Exploration of synthetic terrestrial snow mass estimation via assimilation of AMSR-E brightness temperature spectral differences using the Catchment land surface model and support vector machine regression

JING Wang[1], Barton A Forman[1], and Yuan Xue[2]

[1]University of Maryland, College Park
[2]George Mason University

November 21, 2022

Abstract

This study explores improvements in the estimation of snow water equivalent (SWE) over snow-covered terrain using an ensemble-based data assimilation (DA) framework. The NASA Catchment land surface model is used as the prognostic model in the assimilation of AMSR-E passive microwave (PMW) brightness temperature spectral differences ($\Delta$$T_b$) where support vector machine (SVM) regression is employed as the observation operator. A series of synthetic twin experiments are conducted using different precipitation boundary conditions. The results show, at times, DA degrades modeled SWE estimates (compared to the land surface model without assimilation) over complex terrain. To mitigate this degradation, a physically-constrained approach using different $\Delta$$T_b$ for shallow-to-medium or medium-to-deep snow conditions along with a â\euroœdata-thinningâ\euro strategy are explored. Overall, both strategies improve the model ability to encapsulate more of the evaluation data and mitigate model ensemble collapse. The physically-constrained DA and 3-day thinning DA strategies show marginal improvements of basin-averaged SWE in terms of reduction of bias from 10 mm (baseline DA) to -5.2 mm and $-$2.5 mm, respectively. When the estimated forcings are greater than the truth, the baseline DA, physically-constrained DA, and 3-day thinning DA improve SWE the most with approximately 30\%, 31\%, and 24\% reduction of RMSE (relative to OL), respectively. Overall, these results highlight the limited utility of PMW $\Delta$$T_b$ observations in the estimation of snow in complex terrain, but do demonstrate that a physically-based constraint approach and data thinning strategy can add more utility to the $\Delta$$T_b$ observations in the estimation of SWE.

Exploration of synthetic terrestrial snow mass estimation via assimilation of AMSR-E brightness temperature spectral differences using the Catchment land surface model and support vector machine regression

Jing Wang[1], Barton A. Forman[1], and Yuan Xue[2]

[1]Department of Civil and Environmental Engineering, University of Maryland, College Park, MD, USA.
[2]Department of Geography and GeoInformation Science, George Mason University, VA, USA.

Key Points:

- snow estimation
- physically-constrained assimilation
- data thinning

Corresponding author: Jing Wang, `jwang1@umd.edu`

Abstract

This study explores improvements in the estimation of snow water equivalent (SWE) over snow-covered terrain using an ensemble-based data assimilation (DA) framework. The NASA Catchment land surface model is used as the prognostic model in the assimilation of AMSR-E passive microwave (PMW) brightness temperature spectral differences (ΔT_b) where support vector machine (SVM) regression is employed as the observation operator. A series of synthetic twin experiments are conducted using different precipitation boundary conditions. The results show, at times, DA degrades modeled SWE estimates (compared to the land surface model without assimilation) over complex terrain. To mitigate this degradation, a physically-constrained approach using different ΔT_b for shallow-to-medium or medium-to-deep snow conditions along with a "data-thinning" strategy are explored. Overall, both strategies improve the model ability to encapsulate more of the evaluation data and mitigate model ensemble collapse. The physically-constrained DA and 3-day thinning DA strategies show marginal improvements of basin-averaged SWE in terms of reduction of bias from 10 mm (baseline DA) to -5.2 mm and -2.5 mm, respectively. When the estimated forcings are greater than the truth, the baseline DA, physically-constrained DA, and 3-day thinning DA improve SWE the most with approximately 30%, 31%, and 24% reduction of RMSE (relative to OL), respectively. Overall, these results highlight the limited utility of PMW ΔT_b observations in the estimation of snow in complex terrain, but do demonstrate that a physically-based constraint approach and data thinning strategy can add more utility to the ΔT_b observations in the estimation of SWE.

1 Introduction and Background

Snow is a significant contributor to the Earth's hydrologic cycle (Liston, 1999), energy cycle (Fernandes et al., 2009), and climate system (Curry et al., 1995; Barnett et al., 2005). It accounts for a large fraction of the available freshwater resources in many parts of the northern hemisphere (Barnett et al., 2005). However, direct quantification of snow mass, or snow water equivalent (SWE), across time and space using point-scale, ground-based techniques remains challenging due to the spatial and temporal variability inherent to snow processes. Land surface models are another approach to estimate SWE across regional and continental scales. However, significant uncertainty is commonplace in model-derived SWE estimates due to associated model structure error, model forcing error, model parameterization error, and initial condition error (Lynch-Stieglitz, 1994; Dong et al., 2007; R. H. Reichle, 2008; R. H. Reichle et al., 2017).

Alternatively, characterizing the amount of SWE across regional and continental scales has been attempted using remotely-sensed measurements from space-borne instrumentation, primarily in the form of passive microwave (PMW) brightness temperature, T_b, measurements (e.g., Advanced Microwave Scanning Radiometer for EOS; AMSR-E) (Derksen et al., 2005; Dong et al., 2005; Brucker et al., 2011). However, the accuracy of satellite-based SWE retrievals are adversely impacted by snow morphology (Kelly et al., 2003), stratigraphy (Derksen et al., 2005), snow grain size (Armstrong et al., 1993), ice crusts (Rees et al., 2010), depth hoar (Brucker et al., 2011), and sub-grid scale lake effects (Derksen et al., 2010). PMW T_b-based SWE retrievals are also affected by forest and atmospheric attenuation (Wilheit et al., 1980; Derksen et al., 2005; Savoie et al., 2009), signal attenuation in deep snow (Clifford, 2010), and the assumed (quasi-) linear relationship between the electromagnetic response of the snowpack and the physical characteristics of SWE (Chang et al., 1996; Clifford, 2010). Figure 1 shows a simple comparison of *in situ* measurements of snow depth from the SNOwpack TELemetry (SNOTEL) network along with AMSR-E PMW spectral difference ($\Delta T_{b18V-36V}$, see notations in Equation 1). Although PMW $\Delta T_{b18V-36V}$, in general, captures the accumulation and ablation phase of the snow season, significant high-frequency noise exists and must be carefully considered.

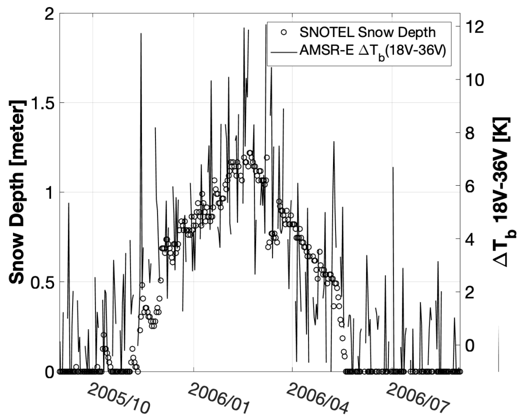

Figure 1. *Comparison between AMSR-E $\Delta T_{b18V-36V}$ observations and SNOTEL snow depth measurements for a location in Western Colorado (40.31° N, 105.65° W) from 1 September 2005 to 1 September 2006. Note how ΔT_b captures the general features of snow depth and SWE, but contains more signals (e.g., snow temperature) not related to snow mass as well as the presence of high-frequency noise.*

Fortunately, data assimilation (DA) is an effective approach to optimally combine information from both observations and model predictions to generate high-quality estimates that are superior to the observations or to the model alone (McLaughlin, 2002; R. H. Reichle, 2008; Forman & Margulis, 2010; Draper & Reichle, 2015). Instead of assimilating snow retrievals, such as SWE or snow depth (SD), PMW T_b or ΔT_b observations (spectral differences, computed as the difference between two T_bs, Equation 1) can be directly assimilated into a land surface model in order to improve model-derived snow mass estimates. Satellite-based PMW T_b observations are sensitive to snow volume scattering, and therefore, contain snow information applicable during all-weather and nighttime conditions.

Multifrequency T_b assimilation was first performed for improvement of point-scale SWE estimation (Durand & Margulis, 2006). Spectral difference (ΔT_b) was first assimilated by Pulliainen (2006) that reduced SWE and SD estimation systematic errors. A follow-on study conducted by Durand et al. (2009) improved SD estimates via assimilation of vertically-polarized T_b ground-based measurements at 18.7 GHz and 36.5 GHz over a snow-covered region. Similarly encouraging results were demonstrated for continental-scale snow storage estimates using T_b assimilation (Kwon et al., 2016, 2017).

The studies mentioned above employed a physically-based microwave radiative transfer model (RTM) as the observation operator to map the relevant land surface model state variables (e.g., SWE or SD) into the corresponding observation space (i.e., T_b). However, it is difficult to apply a RTM over a large-scale snow region due to the nontrivial computational demand (Kwon et al., 2016). In addition, most global land surface models lack the fidelity to accurately represent the snow microstructure (e.g., snow grain size, snow grain shape, internal ice layers) to fulfill the RTM requirements (Kukkonen et al., 2012). Alternatively, a machine learning technique in the form of physically-constrained support vector machine (SVM) regression can be employed (Forman et al., 2014; Forman & Reichle, 2015; Xue & Forman, 2015; Forman & Xue, 2016; Xue & Forman, 2017a, 2017b; Xue et al., 2018; Kwon et al., 2019; Ahmad et al., 2019). It has been shown that a SVM was able to accurately capture the temporal and spatial variability in the modeled T_b or ΔT_b, and thus, holds potential to improve snow estimates across large spatial scales.

Recently, Xue et al. (2018) used SVM regression as the PMW ΔT_b observation operator over North America in the context of snow data assimilation within a land surface model. Xue et al. (2018) showed improvements in snow mass estimation under certain conditions such as shallow, dry snow in the absence of forest cover.

Despite the improvements in snow estimation via radiance assimilation, there are still many deficiencies that must be overcome in order to better optimize its use. For example, snow mass estimation using PMW radiometry is fundamentally an ill-posed, underdetermined system (Durand & Margulis, 2006). That is, there are numerous combinations of snow depth, snow density, snow temperature, snow grain size, and other snow characteristics that collectively yield the same ΔT_b observation. Therefore, the task of assimilating only the snow mass-related portion of the PMW ΔT_b signal from all of the other signals inherent therein (e.g., vegetation, atmosphere, snow temperature, and snow liquid water content) is a challenge. In addition, the efficacy of SVM-based PMW ΔT_b assimilation is often limited by the controllability and reachability of SVM regression (Kwon et al., 2019). All of these issues motivate this study and help answer the question: How can we further improve snow estimation with SVM-based PMW ΔT_b assimilation using a physically-constrained approach? To this end, synthetic AMSR-E PMW ΔT_b observations are assimilated into the Catchment Land Surface Model (i.e., Catchment) (Koster et al., 2000) using SVM regression as the observation operator over snow-covered terrain in Russia.

Unlike previous works that simultaneously assimilating a fixed number of ΔT_b channels (Xue et al., 2018; Kwon et al., 2019), *a priori* modeled SWE is used as an indicator to determine which ΔT_b channels should be assimilated into the model. In addition, a simple "data-thinning" strategy is also explored to help mitigate high-frequency error (e.g., changes in snow temperature *not* related to snow mass) embedded in the synthetic AMSR-E PMW ΔT_b observations. Given the fact that ground-based snow observations of sufficient density and quality are not available in the study area, there is no way to determine the observational baseline for quantifying improvement in snow estimation using real-world measurements. Therefore, a synthetic, identical twin experiment (R. H. Reichle & Koster, 2003) is employed in this study (further discussion in Section 4) in order to provide a systematic means of evaluating land surface model improvements via ΔT_b assimilation.

2 Study Domain

The study domain for this synthetic experiment is the East European Plain spanning from 45°N to 64°N and from 30°E to 62°E (Figure 2), which encompasses the Volga River basin in Russia. The Volga River Basin has an area of 1,390,000 km^2 and occupies about one-third of the East European Plain, and ultimately discharges into the Caspian Sea. The main parts of the basin as delineated in Figure 2 are the upper Volga basin (430,000 km^2), the Moskva Oka River basin (237,000 km^2), the Kama River basin (500,000 km^2), and the lower Volga River basin (223,000 km^2). The relatively large size of the study region allows for an investigation across a range of regional and snow climatologies. The upper and middle parts of the basin are covered by forest and steppe; the lower part of the basin is covered by steppe and desert. The mean annual temperature increases from the north (approximately 3°C) to the south (approximately 9°C), whereas annual precipitation decreases from 750 to 150 mm in the same direction. Average depth of snow cover decreases from 60 cm in the north to about 3 cm in the south and the duration of its persistence is from 240 to 30 days moving from the north to south (Golosov & Belyaev, 2016; Sidorchuk et al., 2009).

