It is good to see some scrutiny of snow on sea ice in CCSM. However, I was unsure how to interpret the comparison with Russian drifting stations. Fig 1 shows that error bars overlap between model and observation. Yet, the text seems to indicate that there is statistically significant differences in some months. It would be helpful to plot the standard error in Fig 1, rather than one sigma for the error bar. Further, it would be useful to explain how the standard error is computed, like whether spatial correlation is considered and how many degrees of freedom are assumed.

I have added some additional text discussing the statistical methods used. Regarding figure 1, the sample sizes are rather large (especially for the snow stake comparison) and switching from standard deviation to standard error basically removes the bars. I would like to retain the information about the variability, especially for the discussion, but can switch the figure if that is required.

"Following the matching process we have matched datasets of snow depth: transect to all ice thicknesses, transect to thin ice, and similarly for snow stake measurements. A paired student's t test was performed on each pair by month. Again, we do not expect the samples to be exactly matched due to the model's inability to simulate a given year. However, by matching time of year, geographical location, and year (mostly for era, not particular year), we produced paired samples. The degrees of freedom vary significantly, from as little as 3 in the summer for the transect to 2000 for the snow stake measurements. With the exception of summer, the DOF for the transect is approximately 50 for each month, and 2000 for the snow stake."

If we could believe the seasonality of the absolute bias, which is higher in summer than winter, does it suggest that the melt rate of snow on sea ice is too low?

I don't think so, a slight high bias persisting into the summer becomes a large high bias. So, if the snow is a little too high in the winter and the "correct" amount of snow is melted off, the remaining amount actually results in a very high bias. For example if you have 50cm of March snow in the model and 40cm in the in situ, and both melt 35cm you now have 15cm and 5cm. However, we could say that as a fraction of snow depth, we melt too little. I would suggest that precipitation biases and lack of sinks to snow mass such as blowing into leads are probably larger concerns. Moreover, there are very few samples in the summer, resulting in difficulty with statistical significance.

p 1508 line 12-14 stumped me at first where you say that it would be hard to conserve mass/energy if the snow density were altered. I guess you mean if the snow density varied in time rather than being held fixed in time it would be hard to conserve properties.

The general issue here is that it is not a simple issue to alter the density, which I believe the reviewer correctly determined.

I noticed Massom et al in you reference list, but it was not cited in your paper

The citation is back in the text.

You may find our GRL Hezel et al, 2012, to be interesting. We compared snow depths in the CMIP5 models with the Russian drifting stations via the Warren climatology and found CMIP5 models on average have snow too deep on sea ice. We also found thatCCSM4 had deeper snow than any of the 10 models we analyzed. We made figures(but did not include them) of the snowfall rate for each of the 10 models. The snowfall rate in CCSM4 was high but not extraordinary

I have added some additional discussion which included a reference to Hezel et al 2012, especially regarding reviewer #1's comment *"2) Have the authors considered examining the long term trends in snow over sea iceand how this may be related to changes in snowfall in the Arctic?"*