Using Records from Submarine, Aircraft and Satellites to Evaluate Climate Model Simulations of Arctic Sea Ice Thickness

Julienne Stroeve[1], Andrew Barrett[1], Mark Serreze[1], Axel Schweiger[2]

[1]National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO USA
[2] Polar Science Center, Applied Physics Laboratory, University of Washington, Seattle, WA, USA

Abstract

Arctic sea ice thickness distributions from models participating in the World Climate Research Programme Coupled Model Intercomparison Project Phase 5 are evaluated against observations from submarines, aircraft and satellites. While it's encouraging that the mean thickness distributions from the models are in general agreement with observations, the spatial patterns of sea ice thickness are poorly represented in most models. The poor spatial representation of thickness patterns is associated with a failure of models to represent details of the mean atmospheric circulation pattern that governs the transport and spatial distribution of sea ice. The climate models as a whole also tend to underestimate the rate of ice volume loss from 1979 to 2013, though the multi-model ensemble mean trend remains within the uncertainty of that from the Pan-Arctic Ice Ocean Modeling and Assimilation System. These results raise concerns regarding the ability of CMIP5 models to realistically represent the processes driving the decline of Arctic sea ice and to project the timing of when a seasonally ice-free Arctic may be realized.

1. Introduction

The last four decades have seen a remarkable decline in the spatial extent of Arctic sea ice at the end of the melt season. Based on sea ice concentrations from the National Snow and Ice Data Center (NSIDC) Sea Ice Index [*Fetterer et al*., 2002), the linear trend for September, as calculated over the 1979 through 2013 period, stands at -14.0% dec^{-1}, or -895,300 km^2 dec^{-1}. The downward trend has been linked to a combination of natural climate variability and warming that is a response to increasing concentrations of atmospheric greenhouse gases [e.g. *Notz and Marotzke*, 2012; *Stroeve et al., 2012a*]. Extent recorded for September 2012 (the record low in the satellite era) was only 50% of values recorded in the late 1970s to early 1980s. Volume losses are even greater showing 80% decline in between September 1979 and 2012 according to the Pan-Arctic Ice Ocean Assimilation System (PIOMAS). While September ice extent rebounded in 2013, partly a result of anomalously cool summer conditions [e.g. *Stroeve et al*., 2014], it was still the 6th lowest in the satellite record.

Coupled global climate models (GCMs) consistently project that if greenhouse gas concentrations continue to rise, the eventual outcome will be a complete loss of the multiyear ice cover, that is, sea ice will become a seasonal feature of the Arctic Ocean [e.g. *Stroeve et al*., 2007; 2012b], presenting both challenges and opportunities to Arctic residents, government agencies and industry. While GCMs can provide useful projections of when a seasonally ice-free

Arctic Ocean may be realized, confidence in these projections depends on their ability to reproduce features of the present-day climate. *Stroeve et al*. [2012b] found that models participating in the World Climate Research Programme Coupled Model Intercomparison Project Phase 5 (CMIP5) are more consistent with observations than those from the previous CMIP3 effort, with 67% of the models (or 16 out of 24) having a 1953-1995 mean September ice extent falling within the minimum and maximum bounds of observed values. However, historical trends from 85% of the model ensemble members examined remain smaller than observed, and the spread in simulated extent between different models remains large.

Realistically simulating the past and future evolution of the Arctic's floating sea ice cover is one of the most challenging facets of climate modeling. Simulating the sea ice thickness distribution has emerged as a key issue. While it follows that climate models with an overly thick initial (early 21st century) ice cover will tend to lose their summer ice later than models with initially thinner ice given the same climate forcing [e.g. *Holland et al.* 2010], the ice thickness distribution strongly determines surface heat fluxes, impacting on both the ice mass budget and ice loss rate, which is in turn a major driver of Arctic amplification - the outsized rise in lower-tropospheric air temperatures over the Arctic Ocean compared to lower latitudes [*Serreze et al.,* 2009].

A major difficulty in evaluating thickness distributions in GCMs is the lack of consistent observations spanning a sufficiently long time period. It was not until 2003 that temporally-limited (autumn and spring) near-Arctic-wide estimates of thickness became available from NASA's Ice, Cloud, and land Elevation Satellite (ICESat) Geoscience Laser Altimeter System (GLAS). Prior to ICESat, information was largely limited to data from upward looking sonars on board British and U.S. submarines collected during the 1980s and 1990s, mainly covering the region near the pole as well as several moorings providing time series in fixed locations [*Lindsay*, 2010]. The first European Remote Sensing satellite (ERS-1) included a radar altimeter that provided fields of estimated sea ice thickness up to latitude 81.5^{o}N, but only for the 1993 to 2001 period *[Laxon et al*., 2003]. Since the failure of ICESat in 2009, additional sea ice thickness measurements have become available from airborne flights as part of NASA's Operation IceBridge program. Arctic-wide coverage has since resumed, starting in 2010 from the radar altimeter on-board the European Space Agency's CryoSat-2. Together, these data provide a valuable source of information for the validation of spatial patterns of sea ice thickness. In addition, satellite and in-situ observations have been used to provide validation of sea ice reanalysis systems such as PIOMAS, which in turn may provide a consistent record of thickness and volume for comparison with climate model long-term trends [*Schweiger et al.,* 2011].

This paper examines biases in contemporary Arctic sea ice thickness and ice volume from the CMIP5 models making use of all of these data sets. Model thicknesses are evaluated for the whole of the Arctic Ocean and on a regional basis depending on data coverage. Since radar measurements are influenced by snowmelt, and IceBridge data are only available in March, we focus on spring (e.g. March) estimates of ice thickness. Modeled ice volume spanning the 1979 to 2013 period is further evaluated against volume estimates simulated from PIOMAS [*Zhang and Rothrock*, 2003] for the months of March and September.

2. Methodology

2.1 Evaluation framework

We evaluate models using three criteria: 1) how well they replicate the statistical distribution of observed mean sea ice thickness fields based on aggregating all available data across the Arctic for each observational data set; 2) how well they replicate the observed spatial pattern of sea ice thickness; and 3) how well they replicate the best estimate of trends in sea ice volume. The first two evaluations make use of the thickness records from in-situ moorings, and submarine, aircraft- and satellite-borne instruments introduced in the previous section. This record is not sufficiently homogeneous to evaluate thickness or volume trends, which is why we also make use of the PIOMAS record. PIOMAS assimilates sea ice concentration, sea surface temperature and ice velocity. While PIOMAS is a model and sensitive to the atmospheric reanalysis used, estimates of thickness compare well with in-situ observations, submarines, airborne measurements, and from satellites [*Zhang and Rothrock*, 2003; *Schweiger et al.*, 2011; *Lindsay et al.,* 2012; *Laxon et al.,* 2013].

A further difficulty in our model evaluation, amplified by the piecemeal nature of the ice thickness record, is that individual years in CMIP5 model time do not correspond with the same years in the observational record. Imprints of intrinsic natural climate variability in the observational record (such as that associated with the phase of the North Atlantic Oscillation) will likely be out of phase with natural variability in the model simulations. Thus, discrepancies in modeled ice thickness can either be due to model biases or natural climate variability. Ideally, climatologies of modeled sea ice thickness need to be compared with observed climatologies that are of similar length and long enough (e.g., 30 years) to average out most of the natural variability.

Monthly mean fields of sea ice thickness for 92 ensemble members of 33 climate models from the CMIP5 archive were downloaded from the Earth System Grid of the Program for Climate Model Diagnosis and Intercomparison data portal (PCMDI) (http://cmip-pcmdi.llnl.gov/cmip5/). The archive consists of both atmosphere-ocean global climate models (AOGCMs) and Earth System Models (ESMs), the latter which incorporate interactive biogeochemical cycles into AOGCMs. Both the historical (1850-2005) and future Representative Concentration Pathway (RCP) 4.5 (2006-2100) emission scenarios were processed and the same number of ensembles for both emission scenarios were used. RCP4.5 is a medium-mitigation scenario that stabilizes CO_2 at ~650 ppm at the end of the century [e.g. *Thompson et al*., 2011], corresponding to a radiative forcing of 4.5 Wm^{-2} by 2100. It is perhaps a conservative scenario given current emission rates. A listing of the models used can be found in Table 2.

Monthly mean thickness fields for the 1981 to 2010 period were calculated for every ensemble member. For models having more than one ensemble member, mean thickness fields from each ensemble for a given model were averaged to form a single ensemble average. Spatial resolutions vary considerably from high-resolution ocean modelling grids to coarse grids with a roughly 1 degree-by-1-degree spacing. To enable comparisons between models and the observations, mean thickness fields were regridded to the 100 km Equal Area Scaleable Earth (EASE) grid [*Brodzik and Knowles*, 2002] using a drop-in-the-bucket approach. The 100 km resolution corresponds to resolution of the coarser model grids.

To compare aggregate mean thickness (evaluation criterion 1), frequency distributions were derived for each model using the regridded mean fields. Separate distributions were produced for each observed thickness field so that model thicknesses could be extracted corresponding to the coverage of each of the observed thickness data sets. For example, only grid cells with thicknesses from both IceBridge and the model were used when evaluating how well the models represent the aggregate thickness distribution during the IceBridge time-period. Regridded model fields were also used to evaluate spatial thickness patterns (criterion 2). To ensure that model ensemble members can be used for validation of spatial patterns, it is important to first assess the natural variability of the sea ice thickness spatial patterns within the models. For models with five or more ensemble members, we evaluated the variability in spatial patterns and Arctic-wide mean thickness from 1981 to 2010 [**Figure 1**]. As expected, higher variability is the rule over the North Atlantic near the sea ice margin. Three of the models (CCSM4, EC-EARTH and HadCM3) stand out because of high local variability, such as in the Beaufort Sea sector in CCSM4. Two of these models (CCSM4 and EC-EARTH) incorporate an ice-thickness distribution (ITD) framework. It could be that models that resolve the statistical sub-grid scale distribution of ice thickness produce grid-cell thicknesses more strongly influenced by natural variability than models without ITD. However, for the models evaluated, variability is less than 8% of the mean over the Arctic Ocean as a whole. In addition, spatial pattern correlations between individual ensembles within a model are above 0.9 (and mostly above 0.98) (not shown). This suggests that the fragmented observational record offers an opportunity to compare characteristics of the thickness patterns, which are less impacted by natural variability.

To evaluate criterion 3 (trends in ice volume using PIOMAS records), March ice volume was calculated for each model ensemble member corresponding to the domain of the PIOMAS estimates. Unlike thickness, ice volume was calculated on the native model grid. Ice thickness in the CMIP5 archive is given as the grid cell mean including ice-free portions of the grid cell. Grid-cell ice volume is simply the product of the mean grid-cell thickness and grid-cell area. Grid cell volumes were summed for the PIOMAS domain, to give a time series of monthly mean ice volume.

2.2 Data: Observations

As previously introduced, the observed record of sea ice thickness is based on a combination of in-situ, submarine, aircraft and satellite data. Although records are available from 1975 through the present, no one data source is spatially or temporally continuous over the whole of this period, making the construction of a homogenous time series from observations alone impossible. To provide a long-term picture, estimates of ice thickness from different sources must be combined. We provide gridded fields at two resolutions on the EASE grid (25- and 100-km) that facilitate comparisons with both PIOMAS (distributed at 25-km spatial resolution) and the CMIP5 mean thickness fields (100-km resolution).

Unclassified sonar data from U.S. Navy and U.K. Royal Navy submarine missions provide the earliest estimates, starting in 1975 and ending in 1993. Ice thickness estimates from submarines and other platforms have been collated and processed into a consistent format by R. Lindsay at the University of Washington Polar Science Center to produce the Unified Sea Ice Thickness Climate Data Record (CDR) [*Lindsay,* 2010]. The most recent version of the submarine data was obtained from the University of Washington, Polar Science Center. An archive version of the CDR, which is updated annually, is also hosted by NSIDC [*Lindsay*, 2013]. Submarine sonars provide measurements of ice draft (the depth of ice below sea level). *Rothrock and Wenshahan* [2007] document the conversion of ice draft into thickness. Briefly, ice thickness is derived from draft estimates using Archimedes principle with assumed ice, snow and water densities, and the depth of snow on the ice. In most cases, snow depth is unknown and the Warren snow climatology [*Warren et al*., 1998] is used. *Rothrock and Wenshahan* [2007] estimate an average thickness bias from the sonar data compared to direct observations of 0.29 m. We subtracted this bias from the submarine data set prior to comparison with the CMIP5 model output. Following *Schweiger et al.,* [2011], we only use data from US cruises because the processing history for UK cruise data is uncertain. Submarine cruises are designated as spring or summer. We use spring cruises, defined as occurring between March and June. Most cruises provide data for the central Arctic Ocean, away from the shallow continental shelves.

Upward Looking Sonar (ULS) instruments on bottom-anchored moorings in the Eastern Beaufort Sea, Beaufort Gyre and Chukchi Sea provide further estimates of ice thickness. Moorings in the Eastern Beaufort Sea and Chukchi Sea are maintained by the Institute of Ocean Sciences [*Melling and Riedel*, 2008]. Data records start in 1990 and end in 2005. Moorings in the Beaufort Gyre region are maintained and data made available by the Beaufort Gyre Exploration Project based at the Woods Hole Oceanographic Institution (http://www.whoi.edu/beaufortgyre). ULS on moorings also measure ice draft. The most recent versions of these in-situ ice draft estimates were also obtained from the Polar Science Center. Thickness was calculated from in-situ ice drafts using the same method as applied to the submarine data.

