

## ***Interactive comment on “Sea-ice extent provides a limited metric of model performance” by D. Notz***

**Anonymous Referee #1**

Received and published: 22 August 2013

The author is on the trail of finding arguments to assess models using the integrated ice area as well as integrated ice extent. There is a valid point in here that invites open discussion.

The paper is too wordy and could make use of equations to define uncertainty terms for brevity and clarity. The main result, which is that use of both extent and area to constrain models (Fig 7) is underplayed whilst other issues relating to passive microwave retrieval inter-comparisons is going over old ground.

Unfortunately the paper also includes distractions, such as fig 5, which lead nowhere. Even fig 3 and 4, although having the potential to tell a story, are rather pointless within the paper objectives – leading to just one (rather ambiguous) conclusion. This topic needs to be pursued separately within a modelling context.

Finally, there is poor treatment of concepts for uncertainty both in models and obser-

C1527

ventions. This is little discussion in comparison with algorithm or model intercomparison studies e.g. Comiso, Stroeve etc.

Specific points.

Introduction : This contains messy concepts to the purpose of the paper. The issue surely is the transfer function between passive microwave (PM) observations and model. The model ice area is explicitly known (albeit with internal variability), but the observations are not – as it is just an empirical fit. Surely the issue is not to pick a particular algorithm which may be differently biased at different stages of the annual cycle, but to understand the observational error (including regridding, landmask etc). Because we are not in a position to forward model the ice area to PM space we must do the inverse. One does not use PM just to calculate the summer or winter extent but to evaluate the phase and amplitude of the seasonal cycle (within observational error). Since over ice PM principally detects water, the only reason that Bootstrap may depict a more dense ice pack than NASA-Team is if it is less sensitive to water – it cannot escape detecting meltponds and cannot discriminate them from leads. The ASI algorithm uses the high frequency channels which are by their nature less sensitive to water. In spatial plots of PM discontinuous ice concentrations, which do not reflect real concentration variations, sometimes occur when the Bootstrap algorithm switches between polarization and frequency schemes. This is of course invisible when integrated.

The discussion here on the difference between integrated extent and area is valid. However, integrated area does not assist in understanding the processes of albedo, heat and turbulent fluxes etc. For such studies one would do spatial plots of ice concentration or perhaps compare with observations of integrated albedo..

3100:22-24. This statement is unclear as you do go on a lot about comparison between Bootstrap and Team throughout section 3. I suggest removing all such references in section 3 and literally just refer to Bootstrap.

3103:29. No model yet has a prognostic floe-size distribution so it is the characteristic

C1528

floe size that is prescribed. The total lateral melt may depend on the modelled extent of the MIZ, and vertical distribution of solar heat absorption by the ocean.

3104:1-15. Not only do HadGEM2 and CCSM4 have multi-category ice but so do MIROC5 (but not MIROC4h) and NorESM1-M.

This is not a sensible method for discrimination between the models. The ice is just as compact if there is lots of 0% ice as well as 100% ice, so simply using a threshold on the 100% is overly simplistic. A better threshold would be on the gradient between the 90-100% and 80-90% bins. If you say that NASA-Team is wrong then this may provide a limit on the gradient threshold. However, a clear discrimination between models with multi-cat ice would still not be possible as MIROC4h does not have it.

Summer winds are rather light in the Arctic so this is unlikely to have a major effect on the ice distribution. More likely it is the initial ice thickness, which determines the ice strength, that has the major effect. Ice rheology may also have an effect as that determines the model spatial distribution of the ice in your rather limited polar zone.

Ironically to your argument here, many of the loose sea ice cover models are those which show best agreement with the observed sea ice decline. This implies that they are loose because of some feedback characteristic.

3105:1-15. This is an example of an emergent constraint from climate models and is a useful outcome of this study worth emphasising in the introduction and conclusions. (see Bracegirdle, Thomas J., David B. Stephenson, 2013: On the Robustness of Emergent Constraints Used in Multimodel Climate Change Projections of Arctic Warming. *J. Climate*, 26, 669–678. doi: <http://dx.doi.org/10.1175/JCLI-D-12-00537.1>). Plotting in this fashion is far less confusing than the obtuse figures 5 and 6.

