Reviewer 2

Review of " Assimilation of airborne gamma observations provide utility for snow estimation in forested environments" by Cho et al.

SUMMARY:

Overall, I think this paper is relevant for publication in HESS with a clear presentation. This paper leverages airborne gamma observations to estimate snow water equivalent using data assimilation. This is great work for the snow community that shows the potential of remote-sensed gamma observations to improve snow estimates. That said, clarification of how localization is implemented with DA would be helpful for others to repeat your work and interpolate the results.

Thank you for your positive feedback and the valuable comments on our manuscript. We have carefully revised our manuscript based on your comments.

Line-by-line comments:

1. L125: For Equations (1) to (4), make sure to use the same format for the uncollided gamma count rates (e.g., $40K_b$ $^{40}K_b$).

Thank you for your keen eye. We edited this.

2. L138: From my understanding, COOP snow depth is first converted to SWE which is assimilated (rather than snow depth) to get UA SWE. Please make sure this sentence won't cause any confusion.

We agreed with the comment. The sentence was edited to clarify the data development procedure as below.

"The UA SWE is the ground observation-based 4-km gridded SWE product developed by consistently assimilating the snow telemetry (SNOTEL) SWE and NWS Cooperative Observer Program (COOP) snow depth measurements (which was first converted to SWE using a newly developed snow density parameterization) with the Parameter-elevation Regressions on Independent Slopes Model (PRISM) temperature and precipitation data over the continental United States"

3. L186: For MERRA2, are bias-corrected precipitation or uncorrected precipitation used as inputs? In line 308, the authors mention that overestimated SWE is likely attributed to precipitation phase partition. Would bias from precipitation contribute to the overestimation?

We used the uncorrected MERRA-2 forcing in the original manuscript because we focused on demonstrating the feasibility of the gamma SWE DA to improve the model estimates of SWE, particularly in forested areas, using the atmospheric forcing as is (i.e., without bias-correction), which is a typical case of operational prediction or monitoring systems. However, in order to analyze the effect of the atmospheric forcing on the snow accumulation, we conducted additionally experiments using the bias-corrected MERRA-2 forcing (as well as different parameterization schemes to address another Reviewer's comments). The figure below shows that the use of bias-corrected MERRA-2 forcing was effective in improving the SWE estimates during the snow accumulation period as compared to the case using the original (i.e., uncorrected) MERRA-2 forcing. Here, it is also worth to note that assimilation of the gamma SWE data (using the uncorrected MERRA-2 forcing) provides similar SWE estimates to the case of using the bias-corrected forcing when the gamma SWE observations are available during the snow accumulation period.

We included these analyses in the revised manuscript as follows:

"These issues can be attributed to parameterization schemes and/or atmospheric forcing employed in Noah-MP. Parameterization options for the precipitation phase partitioning method, ground surface albedo, surface layer drag coefficient, and snow/soil temperature time scheme can affect the snow simulations (You et al., 2020). To further analyze the issues, we conducted additional experiments using different parameterization schemes and atmospheric forcing. That is, the BATS scheme for partitioning precipitation into rainfall and snowfall, CLASS scheme for ground surface albedo, Chen97 scheme for surface layer drag coefficient, fully-implicit snow and soil temperature time scheme, and the bias-corrected MERRA-2 forcing were additionally tested. As shown in ***Figure 8a****, the use of BATS or CLASS schemes do not make a significant difference in the SWE estimates as compared to the original OL results. Although the Chen97 or fully-implicit schemes are effective in delaying the snow removal date, they add considerably more snow during the snow accumulation period and do not help capture snowmelt start date* ***(Figure 8a)****. Furthermore, the effectiveness of each parameterization scheme varies with flight lines and time periods within the study domain as also emphasized by You et al. (2020).*

