Response to Reviewers' Comments

Dear Petty,

We are grateful to receive your valuable and constructive comments in helping us improve this manuscript. According to your comments and suggestions, we have revised the manuscript seriously, including data, algorithm, discussion and conclusion. Please find the point-to-point responses as follows (Reviewer's comments in black and responses in blue). Thank you very much!

Reviewer: 2

Review of 'Estimating snow depth on Arctic sea ice based on reanalysis reconstruction and particle filter assimilation' by Li et al.,

Review by Alek Petty

Summary

This study presents estimates of snow depth on Arctic sea ice from an updated version of the NASA Eulerian Snow on Sea Ice Model (NESOSIM) and a particle filter data assimilation scheme to combine the model estimates with satellite-derived snow depth data (RA-5VLSTM). The results were compared primarily with snow depths collected by NASA's Operation IceBridge and also some more limited Ice Mass Balance, and a MOSAiC snow depth buoys. The results were also compared with a Kilic et al., (2019) snow depth dataset produced from regression of IMBs to passive microwave data, and also the modified version of the Warren climatology.

General comments

In general, I think the approach of this study was good–use a new data assimilation approach to constrain NESOSIM output and potentially improve its ability to simulate snow depth on Arctic sea ice. However, I have a number of concerns about this study which I detail here:

1. NESOSIM is an open-source model (https://github.com/akpetty/NESOSIM) so community development is actively encouraged - e.g., adding new parameterizations, data assimilation modules etc., into the official code base. The framing of an Improved NESOSIM was thus slightly odd, although obviously this could also be a language/communication issue. The 'Improved' nature of this model framework was also somewhat underwhelming. The atmosphere loss term included as 1 of only 2 'improvements' in this version of NESOSIM has already been integrated into NESOSIM (v.1.1, https://github.com/akpetty/NESOSIM/releases/tag/v1.1). The

authors made a note of this term being introduced already but made no link to the official code repo and still included it in your own 'improved' version. This means the only new parameterization introduced here (to generate the Improved NESOSIM framework) was the simple degree day temperature/melt parameterization. I think this parameter inclusion makes broad sense (we've considered something along these lines ourselves) but i) it was not actually clear that this specific parameterization helped improve the simulation of snow depth as most of the validation occurred in winter/spring and ii) this could have been communicated as a simple added parameter to NESOSIM. I think the atmosphere loss term was much more significant and we've found this to be a useful additional tuning factor, although one not well constrained by observations. Indeed most of what this study is doing is bias correcting towards the OIB quicklook data. On that note, I didn't see any information about making the code available (e.g., the degree day melt model or the particle assimilation approach) which was surprising considering the authors utilized extensively an open-source model for much of this work.

Response: Thank you for this thought-provoking suggestion. We are sorry for the misunderstanding caused by the description in the initial manuscript. We redescribe the method in this paper. The atmospheric loss term, proposed by Petty (2020), makes broad sense. However, the melting term is also an essential process. With global warming, the melting process will be more intense and its contribution to the change of snow depth will increase. The melting process we currently consider is relatively simple. In the future, we will consider more complex melting processes to continuously develop the model.

(1) According to the suggestion, we rename the methods in this paper as NESOSIM v1.0 (including four parameterization processes), NESOSIM_M (adding additional atmospheric loss term and melting term, M refers to melting process) and NESOSIM_M-PF (Data assimilation is included).

(2) Then, we give the website of official code, https://github.com/akpetty/NESOSIM, indicating that the NESOSIM is an open-source model. The detailed revisions are as follows:

Petty et al. (2018) developed a two-layer snow depth model (i.e., NESOSIM v1.0), and snow was divided into a new snow layer and an old layer. NESOSIM is an open-source model (https://github.com/akpetty/NESOSIM) that offers public for contributing their efforts to further develop this model.

(3) Finally, when describing the NESOSIM_M model, we point out that the atmospheric loss term was proposed by Petty (2020). On this basis, we propose that the melting term needs to be considered in this model. The detailed revisions are as follows:

3.2.2 NESOSIM_M

In addition to wind packing and blowing snow loss to leads, wind cause snow loss to the atmosphere, resulting in the redistribution in snow depth. In 2020, Petty (2020) updated the

NESOSIM v1.0 and proposed that the snow lost to the atmosphere process should be considered. Similar to the blowing snow lost to leads and wind packing processes, snow is lost to the atmosphere when U exceeds 5 m s^{-1}. The atmospheric loss term is determined by the blowing snow coefficient, atmospheric loss coefficient (γ), wind speed and depth of the new snow layer. The equation is as follows:

$$\Delta h_s^{atm}(t) = -\beta\gamma U T h_s(t, 0) \tag{12}$$

Besides adding the snow lost to the atmosphere term proposed by Petty (2020), we introduce a simple melting term to develop the NESOSIM_M.

With the continuous warming of the Arctic, snow melting becomes increasingly dramatic. In this study, NESOSIM starts to run in mid-August and continues to run until the following mid-May. The mid-August is selected because there is heavy snowfall in the central Arctic, and great snow melting events in June and July have been avoided (Petty et al., 2018). In mid-August, sea ice is mainly distributed in the central Arctic, and snow melting events also mainly occur in the central Arctic. For the sea ice within the central Arctic, Stroeve et al. (2006) revealed that a threshold closer to 0 °C would agree more closely with passive microwave (PMW)-based melt onset (MO) dates. Therefore, we choose 0 °C as the threshold. When the 2-m temperature (T_{air}) is higher than 0 °C, we consider that there occurs a snow melting process on sea ice.

$$\Delta h_s^{melt}(t) = -T_{air}(t) T \tau \rho_w / \rho_s^n \tag{13}$$

where τ is the degree-day factor and ρ_w is the water density. We set τ to 6.3×10^{-8} m °C^{-1} s^{-1} (Kuchment and Gelfan, 1996), which is determined via the degree-day method.

2. A big issue is that quick-look OIB snow depths are used as truth, with bias corrections/model calibration carried out to improve the fit to this dataset, essentially. However, deriving snow depths from Snow Radar data collected by OIB is challenging (Kwok et al., 2017,) and wide differences exist across the different products. We make a big point about this in the original NESOSIM paper (Petty et al., 2018, P2018). More recent research has shown that OIB QL is ~5 cm thinner than the consensus from the three 'final' products analyzed in P2018 (Petty et al., in prep), see preliminary figure below. These are (since 2013) quick-look data, supposed to provide a basic overview of sea ice conditions, not really a reliable dataset for validating models/retrievals.

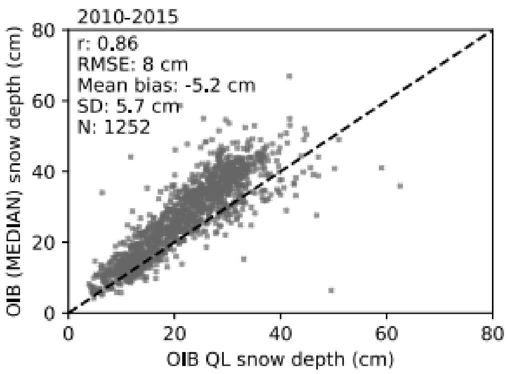

Figure 1: Comparison of the median snow depth from the three different OIB snow depth products used in Petty et al., (2018) and the quick-look (QL) OIB snow depth data. Data are gridded to a 100 km polar stereographic domain before the comparison.

Response: Thank you for this thought-provoking suggestion. We agree that OIB quick look (OIB_{QL}) product has relatively large errors compared with other 'final' products. We are sorry that we do not describe the error of OIB_{QL} clearly. At present, there are three OIB products available to the public: (i) the IceBridge Sea Ice Freeboard, Snow Depth, and Thickness Quick Look, Version 1 (hereafter referred to as OIB_{QL}), covering the period 2012-2019; (ii) the IceBridge L4 Sea Ice Freeboard, Snow Depth, and Thickness, Version 1 (IDCSI4, hereafter referred to as OIB_{IDCSI4}), covering the period 2009-2013; and (iii) the Snow Depth on Arctic Sea Ice Data Set (Newman et al., 2014) which is provided by the NOAA (hereafter referred to as OIB_{NOAA}), covering the periods 2009-2012 and 2014-2015. In the current situation of scarce in-site data, the OIB_{QL} product provides more data than the other two OIB products and can provide us with an intuitive comparison result of snow depth as well. However, before using this data, the error of the product should be clarified. The OIB_{QL} product underestimates the snow depth, and the mean bias is about -5 cm. According to the suggestion, when determining the model, we not only use the OIB_{QL} product to determine the model parameters, but also add the accuracy evaluations based on the OIB_{IDCSI4} product to further determine the model parameters.

In section 4.2, we added OIB_{NOAA} and OIB_{IDCSI4} to evaluate snow depth estimates. The results show that after adding melting term and atmospheric loss term, the accuracy of NESOSIM_M snow depth decreased, but the accuracy of the NESOSIM_M-PF snow depth has been greatly improved compared to NESOSIM v1.0, NESOSIM_M and RA-5VLSTM. When using the MOSAiC product to evaluate snow depth estimates, NESOSIM_M snow depth accuracy is significantly better than NESOSIM v1.0 snow depth. The accuracy of NESOSIM_M-PF snow depth is better than that of satellite-derived snow depth and NESOSIM_M snow depth. In the future, more in-site data are needed to further evaluate the results, continuously optimize the model parameters and improve the snow depth estimates.

The detailed revisions are as follows:

(1) We have added two OIB_{IDCSI4} and OIB_{NOAA} products and more OIB_{QL} information in section 2.6 as follows:

The Operation IceBridge (OIB) mission is proposed for filling the data gap between ICESat and ICESat-2, providing snow depth on sea ice, sea ice thickness, and sea ice type information in the Arctic. These data are widely applied to evaluate satellite-derived or simulated snow depth values. In this study, three OIB products are used which are available to the public: (i) the IceBridge Sea Ice Freeboard, Snow Depth, and Thickness Quick Look, Version 1 (hereafter referred to as OIB_{QL}), covering the period 2012-2019; (ii) the IceBridge L4 Sea Ice Freeboard, Snow Depth, and Thickness, Version 1 (IDCSI4, hereafter referred to as OIB_{IDCSI4}), covering the period 2009-2013; and (iii) the Snow Depth on Arctic Sea Ice Data Set (Newman et al.,

2014) which is provided by the NOAA (hereafter referred to as OIB_{NOAA}), covering the periods 2009-2012 and 2014-2015. OIB_{QL} has a mean bias of about -5 cm, underestimating snow depth (Kwok et al., 2017). OIB_{IDCSI4} product tends to underestimate snow depth (mean bias is about -1 cm) and OIB_{NOAA} tends to overestimate snow depth (mean bias is about 2 cm). OIB_{QL} data from 2014 to 2017 are considered to develop the snow depth model (Fig. 2(a)). OIB_{QL} data from 2018 to 2019, OIB_{IDCSI4} in 2013 and OIB_{NOAA} from 2014 to 2015 are employed to evaluate the established snow depth model (Fig. 2(b), 2(c) and 2(d)).

