the climatic dependence of snow in modeling systems compared to that from observations.

Section 2 discusses the reanalysis and GLDAS products to be evaluated and the two high-resolution SWE validation datasets over the conterminous United States (CONUS) that are used for the validation. Section 3 quantifies the SWE deficiencies of the reanalysis and GLDAS products and identifies the primary reason for these deficiencies. Section 4 discusses the results presented in section 3.

2. Reanalysis and GLDAS products and validation data

a. Reanalysis and GLDAS products

The reanalysis and GLDAS data are from four GLDAS, version 1 (GLDAS1; Rodell et al. 2004), land models (VIC, Noah, Mosaic, and CLM); one GLDAS, version 2 (GLDAS2), land model (Noah); and three recent reanalyses: the Modern-Era Retrospective Analysis for Research and Applications (MERRA; Rienecker et al. 2011) and its second version (MERRA2), the Climate Forecast System Reanalysis (CFSR; Saha et al. 2010), and the European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim, hereafter ERA-I; Dee et al. 2011). ERA-I and MERRA also include their offline versions, ERA-I/Land (Balsamo et al. 2015) and MERRA-Land (Reichle et al. 2011), respectively. In this paper, "GLDAS products" refers to the four GLDAS1 models and GLDAS2-Noah, and "reanalysis products" refers to CFSR, ERA-I, ERA-I/Land, MERRA, MERRA-Land, and MERRA2. The temporal and spatial resolutions vary among these products, as summarized in Table 1. Data for SWE, snowfall, total precipitation, and 2-m air temperature are obtained for each product.

The forcing data and snow models employed by these products vary considerably (also summarized in Table 1). These products (except for GLDAS1-Mosaic) employ mass and energy balance snow models that have prognostic formulations for albedo and snow density, though the complexity of these models differs among products. CFSR, GLDAS1-Noah, and GLDAS2-Noah use the Noah land surface model (LSM; Chen and Dudhia 2001), which has a single-layer snow scheme. CFSR uses the Global Data Assimilation System (GDAS) forcing with precipitation adjustment from the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP; for low latitudes) and the CPC unified global daily gauge analysis (CPCU; for midlatitudes). GLDAS1-Noah is forced with GDAS forcing data with precipitation adjustment from CMAP, and GLDAS2-Noah is forced using the Princeton meteorological forcing dataset (Sheffield et al. 2006).

TABLE 1. High-resolution data products and the reanalysis and GLDAS products that are used in this study.

Product	Spatial resolution	Temporal resolution	Land model	Snow cover mask	External SWE/snow depth	Single/multilayer
SNODAS	30 s	Daily		Yes	Yes	Multiple
PRISM	2 min	Daily				
UA	2 min	Daily		No	Yes	Single
CFSR	0.25°	Hourly	Noah	Yes	Yes	Single
ERA-I	0.5°	6 hourly	ERA-I	Yes	Yes	Single
ERA-I/Land	0.5°	6 hourly	HTESSEL	Yes	No	Single
MERRA	½° × ⅔°	Hourly	Catchment	No	No	Multiple
MERRA-Land	½° × ⅔°	Hourly	Catchment	No	No	Multiple
MERRA2	0.5° × 0.625°	Hourly	Catchment	No	No	Multiple
GLDAS1-VIC	1°	Hourly	VIC	Yes	No	Multiple
GLDAS1-Mosaic	1°	Hourly	Mosaic	Yes	No	Single
GLDAS1-Noah	1°	Hourly	Noah, version 2.7	Yes	No	Single
GLDAS1-CLM	1°	Hourly	CLM	Yes	No	Multiple
GLDAS2-Noah	0.25°	Hourly	Noah, version 3.3	Yes	No	Single

ERA-I uses the relatively simple single-layer snow scheme of Douville et al. (1995), and in ERA-I/Land, this scheme was revised to that in the Hydrology Tiled ECMWF Scheme of Surface Exchanges over Land (HTESSEL; Balsamo et al. 2009) model to include treatment for liquid water storage in the snowpack and revisions to snow density, albedo, and snow cover fraction. ECMWF forcing drives the land components in both cases, though for ERA-I/Land, precipitation data are corrected using the Global Precipitation Climatology Project (GPCP) dataset. MERRA, MERRA-Land, and MERRA2 use the Catchment LSM (Koster et al. 2000), which has a more complex three-layer snow scheme (Stieglitz et al. 2001). The Catchment model is unique in that it uses a model grid that is based on hydrologic boundaries. In addition to MERRA forcing, precipitation for MERRA-Land and the land model in MERRA2 is corrected with CPCU.

GLDAS1 also has three additional LSMs besides Noah: VIC, CLM, and Mosaic. VIC (Liang et al. 1994) and CLM (Zeng et al. 2002) represent the snowpack using multilayer snow models: VIC has a thin surface layer and a thick deeper layer (Andreadis et al. 2009), and CLM has up to five snow layers. To account for subgrid variability, VIC has subgrid tiles based on elevation bands and land-cover type, whereas CLM just has subgrid tiles based on land-cover type. Mosaic (Koster and Suarez 1992) has very simple treatment for snow, but does include subgrid tiles based on land-cover type. These models share the same forcing information as GLDAS1-Noah, but each GLDAS1 model determines the partitioning between rainfall and snowfall individually (so in this study, snowfall is obtained as model output for each GLDAS1 model).

The reanalysis and GLDAS products also differ in terms of the assimilation of snow data. CFSR uses snow thickness data from the U.S. Air Force Weather Agency (Saha et al. 2010), which combines point measurements with satellite microwave estimates of snow depth. ERA-I uses Cressman interpolation to incorporate synoptic (SYNOP) reports of in situ snow depth into their analyses (Drusch et al. 2004; De Rosnay et al. 2014). MERRA, MERRA-Land, and ERA-I/Land do not assimilate any observational snow information (Rienecker et al. 2011; Reichle et al. 2011; Balsamo et al. 2015). The GLDAS models only assimilate information about snow cover from the Moderate Resolution Imaging Spectroradiometer (MODIS; Rodell and Houser 2004; Zaitchik and Rodell 2009).

b. Validation data

The above reanalysis and GLDAS products are evaluated with high-quality observationally constrained gridded precipitation and temperature data for the CONUS from Parameter-Elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 1994), SWE data from the Snow Data Assimilation System (SNODAS; Carroll et al. 2001; Barrett 2003), and SWE and snowfall data that are based on our recent study (Broxton et al. 2016), which are referred to here as the UA data. The PRISM data include daily estimates of air temperature and precipitation on a 2.5-arc-min grid. The PRISM dataset incorporates data from 13 000 stations for precipitation and nearly 10 000 stations for maximum and minimum temperature across the CONUS. The basis for PRISM is a local climate–elevation regression that varies spatially, but it also accounts for a variety of other local factors affecting climate, such as rain shadow affects, temperature inversions, and the effects of nearby water bodies on temperature and precipitation. PRISM is especially well suited for interpolating climate data in physiographically complex landscapes (Daly et al. 1994, 2008).
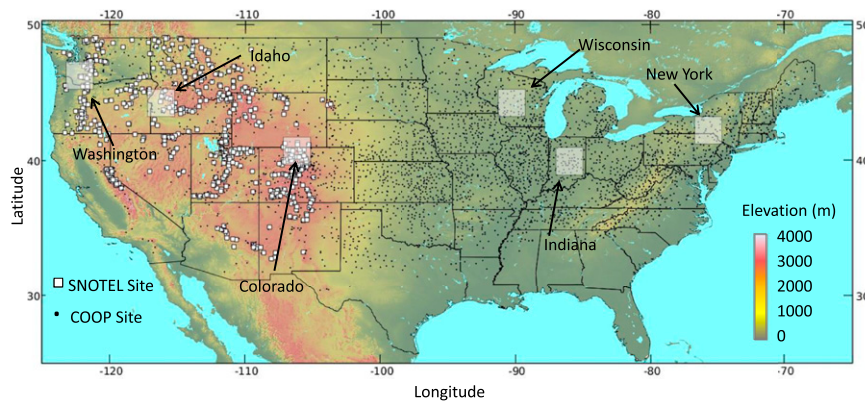
FIG. 1. Site map showing the locations of COOP and SNOTEL stations in the CONUS that are used to generate UA SWE and snowfall data. Also shown are the six 2° × 2° grid boxes referred to in Fig. 5.

Our new SWE and snowfall dataset is generated by combining station data of SWE, snow depth, and precipitation with the gridded PRISM precipitation and temperature data. In this study, we choose to use validation products based on ground data (for precipitation, temperature, SWE, and snow depth) as opposed to satellite data because of the uncertainty of using satellite remote sensing to measure SWE in areas of complex terrain (such as the western United States, where the deepest snowpacks in the CONUS occur). In addition, across the United States, there is a relatively high density of stations that allows for relatively high-quality precipitation and temperature data to constrain our SWE estimates. Note that a recent study (Mudryk et al. 2015) also evaluated many of the products that we are testing, but compared them with a product derived from satellite and in situ snow data. In this study, we feel that constraining our SWE estimates using precipitation and temperature data gives a better estimate for the CONUS, where the station density is relatively high and where there are substantial areas of complex terrain.

Here, in situ data come from the National Resources Conservation Service (NRCS) Snowpack Telemetry (SNOTEL; Serreze et al. 1999) network and the National Weather Service's (NWS) Cooperative Observer Program (COOP) network in the United States. The COOP data used in this study are from the National Climatic Data Center (NCDC) COOP Summary of the Day DSI-3200 dataset, which includes observations from ~8000 active stations, including NWS "first order" or principal climatological stations (operated by trained observers) and "second order" (operated by volunteers) stations (NCDC 2006). Figure 1 shows the locations of the SNOTEL and COOP stations used in this study.

The first step in the creation of our dataset is to determine the precipitation phase at the in situ stations, as

this is the basis of whether to consider PRISM precipitation as rainfall or snowfall. Following Knowles et al. (2006) and Huntington et al. (2004), who perform snowfall/precipitation studies in the western and eastern United States, respectively, precipitation is considered to be snow at COOP stations on days for which newly fallen snow was recorded. Similarly, it is considered to be snow at SNOTEL stations if there is a positive change in snow depth. As pointed out by Huntington et al. (2004), this methodology may sometimes result in positive biases in snowfall, because on days when precipitation falls as both rain and snow, it is recorded entirely as snow. In general, this method has been shown by Knowles et al. (2006) and Huntington et al. (2004) to partition between rainfall and snowfall reasonably well in both the western and eastern CONUS.

