Received: 14 January 2016 | Accepted: 25 August 2016

DOI 10.1002/hyp.11002

WILEY

RESEARCH ARTICLE

Diagnosing snow accumulation errors in a rain-snow transitional environment with snow board observations

Nicholas E. Wayand[1] | Martyn P. Clark[2] | Jessica D. Lundquist[1]

[1] University of Washington, Seattle, WA, USA

[2] National Center for Atmospheric Research, Boulder, CO, USA

Correspondence
Nicholas E. Wayand, Centre for Hydrology, University of Saskatchewan, Saskatoon, Canada.
Email: nic.wayand@usask.ca

Abstract

Diagnosing the source of errors in snow models requires intensive observations, a flexible model framework to test competing hypotheses, and a methodology to systematically test the dominant snow processes. We present a novel process-based approach to diagnose model errors through an example that focuses on snow accumulation processes (precipitation partitioning, new snow density, and snow compaction). Twelve years of meteorological and snow board measurements were used to identify the main source of model error on each snow accumulation day. Results show that modeled values of new snow density were outside observational uncertainties in 52% of days available for evaluation, while precipitation partitioning and compaction were in error 45% and 16% of the time, respectively. Precipitation partitioning errors mattered more for total winter accumulation during the anomalously warm winter of 2014–2015, when a higher fraction of precipitation fell within the temperature range where partition methods had the largest error. These results demonstrate how isolating individual model processes can identify the primary source(s) of model error, which helps prioritize future research.

KEYWORDS

compaction, error diagnosis, precipitation partitioning, rain-snow transition, snow density, snow modeling

Key Points

- Process-based diagnostics identified the density of new snow and precipitation partitioning as the most common sources of model error during accumulation days
- Ubiquitous precipitation partitioning errors during average winters were exposed during the anomalously warm winter of 2015
- Choices of parameter values were more important than competing functions for snow accumulation processes

1 | INTRODUCTION

Physically based models of the hydrological cycle are critical for testing our understanding of the natural world and enabling forecasting of extreme events. Previous intercomparison studies of existing snow models (i.e., SNOWMIP I and II, PILPS) have been hampered by multiple inter-model differences, making it difficult to attribute model-observation discrepancies to individual modeling decisions (Essery et al., 2009). Recent efforts to integrate multiple model hypotheses into a single framework (Clark, Kavetski, & Fenicia, 2011, 2015c; Essery, Morin, Lejeune, & Ménard, 2013; Essery, 2015) have provided the tools to enable for a more rigorous validation of process representation in models. However, there exist few snow observatories that measure sufficient physical states and fluxes to fully constrain the possible combinations within these multiple model frameworks (Essery et al., 2013; Landry, Buck, Raleigh, & Clark, 2014). In practice, observations of bulk snow states, such as snow water equivalent (SWE) or snow depth, are most commonly available. Calibrating a snow model using a single bulk variable can lead to compensatory errors (Essery & Etchevers, 2004), which may hide model deficiencies that matter during extreme/unusual storms.

We present a novel process-based calibration method that takes advantage of multiple observations (including snow board measurements) at the Snoqualmie Pass (SNQ) snow study site, located in the

 Copyright © 2016 John Wiley & Sons, Ltd.

maritime climate of the Washington Cascades, United States (Wayand, Massmann, Butler, Keenan, & Lundquist, 2015b). As an example of this method, we focus on snow accumulation (or lack thereof) because of its importance for regional seasonal water storage and supply (Elsner et al., 2010; Vano, 2015), its projected sensitivity to expected warming given the site's location in the rain-snow transition zone (Elsner et al., 2010; Klos, Link, & Abatzoglou, 2014), and the importance of precipitation phase in controlling runoff generation during flooding (Wayand, Lundquist, & Clark, 2015a; White et al., 2010) and attendant impacts on transportation delays through the Snoqualmie corridor (Barbara et al., 2008).

In this study, we compare simulations of snow pack were performed using the Structure For Unifying Multiple Modeling Alternatives (SUMMA, Clark et al., 2015b, 2015c, 2015c) with historical snow board measurements of daily accumulated SWE and snow depth. These data are used as proxies for daily precipitation partitioning and new snow density, in order to answer the following questions:

1. Which matters more, the functional form of a process (e.g., the choice of equation) or the parameters for that function?
2. What are the most common sources of errors in modeled snow depth accumulation at a maritime rain-snow transitional site?

2 | BACKGROUND

2.1 | Sources of snow model uncertainty

Multiple sources of error affect land surface model predictions of observed states and fluxes. For snow models run uncoupled to an atmosphere, those sources include: upper boundary conditions (meteorological forcing) (Newman et al., 2015; Raleigh, Lundquist, & Clark, 2014; Rössler et al., 2014), model structure (process representations and parameter values) (Clark et al., 2015c; Essery et al., 2013; Gupta, Clark, Vrugt J, Abramowitz, & M., 2012; Pomeroy et al. 2007), and numerical solver errors (Clark & Kavetski, 2010; Kavetski & Clark, 2011). In order to isolate specific areas of model structure that are inadequate, which is the goal of this study, errors within the upper boundary condition must be minimized to prevent them from biasing model evaluations. Numerical errors are not addressed here, but left for future study.

2.2 | Process observations

Unique observations are required to evaluate and improve existing snow models. While bulk snowpack states (e.g., SWE) are most relevant for streamflow predictions (Wood et al., 2015), internal snowpack states (i.e., layered density, liquid water content, temperature, grain size) are critical to evaluate individual process representation (Wever et al., 2015) and are required for some remote sensing applications (e.g., Langlois et al., 2012). Recent advances in internal snowpack observations (Kinar & Pomeroy, 2015; Schmid et al., 2014) now provide information on individual snow processes. Likewise, historical data sets, such as snow board measurements, provide a widely observed but generally under-utilized source of information on individual processes. To improve seasonal snow simulations, we need to improve our *methods* of evaluating model structure to take advantage of multiple internal snow pack observations. In this study, we present a novel methodology to evaluate individual processes that impact snow depth accumulation: partitioning of precipitation into ice or liquid, density of newly fallen snow, and the compaction of existing snow.

2.3 | Isolated processes

Model process representations tested in this study are illustrated in Figure 1 and described below. For an in-depth review of existing model parameterizations, see Essery et al. (2013) and Clark et al. (2015c).

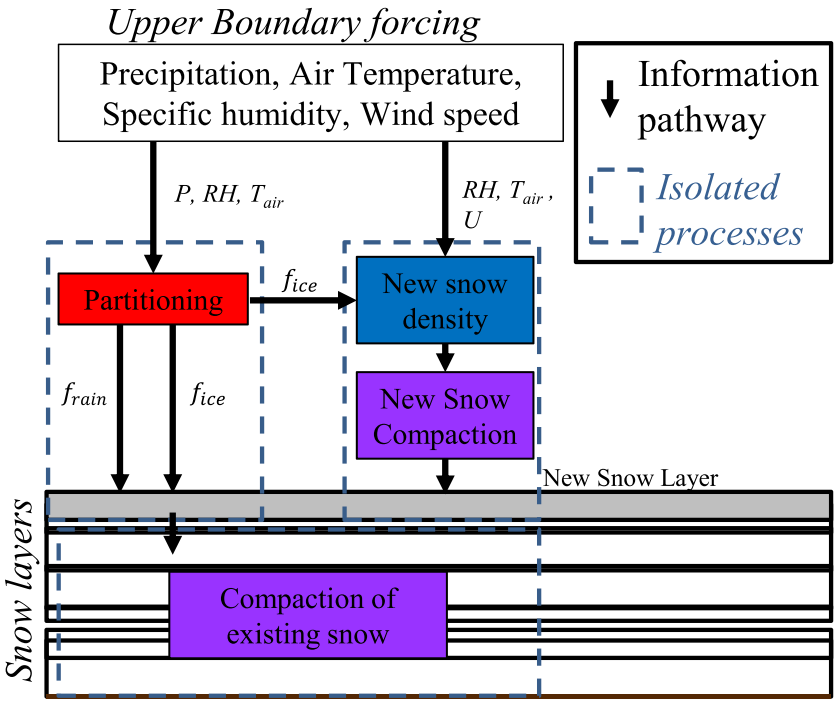

FIGURE 1 Example of isolating new snow accumulation processes. All fluxes into *dashed boxes* were taken directly from observations when possible. Variable abbreviations are defined in Table 2

2.3.1 | Precipitation phase at the surface

The phase of precipitation reaching the surface depends on the atmospheric conditions within the formation cloud as well the air mass it travels through during fallout to the ground (Lundquist et al., 2008). The most reliable way of sampling the instantaneous particle type at the surface is still by human observation, followed by laser disdrometers, both of which are rare in mountainous terrain. On a daily time scale, precipitation type can be estimated from the ratio of accumulated SWE on snow boards compared to total precipitation (Wayand et al., 2015b). Measuring precipitation in the form of snow is difficult and is the subject of intensive field campaigns (Nitu et al., 2012; Rasmussen et al., 2012; Yang & Goodison, 1998). In contrast, observations of total precipitation (i.e., a tipping or weighing bucket) are more common and usually less biased if properly sited, heated, and shielded (Sevruk, 1983). Thus, the partitioning phase determination is most often left to the hydrological model (Harder & Pomeroy, 2013) or the atmospheric model's microphysical scheme.

