An evaluation of surface meteorology and fluxes over the Iceland and Greenland Seas in ERA5 reanalysis: the impact of sea ice distribution

I. A. Renfrew[1*], C. Barrell[1], A. D. Elvidge[1], J. K. Brooke[2], C. Duscha[3], J. C. King[4], J. Kristiansen[5], T. Lachlan Cope[4], G. W. K. Moore[6], R. S. Pickart[7], J. Reuder[3], I. Sandu[8], D. Sergeev[9], A. Terpstra[3,1], K. Våge[3], A. Weiss[4]

[1]*School of Environmental Sciences, University of East Anglia, Norwich, UK*
[2]*Met Office, Exeter, UK*
[3]*Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research, Bergen, Norway*
[4]*British Antarctic Survey, Cambridge, UK*
[5]*Norwegian Meteorological Institute, Oslo, Norway*
[6]*Department of Physics, University of Toronto, Canada*
[7]*Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA*
[8]*European Centre for Medium-range Weather Forecasts, Reading, UK*
[9]*College of Engineering, Mathematics and Physical Science, University of Exeter, Exeter, UK*

[*]*Corresponding Author: Professor Ian Renfrew, Centre for Ocean and Atmospheric Sciences, School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK; i.renfrew@uea.ac.uk*

For submission to: Quarterly J. Royal Met. Soc. on 22nd July 2020; Second submission 8th September 2020; Third submission 20th October 2020.

Key Words

Sea ice, marginal ice zone, surface fluxes, subpolar seas, ERA5, OSTIA

Funding information: Funding for this study was from the NERC AFIS grant (NE/N009754/1), the ALERTNESS (Advanced models and weather prediction in the Arctic: enhanced capacity from observations and polar process representations) project (Research Council of Norway project number 280573), the Trond Mohn Foundation (BFS2016REK01) and the National Science Foundation grant OCE-1558742.

Abstract

The Iceland and Greenland Seas are a crucial region for the climate system, being the headwaters of the lower limb of the Atlantic Meridional Overturning Circulation. Investigating the atmosphere-ocean-ice processes in this region often necessitates the use of meteorological reanalyses – a representation of the atmospheric state based on the assimilation of observations into a numerical weather prediction system. Knowing the quality of reanalysis products is vital for their proper use. Here we evaluate the surface-layer meteorology and surface turbulent fluxes in winter and spring for the latest reanalysis from the European Centre for Medium-Range Weather Forecasts – ERA5. In situ observations from a meteorological buoy, a research vessel and a research aircraft during the Iceland-Greenland Seas Project provide unparalleled coverage of this climatically important region. The observations are independent of ERA5. They allow a comprehensive evaluation of the surface meteorology and fluxes of these subpolar seas and, for the first time, a specific focus on the marginal ice zone. Over the ice-free ocean, ERA5 generally compares well to the observations of surface-layer meteorology and turbulent fluxes. However, over the marginal ice zone the correspondence is noticeably less accurate: for example, the root-mean-square errors are significantly higher for surface temperature, wind speed and surface sensible heat flux. The primary reason for the difference in reanalyses quality is an overly smooth sea-ice distribution in the surface boundary conditions used in ERA5. Particularly over the marginal ice zone, unrepresented variability and uncertainties in how to parameterize surface exchange compromise the quality of the reanalyses. A parallel evaluation of higher resolution forecast fields from the Met Office's Unified Model corroborates these findings.

1. Introduction

The subpolar seas of the North Atlantic are critically important for the global climate system as they are the source of the dense waters of the Atlantic Meridional Overturning Circulation (AMOC). Investigating coupled atmosphere-ocean processes, in particular surface turbulent heat and momentum fluxes, are key steps to improving our understanding of the role that the North Atlantic subpolar seas play within the AMOC (e.g. Buckley and Marshall 2016). The dominant contribution to the AMOC is from east of Greenland (Pickart and Spall 2007), as is the largest variability in volume transport (Lozier et al. 2019), pointing to the Norwegian, Barents, Greenland, Iceland and Irminger Seas as key locations for the formation of dense water masses. Ocean circulation paradigms have shifted over the years: from when it was thought the Iceland and Greenland Seas were the primary source of dense water via open ocean convection (e.g. Swift and Aagaard 1981); to a view that consistent ocean cooling and densification around the rim current of the Nordic Seas was dominant (e.g. Mauritzen 1996); to a shift back to the importance of the Iceland and Greenland Seas due to the discovery of the North Icelandic Jet (Jónsson and Valdimarsson 2004; Våge et al. 2011, 2013; Semper et al. 2019); and of areas of dense water in the northwest Iceland and western Greenland Seas (Våge et al. 2018). Exactly where, when and how the water-mass transformations take place, and how the dense water feeds the AMOC, are active areas of research. These were key questions posed at the inception of the Iceland-Greenland Seas Project (IGP): a coordinated atmosphere-ocean project – encompassing a rare wintertime field campaign – to observe, analyse and model the coupled climate system in this region (see Renfrew et al. 2019a for an overview). Here we make use of several atmospheric data sets gathered during the IGP field campaign, that together provide unparalleled coverage of the region, to evaluate a state-of-the-art meteorological reanalysis product. We focus on the surface-layer meteorology and surface fluxes – the salient fields for atmosphere-ocean-sea-ice coupling.

Meteorological reanalyses are generated from the assimilation of observations into a consistent version of a numerical weather prediction (NWP) forecast system by optimally blending short-range forecasts with millions of observations through data assimilation. As the quality of NWP systems has increased tremendously over recent decades (Bauer et al. 2015), so too has the quality of meteorological reanalysis (Hersbach et al. 2020). They are an excellent tool for the analysis of the climate system (e.g. Papritz and Spengler 2017); especially in regions with a paucity of in situ observations such as the Iceland and Greenland Seas (e.g. Jung et al. 2016). However, it is vital to have knowledge of the quality of reanalyses products before they are used for particular applications. This is particularly important for the polar and subpolar regions, where the NWP systems have numerous well-known weaknesses, e.g. in the representation of stable boundary layers, mixed-phased clouds, sea ice characteristics and surface exchange over heterogeneous surfaces or in the use of observations (Bourassa et al. 2013; Vihma et al. 2014; Jung et al. 2016; Lawrence et al. 2019). All of these processes will impact the quality of surface-layer meteorological variables and surface fluxes, raising questions as to how accurate these fields will be in reanalyses, analyses and forecasts. Here we address this through an evaluation of ERA5, the latest global reanalysis product produced by the European Centre for Medium-range Weather Forecasts (ECMWF), against independent observations from the IGP. Our focus is on ERA5 (Hersbach et al. 2020) as this is a relatively new product that has been produced to replace and improve upon the popular ERA-Interim reanalyses (ERA-I; Dee et al. 2011), using enhanced observations and a recent improved version of the ECMWF Integrated Forecasting System.

A number of recent evaluations of meteorological reanalyses have been carried out for the whole Arctic (Lindsay et al. 2014; Bromwich et al. 2016); for the Arctic Ocean (Lüpkes et al. 2010; Jakobson et al. 2012; Wesslen et al. 2014; Wang et al. 2019) and for the subpolar seas (Renfrew et al. 2009; Harden et al. 2011, 2015; Moore et al. 2016). All of the above evaluations have used ERA-I output (or operational output from the same model cycle in Renfrew et al. 2009) and several also evaluate other products such as the regional Arctic System Reanalyses (ASR; Bromwich et al. 2016; Moore et al. 2015, 2016; Wesslén et al. 2014) or other global reanalyses (e.g. Jakobson et al. 2012; Lindsay et al. 2014; Jones et al. 2016; Graham et al. 2019a, 2019b). A number of errors in surface-layer meteorology have been revealed in these studies, e.g. all reanalyses tend to have wind speeds that are biased low over land stations (Bromwich et al. 2016), especially for moderate-to-strong winds in regions of steep or complex orography (Moore et al. 2015, 2016; Nygård et al. 2016; Jones et al. 2016), although higher resolution partly ameliorates this problem (Renfrew et al. 2009; DuVivier and Cassano 2013; Moore et al. 2016). ERA-I usually performs comparatively well against other global reanalyses (e.g. Jakobson et al. 2012; Lindsay et al. 2014; Jones et al. 2016). The limited evaluations of ERA5 so far indicate it also performs well against independent radiosonde observations and for radiative fluxes in spring and summer over the Arctic Ocean (Graham et al. 2019a, 2019b), and outperforms ERAI for global oceanic wind fields when compared to scatterometer observations (Belmonte Rivas and Stoffelen 2019).

Focusing on evaluations for the subpolar seas, ERA-I generally does well at representing surface-layer temperatures, winds, humidity and turbulent fluxes, although with more scatter in relative humidity and turbulent fluxes (Harden et al. 2015), and with similar findings for the equivalent operational ECMWF analyses evaluated in Renfrew et al. (2009). For example, comparing against two years of meteorological buoy observations in the central Iceland Sea, the ERA-I biases (root-mean-square errors, RMSE) in air temperature, relative humidity, wind speed and sensible heat flux were 0.43

(0.82) K, -5.5 (8.4) %, 0.12 (1.6) m s^{-1} and -8.3 (15.8) W m^{-2} respectively (Harden et al. 2015). Moore et al. (2008) report comparable discrepancies in air temperature and wind speed against 5 months of buoy observations from the Irminger Sea for the North American Regional Reanalyses (NARR). Dukhovskoy et al. (2016) find similar differences in wind speed from the ASR, the Climate Forecast System Reanalysis (CFSR) and satellite-derived products when compared against the same buoy observations. However, if QuikScat scatterometer winds are taken as truth, they find the biases (RMSE) in the ASR and CFSR winds are <0.5 (1) m s^{-1} in the subpolar North Atlantic and Nordic Seas. Closer to the steep orography of coastal Greenland there are challenges in representing orographic flows and 10-m wind speed biases (RMSE) increase to approximately -2 (3-4) m s^{-1} (Renfrew et al. 2009; Harden et al. 2011; Moore et al. 2015, 2016).

Reviewing previous reanalyses evaluations, it is clear that there are some gaps in knowledge. Over the subpolar seas there have been no specific evaluations of reanalysis products over sea ice or the marginal ice zone (MIZ), the zone of more variable sea-ice conditions where waves and swell impact the sea ice. Renfrew et al. (2009) show a handful of aircraft observations over the MIZ that illustrate substantial differences in surface temperature, air temperature and wind speed between the various models and the observations, but there are too few data points for a quantitative analysis. All of the Arctic Ocean evaluations currently available are for near-100% ice concentrations, meaning the quality of reanalyses over any Arctic MIZ is currently unknown.

Our observations come from three separate platforms – a meteorological buoy, a research vessel and a research aircraft – all used during the IGP to make observations of the atmospheric surface layer (Renfrew et al. 2019a). Our meteorological buoy was in the NW Iceland Sea (see **Figure 1**) for 78 days in an open ocean location – closer to the sea ice than the central Iceland Sea buoy of Harden et al. (2015). Our research vessel, the NRV *Alliance*, traversed the Iceland and Greenland Seas for 43 days, penetrating the MIZ on several occasions. While our research aircraft primarily targeted the NW Iceland Sea and the MIZ in particular, with observations from 9 flights included here. Combining data from these three platforms allows us to make a comprehensive evaluation of ERA5 over the winter-to-spring period; and for the first time we are able to contrast ice-free ocean and MIZ conditions.

In Section 2 we describe the observations, model output and the methods employed. Section 3 provides an evaluation of ERA5 for the ice-free ocean and for the MIZ, revealing contrasts in accuracy. In Section 4 we explore why this is the case, aided by an evaluation of higher resolution limited-area analyses and forecasts from the Met Office Unified Model. Section 5 provides conclusions and recommendations.

2. Data sets and methodology

2.1 Observations from a meteorological buoy

A Seawatch Wavescan meteorological buoy was deployed on 17 February 2018 in the NW Iceland Sea at 70°38.38 N, 15°24.58 W. It worked well for 78 days before breaking loose from its anchor. Hourly observations of air temperature, relative humidity (RH), air pressure, solar radiation, wind speed and wind direction were made at a height of ~3 m (see **Table 1** for instrumentation details and estimates of accuracy for all platforms). In addition, observations of sea surface temperature (SST), ocean currents and wave height, period and direction were recorded. All variables were quality controlled with outliers and non-physical measurements removed. Quality control procedures revealed the air pressure to be erroneous for about half of the deployment, so the mean-sea-level pressure from ERA5 is used when needed to derive other variables. Unfortunately, the SST was not measured reliably, so instead we use the shallowest (8 m) ocean temperature from an adjacent ocean mooring (see Renfrew et al. 2019a). At this time of the year, and in this location, the ocean surface layer is generally well mixed, so this substitution is reasonable when comparing mean values (e.g. Våge et al. 2018). However, at 8 m depth the variability in temperature will be reduced compared to that at the surface. A comparison to observations at the meteorological station on Jan Mayen revealed that the air temperature was erroneous during a short cold period in April (likely due to icing). Note that the buoy observations were not made available to meteorological forecast centres, so are independent of ERA5.

