Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations

Julienne C. Stroeve[1], Vladimir Kattsov[2], Andrew Barrett[1], Mark Serreze[1], Tatiana Pavlova[2], Marika Holland[3], and Walter N. Meier[1]

[1]National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA.

[2]Voeikov Main Geophysical Observatory, Roshydromet, St.Petersburg, Russia

[3]National Center for Atmospheric Research, Boulder, CO USA.

Abstract

The rapid retreat and thinning of the Arctic sea ice cover over the past several decades is one of the most striking manifestations of global climate change. Previous research revealed that the observed downward trend in September ice extent exceeded simulated trends from most models participating in the World Climate Research Programme Coupled Model Intercomparison Project Phase 3 (CMIP3). We show here that as a group, simulated trends from the models contributing to CMIP5 are more consistent with observations over the satellite era (1979-2011). Trends from most ensemble members and models nevertheless remain smaller than the observed value. Pointing to strong impacts of internal climate variability, 16% of the ensemble member trends over the satellite era are statistically indistinguishable from zero. Results from the CMIP5 models do not appear to have appreciably reduced uncertainty as to when a seasonally ice-free Arctic Ocean will be realized.

1. Introduction

As assessed from the modern satellite passive microwave data record that began in October 1978, Arctic sea ice extent exhibits downward linear trends in all months. The trend is largest for September (the end of the summer melt season), and appears to have steepened with time [*Stroeve et al.*, 2011]. Through 2011, the linear rate of decline for September is -12.9 $\pm$ 1.47% per decade, equating to a reduction of more than 30% since the late 1970s. This decline has been attended by a transition towards a thinner, younger

ice cover [e.g. *Kwok, 2007*; *Maslanik et al.*, 2007]. At the end of summer 2011, only 25% of the ice was more than two years old, compared to 50 – 60% during the 1980s [*Stroeve et al.,* 2011]. Almost none of the oldest and thickest ice (at least five years old) remains (3% in February 2012 compared to 30-40% in the 1980s). Impacts of ice loss include easier marine access to the Arctic, increased coastal erosion, changing ecosystems and, potentially, altered weather patterns in middle latitudes [e.g. *Francis and Vavrus*, 2012]. Better understanding the observed rate of sea ice loss and the ongoing evolution towards a seasonally ice-free Arctic Ocean has been cited as a "grand challenge" of climate science [*Kattsov et al.*, 2010].

Coupled global climate models not only provide estimates of when a seasonally ice-free Arctic may be realized, but also provide insight into the degree to which the observed trend reflects anthropogenic forcing versus internal climate variability. Based on simulations from models participating in the World Climate Research Programme (WCRP) Coupled Model Intercomparison Project Phase 3 (CMIP3), incorporating observed estimates of climate forcings, *Stroeve et al.* [2007] attributed 47% to 57% of the observed September sea ice extent trend over the period 1979-2006 to anthropogenic forcing. They also found that the observed trend exceeded nearly all of the trends simulated by the individual models and their ensemble members, suggesting that (a) the CMIP3 models, as a group, are "conservative", implying that (b) the anthropogenic component of the trend is larger than their estimate, and (c) a seasonally ice-free Arctic may be realized in only a few decades.

Conclusions that the CMIP3 models are conservative must be tempered by recognition that the comparison of observed and modeled time series is over a fairly short period of record for which trends can be strongly influenced by internal climate variability [e.g. *Stroeve et al.*, 2007; *Kattsov et al.*, 2010; *Kay et al.*, 2011]. It is hence possible that the underlying reason for the observed trend falling outside the envelope of

the CMIP3 trends is that the observed decline is a statistically rare natural event, not captured by the limited size of the CMIP3 ensemble members [*Kattsov et al.*, 2010]. Indeed, *Winton* [2011] concludes that the period of sea ice observations is too short to quantify the impact of anthropogenic forcing on the trend using model output.

Here, we use two metrics to evaluate the performance of climate models contributing to the Phase 5 (CMIP5) project. The distribution of simulated extents over the period of observations is used to assess how well the models capture the observed state of the ice cover. Trends in simulated ice extent are then used as a measure of the ability of the models to capture the response of the ice cover to global climate change. It is possible that a model can capture the historical state but not the trend. We compare results of these evaluations to those based on the CMIP3 simulations. The CMIP5 models will become the main source of climate projections assessed by the International Panel on Climate Change (IPCC) in its 5th Assessment Report.