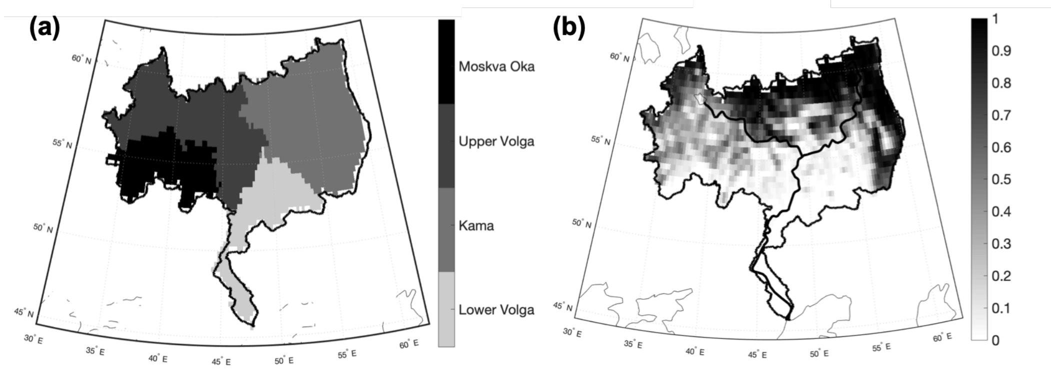

Figure 2. *General maps of (a) Volga Basin and four sub-basins (Moskva Oka river basin, upper Volga river basin, lower Volga river basin, and Kama river basin) and (b) forest cover fraction as derived from the Moderate Resolution Imaging Spectroradiometer (Friedl et al., 2002).*

3 Prognostic Land Surface Model

Following Xue et al. (2018), this work employs the Catchment Land Surface Model (a.k.a. Catchment) (Koster et al., 2000). The basic modeling unit in Catchment is based on the topographic statistics of each hydrologic catchment (or watershed) forced by gridded meteorological forcings that serve as the model boundary conditions (Zaitchik et al., 2008; Kumar, Zaitchik, et al., 2016). Catchment employs three prognostic variables (i.e., surface excess, root zone excess, and catchment deficit) in order to account for soil moisture and shallow groundwater. Snow conditions on the land surface, including snowpack consolidation and metamorphosis, are represented with a three-layer snow model (Stieglitz et al., 2001). One hydraulic limitation in Catchment is the lack of surface water impoundments (e.g., lakes and rivers) along with that of dynamic river routing. In the absence of these hydrologic processes, however, Catchment remains an excellent testbed to explore methods related to the remote sensing and modeling of terrestrial snow.

In the present study, the spatial resolution of the model grid is 25 km on an Equal Area Scalable Earth (EASE version 2) grid (M. J. Brodzik et al., 2012). Realistic initial conditions in the subsurface are generated by looping the model five times over the same 10-year period from 1 September 1992 to 1 September 2002. The model is then initialized and propagated forward from 1 September 2002 in order to initialize the model with minimized initial snowpack and runoff errors. The experiment period covers 1 September 2002 to 1 September 2011, which coincides with the majority of the AMSR-E observation record that is used to train the SVM-based observation operator that is subsequently used to generate the synthetic ΔT_b observations used during assimilation (see section 4.1).

4 Synthetic Identical Twin Experiment Setup

As mentioned in section 1, the lack of available *in-situ* snow, soil moisture, groundwater, and runoff observations in the study region necessitates the use of a simulated version of the synthetic "truth" that serves as a reasonable proxy for the real-world system variability. The synthetic experiment starts with the generation of a synthetic "truth" followed by the generation of synthetic observations (section 4.1). Next, synthetic observations are assimilated (section 4.4) into a degraded version of the same modeling system that is forced by a different (imperfect) set of meteorological boundary conditions (section 4.3).

This type of synthetic experiment is often referred to as an "identical" twin in the sense that the same land surface model is used in all aspects of the experiment. The use of an identical twin implicitly assumes that the majority of errors encountered by a "real-world" system arise from the boundary conditions rather than from the model structure errors, parameter errors, or initial condition errors (Günther et al., 2019). In the context of snow modeling, this assumption is reasonable given the large degree of precipitation error are often found in remote, mountainous terrain such as that as the Volga basin (Rasmussen et al., 2012). Alternatively, a "fraternal" twin could be employed in which different land surface models are used during different phases of the experiment, but this is avoided in this current study in order to focus on model improvements to terrestrial snow (and its subsequent ablation and runoff) associated with erroneous precipitation in the terrestrial environment. The following discussion highlights the different components used in the synthetic identical twin experimental setup.

4.1 Synthetic Truth and Synthetic Observations

The synthetic "truth" is a single replicate simulated by Catchment using boundary conditions defined by the Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) product with an hourly temporal resolution and 0.5° × 0.625° (latitude/longitude) spatial resolution (Gelaro et al., 2017). Hydrologic states and fluxes, including SWE, snow depth, soil moisture, runoff, and surface energy fluxes, from the synthetic "truth" simulation serve as the best Catchment-based representation of the natural environment. It is important to note that the "truth" need not identically match the observed reality. Rather, the "truth" needs to adequately capture the space-time variability of the real-world system.

The synthetic "truth" ΔT_bs are generated using a well-trained SVM (see Appendix A) that maps relevant model states (e.g., SWE, snow temperature) derived from the synthetic truth into the corresponding observation space (i.e., ΔT_b). The synthetic ΔT_b truth, including spectral difference between 10.65 and 36.5 GHz, 10.65 and 18.7 GHz, and 18.7 and 36.5 GHz for both horizontal and vertical polarization (Xue & Forman, 2015, 2017a, 2017b; Xue et al., 2018), is easily expressed as:

$$\Delta T_{b18V-36V} = T_{b18V} - T_{b36V} \tag{1}$$

where T_{b18} represents T_b at 18.7 GHz; T_{b36} represents T_b at 36.5 GHz; and the subscript V represents vertical polarization with a similar equation for horizontal (H) polarization.

The synthetic ΔT_b observations, including $\Delta T_{b10H-36H}$, $\Delta T_{b10V-36V}$, $\Delta T_{b10H-18H}$, $\Delta T_{b10V-18V}$, $\Delta T_{b18H-36H}$, and $\Delta T_{b18V-36V}$, are generated by corrupting the synthetic ΔT_b truth through the inclusion of additive, Gaussian observation noise that is temporally and spatially uncorrelated. The observation noise is assumed to be Gaussian-distributed with zero mean and a standard deviation of 3 K (Burgers et al., 1998; Durand & Margulis, 2007; Kwon et al., 2016; Xue et al., 2018).

4.2 Boundary Condition Correction

Meteorological boundary conditions (a.k.a., forcings) are an important component for snow model simulations in the context of a synthetic, identical twin experiment. Therefore, it is critical to first characterize the boundary condition (precipitation) errors. Boundary condition errors often result in bias or random errors that can be considered representative of the "real-world" errors that could be encountered in an operational assimilation system (Günther et al., 2019). Precipitation (snow) measurement errors frequently range from 20% to 50% in windy conditions (Clifford, 2010). Hence, an error characterization strategy is employed here such that the difference between MERRA-2 forcings (synthetic truth) and GLDAS forcings used in both the open loop (section 4.3) and data assimilation (section 4.4) runs serves as a reasonable proxy for a range of plausible pre-

Table 1. Summary of GLDAS precipitation correction factor γ

Forcing	Neutral	Positively-biased	Negatively-biased
γ	1.9	2.9	1.0

cipitation error scenarios. That is, the cumulative, domain-averaged GLDAS precipitation over the study period is rescaled to match 50% (negatively-biased), 100% (neutral), and 150% (positively-biased) of the MERRA-2 precipitation by multiplying a fixed factor (γ, Table 1) computed as:

$$\gamma = \frac{\alpha \times \sum Precipitation_{(MERRA-2)}}{\sum Precipitation_{(GLDAS)}} \tag{2}$$

where $\sum Precipitation_{(MERRA-2)}$ and $\sum Precipitation_{(GLDAS)}$ are the cumulative, domain-averaged MERRA-2 and GLDAS precipitation over the course of the entire study period, respectively, where α set to 50%, 100% , and 150% to yield the rescaled GLDAS scenarios for negatively-biased, neutral, and positively-biased relative to MERRA-2 (synthetic truth) precipitation, respectively.

These three different scenarios will help explore how data assimilation can improve terrestrial snow estimation where the total amount of precipitation in the study domain is over-, well-, or under-estimated. In addition, the shortwave and longwave radiation boundary conditions are also rescaled proportionally (not shown). This last step is conducted in order to more carefully focus on the first-order control of precipitation (and its error) on snow mass assimilation.

As shown in Figure 3a) and 3b), there is a strong precipitation gradient from the north to south across the study domain where the highest precipitation is in the northwest of the domain for both MERRA-2 and GLDAS. Under the positively-biased scenario (Figure 3d), the "true" precipitation (i.e., MERRA-2) is less than the precipitation forcing field used in both OL and DA (i.e., positively-biased precipitation) with a gradient from the southeast to northwest. A similar pattern is seen for the negatively-biased scenario (Figure 3e) but with more "true" precipitation relative to the OL and DA precipitation. Even though the 9-year cumulative amount of precipitation across the domain is identical between MERRA-2 and the neutral precipitation scenarios, differences still exist at different locations between these two data sets as shown in Figure 3f). As a result, the amount of SWE could be significantly different at different locations due to the nonlinear hydrologic response of forcing even though the domain-averaged precipitation is identical between the two.

4.3 Ensemble Open Loop

As mentioned previously, an ensemble open-loop (OL) simulation is conducted without the assimilation of synthetic observations. The "imperfect" boundary conditions are established using the Global Land Data Assimilation System (GLDAS) product with 3-hourly temporal and $2.0° \times 2.5°$ (latitude/longitude) spatial resolution (Rodell et al., 2004). In this study, the difference between the "truth" and the OL ensemble mean is representative of the system errors. One key in the success of data assimilation (section 4.4) is the appropriate characterization of both model and observation errors (R. H. Reichle, 2008; Kumar, Dong, et al., 2016). An ensemble of perturbations is generally applied to the forcing variables (e.g., precipitation) as a low-rank approximation of the true system errors. The ensemble mean of the model states is typically used as the expected model estimate and the ensemble spread is used as a proxy for the model error variance (Houtekamer & Mitchell, 1998; Burgers et al., 1998). In line with previous work (Xue et al., 2018), the perturbation settings used in this study are summarized in Table 2.

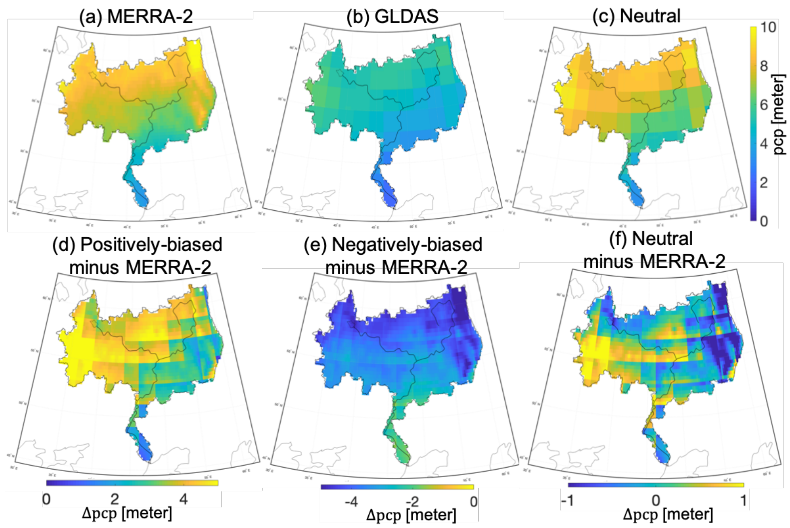

Figure 3. *Cumulative precipitation (pcp [m]; from years 2002 to 2011): (a) MERRA-2 precipitation, (b) GLDAS precipitation, (c) neutral precipitation, (d) positively-biased minus MERRA-2 precipitation, (e) negatively-biased minus MERRA-2 precipitation, and (f) neutral minus MERRA-2 precipitation.*

Table 2. Parameters for meteorological forcings perturbation in the assimilation experiments

					Cross-correlation		
Model Variables	Type	Standard Deviation	x, y_{corr}	t_{corr}(day)	pcp	sw	lw
pcp	M[a]	0.5	2°	3	NA	−0.8	0.5
sw	M	0.3	2°	3	−0.8	NA	−0.5
w	A[b]	20 W m^{-2}	2°	3	0.5	−0.5	NA

[a]Multiplicative (M) or [b]Additive (A) perturbations are applied to precipitation (pcp), downwelling shortwave radiation (sw), downwelling longwave radiation (lw). Spatial correlations are indicated as x,y_{corr} and temporal correlations as t_{corr}.

