How Does Availability of Meteorological Forcing Data Impact Physically Based Snowpack Simulations?*

MARK S. RALEIGH

National Center for Atmospheric Research,[+] Boulder, Colorado

BEN LIVNEH

Cooperative Institute for Research in Environmental Science, University of Colorado Boulder, Boulder, Colorado

KARL LAPO

Atmospheric Sciences Department, University of Washington, Seattle, Washington

JESSICA D. LUNDQUIST

Civil and Environmental Engineering, University of Washington, Seattle, Washington

ABSTRACT

Physically based models facilitate understanding of seasonal snow processes but require meteorological forcing data beyond air temperature and precipitation (e.g., wind, humidity, shortwave radiation, and longwave radiation) that are typically unavailable at automatic weather stations (AWSs) and instead are often represented with empirical estimates. Research is needed to understand which forcings (after temperature and precipitation) would most benefit snow modeling through expanded observation or improved estimation techniques. Here, the impact of forcing data availability on snow model output is assessed with data-withholding experiments using 3-yr datasets at well-instrumented sites in four climates. The interplay between forcing availability and model complexity is examined among the Utah Energy Balance (UEB), the Distributed Hydrology Soil Vegetation Model (DHSVM) snow submodel, and the snow thermal model (SNTHERM). Sixty-four unique forcing scenarios were evaluated, with different assumptions regarding availability of hourly meteorological observations at each site. Modeled snow water equivalent (SWE) and snow surface temperature T_{surf} diverged most often because of availability of longwave radiation, which is the least frequently measured forcing in cold regions in the western United States. Availability of longwave radiation (i.e., observed vs empirically estimated) caused maximum SWE differences up to 234 mm (57% of peak SWE), mean differences up to 6.2°C in T_{surf}, and up to 32 days difference in snow disappearance timing. From a model data perspective, more common observations of longwave radiation at AWSs could benefit snow model development and applications, but other aspects (e.g., costs, site access, and maintenance) need consideration.

---

1. Introduction

Seasonal snow cover serves as a major water source (Viviroli et al. 2003; Barnett et al. 2005), acts as a natural reservoir for hydropower (Madani and Lund 2010; Winther and Hall 1999), impacts ecological activity (Kudo 1991; Trujillo et al. 2012), and alters weather (Hawkins et al. 2002; Jin and Miller 2007) and climate (Qu and Hall 2006) through land–atmosphere interactions.

Despite the importance of seasonal snow, there remains a dearth of automatic weather stations (AWSs) in cold regions (Lundquist et al. 2003; Viviroli et al. 2011; Hijmans et al. 2005; Horel and Dong 2010). Thus, available AWS observations provide only a glimpse into the complexities of the cryosphere, which typically exhibits high spatial variability in snow and meteorological states (Scipión et al. 2013; Deems et al. 2006; Sturm and Wagner 2010; Blöschl 1999; Bales et al. 2006; Dubayah 1992). Physically based snow models are useful for predicting and understanding snow distributions and processes in sparsely instrumented catchments, yet the extensive requisite meteorological forcings to drive these models are often unavailable.

It is often impossible to evaluate forcing estimation methods and to diagnose model deficiencies because few AWSs measure all forcings (Bales et al. 2006). At a minimum, physically based snow models require six near-surface forcings: air temperature T_{air}, precipitation P, wind speed U, humidity (often as relative humidity RH), incoming shortwave radiation Q_{si}, and incoming longwave radiation Q_{li}. A survey of 1318 AWSs across multiple networks that measure either snow water equivalent (SWE) or snow depth in the western United States shows T_{air} and P are most frequently measured, while Q_{li} is least frequently measured (Fig. 1). Nearly 99% do not measure all six forcings. This survey excludes other fluxes measured only at specialized research sites (i.e., fluxes of sensible heat, latent heat, and upwelling radiation) because these are often treated as response variables instead of forcings. Because snow mass is measured more commonly than wind, humidity, and radiation, it is generally only possible to evaluate modeled SWE (Pan et al. 2003; Chen et al. 2014b; Watson et al. 2006; Rasmussen et al. 2011; Livneh et al. 2010) and not all model inputs. Data scarcity limits model diagnostics and may impair advances in hydrology because compensatory errors in key processes can yield the "right answer for the wrong reasons" (Kirchner 2006).

Given the prevalence of T_{air} and P observations, the motivating question here is this: How would physical model output change if less commonly measured forcings (i.e., U, RH, Q_{si}, and Q_{li}) were observed more frequently? In other words, what is the next "best" sensor to install at an AWS for snow modeling? This goes beyond an investigation of the relative importance of snow model forcings (Zuzel and Cox 1975), as investments in new sensors may be unwarranted if a forcing can be estimated reasonably from common data. Indeed, numerous empirical models estimate forcing data in ungauged areas (see section 4b), and these are regularly used within mountain climate and hydrologic
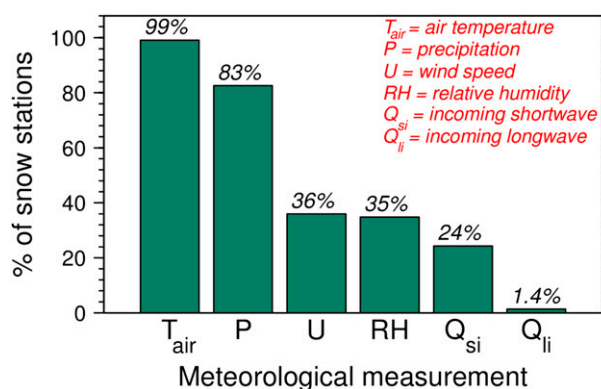

FIG. 1. Meteorological measurements at AWSs ($n = 1318$) in the western United States where SWE or snow depth are also measured. Shown are the percentages of stations measuring each variable. Station operators include (alphabetically): airports (FAA), AmeriFlux/Fluxnet, avalanche centers, Bureau of Reclamation, California Cooperative Snow Surveys, California Department of Water Resources, Desert Research Institute, Long Term Ecological Research sites, NOAA/NWS, NSF Critical Zone observatories, specialized research campaigns (e.g., NASA Cold Land Processes Field Experiment), transportation departments, university research sites, U.S. Army (Corps of Engineers, Cold Regions Research Laboratory), U.S. Department of Agriculture (e.g., Natural Resources Conservation Service SNOTEL, U.S. Forest Service, Agricultural Research Service), and the Western Regional Climate Center. Quality and completeness of the data are not reflected. Survey completed June 2013.

modeling systems [e.g., mountain microclimate simulation model (MTCLIM; Hungerford et al. 1989) and Variable Infiltration Capacity model (Liang et al. 1994)]. It remains unclear how evaluations of forcing data estimation techniques (e.g., Bohn et al. 2013; Flerchinger et al. 2009) relate to specific data availability scenarios and snowpack modeling. To identify future research needs and deficiencies in AWS networks, it is vital to prioritize the relative importance of forcing availability for snow modeling efforts. This is a valuable topic, as adding new sensors to existing AWSs can have lower infrastructure costs than adding a new station.

The purpose of this study is to assess how forcing availability (i.e., observed vs estimated data) controls physically based simulations of snowpack mass and energy states. Seasonal snowpack is modeled with three snow models at well-instrumented AWSs in four climates with 3-yr observed datasets of all six required meteorological forcings. These AWSs permit data-withholding experiments, where hypothetical data availability scenarios are constructed and a single data estimation method is used to replace each withheld forcing. We assume that daily T_{air} (i.e., minimum temperature T_{min} and maximum temperature T_{max}) and daily P are available (at a minimum) at an AWS and

quantify model output divergence as a function of hourly forcing availability.

2. Background and motivation

High-quality hydrometeorological forcing data are important for development, calibration, and simulations of hydrology and snow models (e.g., Mizukami et al. 2014; Schmucki et al. 2014). To our knowledge, no study has examined an extensive suite of data availability scenarios for all required forcings across different climates and snow models.

Waichler and Wigmosta (2003) and Schnorbus and Alila (2004) considered subdaily forcing scenarios for the Distributed Hydrology Soil Vegetation Model (DHSVM) in maritime basins, but they lacked a key measurement (e.g., Q_{li}) and focused only on a single climate and model. Walter et al. (2005) suggested that a simple physical model could adequately simulate SWE using just daily T_{air} and P. Mizukami et al. (2014) tested how different reanalysis datasets changed hydrologic simulations in a continental climate. They found Q_{si} estimation most impacted basin hydrological processes, but surmised other models might respond differently to contrasting forcing datasets. Elsner et al. (2014) examined the impact of forcing dataset selection on VIC calibration and streamflow simulation, finding notable differences in calibrated parameters and monthly streamflow. Differences in T_{air} datasets propagated into their empirical estimates of radiation, impacting snowmelt simulations.

Schmucki et al. (2014) presented the most related work, in which six forcing availability scenarios were evaluated for the SNOWPACK model (Lehning et al. 2002) at three sites in the Swiss Alps. Their scenarios were dictated by whether P data were raw, corrected, or calibrated, and whether Q_{li}, reflected shortwave, and outgoing longwave radiation data were collectively available. They found P data quality was essential while radiation (e.g., Q_{li}) availability was less impactful. We expand on this work by examining a greater number of climatic regions and data availability scenarios, checking for model and climate dependencies. One key difference is that the P data are not calibrated here, as the purpose of the study is to find the next "best" AWS measurement after T_{air} and P. Unlike Schmucki et al. (2014), we do not assume U, RH, and Q_{si} are observed in all cases, which is more likely true for most AWSs globally. Finally, Schmucki et al. (2014) did not isolate the impact of Q_{li} because the availability of other radiation components varied with Q_{li} availability; we provide a more systematic and coherent test of the impact of Q_{li} availability on SWE simulations.

Many studies have used empirical techniques to estimate missing forcings and have compared empirical and observed values (e.g., Bohn et al. 2013; Flerchinger et al. 2009; Feld et al. 2013; Pierce et al. 2013). We examine how synthetic forcing data impacts model behavior, relying on previous studies to identify a reliable estimation method for each forcing. Alternative forcing datasets are available, such as gridded datasets (Maurer et al. 2002; Livneh et al. 2013), geostatistical interpolations (e.g., Jabot et al. 2012), numerical weather model output (Wayand et al. 2013; Rasmussen et al. 2011; Förster et al. 2014), and satellite-based methods (Ma and Pinker 2012; Pinker and Laszlo 1992; Forman and Margulis 2009). This study selects empirical methods (when possible) because of their sustained popularity in distributed hydrological modeling.

3. Study sites and observed forcing data

To assess whether the experiment had regional climate dependencies, we selected four sites (Fig. 2a, Table 1): 1) the tundra Imnavait Creek (IC) site (Sturm and Wagner 2010; Kane et al. 1991; Euskirchen et al. 2012) in Alaska (United States), 2) the maritime Col de Porte (CDP; Morin et al. 2012) site in the Rhône-Alpes (France), 3) the intermountain Reynolds Mountain East (RME) sheltered site (Reba et al. 2011) in Idaho (United States), and 4) the continental Swamp Angel Study Plot (SASP; Landry et al. 2014) in Colorado (United States). Three consecutive water years (WYs; October–September) were considered at each site to capture annual climate variations (Table 1). The study years at SASP (WYs 2006–08) were selected because these years had the lowest annual dust concentration relative to other years with available data (Painter et al. 2012; Skiles et al. 2012). Published (see above citations), serially complete hourly data (all six forcings) were available at CDP, RME, and SASP while data were prepared manually at IC (see the supplemental material).

The sites exhibited distinct snow and meteorological conditions (Figs. 2b–h). Peak SWE was typically highest at SASP and lowest at IC (Fig. 2b). Winter T_{air} was coldest at IC and warmest at CDP (Fig. 2c). CDP, RME, and SASP had similar accumulated P through April (~900 mm), while IC had low P accumulation (<200 mm) in the winter (Fig. 2d). Based on vapor pressure e, CDP was the moistest site while IC and SASP were the driest in the winter (Fig. 2e). Mean wind speeds were highest at IC and typically less than $2 \, \mathrm{m \, s^{-1}}$ at the other sites, which were sheltered (Fig. 2f). Mean Q_{si} was greatest at SASP in the winter and spring and lowest at IC in the winter and CDP in the spring (Fig. 2g). Finally, mean Q_{li} was greatest at CDP through winter and spring and lowest at IC and SASP (Fig. 2h). Mean monthly Q_{si} was only greater than Q_{li} from April to June at SASP, in May at IC, and June at RME; monthly Q_{li} was always greater than Q_{si} at CDP.
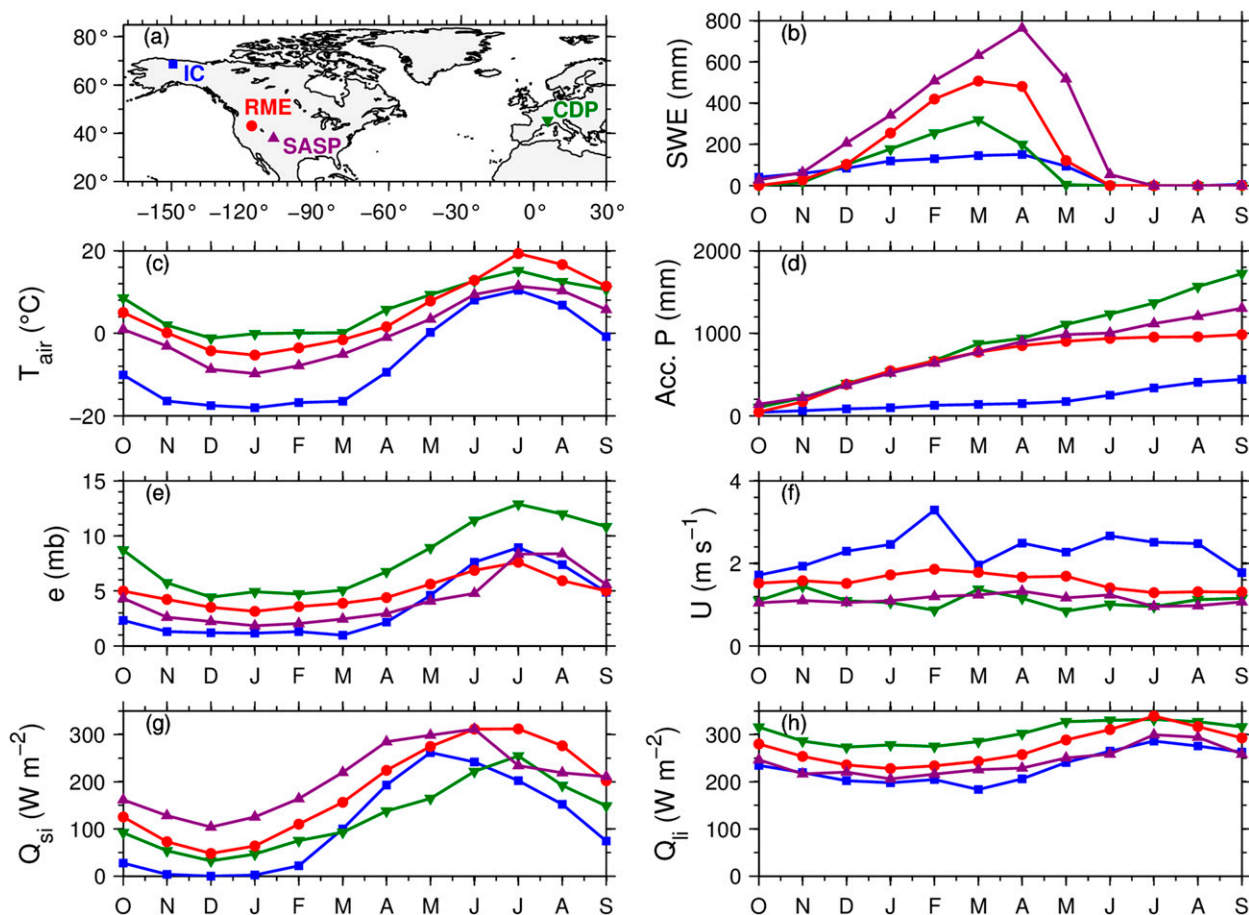
FIG. 2. (a) Global locations of the study sites. Mean monthly values (over 3-yr study period) of observed (b) SWE, (c) T_{air}, (d) accumulated P, (e) vapor pressure (calculated from RH and T), (f) U, (g) Q_{si}, and (h) Q_{li}. The line colors and markers in (b)–(h) correspond to the color and marker of each site in (a).

