

## Evaluation of Six Atmospheric Reanalyses over Arctic Sea Ice from Winter to Early Summer

ROBERT M. GRAHAM,<sup>a</sup> LANA COHEN,<sup>a</sup> NICOLE RITZHAUPT,<sup>b</sup> BENJAMIN SEGGER,<sup>c</sup> RUNE G. GRAVERSEN,<sup>d</sup> ANNETTE RINKE,<sup>c</sup> VON P. WALDEN,<sup>e</sup> MATS A. GRANSKOG,<sup>a</sup> AND STEPHEN R. HUDSON<sup>a</sup>

<sup>a</sup> Norwegian Polar Institute, Fram Centre, Tromsø, Norway

<sup>b</sup> Department of Meteorology, University of Bonn, Bonn, Germany

<sup>c</sup> Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Potsdam, Germany

<sup>d</sup> Department of Physics and Technology, University of Tromsø, Tromsø, Norway

<sup>e</sup> Department of Civil and Environmental Engineering, Washington State University, Pullman, Washington

(Manuscript received 27 September 2018, in final form 7 March 2019)

### ABSTRACT

This study evaluates the performance of six atmospheric reanalyses (ERA-Interim, ERA5, JRA-55, CFSv2, MERRA-2, and ASRv2) over Arctic sea ice from winter to early summer. The reanalyses are evaluated using observations from the Norwegian Young Sea Ice campaign (N-ICE2015), a 5-month ice drift in pack ice north of Svalbard. N-ICE2015 observations include surface meteorology, vertical profiles from radiosondes, as well as radiative and turbulent heat fluxes. The reanalyses simulate surface analysis variables well throughout the campaign, but have difficulties with most forecast variables. Wintertime (January–March) correlation coefficients between the reanalyses and observations are above 0.90 for the surface pressure, 2-m temperature, total column water vapor, and downward longwave flux. However, all reanalyses have a positive wintertime 2-m temperature bias, ranging from 1° to 4°C, and negative (i.e., upward) net longwave bias of 3–19 W m<sup>-2</sup>. These biases are associated with poorly represented surface inversions and are largest during cold-stable periods. Notably, the recent ERA5 and ASRv2 datasets have some of the largest temperature and net longwave biases, respectively. During spring (April–May), reanalyses fail to simulate observed persistent cloud layers. Therefore they overestimate the net shortwave flux (5–79 W m<sup>-2</sup>) and underestimate the net longwave flux (8–38 W m<sup>-2</sup>). Promisingly, ERA5 provides the best estimates of downward radiative fluxes in spring and summer, suggesting improved forecasting of Arctic cloud cover. All reanalyses exhibit large negative (upward) residual heat flux biases during winter, and positive (downward) biases during summer. Turbulent heat fluxes over sea ice are simulated poorly in all seasons.

### 1. Introduction

Temperatures in the Arctic are rising twice as fast as the Northern Hemisphere as a whole, and Arctic sea ice is retreating in all seasons (Serreze and Francis 2006; Bekryaev et al. 2010; Stroeve et al. 2012; Boisvert and Stroeve 2015; Stroeve and Notz 2018). Many studies documenting and attributing these ongoing changes in the Arctic rely heavily on atmospheric reanalyses (Screen and Simmonds 2012; Screen et al. 2013; Mortin et al. 2016; Overland and Wang 2016; Graham et al. 2017a,b; Rinke et al. 2017; Kapsch et al. 2019). Reanalyses are also widely used as boundary conditions for Arctic regional models and ice–ocean models (Dorn

et al. 2009; Schweiger et al. 2011; Rinke et al. 2013; Lindsay and Schweiger 2015).

While these are frequently used for studies in the Arctic, known biases exist that have afflicted several generations of atmospheric reanalyses (Cullather et al. 2016). For example, most reanalyses have a warm and moist bias at the surface in the Arctic (Beesley et al. 2000; Makshtas et al. 2007; Tjernstöm and Graversen 2009; Jakobson et al. 2012; de Boer et al. 2014; Lindsay et al. 2014; Wesslén et al. 2014). This bias is strongest during cold stable periods in winter months, and is associated with simulating surface temperature inversions that are too weak (Tjernstöm and Graversen 2009; Serreze et al. 2012; Graham et al. 2017a). Furthermore, reanalyses simulate clouds poorly in the Arctic (Walsh and Chapman 1998; Makshtas et al. 2007; Walsh et al. 2009; de Boer et al. 2014; Engström et al. 2014; Lindsay et al. 2014; Wesslén et al. 2014). In particular,

Corresponding author: Stephen R. Hudson, stephen.hudson@npolar.no

reanalyses frequently fail to simulate observed persistent clouds during summer months. This results in a poor representation of surface radiative heat fluxes (Walsh et al. 2009; Wesslén et al. 2014). There is also a large spread among reanalyses for both the total precipitation and phase of precipitation in the Arctic, but a lack of observations makes it difficult to assess which products are most accurate (Boisvert et al. 2018). The presence of these biases does not necessarily preclude the use of reanalyses for analyzing interannual variability and trends in temperature (Simmons and Poli 2014). However, the combination of biases, errors, and the spread among products can generate large uncertainties when using reanalyses as boundary conditions to model Arctic sea ice (Lindsay et al. 2014).