3105:4-6. The number of points inside the 'observed' error box appear to be larger for the 'diffuse' models than the 'concentrated' models!

3106:7-26. Thin ice behaves differently than thick ice in the extent-vs-concentration

C1529

regime. In particular, thin ice is likely in free-drift whereas thick ice still has an internal stress component.

3109:1-2. The integrative means are still good for estimating the overall forcing of the ice to generate a seasonal cycle of the right amplitude and phase.. No modelling institute depends on these alone to assess their model performance and spatial characteristics are consequently important. Here, however, you are looking for a means to rank model ice performances. Others have done this and sensibly resorted to thickness pattern as a better assessment. If you must stick with passive microwave, integrated extent or area, then temporal variability (eg. related to NAO) could be assessed against 'observations'.

3109:10-28. This section is poorly expressed. You are talking about initial condition ensembles and the mean value and the standard error on a 27 year mean. This mean includes a trend and hence is not strictly speaking isolating the internal variability as some models may have a large trend and others not. Rather than quote upper and lower bounds for a single. There have been many studies on multidecadal oscillations in sea ice to quote (e.g. J J Day et al 2012 *Environ. Res. Lett.* 7 034011)

3109:17 "...mean September area..." 3109:21-22. What do you mean "estimate of the truth"? Do you mean "observational uncertainty"? You really only have a range here rather than an 'uncertainty'. Do not understand why you are using a range in a model ensemble to justify this uncertainty. Observational uncertainty comes from the observations not the models. What does it matter that one ensemble member is close to the Bootstrap September extent? 3109:22. The so called "117 CMIP5 simulations" are not independent as many of these are ensemble members. In any case you need to mention the model uncertainty first as there is no point in this statement until that is done 3109:25-27. No definition what 'close to Bootstrap' means quantitatively. Instead use 'a member which lies within the observational uncertainty range'. Does your quoted range refer to one specific ensemble or is the model uncertainty range across the subset of model ensembles, one member of which lies within the observed un-

C1530

certainty range? 3110:1-2. This could be misinterpreted. What you mean is that all ensemble members from the CMIP5 archive, not the individual model means. Considering all ensemble members as a group biases any interpretation towards the models with large ensembles. 3110:4. Check your figures – just 6 lines earlier you have given a figure of 6.9 million km<sup>2</sup> for Bootstrap extent. 3110:6-19. The discussion section will need to discuss why your analysis on trends is different from that of Stroeve et al (2012). 3110:20-23. It is not the case that more models lie outside the ‘acceptable’ range in trends for area than for extent. It seems to me that the take-home message from this section is that the trends are the same in area and extent, as each has its own internal consistency. 3111:1-11. I assume that in this paper you have been using monthly mean observed ice concentration products to infer ice extent. This will then be consistent with the calculation from the models. However, if one were to use an ice extent product derived from daily from ice concentration then a comparison with the model would be in error as you describe. 3112:11-12. Since you have no verification of these retrievals from other than passive microwave ‘observations’ this is almost certainly an underestimate of the observational as consequently there will be seasonal systematic biases in the retrievals (eg. water cloud in summer, different snow cover characteristics in winter – possibly wet snow with arctic cycle seasons in spring, thin ice during freeze-up). 3114:10-13. This is not proven in the paper, and is essentially idle speculation. To demonstrate this would require access to diagnostics not available through PCMDI. Apart from my previously expressed objection to your definition of ‘compact’, if you wish to include this then I suggest it is rephrased as ‘It is speculated that the difference in model summer ice distribution is associated with the partitioning of heat between lateral and vertical melting.’ However, since you have demonstrated that model internal variability dominates in the error budget, it becomes increasingly unlikely that this suggested interpretation is valid. 3114:20:23. Since models do not simulate trends which are outside the uncertainty band, this point is meaningless. Indeed, this is said in point 6. 3115:8-10. This is weighted towards models with large ensembles. It would be more accurate to specify how many models have no ensemble

C1531

member within the bounds. 3115:13-14. Specify that this refers to passive microwave satellite retrievals. Also note that the observational error does not include systematic biases associated with PM observations. 3115:19 Correction; ‘area’ rather than ‘are’

Figure 4: Clarity: overlap of x-axis numbering

---

Interactive comment on The Cryosphere Discuss., 7, 3095, 2013.

C1532