2.2 Observations from the research vessel

A time series of surface-layer meteorological variables has been generated from the 43-day cruise of the NATO research vessel *Alliance* in February-March 2018 in the Iceland and southern Greenland Seas (Fig. 1; see Renfrew et al. 2019a). Temperature, pressure and RH were taken from the WeatherPak shipboard systems mounted at ~15 m above sea level on the bow mast (see Table 1). Unfortunately, these systems had some technical problems, so a careful quality control procedure was implemented, with timing, linear regression and bias checks against independent measurements from the boat deck, which were then used to fill in several small gaps in the WeatherPak time series. Due to instrument problems with the WeatherPak anemometers, and to avoid periods of sheltering by the ship's superstructure, here we take wind speed and wind direction from the lowest bin (40 m) of a Doppler wind lidar (a Leosphere WindCube v2 8.66) located on the boat deck. A novel correction algorithm for translational motions of the ship, as well as established corrections for the pitch, roll and yaw of the *Alliance*, based on intertial motion unit measurements was implemented (Duscha et al. 2020). SST was measured by a bow temperature sensor with checks against 2-m measurements from an underway Conductivity-Temperature-Depth (CTD) system. Underway salinity measurements were also used to confirm when a few SST measurements were erroneous and interpolated CTD data were used to cover a few short episodes of missing SST data. Periods of time in port have been removed from the time series. All variables were quality controlled with outliers and non-physical measurements removed. Here we use 10-minute averaged data. Note that data from radiosondes released from the *Alliance* were sent to forecast centres in real time, so were available for assimilation into operational systems and reanalyses. However, the ship-based measurements used here are independent of ERA5.

2.3 Observations from the research aircraft

Surface-layer observations are also available from our coordinated aircraft campaign in February and March 2018. We used the British Antarctic Survey's instrumented DH6 Twin Otter aircraft for 14 science missions (Flights 292-306), several in the vicinity of the *Alliance*, and more than half flying over the MIZ. A summary of the IGP meteorological field campaign is given in Renfrew et al. (2019a). A number of minor technical issues arose during the quality control of the aircraft data: the radar altimeter wasn't functioning on the first three flights, so was substituted by a calibrated GPS altitude; icing on the turbulence probe prevented calculation of 50 Hz winds on flights 292 and 297, so substitute horizontal winds were calculated using pitot tube and inertial navigation unit measurements; the 1 Hz temperature data were not available on flight 297; and high frequency humidity data were missing for part of flight 294 due to a mission scientist blunder. The airborne surface temperature is based on a downward pointing infra-red thermometer which needs to be calibrated. Here we follow Cook and Renfrew (2015) and apply a constant offset for each flight determined by a comparison against ERA5 SSTs over open water. We also checked that the corrected surface temperatures were consistent for co-located data points and physically realistic with respect to the sea-ice cover. It is worth noting that the Heimann infra-red thermometer is only accurate to within $\pm$1 K (c.f. Table 1). Minor flight-dependent timing adjustments were made to the 50 Hz thermistor, humidity sensor and GPS altimeter data on all the flights to account for their positions on the aircraft and the instrument response times, based on lagged correlations with vertical velocity observations. In addition, there was initially a problem with partitioning the horizontal wind into components. A careful analysis of adjacent flight legs with reciprocal headings allowed us to apply a small correction to the true air speed and heading (~1 degree) and thus derive accurate wind components for all flights. The aircraft-based observations are independent of ERA5.

Here we use observations from the 9 successful marine flights. These include over 400 minutes flying in the atmospheric surface layer (over 230 minutes over the MIZ) typically at 20-50 m above sea level. We have divided our surface-layer legs into 'runs' of 150 s (approximately 9 km) and we calculate mean and turbulent quantities for each run. A run length of 150 s was chosen following sensitivity testing – it is a reasonable compromise between capturing the vast majority of the fluxes and accommodating the heterogeneous surface conditions (see Elvidge et al. 2016; Grunwald et al. 1996). The mean variables used here are air temperature, RH, specific humidity, wind speed, surface temperature and ice fraction (derived from albedo and surface temperature – see Elvidge et al. 2016). The turbulent variables used are momentum flux, sensible heat flux and latent heat flux – calculated using the eddy covariance method following Petersen and Renfrew (2009). A strict quality control procedure is applied with covariances, co-spectra and ogives all checked. One concern with this technique is the relatively large sampling error when measuring turbulence for a relatively short time. This sampling error is typically around 30 – 40% of the magnitude of the flux (e.g. Drennan et al. 2007; Petersen and Renfrew 2009; Weiss et al. 2011). To compensate for this, the data are usually averaged together to obtain robust results, e.g. into wind speed bins. Here we will directly compare covariances fluxes to model output. We use this approach because there is not currently a widely accepted bulk flux algorithm for estimating surface fluxes over the MIZ. However, this approach is unusual; more commonly, meteorological observations are used to derive bulk flux estimates from an offline algorithm, with these bulk fluxes then compared to model fluxes (e.g. Renfrew et al. 2002, 2009). Our approach means that the sampling error needs to be taken into account. In other words, for a comparison to be valid there needs to be sufficient data points for the statistical quantities to be robust; we believe this the case, with the possible exception of the aircraft-based fluxes over water.

2.4 ERA5 reanalysis

ERA5 is the fifth generation ECMWF atmospheric reanalysis (Hersbach et al. 2020). ERA5 is produced using cycle 41r2 of the Integrated Forecast System (IFS) model, using a four-dimensional variational data assimilation scheme. The reanalysis benefits from a relatively high-resolution grid with 137 vertical levels and a horizontal grid spacing of 0.28125° (~31 km, or T_L639 triangular truncation). The time frequency of atmospheric reanalyses parameters is 1 h and we use instantaneous meteorological variables and hourly mean surface fluxes. Besides a higher spatiotemporal resolution, ERA5 has a number of additional advantages over its predecessor, ERA-I (whose production stopped in 2019). The ERA5 data assimilation is enhanced by using not only satellite radiances, but also ozone, aircraft and surface pressure data in the variational data assimilation scheme. ERA5 also assimilates a number of humidity-sensitive satellite channels using the all-sky approach instead of the clear-sky approach used in

ERA-I, thus providing new information during cloudy and precipitating conditions. In addition, various reprocessed datasets and recent instruments that could not be ingested in ERA-I are included in ERA5. These improvements result, among other things, in more consistent sea surface temperature and sea ice cover, compared to that for ERA-I. The evolution of SST and sea ice cover in ERA5 is based on a number of products over different periods of time (Hersbach et al. 2020). The Met Office's Operational Sea-surface Temperature and Sea Ice Analysis (OSTIA) dataset is used from 2007 to present; and this uses the EUMETSAT Ocean and Sea Ice Satellite Applications Facility (OSI-SAF) 401 data set for sea ice concentration (Donlon et al., 2012). OSTIA provides daily updated SST and sea ice fields, primarily sourced from satellite observations, with a horizontal resolution of 1/20° (~ 6 km). OSTIA is also used in the ECMWF's operational forecasting system. The ERA5 SST does not vary during the day; although there is not an observable diurnal signal in SST in this region anyway.

2.5 Met Office analyses

The Met Office Unified Model (MetUM) is a state-of-the-art, non-hydrostatic atmospheric model used for operational weather forecasting and as a component in climate models. Here we analyse limited-area simulations made using version 10.6 of the MetUM and a standard parameterization configuration generally following that used operationally in the limited-area km-scale RAL1-M configuration (e.g. Bush et al. 2020). This configuration has proven to be reasonably accurate at simulating cases of cold-air outbreaks and polar lows in this area (e.g. Sergeev et al., 2017; Renfrew et al., 2019b). It employs daily updated sea ice and sea surface temperature fields from OSTIA (as used in ERA5). Here the model domain covers an area approximately 1000 x 1500 km across the Iceland and Greenland Seas (see Fig. 13 in Renfrew et al. 2019a). The setup has a horizontal grid spacing of 0.02 degrees (~2.2 km) and 70 vertical levels – the lowest of which is at a height of 2.5 m over the ocean. The limited area model is forced at its lateral boundaries by a global MetUM simulation which employs a horizontal grid spacing of ∼10 km (N1280) with 70 vertical levels and also generally follows operational settings. We use instantaneous hourly model output from the simulations initialised at 00 UTC that day.

2.6 Comparison methodology

We have used the COARE3 bulk flux algorithm (c.f. Fairall et al. 2003) to adjust the meteorological observations from the buoy and research vessel to standard levels (e.g. 2-m temperature, 10-m wind) and to estimate surface turbulent fluxes. We have matched the model output to the observations as follows. For the buoy and the research vessel observations we use linear spatial interpolation and match hourly observations and model output. For the aircraft observations we use linear interpolation to the height of the observations and for ERA5 the nearest neighbour was used spatially and nearest hour in time, while for the MetUM a linear interpolation was used in both space and time.

The meteorological buoy was located in the ice-free ocean, whereas both the research vessel and aircraft cross from the open ocean into the MIZ on numerous occasions (Fig. 1). For the following comparison we have divided both of these time series into subsets that are *'over water'* and *'over the MIZ'*. For the research vessel, a time series of satellite-derived ice fraction is derived from the nearest OSTIA grid point to the position of the *Alliance* every hour. The *Alliance* is designated as over the MIZ when the ice fraction > 0. This is a pragmatic approach given that in situ observations of ice fraction are not available. For the aircraft, ice fraction is estimated using an albedo derived from shortwave radiation observations (after Elvidge et al. 2016). As above, we designate data as over the MIZ when the ice fraction > 0. Note using an alternative temperature-based ice fraction produces very similar results.

3. An evaluation of ERA5 for the Iceland and Greenland Seas region

Surface-layer meteorology and surface turbulent fluxes generally compare well to observations from the meteorological buoy. **Figure 2** shows a very good correspondence over time for 2-m air temperature (T_{2m}) and 10-m wind speed (U_{10m}). All the major variability is captured, and the timing of the changes is generally in very good agreement. There are a few periods of larger difference, e.g. when the maxima and minima are not captured. The correspondence illustrated here is generally representative of the other variables.

Figure 3 shows scatter plots for the buoy observations versus ERA5 output and **Table 2** gives selected statistics, including the correlation coefficient and slope of a linear regression fit, the bias (model – observations) and the RMSE (root-mean-square error). The correspondence in T_{2m}, RH_{2m}, specific humidity (q_{2m}), and the surface heat fluxes is very good, with low biases and relatively low RMSE (e.g. smaller than the standard deviation of the observations). There is a dry bias of -4.7% in RH_{2m} (or -0.14 g kg^{-1} in q_{2m}). The SST comparison has a small bias (0.07 K); however, recall the observations are from a depth of 8 m, which likely inhibits the observed variability compared to the ERA5 (OSTIA) variability. The correspondence in wind speed and momentum flux (τ) are noticeably poorer. The biases are relatively large, 0.61 m s^{-1} in U_{10m}, and the linear regression slopes deviate from 1. There is evidence that this is partly due to a sheltering of the buoy by waves (c.f. Large et al. 1995). To examine this we have divided the data into quartiles by observed significant wave height and the U_{10m} biases (regression slopes) for each quartile are: 0.14 m s^{-1} (0.8); 0.50 m s^{-1} (0.91); 0.53 m s^{-1} (1.09); 1.52 m s^{-1} (1.13). There is a clear worsening in correspondence with significant wave height, suggesting the wind and momentum flux biases may be entirely due to the buoy sheltering. There is also a bias of -15 degrees in wind direction, i.e. ERA5 has winds coming from a more easterly direction. Overall, our buoy comparison is qualitatively and quantitatively similar to a comparison of ERA-I to a buoy in the central Iceland Sea by Harden et al. (2015); and to a comparison of NARR, ASR and CFSR output to buoy observations in the Irminger Sea and the central Iceland Sea (Moore et al. 2008; Dukhovskoy et al. 2016).

Observations from the *Alliance* as it traversed the Iceland and Greenland Seas are shown in a time series in **Figure 4** and illustrate it penetrating the MIZ on eight occasions (see also Fig. 1). The proximity of the *Alliance* to sea ice results in greater variability in T_{2m} and SST than at the buoy. ERA5 generally captures the timing of this variability well, although it does fail to capture some of the cold extremes and appears poorer for SST at times, especially close to the MIZ. The correspondence in U_{10m} is generally very good.

A quantitative evaluation of ERA5 over water and over the MIZ is presented in **Figure 5** and **Table 3** for the ship-based observations and in **Figure 6** and **Table 4** for the aircraft-based observations. The tables give selected statistics for each time series as well as the bias and RMSE separately for over water and over the MIZ. The scatter plots are shaded to represent open water (blue) or the MIZ (white) for the ship, and ice fraction (blue to white) for the aircraft. Generally, the correspondences – as measured by the correlation coefficient and linear regression slope – are good and similar for the ship- and aircraft-based comparisons. The correspondences for RH and the turbulent fluxes are noticeably worse for the aircraft comparison, partly due to the sampling issues discussed in section 2.4 and the small size of the data subset. We now discuss the comparisons for *over water* and for *over the MIZ* in turn.

Over water, the ERA5 biases against ship-based observations are generally small and the RMSE are small compared to the standard deviation of the observations (see Table 3). Compared to the buoy results: the correlation, slope and RMSE are similar for T_{2m}, RH_{2m}, q_{2m}, the sensible heat flux (SHF) and the latent heat flux (LHF). The bias is higher for T_{2m} (0.49 K compared to 0.05 K) and for the SHF, while there is considerable scatter in the SST comparison, all likely due to the proximity to sea ice. In contrast to the buoy comparison, for U_{10m} the linear regression slope is low due to high wind speeds being under predicted (Fig. 5c) and this contributes to a bias of -0.20 m s^{-1} over water. Similar to the buoy results, there is an easterly bias of -7 degrees in wind direction and high accuracy in the surface flux estimates (the RMSE is less than half the observed standard deviation). In the aircraft comparison, there are 69 data points over water (only 44 for the turbulent fluxes). For the meteorological variables the accuracy is generally similar to that found for the buoy and ship comparisons over water, e.g. the RMSE are generally similar. There is a dry bias in RH, as well as a low slope and negative bias in wind speed, that are consistent with the buoy and ship comparisons. A failure to represent the highest wind speeds over the ocean has been seen in previous studies (e.g. Renfrew et al. 2009; Li et al. 2013). Overall the correspondence over water is very good, largely consistent between the buoy, ship and aircraft comparisons and similar to previous evaluations of ERA-I for the subpolar seas (e.g. Harden et al. 2015).