2. Methods
2.1 Observations
We primarily rely on the passive microwave satellite data record for 1979-present, which provides consistent estimates of sea ice extent combining data from several sensors: the Nimbus-7 Scanning Multichannel Microwave Radiometer (SMMR, 1979-1987), the DMSP Special Sensor Microwave/Imager (SSM/I, 1987-2007) and the Special Sensor Microwave Imager/Sounder (SSMIS, 2008-present) [*Fetterer et al.*, 2002]. Extent is defined as all pixels with an ice concentration of at least 15%. We also compare model output with a longer time series (1953-2011) based on blending the Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST) data record [*Rayner et al.*, 2003], a merged 1972-2002 ESMR-SMMR-SSM/I time-series [*Cavalieri et al.*, 2003] and the *Fetterer et al.* [2002] 1979-2011 record that is part of the National Snow and Ice Data Center (NSIDC) Sea Ice Index (SII). While having more confidence in the shorter passive

microwave record, the longer record captures more of the internal variability of sea ice extent.

To blend the three time series, mean monthly values from the 1972-2002 ESMR-SMMR-SSM/I record were first adjusted to match the 1979-2002 portion of the SMMR-SSM/I-SSMIS SII record, based on the monthly average differences. Next, monthly means from HadISST for the period 1972-1978 were compared to means for the same period from the adjusted ESMR-SMMR-SSM/I record. The average difference for each month was then used to adjust the 1953-1978 HadISST record to fit with the adjusted 1972-1978 data. These adjustments effectively reduced sea ice extent over the period 1953-1978, particularly during summer, resulting in a 1953-2011 September trend of -0.44×10^6 km^2 decade^{-1} (or -6.3% decade^{-1}) compared to a trend of -0.67×10^6 km^2 decade^{-1} (or -9.0% decade^{-1}) when using the unadjusted HadISST record (1953-1979) in combination with NSIDC's SII record (1979-2011) as described in *Meier et al*. [2007].

2.2 Model Output

We processed 56 ensemble members from 20 climate models in the CMIP5 archive spanning both the 20th century (1850-2005), incorporating observed climate forcings, and the representative concentration pathway (RCP) 4.5 future emission scenario. RCP4.5 stabilizes radiative forcing at 4.5 W m^{-2} in the year 2100, resulting in approximately 550 ppm of CO_2 by 2100 [*Meehl et al*. in press]. Use is made of the same CMIP3 20th century and future "business as usual" (SRES A1B emission scenario) model output examined by *Stroeve et al*. [2007]. A1B attains CO_2 levels of 750 ppm by 2100 and is hence a more aggressive scenario than RCP4.5.

Time-series of simulated ice extent for the period 1850 to 2100 were compiled by splicing time-series from matching historical runs and future emission scenario runs. While such splicing allows for comparisons over the entire observational period, it may introduce effects of forcing discontinuities [*Arblaster et al*. 2011]. Homogeneity of the spliced series was visually assessed and found to be reasonable.

4

A subset of models was selected for calculating the multi-model ensemble mean and seasonal cycle. Models were selected by comparing the distribution of the simulated September extent with the observed range as computed over the period 1953-1995. Models with more than 75% of their distribution falling outside the observed range of 6.13 to 8.43×10^6 km^2 were rejected. Of the 20 CMIP5 models, 17 models were retained, resulting in a total of 38 ensemble members. The rejected models were CanESM2, CSIRO Mk3_6 and GISS-E2-R, with GISS-E2-R having a 1953-1995 mean September extent below 5.0×10^6 km^2, or at levels not observed until recently (2007-2011) and CSIRO Mk3_6 having September extents in excess of 11.0×10^6km^2. Extents from NCAR CCSM4 and CESM straddle the mean observed extent. Only three models (IPSL-CM5A-LR, MIROC5 and NorESM1-M) have extents towards the high end of the observed range and four (HadGEM2-ES, MIROC4h, MRI-CGCM3 and INMCM4) have more than 25% of extents below the observed minimum extent. Applying the same criteria to the CMIP3 models resulted in retention of 34 ensemble members from 15 models. Six were excluded (CSIRO Mk3_0, GFDL CM2-1, GISS E-R, INGV ECHAM4, INMCM3, MIROC 3-2-hires), with GISS E-R and CSIRO MK3_0 having a 1953-1995 mean extent in excess of 11.0×10^6 km^2 and the other four models having values below 5.0×10^6 km^2.

While we use a subset of models to compare the multi-model ensemble means between the CMIP3 and CMIP5 archives, a robust approach to model pre-selection has yet to be identified in the literature. For completeness, we hence show September trends for all CMIP5 models, including the rejected ones, and list multi-model ensemble means for both the subset and the full model suite in Table S1.