The most common method of predicting precipitation phase uses ground measurements of air temperature (T_a) (USACE, 1956; *Auer*, 1974), dew point temperature (T_d) (Marks, Winstral, Reba, Pomeroy, & Kumar, 2013), or wet bulb temperature (T_{wet}) (Harder & Pomeroy, 2013; Marks et al., 2013), and sometimes upper-air observations (Sims & Liu, 2015; Wayand, Clark, & Lundquist, 2016c), which are applied at subdaily or daily time scales. See Feiccabrino, Graff, Lundberg, Sandström, and Gustafsson (2015) for an in-depth review of existing methods. For this study, we use the wet-bulb temperature form based on the theory that it best represents the temperature of a falling hydrometer (Marks et al., 2013). T_{wet} was calculated iteratively using the psychrometric equation (Campbell & Norman, 1998) with saturated vapor pressure from (Buck, 1981), using in situ observations of T_a, relative humidity, and surface pressure.

2.3.2 | Density of newly fallen snow

The density of *newly fallen snow* is commonly measured on a surface over a period of 1 hr to 24 hr (Pfister & Schneebeli, 1999; USACE, 1956). Most often, the newly fallen snow is measured manually on a snow board, which is cleared periodically. The longer the period, the greater chance that other processes (i.e., wind redistribution, melting, settlement, etc.) may impact the measurement of newly fallen density, which complicates the evaluation of snow density parameterizations that are applied within a model at a range of time scales.

Methods used to predict newly fallen snow density (Anderson, 1976; Boone, 2002 Hedstrom & Pomeroy, 1998; Oleson, Lawrence, & Gordon, 2010) have been fitted to observational data measured as described above over a range of snow climates (maritime, continental, alpine). However, they all are based on surface air temperature (Anderson, 1976; Hedstrom & Pomeroy, 1998) and wind speed (Pahaut, 1976). The ability of these surface variables to characterize the variability of newly fallen snow density is known to be low (Roebber, Bruening, Schultz, & Cortinas, 2003). In operation, a common rule of thumb of ~100 kg m^{-3} is often simply used, which attempts to capture the mean density.

2.3.3 | Compaction of accumulated snow

Compaction, or densification, of bulk snow is commonly measured by comparing the change in density measured from both bulk SWE and depth observations. The compaction of individual layers has also been measured using settling disks placed post-storm (Morin et al., 2012; USACE, 1956). Compaction of the underlying snowpack (referred to here as "old" snow) during snow accumulation can be measured from the difference between newly accumulated snow depth (i.e., snow board) and the bulk snow depth change (USACE, 1956), assuming no loss of mass through melt. We used this latter method as illustrated in Figure 2a.

Parameterizations of model compaction vary from a simple constant rate to empirical functions that depend on snow viscosity, overburden pressure, metamorphism, and liquid water. We use the commonly applied (Anderson, 1976) function, but only focus on parameters impacting compaction because of overburden (Table 1) as we show that this is the dominant process during snow accumulation events.

2.3.4 | Summary

The above three processes are examined at the Snoqualmie Pass snow study site described in section 3. The snow model framework used here

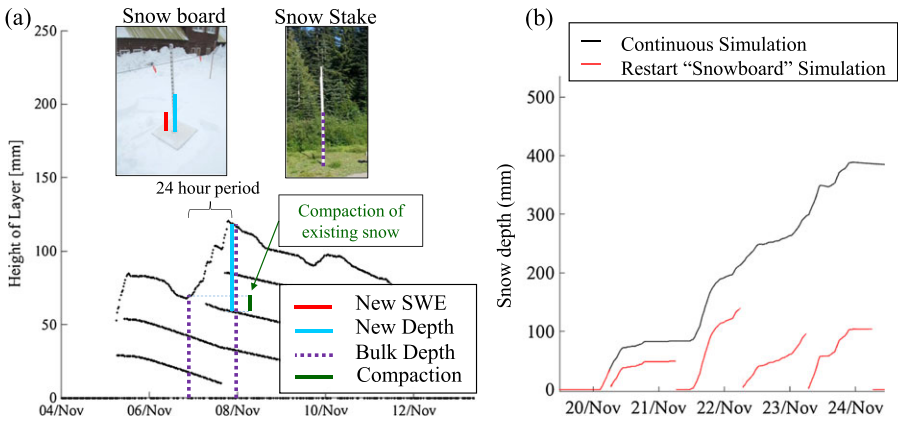

FIGURE 2 (a) Images show available measurements of new snow water equivalent (SWE), new snow depth, and bulk snow depth. *Horizontal black lines* show modeled snow layers (accumulation, compaction, and merging) that are representative of observed layers shown in right image. (b) Example of the "modeled snowboard" simulations, which were compared to observations in (a). Note: the *red line* shows modeled snow depth at the *end* of modeled hourly time step; thus, first step is non-zero

TABLE 1 Model options and parameter values

Process	Option	Parameter full name (symbol) [unit]	Min value	Max value	Default value from literature	Lumped method selected value	Process-based method *without* cascading errors selected value	Process-based method *with* cascading errors selected value
Precipitation partitioning	(USACE, 1956)	T_{wet} threshold for rain-snow (T_{crit}) [°C]	-1	1	1	0	-0.24	-0.24
		Mixed-phase range (δT)[°C]	0.1	5	4	0.5	1.625	1.625
New snow density	(Hedstrom & Pomeroy, 1998)	Minimum new density (ρ_{min}) [kg m^{-3}]	50	100	67.92	100	–	–
		Density multiplier (*A*) [kgm^{-3}]	25	75	51.25	50	–	–
		Density scalar (*B*) [°C]	1	5	2.59	1	–	–
	(Pahaut, 1976)	Density additive (*C*) [kg m^{-3}]	80	120	109	–	–	80
		Density multiplier (*D*) [kg m^{-3} C^{-1}]	1	12	6	–	–	1
		Wind speed factor (*E*) [kg $m^{-7/2}$ $s^{1/2}$]	16	36	26	–	–	21
	(Anderson, 1976)	Minimum new density (ρ_{min}) [kg m^{-3}]	50	100	50	–	–	–
		Density MULTIPLIER (*F*) [$°C^{-1}$]	1	3	1	–	–	–
	Constant new snow density	Constant density (ρ_{const}) [kg m^{-3}]	50	250	100	–	76.8	–
Compaction	(Anderson, 1976)	Overburden density scalar (*G*) [kg^{-1} m^3]	0.02	.036	0.0230	0.02	0.02	0.02
		Overburden temperature scalar (*H*) [$°C^{-1}$]	0.06	0.1	0.0800	0.06	0.095	0.1

and the process-based calibration method are described in section 4. Selected model parameters and diagnosis of the sources of model errors are reported in section 5. The uncertainty and implications of error diagnosis and future uses of the methodology described here are then discussed in section 6, followed by a summary of the main conclusions in section 7. All abbreviations are defined in Table 2.

3 | STUDY SITE AND DATA

The Snoqualmie Snow study site is located within the upper rain-snow transition zone (921 m), receiving ~50% of October through June precipitation as snow. Despite mild temperatures, annual peak snow depths reach 2.6 m on average, accumulated by 12 storms per winter on average. The site is ideal for diagnosing modeled new snow accumulation errors because of its continuous record of snow board observations and meteorological forcing data. Site pictures, time-lapse movies, and a complete description of available data are provided by Wayand et al. (2015b), at http://dx.doi.org/10.6069/H57P8W91. Despite numerous snow observations, no bulk SWE measurements (e.g., a snow pillow) were available or used in this study. Below, we detail all data used to drive and evaluate model simulations in this study.

3.1 | Snow observations

Throughout winter months (typically November to May) since 1974, Washington Department of Transportation avalanche crew took daily snow measurements at approximately 6:00 am PST on a 0.22 m^2 (45 cm × 48 cm) snow board (Figure 2a), which was then cleared and replaced for the next day. Accumulated snow depth was measured from a graduated snow board stake. SWE was separately measured with a snow cutter (289.51 cm^2 area) and weighed with a scale. We refer to the 24-hr accumulations as *new* snow depth and *new* SWE. *Bulk* snow depth was manually read from a second 4 m snow stake each 6:00 am PST. Observational uncertainties were estimated as ±25 mm for bulk snow stake readings of snow depth, ±10 mm for new snow depth, and ±1.3 mm for new SWE measurements (calculated from the new snow depth error of ±10 mm and the average observed new snow density of 130 kg m^{-3}). Additional uncertainty

TABLE 2 Definitions of abbreviations used

KGE	Kling-Gupta Efficiency
SWE	Snow water equivalent
New Snow	The new snow accumulated on a snow board within 24 hr
Old Snow	The existing snowpack underlying the accumulated "New" snow
Bulk Snow	The total snow depth from the ground to the snow surface (New and Old)
P	Precipitation
RH	Relative humidity
T_{wet}	Wet-bulb temperature
T_{air}	Near surface air temperature
U	Near surface wind speed
f_{ice}	Fraction of precipitation ice
f_{rain}	Fraction of precipitation liquid

from horizontal transport of snow onto or off of the snow board was assumed negligible given the low mean wind speeds of 0.6 m s^{-1} at the SNQ site, which is surrounded by forest (see Figure 2 in Wayand et al., 2015b). Finally, we note that no automatic measurements of bulk SWE were available (e.g., a snow pillow), as is common at the majority of snow study sites worldwide.

The above daily snow observations were used as proxies for daily partitioning, new snow density, and compaction of existing snow. The daily fraction of snowfall to total precipitation was calculated as the ratio of accumulated SWE on the snow board to the total precipitation measured by the heated and shielded tipping bucket. The density of the 24-hr accumulation of new snow was calculated directly as the ratio of measured SWE to measured snow depth on the snow board. Compaction of the existing snow pack (underlying the daily new snow accumulation) was estimated as the difference between the 24-hr change in bulk snow depth and the measured new snow depth accumulated on the snow board. This difference is illustrated in Figure 2a as the *green vertical bar*. Note that we were not able to isolate the compaction of new snow, which is further discussed in section 6.2. All three daily proxies are used to evaluate process representation.

