Reply to the Comments by Referee #2 for Manuscript tc-2019-15

In this study, the authors tried to address the uncertainty in predicting the Eurasian snow by inter-comparing the performances of four land surface models coupled to WRF. They have four goals to achieve but the purpose of the work is not clearly articulated. The manuscript can be accepted to be published only after some major concerns have been addressed:

⇒ The authors appreciate the referee's valuable comments on this study, which made significant improvement in the manuscript. We have made our best effort to revise the manuscript based on the referee's comments and suggestions. In the following, we made an item-by-item response to the specific comments by the referee.

(1) *Many studies show errors in the input and validation data, rather than model formulation, seem to be the greatest factor affecting model performance. The manuscript lacks of discussion of the quality for those "observation" used to evaluate the model performance.*

⇒ We appreciate the referee pointing this out. We used these data because they have been widely used for the validation purpose as seen in many studies. Thus we assume that we can use these data without further quality check. We added a description on the quality of observations used in our study, based on the followings:

1) Cooper et al. (2018) assessed snow extent data sets including both CMC data and MODIS Terra data by assuming the surface observations are truth. They compared seven snow extent data sets – CMC, IMS, MAIAC AQUA, MAIAC TERRA, MODIS AQUA, MODIS TERRA, and NISE. They found that CMC represented strong agreement with in situ observations with F score 0.81 while MODIS TERRA represented medium agreement with F score 0.54. More specifically, they calculated accuracy, precision, recall, and F score as the following:

$$\begin{aligned}\text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN}\\ \text{Precision} &= \frac{TP}{TP+FP}\\ \text{Recall} &= \frac{TP}{TP+FN}\\ F &= 2\times\frac{\text{precision}\times\text{recall}}{\text{precision}+\text{recall}}\end{aligned} \tag{1}$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative. The assessment results are summarized in Table R1.

2) The Canadian Meteorological Centre (CMC) snow depth data is the most widely used data in many model evaluation studies, being considered as the surface truth value (e.g., Niu and Yang 2007; Reichle et al., 2011; Yang et al., 2011; Liu et al., 2013; Kumar et al., 2014, 2015; Zhou et al., 2014; Dawson et al., 2016; Toure et al., 2016).

3) The MODIS/Terra Snow Cover Monthly L3 Global 0.05Deg CMG, Version 6 is also very widely used data in validating many model outputs and improving model

Table R1. Evaluation of daily snow extent data set performance for 2015. The GHCN-D surface observations are used as "truth". All products are regridded to a common 4 km resolution. The highest value for each metric is shown in bold. From Cooper et al. (2018).

	Accuracy	Precision	Recall	F
CMC	0.91	0.79	0.83	0.81
IMS	0.93	0.87	0.83	0.85
MAIAC AQUA	0.91	0.90	0.74	0.82
MAIAC TERRA	0.91	0.90	0.75	0.82
MODIS AQUA	0.76	0.51	0.43	0.46
MODIS TERRA	0.82	0.69	0.45	0.54
NISE	0.84	0.83	0.45	0.58

prediction skill (Drusch et al., 2004; Rodell and Houser, 2004; Parajka and Bloschl, 2008; Hall et al., 2010; Nester et al., 2012; Hancock et al., 2013; Zhou et al., 2014).

(2) *Among four goals of this paper only the third one has been addressed sufficiently. The first two haven't been explained clearly. The reader can easily understand the difference between four models if a table is used to show the different treatment of snow albedo, snow density, snow compaction, snow interception, snow age, and etc. The authors need to make some sensitivity tests like the different microphysics schemes to address and make any conclusion on the last goal of this paper.*

⇒ We appreciate these valuable suggestions by the referee. Following the reviewer's suggestion, we have added a table in the revised manuscript to show the characteristics of each model (see Table R2 below):

Table R2. The general feature of each LSM.

Snow	Noah LSM (Liveh et al., 2010)	Noah MP (Niu et al., 2010)	RUC LSM (Benjamin et al., 2004)	CLM4 (Oleson et al., 2010)
Layers	1	3	2	5
Density	Fixed	Calculates variable snow density	Calculates variable snow density	Calculates variable snow density
Liquid	X	O	O	O

⇒ We have also followed the referee's suggestion to do some sensitivity tests, e.g., using different microphysics schemes, to address the last goal of this paper. Although we could not perform extensive tests, we have done some sensitivity experiments using other microphysics scheme. For example, the present study employed the WSM3 scheme, and we conducted sensitivity experiments using the WSM6 scheme. The results indicated that all the LSMs represented better performances with WSM6 than WSM3. We have included such sensitivity results and discussions in the revised manuscript to address the last goal of this study.