4.3.1 Ensemble Size

With a small ensemble size, only a small subset of the error space is sampled, and thus, the statistical (or sampling) error is non-negligible (Keppenne, 2002; R. Reichle et al., 2002; Evensen, 2003). The ensemble size, in part, dictates whether or not the relevant part of the error structure (e.g., error variance) can be captured by the finite ensemble size of model trajectories. In this study, a range of ensemble sizes from $N = 14$ to $N = 74$ was tested. An ensemble size of $N = 24$ is ultimately chosen because $N >$ 24 show no significant change in the ensemble spread (i.e., the ensemble SWE standard deviation over the study domain) compared to $N = 24$. Therefore, it is assumed an ensemble size of 24 replicates could reasonably represent the low-rank approximation of the true error probability distribution.

4.4 Data Assimilation

An existing, one-dimensional (1-D) ensemble Kalman filter (EnKF) framework (R. H. Reichle & Koster, 2003; R. H. Reichle et al., 2010) is employed for daily, synthetic AMSR-E ΔT_b assimilation in this study. In a 1-D EnKF, the computational units are processed independently from one another, which suggests that spatial error correlations between different catchments within the study domain are negligible (R. H. Reichle & Koster, 2003). Only the essential details of the EnKF are discussed here. Further information regarding the EnKF equations can be found in R. H. Reichle et al. (2002).

The EnKF alternates between an ensemble forecast step and an update step. During the forecast step, an ensemble of model state vectors containing the relevant model prognostic variables (e.g., SWE) are propagated forward in time by Catchment. Using the available synthetic ΔT_b observations y_t at time t, the prior state vector, x_t^{i-}, is updated to a new value, x_t^{i+}, based on the relative uncertainties between the state vector (SWE in this particular study) and the predicted observation using appropriate weights expressed in the Kalman gain K_t (R. H. Reichle et al., 2001; Zaitchik et al., 2008; Kumar, Zaitchik, et al., 2016) via:

$$x_t^{i+} = x_t^{i-} + K_t(y_t + v^i - \phi_t(x_t^{i-})), v \sim \mathcal{N}(0, 3^2) \tag{3}$$

where v^i represents the representativeness errors that are assumed here to be Gaussian with zero mean and a spatially- and temporally-uncorrelated covariance of 3^2 K^2 ; $\phi_t(\cdot)$ is the SVM-based observation operator that maps the model states (e.g, SWE, SLWC) into ΔT_b observation space (see Appendix A); and i represents a single replicate drawn from the ensemble of size $N = 24$. The superscripts $-$ and $+$ refer to the *a priori* state vector and *a posterior* state vector, respectively. K_t is the Kalman gain matrix and can be computed as:

$$K_t = C^-_{x_t y_t}[C^-_{y_t y_t} + R]^{-1} \tag{4}$$

where $C^-_{x_t y_t}$ is the error cross-covariance between the modeled SWE estimates and the SVM-based ΔT_b prediction prior to the update; $C^-_{y_t y_t}$ is the error covariance (a.k.a., sample covariance) of the SVM-based ΔT_b prediction prior to the update; and R is the observation error variance.

4.4.1 Observation Operator and SVM Controllability

It is important to highlight the issue of controllability with the SVM-based observation operator (see Appendix A) (Kwon et al., 2019). As an important factor in optimal control theory, controllability demonstrates the skill of a linear or nonlinear model to guide the model output from any physical plausible initial state towards any physically plausible final state over a finite time period (Ogata & Yang, 2002). That is, the system output for a controllable system can be changed by changing the system input (Gelb, 1974). In the context of data assimilation, controllability of the SVM-based observation operator is critical during the analysis update. The reason is that one of the

Table 3. Descriptions of different ΔT_b data assimilation strategies

Name	Description
baseline DA	Simultaneous assimilation of six ΔT_b channels including: $\Delta T_{b10H-18H}$, $\Delta T_{b10V-18V}$, $\Delta T_{b10H-36H}$, $\Delta T_{b10V-36V}$, $\Delta T_{b18H-36H}$, and $\Delta T_{b18V-36V}$
physically-constrained DA	Update SWE based on ***prior* SWE ensemble mean**: If SWE$\leq$120 [mm], use $\Delta T_{b10H-36H}$, $\Delta T_{b10V-36V}$, $\Delta T_{b18H-36H}$, and $\Delta T_{b18V-36V}$ If SWE>120 [mm], use $\Delta T_{b10H-36H}$, $\Delta T_{b10V-36V}$, $\Delta T_{b10H-18H}$, and $\Delta T_{b10V-18V}$
3-day thinning DA	Simultaneous assimilation of six ΔT_b channels **every 3 days** including: $\Delta T_{b10H-18H}$, $\Delta T_{b10V-18V}$, $\Delta T_{b10H-36H}$, $\Delta T_{b10V-36V}$, $\Delta T_{b18H-36H}$, and $\Delta T_{b18V-36V}$

assumptions behind the SVM-based DA framework is that model errors predominately correlate back to errors in the SVM-based ΔT_b predictions (Kwon et al., 2019). An "uncontrollable" SVM is insensitive to changes in the inputs, and thus, leads to the collapse of the ensemble of SVM-based ΔT_b predictions (Kwon et al., 2019). Controllability is related to the set of training data and the inability of the SVM to accurately predict snow ΔT_b when the given inputs that are outside of the prediction space implicit in the training data (Ahmad et al., 2019). As a result, the model error would no longer correlate back to the corresponding error in the SVM-based observation operator, which can lead to spurious error correlations that ultimately degrade the model estimate (Kwon et al., 2019). To avoid this, prior SWE is updated only when the standard deviation of the prior SVM-predicted ΔT_b is greater than 0.05 K based on heuristics outlined in Kwon et al. (2019).

It is worth noting that there is no DA performed around water bodies. This is due to that fact that grid cells with more than 5% coverage by water (ocean or inland water bodies) are excluded from the data assimilation analysis because surface water impoundments are not explicitly accounted for in the Catchment model. Therefore, grid cells containing surface impoundments are not included in the EnKF update.

4.4.2 Shallow-to-Medium versus Medium-to-Deep Snow Algorithm

Following the methods of Xue et al. (2018) and Kwon et al. (2019), the DA experiments start with simultaneous assimilation of six different ΔT_bs, including $\Delta T_{b10H-18H}$, $\Delta T_{b10V-18V}$, $\Delta T_{b10H-36H}$, $\Delta T_{b10V-36V}$, $\Delta T_{b18H-36H}$, and $\Delta T_{b18V-36V}$ (baseline DA; Table 3) in this study. However, these studies demonstrated that DA, at times, could degrade the snow estimation (also see section 7.1). To prevent DA degradation, a physically-constrained DA approach (section 4.4.2) and a data thinning DA approach (section 4.4.3) are introduced.

An important assumption behind spectral difference assimilation is that ΔT_b is positively correlated with SWE and that the T_b at the highest frequency (i.e., 36.5 GHz) decreases as SWE increases while the lower frequency (i.e., 10.65 GHz or 18.7 GHz) T_b is relatively insensitive to increasing snow mass (Kim & England, 2003; Derksen et al., 2010). However, the correlation between SWE and T_b at 36.5 GHz can reverse once SWE is greater than 100 to 200 mm (Schanda et al., 1983; Seve et al., 1986; Mätzler, 1994; Derksen, 2008; Derksen et al., 2010; Kwon et al., 2019). This occurrence is often referred to as signal saturation. Furthermore, ΔT_{b10-18} may introduce errors during shallow snow conditions because the signal is more representative of the soil moisture rather than the snow mass (Roy et al., 2013). In other words, the observation error covariance of ΔT_{b18-36} during deep snow conditions or ΔT_{b10-18} during shallow snow conditions may not be adequately represented by the prescribed error parameters, and hence, may introduce spurious errors during the EnKF update. If an observational data set contains data whose errors are not well represented by the prescribed error parameters, the EnKF will not be able to accurately estimate the true fields (R. H. Reichle, 2008). Therefore, one hypothesis for DA degradation in the snow estimates is due to the simultaneous assimilation of all six ΔT_b when many of the ΔT_bs are not truly representative of snow mass. Further, the presence of signal saturation during deep snow conditions also result in a degraded estimate when the general assumption that ΔT_b is positively correlated with SWE is no longer the case.

Instead of simultaneously assimilating all available multifrequency and polarization spectral differences into a land surface model as done in the works of Xue et al. (2018) and Kwon et al. (2019), a new assimilation strategy based on prior SWE information is explored here such that ΔT_b is assimilated more selectively. In this new approach, the ensemble mean of the prior SWE is used as an indicator to determine which ΔT_b should be assimilated. That is, shallow-to-medium snow conditions now only utilize $\Delta T_{b18H-36H}$, $\Delta T_{b18V-36V}$, $\Delta T_{b10H-36H}$, and $\Delta T_{b10V-36V}$ whereas medium-to-deep snow conditions now only simultaneously utilize $\Delta T_{b10H-18H}$, $\Delta T_{b10V-18V}$, $\Delta T_{b10H-36H}$, and $\Delta T_{b10V-36V}$ (physically-constrained DA; Table 3). The SWE threshold used to differentiate between the shallow-to-medium and medium-to-deep snow is somewhat subjective. In this study, the shallow-to-medium snow refers to SWE$\leq$120 [mm] while SWE>120 [mm] is considered as the medium-to-deep snow based on peer-reviewed literature (Matzler et al., 1982; Mätzler, 1994; De Sève et al., 1997).

4.4.3 Data Thinning

Another common issue with snow ΔT_b assimilation is ensemble collapse (i.e., little or no ensemble spread) (Figure 4). Ensemble collapse results in an under-representation of the true model uncertainty. Given the fact the optimal combination of the observations with the model is predicated on the consideration of the respective uncertainties of each (R. H. Reichle et al., 2008), a poor representation of model uncertainty will often lead to degraded snow estimation.

Ensemble collapse, in part, is exacerbated by the multi-observation nature of the assimilation approach (i.e., multiple observations assimilated daily). In addition, the high-frequency errors embedded in the AMSR-E observations (Figure 1) often overwhelm the snow-related information, and thus, can further degrade DA performance. In an attempt to mitigate such high-frequency noise, the synthetic ΔT_b observations (all six channels) are assimilated every 3-, 5-, 7-, 10-, and 15-day intervals rather than daily in order to explore the impact of using fewer observations on DA performance (Table 3).

5 Normalized Innovation and Filter Optimality Assessments

The optimal operation of the Kalman filter is closely related to the statistical properties of the innovation sequence, which is the difference between the observation and model forecast (R. H. Reichle & Koster, 2002). In theory, the information exchange during the filter update is optimal when the normalized innovation sequence appears as white noise (i.e., mean zero with unit variance and temporally uncorrelated). If the models are unbiased and linear (both the land surface model and the observation operator) and all errors are uncorrelated and Gaussian (and correctly specified), then the normalized innovation sequence, NI, should appear similar in form to a standard normal distribution $\mathcal{N}(0,1)$ (R. H. Reichle & Koster, 2002). Although both the land surface model and observation operator used here are nonlinear, the investigation of the normalized innovation sequence can still provide useful information as to the performance of the DA procedure.

The normalized innovation (NI) at time t can be written as:

$$NI_t = \frac{y_t - \phi_t(x_t^-)}{\sqrt{C_{y_t y_t} + R}} \quad (5)$$

where the numerator is the difference (or innovation) between the synthetic ΔT_b observation (y_t) and SVM-based predicted (prior) ΔT_b observation ($\phi_t(x_t^-)$), and the denominator is the square root of the sum of the background error covariances ($C_{y_t y_t}$) and the observation error covariance (R). The normalized innovation sequence is merely the vector concatenation of all NI_t across the duration of the assimilation experiment that is explored for mean zero, unit variance, temporally-uncorrelated Gaussian-like features that can be used as a proxy for filter optimality.

6 Validation Approach

A synthetic, identical twin experiment is designed such that the "true" values of hydrologic states and fluxes are known. Therefore, the validation is performed against the true states (e.g., SWE) derived from the synthetic truth run. Several goodness-of-fit statistics are used for the validation activities: (1) bias, (2) root mean squared error (RMSE), (3) unbiased root mean squared error (ubRMSE), (4) correlation coefficient (R), (5) Nash-Sutcliffe efficiency (NSE), and (6) containing ratio ($\mathrm{CR}_{2\sigma}$). In addition, normalized information contribution (NIC) is used to quantify the DA improvement (or degradation) relative to the OL (Kumar et al., 2009, 2014). Details about all of these calculations can be found in Appendix B.