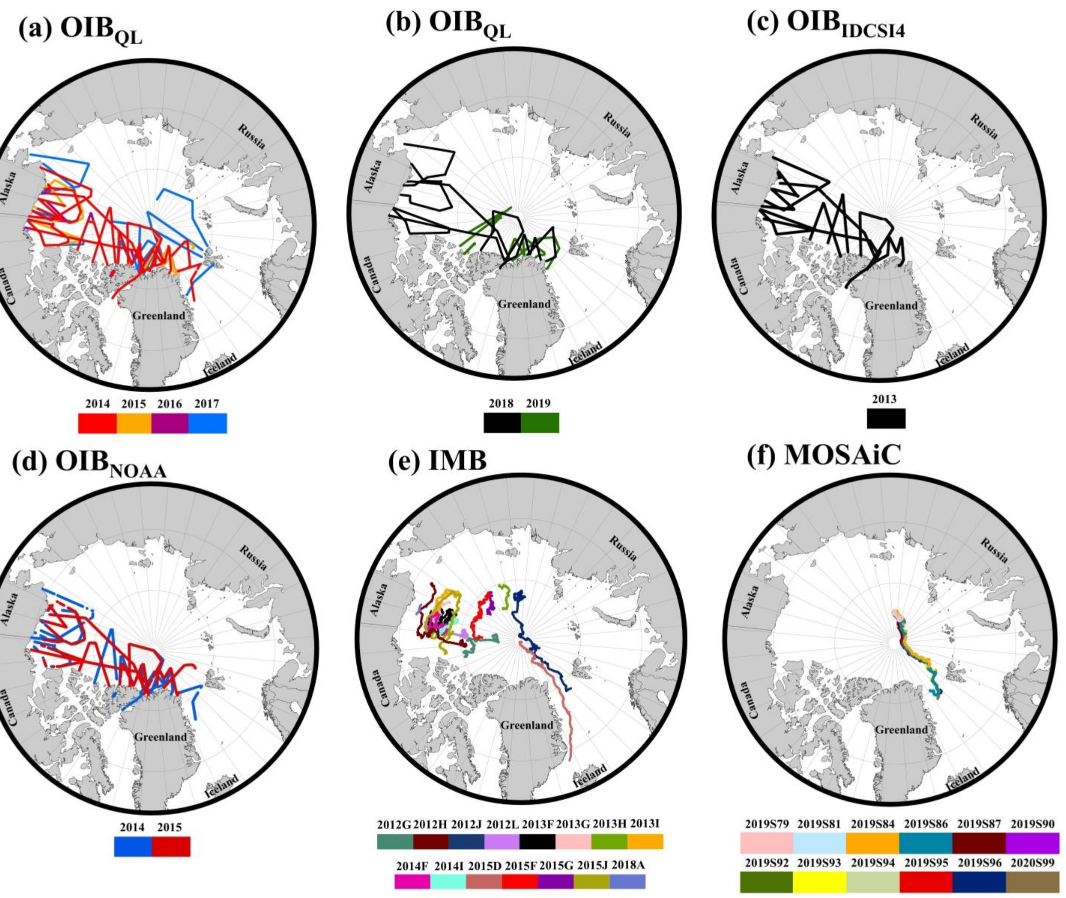

Figure 2. Spatial distribution of (a) the OIB_{QL} track for modeling from 2014 to 2017, (b) the OIB_{QL} track for validation from 2018 to 2019, (c) the OIB_{IDCSI4} track for validation in 2013, (d) the OIB_{NOAA} track for validation from 2014 to 2015, (e) the fifteen IMB tracks for modeling from 2012 to 2018, and (f) the MOSAiC snow buoy track for validation from 2019 to 2020. Note that the legends of (e) and (f) indicate the name of the buoy.

Reference:

Newman, T., Farrell, S. L., Richter-Menge, J., Elder, B., Connor, L., Kurtz, N., and McAdoo, D.: Assessment of radar-derived snow depth measurements over Arctic sea ice. J. Geophys. Res.: Oceans, 119, 8578-8602, https://doi.org/10.1002/2014JC010284, 2014.

(2) We have added related contents based on the OIB_{IDCSI4} product for determining the optimal parameters of models in section 4.1.1 as follows:

When we increase in β value, the peak value of the deviation between derived snow depth and OIB_{QL} moves towards the value of 0, and the frequency of high deviation decreases (Fig. 3). When the default value is applied, the mean bias between the simulated snow depth and OIB_{QL}-measured snow depth is 10.8 cm (Fig. 3). When a value of 5.8×10^{-7} m^{-1} is applied, the deviation from the OIB_{QL} data slightly decreases (Fig. 3). When a value of 11.6×10^{-7} m^{-1} is applied, the mean bias is reduced to 7.7 cm (Fig. 3). We finally choose a value of 11.6×10^{-7} m^{-1} to obtain the most accurate snow depth estimates among the three simulated snow depth vectors. The NESOSIM v1.0 snow depth is generated. Then, we use OIB_{IDCSI4} snow depth to compare the default snow depth with the determined NESOSIM v1.0 snow depth. The NESOSIM v1.0 snow depth is more consistent with OIB_{IDCSI4} snow depth (Fig. 4). The RMSE decreases from 7.1 cm (default snow depth) to 6.0 cm (NESOSIM v1.0 snow depth), and bias decreases from 3.4 cm (default snow depth) to 1.2 cm (NESOSIM v1.0 snow depth).

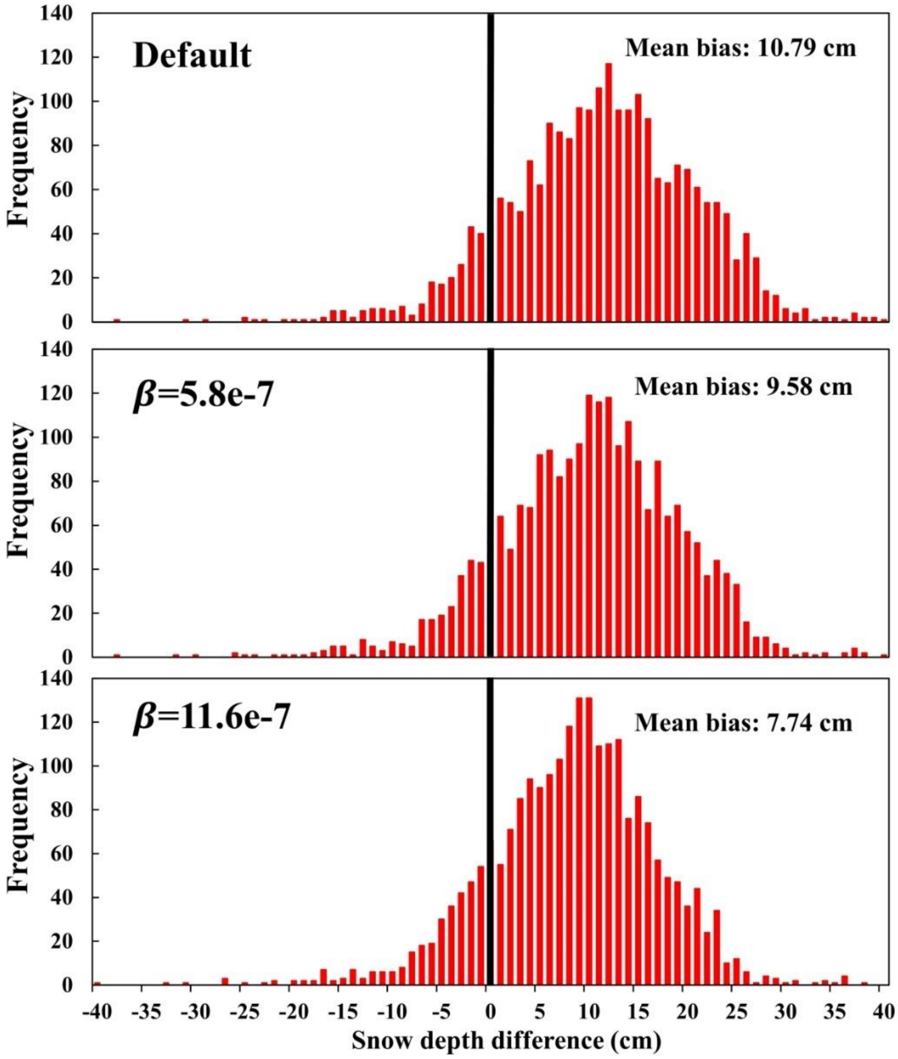

Figure 3. Distribution of deviations between OIB_{QL} snow depth and modeled snow depths (considering different β values) from 2014 to 2017.

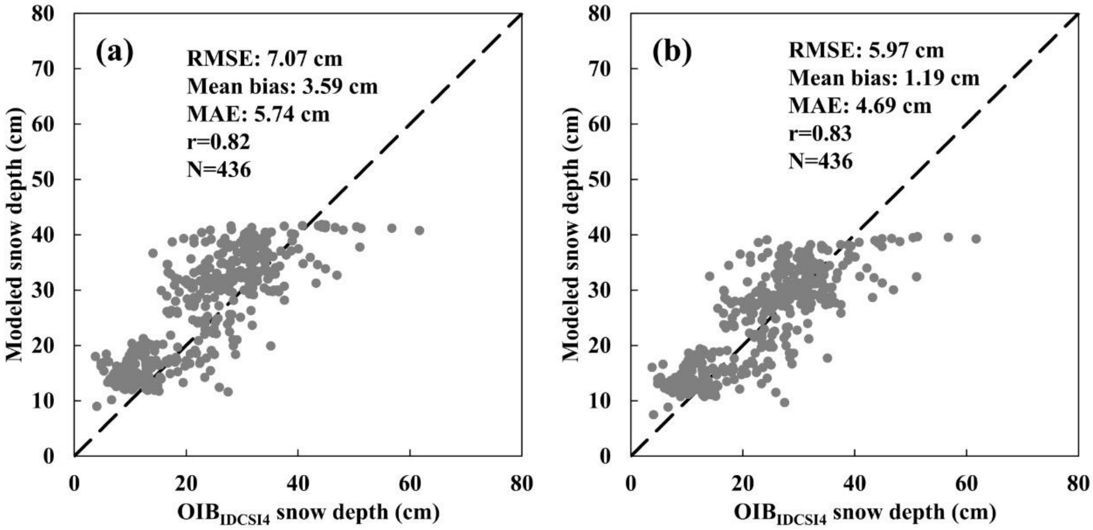

Figure 4. Comparison of the modeled snow depth and the OIB_{IDCSI4} snow depth data. (a) The modeled snow depth using default β value; (b) the modeled snow depth using β value of 11.6×10^{-7} m^{-1}.