Figure 8b *shows that the use of the bias-corrected MERRA-2 forcing is effective in improving the SWE estimates during the snow accumulation period, but it still has the issue of rapid snow melting. The combinational use of the bias-corrected MERRA-2 forcing and fully-implicit scheme leads to improved snow removal timing, but largely overestimated SWE during the snow accumulation period. We originally used the uncorrected MERRA-2 forcing to demonstrate the feasibility of the gamma SWE DA for improving the model estimates of SWE, particularly in forested areas, using the atmospheric forcing as is (i.e., without bias-correction), which is a typical case of operational prediction or monitoring systems. Here, it is worth to note that assimilation of the gamma SWE data provides similar SWE estimates to the case of using the bias-corrected forcing with semi-implicit scheme when the gamma SWE observations are available during the snow accumulation period."*

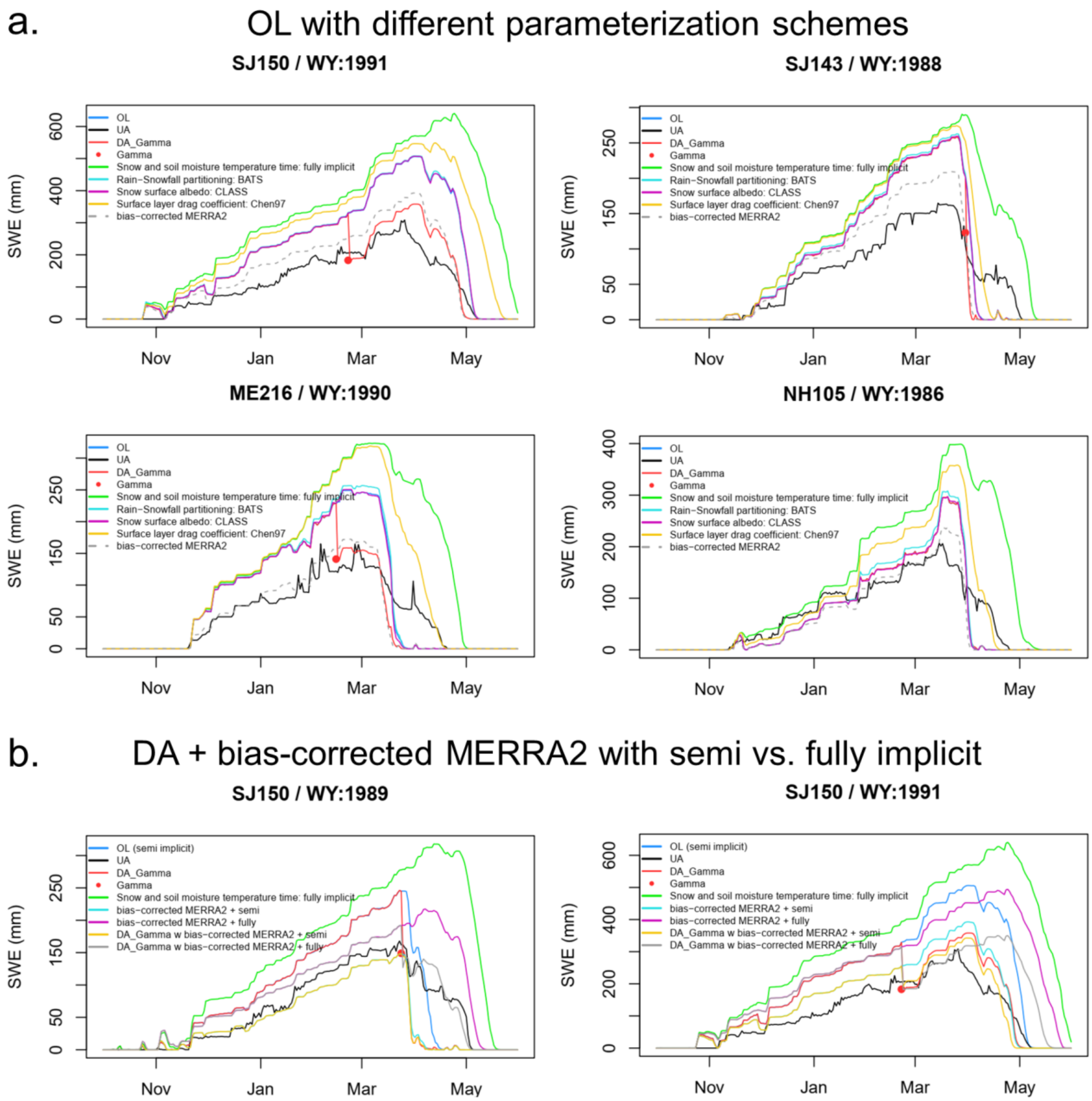

Figure 8. Examples of SWE time series including (a) the additional open-loop (OL) experiments using different parameterization schemes related to snow simulations such as the snow/soil temperature time scheme (semi-implicit vs. fully-implicit), partitioning precipitation phase (Jordan91 vs. BATS), ground surface albedo (BATS vs. CLASS), and surface layer drag coefficient (Monin-Obukhov [M-O] similarity theory vs. Chen97 [original Noah]), and (b) DA runs forced by original vs. bias-corrected MERRA2 forcings with each snow/soil temperature time scheme which is a parameterization option largely affecting snow simulations**.**

4. L196: Could you briefly summarize how Kwon et al. (2021) perturbs the forcings? For example, which forcings are perturbed? How are the parameters chosen? (i.e., Were in situ observations used to

quantify these parameters?) This would be helpful to know if precipitation uncertainties are considered.

Perturbation parameters applied during the OL and DA runs are summarized in the following table, which was included in the revised manuscript. The perturbation parameter values in Kwon et al. (2021) were determined based on previous DA studies (e.g., Forman et al., 2012; Kumar et al., 2009; 2014; 2016; Reichle et al., 2008).