7 Results and Discussion

7.1 SWE Estimates

The DA experiments started with simultaneous assimilation of six different ΔT_bs, including $\Delta T_{b10H-18H}$, $\Delta T_{b10V-18V}$, $\Delta T_{b10H-36H}$, $\Delta T_{b10V-36V}$, $\Delta T_{b18H-36H}$, and $\Delta T_{b18V-36V}$. For illustrative purposes, two relatively ideal locations, Grid #1 and Grid #2, (i.e., long snow season; relatively dry snow conditions; no forest cover; and relatively shallow snow such that $\mathrm{SWE}_{max} < 200$ mm) are selected. Given the fundamental physics of PMW T_b remote sensing of snow, if assimilation does not work at these idealized locations (assuming appropriate specification of input error parameters), then assimilation will likely not work at other locations in the Volga basin. Therefore, we chose to present these locations prior to discussing results across the remainder of the basin.

Figure 4 highlights the performance of DA (denoted as baseline DA as shown in blue) at these idealized locations under the neutral forcing conditions. As shown in Figure 4a) and 4b), simultaneous assimilation of all six ΔT_b channels actually degraded model SWE estimates. Starting in late-January 2010 (Figure 4a), the DA SWE estimates diverged from the OL (gray color) and the synthetic truth (black dots). This divergence resulted in degraded SWE estimates with approximately 82%, 80%, and 85% increases in RMSE, bias, and ubRMSE, respectively, relative to the OL. Similarly, the baseline DA SWE estimates at Grid #2 also diverged from the synthetic truth early in the snow season (Figure 4b) such that the DA routine was unable to recover.

The following sections will discuss whether the physically-constrained DA or the data thinning strategy could serve as a feasible solution in preventing filter divergence, and hence, DA degradation while assimilating PMW ΔT_bs.

7.1.1 *Shallow-to-Medium versus Medium-to-Deep Snow Algorithm*

As shown in Figure 4a), the new assimilation strategy (denoted as physically-constrained DA, red color) improved the Grid #1 SWE estimates with a 58%, 80%, and 41% reduc-

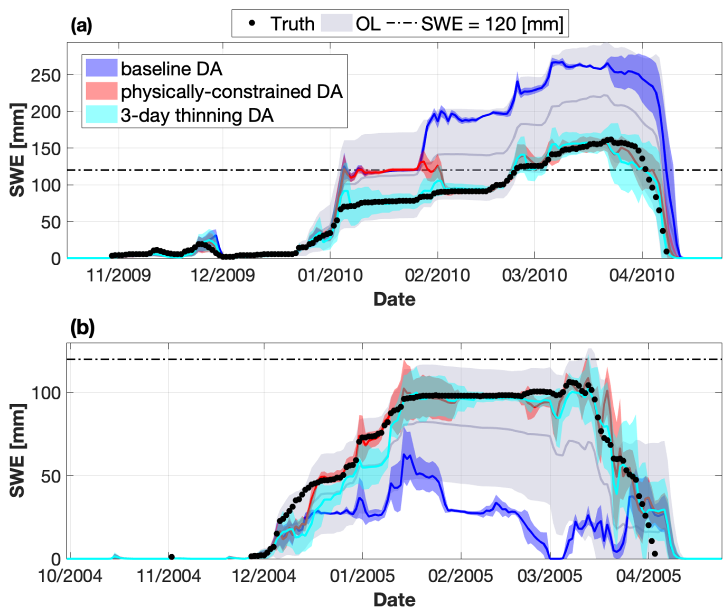

Figure 4. *Example time series of snow water equivalent (SWE) for (a) Grid #1 (54.1685° N, 47.3343° E) from October 2009 to May 2010 and (b) Grid #2 (49.1489° N, 54.0778° E) from October 2004 to May 2005. Physically-constrained DA and data thinning (3-day) improve model results whereas baseline DA (no physical constraint) actually degrades model results relative to the open loop.*

tion in RMSE, bias, and ubRMSE, respectively, relative to the OL. Starting in late-January 2010 (SWE $\approx$ 120 mm), the physically-constrained DA converged toward the synthetic truth (black dots) and was able to encapsulate more of the synthetic truth resulting in a larger containing ratio, $CR_{2\sigma}$ (0.26) than the baseline DA ($CR_{2\sigma} = 0.10$) (see Appendix B). These results suggest that the physically-based shallow-to-medium versus medium-to-deep snow algorithm effectively mitigated much of the negative influence of spurious correlations between SWE and ΔT_{b18-36} during deep snow conditions. Figure 4b) further illustrates the benefits of the physically-constrained shallow-to-medium versus medium-to-deep snow algorithm during the shallow snow conditions.

As shown in Figure 5a) and 5b), the correlation coefficient (R) between the synthetic truth SWE and SVM-based synthetic ΔT_{b10-18} observations for Grid #2 was 0, whereas the correlation coefficients between the synthetic truth SWE and SVM-based synthetic ΔT_{b10-36} (Figure 5c, 5d) and ΔT_{b18-36} (Figure 5e, 5f) observations were greater than 0. This implies that ΔT_{b10-18} contained little or no information about shallow snow conditions (i.e., SWE $\leq$ 120 mm) given the fact that both 10 GHz and 18 GHz undergo little or no scattering across such a shallow snow pack (Durand & Margulis, 2006). That is, these T_b frequencies (and ΔT_b by construct) are effectively transparent through such shallow snow. After removing ΔT_{b10-18} from the observation vector, the physically-constrained DA (red color) was able to correct the model SWE estimates towards the synthetic truth (black dots) during the middle of December 2004, which resulted in a 24%, 92%, and 24% reduction in RMSE, bias, and ubRMSE, respectively, relative to the OL. Further, the $CR_{2\sigma}$ was increased from 0.04 (baseline DA) to 0.29 (physically-constrained DA), which suggested the physically-constrained DA was superior to the more uniformed, baseline approach.

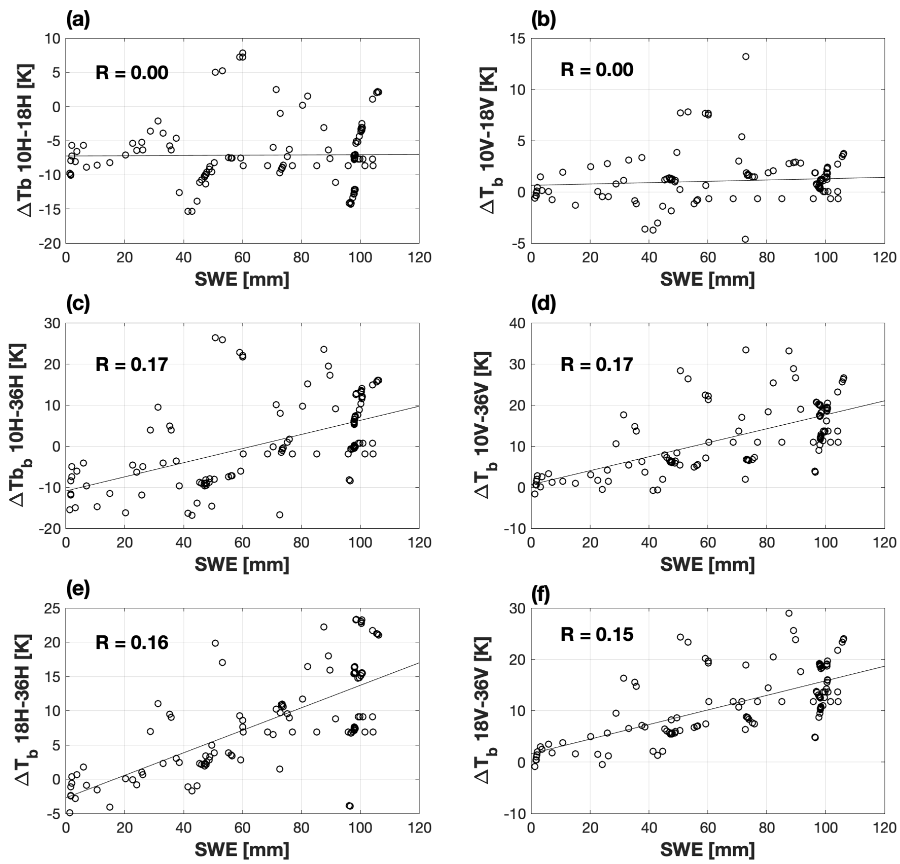

Figure 5. *Scatter plots (with correlation in upper-left corner) between the synthetic truth SWE and the SVM-based brightness temperature spectral difference (ΔT_b) for (a) $10H - 18H$, (b) $10V - 18V$, (c) $10H - 36H$, (d) $10V - 36V$, (e) $18H - 36H$, and (f) $18V - 36V$ estimates for Grid #2 (49.1489°N, 54.0778°E) from 1 September 2002 to 1 September 2011.*

However, such a strategy is far from a panacea and is only effective at some locations. One possible reason is that if the prior SWE estimate is incorrect, it is possible the spectral differences used in the update are not the most appropriate. As an alternative, information from the observations can be used directly to help guide which of the ΔT_bs to assimilate into model at a given point in time and space. For example, if ΔT_{b18-36} observation suggests the spectral difference is nearing saturation, then one can assimilate the longer wavelengths as ΔT_{b10-18}. This approach will be explored in a future study and is considered beyond the current project scope.

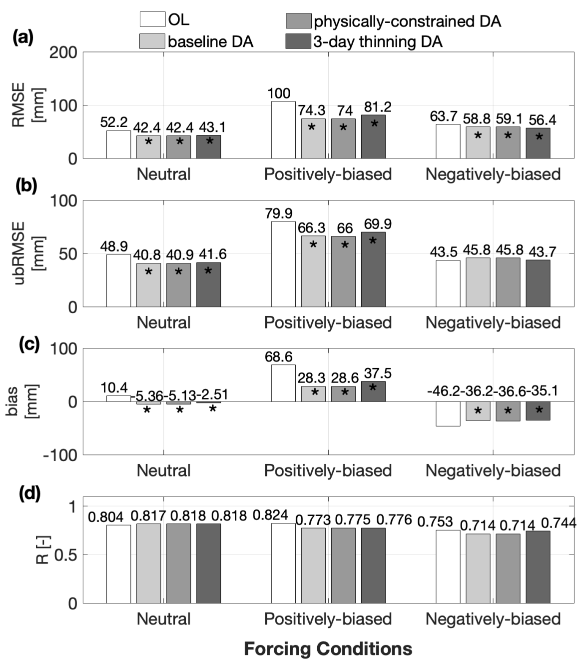

Figure 6. *Histograms of Volga basin-averaged SWE statistics showing (a) RMSE, (b) ubRMSE, (c) bias, and (d) R under the neutral (first set), positively-biased (second set), and negatively-biased (third set) forcing conditions. The white bar is for the Open Loop (OL). The light gray bar is for baseline DA. The medium gray bar is for physically-constrained DA and the dark gray bar is for DA 3-day thinning as listed in Table 3. Bars marked with $*$ indicate DA yields statistically significant statistics with a level of significance of 5%.*

As shown in Figure 6, the Volga basin-averaged results suggested the physically-constrained DA (bias = -5.1 mm; R = 0.818) showed limited improvements in SWE estimation over the baseline DA (bias = -5.4 mm; R = 0.817) under the neutral forcing conditions. Similarly, the $\mathrm{CR}_{2\sigma}$ increased from 0.24 (baseline DA) to 0.25 (physically-constrained DA). Over 57% of the basin grids had improved $\mathrm{CR}_{2\sigma}$. The SWE ensemble spread (σ), which was defined as the long-term time-average of the instantaneous ensemble standard deviation, was also changed from 3.33 mm (baseline DA) to 3.56 mm (physically-constrained DA), which suggested the physically-constrained DA effectively inflated the ensemble spread, and had better ability to capture more of the synthetic truth. All these results suggest that the physically-constrained DA marginally improved the accuracy of SWE estimation relative to the baseline DA under the neutral forcing conditions.

Table 4. Domain-averaged SWE Statistics for DA thinning experiments from 1 September 2002 to 1 September 2011 under the neutral forcing conditions

Statistics	OL	DA baseline	DA thinning 3-day	DA thinning 5-day	DA thinning 7-day	DA thinning 10-day	DA thinning 15-day
RMSE [mm]	52	42	43	44	45	46	47
ubRMSE [mm]	49	41	42	43	43	44	45
bias [mm]	10	-5.4	$\mathbf{-2.5}^{a}$	$\mathbf{-0.73}$	0.16	1.3	3.6
$CR_{2\sigma}$	0.30	0.25	**0.26**	**0.27**	**0.28**	**0.28**	**0.28**

Bold[a] number indicates which experiment yields statistically significant statistics with a level of significance of 5%.

