



# Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations

R. Lindsay and A. Schweiger

Polar Science Center, Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, WA 98105, USA

Correspondence to: R. Lindsay (lindsay@apl.uw.edu)

Received: 6 August 2014 – Published in The Cryosphere Discuss.: 28 August 2014

Revised: 30 December 2014 – Accepted: 11 January 2015 – Published: 10 February 2015

**Abstract.** Sea ice thickness is a fundamental climate state variable that provides an integrated measure of changes in the high-latitude energy balance. However, observations of mean ice thickness have been sparse in time and space, making the construction of observation-based time series difficult. Moreover, different groups use a variety of methods and processing procedures to measure ice thickness, and each observational source likely has different and poorly characterized measurement and sampling errors. Observational sources used in this study include upward-looking sonars mounted on submarines or moorings, electromagnetic sensors on helicopters or aircraft, and lidar or radar altimeters on airplanes or satellites. Here we use a curve-fitting approach to determine the large-scale spatial and temporal variability of the ice thickness as well as the mean differences between the observation systems, using over 3000 estimates of the ice thickness. The thickness estimates are measured over spatial scales of approximately 50 km or time scales of 1 month, and the primary time period analyzed is 2000–2012 when the modern mix of observations is available. Good agreement is found between five of the systems, within 0.15 m, while systematic differences of up to 0.5 m are found for three others compared to the five. The trend in annual mean ice thickness over the Arctic Basin is  $-0.58 \pm 0.07$  m decade<sup>-1</sup> over the period 2000–2012. Applying our method to the period 1975–2012 for the central Arctic Basin where we have sufficient data (the SCICEX box), we find that the annual mean ice thickness has decreased from 3.59 m in 1975 to 1.25 m in 2012, a 65 % reduction. This is nearly double the 36 % decline reported by an earlier study. These results provide additional direct observational evidence of substantial sea ice losses found in model analyses.

## 1 Introduction

In recent years great interest has developed in the changes seen in Arctic sea ice as ice extent and volume have markedly decreased. While ice extent is reasonably well observed by satellites, observations of ice thickness have been, until recently, sparse. Sea ice model reanalyses (e.g., Schweiger et al., 2011) provide useful estimates of thickness and volume loss but so far do not directly incorporate observations of ice thickness. An observational record that does not depend on a sea ice model therefore remains of substantial interest. Historically, a great number of ice thickness measurements have been made at specific locations using drill holes or ground-based electromagnetic methods; however, these point measurements are difficult to translate into area-averaged mean ice thickness because of the highly heterogeneous nature of the ice pack. Estimates of mean ice thickness require a large number of independent samples. In the last 10 years or so a number of different observations of mean sea ice thickness have been made available by different groups using a variety of different methods. The longest historical record is from sporadic observations made by submarines using upward-looking sonar (ULS) to measure ice draft (Rothrock et al., 1999, 2008). These measurements are currently available starting in 1975 and ending in 2005 and include data from 34 cruises. They have broad but incomplete spatial coverage and limited sampling of the seasonal variations. ULS measurements from anchored moorings have been made by a number of different groups (e.g., Vinje et al., 1998; Melling et al., 2005; Krishfield et al., 2014; Hansen et al., 2013). Each has excellent temporal sampling with record lengths of up to 10 years although only for single locations. More recently, airborne and satellite-based observations have become available. Operation IceBridge uses lidar and radar

technology on a fixed-wing aircraft beginning in 2009 (Kurtz et al., 2012) and electromagnetic methods from helicopters have been used to measure the snow plus ice thickness since 2001 (Pfaffling et al., 2007; Haas et al., 2009). Satellite-based lidar techniques began with ICESat during the years 2003–2008 (Kwok et al., 2009; Yi and Zwally, 2009). Radar altimeter techniques are used with data from Envisat (2002–2012; Peacock and Laxon, 2004) and from CryoSat-2 beginning in 2010 (Laxon et al., 2013; Kurtz et al., 2014). However, Envisat and CryoSat-2 estimates are not included in the current study because there are currently few publicly available ice thickness data from these instruments that are not preliminary products.

Observations from submarine ULS instruments have previously been used to establish the time and space variation of sea ice draft using a curve-fitting approach for a limited area of the Arctic Basin (Rothrock et al., 2008). Here we extend this approach by including more recent observations of ice thickness from multiple sources, including satellites, and expand the area to the entire Arctic Basin. In addition, we examine if there are systematic differences between individual data sources. This is important because the data sources differ markedly in their methodologies and sampling characteristics, which may result in systematic errors that can affect the spatial and temporal characteristics of the ice thickness time series.

Differences in mean ice thickness from the various measuring systems vary on a wide range of temporal and spatial scales and even measurements obtained from samples nearly identical in time and space may show differences depending on sampling error, how the measurement is made, and how the systems record small-scale variability. The differences in the results from different measurement systems may also depend on ice type (first-year or multiyear), degree of deformation, ice thickness, snow depth, or season. This study is a first attempt to characterize these differences for a broad range of observing systems with a single number that characterizes the difference between any two observing systems.

## Approach

All available ice thickness observations are fit with a multiple regression least-squares solution of an expression for the mean ice thickness that is a function of time and space. The expression includes non-linear terms that characterize the spatial and temporal variability as well as terms that indicate which observation system is associated with each observation. The observations can be restricted to particular observation systems, geographic regions, or time periods to refine the analysis, with the trade-off of the results being less general. We begin the analysis with a basin-wide selection of all available observations for the time period 2000–2012, then focus on specific observation systems or regions. The trend in the mean ice thickness determined by the regression expression is compared to model-based estimates and

other observational studies. We then expand this analysis to include data back to 1975 to compare with and update the results of Rothrock et al. (2008) and provide an assessment of the 39-year change in ice thickness for the central Arctic Basin from the observational record. An assessment of errors, including sensitivity analyses that examine the role of individual observing systems and focus on subregions of the Arctic, follows.

## 2 Data

The Unified Sea Ice Thickness Climate Data Record (Sea Ice CDR) is a collection of Arctic sea ice draft, freeboard, and thickness observations from many different sources. It includes data from moored and submarine-based upward-looking sonar instruments, airborne electromagnetic (EM) induction instruments, satellite laser altimeters (ICESat), and airborne laser altimeters (IceBridge). The point observations have been averaged spatially for roughly 50 km and temporally for 1 month. The mooring data are averaged only in time, the submarine data only in space, and the airborne and satellite data are averaged both temporally (1 month) and spatially (50 km); e.g., airborne data from one campaign that are taken a few days apart are averaged together. In all data sets except ICESat-J, open water is included in the mean ice thickness estimates. The mean measurements and the probability distributions for all of the sources are collected in a single data set with uniform formatting, allowing the scientific community to better utilize what is now a considerable body of observations. The Sea Ice CDR data are available at the National Snow and Ice Data Center (Lindsay, 2010, 2013; also at <http://psc.apl.washington.edu/seaiccdr>). The data sets used in this study are listed in Table 1 and maps of the data locations and times of the observations from the various systems are shown in Figs. 1 and 2. A short description of the eight different data sets follows.

- *Submarines*: ULS instruments have been deployed on US Navy submarines using either digital or analog recording methods (Polar Science Center, University of Washington; US Navy Arctic Submarine Laboratory; Cold Regions Research and Engineering Laboratory; NSIDC, 1998; Rothrock and Wensnahan, 2007; Wensnahan and Rothrock, 2005; Tucker et al., 2001). The point data are archived at NSIDC. While there are 34 cruises archived for the years 1975–2005, only three are from after 2000: one in 2000 and two in 2005. The draft measured by the ULS instruments is based on the first-return echo. This introduces a positive mean bias in the measured draft that is estimated by Rothrock and Wensnahan (2007, RW07 hereinafter) as  $0.44 \pm 0.09$  m for multiyear ice and typical US submarine depths and beam widths, based on work by Vinje et al. (1998). RW07 also identify an open-water detection bias of  $-0.15 \pm 0.08$  m. Combined, the draft measure-

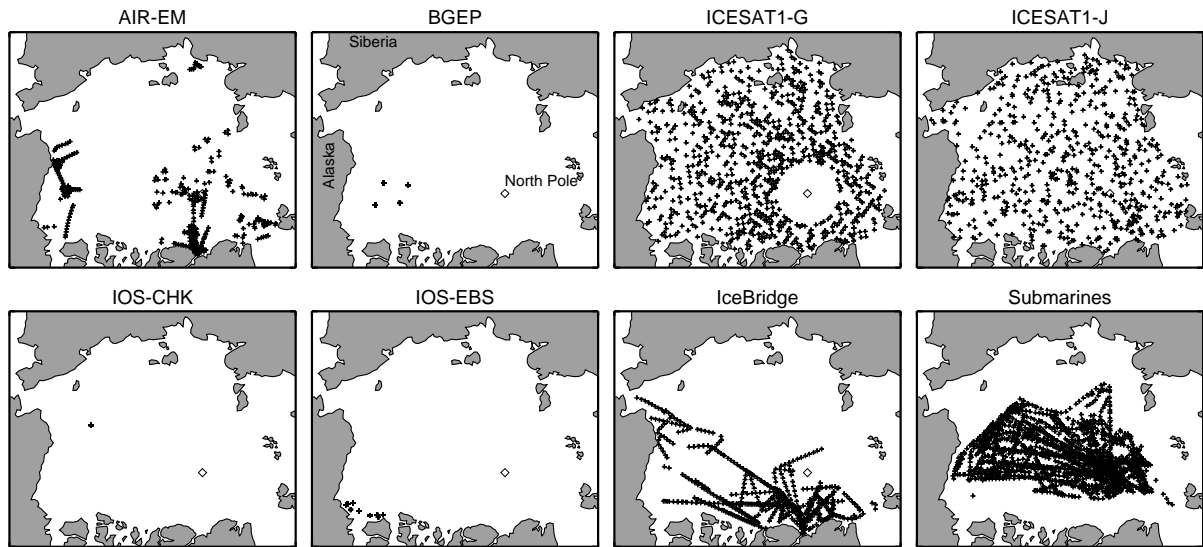


Figure 1. Locations of the observations from different data sources.