(3) We have redrawn Figure 5 (Figure 4 in the initial manuscript) and revised related contents in section 4.1.2 as follows:

First, we incorporate a melting term into the four basic parameterization processes to verify whether the added melting term improves the model accuracy. Compared to NESOSIM v1.0, the mean bias decrease from 7.7 cm to 7.2 cm (based on OIB_{QL} product) by adding the melting term. The peak value of the deviation moves further towards the value of 0 (Fig. 3 and Fig. 5(a)). Under global warming, the effect of the snow melting process will become increasingly obvious. Therefore, considering the melting term is necessary and helpful to understand the snowpack evolution.

Next, we add the atmospheric loss term. The amount of snow lost is determined by the atmospheric loss coefficient γ, and experiments are carried out to determine γ. Then, γ values of 0.0125, 0.015, 0.020 and 0.025 are tested. The results suggest that including the atmospheric loss term greatly reduces the bias between the simulated snow depth and OIB_{QL}-measured snow depth; namely, with increasing coefficient, the bias decreases. When we set γ equal to 0.025, the bias is only 0.6 cm and deviations are evenly distributed on both sides of the 0 value (Fig. 5(b)). According to the deviations between OIB_{QL} snow depth and modeled snow depths, three options (0.015, 0.020 and 0.025) are chosen to determine the corresponding distribution as well as the OIB_{QL} snow depth distribution (Fig. 5(c)). When γ is equal to 0.015, snow depths greater than 48 cm cannot be simulated by NESOSIM_M, but when γ is equal to 0.020, NESOSIM_M is ineffective at snow depths greater than 45 cm, whereas when γ is equal to 0.025, NESOSIM_M is ineffective at snow depths greater than 43 cm. With increasing atmospheric loss coefficient, the retrieval ability of NESOSIM_M for thick snowpacks is weakened.

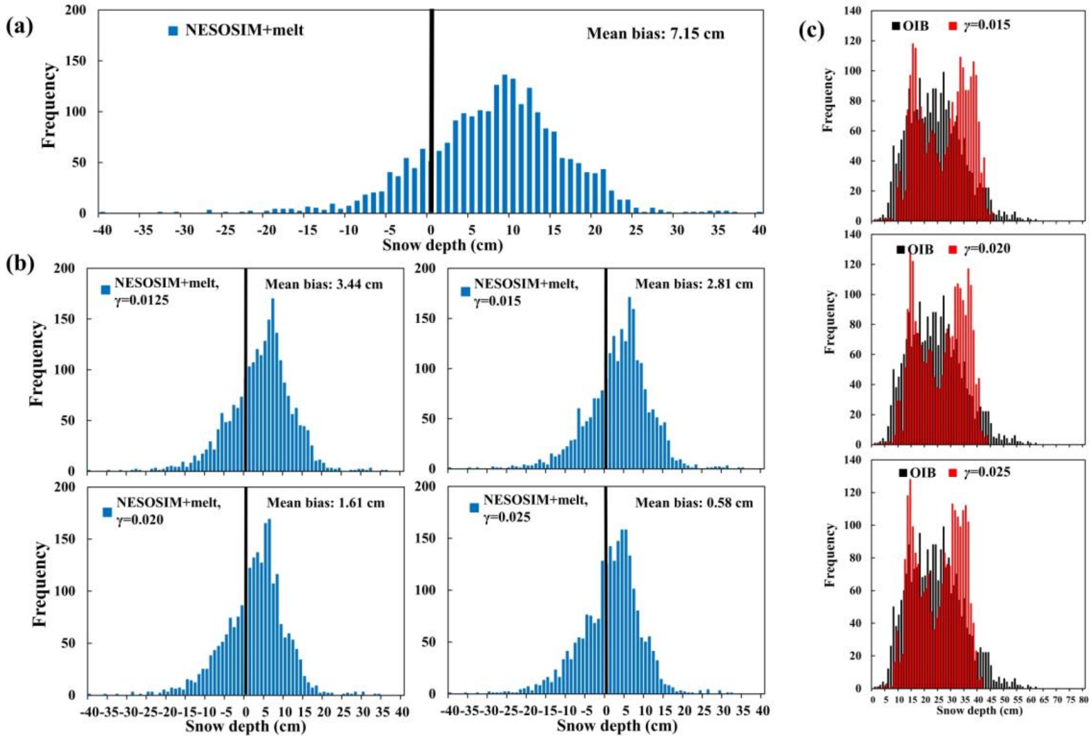

Figure 5. (a) Distribution of deviations between OIB_{QL} snow depth and modeled snow depths (NESOSIM with a melting process); (b) distribution of deviations between OIB_{QL} snow depth and modeled snow depths considering various atmospheric loss coefficient values; (c) distribution of the OIB_{QL} snow depth in black versus that of the simulated snow depth considering different γ values in red. The data cover the period from 2014 to 2017.

(4) We have added accuracy evaluations based on additional two OIB products (OIB_{IDCSI4} and OIB_{NOAA}) and revised related contents in section 4.2 as follows:

To evaluate the accuracy of the assimilation results (NESOSIM_M-PF), we compare the obtained NESOSIM v1.0, NESOSIM_M, RA-5VLSTM, and NESOSIM_M-PF snow depth estimates to independent OIB_{QL} measurements from 2018 to 2019, OIB_{IDCSI4} measurements in 2013, OIB_{NOAA} measurements from 2014 to 2015 and MOSAiC measurements from 2019 to 2020. According to the OIB_{QL} data, NESOSIM_M greatly improves the simulated snow depth over NESOSIM v1.0. The RMSE decreases to 6.6 cm, the MAE decreases to 5.3 cm, and the previously positive bias develops into a negative bias (from 2.6 cm to -2.0 cm) (Table 2). In regard to the RA-5VLSTM snow depth retrievals, the RMSE reaches 6.2 cm, and the correlation coefficient is 0.76, indicating that the accuracy is much higher than that of the NESOSIM v1.0 and NESOSIM_M snow depths. However, RA-5VLSTM snow depth has a less original effective matching number (N0) of 428, while others have N0 of 528. Compared to NESOSIM_M, the accuracy of the NESOSIM_M-PF snow depth has been greatly improved, namely, the RMSE decreases from 6.6 cm to 5.8 cm, the MAE decreases to 4.6 cm, the bias changes from -2.0 cm to -1.4 cm, and the correlation coefficient increases from 0.67 to 0.77 (Table 2). Compared to the RA-5VLSTM method, NESOSIM_M-PF generates a higher accurate snow depth (Table 2).

According to the OIB_{IDCSI4} and OIB_{NOAA} products, the errors of NESOSIM_M snow depth are higher than that of NESOSIM v1.0 snow depth. RA-5VLSTM snow depth has low RMSEs of 4.7 and 6.9 cm for OIB_{IDCSI4} and OIB_{NOAA}, respectively. The NESOSIM_M-PF snow depths are more accurate than snow depth estimates obtained by the other three methods. It has an RMSE of 4.1 cm based on OIB_{IDCSI4} and 6.5 cm based on OIB_{NOAA}. Compared with the other three methods, the RA-5VLSTM has a less N0 (378 for OIB_{IDCSI4} and 535 for OIB_{NOAA}, respectively).

According to the MOSAiC snow buoys, the snow depth estimates obtained with the four methods are generally higher than the MOSAiC-measured snow depths. NESOSIM v1.0 attains the highest RMSE value of 14.1 cm among the four snow depth models (Table 2). Compared to the snow depth estimates retrieved from NESOSIM v1.0, the snow depth estimates obtained with NESOSIM_M-PF are greatly improved (RMSE: decreases by 4.4 cm; bias: decreases by 4.3 cm; MAE: decreases by 3.4 cm; r: increases by 0.05) (Table 2). Compared to RA-5VLSTM, the accuracy of NESOSIM_M-PF is slightly higher (Table 2).

Table 2. Accuracy evaluation of the NESOSIM v1.0, NESOSIM_M, RA-5VLSTM and NESOSIM_M-PF methods through the number of same matching points (Ns), RMSE (cm), bias (cm), MAE (cm) and r based on the OIB_{QL} snow depth from 2018 to 2019, OIB_{IDCSI4} snow depth in 2013, OIB_{NOAA} snow depth from 2014 to 2015 and the MOSAiC-measured snow depth from 2019 to 2020. N0 represents the original effective matching number.

	NESOSIM v1.0	NESOSIM_M	RA-5VLSTM	NESOSIM_M-PF
OIB_{QL}				
N0	528	528	428	528
Ns	428	428	428	428
RMSE (cm)	7.25	6.57	6.24	5.80
Bias (cm)	2.57	-1.96	-1.35	-1.37
MAE (cm)	5.86	5.34	5.02	4.59
r	0.68	0.67	0.76	0.77
OIB_{IDCSI4}				
N0	436	436	378	436
Ns	378	378	378	378
RMSE (cm)	5.51	5.88	4.67	4.61
Bias (cm)	1.55	-2.46	-1.36	-1.12
MAE (cm)	4.40	4.52	3.57	3.50
r	0.83	0.83	0.88	0.89
OIB_{NOAA}				
N0	616	616	535	616
Ns	535	535	535	535
RMSE (cm)	7.28	7.61	6.89	6.50
Bias (cm)	0.34	-4.44	-1.21	-1.66
MAE (cm)	6.10	6.41	5.52	5.18
r	0.71	0.72	0.70	0.70
MOSAiC				
N0	74	74	67	74

Ns	67	67	67	67
RMSE (cm)	14.12	10.31	10.13	9.73
Bias (cm)	11.94	7.60	8.11	7.67
MAE (cm)	12.30	8.65	9.29	8.88
r	0.29	0.29	0.29	0.34

(5) In addition, we have revised related contents in section 6 as follows:

To better understand variations in the snow depth and sea ice, we develop NESOSIM_M-PF based on reanalysis reconstruction and data assimilation methods. First, the coefficients of NESOSIM v1.0 are determined by considering the OIB_{QL} and OIB_{IDCSI4} snow depth. Then, we include a snow melting term and snow lost to the atmosphere term to establish NESOSIM_M. The atmospheric loss coefficient of NESOSIM_M is determined based on OIB_{QL} and IMB-measured snow depth values. Next, the satellite-derived snow depth (RA-5VLSTM snow depth) is assimilated via a particle filter, and the final NESOSIM_M-PF model is established to yield snow depth estimates from August 16 to May 15, 2012–2020. This greatly solves the problem that W99 climatology does not suitably reflect the current changes in snow depth.