Unlike submarine sonar, satellite and aircraft radar and laser altimeters measure the height of bare-ice, snow-covered ice and snow surfaces above the ocean surface, depending on instrument characteristics and surface conditions. By identifying leads between the ice floes, the freeboard (the height of the snow or ice surfaces above sea level) can be derived. Ice freeboard is converted to ice thickness using Archimedes principle in a similar way as the conversion of submarine ice draft to ice thickness, using estimates or assumptions of snow and ice density and snow depth.

Laxon et al. [2003] retrieved ice thickness from the 13.8 GHz radar altimeter onboard the ERS-1 satellite and assessed changes in Arctic sea ice thickness from 1993 to 2001 up to latitude 81.5^{o}N. The winter sea ice area covered by ERS-1 is about 3.08 10^{6} km^{2} and includes the Beaufort, Chukchi, East Siberian, Kara, Laptev, Barents and Greenland seas. ERS-1-derived ice thickness is provided as a single mean field averaged from 1993 to 2001 for the month of March on a 0.1^{o} latitude by 0.5^{o} longitude grid.

ICESat, with its laser altimeter, provided the first thickness data set to cover almost the entire Arctic Ocean. Thicknesses are derived based on the methodology described by *Kwok et al*. [2009]. The ICESat archive provides five years (2004-2009) of gridded fields at 25 km resolution. Estimates of thickness extend up to 86^{o}N. *Kwok et al*. [2009] estimate an uncertainty of 0.5 m for each 25 km grid cell. Operation IceBridge is an ongoing airborne laser altimeter mission aimed at bridging the gap between ICESat and the follow-on ICESat-2 scheduled to launch in 2017. IceBridge provides individual tracks of ice thickness, generally confined to the western Arctic Ocean during March and April from 2009 to present [*Kurtz et al.*, 2012a]. Coverage is sparse in the early years of the program but subsequently improves. Each IceBridge track gives ice thickness estimates at 40 m spacing. Thickness retrievals are detailed by *Kurtz et al*. [2012b]. Finally, CryoSat-2 thickness estimates are derived using a satellite radar altimeter with coverage extending up to 88^{o}N. We use the preliminary thickness product produced by the Alfred Wegner Institute (www.meereisportal.de/cryosat). Data are available for 2011 through 2013 on the EASE-2 25-km grid [*Brodzik et al*, 2012].

Ice thickness is also measured using a combination of airborne electromagnetic (EM) induction instruments and laser altimeter [*Haas et al*, 2009]. The instrument package is flown above the sea ice surface by helicopter. The EM instrument is used to detect the distance between the instrument and ice-water interface. The laser altimeter provides the height of the snow or ice surface. The difference between the two measurements provides the combined snow-ice thickness. Ice thickness can be obtained using information about snow thickness and density. EM derived ice thicknesses are available for the central and western Arctic Ocean between 2002 and 2012. These data are also included in the Unified Sea Ice Thickness CDR and were obtained from the Polar Science Center.

All satellite-derived ice thickness fields were regridded as needed from their original gridded format to 25-km and 100-km EASE grids using a drop-in-the-bucket averaging. This provides a mean 1993-2001 thickness field from ERS-1, a yearly field for each of the five ICESat years (spring 2004 to 2009) and each of the three CryoSat years (2011 to 2013). Period-of-record mean fields from ICESat and CryoSat were additionally calculated, by first averaging on their native grids and then regridding to 25- and100-km resolution.

The in-situ mooring data, Airborne EM, IceBridge and submarine sonar track data needed to be handled differently. For comparison with CMIP5, all observed thickness estimates within 70 km of a 100 km EASE grid box center were averaged to give a grid cell mean thickness. To provide the best coverage to compare with modeled thickness distributions, all thickness estimates for all years were used to calculate a single average field for the period of record. Grids of IceBridge and submarine data at 25-km spatial resolution were additionally produced for individual years by combining multiple flight lines and cruise tracks in a single year. Since the time-periods of coverage vary, composites of ice thickness from IceBridge and submarine data are based on a range of times during the observational intervals and do not exactly correspond to monthly averages. This will introduce a temporal sampling error when making comparisons between the observations from these data sets and the monthly CMIP5 model and PIOMAS output.

Along with temporal sampling problems, the various thickness records have a range of biases due to differences in sensor types and retrieval approaches. Radar and laser technologies use different wavelengths and footprints, and different techniques have been used to estimate snow depth and snow and ice density, which in turn impacts ice thickness retrievals. This creates additional challenges as differences in snow and ice density and snow depth values used can lead to large biases in ice thickness [e.g. *Zygmuntowska et al*., 2014]. For example, for multiyear ice, *Kwok et al.* [2009] use a density of 925 kg m^{-3} while and *Laxon et al.* [2013] use 882 kg m^{-3}. According to *Kurtz et al*. [2014], this could lead to a thickness difference of 1.1m for a typical multiyear ice floe of 60 cm snow-ice freeboard with a 35 cm deep snow cover. Similarly, given an ICESat freeboard of 0.325 m with an estimated 0.25 m of snow (density 300 kg m^{-3}) atop the ice (density of 900 kg m^{-3}), we would compute a sea ice thickness of 1.5 m. Yet if there had been only 0.15 m of snow, the ice would be 2.2 m thick, a change of 0.70 m or 46% of the original estimate.

At present, there is no long-term sea ice thickness data set that applies these parameters in a consistent manner regardless of which instrument is used. It is nevertheless encouraging that all of the records show similar spatial patterns of ice thickness [**Figure 2: left column**], which while lending confidence to the data, also demonstrates persistence of the general spatial pattern of Arctic sea ice thickness from 1979 to present. Mean thicknesses are greater along the northern coasts of the Canadian Arctic Archipelago and Greenland where there is an onshore component of ice motion resulting in strong ridging. Mean thicknesses are lower on the Eurasian side of the Arctic Ocean where there is a persistent offshore ice motion and ice divergence, leading to new ice growth in open water areas. When viewed for the Arctic as a whole, the combined records show a decline through time in ice thickness, although this must be tempered by differences in physical assumptions used to retrieve thickness [*Zygmuntowska et al*., 2014].

2.3 PIOMAS Ice Thickness Patterns and Volume

Since there is not a long-term consistent ice thickness data set with which to evaluate ice volume trends, we assess CMIP5 volume trends from 1979 to 2013 against estimates from PIOMAS [*Zhang and Rothrock*, 2003]. PIOMAS assimilates observed sea ice concentrations, ice motion and sea surface temperatures into a numerical model to estimate ice volume on a continuous basis. The model is forced at the surface by data from the National Centers for Environmental Prediction (NCEP) atmospheric reanalysis.

Schweiger et al. [2011] found that PIOMAS ice thickness estimates agree well with those from ICESat [*Kwok et al*., 2009] and with in-situ and Airborne EM observations from the sea ice thickness CDR. They established uncertainty estimates for PIOMAS ice volume and trends, and concluded that PIOMAS provides useful estimates of changes in ice volume. Comparisons were made for all months in the year. *Laxon et al*. [2013] compared concatenated time series of ICESat and CryoSat data and found that derived trends agree within the established uncertainty limits from PIOMAS, further arguing that PIOMAS is useful for climate model evaluation.

In this paper, our focus is on representation of March ice thickness and volume. It is, therefore, useful to assess PIOMAS for this period in particular. We include data from ERS-1 and IceBridge, which have not been used in previous comparison studies. To this end, the middle column of **Figure 2** (center column) shows the PIOMAS thickness estimates corresponding to the five observational thickness data sets used in this study. The right hand column of **Figure 2** shows corresponding scatter plots between PIOMAS and the observations for each individual year of the observations (plotted as different colors for each year of data, except for the in-situ CDR, which includes 29 years of data, and ERS-1, which was provided as mean field over the entire time-period). The CDR data in the top scatter plot includes thicknesses from in-situ moorings, United States submarines and Airborne EM. Statistics are summarized in **Table 1**.

The observed thickness patterns and magnitudes generally compare well with those simulated by PIOMAS, providing further confidence that PIOMAS can be used to assess the CMIP5 volume trends during winter. However, the scatter plots reveal a general negative (too thin) thickness bias in PIOMAS for higher thickness values (found near the Canadian Archipelago and north of Greenland). The reverse tends to be true for areas of thin ice. In addition, PIOMAS tends to have a tongue of thicker ice (~2.5m) that stretches out across the Arctic Ocean to the Chukchi and East Siberian seas. The observations typically do not depict this feature, especially the ICESat record. PIOMAS also underestimates the ice thickness in the East Greenland Sea. The underestimation of thick ice and overestimation of thin ice by PIOMAS was previously noted in *Schweiger et al*. [2011]. In general the mean errors are smallest with respect to the submarine and ICESat data and are largest for the IceBridge, CryoSat and ERS-1 data.

Based on data comparisons and sensitivity studies, *Schweiger et al.* [2011] estimate an upper bound for the uncertainty of decadal PIOMAS trends of 1×10^3 km^3 dec^{-1}. Given the large observed volume trend of 2.8×10^3 km^3 dec^{-1} in March, PIOMAS is a suitable tool for assessing long-term trends CMIP5 models. Daily ice volume estimates at 25 km spatial resolution from PIOMAS were averaged to create monthly means of ice volume over the 1979 to 2013 record to compare with the CMIP5 output.

3. Results

3.1 Ice Thickness

We first compare observed and CMIP5 mean sea ice thickness fields averaged over the areas of coverage corresponding to each of the different remotely-sensed data sets [**Figure 3**]. The median spring thickness from each data set is shown as a solid red line, together with the 10th and 90th percentiles (green lines) and the interquartile range (grey shading).
Ice thicknesses from the 33 individual CMIP5 models are presented as box and whisker plots based on data for model years 1981 to 2010, where the boxes represent the interquartile range in thickness (25th to 75th percentiles), the whiskers the 10^{th} and 90^{th} percentiles, and the horizontal bars and asterisks within each box define the median and mean, respectively. As mentioned earlier, the 1981 to 2010 averaging time-period for CMIP5 is somewhat arbitrary as we cannot expect the natural variability in the models to be in phase with observed natural variability. This comparison therefore only reflects how well the long-term mean thickness fields in the models compare to the different observational data sets, such that if the spread of the observations for a given platform/instrument falls within the spread for a given model, we conclude the model captures the thickness. If the spread does not overlap, then there is a bias. We may additionally expect that the trend in thickness should be captured in the distributions of model thickness if one exists in those models.

In general, the thickness distributions from the models overlap those from each remotely-sensed data set. There are exceptions. Several models have negative biases in comparison to the in situ, ERS-1 and IceBridge data sets, with means below the 10^{th} percentile of the observations. A negative bias with respect to the in situ and ERS-1 data is not surprising as these observations sample from a thicker ice regime than the more recent two decades. However, some models that show a negative bias compared to the in situ and ERS-1 data also show a negative bias with respect to the IceBridge data (e.g. BCC-CSM1, CanCM4, CanESM2, CNRM-CM5, the GFDL models, MIROC ESM, MIROC-ESM-CHEM, MIROC4h, the MPI models and MRI-CGCM3), suggesting that the models are underestimating in regions of thick ice north of Greenland and the Canadian Archipelago sampled by the IceBridge flights.

The CMIP5 models show the best agreement with the ICESat and CryoSat observations. The ICESat and CryoSat statistics integrate more regions of thin ice along with the thick ice regions north of Greenland and the Canadian Archipelago, resulting in overall smaller mean thickness values compared to the other data sets. The coverage is also from a time period of significant ice thinning throughout most of the Arctic Ocean [e.g. *Kwok and Rothrock,* 2009; *Kwok et al*., 2009; *Laxon et al.,* 2013]. In comparison with ICESat, all but two models (CESM1-WACCM and FGOALS-g2) have a mean thickness within the 10^{th} and 90^{th} percentiles of the observed value. Mean thicknesses during the CryoSat period are slightly smaller than for ICESat, resulting in eight models (CESM-CAM5, CESM1-WACCM, CSIRO-MK3-6-0, EC-EARTH, FGOALS-g2, IPSL-CM5A-MR, MIROC5, NorESM1-M) having mean thicknesses above the 90^{th} percentile from CryoSat.

Given the limited temporal coverage of each observational data set, these comparisons should be regarded as a qualitative assessment. On the other hand, the fairly long PIOMAS record (30 years) brings the advantage of a long and reasonably homogenous data record to compare with the model data. The bottom of Figure 3 compares CMIP5 modeled ice thicknesses with PIOMAS estimates over the same 1981 to 2010 time-period. All but six models (CESM1-WACCM, EC-EARTH, FGOALS-g2, IPSL-CM5A-LR, MIROC5, and NORESM1-M) have mean March ice thickness values falling between the 10^{th} and 90^{th} percentiles of the PIOMAS values, and 70% (23) have mean thicknesses within the PIOMAS interquartile range (i.e. gray shading).