Observed snowpack data included SWE and snow surface temperature T_{surf} and were only used qualitatively to check the basic performance of the three snow models and to conduct a cursory diagnosis of model differences. SWE was measured with a snow pillow at IC and RME. At CDP, SWE was measured with a cosmic ray detector. At SASP, snow depths measured hourly with an ultrasonic sensor were converted to SWE using bulk snow density based on nearby SNOTEL data and weekly snow pit observations of density (Raleigh 2013). Manual SWE measurements were taken from weekly or biweekly snow pits (CDP and SASP) and snow courses (RME). Automatic measurements of T_{surf} were made at IC with infrared radiometers at three towers; we estimated missing values based on linear regressions with the other infrared radiometers and with dewpoint temperatures (Raleigh et al. 2013). At CDP, T_{surf} was derived from infrared radiometer observations and data from a down-looking pyrgeometer. RME T_{surf} was also derived from longwave measurements from a down-looking

pyrgeometer at a nearby exposed site. Finally, SASP T_{surf} data were observed with an infrared radiometer.

Corrected P data were available in the published datasets at CDP, RME, and SASP. Raw P data at IC were corrected assuming 75% undercatch, which was found for Wyoming-type gauges (Yang et al. 2000). Because the study assesses the next "best" forcing to measure at an AWS after T_{air} and P, rigorous correction of P was outside the scope. However, P data quality can critically impact modeled SWE (Schmucki et al. 2014; Raleigh et al. 2015). We allowed each model to calculate the phase of P (rain or snow) at each time step, a delineation that is more important at sites with milder winter climate (e.g., CDP and RME).

4. Methods

a. Experiment design and evaluation metrics

For each site and model, a data-withholding experiment was conducted at hourly resolution, where all observed forcings were available initially, and then sequentially

TABLE 1. Characteristics of the study years and sites, ordered by increasing elevation and decreasing latitude.

Site name	Site ID	Elev (m)	Lat (N)	Site climate classification	Study period (WY)	Mean winter P (mm)[a]	Dec–Feb temp (°C)	Mar–May temp (°C)	Peak SWE (mm)
Imnavait Creek	IC	930	68.6°	Tundra	2009[b]	200	−17.0	−1.9	180
					2010	140	−17.0	−2.3	120
					2011	190	−18.5	+4.9	180
Col de Porte	CDP	1330	45.3°	Mountain (maritime)	2006	790	−3.0	+10.2	440
					2007	940	+1.3	+7.7	220
					2008	1080	+0.5	+11.8	460
Reynolds Mountain East (sheltered site)	RME	2060	43.1°	Mountain (intermountain)	2006	1080	−3.6	+8.7	760
					2007	610	−3.8	+8.7	370
					2008	860	−5.6	+12.7	720
Swamp Angel Study Plot	SASP	3370	37.9°	Mountain (continental)	2006	860	−8.1	+3.5	840
					2007	840	−8.2	+4.0	700
					2008	980	−10.1	+7.6	1010

[a] We sum winter precipitation from 1 Oct to 30 Apr at all sites except IC, where we sum it from 1 Sep to 30 Apr.
[b] Because of early snowfall, simulations in this year began on 15 Sep instead of 1 Oct.

removed and replaced with algorithm-derived values until all forcings (except T_{max}, T_{min}, and daily P) were estimated (Fig. 3a). A binary construction of scenarios with six forcings was used, resulting in $2^6 = 64$ unique data availability scenarios. Each scenario can be envisioned as a "hypothetical AWS," where different sensors are available. The T_{air} and P observations (either hourly or daily) were assumed available in all cases because they are common measurements (Fig. 1); we tested whether temporal resolution in these core data mattered.

Preparation of hourly forcings for the nth scenario proceeded by first examining data availability in the scenario matrix (Fig. 3a), collecting the time series of observed forcings (if any), and then estimating any missing forcings using the methods selected in section 4b. Using daily T_{max}, T_{min}, and P, missing hourly forcings were estimated sequentially in the order T_{air}, P, RH, Q_{si}, and Q_{li} because of data dependencies (Fig. 3b). After assembling 3 years of forcing data for all 64 scenarios, the forcings were input into the three snow models (see section 4c).

The analysis consisted of three parts. In the first part of the analysis, differences between observed O and estimated E hourly forcings were compared graphically and with common metrics, including mean bias difference (MBD), root-mean-square deviation (RMSD), and mean absolute difference (MAD):

$$\text{MBD} = n^{-1} \sum_{i=1}^{n} (E_i - O_i), \qquad (1)$$

$$\text{RMSD} = \sqrt{n^{-1} \sum_{i=1}^{n} (E_i - O_i)^2}, \quad \text{and} \quad (2)$$

$$\text{MAD} = n^{-1} \sum_{i=1}^{n} |E_i - O_i|. \qquad (3)$$

MBD is particularly important because biases can impact snow modeling more than random errors (Raleigh et al. 2015; Lapo et al. 2015). The metrics above are only meaningful to a specific forcing, and thus it is not possible to compare metrics between two variables with different units (e.g., T_{air} MBD vs Q_{si} MBD). To gauge the relative accuracy of estimation methods across forcings, we conducted an auxiliary analysis (see the supplemental material) that considers three dimensionless metrics (Moriasi et al. 2007) and provides insights into the relative skill of all forcings estimated.

In the second part, observed SWE and T_{surf} were qualitatively compared to model outputs (scenario with all observed forcings) to assess baseline differences in the behavior of models when given the same ideal set of forcing data. The impacts of data availability were not examined in this comparison. Observed SWE and T_{surf} were used in this part only.

The third and final part focused on model output spread (from all 64 data scenarios) and assessed how availability of specific forcings contributed to the spread. We compared means of the two data availability ensembles for each forcing (e.g., 32 scenarios with observed Q_{li} vs 32 scenarios with estimated Q_{li}) and identified which forcings exhibited the greatest divergence between ensemble means. We focused on model divergence because differences in model output (for a given model structure and parameter set) guarantees that forcing estimation methods have deficiencies and/or observed forcing data have measurement errors. Differences between ensemble model means reflected 1) differences between observed and estimated forcings and 2) the importance of each forcing for simulating the response variable (e.g., SWE and T_{surf}) with each model. We considered differences in SWE, T_{surf}, and snow disappearance date (SDD). SWE and T_{surf} were summarizing indices of the snowpack mass
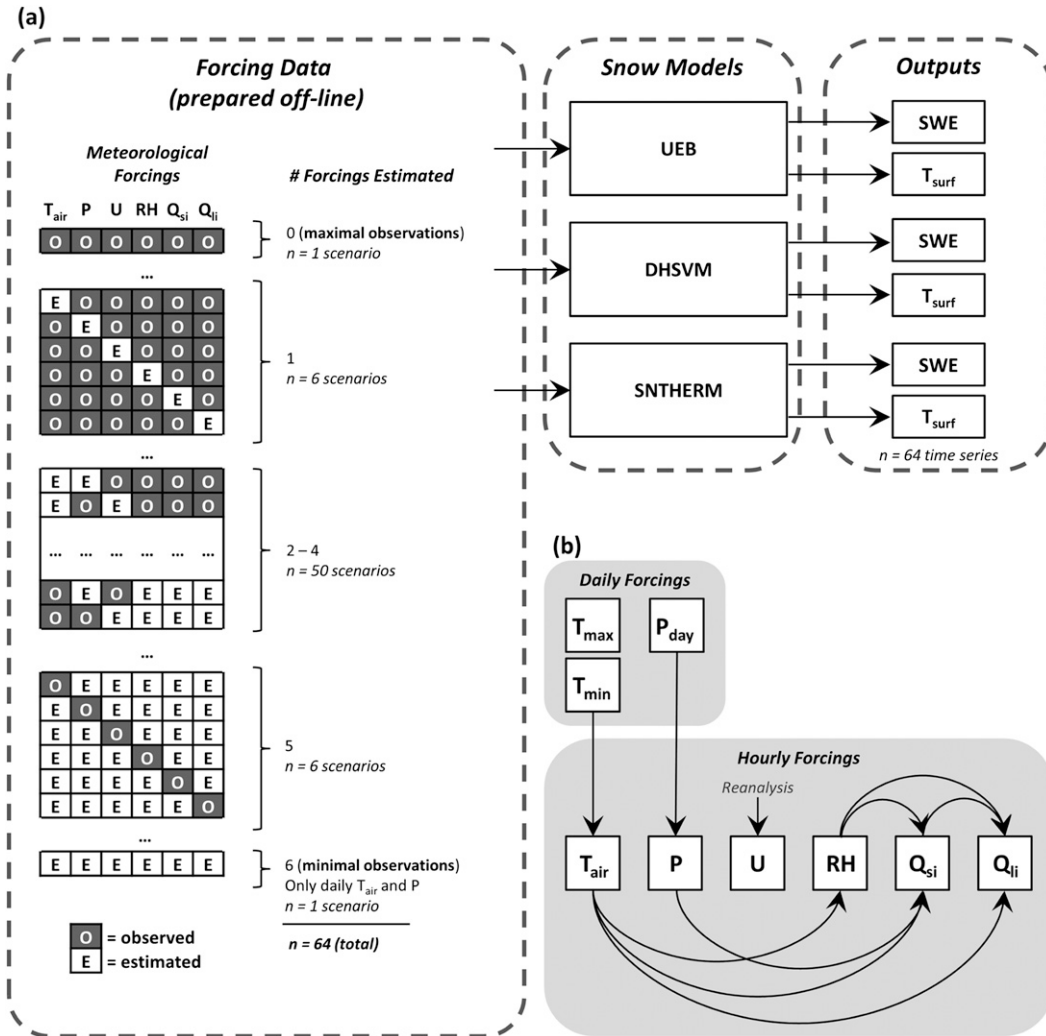
FIG. 3. Conceptual diagrams showing (a) the data assumed available in the 64 forcing scenarios used to force the three snow models to simulate SWE and T_{surf}, and (b) the dependencies among estimated forcing data.

and energy balance, while SDD was an index of both snow accumulation and melt.

To further quantify the impact of forcing availability on the mass and energy balances together, we used modeled SWE and T_{surf} to compute a loss function (Essery et al. 2013):

$$\text{loss function} = \sum_{i=1}^{3} \sum_{j=1}^{2} \sum_{k=1}^{N_{ij}} \frac{(S_{ijk}^{E} - S_{ijk}^{O})^2}{6N_{ij}w_{ij}^2} - 1, \quad (4)$$

where $\mathbf{S}^E$ is a matrix (N_{ij} rows $\times$ 32 columns) of model simulations when a specific forcing is estimated, $\mathbf{S}^O$ is a matrix of the corresponding simulations but with that forcing observed, i is the water year index (3 yr), j is the index of the model output (two outputs, SWE and T_{surf}), k is the time step index, N_{ij} is the number of time steps,

and w_{ij} is a weighting factor. Following Essery et al. (2013), w_{ij} was derived by setting Eq. (4) equal to 0 for the case of lowest difference between $\mathbf{S}^E$ and $\mathbf{S}^O$. We calculated 196 values (32 scenarios $\times$ 6 forcings) of the loss function for each site/model. To aid in interpretability, we normalized loss functions in the [0, 1] range, such that a value of 0 signified availability of that forcing impacted model output the least while a value of 1 indicated the greatest impact.

b. Forcing estimation methods

For each missing forcing variable (Fig. 3), a single estimation method was selected based on prior comparison studies between alternative empirical methods (whenever possible) and on popularity of usage. Literature values were used for any parameters in these

methods. The goal was not to find the "best" forcing estimator, but to apply a single method that might be selected when missing observations. These methods are briefly identified below.

The minimum data assumed available were daily maximum and minimum temperatures and daily P. Hourly T_{air} was estimated from T_{max} and T_{min} with a spline-based model, which was selected because it yielded lower errors relative to other models (Cesaraccio et al. 2001). Hourly P was estimated from daily P by uniformly dividing the daily sum across all hours of the day. This method has precedence in Waichler and Wigmosta (2003) and the widely used Maurer et al. (2002) dataset.

No empirical methods estimate hourly U accurately because of low correlation with other forcings (Parlange and Katz 2000). Therefore, we used NCEP–NCAR reanalyses data (Kalnay et al. 1996), which were the basis for the popular Maurer et al. (2002) dataset and have been used in recent snow model applications (Kang et al. 2014; Park and Markus 2014). We estimated hourly U based on the long-term (1981–2010) mean 6-hourly dataset of scalar wind speed at each site to capture typical diurnal and seasonal cycles. We considered using year-specific wind data from the reanalysis, but found that MBD (relative to observations) was usually lower when using the long-term mean dataset. Terrain-based downscaling of reanalysis data might reduce bias (Winstral et al. 2009), but this was not conducted.