A key source of uncertainty in reanalyses and reason for the spread among products for certain variables is the difference in methods used to parameterize subgrid-scale processes, such as cloud physics and turbulent mixing (Tastula et al. 2013; Engström et al. 2014; Pithan et al. 2014; Klaus et al. 2016; Boisvert et al. 2018; Taylor et al. 2018). Given the global coverage of most atmospheric reanalyses, many parameterization schemes are optimized for lower latitudes (Dee et al. 2011; Saha et al. 2014; Kobayashi et al. 2015; Bosilovich et al. 2015). Regional reanalyses, such as the Arctic systems reanalyses, have been developed with parameterization schemes designed specifically for polar regions (Bromwich et al. 2014, 2016, 2018). This can help to represent certain processes more accurately. However, our understanding of many small-scale processes in the Arctic remains limited (Morrison et al. 2011; Solomon et al. 2014; Boisvert et al. 2018).

One reason for our reliance on reanalyses, despite known biases, is the lack of reliable observations from the central Arctic compared with the midlatitudes (Cullather et al. 2016; Boisvert et al. 2018). In addition to the limited availability of in situ observations, there are large uncertainties with many satellite measurements over the often cloudy and ice-covered Arctic Ocean (Cullather et al. 2016). The lack of observations generates two further issues for reanalyses in the Arctic. First, fewer observations are assimilated into the reanalyses compared with lower latitudes. As a result, the observations have less influence in the final analysis, especially near the surface, which creates a greater reliance on forecast models' "first guess" results (Serreze et al. 2012; Cullather et al. 2016). In addition, there are fewer observations that can be used to evaluate the performance of reanalyses in the Arctic, especially independent observations that were not assimilated into the reanalyses (Jakobson et al. 2012; Wesslén et al. 2014).

The winter and spring months provide the fewest ground-based observations from the central Arctic. This corresponds to the periods of maximum sea ice extent (March) and polar night, when temperatures can plummet to below  $-40^{\circ}\text{C}$ . To date, the primary Arctic datasets used for evaluating reanalyses during the winter season are from North Pole drifting stations (1954–2006), the 1997–98 Surface Heat Budget of the Arctic (SHEBA) experiment, and circumpolar radio-sonde sounding stations on the periphery of the Arctic Ocean (Walsh and Chapman 1998; Makshtas et al. 2007; Liu et al. 2008; Tjernstöm and Graversen 2009; Walsh et al. 2009; Serreze et al. 2012; Naakka et al. 2018). Field campaigns spanning several months, such as the SHEBA experiment, are rare in the central Arctic Ocean.

The Norwegian Young Sea Ice campaign (N-ICE2015) was a 5-month field campaign in which a research ship (R/V *Lance*) drifted with the sea ice from January to June 2015, in the pack ice north of Svalbard (Granskog et al. 2018). N-ICE2015 was the first winter field campaign targeted specifically to study younger and thinner sea ice, which is now ubiquitous in the Arctic (Granskog et al. 2016). The location also coincides with a region of rapid Arctic warming, increased storminess, and significant winter sea ice retreat (Park et al. 2015; King et al. 2017; Graham et al. 2017a; Rinke et al. 2017). In this study, we use N-ICE2015 observations to evaluate the performance of six atmospheric reanalyses over Arctic pack ice during winter, spring, and early summer.

## 2. Data

### a. N-ICE2015 observations

The N-ICE2015 field campaign consisted of four distinct ice drifts, two during winter season and two during the spring and early-summer period (Fig. 1). The two winter drifts covered the dates 15 January–21 February and 24 February–19 March 2015. Both winter drifts began at approximately  $83^{\circ}\text{N}$ . Observations on the first drift were terminated after the floe broke up as it drifted southward into the marginal ice zone. The pause in the campaign between Drift 2 and Drift 3 allowed the ship to refuel and resupply in Longyearbyen. The spring and summer drifts covered the dates 18 April–5 June and 7–21 June. Drift 3 began at  $83^{\circ}\text{N}$  and drifted southward until reaching the ice edge. Subsequently, Drift 4 began at approximately  $81^{\circ}\text{N}$  and followed a path almost parallel to the ice edge (Fig. 1).

At the start of each drift, a meteorological station was built on the ice, approximately 300–400 m from the ship,

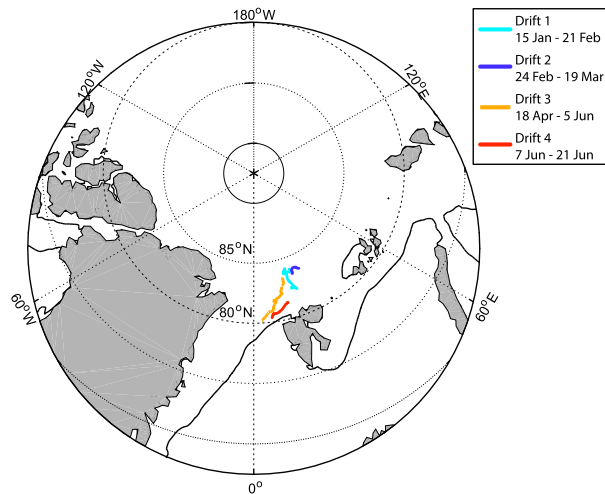


FIG. 1. Overview map of the N-ICE2015 field campaign, with location and dates of four ice drifts. Mean sea ice extent from the National Snow and Ice Data Center for June 2015 is given by a black contour (Fetterer et al. 2017).

to measure the surface meteorology. In this study, we use the mean sea level pressure, 2-m air temperature, and 10-m wind speed (Hudson et al. 2015). The manufacturers stated measurement accuracy of these instruments for the conditions observed during N-ICE2015, are 0.3 hPa, 0.4°C, and 0.4 m s<sup>-1</sup>, respectively. The surface meteorological observations and associated uncertainties are described in detail by Cohen et al. (2017).