This good agreement with PIOMAS must be tempered by recognition of the pronounced inter-model spread in ice thickness aggregated across the Arctic Ocean and large differences in the spatial patterns of thickness [**Figure 4**]. Few models capture the pattern of thin ice close to the Eurasian coast and several additionally fail to place the thickest ice along the Canadian Arctic Archipelago and northern coast of Greenland (i.e. both ACCESS models, BCC-CSM1, CanCM4, CanESM2, CSIRO-Mk3, FIO-ESM, both GISS models, HadCM3, INMCM4, MIROC-ESM-CHEM). Instead, many models show a ridge of thick ice north of Greenland and across the Lomonosov Ridge towards the East Siberian shelf, with thinner ice in the Beaufort/Chukchi and the Kara/Barents seas. As a whole, the models tend to overestimate ice thickness over the central Arctic Ocean and along the Eurasian coast and underestimate ice thickness along the North American coast and north of Greenland and the Canadian Archipelago.

An analysis of spatial pattern correlations and root-mean-square error (RMSE) of ice thickness between CMIP5 models and ICESat observations documents serious model shortcomings. Spatial pattern correlations are less than 0.4 for all but three models (CCSM4, MIROC5 and MRI-GCGM3) [**Figure 5 (left)**] and RMSE values generally exceed 0.7 m [**Figure 5 (right)**]. These spatial pattern correlations are significantly smaller than those between ensembles from the same model, suggesting that the poor correlations cannot be explained by natural variability but rather a bias within the models. Interestingly, the spatial correlations in thickness between the CMIP5 models and PIOMAS are generally higher than those between the CMIP5 models and the ICESat data (not shown). The reason for this is that both PIOMAS and many of the CMIP5 models have a spurious tongue of fairly thick ice extending across the Arctic Ocean towards the Chukchi and East Siberian seas.

Kwok [2011] previously attributed deficiencies in ice thickness fields in the CMIP3 models to their inability to simulate the observed pattern of sea level pressure and hence surface winds. For example, if a model fails to produce a well-structured Beaufort Sea High (BSH) in the correct location north of Alaska, this will adversely affect the Beaufort Gyre ice drift and hence the thickness pattern. Models with overly thick ice offshore of Siberia suggest the presence of a strong anticyclonic drift that extends close to the coast, allowing ice to pile up on the upwind side. However, the presence of thick ice on the Siberian side could also be a result of a higher frequency of occurrence of a specific atmospheric circulation anomaly pattern.

We directly evaluated the annual mean sea level pressure fields and the associated surface geostrophic wind fields in the CMIP5 models [**Figure 6**] against fields from four different atmospheric reanalysis. Note that correlations between the reanalysis themselves range between 0.91 and 0.99 [**Table 3**]. In general, most models feature a closed BSH, though in some it is not well-defined (e.g. MPI-ESM-LR), is shifted towards the pole (e.g. CanCM4, CSIRO-Mk3-6-0, MIROC-ESM), or towards the eastern Arctic (e.g. IPSL-CM5A-LR). Models that do not feature a closed BSH (e.g. bcc-csm1-1, CCSM4, CESM1-WACCM, FGOALS-g2, FIO-ESM, IPSL-CM5A-MR, MIROC-ESM-CHEM and NorESM1) generally also have poor spatial thickness pattern correlations and large RMSEs (Figure 4). The exception is CCSM4. While CCSM4 shows good spatial pattern correlation in ice thickness and the lowest RMSE of all the models (computed with respect to ICESat), the mean sea level pressure pattern does not feature a closed BSH and the mean flow fails to capture the Beaufort Gyre and the Transpolar Drift Stream. Thus, while part of the failure of models to capture the observed thickness distribution can be explained in terms of biases in the surface wind fields, this is not always the case. This points to additional issues such as near surface vertical stability that affects the surface wind stress, sea ice rheology, ocean heat fluxes and the ice thickness itself as this affects ice mobility.

3.2 Ice Volume

Recent studies suggest that because of thinning, sea ice volume is declining faster than ice extent [e.g. *Schweiger et al*. 2011]. Ice volume is also a more important climate indicator than extent through its direct connection with the sea ice energy budget. The rates of ice volume loss for March and September calculated over the 1979 to 2013 period from PIOMAS are -9.9% and -27.9% dec^{-1}, respectively.

The CMIP5 multi-model ensemble mean March ice volume averaged over this period agrees well with PIOMAS, and remains within 1 standard deviation (1σ) throughout the 1979-2013 time-period [**Figure 7**]. When viewed as a group, this indicates that the models realistically capture the last three decades of changes in Arctic ice volume, assuming that PIOMAS provides a good representation of these changes. However, while we find good agreement between PIOMAS ice volume and the CMIP5 multi-model ensemble mean, ice volume varies substantially between different models. Average March ice volume ranges from around 18,000 km^3 (CanESM2) to 48,000 km^3 (CESM1-WACCM) [**Figure 7 – dashed lines**]. Additionally, as noted earlier, few models correctly capture the observed spatial pattern of thickness. Given the wide range of CMIP5 model results, the close match of the ensemble average with the PIOMAS average is somewhat puzzling. We speculate that modeling groups participating in the CMIP5 collection may each individually be working to construct and tune their models to match observed historical ice extent and thicknesses. If the effort or success by these groups is randomly distributed, then a close match of the ensemble mean volume and PIOMAS volume, which assimilates observed sea ice concentrations and is tuned to thickness observations, would be expected.

To evaluate CMIP5 ice volume further, volume trends were computed using linear least squares with a test statistic that combines the standard error of both the model and the observation and accounts for the effects of temporal autocorrelation. This approach, which follows *Santer et al.* [2008], was previously used by *Stroeve et al*. [2012a] to examine ice extent trends in both the CMIP3 and CMIP5 models and how those trends compared to the observed trend. As in *Stroeve et al.* [2012a], the null hypothesis is that the CMIP5 volume trends are consistent with those from PIOMAS. Ice volume trends during March from individual ensemble members range between -0.49x10^3 km^3 dec^{-1} (INMCM3) to -4.28x10^3 km^3 dec^{-1} (MIROC5) as assessed over the period 1979 to 2013 [**Table 2 and Figure 8**]. The corresponding PIOMAS trend is shown in gray shading for one (dark gray) and two standard deviations (light gray). Note that the gray shading does not represent the uncertainty in the PIOMAS volume estimates, which *Schweiger et al.* [2011] estimate to be 1x10^3 km^3. Therefore, the uncertainty in PIOMAS could be larger than we show.

While all model trends are negative, 10 ensemble members have trends that are insignificantly different from zero (i.e. 2σ of the trend overlaps with zero). Neglecting ensemble members with trends indistinguishable from zero, 36 of the remaining ensemble members have mean March volume trends slower, and two faster (IPSL-CM5A-LR and MIROC5) than the 2σ uncertainty of the PIOMAS trend. Nevertheless, the majority of the ensemble member trends cannot be considered incompatible with PIOMAS.

Finally, several ensembles show pronounced interannual variability in ice volume, with periods of increasing volume not captured by PIOMAS (not shown). Interannual variability in the ensembles likely reflects variability in atmospheric forcing. Averaging together the individual ensemble means from each model yields a multi-model ensemble mean trend in March ice volume of -1.95 10^3 km^3 dec^{-1} (or -6.8% dec^{-1} relative to the 1979-2013 mean). This is smaller than the PIOMAS rate of decline of -2.79 10^3 km^3 dec^{-1} (or -10.3% dec^{-1}) but remains within 2σ uncertainty of that value.

It is important to recognize that the difference in trends between PIOMAS and CMIP5 ensemble members can arise from systematic errors in the PIOMAS or CMIP5 models, uncertainties in the atmospheric reanalysis or that the trend in the PIOMAS time series includes significant contributions from natural climate variability. For example, *Day* [2012] attribute about 0.5 to 3.1% of the 1979 to 2010 September sea ice extent trend to changes in the Atlantic Meridional Overturning Circulation. The range of trends for individual models summarized in Table 2 indicates that natural variability maybe a strong contributor to ice volume trends over the last 35 years. However, the models themselves seem to strongly vary in the amount of natural variability in their integrations. The CSIR0-MK3-6-0 trends range from -3.19 to -0.67 10^3 km^3 dec^{-1} between its 10 ensemble members while HadCM3 features a substantially smaller range (-2.34 and -1.01 10^3 km^3 dec^{-1}) for its 10 ensemble members. This makes the identification of model biases or the filtering of models based on how well they represent observed trends difficult.

4. Conclusions

Evaluating model skill is important given the large role that the model projections play in framing the debate on how to address global environmental change. While the CMIP5 models more accurately hindcast sea ice extent than the CMIP3 models [e.g. *Stroeve et al.*, 2012a], trends from most models remain smaller than observed, lending concern that a seasonally ice-free Arctic state may be realized sooner than suggested by such models. Here we have evaluated sea ice thickness and volume from 33 CMIP5 models through comparisons with observed records of sea ice thickness and ice volume simulated by PIOMAS. We find that the CMIP5 models show a general thinning and reduction in ice volume over the period of observations. The CMIP5 ensemble mean ice volume trend over the 1979-2013 is smaller but within the uncertainties of the PIOMAS values. Although the Arctic-wide ensemble mean ice volume and trend is strikingly similar to the PIOMAS sea ice volume and trend, there are large variations among models.

Furthermore, while mean thickness and volume for the Arctic Ocean as a whole appears well represented by many of the models, spatial patterns of sea ice thickness are poorly represented. Many models fail to locate the thickest ice off the coast of northern Greenland and the Canadian Arctic Archipelago and thinner ice over the East Siberian Shelf. Part of the explanation lies in deficiencies in representing the details of the prevailing atmospheric circulation over the Arctic Ocean. This is a critical failure as projections of ice extent are strongly related to the initial ice thickness pattern distribution [e.g. *Holland et al*., 2010; e.g. *Holland and Stroeve*, 2011]. Moreover, *Holland and Stroeve* [2011] suggest that the variance of September sea ice extent anomalies explained by the winter-spring ice thickness increases as the ice-cover thins and transitions towards a seasonal ice cover. Thus as ice thins, the ability of models to represent the spatial thickness distribution, may become more relevant

Several techniques have been advanced in the literature to sub-select models based on different metrics of model performance during the historical time-period, with the aim of reducing uncertainty as to when an ice-free Arctic may be realized [e.g. *Wang and Overland*, 2009, 2012; *Boe et al.*, 2009; *Massonnet et al*., 2012]. It is clear from our study that even if a model captures the seasonal cycle in extent, or trends in extent and/or volume, the model may still poorly represent the prevalent atmospheric circulation patterns and thickness distributions. Indeed, we show that a model may get the trend in ice volume or ice extent reasonably correct, yet fail to locate the thickest ice north of Greenland and the Canadian Archipelago. Only two models capture *both* the spatial pattern of sea ice thickness and the general pattern of atmospheric circulation (MIROC5 and MRI-CGCM3), further reducing confidence in the veracity of future projections based on CMIP5 climate models. The fact that both models display rather different trends in ice volume (-3.6 10^3 km^3 dec^{-1} and -1.15 10^3 km^3 dec^{-1} respectively) does not bode well for constraining climate models based on sea ice thickness patterns alone.

Acknowledgements: This work was funded under NASA Grant NNX12AB75G.

References

Boe, J., A. Hall and X. Qu, (2009), September sea-ice cover in the Arctic ocean projected to vanish by 2100, *Nature Geosc*., 10.1038/NGE0467.

Brodzik, M.J., B. Billingsley, T. Haran, B. Raup and M.H. Savoie, (2012), EASE-Grid 2.0: Incremental but Significant Improvements for Earth-Gridded Data Sets, ISPRS Int. J. Geo-Inf. 2012, 1, 32-45; doi:10.3390/ijgi1010032.

Day, J.J., J.C. Hargreaves, J.D. Annan and A. Abe-Ouchi, (2012), Sources of multi-decadal variability in Arctic sea ice extent, *Environ. Res. Lett.* **7** 034011, doi:10.1088/1748-9326/7/3/034011.

Fetterer, F., K. Knowles, W. Meier, and M. Savoie. 2002, updated daily. Sea Ice Index. Boulder, Colorado USA: National Snow and Ice Data Center. Digital media.

Haas, C., Lobach, J., Hendricks, S., Rabenstein, L., Pfaffling, A. (2009). Helicopter-borne measurements of sea ice thickness, using a small and lightweight, digital EM system, *Journal of Applied Geophysics*, 67(3), 234-241.

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.

Holland, M.M., and J.C. Stroeve (2011), Changing seasonal sea ice predictor relationships in a changing Arctic climate, *Geophys. Res. Lett.,* doi:10.1029/2011GL049303.

Kwok, R. (2011), Observational assessments of Arctic Ocean sea ice motion, export, and thickness in CMIP3 climate simulations, *J. Geophys. Res*., doi:10.1029/2011JC007004

Kwok, R., and D. A. Rothrock (2009), Decline in Arctic sea ice thickness from submarine and ICESat records: 1958 2008, *Geophys. Res. Lett*., 36, L15501, doi:10.1029/2009GL039035.