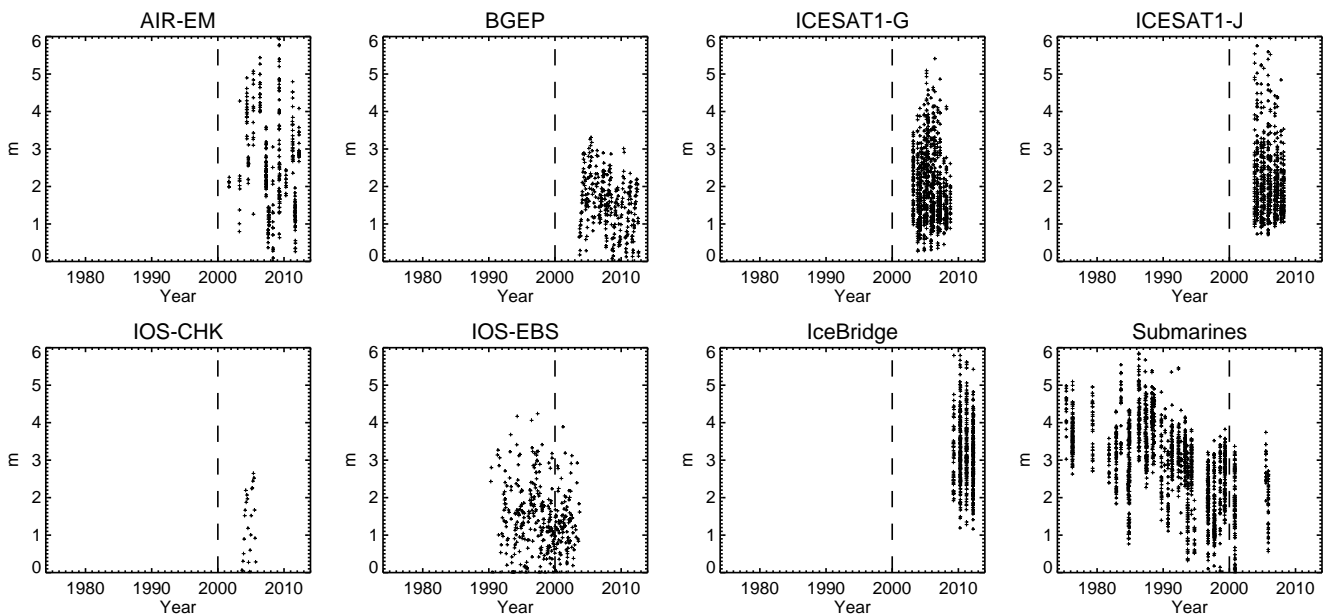


Figure 2. Times and ice thickness of the observations from different data sources. The primary focus is on the years after 2000 (dashed line).

ments reported for the submarines have a likely bias of  $0.29 \pm 0.25$  m. The error range includes the error contributions from other unbiased sources of error (RW07). We have subtracted this bias from the US submarine draft data but not from any of the other ice draft measurements; the bias for these measurement types is unknown and will be accounted for in the multiple regression procedure.

- *Air-EM, Airborne Electromagnetic Induction*: the Air-EM measurements include an electromagnetic induction instrument that determines the distance to the ice-

water interface and a lidar to measure the distance to the top snow surface; consequently the measurements are of the ice + snow thickness. The method is based on measurements of the amplitude and phase of a secondary EM field induced in the water by a primary field transmitted from the EM instrument. Haas et al. (2009) report on the configuration of the EM instruments and give an accuracy of 0.1 m for the ice + snow thickness over level ice. The footprint of the Air-EM system is 40–50 m at common operational altitudes, and as a consequence the thickness of pressure ridges are

Table 1. Observational data sets.

Short name	Long name	Years	Location	Parameter/ instrument
Submarines	US Navy submarines	1975–2005	Arctic Basin	Draft/submarine ULS
BGEP	Beaufort Gyre Exploration Project	2003–2012	Beaufort Sea	Draft/moored ULS
IOS-EBS	Institute of Ocean Sciences	1990–2003	Eastern Beaufort Sea	Draft/moored ULS
IOS-CHK	Institute of Ocean Sciences	2003–2005	Chukchi Sea	Draft/moored ULS
Air-EM	Airborne EM	2001–2012	Arctic Basin	Ice + snow thick/airborne EM
ICESat-G	NASA ICESat–Goddard	2003–2008	Arctic Basin	Ice thickness/satellite lidar
ICESat-J	NASA ICESat–JPL	2003–2008	Arctic Basin	Ice thickness/satellite lidar
IceBridge	NASA Operation IceBridge	2009–2012	western Arctic Basin	Ice thickness/airborne lidar

smoothed and underestimated by as much as 50 % (Haas and Jochmann, 2003; Pfaffling and Reid, 2009). Pfaffling et al. (2007) report mean errors compared to drill holes of the ice + snow thickness of  $-0.04 \pm 0.09$  m over approximately 200 m of level ice in Antarctica. Haas et al. (2010) report that the thickness distributions obtained from the instruments are most accurate with respect to their modal thickness and less so for the mean thickness. Ice + snow thickness samples used here are obtained from various locations around the Arctic Basin (Alfred Wegener Institute for Polar and Marine Research and York University; Haas et al., 2009; Pfaffling et al., 2007). In order to obtain an estimate of the ice thickness alone, the snow depth must be subtracted. The snow depth used here is the mean snow depth estimated from the PIOMAS ice–ocean model, which estimates snow accumulation from the NCEP Reanalysis (Zhang and Rothrock, 2003). The uncertainty in the snow depth from PIOMAS is not well known. Compared to the Warren et al. (1999) climatology of snow depth, it averages between 1 cm greater in May to 7 cm less in October. However, it potentially offers better spatial and interannual variability than using a climatology which may not provide the best estimate for more recent years (Kurtz et al., 2013; Webster et al., 2014). We estimate the uncertainty in the PIOMAS snow depth to be on the order of 0.10 m.

- *BGEP, Beaufort Gyre Exploration Project*: this data set is comprised of a set of three or four (depending on the year) bottom-anchored moorings with top-mounted ULS instruments located in the Beaufort Sea (Woods Hole Oceanographic Institute; Krishfield et al., 2014). These installations use the ASL acoustic Ice Profiler moored at a depth of approximately 50 m below the surface. The Ice Profiler is a 420 kHz ULS instrument with a  $1.8^\circ$  beam width, a precision of 0.05 m, and a sample rate of 2 s. There are a total of 28 station years of data from 2003 to 2012. The data processing procedures are outlined in Krishfield and Proshutinsky (2006) and the point data are available at <http://www.who.edu/page.do?pid=66566>.

The uncertainty in the point ice draft estimates are estimated to be better than 0.10 m (Krishfield et al., 2014).

- *IceBridge, NASA Operation IceBridge*: scanning lidar altimeter, snow radar, and cameras aboard NASA aircraft are used to determine the surface freeboard and snow depth from an altitude of approximately 300 m. These data are then used to determine the ice thickness distribution (Goddard Space Flight Center; Kurtz et al., 2012, 2013; Richter-Menge and Farrell, 2013). The IceBridge mission was initialized after the end of operations of the ICESat-1 satellite in order to partially continue the time series of sea ice and ice sheet observations until the launch of ICESat-2. The data are available at NSIDC (Kurtz et al., 2012) and are provided along the aircraft track at a spacing of 40 m. An estimate of the error is included for each point and is primarily a function of the distance to a lead where the ocean water level needed to compute the freeboard can be determined. The uncertainty in the estimated snow depth is critical because in the freeboard–thickness relationship it is amplified into an ice thickness uncertainty roughly 7 times as large (Kwok and Cunningham, 2008). The mean snow depth uncertainty is not yet well characterized but Kurtz et al. (2013) estimate it as 0.06 m for point estimates. There may also be unknown biases in the snow depth estimates. The data for each spring campaign are aggregated into 50 km samples, combining data from different flight days if they are in close proximity. Points with a thickness uncertainty greater than  $1.0 + 0.25 h$  or 2.0 m, where  $h$  is the ice thickness, are excluded.
- *IOS-CHK, Institute of Ocean Sciences Chukchi Sea*: these are bottom-anchored moorings with ULS instruments located in the Chukchi Sea (Institute of Ocean Sciences; Melling and Riedel, 2008). These moorings also use the ASL acoustic Ice Profiler. Just 2 station years are available, starting in 2003. The measured draft uncertainty is estimated to be 0.10 m (Melling and Riedel, 2008).

