**Supplementary Material**

**S.1** Additional testing of the scatterbin histogram approach

To help capture the range of observed snow water equivalent (SWE) and snow cover fraction (SCF) relationships, a histogram analysis is introduced that allows for reducing the dispersion of data and retaining most measurements within the analysis. A binned scatterplot method is used where SNOwpack TELemetry (SNOTEL) SWE observations (along the x-axis) are grouped into bins of equal length (e.g., 40 mm of SWE) and the corresponding MODIS SCF observations are averaged within the given bin. For each bin, a boxplot can be produced to reflect the spread in MODIS SCF values for that given bin and be represented by that bin's statistics (mean, median, minimum/maximum, etc.). Figure S1 presents a range of some of the statistics per 40 mm SWE bin for the Washington (WA) and Colorado (CO) domains. The median values tend to be higher than the averages, where the median is affected by the amount of 100% SCF values across the many SWE bins. In this study, the average SCF (arithmetic mean) of each SWE bin was used.

To show the robustness of the scatterbin approach, another set of binned scatterplots were applied to the data whereby an equal number of SNOTEL SWE observations were grouped per bin with SCF averaged across each new SWE bin. Here, 1000 SWE values were grouped into each bin. To show the differences in the distribution of observations per bin, Figures S2(a) and S2(c) highlight those observation count distributions for WA and CO, respectively. In adjacent plots, Figures S2(b) and (d), the averaged SCF values are shown for each SWE bin type, revealing tight overlaps between the averages of the two bin approaches, for both WA and CO. One additional test was applied to the binned scatterplot approach and that involved splitting the SNOTEL observation sites into two separate groups, thus two groups of 23 for WA (56 total sites) and 49 for CO (98 total) were set. The equal-length 40 mm SWE bins were applied independently to each group, and again overlaps in the binned averages were found for both regions (not shown). A Student t-test was employed to test whether a significant difference existed between the two groups (null hypothesis: *No difference*), and the final t-statistic values fell within the given two-tailed critical value at $\alpha = 0.05$, not rejecting the null hypothesis of no difference. These tests provide support for the use of the binned scatterplot approach in deriving the snow depletion curves (SDC)-type relationships between the SNOTEL SWE and Terra MODIS SCF observations.

**S.2** Generating temporal and physiographic based SDCs

Physiographic conditions are known to be a major control on larger spatial patterns of snow cover distribution, especially topography (e.g., Jost et al., 2007; Anderton et al., 2004). Also, snow cover patterns vary greatly throughout the year with extensive (lower) snow cover and low (high) snowpack depths in fall (spring), affected by snow accumulation (snowmelt) processes. As derived for the annually representative observation-based SDC equations in Section 4, additional *obs*-**h** SDC relationships are derived and evaluated for different temporal and physiographic conditions, e.g., elevation bands and vegetation type, for the two regions, Washington (WA) and Colorado (CO). We show here how observation-based *obs*-**h** curves, using the methods described in Section 4.1, vary with time of year and physiographic conditions. The binned scatterplot approach is applied, but for varying elevation band, vegetation, and monthly timescale for WA and CO. The SNOTEL SWE and MODIS SCF measurements are used in the same manner, but binned for these specific cases. The SWE bins and corresponding SCF averages are then used to derive each logarithmic equation's slope and intercept coefficients by estimating these beta parameters using a Least Squares fitting-logarithmic approach. We present the predicted SCF values and generated regression lines in the figures found in this section.

For time of year, an *obs*-**h** curve is derived for each month, separately, and for each region. The summer months were found in different calculations to average near zero, so no curves were fit for this season and not shown in subsequent figures. Though for the summer months when a MODIS SCF observation is present, the CLM2 *model*-**h** curve is used instead to derive the predicted model SCF observations. The monthly fitted curves are plotted in Figures S3 and S4 for WA and CO, respectively, as functions of 40 mm interval SWE bins as inputs. For WA, the shape of the binned

points and curves change with each month, with higher predicted SCF values in fall to early winter months, leveling off mostly to 90% SCF at about 1000 mm of SWE (Figure S3). The curves tend to have lower y-intercepts in spring months and lower SCF averages, especially within the lower SWE binned values. The resulting monthly logarithmic curves are similar to the shape of those monthly curves shown in Niu and Yang (2007) Figure 2, where instead they used 1 x 1° gridded SCA maps, which were derived from binary snow cover datasets. For the CO region monthly curves (Figure S4), similar curve shape and features are found as in WA per each month, though with higher SCF values associated with fall to early winter months and lower SCF values into the spring months, as expected snow melt occurs.

The next set of logarithmic functions derived as observation-based observation operators involve differences in elevation height for each region. Based on the range of SNOTEL elevation values, near-average mid-points of elevation are taken as a cutoff value to distinguish higher versus lower elevation bands (based on Andreadis and Lettenmaier, 2006; Su et al., 2008). For WA, the average midpoint elevation is about 1500 m and used to set the lower elevation range as 1000–1500 m and higher elevation range as 1501–2000 m. For CO, the two elevation band ranges are 2500–3000 m for the lower and 3001–3600 m for the upper band. Applying the scatterbin plot approach to each group of elevation points, two sets of logarithmic curves are generated like before for each region, one reflecting the lower and the other higher elevation band, and shown in Figures S5 and S6. For WA (Figure S5), distinctive logarithmic curves emerge between the two bands with much spread or difference between them. As predicted, higher SCF averages are shown with the higher elevation bands. Similar patterns are shown for the CO elevation band curves, but there is less spread shown, especially with lower snowpack SWE values present (Figure S6).

Finally, the scatterbin approach is applied and logarithmic functions are fitted for different vegetation types within each region. For the vegetation classes, to ensure agreement with the original 500m sinusoidally-projected MODIS SCF pixels used here, the MODIS 1km sinusoidal UMD land classification maps are used for consistency purposes. The sinusoidally based MODIS land cover pixels are extracted for each corresponding SNOTEL point and used for grouping the sets of SNOTEL SWE values and MODIS SCF pixel values in deriving the binned scatter points. The resulting vegetation classes included for each region and for which logarithmic functions are derived are shown in Figures S7 and S8, for WA and CO, respectively, including the list of classes for each region.