Kwok, R., G. F. Cunningham, M. Wensnahan, I. Rigor, H. J. Zwally, and D. Yi (2009), Thinning and volume loss of Arctic sea ice: 2003-2008, *J. Geophys. Res*., doi:10.1029/2009JC005312.

Kurtz, N., M. Studinger, J. Harbeck, V. Onana, and S. Farrell. 2012a. *IceBridge Sea Ice Freeboard, Snow Depth, and Thickness*. Version 1. Boulder, Colorado USA: NASA DAAC at the National Snow and Ice Data Center.

Kurtz, N., et al. 2012b. Sea Ice Thickness, Freeboard, and Snow Depth Products from Operation IceBridge Airborne Data. *The Cryosphere Discussions* 6: 4771-4827. http://www.the-cryosphere-discuss.net/6/4771/2012/tcd-6-4771-2012.html.

Kurtz, N. T. and S. L. Farrell. 2011. Large-scale Surveys of Snow Depth on Arctic Sea Ice from Operation IceBridge. *Geophysical Research Letters* 38: L20505, doi:10.1029/2011GL049216.

Laxon, S.W., and others, (2013), CrySat-2 estimates of Arctic sea ice thickness and volume, *Geophys. Res. Lett*., doi:10.1002/grl.50193.

Laxon, S., N. Peacock and D. Smith, (2003), High interannual variability of sea ice thickness in the Arctic region, *Nature*, 425, doi:10.1038/nature02050.

Lindsay, R. 2013. Unified Sea Ice Thickness Climate Data Record Collection Spanning 1947-2012. Boulder, Colorado USA: National Snow and Ice Data Center. *http://dx.doi.org/10.7265/N5D50JXV*.

Massonnet, F., T. Fichefet, H. Goosse, C.M. Bitz, G. Phiippon-Berthier, M.M. Holland and P.-Y. Barriat, (2012), Constraining projections of summer Arctic sea ice, *The Cryosphere*, doi:10.5194/tc-6-1383-2012.

Melling, H. and D.A. Riedel. 2008. Ice Draft and Ice Velocity Data in the Beaufort Sea, 1990-2003. Boulder, Colorado USA: National Snow and Ice Data Center. Digital media.

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

NSIDC, (1998, updated 2006). *Submarine Upward Looking Sonar Ice Draft Profile Data and Statistics*. [indicate subset used]. Boulder, Colorado USA: National Snow and Ice Data Center. http://dx.doi.org/10.7265/N54Q7RWK.

Rothrock, D. A. and M. Wensnahan, (2007), The Accuracy of Sea-Ice Drafts Measured from U.S. Navy Submarines. *Journal of Atmospheric and Oceanic Technology* 24 (11): 1936-49. doi: 10.1175/JTECH2097.1.

Santer, B.D. et al. (2008), Consistency of modeled and observed temperature trends in the tropical troposphere, *Int. J. Climatol*., 28, 1703-1722, doi:10.1002/joc.1756.

Schweiger, A., R. Lindsay, J. Zhang, M. Steele, H. Stern and R. Kwok, (2011), Uncertainty in modeled Arctic sea ice volume, *J. Geophys. Res. Oceans*, doi: 10.1029/2011JC007084.

Serreze, M. C., A. P. Barrett, J. C. Stroeve, D. N. Kindig, and M. M. Holland, (2009). The emergence of surface-based Arctic amplification. *The Cryosphere* 3: 11–19.

Stroeve, J.C., L. Hamilton, C. Bitz, and E. Blanchard-Wigglesworth (2014), Predicting September sea ice: Ensemble skill of the SEARCH sea ice outlook 2008-2013. Geophysical Research Letters, doi: 10.1002/2014GL059388.

Stroeve, J. C., V. Kattsov, A. Barrett, M. Serreze, T. Pavlova, M. Holland, and W. N. Meier, (2012a), Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations, *Geophys. Res. Lett*., 39, L16502, doi:10.1029/2012GL052676.

Stroeve, J.C., M.C. Serreze, J.E. Kay, M.M. Holland, W.N. Meier and A.P. Barrett, (2012b). The Arctic's rapidly shrinking sea ice cover: A research synthesis, *Clim. 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.

Thompson, A.M., K.V. Calvin, S.J. Smith, G.P. Kyle, A. Volke, P. Patel, S. Delgado-Arias, B. Bond-Lamberty, M.A. Wise, L.E. Clark and J.A. Edmonds, (2011), RCP4.5: a pathway for stabilization of radiative forcing by 2100, *Climatic Change*, doi:10.1007/s10584-011-0151-4.

Wadhams, P. (1984), Arctic sea ice morphology and its measurement. *Arctic Technology and Policy*. I. Dyer and C. Chryssostomidis, eds., Washington, D.C., Hemisphere Publishing Corp., 179-195.

Wang, M. and J.E. Overland, (2012), A sea ice free summer Arctic within 30 years – an update from CMIP5 models, *Geophys. Res. Lett*., doi:10.1029/2012GL052868.

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.

Warren, S, I.G. Rigor and N. Untersteiner, (1998), Snow depth on Arctic sea ice, *J. Climate*, 12, 1814-1829.

Zhang, J.L. and D.A. Rothrock, (2003), Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates, *Mon. Weather Rev*., 131, 845-861.

Zygmuntowska, M., P. Rampal, N. Ivanova, and L.H. Smedsrud, (2014), Uncertainties in Arctic sea ice thickness and volume: new estimates and implications for trends, The Cryosphere, doi:10.5194/tc-8-705-2014

Table 1. Mean ice thickness bias, root-mean-square error estimate and correlation between PIOMAS modeled ice thickness and thicknesses from different remotely-sensed data sets.

Observations	**Mean Error (m)**	**RMSE (m)**	**Correlation (r)**
In Situ and Submarine	-0.15	0.78	0.70
ERS-1	-0.36	0.55	0.70
ICESat	0.20	0.50	0.68
IceBridge	-0.47	0.56	0.47
CryoSat-2	-0.37	0.81	0.38

Table 2. Linear trends in Arctic sea ice volume for March based on the period 1979 to 2013 from 33 CMIP5 models and PIOMAS. For models with more than one ensemble member, the mean trend is given along with the range in trend (in parenthesis). Trends are listed as km^3 per decade. Trends statistically different from 0 at 95 and 99% significance are denoted by + and ++, respectively.

Modeling Center (or Group)	**Model Name**	**Trend (10^3 km^3/decade)**	**Range of Trends**	**Numb Ensen**
Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology, Australia	ACCESS-0	-1.77^{++}		1
	ACCESS-3	-2.16^{++}		
Beijing Climate Center, China Meterological Administration	BCC-CSM1-1	-1.83^{++}		1
Canadian Centre for Climate Modelling and Analysis	CanCM4	-0.94^{++}	(-1.23 to -0.68)	9
	CanESM2	-1.03^{++}	(-1.15 to -0.74)	5
National Center for Atmospheric Research	CCSM4	-2.37^{++}	(-2.79 to -1.49)	6
	CESM1-CAM5	-3.13^{++}	(-3.18 to -3.08)	2
	CESM1-WACCM	-3.26^{++}	(-3.63 to -3.00)	3
Centre National de Recherches Meteorologiques/Centre Europeen de Recherche et Formation Avancee en Calcul Scientifique	CNRM-CM5	-2.34^{++}		1
Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence	CSIRO-Mk3-6-0	-2.09^{++}	(-3.19 to -0.67)	1(
EC-EARTH consortium	EC-EARTH	-2.21		1
LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and CESS, Tsinghua University	FGOALS-g2	-3.39^{++}		1
The First Institute of Oceanography, SOA, China	FIO-ESM	-1.25^{++}	(-1.36 to -0.99)	3
NOAA Geophysical Fluid Dynamics Laboratory	GFDL-CM3	-1.68^{++}		1
	GFDL-ESM2G	-1.63^{++}		1

	GFDL-ESM2M	-0.75		1
NASA Goddard Institute for Space Studies	GISS-E2-R	-2.54++	(-3.20 to -1.77)	3
	GISS-E2-H	-1.28++	(-1.40 to -0.81)	5
Met Office Hadley Centre	HadCM3	-1.72++	(-2.34 to -1.01)	1
	HadGEM2-AO	-2.32++		1
	HadGEM2-CC	-2.92++		1
	HadGEM2-ES	-2.26++		1
Institute for Numerical Mathematics	INMCM4	-0.49		1
Institut Pierre-Simon Laplace	IPSL-CM5A-LR	-2.90++	(-3.85 to -2.31)	4
	IPSL-CM5A-MR	-2.48++		1
Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo) and National Institute for Environmental Studies	MIROC-ESM	-0.96++		1
	MIROC-ESM-CHEM	-1.76++		1
	MIROC4h	-1.95++	(-2.34 to -1.27)	3
	MIROC5	-3.63++	(-4.28 to -2.98)	2
Max-Planck-Institut fur Meteorologie	MPI-ESM-LR	-1.37++	(-1.66 to -0.85)	3
	MPI-ESM-MR	-2.48++	(-2.37 to -0.92)	3
Meteorological Research Institute	MRI-CGCM3	-1.15		1
Norwegian Climate Centre	NorESM1-M	-2.41+		1
	Multi-model Mean	**-1.95++**		**2**
	PIOMAS	-2.79++		

Table 3. Spatial correlations between observed mean annual sea level pressure from four different reanalysis data sets and from the CMIP5 models. Ranks of correlations are given in parentheses, running lowest to highest. Because of difficulties in reducing surface pressures to sea level, pressures over Greenland have been screened out. Correlations between the different reanalysis are also included as well as whether or not the models represent a closed Beaufort Sea High (BSH).

Model	ERA-Interim	MERRA	CFSR	NCEP	Closed BSH?
1. ACCESS1-0	0.89 (26)	0.93 (28)	0.86 (25)	0.82 (21)	Y
2. ACCESS1-3	0.89 (28)	0.94 (29)	0.86 (27)	0.82 (23)	Y
3. bcc-csm1-1	0.76 (12)	0.74 (10)	0.73 (13)	0.71 (14)	N
4. CanCM4	0.69 (4)	0.74 (9)	0.65 (3)	0.61 (3)	Y
5. CanESM2	0.72 (7)	0.77 (12)	0.67 (8)	0.63 (7)	Y
6. CCSM4	0.62 (4)	0.51 (1)	0.66 (6)	0.70 (12)	N
7. CESM1-CAM5	0.93 (32)	0.89 (26)	0.93 (33)	0.91 (33)	Y
8. CESM1-WACCM	0.82 (18)	0.83 (19)	0.80 (17)	0.77 (17)	N
9. CNRM-CM5	0.73 (8)	0.79 (14)	0.67 (7)	0.63 (6)	Y
10. CSIRO-Mk3-6-0	0.58 (3)	0.67 (4)	0.52 (3)	0.47 (3)	Y
11. EC-EARTH	0.92 (31)	0.94 (31)	0.89 (30)	0.86 (28)	Y
12. FGOALS-g2	0.43 (1)	0.52 (2)	0.36 (1)	0.31 (1)	N
13. FIO-ESM	0.54 (2)	0.60 (3)	0.49 (2)	0.44 (2)	N
14. GFDL-CM3	0.87 (24)	0.88 (24)	0.85 (22)	0.82 (22)	Y
15. GFDL-ESM2G	0.75 (10)	0.82 (16)	0.70 (10)	0.65 (8)	Y
16. GFDL-ESM2M	0.76 (13)	0.82 (17)	0.71 (11)	0.66 (10)	Y
17. GISS-E2-R	0.81 (15)	0.84 (17)	0.78 (14)	0.74 (14)	Y
18. GISS-E2-H	0.87 (25)	0.88 (23)	0.84 (21)	0.81 (20)	Y
19. HadCM3	0.63 (5)	0.72 (7)	0.58 (4)	0.53 (4)	Y
20. HadGEM2-AO	0.94 (33)	0.97 (33)	0.92 (32)	0.88 (29)	Y
21. HadGEM2-CC	0.89 (27)	0.94 (30)	0.86 (23)	0.81 (19)	Y
22. HadGEM2-ES	0.90 (29)	0.95 (32)	0.87 (28)	0.83 (25)	Y
23. inmcm4	0.86 (21)	0.84 (21)	0.86 (24)	0.83 (26)	Y
24. IPSL-CM5A-LR	0.83 (20)	0.78 (13)	0.84 (20)	0.83 (24)	Y
25. IPSL-CM5A-MR	0.81 (16)	0.73 (8)	0.83 (18)	0.84 (27)	N
26. MIROC4h	0.78 (14)	0.83 (18)	0.74 (14)	0.70 (11)	Y
27. MIROC5	0.80 (15)	0.86 (22)	0.76 (15)	0.71 (15)	Y
28. MIROC-ESM	0.73 (9)	0.73 (9)	0.69 (9)	0.66 (9)	Y
29. MIROC-ESM-CHEM	0.75 (11)	0.71 (5)	0.73 (12)	0.71 (13)	N
30. MPI-ESM-LR	0.86 (23)	0.89 (25)	0.83 (19)	0.81 (18)	Y
31. MPI-ESM-MR	0.91 (30)	0.90 (27)	0.90 (31)	0.88 (30)	Y
32. MRI-CGCM3	0.86 (22)	0.79 (15)	0.87 (29)	0.89 (31)	Y
33. NorESM1-M	0.82 (19)	0.71 (46	0.86 (26)	0.89 (32)	N
ERA-Interim	1.00 (37)	0.96 (35)	0.99 (36)	0.97 (35)	
MERRA	0.96 (34)	1.00 (37)	0.94 (34)	0.91 (33)	
CFSR	0.99 (36)	0.94 (33)	1.00 (3)	0.99 (36)	