While focusing on September, we also examined March, but using the CMIP5 and CMIP3 models selected on the basis of September ice extent. For March, the observed 1953-1995 sea ice extent is 16.10×10^6 km^2, ranging from 15.32 to 16.60×10^6 km^2. Extents for six CMIP5 models (CCSM4, CanESM2, GFDL-CM3, IPSL-CM5A-MR,

INMCM4, NCAR CESM) fall within the maximum and minimum observed values for March. The four MIROC models, as well as NorESM1-M, have 100% of their March extents below the minimum observed value, whereas extents from CNRM-CM5, HadCM3, HadGEM2-CC, IPSL-CM5A-LR and MRI-CGCM3 fall above the maximum observed value.

3. Results
3.1 Seasonal Cycle

Figure 1 shows monthly mean sea ice extents from the CMIP5 and CMIP3 multi-model ensemble means for the period 1979 to 2011 along with the observed values. Results for 1953-2011 are similar. For each month, model data are shown as box and whisker plots, constructed from all runs of all models selected for analysis (38 ensemble members for CMIP5 and 34 ensemble members for CMIP3).

The seasonal cycle of sea ice extent is well represented in the CMIP5 ensemble. Inter-quartile ranges from the CMIP5 ensemble overlap the ranges of observed extents in all months. However, in winter, the mean and median of the CMIP5 ensemble are greater than the observed mean extent. Although the distributions of CMIP3 extents are similar to CMIP5, the mean and median extents have a positive bias with respect to both the CMIP5 models and observations. CMIP3 also has a larger spread of extents in winter months, resulting from an ensemble member with a winter extent exceeding $20.0 \times 10^6 \, \text{km}^2$ (GFDL CM2_0). Since models are selected based on their ability to capture observed extent in September, this can result in a larger spread during other months. This may in part explain why the multi-model ensemble median winter extents exceed the observed value for January through April.

During summer, the situation is different, both in regards to how the CMIP5 and CMIP3 models match each other, as well as how well they compare to the observations. From June to September, mean extents from CMIP5 are close to the observed means. Although the inter-quartile range during these months is smaller than in winter, the

6

overall spread of CMIP5 extents increases. This in part reflects low summer extents in MIROC4h. As for winter, distributions of extents from CMIP3 are biased high in summer months.

3.2 Evolution of September and March Ice Extent

The large panel in **Figure 2a** shows for September the observed sea ice extent (red line) along with the modeled CMIP5 ice extent spanning 1900 to 2100. At the start of the 20th century, the September ice extent from all ensemble members ranges from 4.4×10^6 km^2 (CanESM2) to 11.6×10^6 km^2 (CSIRO Mk3). This is similar to the range found in the full suite of CMIP3 ensemble members (47), 5.1 to 12.1×10^6 km^2. Using only those models that are able to reproduce the observed 1953-1995 September extent (based on the selection criteria discussed earlier) reduces the spread to 5.32 to 10.0×10^6 km^2 for CMIP5, and 6.2 to 10.4×10^6 km^2 for CMIP3. Based on multi-model ensemble mean extents at the beginning of the 20th century, there is a 1.1×10^6 km^2 difference in the mean September extent between CMIP3 and CMIP5 (**Figure 2a inset**).

During the period of satellite observations, the September CMIP5 multi-model ensemble mean tends to be slightly lower than the observed extent until 2007, after which it is higher. By contrast, the CMIP3 multi-model ensemble has a positive bias throughout the period of observations and especially during the most recent decade. Turning to the end of the 21st century, the CMIP5 multi-model ensemble mean never reaches ice-free conditions (defined here as less than 1.0×10^6 km^2), but the minus 1 standard deviation drops below the ice-free threshold around year 2045. Several CMIP5 models (CanESM2, GISS E2-R, GFDL-CM3, NCAR CESM, MIROC-ESM and ESM-CHEM) show essentially ice-free conditions by 2050, with the CanESM2 model having an ensemble member reaching nearly ice-free conditions as early as 2016 (0.54×10^6 km^2). By contrast, despite the more aggressive emission scenario (SRESA1B) driving the CMIP3 models, an overall more extensive sea ice cover is retained, with the minus 1 standard deviation reaching nearly ice-free conditions in 2075.

For March (**Figure 2b and inset**), CMIP3 and CMIP5 have multi-model ensemble mean winter extents at the start of the 20th century within 0.7×10^6 km^2 of each other (16.7 and 17.4×10^6 km^2 for CMIP5 and CMIP3, respectively). However, extents from the two archives begin to depart more from each other by about 1940 such that the CMIP5 multi-model ensemble mean is less than the CMIP3 mean by at least 1.0×10^6 km^2 throughout the observational time-period and beyond. The lower CMIP5 March extent compared to CMIP3 results in good overall agreement with the observations, though the observed values fall below the CMIP5 multi-model ensemble means in recent years.