Based on OIB-measured snow depth values (OIB_{QL}, OIB_{IDCSI4} and OIB_{NOAA} data), the NESOSIM_M-PF-estimated snow depth is improved over the NESOSIM v1.0 and NESOSIM_M-estimated snow depth, namely, the RMSE decreases by 1.5 cm and 0.8 cm, respectively, and the correlation coefficient increases by 0.1. Compared to the RA-5VLSTM snow depth employed for assimilation, the accuracy of the NESOSIM_M-PF-estimated snow depth slightly improves; the matching points of the NESOSIM_M-PF-estimated snow depth and OIB-measured snow depth increase, and the spatial coverage is highly improved. Based on MOSAiC-measured snow depth values, the NESOSIM_M-PF-estimated snow depth is more accurate than the RA-5VLSTM, NESOSIM v1.0 and NESOSIM_M snow depth estimates. The NESOSIM_M-PF-estimated snow depth is insensitive to the selection of the particle number when the particle number is larger than 250. The model is robust and it is less impacted by the import parameters. The average uncertainty due to the uncertainty of input variables is 0.7 cm. In the future, more in-site data are needed to further evaluate the results, continuously optimize the model parameters, and improve the snow depth estimates.

The spatial distribution of the snow depth retrieved from the NESOSIM_M-PF, Kilic19, and modified W99 methods is consistent. Except for the snow depth difference between the NESOSIM_M-PF and modified W99 approaches in the nearshore area smaller than 10 cm, the snow depth difference is smaller than 5 cm in the other sea areas. The Kilic19 snow depth is larger than the NESOSIM_M-PF-estimated snow depth in MYI regions. The monthly and seasonal (referring to autumn and winter) changes in the Kilic19 and NESOSIM_M-PF snow depth estimates are consistent, and the monthly average NESOSIM_M-PF-estimated snow depth is close to the modified W99 climatology. Therefore, the NESOSIM_M-PF snow depth is reliable and can provide high-precision data for sea ice thickness estimation.

There were also plenty of other parts of the study where data uncertainties are vaguely described and, in some cases, described with worrying levels of certainty (‘The satellite-derived snow depth contains an uncertainty of 1 cm,’).

Response: Thank you for this thought-provoking suggestion. We are sorry for the misunderstanding caused by the lack of a clear description of snow depth uncertainty. The RA-5VLSTM model is developed using deep learning, and the total uncertainty cannot be obtained quantitatively. The uncertainty caused by input parameters can be obtained according to the Monte Carlo method, but the uncertainty caused by model training cannot be determined. According to the comments of the reviewer, we added the relevant description of uncertainty.

The detailed revisions are as follows:

RA-5VLSTM model is developed based on deep learning. So, it is difficult to obtain its total uncertainty. The satellite-derived snow depth uncertainty is calculated based on the Monte Carlo method; the estimated uncertainty is 1 cm, and it refers to the uncertainty caused by the input parameters.

3. I was hoping this paper would provide a much deeper explanation and insight into particle filter data assimilation, but the paper provided only really a minimal description of this. In no way is the approach reproducible. It also left me feeling unsure how much the authors understood about the approach and how best to implement this. The particle number sensitivity test did not feel satisfactory.

Response: Thank you for this thought-provoking suggestion. In the revised manuscript, we have added more details about particle filter methodology in section 3.3 and section 4.1.3.

(1) In section 3.3, the revisions are as follows:

3.3 Particle filter assimilation

To obtain a set of snow depths that combines the advantages of simulated snow depth (the high spatial coverage) and satellite-derived snow depth (the high-precision), we will perform data assimilation. By using assimilation, the model-simulated snow depth can be constrained by observations (satellite-derived snow depth). In recent decades, the particle filter has become a popular data assimilation approach. The great advantage of the particle filter is that it can deal with all types of probability distributions and nonlinear models. Here, we provide a simple description of the particle filter, and more details are found in Arulampalam et al. (2002), Smyth et al. (2019) and Magnusson et al. (2017). The particle filter is derived from the sequential Bayesian estimation problem. A posteriori distribution, the conditional distribution of the current state given all observations, is the purpose of the sequence filtering problem. If the posterior distribution is applicable to the previous time step, the prior probability density ($p(x_{k+1}|z_{1:k})$) of the current time step can be calculated. Then, the prior density can be updated using new observations.

$$p(x_{k+1}|z_{1:k})=\int p(x_{k+1}|x_k)\,p(x_k|z_{1:k})dx_k \tag{14}$$

$$p(x_{k+1}|z_{1:k+1})=\frac{p(z_{k+1}|x_{k+1})p(x_{k+1}|z_{1:k})}{p(z_{k+1}|z_{1:k})} \tag{15}$$

where x and z are the state vector and measurement vector, respectively, $p(x_{k+1}|z_{1:k+1})$ is posterior density.

The above problem cannot be solved analytically for most models. So, Monte Carlo samples are introduced to calculate posterior filter density. The core of the particle filter is Monte Carlo simulation and importance resampling. The particle filter assimilation contains four steps: a prediction step, update step, resampling step and output step.

1) Prediction step

A. The initial state variable is set, i.e., x_k. Random noise with an arbitrary distribution is provided to disturb x_k, and the n-dimensional initial state x_k^i is obtained at time step k.

B. The weight of each particle is set to 1/N, and N is the number of particles

C. State and measurement prediction:

$$x_{k+1}^i = f\left(x_k^i, \theta_k^i, u_k\right) + v_k \tag{16}$$

$$z_{k+1}^i = h\left(x_{k+1}^i\right) + n_{k+1} \tag{17}$$

where f and h are the state function and measurement function, respectively, θ is a model parameter, u is the model input, and v and n are the process noise and measurement noise, respectively.

2) Update step

A. The weight of each particle is calculated:

$$w_{k+1}^i = w_k^i p(z_{k+1} | x_{k+1}^i) \tag{18}$$

$$p(z_{k+1} | x_{k+1}^i) = \frac{1}{\sqrt{(2\pi)^N |C_v|}} e^{[-0.5(z_{k+1} - x_{k+1}^i)(z_{k+1} - x_{k+1}^i)/C_v]} \tag{19}$$

where p is the likelihood function and C_v is the measurement error covariance.

B. The weights are normalized, namely, the sum of all weights equals 1.

3) Resampling step

According to the weight, any particles with a low weight are discarded, and the particles with a high weight are duplicated. After importance resampling, the total number of particles remains the same. Then, the weight is reset to 1/N.

4) Output step

The mean of the ensemble (Dong et al., 2015) is selected to define the best estimates of the state vector. Then, the best estimates are output.

(2) In section 4.1.3, the added information is as follows:

Since snow depth is a model output, no observation operator is required. The state function is determined by the NESOSIM_M. To simplify the state function, we use regression analysis to construct the linear equation before and after time step as follows:

$$x_{k+1}^i = 0.9498 \times x_k^i + 1.7205 + v_k \tag{20}$$

(3) There is a little difference between the snow depth when the number of particles is 250 (i.e., P250 snow depth) and the default snow depth (the number of particles is 1000). The snow depth value is not a good idea to show the snow change caused by particle selection, as shown in Fig. BB. To provide more information, we redrawn Figure 9 (Figure 8 in the original manuscript). Firstly, we show the change in default snow depth. Then we show the snow depth difference between snow depth with different particle numbers and default snow depth.

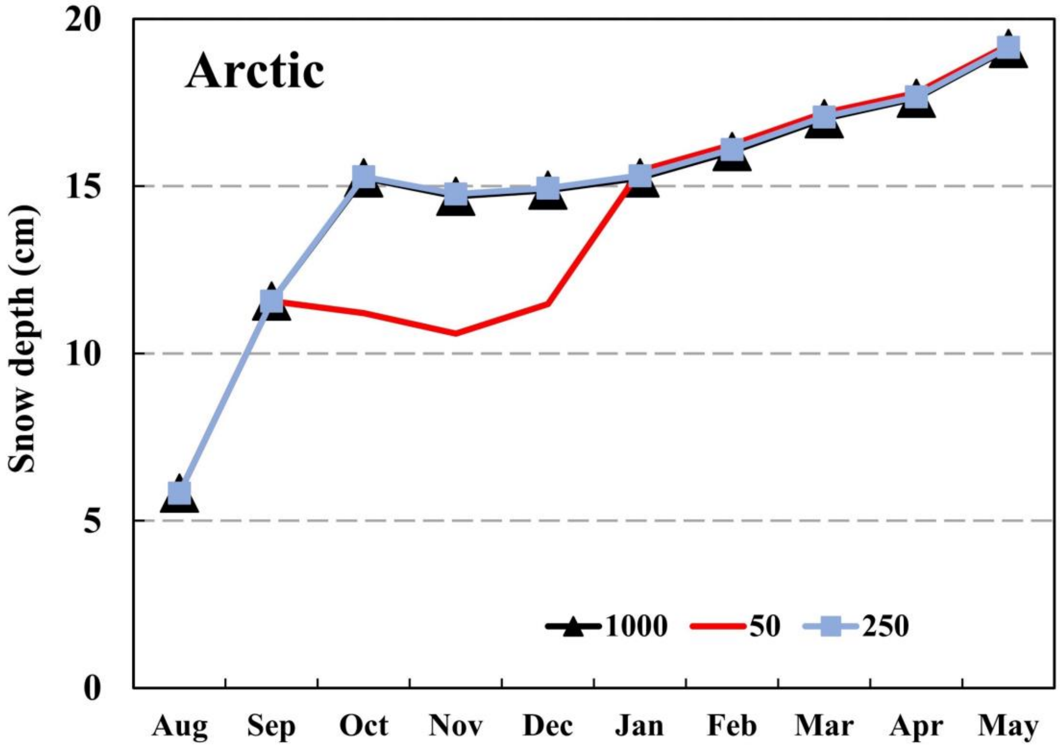

Fig. BB. Variations in the daily average snow depth in the Arctic from 2012 to 2020 using different particle numbers. Note the value of 1000 is the default value.

The detailed revisions are as follows:

5.1 Particle number sensitivity

The sensitivity of NESOSIM_M-PF to the number of particles in the Arctic from 2012–2020 is examined. We show the results for different numbers of particles, i.e., 50, 250, 500, 750, 1000 and 1250 (1000 is the default value). For convenience, the snow depths obtained with 1000, 50, 250, 500, 750, and 1250 particles are defined as the default snow depth, P50 snow depth, P250 snow depth, P500 snow depth, P750 snow depth and P1250 snow depth, respectively.