Radiosondes were launched from the ship twice per day at 1100 and 2300 UTC, providing profiles of temperature, relative humidity, and wind speed (Hudson et al. 2017). The manufacturer states that the uncertainty in these measurements is 0.5°C, 5%, and 0.15 m s<sup>-1</sup>, respectively. These measurements were used to calculate specific humidity and the total column water vapor, using the formula of Hyland and Wexler (1983). For further information on the N-ICE2015 radiosondes we refer to Kayser et al. (2017). Radiosonde data were transmitted directly to the World Meteorological Organization's Global Telecommunication System (WMO-GTS) and were, thus, assimilated into all of the reanalyses products analyzed in this study. Surface observations from the meteorological tower and ship were not transmitted to WMO-GTS.

Surface radiative fluxes (upward and downward shortwave and longwave) were measured at a height of 1.0–1.2 m near the meteorological tower on each floe (Hudson et al. 2016). The measurement uncertainty for these observations is expected to be less than 3% or approximately 5–10 W m<sup>-2</sup>. We also use measurements of surface turbulent sensible and latent heat fluxes (Walden et al. 2017b). These observations, and the

methods used to calculate the fluxes, are described in detail by Walden et al. (2017a). The random uncertainty in the turbulent heat flux measurements was calculated for a clear and cloudy day in both the winter and spring periods, using the method of (Finkelstein and Sims 2001). During winter, sensible heat flux errors are on the order of 2.5 W m<sup>-2</sup> for clear days and 2.0 W m<sup>-2</sup> for cloudy days, while latent heat flux errors are 1.5 W m<sup>-2</sup> for clear days and 0.1 W m<sup>-2</sup> for cloudy days. In spring, errors are approximately 0.5 W m<sup>-2</sup> for the sensible and latent heat flux, on both clear and cloudy days. It should be noted that while the magnitude of these errors are small, as percentage errors they can be relatively large (up to 80%). For sign convention, we define all radiative and turbulent heat fluxes as positive downward from the atmosphere into the snow/ice surface.

### b. Atmospheric reanalyses

The temporal output files for the six reanalyses vary from 1 to 6 h. For consistency, we evaluate all of the reanalysis surface variables using a 6-h temporal window (0000, 0600, 1200, and 1800 UTC). To compare the N-ICE2015 observations with the reanalyses, we chose the nearest horizontal grid point to the mean position of the ship during that 6-h period, using the original reanalysis grid.

The two-dimensional surface analysis fields from the reanalyses and three-dimensional analysis fields are instantaneous values (30-min averages, ±15 min of the analysis time). These include 2-m temperature, 10-m wind speed, mean sea level pressure, and total column water vapor, as well as the vertical profiles of temperature, winds speed, and humidity. We average the N-ICE2015 observations of 2-m temperature, 10-m wind speed, and mean sea level pressure over a 1-h window (i.e., ±30 min), centered on the valid time of the reanalysis analysis field.

To evaluate the radiosonde profiles, we retrieve the reanalyses' three-dimensional analysis fields at 12-h temporal resolution, interpolated onto pressure levels. We use 16 pressure levels below 500 hPa, for all products. These pressure levels have a spacing of 25 hPa up to 750 hPa, and 50 hPa thereafter.

The reanalysis forecast fields, including the turbulent and radiative heat fluxes, are accumulated or average fields over 1-, 3-, or 6-h forecast windows. For the surface radiative heat fluxes we average the N-ICE2015 observations over the 6-h forecast window (e.g., from 0000 to 0600 UTC). Where reanalyses output are available at a higher resolution than 6 h, we average the output over a 6-h window. For the turbulent heat fluxes we use daily average values, due to the high-frequency variability of these fluxes.

### 1) ERA-I

The European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim, herein ERA-I) is a global atmospheric reanalysis product covering the period 1979 to the present (Dee et al. 2011). The horizontal resolution of ERA-I is approximately 79 km (T255 spectral), and there are 60 vertical levels from the surface up to 0.1 hPa. The data assimilation system used to produce ERA-I is based on the 2006 release (Cy31r2) of the ECMWF Integrated Forecasting System (IFS), which includes a four-dimensional variational analysis (4D-Var). The analysis window is 12 h, and analysis fields are available every 6 h (Dee et al. 2011).

### 2) ERA5

ERA5 is a new and updated global reanalysis from ECMWF, released in 2017, that will replace ERA-I. The horizontal resolution of ERA5 is approximately 31 km, compared with 79 km in ERA-I. Similarly, the vertical resolution is increased from 60 to 137 model levels, up to 0.01 hPa. The assimilation system used for ERA5 is the IFS Cycle Cy41r2 with 4D-Var. Analysis fields are available every hour. Some newly reprocessed datasets and data from recent instruments that were not assimilated into ERA-I are included in ERA5 (<https://confluence.ecmwf.int/pages/viewpage.action?pageId=74764925>).