Hourly RH was estimated assuming that daily dewpoint temperature T_{dew} equaled T_{min}, and T_{dew} was constant through each day (Running et al. 1987). Hourly RH was calculated as a function of vapor pressure (at T_{air}) and saturated vapor pressure (at T_{dew}). Kimball et al. (1997) developed an updated approach to address biases in arid regions. However, we used the simpler Running et al. (1987) method because it is easier to implement (i.e., no requirement for simultaneous estimation with Q_{si}) and has had wider usage (Feld et al. 2013).

Time series of Q_{si} were estimated with empirical relationships between daily T_{air} range and daily atmospheric transmissivity (Thornton and Running 1999). We estimated hourly Q_{si} assuming daily atmospheric transmissivity was constant across hours, and with the potential solar radiation changing with solar zenith angle. Estimation of atmospheric transmissivity requires vapor pressure, which we computed from T_{air} and RH. Bohn et al. (2013) evaluated the Thornton and Running (1999) daily model at 50 AWSs, finding biases ranging from -0.7 to $-47\,W\,m^{-2}$. Ball et al. (2004) compared this method to empirical and other mechanistic approaches and found similar accuracy across models,

despite differences in model complexity. The method was selected because it is commonly used in mountain modeling (e.g., MTCLIM; Hungerford et al. 1989) and land surface models (Bohn et al. 2013).

Hourly Q_{li} was empirically estimated with the clear-sky Dilley and O'Brien (1998) and the all-sky Unsworth and Monteith (1975) parameterizations, based on favorable evaluation in prior intercomparisons (Flerchinger et al. 2009; Juszak and Pellicciotti 2013). These were also used in the similar experiment of Schmucki et al. (2014). Hourly Q_{li} was estimated with the Stefan–Boltzmann equation:

$$Q_{li} = \sigma \varepsilon_{all} T_{air}^4, \tag{5}$$

where all-sky emissivity ε_{all} is scaled with cloud cover fraction c and clear-sky emissivity ε_{clr}:

$$\varepsilon_{all} = \varepsilon_{clr}(1 - 0.84c) + 0.84c, \tag{6}$$

and ε_{clr} is a function of precipitable water w and T_{air}:

$$\varepsilon_{clr} = \frac{59.38 + 113.7\left(\dfrac{T_{air}}{273.17}\right)^6 + 96.96\sqrt{\dfrac{w}{2.5}}}{\sigma T_{air}^4}. \tag{7}$$

In Eqs. (5) and (7), T_{air} is in units of kelvins. The Stefan–Boltzmann constant σ is $5.67 \times 10^{-8}\,W\,m^{-2}\,K^{-4}$. The c values in Eq. (6) ranged from 0 to 1 and were calculated based on the hourly solar index of Crawford and Duchon (1999). Nighttime c values were interpolated between the last 3 h prior to sunset and the first 3 h after sunset the next day. Precipitable water (cm) was estimated from vapor pressure and T_{air} (Prata 1996).

c. Snow models

We examined how hourly forcing availability impacted snowpack simulations from three physically based, 1D snow accumulation and melt models. The models included the Utah Energy Balance (UEB), the DHSVM snow submodel, and the snow thermal model (SNTHERM). We used common parameters (e.g., snow surface roughness) across models whenever possible and literature values otherwise. Model calibration was not conducted because it can compensate for forcing errors.

These models were selected because they have been used in snow model intercomparison studies and distributed model applications (Koivusalo and Heikinheimo 1999; Feng et al. 2008; Livneh et al. 2015) and because they span a range of complexity and process parameterizations. We also included multiple models to check for model dependencies in the results. Based on the number of snowpack layers, UEB was least complex, and

TABLE 2. Key differences between the physically based models in this study.

	UEB	DHSVM	SNTHERM
Partitioning of precipitation into rainfall vs snowfall	Based on air temperature (U.S. Army Corps of Engineers 1956), all precipitation is snow when $T_{air} \leq -1°C$, all is rain when $T_{air} \geq 3°C$, linear mix for temperatures in between	Based on wet-bulb temperature T_{wet} (calculated from temperature, humidity, and atmospheric pressure), all precipitation is snow when $T_{wet} < 1.6°C$, all is rain when $T_{wet} \geq 1.6°C$	Based on air temperature, all precipitation is snow when $T_{air} \leq 2.5°C$, all is rain when $T_{air} > 2.5°C$. Liquid water content in snowfall is calculated linearly in the 0%–40% range when $0° \leq T_{air} \leq 2.5°C$
Radiation penetration	Only considered in shallow snowpacks (depth <0.1 m), snow albedo and bare ground albedo interpolated in these cases	Not considered	Considered separately for near-infrared and visible wavelengths, near-infrared extinction limited to top 2 mm using user-specified extinction coefficient; visible extinction coefficient is a function of grain size
Snow albedo	Dependent on time since last snowfall and solar illumination angle, following Biosphere–Atmosphere Transfer Scheme (BATS) methodology (Dickinson et al. 1993)	Dependent on time since last snowfall, based on U.S. Army Corps of Engineers methodology (U.S. Army Corps of Engineers 1956), switches albedo decay curves as a function of T_{surf}	Based on physical calculations of snow grain size
Snow layers	Lumped with infinitesimally thin surface, implicit bulk snow and soil layer	Two layer (thin top layer of ≤100 mm)	Multilayer (up to 127–649 layers here, site dependent), explicit layers (snow and soil separated)
Time stepping	Fixed at resolution of forcing data	Fixed at resolution of forcing data	Adaptive (min = 1 s, max = 900 s)
Turbulent fluxes	Richardson number parameterization with aerodynamic resistance based on wind speed and the surface roughness parameter, stability corrections based on the Richardson number	Similar scheme as UEB, but different functional form	Obukhov length parameterization, includes windless exchange of sensible and latent heat

SNTHERM was most complex. Table 2 summarizes notable model differences, such as differences in albedo parameterization and turbulent heat exchange. Key model parameters are listed in the supplemental material.

1) UEB

UEB is a lumped snowpack (i.e., one layer with snow mass) model developed for prediction of snowmelt contributions to surface hydrology (Tarboton and Luce 1996; You et al. 2014; Mahat and Tarboton 2012). Key snowpack processes represented with UEB include snow accumulation, melt, sublimation, and liquid water retention and refreezing within the snowpack. The UEB snowpack has an infinitesimally thin surface layer where the surface energy balance and T_{surf} are solved. We used the latest version, UEBVeg (Mahat and Tarboton 2012; Mahat 2011; available at http://www.neng.usu.edu/cee/faculty/dtarb/snow/snow.html).

2) DHSVM

DHSVM is a distributed model developed for simulating the effects of topography and vegetation on hydrologic responses in complex terrain (Wigmosta et al. 1994). DHSVM is often applied at spatial scales ranging from 25 to 200 m. Here DHSVM was applied as a point model, with the snow model effectively isolated from other land surface routines. Key snowpack processes represented include snow accumulation, melt, sublimation, liquid water retention, and refreezing. DHSVM discretizes the snowpack into a thin (100 mm maximum) surface layer and lumps the remaining snow in a second layer below the surface. We used DHSVM, version 2.6, as implemented in Livneh et al. (2015).

3) SNTHERM

SNTHERM (Jordan 1991) is a multilayer model developed for prediction of thermal characteristics of seasonal snow and frozen soil. The model accounts for snow accumulation, compaction, densification, metamorphosis, sublimation, liquid water retention, and melt. Using mixture theory, SNTHERM simulates layer-specific density and moisture content. The partial differential equations governing heat and moisture exchange are solved with a finite control volume approach across horizontal layers of snow and soil. As snow

accumulates, the 1D grid compacts while maintaining the original finite element structure. A maximum of from 127 (at IC) to 649 (at SASP) snow layers was generated. SNTHERM provided the most detailed treatment of snowpack processes here, but at the expense of computational efficiency.

5. Results

a. Evaluation of estimated forcings

Figure 4 compares observed and estimated forcings (daily resolution for clarity). Table 3 displays evaluation statistics (hourly resolution). The T_{air} estimates were slightly colder than observations, with MBD ranging from $-0.3°$ to $0.0°$C. Not surprisingly, estimated P had MBD = 0, while RMSD and MAD in P were due to errors in hourly disaggregation (Table 3). Estimated U had MBD ranging from -0.30 to $+0.81\,\mathrm{m\,s^{-1}}$ (Figs. 4e–h, Table 3).

Two synthetic RH datasets were generated, one based on hourly T_{air} observations and the other based on estimated hourly T_{air} from daily temperature observations (Fig. 3b). At each site, the MBD of these estimated RH data were similar, ranging from $+4\%$ to $+16\%$ (observed hourly T_{air}) and from $+2\%$ to $+15\%$ (observed daily temperatures; Table 3). Estimated RH was typically greater than observed RH at all sites and data scenarios, indicating that T_{dew} was usually lower than T_{min}, especially at RME and SASP (Figs. 4i–l).

Two synthetic Q_{si} datasets were created, one assuming availability of observed T_{air}, P, and RH, and the other assuming only observed T_{air} and P (Fig. 3b). Availability of hourly versus daily T_{air} and P did not yield unique Q_{si} datasets because transmissivity was estimated from daily data. MBD in estimated Q_{si} ranged from -18 to $+22\,\mathrm{W\,m^{-2}}$ with RH observed (Figs. 4m–p) and from -19 to $+22\,\mathrm{W\,m^{-2}}$ without RH observed (Table 3).

Eight unique Q_{li} datasets were generated based on different data availability in T_{air}, RH, and Q_{si} (Fig. 3b, Table 3). MBD across the eight datasets ranged from -13 (CDP) to $+18\,\mathrm{W\,m^{-2}}$ (IC). Except at RME, the sign of the Q_{li} MBD was consistent at each site, with positive MBD at IC and SASP and negative MBD at CDP. The Q_{li} MBD was usually closest to 0 at RME (Fig. 4s, Table 3), and the Q_{li} differences were often greatest at IC, which may reflect issues with deriving Q_{li} from other radiation observations (see the supplemental material).

The dependence of estimated Q_{li} on Q_{si} (Fig. 3b) resulted in a relationship between the direction of MBD in estimated Q_{li} and Q_{si} (Table 3). For example, the biases

for estimated Q_{si} and Q_{li} had opposite signs at IC, CDP, and SASP. This made physical sense in that negatively biased Q_{si} resulted in positively biased Q_{li} with cloudier conditions [see c in Eq. (6)]. Likewise, positively biased Q_{si} yielded less cloudy conditions, resulting in negatively biased Q_{li}. This linkage was important in the ensemble analysis (see section 5c).

b. Comparison of observed and modeled snowpack with all forcings observed

Observed SWE and T_{surf} were qualitatively compared to model outputs (all forcings observed) at all sites (Fig. 5). At IC, the models estimated similar SWE and reasonably captured snowpack onset and disappearance (Fig. 5a). However, notable differences in observed and modeled SWE emerged at IC, such as during spring 2009 and the winter of WY 2011 (e.g., potential errors in observed P and/or observed SWE). At CDP (WY 2006 and 2008) and RME (all years), UEB SWE was closer to observations than the other models, which overestimated SWE (Figs. 5b,c). At CDP (WY 2007) and SASP (all years), UEB underestimated SWE, while SNTHERM and DHSVM overestimated SWE, particularly in the spring (Figs. 5b,d). UEB generally melted earlier that the other models, while SNTHERM melted last. At all sites, mean T_{surf} was colder in UEB than in observations and other models.

Examining the modeled snow processes helped identify reasons for similarities and differences in modeled SWE (Fig. 6). First, we considered the differences in rainfall–snowfall partitioning (Table 2). At the colder sites (IC and SASP), all models generally estimated similar snowfall fractions (Figs. 6a,d). At the warmer sites (CDP and RME), UEB generally had the lowest snowfall while DHSVM consistently had the most (Figs. 6b,c). SNTHERM generated similar (but 1%–3% lower than) snowfall totals as DHSVM. DHSVM partitioned 11% more snowfall than UEB at CDP, but only 5% more at RME and SASP and 3% more at IC. As expected, these snowfall estimates translated to monthly patterns in SWE accumulation (Figs. 6e–h).

