

Interactive comment on “The NASA Eulerian Snow on Sea Ice Model (NESOSIM): Initial model development and analysis” by Alek A. Petty et al.

Anonymous Referee #3

Received and published: 11 July 2018

The authors present a new open source model, the NASA Eulerian Snow on Sea Ice Model, for estimating daily depth and density of snow on sea ice. The authors note at a few points in the paper that the model is being developed primarily with application to altimetry-based ice thickness determination in mind, though other applications are likely. The model is a simple representation of the snow that is largely an accounting of snowfall produced by reanalysis data, similar to prior efforts (e.g. Maksym and Markus 2008; Kwok and Cunningham, 2008), with terms for snow compaction, loss to leads, and transport on sea ice. It is Eulerian, but features pseudo transport by exchange between grid cells, features only 2 layers, and is forced with available spatially and temporally complete datasets that are known to be of limited accuracy (e.g. Reanalysis, passive microwave concentration). The model is calibrated/validated against limited

C1

available snow on sea ice data from Operation Ice Bridge and from 1980s era Soviet drifting stations. The description of the model is complete and in this regard the model is publishable with minor revisions – but reviewer doesn't feel the model is very good or useful in its current form for its intended purpose. Reviewer focuses most of this review on highlighting its shortcomings. In fact, a possible conclusion of this data presented would be that simple treatment of snow on sea ice will not meet the accuracy levels required for altimetry applications. The reviewer encourages the early career team to put the paper aside for awhile and take the time write a model that would actually be highly used.

The reviewer feels that the key issues are that the model is excessively simplistic, not representative of known physical process (even at the level of simplicity targeted), and that its results show it is inadequate for the intended purpose. There are errors in the equations presented, many compromises appear to have been made that make accuracy and/or realism lower in favor of rapid release, and as a result the work is unlikely to have much impact as presented. The presentation in the paper is quite long, and focuses on trying to convince the reader that the model is good, rather than taking a hard look and comparing against a reasonable standard.

The development of a snow product for improving retrieval of sea ice thickness from altimetry is critical for ICESAT 2 to be useful and this team should have NASA's support to do just that. Such a snow model's accuracy goal must be based on a desired accuracy in thickness retrievals. (e.g. retrieval of ice thickness accurate to ± 0.5 m over a given domain demands snow depth accurate to 0.5 cm over the same domain). The model presented here is not up to meeting these kinds of needs, and does not leverage the existing (more sophisticated) models of snow on sea ice (e.g. LIM, SnowModel, CICE).

Some major issues include: Model design relative to state of knowledge: 1. The two layers used in the two layer model (new snow and windslab) do not represent the two layers of the snowpack discussed in literature (wind slab and depth hoar). Authors

C2

cite and discuss the literature indicating that windslab and depth hoar dominate the mass of the pack and have quite different density – then ignore these decades of observation to invent a new scheme unsupported by observations. Respecting the effort to create a simple, 2 layer model, new snow should not be one of the two layers. The references cited clearly state that new snow rarely comprises much of the Arctic snowpack, because it is very rapidly converted to windslab. The preservation of a new snow layer appeared to be designed for modeling loss of snow into leads – but little is known about the magnitude of this flux, and it was minor in this model. 2. The model is operated on a 100x100km grid, which is very coarse relative to the variability in ice – which is shown to be important in impacting the accumulation of snow. The data sets used provide much higher ice concentration, and movement information – this data should be used at full resolution and atmospheric data can be downsampled. 3. Melt is neglected despite it being important during part of the timeframe and having significant impact on results.

Quality of the Model Results and Characterization thereof 1. Validation shown indicates the model produces results that do not capture the variability in observed snow depth or density reliably. Authors focus on averages of model output over decadal timeframes, which can be made to match observations by tuning of the arbitrary, non-physical constants in the model. This focus fails to acknowledge the inability of the model to capture interannual or spatial variability. 2. Prediction intervals are not provided, but scatter plots show little relationship between individual observations of snow depth and modeled snow depth. No discussion is provided of how these errors would propagate in the intended use (altimetry retrievals of ice thickness) but it appears errors are sufficient to radically alter retrievals of depth and appear to indicate the data would not be useful for altimetry retrievals of ice thickness from ICESAT2. Authors fail to acknowledge any of these shortcomings and go to great pains to make the results appear good. 3. Modeled variability in density appears to have very little relationship to observations. 4. Comparison with the southern ocean, are pushed to a future effort, but validation statements in the paper suggest the model applies to ‘polar oceans’. 5.