NCEP	0.97 (35)	0.91 (28)	0.99 (35)	1.00 (37)	

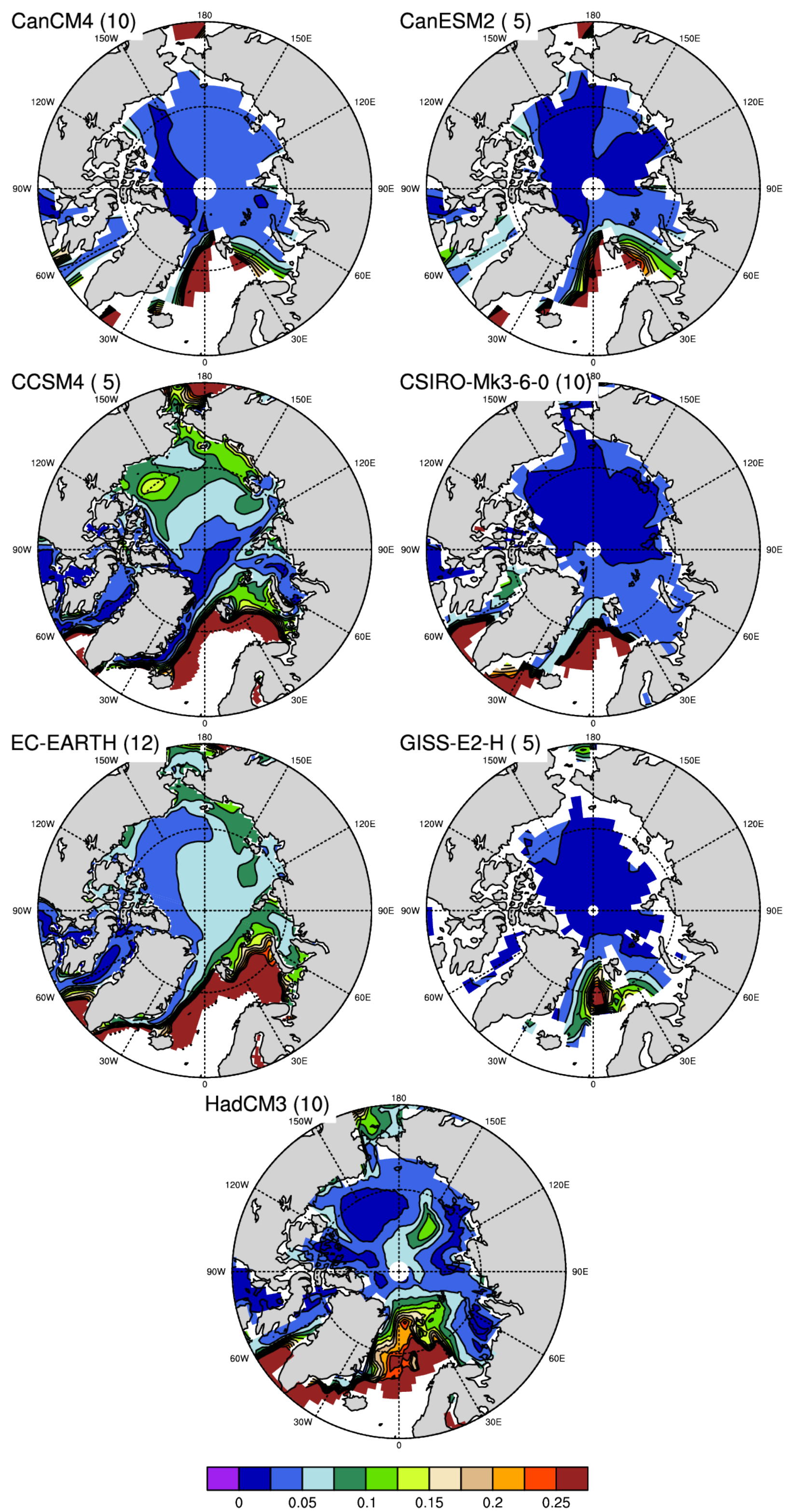


Figure 1. Variability of thicknesses in seven models is attached. The values are coefficient of variability (stddev/average). This is a normalized measure of variability so that variability can be compared spatially and between models.

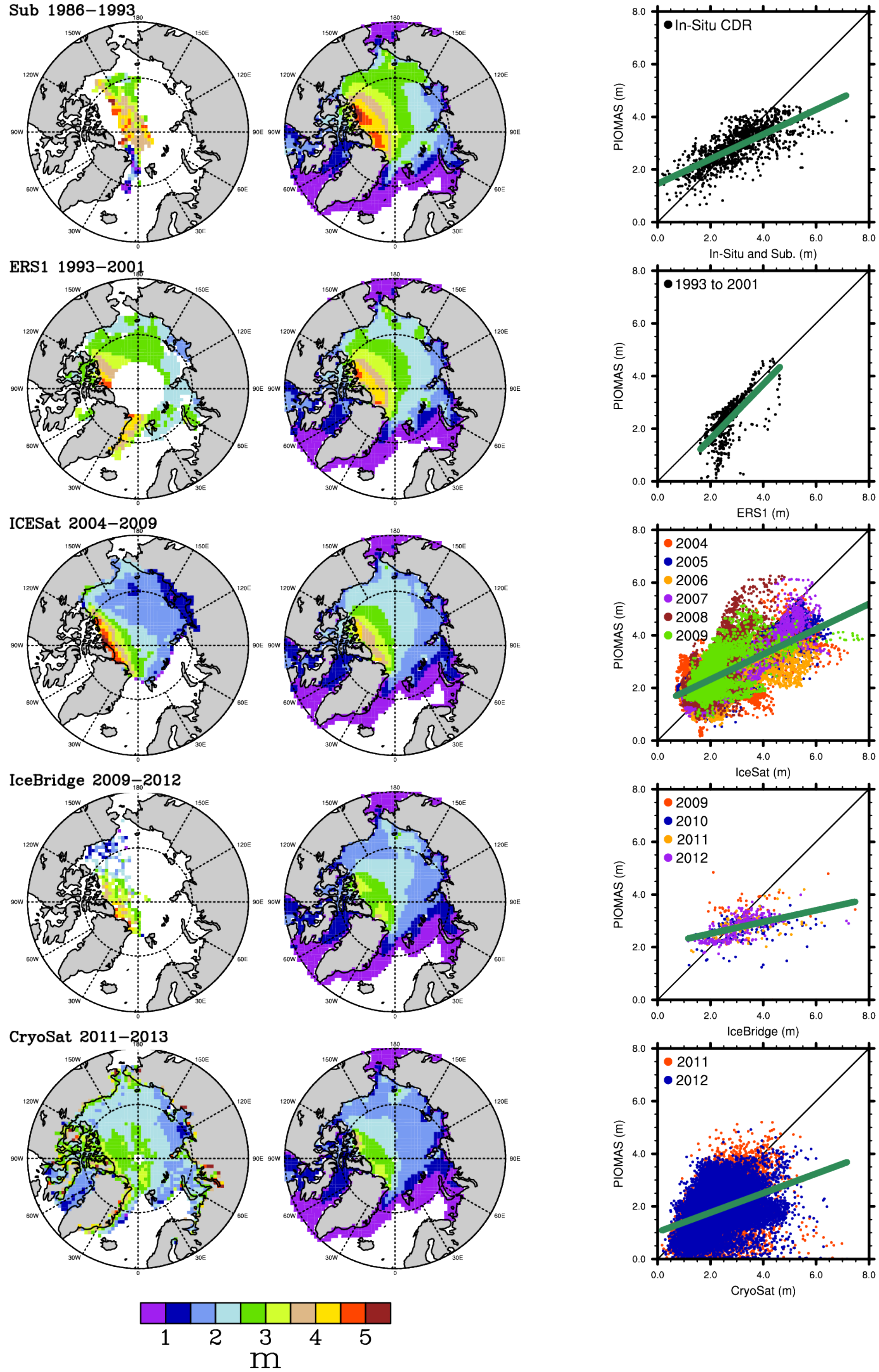


Figure 2. Comparison of submarine, ERS-1, ICESat, IceBridge and CryoSat-2 sea ice thickness fields (left column), for each campaigns period of record, with ice thickness fields simulated by PIOMAS (middle column) and corresponding scatter plots (right column). PIOMAS fields are the average March thicknesses for the same periods as corresponding observed records. In the scatter plots, individual years are shown in different colors, except for ERS-1, which was provided as a mean field for the entire time-period.