Next, we considered differences in modeled ablation through decreases in monthly SWE (Figs. 6i–l). At IC, the models were similar; all had negligible snowmelt before April, and snowmelt in April was marginal (<25 mm; Fig. 6i). Starting in May at IC, UEB melted 26%–34% more snow than the other models (Fig. 6i), contributing to lower late season SWE and earlier snow disappearance in UEB (Fig. 5a). At CDP, all models simulated melt from November to May, but the monthly magnitudes varied across models (Fig. 6j). UEB had the greatest snowmelt from December through March (Fig. 6j), helping to explain why UEB SWE was lower
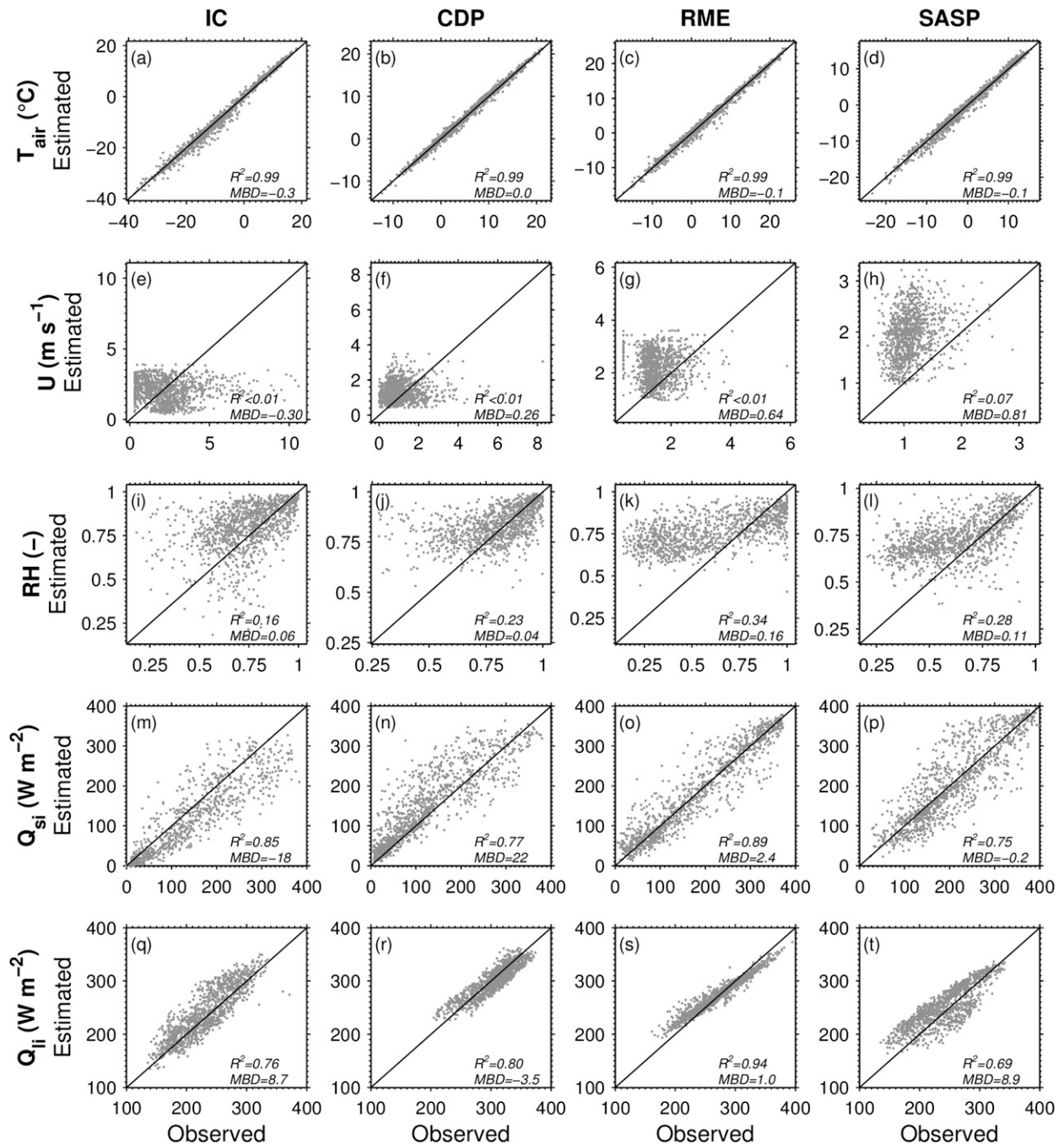
FIG. 4. Scatterplots of 3 yr of mean daily observed and estimated forcings at (from left to right) the sites (IC, CDP, RME, and SASP): (from top to bottom) T_{air}, U, RH, Q_{si}, and Q_{li}. Shown are estimates of each forcing when all other forcings are available. Precipitation observations and estimates are not shown because they are equivalent at the daily scale (see section 4b). The coefficient of determination R^2 and MBD (estimated minus observed, units on left axis) are provided. Note that the scales change between sites for some forcings.

than other models (Fig. 5b). At RME, all models had low snowmelt (<25 mm) in November and December, negligible melt in January and February, and more substantial melt after March (Fig. 6k). Again, UEB melted more snow than the other models in early spring

(March–April), contributing to lower SWE and earlier snow disappearance (Fig. 5c). At SASP, UEB had snowmelt in November (<25 mm; Fig. 6l), typically delaying the snow season and creating an offset from the other models throughout most of the year (Fig. 5d).

TABLE 3. Hourly error statistics of MBD, RMSD, and MAD for different estimated forcings for different data availability scenarios. Positive (negative) MBD indicates that the estimated forcing tends to have greater (lower) values than the observed forcing.

Target forcing (hourly)	Units	Available observed data[a]	MBD					RMSD					MAD				
			IC	CDP	RME	SASP	Mean	IC	CDP	RME	SASP	Mean	IC	CDP	RME	SASP	Mean
T_{air}	°C	T_{max}, T_{min}	−0.3	0.0	−0.1	−0.1	−0.1	2.9	1.7	2.1	2.2	2.3	2.0	1.2	1.4	1.6	1.5
P	mm h^{-1}	P_{daily}	0.0	0.0	0.0	0.0	0.0	0.1	0.6	0.4	0.5	0.4	0.0	0.2	0.1	0.2	0.1
U	m s^{-1}	−(NCEP$_{6hr}$)	−0.30	0.26	0.64	0.81	0.35	2.17	1.25	1.33	1.37	1.53	1.69	0.96	1.08	1.09	1.20
RH	—	T_{air}	0.06	0.04	0.16	0.11	0.09	0.20	0.16	0.29	0.22	0.22	0.15	0.12	0.22	0.17	0.17
RH	—	T_{max}, T_{min}	0.06	0.02	0.15	0.08	0.08	0.23	0.18	0.30	0.23	0.23	0.18	0.13	0.23	0.18	0.18
Q_{si}	W m^{-2}	T_{air}, P, RH	−18	22	2.4	−0.2	1.6	115	103	148	188	139	66	55	92	118	83
Q_{si}	W m^{-2}	T_{air}, P	−19	22	−1.4	−1.5	0.0	116	103	147	188	139	66	55	92	118	83
Q_{li}	W m^{-2}	T_{air}, RH, Q_{si}	8.7	−3.5	1.0	8.9	3.8	36	25	19	34	29	28	21	15	30	24
Q_{li}	W m^{-2}	T_{air}, RH	16	−13	−0.4	9.5	3.0	43	31	30	40	36	33	26	25	34	29
Q_{li}	W m^{-2}	T_{air}, Q_{si}	9.9	−2.4	8.6	12	7.2	37	26	20	38	30	29	22	16	33	25
Q_{li}	W m^{-2}	RH, Q_{si}	7.8	−3.3	0.40	8.7	3.4	37	27	21	36	30	30	23	16	31	25
Q_{li}	W m^{-2}	T_{air}	18	−12	8.1	14	6.9	44	31	29	43	37	34	26	24	37	30
Q_{li}	W m^{-2}	RH	15	−13	−1.0	9.4	2.6	44	34	32	43	38	34	28	26	35	31
Q_{li}	W m^{-2}	Q_{si}	8.6	−2.4	7.6	12	6.4	38	28	20	39	31	30	23	16	34	26
Q_{li}	W m^{-2}	T_{max}, T_{min}	16	−12	7.2	13	6.2	45	34	30	45	38	35	27	25	38	31

[a] All data are hourly unless otherwise noted with "6hr," "daily," "max," or "min" subscripts.

UEB was also the only model to yield snowmelt exceeding 50 mm before May at SASP, and UEB monthly melt totals were 50–240 mm greater than the other models in March–May (Fig. 6l). These factors caused UEB to have the lowest SWE (Fig. 5d).

UEB had lower SWE and earlier snow disappearance than DHSVM and SNTHERM because of both lower snow accumulation (at CDP and RME) and higher melt (all sites) during winter and early spring. Differences in albedo (Figs. 6m–p) and turbulent flux parameterizations (Figs. 6q–t) have caused major intermodel differences in prior studies (Lapo et al. 2015; Chen et al. 2014a). Here, patterns in modeled snowmelt were most consistent with modeled net shortwave radiation Q_{sn}. Differences in modeled Q_{sn} were due to albedo parameterization and radiation penetration because Q_{si} was the same across models. UEB generally had higher Q_{sn} and higher melt compared to the other models (Figs. 6m–p). The main exception occurred at CDP and RME when UEB had melted out and the other models continued melting (Figs. 6j,k). Albedo was typically 10% lower in UEB than SNTHERM. UEB accounted for radiation penetration in shallow snowpacks (<100 mm), making it more susceptible to early season melt (e.g., November in SASP; Fig. 5d), while DHSVM did not account for radiation penetration (Table 2). UEB and DHSVM modeled Q_{sn} data were generally closer to observations in midwinter than in spring, while SNTHERM Q_{sn} was consistently too low (Figs. 6n,p). All models underestimated Q_{sn} at SASP because they did not account for dust-on-snow effects. A more detailed assessment of mechanistic differences between models was beyond the scope; a modular approach (Essery et al. 2013; Clark et al. 2015a,b) could better pinpoint reasons for differences in model behavior.

c. Comparison of modeled snowpack with variable forcing availability

The remainder of the study focused on how data availability governed changes in modeled snow variables, and observed response variables were no longer compared. To first illustrate the spread in snow simulations across the 64 forcing scenarios, Fig. 7 shows SWE and T_{surf} ensembles from each model during an example water year. For each model, the greatest spread in SWE generally emerged in the spring. In some cases (e.g., all models at CDP and SASP), SWE simulations tended to cluster into distinct groups, with greatest coherency in the spring (Figs. 7b,d). In other cases (e.g., all models at RME), modeled SWE emerged as a continuum of responses (i.e., a single grouping). Like SWE, clustering was observed in T_{surf}, particularly at CDP and SASP (Figs. 7f,h). At some sites, the spread in model
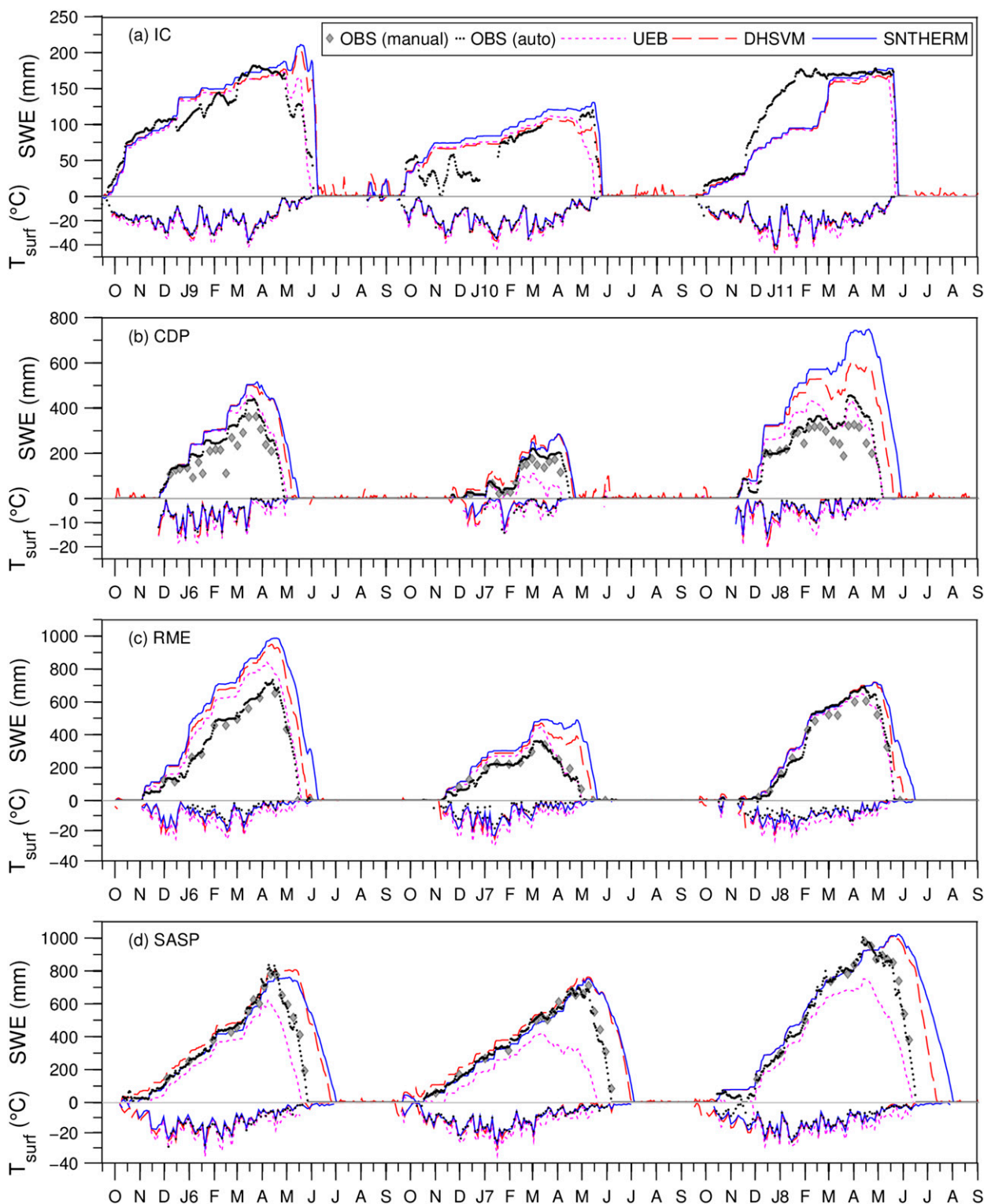
FIG. 5. Observed (manual OBS, gray diamonds; auto OBS, black dots) and modeled (black dotted) SWE and snow surface temperature (T_{surf}) with the three models [UEB (magenta dotted), DHSVM (red dashed), and SNTHERM (solid blue)] from October 2008 to September 2011 at (a) IC, (b) CDP, (c) RME, and (d) SASP. For all models, only the simulation generated with all observed forcings is shown. For clarity, daily mean SWE and 3-day mean T_{surf} are shown. The number next to "J" (January) represents the year (e.g., 6 signifies 2006).
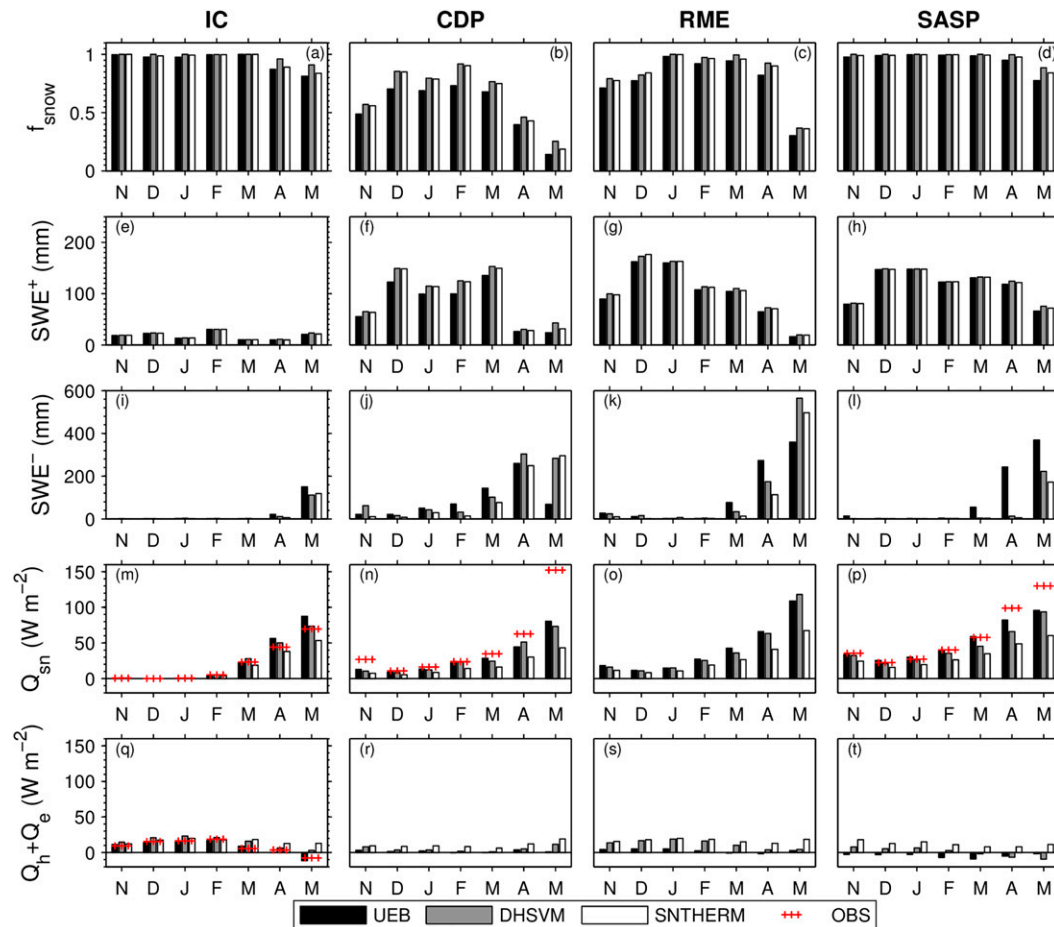
FIG. 6. Mean monthly summaries of simulated (a)–(d) snowfall fraction, (e)–(h) SWE increases, (i)–(l) SWE decreases, (m)–(p) net shortwave radiation, and (q)–(t) net turbulent fluxes (sensible flux + latent flux) from each model—UEB (black), DHSVM (gray), and SNTHERM (blank)—at (from left to right) the four sites. Values are averaged over the 3-yr study period in each month and only during times with modeled SWE > 25 mm. Also shown are observed values of net shortwave (red plus signs) at IC, CDP, and SASP and turbulent fluxes at IC.

simulations from the forcing data scenarios obscured intermodel differences (e.g., CDP; Figs. 7b,f), while at other sites the intermodel differences were larger (e.g., SASP; Fig. 7d).