### 3) JRA-55

The Japanese 55-yr Reanalysis (JRA-55) is a global reanalysis that was released in 2013. JRA-55 is produced using the TL319 version of Japan Meteorological Agency's (JMA) operational data assimilation system, as of December 2009 (Kobayashi et al. 2015; Harada et al. 2016). This system was extensively improved following the earlier Japanese 25-yr Reanalysis (JRA-25), and now includes 4D-Var. JRA-55 also assimilates several newly available and improved past observations, compared with JRA-25, including atmospheric motion vectors and clear-sky radiances from Geostationary Meteorological Satellite (GMS) and Multifunctional Transport Satellite (MTSAT) imagery (Kobayashi et al. 2015). JRA-55 has a horizontal resolution of approximately 55 km and 60 vertical levels up to 0.1 hPa. Analysis fields are available at 6-hourly resolution. JRA-55 has a relatively crude classification of sea ice, and considers all regions with an observed sea ice concentration greater than 55% to have an ice fraction of 1.00.

### 4) CFSv2

The National Centers for Environmental Prediction's (NCEP) Climate Forecast System, version 2

(CFSv2), is an operational analysis that began in 2011 and is available in near real time (Saha et al. 2014). CFSv2 provides a continuation of the 2010 NCEP Climate Forecast System Reanalysis (CFSR) (Saha et al. 2010). The analysis system used in CFSR is the Grid-point Statistical Interpolation (GSI), with 3D-Var. The atmospheric model used is the NCEP Global Forecast System (GFS). The horizontal resolution is approximately 38 km (T382) with 64 vertical levels, up to 0.2 hPa (Saha et al. 2010). Analysis fields are available every 6 h and forecast fields are available every hour.

In contrast to the other reanalyses included in this study, CFSv2 is a weakly coupled reanalysis with an ocean component and interactive sea ice model. The ocean model is the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model, version 4 (MOM4), which uses the Global Ocean Data Assimilation System (Saha et al. 2010). Simultaneous coupled data assimilation for the atmosphere and ocean is not performed.

### 5) MERRA-2

MERRA-2 is produced with version 5.12.4 of the Goddard Earth Observing System (GEOS5.12.4) atmospheric data assimilation system (Bosilovich et al. 2015). The GEOS-5 atmospheric model is used together with the GSI analysis scheme with 3D-Var. The model has a horizontal resolution of  $0.5^\circ$  latitude  $\times$   $0.625^\circ$  longitude, and 72 vertical levels up to 0.01 hPa. Analysis fields are available at 3-h resolution.

### 6) ASRv2

The Arctic System Reanalysis version 2 (ASRv2) is a regional reanalysis for the Arctic produced using a high-resolution version of the Weather Research and Forecasting (WRF) Model that is optimized for polar environments (Polar-WRF) (Bromwich et al. 2018). Polar optimizations are mainly within the Noah land surface model, and include improved heat transfer through snow and ice, the inclusion of fractional ice, and the ability to specify variable snow depth on sea ice, albedo, and ice thickness (Hines and Bromwich 2008; Bromwich et al. 2009; Hines et al. 2015; Bromwich et al. 2018).

ASRv2 follows the earlier coarser-resolution Arctic System Reanalysis (Bromwich et al. 2016). The inner domain of the model covers approximately half of the Northern Hemisphere, with a horizontal resolution of 15 km and 71 vertical layers up to 10 hPa. ASRv2 uses the WRF Data Assimilation system (WRFDA) with 3D-Var. Initial and lateral boundary conditions for the model are provided by ERA-I. ASRv2 fields are available at 3-h resolution.

