Response to Referee #1

Comment: This manuscript evaluated the sea ice simulations by 49 CMIP5 models over the Arctic and Antarctic for the period of 1979-2005. The authors compared the model simulated sea ice extent (SIE) with that derived from satellite-based observations (NASA algorithm), and sea ice volume (SIV) with output from the Global Ice-Ocean Modeling and Assimilation System (GIOMAS). The assessment was done for both the Arctic and Antarctic region. Annual cycle and linear trend from multi-model ensemble means are compared with the observation. This is a nice summary about the CMIP5 model sea ice simulations, and could have added value to the pile of publications in this research area, especially considering that most of the preciously published paper deal with part of the CMIP5 models due to their publication date. The work in the manuscript is very similar to what Parkinson et al. published in 2006 about CMIP3 models. However only the "multi-model ensemble mean" (MME) were presented in the manuscript. It would be nice to show each individual model's performance as other peoples do, for example, Massonnet et al. (2012), Stroeve et al. (2012), Wang and Overland (2012) and Liu et al. (2013). While most previous studies focused their discussion on SIE simulations, this manuscript also evaluated SIV, which is an importance variable to be considered. This is an very important addition.

Reply: We are very grateful to the above comments and all the following thoughtful suggestions. The quality of this manuscript is much improved through considering these comments. The revised manuscript has been modified according to the reviewer's suggestions.

Comment: In the introduction the authors state "assessment of the performance of CMIP5 model outputs is necessary for scientists to decide which model outputs to use in their research". Yet throughout the text we could not tell which model(s) perform better in their standard since only the MME were presented. We should NOT compare

the MME with observation as observation represents only ONE-single realization.

Reply: Thank you for the suggestion. We agree with you that observation is only ONE-single realization and we should compare each model simulation with the observation and give the performance of each CMIP5 model. The performance of each CMIP5 model is listed in Table 1. But detail information of Table 1 in the old version of manuscript is not enough as mentioned in your next suggestion. In the revised manuscript we added more description of Table 1. The performance of each model is evaluated by climatological sea ice extent (SIE) bias (root mean square (RMS) error) compared with satellite observation, simulated SIE linear trend, climatological sea ice volume (SIV) bias (RMS error) compared with GIOMAS output, and simulated SIV linear trend for both Arctic and Antarctic. Climatological SIE and SIV biases can evaluate the performance of long-term mean state of sea ice. SIE and SIV linear trends can evaluate the performance of sea ice changing speed.

Revision in manuscript:

Line 25 of P.8: "From Tables 1 and 2 we can conclude that the performance of each model is different. For the Antarctic, ACCESS1.0, BCC-CSM1.1, CESM1-CAM5-1-FV2, CMCC-CM, EC-EARTH, GISS-E2-H-CC, MIROC-ESM, MIROC-ESM-CHEM, MRI-CGCM3, MRI-ESM1 and NorESM1-M can give better SIE and SIV mean state. For the Arctic, ACCESS1.3, CCSM4, CESM1-BGC, CESM1-CAM5, CESM1-CAM5-1-FV2, CESM1-FASTCHEM, EC-EARTH, MIROC5, NorESM1-M and NorESM1-ME can give better mean state of SIE and SIV. The Arctic SIE linear trends of BNU-ESM, CanCM4, CESM1-FASTCHEM, EC-EARTH, GFDL-CM2p1, HadCM3, HadGEM2-AO, MIROC-ESM-CHEM, MPI-ESM-MR and MRI-ESM1 are closed to the observations.".

Comment: In Table 1, CMIP5 simulated errors and trends are listed for each individual model. What are these RMS errors and linear trend? Annual mean? Seasonal mean? or what? It is unclear from the text. In Stroeve et al. paper, 20 CMIP5 models' trends were presented already. How does your result compare with their

result?

Reply: Thank you for your valuable suggestion. In Table 1, RMS errors are seasonal mean RMS errors. Linear trends are the trends of the monthly anomalies. We added these two sentences in the revised manuscript. In Stroeve et al. paper, Arctic SIE trends are for the period of 1979-2011, but here the SIE trends are for the period of 1979-2005 since CMIP5 historical run is from 1850 to 2005. So it is hard to compare these trends for two different periods.