Among the ensembles in Fig. 7, the scenarios with all hourly forcings and with only daily T_{air} and P were highlighted. At IC and CDP, these scenarios tended toward the middle of the ensemble in the spring (Figs. 7a,b), suggesting compensatory errors in the forcing estimation methods (Table 3) yielded similar snowpack simulations. In contrast, these scenarios enveloped the SWE and T_{surf} ensembles at the other sites, particularly with UEB and SNTHERM at RME (Figs. 7c,g) and SNTHERM at SASP (Figs. 7d,h). At RME and SASP, the minimal forcing data scenario (i.e., daily T_{air} and P only) tended to lower SWE relative to the scenario with all forcings observed (Figs. 7c,d). This

was consistent with higher T_{surf} in the minimal forcing data scenario (Figs. 7g,h), suggesting greater energy input and greater snowmelt sensitivity.

Comparing the ensemble mean of model output as a function of forcing availability clarified the major controls on model spread. Results from an example season are shown for SWE (Fig. 8). Maximum absolute SWE differences and mean absolute T_{surf} differences in the ensemble means over all study years appear in Table 4. At IC, CDP, and SASP, maximum SWE differences were often associated with Q_{li} availability, and in several cases, Q_{si} availability exerted an opposite effect on SWE (Fig. 8). Across models, Q_{li} availability caused the largest SWE differences at IC (65–67 mm) and CDP (174–234 mm; Table 4). At SASP, Q_{li} availability also contributed the largest SWE differences in SNTHERM (221 mm), but the second largest in UEB (130 mm) and
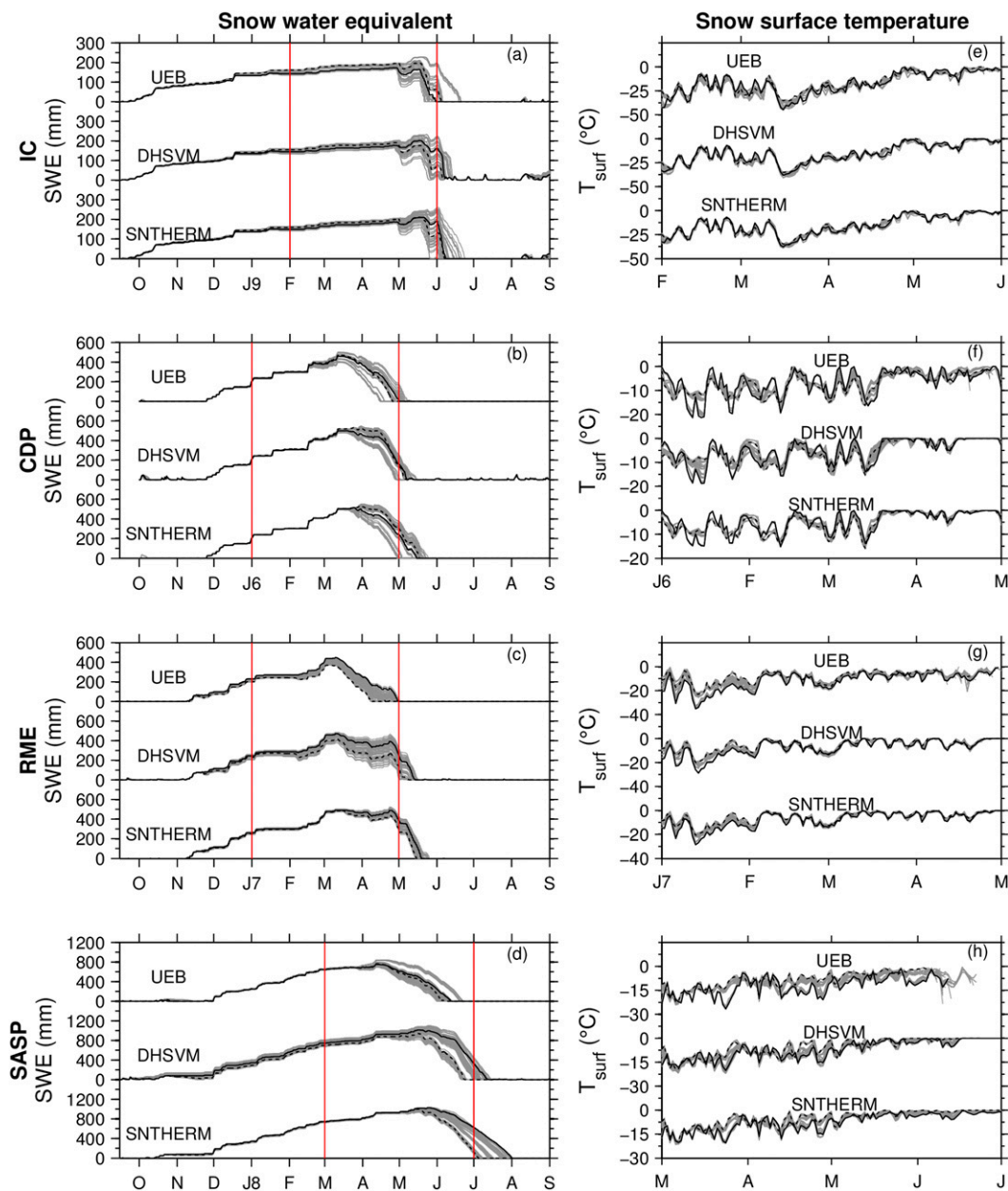
FIG. 7. Example time series of (a)–(d) SWE and (e)–(h) mean daily T_{surf} from three snow models (UEB, DHSVM, and SNTHERM) across 64 forcing data scenarios (gray lines) at (a),(e) IC in WY2009; (b),(f) CDP in WY2006; (c),(g) RME in WY2007; and (d),(h) SASP in WY2008. The solid black line denotes the scenario with all hourly forcings available, while the dashed line denotes the scenario with only daily T_{air} and P available. The period between the vertical red lines in (a)–(d) is shown in (e)–(h).

DHSVM (218 mm). Availability of RH caused the largest SWE differences at SASP with DHSVM (225 mm), and at RME with DHSVM (166 mm) and SNTHERM (71 mm). RH availability was more important for DHSVM than the other models (Table 4). Availability of Q_{si} only yielded the largest mean SWE differences at SASP with UEB (160 mm). Expressing SWE differences as a percentage of peak SWE (from the

ensemble mean with the observed forcing), Q_{li} yielded differences of 4%–93%, RH yielded 4%–25%, and Q_{si} yielded 4%–40%. The U availability created notable (but not the greatest) SWE differences at IC (all models), likely due to high wind speeds (Fig. 4e) and higher turbulent fluxes (Fig. 6q). For all sites and models, Q_{li} availability exhibited the strongest control on T_{surf} (Table 4). The MAD between the ensemble
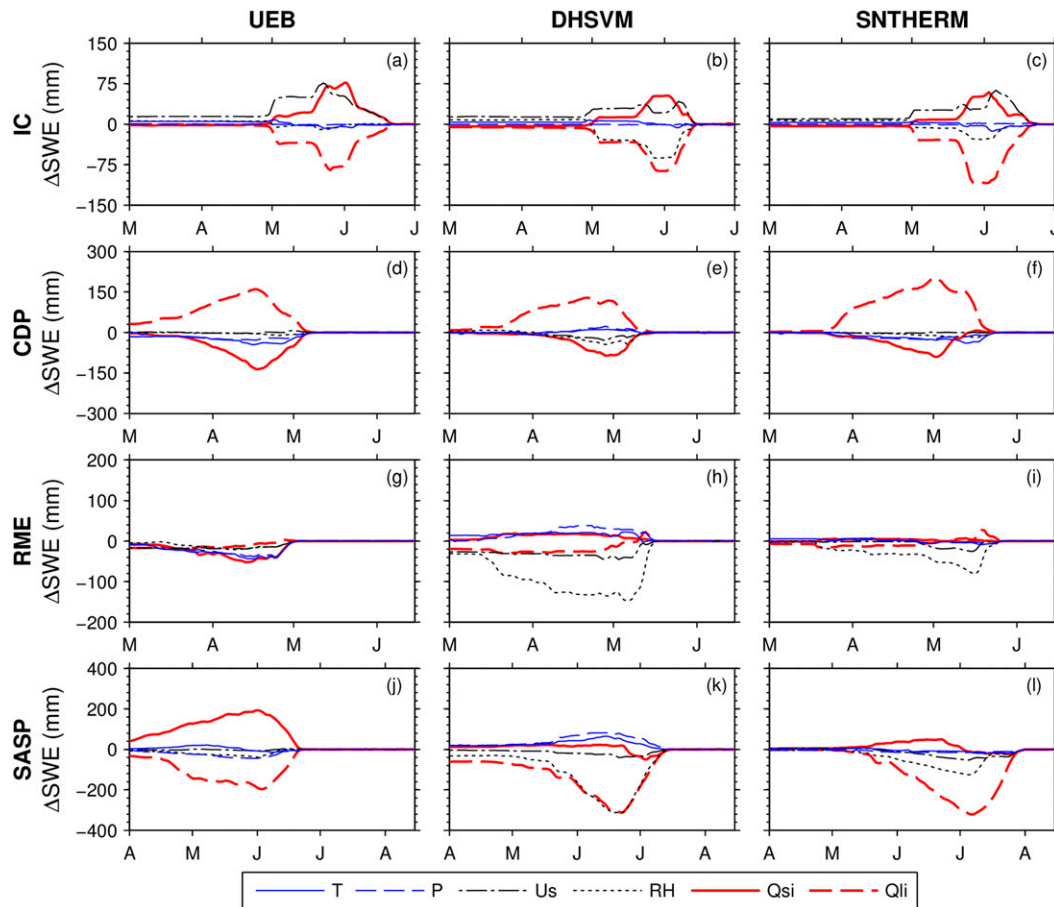
FIG. 8. Differences in mean SWE simulations for (from left to right) the three models at (from top to bottom) the four sites based on whether each forcing (lines) was observed or estimated. The difference was taken as mean SWE (across 32 scenarios with that forcing estimated) minus mean SWE (across 32 scenarios with that forcing observed). Positive differences indicated the model had higher SWE when the forcing was unavailable. Negative differences indicated the model had lower SWE when the forcing was unavailable. Only the late snow season is shown for a single year (2009 at IC, 2006 at CDP, 2007 at RME, and 2008 at SASP); similar results were found for other years.

mean T_{surf} with Q_{li} available versus unavailable ranged from 1.2° to 6.2°C (Table 4). Availability of T_{air}, U, and Q_{si} had secondary importance on simulations of T_{surf}.

When comparing differences in ensemble mean SDD based on forcing availability, three key results emerged (Fig. 9). First, Q_{li} availability often dictated the greatest or second greatest change in SDD. This was found for all models at IC, CDP, and SASP. Averaged across the study years, Q_{li} caused absolute differences in SDD ranging from 0 to 18 days (maximum = 32 days), whereas Q_{si} caused differences of 0–8 days (maximum = 9 days) and RH caused differences of 0–9 days (maximum = 10 days). Second, the models and the study years both exhibited qualitatively similar signatures (though different magnitudes) in the change in SDD because of forcing availability. In other words, the shape of the lines in Figs. 9a–d was similar to the lines in Figs. 9e–h and

Figs. 9i–l. Third, linked forcings, such as Q_{si} and Q_{li}, had opposing effects on modeled SDD in several cases (all models at IC and CDP, and UEB at SASP). This was related to the opposite sign of MBD in estimated Q_{si} and Q_{li} (section 5a), implying error compensation when both were unavailable.

The loss functions [Eq. (4)] summarized the impact of forcing data availability simultaneously on the mass and energy balances (Fig. 10). Availability of Q_{li} generally resulted in the highest loss function values, signifying this forcing contributed most to the spread in both SWE and T_{surf}. At IC and CDP (all models), RME (UEB and SNTHERM), and SASP (DHSVM and SNTHERM), Q_{li} had the highest median loss function. This was significantly higher than all other forcings (95% level, Wilcoxon rank-sum test) for DHSVM and SNTHERM at IC and at SASP, all models at CDP, and

TABLE 4. Differences between the model ensemble mean ($n = 32$) when a given forcing (T, P, U, RH, Q_{si}, Q_{li}) is observed hourly vs the model ensemble mean ($n = 32$) when that hourly forcing is unavailable. Shown are the maximum absolute SWE differences (expressed as mm and % of peak SWE from mean ensemble with that forcing observed) and mean absolute T_{surf} differences, averaged across the three study years at each site. The forcing that caused the greatest difference is in boldface for each site and model.