- *IOS-EBS, Institute of Ocean Sciences, Eastern Beaufort Sea*: this collection includes data from bottom-anchored moorings with ULS instruments located near the coast at nine different locations in the eastern Beaufort Sea near the Mackenzie River delta and Banks Island (Institute of Ocean Sciences; Melling et al., 2005). The data are available at NSIDC (Melling and Riedel, 2008). We use data from 1990 to 2003. The moorings use various models of the ASL acoustic Ice Profiler. The ice draft uncertainty for point measurements is about 0.10 m (Melling and Riedel, 2008).
- *ICESat-G, ICESat measurements processed by NASA Goddard Space Flight Center*: satellite laser altimeter measurements of freeboard are used to compute ice thickness (Yi and Zwally, 2009; Zwally et al., 2008). Snow depth is from climatology (Warren et al., 1999). Snow density, including its time variation, is based on Kwok and Cunningham (2008). Fifteen 1-month measurement campaigns are included in this data set. The track data of position and ice thickness have a resolution of about 170 m in the along-track direction. Portions of track data from each campaign are aggregated to form nearly 30 000 50 km mean ice thickness samples. In order to not overly fit the multiple regression procedure to the satellite data, 900 randomly selected samples from the Arctic Basin from all campaigns are used. This accounts for the high spatial autocorrelation of these data and makes the ICESat data have roughly the same number of points as the submarine data. The autocorrelation length scale of the residuals from the regression procedure of the ICESat-G aggregated samples is about 300 km. There are no published estimates of the expected ice thickness errors for this system.
- *ICESat-J, ICESat measurements processed by the Jet Propulsion Laboratory*: these data use different processing methods from ICESat-G and cover just 10 measurement campaigns. In particular the methods of determining the freeboard and the snow depths are different (Kwok et al., 2009). Snow depth was estimated from daily snow accumulation data from the ECMWF Reanalysis. The data gap at the pole due to the satellite orbital configuration is filled by interpolation. Kwok and Cunningham (2008) find the overall uncertainty of ice thickness estimates within 25 km track segments is  $\sim 0.7$  m but varies with the total freeboard and the snow depth. In a second study, Kwok et al. (2009) find their ICESat estimates of ice draft are  $0.1 \pm 0.42$  m thinner than those from a submarine cruise in 2005. For this gridded data set there is no accounting for the overall ice concentration within a grid cell after data accumulation and interpolation. A weighting by passive-microwave-derived ice concentration to address this is sometimes applied to this data set (e.g., Kwok and Cunningham, 2008; Schweiger et al., 2011; Laxon et al., 2013), but

this adjustment is not made here. Weighting by ice concentration reduces the average ICESat-J ice thickness by just 0.05 m in October/November and 0.02 m in February/March. The data are provided on a 25 km grid for each 1-month campaign, but they have been aggregated here to a 50 km grid to make them compatible with the other data sets. Similar to the ICESat-G data, a subsample of 600 randomly selected points from all campaigns (proportional to the number of measurement campaigns available) is used in order to account for the high spatial autocorrelation (also about 300 km) of these data. This data set is not in the Sea Ice CDR but may be obtained from JPL (<http://rkwok.jpl.nasa.gov/icesat/index.html>).

The submarine and mooring observations of ice draft are converted to ice thickness following Rothrock et al. (2008) using a density of water of  $1027 \text{ kg m}^{-3}$ , a density of ice of  $928 \text{ kg m}^{-3}$ , and the weight of the snow. The ice thickness  $h$  is then related to the ice draft  $D$  by

$$h = 1.107D - f(m), \quad (1)$$

where  $f(m)$  is the monthly mean ice equivalent of the snow on the surface. We use the monthly values of  $f(m)$  determined by Rothrock et al. (2008, RPW08 hereafter), who found  $f(m)$  ranges up to 0.12 m in May based on the snow climatology of Warren et al. (1999) for multiyear ice. First-year ice may have substantially less snow than multiyear ice (Kurtz et al., 2013) but, because the total snow accumulation depends on freeze-up dates, this difference is likely to be variable and difficult to estimate. This uncertainty in the snow depth plus some uncertainty in the density of the ice add to the uncertainty of the conversion of ice draft to ice thickness.

We have little information on the absolute accuracy of the averaged samples because we do not know the degree to which the reported measurement errors are uncorrelated. Clearly if the errors are uncorrelated, the many thousands of point observations that typically comprise a sample would result in very small sample errors (Kwok et al., 2008). However, this assumption is unrealistic (Kwok et al., 2009) since the sea ice characteristics that affect these errors (e.g., thickness variability, snow cover, ridging) likely have spatial autocorrelations substantially larger than the distance between samples (Zygmuntovska et al., 2014).

### 3 Methodology

Following RPW08, who developed a regression model to fit ice draft observations from US submarine data for a sub-area of the Arctic Basin, a smooth function of space and time,  $h(x, y, t)$ , is fit to all of the selected observations simultaneously using a least-squares multiple linear regression procedure. We refer to this as the *Ice Thickness Regression Procedure*, or ITRP. This function can be evaluated at all locations and times to yield a complete time and space record of

Arctic Basin ice thickness. However, an additional complication, the fact that different observation systems may have unknown biases relative to each other, needs to be accounted for. In order to do this, an indicator variable  $I$  is included for each observation system in the multiple regression procedure except for the reference system.  $I$  is 1 for observations from the corresponding source and 0 otherwise. The regression equation becomes ill posed if all systems have an associated indicator, so one of the observations systems needs to be excluded and therefore implicitly becomes the reference system. We chose the ICESat-G data as a reference data set in this study because of ICESat-G's extensive spatial and temporal coverage, but we emphasize that this does not mean it is assumed to be more accurate than the other systems. The choice of the reference does not change the form or the goodness of fit of the regression equation or the relative magnitudes of the indicator variable coefficients. However, it does help determine the constant  $a_0$  so that predictions where there are no data (all  $I = 0$ ) depend on this choice and we need to reexamine the choice after the fit is made. The regression equation for the ice thickness is

$$h(x, y, t) = a_0 + \sum a_i T_i(x, y, t) + \sum b_j I_j + \text{error}, \quad (2)$$

where  $T_i(x, y, t)$  are the spatial and temporal terms of the regression equation,  $I_j$  are the indicator variables for each of the observation systems (excluding the reference), and "error" is the residual of the fit. Positive coefficients  $b_j$  for the indicator variables  $I_j$  of a particular observation system indicate that the error in the regression is reduced if a constant value (the coefficient  $b_j$ ) is added to the regression expression (not to the observations) for all observations from that system, so positive coefficients indicate that measurements from the system are systematically thicker relative to the reference measurements. Different observations are not weighted by their uncertainties because the uncertainties of the time and space averaged observations are unknown.

The choice of terms in the regression follows the methods of RPW08. The spatial coordinate system  $x, y$  is based on a Cartesian grid in units of 1000 km and the time coordinate  $t$  is in years relative to 2000. Spatial and temporal terms are included in sequence in a forward selection procedure, starting with the one that is most correlated with the observed thickness. Additional terms are then added one-by-one and at each step the variable that is most correlated with the residuals is added to the list of terms. Terms considered for the expression are up to third order in space and time, including mixed terms involving both space and time. The seasonal cycle of the thickness is estimated by including  $\text{COS} = \cos(2\pi \text{ year-fraction})$  and  $\text{SIN} = \sin(2\pi \text{ year-fraction})$  as the first harmonic of the annual period. The second and third harmonics ( $\text{COS2}$ ,  $\text{SIN2}$ ,  $\text{COS3}$ , and  $\text{SIN3}$ ) are also included. The linear time variable is introduced before the quadratic and all sine and cosine seasonal terms are always included. The partial- $p$  values of all coefficients are assessed at each step and any term with a value less than 0.90 is dropped unless it is one

of the indicator functions or SIN or COS. The procedure is stopped when a new coefficient has a partial  $p$  value of less than 0.90.

The multiple regression procedure provides an estimate of the standard error of each of the coefficients:  $\sigma_i$  for the space and time terms or  $\sigma_j$  for the indicator terms. For the reference source we say the coefficient is zero and the standard error is taken as the standard error of the constant term  $a_0$ . Without the indicator variables, the RMS error of the fit for the Arctic Basin increases slightly from 0.62 to 0.64 m and the RMS difference in the fit values at the data locations is 0.20 m, indicating that these variables play a minor role in determining the shape of the regression function while at the same time providing an estimate of the relative bias of the different observational data sets.