Next, a temperature threshold is used to define the phase (rainfall vs snowfall) of the PRISM precipitation data. This threshold is determined individually for each station that records precipitation. For stations where precipitation occurs as rainfall (based on the above methodology) on days when the daily average air temperature is below freezing, the temperature threshold is set as the station temperature minus 1°C. For stations where precipitation occurs as snowfall (based on the above methodology) on days where the daily average air temperature is above freezing, the temperature threshold is set as the station temperature plus 1°C. Otherwise, for stations that record snowfall at subfreezing temperatures and rainfall at temperatures warmer than 0°C (which occur a majority of the time), the threshold is set to 0°C. These threshold temperatures are then interpolated from each station location to the PRISM 2.5-arc-min grid using ordinary kriging. PRISM precipitation that occurs when the air temperature below this interpolated threshold is considered snow; otherwise, it is considered rain.
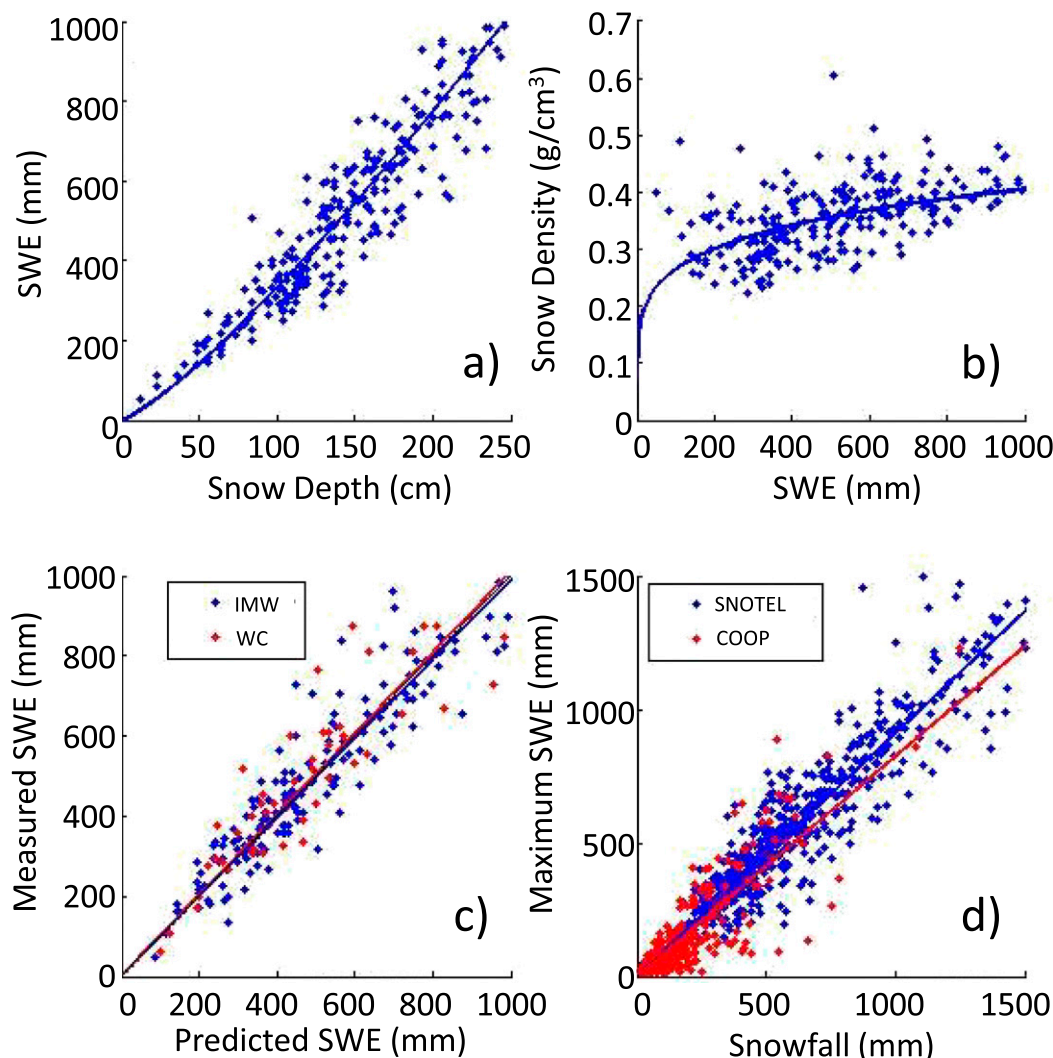
FIG. 2. (a),(b) Regression used to convert snow depth to SWE for COOP data. In (a) each blue dot represents a SNOTEL observation at SWE_{max} and the solid line represents the best-fit regression. In (b) the same is shown except plotted as bulk snow density vs SWE. (c) Observed vs predicted SWE_{max} recorded for the half of the SNOTEL stations that are not used in the calibration. The red dots correspond to West Coast (WC) sites (west of 120°W) and the blue dots correspond to Intermountain West (IMW) sites (east of 120°W). (d) Cool season accumulated snowfall vs SWE_{max} at SNOTEL (blue dots) and COOP (red dots) stations and their respective regressions (solid lines). In (d), COOP SWE is obtained from the observed snow depth and the regression line in (a).

The generation of SWE estimates follows the methodology of Broxton et al. (2016). Namely, SWE at each station is divided by the difference between accumulated snowfall (from the time that no snow was on the ground to the time of the SWE estimate) measured at the station and an estimate of accumulated ablation (for the same period), which is given by a simple temperature index snow model. We then interpolate these ratios to the PRISM 2.5-arc-min grid and multiply them by the difference between accumulated snowfall within each PRISM pixel and our estimate of accumulated ablation (for each pixel). For this dataset, we are essentially using forward modeling to predict accumulated snowfall and accumulated ablation. This approach (but with a different model) has been shown to produce reliable SWE estimates in areas with dense precipitation networks (Livneh et al. 2014); however, in places with fewer gauges, it has also been shown that incorporation of satellite data (e.g., using inverse modeling to work backward from snow-free conditions; Molotch 2009) can generate comparable measurements, or even improve upon SWE measurements generated from forward modeling (Raleigh and Lundquist 2012; Livneh et al. 2014).
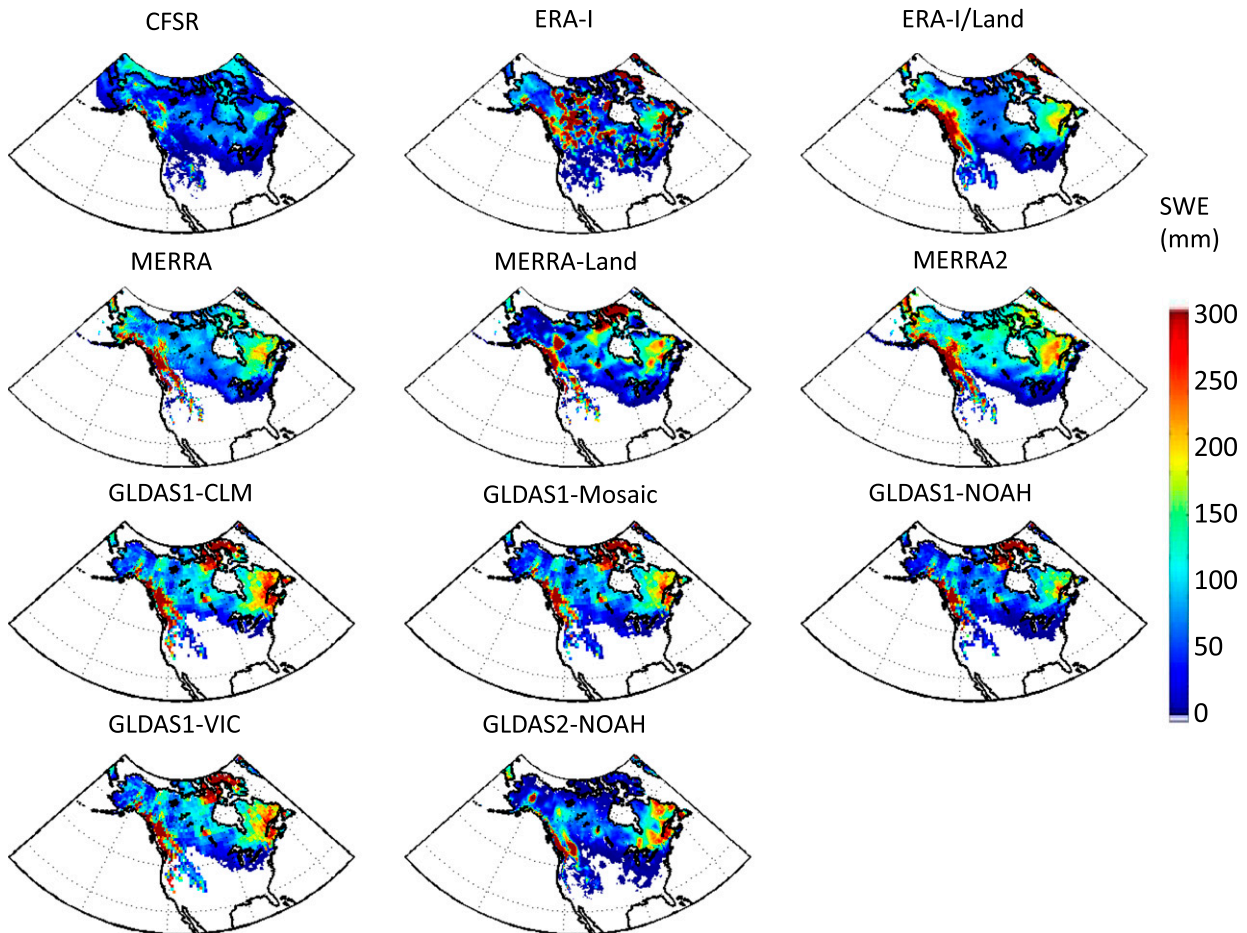
FIG. 3. Spatial patterns of SWE on 1 Mar 2008. Areas with values <1 mm appear as white. Note that CFSR shows snow over frozen water bodies.

While the SNOTEL stations measure SWE directly, COOP stations do not. Here the COOP snow depths are converted to SWE using a relationship between the SNOTEL SWE and snow depth data at the time of maximum snow accumulation (Fig. 2a). In general, snow density increases over time because of settling and metamorphosis of grains, compaction due to the weight of overlying snow, and refreezing of meltwater (Essery et al. 2013), which occurs at different rates in different environments. Figure 2b shows that, commonly, when there is more snow on the ground, the bulk snow density is higher. Furthermore, this relationship generally holds even when half of the data are withheld during calibration and used as evaluation data instead. This also holds even when the data are stratified so that the calibration data are from a different environment than the validation data. For example, Fig. 2c shows that this relationship holds for sites that are in the Cascades and Sierra Nevada (which generally have a higher snow density through much of the winter) versus those in the

Intermountain West (which generally have a lower snow density through much of the winter). This transformation, along with the snowfall calculations below, results in a nonlinear regression slope between computed maximum SWE and snowfall at COOP stations that is quite similar to the SNOTEL record, albeit with more scatter (the square of the correlation coefficient R^2 in Fig. 2d is 0.74 for the COOP data vs 0.92 for the SNOTEL data).

Besides the UA dataset, we also use the SNODAS SWE data. SNODAS provides data for SWE on a 30-arc-s grid. It is the National Operational Hydrologic Remote Sensing Center's (NOHRSC) operational snow monitoring system that provides the SWE and snow depth estimates in the CONUS (Barrett 2003). The assimilation system uses downscaled forcing data from a regional weather model to drive a multilayered, uncoupled energy and mass balance model. SNODAS can then be nudged using a large amount of snow measurements from ground, airborne, and satellite sources if
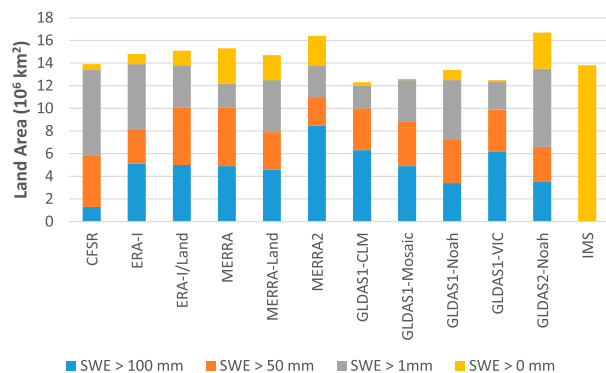
FIG. 4. Snow-covered land area ($\times 10^6$ km^2) on 1 Mar 2008 based on SWE > 0, 1, 50, and 100 mm for the reanalysis and GLDAS products, as well as SCA based on the IMS product. Note that IMS data only distinguish between snow and no snow for land areas.

it is determined by an analyst that such nudging is necessary (Barrett 2003).