Over the MIZ, there are typically 84 data points in the ship-based comparison and 88 in the aircraft comparison. Figures 5 and 6 illustrate that ERA5 is less accurate over the MIZ than over water. For example, there is a clear increase in scatter with increasing ice fraction (paler dots) in Fig. 6. Examining the statistics (Tables 3 and 4) the RMSE are greater over the MIZ than over water for all meteorological variables (except for RH/q in the aircraft comparison) and for all the surface fluxes (except for momentum in the aircraft comparison)[1]. For some variables the RMSE over the MIZ are particularly large; e.g. 2.94 K for T_{sfc} and 2.42 m s^{-1} for U from the aircraft. Note the RMSE for SST are similar over the MIZ and over water in the ship comparison. This difference reflects that the *Alliance* was on the fringes of the MIZ and actively avoiding sea ice, whereas the aircraft went much deeper into the MIZ. In general, the accuracies between the aircraft- and ship-based comparisons over the MIZ are consistent, but there are quantitative differences due to the aircraft observations being from flight level (30-70 m) or derived differently. Turning to the biases, these are larger over the MIZ for all of the ship-based comparisons, except T_{2m} and wind direction; but this finding is not consistent with the aircraft-based comparison.

In short, combining the comparisons from the three observing platforms demonstrates that ERA5 is significantly less accurate over the MIZ than over water for both the surface-layer meteorology and surface turbulent fluxes. This is clearly demonstrated by contrasting the RMSE *over water/over the MIZ:*

- for air temperature, surface temperature, wind speed: 0.78/1.00 K, 0.47/2.94 K, 1.77/2.42 m s^{-1} from the aircraft comparison (Table 4); and
- for momentum flux, sensible heat flux, and latent heat flux: 0.077/0.098 N m^{-2}, 20.9/35.2 W m^{-2}, and 14.2/20.1 W m^{-2} from the ship comparison (Table 3).

In the next section we examine the causes of this lower accuracy over the MIZ.

[1] The two exceptions, for RH/q and the momentum flux, are both due to the aircraft RMSE over water being surprisingly large (when compared to the ship or buoy comparisons), primarily due to the large variances for these variables over water and the relatively small size of the data subset.

4. Investigating the reduced accuracy of ERA5 over the marginal ice zone

There are a number of possible reasons why the surface-layer meteorology and the surface fluxes from ERA5 are less accurate over the MIZ. There is an increase in the heterogeneity of many surface properties over the MIZ compared to over the ice-free ocean; for example, in surface temperature, surface roughness and albedo as is evident from our aircraft-based observations (e.g. Fig. 6f). Perhaps ERA5 cannot represent this heterogeneity due to limitations in the data assimilated. Or perhaps there are deficiencies in model parameterizations (e.g. in surface exchange) which may be more acute during meteorological conditions that are more prevalent over the MIZ. Here we investigate these possibilities primarily by focussing on some of the aircraft observations.

It is instructive to consider a case study and here we compare the observations and ERA5 output off SE Greenland on 8 March 2018, when the aircraft spent a considerable amount of time over the pack-ice and twice crossed the ice edge. **Figure 7a** shows the aircraft-observed ice fraction plotted over a Sentinel Synthetic Aperture Radar (SAR) backscatter image, while Figure 7b shows the same data overlaid on ice fraction derived from the Advanced Microwave Scanning Radiometer 2 (AMSR2) – Spreen et al. (2008). The ice fraction is plotted using the same colour bar for the aircraft and satellite-derived observations. Most of the pack-ice is highly concentrated, with some leads and polynyas, as well as some narrow filaments of sea ice at an otherwise very narrow ice edge zone. The AMSR2 data correspond well to the SAR image, capturing the shape of the well-defined ice edge and the coherent patches of lower ice fraction, and also match the aircraft ice fraction observations reasonably well. Note the seemingly different observations from the easternmost SW to NE leg are only just below an ice fraction of 0.8. In contrast to this, Figures 7d,f,h show sea ice from the satellite-derived OSTIA analysis that is assimilated into ERA5. This has a much smoother sea-ice distribution. The gradient in ice fraction across the OSTIA MIZ is spread out over 50-80 km and does not match the abrupt ice edge seen in the aircraft observations, the SAR image or the AMSR2 data. The OSTIA product has a grid size of 1/20th degree (~6 km) and has recently undergone an upgrade in its data assimilation algorithm to capture fine-scale fronts in SST (Fiedler et al. 2019), so it should be able to resolve the observed MIZ gradient. The smoothness of the sea-ice field is due to the relatively coarse resolution of the input data, i.e. the OSI-SAF 401 data (Tonboe et al. 2016), which is based on SSMI observations from the 19 and 37 GHz channels which have along-track resolutions of 69 and 37 km, respectively.

The aircraft observations illustrate a clear division between conditions over the sea ice and over water. There is a sharp increase in T, U and SHF progressing across the ice edge into open water, with the SHF rising from 0 to ~100 W m^{-2} over 30 km for example. There are also sharp increases in T_{sfc}, RH, q and LHF (not shown). In contrast, the gradients from ERA5 are much weaker and smoother; for example, the SHF rises from 0 to ~100 W m^{-2} over ~80 km. It is evident that the overly smooth sea-ice field in ERA5 leads to overly smooth surface-layer meteorology and flux fields.

Figure 8 illustrates another case study from 16 March 2018. As before, the AMSR2 sea-ice distribution matches the SAR image and aircraft observations well, whereas the OSTIA sea-ice distribution is too smooth, with an ice edge that is smeared out over 60-100 km instead of 10-20 km. Again, there is an increase in observed U, T and SHF across the MIZ, with a sharp increase at the ice edge in the southernmost leg. The pattern is broadly captured in ERA5, but with weaker gradients and an overly smooth distribution. These cases illustrate that ERA5 does not represent the sea-ice distribution across the MIZ very well and that this directly impacts the simulated surface-layer meteorology and fluxes. Looking across all the aircraft data over the MIZ, the linear regression slope for ice fraction is only 0.64 – confirming the smearing out of ice fraction seen in Figs. 7 and 8 – and there are also low regression slopes for T, T_{sfc}, U, momentum flux, SHF and LHF (not shown). Using all of the IGP aircraft observed ice fraction data it is clear the AMSR2 ice fraction is more accurate than the OSTIA ice fraction; for example, the RMSE and linear regression slopes are 0.17/0.19 and 1.00/0.75 respectively.

ERA5 has a grid size resolution of about 30 km and is thus limited in its representation of spatial gradients. To examine whether this was the decisive factor, we have carried out a parallel evaluation of output from a set of MetUM forecasts that have a grid size of 2.2 km (see Section 2.5 for model details). **Figure 9** shows MetUM output for the 8 and 16 March 2018 case studies. Note the OSTIA surface boundary conditions used in these forecasts are from two days earlier than those used in ERA5 (although this makes little qualitative difference). In both cases studies, the MetUM suffers from similar problems as ERA5: the spatial gradients are smeared out into an overly smooth distribution and the abrupt increases in T, U and SHF at the ice edge are not captured (compare Fig. 9 to Figs. 7 and 8). Note in the 8 March case, the MetUM is uniformly about 1 K too cold and the winds are too strong over the ice (also the case for ERA5), although the MetUM does capture the high winds over water in the easternmost leg (unlike ERA5). In the 16 March case, the MetUM is uniformly about 2 K too warm (also the case for ERA5).

Table 5 provides an evaluation of the MetUM for all the marine flights. The mean, standard deviation, correlation coefficient and linear regression slope are generally very similar to those of the ERA5 comparison for the meteorological variables (c.f. Table 4). The mean fluxes are higher, giving a better match for the momentum flux, but a worse match for the SHF. The bias and RMSE are shown separately for over water and over the MIZ and they generally follow the same qualitative pattern as those of ERA5. For example, there is a negative bias in T over water and a positive bias in T over the MIZ. As for ERA5, the RMSE are greater over the MIZ than over water for all the meteorological variables (except q and wind direction). For the turbulent heat fluxes, the RMSE over the MIZ are large but, as discussed earlier, over water the large variance and relatively small dataset make this comparison unreliable. Note an evaluation of the MetUM forecasts against the buoy observations gives RMSE over water of 21 and 18 W m^{-2} for the SHF and LHF respectively, compared to 48 and 31 W m^{-2} over the MIZ (Table 5). This implies that the MetUM heat fluxes are less accurate over the MIZ than over water, in keeping with our findings for ERA5.

In short, the ERA5 and MetUM comparisons over the MIZ are remarkably similar and have the same major deficiencies. This suggests a common cause: the overly smooth sea-ice distribution in the surface boundary conditions. The evidence points to this being the primary reason for less accurate simulations over the MIZ. However, there are other factors to consider:

- The biases in the SHF and LHF over the MIZ are relatively large in magnitude for both ERA5 (Table 3) and the MetUM (Table 5). This raises questions about the surface exchange parameterization over the MIZ, which are being pursued in a separate study. Recent work has demonstrated that an improved surface exchange scheme for momentum can significantly improve forecasts for surface-layer meteorology and fluxes over the MIZ, regionally and globally (Renfrew et al. 2019b).
- The atmospheric conditions (e.g. static stability) may be different over the MIZ and over water. Even if this is the situation it seems unlikely to be the dominant factor, especially as the aircraft-based comparison uses data from legs that often cover both the MIZ and open water.
- The models may not properly resolve the heterogeneity and sharp contrasts of the MIZ. The ERA5 grid size makes it impossible to fully represent the detailed sea-ice distribution seen in the AMSR2 product; however, it should be able to represent more detail than it currently does. The limiting factor appears to be the very smooth OSTIA sea-ice distribution, which is based on the OSI-SAF 401 product.

5. Conclusions

A comprehensive evaluation of surface-layer meteorology and surface turbulent fluxes in ERA5 for winter conditions over the Iceland and Greenland Seas has been presented. Observations from three platforms – a meteorological buoy, a research vessel and a research aircraft – provide unparalleled coverage of both the ice-free ocean and the marginal ice zone (MIZ) that is independent from the reanalyses and forecasts. These observations allow the first evaluation of meteorological reanalyses that focuses on the MIZ. In general, ERA5 performs well: it captures the temporal variability very well and the spatial variability qualitatively well. The biases are significantly less than the observed standard deviations for all variables. Over water, ERA5 performs very well and broadly in line with previous evaluations of ERA-I for the subpolar seas (e.g. Harden et al. 2015). Over the MIZ, ERA5 is less accurate for almost all variables. This is clearly demonstrated by contrasting the RMSE *over water/over the MIZ:*

- for air temperature, surface temperature, wind speed: 0.78/1.00 K, 0.47/2.94 K, 1.77/2.42 m s^{-1} from the aircraft comparison (Table 4); and
- for momentum flux, sensible heat flux, and latent heat flux: 0.077/0.098 N m^{-2}, 20.9/35.2 W m^{-2}, and 14.2/20.1 W m^{-2} from the ship comparison (Table 3).

There is also a bias in surface temperature over the MIZ of about 1 K in the aircraft comparison. A parallel evaluation of a set of forecasts from a 2-km configuration of the MetUM has similar findings.

The primary cause of the lower accuracy over the MIZ is an overly smooth sea-ice distribution in the prescribed surface boundary conditions. These use the OSTIA SST and sea-ice analysis which takes sea ice concentration from the OSI-SAF 401 product. The OSTIA sea-ice concentration gradient is too weak compared to aircraft observations, SAR imagery or satellite observations from AMSR2. This has an impact on the surface-layer meteorology and fluxes, which also have gradients that are too weak across the MIZ. It is likely that the surface exchange parameterization over the MIZ also has some limitations, but these appear secondary.

Our findings suggest the hypothesis that a more accurate and precise sea-ice concentration would yield a better performance from meteorological reanalyses or forecasts for surface-layer meteorology and fluxes in the marginal ice zone. There is evidence from idealised modelling studies that the atmospheric surface layer is strongly impacted by the sea-ice distribution both locally and for hundreds of kilometres downstream (e.g. Liu et al. 2006; Chechin et al. 2013; Gryschka et al. 2008; Müller et al. 2017; Batrak and Müller 2018); and case studies have shown an improved sea-ice distribution can improve the surface-layer meteorology (e.g. Outten et al. 2009; Smith et al. 2013). Verifying this hypothesis for the IGP data or, more generally, for the subpolar North Atlantic region should be a next step and would provide further motivation for improving the sea-ice data used as initial conditions in meteorological reanalyses and forecasts.