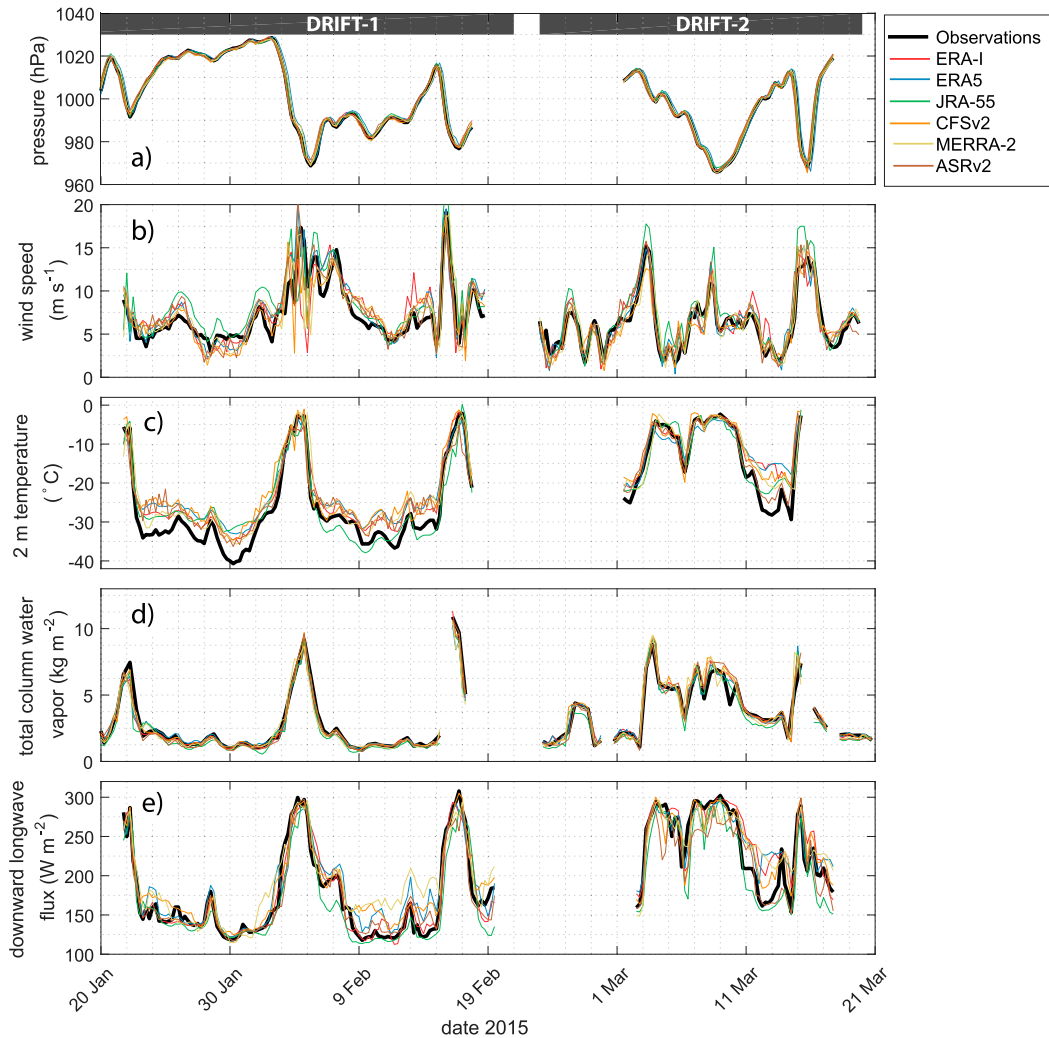


FIG. 2. Time series of N-ICE2015 observations (black) for the two winter drifts (Drifts 1–2), compared with values from the six reanalyses (colors). (a) Mean sea level pressure, (b) 10-m wind speed, (c) 2-m air temperature, (d) total column water vapor, and (e) downward longwave radiative flux at surface.

### 3. Results

#### a. Winter season

Here we compare N-ICE2015 observations with the six reanalyses, for the first two ice drifts. These drifts cover the dates 15 January–21 February and 24 February–19 March 2015 (Fig. 1). This period corresponds mostly to the polar night, with negligible short-wave radiative fluxes.

#### 1) ANALYSIS FIELDS: SURFACE METEOROLOGY AND VERTICAL PROFILES

Overall, the reanalyses perform well for the surface meteorology and water vapor profiles during the winter season (Figs. 2a–d). Correlation coefficients ( $R$ ) between the reanalyses and observations are above 0.84

for the 2-m temperature, 10-m wind speed, and total column water vapor (Table 1). The exceptional performance of the mean sea level pressure and total column water vapor is reasonable to expect, with the assimilation of data from radiosondes (Figs. 2a,d). Nonetheless, JRA-55 has a significant dry bias compared with the other reanalyses (Figs. 3e,f; Table 1).

Correlation coefficients for the 10-m wind speed range from 0.84 in ERA-I to 0.92 in ERA5 (Table 1; Fig. 2b). Most reanalyses have a positive 10-m wind speed bias, although the bias is not always significant (Table 1). CFSv2 has no detectable bias. The largest bias ( $+1.0 \text{ m s}^{-1}$ ) and RMSE ( $2.2 \text{ m s}^{-1}$ ) are found in JRA-55. ERA5 has the smallest RMSE of  $1.4 \text{ m s}^{-1}$ . Most reanalyses have too broad of a distribution of wind speeds during calm periods; that is, they underestimate

TABLE 1. Table shows correlations coefficients ( $R$ ), biases, and root-mean-square errors (RMSEs) for the six reanalyses, based on N-ICE2015 observations. Biases that are not statistically significant at the 95% confidence interval are shown in italics. To test the significance, we assume that the bias is stationary during each season and calculate the confidence interval using a moving block bootstrap approach (Künsch 1989). The optimal block length is estimated using the method described in Patton et al. (2009). The statistics are calculated for winter (January–March), spring (April–May), and summer (June). For every variable and season, we give each reanalysis a score out of 10. This score is defined as:  $\text{Score} = \{R + [1 - \text{abs}(\text{bias})/\text{range}] + (1 - \text{RMSE}/\text{RANGE})\} \times 10/3$ , where RANGE is the maximum range of the N-ICE2015 observations for that variable during the given season. Statistics for the shortwave radiative fluxes and sensible heat flux are based on daily mean values. All other variables are based on 6-hourly values.