SNODAS assimilates much more snow and near-surface atmospheric forcing data than our method, but its method (Carroll et al. 2001; Barrett 2003) was not published in peer-reviewed journals and the subjective adjustment by analysts during the data development is not reproducible. The SNODAS data are available beginning 30 September 2003. In contrast, our method is based directly on daily PRISM precipitation and temperature data (and hence relies on more in situ precipitation and temperature data) and relies on station reports of SWE, snow depth, and precipitation that go back further than 30 years, enabling the creation of a dataset spanning 30+ years. Broxton et al. (2016) found that the two SWE datasets are comparable, and their differences can be interpreted as a rough estimate of our SWE data uncertainty.

In this study, we use data from water year (WY) 2004 to WY 2009 (meaning that the data go from 1 October 2003 to 30 September 2009), as this covers most of the overlap period for our products. SNODAS begins in WY 2004 and CFSR is only available until March 2011 (after which it was superseded by the CFSv2 operational analysis).

3. Evaluation of SWE in the reanalysis and GLDAS products

In this section, we first present a comparison between SWE values from each product over all of North America to give a sense of which products predict the most snow and which products predict the least (section 3a). Because this analysis partly examines snow cover extent, we also compare the overall snow-covered extent to that indicated by the 24-km Ice Mapping System

(IMS) product (Ramsay 1998; National Ice Center 2008) over North America. Next, we evaluate reanalysis and GLDAS SWE with the SNODAS and UA SWE data (section 3b). We evaluate both the temporal evolution of the reanalysis and GLDAS SWE for six 2° × 2° grid boxes, as well as the spatial distribution of maximum SWE (SWE$_{max}$) across the entire CONUS. Next, to understand how deficiencies in the reanalysis and GLDAS SWE relate to deficiencies of the forcing data, we evaluate the reanalysis and GLDAS products in terms of how well they represent "snow season" precipitation (both the total amount and the fraction that falls as snow) and snow season temperatures (section 3c). The snow season is defined, for grid cells of each product separately, as the period between the first day of snow on the ground and the last. Reanalysis and GLDAS precipitation amounts are also compared to PRISM precipitation data, and their snowfall fractions are compared to the UA snowfall data and PRISM precipitation data.

Last, to understand how deficiencies in reanalysis and GLDAS SWE might be related to the amount of snow ablation in each product, we evaluate the SWE differences between the reanalysis and GLDAS products versus the SNODAS and UA data in terms of the rate of snow ablation predicted in each product (section 3d). For consistency among all products, we treat the reduction of SWE on precipitation-free days as representing the amount of daily snow ablation due to evaporation/sublimation losses and snowmelt. The results in sections 3a–d focus on a single water year for easy interpretation of the results, and these results are extended to six water years (from WY 2004 to WY 2009) in section 3e.

a. Spatial patterns of GLDAS and reanalysis SWE over North America

Generally, the spatial patterns of SWE are relatively consistent among the reanalysis and GLDAS products that we tested over North America. This can be seen in Fig. 3, which shows a SWE "snapshot" for each product on 1 March 2008, which is near the time of maximum snow cover extent. The two products that look the most dissimilar from the rest are CFSR and ERA-I. For CFSR, the total snow-covered area is relatively consistent with the other products, though there is a much smaller area with higher amounts of SWE. For example, it only has 1.3×10^6 km^2 of land area with more than 100 mm of SWE, compared to areas from 3.4×10^6 to 8.5×10^6 km^2 for the other GLDAS or reanalysis products (Fig. 4). For ERA-I, there are many areas with large amounts of SWE that are not reflected in the other reanalysis or GLDAS products. For example, ERA-I is
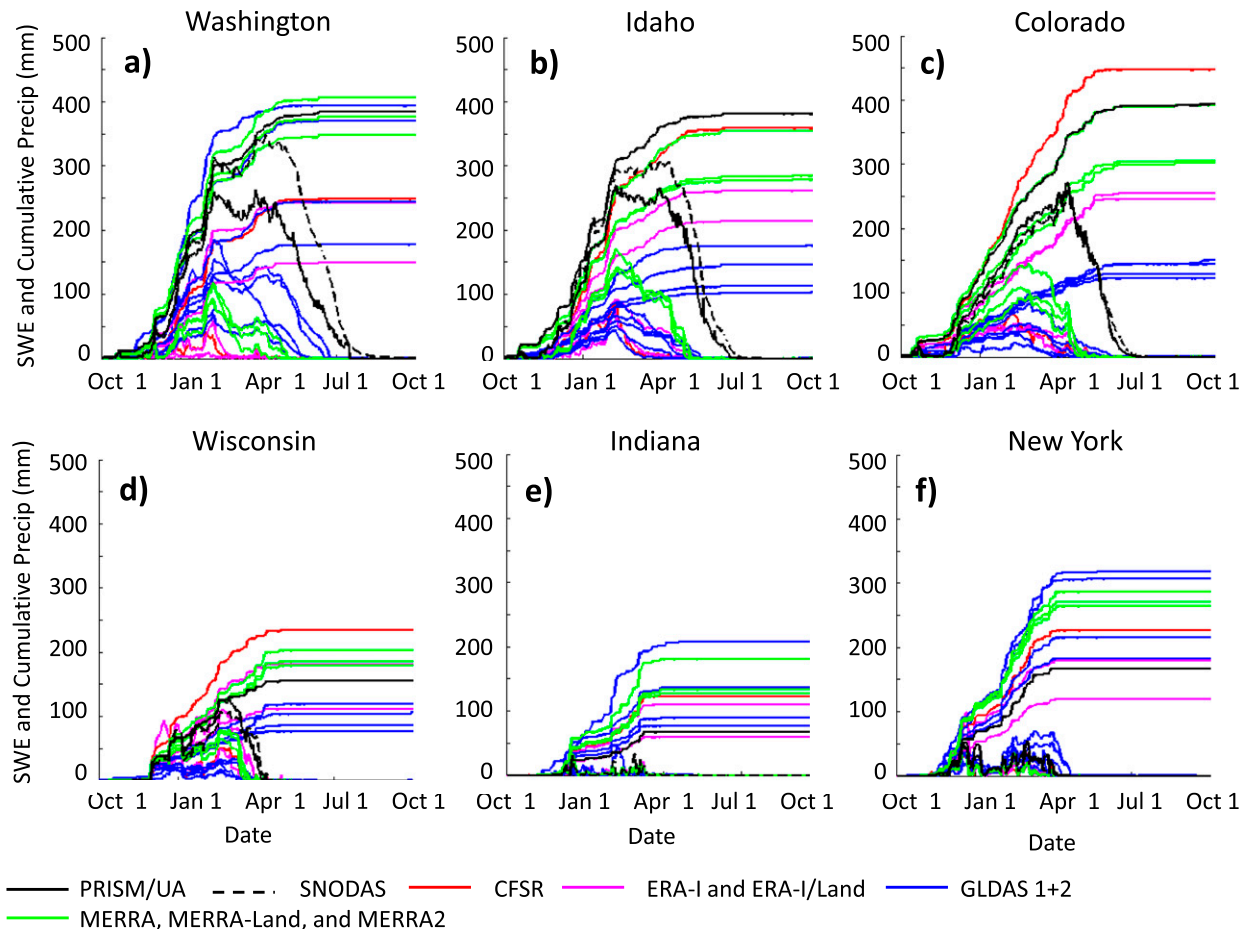
FIG. 5. Time series showing the area averages of SWE and cumulative snowfall for the UA/PRISM datasets (black lines); CFSR (red line); ERA-I and ERA-I/Land (magenta lines); MERRA, MERRA-Land, and MERRA2 (green lines); GLDAS1 and GLDAS2 products (blue lines); and SNODAS SWE (black dashed lines) for the (a) Washington, (b) Idaho, (c) Colorado, (d) Wisconsin, (e) Indiana, and (f) New York 2° grid boxes (shown in Fig. 1) during WY 2008 (from 1 Oct 2007 to 30 Sep 2008). The monotonically increasing lines are gridbox averages of cumulative snowfall, and the lines that rise and fall are gridbox averages of SWE.

the only product depicting large amounts of SWE in pockets through central Canada east of the Rocky Mountains (Fig. 3). These patches are reminiscent of artifacts in the ERA-I's snow analysis that have been identified in the past. ERA-I/Land looks more consistent with the other products than ERA-I (Fig. 3). GLDAS1-CLM and GLDAS1-VIC are the snowiest GLDAS models, while Noah (both GLDAS1 and GLDAS2) and GLDAS1-Mosaic have fewer areas of deeper snow (Fig. 4). MERRA2 also has considerably more area with >100 mm of SWE than the other products (8.5×10^6 km^2). This is much more than MERRA, which has 4.6×10^6 km^2, and MERRA-Land, which has 4.9×10^6 km^2 with SWE > 100 mm. Note that products that have more SWE as of 1 March also have more SWE through the entire course of the winter (not shown).

Figure 4 also shows the sensitivity of the estimated snow-covered area (SCA) to the threshold SWE value