7.1.2 Data Thinning

An example of data thinning to once every three days (cyan color) for Grid #1 under the neutral forcing conditions is shown in Figure 4a). SWE estimates were improved with a 80%, 98%, and 70% reduction in RMSE, bias, and ubRMSE, respectively, relative to the OL. $CR_{2\sigma}$ increased from 0.055 (OL) to 0.25 implying that using fewer observations in time during the DA update can help better capture the synthetic truth. In addition, the SWE ensemble spread increased from 1.86 mm (baseline DA) to 3.02 mm (3-day thinning DA), but was significantly less than the OL (9.79 mm). The bigger SWE ensemble spread indicated the 3-day thinning DA effectively prevented ensemble collapse from January to February. The 3-day thinning strategy also significantly improved SWE estimates for Grid #2 as shown in Figure 4b). The 3-day thinning strategy helped prevent SWE divergence in December 2004, and as a result, yielded a 50%, 72%, and 25% reduction in RMSE, bias, and ubRMSE, respectively, relative to the OL. As a measure of the standard deviation of the errors, the decrease in ubRMSE suggested that the data thinning strategy effectively mitigated some of the introduction of high-frequency noise (random error) at both locations.

The Volga basin-averaged results under the neutral forcings are summarized in Table 4. SWE RMSE increased from 43 mm to 47 mm as fewer observations were assimilated into the model from once a day to every 15 days. This corroborated the earlier results that synthetic assimilation indeed added utility to the model; in the absence of assimilation (i.e., if data thinning approached an infinite amount of time) the results would revert back to the original OL results. Further, there was no statistically significant difference between the baseline DA RMSE and 3-day thinning DA RMSE. In aggregate, these results suggest assimilating the synthetic ΔT_b every 3 days yielded the same amount of SWE errors with daily assimilation.

As fewer observations were assimilated beyond every 3 days, ubRMSE increased from 41 mm (baseline DA) to 45 mm (15-day thinning DA) indicating that assimilating the noisy observations every few days did not help mitigate the random noise embedded in the synthetic ΔT_b observations. The bias, however, were statistically significant (at a level of significance of 5%) and decreased from 10 mm (OL) to -2.5 mm and -0.73 mm when the model simultaneously assimilated with all six channels every three and five days, respectively, rather than once a day. These results imply that daily assimilation using all six channels tended to underestimate SWE (in part due to filter divergence) and may have overconstrained the model, and hence, often resulted in degraded SWE estimation in terms of bias.

Compared to the baseline DA, all DA thinning strategies enhanced the ability to better capture the synthetic truth (i.e., larger $CR_{2\sigma}$) in part by preventing ensemble collapse, which was also proven by the bigger SWE ensemble spread of 4.6 mm (3-day thinning DA) relative to 3.3 mm (baseline DA). It can be reasonably argued that the 3-day thinning data assimilation strategy was better for SWE estimation under the neutral forcing conditions given the statistical results along with the benefit of a reduction in computational demand.

7.1.3 Effects of Precipitation Bias

Figure 6 shows the spatially-averaged statistics for the Volga river basin using three different sets of boundary (forcing) conditions. All the DA strategies (including the baseline DA, physically-constrained DA, and 3-day thinning DA) had the best performance in terms of SWE estimation under the positively-biased forcing conditions. Compared to the OL, the RMSE was reduced by approximately 30%, 31%, and 24% with the baseline DA, physically-constrained DA, and 3-day thinning DA, respectively. On the contrary, DA with the negatively-biased forcing conditions had relatively smaller improvements with approximately 7.6% (baseline DA), 7.2% (physically-constrained DA), and 11% (3-day thinning DA) reduction in RMSE relative to the OL. Under the negatively-biased forcing conditions, all the DA strategies even degraded the SWE estimation in terms of more ubRMSE relative to the OL. The same results were found for bias and ubRMSE.

These results highlights a unique facet of snow assimilation – it is easier for the DA system to remove excess mass than to add missing mass. That is, in part, because the SVM can only make a prediction when snow exists, and hence, can only update the land surface model when snow is present in the model. This behavior is not unique to the SVM, but could also be said when a radiative transfer model is used as the observation operator as part of an ensemble-based DA approach for snow.

7.2 Filter Diagnostics

Figure 7 shows the temporal mean ($\overline{NI}$) and standard deviation (σ_{NI}) of the normalized innovation sequence (NI) over the study domain under the neutral forcing conditions. In general, the negative $\overline{NI}$s computed at $\Delta T_{b10H-18H}$ and $\Delta T_{b10V-18V}$ suggest the SVM-based ΔT_b forecasts had a small negative bias relative to the synthetic ΔT_b observations. On the contrary, SVM-based $\Delta T_{b10H-36H}$, $T_{b10V-36V}$, $\Delta T_{b18H-36H}$ and $\Delta T_{b18V-36V}$ forecasts had positive biases relative to the synthetic ΔT_b observations.

The σ_{NI} (i.e., the standard deviation of the NI) values computed from horizontally polarized spectral differences were greater than for the vertically polarized spectral differences. In addition, the spatially-averaged σ_{NI} were greater than 1, implying that all DA strategies underestimated the observation and/or forecast errors for each frequency and polarization combination under the neutral forcing conditions. Such underestimation could be corrected using a fraternal twin experiment (rather than an identical twin experiment), but is considered well beyond the scope of this study. The σ_{NI} computed at $\Delta T_{b10H-36H}$ and $T_{b10V-36V}$ were the smallest, which can be explained by the fact $\Delta T_{b10H-18H}$, $T_{b10V-18V}$ are not as sensitive to shallow snow conditions, and hence, typically exhibit smaller variability during the entire study period.

Compared to the baseline DA (blue color), the physically-constrained DA (red color) and 3-day thinning DA (black color) had a smaller σ_{NI} for each frequency and polarization combination. It suggested that the prescribed observation error characteristics are more optimal for the physically-constrained DA and 3-day thinning DA compared to the baseline DA. However, the SVM-based $\Delta T_{b10H-36H}$, $T_{b10V-36V}$, $\Delta T_{b18H-36H}$ and $\Delta T_{b18V-36V}$ forecasts within the physically-constrained DA and 3-day thinning DA had a relatively larger bias relative to the synthetic ΔT_b observations. The $\overline{NI}$ and σ_{NI} computed at $\Delta T_{b10H-18H}$ and $\Delta T_{b10V-18V}$ for the 3-day thinning DA were the closest to 0 and 1, respectively, relative to baseline DA and physically-constrained DA. These results

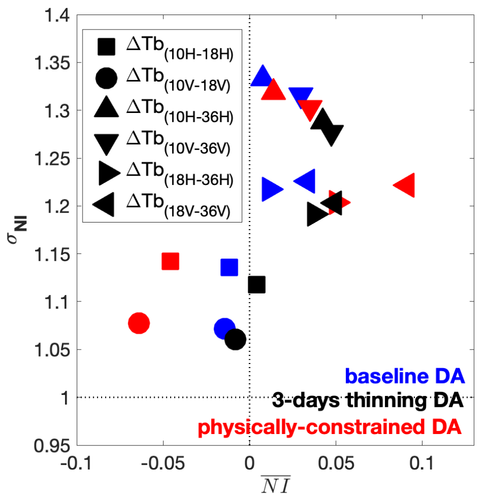

Figure 7. *Innovation statistics for $\Delta T_{b10H-18H}$, $\Delta T_{b10V-18V}$, $\Delta T_{b10H-36H}$, $\Delta T_{b10V-36V}$, $\Delta T_{b18H-36H}$, and $\Delta T_{b18V-36V}$ shown as different marker shapes. The different marker colors represent different DA strategies as listed in Table 3.*

suggest that the observation error characteristics for ΔT_b assimilation in this study may be too simplistic. Observation error standard deviations as a function of frequency, polarization, and land cover type (i.e., forested versus non-forested) should be explored in the future.

7.3 Seasonality

To further investigate DA performance over the Volga Basin, the basin-averaged bias and RMSE as a function of season for positively-biased, negatively-biased, and neutral forcing conditions are presented in Figure 8. Similar patterns were found for ubRMSE (not shown). As expected, the baseline DA performance showed a strong seasonal component under all three forcing conditions. During the snow accumulation period, generally from September to March, DA SWE estimates outperformed OL in terms of smaller bias and RMSE.

Staring in April, DA performance waned in terms of SWE estimation due to deep snow conditions and/or wet snow conditions given the limited skill of PMW remote sensing of snow (Clifford, 2010). As a measure of the presence of random error, ubRMSE had the largest value during April for DA SWE (not shown). The increase in ubRMSE can be explained, in large part, by the introduction of high-frequency errors originating from the synthetic ΔT_b observations along with the fact that PMW remote sensing skill is least when the snow is deep and/or wet (Clifford, 2010). One main reason for the degradation via DA during April was that snow liquid water (i.e., liquid water coating the snow grains) was commonplace during the snow ablation period.

It has been shown that wet snow introduces additional uncertainties in the estimation of SWE (Walker & Goodison, 1993; Clifford, 2010). The presence of liquid water within the snowpack alters the electromagnetic response from a dry microwave scatter to a wet microwave emitter (Walker & Goodison, 1993; R. L. A. Brodzik & J., 2001). When the snow is wet, the general assumptions implicit in ΔT_b-based remote sensing of snow are violated (Walker & Goodison, 1993), and hence, the information content in the ΔT_b observations need not be related to snow mass. As an example shown in Figure 9,

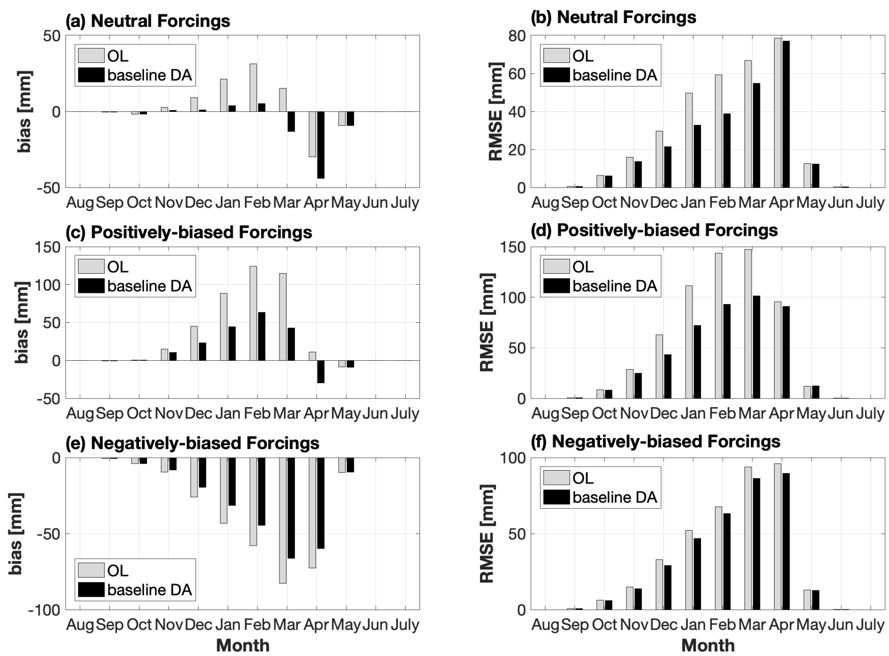

Figure 8. *Histograms of monthly Volga basin-averaged SWE bias (first column) and RMSE (second column) under the neutral (first row), positively-biased (second row), and negatively-biased (third row) forcing conditions. Bias and root mean squared error (RMSE) were computed by comparing OL or DA SWE ensemble mean against the synthetic truth. The light gray bar is for the Open Loop (OL) and the black bar is for the baseline DA as listed in Table 3.*

the correlations between dry (gray plus signs) or wet snow (black dots) and the SVM-based $\Delta T_{b18V-36V}$ synthetic observations changed dramatically. Namely, $\Delta T_{b18V-36V}$ increased as SWE increased for dry snow. Alternatively, $\Delta T_{b18V-36V}$ transitioned to a zero (Figure 9a) or negative (Figure 9b) correlation with SWE when the snow pack ripens.