For WA, five UMD classes resulted from the group of SNOTEL sites, so bins and curves were derived for those classes. Vegetation class 1, evergreen needleleaf forest, is shown to correspond to the highest observation counts per bin, and the grassland type (UMD class 10) has the least amount of counts for both regions. However as expected, higher SCF averages are associated with grassland points, and lowest values associated with mixed forest type (class 5), as shown for WA (Figures S7). The CO vegetation class logarithmic functions also similarly exhibit expected patterns as WA, but the evergreen needleleaf class function is associated with the lowest SCF averages across all 40 mm SWE bin inputs. This result is not too surprising, since evergreen needleleaf forests and similar vegetation types contribute to lower MODIS SCF detection ability through obscuring the underlying snow from the satellite's view and impacts on the MODIS Normalized Difference Snow Index (NDSI) algorithm for discriminating snow cover from the vegetation (Hall and Riggs, 2007).

Since the evergreen needleleaf forest class is the most dominant type at or near the SNOTEL sites, it can be important to show whether these derived logarithmic functions, or observation operators to the Ensemble Kalman filter (EnKF) scheme, are actually representative of these MODIS SCF averages, and do not just reflect the associated large observation counts. An independent check is applied here to test the curves associated with this land class. Since it was shown in Figures S5 and S6 that higher binned SCF averages occur at higher elevations, an elevation threshold was selected for each region, which is weighted toward higher elevation points, to derive a new set of scatterplot bins and a new logarithmic curve for observation points associated only with evergreen needleleaf points. For WA, the threshold selected was 1600 m, and 3200 m for CO, above which only snow observations are included in the bins. The results are highlighted in Figures S9 and S10 for WA and

CO, respectively, and indicate pretty close agreement between the elevation-screened bins and bins that included all snow observation values associated with evergreen needleleaf class, for both regions. For CO, the bins remain much lower than the annual averaged bin values and in close agreement with the bins, including all evergreen needleleaf points. This suggests that the evergreen needle class-based curves do reflect "realistic" SCF estimates, as seen from the satellite and determined with the NDSI algorithm.

**S.3** Temporal and physiographic SDC effects on the data assimilation (DA) experiments

As for the SDC functions derived for the different months and physiographic conditions, Figure S11(a)–(d) shows a comparison of the spatially averaged predicted SCF from the EnKF experiments with the different observation operators applied and MODIS SCF observations for the four different melt seasons for WA (a,b), an CO (c,d). The four EnKF experiments include the original *model*-**h** one, and the annual, monthly, vegetation type and elevation band based *obs*-**h** experiments. SNOTEL estimated SCF is also derived using the CLM2 *model*-**h** curve and presented as spatial averages with the other SCF averages. The impact of the monthly *obs*-**h** curves can be seen, such as the drops that occur in the averaged SCF values at the start of each month, e.g., April 1st for WA's WY2005 case. The vegetation based EnKF experiment shows predicted SCF values to be higher for all four snowmelt season cases, revealing the impact of 100% MODIS SCF on the analysis, due to the large discrepancies in the curve-predicted SCF values and observed values, especially for the evergreen needleleaf case in CO. Such results indicate that MODIS can be effective in adding snow, depending on the observation operator's characteristics and how much model SWE and snow depth are present.

To further show an example of the difference between the *model*-**h** and *obs*-**h** curves in the DA experiments, we present two individual SNOTEL site location comparisons from the WA domain (Salmon Meadows and Ollalie Meadows) for WY2004, focusing on just the impact of the vegetation *obs*-**h** equations and observational standard errors (Figure S12). The SNOTEL site comparisons include two evergreen needeleaf locations in the WA domain to highlight the model SWE analysis response to the model, observed annual, and vegetation class-based observation operators. The less steep nature of the two *obs*-**h** curves, both for the annual and vegetation-based, compared to the CLM2 *model*-**h** curve, produced higher SWE analysis due to the *obs*-**h** curves never reaching above 90% SCF, allowing the 100% MODIS SCF values to "pull" the predicted SWE higher. This influence improves the peak SWE and early melt-off issue for the Olallie Meadows site (21b55s) in relation to the SNOTEL SWE (Figure S12(b)). However, the peak SWE for the Salmon Meadows site (19a02s) improves between March and April, but due to the model's snowmelt and lack of windblown physics, retains too much SWE into May (Figure S12(a)). These two examples show how the shape of the SDC function can play a major role in the snow cover assimilation updates, leading to either late snowmelt or improved snowmelt conditions, as the observational standard errors were considered similar enough to not have made a large impact on the DA-based update steps.

