Hydrology and
Earth System
Sciences

Assessment of a multiresolution snow reanalysis framework: a multidecadal reanalysis case over the upper Yampa River basin, Colorado

Elisabeth Baldo and Steven A. Margulis

Department of Civil and Environmental Engineering, University of California,
Los Angeles, California, USA

Correspondence: Elisabeth Baldo (ebaldo@g.ucla.edu)

Abstract. A multiresolution (MR) approach was successfully implemented in the context of a data assimilation (DA) framework to efficiently estimate snow water equivalent (SWE) over a large head water catchment in the Colorado River basin (CRB), while decreasing computational constraints by 60 %. A total of 31 years of fractional snow cover area (fSCA) images derived from Landsat TM, ETM+, and OLI sensor measurements were assimilated to generate two SWE reanalysis datasets, a baseline case at a uniform 90 m spatial resolution and another using the MR approach. A comparison of the two showed negligible differences in terms of snow accumulation, melt, and timing for the posterior estimates (in terms of both ensemble median and coefficient of variation). The MR approach underestimated the baseline peak SWE by less than 2 % and underestimated day of peak and duration of the accumulation season by a day on average. The largest differences were, by construct, limited primarily to areas of low complexity, where shallow snowpacks tend to exist. The MR approach should allow for more computationally efficient implementations of snow data assimilation applications over large-scale mountain ranges, with accuracies similar to those that would be obtained using ∼100 m simulations. Such uniform resolution applications are generally infeasible due to the computationally expensive nature of ensemble-based DA frameworks.

1 Introduction

Spatial resolutions of 100 m or less are more commonly being recommended when using land surface models (Wood et al., 2011; Bierkens et al., 2015; Beven et al., 2015), especially when trying to capture the heterogeneity of snowpack states in montane regions (Clark et al., 2011; Winstral et al., 2014). Previous work using hydrologic response units (HRUs; Beven and Kirby, 1979; US Geological Survey et al., 1983; Sivapalan et al., 1987; Chaney et al., 2016), or triangulated irregular networks (TINs; Tucker et al., 2001; Vivoni et al., 2004; Mascaro et al., 2015) showed that simulating in a "one size fits all" (uniform grid) approach is not only computationally expensive, but also suboptimal since only small subsets of watersheds actually require being resolved at fine spatial resolutions. Along these lines, Baldo and Margulis (2017) developed a multiresolution (MR) scheme for raster-based models and tested it in the context of deterministic snow modeling. By adapting the grid size to the physiographic complexity of the terrain, runtime and storage needs were cut in half while preserving the accuracy of a 90 m baseline simulation.

Deterministic forward modeling itself, even at high resolution, is often insufficient due to errors in model inputs (most notably precipitation) that are poorly characterized in montane regions. In lieu of deterministic modeling techniques, ensemble-based data assimilation (DA) methods are now frequently used to estimate snow states (Clark et al., 2006; Andreadis and Lettenmaier, 2006; Su et al., 2008; De Lannoy et al., 2010; Liu et al., 2013; Arsenault et al., 2013; Girotto et al., 2014; Margulis et al., 2015; Kumar et al., 2015). The

advantage of such approaches is to offer spatially and temporally continuous estimates, while also providing a measure of their uncertainty. However, due to their ensemble nature, such methods can be extremely expensive to run at high spatial resolutions, which at least partly explains why many of the large-scale studies cited above simulate snow processes at resolutions on the order of 1 km or greater. Simulating at these scales can solve the computational issue, but inherently sacrifices valuable information related to subgrid heterogeneities in montane regions. This is undesirable since relevant remote sensing data streams that can act as model constraints (e.g., lidar, Landsat, MODIS) are available at higher resolution (on scales of meters to hundreds of meters).

The recently developed 30+ year Sierra Nevada and Andes snow reanalysis datasets by Margulis et al. (2016) and Cortés and Margulis (2017) successfully leveraged high-resolution Landsat data using a data assimilation framework applied at uniform resolutions of 90 and 180 m respectively. For these regional-scale domains, this resulted in ∼ 6 million and 5.5 million simulation pixels respectively, which were run in the context of a 100-member ensemble. For reference, given that Northern Hemisphere snow covered area is on the order of 8 million km^2 (Derksen and Brown, 2012), using a 100 m resolution would require the simulation of 8 billion pixels, an increase of nearly 4 orders of magnitude relative to the combined effort for the Sierra Nevada and Andes. Hence, extending these ensemble-based reanalysis methods to much larger scales using a uniform resolution on the order of 100 m is computationally prohibitive. Taking advantage of a MR approach to significantly reduce computational constraints might therefore greatly benefit ensemble-based DA frameworks and allow for applications at much larger scales. This paper aims to test the performance of the MR approach from Baldo and Margulis (2017) in the context of a probabilistic DA framework (Margulis et al., 2015).

The MR approach as applied by Baldo and Margulis (2017) only impacted prior (model-based) snow estimates as a result of aggregation of model inputs. In the context of the DA framework used by Margulis et al. (2016) and Cortés et al. (2016), the MR approach will also coarsen the fractional snow cover area (fSCA) observations derived from raw Landsat images (Cortés et al., 2014), which can potentially additionally impact the accuracy of the posterior snow state estimates. We hypothesize that this additional source of aggregation error will have minimal impact on the posterior estimates because it is expected a priori that the heterogeneity of fSCA in areas of low complexity will be minimal. Areas of high physiographic complexity typically correspond to areas of spatially heterogeneous snow accumulation and melt patterns, which drive fSCA evolution. Applying the MR approach to fSCA observations will therefore coarsen regions of the image where fSCA is most likely homogeneous and refine regions where fSCA is most likely heterogeneous, and

should therefore mitigate the impact of reducing the number of pixels on the reanalysis accuracy.

In this paper, a high-resolution (90 m) uniform grid baseline snow water equivalent (SWE) reanalysis dataset was compared to one derived using the MR scheme to address the following questions: (1) how does the MR approach impact the assimilated fSCA observations? (2) How well does the MR approach perform in estimating the central tendency (i.e., ensemble median) of the posterior snow state distribution in space and time? (3) How well does the MR approach perform in estimating the uncertainty of the posterior snow state distribution in space and time?

The rest of the paper is organized as follows: Sect. 2 illustrates the study area and the methodology used in this work, Sect. 3 compares the MR approach to the 90 m baseline case in order to answer the questions listed above, and, finally, Sect. 4 summarizes the key points of this work.

2 Methodology

2.1 Study area

In order to maintain consistency with the work of Baldo and Margulis (2017), this study also used the upper Yampa River basin (UYRB, outlined in black in Fig. 1) as a representative test domain of the Colorado River basin (CRB). The CRB is large (6770 km^2) and snow-dominated, which makes it a critical source of fresh water for the 20 million people living downstream (Christensen et al., 2004).

In this study, the physiographic complexity metric (CM) was calculated for each 90 m pixel i across the CRB (Fig. 1) following the approach described in Baldo and Margulis (2017):

$$\mathrm{CM}_i = \hat{\sigma_{Z_i}} + \hat{\sigma_{\mathrm{NI}_i}} + \hat{\sigma_{\mathrm{fveg}_i}}, \tag{1}$$

where the normalized standard deviations of elevation ($\hat{\sigma_{Z_i}}$), and northness index ($\hat{\sigma_{\mathrm{NI}_i}}$, Molotch et al., 2004) were derived from the advanced spaceborne thermal emission and reflection (ASTER) global digital elevation model (DEM, JPL, 2009), and the normalized standard deviation of forested fraction ($\hat{\sigma_{\mathrm{fveg}_i}}$) was derived from the National Land Cover Dataset (NLCD, Homer et al., 2007). Across the CRB, CM varies from 0 (bare and flat areas) to over 0.8 (steep and forested areas), with the UYRB sampling a similar range of complexity (Fig. 1).