3.2 | Meteorological forcing data

The meteorological data set used to drive snow model simulations was taken directly from Wayand et al. (2015b), except as follows. We restrict our study period to water years (Oct. – Sept.) 2004 through 2015, when all forcing variables critical for simulating snow accumulation (air temperature, relative humidity, wind speed, and precipitation) were measured in situ (see Figure 3 in Wayand et al., 2015b). Air pressure and incoming irradiance measurements were available only after 2008 and 2012, respectively. Missing forcing data were primarily filled from the National Land Data Assimilation Systems (NLDAS) data (Cosgrove, 2003), bias-corrected to available in situ observations at SNQ as described in Wayand et al. (2015b). Precipitation gauge undercatch of snowfall was corrected using snow board SWE observations. Finally, half-hourly data for water years 2013 through 2015 were aggregated to hourly time steps to create a consistent meteorological forcing data set over water years 2004–2015.

4 | METHODS

We present a novel method for diagnosing sources of errors in modeled snow accumulation, which we refer to as the process-based method. This approach is contrasted with a more common method of calibrating a snow model using only observations of bulk snow depth. All simulations were performed using the Structure for Unifying Multiple Modeling Alternatives (SUMMA) model, with modifications as described below.

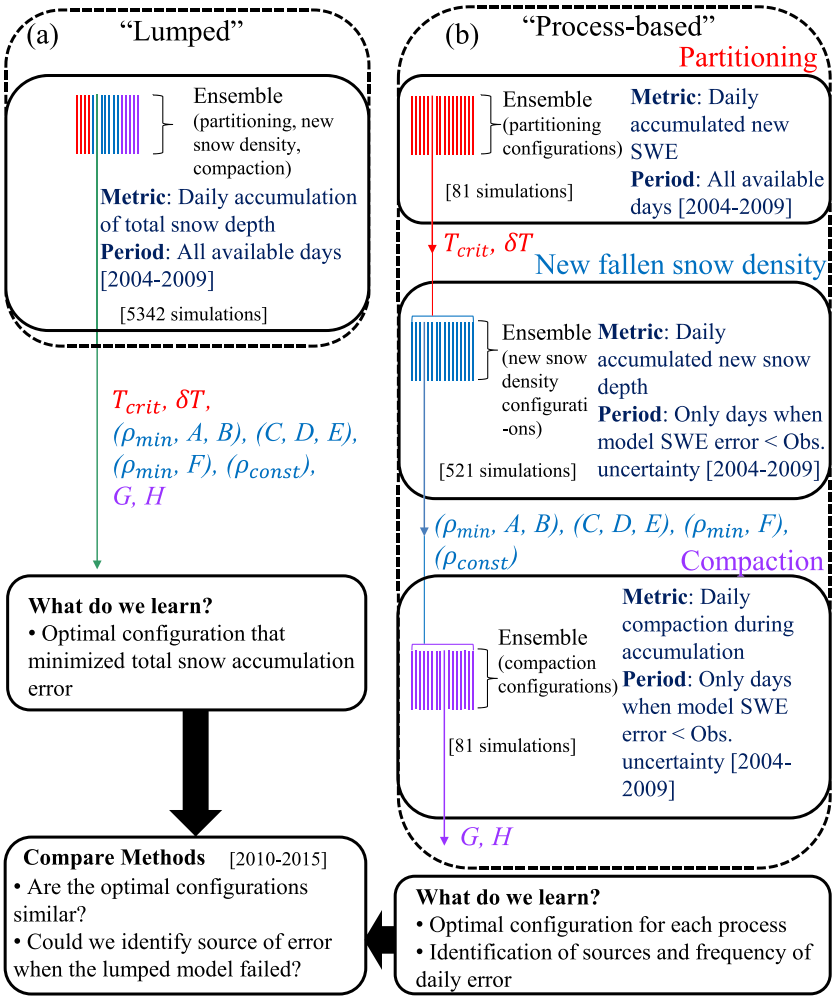

FIGURE 3 Methodology for both (a) Typical and (b) Process-based methods. The typical method uses a lumped calibration approach by running all model option/parameter configurations. In contrast, the process-based method evaluates one process at a time, passing on the selected option/parameters to the next process evaluation. Selected model option/parameters are defined in Table 1

4.1 | Snow model (SUMMA)

SUMMA is a physically based hydrological model that simulates the energy balance and transport of water through the canopy, snowpack, soil, and ground water (Clark et al., 2015a, 2015c, 2015b). For each physical process, the model includes multiple hypotheses of process representations that are currently used in existing snow and land surface models. Required meteorological forcing data consist of: air temperature, precipitation rate, wind speed, specific humidity, air pressure, and incoming short- and long-wave irradiance. Complete model configurations are provided in Table S1, and equations for each process are provided in Appendix S1.

4.2 | Simulating snow board measurements with SUMMA

SUMMA simulations were modified in order to allow a direct comparison to the snow board observations of 24-hr accumulated SWE and snow depth (Figure 2a). This was a critical step for a fair model evaluation. Because the snow board measurements were made daily at 6:00 am PST, and then the board was cleared of all snow, SUMMA simulations were restarted every 6:00 am PST with the previous accumulated snowpack removed (Figure 2b). Each daily restart simulation used initial soil conditions from a continuous SUMMA simulation with default parameter settings with two significant modifications: (a) the soil albedo was set at 0.8 to match the albedo of a snow board, and (b) the upper soil layer temperatures were set to the upper snow layer temperatures (if present in the continuous simulation) to match the surface temperature of the actual snow board prior to snow fall. A sensitivity analysis found that the choice of plausible snow board albedo (0.6–0.9) or reasonable initial temperature values did not impact simulated daily SWE or snow depth accumulation significantly.

4.3 | Lumped calibration method

A lumped calibration method (Figure 3a) was used to set an upper bound of SUMMA performance based on previous work that has shown that allowing free-ranging parameters will always identify the optimal fit to any designated criteria (Gong, Shen, Hong, Liu, & Liao, 2011; Nearing & Gupta, 2014). A 5342 member ensemble of continuous SUMMA simulations was created by varying function and parameter combinations for partitioning, new snow density, and compaction simultaneously, as described in section 4.2. Parameter values were selected through uniform random sampling through feasible parameter space as defined in Table 1 and illustrated in Figures S1–S6.

We only used the bulk daily accumulation of snow depth for calibration to be representative of a typical observational snow site (i.e., without manual snow board or bulk SWE measurements). Simulated daily accumulation of bulk snow depth for each ensemble member was evaluated during the 2004 to 2009 water years using the modified Kling-Gupta efficiency (KGE) (Gupta, Kling, Yilmaz, & Martinez, 2009; Kling, Fuchs, & Paulin, 2012),

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (1)$$

where r is the correlation coefficient, α is the ratio of the modeled to observed coefficient of variations, and β is the ratio of model mean to observed mean. This metric was chosen because it can be separated into a correlation term, a variance term, and a bias term (Magnusson & Wever, 2015). We note that the variance term did not significantly impact results here, but keep the formula to remain consistent with previous snow model evaluation studies. A perfect simulation has a KGE value of one.

4.4 | Process-based calibration

A process-based method to calibrate snow model options/parameters (referred to as "model configuration" from here on) relating to snow accumulation was designed to minimize the propagation of forcing and model errors that impact calibration (Figure 3b). We evaluated one process at a time and attempted to remove cascading errors into the next process evaluated (e.g., information pathways in Figure 1).

First, we minimized meteorological forcing errors by selecting the 2004 to 2015 water years where the majority (98%) of forcing data that impact new snow accumulation (air temperature, precipitation, wind speed, and relative humidity) were taken from in situ observations. Next, an ensemble of simulations (Table 1) was run for modeled processes impacting snow accumulation examined in this study (partitioning, new snow density, and compaction). Process parameterizations were selected from a range of the most commonly used functions in snow hydrology models (Clark et al., 2015c; Essery et al., 2013). Parameter ranges for each function were determined by choosing values that resulted in physically possible estimated states or fluxes, as illustrated in Figures S1–S6. Each ensemble of simulations, referred to as a process experiment, was run in a step-wise approach as illustrated in Figure 3b and described below.

For the partitioning experiment, we evaluated all days when both precipitation occurred and observations of daily accumulated SWE were available (N = 564). Figure 4a shows all model simulations compared to snow board observed SWE for an example period where the sensitivity of the linear partitioning method varied depending on the wet-bulb temperature (T_{wet}). The ensemble member with the highest KGE value was selected, and its parameter values were applied to the snow density experiment, as illustrated in Figure 3b.

For the snow density experiment, only days when the simulated SWE error was less than observational uncertainty (± 1.3 mm) were evaluated to prevent large partitioning errors from propagating into new snow density evaluations. Again, the ensemble member with the highest KGE value was selected, and its option/parameter values were applied to the compaction. Figure 4b illustrates modeled new snow depth sensitivity to density model configuration. This direct comparison is only possible because we modified SUMMA to match observations. For instance, if we had instead only used the difference in the bulk snow depth between days, compaction of the underlying snowpack would have biased our measure of new snow density. The impact of snow density changes between the time of snowfall and the time of the snow board measurement are discussed in section 6.2.