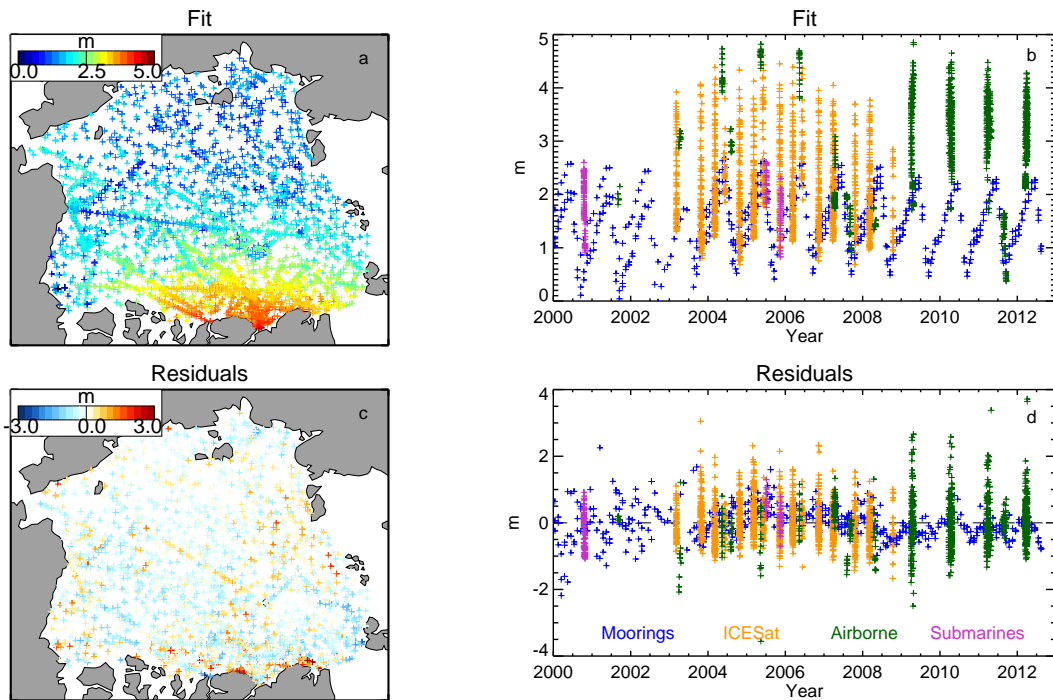
## 4 Results

### 4.1 Fit for the Arctic Basin

For the entire Arctic Basin, 2000–2012, the ITRP outlined above selected 21 terms: 7 for indicator variables and 14 for time and space variability of the ice thickness. Table 2 shows all of the terms and coefficients for this fit. The multiple regression coefficient is  $R_{\text{mul}} = 0.84$  ( $R_{\text{mul}}^2 = 0.70$ ) and the RMS error of the fit is 0.62 m. Summaries of the values of the fit predictions at the time and location of the observations and the residuals for the fit are depicted in Fig. 3. The observations are grouped into four types for this figure: submarines, moorings, aircraft, and satellite. The scatter in the temporal plot for the predictions is due to the spatial distribution of the observations. This mixture of time and space variability is also seen in the maps. The residuals have little temporal or spatial structure, as we would expect, because the terms have been selected to largely account for the systematic spatial and temporal variability.

### 4.2 Systematic differences between ice thickness estimates

As a step towards generating a time series of sea ice thickness from observations alone, we need to determine what, if any, the mean differences are between the ice thickness estimates from the different measurement systems. The ITRP provides a method to do this even when the observations are not coincident. In this analysis the observation sources with indicator coefficients not significantly different from zero are Air-EM, BGEP, IOS-CHK, ICESat-G, and the submarines, indicating that these sources are all consistent in the mean with each other over the region and period analyzed. There is just a 0.11 m spread in the mean between the five systems. Ice thickness data from the three submarine cruises agree in the mean with the ICESat-G data very closely, with a bias coefficient of  $-0.05 \pm 0.06$  m (error brackets are 1 standard deviation). Two indicator coefficients are significantly different



**Figure 3.** Fit to ice thickness observation data from the Arctic Basin for 2000–2012. (a) Map of ice thickness of the fit predictions at the data locations regardless of time; (b) the fit predictions at the data times regardless of location; (c) map of the residuals; (d) residuals as a function of time. The observational sources are grouped into four different types and color-coded as shown in (d).

from zero: ICESAT-J and IceBridge. This means they are significantly larger or smaller than the reference data set and, in this case, from the cluster of five observation sets that agree with each other.

The ICESat-J coefficient, 0.42, indicates that on average the JPL thickness product is 0.42 m thicker than the Goddard product. A small portion of this difference is due to the lack of inclusion of open water in the ice thickness estimates but the bulk of the difference between the ICESat-G and ICESat-J values may be related to the different techniques of determining the sea level in order to obtain the freeboard and the different methods for estimating snow depth. The ITRP shows the ICESat-J estimates are on average 0.47 m thicker than the submarine-based estimates. In contrast, Kwok et al. (2009) found that the ICESat track estimates of ice draft were  $0.1 \text{ m} \pm 0.4 \text{ m}$  thinner than the fall 2005 submarine ice draft data.

The estimation of the submarine coefficient is sensitive to the inclusion of a particular cruise. The large difference between the submarines and the ICESat-J estimates for the entire basin stems from the inclusion of the 2000 submarine cruise when there is no overlap with the ICESat data. If the analysis period is chosen as 2001–2012 with all sources included, the ICESat-J product is found to be just  $0.05 \text{ m} \pm 0.09 \text{ m}$  thicker than the submarine-based estimates and in line with the ICESat-J validation results reported by Kwok et al. (2009). The very sparse submarine data do not

provide a robust estimate of their mean bias relative to the other measurements.

The IceBridge data are also significantly thicker than the reference data, in this case by  $0.59 \pm 0.06 \text{ m}$ , and hence also thicker than the submarine, BGEP, IOS-CHK, and Air-EM data. We will examine the IceBridge and Air-EM data sets below to show that this large difference is robust. The IOS-EBS data are estimated to be  $0.20 \pm 0.10 \text{ m}$  thinner than the reference. However, we have less confidence in this result since the IOS-EBS moorings are near the coast in the extreme southeast corner of the Beaufort Sea and may not be well represented by the spatial terms of the regression model. Further discussion of the uncertainties of the indicator coefficients is found in the error assessments section.

### 4.3 Evaluation of ice thickness trends

#### 4.3.1 Arctic Basin for 2000–2012

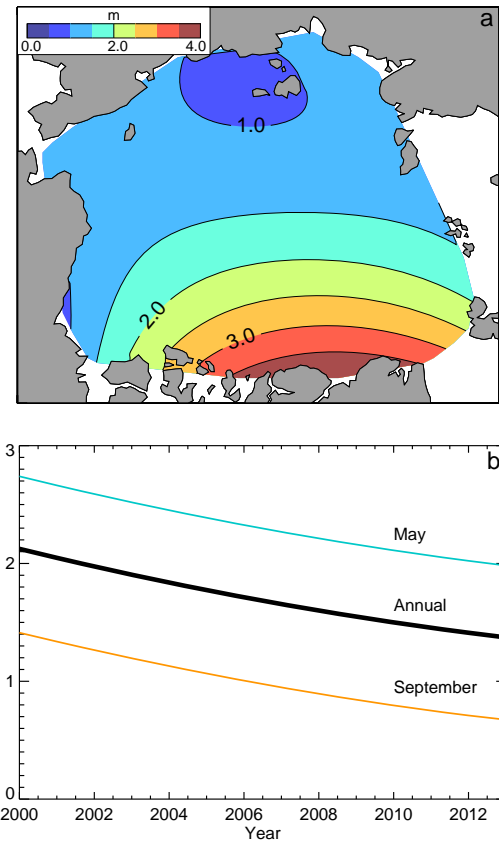
The ITRP expression for the whole basin can be used to evaluate the spatial and temporal patterns of ice thickness change. To do this, the expression was evaluated at every location within the basin on a 40 km grid with all of the indicator variables set to zero. Here it is important to reconsider the choice of the reference system, ICESat-G. Table 2 shows that the ICESAT-G coefficient, zero by its selection as the reference, is very close to the median value of the coefficients of the cluster of five observation systems that have quite sim-

**Table 2.** ITRP coefficients for the Arctic Basin for all observational sources, 2000–2013. Sigma is the standard error of the coefficient and the  $p$  value is the probability of being non-zero. The  $X$  and  $Y$  spatial coordinates are oriented as in the map in Fig. 4 and are in units of 1000 km. The time  $T$  is in years relative to 2000. The indicator coefficients are ordered by the magnitude of the coefficients.

Term	Coefficient	Sigma	$p$ value
Indicator variables ( $b_j$ in Eq. 2)			
IOS-EBS	−0.204	0.103	0.000
Submarine	−0.049	0.061	0.000
BGEP	−0.045	0.058	0.000
IOS-CHK	−0.007	0.130	0.000
ICESat-G	0.000	0.066	0.000
Air-EM	0.063	0.061	0.000
ICESat-J	0.420	0.034	1.000
IceBridge	0.590	0.057	1.000
Time and space variables ( $a_i$ in Eq. 2)			
$T$	−0.079	0.007	1.000
COS	−0.233	0.032	1.000
SIN	0.296	0.024	1.000
COS2	0.162	0.028	0.953
SIN2	−0.226	0.021	1.000
COS3	−0.140	0.030	0.582
SIN3	0.015	0.025	0.000
$Y$	−1.767	0.038	1.000
$X^2$	−0.329	0.017	1.000
$XY^2$	0.253	0.040	0.991
XSIN	−0.199	0.019	1.000
$Y^2$	0.674	0.037	1.000
$X^2Y$	0.398	0.024	1.000
$XT^2$	−0.002	0.000	1.000

ilar coefficients: submarines, BGEP, IOS-CHK, ICESat-G, and Air-EM. These systems have a range of coefficients of 0.11 m, indicating that when spatial and temporal variability is accounted for there is little mean difference in the observations. The coefficients for these five are not significantly different from each other since the sigma values are between 0.06 and 0.13 m (Table 2). This suggests that using ICESat-G as a reference predicts an ice thickness that is consistent with observations from these five systems but not with the unadjusted observations from IOS-EBS, ICESat-J, or IceBridge.