Variable	Reanalyses	Winter				Spring				Summer			
		$R$	Bias	RMSE	Score	$R$	Bias	RMSE	Score	$R$	Bias	RMSE	Score
2-m air temperature ( $^{\circ}\text{C}$ )	ERA-I	0.96	+3.0	4.5	9.24	0.97	+1.3	1.8	9.65	0.70	+1.6	1.7	6.72
	ERA5	0.96	+3.4	5.3	9.13	0.96	+1.7	2.2	9.55	0.57	+0.8	1.0	7.32
	JRA-55	0.93	+1.1	4.4	9.30	0.94	+0.2	2.6	9.57	0.68	+0.7	1.0	7.76
	CFSv2	0.97	+3.8	5.0	9.16	0.93	-0.2	2.2	9.57	0.64	+0.1	0.6	8.32
	MERRA-2	0.97	+3.0	4.2	9.30	0.96	+0.9	1.6	9.66	0.81	+0.3	0.5	8.81
	ASRv2	0.97	+1.9	3.5	9.45	0.98	+0.8	1.2	9.77	0.76	+0.1	0.6	8.72
10-m wind speed ( $\text{m s}^{-1}$ )	ERA-I	0.84	+0.4	1.9	9.01	0.85	-0.2	1.5	9.04	0.96	+0.3	1.1	9.53
	ERA5	0.92	+0.4	1.4	9.37	0.91	+0.1	1.1	9.37	0.97	-0.2	0.9	9.63
	JRA-55	0.87	+1.0	2.2	8.93	0.87	0.0	1.4	9.18	0.96	+0.5	1.2	9.45
	CFSv2	0.87	0.0	1.7	9.23	0.88	-0.6	1.4	9.05	0.97	-0.8	1.3	9.39
	MERRA-2	0.87	+0.2	1.7	9.19	0.88	-0.4	1.3	9.14	0.97	+0.1	0.9	9.66
	ASRv2	0.89	+0.3	1.7	9.23	0.86	-0.3	1.4	9.07	0.94	-0.2	1.2	9.46
Water vapor path ( $\text{kg m}^{-2}$ )	ERA-I	0.98	0.0	0.5	9.77	0.97	+0.1	0.5	9.67	0.89	+0.5	1.3	9.17
	ERA5	0.98	0.0	0.5	9.77	0.97	+0.1	0.5	9.67	0.92	+0.3	1.1	9.37
	JRA-55	0.97	-0.3	0.6	9.60	0.95	-0.1	0.7	9.53	0.90	-0.2	1.1	9.33
	CFSv2	0.98	0.1	0.5	9.74	0.96	+0.1	0.6	9.60	0.88	+0.2	1.2	9.24
	MERRA-2	0.96	0.0	0.6	9.67	0.94	+0.1	0.8	9.46	0.80	+0.3	1.6	8.85
	ASRv2	0.99	-0.1	0.4	9.80	0.99	-0.4	0.5	9.62	0.94	-0.2	1.0	9.49
Downward longwave radiative flux ( $\text{W m}^{-2}$ )	ERA-I	0.95	+4	20	9.41	0.67	-24	36	7.45	0.59	-9	17	7.46
	ERA5	0.94	+9	22	9.26	0.62	+2	27	8.03	0.80	-7	13	8.43
	JRA-55	0.95	-13	24	9.19	0.72	-38	47	7.01	0.53	-17	27	6.45
	CFSv2	0.96	+9	20	9.36	0.67	-23	39	7.40	0.64	-3	17	7.90
	MERRA-2	0.92	+13	28	9.02	0.57	-19	36	7.24	0.64	-19	23	6.90
	ASRv2	0.95	-6	20	9.38	0.61	-46	54	6.28	0.38	-31	40	4.73
Net longwave radiative flux ( $\text{W m}^{-2}$ )	ERA-I	0.83	-9	19	8.20	0.41	-25	34	5.66	0.53	-9	18	7.16
	ERA5	0.82	-5	15	8.52	0.15	-5	24	6.00	0.80	-8	14	8.30
	JRA-55	0.79	-16	21	7.67	0.36	-36	43	4.69	0.55	-19	28	6.29
	CFSv2	0.84	-8	14	8.50	0.41	-22	32	5.86	0.65	-2	16	7.99
	MERRA-2	0.65	-3	16	8.00	0.34	-22	33	5.59	0.66	-19	22	6.94
	ASRv2	0.79	-19	26	7.31	0.28	-43	49	3.90	0.39	-28	25	5.47
Downward shortwave radiative flux ( $\text{W m}^{-2}$ )	ERA-I	—	—	—	—	0.75	+26	49	8.15	0.62	-13	44	7.62
	ERA5	—	—	—	—	0.70	+5	36	8.44	0.82	-12	30	8.58
	JRA-55	—	—	—	—	0.77	+56	65	7.56	0.67	+30	51	7.32
	CFSv2	—	—	—	—	0.70	+36	53	7.79	0.79	-15	32	8.38
	MERRA-2	—	—	—	—	0.76	-2	43	8.59	0.69	-4	33	8.25
	ASRv2	—	—	—	—	0.71	+79	88	6.76	0.45	+93	102	4.37
Net shortwave radiative flux ( $\text{W m}^{-2}$ )	ERA-I	—	—	—	—	0.65	+24	30	5.85	0.60	+52	60	2.77
	ERA5	—	—	—	—	0.53	+8	16	7.11	0.89	+44	52	4.85
	JRA-55	—	—	—	—	0.45	+25	37	4.74	0.66	+65	67	1.92
	CFSv2	—	—	—	—	0.58	+24	30	5.62	0.87	+55	58	3.62
	MERRA-2	—	—	—	—	0.52	+13	29	6.08	0.72	+47	55	3.70
	ASRv2	—	—	—	—	0.68	+38	40	4.63	0.74	+41	43	4.71
Sensible heat flux ( $\text{W m}^{-2}$ )	ERA-I	0.18	-17	39	4.01	0.36	+4	8	5.91	0.34	+9	12	5.65
	ERA5	0.32	-14	32	5.06	0.48	+1	7	6.96	0.61	+7	10	6.96
	JRA-55	0.74	+4	11	8.26	0.45	+12	14	3.93	0.47	+8	11	6.29
	CFSv2	0.11	-46	70	0.30	0.18	-1	13	4.99	0.57	+19	21	4.47
	MERRA-2	0.23	-31	38	3.43	0.24	+1	7	6.16	0.18	+2	9	6.14
	ASRv2	0.44	-25	41	4.30	-0.02	+6	11	3.83	0.18	+5	10	5.73