Revision in manuscript:

Line 1 of P.12: "RMS errors are seasonal mean RMS errors. Linear trends are the trends of the monthly anomalies.".

Comment: Comparing CMIP5 model simulated SIV with output from GIOMAS is fine since we do not have systematic pa-Arctic observed ice thickness yet. But please do remember that those are model simulations, and therefore it is not true "observation".

Reply: Thank you for your reminding.

Revision in manuscript:

Line 4 of P.2: changed "Global Ice–Ocean Modeling and Assimilation System (GIOMAS) data in this study" to "Global Ice–Ocean Modeling and Assimilation System (GIOMAS) output data in this study".

Line 20 of P.4: "SIV data from GIOMAS are not observations but model simulations with data assimilation."

Finally, we hope to express our sincere thanks again for all these valuable comments and suggestions.

Response to Referee #2

Overall assessment: The authors present a broad overview assessment of the performance of the CMIP5 simulations of sea ice extent and sea ice volume in both the Antarctic and Arctic regions. 49 different CMIP5 models are used in this assessment and comprises the most inclusive set of CMIP5 results for sea ice in publications to date and represents a significant effort. Beyond this however, the paper lacks a clear focus or purpose, and fails to provide new insight or information beyond that already contained in previous assessments of CMIP5 results from a subset of the models evaluated here, (e.g. Stroeve et al, 2012, Massonnet et al, 2012). The introduction is missing citations for key references for previous assessments of sea ice in the Arctic (e.g. Massonnet et. al., 2012) and the Antarctic (e.g. Zunz et al., 2013), and these are listed in Chapters 9 and 12 of the IPCC AR5 report published in Fall 2013.

Reply: We are very grateful to the above comments and all the following thoughtful suggestions. The quality of this manuscript is much improved through considering these comments. The revised manuscript has been modified according to the reviewer's suggestions. And the key references are citied in the revised manuscript.

Comment: Scientifically, the analysis focuses on a simple comparison of the multi-model ensemble mean to the satellite observed sea ice extent, as well as reanalyzed sea ice volume from the GIOMAS model. The strategy of assessing the multi-model ensemble mean to observations yields no insights into the behavior of any particular models, or assessment of which models do a better job at producing the mean state and trends over the satellite era based only on the historical period of CMIP5 (1979-2005). Though tedious, a more detailed evalutation of the model mean state, seasonal cycle, trends, and variability, would actually be a more useful reference for the community. This might involve expanding the number of fields in Table 1 to include more metrics, and indicating an assessment of model performance for each

metric.

Reply: Thank you for your valuable suggestions. For the Antarctic sea ice, the model internal variability is an important metric to evaluate the observed positive SIE trend (Zunz et al., 2013), so variability is included in Table1. For the Arctic sea ice, model mean state and seasonal cycle are important to Arctic sea ice projection (Massonnet et al., 2012), so model mean SIE and cycle amplitude are also included in Table 2.

My comments below contain some ideas that might lead to a more useful paper, and would expect that an expanded discussion of these would lead to a completely revised manuscript. Suggestions: If the goal of the paper is to identify CMIP5 models that do a reasonable job of reproducing sea ice characteristics, then it would be helpful to have (a) a clear set of criteria that can be evaluated for each model, and (b) the assessment of each model performance against those criteria. Massonnet et al, (2012), does this to answer a specific question related to the timing of the disappearance of Arctic ice. The idea might be not to find the best sea ice models, but rather the best models to address a particular question.

Reply: The main goal of this manuscript is to identify CMIP5 models that do a reasonable job of reproducing sea ice characteristics, and the metrics we used is SIE and SIV mean state, linear trend and model error. In fact it is hard to find the best sea ice models if we do not address a particular question as the reviewer pointed. In order to give more information of CMIP5 models' performance, we added more metrics in the revised manuscript. But these metrics do not address a particular question.

Revision in manuscript:

Line 2 of P.12: Model mean SIE and SIV, cycle amplitude, and variability of each model are added in the table.