		UEB						DHSVM						SNTHERM					
	Site	T_{air}	P	U	RH	Q_{si}	Q_{li}	T_{air}	P	U	RH	Q_{si}	Q_{li}	T_{air}	P	U	RH	Q_{si}	Q_{li}
Max SWE	IC	16	2	60	16	62	**67**	14	9	51	34	57	**65**	27	40	49	32	58	**67**
difference (mm)	CDP	60	33	25	19	103	**205**	48	51	24	34	71	**174**	60	42	19	31	113	**234**
	RME	**47**	35	15	25	42	25	42	37	40	**166**	40	50	34	19	28	**71**	32	62
	SASP	33	51	14	36	**160**	130	63	121	51	**225**	80	218	43	26	59	82	44	**221**
Max SWE	IC	10	1	38	10	40	**41**	9	5	36	20	36	**38**	15	23	27	18	33	**35**
difference (% of	CDP	19	10	8	6	29	**93**	9	10	5	7	14	**40**	12	9	4	7	24	**57**
peak SWE)	RME	**8**	6	3	4	7	4	6	6	6	**24**	5	8	5	3	4	**10**	4	9
	SASP	6	9	2	6	**29**	21	8	15	6	**25**	10	24	5	3	7	9	5	**25**
Mean T_{surf}	IC	1.2	0.0	2.9	0.5	1.5	**4.6**	1.1	0.0	1.7	0.6	0.7	**2.1**	1.3	0.2	1.2	0.4	0.7	**2.5**
difference (°C)	CDP	0.7	0.4	0.6	0.3	1.3	**3.2**	0.3	0.1	0.4	0.3	0.3	**1.4**	0.4	0.2	0.1	0.1	0.5	**1.8**
	RME	0.7	0.4	1.4	0.6	1.6	**3.0**	0.5	0.1	**1.2**	0.9	0.4	**1.2**	0.7	0.3	0.9	0.5	0.6	**1.4**
	SASP	0.8	0.4	0.8	0.5	2.1	**6.2**	0.6	0.3	0.7	0.8	0.7	**3.0**	0.8	0.3	0.5	0.4	1.0	**3.9**

was significantly higher than four out of the other five forcings in all cases except DHSVM at RME. At RME, the loss function for Q_{li} was only significantly lower than the median loss function from RH in DHSVM (Fig. 10h). Qualitatively, the hierarchy of loss functions was similar when comparing models at IC (Figs. 10a–c) and CDP (Figs. 10d–f). However, there was more intermodel variability at RME (Figs. 10g–i) and SASP (Figs. 10j–l). The impact of forcing availability was not consistent across sites because the accuracy of the estimation methods varied (Table 3, Fig. 4).

6. Discussion

We aimed to identify which uncommonly measured forcings (Fig. 1) would most impact physical simulations of snowpack mass and energy states based on forcing data availability (i.e., through new instrumentation). For SWE and SDD, no single forcing was dominant across all climates and models, but Q_{li} availability was often among the most important (Figs. 8, 10 and Table 4). For T_{surf}, Q_{li} availability was consistently the largest contributor to model spread (Table 4). When considering both SWE and T_{surf}, Q_{li} availability often dictated model behavior (Fig. 10). These results suggest snow modeling in many climates would benefit from expanded observations of Q_{li} (measured at only 1.4% of AWSs in the western United States) or from improved Q_{li} estimation methods. This important finding was not encountered in prior studies because of a lack of Q_{li} observations (e.g., Waichler and Wigmosta 2003; Mizukami et al. 2014) or confounding factors in experiment designs (Schmucki et al. 2014).

When modeling snowpacks, it is important to consider how forcing availability and selected forcing estimation methods control model output. In general, differences between estimated and observed forcings in this study (Table 3) were similar to those found in previous studies. For example, Q_{li} estimation bias ranged from -12 to $+18\,W\,m^{-2}$ here, compared to -10 to $+16\,W\,m^{-2}$ in Flerchinger et al. (2009). Hourly RMSD in Q_{li} ranged from $+19$ to $+45\,W\,m^{-2}$, comparable in range with $+25$ to $+42\,W\,m^{-2}$ found by Juszak and Pellicciotti (2013). The Q_{si} bias ranged from -19 to $+22\,W\,m^{-2}$ here, while Bohn et al. (2013) found negative bias at their sites, ranging from -1 to $-47\,W\,m^{-2}$. The RH bias ranged from 2% to 11%, whereas Bohn et al. (2013) found bias ranging from -5.5% to 6.6% for a similar RH estimation method. Despite having positive bias, RH usually did not cause the greatest divergence in model output. The analysis across models provided context for how differences between observed and estimated forcings propagate to snow model output (Figs. 8–10).

Because only one estimation method was selected for each forcing across climates, the results strongly depend on the skill of that method for each climate and the sensitivity of snow processes to that forcing. Different methods do not always yield similar forcing estimates [e.g., Q_{li} (Flerchinger et al. 2009) and RH (Feld et al. 2013)], and performance of a method can be inconsistent across sites (Gubler et al. 2012). For example, estimated Q_{li} had negative MBD at CDP but positive MBD at IC and SASP (Table 3), and modeled SWE and SDD reflected these differences (Figs. 8, 9). Likewise, MBD at RME was often closest to 0 (Figs. 4q–t, Table 3), which explained why Q_{li} availability did not appear as important there as it did at the other sites. When considering normalized error metrics (see the supplemental material), Q_{li} was typically estimated with more skill than RH
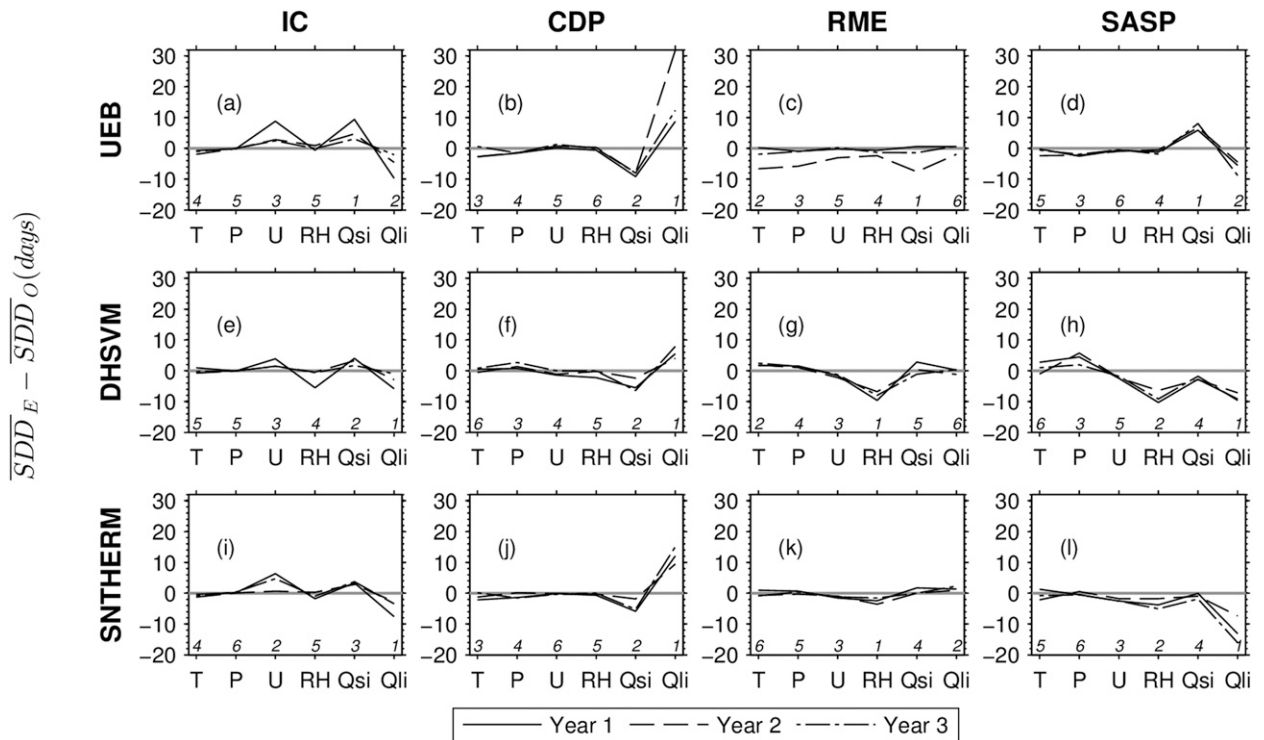
FIG. 9. Differences in mean SDDs by year 1, 2, and 3 at (from left to right) the four sites and for (from top to bottom) the three snow models as a function of estimated minus observed forcings. For each of the six forcings (from left to right on x axes), mean SDD is computed separately for all 32 scenarios where that forcing is estimated (E) and for all 32 scenarios where that forcing is observed (O). A positive value indicates SDD with the estimated forcing occurs later than the SDD with the observed forcing, while a negative value indicates SDD with the estimated forcing occurs earlier. The number at the bottom of each plot is the rank of the absolute SDD difference compared to that of the other forcing variables based on a 3-yr average.

and U but less skill than Q_{si}; hence, there is greater potential for improved modeling through more reliable Q_{li} data. Given incomplete forcings (Fig. 1), hydrologic modeling might benefit from considering multiple working hypotheses (Clark et al. 2008, 2011) about how unmonitored forcings may be modeled (i.e., multiple forcing estimation methods), which would provide insights into uncertainty in model forcing. Other forcing data estimators [e.g., reanalysis data (Mizukami et al. 2014) and numerical weather model output (Wayand et al. 2013)] provide alternative, albeit imperfect approaches (Slater et al. 2013; Shook and Pomeroy 2011) to empirical methods, but could be included in this framework.

The models had contrasting parameterizations and model complexity (Table 2), thereby yielding differences in modeled SWE and T_{surf} at CDP, RME, and SASP (Figs. 5, 6). The models had dissimilar portrayals of the energy balance (Fig. 6), and these likely affected each model's sensitivity to snowmelt, a major control on SWE (Fig. 6). Delayed snowmelt in DHSVM and SNTHERM (Figs. 5, 6) was likely caused by albedo specification; this is supported by Feng et al. (2008), who found SNTHERM and VIC (similar to DHSVM) had

high albedo. Koivusalo and Heikinheimo (1999) also found albedo is generally higher in SNTHERM than UEB, consistent with UEB melting first and SNTHERM last. These results support the prevailing understanding that albedo parameterizations can create significant model differences (e.g., Chen et al. 2014a). Even when controlling for albedo differences, however, differences in the energy balance can result because of factors such as turbulent flux parameterization, as recently demonstrated by Lapo et al. (2015) with SNTHERM and UEB. Despite the above differences, qualitatively consistent patterns emerged in how the models responded to forcing availability (Figs. 8–10), suggesting the results may apply to other snow models. For example, we might expect that the availability and quality of Q_{li} data at CDP would also have a major impact on SWE, T_{surf}, and SDD from other 1D physically based snow models (e.g., Lapo et al. 2015).

The study addresses technical aspects but not practical considerations, such as costs of purchasing, powering, and maintaining new sensors. Sensor costs can become progressively higher for radiation and humidity measurements. A joint T_{air} and RH sensor (e.g., HMP155A)
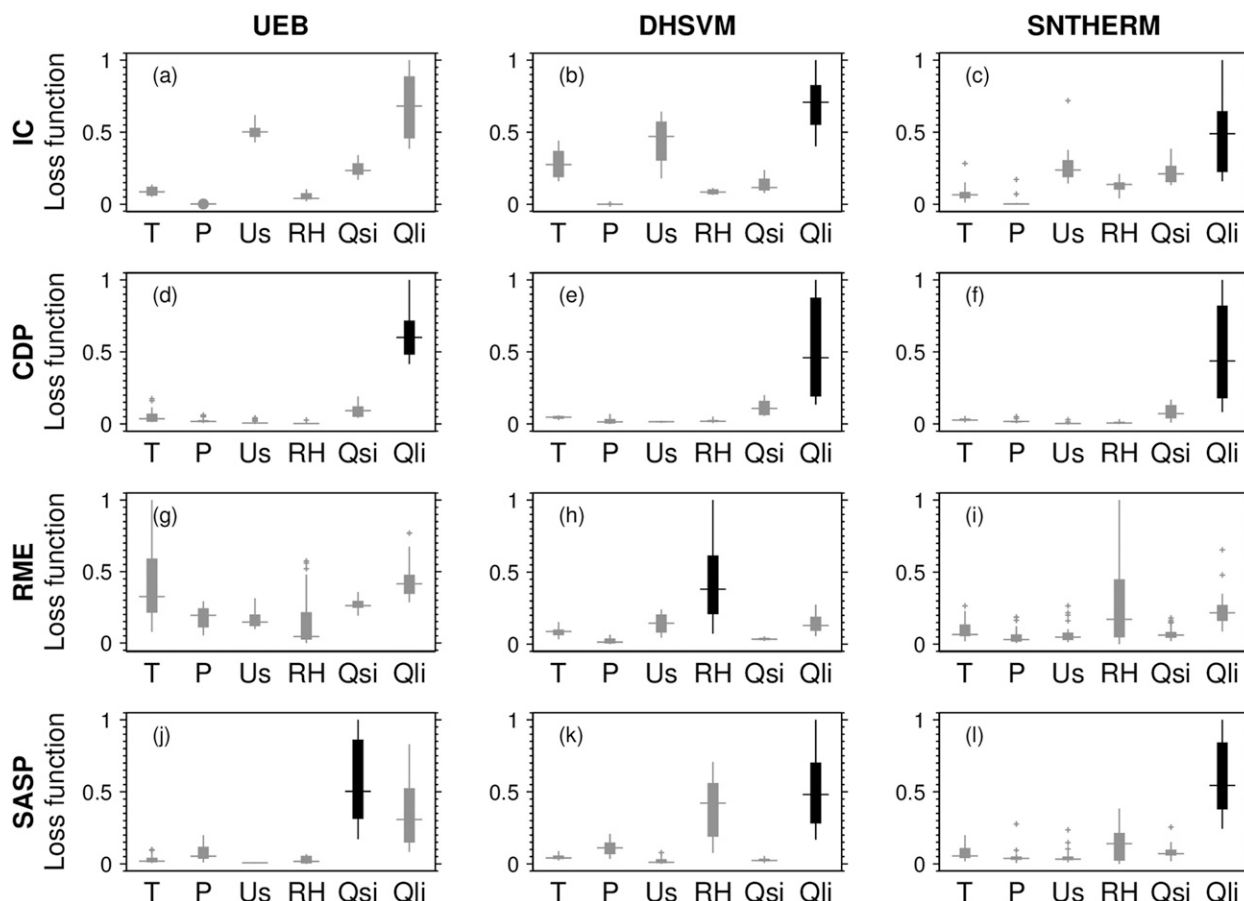
FIG. 10. Boxplots of the loss function by (from top to bottom) site, (from left to right) model, and (*x* axis) forcing. The loss function quantifies the importance of forcing availability for simulations of both SWE and snow surface temperature and is normalized such that values of 0 and 1 indicate the smallest and greatest impact of data availability on model simulations, respectively. All panels have a box for each of the six forcings, summarizing the distribution of loss functions ($n = 32$) for that forcing. The horizontal line in the box is the median, the filled box spans the 25–75th quartiles, the vertical whiskers extend to 1.5 times the interquartile range, and the plus signs indicate outliers. The forcing with the greatest median loss function appears in black if it is significantly different (95% level, Wilcoxon rank-sum test) than that of all other forcings. Forcings appear from most (left) to least (right) commonly available (Fig. 1).

can cost twice as much as a Met One 014A three-cup anemometer (to measure *U*), while a Kipp & Zonen CNR4 four-stream radiometer (to measure both Q_{si} and Q_{li} and upwelling radiation) can cost 20 times as much as an anemometer (D. Neff, Campbell Scientific, 2013, personal communication). These costs depend on sensor models and are reported for context. Coincidentally, this study recommends more frequent installation of the most expensive sensors, but financial and other considerations may govern ASW design. For example, site access and the frequency of site visits (maintenance and recalibration) are key considerations. More frequent (e.g., daily or weekly) site visits may be required to clean and inspect radiometer domes to ensure data quality (Augustine et al. 2000; World Meteorological Organization 2008), and these additional labor costs can become significant over time. Calibrations may incur additional costs and can

significantly influence data quality (Lundquist et al. 2015). The question also remains whether a forcing such as Q_{li} might be approximated using related data (e.g., observed Q_{si} for cloud factors or observed T_{surf}, which is tightly coupled with Q_{li}). Hence, many factors must be considered before upgrading existing AWS networks.