The modeled compaction of the existing snow pack ("old snow") was only evaluated on days with new snow accumulation and not in between snow fall events. On days with snowfall, we assumed that

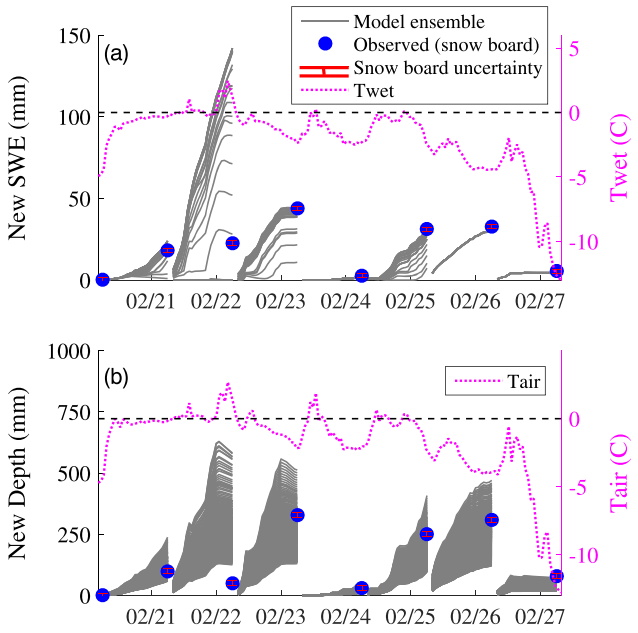

FIGURE 4 Example of model ensemble of (a) accumulated SWE, and (b) accumulated snow depth, compared to independent snow board measurements of 24-hr accumulated SWE and snow depth. Observed wet-bulb (T_{wet}) and air (T_{air}) temperatures are also shown because modeled partition depended on T_{wet} and new snow density on T_{air}

overburden was the dominant factor contributing to compaction of the old snow (discussed in section 6.2). Therefore, it follows that compaction errors will be largest for days when the accumulation of SWE for that day was biased. Thus, we only evaluate model compaction values during days when the simulated partitioning (modeled new SWE) values were within observational uncertainty.

Ignoring melt, the time evolution of snow depth can be given as

$$\frac{dh_s}{dt} = h_s^{new} - h_s^{compact} \tag{2}$$

where h_s(m) is the bulk snow depth, h_s^{new} (m s^{-1}) is the rate of new snowfall, and $h_s^{compact}$ (m s^{-1}) is the rate of snow compaction for old snow (Table 2). The snow compaction and the snowfall rate are estimated as

$$h_s^{compact} = h_s^{new} - \left[\frac{h_s(t+\Delta t) - h_s(t)}{\Delta t}\right]_{cont} \tag{3}$$

$$h_s^{new} = \left[\frac{h_s(t+\Delta t) - h_s(t)}{\Delta t}\right]_{sb} \tag{4}$$

where Δt (s) is the time interval for successive snow depth measurements (Δt is 24 hr in this study). For modeled compaction, the change in total snow depth is from the continuous simulation (*cont*) while the change in new snow depth is from the snow board simulation (*sb*). Finally, the two parameters that control compaction because of over burden, *G and H* as implemented in the Anderson (1976) function (see Equations 8 and 9 in Appendix S1), were varied in an ensemble of 81 simulations. Alternative compaction functions (e.g., Verseghy (1991)) were not included in this study.

4.5 | Diagnosing daily snow accumulation errors

From the process-based methods above, we kept track of the most likely source of model error during each day using the best-fit parameters provided in Table 1. Given the direction of error propagation from partitioning, to new snow density, into compaction (Figure 1), we were only able to isolate the error source for the subset of days when all preceding errors were within observational uncertainty (see Figure 1). For example, of the 564 snow accumulation days between 2004 and 2009, the simulated daily partition was within the observational uncertainty on 333 days. Thus, the modeled newly fallen snow density skill could only be assessed for these 333 days, with the remaining 231 assumed to be in error because of partitioning. Uncertainties in this error diagnostic approach are discussed in section 6.2.

4.6 | Case study of water year 2015

As an example of the application of the process-based error diagnostic method, we focused on water year 2015 because its winter average temperature anomaly (+2.1°C) was equal to projected winter temperature increases in the 2040s (Elsner et al., 2010; Klos et al., 2014; Mauger et al., 2015; Vano, 2015). During historically low snow years (e.g., *California* 2006–2015 drought, Kogan and Guo (2015)), the water content of snow has become more valuable per unit volume; thus, errors in model predictions of snow accumulation will have a larger impact on water resources. Characterizing errors in modeled new snow accumulation will help focus efforts to improve model robustness in the current and future rain-snow transitional climates.

5 | RESULTS

5.1 | Lumped method results

From the ensemble of 5342 model configurations, the "best" simulation of bulk snow depth accumulation was selected for water years 2004 to 2009 using the highest KGE metric (Equation 1). Selected parameter values are provided in Table 1. Because SUMMA was purposefully calibrated to the accumulation of bulk snow depth, the best model run had a small bias of 1% of observations and a KGE value of 0.78. The model skill from the calibration represented an upper bound for this site and these water years given the abilities of the new snow accumulation functions included here (Gong et al., 2011; Nearing & Gupta, 2014). Hidden model errors in bulk snow accumulation that had a compensatory impact throughout the study period were diagnosed using the process-based approach as reported below.

5.2 | Process-based method results

5.2.1 | Precipitation partitioning experiment

Simulated daily SWE accumulation (new SWE) was sensitive to the range of parameters because the SNQ winter wet-bulb temperature (T_{wet}) was frequently near 0°C during snowfall events. Modeled new SWE sensitivity is illustrated in Figure 5a, where all model configurations (*grey circles*) are compared to observations of new SWE for the calibration period. The parameter options with the lowest model error

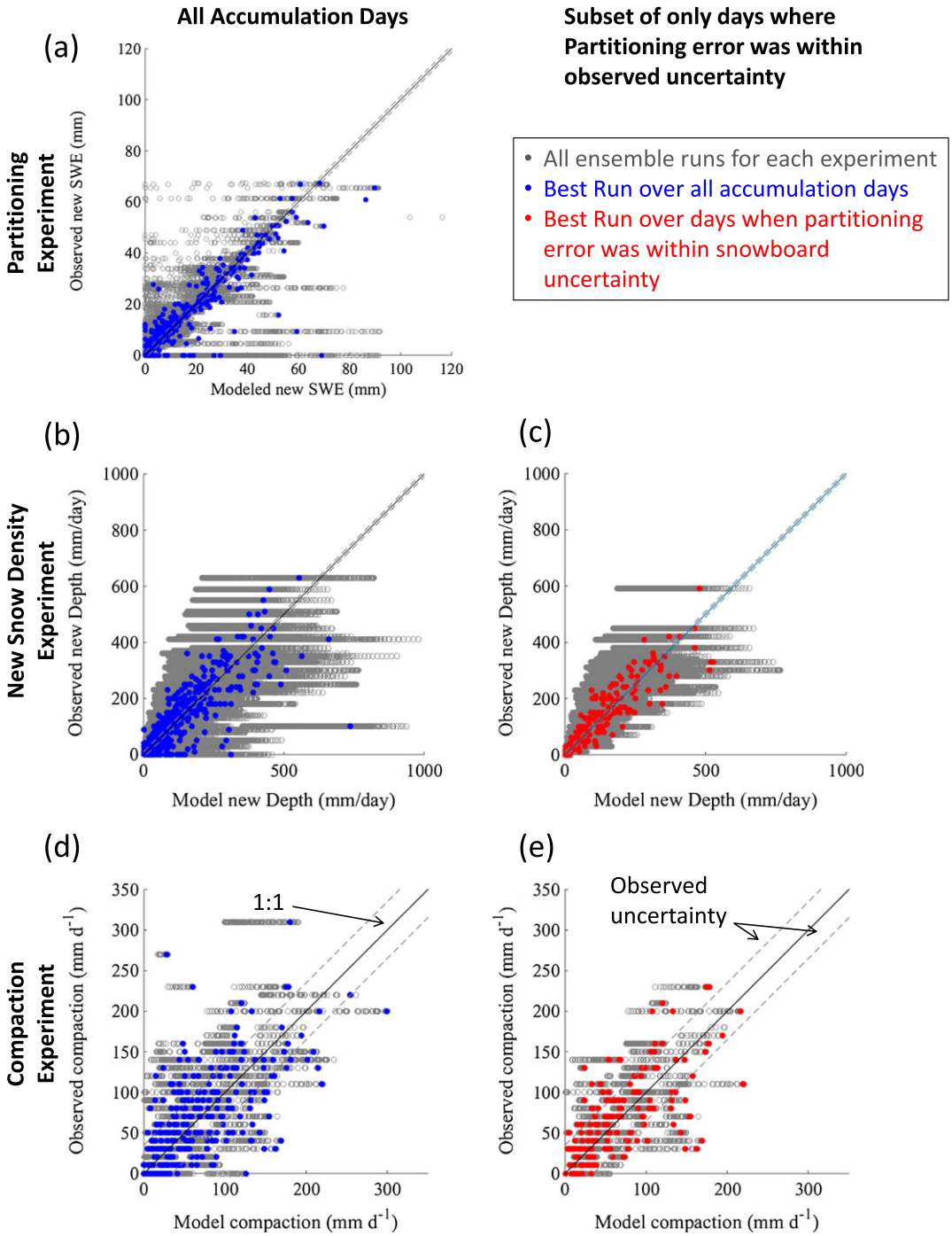

FIGURE 5 Modeled and observed daily accumulated new SWE (a), new snow depth (b,c), and compaction (d,e), shown for each day. *Gray circles* show all ensemble simulations. *Blue and red filled circles* show ensemble member selected from the highest Kling-Gupta efficiency (KGE) value

(*blue dots* in Figure 5a) had a T_{wet} critical value (P_c) of −0.24°C with a mixed-phase range (P_r) of 1.625°C (Table 1). Out of the 564 days with observed or modeled snowfall during the calibration period (water years 2004–2009), the model was within the uncertainty of the snow board SWE measurements (±1.3 mm) 333 days (59%), which was 64% of total accumulated SWE (Table 3). The remaining 231 (41%) days highlighted the difficulty of predicting precipitation phase at this maritime site using near-surface T_{wet}.

