R2 comments:

I reviewed the technical note “Accounting for snow in the estimation of root water storage capacity from precipitation and evapotranspiration fluxes”. The paper is well written and easy to follow but lacks a significant scientific contribution in its current form. I would first like to commend the authors for the open-source nature of the analysis and dataset.

R2.1: Thanks!

The paper addresses an important need, namely to quantify root zone water storage in places receiving snow, but has several limitations that concern me with its application. As I understand it, the main contribution of the paper is to improve the Wang-Eralandsson et al. (2016) approach in places with snow.

R2.2: Correct, that is the goal.

To do this the authors make three big assumptions.

R2.3: Before our point-by-point responses, we want to stress that we completely agree with the reviewer on their points regarding these processes. In particular, we agree that 1) vegetation may transpire when snow is on the ground, 2) snow may alter the timing of infiltration of precipitation into the root zone, and 3) runoff/leakage losses from the rooting zone can be significant and do not necessarily occur in a threshold “field-capacity” like manner.

In our point by point responses below, we outline how our contribution does not rely on these assumptions and instead, takes an identical approach to Wang-Erlandsson et al. (2016) to arrive at a conservative estimate or lower bound on the root zone storage capacity. Generally speaking, the method occasionally enforces zero values for Fout and Fin (for example, changing Fout=ET to Fout=0 when snow is present) to ensure that deficit calculations remain conservative in light of uncertainty in the magnitude and timing of fluxes used to determine Fout and Fin. For example, Wang-Erlandsson et al (2016) set runoff/leakage fluxes from the rooting zone to zero, not because runoff/leakage do not occur, but because the magnitude and timing of these fluxes are difficult to estimate with remotely sensed data products. We hope that our responses below clarify this strategy.

First that vegetation does not transpire when snow is on the ground.

R2.4: We do not assume vegetation does not transpire when snow is on the ground. On the contrary, the potential for vegetation transpiration when snow is on the ground is precisely our concern; it is a primary reason to set Fout=0 when snow is present (we note, as described in R2.3, that setting Fout=0 is not the same as claiming ET=0 when snow is present). For if an unknown snowmelt flux enters the rooting zone as vegetation transpires, one cannot be certain whether the transpiration flux results in an increasing storage deficit, or if the deficits are being constantly replenished by snowmelt. Consequently, one might *overestimate storage deficits by letting Fout=ET when snow is present*. Since we cannot know the magnitude of snowmelt flux into the rooting zone, it is therefore a more conservative choice (with regards to calculating the root zone storage deficit) to set Fout=0 when snow is present, thereby ensuring that an artifactual deficit of the type described above cannot accrue.

The potential to overestimate the storage deficit is a general concern stated in the method originally established by Wang-Erllandson et al (2016): When determining Fout and Fin to calculate root-zone storage deficits, it is important to make sure not to overestimate fluxes leaving the rooting zone (e.g., to overcount ET) or underestimate fluxes entering the rooting zone (to undercount P). In Wang-Errlandson et al (2016),the authors were primarily concerned with overestimation of deficits due to unaccounted-for irrigation, which would have the effect of underestimating inputs into the rooting zone.

In response to the Reviewer's collection of comments regarding flux assumptions, we propose to re-write the first paragraph of the Methods section to more clearly describe the strategies employed to ensure deficit calculations remain conservative in light of uncertainty in flux timing and magnitude:

"To estimate S_R, \citet{Wang_Erlandsson_2016} compute a running root-zone storage deficit (more positive means larger capacity in the subsurface for storage) using differences between fluxes exiting (F_{out}) and entering (F_{in}) the root zone during a given time interval (typically equal to the sampling period of the remotely sensed evapotranspiration dataset). Typically, F_{in} and F_{out} are set equal to precipitation (P) and evapotranspiration (ET), respectively. However, to obtain a conservative estimate of S_R (that is, a robust lower bound), it is important to make sure that F_{in} is not underestimated (when in doubt, assume all precipitation enters the rooting zone), and that F_{out} is not overestimated (when in doubt on the amount of F_{out} that contributes to increases in the root zone storage deficit, simply set $F_{out}=0$). This is a general strategy also employed in the original method developed by \citet{Wang_Erlandsson_2016}. In particular, the method occasionally enforces zero values for F_{out} and F_{in} to ensure that deficit calculations remain conservative in light of uncertainty in the timing or magnitude of fluxes; this is not equivalent to assuming that these fluxes are zero. For example, \citet{Wang_Erlandsson_2016} set runoff/leakage fluxes from the root zone to zero, not because runoff/leakage do not occur, but because the magnitude and timing of these fluxes are difficult to estimate with remotely sensed data products. “

Second, that snow doesn’t modify the timing or intensity of precipitation at a daily time step.