used. Some products have substantial areas with <1 mm of SWE, giving unrealistically high estimates of total SCA when the threshold is very low. For example, using a threshold of 0 mm gives total SCA that is from 0.1×10^6 to 3.2×10^6 km^2 greater than using a threshold of 1 mm. GLDAS2-Noah, MERRA, MERRA-Land, and MERRA2 have particularly large differences when these two thresholds are used ($>1.5 \times 10^6$ km^2), while ERA-I, ERA-I/Land, GLDAS1-Noah, and CFSR have moderately large differences (from 0.5×10^6 to 1.5×10^6 km^2). The other GLDAS1 models have smaller differences. Compared with the IMS data, using a threshold of 1 mm seems to give a better estimate of SCA than 0 mm for the products with the largest differences between SCA computed with the 0- and 1-mm thresholds (e.g., GLDAS2-Noah, ERA-I, ERA-I/ Land, MERRA, and MERRA-Land), as the 0-mm threshold results in SCA that is significantly higher
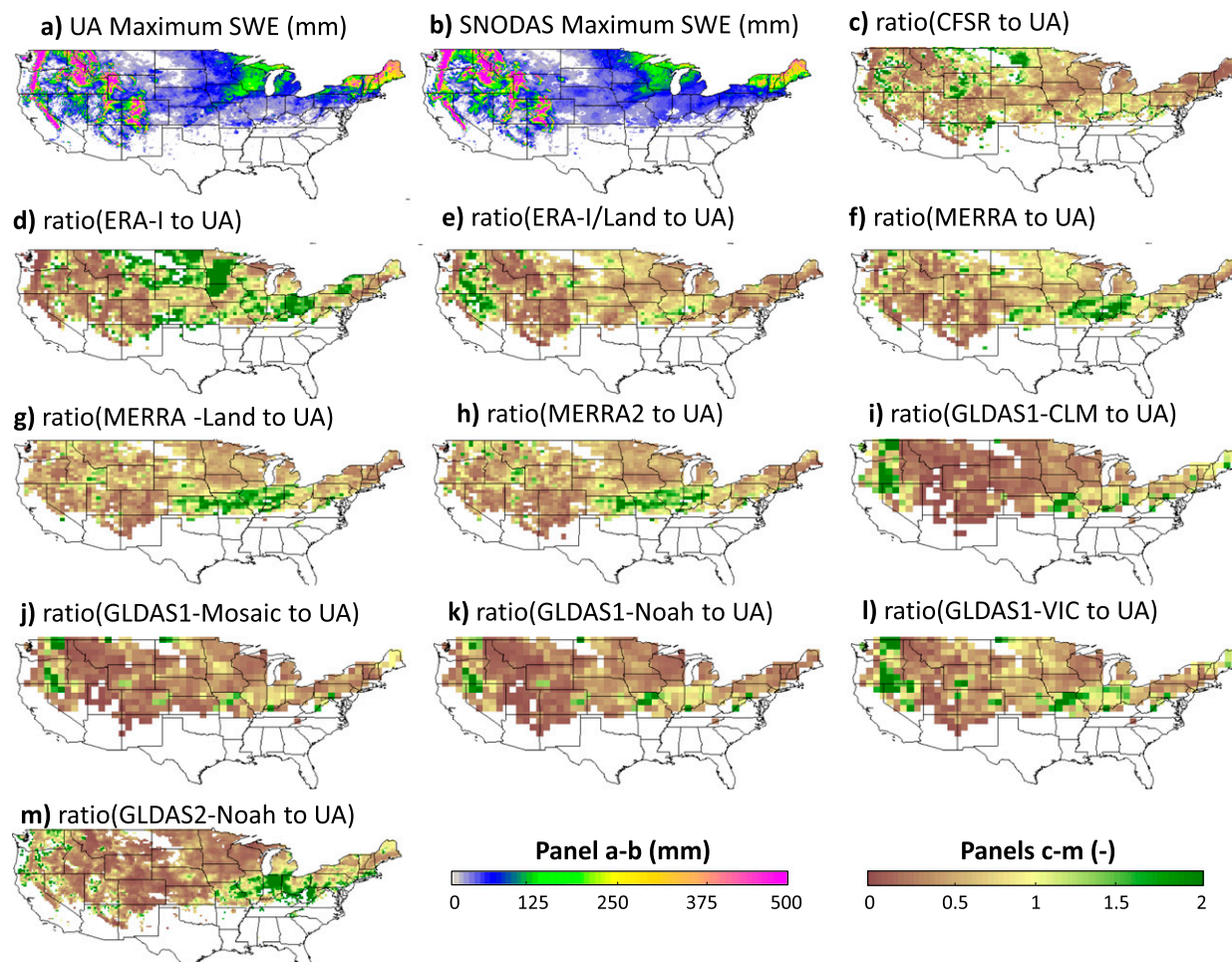
FIG. 6. (a),(b) Values of SWE$_{max}$ for WY 2008 from the UA and SNODAS datasets (with a grid size of 2.5 arc min), respectively. (c)–(m) The ratio between SWE$_{max}$ for each reanalysis and GLDAS product and that in (a) after first averaging to each product's grid cell (see Table 1). Grid cells with less than 10 days of continuous snow cover appear as white.

than the IMS data in these products (Fig. 4). Using a threshold of 1 mm slightly underpredicts SCA compared to the IMS data across most products.

b. Comparison with high-resolution SWE data over the CONUS

In comparison with the validation data (SNODAS SWE and our UA SWE dataset), SWE is too low in the reanalysis and GLDAS products in the CONUS. For example, Fig. 5 shows that through the winter, area-averaged SWE is commonly higher in the high-resolution data than the GLDAS and reanalysis datasets in the six 2° × 2° boxes shown in Fig. 1 (centered on northern Colorado, southern Idaho, western Washington, central Wisconsin, central Indiana, and central New York). This is particularly true of the Colorado and Idaho boxes. Furthermore, this underestimation of SWE persists for much of the winter, and particularly in the

Colorado, Idaho, and Washington grid boxes, snow persists much longer in the gridded observational data than in the reanalysis and GLDAS products.

This general underestimation of SWE by the reanalysis and GLDAS products compared to the higher-resolution SWE data can also be seen across the CONUS at maximum snow accumulation (Fig. 6). In many areas where there are deeper snowpacks, the reanalysis and GLDAS products predict less snow than the UA (and SNODAS) data. Many of the products (including SNODAS), however, predict more snow than the UA data along the southern margin of the extent of seasonal snow, as well as in some portions of the western United States and northern Great Plains, primarily in areas where the snow is shallow (Fig. 6). However, as can be seen in Fig. 6a, the magnitudes of SWE$_{max}$ in these regions are often quite low, so even though there are substantial relative differences in these areas (shown

TABLE 2. Values of SWE_{max}, cumulative precipitation (i.e., P), cumulative snowfall (i.e., S) divided by P, average temperature (i.e., T_{Air}), SWE_{max} divided by S, and median daily snow ablation for air temperatures between $-1°$ and $+1°C$ (i.e., ΔSWE at $0°C$) averaged across the CONUS for the snow season in WY 2008 (from 1 Oct 2007 to 30 Sep 2008).

Product	SWE_{max} (mm)	P (mm)	S/P	T_{Air} (°C)	SWE_{max}/S	ΔSWE at 0°C (mm)
UA/PRISM	113.1	292.3	0.52	−2.49	0.65	0.0
SNODAS	113.2	—	—	—	—	0.0
CFSR	43.8	451.9	0.47	−1.37	0.23	−6.5
ERA-I	53.6	289.8	0.40	−1.11	0.49	−1.3
ERA-I/Land	41.3	381.5	0.33	−1.26	0.38	−1.6
MERRA	55.0	303.0	0.62	−1.70	0.25	−10.6
MERRA-Land	55.3	283.4	0.59	−1.72	0.30	−9.5
MERRA2	52.7	292.7	0.59	−2.06	0.27	−6.0
GLDAS1-CLM	50.5	135.3	0.82	−4.41	0.46	−1.9
GLDAS1-Mosaic	38.2	127.3	0.64	−4.26	0.48	−1.6
GLDAS1-Noah	31.5	157.6	0.58	−3.45	0.40	−3.3
GLDAS1-Vic	57.6	152.5	0.64	−3.36	0.57	−1.6
GLDAS2-Noah	45.6	235.1	0.57	−1.93	0.31	−4.6

in Figs. 6b–m), the absolute differences are quite small. Overall, averaged across the entire CONUS, both the UA and SNODAS SWE data have considerably higher SWE than the reanalysis and GLDAS products.

This conclusion holds for all of the reanalysis and GLDAS products, regardless of their resolution. For example, GLDAS2-Noah has a spatial resolution of $0.25° \times 0.25°$, but it underestimates SWE_{max} more than many GLDAS1 models, which have a spatial resolution of $1° \times 1°$. Furthermore, there is more spread in the CONUS average SWE_{max}, predicted by the GLDAS1 models (31.5–57.6 mm, or 28%–51% of UA SWE_{max}), which all share the same spatial resolution of $1° \times 1°$, than that predicted by the other products (41.3–55.3 mm, or 37%–49% of UA SWE_{max}), which have spatial resolutions ranging from $0.25° \times 0.25°$ to $0.50° \times 0.67°$ (column 1 of Table 2).

The same conclusions can also be drawn if the reanalysis and GLDAS SWE_{max} are compared to SNODAS (rather than UA) SWE_{max}. Not only is the seasonal evolution of SNODAS SWE similar to that in the UA dataset for the $2° \times 2°$ boxes (Fig. 5), but averaged across the entire country, maximum SWE predicted by the UA dataset is nearly the same as that predicted by SNODAS (column 1 of Table 2). This gives us confidence that the underestimation of SWE by the reanalysis and GLDAS products are real and substantial.

c. Evaluating the role of atmospheric forcing

One of the first things that comes to mind when trying to explain this widespread underestimation of SWE are deficiencies in the model precipitation data, as the amount of SWE is particularly sensitive to the amount of snow that has fallen (Broxton et al. 2016). However, this widespread underestimation of SWE (and SWE_{max}) occurs in the reanalysis and GLDAS products despite the fact that some products have more precipitation than is suggested by the PRISM data, and some have less. CFSR generally has more precipitation than the PRISM data do while the GLDAS models have less. ERA-I, ERA-I/Land, MERRA, MERRA2, and MERRA-Land have smaller precipitation biases compared to PRISM. This is generally true for the $2° \times 2°$ boxes (Fig. 5) as well as across the entire CONUS (Fig. 7 and column 2 of Table 2). For example, for WY 2008, the reanalysis products have 97%–155% of the PRISM snow season precipitation, averaged across the entire CONUS (column 2 of Table 2), while the GLDAS products have 44%–80% of the PRISM snow season precipitation (column 2 of Table 2).