Comment: GIOMAS sea ice volume data for the Antarctic has not been tested against the limited set of observations, but is the best available time series available now. Whether it represents a useful set of 'observations' to test model performance is

another question. For the Arctic, I suggest use of the PIOMAS data, which has been more extensively investigated (e.g., Schweiger et al, 2011). Since SIV is a poorly observed quantity, It would be also be worth mentioning how PIOMAS/GIOMAS SIV estimates compare against independent satellite estimates of SIV (e.g. Kurtz and Markus, 2013), especially for the Antarctic.

Reply: For sea ice volume data, we totally agree with your common, and in the revised manuscript we delete the word of 'observations' when discuss sea ice volume. In Figure 5b the legends of 'Observation anomaly' and 'Observation trend' are replaced by 'GIOMAS anomaly' and 'GIOMAS trend'. The legends in Figure 10b are also replaced. Following your suggestion, we used PIOMAS for the Arctic in the revised manuscript and all the models are assessed again.

One of the more interesting points in the paper is contained in the final paragraph, which assesses the number of models necessary to reduce the error between multimodel ensemble mean and observations. As the authors point out, the RMS error between the MME of both SIE and SIV compared to observations is minimized by the inclusion of about 22 models, which indicates that previous assessments of the MME (e.g., Turner et al, 2013) are not enhanced by the inclusion of additional ensemble members.

Reply: Figure 11 shows that the more model we used the smaller MME error we will get. But the conclusion that the CMIP5 MME cannot reproduce the observed slight increase of SIE is the same. Turner et al, 2013 used 18 models. Figure 11 shows that the ratios of 18 models and 22 models are very closed, and they are both close to the ratio all the models used.

An understanding of what causes the spread in SIV estimates in CMIP5 models would be a potentially useful line of inquiry. Perhaps models with a more realistic mean state or seasonal cycle results in a convergence of estimates of SIV.

Reply: The spread of CMIP5 SIV is large, especially for the Antarctic. We checked the correlation between SIE RMS error and SIV RMS error, and we can find that for

the Antarctic the models with small SIE RMS errors always have small SIV RMS errors. So for the Antarctic, the reviewer's point that models with a more realistic mean state result in a convergence of estimates of SIV is correct. But for the Arctic, this conclusion is not clear.

Revision in manuscript:

Line 2 of P.7: the sentences "We checked the correlation between SIE RMS error and SIV RMS error, and we can find that for the Antarctic the models with small SIE RMS errors always have small SIV RMS errors. It means that for the Antarctic models with a more realistic SIE mean state may result in a convergence of estimates of SIV" are added.

There is no reason one would expect the models to capture the observed trends in the exact time period 1979-2005 given the contribution of natural variability (roughly half) to the observed trend (see Kay et al 2011). The authors could explore the ability of the models to reproduce the observed 27 year trends in the vicinity of the same time period in the models. They would still need to address the potential confounding influence of differing sensitivity of Arctic/Antarctic sea ice loss per degree global warming.

Reply: Thank you for your suggestions. Exploring the ability of CMIP5 models to reproduce the observed 27 year trends in the vicinity period in the models and addressing the potential confounding influence of differing sensitivity of Arctic/Antarctic sea ice loss per degree global warming are not included in this manuscript. We will study in these directions in the following work.

Specific comments: Observed Antarctic SIE trends of 1.56 x 10^5 km^2/decade are not consistent with other literature, and it's not clear where this value comes from. My estimate using NSIDC sea ice index is trends of 1.12 x 10^5 km^2/dec if based on annual mean SIE or 1.29x10^5 km^2/dec if based on monthly anomalies for 1979-2005 (crudely ignoring missing data values). Turner et al (2013) quotes 1.27 x10^5 km^2/decade. Uncertainties should be calculated for all trends. See Stroeve et al,

2012 for suggestions.