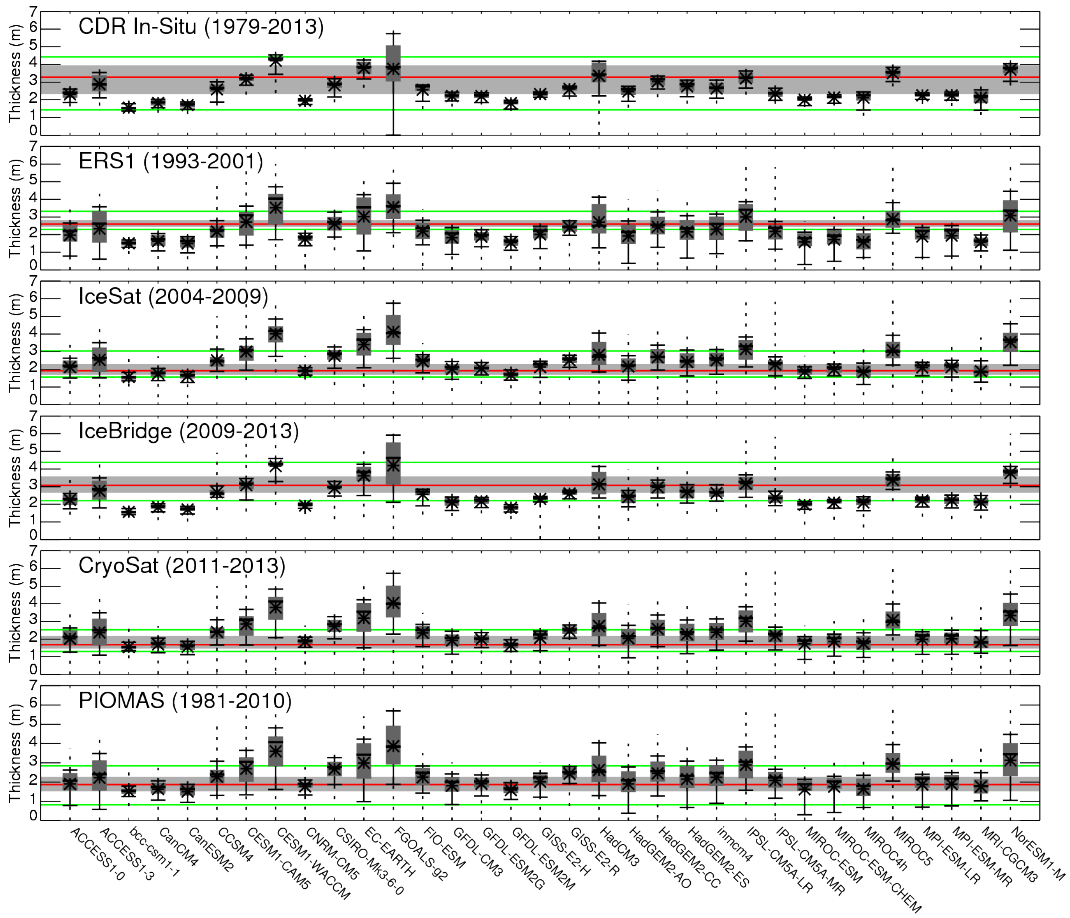
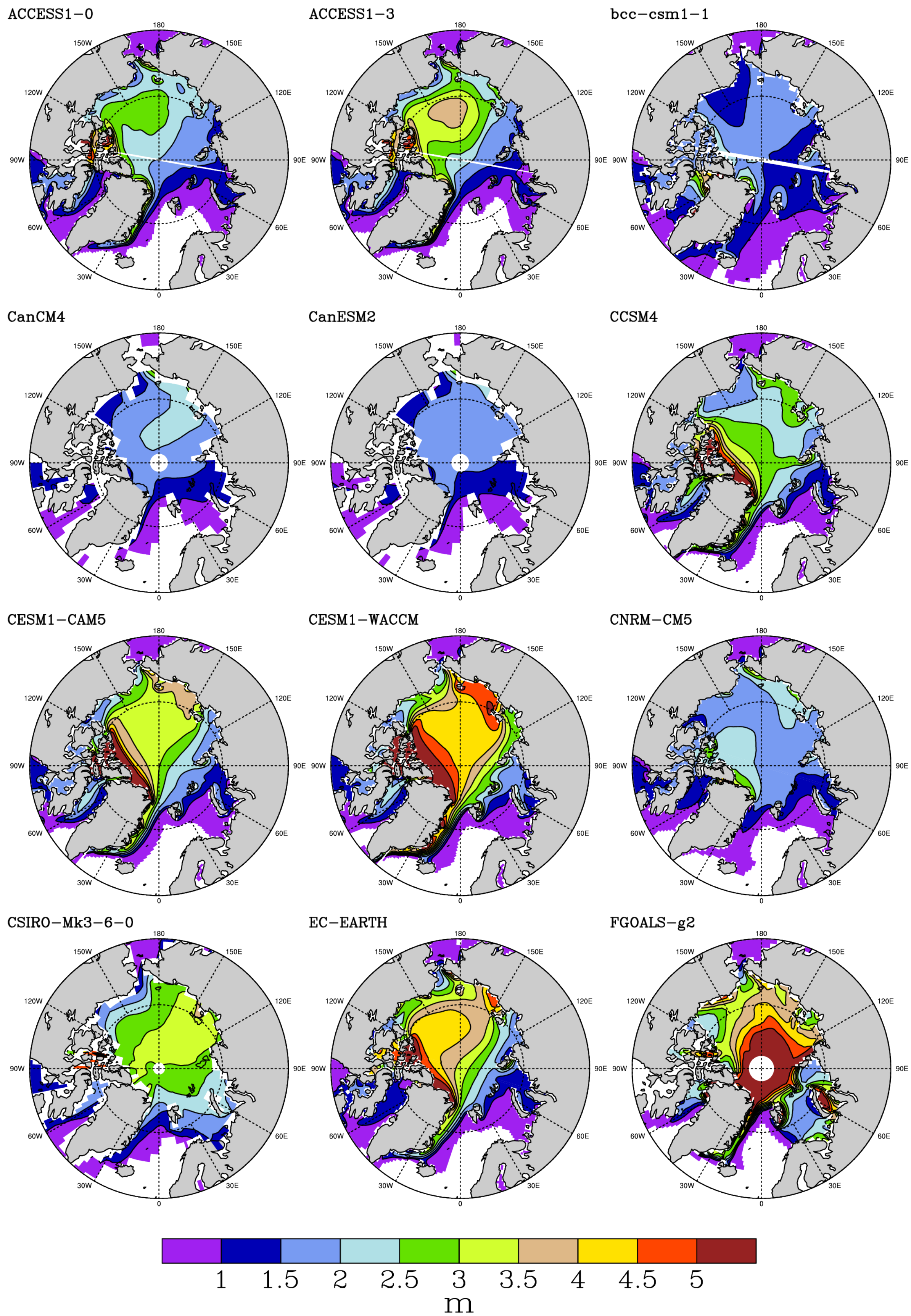
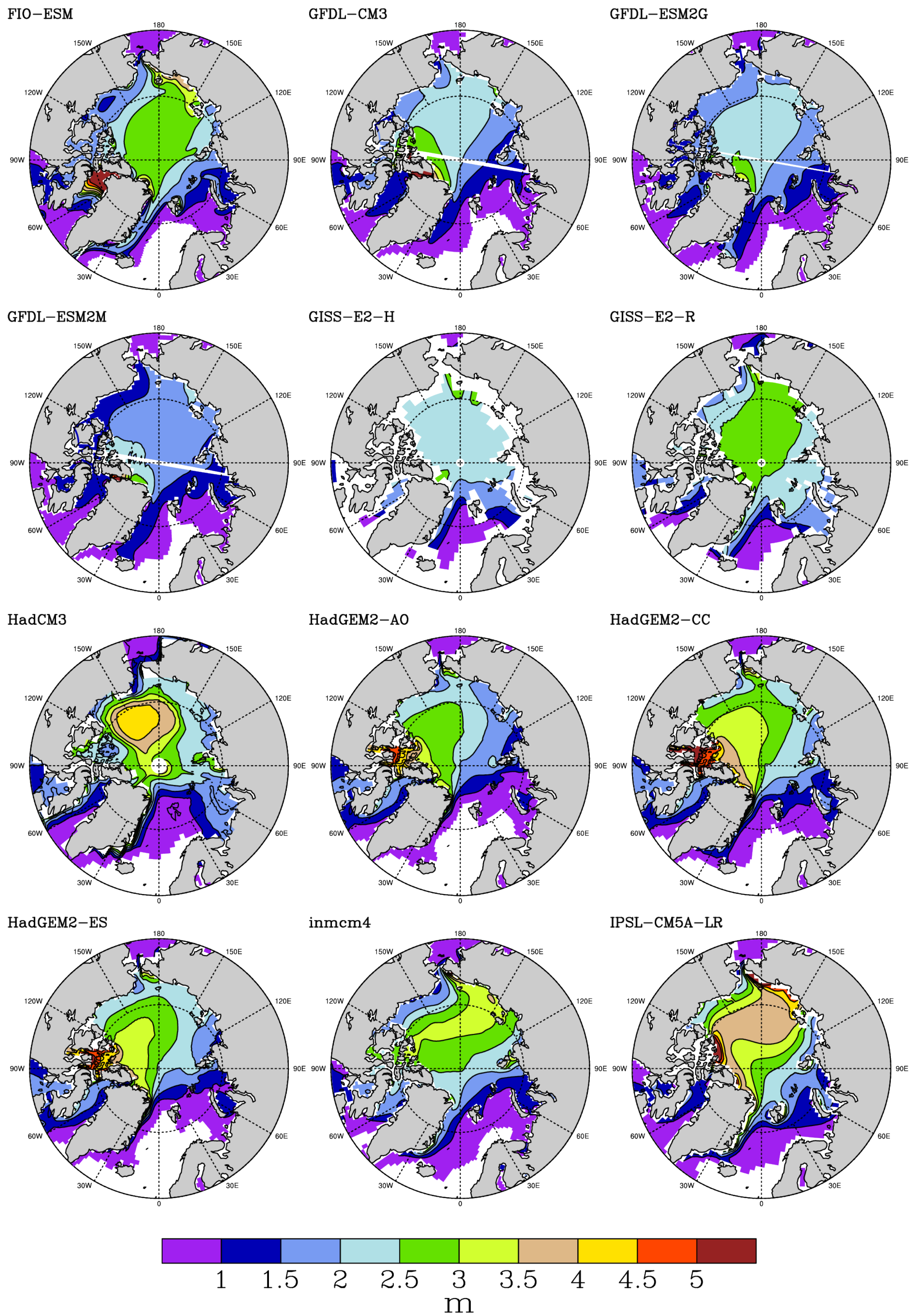


Figure 3. Comparison of thickness distributions between five observational data sets, PIOMAS and 33 individual CMIP5 models. Model results are presented as box and whisker plots from 1981 to 2010, where the boxes represent the interquartile range (25th to 75th percentiles) and the horizontal bars and asterisks within each box define the median and mean, respectively. The median spring thicknesses from each observational data set and PIOMAS are shown as a solid red line, together with the 10th and 90th percentiles (green lines) and the interquartile range (grey shading).


ACCESS1−0
ACCESS1−3
bcc−csm1−1
CanCM4
CanESM2
CCSM4
CESM1−CAM5
CESM1−WACCM
CNRM−CM5
CSIRO−Mk3−6−0
EC−EARTH
FGOALS−g2
1 1.5 2 2.5 3 3.5 4 4.5 5
m


FIO−ESM
GFDL−CM3
GFDL−ESM2G
GFDL−ESM2M
GISS−E2−H
GISS−E2−R
HadCM3
HadGEM2−AO
HadGEM2−CC
HadGEM2−ES
inmcm4
IPSL−CM5A−LR
180
150W
150E
120W
120E
90W
90E
60W
60E
30W
30E
0
1
1.5
2
2.5
3
3.5
4
4.5
5
m

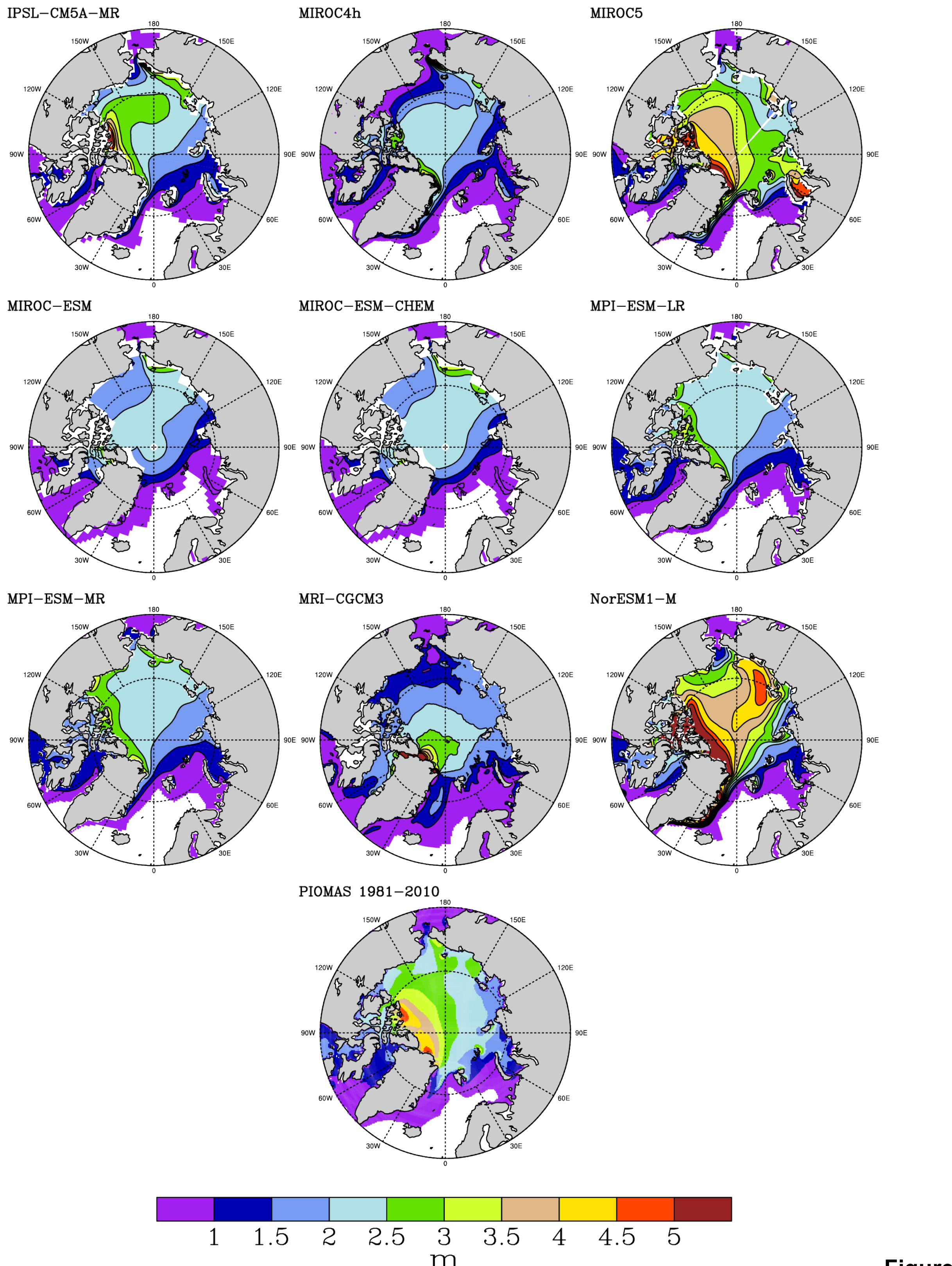
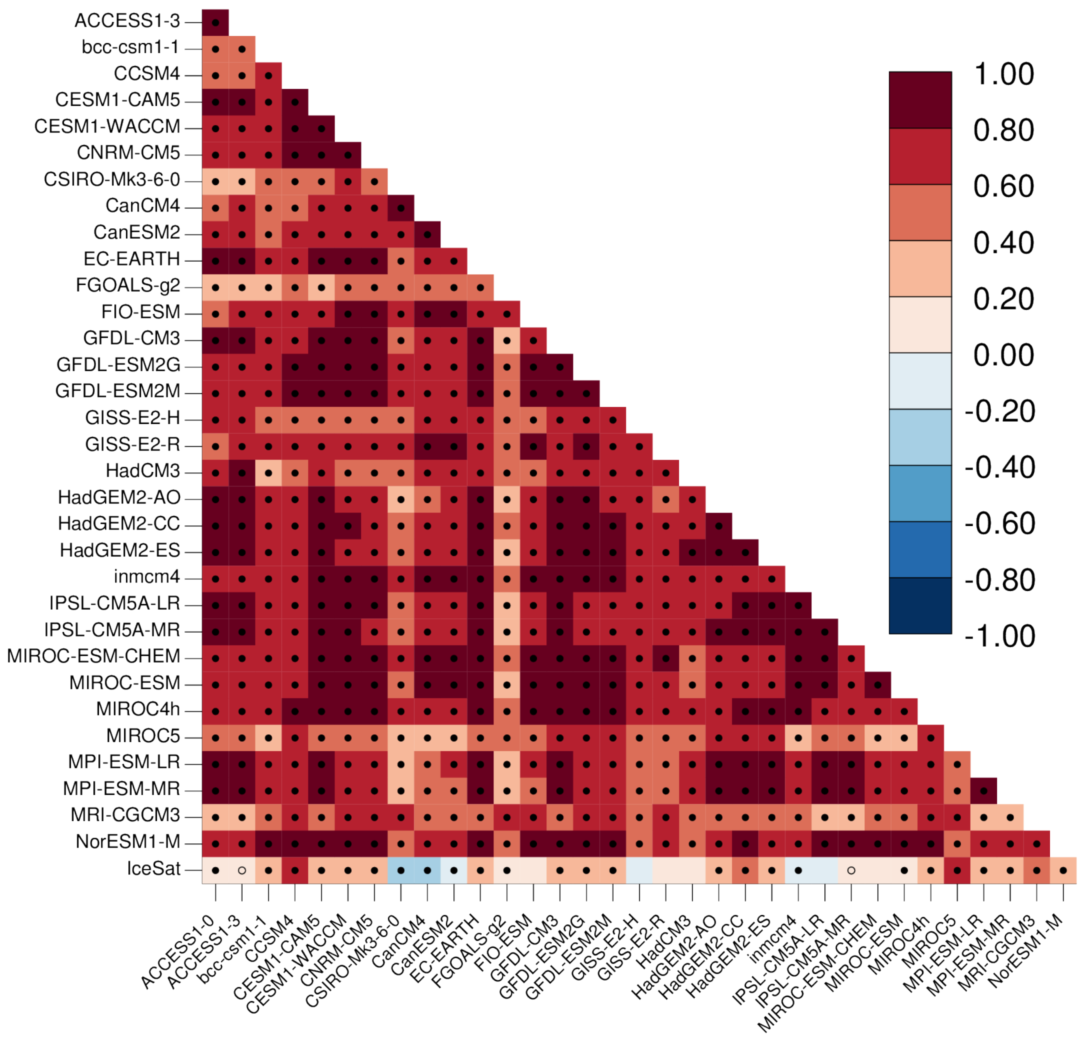