C3

Results from the median of the three reanalysis products are declared ‘better’ repeatedly with no reasonable support. Taking the median of atmospheric reanalysis models would result in nonphysical jumps between atmospheric states and the removal of extreme events from the record, and is challenging to support physically.

DETAILED COMMENTS Page 1, line 16. “very strong agreement” Delete “very strong”

Page 1 line 22 descriptions of agreement too subjective. The use here is altimetry. Tell the reader about the error in estimates implied.

Page 2 line 5-8. Poorly worded sentence. Consider modifying. One suggestion is: The altimetry technique involves measurements of freeboard, the extension of sea ice or snow surface above a local sea level. Estimates of snow depth are required to derive sea ice thickness from either snow surface freeboard or ice freeboard, because snow depresses ice freeboard and adds to snow surface freeboard. Snow depth is one of the primary sources of uncertainty for both laser and radar altimetry (e.g. Giles et al., 2007).

Page 2 line 10. Replace ‘lacking’ with something more descriptive/accurate (they aren’t lacking they are just not complete/good enough).

Page 2 line 22-24. The sea ice community often relies on simple models of snow depth forced by reanalyses – please clarify how this is different. To the reader, it still looks like a simple model forced by reanalyses!

P 3 Line 16 “and two snow layers to broadly represent the evolution of both old/compacted snow and new/fresh snow.” The assignment of the two layers in this two layer model is not consistent with the widespread understanding of the primary two layers on sea ice as depth hoar and windslab. New snow is occasionally present but usually rapidly transformed to windslab. It may be an acceptable third layer. See many of the snow on sea ice references cited here, such as Sturm et al., 2002 – generally the snow is treated in these two layers. The author’s choice here to take the

C4

two layers to represent layers that the extensive literature reviewed does not discuss is perplexing.

P2 line 18 replace “detailed” with “iterative”. The simplified scheme does not permit a ‘detailed’ assessment of connection between input data and snow depth given its lack of physical complexity – it permits an easier iteration of possibilities.

P4 line 13 Input data from passive microwave higher resolution than 100x100km, even if atmospheric data is not. Since ice concentration is so important, reviewer questions if 100km resolution is adequate. Further - does observed snow depth vary over 100km resolution? Since this is the motivation, what resolution is needed for useful for altimetry based determination of sea ice thickness?

Page 4 line 14 add “from reanalysis data” after the word ‘drift’.

Page 4 line 16 – (volume of snow per unit grid cell in units of meters) – doesn’t make sense volume is meters cubed. Throughout the treatment of snow varies between depth and volume freely, but this free transition between volume and depth is challenged for some considerations of snow – particularly convergence/divergences. Since the goal here is to understand depth for altimetry retrieval, a convergence, which moves volume into a cell, is not the same as a change in depth.

Page 5 table one – put formal references to data sources, e.g. “bootstrap” is not sufficient.

Page 5 delete “snow pit and density data. . . helped guide. . . parameterization . . . seasonal evolution.” There is no prescribed seasonal evolution of density, use of snow pit data etc. in this model. Two constant snow densities are selected and declared. This sentence obfuscates the very simple, non-experimentally supported nature of the scheme.

Page 6 line 8 replace bulk density with mass.

Page 6 – here authors note that the community of snow science experts and prior

C5

literature they have created generally group the snow into two layers (wind slab and depth hoar). They further note substantial differences observed in density of these two layers, and that these two layers comprise the majority of the snowpack. Not noted, but available in the literature is data showing that the contribution of the two layers to the overall snowpack varies from the approximately 50-50% contribution seen at SHEBA. So it seems windslab and depth hoar are the two layers to model. But. . . these two layers are different than the layers the authors have chosen (new snow/old snow). It seems a major departure from decades of snow research is being made here and it is not being well defended. Why?

Page 6 line 12 “for this reason we use the average of higher end values of ws and dh”. Reviewer sees no reason provided supporting the use of the higher end of the range of values for each of the two common layers. The mean density of each layer, multiplied by the mean fraction of each layer should provide a more representative density for the combined wind slab and depth hoar. Further, the value selected is not the average of the higher end of the range of values for each of the two common layers, leaving it unclear how it was determined.

Page 6 Line 16 “Our simple parameterization is thus expected to be generally representative” No reasonable evidence provided supports this. Statements like this are found throughout this paper. Delete or support with concrete evidence that quantifies what the range of uncertainty they will work within.

Page 6, Line 23 (default of 5m/s). Default or for the purposes of this work is it simply always set to this?