Table 1. Perturbation parameters applied to model prognostic state variables and atmospheric forcing fields during the OL and DA runs.

Variable	Perturbation Types	Std dev	AR(1)	Cross correlations		
Model prognostic state variables				*SWE*	*SD*	
SWE	M	0.01	3 hr	–	0.9	
Snow depth (*SD*)	M	0.02	3 hr	0.9	–	
Atmospheric forcing fields				*SW*	*LW*	*P*
Shortwave radiation (*SW*)	M	0.3	1 day	–	-0.5	-0.8
Longwave radiation (*LW*)	A	50 W m^{-2}	1 day	-0.5	–	0.5
Precipitation (*P*)	M	0.5	1 day	-0.8	0.5	–

M: multiplicative; A: additive; AR(1): first-order autoregressive temporal correlation.

*** References**

Forman, B.A., Reichle, R.H., & Rodell, M.: Assimilation of terrestrial water storage from GRACE in a snow-dominated basin. Water Resources Research, 48(1), W01507, https://doi.org/10.1029/2011WR011239, 2012.

Kumar, S.V., Reichle, R.H., Koster, R.D., Crow, W.T., & Peters-Lidard, C.D.: Role of subsurface physics in the assimilation of surface soil moisture observations. Journal of Hydrometeorology, 10 (6), 1534–1547, https://doi.org/10.1175/2009JHM1134.1, 2009.

Kumar, S.V., Peters-Lidard, C.D., Mocko, D., Reichle, R., Liu, Y., Arsenault, K.R., Xia, Y., Ek, M., Riggs, G., Livneh, B., & Cosh, M.: Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. Journal of Hydrometeorology, 15 (6), 2446–2469, https://doi.org/10.1175/JHM-D-13-0132.1, 2014.

Kumar, S.V., Zaitchik, B.F., Peters-Lidard, C.D., Rodell, M., Reichle, R., Li, B., Jasinski, M., Mocko, D., Getirana, A., De Lannoy, G., Cosh, M.H., Hain, C.R., Anderson, M., Arsenault, K.R., Xia, Y., & Ek, M.: Assimilation of gridded GRACE terrestrial water storage estimates in the North American Land Data Assimilation System. Journal of Hydrometeorology, 17 (7), 1951–1972. https://doi.org/10.1175/JHM-D-15-0157.1, 2016.

Reichle, R.H., Crow, W.T., & Keppenne, C.L.: An adaptive ensemble Kalman filter for soil moisture data assimilation. Water Resources Research, 44 (3), W03423, https://doi.org/10.1029/2007WR006357, 2008.