2.2 Multiresolution approach

The MR algorithm begins with a pre-defined set of resolutions across which a raster-based model implementation will be applied. The finest baseline resolution is chosen to correspond to that deemed important for representing processes in high-complexity areas of a basin. Factor 2 multiples of a 90 m baseline up to 720 m are chosen in this study as the
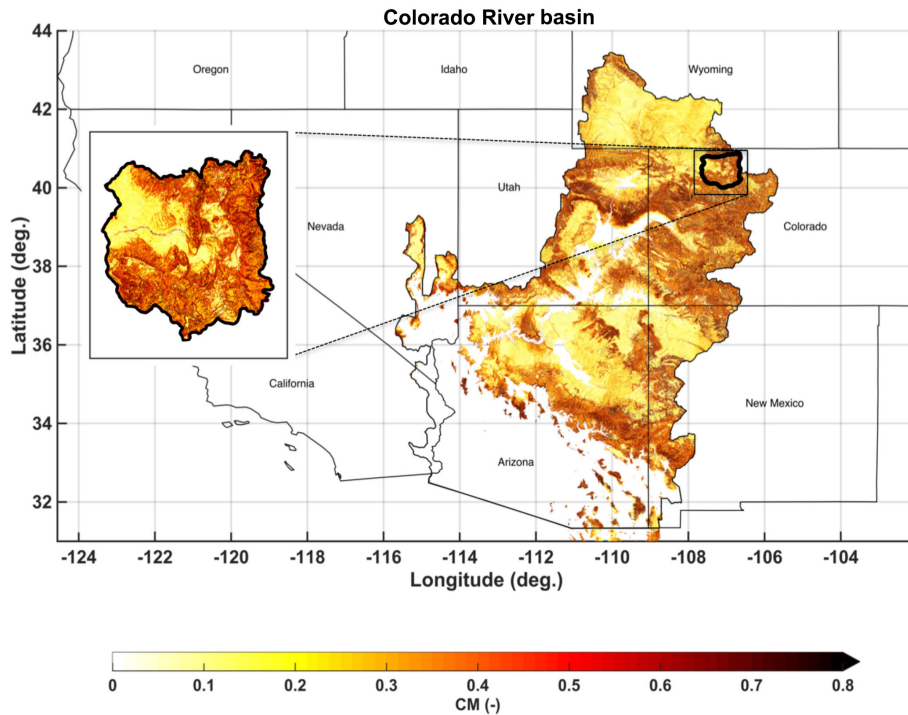
Figure 1. Complexity metric (CM) map of the Colorado River basin (CRB) with the upper Yampa River basin (UYRB) outlined in black and displayed in more detail in the subpanel.
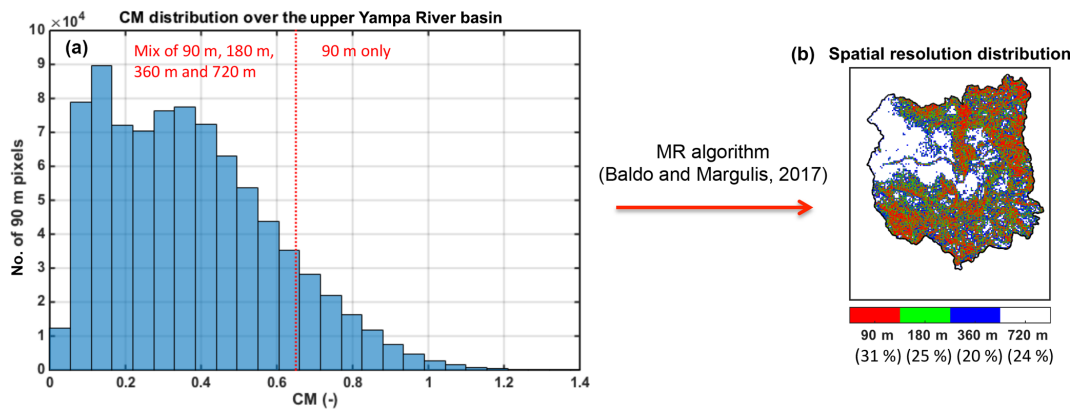
Figure 2. (a) Complexity metric distribution for the upper Yampa River basin. The choice of the maximum threshold CM_{max} of 0.65 represented as the red vertical line leads to **(b)** the spatial resolution distribution map.

specific set of resolutions. The final spatial distribution of resolutions depends on the choice of a maximum CM threshold (CM_{max}), above which pixels are simulated at the finest resolution and below which pixels are simulated at a mix of coarser resolutions. The threshold is chosen based on available computational resources for an application. In this study we chose to use a CM_{max} of 0.65, which corresponds to the 90th percentile of the CRB CM values (Fig. 2). Based on the benchmarking tests performed by Baldo and Margulis (2017), such a threshold leads to a decrease in total pixel numbers on the order of 60 to 70 %, which corresponds to

reasonable computational costs for a full CRB snow reanalysis.

By construct, all of the UYRB pixels with a CM value larger than 0.65 were resolved at the baseline spatial resolution of 90 m, while the less complex ones were assigned either 720, 360, 180 or 90 m by the MR algorithm developed by Baldo and Margulis (2017). The majority of the 720 m pixels are located in the northwestern part of the basin (Fig. 2) corresponding to flat and grassy areas. Modeling almost a quarter of the pixels at this coarse resolution represents the main source of computational savings, while minimizing the

impact on snow accumulation and melt patterns given the homogeneous physiography of the terrain. The remaining low CM pixels were assigned either 360 or 180 m depending on the complexity of their neighbors. In terms of the most complex pixels, 31 % of the pixels are resolved at 90 m in order to preserve the accuracy of SWE estimates. In UYRB, these pixels tend to be located at higher elevations, where the terrain is rugged and densely forested as described in Baldo and Margulis (2017) (Fig. 2).

2.3 SWE reanalysis framework

2.3.1 Model framework and forcings

The modeling setup used in this study is the same as described in Margulis et al. (2016). The Simplified Simple Biosphere (SSiB) model developed by Xue et al. (1991), coupled with a three-layer snow and atmosphere soil transfer (SAST) model (Sun and Xue, 2001; Xue et al., 2003) was used as the land surface model (LSM) to represent the interactions between the atmosphere, vegetation, and snow. A snow depletion curve (SDC) (Liston, 2004) was used to represent the subgrid heterogeneity in SWE and the resulting fSCA. The coupled LSM–SDC generates time series of SWE and fSCA as a function of the subgrid coefficient of variation (CV) and pixel-averaged cumulative snowfall and snowmelt.

The static inputs required by the LSM are latitude, longitude, elevation, slope, and aspect, which were derived from the ASTER DEM (JPL, 2009), as well as land cover derived from the NLCD (Homer et al., 2007). The static inputs were aggregated from their original 30 m resolution to the model resolution (either 90 m for the baseline or a mix of 90, 180, 360, and 720 m for the MR case). The dynamic meteorological forcings were obtained from the Phase 2 North American Land Data Assimilation System (NLDAS-2, Cosgrove et al., 2003; Xia et al., 2012) hourly forcing dataset. NLDAS-2 variables include precipitation, incident shortwave radiation, near-surface air temperature, humidity, wind speed, and pressure at a coarse spatial resolution of 1/8°. The NLDAS-2 forcings were downscaled to the model resolution using topographic correction methods that have been previously applied over the Sierra Nevada and the Andes (Girotto et al., 2014; Girotto et al., 2014a; Margulis et al., 2016; Cortés et al., 2016) as well as the upper Yampa in Baldo and Margulis (2017). Lapse rates of 6.5 and 4.1°K km^{-1} were used for air temperature and dew-point temperature respectively. Downscaling approaches for atmospheric pressure, specific humidity, and the incoming longwave and shortwave radiation fluxes are explained in detail in Girotto et al. (2014) (Appendix A). The downscaling is not deterministic, but also incorporates a priori uncertainty in the forcings (Girotto et al., 2014; Appendix A). It is important to note that the precipitation is not downscaled a priori, but treated as an uncertain random variable following a lognormal distribution with a mean of 2.25 and a standard deviation of 0.5 that is then implicitly downscaled and updated as part of the data assimilation framework.

2.3.2 Assimilation of Landsat-based fractional snow cover area using a particle batch smoother

The probabilistic DA framework used in this study is referred to as the Particle Batch Smoother, or PBS, and was developed by Margulis et al. (2015) in order to improve the probabilistic reanalysis framework used previously for SWE reanalysis in Durand et al. (2008) and Girotto et al. (2014); Girotto et al. (2014a). The coupled LSM–SDC provides a prior ensemble estimate for all snow states and fluxes based on the specified input uncertainty and its propagation through the model. The prior ensemble treats each replicate as an equally likely (equal weight) realization based on the postulated input uncertainty. The goal of the PBS approach is to optimally weight the different uncertainty sources coming from the meteorological forcing and fSCA retrievals in order to generate posterior snow estimates. Specifically, the reanalysis step is applied to a batch of the full set of fSCA measurements (retrospectively) over the water year. A likelihood function updates the prior weights whereby the posterior weights can be used to determine the probability density function or moments (i.e., mean, median, variance, interquartile range, etc.) of any of the snow states and fluxes. The mathematical framework is presented in detail in Margulis et al. (2015).