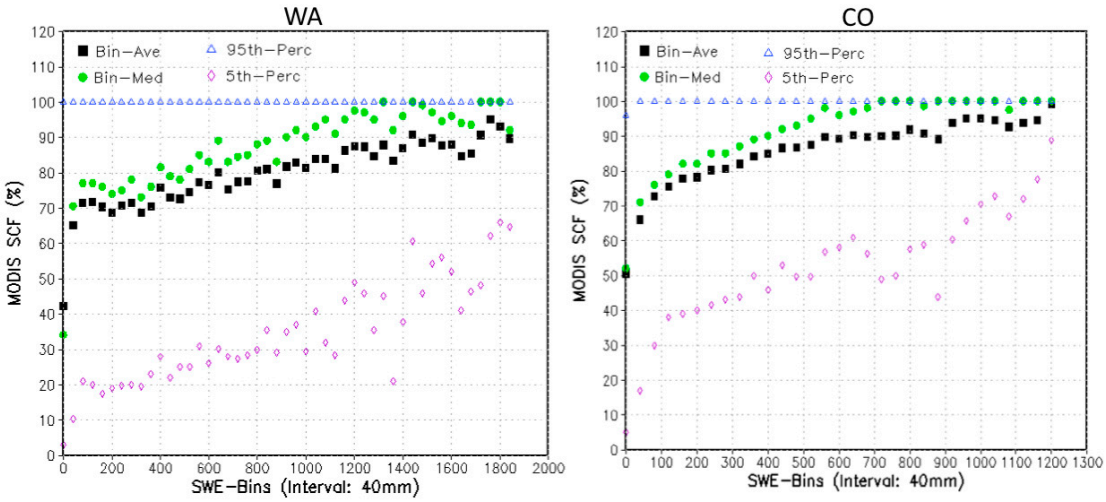


**Figure S1.** Binned scatterplots and relevant bin statistics are shown for Washington (WA) (**left**) and Colorado (CO) (**right**) domains. The statistics per each 40 mm snow water (SWE) bin include average (black squares), median (green circles), 5th and 95th percentile ranks (purple diamonds and blue triangles, respectively) of MODIS snow depletion curves (SCF) values.

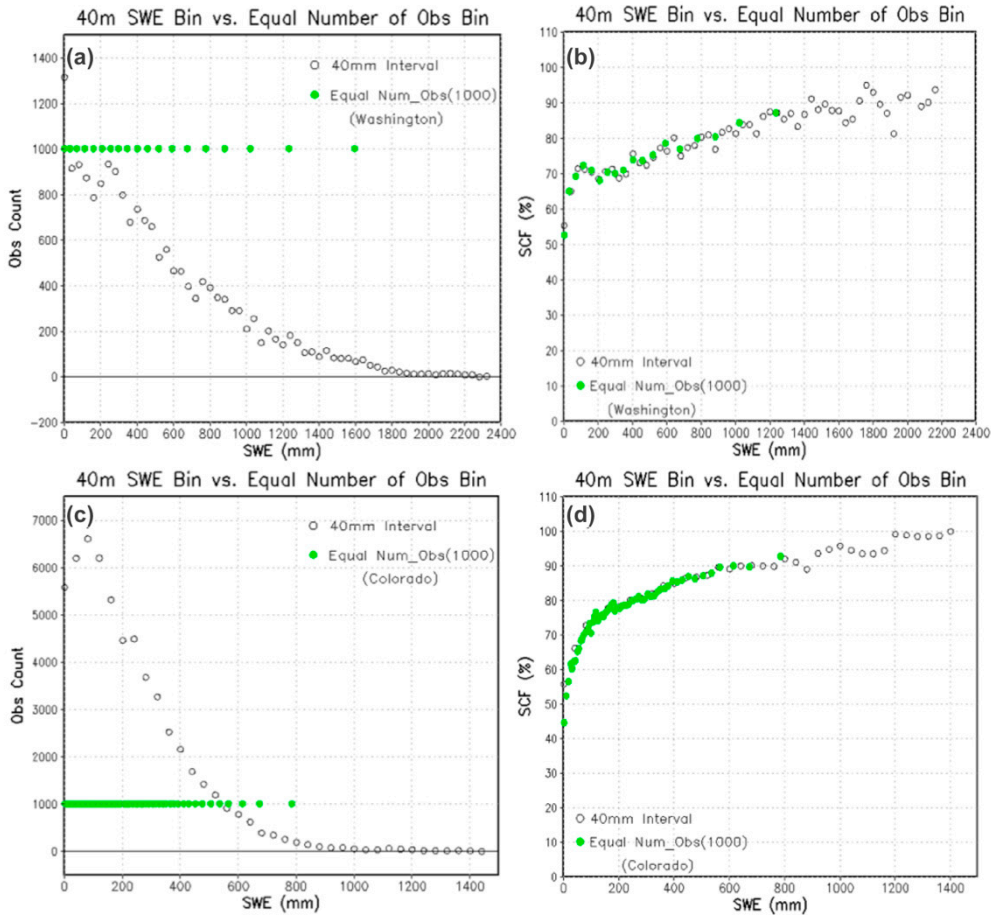


**Figure S2.** Comparison of equal-length SWE histogram bins (open circles) to equal-number of SWE observations per histogram bin (solid green circles) for Washington (**a,b**) and Colorado (**c,d**). The distributions of observations per bin are shown in plots (**a**) and (**c**), and plots (**b**) and (**d**) present average SCF values (in %) for both 40 mm SWE bins and an equal number (1000 count) of SWE observations.