Acknowledgments

This study was part of the Iceland Greenland Seas Project. Funding was from the NERC AFIS grant (NE/N009754/1), the ALERTNESS (Advanced models and weather prediction in the Arctic: enhanced capacity from observations and polar process representations) project (Research Council of Norway project number 280573), the Trond Mohn Foundation (BFS2016REK01) and the National Science Foundation grant OCE-1558742. The Leosphere WindCube v2 and the Wavescan buoy are part of the OBLO (Offshore Boundary Layer Observatory) infrastructure funded by the Research Council of Norway (project number 227777). We would like to thank all those involved in the field work associated with the IGP, including L. Papritz, E. Kolstad, and M. Hallerstig. We would also like to thank the reviewers for their comments. Fully quality-controlled meteorological data sets from the buoy, research vessel and aircraft are available at CEDA (www.ceda.ac.uk); and for the buoy data at thredds.met.no. This paper contains modified Copernicus Climate Change Service Information (2020). The MetUM simulations were carried out on MONSooN, a collaborative facility supplied under the Joint Weather and Climate Research Programme, a strategic partnership between the Met Office and the NERC. Visualisation software used: Matlab, Matplotlib: A 2D Graphics Environment (J.D. Hunter, Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, 2007), Cartopy v0.17 (16-Nov-2018, Met Office, UK, https://github.com/SciTools/cartopy/archive/v0.17.0.tar.gz) and Iris v2.2 (11-Oct-2018, Met Office, UK, https://github.com/SciTools/iris/archive/v2.2.0.tar.gz). This study is a contribution to the Year of Polar Prediction (YOPP), a flagship activity of the Polar Prediction Project, initiated by the World Weather Research Programme of the World Meteorological Organisation.

References

Batrak, Y., & Müller, M. 2018: Atmospheric response to kilometer-scale changes in sea ice concentration within the marginal ice zone. *Geophysical Research Letters*, **45,** 6702–6709 doi:10.1029/2018GL078295

Bauer, P., Thorpe, A. and Brunet, G., 2015. The quiet revolution of numerical weather prediction. *Nature*, **525**,47-55.

Belmonte Rivas, M. and Stoffelen, A., 2019. Characterizing ERA-Interim and ERA5 surface wind biases using ASCAT. *Ocean Science*, 15, 831–852. doi: 10.5194/os-15-831-2019

Bourassa, M.A., S.T. Gille, C. Bitz, D. Carlson, I. Cerovecki, C. A. Clayson, M. Cronin, W.M. Drennan, C.W. Fairall, R.N. Hoffman, G. Magnusdottir, R.T. Pinker, I.A. Renfrew, M. Serreze, K. Speer, L.D. Talley, G.A. Wick, 2013: High-latitude ocean and sea ice surface fluxes: Challenges for climate research, *Bulletin of the American Meteorological Society*, **94,** 403-423. doi:10.1175/BAM S-D-11-00244.1

Bromwich, D.H., A.B. Wilson, L.S. Bai, G.W. Moore and P. Bauer 2016: A comparison of the regional Arctic System Reanalysis and the global ERA-Interim Reanalysis for the Arctic. *Quarterly Journal of the Royal Meteorological Society,* **142**, 644-658.

Buckley, M.W. and J. Marshall, 2016: Observations, inferences, and mechanisms of the Atlantic Meridional Overturning Circulation: A review. *Reviews of Geophysics*, **54**, 5-63.

Bush, M., et al. 2020: The first Met Office Unified Model–JULES Regional Atmosphere and Land configuration, RAL1, *Geosci. Model Dev.*, 13, 1999–2029, doi:10.5194/gmd-13-1999-2020

Chechin, D. G., C. Lüpkes, I. A. Repina, and V. M. Gryanik 2013: Idealized dry quasi 2-D mesoscale simulations of cold-air outbreaks over the marginal sea ice zone with fine and coarse resolution, *J. Geophys. Res. Atmos*., **118**, 8787–8813, doi:10.1002/jgrd.50679.

Dee, D. P., et al. 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society,* **137**, 553-597.

Donlon, C.J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E. and Wimmer, W., 2012. The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sensing of Environment*, **116**, 140-158.

Drennan W.M., Zhang J.A., French J.R., McCormick C., Black P.G. 2007: Turbulent fluxes in the hurricane boundary layer. Part II: Latent heat flux. *J. Atmospheric Sciences*, **64**, 1103–1115.

Dukhovskoy, D.S., M.A. Bourassa, G.N. Petersen and J. Steffen 2017: Comparison of the ocean surface vector winds from atmospheric reanalysis and scatterometer-based wind products over the Nordic Seas and the northern North Atlantic and their application for ocean modeling. *J. Geophysical Research: Oceans*, **122**, 1943-1973. doi:10.1002/2016JC012453

Duscha, C., M. Bakhoday Paskyabi and J. Reuder 2020: Statistic and coherence response of ship-based Lidar observations to motion compensation, accepted for publication in Journal of Physics: Conference Series.

DuVivier, A.K. and Cassano, J.J., 2013. Evaluation of WRF model resolution on simulated mesoscale winds and surface fluxes near Greenland. *Monthly Weather Review*, *141*(3), 941-963.

Elvidge, A.D., I.A. Renfrew, A.I. Weiss, I.M. Brooks, T.A. Lachlan-Cope, and J.C. King 2016: Observations of surface momentum exchange over the marginal ice zone and recommendations for its parameterization, *Atmospheric Chemistry and Physics,* **16,** 1545-1563. doi:10.5194/acp-16-1545-2016

Fiedler, E., T. Lachlan-Cope, I. A. Renfrew, J. C. King, 2010: Convective heat transfer of thin-ice covered polynyas, *J. Geophysical Research,* **115**, C10051, doi:10.1029/2009JC005797.

Fiedler, E.K., Mao, C., Good, S.A., Waters, J. and Martin, M.J., 2019. Improvements to feature resolution in the OSTIA sea surface temperature analysis using the NEMOVAR assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, **145**, 3609-3625.

Graham, R.M., S.R. Hudson and M. Maturilli 2019a: Improved performance of ERA5 in Arctic gateway relative to four global atmospheric reanalyses. *Geophysical Research Letters*, **46**, 6138-6147. doi: 10.1029/2019GL082781

Graham, R.M., Cohen, L., Ritzhaupt, N., Segger, B., Graversen, R.G., Rinke, A., Walden, V.P., Granskog, M.A. and Hudson, S.R., 2019b. Evaluation of six atmospheric reanalyses over Arctic sea ice from winter to early summer. *J. Climate*, **32**, 4121-4143.

Grunwald, J., Kalthoff, N., Corsmeier, U., & Fiedler, F. 1996: Comparison of areally averaged turbulent fluxes over non-homogeneous terrain: Results from the EFEDA-field experiment. *Boundary-Layer Meteorology*, **77**, 105–134. doi:10.1007/BF00119574

Gryschka, M., C. Drüe, D. Etling, and S. Raasch 2008: On the influence of sea-ice inhomogeneities onto roll convection in cold-air outbreaks, *Geophys. Res. Lett*., **35**, L23804, doi:10.1029/2008GL035845

Harden, B. E., I. A. Renfrew and G. N. Petersen, 2011: A climatology of wintertime barrier winds off southeast Greenland, *J. Climate*, **24,** 4701-4717.

Harden, B. E., I. A. Renfrew and G. N. Petersen, 2015: Meteorological buoy observations from the central Iceland Sea, *J. Geophysical Research,* **120,** doi:10.1002/2014JD022584

Hersbach, H. et al. 2020: The ERA5 Global Reanalysis, *Quarterly Journal of the Royal Meteorological Society,* doi: 10.1002/qj.3803.

Jakobson, E., T. Vihma, T. Palo, L. Jakobson, H. Keernik, and J. Jaagus 2012: Validation of atmospheric reanalyses over the central arctic ocean, *Geophysical Research Letters*, **39**, L10,802, doi:10.1029/2012GL051591.

Jones, R.W., I.A. Renfrew, A. Orr, B.G.M. Webber, D.M. Holland and M.A. Lazzara, 2016: Evaluation of four global reanalysis products using in situ observations in the Amundsen Sea Embayment, Antarctica. *J. Geophysical Research*, **121**, 6240-6257, doi:10.1002/2015JD024680

Jónsson and Valdimarsson 2004: A new path for the Denmark Strait overflow water from the Iceland Sea to Denmark Strait. *Geophysical Research Letters,* **31,** doi:10.1029/2003GL019214

Jung, T., N. Gordon, P. Bauer, D. Bromwich, M. Chevallier, J. Day, J. Dawson, F. Doblas-Reyes, C. Fairall, H. Goessling, M. Holland, J. Inoue, T. Iversen, S. Klebe, P. Lemke, M. Losch, A. Makshtas, B. Mills, P. Nurmi, D. Perovich, P. Reid, I. A. Renfrew, G. Smith, G. Svensson, M. Tolstykh, and Q. Yang, 2016: Advancing polar prediction capabilities on daily to seasonal time scales. *Bulletin of the American Meteorological Society,* **97,** 1631-1647. doi:10.1175/BAMS-D-14-00246.1

Large, W. G., J. Morzel, and G. B. Crawford 1995: Accounting for surface wave distortion of the marine wind profile in low-level ocean storms wind measurements, *J. Physical Oceanography*, **25**, 2959–2971.

Lawrence, H., Bormann, N., Sandu, I., Day, J., Farnan, J. and Bauer, P., 2019: Use and impact of Arctic observations in the ECMWF numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society,* **145**, 3432-3454.

Li, M., Liu, J., Wang, Z., Wang, H., Zhang, Z., Zhang, L. and Yang, Q., 2013: Assessment of sea surface wind from NWP reanalyses and satellites in the Southern Ocean. *J. Atmospheric and Oceanic Technology,* **30**, 1842-1853.

Lindsay, R., M. Wensnahan, A. Schweiger and J. Zhang, J. 2014: Evaluation of seven different atmospheric reanalysis products in the Arctic. *J. Climate*, **27**, 2588-2606.

Liu, A.Q., G.W.K. Moore, K. Tsuboki and I. A. Renfrew, 2006: The effect of the sea-ice zone on the development of boundary-layer roll clouds during cold-air outbreaks, *Boundary-Layer Meteorology*, **118**, 557-581.

Lozier, M.S., et al. 2019: A sea change in our view of overturning in the subpolar North Atlantic. *Science*, **363**, 516-521.

Lüpkes, C., T. Vihma, E. Jakobson, G. König-Langlo, and A. Tetzlaff 2010: Meteorological observations from ship cruises during summer to the central Arctic: A comparison with reanalysis data, *Geophysical Research Letters,* **37**, L09810, doi:10.1029/2010GL042724.

Mauritzen, C. 1996: Production of dense overflow waters feeding the N Atlantic across the Greenland-Scotland Ridge, Evidence for a revised circulation scheme, *Deep-Sea Research I,* **43,** 769-806.

Moore, G. W. K., R. S. Pickart, and I. A. Renfrew 2008: Buoy observations from the windiest location in the world ocean, Cape Farewell, Greenland, *Geophysical Research Letters*, 35, L18,802, doi:10.1029/2008GL034845

Moore, G.W.K, I.A. Renfrew, B.E. Harden, S.H. Mernild, 2015: The impact of resolution on the representation of Greenland barrier winds and katabatic flows, *Geophysical Research Letters*, **42**, doi:10.1002/2015GL063550

Moore, G.W.K., D.H. Bromwich, A.B. Wilson, I. Renfrew, L. Bai, 2016: Arctic System Reanalysis improvements in topographically-forced winds near Greenland, *Quarterly Journal of the Royal Meteorological Society*, 142, 2033-2045. doi:10.1002/qj.2798

Müller, M., Batrak, Y., Kristiansen, J., Køltzow, M.A., Noer, G. and Korosov, A., 2017: Characteristics of a convective-scale weather forecasting system for the European Arctic. *Monthly Weather Review*, **145**, 4771-4787.

Nygård, T., T. Vihma, G. Birnbaum, J. Hartmann, J. King, T. Lachlan-Cope, R. Ladkin, C. Lüpkes, and A. Weiss 2016: Validation of eight atmospheric reanalyses in the Antarctic Peninsula region. *Quarterly Journal of the Royal Meteorological Society*, **142**, 684-692.

Outten, S.D., Renfrew, I.A. and Petersen, G.N., 2009: An easterly tip jet off Cape Farewell, Greenland. II: Simulations and dynamics. *Quarterly Journal of the Royal Meteorological Society*, **135**, 1934-1949.

Papritz, L. and T. Spengler, 2017: A Lagrangian climatology of wintertime cold air outbreaks in the Irminger and Nordic Seas and their role in shaping air–sea heat fluxes. *J. Climate*, **30,** 2717-2737.

Petersen, G.N. and I.A. Renfrew, 2009: Aircraft-based observations of air-sea fluxes over Denmark Strait and the Irminger Sea during high wind speed conditions, *Quarterly Journal of the Royal Meteorological Society,* **135,** 2030-2045.

Pickart, R. S. and Spall, M. A. 2007: Impact of Labrador Sea convection on the North Atlantic meridional overturning circulation. *J. Physical Oceanography*, **37**, 2207-2227.

Renfrew, I.A., G.W.K. Moore, P.S. Guest, and K. Bumke, 2002: A comparison of surface-layer and surface turbulent-flux observations over the Labrador Sea with ECMWF analyses and NCEP reanalyses, *J. Physical Oceanography*, **32**, 383-400.

Renfrew, I.A., G.W.K. Moore, J.E. Kristjánsson, H. Ólafsson, S.L. Gray, G.N. Petersen, K. Bovis, P.R.A. Brown, I. Føre, T. Haine, C. Hay, E.A. Irvine, A. Lawrence, T. Ohigashi, S. Outten, R.S. Pickart, M. Shapiro, D. Sproson, R. Swinbank, A. Woolley, S. Zhang, 2008: The Greenland Flow Distortion experiment, *Bulletin of the American Meteorological Society,* **89**, 1307-1324.