Landsat-5 thematic mapper (TM), Landsat-7 enhanced thematic mapper (ETM+), and Landsat-8 operational land imager (OLI) images from water year 1985 to 2015 were used to calculate fSCA and fractional vegetation cover over each pixel. For a given sensor, measurements are available every 16 days at a spatial resolution of 30 m, and only clear-sky images were processed to obtain fSCA. The raw data consist of multispectral top-of-atmosphere radiance measurements that are transformed into top-of-atmosphere reflectance before being atmospherically corrected. The spectral unmixing algorithm validated by Cortés et al. (2014) and based on Painter et al. (2009) then retrieves the fraction and type of constituent (snow, vegetation, or bare rock or soil) within each pixel through a least-square-error optimization. The linear unmixing model estimates reflectances from each constituent and selects the combination of constituents leading to the lowest root mean square error (RMSE) between the modeled reflectance and a library of snow reflectances that have previously been calculated for different combinations of constituents within each pixel. The validation of the algorithm by Cortés et al. (2014) showed an fSCA retrieval error of approximately 15 %. The vegetation cover fraction (fVEG) was also retrieved from the spectral unmixing algorithm and annually averaged and used within the LSM–SDC. The fVEG derived from Landsat observations was chosen over the static NLCD for use in the LSM–SDC model to allow for interannual variability and because it is

also, by construct, more consistent with the fSCA observations used in the assimilation step. Similar to the static input data, the fSCA and fVEG images at 30 m were then aggregated to either 90 m for the baseline case, or a mix of 90, 180, 360, and 720 m for the MR case.

2.3.3 Verification of posterior SWE estimates

A posterior set of SWE reanalysis estimates was first generated for 31 years (WY 1985–WY 2015) at the baseline resolution of 90 m, and compared to in situ measurements to assess its accuracy. A total of 203 peak SWE measurements from six SNOTEL stations and 1421 monthly manually sampled SWE from seven snow courses were used. Not all locations have full records for the full period, with two snow pillows or courses starting in 1986 and one in 1998. All snow pillows are collocated with snow courses and station no. 5 is a snow course only (Fig. 3). All in situ observations are taken at high elevations, between 2500 and 3200 m, in densely forested clearings; some representativeness errors are therefore expected when compared to grid-averaged SWE estimates.

The prior SWE estimates are highly uncertain by construct and overestimated in situ observations from both snow courses and pillow (Fig. 4). Prior estimates had a mean difference (MD) of 30 cm with a root mean square difference (RMSD) of 41 cm for snow courses, and a MD of 43 cm with a RMSD of 51 cm for snow pillows. Both showed a similar correlation coefficient (R^2) of 0.86. Note that, based on previous work (Luo et al., 2003; Girotto et al., 2014), the NLDAS-2 precipitation was assumed biased and therefore bias-corrected using the prior distribution (using a mean of 2.25 as indicated above). The fact that the prior SWE overestimates in situ data is an indication that there is likely an overcorrection in the prior precipitation (at least at these sites). In contrast, the reanalysis generated posterior SWE estimates that are much more consistent with the in situ data, are extremely well correlated to in situ measurement, and show limited mean differences. The MD is less than 2 cm for snow courses and less than 5 cm for snow pillows, with RMSD of 10 cm and R^2 higher than 0.95 for both. The small differences observed may be partly explained by undercatch problems with SNOTEL pillows measurements, and also by the fact that in situ SWE measurements are usually made in easily accessible areas such as clearings and therefore not fully representative of the collocated 90 m pixel-average values. The difference in errors between the prior and posterior is primarily indicative of the data assimilation method of properly selecting ensemble members with precipitation forcing that is consistent with the fSCA observations. Based on the comparison with in situ data, the posterior SWE estimates generated at 90 m can be considered to be an accurate representation of the true underlying SWE for the UYRB and are thus used as a baseline throughout.
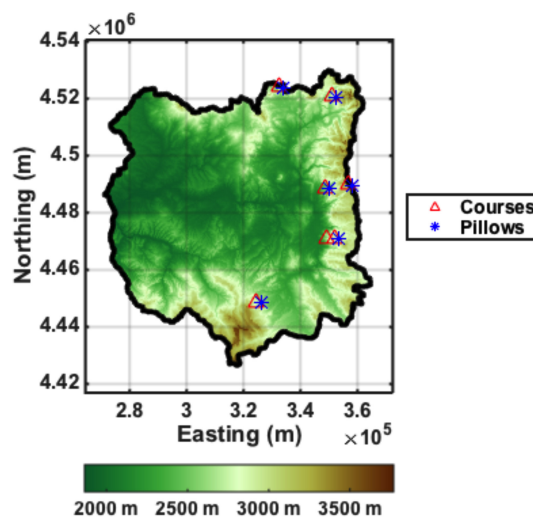
Figure 3. Elevation map of the upper Yampa River basin with the location of the seven snow courses shown in red and the location of the six snow pillows shown in blue.

3 Performance of the MR SWE reanalysis compared to the 90 m baseline

As shown previously in Sect. 2.3.3, performing a SWE reanalysis at 90 m yields an accurate reference solution for our test basin. However, such a simulation is very expensive in terms of computational resources. Modeling the basin uniformly at 90 m meant running almost 840 000 pixels with an ensemble size of 100 replicates, which took over a month on the UCLA computer cluster and required 850G of space to store the resulting outputs. By contrast, the MR approach decreased the number of pixels and storage needed by 59 %. Since pixels are simulated independently from each other, they are run in parallel, which is why runtime also decreased by 59 % and took less than 2 weeks. Knowing that the MR SWE reanalysis can decrease computational constraints by a factor of 2 or more, the following section aims to assess its performance in terms of accuracy.

3.1 Impact of the MR approach on the assimilated fSCA observations

The MR modeling approach as applied previously in Baldo and Margulis (2017) impacts the prior snow simulations, but in the context of a DA (reanalysis) framework as done herein, it also coarsens the fSCA observations that provide the key constraint that generates the posterior estimates. Assessing the difference between the baseline and MR fSCA is therefore crucial to understanding the full effect of the MR approach on the data assimilation step.

In order to first understand the seasonality of the fSCA differences, all observations were binned by month and averaged over the 31 years of record (Fig. 5a). The differences are negligible between the 90 m baseline and the MR case
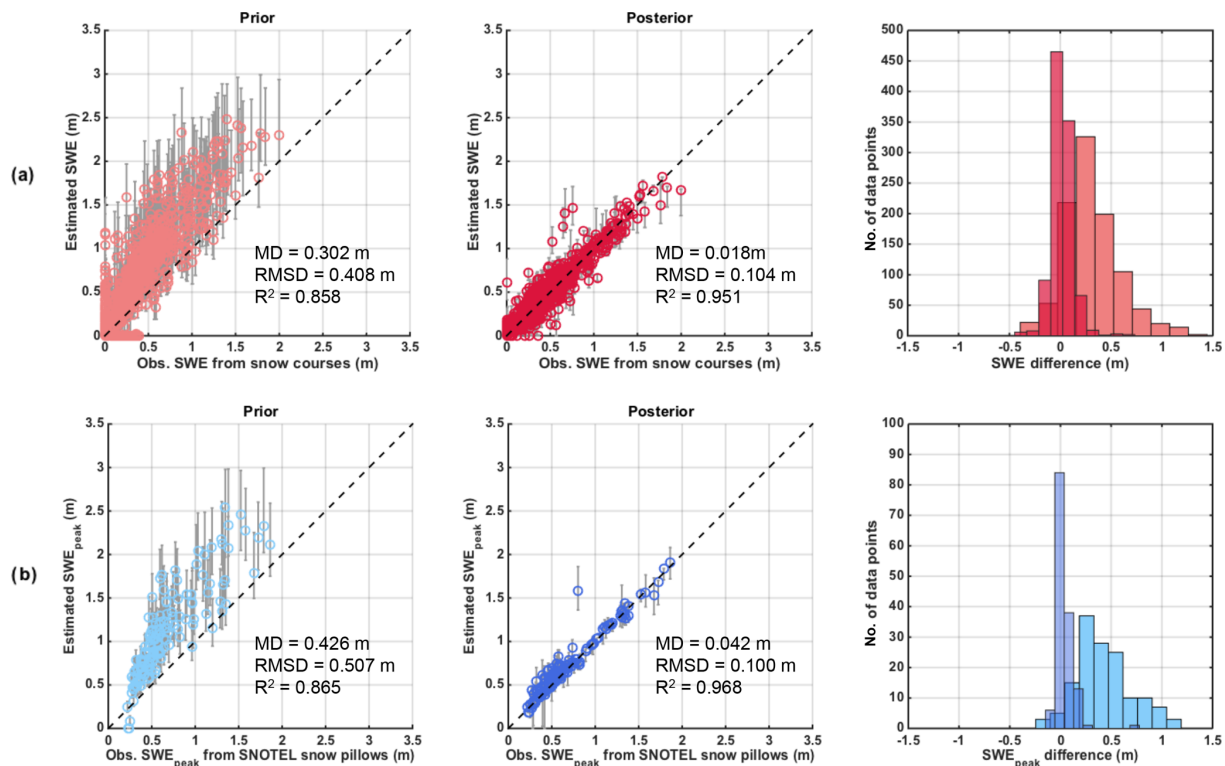
Figure 4. Left panels show scatter plots of prior estimated snow water equivalent (SWE) vs. in situ, middle panels show posterior estimated SWE vs. in situ, and right panels show histograms of the difference for **(a)** all snow courses and **(b)** snow pillows. The markers represent ensemble medians while the intervals represent the interquartile range (IQR). The mean difference (MD), root mean square difference (RMSD), and correlation coefficient (R^2) are displayed.