The mean ice thickness for the 2000–2012 period is shown in Fig. 4. The map shows a maximum along the Canadian coast and a minimum in the vicinity of the New Siberian Islands. The ITRP annual mean basin-average ice thickness has declined from 2.12 to 1.41 m (34 %) with a linear trend of  $-0.58 \pm 0.07$  m decade $^{-1}$ . A quadratic time term in the fit,  $xT^2$  (Table 2), creates a slight curvature in the basin-wide mean thickness seen in Fig. 4. The September thickness has declined from 1.41 to 0.71 m (50 %). This observationally based trend can be compared to that of an ice–ocean model



**Figure 4.** (a) Mean annual ice thickness from the ITRP for the period 2000–2012. (b) Mean ice thickness for the Arctic Basin in May, in September, and for the annual mean.

commonly used for ice volume estimates. The PIOMAS model (Version 2.1, Zhang and Rothrock, 2003) has an annual mean thickness trend of  $-0.60 \pm 0.04$  m decade $^{-1}$  for the same area and time period, and thus its trend is quite consistent with that of the observations. In another observational study, Laxon et al. (2013) computed the ice volume in the Arctic Basin from CryoSat-2 data for 2 years, 2010 and 2011, and computed volume trends by concatenating the ICESAT-J estimates to compute a trend from 2003 to 2011. They found a thickness trend for fall and spring of  $0.75$  m decade $^{-1}$ . A recent study of ice thickness measurements in Fram Strait using both surface-based and helicopter-based EM methods (Renner et al., 2014) also found a decline in the mean thickness. They found a decrease of  $2.0$  m decade $^{-1}$  in late summer for the period 2003–2012, a decline of over 50 %, for ice exiting the Arctic Basin.

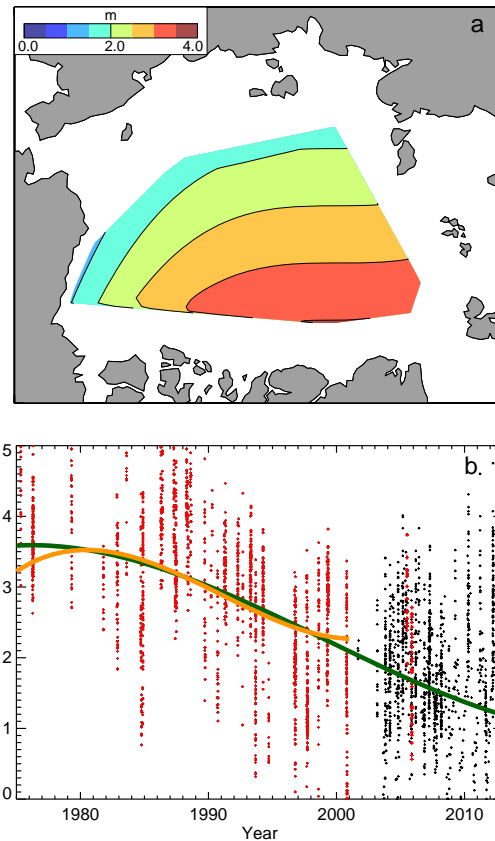
#### 4.3.2 SCICEX box for 1975–2012

The regression analysis of RPW08 concentrated on submarine ice draft data from 1975 to 2000 within the SCICEX box. They determined that the best fit included terms up to fifth order in space and up to third order in time. The fit



showed a maximum in 1980 followed by a steep decline and then a leveling off at the end of the period. Kwok and Rothrock (2009) used 5 years of ICESat data to analyze the fall and winter changes in the ice draft for an additional 5 years, to 2008; however, their regression procedure did not take advantage of the spatial information in the ICESat data but simply concatenated submarine and satellite records. They found the ICESat data showed an additional modest thinning. In order to estimate the temporal variation of ice thickness from 1975 to 2013 and to compare our results to those of RPW08, the ITRP is extended back to 1975 in this region. The fit procedure was performed using all of the data available from all sources that fall within the box, 3017 observations in all. Figure 5 shows the third-order fit from this study and the third-order curve from RPW08 that is computed for the years 1975–2001. The ITRP fit includes indicator variables as before and 12 additional terms:  $T$ ,  $T^3$ ,  $X^3$ ,  $Y$ ,  $\text{COS}$ ,  $\text{SIN}$ ,  $\text{COS2}$ ,  $\text{SIN2}$ ,  $\text{COS3}$ ,  $\text{SIN3}$ ,  $X*\text{SIN}$ , and  $T*\text{SIN2}$ . It explains 80% of the variance and the RMS error is 0.49 m, while the fit in RPW08 study explained 79% of the variance and has an RMS error of 0.49 m as well, so the two are very similar in the fit properties. With an additional 13 years of data it is apparent that the annual mean ice thickness in the central Arctic Basin has continued to decline at an approximately linear rate and the short leveling off at the end of the RPW08 and Kwok and Rothrock (2009) time periods did not persist. We find that the annual mean ice thickness for the SCICEX box has thinned from 3.59 m in 1975 to 1.25 m in 2012, a 65% decline. This is nearly double the decline reported by RPW08, 36%, for the period ending in 2000. In September the mean ice thickness has thinned from 3.01 to 0.44 m, an 85% decline. The linear trend of the annual average thickness over this period is  $-0.69 \pm 0.03 \text{ m decade}^{-1}$ . This is double the rate of ice thickness loss computed from PIOMAS for the same area for the period 1979–2012,  $-0.34 \text{ m decade}^{-1}$ , showing that for the central Arctic Basin and for the longer time period the PIOMAS trend in ice volume is too conservative, as also shown by Schweiger et al. (2011). This is in contrast to the good match for the trends from PIOMAS and the ITRP we found for the whole basin for just the most recent 13 years.

The difference in the trends between the observations and the model for the 1979–2012 period may possibly be due in part to a time-varying bias of the submarine observations. The early part of the record has much thicker ice in this region than the later part. The thicker ice has much larger variability in the ice draft and hence the bias related to the first-return correction (see also below) may be much larger for the earlier thicker ice. If this is the case, the early ice thickness is overestimated by the draft measurements and the magnitude of the ice thickness trend is smaller than estimated here.



**Figure 5.** (a) Map of the annual mean ice thickness in the SCICEX box and (b) time series of the annual mean. The orange line is the third-order polynomial from RPW08 for which the draft was converted to thickness with a factor of 1.107. The green line is a third-order polynomial from this study. The dots show the observations from within the box; red are from the submarines.

## 5 Error assessments

### 5.1 Long-memory processes

Percival et al. (2008) find that the spatial autocorrelation of 1 km ice draft measurements from submarines exhibits what is known as a long-memory process, in which the spatial autocorrelation does not drop off as quickly as for an autoregressive process at length scales up to 80 km. This means that the sampling error drops off with the track length  $L$  as  $L^{-0.49}$  rather than  $L^{-1}$ . However, RPW08 found that accounting for this long-memory correlation has only a small effect on the multiple regression coefficients determined from submarine ice draft data. Hence we have not accounted for this process in our analysis.

### 5.2 ULS first-return bias

As mentioned above, the submarine ice draft data have all been corrected with a constant  $-0.29 \text{ m}$  to account for the first-return and open-water-detection errors of ULS draft

measurements as done by Rothrock and Wensnahan (2007). This first-return bias is a function of the roughness of the underside of the sea ice and of the footprint width of the region insonified by the sonar beam (Vinje et al., 1998). For the submarines, the spatial sampling is typically 2 m and the footprint size is 2 to 5 m (Rothrock and Wensnahan, 2007), which, according to the analysis of Vinje et al. (1998), corresponds to a first-return correction of  $-0.44$  m for multiyear ice. However, it is likely an over-simplification to assume this correction is constant. It increases as the roughness or the footprint size increases (Vinje et al., 1998; Moritz and Ivakin, 2012). In addition, our analysis shows a strong positive correlation for all data sources between the mean thickness and the within-sample standard deviation determined from the point values. Similarly, Moritz and Ivakin (2012) show a strong correlation ( $R = 0.81$ ) between the within-footprint roughness for a set of ULS observations and the standard deviation of the sample thickness values for 256 profiles of length 50 to 150 m. Future research may show it is possible to determine a correction for first return that is based on the sample standard deviation. Clearly for smooth ice, for which there is no variation in the bottom topography, it should be zero. Not accounting for this dependence on bottom roughness may create an artificially thin bias for thin ice and a thick bias for thick ice as was mentioned above in regards to the thickness trend.