Our results apply most to sites where vegetation and wind redistribution effects are negligible and radiation drives snowmelt. In forests, the increasing contribution of Q_{li} and decreasing contribution of turbulent fluxes to the energy balance (Varhola et al. 2010; Lundquist et al. 2013) might emphasize our findings. Availability of Q_{si} may be more critical when dust on snow enhances the radiative forcing (e.g., Skiles et al. 2012) and the model accounts for this effect. Availability of *U* is likely more important in exposed areas where wind redistribution is important (Liston and Sturm 1998; Groot Zwaaftink

et al. 2013; Winstral et al. 2002). These accumulation processes were not considered because CDP, RME, and SASP are sheltered sites, and the effects of wind drift on P at the windiest site (IC) were reduced with correction. Thus, U primarily impacted modeled SWE through turbulent fluxes, which was less important than radiation at all sites (Fig. 6). Forcings related to turbulent fluxes (e.g., U and RH) are likely more important under special ablation conditions, such as rain on snow (Marks et al. 1998), chinook wind events, ephemeral low-elevation snowmelt, or wind-enhanced sublimation in dry climates. Turbulent fluxes also feedback with T_{surf} variations, which are tightly coupled with Q_{li} and atmospheric stability (e.g., Lapo et al. 2015; Raleigh et al. 2013; Slater et al. 2001). Therefore, the availability and quality of one or more forcings can have critical implications for the surface energy balance.

7. Conclusions

We conducted data-withholding experiments at four well-instrumented AWSs in contrasting snow climates. We were motivated to understand which meteorological forcings, beyond commonly measured T_{air} and P, most impact physically based snow model output and therefore need expanded observation or improved estimation in cold regions. While the hierarchy of forcings varied somewhat between models and sites, we found Q_{li}, which is measured least frequently in the western United States, most often caused the greatest divergences in model behavior. A more general conclusion that can be drawn from the results is that the particular configuration (i.e., collection of sensors) and location of an AWS impacts model performance.

Based on our snowpack modeling results, we conclude that the priority of expanded forcing observations at existing AWSs is (in order of decreasing importance) Q_{li} (primary), RH and Q_{si} (secondary), and U (tertiary). This recommendation demands detailed consideration of factors not assessed here, such as sensor costs and power, maintenance, model spatial scale, spatial variability, other purposes of AWSs (e.g., importance of U and RH for fire risk monitoring), and dominant meteorological conditions. Future research should continue evaluating methods for estimating forcings and for detecting and correcting errors in observations. Work is also needed to examine how data availability impacts modeled snow states across a wider range of landscapes where dominant snow processes differ. Finally, because this study only considered the case of adding sensors to an existing AWS, work is needed to understand how estimation errors propagate spatially to unmonitored locations in basinwide modeling applications.

REFERENCES

Augustine, J. A., J. J. DeLuisi, and C. N. Long, 2000: SURFRAD— A national surface radiation budget network for atmospheric research. *Bull. Amer. Meteor. Soc.*, **81**, 2341–2357, doi:10.1175/1520-0477(2000)081<2341:SANSRB>2.3.CO;2.

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.

Ball, R., L. Purcell, and S. Carey, 2004: Evaluation of solar radiation prediction models in North America. *Agron. J.*, **96**, 391–397, doi:10.2134/agronj2004.0391.

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

Blöschl, G., 1999: Scaling issues in snow hydrology. *Hydrol. Processes*, **13**, 2149–2175, doi:10.1002/(SICI)1099-1085(199910)13:14/15<2149::AID-HYP847>3.0.CO;2-8.

Bohn, T. J., B. Livneh, J. W. Oyler, S. W. Running, B. Nijssen, and D. P. Lettenmaier, 2013: Global evaluation of MTCLIM and related algorithms for forcing of ecological and hydrological models. *Agric. For. Meteor.*, **176**, 38–49, doi:10.1016/j.agrformet.2013.03.003.

Cesaraccio, C., D. Spano, P. Duce, and R. L. Snyder, 2001: An improved model for determining degree-day values from daily temperature data. *Int. J. Biometeorol.*, **45**, 161–169, doi:10.1007/s004840100104.

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

——, C. Liu, J. Dudhia, and M. Chen, 2014b: A sensitivity study of high-resolution regional climate simulations to three land surface models over the western United States. *J. Geophys. Res. Atmos.*, **119**, 7271–7291, doi:10.1002/2014JD021827.

Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay, 2008: Framework for Understanding Structural Errors (FUSE): A modular framework to

diagnose differences between hydrological models. *Water Resour. Res.*, **44**, W00B02, doi:10.1029/2007WR006735.

——, D. Kavetski, and F. Fenicia, 2011: Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resour. Res.*, **47**, W09301, doi:10.1029/2010WR009827.

——, and Coauthors, 2015a: A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resour. Res.*, **51**, 2498–2514, doi:10.1002/2015WR017198.

——, and Coauthors, 2015b: A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies. *Water Resour. Res.*, **51**, 2515–2542, doi:10.1002/2015WR017200.

Crawford, T., and C. Duchon, 1999: An improved parameterization for estimating effective atmospheric emissivity for use in calculating daytime downwelling longwave radiation. *J. Appl. Meteor.*, **38**, 474–480, doi:10.1175/1520-0450(1999)038<0474:AIPFEE>2.0.CO;2.

Deems, J. S., S. R. Fassnacht, and K. J. Elder, 2006: Fractal distribution of snow depth from Lidar data. *J. Hydrometeor.*, **7**, 285–297, doi:10.1175/JHM487.1.

Dickinson, R., A. Henderson-Sellers, and P. Kennedy, 1993: Biosphere–Atmosphere Transfer Scheme (BATS) Version 1e as coupled to the NCAR Community Climate Model. NCAR Tech. Note NCAR-TN-387+STR, 72 pp., doi:10.5065/D67W6959.

Dilley, A., and D. O'Brien, 1998: Estimating downward clear sky long-wave irradiance at the surface from screen temperature and precipitable water. *Quart. J. Roy. Meteor. Soc.*, **124**, 1391–1401, doi:10.1002/qj.49712454903.

Dubayah, R., 1992: Estimating net solar radiation using Landsat Thematic Mapper and digital elevation data. *Water Resour. Res.*, **28**, 2469–2484, doi:10.1029/92WR00772.

Elsner, M. M., S. Gangopadhyay, T. Pruitt, L. D. Brekke, N. Mizukami, and M. P. Clark, 2014: How does the choice of distributed meteorological data affect hydrologic model calibration and streamflow simulations? *J. Hydrometeor.*, **15**, 1384–1403, doi:10.1175/JHM-D-13-083.1.

Essery, R., 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.

Euskirchen, E. S., M. S. Bret-Harte, G. J. Scott, C. Edgar, and G. R. Shaver, 2012: Seasonal patterns of carbon dioxide and water fluxes in three representative tundra ecosystems in northern Alaska. *Ecosphere*, **3**, doi:10.1890/ES11-00202.1.

Feld, S. I., N. C. Cristea, and J. D. Lundquist, 2013: Representing atmospheric moisture content along mountain slopes: Examination using distributed sensors in the Sierra Nevada, California. *Water Resour. Res.*, **49**, 4424–4441, doi:10.1002/wrcr.20318.

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

Flerchinger, G. N., W. Xaio, D. Marks, T. J. Sauer, and Q. Yu, 2009: Comparison of algorithms for incoming atmospheric longwave radiation. *Water Resour. Res.*, **45**, W03423, doi:10.1029/2008WR007394.

Forman, B. A., and S. A. Margulis, 2009: High-resolution satellite-based cloud-coupled estimates of total downwelling surface radiation for hydrologic modelling applications. *Hydrol. Earth Syst. Sci.*, **13**, 969–986, doi:10.5194/hess-13-969-2009.

Förster, K., G. Meon, T. Marke, and U. Strasser, 2014: Effect of meteorological forcing and snow model complexity on hydrological simulations in the Sieber catchment (Harz Mountains, Germany). *Hydrol. Earth Syst. Sci.*, **18**, 4703–4720, doi:10.5194/hess-18-4703-2014.

Groot Zwaaftink, C. D., R. Mott, and M. Lehning, 2013: Seasonal simulation of drifting snow sublimation in alpine terrain. *Water Resour. Res.*, **49**, 1581–1590, doi:10.1002/wrcr.20137.

Gubler, S., S. Gruber, and R. S. Purves, 2012: Uncertainties of parameterized surface downward clear-sky shortwave and all-sky longwave radiation. *Atmos. Chem. Phys.*, **12**, 5077–5098, doi:10.5194/acp-12-5077-2012.

Hawkins, T., A. Ellis, J. Skindlov, and D. Reigle, 2002: Intra-annual analysis of the North American snow cover–monsoon teleconnection: Seasonal forecasting utility. *J. Climate*, **15**, 1743–1753, doi:10.1175/1520-0442(2002)015<1743:IAAOTN>2.0.CO;2.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis, 2005: Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.*, **25**, 1965–1978, doi:10.1002/joc.1276.

Horel, J. D., and X. Dong, 2010: An evaluation of the distribution of Remote Automated Weather Stations (RAWS). *J. Appl. Meteor. Climatol.*, **49**, 1563–1578, doi:10.1175/2010JAMC2397.1.

Hungerford, R., R. Nemani, S. Running, and J. Coughlan, 1989: MTCLIM: A mountain microclimate simulation model. U.S. Forest Service Intermountain Research Station Research Paper INT-414, 52 pp. [Available online at http://www.fs.fed.us/rm/pubs_int/int_rp414.pdf]

Jabot, E., I. Zin, T. Lebel, A. Gautheron, and C. Obled, 2012: Spatial interpolation of sub-daily air temperatures for snow and hydrologic applications in mesoscale alpine catchments. *Hydrol. Processes*, **26**, 2618–2630, doi:10.1002/hyp.9423.

Jin, J., and N. L. Miller, 2007: Analysis of the impact of snow on daily weather variability in mountainous regions using MM5. *J. Hydrometeor.*, **8**, 245–258, doi:10.1175/JHM565.1.

Jordan, R., 1991: A one-dimensional temperature model for a snow cover: Technical documentation for SNTERERM.89. Special Rep. 91-16, Cold Region Research and Engineers Laboratory, U.S. Army Corps of Engineers, Hanover, NH, 61 pp.

Juszak, I., and F. Pellicciotti, 2013: A comparison of parameterizations of incoming longwave radiation over melting glaciers: Model robustness and seasonal variability. *J. Geophys. Res. Atmos.*, **118**, 3066–3084, doi:10.1002/jgrd.50277.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kane, D. L., L. D. Hinzman, C. S. Benson, and G. E. Liston, 1991: Snow hydrology of a headwater Arctic basin: 1. Physical measurements and process studies. *Water Resour. Res.*, **27**, 1099–1109, doi:10.1029/91WR00262.

Kang, D. H., X. Shi, H. Gao, and S. J. Déry, 2014: On the changing contribution of snow to the hydrology of the Fraser River basin, Canada. *J. Hydrometeor.*, **15**, 1344–1365, doi:10.1175/JHM-D-13-0120.1.

Kimball, J., S. Running, and R. Nemani, 1997: An improved method for estimating surface humidity from daily minimum temperature. *Agric. For. Meteor.*, **85**, 87–98, doi:10.1016/S0168-1923(96)02366-0.

Kirchner, J. W., 2006: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resour. Res.*, **42**, W03S04, doi:10.1029/2005WR004362.

Koivusalo, H., and M. Heikinheimo, 1999: Surface energy exchange over a boreal snowpack: Comparison of two snow energy balance models. *Hydrol. Processes*, **13**, 2395–2408, doi:10.1002/(SICI)1099-1085(199910)13:14/15<2395::AID-HYP864>3.0.CO;2-G.

Kudo, G., 1991: Effects of snow-free period on the phenology of alpine plants inhabiting snow patches. *Arct. Alp. Res.*, **23**, 436–443, doi:10.2307/1551685.

Landry, C. C., K. A. Buck, M. S. Raleigh, and M. P. Clark, 2014: Mountain system monitoring at Senator Beck basin, San Juan Mountains, Colorado: A new integrative data source to develop and evaluate models of snow and hydrologic processes. *Water Resour. Res.*, **50**, 1773–1788, doi:10.1002/2013WR013711.

Lapo, K. E., L. M. Hinkelman, M. S. Raleigh, and J. D. Lundquist, 2015: Impact of errors in the downwelling irradiances on simulations of snow water equivalent, snow surface temperature, and the snow energy balance. *Water Resour. Res.*, **51**, 1649–1670, doi:10.1002/2014WR016259.