during the accumulation season (October to January), while the MR method slightly overestimates the baseline fSCA by 4 % or less during the ablation season (February to August). The annual average difference is 0.87 %. The expected impact of assimilating larger fSCA values during the ablation season is an overestimation of the length of the snowmelt period, which, for the same amount of melt season energy inputs, would translate into larger posterior SWE estimates. As seen in Fig. 5b and c, fSCA from both the baseline and the MR case share a similar distribution with respect to CM and Peak SWE (SWE_{peak}). As expected, areas of high fSCA correspond to areas of high SWE accumulation at the higher elevation of the basin, which also tend to be the most complex. By design, the MR approach does not coarsen areas of high physiographic complexity that can experience sharp differences in accumulation and/or ablation from one pixel to another. Hence, by construct, the MR fSCA is identical to the baseline for CM larger than 0.65 and differs slightly from the baseline in low-complexity areas as seen in Fig. 5b. In addition, Fig. 5c shows that the difference in fSCA over regions of high SWE accumulation is negligible as well (1.3 % or less). Given the small differences observed, the effect of the MR approach on the assimilated fSCA observations is minimal and therefore is not expected to significantly alter

the performance of the data assimilation scheme (discussed in more detail below).

3.2 Impact of the MR approach on snow climatology metrics

The following analysis focuses on the comparison of the posterior ensemble median SWE estimates for the baseline and MR cases. Peak SWE (SWE_{peak}), day of peak (DOP), and duration of melt (DOM) were chosen for analysis. SWE_{peak} is defined as the maximum daily SWE in a given WY. DOP is defined for each WY as the day when SWE is equal to SWE_{peak}. DOM is the difference between the melt-out day, defined as the day when only 1 % of the original SWE_{peak} remains, and DOP, which effectively quantifies the duration of the ablation season. These metrics can be defined either pixel-wise or for basin-averaged values.

3.2.1 Mean spatial distribution

Figures 6a–8a show maps of the 31-year-average pixel-wise SWE_{peak}, DOP, and DOM, while Figs. 6b–8b show the distribution of the respective 31-year-average relative differences binned by CM, elevation (Z), slope, fVEG, and SWE_{peak}. In these figures, the baseline estimates were al-
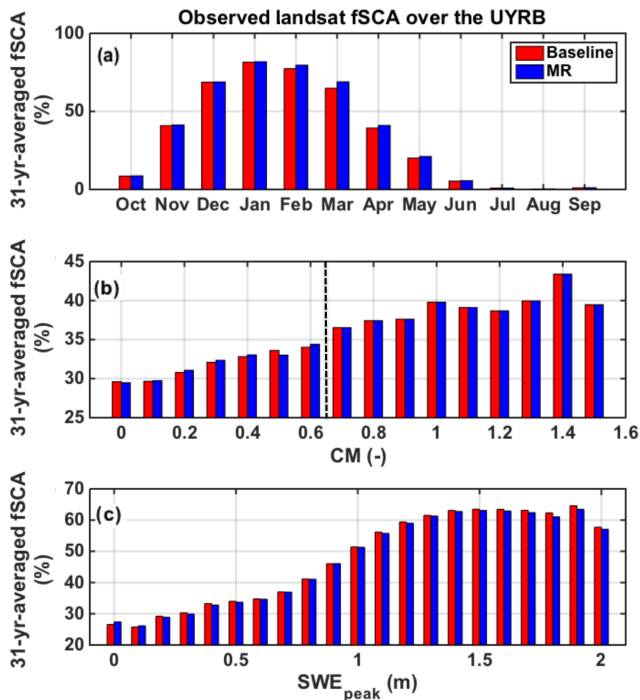
Figure 5. The fSCA climatology derived from the 31-year record of Landsat observations over the upper Yampa River basin: bin-averaging of all observations across the range of **(a)** months of the water year, **(b)** CM values, and **(c)** peak SWE (SWE_{peak}) values for the 90 m baseline and MR case. The CM maximum threshold CM_{max} of 0.65 is represented by the vertical dashed line **(b)**.

ways subtracted from the MR estimates, which means that a positive difference represents an overestimation of the baseline by the MR case and vice versa.

As expected, the climatological SWE_{peak} shows significant spatial variability for both the MR and 90 m baseline with values ranging from zero to well over 1 m of SWE (Fig. 6a). The middle and western parts of the basin that are not physiographically complex (see Fig. 1) receive 25 cm or less on average. Given their location and relatively low elevation (less than 2000 m) the SWE accumulation is not orographically driven, but more heavily influenced by the few winter snowstorms occurring over the basin. The more complex areas in the eastern and southern edges of the basin accumulate a much larger amount of SWE (on the order of 1 m or more). On average, the MR approach underestimated pixel-wise SWE_{peak} by 7.2 mm or 1.6 %, with the most complex areas showing no difference since they were modeled at 90 m by design, and the less complex but high-elevation areas showing larger differences on the order of 10 cm, or roughly 10 % of SWE_{peak}. As seen in the density scatter plot, the majority of pixels have a SWE_{peak} around 20 cm, and the correlation between the baseline and the MR case is very strong, with a correlation coefficient of 0.96. Figure 6b shows that the bin-averaged relative differences be-

tween the pixel-wise MR and baseline SWE_{peak} are constrained between −5 and 5 %. By construct, the CM bands larger than 0.65 show no difference because all the MR pixels were simulated at the baseline resolution. All elevation bands show an underestimation of SWE_{peak}, with the largest differences observed at middle elevations between 2600 and 3200 m. Since the UYRB is densely forested at these elevations, this is consistent with the largest underestimation occurring for the highest fVEG bands. Regarding the distribution of the differences with slope, the lower slope bands (0–15°) underestimate SWE_{peak} while the higher slope bands (20–35°) show overestimation. As discussed in Baldo and Margulis (2017), the coarsening of pixel properties by the MR method leads to a slight increase in fVEG for densely vegetated pixels, as well as an increase in more gently sloped and north-facing pixels. In the context of the SWE reanalysis, the magnitude of melt energy flux largely dictates the peak SWE that is consistent with a given fSCA depletion time series. The increase in fVEG as a result of the MR approach leads to an underestimation of the melt (energy) flux at the snow surface (as a result of attenuation of solar radiation), which decreases the posterior MR SWE_{peak} for these pixels. Since the minimum solar zenith angle during the ablation season over the UYRB is 16°, reducing gentle slopes (0–15°) leads to an underestimation of the melt flux (as a result of becoming less perpendicular to the incoming direct beam solar radiation), which decreases the posterior MR SWE_{peak} for these pixels. Reducing steeper slopes (20–35°) has the opposite effect and overestimates the melt flux, increasing the posterior MR SWE_{peak} for these pixels.

The posterior SWE_{peak} estimates are therefore impacted by the MR approach in two ways: (i) an overestimation of the assimilated fSCA during the ablation season and (ii) a general underestimation of the melt flux due to the coarsening of the basin physiography, with the exception of steep pixels where the melt flux is overestimated. The basin-averaged underestimation of SWE_{peak} observed in Fig. 6 suggests that the effect of coarsening the static inputs and meteorological forcing on SWE_{peak} is more important than the effect from the coarsened assimilated fSCA images. More importantly, the differences are the largest for the lowest SWE_{peak} band (less than 15 cm). The MR approach therefore concentrated the largest SWE_{peak} differences to areas of low CM that tend to accumulate less SWE.