Reply: Antarctic SIE we used is based on monthly mean sea ice concentration for both satellite observations (http://nsidc.org/data/seaice/) and CMIP5 models (http://pcmdi9.llnl.gov/esgf-web-fe/). SIE is computed as the total area of all grid cells where monthly SIC exceeds 15%. The trend is based on monthly mean SIE anomaly. If we use daily satellite observed sea ice concentration to get daily SIE and then calculate monthly SIE anomaly and Antarctic SIE trend, we will get the trend of 1.36 x 10^5 km^2/decade. If we use NSIDC sea ice daily index (ftp://sidads.colorado.edu/DATASETS/NOAA/G02135/) to calculated monthly SIE anomaly and then calculated trend, we will get the trend of 1.29 x 10^5 km^2/decade which is the same as the reviewer's result. So we think different datasets can give different trends. In the revised manuscript, NSIDC sea ice index is used with the linear trend of 1.29 x 10^5 km^2/decade. All the models are re-assessed again, and the related figures are updated.

Further discussion of the increase in Antarctic SIE should incorporate the recent revelation of Eisenman et al, (2014), which suggests that the trend may not be as strong as quoted recently.

Reply: Eisenman et al. (2014) give us a good reference when we use satellite observed sea ice concentration. They mainly focus on the Bootstrap algorithm. The satellite observed sea ice concentration we used is based on NASA Team algorithm, and the above-mentioned NSIDC sea ice index is also based on NASA Team algorithm. Although we don't know whether the Antarctic sea ice based on NASA Team algorithm has a jump, we also incorporate the recent revelation of Eisenman et al, (2014).

Revision in manuscript:

Line 6 of P.9: "Although satellite observed Antarctic SIE has increased trends, when we use satellite observed sea ice record, we should also keep in mind that satellite observed sea ice record may also has large uncertainty. Eisenman et al. (2014) point out that sensor transition may cause a substantial change in the long-term trend." is

added.

Correlations of the seasonal cycle of the MME compared to observations are not informative unless they are not highly correlated and would therefore indicate a substantial problem.

Reply: we delete the descriptions about the correlations of the seasonal cycle of the MME compared to observations.

Revision in manuscript:
Line 9 of P.5: "; the correlation coefficient between observations and MME is 0.996" is deleted.
Line 11 of P.7: "with a correlation coefficient of 0.997" is deleted.

In Table 1, it should be made clear how the RMS error of climatologies is computed. It would be useful to distinguish the error on the winter/summer means from the annual mean error.

Reply: Thank you for your valuable suggestion. In Table 1, RMS errors are monthly RMS errors. In the revised manuscript, sentences "Column (e) is monthly SIE room mean square error" and "Column (j) is monthly SIV room mean square error" are added.

References

Eisenman, I., Meier, W. N., and Norris, J. R.: A spurious jump in the satellite record: has Antarctic sea ice expansion been overestimated?, The Cryosphere, 8, 1289-1296, doi:10.5194/tc-8-1289-2014, 2014.

Kay, J. E., M. M. Holland, and A. Jahn (2011), InterˇAˇ Rannual to multiˇARˇ decadal Arctic sea ice extent trends in a warming world, Geophys. Res. Lett., 38, L15708, doi:10.1029/ 2011GL048008.

Kurtz, N. T., and T. Markus (2012), Satellite observations of Antarctic sea ice thickness and volume, J. Geophys. Res., 117, C08025, doi:10.1029/2012JC008141.

Massonnet, F., Fichefet, T., Goosse, H., Bitz, C. M., Philippon-Berthier, G., Holland, M. M., and Barriat, P.-Y.: Constraining projections of summer Arctic sea ice, The

Cryosphere, 6, 1383-1394, doi:10.5194/tc-6-1383-2012, 2012.

John Turner, Thomas J. Bracegirdle, Tony Phillips, Gareth J. Marshall, and J. Scott Hosking, 2013: An Initial Assessment of Antarctic Sea Ice Extent in the CMIP5 Models. J. Climate, 26, 1473–1484. doi: http://dx.doi.org/10.1175/JCLI-D-12-00068.1

Zunz, V., Goosse, H., and Massonnet, F.: How does internal variability influence the ability of CMIP5 models to reproduce the recent trend in Southern Ocean sea ice extent?, The Cryosphere, 7, 451-468, doi:10.5194/tc-7-451-2013, 2013.

Finally, we hope to express our sincere thanks again for all these valuable comments and suggestions.