(3) *For fair inter-comparison the forcing should be as close as possible. For snow predictions how to differenciate snowfall or rainfall from the total precipitation is critical. This process should be mainly determined by microphysics scheme. The authors should use the same MP for all models to determine the amount of snowfall and rainfall rather than the empirical method by each model.*

⇒ We agree with the referee on the statement "*For snow predictions, how to differenciate snowfall or rainfall from the total precipitation is critical*". But we cannot fully accept the referee's suggestion to use the same MP for all models to determine the amount of snowfall/rainfall. We believe that each model has its own characteristics, including the way to classify the total precipitation into liquid versus solid forms. Although these schemes might be based on empirical methods, we believe that they are devised to balance with other physical/dynamical processes in the model; thus, formulating their own characteristic ways to calculate each process in the model. The referee's speculation may be right in terms of fair intercomparison; however, by doing so, we are afraid that the model balance would be broken and/or the model characteristics might be lost. Although we respect the referee's suggestion, it is another important thing in this study to examine the characteristic behavior of each model, when coupled to a regional climate model; thus, we decided to use the MP as it is in each model.

(4) *For the most part, snow models are built on similar principles. The greatest differences are found in how each model parameterizes individual processes (e.g., surface albedo and snow compaction). Parameterization choices naturally span a wide range of complexities. Ensemble is a promising way to reduce these uncertainties. It would be greatly beneficial to the title of this paper if the authors can also compare the performance of the ensemble mean of four models.*

⇒ We agree to the referee's comment. Due to the uncertainty involved in parameterizations, it is desirable to employ an ensemble approach. On the other hand, we think that the characteristics of each LSM, coupled to a regional climate model, may be better understood through deterministic approach and various sensitivity experiments. The referee's suggestion on the ensemble approach is fantastic, but it is left for a further study as the next step. We do appreciate this valuable advice by the referee.

(5) *Noah MP itself has so many options to choose. Many of them can have significant impact on the snow prediction. The authors should have optimal options before comparing it to the other three models.*

⇒ We think the referee pointed out an important issue. As the referee mentioned, the Noah-MP has many parameterization schemes in calculating leaf are index (dveg), surface layer drag coefficient (opt_sfc), canopy stomal resistance (opt_crs), snow surface albedo (opt_alb), frozen soil permeability (opt_inf), supercooled liquid water (opt_frz), radiation transfer (opt_rad), partitioning of precipitation to snowfall and rainfall (opt_snf) and, runoff and ground water (opt_run) (more details in the Noah-MP technical description note, `http://www.jsg.utexas.edu/noah-mp/files/Noah-MP_Technote_v0.2.pdf`).

We have performed some preliminary tests with multiple scheme combinations – 3 options in dveg, 2 options in opt_rad, and 1 option in opt_alb. Table R3 summarizes the best results from each scheme in terms of snow depth and RMSE. We noticed that the accuracy of snow depth prediction is highly influenced by the vegetation option compared to other options. Turning off the dynamic vegetation option and calculating the vegetation fraction inside Noah-MP results in the highest accuracy in predicting the Eurasian snow.

Note that finding the optimal scheme set requires an enormous amount of computing resources. The Noah-MP has 9 parameterization schemes, each having 2 4 options – 4 schemes with 2 options, 3 schemes with 3 options and 2 schemes with 4 options –

a total of 6912 combinations. This implies that we should run the model 6912 times to find the optimal set of parameterization schemes/options. In our study, as the Noah-MP is coupled to WRF, the computation time will tremendously increase to find the optimal scheme set, which is almost impossible. Therefore, we decided to adopt the default set of parameterization schemes/options in Noah-MP by assuming that this default set is the most general setting.

Table R3. The result of physical snow depth (m) and the RMSE of snow depth for each option. The name of options in this table "albedo", "radiation", "vegetation" mean that we modulated these options.

	Default	Albedo	Radiation	Vegetation
Snow Depth [m]	0.200 ($\pm$0.060)	0.203 ($\pm$0.062)	0.193 ($\pm$0.058)	0.193 ($\pm$0.058)
RMSE	0.173	0.174	0.166	0.165

(6) *The quality of figure need to be improved. Some of them are very blurry and hard to read. Like figure 4a only 6 curves can be seen.*

$\Rightarrow$ We appreciate the careful check and suggestion by the referee on the quality of figures. We carefully checked all the figures and improved them by uploading the high-resolution files.