The selected partitioning function and the observed fraction of daily precipitation measured as ice are shown in Figure 6a. The daily fraction as ice was estimated from the ratio of daily accumulated SWE on the snow board to the daily accumulated precipitation (rain and snow). We note that retention of rainfall within recently accumulated snow is included in our estimated daily fraction as ice, which may positively bias daily fractions up to the maximum liquid water holding capacity of snowpack (~10% of bulk SWE) (Boone & Etchevers, 2001; Essery et al., 2013). Total precipitation amount is indicated by the relative circle size. Because wet-bulb temperature is most important here in its use for distinguishing precipitation phase, wet-bulb temperature in each hour was weighted by the fraction of daily total precipitation falling within that hour. Therefore, wet-bulb temperatures during the periods of heaviest precipitation are weighted the most.

TABLE 3 Number of days attributed to each source of modeled[a] error

	Variable evaluated (process represented)	# days snow fall	# days unknown because of upstream error or missing observations required[b]	# available days to evaluate process	# days model error *within* observed uncertainties (# days available) [% of available days]	# days model error *outside* observed uncertainties (# days available) [% of available days]
Calibration period (2004–2009)	New SWE (precipitation partitioning)	564	0	564	333 (564) [59%]	231 (564) [41%]
	New depth (snow density)	564	231	333	51 (333) [45%]	182 (333) [55%]
	Old compaction (overburden compaction)	564	231	333	266 (333) [80%]	67 (333) [20%]
	No error				144 (25%)	
Evaluation period (2010–2015)	New SWE (precipitation partitioning)	552	0	552	304 (552) [55%]	248 (552) [45%]
	New depth (snow density)	552	248	304	147 (304) [48%]	157 (304) [52%]
	Old compaction (overburden compaction)	552	248	304	256 (304) [84%]	48 (304) [16%]
	No Error				138 (25%)	

[a]Statistics shown for the process-based SUMMA configuration without cascading errors.

[b]For example, modeled snow density values were not evaluated for days where a precipitation partitioning error was observed.

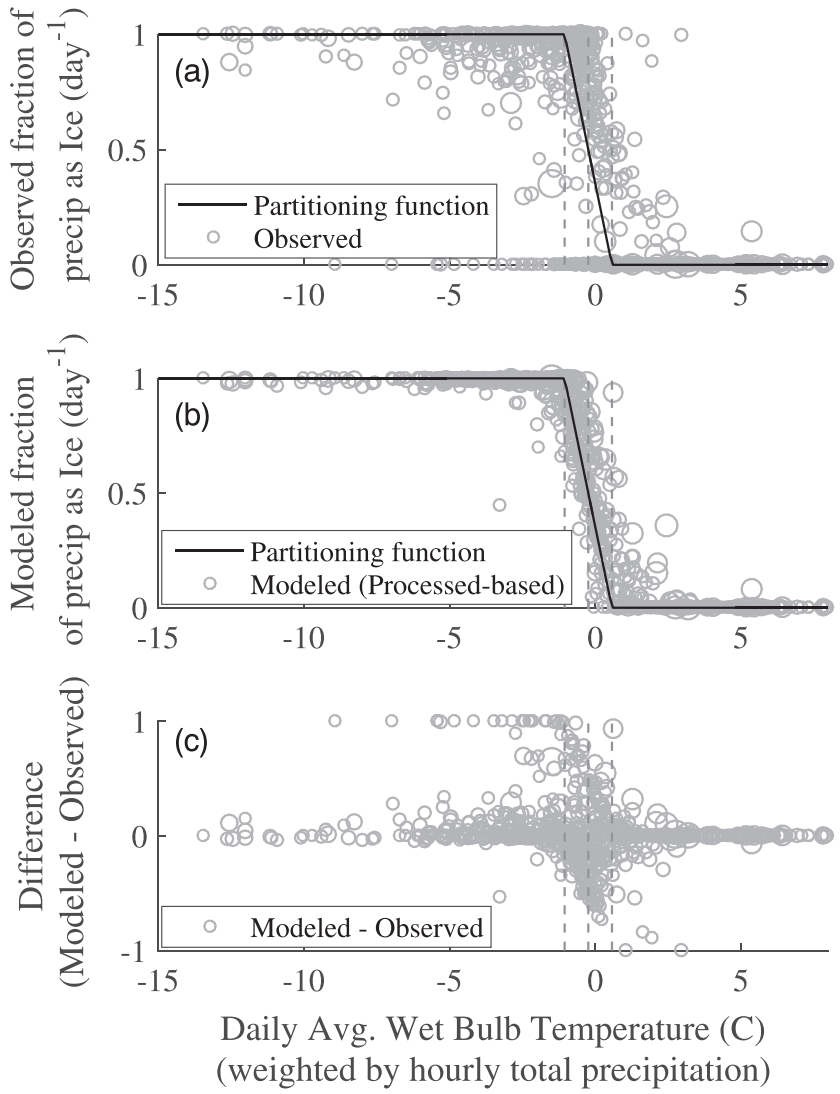

FIGURE 6 Daily fraction of precipitation (a) observed and (b) modeled as ice. Observed fraction was calculated from the ratio of snow board SWE to total gauge precipitation (corrected for undercatch). Modeled fraction was calculated from the ratio of simulated new SWE (restart simulations) to the observed total gauge precipitation. Total precipitation amount is indicated by the relative circle size. Because wet-bulb temperature is most important here in its use for distinguishing precipitation phase, wet-bulb temperature in each hour was weighted by the fraction of daily total precipitation falling within that hour. Therefore, wet-bulb temperatures during the periods of heaviest precipitation are weighted the most

The selected partitioning function does not appear to match the daily observations well. However, this discrepancy is partly explained by the difference in time steps between the observations (daily) and model (hourly). The daily fraction of ice derived from the modeled snow board SWE using the above partitioning function is shown in Figure 6b, which more closely matches the daily observations. Using this more direct comparison (daily fraction to daily fraction), we found the largest model errors in accumulated SWE occurred near a T_{wet} of 0°C (Figure 6c) but were unbiased as a result of calibration (Figure 5a).

5.2.2 | New snow density experiment

The above-identified errors in partitioning directly impacted simulations of bulk snow depth and the calibration of best new snow density function. Evaluation of the ensemble of new snow depth simulations using all days (N = 564) and only days where partitioning errors were within observed uncertainty (N = 333) are shown in Figure 5b, c, respectively. The exclusion of partitioning errors reduced the largest new snow depth errors (reduced scatter of grey dots in Figure 5c). The ensemble member with the lowest error (highest KGE value) in daily-accumulated snow depth was different for the two sets of days. Excluding partitioning error days resulted in the constant density method with a value of 78 kg m^{-3} being chosen, while including all days selected the Pahaut (1976) method (Table 1).

In general, the choice of parameter values was more important than the choice of new snow density function, as illustrated by Figure 7a. All new snow density functions could have a KGE value greater than 0.8 (*red* dots) so long as the correct parameter values were selected for this location. However, if the optimal parameter values were unknown and selected at random, the Hedstrom and Pomeroy (1998) and Pahaut (1976) functions would have a higher likelihood of a higher KGE value. If default parameter values from literature (Table 1) were used, the Hedstrom and Pomeroy (1998) function had the highest KGE value (blue square). The resulting functions with optimal parameter values compared to observations are shown in Figure 7b, which further illustrates the similarity in moderate skill between all competing hypotheses. The large variability of measured density for a given average temperature was a combination of variations in falling density, compaction, and melting, which are further discussed in section 6.2.

5.2.3 | Compaction experiment

The modeled compaction of the existing snowpack exhibited large sensitivity to the range of parameter values used here (Figure 5d, e). Excluding days where the partitioning error was outside observational uncertainty (Figure 5e) removed many of the largest errors in modeled snow compaction. For the remaining subset of days where partitioning was correct (333 days), modeled compaction was in error 67 days (Table 3).

5.3 | Bulk snow depth

Simulations of bulk snow depth using the three model configurations in Table 1 are compared to observations for the calibration period (Figure 8a–d) and evaluation period (Figure 8e–h). All simulations produced generally high quality simulations of the total bulk snow depth

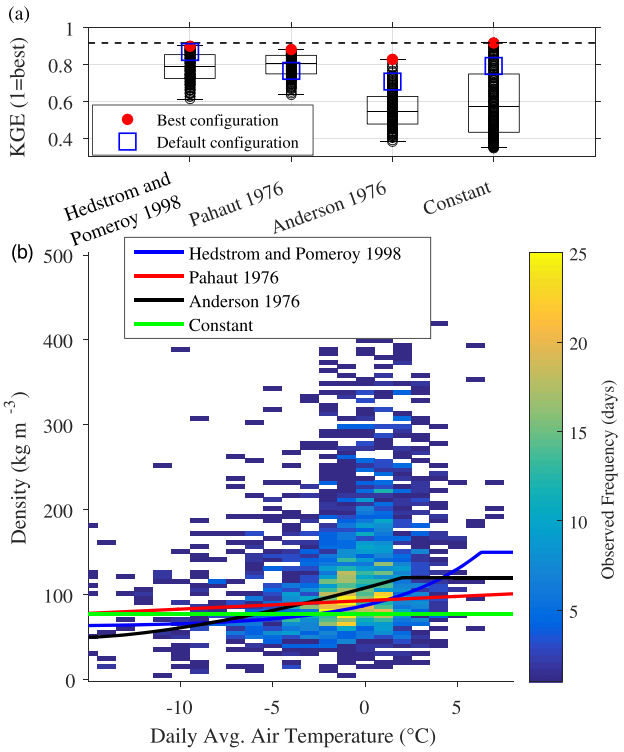

FIGURE 7 (a) Calibration results for each new snow density function, showing near-equal performance given the right parameter values. A KGE value of unity is best. (b) Observed newly fallen snow density compared to the daily average air temperature during water years 1980–2015. Overlaid lines show each function in (a) with the highest KGE value (*red circles*). Blue squares in (a) show KGE values with default parameter values taken from literature (Table 1)

for most water years, despite having different model configurations (Table 1). A notable result was that both process-based configurations had similar skill in simulating the bulk snow depth as the lumped configuration, which represented the upper bound in this study. However, intermodel simulation differences were found in simulated new SWE (Figure 8b,f), new snow depth (Figure 8c,g), and old snow compaction (Figure 8d,h), which had compensatory effects on bulk snow depth accumulation.