5. L215: It is not clear to me how localization is applied in the assimilation. Does it weigh the covariance matrix? It might be better to link equation (7) with the relevant equation mentioned above.

I might not fully understand it, but why localization used to update SWE estimates would impact the open loop results shown in figure 6? I assume localization would only impact DA SWE.

The localization is applied in the assimilation by weighting distances from the flight lines (up to a specified localization distance; *r*) using the Gaussian decay-based function. The degree of the SWE updates for a grid cell from the assimilation is calculated using a localization weight (W) which is calculated based on the distance (*d*) from the updated grid cells overlapped with the flight line.

Yes, the localization only impacts DA SWE. The reason why the OL statistics were changed with increasing the localization distances in Figure 6 is that the effective distance used to calculate the domain-averaged OL SWE was changed according to the localization distance. For example, at a localization distance of 4 km, the domain-averaged OL SWE is calculated using gridded SWE values within an effective surrounding area (4 km distance from the gamma flight lines). At 32 km, the domain-averaged OL SWE was calculated based on larger number of the gridded SWE values within surrounding areas up to 32-km distance from the gamma flight line. The OL/DA statistics in Figure 6 are calculated using domain-averaged time series of OL/DA SWE over the effective surrounding areas by localization distances with the corresponding UA SWE.

We edited and added descriptions in Section 4.3 as below.

"To quantify if the spatially sparse gamma SWE observations can improve the SWE estimates in the surrounding areas, where the observations are not available, we apply a distance-based localization method into the assimilation procedure. The localization is applied in the assimilation by weighting distances from the flight lines (up to a specified localization distance; r) using the Gaussian decay-based localization method as follows:

$$W = exp\left\{\frac{-d^2}{2\cdot\left(\frac{r}{2}\right)^2}\right\} \tag{7}$$

where d is the distance between the updated grid cells (i.e., flight lines) and grid cells without observations within a specified localization radius r. The degree of the SWE updates for a grid cell from the assimilation is calculated using the localization weight (W) which is calculated based on the distance (d) from the updated grid cells overlapped with the flight line."

When double-checking our calculation for Figure 6, we found a script error when calculating the error metrics. The error metrics were miscalculated using a "static" domain-averaged OL SWE time series with DA SWE time series at different localization distances. They should have been calculated with a domain-averaged OL SWE time series at a "corresponding" localization distance. We reviewed throughout the script, corrected the part, and recalculated the metrics which are shown in the new figure below.