### 5.3 Snow

The snow depth or snow water equivalent needs to be taken into account in determining the ice thickness in all of the measurement systems. The error in the estimated snow depth then contributes to the error of the thickness estimate. However, the error in the snow depth is much less important for the ULS observations of ice draft from submarines and moorings than for the systems that measure the freeboard of the snow surface such as ICESat and IceBridge. For the ULS, the snow correction for ice draft,  $f(m)$  in Eq. (1), is based on the Warren et al. (1999) climatology and has an uncertainty on the order of 20 %, or up to just 0.02 m. The snow depth used to correct the Air-EM ice + snow measurements is taken from PIOMAS and has an uncertainty of about 0.10 m, which contributes the same amount to the uncertainty in the thickness estimate. The ICESat-J thickness estimates use a snow depth estimated from the accumulation of snowfall from the ECMWF Reanalysis. Kwok and Cunningham (2008) estimate that the snow depth uncertainty is 0.05 m and contributes 0.35 m to the uncertainty in the ice thickness while the snow density uncertainty contributes 0.10 to 0.36 m, depending on the freeboard and snow depth. The ICESat-G thickness estimates use the Warren et al. (1999) climatology. This climatology has an RMS error of between 0.05 and 0.14 m, depending on the month. The associated uncertainty in the ice thickness is a factor of 6.96 larger (Kwok and Cunningham, 2008), or 0.35 to 0.97 m. The Warren climatology

may be biased for recent years. Webster et al. (2014) find a  $0.029$  m decade<sup>-1</sup> decline in the spring snow depth in the western Arctic now dominated by first-year ice over the period 1950–2013. This would mean a mean decline of 0.14 m from 1960 to 2010 or roughly one-third of the spring snow depth.

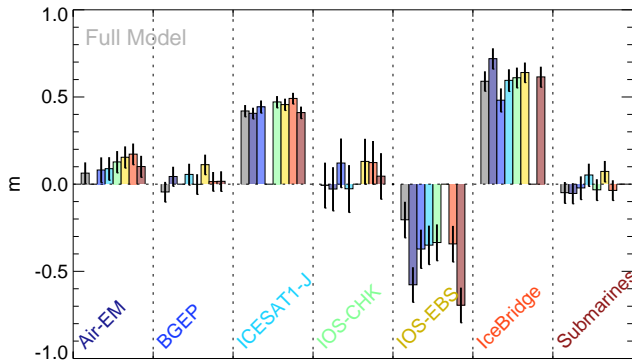
### 5.4 Sampling error

As we have alluded to above, sampling error can be a significant and serious source of uncertainty in comparing different ice thickness observations. All of the samples are from different times and/or places, so there are real differences in the nature of the ice sampled by the different measurements. The method used here depends on obtaining a large number of observations from a broad range of ice conditions so that comparisons in the mean can be made while accounting for large-scale variations in the mean ice thickness. The error in the fit includes random measurement errors, systematic measurement errors, sampling errors, and errors related to the inadequacy of the ITRP expression to fully represent the thickness variability.

One way to address the robustness of the results is to randomly withhold some of the data and repeat the fits to see if the coefficients change significantly. A set of 100 fits were computed for the entire Arctic Basin, 2000–2012, for each of which only half of the data, randomly selected for each system, was used. The mean of the resulting indicator coefficients is very similar to that found using all of the data and the variability of the coefficients from this ensemble is comparable to the standard error,  $\sigma_j$ , of the coefficients computed as part of the fit procedure. For example, we can conclude that the IceBridge data for the full Arctic Basin are significantly thicker than Air-EM, BGEP, ICESat-G, IOS-EBS, and the submarines but perhaps not thicker than ICESat-J.

### 5.5 Leave one source out

The importance of the individual data sources for computing the bias coefficients can be explored by repeating the analysis while leaving out each of the sources in turn. Do the bias coefficients change significantly? Figure 6 shows a bar chart of the indicator coefficients when just one data source is left out. The coefficients for most of the sources are quite similar for all of the ITRP fits. The largest variability is seen for the coefficients for IOS-EBS, which is not surprising given the isolated location of these measurements. The IOS-EBS coefficient is particularly sensitive to the exclusion of the BGEP or submarine data. There is also a fair amount of variability for the IceBridge coefficients, but in all cases the coefficients are still large. However, if both ICESat data sets are excluded and the submarines are used as a reference, we find very large changes in the relative magnitudes all of the remaining coefficients (not shown). This indicates the great importance of the satellite data in establishing the spatial structure of the



**Figure 6.** Coefficients of the ITRP indicator variables for fits that leave one data source out at a time for the Arctic Basin, 2000–2012. The coefficients for each source are grouped together. Grey bars show the coefficients for a fit that includes all of the observations, and bars in other colors indicate which source has been left out as shown by the colors of the diagonal labels (same order as the bars). The black lines give the  $1\sigma$  interval for the coefficients. ICESat-G is always the reference.

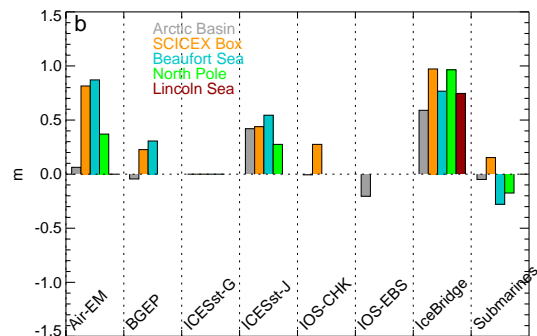
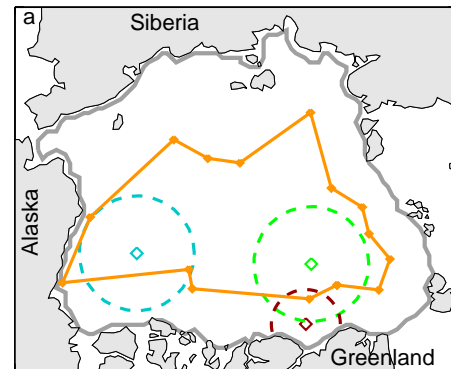
ice thickness fields when performing broad analyses of observing system differences.

### 5.6 Regional fits

The comparisons between data sets depend very much on the nature of the samples available for each. If they are far removed from each other in space or time, the true variability of the ice thickness may contaminate the difference estimates. For example, a bias between the observations could be partially resolved by the regression procedure with a spatial term if there is no spatial overlap. In addition, the differences between measurement systems may not be constant because the source of the bias, for example snow thickness or small-scale sea ice variability, is not constant. One way of addressing these uncertainties is to examine subsets of the data to see if differences observed between the systems are more or less robust. We look at five different regions, all for the period 2000–2012: (1) the entire Arctic Basin and using all measurement systems (the fit mentioned above), (2) the so-called SCICEX box in a broad region of the central basin that includes all submarine observations, (3) a 500 km radius circle centered on the BGEP moorings in the Beaufort Sea, (4) a 500 km circle centered on the North Pole, where a variety of observations are concentrated, and (5) a 300 km circle in the Lincoln Sea to evaluate Air-EM and IceBridge observations. Table 3 lists the summary information for each fit and Fig. 7 shows their locations. The coefficients of the indicator variables provide an estimate of the mean difference between each set of observations and the reference set in the sense that the RMS error of the fit is minimized if this difference is accounted for. Table 4 lists the values of the indicator coefficients for each fit and the RMS error of the fit for each observation source. Figure 7 shows the relative magnitudes

**Table 3.** The region, time period, number of observations used, number of terms, multiple regression coefficient, and RMS error (m) for each ITRP fit.

Region	Years	$N_{\text{obs}}$	$N_{\text{terms}}$	$R_{\text{mul}}$	$\text{RMS}_{\text{err}}$
Arctic Basin	2000.0–2012.6	3070	21	0.84	0.62
SCICEX box	2000.8–2012.6	1440	16	0.80	0.49
Beaufort Sea	2000.8–2012.6	725	15	0.76	0.49
North Pole	2000.8–2012.3	508	10	0.75	0.56
Lincoln Sea	2009.3–2012.3	127	3	0.62	0.69
SCICEX box	1975.3–2012.6	3017	18	0.89	0.49



**Figure 7.** (a) Locations of five regional fits for the period 2000–2012 and (b) relative magnitudes of the ITRP indicator coefficients. The magnitudes of the coefficients are grouped by observation source and color-coded by region (the order of the bars is the same as that of the region names). Grey depicts the coefficients for the fit for the entire basin.

of the coefficients for easy intercomparison of the bias terms determined for the different regions.

#### 5.6.1 SCICEX box

Data from US submarines are available mostly from a data release area defined by the US Navy (RPW08), the so-called “SCICEX box” (taken from the project name Scientific Ice Expeditions). Of the 34 submarine cruises available since 1975, there are only three cruises after 2000. However, the box is a convenient way to restrict the geographic extent of the data considered to a broad region in the central basin and

**Table 4.** Number of observations, the indicator coefficients and their  $\sigma$  values, and the RMS error for each source for each regional fit for the period 2000–2012.