Regarding DOP, Fig. 7a shows that SWE in the middle and western regions of the basin that are not physiographically complex peaks early during the winter between January and March. In contrast, the more complex regions in the eastern and southern parts of the UYRB accumulated SWE until much later during the spring (April to June). These complex regions show very good agreement between the baseline and MR case in terms of timing, with larger differences over the rest of the basin. The average underestimation of 0.8 days or −0.5 % is negligible. As seen in the density scatter plot, the majority of pixels have peak values around 1 March, with a
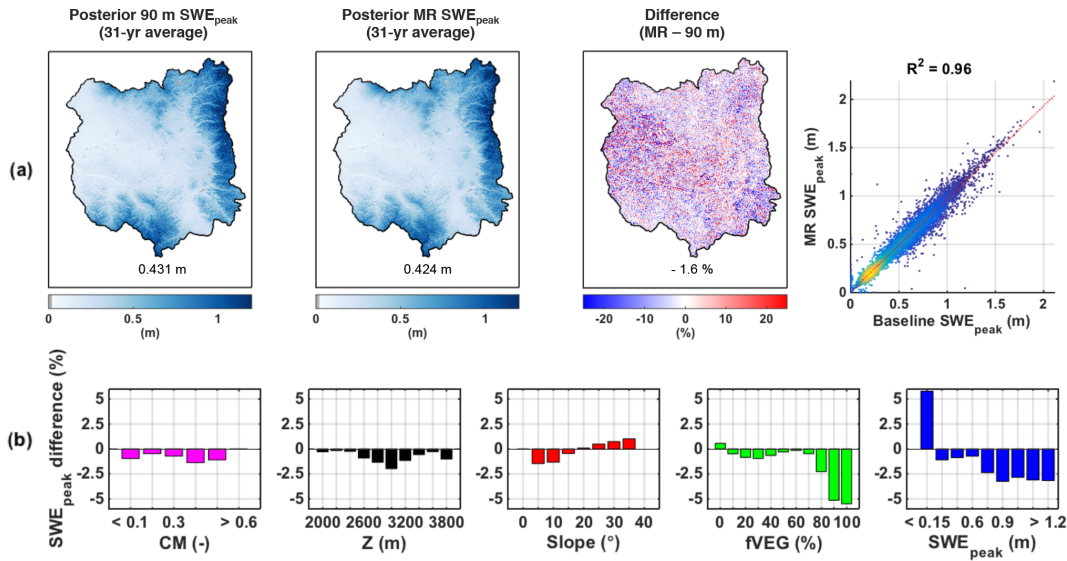
Figure 6. (a) Maps of pixel-wise 31-year-average posterior peak SWE (SWE$_{peak}$) over the upper Yampa River basin for the 90 m baseline, the MR case, the percentage difference between the two approaches (MR − baseline), and the corresponding scatter plot. Basin averages are displayed at the bottom of each map. **(b)** Distribution of SWE$_{peak}$ relative difference with complexity metric (CM), elevation (Z), slope, forested fraction (fVEG), and SWE$_{peak}$. Pixels with a 31-year-average SWE$_{peak}$ lower than 5 cm were discarded from the analysis.
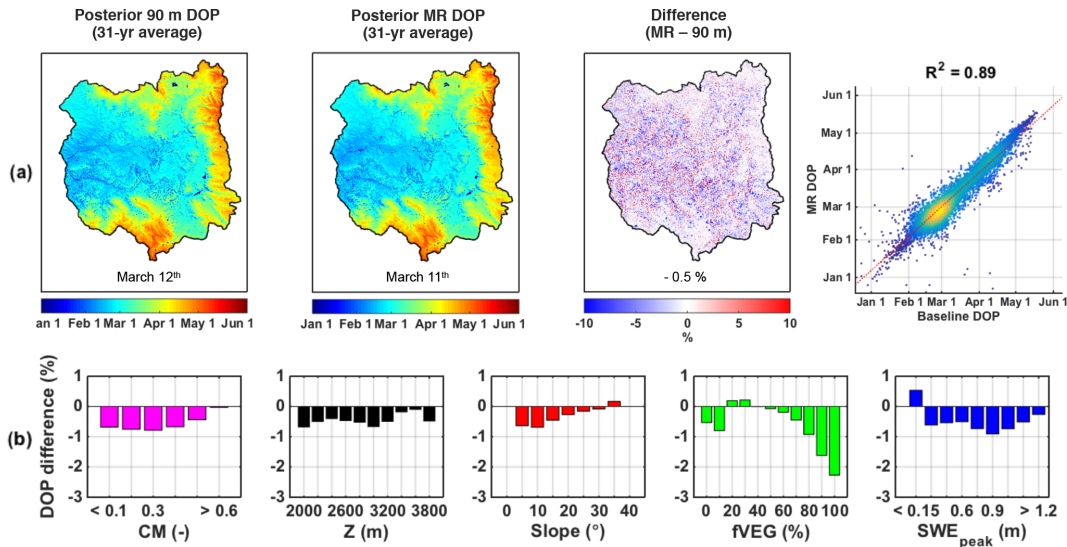
Figure 7. (a) Maps of pixel-wise 31-year-average day of peak (DOP) over the upper Yampa River basin for the 90 m baseline, the MR case, the percentage difference between the two approaches (MR − baseline), and the corresponding scatter plot. Basin averages are displayed at the bottom of each map. **(b)** Distribution of DOP relative difference with complexity metric (CM), elevation (Z), slope, forested fraction (fVEG), and SWE$_{peak}$. Pixels with a 31-year-average SWE$_{peak}$ lower than 5 cm were discarded from the analysis.

strong correlation coefficient of 0.89. Figure 7b shows DOP difference distributions with CM, elevation, slope, fVEG, and SWE$_{peak}$ similar to SWE$_{peak}$ (Fig. 6b), while the magnitude of the DOP differences is much smaller and ranges between 0.5 and −2 %. The MR approach therefore preserves the accuracy of areas accumulating large amounts of SWE, that peak later in the spring.

Regarding the duration of the ablation season, DOM can vary from less than 1 month over the areas that accumulated little SWE and started melting as soon as the snowstorm events ended, to almost 5 months over the southwestern edge of the basin (Fig. 8a). The average DOM is 61.7 days, or 2 months, for the MR case, which overestimates the 90 m baseline by 1 day or 1.6 %. The density scatter plot shows that the majority of pixels have a DOM between 1 and 2
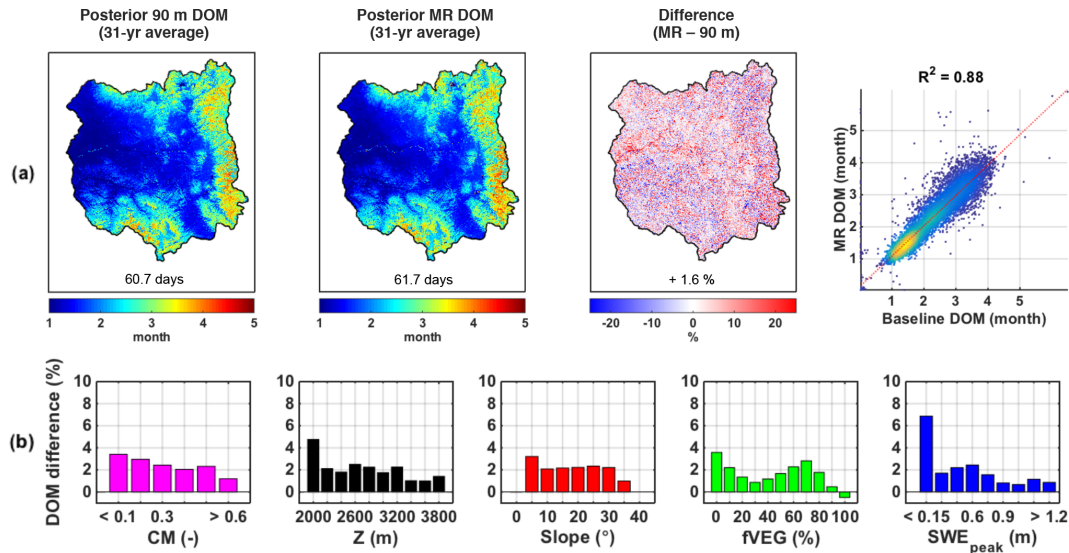
Figure 8. (a) Maps of pixel-wise 31-year-average duration of melt (DOM) over the upper Yampa River basin for the 90 m baseline, the MR case, the percentage difference between the two approaches (MR − baseline), and the corresponding scatter plot. Basin averages are displayed at the bottom of each map. **(b)** Distribution of DOM relative difference with complexity metric (CM), elevation (Z), slope, forested fraction (fVEG), and SWE_{peak}. Pixels with a 31-year-average SWE_{peak} lower than 5 cm were discarded from the analysis.