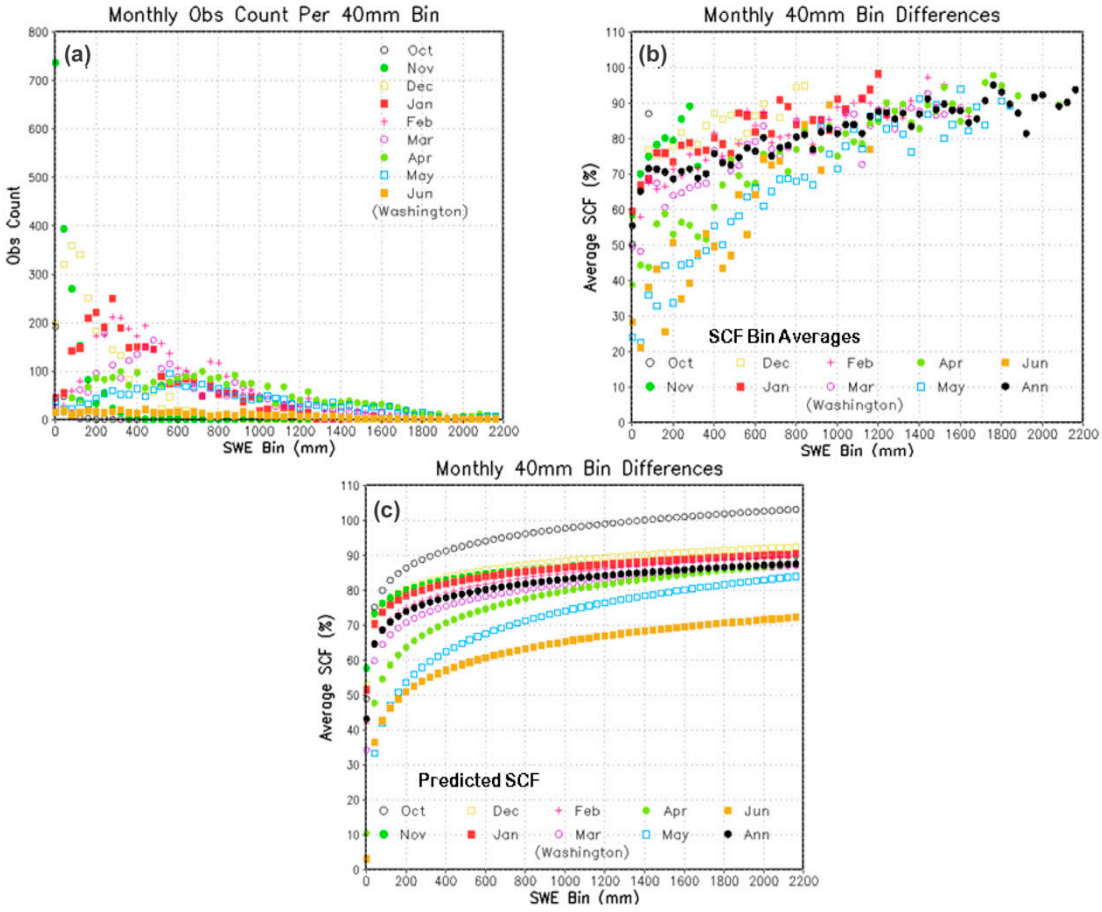


**Figure S3.** Monthly SDC relationships derived from binned scatterplot approach for WA, including water years 2000-03 and 2006-10. (**a**) The observation counts per bin are shown for each month category along with (**b**) the bin-averaged SCF values, and **c**) the predicted SCF values for each 40-mm SWE bin from the logarithmic functions fitted to the scatter-bin points in plot (**b**). Summer months from July to September are not included.

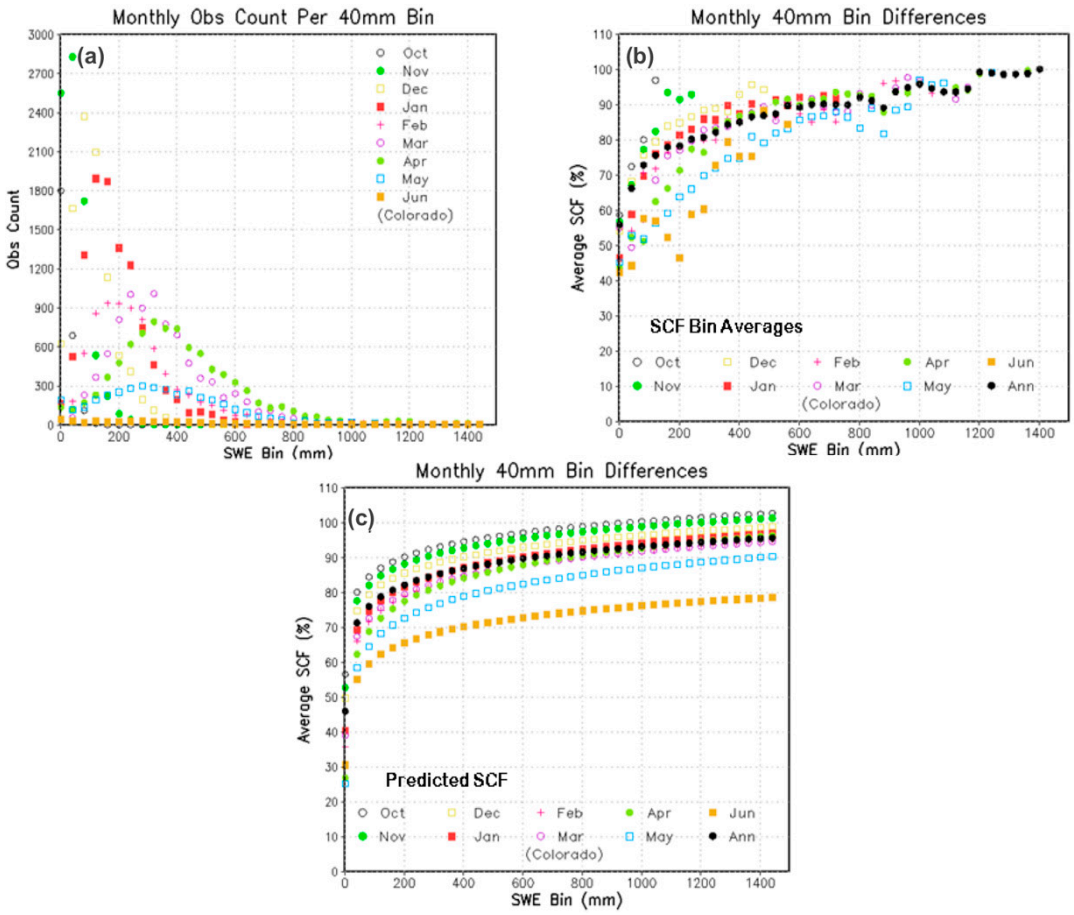


**Figure S4.** Monthly SDC relationships derived from binned scatterplot approach for CO, including water years 2000-03 and 2006-10. (**a**) The observation counts per bin are shown for each month category along with (**b**) the bin-averaged SCF values, and (**c**) the predicted SCF values for each 40-mm SWE bin from the logarithmic functions fitted to the scatter-bin points in plot (**b**). Summer months from July to September are not included.