Furthermore, this widespread underestimation of SWE probably cannot be explained by incorrect partitioning of precipitation between rainfall and snowfall. Snow season snowfall S divided by snow season precipitation P is, in general, lower in CFSR, ERA-I, and ERA-I/Land than is suggested by the high-resolution data. At the same time, this ratio is higher in MERRA, MERRA-Land, MERRA2, and GLDAS products than is suggested by the high-resolution data (column 3 of Table 2; Fig. 8). For example, averaged across the CONUS during WY 2008, for CFSR, ERA-I, and ERA-I/Land, this ratio ranges from 63% to 90% of S/P determined from the PRISM precipitation and UA snowfall datasets. For the remainder of the reanalysis and GLDAS products, this ratio ranges from 109% to 158% of S/P determined from the PRISM precipitation and UA snowfall datasets (column 3 of Table 2). This means that for these products, snowfall makes up a greater proportion of total precipitation when snow is on
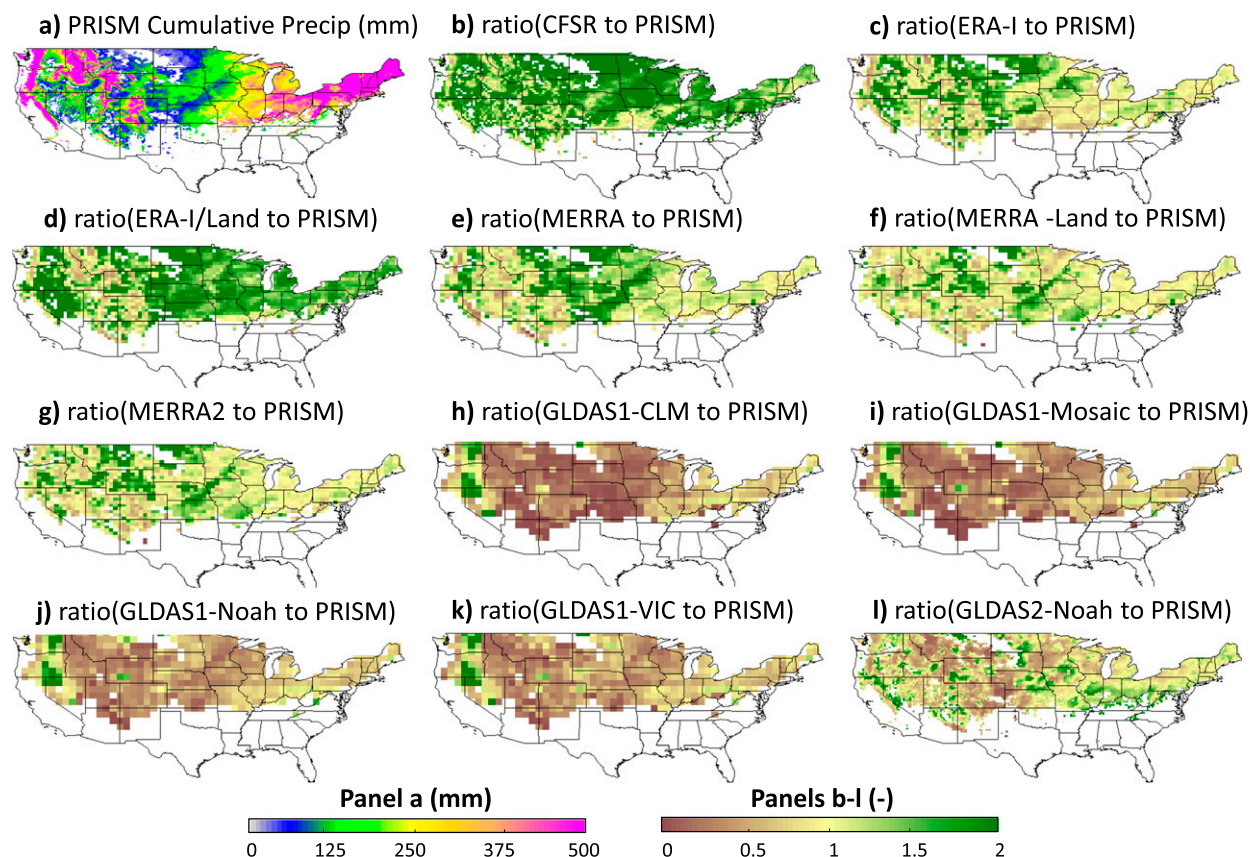
FIG. 7. (a) Total precipitation that falls during the snow season in WY 2008 based on PRISM data (with a grid size of 2.5 arc min). (b)–(l) The ratio between the total snow season precipitation for each reanalysis and GLDAS product and that in (a) after first averaging to each product's grid cell (see Table 1). As mentioned in section 3, the snow season is defined independently for each grid box of each product as the period between the first and last day of nonzero SWE. Grid cells with less than 10 days of continuous snow cover appear as white.

the ground than is suggested by the high-resolution datasets.

It is also unlikely that temperature T_{Air} biases can account for the widespread underestimation of SWE. For example, some products have higher temperatures than are suggested by the PRISM dataset, and some have lower temperatures. In general, the reanalysis products (especially CFSR, ERA-I, and ERA-I/Land) and GLDAS2-Noah have higher temperatures than are suggested by the PRISM data, and the GLDAS1 products have lower temperatures (column 4 of Table 2; Fig. 9). These results are complimentary with the above results because, in general, the same products that have lower S/P than is suggested by the high-resolution data have higher temperatures, and vice versa (cf. columns 3 and 4 of Table 2).

d. Evaluating the role of snow ablation

Section 3c shows that the significant underestimation of reanalysis and GLDAS SWE occurs irrespective of

biases in atmospheric forcing. This points to a deficiency in the land components of these reanalyses and GLDAS products. In particular, SWE_{max} is too low in the reanalysis and GLDAS products for the amount of snow that falls. This can be seen by comparing the ratio of SWE_{max} to accumulated snowfall (SWE_{max}/S) for the reanalysis and GLDAS products with that for the UA data. Figure 10 clearly shows that this ratio is much lower when computed for the reanalysis and GLDAS products than for the high-resolution data. Averaged across the CONUS, SWE_{max}/S for the reanalysis products ranges from 36% to 76% of that computed using the UA dataset, and SWE_{max}/S for the GLDAS products ranges from 47% to 88% of that computed using the UA dataset (column 5 of Table 2). Higher-resolution products are not necessarily better at representing this ratio, as the two highest-resolution products (CFSR and GLDAS2-Noah) perform the worst and GLDAS1-VIC represents this ratio better than the other reanalysis and GLDAS products (column 5 of Table 2).
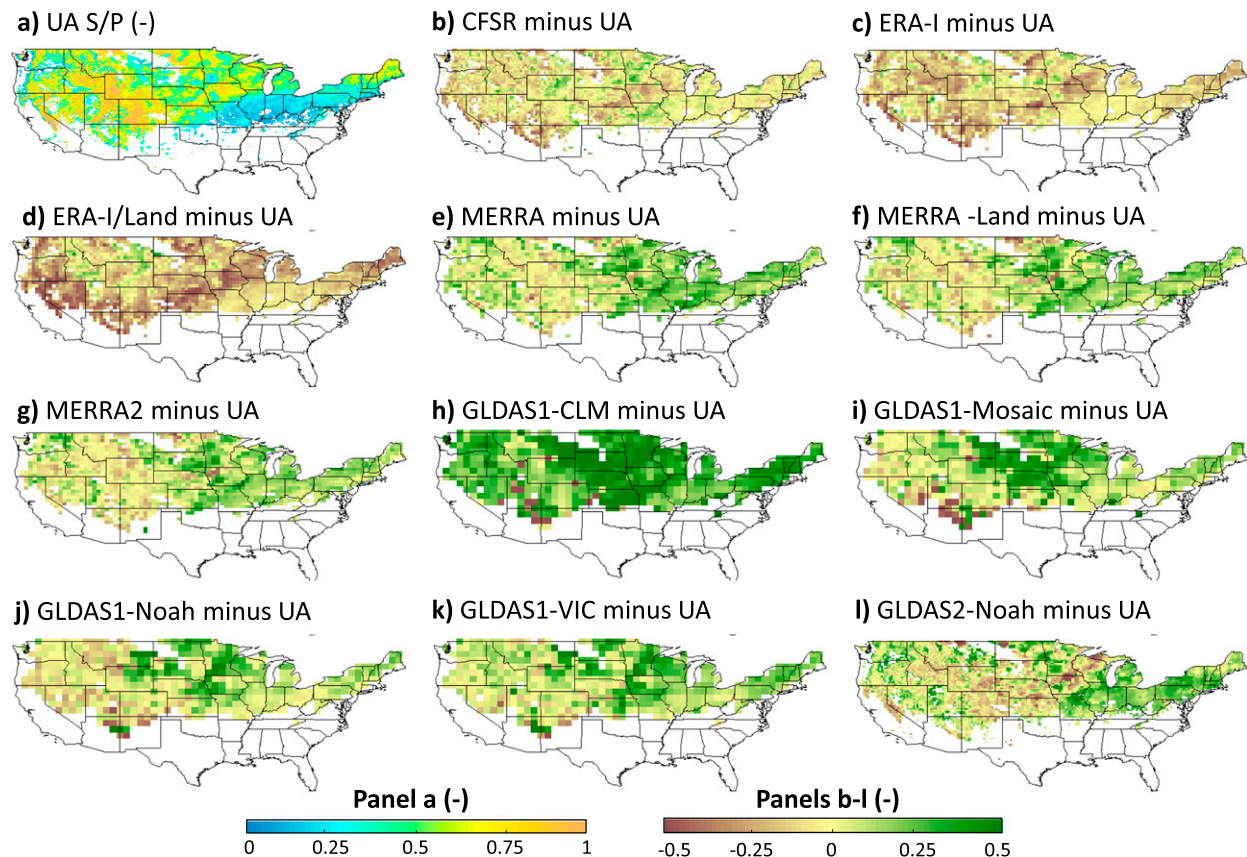
FIG. 8. (a) Values of S/P (with a 2.5-arc-min grid size) for the snow season during WY 2008 based on the UA snowfall and PRISM precipitation data. (b)–(l) The difference between S/P for each reanalysis and GLDAS product and that in (a) after first averaging to each product's grid cell (see Table 1).

At the same time, many of the reanalysis and GLDAS products have more ablation than the UA or SNOTEL SWE datasets. Figure 11a plots, in 2° daily temperature bins, the SWE change (ΔSWE) on precipitation-free days for all products. This quantity can be thought of as approximately equivalent to daily snow ablation. The reanalysis and GLDAS products have different ΔSWEs at particular temperatures. MERRA, MERRA-Land, MERRA2 (green lines in Fig. 11a), and CFSR (red line in Fig. 11a) have the most ΔSWE at subfreezing temperatures. Under these conditions, the UA and SNODAS SWE data show very little ΔSWE, as even the 25th percentile of ΔSWE is close to zero. Even at temperatures above 0°C, most of the reanalysis and GLDAS products have ΔSWEs that are larger in magnitude (more negative) than the UA SWE and SNODAS SWE data. Only ERA-I and ERA-I/Land have median values of ΔSWE that are consistently within the 25th–75th percentile range of ΔSWE for the UA SWE and SNODAS SWE data. At temperatures between 0° and +5°C, GLDAS1-VIC also has less ΔSWE than the remainder

of the reanalysis and GLDAS products. Many of the differences between the GLDAS and reanalysis products can be captured by comparing the median ΔSWE for the temperature bin from −1° to +1°C for different products (Table 2, column 6).

There is a good relationship between products that have too much ablation at near-freezing temperatures and those that have ratios of SWE_{max}/S that are too low (cf. columns 5 and 6 of Table 2). Figure 11b shows this graphically by plotting the differences between SWE_{max}/S for each reanalysis and GLDAS product and that computed from the UA SWE and snowfall data (these values are determined by subtracting rows 3–13 from row 1 in column 5 of Table 2) versus the median ΔSWE for the temperature bin from −1° to +1°C for each product (column 6 of Table 2). Products that have more ablation at near-freezing temperatures clearly have smaller values of SWE_{max}/S.

