Review of paper by Lindsay and Schweiger

General Comments:

This is an interesting paper. The authors attempt to combine disparate data sets to tease out the spatial and temporal trends in Arctic sea ice thickness. The primary assumption in their regression is that there is one number that describes the bias of a given data set relative to a selected reference data set. And, once these biases are identified then all the estimates from different instruments could be combined into a 'rationalized' data set for understanding spatial and temporal trends.

I think, this assumption is only valid if there were no space- and time-varying biases in the reference data set. That is, the reference data set has to be internally consistent: no relative biases in time or space within the data set itself. Take ICESat:G for example, if the thickness of multiyear ice were systematically under-estimated compared to the thickness of seasonal ice and in addition their relative biases changes in time (t) and space (X) due to snow loading, then a single parameter will not adequately describe these biases, since b_j has to be represented as $b_j(X, t) = \text{const} + f_j(X, t)$. And, this is probably the case for ICESat:G. In these cases, the time- and space-varying biases are subsumed by other terms in the regression, which could cause biases in spatial and temporal trends, and in the magnitude of the relative biases. In fact, the authors alluded to this in the discussion of systematic biases in Section 4.1 that the intercomparisons in one region may not be applicable elsewhere.

In the end, I am not entirely convinced that b_j is adequate for describing the biases of each data set relative to the reference data set. And, the data set may not be rich enough to assume that $f_j(X, t)$ somehow averages to zero. In the conclusion section, the authors cautioned that the bias term should be interpreted only in a relative sense, I would add that that there are hidden (unmodeled) biases in a regression analysis of this sort although the magnitude is unknown.

In summary, this represents an interesting attempt at using all available data set, the results seem reasonable. I recommend publication after revision. I should add that I had fun reading this paper.

R. Kwok

Detailed Comments:

Page:Line number

4547: 23 – Kwok et al., 2008 offer a preliminary examination of four campaign and does not cover 2003-2008. Please use Kwok et al., 2009.

4549: 13 – re:Submarines. For the years after 2000, substantial portions of the submarine cruises are not under multiyear ice. Should/would the biases due to beamwidths and depths be different (time and space-varying)? Perhaps a short discussion is warranted because the impact of these biases could be significant when compared with other data sets. Some of these issues are discussed in Section 5.

4550:3 re: Air-EM. The uncertainty of the ice thickness from Air-EM is not clearly stated as uncertainty in snow depth from PIOMAS has not been assessed. I suppose the impact of errors in snow depth should be relatively small for the purposes here. In any case, this should be addressed the NCEP-NCAR precipitation could be biases. (why not use Warren?)

4551:1 re: BGEP data: there is no comment on data quality?

4551: IceBridge data – the error is dependent on the snow depth as well and could be a significant source of error.

4551: IOS-CHK, IOS-EBS: should comment on the data quality. Melling claims 0.1 cm.

4552: ICESat-J. A clarification:
1. The results in Kwok and Cunningham, 2008 were based on a preliminary analysis and do not include the improvements in the estimates as discussed in Kwok et al., 2009.
2. Within the 25-km segments, open water samples are included in the creation of the gridded field. However, in there is no accounting for the overall ice concentration within a grid cel after data accumulation and interpolation.

Question: would it be useful to tabulate the expected error (well, our best understanding) of each data source even though they are not used in the regression.

4553:10 Comments:
1. f(m) is probably lower over first-year ice. (reduced accumulation)
2. The lower density of older ice (could be down to 880 kg m-3) would introduce variability in your thickness estimates.
So, what is the overall uncertainty in the submarine-derived ice thickness estimates after all this and including the uncertainty in ice draft? adding to the uncertainty of ~0.25 m in ice draft?

4554:20 In the regression, the uncertainties of the measurement source are not considered?

4553:23 I think Kwok et al. (2009, not 2008) pointed out this was unrealistic. The following is probably better phrased: "…. this assumption is unrealistic (Kwok et al., 2009) since the sea ice….'

4554:25 It would be useful to show all the terms T(x,y,t) used in the regression.

Table 2. The indicator coefficients are NOT ordered by the magnitude of the coefficients.

Figure 3. It would be interesting to show the RMSerror of the fits here as well.

4557:18-27 The submarine ice thickness used here is different from submarine-derived ice thickness used in Kwok et al. (2009), in which the snow was not accounted for. With the snow loading in your sub data, it would make sense that the resulting comparison would be different (signs). There is no inconsistency in the two comparisons. So, what do you make of the difference between the comparison between the ITRP ice thickness and ICESat:J vs between ICESat:J and ICESat:G?
It seems that if the indicators were correct that I should expect the difference between ICESat:J (b = 0.42 m) and submarine (b = ~ -0.1 m) to be more like 0.5 m, but the difference is only ~0.1-0.2 m if directly comparisons were made. I am somewhat puzzled. So, there must be spatial gradients or some other differences between the two data sets that caused the differences?

4557:18-27 Aren't these comparisons (2000-2013) dominated by the large populations of ICESat:J and ICESat:G (only 2005-2008) estimates and thus just a measure of the difference between the two? And if the trend in 2000-2013 was not exactly linear, the ITRP would pick the a trend that fits through ICESat:G? Looking at Figure 1, between 2000 and 2013, the population is dominated by the satellite retrievals. There is submarine data only in 2005 and 2013. If this were correct then it would explain my remarks in my previous point.

4558:20 …. as well as the size of the sample population?

4558:21 My remarks above for the other regions as well: Beaufort, North Pole, Lincoln Sea.

Figure 4. It was indicated that observations were adjusted for the bias of each. This was not discussed, but perhaps should be mentioned, in Section 4.2.

4563:15 Withholding data is a good idea although it does not change the fact that the sample populations are biased to a given sensor.

Figure 6: I should expect that much difference at all, so yes it is robust in this sense.