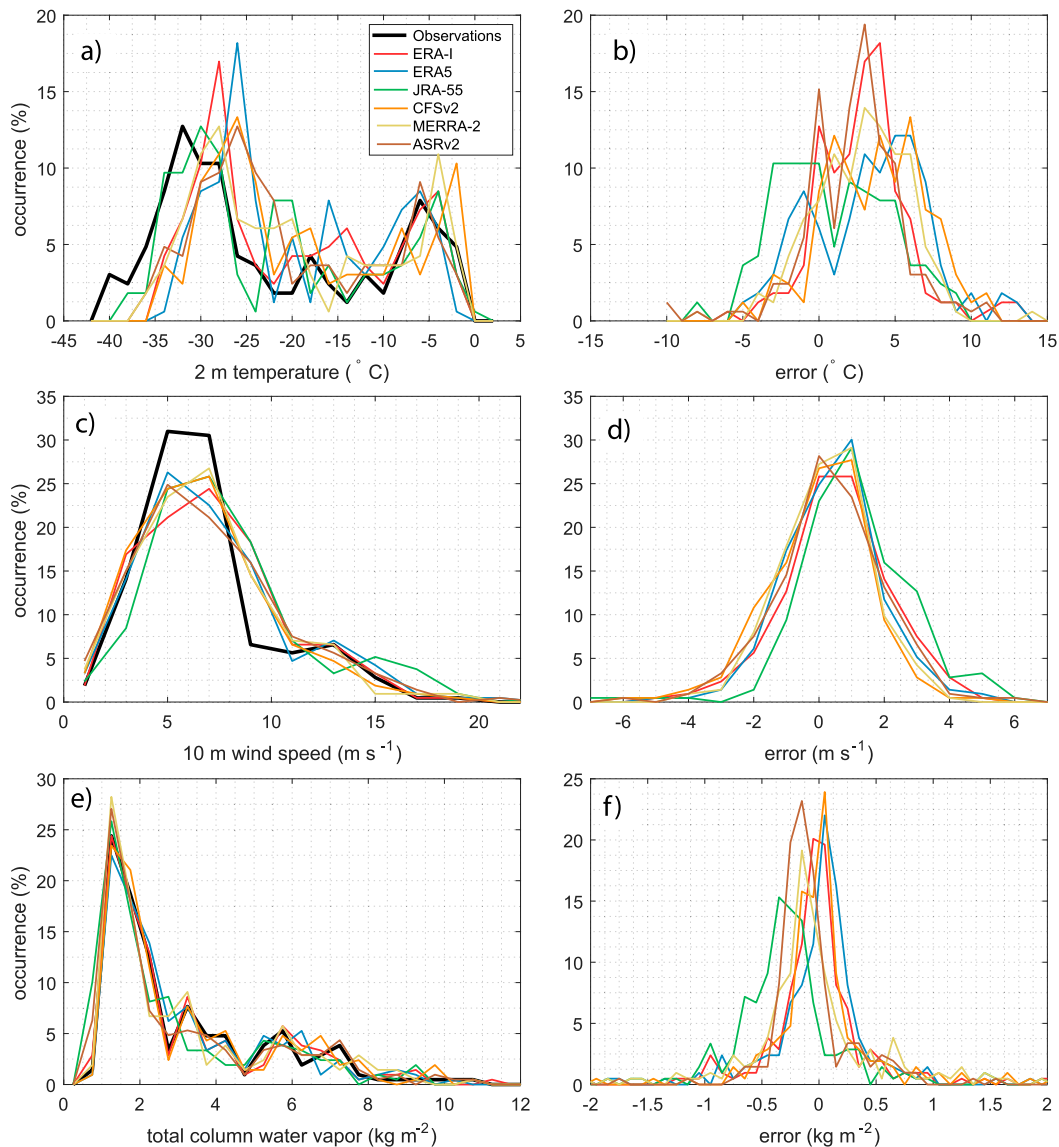


FIG. 3. Frequency distributions of N-ICE2015 observations (black) during winter compared with reanalyses (colors), and associated errors. (a) 2-m air temperature ( $2^{\circ}\text{C}$  bins), (b) 2-m temperature errors ( $1^{\circ}\text{C}$  bins), (c) 10-m wind speed ( $2\text{ m s}^{-1}$  bins), (d) 10-m wind speed errors ( $1\text{ m s}^{-1}$  bins), (e) total column water vapor ( $0.5\text{ kg m}^{-2}$  bins), and (f) total column water vapor errors ( $0.1\text{ kg m}^{-2}$  bins).

the occurrence of light winds ( $3\text{--}8\text{ m s}^{-1}$ ) and overestimate the occurrence of moderate wind ( $8\text{--}10\text{ m s}^{-1}$ ) (Fig. 3c). Most reanalyses display a small negative bias during storm periods, while JRA-55 has a distinct positive bias for strong ( $>15\text{ m s}^{-1}$ ) wind speeds (Figs. 3c,d).