In the Arctic, the default snow depth begins to increase rapidly in mid-August, which continues until October (Fig. 9(a)). In October, the increase rate of the snow depth decreases, and the snow depth begins to decline in late October. This change can also be observed among the results of Petty et al. (2018) using the snowfall products of JAR-55 and MEDIAN-SF to run NESOSIM. However, this change is more notable in our results. Then, the snow depth slowly increases until May due to the continuous accumulation of snow and less snow melting during the cold season. The interannual variability in the snow depth from August to September is low, basically between -1 cm and 1 cm (Fig. 9(a)). The interannual variability in the snow depth from October to May ranges from -2.0 cm to 2.0 cm. The absolute differences between the monthly P50 snow depth and default snow depth are large from October to December, reaching a maximum value of 4.1 cm in November (Fig. 9(b)). The differences between the monthly snow depth estimates (P250, P500 and P750 snow depths) and default snow depth are larger than 0. The closer the number of applied particles is to the default value, the smaller the snow depth difference. The differences between the monthly P1250 snow depth and default snow depth are smaller than 0. The largest snow depth difference is smaller than 0.01 cm (Fig. 9(b)). All the absolute values of the five average monthly snow depth difference vectors increase rapidly in September, a maximum value is reached in November-January, and the value subsequently decreases (Fig. 9(b)).

The Central Arctic is an area largely covered with thick snow. The default snow depth begins to increase rapidly in mid-August, which continues until November (Fig. 9(c)). Then, the default snow depth shows a slowly decreasing trend, which continues until May. Through the sensitivity analysis in this area, we can elucidate the influence of the particle number on the area with a large snow depth. When the particle number is small (i.e., 50), the choice of the particle number imposes a dramatic effect on the snow depth in thick-snow areas, with the largest deviation of 10.3 cm in November (Fig. 9(d)). When the particle number is larger than 250, the choice of the particle number yields no effect on the snow depth in thick-snow areas, and the absolute difference between the four snow depth vectors and the default snow depth is smaller than 0.014 cm (Fig. 9(d)). After January, the absolute value of the five average monthly snow depth differences remains small (0.002 cm) (Fig. 9(d)).

The Chukchi Sea is covered with thin FYI. The default snow depth increases slightly in September and October, decreases for a short period in November. From December, the increase rate of snow depth accelerates and continues until May, with an increase of nearly 10 cm (Fig. 9(e)). The interannual variability in the snow depth from August to May is high, with a maximum value of more than 5 cm (Fig. 9(e)). The influence of the particle number on areas with small snow depths can be revealed by analyzing the Chukchi Sea. The absolute differences between the monthly P50 snow depth and default snow depth are the largest in October, with an absolute value of 1.2 cm (Fig. 9(f)). The snow depth difference based on the other four numbers of particles (250, 500, 750, and 1250) is smaller than 0.1 cm (Fig. 9(f)). This verifies that when the number of particles is large, the choice of the particle number imposes little effect on snow depth estimation.

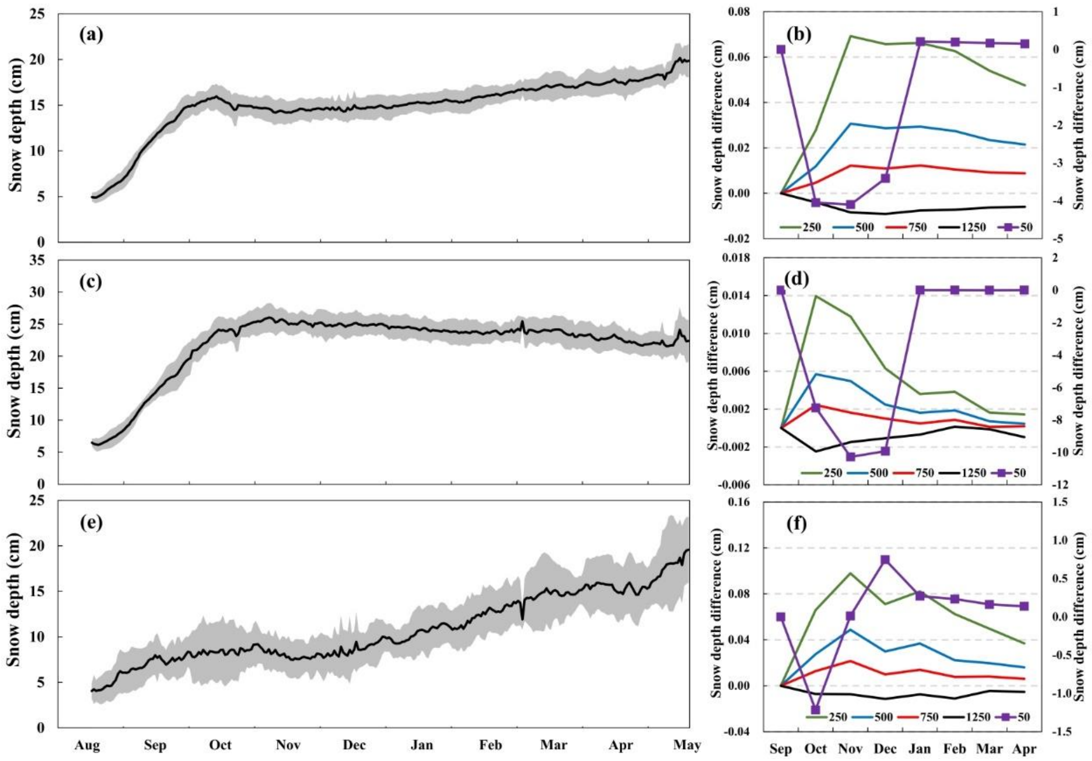

Figure 9. Variations in the daily average default snow depth in the (a) Arctic, (c) Central Arctic and (e) Chukchi Sea from 2012 to 2020, and the shaded areas represent the interannual variability using 1 standard deviation; variations in the monthly average snow depth difference (snow depth estimates based on the different particle numbers minus the default snow depth) in the (b) Arctic, (d) Central Arctic and (f) Chukchi Sea. Note that the secondary axes of (b), (d) and (f) indicate the difference between the P50 snow depth and the default snow depth.

4. The RA-5VLSTM dataset was used as the only input to the data assimilation system but the citation linked to is just a data portal that I was unable to translate, so really there is no background to how this data was obtained and how well it agrees with other snow depth datasets that exist. My guess is that the INESOSIM-PF run tracks this observational dataset quite closely, but it's unclear if that's a good thing or not.

Response: Thank you for this thought-provoking suggestion. We are sorry to use the wrong link when referencing the dataset. The correct link is http://data.tpdc.ac.cn/en/disallow/8b4b5f67-ee01-45ee-96a6-db8f712a2e0c. In the revised manuscript, we have changed the reference of the dataset to the reference of the paper.

According to the suggestion, we have introduced the RA-5VLSTM model in section 2.4 as follows:

Using the IMB data, Li et al. (2022) developed the snow depth model for MYI and FYI based on regression analysis (RA) and long short-term memory (LSTM: a deep learning method). Then, the RA-5VLSTM was proposed. For the FYI, the regression relationship between the GR_h (37/7) and IMB-measured snow depth was determined:

$$h_{FYI}=11.01\text{-}352.17\times\frac{T_{bice}(37H)\text{-}T_{bice}(7H)}{T_{bice}(37H)+T_{bice}(7H)} \tag{3}$$

where h_{FYI} is the snow depth atop FYI.

For the MYI, the snow depth was calculated using LSTM by imputing five parameters (GR_v (37/19), GR_v (19/7), GR_v (19/10), GR_h (37/19) and GR_v (37/7)). The RMSE of snow depth obtained by RA-5VLSTM is 7.16 cm based on the OIB_{QL} product from 2015 to 2019, and RMSE is 11.27 cm based on the IMB data.

We obtain snow depth estimates using the RA-5VLSTM model from 2012 to 2020, which is used for assimilation. We obtain modified W99 climatology (come from the NSIDC CryoSat-2 Level-4 product) and Kilic19 snow depth estimates from 2012 to 2020 to compare the performance of different snow depth estimates.

Reference:

Li, H., Ke, C., Zhu, Q., Li, M., Shen, X.: A deep learning approach to retrieve cold-season snow depth over Arctic sea ice from AMSR2 measurements. Remote Sens. Environ., 269, 112840, https://doi.org/10.1016/j.rse.2021.112840, 2022.

Specific comments

The statistics of RMSE and MAE include the bias – so really all the statistics presented are highly sensitive to the presence of a bias. Most of this study seemed to involve basic bias correction (which is somewhat understandable considering the large uncertainties in snow) but limits the impact of the results presented. Generally I think it is not a good idea to express RMSE/bias changes as percentages. Just stating the change in absolute terms is easier for the reader to assess.

Response: Thank you for this thought-provoking suggestion. In the revised manuscript, we have used absolute term to express RMSE/bias changes.

L73-74: this particle filter methodology and motivation needs to be much better described.

Response: Thank you for this thought-provoking suggestion. In the revised manuscript, we have added more details about particle filter methodology and motivation in section 3.3 as follows:

3.3 Particle filter assimilation

To obtain a set of snow depths that combines the advantages of simulated snow depth (the high spatial coverage) and satellite-derived snow depth (the high-precision), we will perform data assimilation. By using assimilation, the model-simulated snow depth can be constrained by observations (satellite-derived snow depth). In recent decades, the particle filter has become a popular data assimilation approach. The great advantage of the particle filter is that it can deal with all types of probability distributions and nonlinear models. Here, we provide a simple description of the particle filter, and more details are found in Arulampalam et al. (2002), Smyth et al. (2019) and Magnusson et al. (2017). The particle filter is derived from the sequential Bayesian estimation problem. A posteriori distribution, the conditional distribution of the current state given all observations, is the purpose of the sequence filtering problem. If the posterior distribution is applicable to the previous time step, the prior probability density ($p(x_{k+1}|z_{1:k})$) of the current time step can be calculated. Then, the prior density can be updated using new observations.

$$p(x_{k+1}|z_{1:k})=\int p(x_{k+1}|x_k)\,p(x_k|z_{1:k})dx_k \tag{14}$$

$$p(x_{k+1}|z_{1:k+1})=\frac{p(z_{k+1}|x_{k+1})p(x_{k+1}|z_{1:k})}{p(z_{k+1}|z_{1:k})} \tag{15}$$

where x and z are the state vector and measurement vector, respectively, $p(x_{k+1}|z_{1:k+1})$ is posterior density.