Lehning, M., P. Bartelt, B. Brown, and C. Fierz, 2002: A physical SNOWPACK model for the Swiss avalanche warning: Part III: Meteorological forcing, thin layer formation and evaluation. *Cold Reg. Sci. Technol.*, **35**, 169–184, doi:10.1016/S0165-232X(02)00072-1.

Liang, X., D. Lettenmaier, E. Wood, and S. 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.

Liston, G., and M. Sturm, 1998: A snow-transport model for complex terrain. *J. Glaciol.*, **44**, 498–516.

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.

——, E. A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier, 2013: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *J. Climate*, **26**, 9384–9392, doi:10.1175/JCLI-D-12-00508.1.

——, J. S. Deems, B. Buma, J. J. Barsugli, D. Schneider, N. P. Molotch, K. Wolter, and C. A. Wessman, 2015: Catchment response to bark beetle outbreak and dust-on-snow in the Colorado Rocky Mountains. *J. Hydrol.*, **523**, 196–210, doi:10.1016/j.jhydrol.2015.01.039.

Lundquist, J. D., D. R. Cayan, and M. D. Dettinger, 2003: Meteorology and hydrology in Yosemite National Park: A sensor network application. *Information Processing in Sensor Networks*, F. Zhao and L. Guibas, Eds., Lecture Notes in Computer Science, Vol. 2634, Springer, 518–528, doi:10.1007/3-540-36978-3_35.

——, S. E. Dickerson-Lange, J. A. Lutz, and N. C. Cristea, 2013: Lower forest density enhances snow retention in regions with warmer winters: A global framework developed from plot-scale observations and modeling. *Water Resour. Res.*, **49**, 6356–6370, doi:10.1002/wrcr.20504.

——, N. E. Wayand, A. Massmann, M. P. Clark, F. Lott, and N. C. Cristea, 2015: Diagnosis of insidious data disasters. *Water Resour. Res.*, **51**, 3815–3827, doi:10.1002/2014WR016585.

Ma, Y., and R. T. Pinker, 2012: Modeling shortwave radiative fluxes from satellites. *J. Geophys. Res.*, **117**, D23202, doi:10.1029/2012JD018332.

Madani, K., and J. R. Lund, 2010: Estimated impacts of climate warming on California's high-elevation hydropower. *Climatic Change*, **102**, 521–538, doi:10.1007/s10584-009-9750-8.

Mahat, V., 2011: Effect of vegetation on the accumulation and melting of snow at the TW Daniels Experimental Forest. Utah State University, Ph.D. thesis, 181 pp. [Available online at http://digitalcommons.usu.edu/etd/1078.]

——, and D. G. Tarboton, 2012: Canopy radiation transmission for an energy balance snowmelt model. *Water Resour. Res.*, **48**,

W01534, doi:10.1029/2011WR010438; Corrigendum, **48**, W03901, doi:10.1029/2012WR011964.

Marks, D., J. Kimball, D. Tingey, and T. Link, 1998: The sensitivity of snowmelt processes to climate conditions and forest cover during rain-on-snow: A case study of the 1996 Pacific Northwest flood. *Hydrol. Processes*, **12**, 1569–1587, doi:10.1002/(SICI)1099-1085(199808/09)12:10/11<1569::AID-HYP682>3.0.CO;2-L.

Maurer, E., A. Wood, J. Adam, D. Lettenmaier, and B. Nijssen, 2002: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *J. Climate*, **15**, 3237–3251, doi:10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2.

Mizukami, N., M. P. Clark, A. G. Slater, L. D. Brekke, M. M. Elsner, J. R. Arnold, and S. Gangopadhyay, 2014: Hydrologic implications of different large-scale meteorological model forcing datasets in mountainous regions. *J. Hydrometeor.*, **15**, 474–488, doi:10.1175/JHM-D-13-036.1.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Binger, R. D. Harmel, and T. L. Veith, 2007: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE*, **50**, 885–900, doi:10.13031/2013.23153.

Morin, S., Y. Lejeune, B. Lesaffre, J.-M. Panel, D. Poncet, P. David, and M. Sudul, 2012: An 18-yr long (1993–2011) snow and meteorological dataset from a mid-altitude mountain site (Col de Porte, France, 1325 m alt.) for driving and evaluating snowpack models. *Earth Syst. Sci. Data*, **4**, 13–21, doi:10.5194/essd-4-13-2012.

Painter, T. H., S. M. Skiles, J. S. Deems, A. C. Bryant, and C. C. Landry, 2012: Dust radiative forcing in snow of the Upper Colorado River basin: 1. A 6 year record of energy balance, radiation, and dust concentrations. *Water Resour. Res.*, **48**, W07521, doi:10.1029/2012WR011985.

Pan, M., 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.

Park, D., and M. Markus, 2014: Analysis of a changing hydrologic flood regime using the Variable Infiltration Capacity model. *J. Hydrol.*, **515**, 267–280, doi:10.1016/j.jhydrol.2014.05.004.

Parlange, M., and R. Katz, 2000: An extended version of the Richardson model for simulating daily weather variables. *J. Appl. Meteor.*, **39**, 610–622, doi:10.1175/1520-0450-39.5.610.

Pierce, D. W., A. L. Westerling, and J. Oyler, 2013: Future humidity trends over the western United States in the CMIP5 global climate models and variable infiltration capacity hydrological modeling system. *Hydrol. Earth Syst. Sci.*, **17**, 1833–1850, doi:10.5194/hess-17-1833-2013.

Pinker, R., and I. Laszlo, 1992: Modeling surface solar irradiance for satellite applications on a global scale. *J. Appl. Meteor.*, **31**, 194–211, doi:10.1175/1520-0450(1992)031<0194:MSSIFS>2.0.CO;2.

Prata, A., 1996: A new long-wave formula for estimating downward clear-sky radiation at the surface. *Quart. J. Roy. Meteor. Soc.*, **122**, 1127–1151, doi:10.1002/qj.49712253306.

Qu, X., and A. Hall, 2006: Assessing snow albedo feedback in simulated climate change. *J. Climate*, **19**, 2617–2630, doi:10.1175/JCLI3750.1.

Raleigh, M. S., 2013: Quantification of uncertainties in snow accumulation, snowmelt, and snow disappearance dates. Ph.D. thesis, University of Washington, 206 pp. [Available online at https://digital.lib.washington.edu/researchworks/handle/1773/24108.]

——, C. C. Landry, M. Hayashi, W. L. Quinton, and J. D. Lundquist, 2013: Approximating snow surface temperature

from standard temperature and humidity data: New possibilities for snow model and remote sensing evaluation. *Water Resour. Res.*, **49**, 8053–8069, doi:10.1002/2013WR013958.

——, J. D. Lundquist, 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.

Rasmussen, R., and Coauthors, 2011: High-resolution coupled climate runoff simulations of seasonal snowfall over Colorado: A process study of current and warmer climate. *J. Climate*, **24**, 3015–3048, doi:10.1175/2010JCLI3985.1.

Reba, M. L., D. Marks, M. Seyfried, A. Winstral, M. Kumar, and G. Flerchinger, 2011: A long-term data set for hydrologic modeling in a snow-dominated mountain catchment. *Water Resour. Res.*, **47**, W07702, doi:10.1029/2010WR010030.

Running, S. W., R. R. Nemani, and R. D. Hungerford, 1987: Extrapolation of synoptic meteorological data in mountainous terrain and its use for simulating forest evapotranspiration and photosynthesis. *Can. J. For. Res.*, **17**, 472–483, doi:10.1139/x87-081.

Schmucki, E., C. Marty, C. Fierz, and M. Lehning, 2014: Evaluation of modelled snow depth and snow water equivalent at three contrasting sites in Switzerland using SNOWPACK simulations driven by different meteorological data input. *Cold Reg. Sci. Technol.*, **99**, 27–37, doi:10.1016/j.coldregions.2013.12.004.

Schnorbus, M., and Y. Alila, 2004: Generation of an hourly meteorological time series for an alpine basin in British Columbia for use in numerical hydrologic modeling. *J. Hydrometeor.*, **5**, 862–882, doi:10.1175/1525-7541(2004)005<0862:GOAHMT>2.0.CO;2.

Scipión, D. E., R. Mott, M. Lehning, M. Schneebeli, and A. Berne, 2013: Seasonal small-scale spatial variability in alpine snowfall and snow accumulation. *Water Resour. Res.*, **49**, 1446–1457, doi:10.1002/wrcr.20135.

Shook, K., and J. Pomeroy, 2011: Synthesis of incoming shortwave radiation for hydrological simulation. *Hydrol. Res.*, **42**, 433–446, doi:10.2166/nh.2011.074.

Skiles, S. M., T. H. Painter, J. S. Deems, A. C. Bryant, and C. C. Landry, 2012: Dust radiative forcing in snow of the Upper Colorado River basin: 2. Interannual variability in radiative forcing and snowmelt rates. *Water Resour. Res.*, **48**, W07522, doi:10.1029/2012WR011986.

Slater, A. G., and Coauthors, 2001: The representation of snow in land surface schemes: Results from PILPS 2(d). *J. Hydrometeor.*, **2**, 7–25, doi:10.1175/1525-7541(2001)002<0007:TROSIL>2.0.CO;2.

——, A. P. Barrett, M. P. Clark, J. D. Lundquist, and M. S. Raleigh, 2013: Uncertainty in seasonal snow reconstruction: Relative impacts of model forcing and image availability. *Adv. Water Resour.*, **55**, 165–177, doi:10.1016/j.advwatres.2012.07.006.

Sturm, M., and A. M. Wagner, 2010: Using repeated patterns in snow distribution modeling: An Arctic example. *Water Resour. Res.*, **46**, W12549, doi:10.1029/2010WR009434.

Tarboton, D., and C. Luce, 1996: Utah Energy Balance Snow Accumulation and Melt Model (UEB): Computer model technical description users guide. Utah Water Research Laboratory/USDA Forest Service Intermountain Research Station, 64 pp. [Available online at http://www.fs.fed.us/rm/boise/publications/watershed/rmrs_1996_tarbotond001.pdf.]

Thornton, P. E., and S. W. Running, 1999: An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. *Agric. For. Meteor.*, **93**, 211–228, doi:10.1016/S0168-1923(98)00126-9.

Trujillo, E., N. P. Molotch, M. L. Goulden, A. E. Kelly, and R. C. Bales, 2012: Elevation-dependent influence of snow accumulation on forest greening. *Nat. Geosci.*, **5**, 705–709, doi:10.1038/ngeo1571.

Unsworth, M. H., and J. L. Monteith, 1975: Long-wave radiation at the ground I. Angular distribution of incoming radiation. *Quart. J. Roy. Meteor. Soc.*, **101**, 13–24, doi:10.1002/qj.49710142703.

U.S. Army Corps of Engineers, 1956: Snow hydrology: Summary report of the snow investigations. North Pacific Division, USACE, 437 pp.

Varhola, A., N. C. Coops, M. Weiler, and R. D. Moore, 2010: Forest canopy effects on snow accumulation and ablation: An integrative review of empirical results. *J. Hydrol.*, **392**, 219–233, doi:10.1016/j.jhydrol.2010.08.009.

Viviroli, D., R. Weingartner, and B. Messerli, 2003: Assessing the hydrological significance of the world's mountains. *Mt. Res. Dev.*, **23**, 32–40, doi:10.1659/0276-4741(2003)023[0032:ATHSOT]2.0.CO;2.

——, and Coauthors, 2011: Climate change and mountain water resources: Overview and recommendations for research, management and policy. *Hydrol. Earth Syst. Sci.*, **15**, 471–504, doi:10.5194/hess-15-471-2011.

Waichler, S. R., and M. S. Wigmosta, 2003: Development of hourly meteorological values from daily data and significance to hydrological modeling at H. J. Andrews Experimental Forest. *J. Hydrometeor.*, **4**, 251–263, doi:10.1175/1525-7541(2003)4<251:DOHMVF>2.0.CO;2.

Walter, T. M., E. S. Brooks, D. K. McCool, L. G. King, M. Molnau, and J. Boll, 2005: Process-based snowmelt modeling: Does it require more input data than temperature-index modeling? *J. Hydrol.*, **300**, 65–75, doi:10.1016/j.jhydrol.2004.05.002.

Watson, F. G. R., W. B. Newman, J. C. Coughlan, and R. A. Garrott, 2006: Testing a distributed snowpack simulation model against spatial observations. *J. Hydrol.*, **328**, 453–466, doi:10.1016/j.jhydrol.2005.12.012.

Wayand, N. E., A. F. Hamlet, M. Hughes, S. I. Feld, and J. D. Lundquist, 2013: Intercomparison of meteorological forcing data from empirical and mesoscale model sources in the North Fork American River basin in northern Sierra Nevada, California. *J. Hydrometeor.*, **14**, 677–699, doi:10.1175/JHM-D-12-0102.1.

Wigmosta, M. S., L. W. Vail, and D. P. Lettenmaier, 1994: A distributed hydrology–vegetation model for complex terrain. *Water Resour. Res.*, **30**, 1665–1679, doi:10.1029/94WR00436.

Winstral, A., K. Elder, and R. Davis, 2002: Spatial snow modeling of wind-redistributed snow using terrain-based parameters. *J. Hydrometeor.*, **3**, 524–538, doi:10.1175/1525-7541(2002)003<0524:SSMOWR>2.0.CO;2.

——, D. Marks, and R. Gurney, 2009: An efficient method for distributing wind speeds over heterogeneous terrain. *Hydrol. Processes*, **23**, 2526–2535, doi:10.1002/hyp.7141.

Winther, J.-G., and D. Hall, 1999: Satellite-derived snow coverage related to hydropower production in Norway: Present and future. *Int. J. Remote Sens.*, **20**, 2991–3008, doi:10.1080/014311699211570.

World Meteorological Organization, 2008: Guide to meteorological instruments and methods of observation. 7th ed. WMO 8, 681 pp.

Yang, D., and Coauthors, 2000: An evaluation of the Wyoming Gauge System for snowfall measurement. *Water Resour. Res.*, **36**, 2665–2677, doi:10.1029/2000WR900158.

You, J., D. G. Tarboton, and C. H. Luce, 2014: Modeling the snow surface temperature with a one-layer energy balance snowmelt model. *Hydrol. Earth Syst. Sci.*, **18**, 5061–5076, doi:10.5194/hess-18-5061-2014.

Zuzel, J. F., and L. M. Cox, 1975: Relative importance of meteorological variables in snowmelt. *Water Resour. Res.*, **11**, 174–176, doi:10.1029/WR011i001p00174.