References

Cooper, M. J., R. V. Martin, A. I. Lyapustin, and C. A. McLinden: Assessing snow extent data sets over North America to inform and improve trace gas retrievals from solar backscatter, Atmospheric Meas. Tech., 11, 2983-2994, `https://doi.org/10.5194/amt-11-2983-2018`, 2018.

Dawson, N., P. Broxton, X. Zeng, M. Leuthold, M. Barlage, and P. Holbrook: An evaluation of snow initializations in NCEP global and regional forecasting models, J. Hydrometeor., 17, 1885-1901, 2016.

Drusch, M., D. Vasiljevic, and P. Viterbo: ECMWFs global snow analysis: Assessment and revision based on satellite observations, J. Appl. Meteorol. Climatol., 43, 1282-1294, https://doi.org/10.1175/1520-0450(2004)043<1282:EGSAAA>2.0.CO;2, 2004.

Hall, D. K., and G. A. Riggs: Accuracy assessment of the MODIS snow products, Hydrol. Process., 21, 1534-1547, `https://doi.org/10.1002/hyp.6715`, 2007.

Hall, D. K., G. A. Riggs, J. L. Foster, and S. V. Kumar: Development and evaluation of a cloud-gap-filled MODIS daily snow-cover product, Remote Sens. Environ., 114, 496-503, 2010.

Hancock, S., R. Baxter, J. Evans, and B. Huntley: Evaluating global snow water equivalent products for testing land surface models, Remote Sens. Environ., 128, 107-117, `https://doi.org/10.1016/j.rse.2012.10.004`, 2013

Kumar, S., and Coauthors: Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation, J. Hydrometeor., 15, 2446-2469, https://doi.org/10.1175/JHM-D-13-0132.1, 2014.

Kumar, S., C. D. Peters-Lidar, K. R. Arsenault, A. Getirana, D. Mocko, and Y. Liu: Quantifying the added value of snow cover area observations in passive microwave snow depth data assimilation, J. Hydrometeor, 16, 1736-1741, https://doi.org/10.1175/JHM-D-15-0021.1, 2015.

Liu, Y., C. D. Peters-Lidard, S. Kumar, J. L. Foster, M. Shaw, Y. Tian, and G. M. Fall: 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.

Nester, T., R. Kirnbauer, J. Parajka, and G. Bloschl: Evaluating the snow component of a flood forecasting model, Hydrol. Res., 43, 762-779, https://doi.org/10.2166/nh.2012.041, 2012.

Niu, G.-Y., and Z.-L. Yang: An observation-based formulation of snow cover fraction and its evaluation over large North American river basins, J. Geophys. Res., 112, D21101, https://doi.org/10.1029/2007JD008674, 2007.

Parajka, J., and G. Bloschl: The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models, J. Hydrol., 358, 240-258, 2008.

Reichle, R. H., R. D. Koster, G. J. M. Lannoy, B. A. Forman, Q. Liu, and S. P. P. Mahanama: Assessment and enhancement of MERRA land surface hydrology estimates, J. Climate, 24, 6322-6338, https://doi.org/10.1175/JCLI-D-10-050331.1, 2011.

Riggs, G. A., D. K. Hall, and M. O. Roman: Overview of NASAs MODIS and visible infrared imaging radiometer suite (VIIRS) snow-cover earth system data records, Earth Syst. Sci. Data, 9, 765-777, https://doi.org/10.5194/essd-9-765-2017, 2017.

Rodell, M., and P. R. Houser: Updating a land surface model with MODIS-derived snow cover, J. Hydrometeor., 5, 1064-1075, https://doi.org/10.1175/JHM-395.1, 2004.

Toure, A. M., M. Rodell, Z. L. Yang, H. Beaudoing, E. Kim, Y. Zhang, and Y. Kwon: Evaluation of the snow simulations from the Community Land Model, version 4 (CLM4), J. Hydrometeor., 17, 153-170, https://doi.org/10.1175/JHM-D-14-0165.1, 2016.

Yang, Z.-L., and Coauthors: The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. J. Geophys. Res., 116, D12100, https://doi.org/10.1029/2010JD015140, 2011.

Zhou, X., and Coauthors: Evaluation of Arctic land snow cover characteristics, surface albedo, and temperature during the transition seasons from regional climate model simulations and satellite data, Adv. Meteorol., https://doi.org/10.1155/2014/604157, 2014.