3.3 Trend Analysis

While the overall lower September extent at the beginning of the 20th century in the CMIP5 models helps to explain why they match the observed extent more closely than the CMIP3 models, it is important to assess whether the observed trend falls within the distribution of simulated trends at a specified level of statistical significance. Thus we evaluate whether the CMIP5 model trends (β_m) are smaller or larger with respect to observed trends (β_o). The null hypothesis is that the trend from any given model ensemble member is consistent with the observed trend. Following *Santer et al.* [2008] the test statistic based on combining the standard error of both the model and the observations is:

$$d = (\beta_m - \beta_o) \Big/ \sqrt{s(\beta_m)^2 + s(\beta_o)^2}$$

The linear trends were estimated using the standard least-squares approach and are reported as 10^6 km^2 decade^{-1}. An effective sample size (n_{eff}) was calculated to adjust the standard error ($s(\beta_m)$ or $s(\beta_o)$) of the modeled (β_m) or observed (β_o) trend for the effects of temporal autocorrelation:

$$n_{eff} = n_{tot}(1-AR1)/(1+AR1)$$

where n_{tot} is the number of years over which the trend is evaluated and AR1 is the lag-1 temporal autocorrelation coefficient. Autocorrelation coefficients for regression residuals

of model time series are between 0.1 and 0.8, whereas for the observed September extent the autocorrelation coefficient is 0.4 from 1953-2011 and 0.1 from 1979-2011. **Figure 3** shows September trends from 1953-2011 (a) and 1979-2011 (b), together with their adjusted 2σ confidence interval, and the multi-model ensemble mean. The 1σ and 2σ observed trends are shown in dark gray shading (1σ) and light gray shading (2σ). Table S1 lists the mean trends for all CMIP5 models examined along with the range for models with more than one ensemble member.

From 1953 to 2011, the observed rate of decline is -0.44×10^6 km^2 decade^{-1} and is statistically different from zero. Rates of decline for most models are slower than observed, though trends vary considerably, not only between models but also between ensemble members for a given model. Thirty out of 56 ensemble trends fall outside the 2σ bound for the observed trend, and 15 of those also have their 2σ confidence interval outside 2σ of the observed trend. Three-quarters of the ensemble members (including all 9 HadCM3 ensembles) have trends outside 1σ of the observed trend, and in some cases the 2σ confidence interval of these ensembles is also outside 1σ of the observed trend (24 ensemble members). Only two ensemble members (from GISS-ER-2 and CNRM-CM5) have a mean trend larger than 1σ of the observed trend. Some model time series exhibit significant autocorrelation, reflected by the large 2σ error bars. This in part explains why 9 ensemble members have trends that are not statistically different from zero at the 90% confidence level.

The overlap of the modeled trends $\pm\, 2\sigma$ with the observed trend $\pm\, 2\sigma$ provides a measure of the level of confidence that the trends are from the same distribution, which is essentially what the *d*-statistic provides. Calculated *p*-values for the *d* statistic were compared for *p*-values of 0.10. Table S1 lists the percentage of ensemble members for which we can reject the null hypothesis that the model trends differ from the observed trend. For the 1953-2011 time-period, the null hypothesis is rejected for 50% of the

9

CMIP5 ensemble members. In the CMIP3 archive, the corresponding rejection rate after similarly accounting for temporal autocorrelation is 79%. The CMIP5 multi-model ensemble mean trend is -0.27 km^2 decade^{-1}, which is less than a third of the observed trend, and slightly larger than the CMIP3 value of -0.22 km^2 decade^{-1}.

Turning to the modern satellite era, 1979-2011, more CMIP5 ensembles have a smaller rate of decline than observed, which at -0.84x10^6 km^2 decade^{-1} is nearly twice as large as the trend for 1953-2011. Forty-six of 56 ensemble members have trends outside of the 2σ bound for the observations and 9 ensemble members have trends that are not statistically different from zero at the 90% confidence level. Although most model trends remain slower than observed, six ensemble members have rates of decline larger than observed. Overall, 64% of the ensemble member trends are statistically different from the observed trend at the 90% confidence level. In contrast, 85% of the CMIP3 ensemble members have trends that are statistically different from observed. The multi-model ensemble mean trend over the satellite period is -0.50 km^2 decade^{-1}, which is 70% larger than the CMIP3 multi-model mean value of -0.35 km^2 decade^{-1}.