The above problem cannot be solved analytically for most models. So, Monte Carlo samples are introduced to calculate posterior filter density. The core of the particle filter is Monte Carlo simulation and importance resampling. The particle filter assimilation contains four steps: a prediction step, update step, resampling step and output step.

1) Prediction step

A. The initial state variable is set, i.e., x_k. Random noise with an arbitrary distribution is provided to disturb x_k, and the n-dimensional initial state x_k^i is obtained at time step k.

B. The weight of each particle is set to 1/N, and N is the number of particles

C. State and measurement prediction:

$$x_{k+1}^i=f\left(x_k^i,\theta_k^i,u_k\right)+v_k \tag{16}$$

$$z_{k+1}^i=h\left(x_{k+1}^i\right)+n_{k+1} \tag{17}$$

where f and h are the state function and measurement function, respectively, θ is a model parameter, u is the model input, and v and n are the process noise and measurement noise, respectively.

2) Update step

A. The weight of each particle is calculated:

$$w_{k+1}^i=w_k^i p(z_{k+1}|x_{k+1}^i) \tag{18}$$

$$p(z_{k+1}|x_{k+1}^i)=\frac{1}{\sqrt{(2\pi)^N|C_v|}}e^{[-0.5(z_{k+1}-x_{k+1}^i)(z_{k+1}-x_{k+1}^i)/C_v]} \tag{19}$$

where p is the likelihood function and C_v is the measurement error covariance.

B. The weights are normalized, namely, the sum of all weights equals 1.

3) Resampling step

According to the weight, any particles with a low weight are discarded, and the particles with a high weight are duplicated. After importance resampling, the total number of particles remains the same. Then, the weight is reset to 1/N.

4) Output step

The mean of the ensemble (Dong et al., 2015) is selected to define the best estimates of the state vector. Then, the best estimates are output.

'Section 3.2 Two snow depth retrieval methods' – why are these not in the data section?

They are previous data not really created in this study - one 'retrieval' - the multi-linear regression to passive microwave data from Kilic et al., (2019) and then the Warren 1999

(W99) quadratic fit to in-situ snow depths.

Response: Thank you for this thought-provoking suggestion. According to the suggestion, we have placed these two retrieval methods in section 2.4 as follows:

Based on the snow depth data measured at Soviet stations from 1954 to 1991, Warren et al. (1999) constructed a two-dimensional quadratic function as follows (hereinafter W99):

$$h_s(x,y)=h_0+Ax+By+Cxy+Dx^2+Ey^2 \quad (1)$$

where h_0 is the fitted snow depth at the North Pole and A, B, C, D, and E are coefficients of Eq. (1). The coefficients and h_0 are different in the different months. All coefficient and h_0 values for all 12 months are obtained from Warren et al. (1999).

We select CryoSat-2 Level-4 Sea Ice Elevation, Freeboard, and Thickness Version 1 data pertaining to the ten subregions, which are provided by the NSIDC. These data provide the 30-day average snow depth on sea ice from 2010 to the present in a 25 km grid. A snow depth dataset is constructed based on a modified W99 climatology (Laxon et al., 2013), i.e., the snow depth estimates on FYI are equal to half the W99 climatology, and those on MYI are equal to the W99 climatology.

Kilic et al. (2019) (hereinafter Kilic19) developed a multilinear regression approach for snow depth estimation based on four ice mass balance buoys (IMB), i.e., 2012G, 2012H, 2012J and 2012L. The multilinear regression relationship between the vertically polarized brightness temperatures of AMSR2 (7, 19 and 37 GHz) and the IMB-measured snow depth was established, and Eq. (2) was determined:

$$h_s=177.01+1.75\times T_b(7V)-2.80\times T_b(19V)+0.41\times T_b(37V) \quad (2)$$

where h_s is the snow depth atop sea ice.

Reference:

Laxon, S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang, J., Haas, C., Hendricks, S., Krishfield, R., Kurtz, N., Farrell, S., and Davidson, M.: CryoSat-2 estimates of Arctic sea ice thickness and volume, Geophys. Res. Lett., 40, 732-737, https://doi.org/10.1002/grl.50193, 2013.

It is also confusing that you use the W99 with snow depths halved over FYI as well, and refer to this as an NSIDC product (taken from the CryoSat-2 implementation of this). This was also used in P2018 and is typically referred to as the modified Warren climatology. This was referenced in P2018 (and I think was first introduced by Laxon et al., 2013). I don’t think it should be referred to as an NSIDC product particularly.

Response: Thank you for this thought-provoking suggestion. We are sorry for inappropriately referring to the modified W99 climatology as NSIDC snow depth. In the revised manuscript, we have changed “NSIDC snow depth” to “modified W99 climatology”.

L201: ‘However, the updated algorithm has not been debugged’ is a bit of a strange way of framing this. The code is on GitHub (version 1.1) so you should ideally cite that more clearly, as it is exactly the same as the ‘improved’ atmosphere snow loss term used in this study.

Response: We are sorry we do not cite clearly and there is an inappropriate description. In the revised manuscript, we have deleted the “However, the updated algorithm has not been debugged” and ideally cite “atmosphere snow loss term” more clearly.

The detailed revisions in section 3.2.2 are as follows:

In addition to wind packing and blowing snow loss to leads, wind cause snow loss to the atmosphere, resulting in the redistribution in snow depth. In 2020, Petty (2020) updated the NESOSIM v1.0 and proposed that the snow lost to the atmosphere process should be considered. Similar to the blowing snow lost to leads and wind packing processes, snow is lost to the atmosphere when U exceeds 5 m s^{-1}. The atmospheric loss term is determined by the blowing snow coefficient, atmospheric loss coefficient (γ), wind speed and depth of the new snow layer. The equation is as follows:

$$\Delta h_s^{atm}(t) = -\beta\gamma U T h_s(t, 0) \qquad (12)$$

Besides adding the snow lost to the atmosphere term proposed by Petty (2020), we introduce a simple melting term to develop the NESOSIM_M.

L240 – ‘Therefore, the model was insensitive to β.’ – really, it’s just insensitive in the regions where we have observations (e.g. in the central Arctic). P2018 showed how in more marginal seas it has a bigger impact.

Response: Thank you for this thought-provoking suggestion. In the revised manuscript, we have corrected the related descriptions as follows:

When the blowing snow coefficient was set to twice the default value (i.e., β was 5.8×10^{-7} m^{-1}), negligible effects were observed based on drifting station data. However, in more marginal seas, it had a bigger impact on the model.

L243 – now a bit confused regarding the parameter you’re looking at here. I think it’s beta but why choose that if NESOSIM is less sensitive to this parameter?

Response: When α and ω are set to default values, the difference between the snow depth (snow density) obtained by the model and snow depth (snow density) obtained by the drifting station is small. So, for the α and ω, we select the default value. For the β, we perform experiments to choose an appropriate value.

Section 4.1.1. – the problem here is that you’re fitting to quick-look OIB now depths that are likely biased. To accommodate the product uncertainty in P2018 we looked at the different algorithms and noted the wide-spread made it hard to calibrate.

Response: Thank you for this thought-provoking suggestion. According to the suggestion, we added the result of OIB_{IDCSI4} in section 4.1.1 to further prove that the error of snow depth obtained by using the parameters we selected is reduced.

4.1.1 Determination of NESOSIM v1.0

Petty et al. (2018) performed sensitivity analysis of three model parameters (α, β, and ω). The results indicated that NESOSIM v1.0 was sensitive to α. At an α value of 5.8×10^{-7} s^{-1}, the simulated snow depth was the most consistent with the obtained data from Soviet drifting stations. When α equaled 11.6×10^{-7} s^{-1}, the snow density was greatly influenced. When the blowing snow coefficient was set to twice the default value (i.e., β was 5.8×10^{-7} m^{-1}), negligible effects were observed based on drifting station data. However, in more marginal seas, it had a bigger impact on the model. When the wind threshold was 10 m s^{-1}, the difference between the snow depth and drifting station data notably increased. When we run NESOSIM v1.0, α and ω are set to default values. Then, adopting OIB_{QL} data from 2014 to 2017 and OIB_{IDCSI4} in 2013, we perform experiments to select an appropriate β value.

When we increase in β value, the peak value of the deviation between derived snow depth and OIB_{QL} moves towards the value of 0, and the frequency of high deviation decreases (Fig. 3). When the default value is applied, the mean bias between the simulated snow depth and OIB_{QL}-measured snow depth is 10.8 cm (Fig. 3). When a value of 5.8×10^{-7} m^{-1} is applied, the deviation from the OIB_{QL} data slightly decreases (Fig. 3). When a value of 11.6×10^{-7} m^{-1} is applied, the mean bias is reduced to 7.7 cm (Fig. 3). We finally choose a value of 11.6×10^{-7} m^{-1} to obtain the most accurate snow depth estimates among the three simulated snow depth vectors. The NESOSIM v1.0 snow depth is generated. Then, we use OIB_{IDCSI4} snow depth to compare the default snow depth with the determined NESOSIM v1.0 snow depth. The NESOSIM v1.0 snow depth is more consistent with OIB_{IDCSI4} snow depth (Fig. 4). The RMSE decreases from 7.1 cm (default snow depth) to 6.0 cm (NESOSIM v1.0 snow depth), and bias decreases from 3.4 cm (default snow depth) to 1.2 cm (NESOSIM v1.0 snow depth).