Page 6, Line 24 “determines the fraction. . . transferred. . .” Over what time? (seems that the coefficient is model timestep dependent. . . and perhaps shouldn’t be)

Page 6 line 26 ‘Wind threshold of 5m/s was determined based on. . .’ studies. Please add a description of the range of wind thresholds indicated by these studies, and why 5m/s was selected from within that range.

C6

Page 6 Line 8 Daily gridded ice drift is still required in this Eulerian scheme, eliminating it as a reason for choosing Eulerian over Lagrangian, discussed above.

Page 7, line 19. Reviewer is not aware of any evidence indicating that the loss of snow to leads in the North Atlantic sector of the Arctic is significant relative to the thick snowpack in that region. No evidence seems to be coming out of the N-ICE experiment to that effect. Some quantification of loss to leads in the Antarctic has been made by Leonard and Maksym as noted, but this was in the southern ocean. Please cite appropriate literature or delete speculation.

Page 8 line 4 – This parameterization doesn't make sense and is under supported for several reasons. 1. It appears that a constant coefficient beta is multiplied by 10m windspeed NOT by the amount which the wind speed exceeds the threshold velocity! So snow is lost to leads even when winds are too slow to move snow. 2. The amount of the snow lost to leads increases linearly with windspeed, when the drifting snow volume is well known to vary more rapidly than linearly 3. The loss to leads varies linearly with open water area, again this is likely more rapid than linear, and a thought experiment with random lead spacing/size could arrive at a better approximation. 4. The parameterization removes a fraction (2.5%) of the new snow layer to leads on each windy timestep – timestep is then important due to compounding what timestep is this defined for? 5. Is this parameterization/ value supported by any field quantification of loss to leads or is it simply made up due to lack of available observation. Either is fine, but state which it is. Page 6 line 9 – missing parenthesis on equation

Page 6 Equation 7 – appears incorrect. Change due to blowing snow is added (last term), but this should be a loss term (loss into leads). It appears that the term calculated in Eq 5 is always positive, so adding here will result in addition of snow, not loss. Similarly, how signs are handled on dynamics, convergence and divergence as well as advection depends on how (+-) u_i is defined in equation 3 and 4, and this is not (but should be) specified above... so the reviewer is unsure if the sign here is handled correctly.

C7

Page 8 line 21 August is mid- late summer. Change "early" to 'late' or delete.

Page 9 line 2-3 Do these melt events invalidate the results here? Is this model useful before these 'hoped for' additions occur? It sounds like this is being hurried along.

Page 9 line 8-13 This paragraph appears to handle a specific test case, not discussed here. Seems out of place possibly a draft fragment. Unclear what tests this new density applies to, or how this test relates to the model released for community use. (update after later reading, now understand what this refers to, but still feel it was out of place and not well enough contextualized here)

Page 9 line 16 Soviet - capitalize.

P 9 line 26 one OF the

P9 – would be appropriate to acknowledge the lack of validation sites or validation data over Arctic sea ice, and uncertain accuracy of the products in that region.

P11- Taking the median of the reanalysis products is an interesting idea if one has no idea which of the different products is best, but don't authors have better information about which is doing best from the comparison studies in literature?

P14 L10 – Initial conditions the Warren climatology is quite outdated. It is good you are trying to update them somehow. Is there evidence, e.g. from current autonomous ice mass balance buoys, that snow still regularly survives summer? Can you 'calibrate' this adjustment scheme based on those observations? Would a degree-day model be better than number of melting days? Also, what category is this snow placed in? Does it have a density reflective of melting snow (i.e. 400-500 kg/m³)?

P14 L22 – explain how this is 'linearly scaled' a bit better. Provide an equation. Is the fraction by which duration of melt is different from mean simply multiplied by snow depth? Does it mean that at 2x duration no snow is left and at 0x duration 2x snow is left?

C8

P14 L 28 – were necessary. . . Could this be because the model doesn't handle melt processes?

P15Fig2 – These substantial August snow depths in 2012 and 2013 should be compared against available buoy data to determine if they are reasonable. The reviewer believes they are not and that this is ultimately a nonphysical tuning mechanism that helps account for lack of melt processes and poor representation of precipitation phase at this time of year in reanalyses.

P 16 L 17 – this section is missing a clear statement of how accurate OIB data is expected to be.

P16 L 19-27 – pretty hand wavey – not rigorous.

Show plots of how snow evolves in model – what fraction of the snowpack is new snow layer vs time (it would have to be small to be realistic.)