4. Discussion and Conclusions

Is the observed evolution of the Arctic sea ice cover better simulated in the newer CMIP5 models? While there remains considerable scatter in modeled sea ice extents, compared to CMIP3, the CMIP5 multi-model ensemble mean is more consistent with the historical observations, both in terms of the mean observed climatological state of the ice cover and the rate of decline during the past several decades. In this regard, one can argue the CMIP5 models are "less conservative" than the earlier CMIP3 models. However, does this imply increased confidence in projected conditions through the 21st century and the timing of a seasonally ice-free Arctic Ocean? If we constrain the analysis to models that (based on criteria described earlier) can reproduce the observed climatological mean September ice extent, the spread in projected ice extent throughout the 21st century from the CMIP5 and CMIP3 models is similar. This holds despite the overall better match of

the CMIP5 models with the observed seasonal cycle and trend. Furthermore, a similar number of these ensemble members reach nearly ice-free conditions (less than 1.0×10^6 km^2) before the end of the 21st century. Viewed in this context, conclusions drawn from CMIP5 are not much different than those from CMIP3.

Conclusions based on climate models are only as reliable as their ability to capture the important underlying processes through their physical and numerical formulations. *Stroeve et al.* [2007] showed that models with more sophisticated sea ice processes (such as those incorporating the Los Alamos sea ice model, CICE; *Hunke and Lipscomb*, 2008) were better able to represent the present state of the sea ice cover. While detailed documentation of the sea-ice components from many of the CMIP5 models is not readily available, it is likely that in some cases, model improvements, such as new sea-ice albedo parameterization schemes that allow for melt ponds [e.g. *Pedersen et al*, 2009; *Holland et al., 2012*] have led to better representation of the historical ice conditions (with respect to the CMIP3 models originating from the same centers). Additionally it is possible that more attention has been given to tuning sea ice simulations in present-day models, through parameterization optimization efforts [e.g. *Miller et al.*, 2006; *Uotila et al.*, 2012] and other means.

While we focus on representation of ice extent, other features of the models, such as their ice thickness distributions, rates of ice transport and ice melt/growth, are important. *Holland et al.* [2010] showed that models with initially thicker ice generally retain more extensive ice throughout the 21st century despite larger increases in net ice melt. Evaluation of thickness fields from CMIP5 (not shown) indicate that part of the explanation for the better representation of the observed September ice extent is that several of these models start the 20th century with rather thin winter ice cover, even though the winter extent is similar to that observed. For example, CanESM2 starts with only a 2 m winter ice cover as averaged over the Arctic Ocean so that although the winter

extent is consistent with observations, summer extent is significantly underestimated. As a next step, we will compare early 21st century modeled ice thickness distributions with thickness distributions based on NASA ICESat and Operation IceBridge missions.

Other factors influencing model representation of historical sea ice changes include errors in the specified values of the external forcings, inability of the models to properly respond to the imposed external forcings and/or reproduce natural climate variability, discontinuities in the observational record and an insufficient number of ensemble members for sampling the natural variability. Because of natural variability, the timing of rapid ice loss (and temporary recovery) will vary between different models and their ensemble members. Another issue that needs to be addressed is the effects of interpolating sea ice concentration from native ocean model grids to the grids the data are archived on. While the CMIP5 output is provided on native model grids, the CMIP3 data have been interpolated.

Kay et al. [2011] concluded that approximately half (56%) of the observed rate of decline from 1979 to 2005 was externally forced based on 6 ensemble members from CCSM4. The use of a multi-model ensemble mean provides another means to quantify the expected value of sea ice changes due to external forcing alone, with the range of ensemble members as a measure of spread related to internal variability. Based on the CMIP5 multi-model ensemble mean, approximately 60% of the observed rate of decline from 1979-2011 is externally forced (compared to 41% from CMIP3, based on about the same number of ensemble members), ranging from 52% to 67% for +/- one standard deviation from the ensemble mean. However this value is based on the assumption that the model simulations are members from the same distribution and that we have enough members to resolve the distribution. This assumption does not exactly hold since some models better represent the climate system than others.

Looking to the future, 32% of the 56 CMIP5 ensemble members evaluated for the RCP4.5 emission scenario reach nearly ice free conditions (less than $1.0 \times 10^6 \, km^2$) by the end of this century, with some showing a nearly ice-free state as early as 2020. However, we must acknowledge the large uncertainty in future greenhouse gas emissions. While on time-scales of decades the sea ice projections are relatively insensitive to the prescribed greenhouse gas forcing, this is not the case by the end of the century. Thus, the range of available results does not necessarily represent the "true" uncertainty as to when a seasonally ice-free Arctic may be realized.