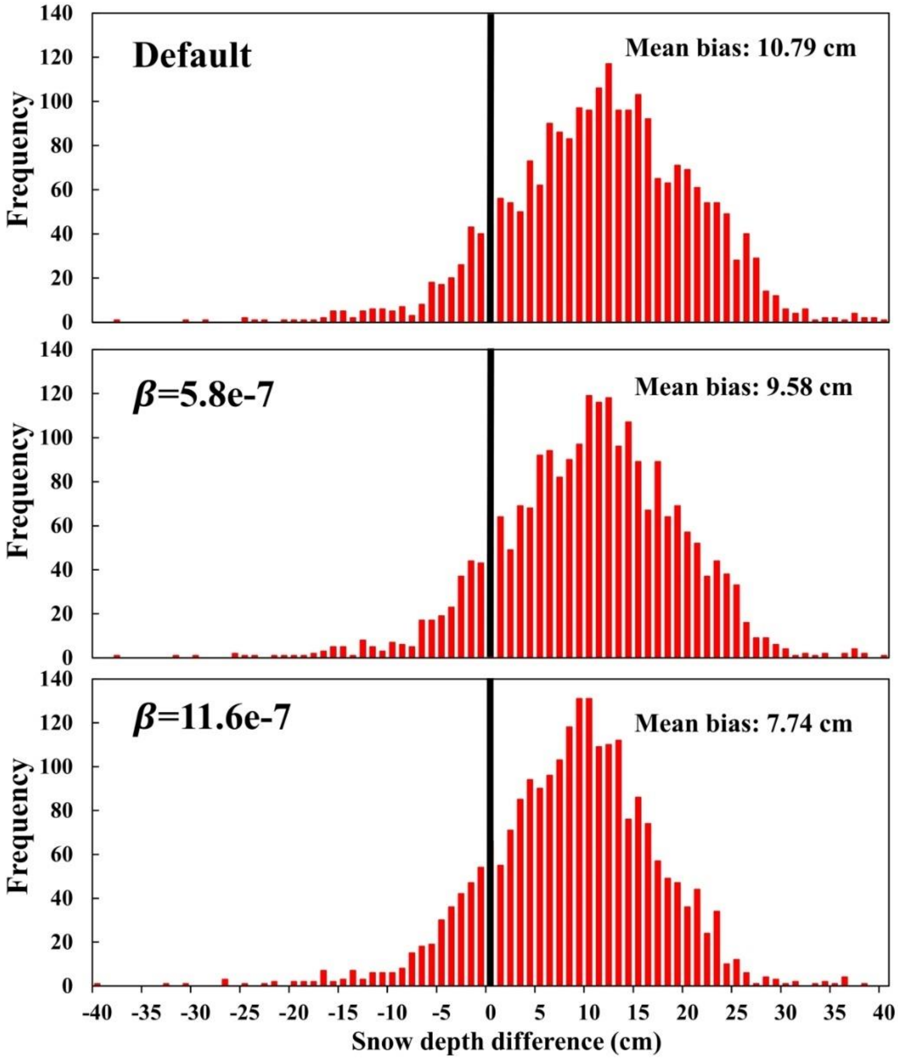

Figure 3. Distribution of deviations between OIB_{QL} snow depth and modeled snow depths (considering different β values) from 2014 to 2017.

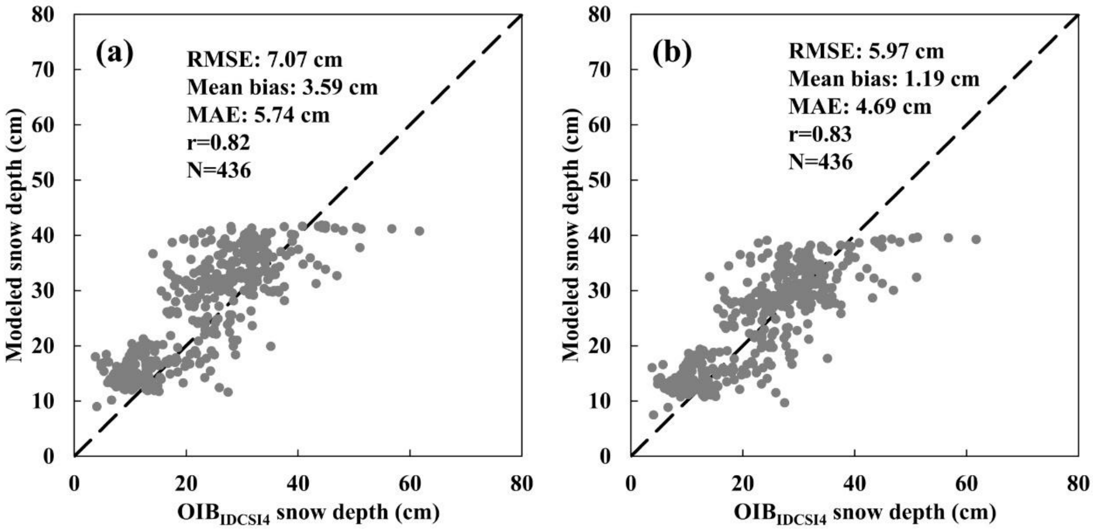

Figure 4. Comparison of the modeled snow depth and the OIB_{IDCSI4} snow depth data. (a) The modeled snow depth using default β value; (b) the modeled snow depth using β value of

11.6×10^{-7} m^{-1}.

Figure 3 – this is not really a great way of showing differences/biases between runs as the lines all look basically the same.

Response: Thank you for this thought-provoking suggestion. In the revised manuscript, we have redrawn Figure 3.

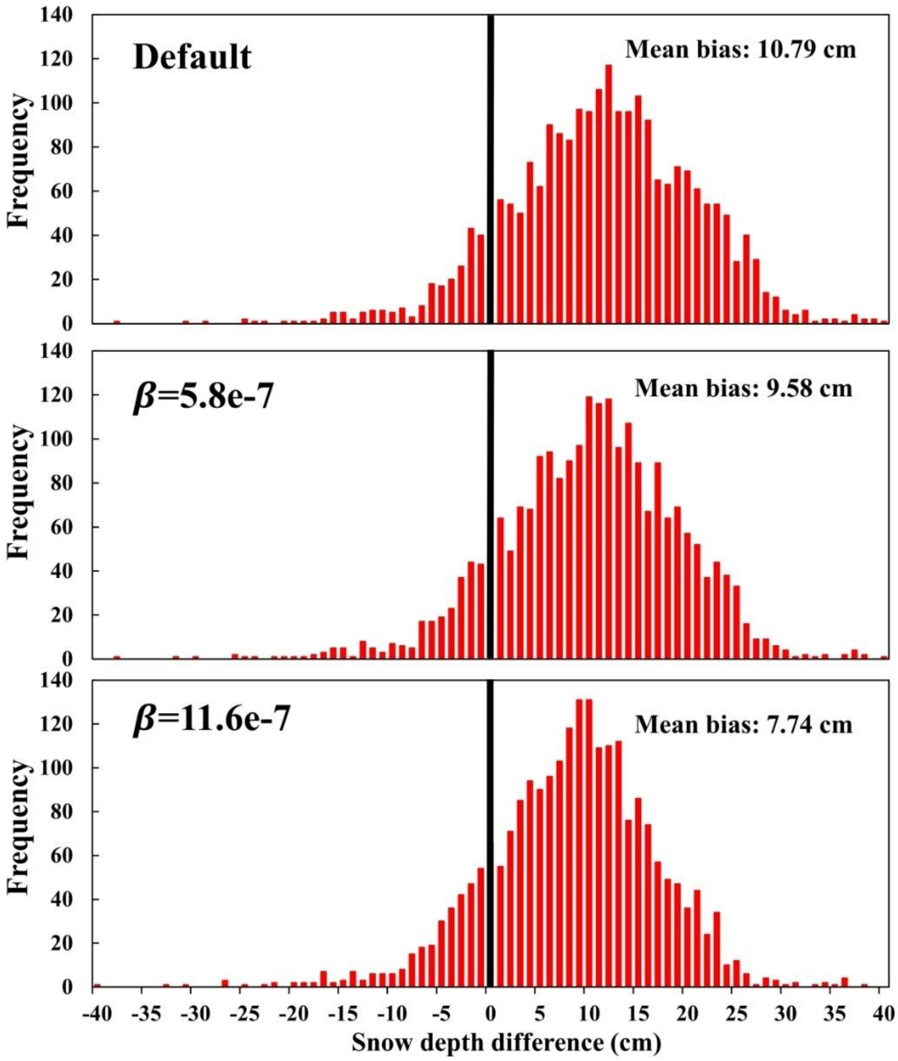

Figure 3. Distribution of deviations between OIB_{QL} snow depth and modeled snow depths (considering different β values) from 2014 to 2017.

Figure 4 - Bimodal NESOSIM output is interesting, what's going on there? I think P2018 showed weak evidence of bimodality.

Response: Thank you for this thought-provoking suggestion. Regardless of the value of γ, the estimated snow depth presents a bimodal distribution. We take γ=0.015 as an example. There may be two reasons for the bimodal distribution of snow depth estimates: (i) due to the difference in the distribution of snow depth estimates over FYI and MYI. The peak of the snow depth distribution outputted by the model is about 15 cm over FYI, and about 40 cm over MYI

(Fig. S1(a) and (b)); (ii) there are two concentrated distribution intervals for the estimated snow depth over MYI, one range is 30-40 cm and the other is 15-25 cm (Fig. S1(b)). This may be due to regional differences caused by different regions. The mean snow depth is concentrated at 30 cm in the Central Arctic, concentrated at 20 cm in the Beaufort Sea and concentrated at 10 cm in the Canadian Archipelago (Fig. S1(c)).

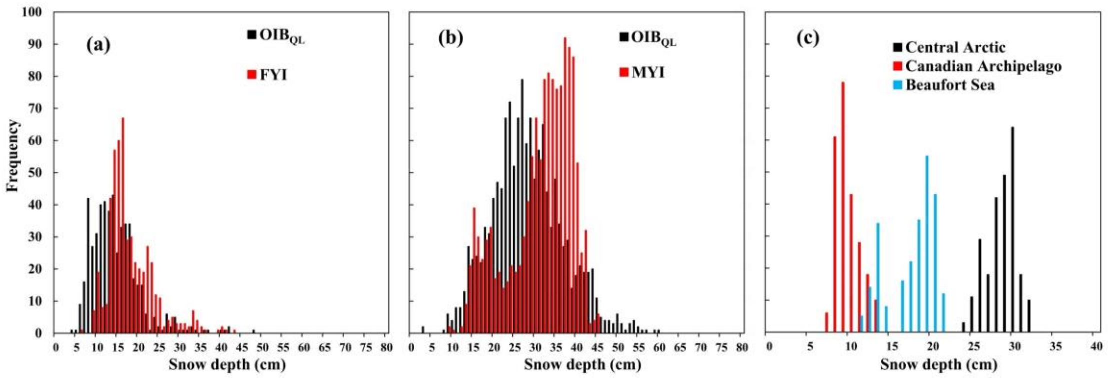

Fig. S1. Distribution of the OIB_{QL} snow depth over FYI (a) and MYI (b) in black versus that of the corresponding simulated snow depth (γ=0.015) in red. (c) Distribution of the simulated snow depth in the Central Arctic, Canadian Archipelago and the Beaufort Sea from March and April. The data cover the period from 2014 to 2017.

L280-287 – but you don't seem to use the F labelling in the figures/tables?

Response: Thank you for this thought-provoking suggestion. In the revised manuscript, we have added F labelling in Table 1.

Table 1. Accuracy of NESOSIM_M with different atmospheric loss coefficient values (γ) based on the IMB-measured snow depth (number of same matching points (Ns), RMSE (cm), bias (cm), MAE (cm) and r).