e. Consistency of results for additional water years

For the remainder of the water years in the study period (from WY 2004 to WY 2009), the results were
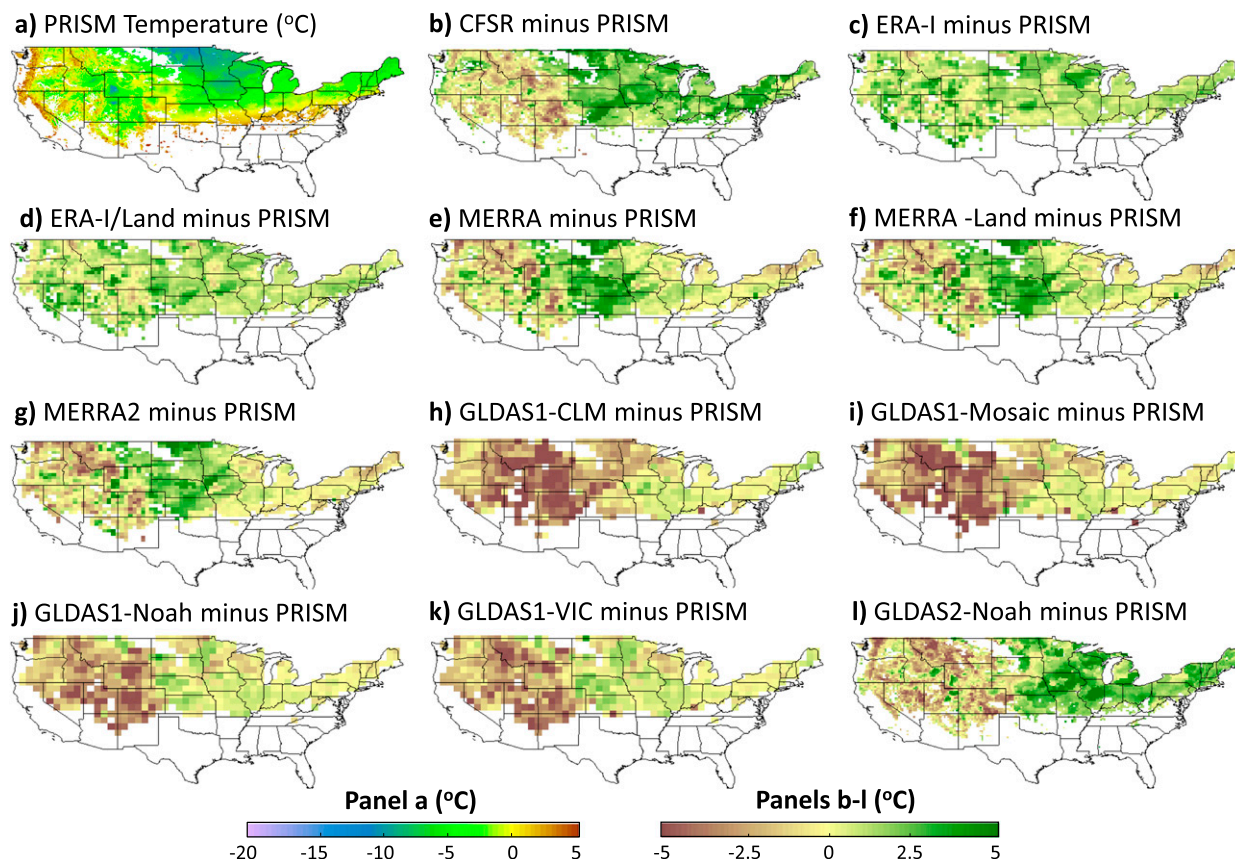
FIG. 9. (a) The average temperature for the snow season during WY 2008 based on the PRISM data (with a 2.5-arc-min grid size). (b)–(l) The difference between the snow season T_{Air} for each reanalysis and GLDAS product and that in (a) after first averaging to each product's grid cell (see Table 1).

similar, that is, SWE$_{max}$ was underpredicted by all of the reanalysis and GLDAS products (Fig. 12a), and the underestimation was a little bit worse in comparison to SNODAS because SNODAS predicted slightly more SWE than the UA dataset for many of the years. Despite the interannual climatic variability, CFSR consistently overestimated precipitation, and GLDAS1 products consistently underestimated precipitation (Fig. 12b). Often, the same products that predicted higher or lower snow season snowfall fraction (S/P) and average snow season temperature than the PRISM and UA data were consistent in doing so (Figs. 12c,d). Furthermore, during the entire period, the only metric that was consistently lower than the UA data was SWE$_{max}$/S (Fig. 12e). This interannual consistency indicates that the results in sections 3a–d are robust.

4. Discussion and conclusions

Similar to another recent study (Mudryk et al. 2015), we find that there are considerable differences of the climatological values of SWE among the GLDAS and reanalysis products that we tested, even when the forcing for those models is consistent (e.g., for GLDAS1). However, we find that all of the reanalysis and GLDAS products that we tested generally underpredict SWE in the CONUS. They have considerably less snow than higher-resolution, observationally based gridded SWE data, especially in the western CONUS. This is true when these products are compared to either the UA or the SNODAS data, indicating that these results are probably fairly robust. Even though there is a wide spread in the model precipitation, with GLDAS1 products (which all share the same precipitation amount) having the least amount of precipitation (well below what is suggested by PRISM) and CFSR having the most (more than what is suggested by PRISM), SWE$_{max}$, normalized by precipitation, is significantly lower in all of the reanalysis and GLDAS products than is suggested by the high-resolution data.

The primary reason for this underestimation is that most reanalysis and GLDAS products have too much snow ablation at near-freezing temperatures. In fact, the UA and SNODAS data suggest that, on average,
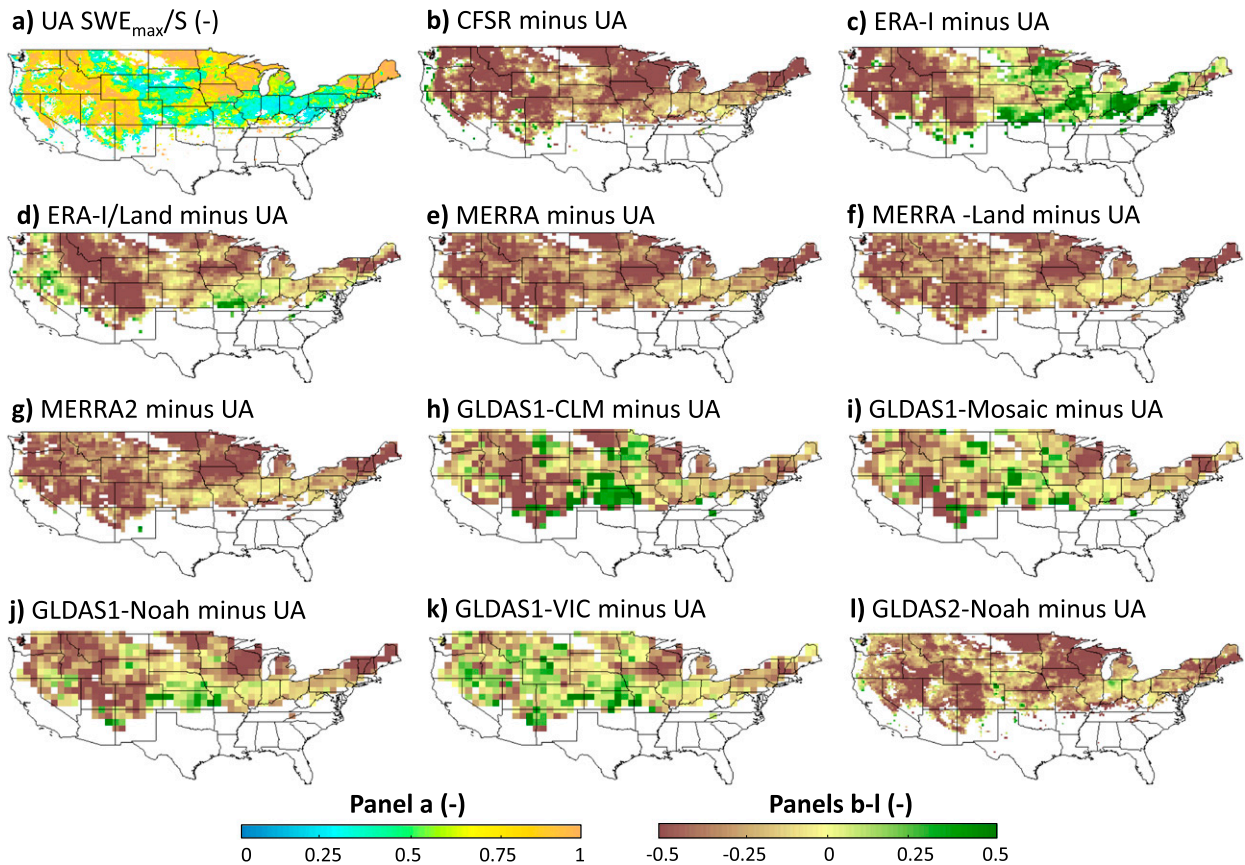
FIG. 10. (a) Values of SWE_{max}/S for the snow season during WY 2008 based on the UA SWE and snowfall data (with a 2.5-arc-min grid size). (b)–(l) The difference between SWE_{max}/S for each reanalysis and GLDAS product and that in (a) after first averaging to each product's grid cell (see Table 1).

significant snow ablation does not occur until the daily temperature is above freezing, though some of the GLDAS and reanalysis products that we tested have significant snow ablation when the daily average temperatures are well below freezing. In particular, CFSR, MERRA, MERRA-Land, and MERRA2 have the most snow ablation at subfreezing temperatures. In general, the more ablation that the reanalysis or GLDAS products have at near-freezing temperatures, the larger the difference between SWE_{max}/S for the UA dataset versus that computed from reanalysis or GLDAS data (Fig. 11b).

This does not mean that other factors do not contribute to the widespread underestimation of SWE, especially on a case-by-case basis. For example, as mentioned in section 2, CFSR and ERA-I include initialization that affects snow thickness. This certainly affects SWE for these products. CFSR uses a similar initialization procedure as the operational CFS, which was shown in Dawson et al. (2016) to lead to significant underestimations of SWE. The MERRA and GLDAS products do not initialize snow thickness, so it is not

expected that snow initialization could dramatically alter the amount of SWE in these products. In addition, forcing biases (e.g., in precipitation and temperature) certainly contribute to either worsening or alleviating the underestimation of SWE in the reanalysis and GLDAS products. It is well known that forcing biases can lead to large uncertainties in SWE simulation, especially if those biases are large (Raleigh et al. 2015). For example, Brun et al. 2013 find that the Crocus snow model yields much better snow simulations with relatively small biases over Eurasia if the model is forced with ECMWF forcing versus the Princeton meteorological forcing dataset. In this study, we find that GLDAS products, especially GLDAS1, underestimate precipitation, which undeniably contributes to these products having low SWE. However, at the same time, some of the GLDAS1 products, particularly GLDAS1-CLM, have higher snowfall ratios (S/P) than is suggested by the UA data, which would likely increase SWE. Note that even though GLDAS1 models share nearly identical precipitation forcing, each model decides how to partition between rainfall and snowfall individually.
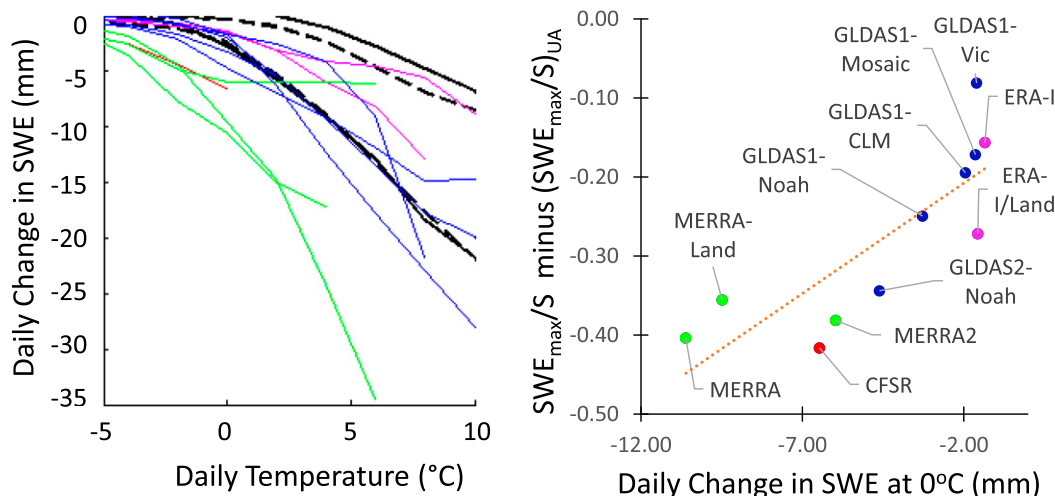
FIG. 11. (a) Plot of the median daily snow ablation computed as the difference between SWE on consecutive precipitation-free days vs daily temperature for each product. The black solid lines show the 25th and 75th percentiles of ablation computed for the UA SWE and PRISM temperature data within 2° temperature bins (e.g., from −3° to −1°C, −1° to +1°C, etc.) and the black dashed lines show the 25th and 75th percentiles of ablation computed for the SNODAS SWE and PRISM temperature data. The colored solid lines show the median ablation for each GLDAS or reanalysis product within each temperature bin [the colors scheme is the same in (b)]. (b) Plot of the CONUS average value of SWE_{max} divided by total snowfall (SWE_{max}/S) for each reanalysis and GLDAS product minus that based on the UA SWE and snowfall data (from column 5 of Table 2) as a function of the median amount of snow ablation for the temperature bin in (a) that goes from −1° to +1°C (column 6 of Table 2).