R2.5: As stated above, we agree that snow may alter the timing/intensity of fluxes entering the rooting zone (different from the arrival of precipitation); we do not assume otherwise. Because of this potential for (generally unknown) shifts in the timing of delivery of the precipitation into the root zone, from a deficit-calculation standpoint, the most conservative choice (that is, the choice that will definitely not undercount fluxes entering the root zone, which might lead to an overestimate of the actual deficit) is to assume that all precipitation enters the rooting zone as it arrives (even if it is snow and does not immediately enter the root zone), and then to not resume accruing a deficit until all snow from that precipitation event (and potentially from others) has melted.

And third, that runoff losses are minimal (or occur only at soil saturation).

R2.6: Please see proposed new wording in R2.4, and the response below, R2.11.

We know that these assumptions are inaccurate, as I detail below. Therefore, I challenge the authors to better quantify improvement using additional observations and an uncertainty analysis. The concern that a method can be overly simple and therefore, gives erroneous results at the time and spaces scales applied (i.e. daily and 4 km) should be dispelled by the authors. To this end, I try to give constructive ways forward in three major comments and also some minor points to help improve the readability of the paper.

R2.7: We appreciate the reviewer’s attempt to improve the readability of the manuscript.

Major comment 1: The paper uses the justification of making conservative estimates of storage for many of the decisions. However, it is unclear to me that deeper storage estimates are in fact more conservative. One can think of many situations where shallower storage steps would be more conservative, e.g. flood prediction.

R2.8: This is a good point. We agree that overestimating the size of the “bucket” might be problematic. As the reviewer points out, for conservative flood prediction, one might be safer erring on the side of a *smaller* bucket (thus potentially obtaining an overestimation of flood frequency and, presumably, more conservative management choices related to flood mitigation). However, we stress that our proposed method does in fact produce a smaller estimate of root-zone storage than the original method outlined by Wang-Erllandson et al (2019), and is therefore “more conservative” in the sense that the reviewer describes.

Perhaps a range of possible storages would be more consistent with the information available. It doesn’t matter how conservative or consistent the method is if it misinterprets important processes. To rephrase this concern, there are three major assumptions that the authors make in their formulation that need to be explored and discussed more. The first is that trees don’t transpire with snow cover. That is clearly not the case in the Sierra Nevada and more care needs to be taken in justifying this assumption. Evidence of substantial transpiration prior to snow disappearance in the Sierra Nevada are shown in the following papers. Royce, E.B., Barbour, M.G., 2001. Mediterranean climate effects. I. Conifer water use across a Sierra Nevada ecotone. Am. J. Bot. 88 (5), 911–918. https://doi.org/10.2307/2657044. Kelly, A.E., Goulden, M.L., 2016. A montane Mediterranean climate supports yearround photosynthesis and high forest biomass. Tree Physiol. 36 (4), 459–468. Cooper, A.E., Kirchner, J.W., Wolf, S., Lombardozzi, D.L., Sullivan, B., Tyler, S.W. and A.A. Harpold . Snowmelt causes different limitations on transpiration in a Sierra Nevada conifer forest. Agricultural and Forest Meteorology. 291. https://doi.org/10.1016/j.agrformet.2020.108089

R2.9: We thank the reviewer for these references. As noted in the detailed responses to the reviewer’s three concerns (R2.4, R2.5, and R2.11), we do not make these assumptions.

The second assumption is that snow doesn’t modify the timing of intensity of water Inputs. I find this assumption to be troubling, particularly because it is not discussed in any detail. While one can understand that this method might capture the overall/annual fluxes of water the daily dynamics are critical to how water is partitioned in the root zone (see next point). Here are some recent citations on how snow modifies water input timing and intensity: Yan, H., Sun, N., Wigmosta, M., Skaggs, R., Hou, Z., & Leung, R. (2018). NextâA˘ Rgeneration intensityâ ˘ A˘ Rdurationâ ˘ A˘ Rfrequency curves for hydrologic design in ˘ snowâA˘ Rdominated environments. Water Resources Research, 54, 1093– 1108. ˘ https://doi.org/10.1002/2017WR021290 Harpold, A. A., & Kohler, M. (2017). Potential for changing extreme snowmelt and rainfall events in the mountains of the western United States. Journal of Geophysical Research: Atmospheres, 122, 13,219– 13,228. https://doi.org/10.1002/2017JD027704

R2.10: We thank the reviewer for these additional citations. As noted in R2.5, we do not make this assumption.