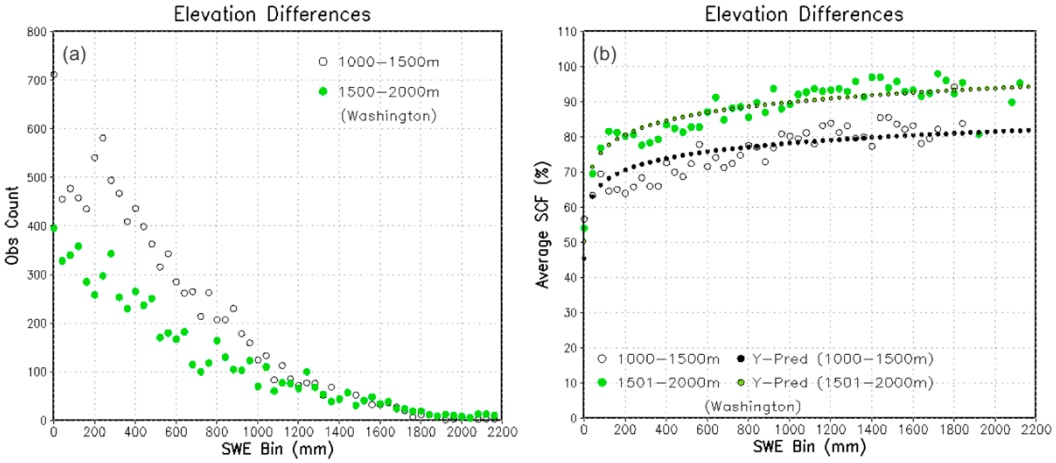


**Figure S5.** SDC relationships derived for two different elevation bands, 1000–1500m and 1501–2000m, from binned scatterplot approach for WA region. (**a**) The observation counts per bin are shown for each elevation band category along with (**b**) the bin-averaged SCF values overlaid with the predicted SCF values for each 40 mm SWE bin.

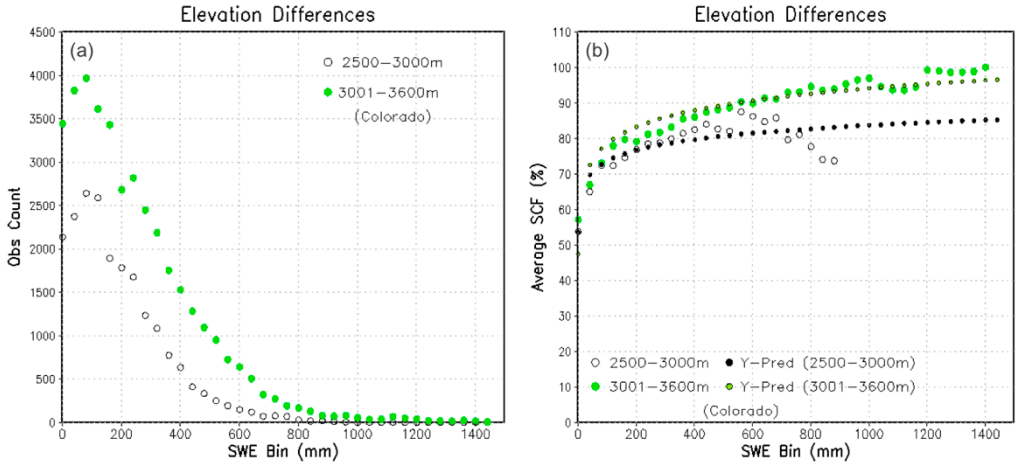


**Figure S6.** SDC relationships derived for two different elevation bands, 2500-3000m and 3001–3600m, from binned scatterplot approach for CO region. (**a**) The observation counts per bin are shown for each elevation band category along with (**b**) the bin-averaged SCF values overlaid with the predicted SCF values for each 40 mm SWE bin.