γ	0.015 (F1)	0.020 (F2)	0.025 (F3)
Ns	443	443	443
RMSE (cm)	16.58	16.85	17.11
Bias (cm)	-6.67	-7.36	-7.97
MAE (cm)	11.06	11.28	11.51
r	0.12	0.12	0.11

Section 4.2–these descriptions were generally quite unclear 'superiority of the assimilation results'

Response: Thank you for this thought-provoking suggestion. In section 4.2, the accuracy evaluations are implemented. There are no contents to show superiority of the assimilation results. Therefore, in the revised manuscript, we have deleted the "superiority of the assimilation results".

L353 – 355: 'We obtain the error in the Kilic19 snow depth based on the OIB-measured snow depth from 2018 to 2019.' This is a very bad idea!

Response: Thank you for this thought-provoking suggestion. Indeed, it is inappropriate to use this expression. According to the suggestion, we have changed this sentence to "We evaluate the Kilic19 snow depths based on three OIB-measured snow depth records, i.e., OIB_{QL}, OIB_{IDCSI4} and OIB_{NOAA}". Then, we added the result of OIB_{IDCSI4} and OIB_{NOAA}. The added content is as follows:

We evaluate the Kilic19 snow depths based on three OIB-measured snow depth records, i.e., OIB_{QL}, OIB_{IDCSI4} and OIB_{NOAA}. According to the OIB_{QL} product, the RMSE of the Kilic19 snow depth is 10.8 cm, which is approximately 1.6 times that of the NESOSIM_M-PF snow depth (Tables 2 and 3, respectively). Based on the OIB_{IDCSI4} product, the NESOSIM_M-PF snow depth (RMSE: 4.6 cm; bias: -1.1 cm) has a lower error than the Kilic19 snow depth (RMSE: 6.7 cm; bias: 1.9 cm). Using the OIB_{NOAA} product as a true value, the Kilic19 snow depth has a high RMSE of 13.9 cm and a bias of 5.0 cm. According to the MOSAiC snow buoys, the Kilic19 model generates a slightly smaller RMSE and a lower correlation coefficient than does NESOSIM_M-PF (Table 3). The RMSE of the modified W99 model reaches 13.4 cm, and the correlation coefficient is the highest of 0.37 among all six methods (Tables 2 and 3, respectively). The above results indicate that the Kilic19 method performs well in snow buoy distribution areas, while the NESOSIM_M-PF method performs well in both snow buoy and OIB distribution areas (e.g., the Beaufort Sea and Chukchi Sea).

Table 3. Accuracy evaluation of the Kilic19 and modified W99 methods through the number of same matching points (Ns), RMSE (cm), bias (cm), MAE (cm) and r based on the OIB_{QL} snow depth from 2018 to 2019, OIB_{IDCSI4} snow depth in 2013, OIB_{NOAA} snow depth from 2014 to 2015, and the MOSAiC-measured snow depth from 2019 to 2020.

	OIB_{QL}		OIB_{IDCSI4}	
	Kilic19	modified W99	Kili19	modified W99
Ns	428		378	/
RMSE (cm)	10.78	/	6.66	
Bias (cm)	2.12	/	1.89	/
MAE (cm)	9.07	/	4.50	/
r	0.72	/	0.85	/
	OIB_{NOAA}		MOSAiC	
	Kilic19	modified W99	Kili19	modified W99
Ns	535	/	67	67
RMSE (cm)	13.87	/	9.39	13.38
Bias (cm)	5.01	/	6.50	11.19
MAE (cm)	10.83	/	8.35	11.51
r	0.62	/	0.01	0.37

Section 4.3 seemed superfluous. You compared against one snow depth dataset and the

modified Warren climatology but without much context to guide this.

Response: Thank you for this thought-provoking suggestion. According to the suggestion, we have added some content to guide the comparisons against the Kilic19 snow depth and modified W99 climatology.

The detailed revisions are as follows:

The modified W99 climatology, a public product, is often used to calculate sea ice thickness, and it is widely recognized by the public. So, we compare the snow depth obtained in this study with modified W99 climatology. Kilic19 model is a relatively new snow depth model at present. The Kilic19 snow depth is calculated using the Kilic19 model. Therefore, we compare the three snow depth datasets obtained in this study with Kilic19 snow depth as well.

Figure 8a - I think something odd is happening in Figure 8a for that big start of October jump in snow depth. This needs to be looked into.

Response: Thank you for this thought-provoking suggestion. Except for August and September, the satellite-derived snow depth has been used for assimilation. There is no satellite-derived snow depth in August and September. Therefore, the estimated snow depth in August and September is the NESOSIM_M snow depth, resulting in the increased jump at the end of September.

We are sorry we ignored this increased jump earlier. To solve the increased jump at the end of September, we use the NESOSIM_M-PF snow depth and NESOSIM_M snow depth at the same time and location in October to establish the linear regression equation as follows:

$$h_{NESOSIM\_M\text{-}PF}=1.2138\times h_{NESOSIM\_M}+0.9214 \tag{21}$$

We use Eq. (21) to obtain NESOSIM_M-PF snow depths in August and September. The results show that the increased jump at the end of September disappears and variation in snow depth from September to May is more reasonable (Fig. AA).

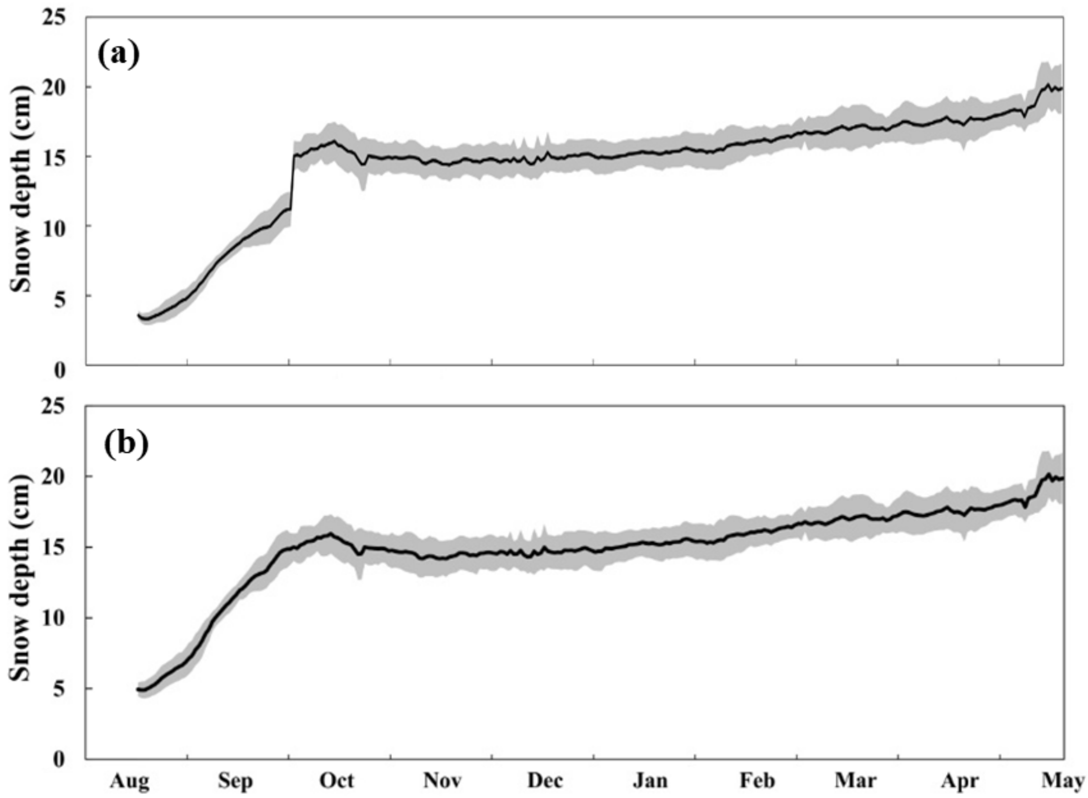

Figure AA. (a) Variations in the daily average Arctic default snow depth (no post-processing). (b) Variations in the daily average Arctic default snow depth (with post-processing).

In the revised manuscript, we have added the additional processing for eliminating the increased jump at the end of September. The revisions are as follows:

Except for August and September, the satellite-derived snow depth has been used for assimilation. There is no satellite-derived snow depth in August and September. Therefore, the estimated snow depth in August and September is the NESOSIM_M snow depth, resulting in the increased jump at the end of September. To solve this problem, we use the NESOSIM_M-PF snow depth and NESOSIM_M snow depth at the same time and location in October to establish the linear regression equation as follows:

$$h_{NESOSIM\_M\text{-}PF}=1.2138\times h_{NESOSIM\_M}+0.9214 \quad (21)$$

We use Eq. (21) to obtain NESOSIM_M-PF snow depths in August and September.

L444-445: 'The satellite-derived snow depth contains an uncertainty of 1 cm, and the NESOSIM snow depth uncertainty 445 reaches 5 cm (Petty et al., 2020).' Not sure where this is from. An uncertainty of 1 cm on what I assume are your snow depth measurements can't be right.

Response: Thank you for this thought-provoking suggestion. We are sorry that we do not clarify the uncertainty of RA-5VLSTM snow depth. The uncertainty of 1 cm only indicates the uncertainty caused by the uncertainty of input parameters. In the revised manuscript, we have explained the meaning of RA-5VLSTM snow depth uncertainty.

The detailed revisions are as follows:

We adopt the Monte Carlo method (Braakmann-Folgmann and Donlon, 2019) to determine the NESOSIM_M-PF snow depth uncertainty. RA-5VLSTM model is developed based on deep learning. So, it is difficult to obtain its total uncertainty. The satellite-derived snow depth uncertainty is calculated based on the Monte Carlo method; the estimated uncertainty is 1 cm, and it refers to the uncertainty caused by the input parameters.

References

Kwok, R., Kurtz, N. T., Brucker, L., Ivanoff, A., Newman, T., Farrell, S. L., King, J., Howell, S., Webster, M. A., Paden, J., Leuschen, C., MacGregor, J. A., Richter-Menge, J., Harbeck, J., and Tschudi, M.: Intercomparison of snow depth retrievals over Arctic sea ice from radar data acquired by Operation IceBridge, The Cryosphere, 11, 2571–2593, https://doi.org/10.5194/tc11-2571-2017, 2017.

Laxon, S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang, J., Haas, C., Hendricks, S., Krishfield, R., Kurtz, N., Far- rell, S., and Davidson, M.: CryoSat-2 estimates of Arctic sea ice thickness and volume, Geophys. Res. Lett., 40, 732–737, https://doi.org/10.1002/grl.50193, 2013.

Petty, A. A., M. Webster, L. N. Boisvert, T. Markus (2018), The NASA Eulerian Snow on Sea Ice Model (NESOSIM) v1.0: Initial model development and analysis, Geosci. Model Dev., doi: 10.5194/gmd-11-4577-2018.