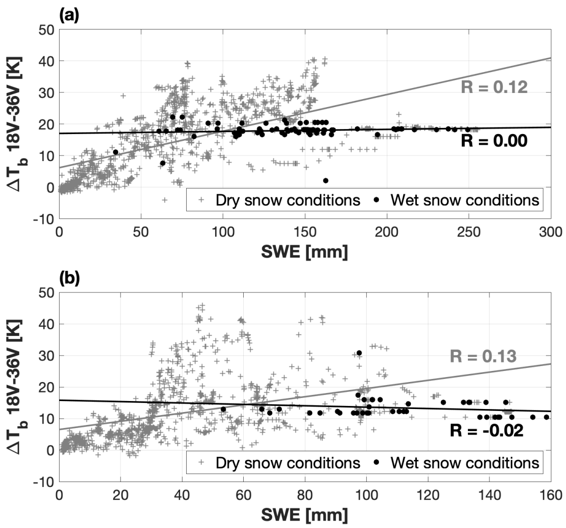

Figure 9. *Scatter plots (with correlations) between the model dry snow (gray plus signs) and wet snow (black dots) along with SVM-based brightness temperature spectral difference* $\Delta T_{b18V-36V}$ *estimates for (a) Grid #1 (54.1685° N, 47.3343° E) and (b) Grid #2 (49.1489° N, 54.0778° E) from 1 September 2002 to 1 September 2011.*

It is worth noting that DA had the worst performance in terms of SWE estimation under the negatively-biased forcing conditions. The relatively small change in RMSE between the OL and DA suggested that DA could not significantly improve SWE estimates. In addition, DA had a larger ubRMSE than OL across the entire snow season. It suggested that ΔT_b assimilation under the negatively-biased forcing conditions was suboptimal. This latter point highlights the fact that assimilation works better at ameliorating a positive bias (positively-biased forcings) more so than a negative bias.

7.4 Effects of Forest Attenuation

The performance of the snow DA framework in forested regions is explored here in more detail because the presence of forest canopy can significantly alter the PMW ΔT_b signal as measured at the top of the atmosphere. More specifically, a low sensitivity of PMW ΔT_b from terrestrial snow is often observed in densely-forested areas. Overlying vegetation attenuates the PMW radiation emitted from the underlying snowpack while simultaneously adding its own contribution to the signal that is measured by the radiometer (Derksen et al., 2005). Among all these three frequency channels (i.e., 10.65 GHz, 18.6 GHz, and 36.5 GHz), microwave emission at 36.5 GHz is most strongly absorbed by standing vegetation (Derksen, 2008). Consequently, the scattering signal from the underlying snowpack can be overwhelmed by upwelling microwave radiation from the canopy (Derksen, 2008).

Figure 10 shows NIC_{RMSE} as a function of forest fraction under the neutral forcing conditions for baseline DA. Similar results were found for other DA strategies under both the positively-biased and negatively-biased forcing conditions (not shown). Over-

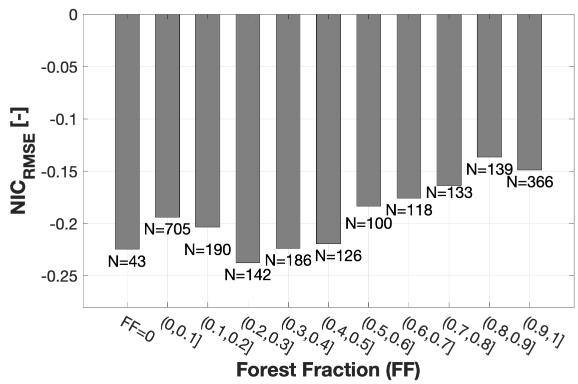

Figure 10. *Histograms of the domain-averaged SWE NIC_{RMSE} as a function of forest fraction under the neutral forcing condition across the study domain. N is the number of model grid cells. A negative value of NIC_{RMSE} indicates data assimilation (DA) improves SWE estimates relative to the open loop (OL). Note that the largest improvements occur in the relatively sparsely-forested region where PMW attenuation is less pronounced.*

all, DA improved the SWE estimates relative to OL in the most sparsely-forested regions (i.e., forest fraction≤0.4). A hypothesis test at a level of significance of 5% was conducted to investigate whether the forest cover had a significant effect on the DA performance. The null hypothesis was that the mean of NIC_{RMSE} for sparsely-forested areas (FF≤0.4) was significantly smaller than the mean of NIC_{RMSE} for densely forested areas (FF>0.4) (i.e., the forest cover has a negative impact on DA performance). The results suggested the negative effect of forest was statistically significant for the DA algorithm.

7.5 Runoff Estimates

Monthly domain-averaged runoff estimates from the OL and DA were compared against true (synthetic) runoff from September 2002 to August 2011. It is encouraging to see that all basins improved runoff estimation skill with the baseline DA, physically-constrained DA, and 3-day thinning DA relative to the OL under the neutral, positively-biased, and negatively-biased forcing conditions.

In general, monthly runoff in the Moskva Oka (OL bias = 0.46 mm) and lower Volga (OL bias = 0.12 mm) basins were overestimated whereas runoff in the upper Volga (OL bias = −3.8 mm) and Kama basins (OL bias = −5.5 mm) were underestimated under the neutral forcing conditions. This behavior can be explained by the spatial pattern of precipitation as shown in Figure 3f). MERRA-2 (synthetic truth) precipitation was greater than the neutral scenario for the OL run precipitation in the Kama and upper Volga basins, and hence, the runoff from the synthetic truth run was greater than the OL run in the Kama and upper Volga basins.

As a measure of overall hydrograph fit, Nash-Sutcliffe efficiency (NSE) was calculated for all monthly instances when either the synthetic truth or OL/DA runoff estimation was nonzero (Nash & Sutcliffe, 1970). All three DA strategies had greater NSE ($NSE > 0.84$) than the OL ($NSE = 0.82$) for all four sub-basins thereby highlighting the DA skill in runoff estimation beyond simply estimating the mean of the synthetic truth under the neutral forcing conditions. In addition, DA (baseline) had better performance in the Moskva Oka (RMSE = 8.43 mm) and lower Volga (RMSE = 3.55 mm) than the upper Volga (RMSE = 15.6 mm) and Kama basins (RMSE = 15.4 mm).

For the Volga basin runoff estimation, the physically-constrained DA had the best performance in terms of the greatest reduction of RMSE (relative to the OL, 30.3%) compared to the baseline DA (30.2%) and 3-day thinning DA (23.7%) under the positively-biased forcing conditions. This result further illustrated the fact that assimilation worked better when forced with a positive precipitation bias more so than a negative precipitation bias.

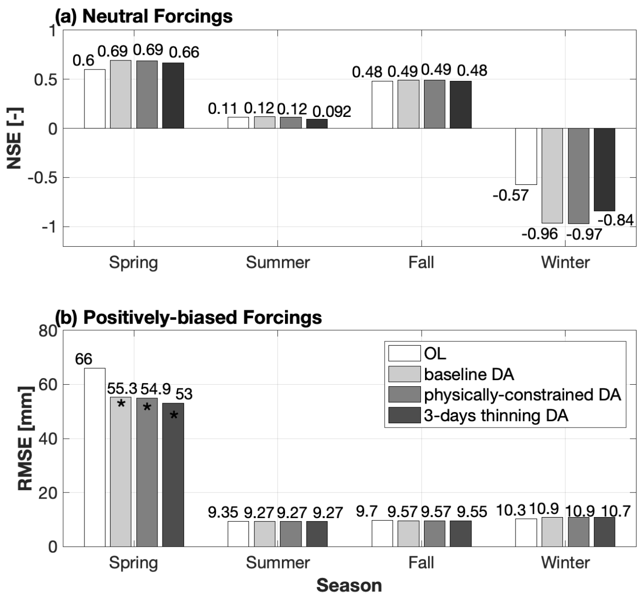

Figure 11. *(a) histogram of the Volga basin monthly runoff Nash-Sutcliffe efficiency (NSE) under the neutral forcing conditions and (b) RMSE under the positively-biased forcing conditions. Bars marked with $*$ indicate which experiment yields statistically significant statistics with a level of significance of 5%.*

It is worth noting that monthly runoff estimation showed a strong seasonality effect. The spring season had the largest magnitude of runoff among the four different seasons due to the snow melt. All three DA strategies yielded better performance in the runoff estimation during the spring season compared to the OL in terms of bigger NSE and smaller RMSE as shown in Figure 11a) and 11b), respectively. Most notably during the positively-biased forcing conditions, DA strategies showed significant improvement over the OL (Figure 11b). These results suggest DA effectively improved the model performance in capturing relatively high runoff.

8 Conclusions

A series of synthetic twin experiments were conducted to explore improvements in the estimation of SWE in the Volga basin based on prescribed precipitation errors. An ensemble Kalman filter (EnKF) was used to merge synthetic PMW brightness temperature spectral differences (ΔT_b) into the NASA Catchment land surface model where well-trained support vector machines served as the observation operator.

The results suggested that simultaneous assimilation of $\Delta T_{b10H-36H}$, $\Delta T_{b10V-36V}$, $\Delta T_{b10H-18H}$, $\Delta T_{b10V-18V}$, $\Delta T_{b18H-36H}$, and $\Delta T_{b18V-36V}$ could degrade SWE estimation due to divergence from the synthetic truth at some experimental locations. One reason for DA degradation was due to simultaneous assimilation of all six ΔT_b channels and the presence of signal saturation during deep snow conditions. To help mitigate this degradation, a physically-constrained approach that used the prior SWE ensemble mean as an indicator was explored. That is, $\Delta T_{b18H-36H}$, $\Delta T_{b18V-36V}$, $\Delta T_{b10H-36H}$, and $\Delta T_{b10V-36V}$ were assimilated during shallow-to-medium snow conditions (i.e., SWE $\leq$ 120 mm), while simultaneously assimilating $\Delta T_{b10H-18H}$, $\Delta T_{b10V-18V}$, $\Delta T_{b10H-36H}$, and $\Delta T_{b10V-36V}$ during medium-to-deep snow conditions (i.e., SWE $>$ 120 mm). The physically-constrained assimilation approach helped improve SWE estimation at some locations but not all.

In addition, a simple data thinning assimilation strategy was explored to further mitigate the high-frequency noise embedded in synthetic AMSR-E ΔT_b observations. That is, the ΔT_b channels were assimilated every 3-, 5-, 7-, 10-, and 15 days rather than daily. The results suggested DA with 3-day data thinning modestly reduced Volga basin averaged bias from -5.5 mm to -2.5 mm under the neutral forcing conditions. $\mathrm{CR}_{2\sigma}$ was slightly increased from 0.25 (baseline DA) to 0.26 (3-day thinning DA).

DA performance under the neutral, positively-biased, and negatively-biased forcing conditions were investigated. The results suggest AMSR-E ΔT_b DA performed the best under the positively-biased conditions in terms of SWE estimation. This highlights a unique facet of snow assimilation that it is easier for the DA system to remove excess mass than to add missing mass. This is, in part, due to the fact that the snow-centric DA update can only happen when snow exists in the land surface model.

The investigation in forested regions highlighted the significant negative impact of dense forest on SWE estimation. This is due to the fact the presence of forest canopy can further alter the PMW ΔT_b signal as measured at the top of the atmosphere. Given the physical limitations of coarse-scale PMW radiometry of snow in forested areas, such scenarios should likely be excluded from the snow DA update in densely-forested areas.

SWE estimation demonstrated a strong seasonality. That is, DA SWE estimates outperformed OL in terms of smaller RMSE, bias, and ubRMSE during the snow accumulation period. However, DA SWE estimates were often degraded during the ablation period due to the presence of liquid water coating the snow grains. The reason for this is that the presence of liquid water within the snowpack elicits a shift in the electromagnetic response from a dry microwave scatter to a wet microwave emitter, and hence, the assumptions implicit in ΔT_b-based remote sensing of snow are regularly violated. The results of runoff estimation also showed a seasonal pattern. Among all four seasons, DA runoff estimates had the best performance relative to the OL during the spring season. This was consistent with the fact that DA SWE estimates were the best during the winter season, and therefore, the runoff derived from snowmelt was vastly improved during the spring season.

Appendix A SVM Training and Prediction

SVM regression served as the observation operator for mapping the geophysical states (e.g., SWE, snow temperature) into observational (i.e., PMW spectral difference) space. Following Forman et al. (2014) and Forman and Reichle (2015), SVM training used a split-sample, jackknifing procedure where observations used for validation were excluded from the training dataset. The training period was from 1 September 2002 to 1 September 2011. A fortnightly (two weeks) training period was selected to best capture seasonal variability while still providing a sufficiently large enough set for training. The inputs to SVM training were four Catchment model states relevant to PMW remote sensing of snow: (1) SWE, (2) snow liquid water content, (3) top-layer soil temperature, and (4) skin temperature. They were selected based on the results of an extensive sensitivity analysis (Xue & Forman, 2017b). The SVM outputs were the synthetic ΔT_b truth, including $\Delta T_{b10H-36H}$, $\Delta T_{b10V-36V}$, $\Delta T_{b10H-18H}$, $\Delta T_{b10V-18V}$, $\Delta T_{b18H-36H}$, and $\Delta T_{b18V-36V}$.