Figure 4. Spatial patterns of sea ice thickness from 1981 to 2010 from 33 CMIP5 models and PIOMAS.


ACCESS1-3
bcc-csm1-1
CCSM4
CESM1-CAM5
CESM1-WACCM
CNRM-CM5
CSIRO-Mk3-6-0
CanCM4
CanESM2
EC-EARTH
FGOALS-g2
FIO-ESM
GFDL-CM3
GFDL-ESM2G
GFDL-ESM2M
GISS-E2-H
GISS-E2-R
HadCM3
HadGEM2-AO
HadGEM2-CC
HadGEM2-ES
inmcm4
IPSL-CM5A-LR
IPSL-CM5A-MR
MIROC-ESM-CHEM
MIROC-ESM
MIROC4h
MIROC5
MPI-ESM-LR
MPI-ESM-MR
MRI-CGCM3
NorESM1-M
IceSat
ACCESS1-0
ACCESS1-3
bcc-csm1-1
CCSM4
CESM1-CAM5
CESM1-WACCM
CNRM-CM5
CSIRO-Mk3-6-0
CanCM4
CanESM2
EC-EARTH
FGOALS-g2
FIO-ESM
GFDL-CM3
GFDL-ESM2G
GFDL-ESM2M
GISS-E2-H
GISS-E2-R
HadCM3
HadGEM2-AO
HadGEM2-CC
HadGEM2-ES
inmcm4
IPSL-CM5A-LR
IPSL-CM5A-MR
MIROC-ESM-CHEM
MIROC-ESM
MIROC4h
MIROC5
MPI-ESM-LR
MPI-ESM-MR
MRI-CGCM3
NorESM1-M
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
-0.40
-0.60
-0.80
-1.00

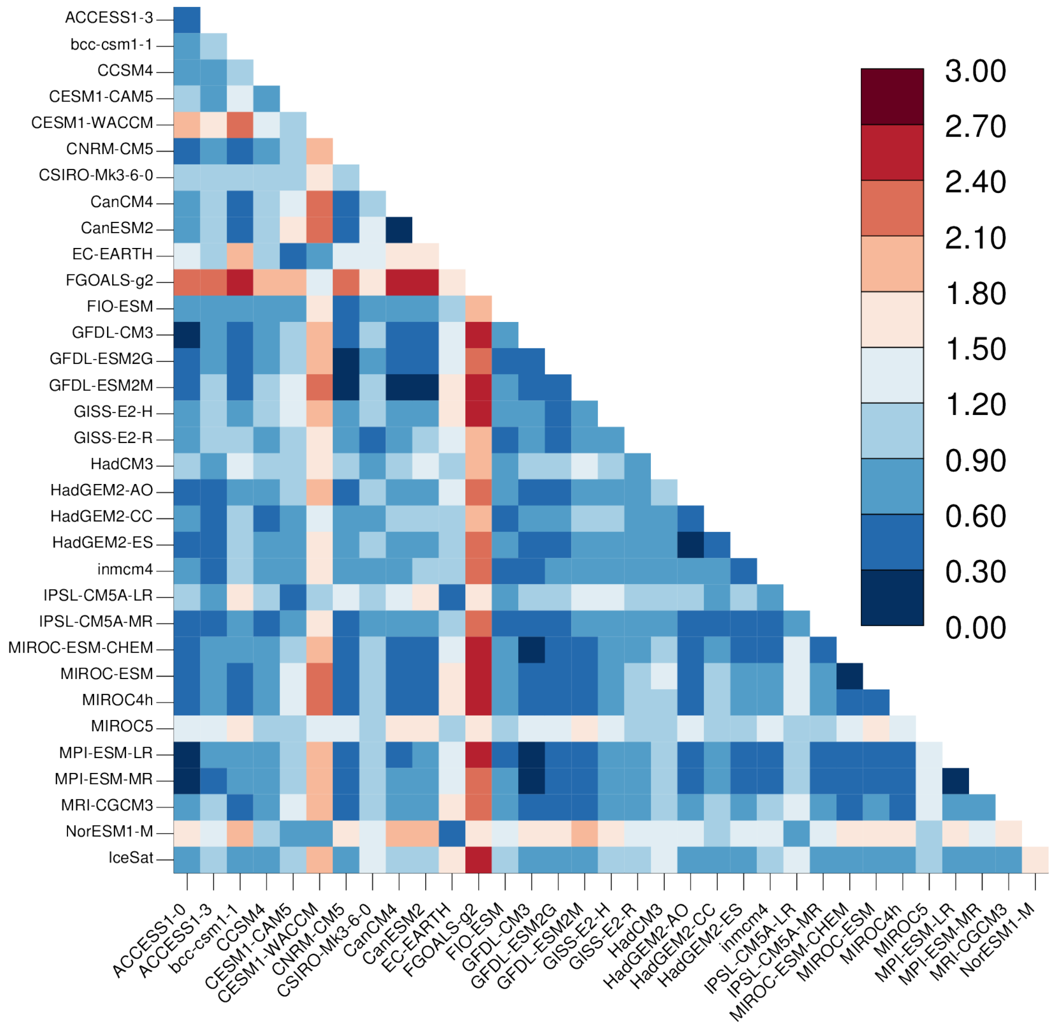
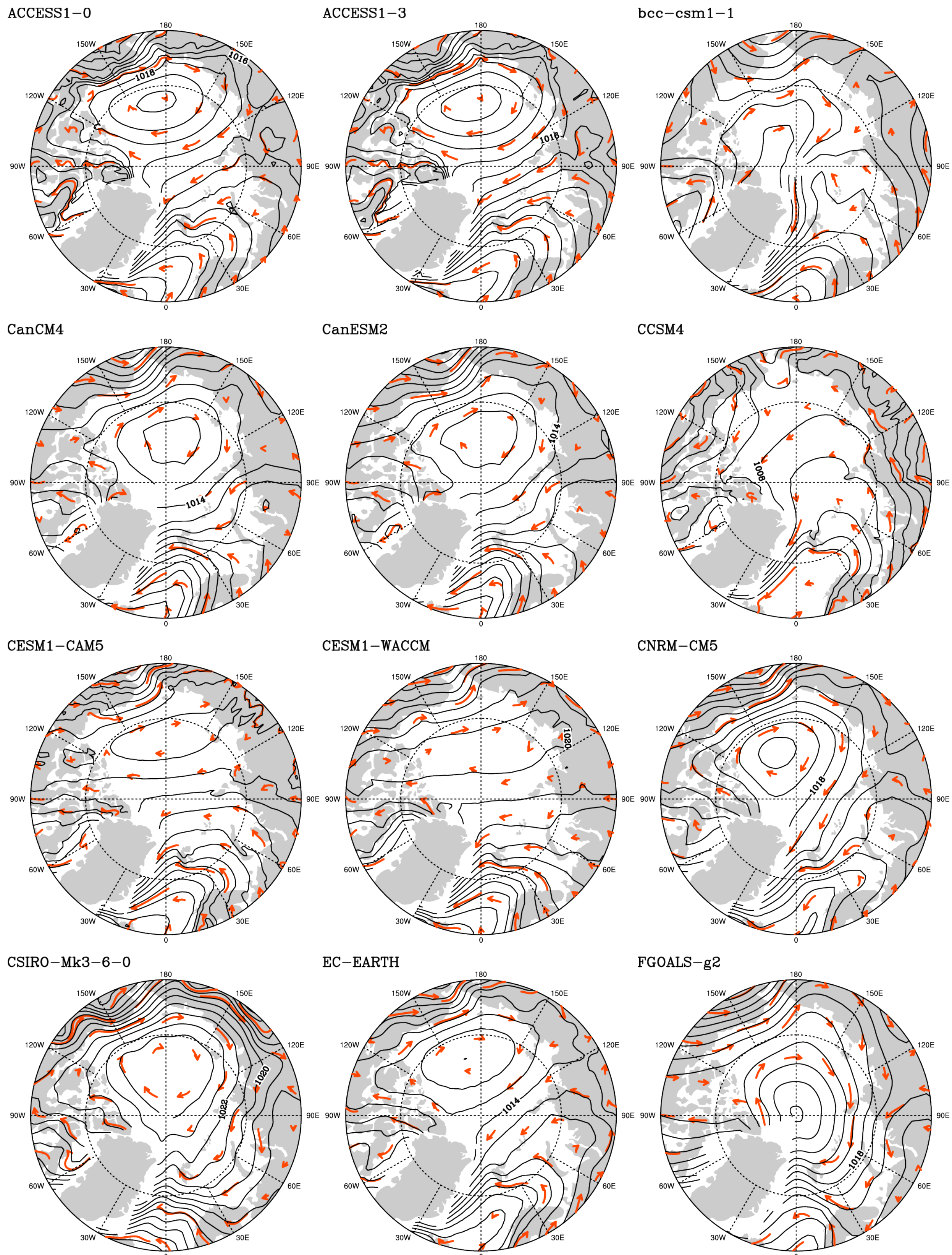
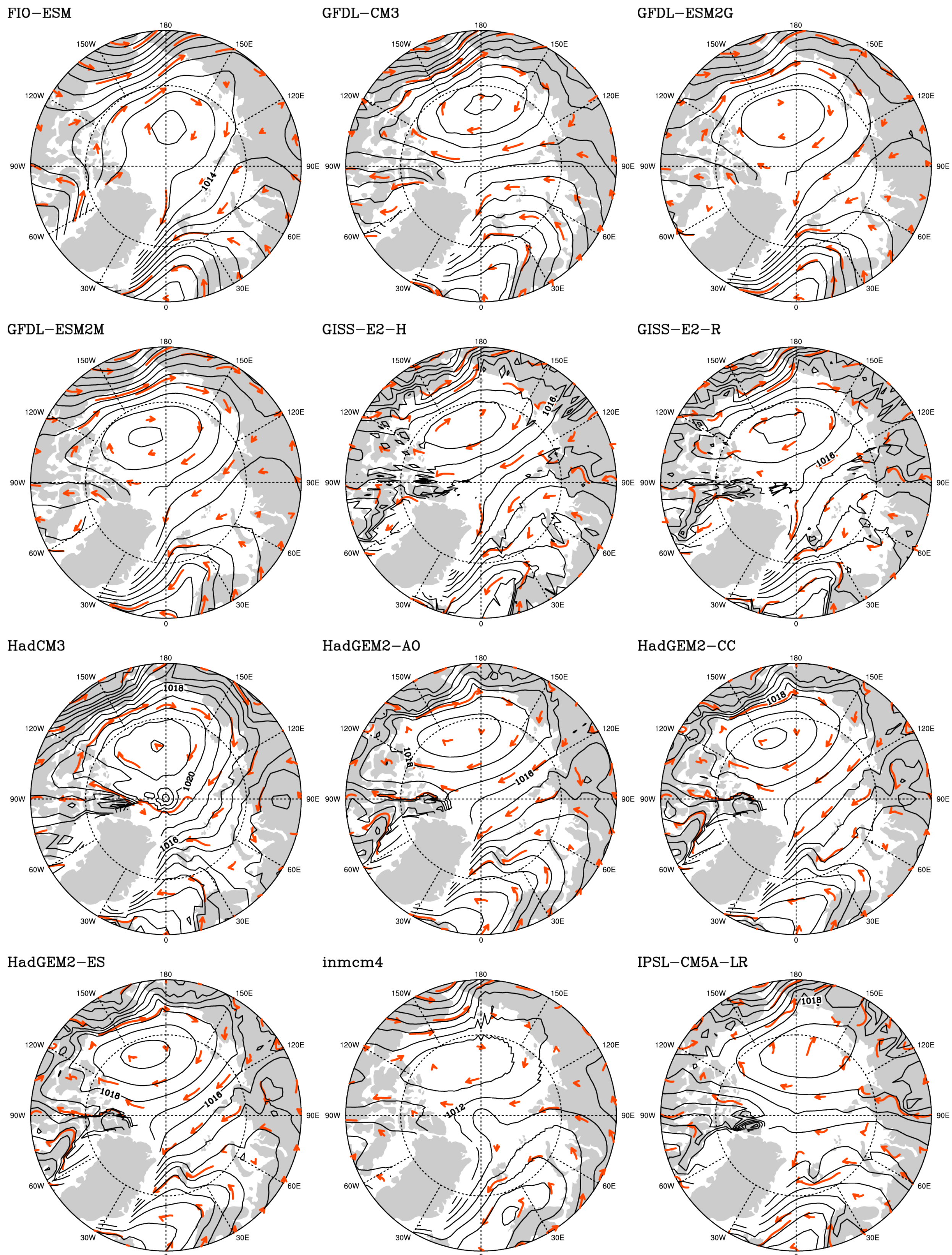


Figure 5. Spatial pattern correlations (top) and root-mean-square error (RMSE) (bottom) of ice thickness in 27 CMIP5 models and ICESat. Filled and hollow circles indicate correlations that are significant at the 99% and 95% level.