Region	Air-EM	BGEP	ICESat-G	ICESat-J	IOS-CHK	IOS-EBS	IceBridge	Submarines	
Arctic Basin	354	334	900	600	26	107	588	161	$N_{\text{obs}}$
	0.06	-0.04	0.00	0.42	-0.01	-0.20	0.59	-0.05	Coef
	0.06	0.06	0.07	0.03	0.13	0.10	0.06	0.06	$\sigma$
	0.80	0.40	0.55	0.60	0.45	0.73	0.73	0.56	RMS <sub>err</sub>
SCICEX box	131	334	366	225	26	0	170	1765	
	0.42	0.26	0.00	0.42	0.31	0.98	0.09		
	0.06	0.05	0.04	0.04	0.10	0.05	0.05		
	0.50	0.37	0.52	0.44	0.41	0.66	0.51		
Beaufort Sea	48	334	150	100	0	0	64	29	
	0.87	0.31	0.00	0.54	0.77	-0.28			
	0.11	0.07	0.17	0.07	0.10	0.11			
	0.37	0.39	0.50	0.64	0.72	0.51			
North Pole	66	0	150	100	0	0	139	53	
	0.37	0.00	0.28	0.96	-0.17				
	0.13	0.13	0.07	0.14	0.11				
	0.30	0.57	0.44	0.69	0.57				
Lincoln Sea	51	0	0	0	0	0	76	0	
	0.00						0.75		
	0.63						0.14		
	0.61						0.74		

to also compare our results to those of RPW08. For the 2000–2012 period the submarine data are still in good agreement with the reference,  $0.14 \pm 0.05$  m; however, the coefficients for Air-EM ( $0.81 \pm 0.08$  m) and IceBridge ( $0.98 \pm 0.07$  m) are both notably thicker than the reference and submarines when compared to the full-basin fit. In addition, if 2000, when the first submarine cruise of the period occurred, is excluded the coefficient for the submarines increases to 0.42 m and the difference between it and that of ICESat-J is greatly reduced, similar to what we found earlier for the whole basin. These changes illustrate the fact that the differences between observation systems are not constant and may depend on the sample populations, the region, and the time periods included in the analysis.

### 5.6.2 Beaufort Sea

In the Beaufort Sea the four BGEP moorings provide abundant data for the entire annual cycle, and this is a good location to further assess the mean differences between the data sets while restricting the amount of spatial variability that is encountered. Within a 500 km circle of the center of the mooring array there are Air-EM and IceBridge observations as well as the satellite-based estimates. Compared to the reference, ICESat-J estimates are  $0.54 \pm 0.07$  m thicker and IceBridge estimates are  $0.77 \pm 0.10$  m thicker, similar to what we found for the full basin (Table 4). The Air-EM and BGEP coefficients are both substantially larger than for

the full basin,  $0.87 \pm 0.11$  m and  $0.31 \pm 0.07$  m, respectively. The difference between the two, 0.56 m, is much larger than the difference between them for the fit for the basin, 0.10 m, and is likely due to regional changes in the bias of the Air-EM data. This again illustrates that comparisons between data sets can be highly sensitive to the particular ice conditions encountered and that caution is recommended in assuming that intercomparisons and validation results for one area are applicable elsewhere.

### 5.6.3 North Pole

Abundant observations from submarines, IceBridge, Air-EM, and ICESat are available in the vicinity of the North Pole. ICESat-G has no observations closer than 400 km because of the nadir viewing of the satellite lidar while the ICESat-J data set has estimates within this circle based on interpolation from adjacent data points. A 500 km circle centered on the pole includes observations from both data sets. Note that data from a mooring at the pole, part of the North Pole Environmental Observatory, are still being reprocessed (Moritz personal communication) and is not included. Within this circle 508 observations are used for the fit. In this region the IceBridge estimates are 1.13 m thicker than the submarine estimates and 0.59 m thicker than the Air-EM estimates. ICESat-J estimates are 0.28 m thicker than the ICESat-G estimates. The coefficients from this fit are in general consistent with those for the entire basin (Fig. 7b and Table 4).

#### 5.6.4 Lincoln Sea

Is the large thickness bias in the IceBridge observations seen in the previous analyses robust? IceBridge observations have a coefficient larger than that of any of the other measurement systems in each of the fits except for the Beaufort Sea, where it is smaller than the Air-EM coefficient. Perhaps the IceBridge data are not well represented in the regression equation because they are concentrated in thick ice near the Canadian coast. We can partially address the IceBridge bias by examining only IceBridge and Air-EM measurements in a limited region in the Lincoln Sea, where there are 50 Air-EM and 76 IceBridge measurements within 100 km and one month of each other during the springs of 2009, 2011, and 2012. The ITRP shows that for this sample the IceBridge data are  $0.75 \pm 0.13$  m thicker than the Air-EM data. This is larger than the difference computed for the entire basin where the difference between the two is  $0.59 - 0.06 = 0.53$  m (Table 4). It is also larger than for the ITRP fits for the SCICEX box and for the Beaufort Sea where the differences between the two are smaller, 0.17 and  $-0.10$  m, respectively. While we cannot be confident of the exact magnitude of the bias and indeed as we have seen it changes considerably from place to place, it is likely that the IceBridge estimates are systematically thicker than any of the other measurements by up to 1.0 m (Table 4).

## 6 Conclusions

There is no gold standard for the estimation of the mean thickness of sea ice. All of the existing measurement techniques have one or more large sources of uncertainty. In situ measurements from the surface cannot sample the full thickness distribution. The submarine ULS measurements depend of the first-return echo to determine the ice draft, which is a potential source of unknown bias that may be a function of the bottom roughness. The mooring ULS measurements may also be subject to this same source of error. Both have potential errors in determining the open water level and accounting for the correct snow water equivalent. The satellite and airborne lidar observations depend on reliable detection of the surface height of nearby leads to accurately determine the height of the ocean surface and hence the total freeboard. The Air-EM measurements require an independent estimate of the snow depth, as do the satellite lidar measurements. All of the measurements struggle with obtaining an accurate mean value when the thickness is highly variable within the sensor footprint due to ridging. Finally, none of the measurements have been verified against other observations over regions that encompass the full ice thickness distribution of the area.

This study has determined some broad measures of the relative bias of the different systems. The ITRP method is dependent on having a large number of independent obser-

vations from each system so that a function can be fit to the thickness observations to account for the large-scale variability of the ice thickness. In addition to the nonlinear space and time variables, a bias term is included for each system that can contribute to the minimization of the error of the fit by adding or subtracting a constant value to all observations from a given system. This bias term can only be interpreted in a relative sense: how much thicker or thinner, in the mean, is one system compared to another? While we have typically used the ICESat-G system as a reference here, that does not mean it is a priori considered to be more accurate than the others. Indeed, nothing in the study speaks to the absolute accuracy of the measurements.

When ordered by relative magnitude of the coefficient of each system (Table 2), we see that the coefficient for IOS-EBS has the largest negative value relative to ICESat-G. However, because these measurements are in a small corner of the southeastern Beaufort Sea, we have little confidence that this result is a good indicator of the bias of the ULS measurements in this location compared to the other measurements. Of the others, ICESAT-G, submarines, IOS-CHK, BGEP, and Air-EM are all in broad agreement and in the mean are within 0.11 m of each other. However, we saw that the submarine bias coefficient is sensitive to the inclusion of the 2000 cruise. ICESat-J is 0.42 m thicker than ICESat-G but in good agreement with the submarine measurements in 2005. Finally, the IceBridge measurements average 0.59 m thicker than ICESat-G measurements.

It is beyond the scope of this study to determine why some of the observation systems appear to have biases, sometimes very significant, compared to the others. Possible sources of these discrepancies are the interpretation of ULS echo data, assumptions about snow depth or snow water equivalent, and methods of determination of the ocean water level for the lidars. While it is possible that there are systematic errors in determining the measurement differences introduced by the different times and locations of the observations, so called sampling errors, all of the systems, with the possible exception of IOS-EBS and the submarines, have sufficient observations spread over large spatial or temporal ranges to make this unlikely. Figure 7 shows the range of the coefficients determined with various spatial subsets of the data. For the entire basin, the experiment in which only a random half of the data from each system was used in a large set of fits gives very similar results to that when using the full data set. The leave-one-out experiment showed that the satellite measurements had a greater impact on the bias coefficients than the other systems. While our results provide an estimate of the relative biases of the measurement systems, they also point to the fact that more research to understand, characterize, and correct these errors is clearly required before we can homogenize the observational ice thickness record.

The ITRP annual mean basin-average ice thickness over the period 2000–2012 has declined 34%, a trend of  $-0.58 \pm 0.07$  m decade<sup>-1</sup>, while the September thickness

has declined by 50%. Finally, all of the observations in the central Arctic Basin within the SCICEX box for the period 1975–2012 indicate that the annual mean ice thickness in this region has decreased from 3.59 to 1.25 m, a 65% decline. In September the mean ice thickness has declined from 3.01 to 0.44 m, an 85% decline.

*Acknowledgements.* This study was supported by the NASA Cryospheric Sciences Program under grant number NNX11AF45G and by the National Science Foundation Division of Polar Programs under grant number 1023283. We thank all of the data providers for sharing ice thickness observations, often obtained under difficult conditions. We thank Harry Stern and two reviewers for a careful review of the manuscript.