While quantification of the role of external forcing depends on many assumptions, it is nevertheless becoming increasingly clear in both the observations [e.g. *Notz and Marotzke*, 2012] and model studies [*Stroeve et al.*, 2007; *Kay et al.*, 2011; *Min et al.*, 2008; *Wang and Overland*, 2009] that if greenhouse gas concentrations continue to rise, the Arctic Ocean will eventually become seasonally ice-free. However, results from the CMIP5 models do not appear to have appreciably reduced uncertainty as to when a seasonally ice-free Arctic Ocean will be realized. Nevertheless, CMIP5 arrives at a seasonally ice-free Arctic sooner than CMIP3, leading to the conclusion that a seasonally ice-free Arctic Ocean within the next few decades is a distinct possibility.

Acknowledgements

Earth System Science Portals. The Editor thanks 2 anonymous reviewers for their assistance in evaluating this paper.

References

Arblaster, J.M., G.A. Meehl, and D.J. Karoly (2011), Future climate change in the Southern Hemisphere: Competing effects of ozone and greenhouse gases, *Geophys. Res. Lett*, 38, L02701, doi:10.1029/2010GL045384.

Cavalieri, D. J., C. L. Parkinson, and K. Y. Vinnikov (2003), 30-year satellite record reveals contrasting Arctic and Antarctic decadal sea ice variability, *Geophys. Res. Lett.* 30, doi: 10.1029/2003GL018031.

Fetterer, F., K. Knowles, W. Meier, and M. Savoie (2002), updated 2009. *Sea Ice Index*, Boulder, CO: National Snow and Ice Data Center. Digital media.

Francis, J.A. and S.J. Vavrus (2012), Evidence linking Arctic amplification to extreme weather in mid-latitudes, *Geophys. Res. Lett.*, doi:101029/2012GL051000.

Holland, M., D. Bailey, B. Briegleb, B. Light, and E. Hunke (2012), Improved sea ice shortwave radiation physics in CCSM4: The impact of melt ponds and aerosols on arctic aea ice, *J. Climate*, 25, 1413-1430.

Holland, M.M., M.C. Serreze, and J. Stroeve (2010), The sea ice mass budget of the Arctic and its future change as simulated by coupled climate models, *Climate Dynamics*, doi:10.1007/s00382-008-0493-4.

Hunke, E. C., and W. H. Lipscomb (2008), *CICE: The Los Alamos Sea Ice Model. Documentation and Software User's Manual Version 4.0*, T-3 Fluid Dynamics Group, Los Alamos National Laboratory, Tech. Rep. LA-CC-06-012.

Kattsov, V., V. Ryabinin, J. Overland, M. Serreze, M. Visbeck, J. Walsh, W. Meier, and X. Zhang (2010), Arctic sea ice change: a grand challenge of climate science, *J. Glaciol.*, 56(200), 1115-1121.

Kay, J.E., M.M. Holland, and A. Jahn (2011), Inter-annual to multi-decadal Arctic sea ice extent trends in a warming world, *Geophys. Res. Lett.*, L15708, doi:10.1029/2011GL048008.

Kwok, R. (2007), Near zero replenishment of the Arctic multiyear sea ice cover at the end of 2005 summer, *Geophys. Res Lett.*, 34, L05501, doi:10.1029/2006GL028737.

Maslanik, J.A., C. Fowler, J. Stroeve, S. Drobot, J. Zwally, D. Yi, and W. Emery (2007), A younger, thinner Arctic ice cover: Increased potential for rapid, extensive sea-ice loss, *Geophys. Res. Lett.*, 34, L24501, doi:10.1029/2007GL032043.

Meehl, G.A., W.M. Washington, J.M. Arblaster, A. Hu, H. Teng, C. Tebaldi, W.G. Strang, and J.B. White III (2012), Climate system response to external forcing and climate change projections in CCSM4, *J. Climate CCSM4 Special Collection*, in press.

Meier, W.N., J. Stroeve, and F. Fetterer (2007), Whither Arctic sea ice? A clear signal of decline regionally, seasonally and extending beyond the satellite record, *Ann. Glaciol.*, 46, 428-434.

Miller, R.L., G.A. Schmidt, and D.T. Shindell, (2006), Forced annular variations in the 20th century Intergovernmental Panel on Climate Change Fourth Assessment Report models. *J. Geophys. Res.*, 111, D18101, doi:10.1029/2005JD006323.

Min, S-K., X. Zhang, F.W. Zwiers, and T. Agnew (2008), Human influence on Arctic sea ice detectable from early 1990s onwards, *Geophys. Res. Lett.*, 35, L21701, doi:10.1029/2008GL035725.