Renfrew, I.A., G.N. Petersen, D.A.J. Sproson, G.W.K. Moore, H. Adiwidjaja, S. Zhang, and R. North, 2009: A comparison of aircraft-based surface-layer observations over Denmark Strait and the Irminger Sea with meteorological analyses and QuikSCAT winds, *Quarterly Journal of the Royal Meteorological Society,* **135,** 2046-2066.

Renfrew, I.A., R. S. Pickart, K. Våge, G. W. K. Moore, et al. 2019a: The Iceland Greenland Seas Project, *Bulletin of the American Meteorological Society*, **100**, 1795–1817. doi:10.1175/BAMS-D-18-0217.1

Renfrew, I.A., Elvidge, A.D. and Edwards, J.M., 2019b: Atmospheric sensitivity to marginal-ice-zone drag: Local and global responses. *Quarterly Journal of the Royal Meteorological Society*, **145**, 1165-1179.

Semper, S., K. Våge, R. S. Pickart, H. Valdimarsson, D. J. Torres, and S. Jónsson 2019: The emergence of the North Icelandic Jet and its evolution from northeast Iceland to Denmark Strait, *J. Physical Oceanography,* **49,** 2499-2521.

Sergeev, D. E., I. A. Renfrew, T. Spengler, S. R. Dorling, 2017: Structure of a shear-line polar low, *Quarterly Journal of the Royal Meteorological Society*, 143, 12-26. 10.1002/qj.2911

Spreen, G., L. Kaleschke, and G.Heygster, 2008: Sea ice remote sensing using AMSR-E 89 GHz channels. *J. Geophysical Research*, 113, C02S03, doi:10.1029/2005JC003384

Swift, J.H. and K. Aagaard, 1981: Seasonal transitions and water mass formation in the Iceland and Greenland seas. *Deep Sea Research I*, **28**, 1107-1129.

Tonboe, R., J. Lavelle, R.-H. Pfeiffer and E. Howe 2017: Product User Manual for OSI SAF Global Sea Ice Concentration, Product OSI-401-b, Version 1.6. Available at http://www.osi-saf.org/

Våge, K., R.S. Pickart, M.A. Spall, H. Valdimarsson, S. Jónsson, D.J. Torres, S. Østerhus, and T. Eldevik, 2011: Significant role of the North Icelandic Jet in the formation of Denmark Strait overflow water. *Nature Geoscience*, **4**, 723-727.

Våge, K., R.S. Pickart, M.A. Spall, G.W.K. Moore, H. Valdimarsson, D.J. Torres, S.Y. Erofeeva, and J.E.Ø. Nilsen, 2013: Revised circulation scheme north of the Denmark Strait. *Deep Sea Research I*, **79**, 20-39.

Våge, K., L. Papritz, L. Håvik, M.A. Spall, and G.W.K Moore, 2018: Ocean convection linked to the recent ice edge retreat along east Greenland. *Nature Communications*, **9**, 1287.

Vihma, T., R. Pirazzini, I. Fer, I. A. Renfrew, J. Sedlar, M. Tjernström, C. Lüpkes, T. Nygård, D. Notz, J. Weiss, D. Marsan, B. Cheng, G. Birnbaum, S. Gerland, D. Chechin, and J. C. Gascard, 2014: Advances in understanding and parameterization of small-scale physical processes in the marine Arctic climate system: a review, *Atmospheric Chemistry and Physics*, **14,** 9403-9450, doi:10.5194/acp-14-9403-2014

Wang, C., Graham, R.M., Wang, K., Gerland, S. and Granskog, M.A., 2019. Comparison of ERA5 and ERA-Interim near-surface air temperature, snowfall and precipitation over Arctic sea ice: effects on sea ice thermodynamics and evolution. *The Cryosphere*, 13, 1661-1679. doi: 10.5194/tc-13-1661-2019

Weiss, A. I., King, J., Lachlan-Cope, T., and Ladkin, R. 2011: On the effective aerodynamic and scalar roughness length of Weddell Sea ice, *J. Geophysical Research,* **116**, D19119, doi:10.1029/2011JD015949

Wesslén, C., Tjernström, M., Bromwich, D.H., Boer, G.D., Ekman, A.M., Bai, L.S. and Wang, S.H., 2013. The Arctic summer atmosphere: an evaluation of reanalyses using ASCOS data. *Atmospheric Chemistry and Physics*, **14**, 2605–2624. doi:10.5194/acp-14-2605-2014

Table 1 | Accuracy estimates for selected instruments on each observing platform.

Platform	Instruments	Measured quantity and (estimated accuracy)
Seawatch Wavescan buoy	Vaisala HMP155 Vaisala PTB330A Young Ultrasonic Anemometer (86106)	temperature (±0.2 K) and RH (±2%) pressure (±0.25 hPa) wind speed (±2% of value ±0.1 m s^{-1}, so ±0.3 m s^{-1} at 10 m s^{-1})
Mooring	Sea-Bird Scientific SBE37 MicroCat	8-m temperature (±0.002 K)
Research vessel	WeatherPak Leosphere Windcube Lidar Seabird SBE38 bow thermometer	temperature (±0.2 K), RH (±2 %) and windspeed (±0.3 m s^{-1}) horizontal wind speed (±0.2 m s^{-1}, after averaging) sea surface temperature (±0.001 K)
Research aircraft	Rosemount thermometer BAT turbulence probe Buck cooled mirror hygrometer Heimann infrared thermometer Radar altimeter Eppley PSP pyranometers	temperature (±0.3 K) wind speed (±0.3 m s^{-1}) dewpoint T (±0.25 K to ±1 K with decreasing T) surface temperature (±1 K) altitude (±3 m) shortwave radiation (±3%)

Note to obtain wind speeds from the research vessel and aircraft, data on the location and platform motion has to be combined with the measurements from the named instruments (e.g. Duscha et al. 2020; Renfrew et al. 2008). For brevity accuracy estimates are only given for the derived wind speed. The uncertainty in aircraft winds given here is higher than previous studies – Fiedler et al. (2010), Weiss et al. (2011) – due to the post-flight calibration that was required. Note an uncertainty in dewpoint temperature of ±0.5 K is equivalent to an uncertainty of ±0.08 g kg^{-1} in specific humidity and ±3% in relative humidity (RH) at the air temperatures observed.

Table 2 | Comparison of hourly buoy observations and ERA5 output.

	T_{2m} (^{o}C)	**SST** (^{o}C)	**q_{2m}** (g kg^{-1})	**RH_{2m}** (%)	**U_{10m}** (m s^{-1})	**WD** (deg)	**τ** (N m^{-2})	**SHF** (W m^{-2})	**LHF** (W m^{-2})
Mean: buoy	-1.43	(0.26)	2.92	83.0	7.54	122	0.112	23.6	28.1
ERA5	-1.38	0.33	2.78	78.3	8.15	106	0.164	21.6	34.9
Std. dev.: buoy	2.76	(0.11)	0.85	11.7	3.00	94	0.095	39.9	29.4
ERA5	2.55	0.28	0.86	12.8	3.65	96	0.179	41.9	33.6
Correlation coefficient	*0.92*	-	*0.95*	*0.85*	*0.92*	*0.97*	*0.89*	*0.90*	*0.91*
Slope	*0.85*	-	*0.95*	*0.93*	*1.12*	*0.99*	*1.69*	*0.94*	*1.04*
N	1156	1665	1556	1665	1556	1484	1556	1556	1556
Bias error	0.05	(0.07)	**-0.14**	**-4.7**	**0.61**	**-15**	**0.052**	-1.7	**6.9**
RMSE	1.11	(0.31)	0.31	8.2	1.62	28	0.116	18.7	15.7

The variables are: temperature at 2 m, T_{2m} (^{o}C); sea surface temperature, SST (^{o}C); specific humidity at 2 m, q_{2m} (g kg^{-1}); relative humidity at 2 m, RH_{2m} (%); wind speed at 10 m, U_{10m} (m s^{-1}); wind direction, WD (deg.); surface momentum flux, τ (N m^{-2}); surface sensible heat flux, SHF (W m^{-2}); and surface latent heat flux, LHF (W m^{-2}). Note the wind direction time series are filtered to remove data where the difference is greater than 270^{o}. The observed surface turbulent fluxes are calculated using the COARE3 algorithm. Recall the observations of SST are from a depth of 8 m and so are shown with brackets. The non-dimensional *correlation coefficient* and *linear regression slope* are shown in italics when statistically significant. The number of data points, N, plus the bias (model – observations) and the root-mean-square error (RMSE) are shown for all data. The bias is bold when statistically significant at the 95% level using a one-sided T-test.

Table 3 | Comparison of ship-based observations and ERA5 output.

	T_{2m} (°C)	SST (°C)	q_{2m} (g kg^{-1})	RH_{2m} (%)	U_{10m} (m s^{-1})	WD (deg)	τ (N m^{-2})	SHF (W m^{-2})	LHF (W m^{-2})
Mean: ship	-2.09	0.62	2.80	83.7	8.84	125	0.182	42.2	41.1
ERA5	-1.63	0.61	2.81	80.0	8.54	118	0.177	28.8	37.8
Std dev: ship	3.16	1.12	0.88	12.2	3.66	96	0.173	55.6	36.3
ERA5	3.57	1.10	1.01	12.6	3.50	99	0.170	54.8	37.8
Correlation coefficient	*0.97*	*0.80*	*0.97*	*0.85*	*0.92*	*0.90*	*0.89*	*0.94*	*0.93*
Slope	*1.09*	*0.78*	*1.10*	*0.88*	*0.88*	*0.94*	*0.87*	*0.93*	*1.00*
N: Over water	527	658	527	527	527	530	527	527	527
Over MIZ	84	92	84	84	84	76	84	84	84
Bias: Over water	**0.49**	-0.02	0.02	**-3.4**	**-0.20**	**-7**	-0.004	**-11.8**	**-2.7**
Over MIZ	0.28	(0.11)	-0.05	**-5.6**	**-0.88**	**-4**	-0.004	**-23.2**	**-7.3**
RMSE: Over water	0.99	0.71	0.27	7.4	1.33	42	0.077	20.9	14.2
Over MIZ	1.41	(0.73)	0.31	9.1	2.02	58	0.098	35.2	20.1

Variables and statistics are the same as Table 2. Note the ship-based SST observations over the MIZ will not be representative of a grid-box value so are bracketed. The number of data points, N, the bias (model – observations) and the RMSE are shown separately for observations over water (i.e. ice-free ocean) and over the MIZ. The bias is bold when statistically significant at the 95% level using a one-sided T-test.

Table 4 | Comparison of aircraft-based observations and ERA5 output for all marine flights.

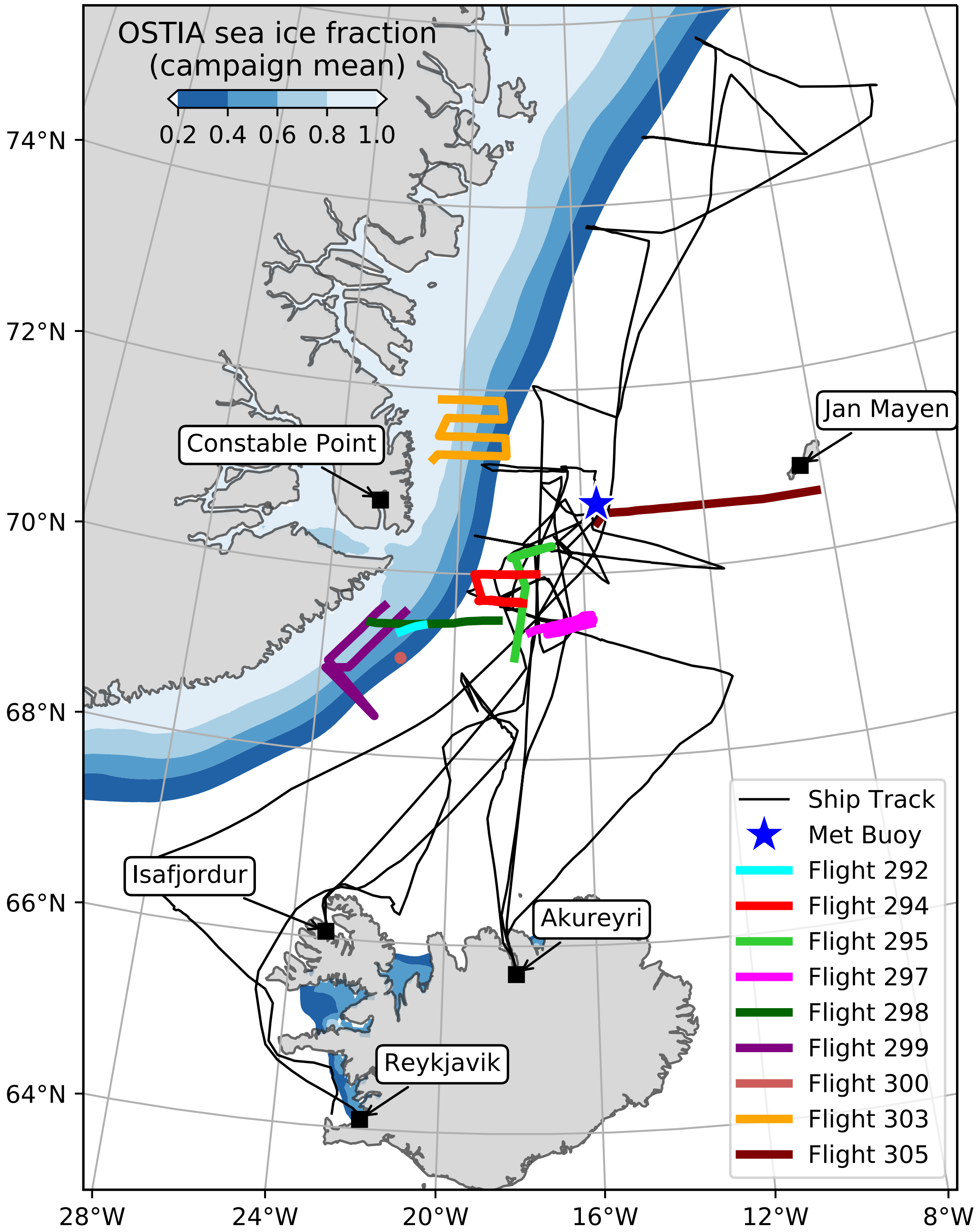
	T (°C)	**T_{sfc}** (°C)	**q** (g kg^{-1})	**RH** (%)	**U** (m s^{-1})	**WD** (deg)	**τ** (N m^{-2})	**SHF** (W m^{-2})	**LHF** (W m^{-2})	**Ice Frac.**
Mean: aircraft	263.7	267.3	1.51	80.0	9.27	6.6	0.214	41.3	22.4	0.40
ERA5	264.0	267.8	1.41	71.2	9.28	10.8	0.180	43.0	-	0.37
Std dev: aircraft	5.6	6.3	0.59	10.2	3.63	35.5	0.151	57.1	28.4	0.42
ERA5	5.1	5.7	0.54	7.6	3.61	31.5	0.128	49.6	-	0.38
Correlation coefficient	*0.99*	*0.94*	*0.96*	*0.51*	*0.82*	*0.96*	*0.70*	*0.83*	-	*0.93*
Slope	*0.89*	*0.85*	*0.89*	*0.38*	*0.82*	*0.85*	*0.59*	*0.72*	-	*0.82*
N: Over water	69	69	69	69	69	69	44	44	-	69
Over MIZ	88	88	88	88	88	86	83	88	-	88
Bias: Over water	-0.10	-	**-0.16**	**-6.4**	**-0.72**	**5.8**	(-0.057)	(-0.6)	**-**	0.01
Over MIZ	**0.65**	**1.02**	**-0.04**	**-7.1**	**0.58**	**2.8**	-0.022	2.9	**-**	**-0.06**
RMSE: Over water	0.78	0.47	0.25	12.3	1.77	9.1	(0.129)	(30.8)	-	0.04
Over MIZ	1.0	2.94	0.13	10.4	2.42	12.2	0.107	31.9	-	0.21