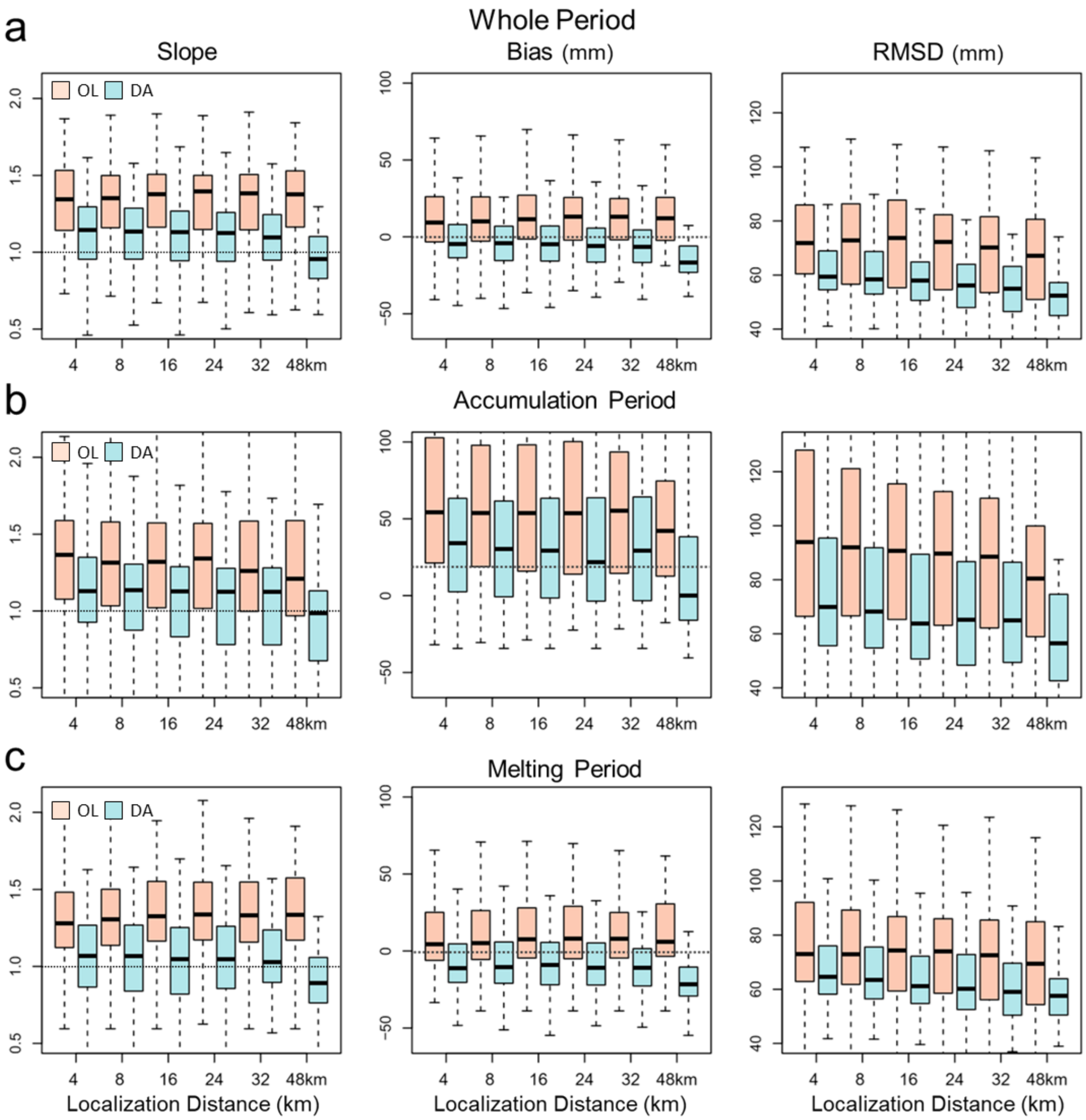

Figure 6. Localized data assimilation (DA) and open-loop (OL) Noah-MP SWE performances as compared to the UA SWE with different localization distances (e.g., 4, 8, 16, 24, 32, and 48 km) for the whole (accumulation and melting periods), accumulation, and melting periods, respectively.

The result parts of the manuscript related to the revised figure 6 were also revised as below.

"For the whole snow season that includes both accumulation and melting periods, the boxplot of the 1:1 slope shows that the localized DA SWE were improved as compared to OL. The slopes of the DA SWE are closer to 1 than the OL's slopes. The bias and RMSD boxplots also show that the DA SWE has lower errors than the OL SWE for all localization distances, expect for bias at 48 km which is too low (median: – 23 mm). The OL's RMSDs slightly increased at the distances up to 16 km (median: 72 mm) and decreased after that, while the DA's RMSD values continually decreased with increasing the distances up to 48 km (median: 53 mm). When the statistics were calculated for the accumulation and melting periods separately, the lower RMSDs and slopes closer to 1 of the localized DA SWE were found consistently. As previously discussed, the efficacy of assimilating the airborne gamma SWE is greater during the

accumulation period, especially for bias and RMSD, than during the melting period. In the melting period, the improvements in the bias, RMSD and 1:1 slope are also achieved up to 32 km"

6. L226: maybe use lowercase "A" in the parenthesis.

Edited.

7. L236: it seems peak SWE might not be correctly estimated if only one data point exists after the accumulation season (Figure 4 SJ150 in WY 1989 and NH106 in WY 1997). It might be worth pointing that out and/or discussing this issue.

We agreed with your comment. To point out the issue, we added a statement as below.
"Also, the peak SWE cannot be corrected if a single gamma SWE exists only after the accumulation period."

8. Please be consistent with either RMSD or RMSE throughout the manuscript.

We edited and consistently used RMSD throughout the manuscript.