months, with a strong correlation coefficient of 0.88. The slight overestimation of DOM by the MR case was expected, given the underestimation of melt fluxes from the increase in gently north-facing and densely forested pixels (Baldo and Margulis, 2017) and the higher assimilated fSCA observations in the MR case. Figure 8b shows that the largest DOM overestimation occurs at the lowest band for all five variables. When looking at the distribution with SWE_{peak} specifically, pixels accumulating 15 cm of SWE or less show a DOM difference of 7 %, while pixels accumulating 1 m or more only show a DOM difference of 1 % or less. Pixels accumulating low amounts of SWE can be very intermittent in nature, without a clear SWE_{peak} or DOP, which can explain the higher difference seen in Fig. 8b.

Based on these results, when applying the MR approach to the SWE reanalysis framework, we therefore expect the largest differences to occur over areas of low physiographic complexity. These types of areas tend to peak early during the winter, accumulate less SWE, and melt within a month and display lower levels of spatial variability that are easier to model at coarser resolutions.

3.2.2 Basin-average mean seasonal cycle

The mean seasonal cycle of MR SWE underestimates the baseline case by less than 1 cm, as shown by the 31-year-average difference displayed in black in Fig. 9b. Figure 9a and b show that the seasonal cycles for both the MR and baseline case closely match during the accumulation season (November to March) with differences in the range of +1/−1 cm and a negative mean around −5 mm as shown
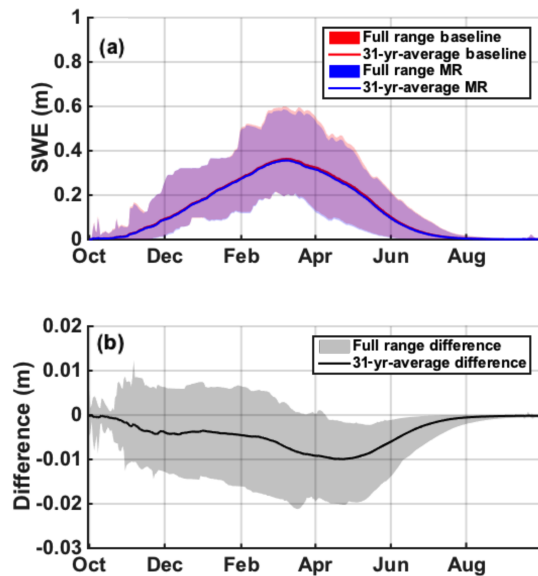
Figure 9. (a) Daily time series of basin-averaged posterior SWE from WY 1985 to WY 2015. The 31-year averages are displayed in solid lines, while the shaded regions represent the full range across WYs. **(b)** The 31-year-averaged difference between the MR case and the baseline is displayed in black, with the full range of differences shaded in grey.

by the grey shaded area and the black line respectively in Fig. 9b. The basin-averaged mean SWE_{peak} is 0.374 m for the MR case, and 0.381 m for the 90 m baseline (Fig. 9a), which leads to a mean difference of −7.1 mm (or −1.95 %, Fig. 9b).
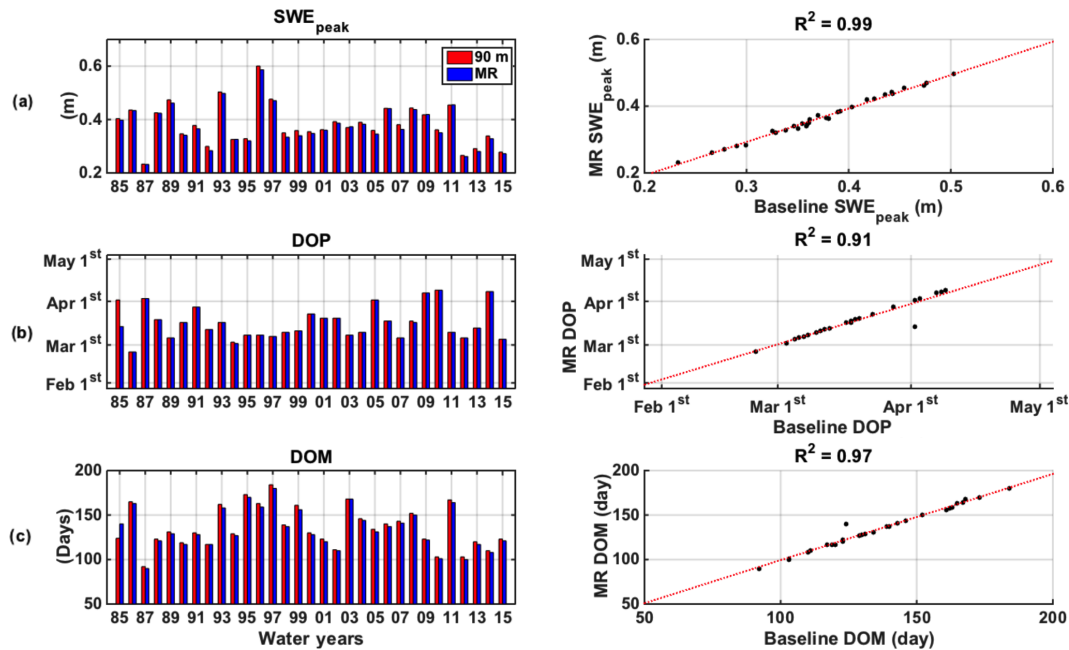
Figure 10. Left panels show annual time series and right panels show scatter plots with linear regressions of basin-averaged **(a)** peak SWE (SWE_{peak}), **(b)** day of peak (DOP), and **(c)** duration of melt (DOM) for the 90 m baseline and the MR case.

In terms of timing, DOP based on the mean seasonal cycle fits almost perfectly within a day, with the MR case peaking on 15 March and the baseline case on 16 March on average (Fig. 9a). The underestimation is more pronounced during the early ablation season (March to June), where the difference in assimilated fSCA observations is the largest (Fig. 5a), with the entire range of WYs showing negative differences, and a maximum of -2.1 cm (or -5.4 %) observed for the wettest year, WY 1996. Even though the MR case is assimilating slightly larger fSCA observations during the ablation season (Fig. 5a), the coarsening of the static inputs shown in Baldo and Margulis (2017) decreases the energy inputs, which ultimately lowers the posterior MR SWE estimates, and therefore explains the slight underestimation observed during the ablation season in Fig. 9.

3.2.3 Interannual variability

The baseline and MR annual time series of SWE_{peak} show close agreement in interannual variations (Fig. 10a). The scatter plot illustrates the positive performance of the MR case, including at both ends of the spectrum, which confirms that the MR case is estimating dry and wet years accurately. Figure 10b and c also illustrates the similarities in DOP and DOM interannual variability. WY 1985 shows the largest differences because there were two similar values of maximum SWE within 1 cm that occurred 15 days apart. The MR case identified the first peak as SWE_{peak}, while the baseline did the opposite, which does not impact the SWE_{peak} estimate, but does impact both DOP and DOM. Beyond this

single year, the MR case closely represents the interannual variability in the timing and length of accumulation and ablation seasons over the reanalysis period.

### 3.3 Impact of the MR approach on spatial variations of SWE uncertainty

The previous analysis focused on the impact of the MR approach on the posterior ensemble SWE median (i.e., a metric of central tendency). However, another strength of the reanalysis framework is to also provide a measure of uncertainty via the posterior ensemble. In this section the impact of the MR approach on the posterior ensemble SWE_{peak} coefficient of variation ($<CV>$) is examined, where the angle brackets ($<>$) are used to emphasize the ensemble operator.