The variables are: temperature, T (K), surface temperature, T_{sfc} (K), specific humidity, q (g kg^{-1}), relative humidity, RH (%), wind speed, U (m s^{-1}), momentum flux, τ (N m^{-2}), sensible heat flux, SHF (W m^{-2}), latent heat flux, LHF (W m^{-2}) and ice fraction. All variables are at flight-level, except for T_{sfc} and the ice fraction. Flight-level ERA5 LHF are not available. The observed surface turbulent fluxes are calculated using the eddy covariance method; there is higher uncertainty in the comparison of these over water due to there being relatively few data points (hence the bias and RMSE are bracketed). The mean, standard deviation and non-dimensional *correlation coefficient* and *linear regression slope* (in italics) are shown for all of the observations. The number of data points, N, the bias (model – observations) and the RMSE are shown separately for observations over water (i.e. ice-free ocean) and over the MIZ. Points are defined as over the MIZ when the observed ice fraction > 0. The bias is bold when statistically significant at the 95% level using a one-sided T-test. The T_{sfc} bias over water is not shown because ERA5 data are used for calibrating the airborne observations.

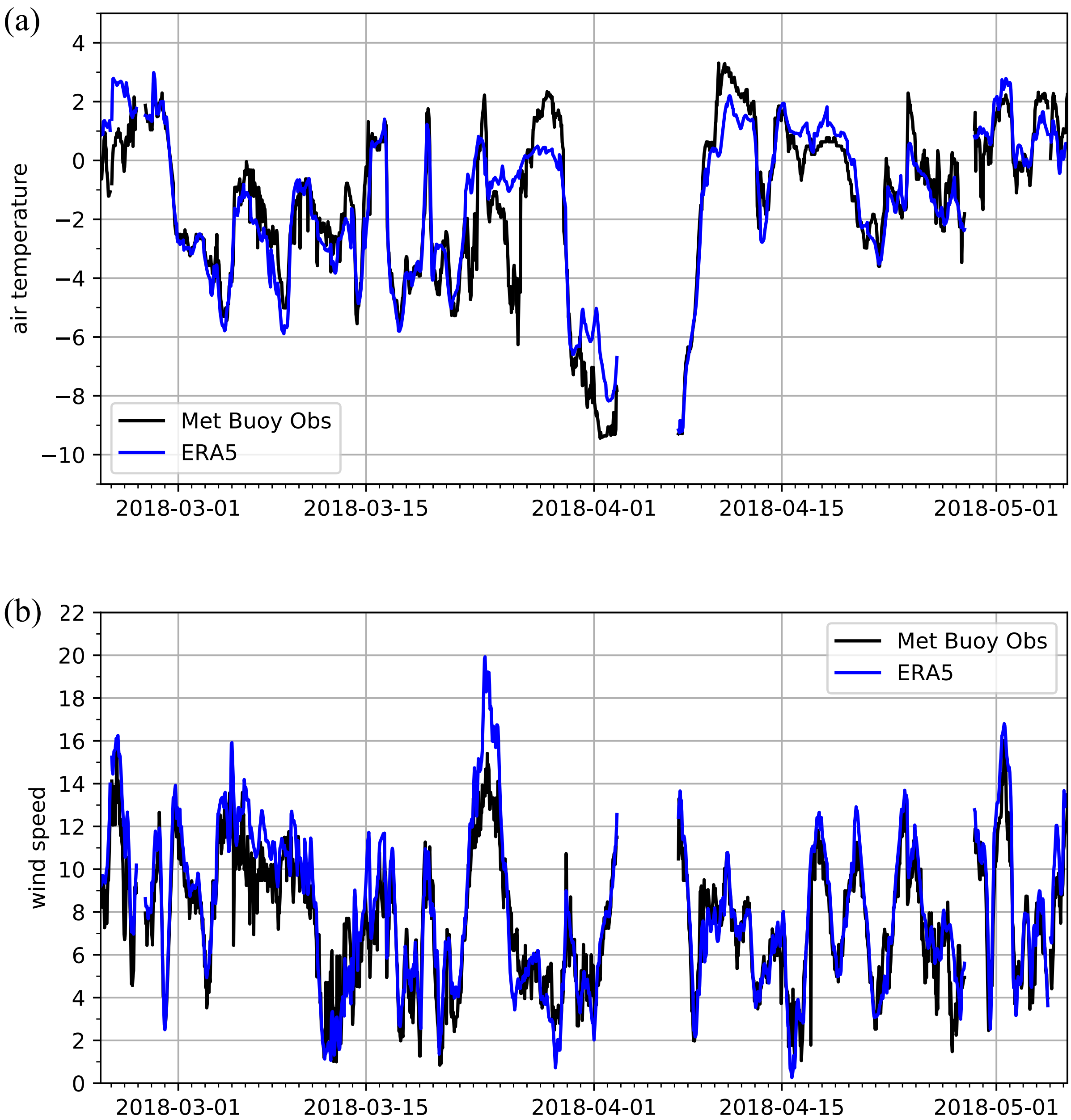
Table 5 | Comparison of aircraft-based observations and MetUM output for all marine flights.

	T (°C)	**T_{sfc}** (°C)	**q** (g kg^{-1})	**RH** (%)	**U** (m s^{-1})	**WD** (deg)	**τ** (N m^{-2})	**SHF** (W m^{-2})	**LHF** (W m^{-2})	**Ice Frac.**
Mean: aircraft	263.7	267.3	1.51	80.0	9.27	6.6	0.214	41.3	22.4	0.40
MetUM	263.6	267.8	1.35	70.0	9.32	9.3	0.205	61.0	49.2	0.34
Std dev: aircraft	5.6	6.3	0.59	10.2	3.63	35.5	0.151	57.1	28.4	0.42
MetUM	5.1	5.7	0.54	9.8	3.89	33.2	0.157	46.7	31.3	0.36
Correlation coefficient	*0.99*	*0.95*	*0.94*	*0.53*	*0.83*	*0.92*	*0.65*	*0.73*	*0.47*	*0.93*
Slope	*0.91*	*0.86*	*0.86*	*0.50*	*0.89*	*0.86*	*0.67*	*0.59*	*0.52*	*0.79*
N: Over water	69	69	69	69	69	69	44	44	44	69
Over MIZ	88	88	88	88	88	86	83	88	82	88
Bias: Over water	**-0.61**	-	**-0.23**	**-6.0**	**-1.06**	2.8	(-0.043)	(2.0)	**(35.4)**	**0.02**
Over MIZ	**0.39**	**1.18**	**-0.11**	**-9.6**	**0.91**	2.5	0.009	**28.6**	**22.2**	**-0.11**
RMSE: Over water	0.93	0.78	0.32	11.4	1.79	16.0	(0.159)	(35.1)	(54.3)	0.05
Over MIZ	0.97	2.60	0.20	13.5	2.46	12.4	0.111	48.1	31.1	0.22

Variables and statistics are the same as Table 4. In brief, the mean, standard deviation and non-dimensional *correlation coefficient* and *linear regression slope* (in italics) are shown for all of the observations. The number of data points, N, the bias (model – observations) and the RMSE are shown separately for observations over water and over the MIZ. The bias is bold when statistically significant at the 95% level using a one-sided T-test.



785 **Figure 1 |** Map of the Iceland and Greenland Seas with sea-ice fraction averaged over the field campaign period.
786 Overlaid are the positions of the low-level components of the research flights (coloured by flight number); the
787 track of the research vessel (black) and the position of the meteorological buoy (blue star). Some key locations are
788 noted.



789 **Figure 2 |** Time series of (a) 2-m air temperature (°C) and (b) 10-m wind speed (m s^{-1}) derived from the
790 meteorological buoy (black) and extracted from ERA5 (blue) from 23 February to 7 May 2018.

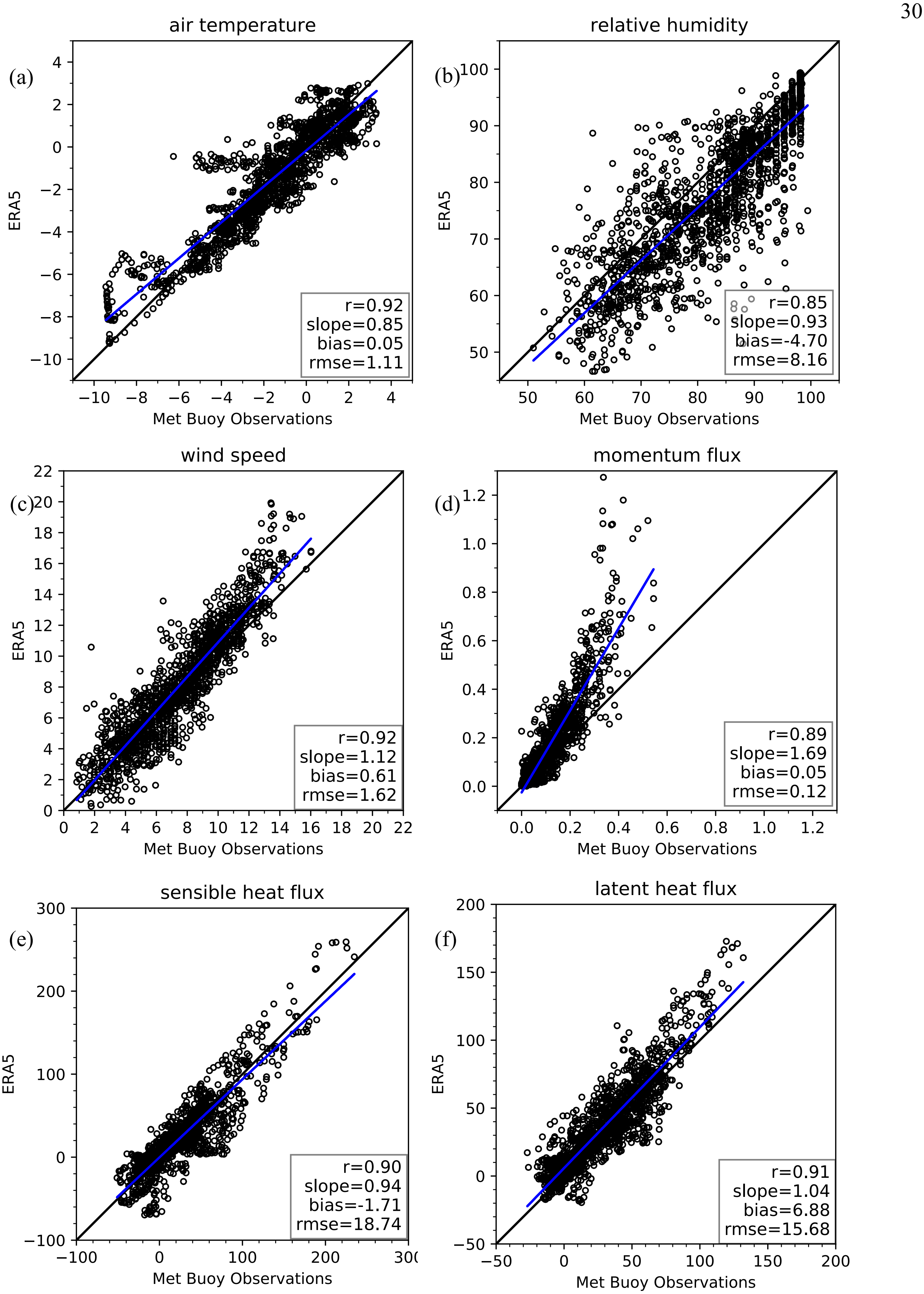


Figure 3 | Scatter plots of buoy observations versus ERA5 data for (a) 2-m air temperature (°C), (b) 2-m relative humidity (%), (c) 10-m wind speed (m s^{-1}), (d) surface momentum flux (N m^{-2}), (e) surface sensible heat flux (W m^{-2}), and (f) surface latent heat flux (W m^{-2}). The correlation coefficient (r), linear regression slope, bias and root-mean-square error (rmse) are noted in each panel.