Appendix B Goodness-of-Fit Statistics

Goodness-of-fit statistics used in this study include bias, root mean squared error (RMSE), unbiased root mean squared error (ubRMSE), correlation coefficient (R), Nash-Sutcliffe efficiency (NSE), normalized information contribution (NIC), and containing ration ($CR_{2\sigma}$). The symbol x_{est} denotes the OL or DA ensemble mean and the symbol x_{truth} denotes the synthetic truth. The bias was computed as:

$$bias = \frac{1}{N_t}\sum_{i=1}^{N_t}(x_{est,i} - x_{truth,i}), \tag{B1}$$

where x_i is the state variable (e.g., SWE) at time i and N_t is the sample size over the time period t. The RMSE was computed as:

$$RMSE = \sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}(x_{est,i} - x_{truth,i})^2}, \tag{B2}$$

where x_i is the state variable (e.g., SWE) at time i and N_t is the sample size over the time period t. The ubRMSE was computed as:

$$ubRMSE = \sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}(x_{est,i} - x_{truth,i})^2 - (\overline{x}_{est} - \overline{x}_{truth})^2}, \tag{B3}$$

where $\overline{x}_{est}$ is the time-averaged estimate of the model state variable (e.g., SWE) and $\overline{x}_{truth}$ is the time-averaged synthetic truth. The R was computed as:

$$R = \frac{\sum_{i=1}^{N_t}(x_{est,i} - \overline{x}_{est})(x_{truth,i} - \overline{x}_{truth})}{\sqrt{\sum_{i=1}^{N_t}(x_{est,i} - \overline{x}_{est})^2}\sqrt{\sum_{i=1}^{N_t}(x_{truth,i} - \overline{x}_{truth})^2}} \tag{B4}$$

The NSE was computed as:

$$NSE = 1 - \frac{\sum_{i=1}^{N_t}(x_{truth,i} - x_{est,i})^2}{\sum_{i=1}^{N_t}(x_{truth} - \overline{x}_{est,i})^2} \tag{B5}$$

The NIC for RMSE, NIC_{RMSE}, was computed as

$$NIC_{RMSE} = \frac{RMSE_{OL} - RMSE_{DA}}{RMSE_{OL}} \tag{B6}$$

where the $RMSE_{OL}$ is the OL-based RMSE and $RMSE_{DA}$ is the DA-based RMSE. The containing ratio, $CR_{2\sigma}$, is the number of synthetic truth that fall within the ensemble mean ± 2 times the ensemble standard deviation normalized by the total number of synthetic truth (N_t), and was computed as

$$CR_{2\sigma} = \frac{\sum_{i=1}^{N_t} I[O(x,i)]}{N_t} \tag{B7}$$

where $I[O(x,i)] = 1$ if $x_{min,i} \leq x_{truth,i} \leq x_{max,i}$. In other words, if the synthetic truth at time i, $x_{truth,i}$, is equal to or greater than the minimum of OL or DA ensemble estimates, $x_{min,i}$, and also is less than or equal to the maximum of OL or DA ensemble estimates, $x_{max,i}$, the $I[O(x,i)] = 1$. Otherwise, $I[O(x,i)] = 0$.

Acknowledgments

This work was supported by a grant from the the NASA's GRACE-FO Science Team (NNX16AF17G). Computational resources were provided by the University of Maryland High-Performance Computing Clusters (HPCC). SNOTEL data used in this study were obtained from https://www.wcc.nrcs.usda.gov/snow/. The authors would like to expressly thank Yonghwan Kwon for many constructive comments.

References

Ahmad, J. A., Forman, B. A., & Kwon, Y. (2019). Analyzing machine learning predictions of passive microwave brightness temperature spectral difference over snow-covered terrain in High Mountain Asia. *Frontiers in Earth Science*, *7*, 212.

Armstrong, R. L., Chang, A., Rango, A., & Josberger, E. (1993). Snow depths and grain-size relationships with relevance for passive microwave studies. *Annals of Glaciology*, *17*, 171–176.

Barnett, T. P., Adam, J. C., & Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, *438*(7066), 303–309. doi: 10.1038/nature04141

Brodzik, M. J., Billingsley, B., Haran, T., Raup, B., & Savoie, M. H. (2012). EASE-Grid 2.0: Incremental but Significant Improvements for Earth-Gridded Data Sets. *ISPRS International Journal of Geo-Information*, *1*(3), 32–45. Retrieved from http://www.mdpi.com/2220-9964/1/1/32/ doi: 10.3390/ijgi1010032

Brodzik, R. L. A., & J., M. (2001). Recent northern hemisphere snow extent: A comparison of data derived from visible and microwave satellite sensors. , *28*(19), 3673–3676.

Brucker, L., Royer, A., Picard, G., Langlois, A., & Fily, M. (2011). Hourly simulations of the microwave brightness temperature of seasonal snow in Quebec, Canada, using a coupled snow evolution-emission model. *Remote Sensing of Environment*, *115*(8), 1966–1977. Retrieved from http://dx.doi.org/10.1016/j.rse.2011.03.019 doi: 10.1016/j.rse.2011.03.019

Burgers, G., Jan van Leeuwen, P., & Evensen, G. (1998). Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review*, *126*(6), 1719–1724. Retrieved from http://journals.ametsoc.org/doi/abs/10.1175/1520-0493{\%}281998{\%}29126{\%}3C1719{\%}3AASITEK{\%}3E2.0.CO{\%}3B2 doi: 10.1175/1520-0493(1998)126⟨1719:ASITEK⟩2.0.CO;2

Chang, A. T., Foster, J. L., & Hall, D. K. (1996). Effects of forest on the snow parameters derived from microwave measurements during the BOREAS Winter Field Campaign. *Hydrological Processes*, *10*(12), 1565–1574. doi: 10.1002/(SICI)1099-1085(199612)10:12⟨1565::AID-HYP501⟩3.0.CO;2-5

Clifford, D. (2010). Global estimates of snow water equivalent from passive microwave instruments: History, challenges and future developments. *International Journal of Remote Sensing*, *31*(14), 3707–3726. doi: 10.1080/01431161.2010.483482

Curry, J. a., Schramm, J. L., & Ebert, E. E. (1995). *Sea-Ice Albedo Climate Feedback Mechanism* (Vol. 8) (No. 2). doi: Doi10.1175/1520-0442(1995)008⟨0240:Siacfm⟩2.0.Co;2

De Lannoy, G. J., & Reichle, R. H. (2016). Global assimilation of multiangle and multipolarization SMOS brightness temperature observations into the GEOS-5 catchment land surface model for soil moisture estimation. *Journal of Hydrometeorology*, *17*(2), 669–691. doi: 10.1175/JHM-D-15-0037.1

De Sève, D., Bernier, M., Fortin, J.-P., & Walker, A. (1997). Preliminary analysis of snow microwave radiometry using the SSM/I passive-microwave data: the case of La Grande River watershed (Quebec). *Annals of Glaciology*, *25*, 353–361.

Derksen, C. (2008). The contribution of AMSR-E 18.7 and 10.7 GHz measurements to improved boreal forest snow water equivalent retrievals. *Remote Sensing of Environment*, *112*(5), 2701–2710. doi: 10.1016/j.rse.2008.01.001

Derksen, C., Toose, P., Rees, A., Wang, L., English, M., Walker, A., & Sturm, M. (2010). Development of a tundra-specific snow water equivalent retrieval algorithm for satellite passive microwave data. *Remote Sensing of Environment*, *114*(8), 1699–1709. Retrieved from http://dx.doi.org/10.1016/j.rse.2010.02.019 doi: 10.1016/j.rse.2010.02.019

Derksen, C., Walker, A., & Goodison, B. (2005). Evaluation of passive microwave snow water equivalent retrievals across the boreal forest/tundra transition of western Canada. *Remote Sensing of Environment*, *96*(3-4), 315–327. doi: 10.1016/j.rse.2005.02.014

Dong, J., Walker, J. P., & Houser, P. R. (2005). Factors affecting remotely sensed snow water equivalent uncertainty. *Remote Sensing of Environment*, *97*(1), 68–82. doi: 10.1016/j.rse.2005.04.010

Dong, J., Walker, J. P., Houser, P. R., & Sun, C. (2007). Scanning multichannel microwave radiometer snow water equivalent assimilation. *Journal of Geophysical Research Atmospheres*, *112*(7), 1–16. doi: 10.1029/2006JD007209

Draper, C., & Reichle, R. (2015). The impact of near-surface soil moisture assimilation at subseasonal, seasonal, and inter-annual timescales. *Hydrology and Earth System Sciences*, *19*(12), 4831–4844. doi: 10.5194/hess-19-4831-2015

Durand, M., Kim, E. J., & Margulis, S. A. (2009). Radiance assimilation shows promise for snowpack characterization. *Geophysical Research Letters*, *36*(2), 1–5. doi: 10.1029/2008GL035214

Durand, M., & Margulis, S. A. (2006). Feasibility Test of Multifrequency Radiometric Data Assimilation to Estimate Snow Water Equivalent. *Journal of Hydrometeorology*, *7*(3), 443–457. doi: 10.1175/jhm502.1

Durand, M., & Margulis, S. A. (2007). Correcting first-order errors in snow water equivalent estimates using a multifrequency, multiscale radiometric data assimilation scheme. *Journal of Geophysical Research Atmospheres*, *112*(13), 1–16. doi: 10.1029/2006JD008067

Evensen, G. (2003). The Ensemble Kalman Filter : theoretical formulation and practical implementation. , 343–367. doi: 10.1007/s10236-003-0036-9

Fernandes, R., Zhao, H., Wang, X., Key, J., Qu, X., & Hall, A. (2009). Controls on Northern Hemisphere snow albedo feedback quantified using satellite Earth observations. *Geophysical Research Letters*, *36*(21), 1–6. doi: 10.1029/2009GL040057

Forman, B. A., & Margulis, S. A. (2010). Assimilation of multiresolution radiation products into a downwelling surface radiation model: 2. Posterior ensemble implementation. *Journal of Geophysical Research Atmospheres*, *115*(22), 1–13. doi: 10.1029/2010JD013950

Forman, B. A., & Reichle, R. H. (2015). Using a Support Vector Machine and a Land Surface Model to Estimate Large-Scale Passive Microwave Brightness Temperatures over Snow-Covered Land in North America. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(9), 4431–4441. doi: 10.1109/JSTARS.2014.2325780

Forman, B. A., Reichle, R. H., & Derksen, C. (2014). Estimating Passive Microwave Brightness Temperature Over Snow-Covered Land in North America Using a Land Surface Model and an Artificial Neural Network. *Geoscience and Remote Sensing, IEEE Transactions on*, *52*(1), 235–248. doi: 10.1109/TGRS.2013.2237913

Forman, B. A., Reichle, R. H., & Rodell, M. (2012). Assimilation of terrestrial water storage from GRACE in a snow-dominated basin. *Water Resources Research*, *48*(1), 1–14. doi: 10.1029/2011WR011239

Forman, B. A., & Xue, Y. (2016). Machine learning predictions of passive microwave brightness temperature over snow-covered land using the special sensor microwave imager (SSM/I). *Physical Geography*, *3646*(February), 1–21. Retrieved from `https://www.tandfonline.com/doi/full/10.1080/02723646.2016.1236606` doi: 10.1080/02723646.2016.1236606

Friedl, M. A., McIver, D. K., Hodges, J. C., Zhang, X. Y., Muchoney, D., Strahler, A. H., . . . Schaaf, C. (2002). Global land cover mapping from MODIS: Algorithms and early results. *Remote Sensing of Environment*, *83*(1-2), 287–302. doi: 10.1016/S0034-4257(02)00078-0

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., . . .

Zhao, B. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. doi: 10.1175/JCLI-D-16-0758.1

Gelb, A. (1974). *Applied optimal estimation.* MIT press.

Girotto, M., De Lannoy, G. J., Reichle, R. H., & Rodell, M. (2016). Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model. *Water Resources Research*, *52*(5), 4164–4183. doi: 10.1002/2015WR018417

Golosov, V., & Belyaev, V. (2016). *The Volga River Basin Report.* UNESCO International Sediment Initiative. http://www. irtces. org/isi/isi˜. . . .

Günther, D., Marke, T., Essery, R., & Strasser, U. (2019). Uncertainties in Snowpack Simulations—Assessing the Impact of Model Structure, Parameter Choice, and Forcing Data Error on Point-Scale Energy Balance Snow Model Performance. *Water Resources Research*, *55*(4), 2779–2800. doi: 10.1029/2018WR023403

Houborg, R., Rodell, M., Li, B., Reichle, R., & Zaitchik, B. F. (2012). Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resources Research*, *48*(7). doi: 10.1029/2011WR011291

Houtekamer, P. L., & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, *126*(3), 796–811.