ACCESS1−0
ACCESS1−3
bcc−csm1−1
CanCM4
CanESM2
CCSM4
CESM1−CAM5
CESM1−WACCM
CNRM−CM5
CSIRO−Mk3−6−0
EC−EARTH
FGOALS−g2


FIO−ESM
1014
GFDL−CM3
GFDL−ESM2G
GFDL−ESM2M
GISS−E2−H
1016
GISS−E2−R
1016
HadCM3
1018
1020
1016
HadGEM2−AO
1018
1016
HadGEM2−CC
1018
HadGEM2−ES
1018
1016
inmcm4
1012
IPSL−CM5A−LR
1018
180
150W
150E
120W
120E
90W
90E
60W
60E
30W
30E
0

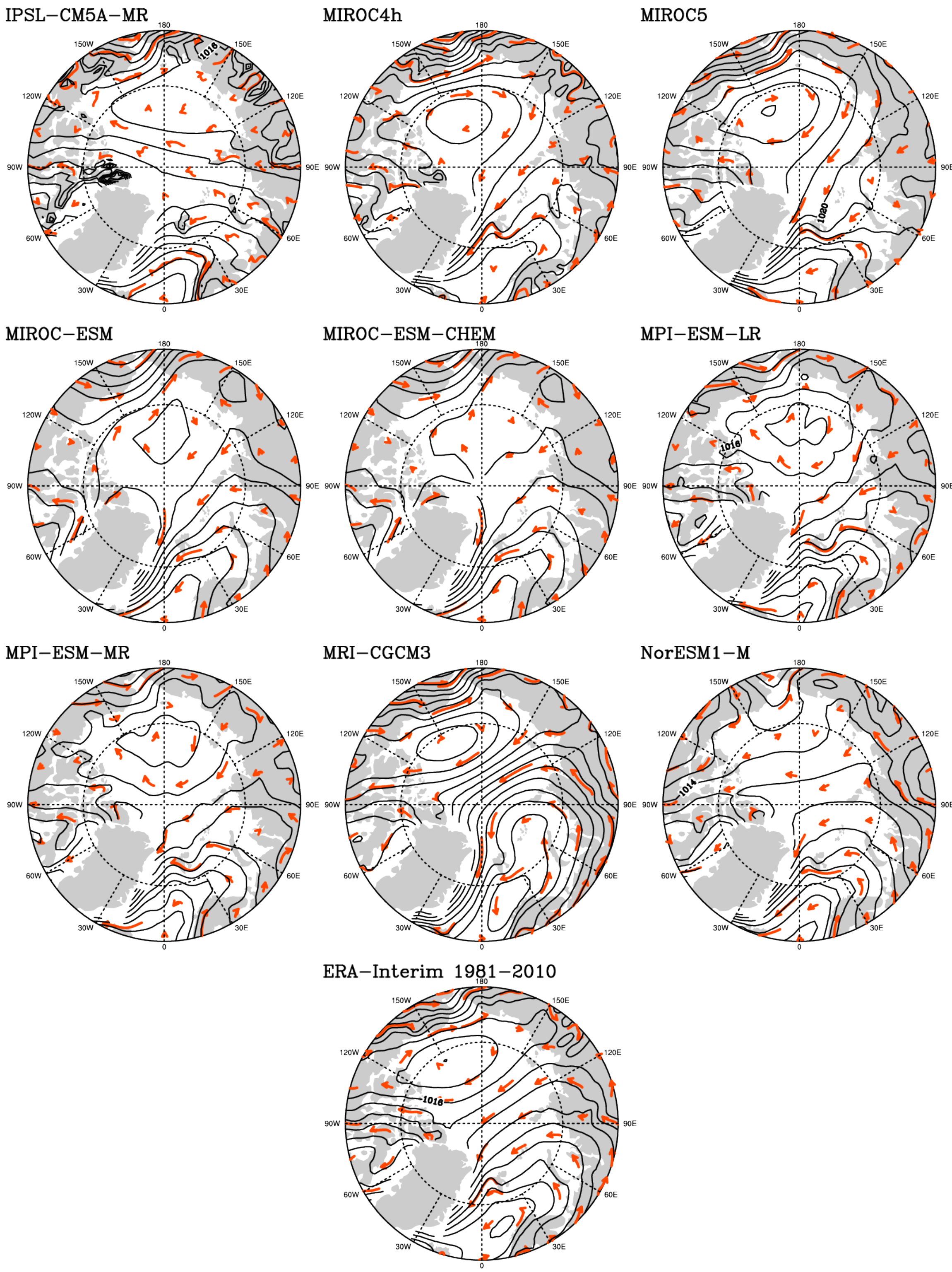


Figure 6. Mean annual sea level pressure and geostrophic wind from 27 CMIP5 models and from ERA-Interim spanning 1981-2010. Contour interval is 1 hPa. Near-surface geostrophic wind is used as a proxy for sea ice motion and is shown by red vectors. Vector length is proportional to wind speed. Vectors are curved tangent to the instantaneous flow.

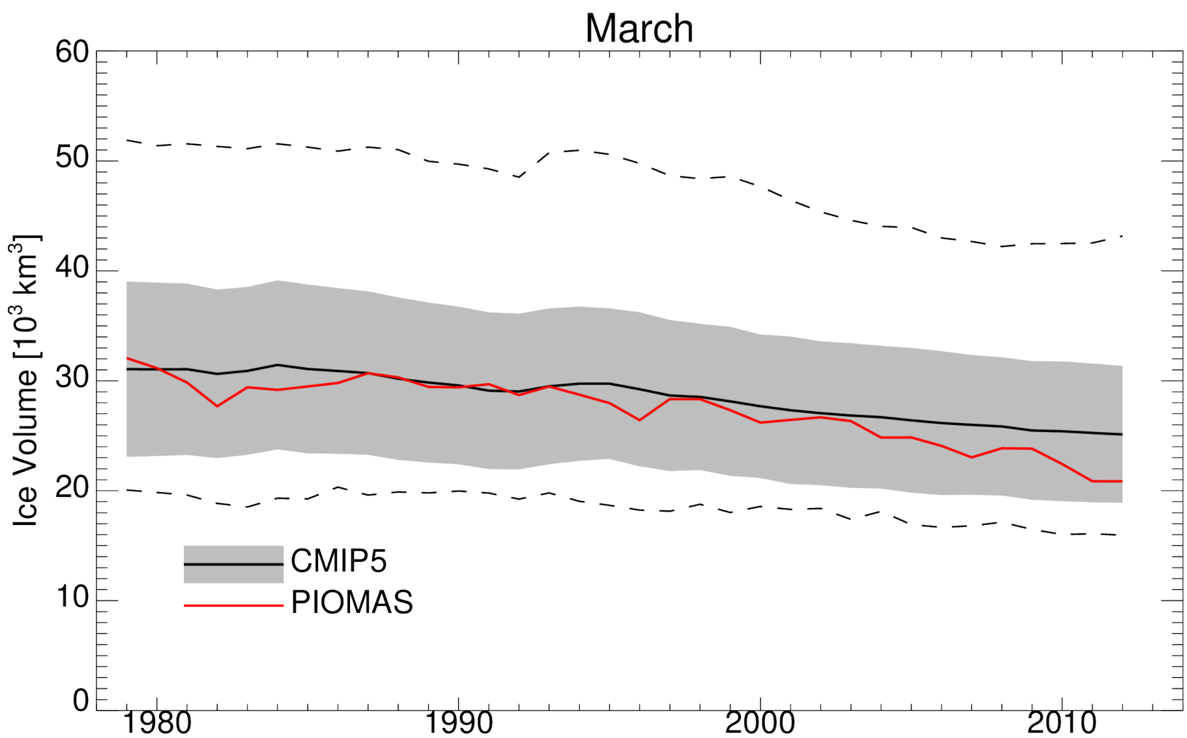


Figure 7. Change in Arctic sea ice volume as shown from the CMIP5 ensemble and from PIOMAS for the period 1979 to 2012, for March. Grey shading shows the ±1 standard deviation of CMIP5 ensemble. Upper and lower pecked lines show maximum and minimum ice volume of the model ensemble. Multi-model ensemble mean ice volume is shown as the black line.

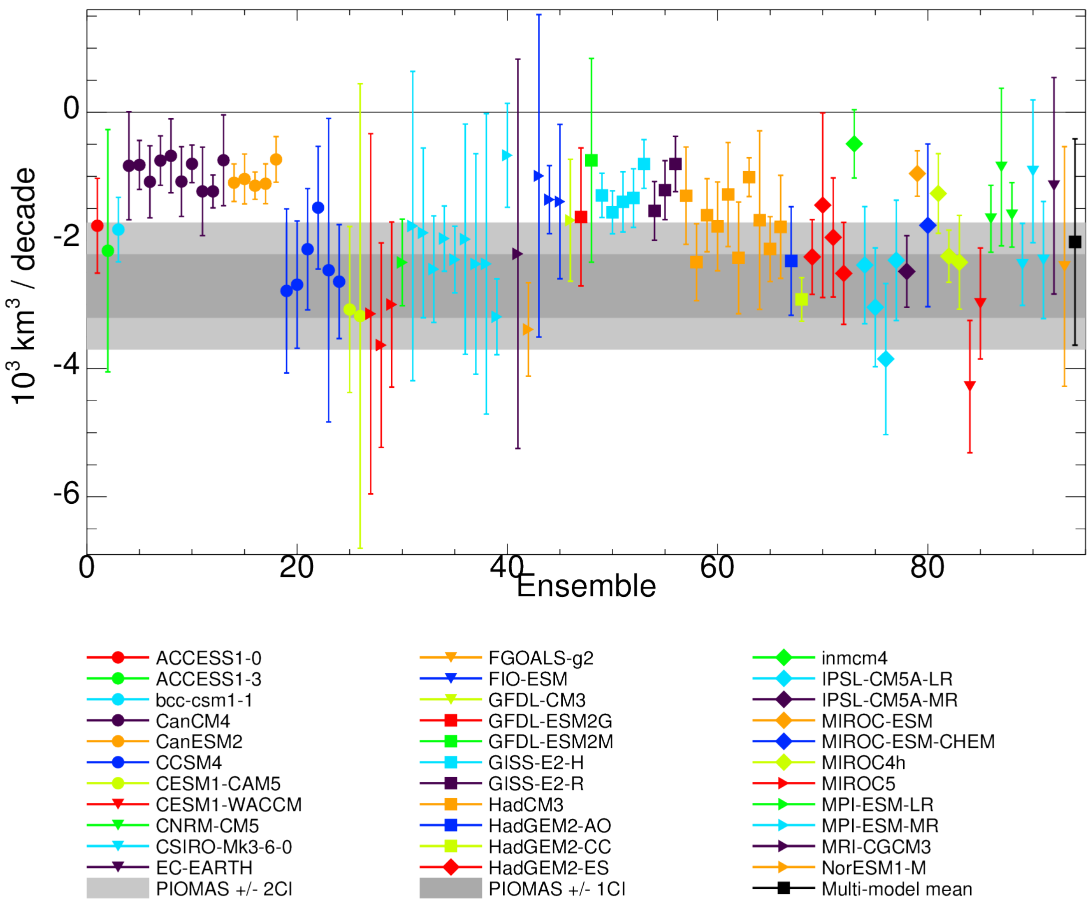


Figure 8. March ice volume trends from 1979 to 2013 for all 92 individual CMIP5 model ensembles as well as the multi-model ensemble mean (shown in black) with confidence intervals (vertical lines). The 1σ and 2σ confidence intervals of PIOMAS trends are shown in dark gray shading (1σ) and light gray shading (2σ).