5.4 | Diagnosis of bulk snow accumulation errors

For both the calibration and evaluation periods, errors because of partitioning of precipitation resulted in the largest absolute number of days compared to other processes (Figure 9). However, the processes of new snow density and compaction were only evaluated for days where partitioning errors were small (to prevent cascading errors). Therefore, as a percentage of all available days, each process was evaluated (*outlined boxes* Figure 9), new snow density errors were most common (55% calibration, 52% evaluation), followed by partitioning errors (41% calibration, 45% evaluation), and compaction of old snow errors during accumulation days (20% calibration, 16% evaluation) (Table 3). Only 25% of days were identified as having none of the above errors, which meant that during the majority of snow accumulation days (75%) at least one process representation failed, most commonly new snow density and precipitation partitioning. However, the individual daily errors identified canceled out over annual time scales

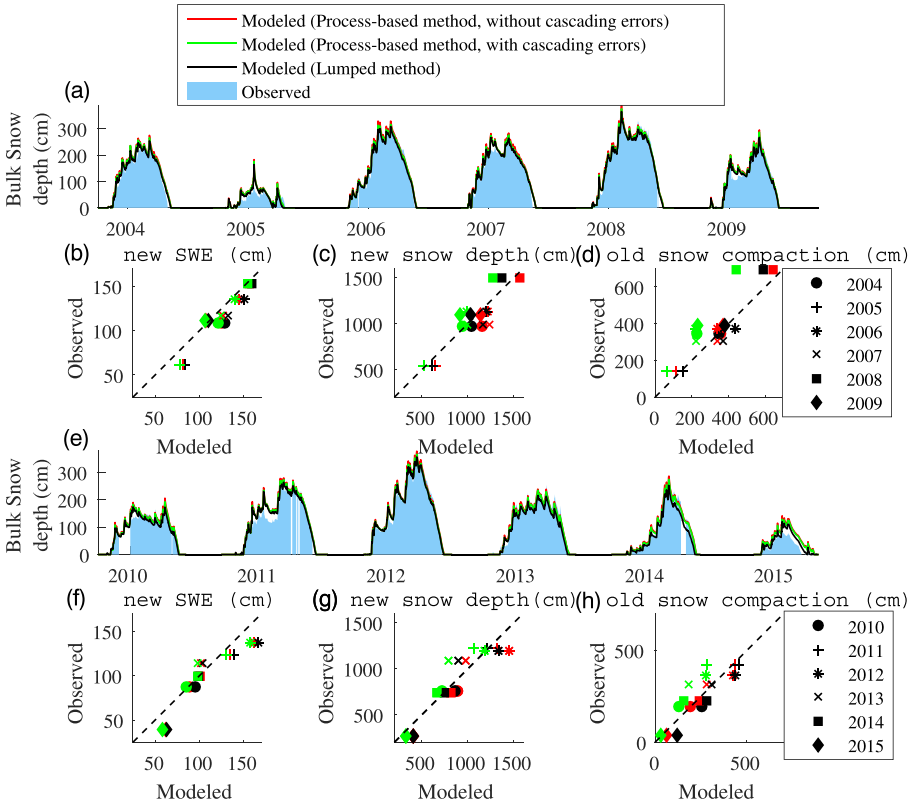

FIGURE 8 Modeled and observed bulk snow depth during the calibration period (a) (water years 2004–2009) and (e) evaluation period (2010–2015). Vertical white periods represent missing snow stake measurements. Scatter plots of modeled and observed *total accumulated* new SWE (b,f), new snow (c,g), and compaction of old snow during accumulation days (d,h) for each water year. Symbol type indicates each water year, while color refers to model simulations defined in legend (a)

during most years, resulting in high skill in simulating bulk snow depth (Figure 8). The exception to this finding was water year 2015, which had an anomalously warm winter (December through February) temperatures 2°C higher than the 1980 to 2014 average.

Figure 10 illustrates the identified source of bulk snow depth accumulation errors during water year 2015 for each day. While on average, about 12 storms build the annual snowpack at Snoqualmie, observations show that the 2015 seasonal snowpack was built with only seven storms. The dominant source of new snow accumulation error during the seven major storms of water year 2015 was incorrect partitioning, based on the process-based evaluation. Four available manual full snowpit measurements of bulk SWE (Wayand et al., 2015b) corroborate that the model accumulated too much mass, in contrast to a density error (not shown). Interestingly, water year 2015 had precipitation partitioning errors as often as other years (Figure 9), but those errors mattered more toward bulk snow accumulation because a larger fraction of winter precipitation occurred near the freezing point, where partitioning parameterizations were most uncertain (Figure 6c). This example illustrates the utility of the process-based method, which is not to select the best-fit model, but that when the best-fit model fails, to identify the process(es) responsible for that failure (e.g., precipitation partitioning in water year 2015). In addition, the process-based method identified the need for new partitioning methods that will not fail during warm winters, which may become more common in the future for the Washington Cascades (Elsner et al., 2010; Klos et al., 2014; Mauger et al., 2015; Vano, 2015).

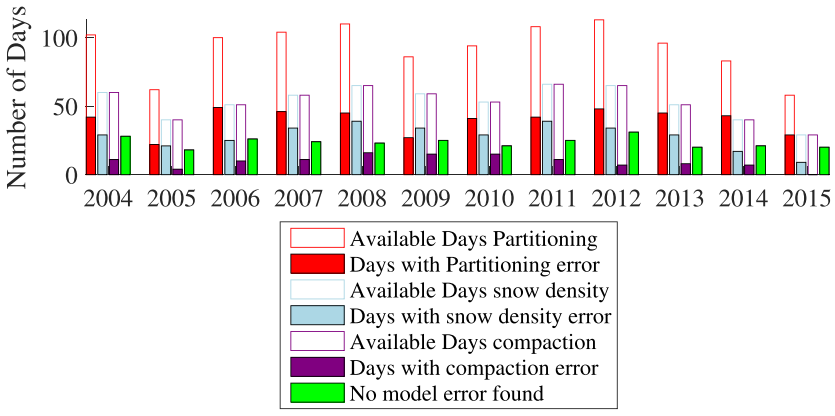

FIGURE 9 Frequency of each source of model daily error (*solid bars*) out of available days where cascading errors were removed (*outline bars*), for each water year

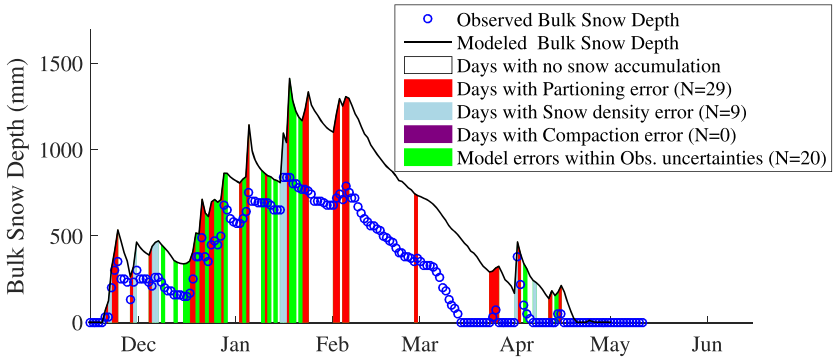

FIGURE 10 Time series of observed and modeled total snow depth for water year 2015. Colors for each day show the identification of the dominate source of model error using the process-based model configuration (without cascading errors). *White* areas show non-accumulation periods that not included in analysis

6 | DISCUSSION

6.1 | Robustness of simulations of bulk snow accumulation

The lumped method identified a model configuration that resulted in very low error simulating bulk snow depth during both calibration and evaluation periods (Figure 8). However, the process-based approach presented here revealed that on a daily time scale, one of the three snow accumulation processes was in error 75% of the time, indicating the models did not get the right answer (i.e., bulk snow depth accumulation) for the right reasons (Kirchner, 2006). This level of skill may be acceptable for some applications (i.e., predicting peak basin-averaged SWE for seasonal streamflow forecasts) but will not be acceptable for others (i.e., microwave remote sensing, assessing climate change sensitivity, forecasting of rain-on-snow floods, or backing out SWE from observed snow depth (Egli et al., 2009; Sturm et al., 2010)).

Likewise, the most important source of model error (where to focus model development) depends on the application. For example, simulating the annual SWE accumulation requires a precipitation partitioning method that will not fail during anomalous years (e.g., water year 2015 at SNQ). On an individual storm basis, we found that the most frequent source of error in total snow depth was modeled new snow density (52%) or precipitation partitioning (45%). Thus, methods that derive SWE from observations of snow depth (e.g., airborne LIDAR) and modeled bulk snow density for a single storm would be sensitive to the frequent errors in new snow density found here. Fortunately, low snow density errors are compensated by higher compaction rates with time, and high snow density errors are compensated by lower compaction rates with time; thus bulk density errors should decrease with time following a snow event (not shown). Finally, the process-based method used here provides a way to quantity uncertainty of individual modeled processes, which would be useful within a data assimilation framework (Reichle, 2008).

6.2 | Uncertainty in diagnosis of daily model error source

Processes of precipitation phase, density of newly fallen snow, and compaction of existing snowpack were measured indirectly with available daily observations at the SNQ site (i.e., snow board measurements). An assessment of the degree to which the daily time scale impacted the diagnostic of sources of error was done to ascertain the uncertainty of this study's results.