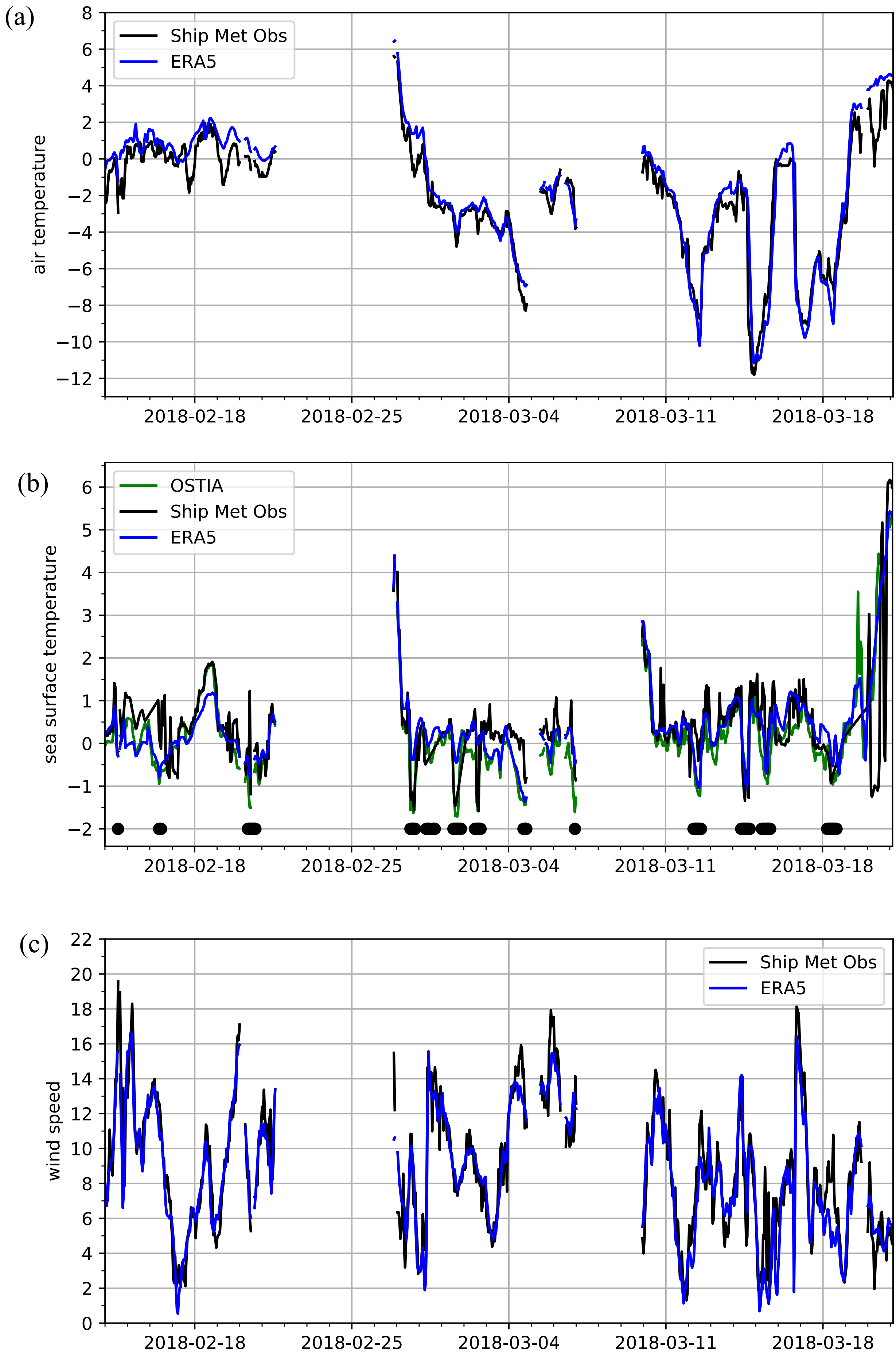


Figure 4 | Time series of (a) 2-m air temperature (°C), (b) sea surface temperature (°C), and (c) 10-m wind speed (m s^{-1}) derived from the ship-based observations (black) and extracted from ERA5 (blue) from 14 February to 21 March 2018. OSTIA SST (green) and times when the ice fraction > 0 (black dots) are shown in (b). Periods in port are not shown.

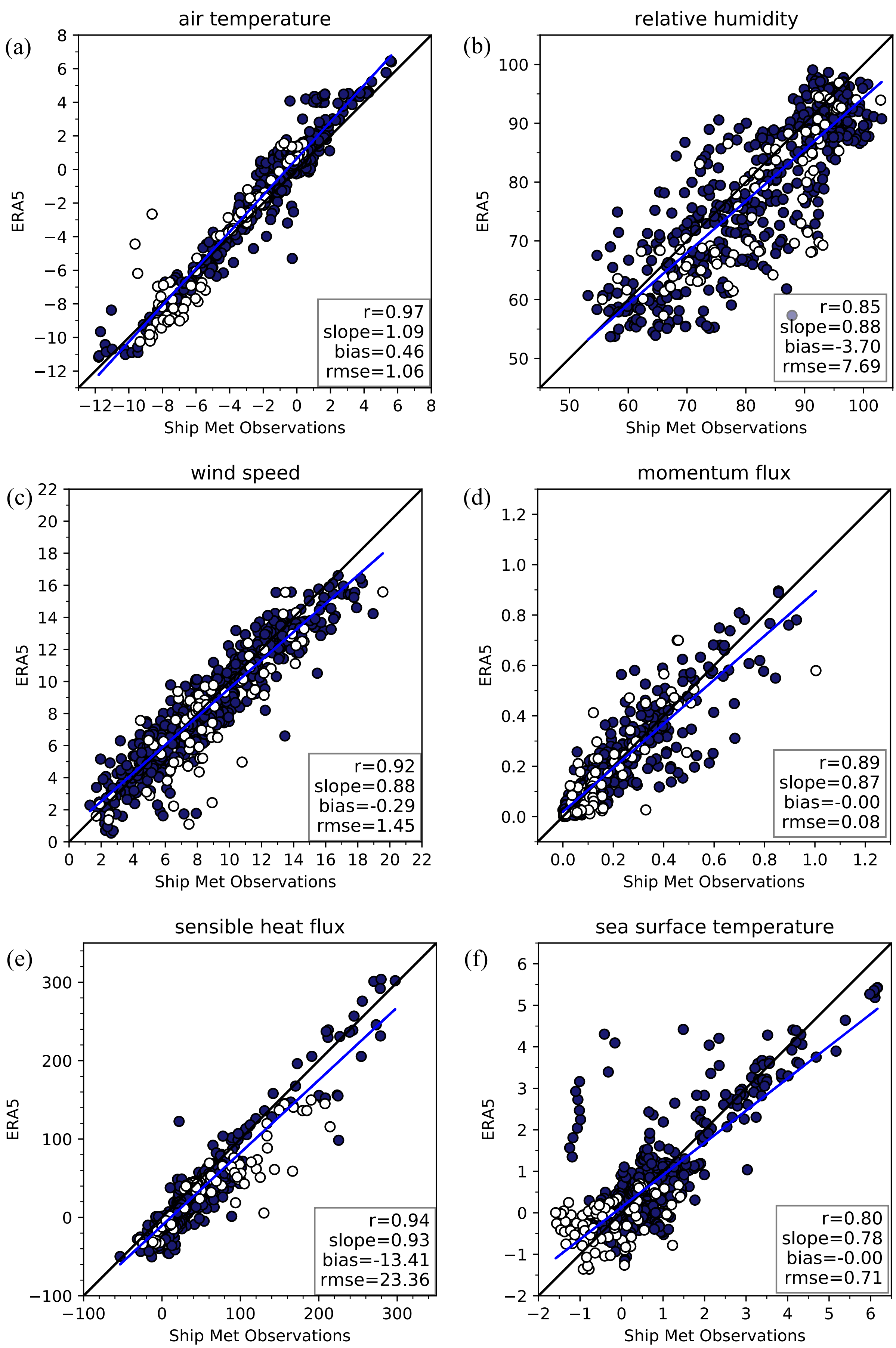


Figure 5 | Scatter plots of ship-based observations versus ERA5 data for (a) air temperature (°C), (b) relative humidity (%), (c) wind speed (m s^{-1}), (d) momentum flux (N m^{-2}), (e) sensible heat flux (W m^{-2}), and (f) sea surface temperature (°C). Dots are shaded blue over water and white over the MIZ. The correlation coefficient (r), linear regression slope, bias and root-mean-square error (rmse) are noted.

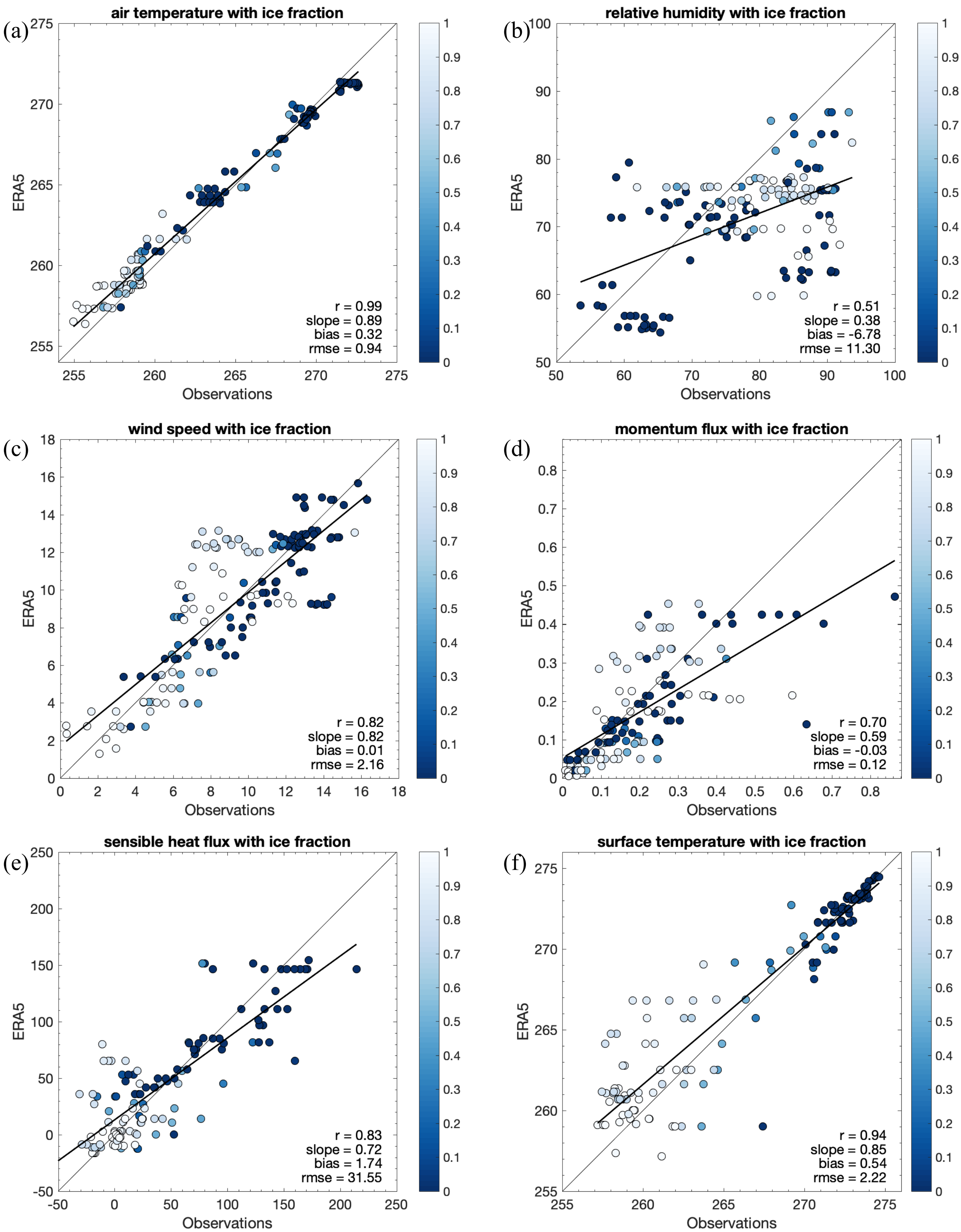


Figure 6 | Scatter plots of aircraft-based observations versus ERA5 data for flight-level (a) air temperature (K), (b) relative humidity (%), (c) wind speed (m s^{-1}), (d) momentum flux (N m^{-2}), (e) sensible heat flux (W m^{-2}), and (f) surface temperature (K). The observed ice fraction is shaded. The correlation coefficient (r), linear regression slope, bias and root-mean-square error (rmse) are noted. The comparison is for all the IGP marine flights.