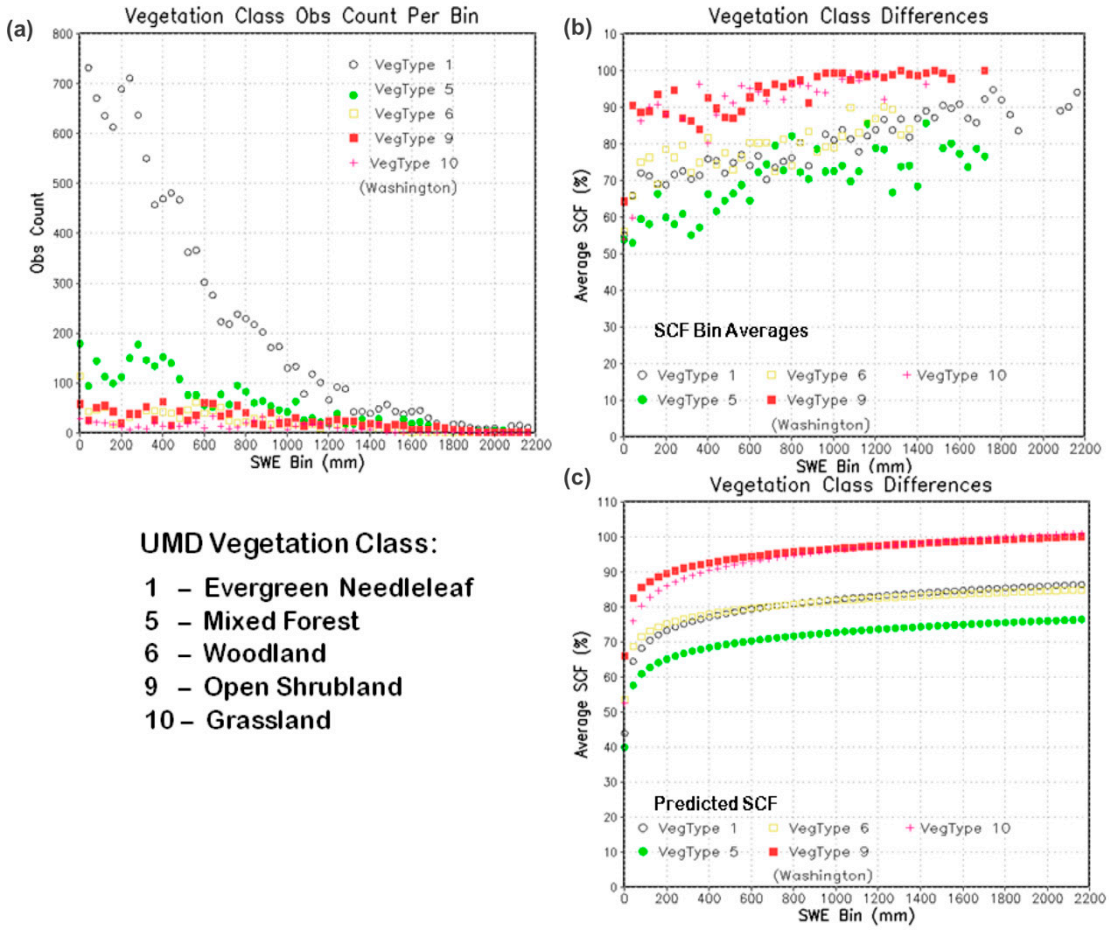


**Figure S7.** SDC relationships derived for different vegetation types from binned scatterplot approach for WA. (**a**) The observation counts per bin are shown for each vegetation category along with (**b**) the bin-averaged SCF values, and (**c**) the predicted SCF values for each 40 mm SWE bin from the logarithmic functions fitted to the scatter-bin points in plot (b).

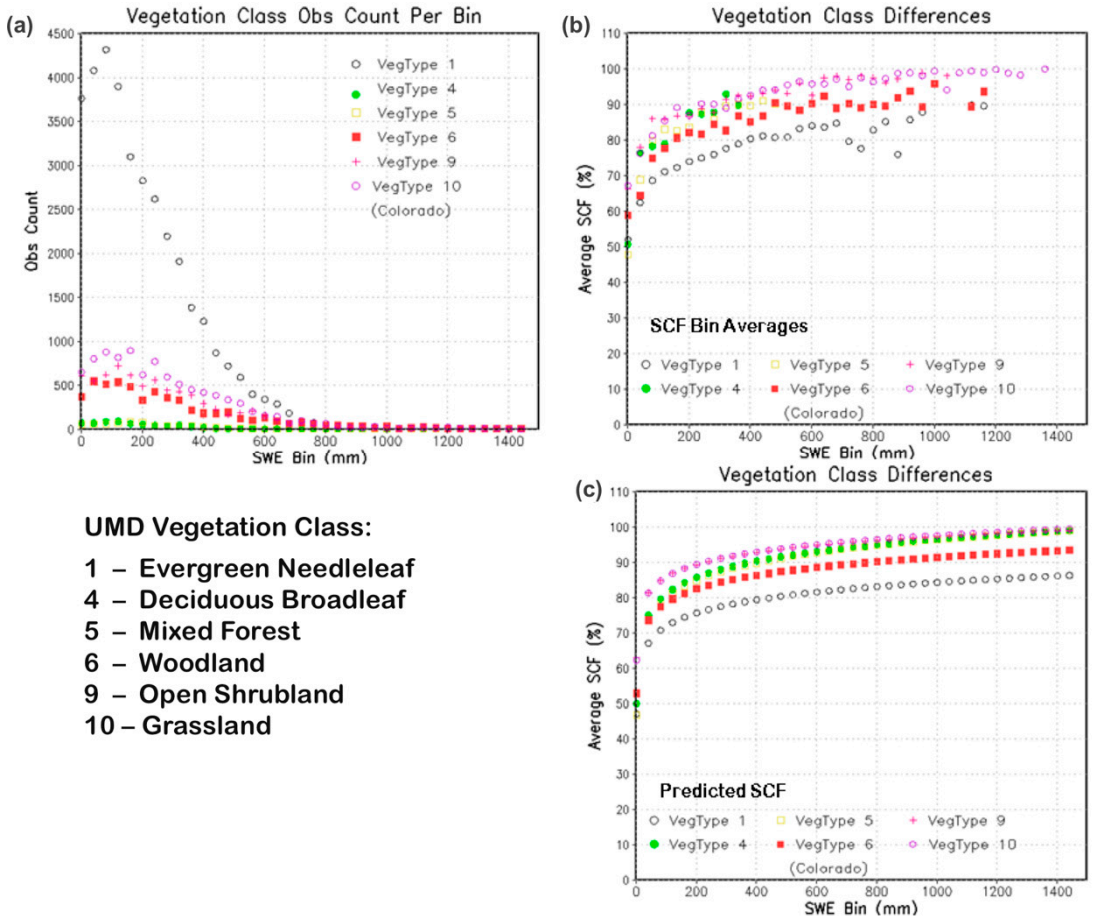


**Figure S8.** SDC relationships derived for different vegetation types from binned scatterplot approach for CO. (**a**) The observation counts per bin are shown for each vegetation category along with (**b**) the bin-averaged SCF values, and (**c**) the predicted SCF values for each 40 mm SWE bin from the logarithmic functions fitted to the scatter-bin points in plot (b).