The measured snow board SWE is a function of the accumulated snowfall and retained rainfall minus any drained snowmelt over the prior 24 hr. Retention of rainfall in newly accumulated snow (i.e., during the passage of warm fronts when snow changes to rain) should be less than ~10% by mass (as previously discussed in section 5.2.1), which suggests that our daily partitioning fractions calculated in this study (Figure 6) could be biased high by 10% during these events. Quantifying the additional error introduced by partial melting of the accumulated snow board SWE is more difficult. Although snowmelt lysimeter observations were available, they are (a) impacted by hydrological processes taking place throughout the bulk snow back, and (b) did not always provide accurate quantitative estimates of snowmelt (Wayand et al., 2015b). The overall accuracy of the daily phase from the snow board was assessed using 10-s phase observations from a laser disdrometer, which was only available for water year 2015 (Wayand, et al., 2016c). The daily precipitation fraction as ice derived from the snow board was found to have a bias ratio of 0.65 and a correlation to the disdrometer of 0.6 (Figure S7).

Besides the density of newly fallen snow, additional processes that impacted the measured snow density after 24 hr include the compaction of accumulated snow because of overburden and metamorphism. Thus, some of the scatter in Figure 7 is likely because of a combination of these processes and variability in the newly fallen snow density (Wayand et al., 2015b). Very high snow density measurements may also have been because of hail and graupel, which occur frequently during the passage of vigorous frontal systems. Without sub-daily observations of newly fallen snow density, it is difficult to separate out these impacts and quantify their effect on the calculated density values shown in Figure 7.

Compaction of the underlying ("old") snowpack was calculated here based on measurements of the change in total bulk snow depth (from a snow stake) and accumulation of new snow depth (from a snow board). Although the compaction of *new snow* could not be isolated given available observations, this did not impact our calculation of old snow compaction because compacted new snow would equally reduce snow depths measured via the snow board and the snow stake (see Figure 2).

The assumed control of overburden on compaction rates was supported by the fact that observed bulk compaction showed a high correlation (R^2 = 0.6) with the amount of accumulated new SWE. Further, temperature-driven metamorphism and resulting compaction

were likely small given small internal snowpack temperature gradients in a maritime climate during snow accumulation. However, the impact of liquid water was likely significant, especially during frequent warm frontal systems when snow switched to rain within a few hours. We were unable to isolate this process given available observations.

6.3 | Applicability of process-based method

The process-based approach presented here used multiple snow observations to constrain model choices, as has been previously advocated (e.g., Essery et al., 2013; Magnusson & Wever, 2015). Our methodology differed from previous approaches in that we focused on removing errors (e.g., biased meteorological forcing, preceding processes, etc.) from cascading into the evaluation given modeled processes. Isolating process evaluation from preceding errors is critical for understanding whether or not models get the right answer for the right reasons (Kirchner, 2006).

The methodology used here to identify process representation errors and limit propagation of model errors can be generalized to other areas of snow hydrology. The requirements for such application are observations of the boundary conditions (here meteorological forcing data) and internal snow pack observations (here daily new snow accumulation). A lack of the required observations is likely the greatest limitation to applying this method to other processes and at other locations, which makes efforts to connect and share available mountainous data sets (Pomeroy et al., 2015) critical for model development.

7 | SUMMARY

This study presented a novel process-based approach to diagnose model errors through an example that focused on snow accumulation processes. For the Snoqualmie Pass study site, we found that errors in new snow density and precipitating partitioning occurred during 52% and 45% of available snow accumulation days, respectively, while compaction errors were less frequent (16%). Therefore, for applications where the change in total snow depth during a snow storm matters (e.g., estimating storm-specific snowfall mass from LIDAR derived snow depth), the model parameterizations of new snow density and precipitation partitioning need to be improved. In addition, we found that the choices of new snow density function and parameter values were impacted by preceding precipitation partitioning errors, suggesting caution must be used when evaluating future new snow density parameterizations. Finally, the main benefit of the process-based method is not that it results in a better fit than traditional calibration methods (Figure 8), but that when the model fails, we can identify the process responsible for that failure (i.e., partitioning in water year 2015, Figure 10), which is critical for model development and advancement.

ACKNOWLEDGMENTS

All meteorological raw, quality controlled, and filled data are available at http://dx.doi.org/10.6069/H57P8W91. SUMMA model code is available at https://github.com/NCAR/summa, and scripts to create SUMMA configuration files used in this study are available at https://github.com/NicWayand/summa_scripts. This work was supported by the National Science Foundation (EAR-1215771). We would like to thank two anonymous reviewers for their helpful comments.

REFERENCES

Anderson, E. A. (1976). A point energy and mass balance model of a snow cover. Silver Spring, Md.: Office of Hydrology, National Weather Service. Retrieved from http://www.csa.com/partners/viewrecord.php?requester=gs&collection=ENV&recid=7611864\nhttp://www.agu.org/pubs/crossref/2009/2009JD011949.shtml

Auer, A. H. (1974). The rain versus snow threshold temperatures. *Weatherwise*, *27*, 67.

Barbara, L., George, X., Buell, T., Moore, D., Austin, B., Wang, Y.-J. (2008). Storm-related closures of I-5 and I-90: Freight transportation economic impact assessment report. Retrieved from http://www.wsdot.wa.gov/research/reports/fullreports/708.1.pdf

Boone, A. (2002). Description du schema de neige ISBA-ES (Explicit Snow). Centre National de Recherches Retrieved from http://www.cnrm-game-meteo.fr/IMG/pdf/snowdoc_v2.pdf

Boone, A., & Etchevers, P. (2001). An intercomparison of three snow schemes of varying complexity coupled to the same land surface model: Local-scale evaluation at an Alpine site. *Journal of Hydrometeorology*, class 2374–394 .Retrieved from http://journals.ametsoc.org/doi/abs/10.1175/1525-7541(2001)002%3C0374:AIOTSS%3E2.0.CO%3B2

Buck, A. L. (1981). New equations for computing vapor pressure and enhancement factor. *Journal of Applied Meteorology*, *20*(12), 1527–1531. doi:10.1175/1520-0450(1981)020<1527:NEFCVP>2.0.CO;2

Campbell, G. S., & Norman, J. M. (1998). *An introduction to environmental biophysics*. New York: Springer-Verlag.

Clark, M. P., & Kavetski, D. (2010). Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes. *Water Resources Research*, *46*(10). doi:10.1029/2009WR008894

Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, *47*(9). doi:10.1029/2010WR009827

Clark, M. P., Lundquist, J. D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., … Marks, D. (2015a). The structure for unifying multiple modeling alternatives (SUMMA), version 1: Technical Description, NCAR Technical Note NCAR/TN-514 + STR, 54 pp. National Center for Atmospheric Research, Boulder, Colo., doi:10.5065/D6WQ01TD

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., … Marks, D. (2015b). A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies. *Water Resources Research*, *51*, 2515–2542. doi:10.1002/2015WR017200

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., … Rasmussen, R. (2015c). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*, *51*(4), 1–17. doi:10.1002/2015WR017200.A

Cosgrove, B. (2003). Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research*, *108*(D22), 8842. doi:10.1029/2002JD003118.

Egli, L., Jonas, T., Meister, R. (2009). Comparison of different automatic methods for estimating snow water equivalent. *Cold Regions Science and Technology* *57*(2-3), 107–115. doi:10.1016/j.coldregions.2009.02.008

Elsner, M. M., Cuo, L., Voisin, N., Deems, J. S., Hamlet, A. F., Vano, J. A., … Lettenmaier, D. P. (2010). Implications of 21st century climate change for the hydrology of Washington State. *Climatic Change*, *102*(1–2), 225–260. doi:10.1007/s10584-010-9855-0

Essery, R. (2015). A factorial snowpack model (FSM 1.0). *Geoscientific Model Development*, *8*(12), 3867–3876. doi:10.5194/gmd-8-3867-2015.

Essery, R., & Etchevers, P. (2004). Parameter sensitivity in simulations of snowmelt. *Journal of Geophysical Research, D: Atmospheres*, *109*(20), 1–15. doi:10.1029/2004JD005036

Essery, R., Rutter, N., Pomeroy, J., Baxter, R., Stahli, M., Gustafsson, D., … Elder, K. (2009). SNOWMIP2: An evaluation of forest snow process simulations. *Bulletin of the American Meteorological Society*, *90*(8), 1120–1135. doi:10.1175/2009BAMS2629.1

Essery, R., Morin, S., Lejeune, Y., & Ménard, C. (2013). A comparison of 1701 snow models using observations from an alpine site. *Advances in Water Resources*, *55*, 131–148. doi:10.1016/j.advwatres.2012.07.013

Feiccabrino, J., Graff, W., Lundberg, A., Sandström, N., & Gustafsson, D. (2015). Meteorological knowledge useful for the improvement of snow rain separation in surface based models. *Hydrology*, *2*(4), 266–288. doi:10.3390/hydrology2040266

Gong, Y. W., Shen, Z. Y., Hong, Q., Liu, R. M., & Liao, Q. (2011). Parameter uncertainty analysis in watershed total phosphorus modeling using the GLUE methodology. *Agriculture, Ecosystems & Environment*, *142*, 246–255.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. doi:10.1016/j.jhydrol.2009.08.003

Gupta, H. V., Clark, M. P., Vrugt J, A., Abramowitz, G., & M., Y. (2012). Towards a comprehensive assessment of model structural adequacy. *Water Resources Research*, *48*(8), 1–16. doi:10.1029/2011WR011044

Harder, P., & Pomeroy, J. (2013). Estimating precipitation phase using a psychrometric energy balance method. *Hydrological Processes*, *27*(13), 1901–1914. doi:10.1002/hyp.9799

Hedstrom, N. R., & Pomeroy, J. W. (1998). Measurements and modelling of snow interception in the boreal forest. *Hydrological Processes*, *1625* (March), 1611–1625.