Kelly, R. E., Chang, A. T., Tsang, L., & Foster, J. L. (2003). A prototype AMSR-E global snow area and snow depth algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(2 PART 1), 230–242. doi: 10.1109/TGRS.2003.809118

Keppenne, C. L. (2002). Data Assimilation into a Primitive-Equation Model with a Parallel Ensemble Kalman Filter. *Monthly Weather Review*, *128*(6), 1971–1981. doi: 10.1175/1520-0493(2000)128⟨1971:daiape⟩2.0.co;2

Kim, E. J., & England, A. W. (2003). A yearlong comparison of plot-scale and satellite footprint-scale 19 and 37 GHz brightness of the Alaskan North Slope. *Journal of Geophysical Research D: Atmospheres*, *108*(13). doi: 10.1029/2002JD002393

Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., & Kumar, P. (2000). A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *Journal of Geophysical Research*, *105*(D20), 24809. Retrieved from `http://www.scopus.com/inward/record.url?eid=2-s2.0-0034535097{\&}partnerID=tZOtx3y1{\%}5Cnhttp://doi.wiley.com/10.1029/2000JD900327` doi: 10.1029/2000JD900327

Kukkonen, J., Olsson, T., Schultz, D. M., Baklanov, A., Klein, T., Miranda, A. I., . . . Eben, K. (2012). A review of operational, regional-scale, chemical weather forecasting models in Europe. *Atmospheric Chemistry and Physics*, *12*(1), 1–87. doi: 10.5194/acp-12-1-2012

Kumar, S. V., Dong, J., Peters-Lidard, C. D., Mocko, D., & Gómez, B. (2016). Role of forcing uncertainty and model error background characterization in snow data assimilation. *Hydrology and Earth System Sciences Discussions*, 1–24. Retrieved from `http://www.hydrol-earth-syst-sci-discuss.net/hess-2016-581/` doi: 10.5194/hess-2016-581

Kumar, S. V., Peters-Lidard, C. D., Mocko, D., Reichle, R., Liu, Y., Arsenault, K. R., . . . Cosh, M. (2014). Assimilation of Remotely Sensed Soil Moisture and Snow Depth Retrievals for Drought Estimation. *Journal of Hydrometeorology*, *15*(6), 2446–2469. Retrieved from `http://journals.ametsoc.org/doi/abs/10.1175/JHM-D-13-0132.1` doi: 10.1175/JHM-D-13-0132.1

Kumar, S. V., Reichle, R. H., Koster, R. D., Crow, W. T., & Peters-Lidard, C. D. (2009). Role of Subsurface Physics in the Assimilation of Surface Soil Moisture Observations. *Journal of Hydrometeorology*, *10*(6), 1534–1547. Retrieved from http://journals.ametsoc.org/doi/abs/10.1175/2009JHM1134.1 doi: 10.1175/2009JHM1134.1

Kumar, S. V., Zaitchik, B. F., Peters-Lidard, C. D., Rodell, M., Reichle, R., Li, B., ... Ek, M. (2016). Assimilation of Gridded GRACE Terrestrial Water Storage Estimates in the North American Land Data Assimilation System. *Journal of Hydrometeorology*, *17*(7), 1951–1972. Retrieved from http://journals.ametsoc.org/doi/10.1175/JHM-D-15-0157.1 doi: 10.1175/JHM-D-15-0157.1

Kwon, Y., Forman, B. A., Ahmad, J. A., Kumar, S. V., & Yoon, Y. (2019). Exploring the Utility of Machine Learning-Based Passive Microwave Brightness Temperature Data Assimilation over Terrestrial Snow in High Mountain Asia. *Remote Sensing*, *11*(19), 2265.

Kwon, Y., Yang, Z. L., Hoar, T. J., & Toure, A. M. (2017). Improving the radiance assimilation performance in estimating snow water storage across snow and land-cover types in North America. *Journal of Hydrometeorology*, *18*(3), 651–668. doi: 10.1175/JHM-D-16-0102.1

Kwon, Y., Yang, Z.-L., Zhao, L., Hoar, T. J., Toure, A. M., & Rodell, M. (2016). Estimating Snow Water Storage in North America Using CLM4, DART, and Snow Radiance Data Assimilation. *Journal of Hydrometeorology*, *17*(11), 2853–2874. Retrieved from http://journals.ametsoc.org/doi/10.1175/JHM-D-16-0028.1 doi: 10.1175/JHM-D-16-0028.1

Liston, G. E. (1999). Interrelationships among Snow Distribution, Snowmelt, and Snow Cover Depletion: Implications for Atmospheric, Hydrologic, and Ecologic Modeling. *Journal of Applied Meteorology*, *38*(10), 1474–1487. Retrieved from http://journals.ametsoc.org/doi/abs/10.1175/1520-0450(1999)038{\%}3C1474:IASDSA{\%}3E2.0.CO;2 doi: 10.1175/1520-0450(1999)038⟨1474:IASDSA⟩2.0.CO;2

Lynch-Stieglitz, M. (1994). The development and validation of a simple snow model for the GISS GCM. *Journal of Climate*, *7*(12), 1842–1855.

Mätzler, C. (1994). Passive microwave signatures of landscapes in winter. *Meteorology and Atmospheric Physics*, *54*(1-4), 241–260. doi: 10.1007/BF01030063

Matzler, C., Schanda, E., & Good, W. (1982). Towards the definition of optimum sensor specifications for microwave remote sensing of snow. *IEEE Transactions on Geoscience and Remote Sensing*(1), 57–66.

McLaughlin, D. (2002). An integrated approach to hydrologic data assimilation: Interpolation, smoothing, and filtering. *Advances in Water Resources*, *25*(8-12), 1275–1286. doi: 10.1016/S0309-1708(02)00055-6

Nash, J., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290.

Ogata, K., & Yang, Y. (2002). *Modern control engineering* (Vol. 4). London.

Pulliainen, J. (2006). Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations. *Remote Sensing of Environment*, *101*(2), 257–269. doi: 10.1016/j.rse.2006.01.002

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., ... Gutmann, E. (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

Rees, A., Lemmetyinen, J., Derksen, C., Pulliainen, J., & English, M. (2010). Observed and modelled effects of ice lens formation on passive microwave brightness temperatures over snow covered tundra. *Remote Sensing of Environment*, *114*(1), 116–126. Retrieved from http://dx.doi.org/10.1016/j.rse.2009.08.013 doi: 10.1016/j.rse.2009.08.013

Reichle, R., McLaughlin, D. B., & Entekhabi, D. (2002). Hydrologic Data Assimilation with the Ensemble Kalman Filter. *Journal of Hydrometeorology*, *130*, 103–114. doi: 10.1175/1520-0493(2002)130⟨0103:HDAWTE⟩2.0.CO;2

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

Reichle, R. H., Crow, W. T., & Keppenne, C. L. (2008). An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resources Research*, *44*(3), 1–13. doi: 10.1029/2007WR006357

Reichle, R. H., Draper, C. S., Liu, Q., Girotto, M., Mahanama, S. P., Koster, R. D., & De Lannoy, G. J. (2017). Assessment of MERRA-2 land surface hydrology estimates. *Journal of Climate*, *30*(8), 2937–2960. doi: 10.1175/JCLI-D-16-0720.1

Reichle, R. H., Entekhabi, D., Mclaughlin, D. B., & Parsons, R. M. (2001). Downscaling of radio brightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach. *Water Resources*, *37*(9), 2353–2364. doi: 10.1029/2001WR000475

Reichle, R. H., & Koster, R. D. (2002). Land data assimilation with the ensemble Kalman filter: assessing model error parameters using innovations. *Developments in Water Science*, *47*, 1387–1394.

Reichle, R. H., & Koster, R. D. (2003). Assessing the Impact of Horizontal Error Correlations in Background Fields on Soil Moisture Estimation. *Journal of Hydrometeorology*, *4*(6), 1229–1242. doi: 10.1175/1525-7541(2003)004⟨1229:ATIOHE⟩2.0.CO;2

Reichle, R. H., Kumar, S. V., Mahanama, S. P. P., Koster, R. D., & Liu, Q. (2010). Assimilation of Satellite-Derived Skin Temperature Observations into Land Surface Models. *Journal of Hydrometeorology*, *11*(5), 1103–1122. Retrieved from http://journals.ametsoc.org/doi/abs/10.1175/2010JHM1262.1 doi: 10.1175/2010JHM1262.1

Reichle, R. H., Walker, J. P., Koster, R. D., & Houser, P. R. (2002). Extended versus Ensemble Kalman Filtering for Land Data Assimilation. *Journal of Hydrometeorology*, *3*(6), 728–740. Retrieved from http://journals.ametsoc.org/doi/abs/10.1175/1525-7541{\%}282002{\%}29003{\%}3C0728{\%}3AEVEKFF{\%}3E2.0.CO{\%}3B2 doi: 10.1175/1525-7541(2002)003⟨0728:EVEKFF⟩2.0.CO;2

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., . . . Toll, D. (2004). The Global Land Data Assimilation System. *Bull. Am. Meteorol. Soc.*, *85*(March), 381–394. doi: 10.1175/BAMS-85-3-381

Roy, A., Picard, G., Royer, A., Montpetit, B., Dupont, F., Langlois, A., . . . Champollion, N. (2013). Brightness temperature simulations of the Canadian seasonal snowpack driven by measurements of the snow specific surface area. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(9), 4692–4704. doi: 10.1109/TGRS.2012.2235842

Savoie, M. H., Armstrong, R. L., Brodzik, M. J., & Wang, J. R. (2009). Atmospheric corrections for improved satellite passive microwave snow cover retrievals over the Tibet Plateau. *Remote Sensing of Environment*, *113*(12), 2661–2669. Retrieved from http://dx.doi.org/10.1016/j.rse.2009.08.006 doi: 10.1016/j.rse.2009.08.006

Schanda, E., Mätzler, C., & Künzi, K. (1983). Microwave remote sensing of snow cover. *International Journal of Remote Sensing*, *4*(1), 149–158. doi: 10.1080/01431168308948536

Seve, D. A. N. I. D. E., Tier, I. M. E. B. E. R., Fortln, I. J.-p., & Valker, I. A. N. (1986). Prelitninary analysis of snow tnicrowave radiotnetry using the SSMjI passive-microwave data : the case of La Grande River watershed (Quebec). (1 994), 353–361.

Sidorchuk, A. Y., Panin, A. V., & Borisova, O. K. (2009). Morphology of river channels and surface runoff in the Volga River basin (East European Plain) during the Late Glacial period. *Geomorphology*, *113*(3-4), 137–157. Retrieved from http://dx.doi.org/10.1016/j.geomorph.2009.03.007 doi: 10.1016/j.geomorph.2009.03.007

Stieglitz, M., Ducharne, A., Koster, R., & Suarez, M. (2001). The Impact of Detailed Snow Physics on the Simulation of Snow Cover and Subsurface Thermodynamics at Continental Scales. *Journal of Hydrometeorology*, *2*(3), 228–242. doi: 10.1175/1525-7541(2001)002⟨0228:TIODSP⟩2.0.CO;2

Walker, A. E., & Goodison, B. E. (1993). Discrimination of a wet snow cover using passive microwave satellite data. *Annals of Glaciology*, *17*(1989), 307–311.

Wilheit, T., Chang, A. T., & Milman, A. S. (1980). Atmospheric corrections to passive microwave observations of the ocean. *Boundary-Layer Meteorology*, *18*(1), 65–77. doi: 10.1007/BF00117911

Xue, Y., & Forman, B. A. (2015). Comparison of passive microwave brightness temperature prediction sensitivities over snow-covered land in North America using machine learning algorithms and the Advanced Microwave Scanning Radiometer. *Remote Sensing of Environment*, *170*, 153–165. Retrieved from http://dx.doi.org/10.1016/j.rse.2015.09.009 doi: 10.1016/j.rse.2015.09.009

Xue, Y., & Forman, B. A. (2017a). Atmospheric and Forest Decoupling of Passive Microwave Brightness Temperature Observations Over Snow-Covered Terrain in North America. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(7), 3172–3189. doi: 10.1109/JSTARS.2016.2614158

Xue, Y., & Forman, B. A. (2017b). Integration of satellite-based passive microwave brightness temperature observations and an ensemble-based land data assimilation framework to improve snow estimation in forested regions. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2017-July*, 311–314. doi: 10.1109/IGARSS.2017.8126958

Xue, Y., Forman, B. A., & Reichle, R. H. (2018). Estimating snow mass in North America through assimilation of AMSR-E brightness temperature observations using the Catchment land surface model and support vector machines. *Water Resources Research*, 1–22. Retrieved from http://doi.wiley.com/10.1029/2017WR022219 doi: 10.1029/2017WR022219

Zaitchik, B. F., Rodell, M., & Reichle, R. H. (2008). Assimilation of GRACE Terrestrial Water Storage Data into a Land Surface Model: Results for the Mississippi River Basin. *Journal of Hydrometeorology*, *9*(3), 535–548. doi: 10.1175/2007JHM951.1