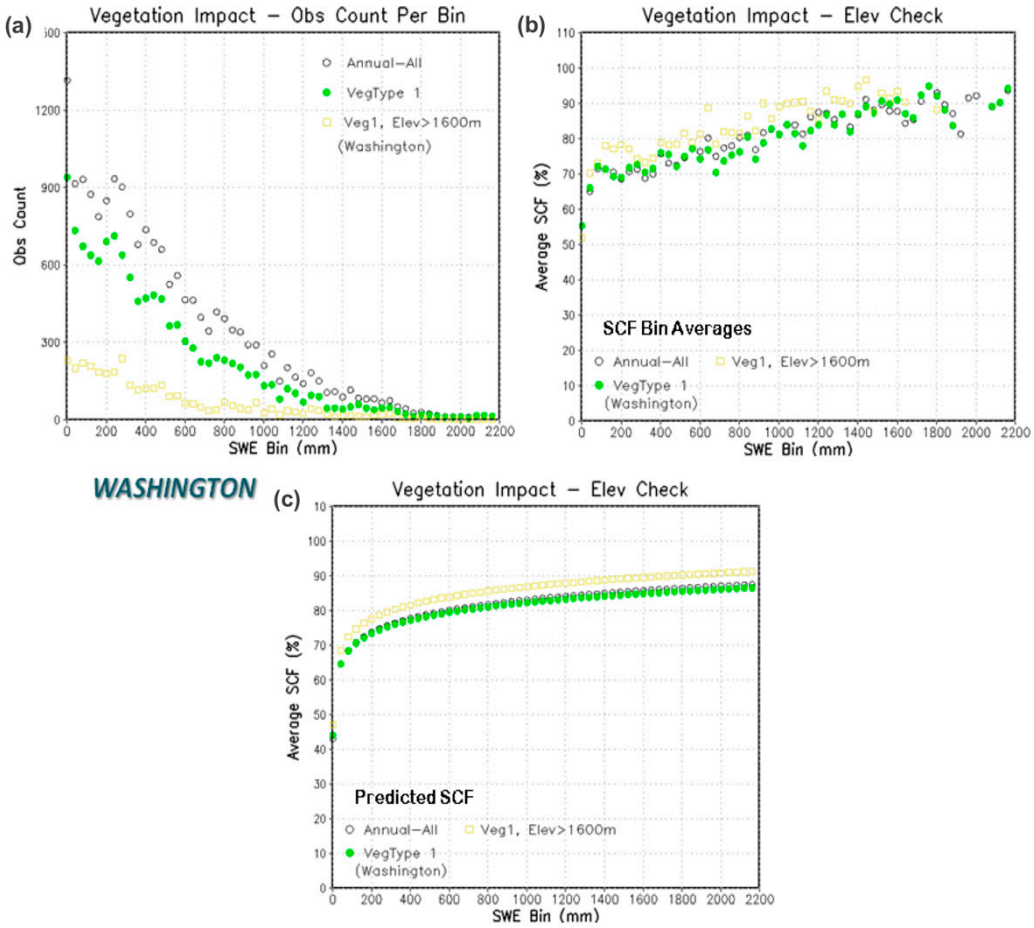


**Figure S9.** An independent check was performed on SWE bins used for SCF averages for the dominant vegetation type, needleleaf evergreen, to see if lower bin averages really do occur with such forest type. For WA domain, an elevation of 1600 m was chosen to generate an independent set of SWE bins above which SCF values are averaged if the evergreen needleleaf vegetation type is met (yellow boxes). The (**a**) counts, (**b**) binned scatterplots, and (**c**) predicted SCF estimates are compared with the original set (green circles) and with the annually based curves (open-circles).

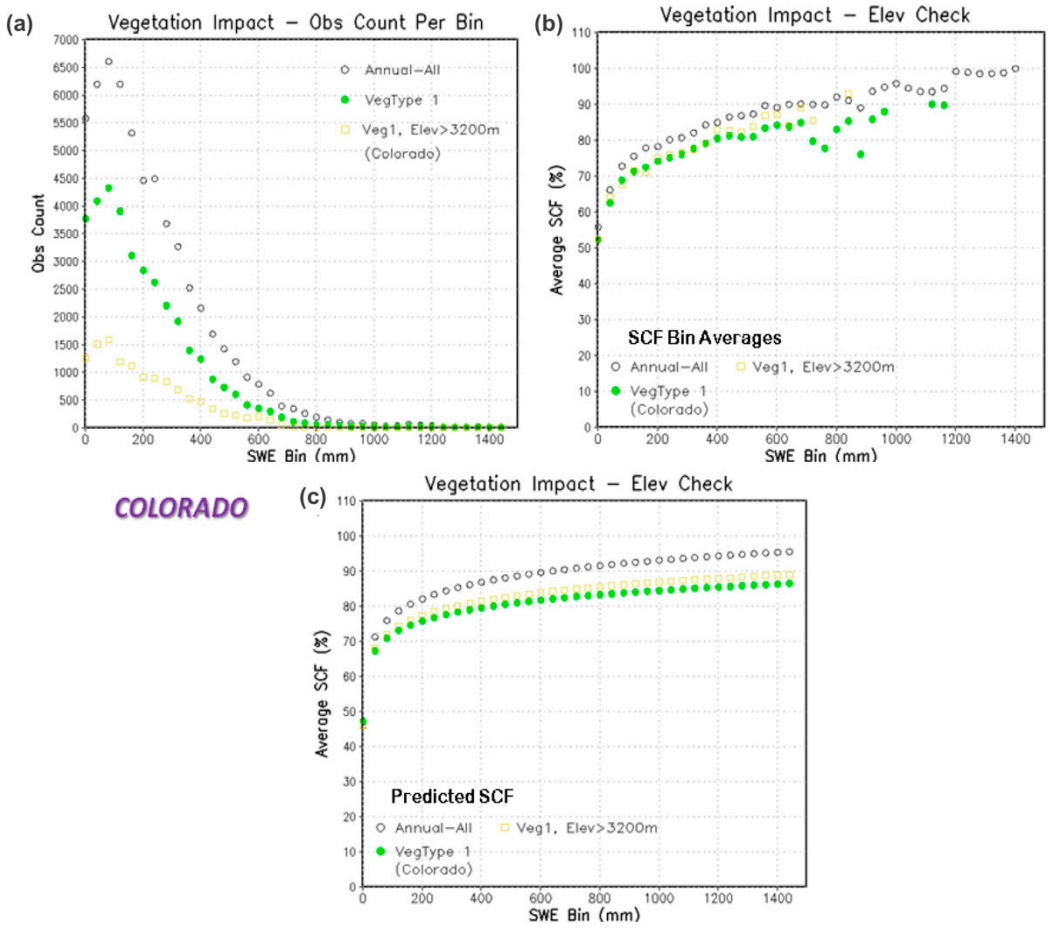


**Figure S10.** An independent check was performed on SWE bins used for SCF averages for the dominant vegetation type, needleleaf evergreen, to see if lower bin averages really do occur with such forest type. For the CO domain, an elevation of 3200 m was chosen to generate an independent set of SWE bins above which SCF values are averaged if the evergreen needleleaf vegetation type is met (yellow boxes). The (**a**) counts, (**b**) binned scatterplots, and (**c**) predicted SCF estimates are compared with the original set (green circles) and with the annually based curves (open-circles).