Kavetski, D., & Clark, M. P. (2011). Numerical troubles in conceptual hydrology: Approximations, absurdities and impact on hypothesis testing. *Hydrological Processes*, *25*(4), 661–670. doi:10.1002/hyp.7899

Kinar, N. J., & Pomeroy, J. W. (2015). SAS2: The system for acoustic sensing of snow. *Hydrological Processes DOI*. doi:10.1002/hyp.10535

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, *42*(3). doi:10.1029/2005WR004362

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, *424-425*, 264–277. doi:10.1016/j.jhydrol.2012.01.011

Klos, P., Link, T., & Abatzoglou, J. (2014). Extent of the rain–snow transition zone in the western US under historic and projected climate. *Geophysical Research*, 4560–4568. doi:10.1002/2014GL060500

Kogan, F., & Guo, W. (2015). 2006–2015 mega-drought in the western USA and its monitoring from space data. *Geomatics Natural Hazards and Risk*, *6*(8), 651–668. doi:10.1080/19475705.2015.1079265

Landry, C., Buck, K., Raleigh, M., & Clark, M. (2014). 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 Resources*, 1773–1788. doi:10.1002/2013WR013711

Langlois, A., Royer, A., Derksen, C., Montpetit, B., Dupont, F., & Goita, K. (2012). Coupling of the snow thermodynamic model SNOWPACK with the Microwave Emission Model for Layered Snowpacks (MEMLS) for subarctic and arctic Snow Water Equivalent retrievals. *Water Resources Research*, *48*, .W12524, 14pp

Lundquist, J. D., Neiman, P. J., Martner, B., White, A. B., Gottas, D. J., & Ralph, F. M. (2008). Rain versus snow in the Sierra Nevada, California: Comparing Doppler profiling radar and surface observations of melting level. *Journal of Hydrometeorology*, *9*(2), 194–211. doi:10.1175/2007JHM853.1

Magnusson, J., & Wever, N. (2015). Evaluating snow models with varying process representations for hydrological applications. *Water Resources*, 1–17 .Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/2014WR016498/full

Marks, D., Winstral, A., Reba, M., Pomeroy, J., & Kumar, M. (2013). An evaluation of methods for determining during-storm precipitation phase and the rain/snow transition elevation at the surface in a mountain basin. *Advances in Water Resources*, *55*, 98–110.

Mauger, G. S., Casola, J. H., Morgan, H. A., Strauch, R. L., Jones, B., Curry, B., … Snover, A. K. (2015). State of knowledge: Climate change in Puget Sound. Report prepared for the Puget Sound Partnership and the National Oceanic and Atmospheric Administration. Climate Impacts Group, University of Washington, Seattle. doi:10.7915/CIG93777D

Morin, S., Lejeune, Y., Lesaffre, B., Panel, J.-M., Poncet, D., David, P., & Sudul, M. (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 System Science Data*, *4*(1), 13–21. doi:10.5194/essd-4-13-2012

Nearing, G. S., & Gupta, H. V. (2014). The quantity and quality of information in hydrologic models. *Water Resources Research*, *51*, 524–538. doi:10.1002/2013WR014956

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., … Arnold, J. R. (2015). Gridded ensemble precipitation and temperature estimates for the contiguous United States. *Journal of Hydrometeorology*, *16*(6), 2481–2500. doi:10.1175/JHM-D-15-0026.1

Nitu, R., Rasmussen, R., Baker, B., Lanzinger, E., Joe, P., Yang, D., Smith, C., Roulet, Y. A., Goodison, B., Liang, H., et al. (2012). WMO intercomparison of instruments and methods for the measurement of solid precipitation and snow on the ground: Organization of the experiment. Preprints, TECO-2012: WMO Technical Conf. on Meteorological and Environmental Instruments and Methods of Observations, Brussels, Belgium,WMO, 10 pp. Retrieved from https://www.wmo.int/pages/prog/www/IMOP/publications/IOM-109_TECO-2012/ Session1/O1_01_Nitu_SPICE.pdf

Oleson, K., Lawrence, D., & Gordon, B. (2010). Technical description of version 4.0 of the Community Land Model (CLM). Boulder. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.172.7769

Pahaut. (1976). La métamorphose des cristaux de neige (Snow crystal metamorphosis).

Pfister, R., & Schneebeli, M. (1999). Snow accumulation on boards of different sizes and shapes. *Hydrological Processes*, *2355*(May 1998), –2355. doi:10.1002/(SICI)1099-1085(199910)13:14/15<2345::AID-HYP873>3.0.CO;2-N/full

Pomeroy, J., Bernhardt, M., & Marks, D. (2015). Research network to track alpine water. *Nature*, *521*, 32.

Pomeroy, J. W., Gray, D. M., Brown, T., Hedstrom, N. R., Quinton, W. L., Granger, R. J., Carey, S. K. (2007). The cold regions hydrological process representation and model: A platform for basing model structure on physical evidence. *Hydrological Processes 21*(19), 2650–2667. doi:10.1002/Hyp.6787

Raleigh, M. S., Lundquist, J. D., & Clark, M. P. (2014). Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. *Hydrology and Earth System Sciences Discussions*, *11*(12), 13745–13795. doi:10.5194/hessd-11-13745-2014

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., et al. (2012). How well are we measuring snow: The NOAA/FAA/NCAR Winter Precipitation Test Bed. *Bulletin of the American Meteorological Society*, *93*(6), 811–829. doi:10.1175/BAMS-D-11-00052.1

Reichle, R. H. (2008). Data assimilation methods in the Earth sciences. *Advances in Water Resources*, *31*(11), 1411–1418. doi:10.1016/j.advwatres.2008.01.001

Roebber, P. J., Bruening, S. L., Schultz, D. M., & Cortinas, J. V. (2003). Improving snowfall forecasting by diagnosing snow density. *Weather and Forecasting*, *18*, 264–287. doi:10.1175/1520-0434(2003)018<0264:ISFBDS>2.0.CO;2

Rössler, O., Froidevaux, P., Börst, U., Rickli, R., Martius, O., & Weingartner, R. (2014). Retrospective analysis of a nonforecasted rain-on-snow flood in the Alps—A matter of model limitations or unpredictable nature? *Hydrology and Earth System Sciences*, *18*(6), 2265–2285. doi:10.5194/hess-18-2265-2014

Schmid, L., Heilig, A., Mitterer, C., Schweizer, J., Maurer, H., Okorn, R., & Eisen, O. (2014). Continuous snowpack monitoring using upward-looking ground-penetrating radar technology. *Journal of Glaciology*, *60* (221), 509–525. doi:10.3189/2014JoG13J084

Sevruk, B. (1983). Correction of measured precipitation in the Alps using the water equivalent of new snow. *Weather*, *14*(2), 49–58 .Retrieved from http://www.iwaponline.com/nh/014/0049/0140049.pdf

Sims, E. M., & Liu, G. (2015). A parameterization of the probability of snow–rain transition. *Journal of Hydrometeorology*, *16*(4), 1466–1477. doi:10.1175/JHM-D-14-0211.1

Sturm, M., Taras, B., Liston, G. E., Derksen, C., Jonas, T., & Lea, J. (2010). Estimating snow water equivalent using snow depth data and climate classes. *Journal of Hydrometeorology*, *11*(6), 1380–1394. doi:10.1175/2010JHM1202.1

USACE. (1956). Snow hydrology summary report of the snow investigations. North Pacific Division Portland.

Vano, J. (2015). Seasonal hydrologic responses to climate change in the Pacific Northwest. *Water Resources*, *51*, 1959–1976. doi:10.1002/2014WR015909

Wayand, N. E., Lundquist, J. D., & Clark, M. P. (2015a). Modeling the influence of hypsometry, vegetation, and storm energy on snowmelt contributions to basins during rain-on-snow floods. *Water Resources Research*. doi:10.1002/2014WR016576

Wayand, N. E., Massmann, A., Butler, C., Keenan, E., & Lundquist, J. D. (2015b). A meteorological and snow observational data set from Snoqualmie Pass (921 m), Washington Cascades, U.S. *Water Resources Research*. doi:10.1002/2015WR017773

Wayand, N. E., Clark, M. P., & Lundquist, J. D. (2016c). Diagnosing snow accumulation errors in a rain-snow transitional environment with snow board observations. *Hydrological Processes*. doi:10.1002/hyp.11002

Wever, N., Schmid, L., Heilig, A., Eisen, O., Fierz, C., Lehning, M. (2015). Verification of the multi-layer SNOWPACK model with different water transport schemes. *The Cryosphere Discussions 9*(2), 2655–2707. doi:10.5194/tcd-9-2655-2015

White, A. B., Gottas, D. J., Henkel, A. F., Neiman, P. J., Ralph, F. M., & Gutman, S. I. (2010). Developing a performance measure for snow-level forecasts. *Journal of Hydrometeorology*, *11*(3), 739–753. doi:10.1175/2009JHM1181.1

Wood, A. W., Hopson, T., Newman, A., Brekke, L., Arnold, J., & Clark, M. (2015). Quantifying streamflow forecast skill elasticity to initial condition and climate prediction skill. *Journal of Hydrometeorology*, (OCTOBER): 151030135123004. doi:10.1175/JHM-D-14-0213.1

Yang, D., & Goodison, B. (1998). Accuracy of NWS 8″ standard nonrecording precipitation gauge: Results and application of WMO intercomparison. *Journal of Atmospheric and Oceanic Technology*, 54–68. doi:10.1175/1520-0426(1998)015<0054:AONSNP>2.0.CO;2

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Wayand NE, Clark MP, and Lundquist JD. Diagnosing snow accumulation errors in a rain-snow transitional environment with snow board observations. *Hydrological Processes*. 2017;31:349–363. doi: 10.1002/hyp.11002.