P 17 Fig 3 – modeled data appears systemically low by about 5 cm depth. Snow density has essentially no relationship between modeled and observed. Individual year data – which is how this data would be used to derive altimetry based estimates of ice thickness – appear poor.

P17 L 13 delete “extremely” – an error of ~5cm on a snowpack of ~20cm is still a 25% error.

P18 L 1 – r of 0.74 would not generally be characterized as ‘strong’ P18 L2 - delete ‘more’ . . . its just moderate. Also, what is actually suggested here is that the model is good at predicting the MEAN – because you can tune your constants to make the mean look very nice, but not very good at capturing the interannual variability that is key to getting snow depth right for altimetry. P18,L4-5 “In General, the moderate/high correlations. . . provide confidence. . .” This statement is hand waving and cheerleader-y without content. Delete this statement and replace it with a statement that articulates the degree of certainty with which output of the model should be treated. Suggest

C9

authors calculate the +-95% prediction interval for a modeled density or depth relative to this dataset. Suggest authors do this for individual locations/months on individual years, as well as mean.

P18, Figure 4 – This comparison is really just showing how well tuned the models are on average. Since the model is not presented as a mean climatology, but rather is presented as a deterministic snow product for specific locations on specific years, this comparison is inadequate.

P19 L 3-6 Here authors make an odd argument. The model does not reproduce the climatology observations as well as a single mean over the entire timeframe. They argue this is OK because the model will handle interannual variability better because of its ‘more advanced’ density parameterization. The density parameterization is not particularly physically realistic, however, and fails to meaningfully capture interannual variability of the climatology density data (figure 3d). Reviewer therefore finds this statement lacking.

P19 I10 – not all marginal ice zones have low concentration – clarify that low concentration areas are where greater impact is expected

P 19 L 19 – Reviewer disagrees that wind threshold velocity for blowing snow is unconstrained. Resources reviewed can establish that under all but extreme conditions (e.g. recent rain on snow) a threshold of 10 m/s is pretty high, maybe unreasonable. It would be better to range this within the values observed in the references cited earlier.

P 20 L 15-18 – Speculative. Reviewer finds no reason to believe the median should be superior in regions of heavy snowfall. Defend or delete.

P21 L 5 – again this represents the mean over the decade being presented/compared. The model performance over this timeframe is highly tunable and not the performance metric of interest to an end user taking this data as an input to altimetry – that user would want to know the prediction interval for individual or moderate size groups of

C10

snow datapoints, and probably also whether there is any change in mean bias over time.

P 22 Fig 6 – are standard deviations in depths this low comparable to any observations? If so they suggest a single climatology would be adequate for most end uses.

P22 L 15 plurals

P23 L4 The fact that there is scatter among the reanalyses is not necessarily an argument for taking the median of them. Delete.

P23 L19 providing significant VOLUME REDUCTIONS and sinks of snow (wind packing is not a sink)

P24 L1 – Convergence really causes an increase in snow volume, not an increase in depth. The use of depth vs volume for a cell needs to be sorted out and treated consistently throughout this paper. Reviewer recalls a section way up at the top saying snow would be handled in volume throughout the paper, but has seen treatment vary.

P25 Fig 9b – unclear what “Ocean” refers to. Snowfall directly into water? Caption needs to be more descriptive and the figure subcomponents should be linked back to which equation # they represent. Fig 9f – this underscores the issue with treating convergence as a change in depth- in reality convergence/divergence of the ice at scale does not change depth in the sense that such would be used to interpret remote sensing. It changes snow volume, further, it appears the impact of convergence/divergence is noisy at best. Fig 9 l – density map warrants discussion. For example, density appears highest in central arctic (far from melt) This is likely untrue and an issue with not including melt/rain on snow processes. Very low density is indicated in marginal areas around N Greenland and in Baffin bay/Canadian islands. Can these be supported at all? Fig 9 h-k colors appear to fade toward lower value indication near land in general. Is this valid or a plotting artifact? Again volume and depth are used interchangeably and plotted on a depth scale. This needs to be resolved consistently. Page 24 L 5-7

C11

– Authors state these measurements are explicitly for altimetry retrievals, so they must have characteristics useful for such, (more than matching seasonal evolution on average) including: 1. Capture interannual variability 2. Capture spatial variability 3. No long term bias or trends in error (that could be mistaken as trends in ice thickness). If these cannot be shown, perhaps a discussion about whether this approach is viable is needed.