Furthermore, the GLDAS1 products tend to underestimate snow season temperature, especially in the western CONUS (Fig. 9, Table 2).

In addition, the above conclusions are not much affected by the different spatial resolutions of these products. For example, GLDAS1-Noah (which has a grid size of 1°) predicts SWE_{max}/S that is closer to the UA data than GLDAS2-Noah (which has a grid size of 0.25°). In addition, even though GLDAS1-VIC has a grid size of 1° (among the coarsest resolution), it was most similar to the UA data in terms of SWE_{max}/S and to the UA and SNODAS data in terms SWE_{max}. This does not mean that resolution is not an issue because all products that we evaluated have resolutions that are too coarse to resolve complex topography (mountains) in some regions. However, what we have shown is that even after removing this effect (i.e., evaluating how each product predicts SWE given a set of climatological conditions within each product), the GLDAS and reanalysis products still underestimate SWE.

We have shown that the degree to which each product underestimates SWE is related to the propensity of each product to predict more or less ablation at near-freezing temperatures. Although full model sensitivity studies are required to fully understand, on a case by case basis, why this may be so (a task that is beyond the scope of the current manuscript), it is reasonable to assume that it can potentially be understood in terms of the differences in model physics or the way in which the models are implemented. Consistent with previous studies (e.g., Essery et al. 2013), more complex models do not necessarily predict SWE better than simpler models. For example, MERRA, which employs a multilayer snow model, predicts SWE_{max}/S that is lower than GLDAS1-Mosaic, which employs a single-layer snow model (and with much simpler snow physics). However, deficiencies in the representation of certain aspects of the model physics (e.g., for albedo) can have a big influence on snow simulation (Wang et al. 2010; Feng et al. 2008). Furthermore, other factors, such as how individual models are influenced by forest cover or whether they include subgrid tiling of topography and vegetation, may also have a large influence on the snow simulation (Chen et al. 2014). It is striking that GLDAS1-VIC is the best at representing SWE_{max}/S and yet is the only model to include subgrid elevation bands. Representing this sort of subgrid variability may be particularly appropriate for snow simulation because of SWE's strong dependence on elevation. It is possible that better representation of subgrid topography (e.g., elevation and aspect) may lead to more realistic snow simulation in large-scale reanalysis and GLDAS.

While the UA dataset is likely much more accurate at predicting SWE than the GLDAS and reanalysis products, it still contains a number of uncertainties. One potential uncertainty is the statistical model to
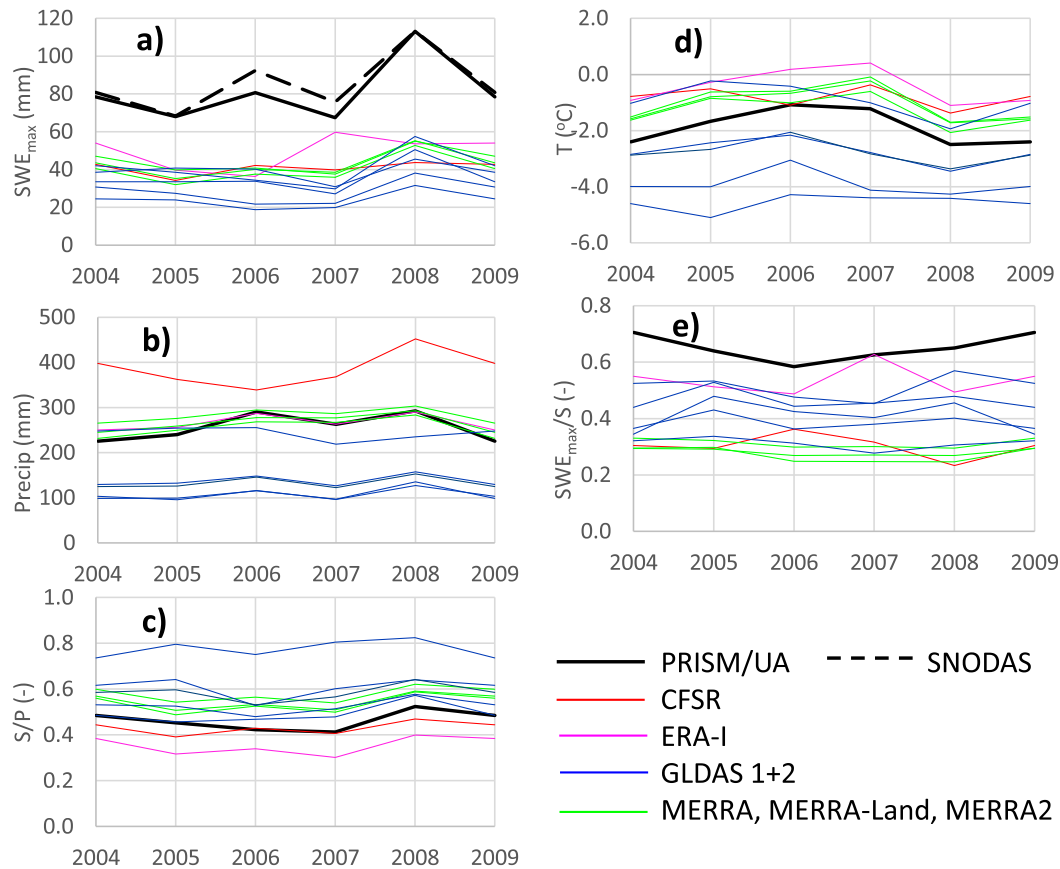
FIG. 12. Plots showing, for each water year (from WY 2004 to WY 2009), (a) SWE$_{max}$, (b) P, (c) S/P, (d) average temperature, and (e) SWE$_{max}/S$ for the snow season averaged across the CONUS. The values of each line for WY 2008 correspond with those given in columns 1–5 of Table 2. A line for ERA-I/Land is not shown because we were unable to obtain data for ERA-I/Land for all variables for the whole period.

convert COOP snow depths into SWE (Fig. 2a) prior to their inclusion into the dataset. In general, this relationship potentially underestimates SWE at COOP stations because it is derived from SNOTEL stations, which are often located in cold environments with continuous snowpacks, and applied to COOP stations, which are located in warmer environments with ephemeral snowpacks. Ephemeral snowpacks often have a higher density, at a given SWE, than cold snowpacks (Brown et al. 2003). Furthermore, there is a substantial difference between the scale of in situ measurements and the grid scale of the UA data (4 km). However, the effect of this difference on the SWE assimilation is mediated somewhat in the UA data by the fact that a normalized quantity (SWE/S), instead of SWE itself, is interpolated between the point observations. The normalized quantity was shown in Broxton et al. (2016) to be spatially more consistent than the unnormalized quantity. Furthermore, the dataset is strongly constrained by the gridded temperature and precipitation data, which makes it relatively consistent, regardless of what SWE data are included (and their uncertainty). For example, Figure S1 in the supplemental material shows the area-averaged SWE through time for WY 2008 for the Colorado and Wisconsin grid boxes where 0%, 50%, 75%, and 90% of the SNOTEL and COOP SWE and snow depth data have been withheld (i.e., are not used in the interpolation). These area-averaged quantities rarely differ by more than 5%. It is clear that withholding large amounts of SWE and snow depth data have relatively little impact on the overall averaged SWE for the grid boxes, demonstrating the robustness of the UA product.

Since the UA SWE dataset is based on daily PRISM precipitation and temperature data and station SWE and snow depth data, it covers a period that goes back much further than the SNODAS data. Because of the potential use of such a dataset, we are in the process of documenting and releasing the entire dataset (from WY 1981 to WY 2010) in a separate study. Furthermore, recognizing the uncertainty in the simple relationship between SWE and snow depth in Fig. 2, we are in the

process of developing a new observational data-based snow density model in a separate study.

REFERENCES

Andreadis, K. M., P. Storck, and D. P. Lettenmaier, 2009: Modeling snow accumulation and ablation processes in forested environments. *Water Resour. Res.*, **45**, W05429, doi:10.1029/2008WR007042.

Bales, R. C., N. P. Molotch, T. H. Painter, M. D. Dettinger, R. Rice, and J. Dozier, 2006: Mountain hydrology of the western United States. *Water Resour. Res.*, **42**, W08432, doi:10.1029/2005WR004387.

Balsamo, G., A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, and A. K. Betts, 2009: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System. *J. Hydrometeor.*, **10**, 623–643, doi:10.1175/2008JHM1068.1.

——, and Coauthors, 2015: ERA-Interim/Land: A global land surface reanalysis data set. *Hydrol. Earth Syst. Sci.*, **19**, 389–407, doi:10.5194/hess-19-389-2015.

Barlage, M., and Coauthors, 2010: Noah land surface model modifications to improve snowpack prediction in the Colorado Rocky Mountains. *J. Geophys. Res.*, **115**, D22101, doi:10.1029/2009JD013470.

Barrett, A. P., 2003: National Operational Hydrologic Remote Sensing Center Snow Data Assimilation System (SNODAS) products at NSIDC. NSIDC Special Rep. 11, 19 pp. [Available online at http://nsidc.org/pubs/documents/special/nsidc_special_report_11.pdf.]

Bowling, L. C., and Coauthors, 2003: Simulation of high-latitude hydrological processes in the Torne–Kalix basin: PILPS Phase 2(e): 1: Experiment description and summary intercomparisons. *Global Planet. Change*, **38**, 1–30, doi:10.1016/S0921-8181(03)00003-1.

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

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

Broxton, P. D., N. Dawson, and X. Zeng, 2016: Linking snowfall and snow accumulation to generate spatial maps of SWE and snow depth. *Earth Space Sci.*, **3**, 246–256, doi:10.1002/2016EA000174.

Brun, E., V. Vionnet, A. Boone, B. Decharme, Y. Peings, and R. Valette, 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.

Carroll, T., D. Cline, G. Fall, A. Nilsson, L. Li, and A. Rost, 2001: NOHRSC operations and the simulation of snow cover properties for the coterminous U.S. *Proc. 69th Annual Western Snow Conf.*, Sun Valley, ID, Western Snow Conference, 14 pp. [Available online at http://www.westernsnowconference.org/node/185.]

Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.*, **129**, 569–585, doi:10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

——, and Coauthors, 2014: Modeling seasonal snowpack evolution in the complex terrain and forested Colorado Headwaters region: A model intercomparison study. *J. Geophys. Res. Atmos.*, **119**, 13 795–13 819, doi:10.1002/2014JD022167.

Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.*, **33**, 140–158, doi:10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2.

——, M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.*, **28**, 2031–2064, doi:10.1002/joc.1688.

Dawson, N., P. D. Broxton, X. Zeng, M. Leuthold, and P. Holbrook, 2016: An evaluation of snow initializations for NCEP global and regional forecasting models. *J. Hydrometeor.*, **17**, 1885–1901, doi:10.1175/JHM-D-15-0227.1.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.

De Goncalves, L. G. G., W. J. Shuttleworth, S. C. Chou, Y. Xue, P. R. Houser, D. L. Toll, J. Marengo, and M. Rodell, 2006: Impact of different initial soil moisture fields on Eta model weather forecasts for South America. *J. Geophys. Res.*, **111**, D17102, doi:10.1029/2005JD006309.

De Rosnay, P., G. Balsamo, C. Albergel, J. Muñoz-Sabater, and L. Isaksen, 2014: Initialisation of land surface variables for numerical weather prediction. *Surv. Geophys.*, **35**, 607–621, doi:10.1007/s10712-012-9207-x.

Dixon, D., S. Boon, and U. Silins, 2014: Watershed-scale controls on snow accumulation in a small montane watershed, southwestern Alberta, Canada. *Hydrol. Processes*, **28**, 1294–1306, doi:10.1002/hyp.9667.

Douville, H., J.-F. Royer, and J.-F. Mahfouf, 1995: A new snow parameterization for the Météo-France climate model. Part I: Validation in stand-alone experiments. *Climate Dyn.*, **12**, 21–35, doi:10.1007/BF00208760.

Drusch, M., D. Vasiljevic, and P. Viterbo, 2004: ECMWF's global snow analysis: Assessment and revision based on satellite observations. *J. Appl. Meteor. Climatol.*, **43**, 1282–1294, doi:10.1175/1520-0450(2004)043<1282:EGSAAA>2.0.CO;2.

Erxleben, J., K. Elder, and R. Davis, 2002: Comparison of spatial interpolation methods for estimating snow distribution in the Colorado Rocky Mountains. *Hydrol. Processes*, **16**, 3627–3649, doi:10.1002/hyp.1239.

Essery, R., and Coauthors, 2009: SNOWMIP2: An evaluation of forest snow process simulations. *Bull. Amer. Meteor. Soc.*, **90**, 1120–1135, doi:10.1175/2009BAMS2629.1.

——, S. Morin, Y. Lejeune, and C. B. Ménard, 2013: A comparison of 1701 snow models using observations from an alpine site. *Adv. Water Resour.*, **55**, 131–148, doi:10.1016/j.advwatres.2012.07.013.

Etchevers, P., and Coauthors, 2002: SnowMIP, an intercomparison of snow models: First results. *Proc. of the Int. Snow Science Workshop*, Penticton, BC, Canada, International Snow Science Workshop, 353–360. [Available online at http://arc.lib.montana.edu/snow-science/objects/issw-2002-353-360.pdf.]

Feng, S., A. Sahoo, K. Arsenault, P. Houser, Y. Luo, and T. Troy, 2008: The impact of snow model complexity at three CLPX sites. *J. Hydrometeor.*, **9**, 1464–1481, doi:10.1175/2008JHM860.1.

Gutzler, D. S., and J. W. Preston, 1997: Evidence for a relationship between spring snow cover in North America and summer rainfall in New Mexico. *Geophys. Res. Lett.*, **24**, 2207–2210, doi:10.1029/97GL02099.

Hahn, D. G., and J. Shukla, 1976: An apparent relationship between Eurasian snow cover and Indian monsoon rainfall. *J. Atmos. Sci.*, **33**, 2461–2462, doi:10.1175/1520-0469(1976)033<2461:AARBES>2.0.CO;2.

Huntington, T. G., G. A. Hodgkins, B. D. Keim, and R. W. Dudley, 2004: Changes in the proportion of precipitation occurring as snow in New England (1949–2000). *J. Climate*, **17**, 2626–2636, doi:10.1175/1520-0442(2004)017<2626:CITPOP>2.0.CO;2.

Knowles, N., M. D. Dettinger, and D. R. Cayan, 2006: Trends in snowfall versus rainfall in the western United States. *J. Climate*, **19**, 4545–4559, doi:10.1175/JCLI3850.1.

Koster, R. D., and M. J. Suarez, 1992: Modeling the land surface boundary in climate models as a composite of independent vegetation stands. *J. Geophys. Res.*, **97**, 2697–2715, doi:10.1029/91JD01696.

——, ——, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *J. Geophys. Res.*, **105**, 24 809–24 822, doi:10.1029/2000JD900327.

——, and Coauthors, 2004: Regions of strong coupling between soil moisture and precipitation. *Science*, **305**, 1138–1140, doi:10.1126/science.1100217.

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

Livneh, B., Y. Xia, K. E. Mitchell, M. B. Ek, and D. P. Lettenmaier, 2010: Noah LSM snow model diagnostics and enhancements. *J. Hydrometeor.*, **11**, 721–738, doi:10.1175/2009JHM1174.1.

——, J. S. Deems, D. Schneider, J. J. Barsugli, and N. P. Molotch, 2014: Filling in the gaps: Inferring spatially distributed precipitation from gauge observations over complex terrain. *Water Resour. Res.*, **50**, 8589–8610, doi:10.1002/2014WR015442.

López-Moreno, J. I., and D. Nogués-Bravo, 2006: Interpolating local snow depth data: An evaluation of methods. *Hydrol. Processes*, **20**, 2217–2232, doi:10.1002/hyp.6199.

Magnusson, J., N. Wever, R. Essery, N. Helbig, A. Winstral, and T. Jonas, 2015: Evaluating snow models with varying process representations for hydrological applications. *Water Resour. Res.*, **51**, 2707–2723, doi:10.1002/2014WR016498.

Molotch, N. P., 2009: Reconstructing snow water equivalent in the Rio Grande headwaters using remotely sensed snow cover data and a spatially distributed snowmelt model. *Hydrol. Processes*, **23**, 1076–1089, doi:10.1002/hyp.7206.

——, M. T. Colee, R. C. Bales, and J. Dozier, 2005: Estimating the spatial distribution of snow water equivalent in an alpine basin using binary regression tree models: The impact of digital elevation data and independent variable selection. *Hydrol. Processes*, **19**, 1459–1479, doi:10.1002/hyp.5586.

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

National Ice Center, 2008: IMS daily Northern Hemisphere snow and ice analysis at 1 km, 4 km, and 24 km resolutions, version 1. Subset used: 24 km data, National Snow and Ice Data Center, accessed 12 December 2014, doi:10.7265/N52R3PMC.

NCDC, 2006: Data Documentation for Data Set 3200 (DSI-3200). National Climatic Data Center, 19 pp. [Available online at http://www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/td3200.pdf.]

Pan, N., and Coauthors, 2003: Snow process modeling in the North American Land Data Assimilation System (NLDAS): 2. Evaluation of model simulated snow water equivalent. *J. Geophys. Res.*, **108**, 8850, doi:10.1029/2003JD003994.

Raleigh, M. S., and J. D. Lundquist, 2012: Comparing and combining SWE estimates from the SNOW-17 model using PRISM and SWE reconstruction. *Water Resour. Res.*, **48**, W01506, doi:10.1029/2011WR010542.

——, ——, and M. P. Clark, 2015: Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. *Hydrol. Earth Syst. Sci.*, **19**, 3153–3179, doi:10.5194/hess-19-3153-2015.

Ramsay, B. H., 1998: The interactive multisensor snow and ice mapping system. *Hydrol. Processes*, **12**, 1537–1546, doi:10.1002/(SICI)1099-1085(199808/09)12:10/11<1537::AID-HYP679>3.0.CO;2-A.

Reichle, R. H., R. D. Koster, G. J. de Lannoy, B. A. Forman, Q. Liu, S. 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.

Robinson, D. A., and G. Kukla, 1985: Maximum surface albedo of seasonally snow-covered lands in the Northern Hemisphere. *J. Climate Appl. Meteor.*, **24**, 402–411, doi:10.1175/1520-0450(1985)024<0402:MSAOSS>2.0.CO;2.

Rodell, M., and P. R. Houser, 2004: Updating a land surface model with MODIS-derived snow cover. *J. Hydrometeor.*, **5**, 1064–1075, doi:10.1175/JHM-395.1.

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

——, J. Chen, H. Kato, J. S. Famiglietti, J. Nigro, and C. R. Wilson, 2007: Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE. *Hydrogeol. J.*, **15**, 159–166, doi:10.1007/s10040-006-0103-7.

Rutter, N., and Coauthors, 2009: Evaluation of forest snow processes models (SnowMIP2). *J. Geophys. Res.*, **114**, D06111, doi:10.1029/2008JD011063.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, doi:10.1175/2010BAMS3001.1.

Schlosser, C. A., and Coauthors, 2000: Simulations of a boreal grassland hydrology at Valdai, Russia: PILPS phase 2(d). *Mon. Wea. Rev.*, **128**, 301–321, doi:10.1175/1520-0493(2000)128<0301:SOABGH>2.0.CO;2.

Serreze, M. C., M. P. Clark, R. L. Armstrong, D. A. McGinnis, and R. S. Pulwarty, 1999: Characteristics of the western United States snowpack from Snowpack Telemetry (SNOTEL) data. *Water Resour. Res.*, **35**, 2145–2160, doi:10.1029/1999WR900090.

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**, 3088–3111, 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, doi:10.1175/1525-7541(2001)002<0228:TIODSP>2.0.CO;2.

Syed, T. H., J. S. Famiglietti, M. Rodell, J. Chen, and C. R. Wilson, 2008: Analysis of terrestrial water storage changes from GRACE and GLDAS. *Water Resour. Res.*, **44**, W02433, doi:10.1029/2006WR005779.

Takala, M., K. Luojus, J. Pulliainen, C. Derksen, J. Lemmetyinen, J. P. Kärnä, J. Koskinen, and B. Bojkov, 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**, 3517–3529, doi:10.1016/j.rse.2011.08.014.

Wang, Z., X. Zeng, and M. Decker, 2010: Improving snow processes in the Noah land model. *J. Geophys. Res.*, **115**, D20108, doi:10.1029/2009JD013761.

Zaitchik, B. F., and M. Rodell, 2009: Forward-looking assimilation of MODIS-derived snow-covered area into a land surface model. *J. Hydrometeor.*, **10**, 130–148, doi:10.1175/2008JHM1042.1.

Zeng, X., M. Shaikh, Y. Dai, R. E. Dickinson, and R. Myneni, 2002: Coupling of the Common Land Model to the NCAR Community Climate Model. *J. Climate*, **15**, 1832–1854, doi:10.1175/1520-0442(2002)015<1832:COTCLM>2.0.CO;2.

Zhang, J., W.-C. Wang, and J. Wei, 2008: Assessing land–atmosphere coupling using soil moisture from the Global Land Data Assimilation System and observational precipitation. *J. Geophys. Res.*, **113**, D17119, doi:10.1029/2008JD009807.