The Cryosphere Discuss., 7, C1857–C1861, 2013
www.the-cryosphere-discuss.net/7/C1857/2013/
© Author(s) 2013. This work is distributed under
the Creative Commons Attribute 3.0 License.

The Cryosphere
Discussions
Open Access

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

Anonymous Referee #2

Received and published: 24 September 2013

In "Sea-ice extent provides a limited metric of model performance" Dirk Notz provides a long overdue, comprehensive discussion on how different a metric sea ice area and extent are. He nicely presents the issues arising from focusing on either one in determining model skill. In the introduction he points out that sea ice area is the physically more meaningful measure; sea ice extent is, however, preferred by the remote sensing community since the estimation of the sea ice edge is associated with smaller errors than that of the overall ice concentration (or total area).

The paper is well structured and written. The author provides an easy to understand introduction to the two different quantities sea ice area and extent, studies uncertainties associated with different remote sensing data sets, and provides ample proof of how misleading skill assessment can be if only ice extent is used. Although the paper basically provides no new scientific insight the discussed topic is very important. Model evaluation is essential and sea ice coverage has been in the focus due to the availability

of extensive remote sensing coverage. However, as this study nicely demonstrates, sea ice extent is not the right metric for model evaluation although it is the preferred metric of the remote sensing community and has become publicly known as a measure of the annual sea ice minimum. It will be very helpful to have a decent study like this one as a reference.

I recommend the discussion paper for publication in The Cryosphere. Nevertheless, I have a few minor comments the author may want to consider before final publication.

Detailed comments:

p.3097 l. 28 ". . . shift in the location or its spatial distribution of the sea-ice cover . . ."

p.3099 l.19f ". . . time series of monthly mean sea-ice extent . . . from their monthly mean sea-ice concentration . . ."

p.3100 l.6f please mention here that the time series based on the ASI algorithm cover a shorter time period; specifically state the covered periods.

p.3102 l.19 I wonder if there is a more objective measure for "compact" and "loose" ice covers based on the red line in Fig. 7a, for example using a standard deviation or squared error to measure an acceptable deviation from this line.

p.3103 l.4ff With respect to an earlier statement in the paper that part of the "open water" seen in the NASA Team ice concentrations is actually meltwater-covered ice, I wonder about the value of the differentiated statement made here. I think it needs to be stated even more clearly, that lower summer ice concentrations in "loose" models vs. "compact" models really mean more open water whereas the difference between Bootstrap and NASA Team ice concentrations may mostly refer to wet ice/snow and water on ice but not open water—physically this has very different effect.

p.3107 l.16 I think that September 2007 is not a good example for testing the effect of grid resolution. If I am not mistaken grid resolution should matter most for a generally loose ice cover. However, summer 2007 featured a record low ice concentration be-

cause the atmospheric circulation compacted the otherwise loose ice pack to have a record low extent. Therefore, I assume that extent and area were not that much different in September 2007. I think a summer with a great difference in area and extent would suit the test better.

p.3107 l.24 what does “small” mean here, please give a number

p.3109 l.10ff I am confused by the definition of “internal variability” and the yellow shading applied to Fig. 6. The text says that internal variability is measured in terms of the spread of ensemble members per model. However, the yellow shading seems to be constant across models, i.e. indicates the spread considering all 117 simulations at once. Both definitions could work as a measure of internal variability, just be clear.

p.3110 l.11 This sentence could be interpreted as a strong argument against anthropogenic (or CO2) forced climate change since the spread in trends is huge and the observed trends is within the range of modeled internal variability as are even positive trends shown by some simulations. This would not be the case if internal variability would be assessed for each model individually. Then some models would fail to reproduce the observed trend within their range of internal variability, which would be much more a statement about model quality, which the author is aiming for, I guess.

p.3110 l.21 replace “much” by “many”

p.3112 l.9 “(green curve in Fig. 13e)”

p.3111 l.24/27 (anomalies/offset), p.3112 l.21ff (absolute bias) In general I agree with the statements made in these lines. However, we should take into account that a fully coupled model system has many degrees of freedom and take into consideration that different models are designed for different purposes. I expect a good coupled model to correctly (say within the observed variability) simulate first the mean state of total ice area(!) and volume and second the observed variability (range of anomalies though not necessarily in the observed sequence) and third the observed trend. In this sense

the absolute bias is a very hard to meet criterion that unnecessarily may make many models look to perform badly. If I want to study regional changes in particular with a focus on the observational period I would rather use a forced ice-ocean model, which I indeed expect to meet the absolute bias criterion. This said, I wouldn’t say that the absolute bias is the most meaningful measure, it really depends on the purpose; it’s probably the most precise measure of model performance.

Figure 1 nice example! However, I would rather have a “top view” perspective than a “side view”, i.e. ice floes should be fully contained in blue ocean boxes covering the indicated area; all boxes should be blue independent of having ice in them or not (currently left box in panel (a) and left and right boxes in panel (c) are white) to be related to open water.

Figure 2 legend: name satellites/sensors which are used for Bootstrap and NASA Team

Figures 2, 6, 7, 10, 13 use same color key for satellite data in all four figures, i.e. red=Bootstrap, green=NASA Team, blue=ASI SSMI, and black=ASI AMSR

Figures 3 and 4 Would be great to note the ice-covered (>15%) area fraction in each panel (print a number in each panel). This would provide a quick access to differences in ice extent.

Figure 5 This Figure is awfully dense and very difficult to read. Considering how it is used on p.3103 I suggest showing 4 scatter plots instead: concentration change vs. volume loss, thickness change vs. volume loss, concentration change vs. September concentration, and thickness change vs. September concentration with a single dot for each simulation. The “loose” models should then accumulate in a different part of each plot than the compact models.

Figure 6 I suggest to group models by “compact” and “loose” ice conditions and then sort them alphabetically within each group. As they are listed now I have a hard time to get a quick impression on how the one or the other group performs. Moreover, internal

variability should be assessed per model not across models (see comment above).

Figure 7 I suggest to add a 1:1 line (slope 1) to both panels to get a quick sense of the difference between area and extent and for better orientation. And I recommend to use vertical and horizontal error bars centered on the red dot (or a gray shaded cross as in Figure 10) to indicate a reasonable deviation fro the observation instead of the gray block on the right. Also, please mention in the caption that blue “loose ice” dots are only shown in panel (a).

Figure 8 I don’t think this figure is necessary. I think it is obvious from Fig. 7a that dots (models) along the green line with excessive extent have a small bias in area and that dots at smaller than observed ice extents will have even greater area bias. This is very straight forward. In case the author should remove this figure, text between p.3105 l.16 and p.3106 l.6 should be shortened considerably; in case the author wants to keep this figure, I recommend to add a statement in the figure or its caption that the distance between the blue and read lines of Model 1 and Model 2 are the same (at least they should be).

Figure 9 print names of “compact” models in red as in Figure 6

Figure 11 what is the threshold for “ice-covered” here, 15%? Please state this in the caption

Figure 12 typo: panel (c) is labeled (b)

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