Edited by: C. Haas

## References

- Haas, C. and Jochmann, P.: Continuous EM and ULS thickness profiling in support of ice force measurements, in: Proceedings of the 17th International Conference on Port and Ocean Engineering under Arctic conditions (POAC'03), 16–19 June 2003, Trondheim, Norway, edited by: Loeset, S., Bonnemaire, B., and Bjerkas, M., Norwegian University of Science and Technology, Trondheim, 849–856, 2003.
- Haas, C., Lobach, J., Hendricks, S., Rabenstein, L., and Pfaffling, A.: Helicopter-borne measurements of sea ice thickness, using a small and lightweight, digital EM system, *J. Appl. Geophys.*, 67, 234–241, 2009.
- Haas, C., Hendricks, S., Eicken, H., and Herber, A.: Synoptic airborne thickness surveys reveal state of Arctic sea ice cover, *Geophys. Res. Lett.*, 37, L09501, doi:10.1029/2010GL042652, 2010.
- Hansen, E., Gerland, S., Granskog, M. A., Pavlova, O., Renner, A. H. H., Haapala, J., Lyring, T. B., and Tschudi, M.: Thinning of Arctic sea ice observed in Fram Strait: 1990–2011, *J. Geophys. Res.-Oceans*, 118, 5202–5221, doi:10.1002/jgrc.20393, 2013.
- Krishfield, R. A. and Proshutinsky, A.: BGOS ULS Data Processing Procedure. Woods Hole Oceanographic Institute report, available at: <http://www.whoi.edu/files/server.do?id=85684&pt=2&p=100409> (last access: 26 April 2013), 2006.
- Krishfield, R. A., Proshutinsky, A., Tateyama, K., Williams, W. J., Carmack, E. C., McLaughlin, F. A., and Timmermans, M.-L.: Deterioration of perennial sea ice in the Beaufort Gyre from 2003 to 2012 and its impact on the oceanic freshwater cycle, *J. Geophys. Res.-Oceans*, 119, 1271–1305, doi:10.1002/2013JC008999, 2014.
- Kurtz, N., Studinger, M., Harbeck, J., Oana, V.-D.-P., and Farrell, S.: IceBridge Sea Ice Freeboard, Snow Depth, and Thickness, Digital media, NASA Distributed Active Archive Center at the National Snow and Ice Data Center, Boulder, Colorado, USA, <http://nsidc.org/data/idsi2.html> (last access: 6 December 2013), 2012.
- Kurtz, N. T., Farrell, S. L., Studinger, M., Galin, N., Harbeck, J. P., Lindsay, R., Onana, V. D., Panzer, B., and Sonntag, J. G.: Sea ice thickness, freeboard, and snow depth products from Operation IceBridge airborne data, *The Cryosphere*, 7, 1035–1056, doi:10.5194/tc-7-1035-2013, 2013.
- Kurtz, N. T., Galin, N., and Studinger, M.: An improved CryoSat-2 sea ice freeboard retrieval algorithm through the use of waveform fitting, *The Cryosphere*, 8, 1217–1237, doi:10.5194/tc-8-1217-2014, 2014.
- Kwok, R. and Cunningham, G. F.: ICESat over Arctic sea ice: Estimation of snow depth and ice thickness, *J. Geophys. Res.*, 113, C08010, doi:10.1029/2008JC004753, 2008.
- Kwok, R. and Rothrock, D. A.: Decline in Arctic sea ice thickness from submarine and ICESat records: 1958–2008, *Geophys. Res. Lett.*, 36, L15501, doi:10.1029/2009GL039035, 2009.
- Kwok, R., Cunningham, G. F., Wensnahan, M., Rigor, I., Zwally, H. J., and Yi, D.: Thinning and volume loss of the Arctic Ocean sea ice cover: 2003–2008, *J. Geophys. Res.*, 114, C07005, doi:10.1029/2009JC005312, 2009.
- Laxon, S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang, J., Haas, C., Hendricks, S., Krishfield, R., Kurtz, N., Farrell, S., and Davidson, M.: CryoSat-2 estimates of Arctic sea ice thickness and volume, *Geophys. Res. Lett.*, 40, 732–737, doi:10.1002/grl.50193, 2013.
- Lindsay, R. W.: A new sea ice thickness climate data record, *Eos*, 44, 405–406, 2010.
- Lindsay, R.: Unified Sea Ice Thickness Climate Data Record Collection Spanning 1947–2012, National Snow and Ice Data Center, Boulder, Colorado, USA, doi:10.7265/N5D50JXV, 2013.
- Melling, H. and Riedel, D. A.: Ice Draft and Ice Velocity Data in the Beaufort Sea 1990–2003, National Snow and Ice Data Center, Boulder, Colorado, USA, doi:10.7265/N58913S6, 2008.
- Melling, H., Riedel, D. A., and Gedalof, Z.: Trends in the draft and extent of seasonal pack ice, Canadian Beaufort Sea, *Geophys. Res. Lett.*, 32, L24501, doi:10.1029/2005GL024483, 2005.
- Moritz, R. E. and Ivakin, A. N.: Retrieving sea-ice thickness from ULS echoes: methods and data analysis, Proceedings of the 11th European Conference on Underwater Acoustics, Institute of Acoustics, St. Albans, UK, 8 pp., 2012.
- NSIDC: updated 2006: Submarine Upward Looking Sonar Ice Draft Profile Data and Statistics, National Snow and Ice Data Center, Boulder, Colorado, USA, doi:10.7265/N54Q7RWK, 1988.
- Peacock, N. R. and Laxon, S. W.: Sea surface height determination in the Arctic Ocean from ERS altimetry, *J. Geophys. Res.*, 109, C07001, doi:10.1029/2001JC001026, 2004.
- Percival, D. B., Rothrock, D. A., Thorndike, A. S., and Gneiting, T.: The variance of mean sea-ice thickness: Effect of long-range dependence, *J. Geophys. Res.*, 113, C01004, doi:10.1029/2007JC004391, 2008.
- Pfaffling, A. and Reid, J. E.: Sea ice as an evaluation target for HEM modelling and inversion, *J. Appl. Geophys.*, 67, 242–249, 2009.
- Pfaffling, A., Haas, C., and Reid, J. E.: A direct helicopter EM sea ice thickness inversion, assessed with synthetic and field data, *Geophysics*, 72, F127–F137, 2007.
- Renner, A. H. H., Gerland, S., Haas, C., Spreen, G., Beckers, J. F., Hansen, E., Nicolaus, M., and Goodwin, H.: Evidence of Arctic sea ice thinning from direct observations, *Geophys. Res. Lett.*, 41, 5029–5036, doi:10.1002/2014GL060369, 2014.
- Richter-Menge, J. A. and Farrell, S. L.: Arctic sea ice conditions in spring 2009–2013 prior to melt, *Geophys. Res. Lett.*, 40, 5888–5893, doi:10.1002/2013GL058011, 2013.

- Rothrock, D. A. and Wensnahan, M.: The accuracy of sea-ice drafts measured from U.S. Navy submarines, *J. Atmos. Ocean. Tech.*, 24, 1936–1949, doi:10.1175/JTECH2097.1, 2007.
- Rothrock, D. A., Yu, Y., and Maykut, G. A.: Thinning of the Arctic sea-ice cover, *Geophys. Res. Lett.*, 26, 3469–3472, 1999.
- Rothrock, D. A., Percival, D. B., and Wensnahan, M.: The decline in arctic sea-ice thickness: Separating the spatial, annual, and interannual variability in a quarter century of submarine data, *J. Geophys. Res.*, 113, C05003, doi:10.1029/2007JC004252, 2008.
- Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H., and Kwok, R.: Uncertainty in modeled Arctic sea ice volume, *J. Geophys. Res.*, 116, C00D06, doi:10.1029/2011JC007084, 2011.
- Tucker III, W. B., Weatherly, J. W., Eppler, D. T., Farmer, D., and Bentley, D. L.: Evidence for the rapid thinning of sea ice in the western Arctic Basin at the end of the 1980s, *Geophys. Res. Lett.*, 28, 2851–2854, 2001.
- Vinje, T., Nordlund, N., and Kvambekk, A.: Monitoring ice thickness in Fram Strait, *J. Geophys. Res.*, 103, 10437–10450, 1998.
- Warren, S. G., Rigor, I. G., Untersteiner, N., Radionov, V. F., Bryazgin, N. N., Aleksandrov, Y. I., and Colony, R.: Snow depth on arctic sea ice, *J. Climate*, 12, 1814–1829, 1999.
- Webster, M. A., Rigor, I. G., Nghiem, S. V., Kurtz, N. T., Farrell, S. L., Perovich, D. K., and Sturm, M.: Interdecadal changes in snow depth on Arctic sea ice, *J. Geophys. Res.-Oceans*, 119, 5395–5406, doi:10.1002/2014JC009985, 2014.
- Wensnahan, M. and Rothrock, D. A.: Sea-ice draft from submarine-based sonar: Establishing a consistent record from analog and digitally recorded data, *Geophys. Res. Lett.*, 32, L11502, doi:10.1029/2005GL022507, 2005.
- Yi, D. and Zwally, J.: Arctic Sea Ice Freeboard and Thickness, National Snow and Ice Data Center, Boulder, Colorado, USA, 2009.
- Zhang, J. and Rothrock, D. A.: Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates, *Mon. Weather Rev.*, 110., 845–861, 2003.
- Zwally, H. J., Yi, D., Kwok, R., and Zhao, Y.: ICESat Measurements of Sea Ice Freeboard and Estimates of Sea Ice Thickness in the Weddell Sea, *J. Geophys. Res.*, 113, C02S15, doi:10.1029/2007JC004284, 2008.
- Zygmuntowska, M., Rampal, P., Ivanova, N., and Smedsrud, L. H.: Uncertainties in Arctic sea ice thickness and volume: new estimates and implications for trends, *The Cryosphere*, 8, 705–720, doi:10.5194/tc-8-705-2014, 2014.