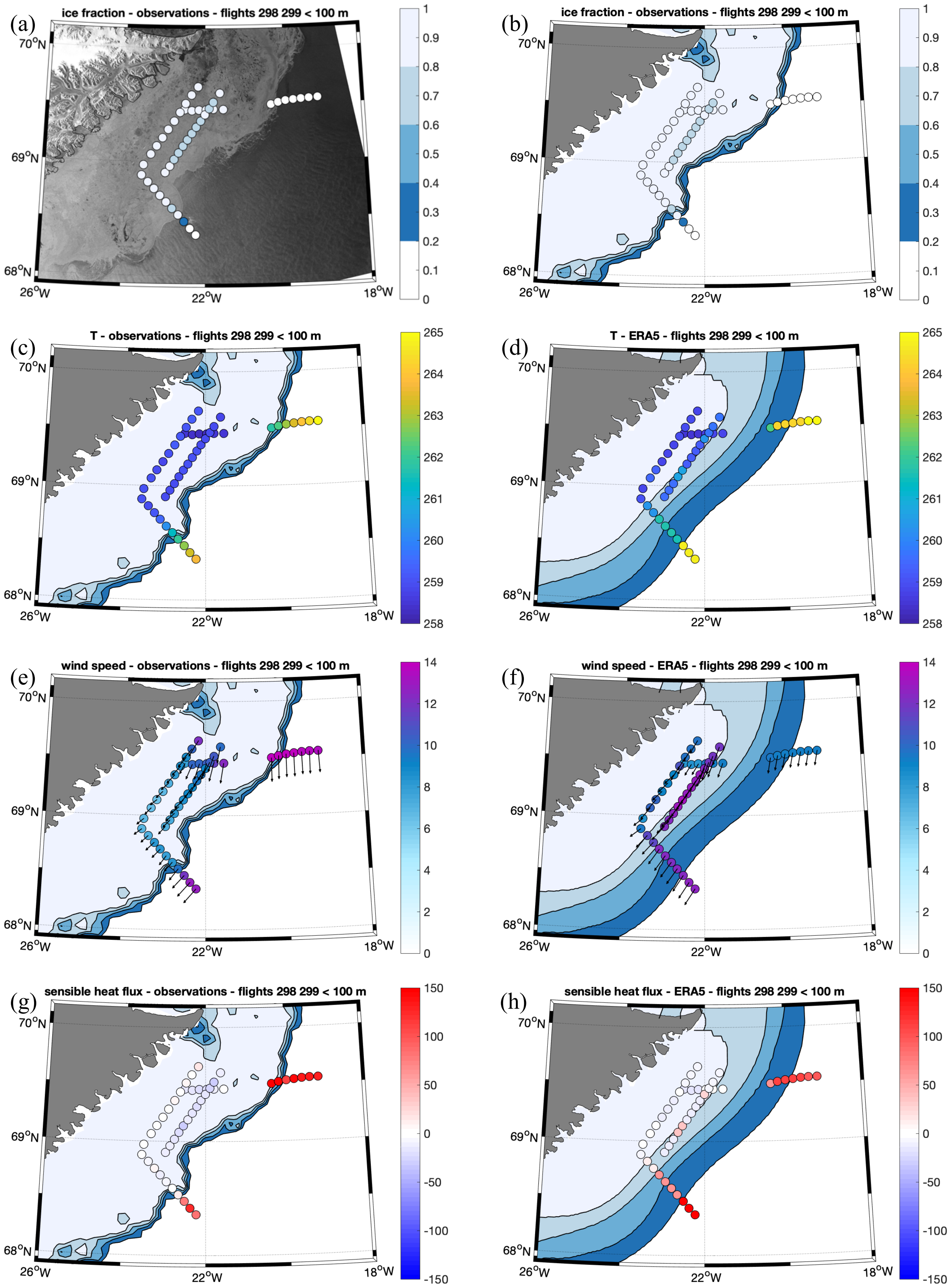
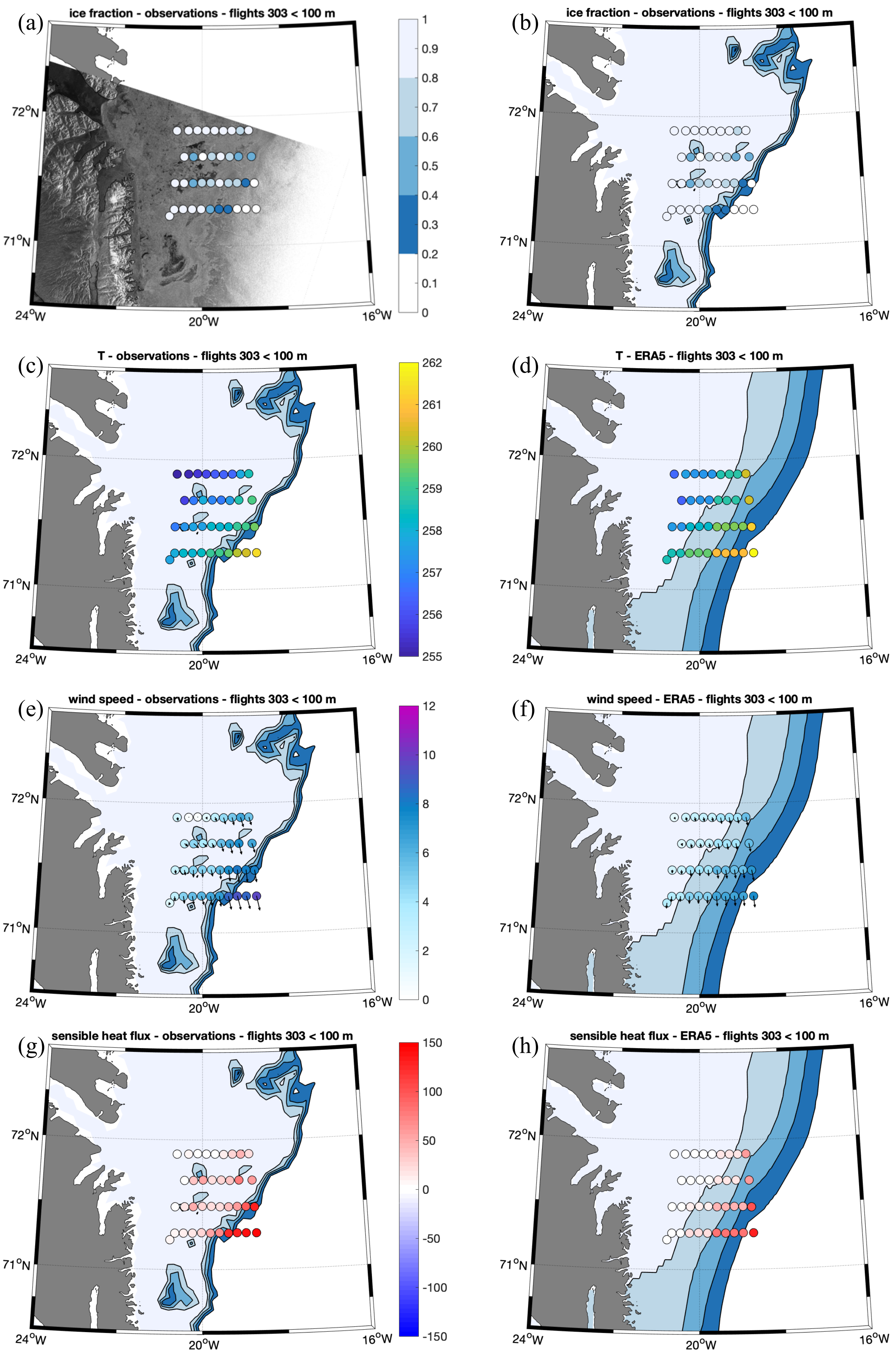


Figure 7 | Spatial maps of sea-ice distribution from 8 March 2018 with aircraft observations or ERA5 output overlaid. Panel (a) is a SAR image, with brighter shading indicating higher reflectively over the ocean implying sea-ice; all other panels show satellite-derived sea-ice fraction contours from AMSR2 (panels b,c,e,g) or from OSTIA (panels d,f,h) using the colour-bar shading of panels (a) & (b). Overlaid are flight-level aircraft observations from runs < 100 m altitude – mean altitude of 35 m – (panels a,b,c,e,g) or ERA5 data extracted to the same locations (panels d,f,h) of ice fraction, air temperature, T, (K), wind speed (m s^{-1}) and sensible heat flux (W m^{-2}) as indicated.



814 **Figure 8 |** Spatial maps of sea-ice distribution from 16 March 2018 with aircraft observations or ERA5 output
815 overlaid – see Fig. 7 for details. Satellite-derived ice fraction contours are from AMSR2 (panels b,c,e,g) or from
816 OSTIA (panels d,f,h). The mean altitude of the runs shown is 24 m.

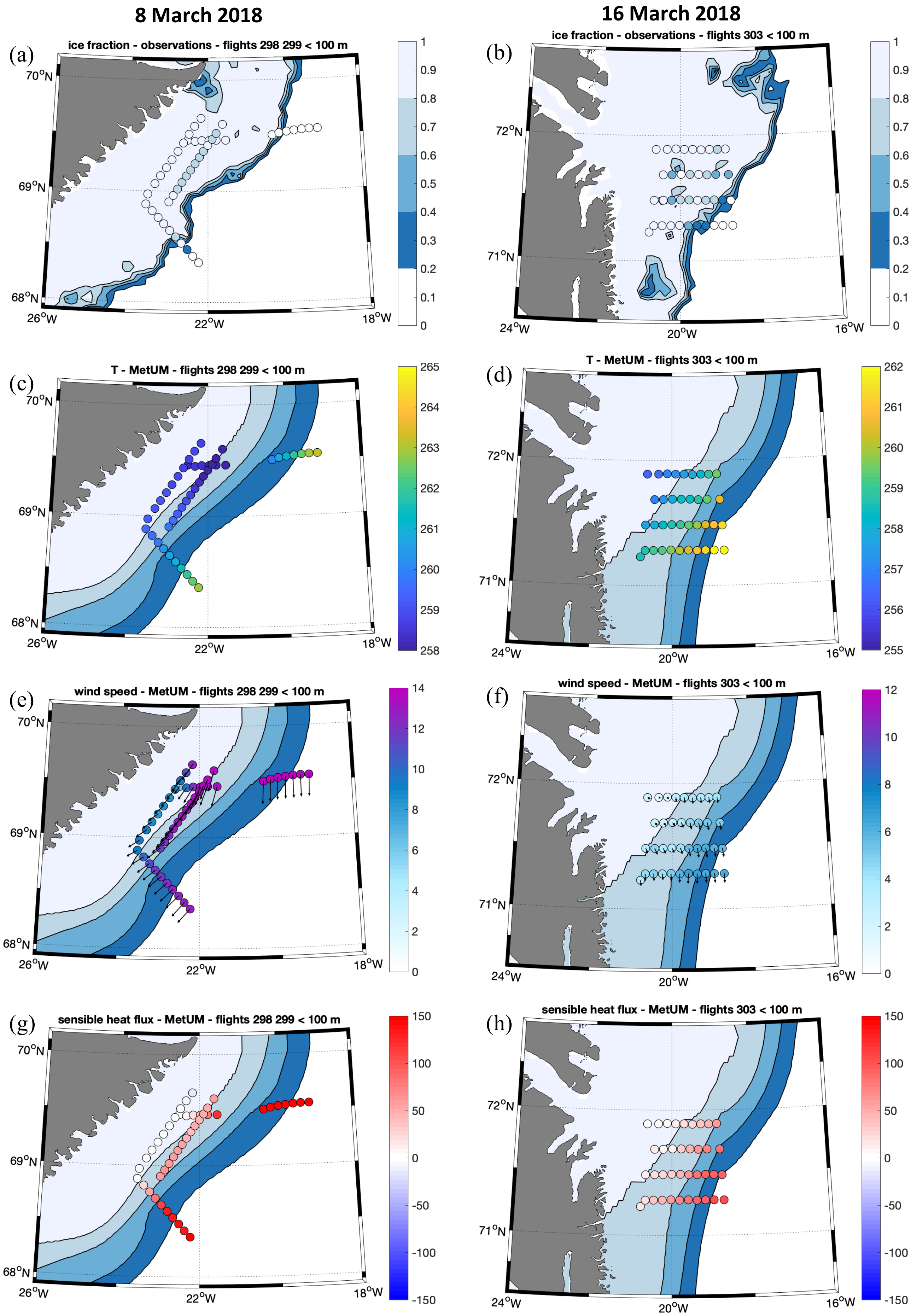


Figure 9 | Spatial maps of sea-ice distribution from 8 March (left) and 16 March 2018 (right) with observations or MetUM output overlaid. Panels shows satellite-derived sea-ice fraction contours: (a) from AMSR2, with ice fraction observations overlaid; and (b-d) from OSTIA, with MetUM output overlaid for flight-level air temperature, T, (K), wind speed (m s^{-1}) and sensible heat flux (W m^{-2}) as indicated. Recall Figs. 7 & 8 shows aircraft observations of the same quantities.