My third concern is the assumption of zero outflow during snow covered periods, which is easily shown to be false with observational data. Additionally, these high elevation snow-covered zones are known to be runoff generation areas for the large downstream rivers. Consider some discussion of vertical drainage processes. It should be noted the method makes an implicit assumption that runoff is generated when the storage is full (i.e. saturation excess). Here is some relevant modeling work: Tague, C., and Peng, H. (2013), The sensitivity of forest water use to the timing of precipitation and snowmelt recharge in the California Sierra: Implications for a warming climate, J. Geophys. Res. Biogeosci., 118, 875– 887, doi:10.1002/jgrg.20073. Hammond, J. C., Harpold, A. A., Weiss, S., & Kampf, S. K. (2019). Partitioning snowmelt and rainfall in the critical zone: effects of climate type and soil properties. Hydrology and Earth System Sciences, 23(9), 3553-3570.

R2.11: We believe the reviewer is referring to the requirement that the deficit cannot be negative; that is, additional precipitation does not decrease the deficit below zero. It's important to note, however, that a zero deficit with this method does not imply storage is full or that runoff occurs. Instead, a zero deficit only means that storage must be equal to or greater than the value of storage when the deficit accrual began at $t=t_0$. Whether or not there is additional "space" in the root zone for storage to accrue is not something this particular method can resolve. In this regard, our approach is identical to that of Wang-Errlandson et al. (2016); the method makes no assumptions about how or when runoff/outflow events occur.

The previous application of this method was in drier locations where issues of lateral and vertical subsidy of water are less important. In order to remedy these concerns, the authors need to compare their findings against observational data or somehow quantify improvement (not just difference with the previous method). The authors should consider using a recent paper (Rungee et al., 2019) in Hydrological Processes that has data for validation in the Sierra Nevada. It is the author's responsibility to show that this method reproduces some parts of reality and does not introduce artifacts via these three assumptions. Rungee, J, Bales, R, Goulden, M. Evapotranspiration response to multiyear dry periods in the semiarid western United States. Hydrological Processes. 2019; 33: 182– 194. https://doi.org/10.1002/hyp.13322

R2.12: Originally, the method was applied globally (see the original Wang-Erlandsson et al. (2016) manuscript) and was not constrained to dry regions. Regarding the reviewer's reference to artifacts introduced via certain methodological assumptions, please see responses R2.4, R2.5, and R2.11.

Major comment 2: I am also concerned about the lack of uncertainty analysis given the reliance on modeled precipitation and evapotranspiration products. In particular, the authors should consider using multiple precipitation and evapotranspiration products in order to assess how much storage is sensitive to the products versus the method itself. For example, the previous work suggest that evapotranspiration uncertainty is more important than precipitation. However, these analyses were not done in wetter places like the Sierra Nevada. Precipitation products are inherently uncertain and need to be considered in the interpretation of the storage results. An easy alternative would be to use multiple precipitation products. Evapotranspiration is more challenging since the products are less validated, however, there are sufficient products to consider this or put arbitrary error terms on the existing data set. My concern is that this method portrays a certain level of sensitivity that will be specific to the climatology of the Sierra Nevada as well as the errors and uncertainties in the products themselves. Given that snowmelt strongly modifies the intensity of terrestrial water input, it's unclear to me how a non-explicit treatment of snow melt is sufficient for a paper that is trying to include snow in subsurface water storage. Again, I think some type of sensitivity analysis is needed to justify that this modulation of precipitation intensity is not a driver of soil saturation and excess runoff (related to major concern 1).

R2.13

- This contribution introduces an extension to an existing method, and provides the source code to carry out the method with any E or P dataset as improved datasets become available. While we provide an example of a deficit calculation using available P and ET datasets, it is not our intention to suggest that these are the only or best datasets, and it is outside of the scope of a methods-oriented Technical Note, we believe, to identify the best datasets. For these reasons, we do not conduct sensitivity analyses with multiple datasets.
- In the original manuscript introducing this general approach, Wang-Erlandsson et al. (2016) perform a global analysis across biomes and regions, including the Sierra Nevada, as well as tropical rainforest (e.g. the Amazon). We are not sure which study constrained only to more arid regions the reviewer is referring to.
- Soil saturation and excess runoff are not explicitly treated in this manuscript or the original Wang-Erlandsson et al. manuscript. We respond to this concern in R2.11.

Major comment 3: Consistency and differences with the Wang- Erlandsson et al. (2016) approach. I am a bit confused about the attempt to modify this previous work but also changing the methods. In particular, the previous work includes interception in the outgoing water flux term.

R2.14: It is true that the Wang-Erlandsson et al. (2016) ET dataset includes interception, but they themselves acknowledge that this is a drawback of their dataset, and that it would be more accurate to use transpiration and soil evaporation only (leaving out interception) in calculation of root-zone storage deficits. This makes sense; interception fluxes should not, technically, increase storage deficits in the root zone, as interception fluxes are sourced from above-ground, not below-ground, sources. Specifically, Wang-Erlandsson et al. (2016) state, “More sophisticated two-layer surface energy balance models also have the capacity to distinguish transpiration from other forms of evaporation. This implies that local root zone storage capacity can be computed, based on transpiration fluxes, which is preferred from a bio-physical point of view (although it would require estimate of interception evaporation to calculate effective precipitation). As new evaporation data sets become available, the SR estimates can easily be updated.”