In order to focus on the spatial distribution of the ensemble posterior SWE_{peak} uncertainty, the 31-year-average maps of $<CV>$ (Fig. 11) were created by pooling $<CV>$ for each pixel from its mean and standard deviation over all 31 WYs as follows (Bingham and Fry, 2010):

$$\overline{<\sigma>_i} = \sqrt{\frac{\sum_{y=1}^{31}\left(<\sigma>_i^y\right)^2 + \sum_{y=1}^{31}\left(<\mu>_i^y - \overline{<\mu>_i}\right)^2}{31}},$$

$$\overline{<CV>_i} = \frac{\overline{<\sigma>_i}}{\overline{<\mu>_i}}, \qquad (2)$$

where the overbar notation denotes the 31-year average, $\overline{<\sigma>_i}$ is the 31-year-average ensemble SWE_{peak} standard deviation for pixel i, $\overline{<\mu>_i}$ is the 31-year-average ensem-
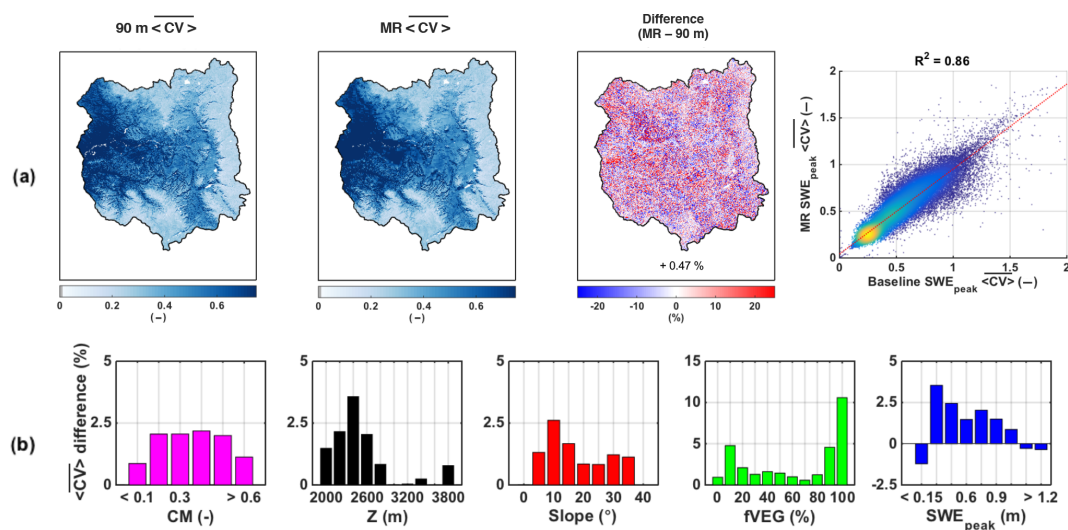
Figure 11. (a) Maps of pixel-wise 31-year-average SWE_{peak} coefficient of variation ($\overline{<CV>}$) over the upper Yampa River basin for the 90 m baseline, the MR case, the percentage difference between the two approaches (MR − baseline), and the corresponding scatter plot. **(b)** Distribution of $\overline{<CV>}$ relative difference with complexity metric (CM), elevation (Z), slope, forested fraction (fVEG), and SWE_{peak}. Pixels with a 31-year-average SWE_{peak} lower than 5 cm were discarded from the analysis.

ble SWE_{peak} mean for the same pixel i, and $<\sigma>_i^y$ and $<\mu>_i^y$ are respectively the ensemble SWE_{peak} standard deviation and mean for each individual WY y. The 31-year-average SWE_{peak} coefficient of variation ($\overline{<CV>_i}$) for each pixel i was calculated as the ratio between the pixel 31-year-average ensemble SWE_{peak} standard deviation and mean.

As seen in Fig. 11a, the spatial distributions of $\overline{<CV>}$ is highly variable. For both the baseline and MR cases, the high-elevation areas accumulating large amounts of SWE (see Fig. 6a) show a $\overline{<CV>}$ on the order of 10–20 %, while the lower parts of the UYRB have a $\overline{<CV>}$ higher than 60 %. Regarding the relative difference between the MR and baseline cases (Fig. 11a), regions accumulating the most SWE with the lowest $\overline{<CV>}$ also have the lowest relative difference between the MR and baseline cases (white areas on the eastern and southern edges of the UYRB). The basin-average difference in $\overline{<CV>}$ of 0.47 % is, however, negligible.

The spatial distributions of these differences are shown in Fig. 11b. Besides highlighting again the low magnitude of the difference, its distribution is also in accordance with the way the MR was designed. Most of the difference observed in SWE_{peak} uncertainty is concentrated over areas of low complexity, elevation, slope, heterogeneously dense forests, and, most importantly, low SWE_{peak}.

4 Conclusions

This study demonstrated the performance of a new MR terrain discretization approach in the context of a snow reanalysis framework using the assimilation of Landsat-derived fSCA observations. The MR approach was shown to have an insignificant impact on the fSCA observations assimilated and the reanalysis framework led to posterior SWE ensembles similar to the high-resolution 90 m baseline. The SWE reanalysis dataset generated with the MR approach matched the 90 m baseline ensemble median within 1 cm on average for peak SWE magnitude and within 1 day on average for timing of the accumulation and melt seasons. Most of the difference between the two approaches occurs in areas accumulating less than 15 cm of SWE, while areas accumulating more than that are estimated with a high degree of accuracy. In addition, the MR approach also preserved the SWE uncertainty, where the coefficient of variation showed differences on the order of 0.5 %. This study has demonstrated the feasibility of the MR approach in the context of a snow reanalysis framework, where the significant decrease in computational costs will allow much larger scale implementations of the SWE reanalysis over full mountain ranges, while preserving the accuracy of fine spatial resolution simulations.

References

Andreadis, K. M. and Lettenmaier, D. P.: Assimilating remotely sensed snow observations into a macroscale hydrology model, Adv. Water Resour., 29, 872–886, https://doi.org/10.1016/j.advwatres.2005.08.004, 2006.

Arsenault, K. R., Houser, P. R., De Lannoy, G. J. M., and Dirmeyer, P. A.: Impacts of snow cover fraction data assimilation on modeled energy and moisture budgets, J. Geophys. Res.-Atmos., 118, 7489–7504, https://doi.org/10.1002/jgrd.50542, 2013.

Baldo, E. and Margulis, S. A.: Implementation of a physiographic complexity-based multiresolution snow modeling scheme, Water Resour. Res., 53, 3680–3694, https://doi.org/10.1002/2016WR020021, 2017.

Beven, K. J., Cloke, H., Pappenberger, F., Lamb, R., and Hunter, N.: Hyperresolution information and hyperresolution ignorance in modelling the hydrology of the land surface, Sci. China Earth Sci., 58, 25–35, https://doi.org/10.1007/s11430-014-5003-4, 2015.

Beven, K. J. and Kirby, M. J.: A physically based, variable contributing area model of basin hydrology, Hydrolog. Sci. Bull., 24, 43–69, https://doi.org/10.1080/02626667909491834, 1979.

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., van de Giesen, N., Winsemius, H., and Wood, E. F.: Hyper-resolution global hydrological modelling: what is next?, Hydrol. Process., 29, 310–320, https://doi.org/10.1002/hyp.10391, 2015.

Bingham, N. H. and Fry, J. M.: Regression Linear Models in Statistics, Springer-Verlag, London, 2010.

Chaney, N. W., Metcalfe, P., and Wood, E. F.: HydroBlocks: a field-scale resolving land surface model for application over continental extents, Hydrol. Process., 30, 3543–3559, https://doi.org/10.1002/hyp.10891, 2016.

Christensen, N., Wood, A., Voisin, N., Lettenmaier, D., and Palmer, R.: The Effects of Climate Change on the Hydrology and Water Resources of the Colorado River Basin, Climatic Change, 62, 337–363, https://doi.org/10.1023/B:CLIM.0000013684.13621.1f, 2004.

Clark, M. P., Slater, A. G., Barrett, A. P., Hay, L. E., McCabe, G. J., Rajagopalan, B., and Leavesley, G. H.: Assimilation of snow covered area information into hydrologic and land-surface models, Adv. Water Resour., 29, 1209–1221, https://doi.org/10.1016/j.advwatres.2005.10.001, 2006.

Clark, M. P., Hendrikx, J., Slater, A. G., Kavetski, D., Anderson, B., Cullen, N. J., Kerr, T., Örn Hreinsson, E., and Woods, R. A.: Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review, Water Resour. Res., 47, W07539, https://doi.org/10.1029/2011WR010745, 2011.