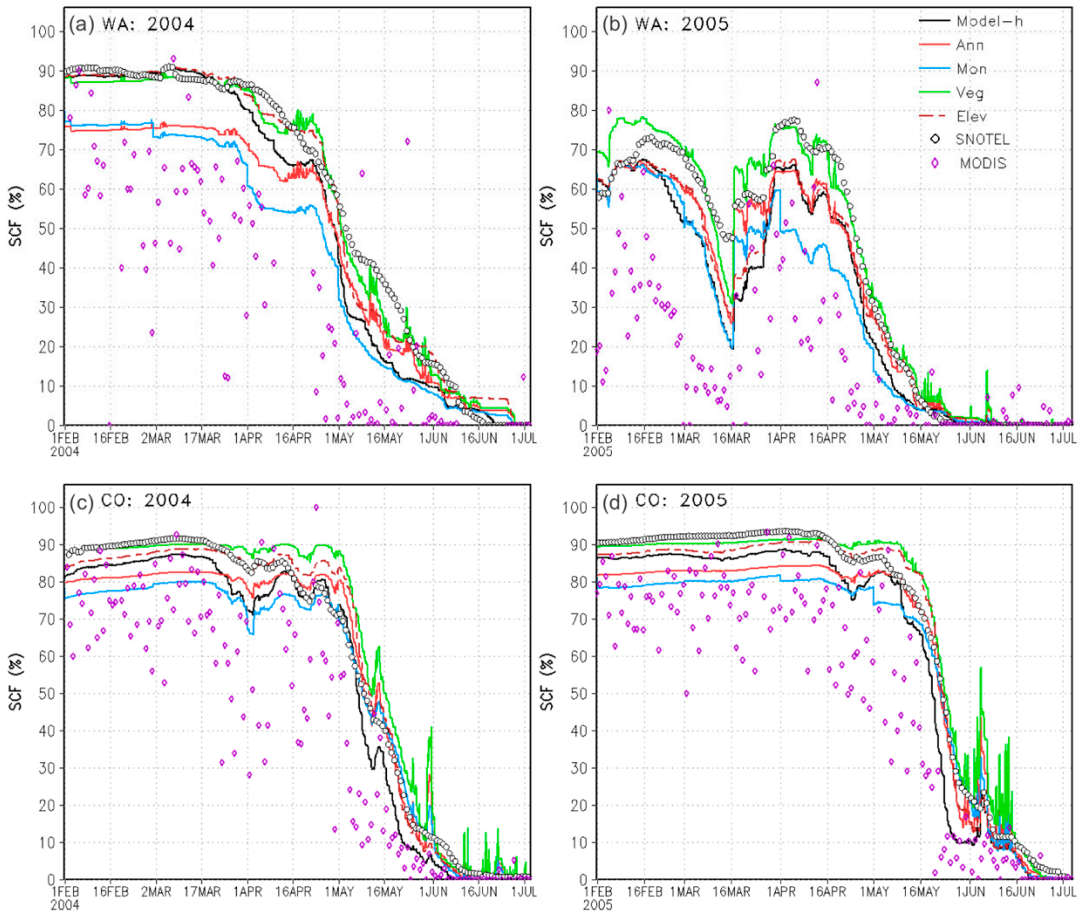


**Figure S11.** Comparison of spatially averaged predicted SCF (%) with the different observation operators and MODIS SCF observations (single pixel; purple diamonds) for the four different melt seasons for WA (**a,b**), an CO (**c,d**). The four EnKF experiments include the *model*-**h** EnKF experiment (black line), and the annual (red), monthly (blue), vegetation type (green), and elevation band (brown long-dashed) based *obs*-**h** experiments.

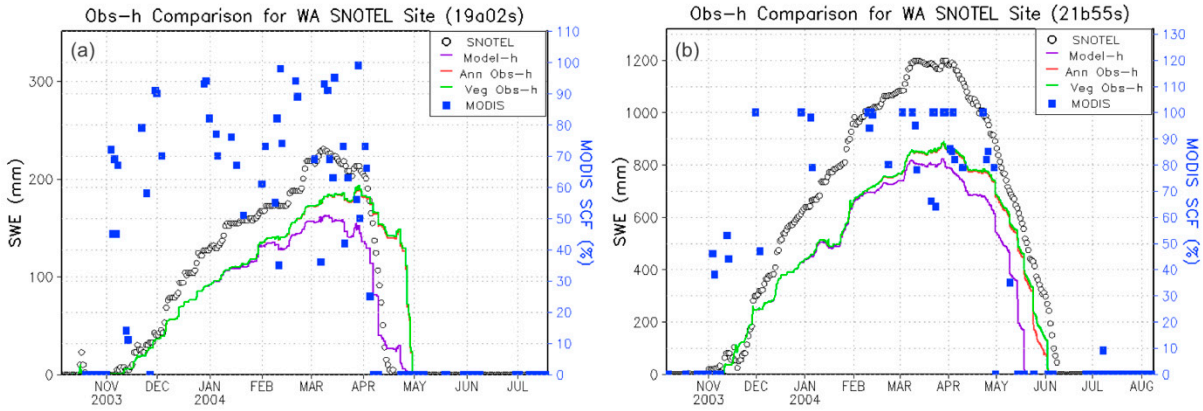


**Figure S12.** Comparison of the *model*-**h** (purple line), annual *obs*-***h*** (red line) and vegetation-based *obs*-***h*** (green line) based DA experiments with the MODIS SCF estimates (in %; blue squares) and SNOTEL SWE observations (in mm; open circle) for WY2004 at two WA SNOTEL sites: (**a**) Salmon Meadows (19a02s) and (**b**) Olallie Meadows (21b55s). Both sites are collocated with MODIS-based UMD evergreen needleleaf forest classes. MODIS SCF (%) values range from 0 to 100% and corresponds with the right y-axis (in blue).