In summary, as the reviewer notes, we do not include interception in our outgoing water flux term. However, we stress that this is a benefit, not a drawback, of the PML evapotranspiration dataset we used.

Additionally, the previous work corrects for loss terms by adding back in the difference between long-term precipitation and evapotranspiration. Again the justification of a conservative estimate is used but that is unclear in terms of what that means or whether these decisions support that.

R2.15: As Wang-Erlandsson et al. (2016) note, this correction technique can be applied in places where long run ET is greater than P, which will arise either due to biases in the underlying datasets and/or significant unaccounted for fluxes into the rooting zone (e.g. irrigation or inter-basin transfers of water). As opposed to making this correction, we instead opt for a more conservative approach, leaving these areas out of our analysis entirely, as it is not possible to determine the origin of the long-term imbalance in fluxes without additional information.

Minor comments: âA˘ c I am curious about the argument on line 5: If warming sce- ´ narios show decreased snowpack and increased rain as snow, wouldn’t this effectively make the Wang-Erlandsson et al. (2016) more viable over time? It seems like this proposed method is most important now and would likely diminish in its ability to represent what is happening as more precipitation shifts from snow to rain.

R2.16: The reviewer is correct: Without snow, our method would be unnecessary as it would reduce to the original method of Wang-Erlandsson et al. However, our hope is to improve estimates of root zone storage capacity now, as we are concerned with what stresses these systems are under currently. Moreover, these estimates of “true” subsurface S_R might help us (in a modeling context, for example) to better predict how snow-dominated systems might cope with only subsurface storage in a future with decreased snowpack.

Lines 27-29: It would be helpful to understand how other estimates rely on only soil moisture and what sets this method apart in partitioning plant-accessible water. I’m not clear on how this marks a distinction.

R2.17: We only intend to point out that most modeling and analysis approaches only account for plant water storage in soils. The distinction here is that we do not constrain our analysis to upper soil layers, typically within the upper 1.5 m of the subsurface in most available soils datasets (e.g. gNATSGO).

Selection of fSCA threshold– Line ´ 90: I don’t fully understand the selection criteria for the threshold below which snow contribution to the pixel isn’t assumed to be meaningful. 10% seems sort of arbitrary aside from being the uncertainty for MODIS. What if that 10% is deep and holds significant water? Would be great to see more justification/explanation of how to derive this threshold since it exerts powerful control over Eq. 3. How sensitive is the result on your threshold? Uncertainty based on the selection of a threshold. How much impact does the selection of a threshold have and if its user defined, how much would that swing the significance of the results as presented in Figure 1?

R2.18: The threshold is chosen primarily as a function of the underlying snow cover dataset, which has a precision of 10%. Like all other choices in this manuscript, we have opted for the conservative choice setting the threshold to the lowest non-zero value available in the dataset. In this way, our method represents an end-member case.

4) Assumption about groundwater losses/contributions how much does this assumption alter the bound if it looks fairly reasonable to assume that there are significant losses/contributions on a daily timestep in this particular region? Enzminger, T. L., Small, E. E., & Borsa, A. A. (2019). Subsurface water domi- nates Sierra Nevada

seasonal hydrologic storage. Geophysical Research Letters, 46, 11993– 12001. https://doi-org.unr.idm.oclc.org/10.1029/2019GL084589

R2.19: This is a reasonable concern, but it is beyond the scope of this technical note (that is concerned with the role of snow), which extends the method by Wang-Erlandsson et al. (2016) that is subject to the same problems if lateral transport is a major factor. Proper exploration of this issue would require estimates of inter-pixel transfers of groundwater.

Definition ´ of winter months –line 115: Justification of the time-period selected (January- April) for analysis. Why not expand to include Oct – June since that would be more reflective of a true winter especially with the focus on higher elevation sites?

R2.20: We do not constrain analysis to this time period alone (as noted in the manuscript, it is run continuously from 2003 to 2017). The choice to show average snowpack in January-April is purely for the purposes of illustration, to demonstrate general patterns of snow cover that might be expected to impact the underlying calculations of the method.

Line 135: I'm still ´ having a hard time grasping why this method is more important in a warming climate with more precipitation as rain. It seems like those are the exact conditions where Wang-Erlandsson is valid and where you have minimal differences (30 mm) in Figure 1. As we shift to rain dominated systems with less snow, it seems like it would reduce the importance of this contribution, not enhance it. âA˘ c Line 110: Smax not previously ´ introduced

R2.21: Please see our explanation in R2.16.