Cortés, G. and Margulis, S.: Impacts of El Niño and La Niña on interannual snow accumulation in the Andes: Results from a high-resolution 31 year reanalysis, Geophys. Res. Lett., 44, 6859–6867, https://doi.org/10.1002/2017GL073826, 2017.

Cortés, G., Girotto, M., and Margulis, S. A.: Analysis of sub-pixel snow and ice extent over the extratropical Andes using spectral unmixing of historical Landsat imagery, Remote Sens. Environ., 141, 64–78, https://doi.org/10.1016/j.rse.2013.10.023, 2014.

Cortés, G., Girotto, M., and Margulis, S.: Snow process estimation over the extratropical Andes using a data assimilation framework integrating MERRA data and Landsat imagery, Water Resour. Res., 52, 2582–2600, https://doi.org/10.1002/2015WR018376, 2016.

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Marshall, C., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., and Meng, J.: Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project, J. Geophys. Res.-Atmos., 108, 8842, https://doi.org/10.1029/2002JD003118, 2003.

De Lannoy, G. J. M., Reichle, R. H., Houser, P. R., Arsenault, K. R., Verhoest, N. E. C., and Pauwels, V. R. N.: Satellite-Scale Snow Water Equivalent Assimilation into a High-Resolution Land Surface Model, J. Hydrometeorol., 11, 352–369, https://doi.org/10.1175/2009JHM1192.1, 2010.

Derksen, C. and Brown, R.: Spring snow cover extent reductions in the 2008–2012 period exceeding climate model projections, Geophys. Res. Lett., 39, 119504, https://doi.org/10.1029/2012GL053387, 2012.

Durand, M., Molotch, N. P., and Margulis, S. A.: A Bayesian approach to snow water equivalent reconstruction, J. Geophys. Res.-Atmos., 113, D20117, https://doi.org/10.1029/2008JD009894, 2008.

Girotto, M., Cortés, G., Margulis, S. A., and Durand, M.: Examining spatial and temporal variability in snow water equivalent using a 27 year reanalysis: Kern River watershed, Sierra Nevada, Water Resour. Res., 50, 6713–6734, https://doi.org/10.1002/2014WR015346, 2014a.

Girotto, M., Margulis, S. A., and Durand, M.: Probabilistic SWE reanalysis as a generalization of deterministic SWE reconstruction techniques, Hydrol. Process., 28, 3875–3895, https://doi.org/10.1002/hyp.9887, 2014b.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J. N., and Wickham, J.: Completion of the 2001 National Land Cover Database for the conterminous United States, 337–341, http://pubs.er.usgs.gov/publication/70029996 (last access: 9 November 2017), 2007.

JPL: ASTER Global Digital Elevation Model, NASA JPL, http://doi.org/10.5067/ASTER/ASTGTM.002 (last access: 9 November 2017), 2009.

Kumar, S. V., Peters-Lidard, C. D., Arsenault, K. R., Getirana, A., Mocko, D., and Liu, Y.: Quantifying the Added Value of Snow Cover Area Observations in Passive Microwave Snow Depth Data Assimilation, J. Hydrometeorol., 16, 1736–1741, https://doi.org/10.1175/JHM-D-15-0021.1, 2015.

Landsat: Landsat-5, Landsat-7, and Landsat-8 images courtesy of the US Geological Survey, available at: http://earthexplorer.usgs.gov/, last access: 9 November 2017.

Liston, G. E.: Representing Subgrid Snow Cover Heterogeneities in Regional and Global Models, J. Climate, 17, 1381–1397, https://doi.org/10.1175/1520-0442(2004)017<1381:RSSCHI>2.0.CO;2, 2004.

Liu, Y., Peters-Lidard, C. D., Kumar, S., Foster, J. L., Shaw, M., Tian, Y., and Fall, G. M.: Assimilating satellite-based snow depth and snow cover products for improving snow predictions in Alaska, Adv. Water Resour., 54, 208–227, https://doi.org/10.1016/j.advwatres.2013.02.005, 2013.

Luo, L., Robock, A., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Lohmann, D., Cosgrove, B., Wen, F., Sheffield, J., Duan, Q., Higgins, R. W., Pinker, R. T., and Tarpley, J. D.: Validation of the North American Land Data Assimilation System (NLDAS) retrospective forcing over the southern Great Plains, J. Geophys. Res.-Atmos., 108, 8843, https://doi.org/10.1029/2002JD003246, 2003.

Margulis, S. A., Girotto, M., Cortés, G., and Durand, M.: A Particle Batch Smoother Approach to Snow Water Equivalent Estimation, J. Hydrometeorol., 16, 1752–1772, https://doi.org/10.1175/JHM-D-14-0177.1, 2015.

Margulis, S. A., Cortés, G., Girotto, M., and Durand, M.: A Landsat-Era Sierra Nevada Snow Reanalysis (1985–2015), J. Hydrometeorol., 17, 1203–1221, https://doi.org/10.1175/JHM-D-15-0177.1, 2016.

Mascaro, G., Vivoni, E. R., and Méndez-Barroso, L. A.: Hyperresolution hydrologic modeling in a regional watershed and its interpretation using empirical orthogonal functions, Adv. Water Resour., 83, 190–206, https://doi.org/10.1016/j.advwatres.2015.05.023, 2015.

Molotch, N. P., Painter, T. H., Bales, R. C., and Dozier, J.: Incorporating remotely-sensed snow albedo into a spatially-distributed snowmelt model, Geophys. Res. Lett., 31, l03501, https://doi.org/10.1029/2003GL019063, 2004.

Painter, T. H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R. E., and Dozier, J.: Retrieval of subpixel snow covered area, grain size, and albedo from MODIS, Remote Sens. Environ., 113, 868–879, https://doi.org/10.1016/j.rse.2009.01.001, 2009.

Sivapalan, M., Beven, K., and Wood, E. F.: On hydrologic similarity: 2. A scaled model of storm runoff production, Water Resour. Res., 23, 2266–2278, https://doi.org/10.1029/WR023i012p02266, 1987.

Su, H., Yang, Z.-L., Niu, G.-Y., and Dickinson, R. E.: Enhancing the estimation of continental-scale snow water equivalent by assimilating MODIS snow cover with the ensemble Kalman filter, J. Geophys. Res.-Atmos., 113, d08120, https://doi.org/10.1029/2007JD009232, 2008.

Sun, S. and Xue, Y.: Implementing a new snow scheme in Simplified Simple Biosphere Model, Adv. Atmos. Sci., 18, 335–354, https://doi.org/10.1007/BF02919314, 2001.

Tucker, G. E., Lancaster, S. T., Gasparini, N. M., Bras, R. L., and Rybarczyk, S. M.: An object-oriented framework for distributed hydrologic and geomorphic modeling using triangulated irregular networks, Comput. Geosci., 27, 959–973, https://doi.org/10.1016/S0098-3004(00)00134-5, 2001.

US Geological Survey, W. R. D., Leavesley, G. H., Lichty, R. W., Troutman, B. M., and Saindon, L. G.: Precipitation-runoff modeling system; user's manual, Tech. rep., http://pubs.er.usgs.gov/publication/wri834238 (last access: 9 November 2017), 1983.

Vivoni, E., Ivanov, V., Bras, R., and Entekhabi, D.: Generation of triangulated irregular networks based on hydrological similarity, J. Hydrol. Eng.-ASCE, 9, 288–302, https://doi.org/10.1061/(ASCE)1084-0699(2004)9:4(288), 2004.

Winstral, A., Marks, D., and Gurney, R.: Assessing the Sensitivities of a Distributed Snow Model to Forcing Data Resolution, J. Hydrometeorol., 15, 1366–1383, https://doi.org/10.1175/JHM-D-13-0169.1, 2014.

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, Water Resour. Res., 47, w05301, https://doi.org/10.1029/2010WR010090, 2011.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., and Mocko, D.: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, J. Geophys. Res.-Atmos., 117, d03109, https://doi.org/10.1029/2011JD016048, 2012.

Xue, Y., Sellers, P. J., Kinter, J. L., and Shukla, J.: A Simplified Biosphere Model for Global Climate Studies, J. Climate, 4, 345–364, https://doi.org/10.1175/1520-0442(1991)004<0345:ASBMFG>2.0.CO;2, 1991.

Xue, Y., Sun, S., Kahan, D. S., and Jiao, Y.: Impact of parameterizations in snow physics and interface processes on the simulation of snow cover and runoff at several cold region sites, J. Geophys. Res.-Atmos., 108, 8859, https://doi.org/10.1029/2002JD003174, 2003.