Notz, D., and J. Marotzke (2012), Observations reveal external driver for Arctic sea-ice retreat, *Geophys. Res. Lett.,* 39, L08502, doi:10.1029/2012GL051094.

Pedersen, C.A., E. Roeckner, M. Lüthje, and J.-G. Winther (2009), A new sea ice albedo scheme including melt ponds for ECHAM5 general circulation model, *J. Geophys. Res.*, 114, D08101, doi:10.1029/2008JD010440.
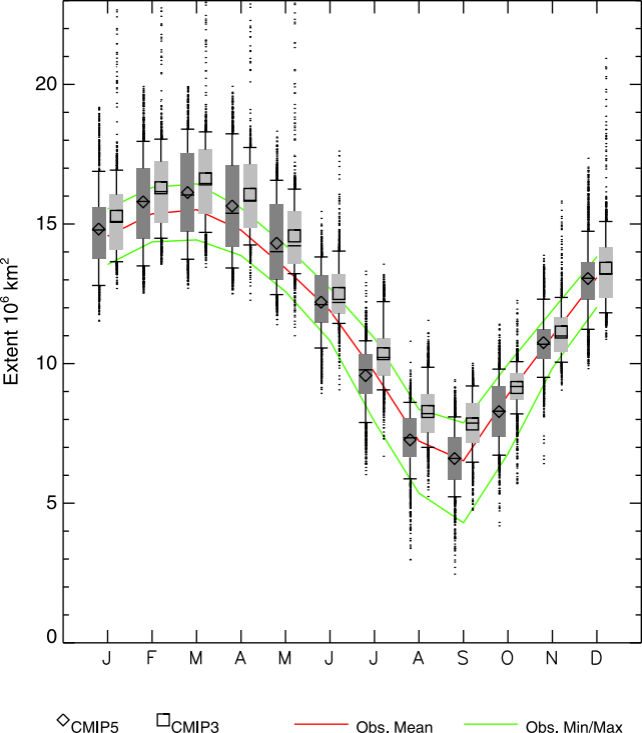
Rayner, N.A., D.E. Parkler, E.B. Horton, C.K. Folland, L.V. Alexander, D.P. Rowell, E.C. Kent, and A. Kaplan (2003), Global analyses of sea surface temperature, sea ice, and night marine air temperatures since the late nineteenth century, *J. Geophy. Res.*, 108(D14), 4407, doi:10.1029/2002JD002670.

Santer, B.D., P.W. Thorne, L. Haimberger, K.E. Taylor, T.M.L. Wigley, J.R. Lanzante, S. Solomon, M. Free, P.J. Gleckler, P.D. Jones, T.R. Karl, S.A. Klein, C. Mears, D. Nychka, G.A. Schmidt, S.C. Sherwood, and F.J. Wentz (2008), Consistency of modelled and observed temperature trends in the tropical troposphere, *Intl. J. Climatol.*, 28, 1703-1722, doi:10.1002/joc.1756.

Stroeve, J.C., M.C. Serreze, J.E. Kay, M.M. Holland, W.N. Meier, and A.P. Barrett, (2011), The Arctic's rapidly shrinking sea ice cover: A research synthesis, *Climatic Change*, doi: 10.1007/s10584-011-0101-1.

Stroeve, J., M. M. Holland, W. Meier, T. Scambos and M. Serreze, (2007), Arctic sea ice decline: Faster than forecast, *Geophys. Res. Lett.*, 34, doi: 10.1029/2007GL029703.

Uotila, P., S. O'Farrell, S.J. Marsland, and D. Bi (2012), A sea-ice sensitivity study with a global ocean-ice model, *Ocean Modelling*, 51, 1-18.

Wang, M. and J.E. Overland (2009), A sea ice free summer Arctic within 30 years?, *Geophys. Res. Lett.*, 36, L07502, doi:10.1029/2009GL037820.

Winton, Michael (2011), Do climate models underestimate the sensitivity of Northern Hemisphere sea ice cover?, *J. Climate*, 24, 3924–3934, doi: http://dx.doi.org/10.1175/2011JCLI4146.1

Figure 1. Comparison of the observed seasonal cycle with the multi-model ensemble mean seasonal cycle from CMIP5 and CMIP3 from 1979 to 2011. For each month, the model data are shown as box and whisker plots. Boxes represent inter-quartile ranges (25th to 75th percentiles). Median (50th percentile) extents are shown by the thick horizontal bar in each box. Whiskers (vertical lines and thin horizontal bars) represent

the 10th and 90th percentiles. Extents less than the 10th percentile or greater than the 90th percentile are shown as individual dots. 1979-2011 mean monthly extents are shown as diamonds and squares for CMIP5 and CMIP3 respectively. Corresponding mean, and minimum and maximum observed extents are shown as red and green lines respectively.