Page 26 Figure 10 – “As in figure 6” This is a nice addition to convey consistency, but pls also provide full description in caption, don’t make reader hunt back several pages.

Page 26 L15 Without melt processes included, what explains the loss in depth?

Page 26 L 5 Why the depths are different during the later time period IN THIS MODEL is within the scope of demonstrating a model, even if understanding why they are changing in reality is not. Please answer: Are depths less due to less ice for snow to fall on? Or due to less precip? The reader must know if the model is representing the changing Arctic – since it is calibrated on old data. This cannot reasonably be scoped out of the study.

P 27 Fig 11 – see comment on Figure 10

P 29 Fig 12 – please clarify what positive and negative deviation mean (is the product higher or lower snow than the median product)?

P 30 line 3 – it is not clear that the median provides a result any more useful than the others. One should note which product compared best to coastal stations data and any other indications from literature which might be best.

P 30 line 20 – this is not surprising and should be noted as such. Advection/convergence/divergence was much less important than snowfall in the plots above.

P30 L 23-24 – here is where the idea of snow depth vs snow volume is really important. Dynamics are perhaps not important in depth over a 100km cell average, but they are important to the DEPTH on the actual subgrid ice, since divergence creates new

C12

ice with no snow, rather than rearranging all the snow into a gridcell average. The averaging over the 100km cell at each timestep may be particularly important in ice generating areas, where snow is continually averaged back into source regions, rather than being advected out entirely. Tracking ice classes within the cells, as is done in CICE may be critically important.

P 31 figure 13 – the drift scheme matters little over huge areas because convergence and divergence cancel. This plot is just not the right way to consider this, particularly in the context of use for spatially distributed altimetry observations. Figure 14 suggests that the drift products don't differ that much between them in the central basin, but that having drift represented at all is very important, altering snowpack by O50% in large areas of the Arctic.

P32 L6 – There are actually substantial biases in the peripheral seas – which may not be important overall, but cannot be ignored in the statement about biases.

P33 – given the importance of concentration product, better understanding the role of changing concentration in the changing modeled snow depth above is important.

P34 – Tough to compare to observational data this noisy. Reviewer agrees they can be considered 'in agreement' within the bounds of the error of either. . . both of which are large. Are any of the OIB algorithms emerging as superior? Must all three be treated as equally likely?

P35 – comparison of 100km grid cells still includes substantial averaging, but already shows poor agreement. Agreement should be presented in terms of a 95 % prediction interval so user knows the capability of the method in useful terms – if the model says snow was xx OIB will say snow was xx +/- yy 95% of the time.

P36 – this discussion of the comparison of the scatter plots goes to great lengths to avoid describing the obvious. The model isn't very good at reproducing variability on OIB data, and if you believe OIB snow data is in any way representative of the variability

C13

in snow depth on ice, the modeled snow depth isn't very good at capturing spatial or interannual variability. The conclusion should then be that more sophisticated model representations are needed or that OIB data is trash. Since the model didn't agree with the Soviet drift station data scatter plot very well either, I don't think you can conclude that the model is adequate but OIB is trash.

P37 – "in general, however, the moderate to strong correlations. . . gives us confidence" Reviewer cussed in exasperation when reading this. This is a science paper not an opinion piece. These are not moderate to strong correlations! They clearly show NE-SOSIM cannot capture the variability observed well. Get this subjective language out of the paper and replace it with quantifications of how well the model does at both representing means (where performance is good because of tuning) and variability (where the model is not working so well). Talk about whether the model is good enough to be used in altimetry honestly and present some paths forward to getting there if it isn't.

P 37 L 19 – data is yet to be released in parenthesis? Thin its out now. . .

P 37 L 20 – There must be some field data available that you could at least spot check it against!

P 37, L 26 delete very strong, delete good

P 37 L 28 contributing to the MODELED seasonal evolution in snow depth

P38 L5 uncapitalize New, consider replacing with 'more recent.'

P38 L7 There is no evidence presented that this median product is better, and good reason to believe it just averages in erroneous values and non-physically jumps between atmospheric states toward limited representation of extreme events. Defend the use on scientific merit or consider deleting the median product.

P38 L10 use consistent language. . . it is 2nd order on mean, but first order in some regions.

C14

P 38 L 14 “moderate/strong correlations” This statement is flatly unsupported by the results shown, and authors ‘confidence’ in line 16 is unfounded. The product does not represent the OIB data well in terms of the intended use – in retrieving thickness from freeboard.

Please provide a variable list

Interactive comment on Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2018-84>, 2018.