Figure 2. Time-series of modeled (colored lines) and observed (solid red line) September (a) and March (b) sea ice extent from 1900 to 2100. All 56 individual ensemble members from 20 CMIP5 models are included as dotted colored lines, with their individual model ensemble means in solid color lines. The multi-model ensemble mean is based on 38 ensemble members from 17 CMIP5 models (shown in black), with +/- 1 standard deviation shown as dotted black lines. Figure 2a and 2b insets are based on the multi-model ensemble mean from CMIP5 and CMIP3, +/- 1 standard deviation.

Figure 3. September trends from 1953-2011 (top) and 1979-2011 (bottom) for all individual model ensembles as well as the multi-model ensemble mean with confidence intervals (vertical lines). The 1σ and 2σ observed trends are shown in dark gray shading (1σ) and light gray shading (2σ).

(a) **September Ice Extent (RCP4.5)**

(b) **March Ice Extent (RCP4.5)**
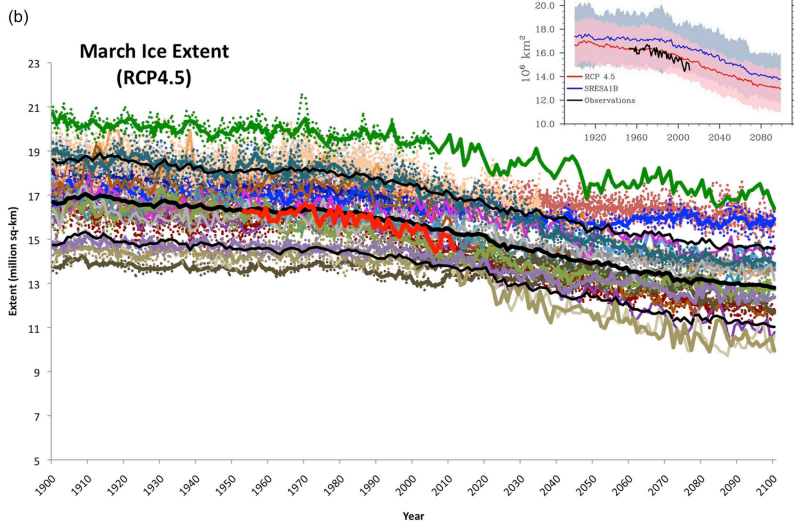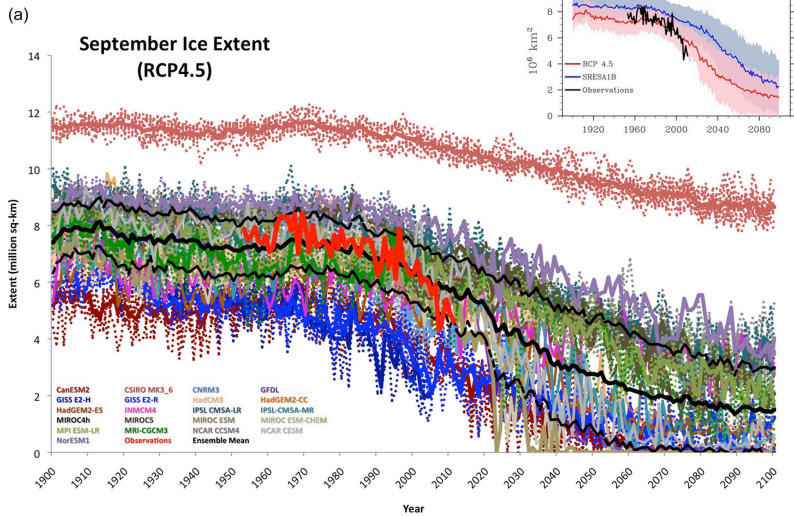
(a)

1953 to 2011

(b)

1979 to 2011

Legend:
- CanESM2
- CESM
- CCSM4
- CNRM-CM5
- CSIRO-Mk3-6-0
- GFDL-CM3
- Observed +/- 2CI
- GISS-E2-R
- HadCM3
- HadGEM2-CC
- HadGEM2-ES
- inmcm4
- IPSL-CM5A-LR
- Observed +/- 1CI
- IPSL-CM5A-MR
- MIROC4h
- MIROC5
- MIROC-ESM-CHEM
- MIROC-ESM
- MPI-ESM-LR
- Multi-model mean

Axis labels: 10^6 km^2 / decade, Ensemble