All reanalyses exhibit high correlation coefficients (0.93–0.97) with the observed 2-m temperature in winter (Fig. 2c; Table 1). Nonetheless, there are periods when some reanalyses are more than  $10^{\circ}\text{C}$  warmer than the observations (Figs. 2c and 3a,b). RMSEs for the 2-m temperature are large, ranging from  $3.5^{\circ}\text{C}$  in ASRv2 to  $5.3^{\circ}\text{C}$  in ERA5. All reanalyses have a warm bias during

winter, ranging from  $+1.1^{\circ}\text{C}$  in JRA-55 to  $+3.8^{\circ}\text{C}$  in CFSv2 (Figs. 3a,b; Table 1). The near-surface warm bias in reanalyses is confined foremost to cold periods, when the observed temperature is below  $-25^{\circ}\text{C}$  (Figs. 2c and 3a). At warmer temperatures ( $>-10^{\circ}\text{C}$ ) the bias is much smaller. A winter warm bias over sea ice has persisted through several generations of different reanalyses (Beesley et al. 2000; Makshtas et al. 2007; Liu et al. 2008; Tjernstöm and Graversen 2009; Lindsay et al. 2014; Graham et al. 2017a). Reanalyses continue to have difficulties resolving strong vertical temperature gradients in highly stable surface boundary layers (Serreze et al.

TABLE 2. Seasonal mean values for individual components of the surface energy budget during the winter (January–March), spring (April–May), and summer (June) seasons of the N-ICE2015 campaign. Mean values and biases are also shown for the six reanalyses. Biases are based on daily means. Sensible and latent heat fluxes that are not statistically significant are shown in italics. For reference, we give the seasonal mean sea ice concentration in each reanalysis. Observed ice concentration is not provided, as measurements relating to the energy budget are point observations over sea ice (i.e., 1.00 ice fraction).

Variable	Reanalyses	Winter		Spring		Summer	
		Mean	Bias	Mean	Bias	Mean	Bias
Ice concentration	ERA-I	0.96	—	0.96	—	0.67	—
	ERA5	0.98	—	0.98	—	0.81	—
	JRA-55	1.00	—	1.00	—	1.00	—
	CFSv2	0.93	—	0.95	—	0.75	—
	MERRA-2	0.96	—	0.97	—	0.69	—
	ASRv2	0.94	—	0.94	—	0.92	—
Net radiative flux ( $\text{W m}^{-2}$ ) (shortwave + longwave)	Observations	−29	—	+12	—	+39	—
	ERA-I	−37	−9	+11	−1	+81	+43
	ERA5	−33	−5	+15	+3	+75	+36
	JRA-55	−44	−16	+1	−11	+85	+47
	CFSv2	−36	−7	+14	+2	+92	+53
	MERRA-2	−31	−2	+2	−10	+67	+29
Sensible heat flux ( $\text{W m}^{-2}$ )	ASRv2	−47	−19	+8	−4	+52	+14
	Observations	+14	—	−7	—	−3	—
	ERA-I	−3	−17	−4	+4	+7	+9
	ERA5	0	−14	−6	+1	+5	+7
	JRA-55	+17	+4	+5	+12	+5	+8
	CFSv2	−32	−46	−9	−1	+16	+19
Latent heat flux ( $\text{W m}^{-2}$ )	MERRA-2	−17	−31	−6	+1	0	+2
	ASRv2	−11	−25	−1	+6	+2	+5
	Observations	0	—	0	—	0	—
	ERA-I	−3	−3	−7	−6	−9	−9
	ERA5	−3	−3	−11	−11	−12	−13
	JRA-55	−1	−1	−6	−5	−13	−13
Residual heat flux ( $\text{W m}^{-2}$ )	CFSv2	−22	−22	−15	−14	−18	−18
	MERRA-2	−5	−5	−7	−6	−11	−11
	ASRv2	−3	−3	−3	−2	−9	−9
	Observations	−13	—	+3	—	+36	—
	ERA-I	−31	−18	−3	−6	+80	+43
	ERA5	−29	−15	−5	−9	+68	+32
	JRA-55	−26	−13	−7	−10	+77	+41
	CFSv2	−88	−76	−12	−16	+88	+51
	MERRA-2	−42	−29	−14	−17	+58	+22
	ASRv2	−60	−47	+3	0	+45	+9

2012). Interestingly, we find that despite having twice as many model levels (20 vs 10) below 900 hPa, the near-surface winter warm bias and RMSE in the newly released ERA5 are larger than ERA-I (Table 1). In contrast, ASRv2, which is optimized for polar regions, has the smallest RMSE out of all the reanalyses and the mean warm bias is more than 1°C smaller than most of the global products (Table 1). JRA-55 clearly simulates the best near-surface temperature distribution for winter (Fig. 3a). However, it has the lowest correlation coefficient among all reanalyses and a relatively large RMSE (Table 1). It is noteworthy that the reanalysis with the smallest 2-m temperature bias in winter (JRA-55) has the highest mean ice-covered fraction (1.00), and the reanalysis with the largest temperature bias (CFSv2) has the lowest (0.93) ice-covered fraction (Tables 1 and 2).

Findings from previous studies and our analyses above suggest that the lowest skill for reanalyses in winter is during cold-stable periods (Figs. 2 and 3). To study these periods in more detail, we average 60 N-ICE2015 radiosonde profiles that were launched when the surface air temperature was below  $-25^{\circ}\text{C}$  and compare these to the reanalyses (Fig. 4).

The six reanalyses generally capture the shape of the cold winter profiles well (Fig. 4). However, all reanalyses underestimate the near-surface stability, which we define as the temperature difference between the 850- and 1000-hPa levels. The observed value for this parameter is approximately  $7^{\circ}\text{C}$ , while in the reanalyses values range from  $3^{\circ}\text{C}$  in ERA5 and MERRA-2 to  $6.5^{\circ}\text{C}$  in JRA-55 (Fig. 4a). The weak static stability in reanalyses is associated foremost with a large near-surface warm



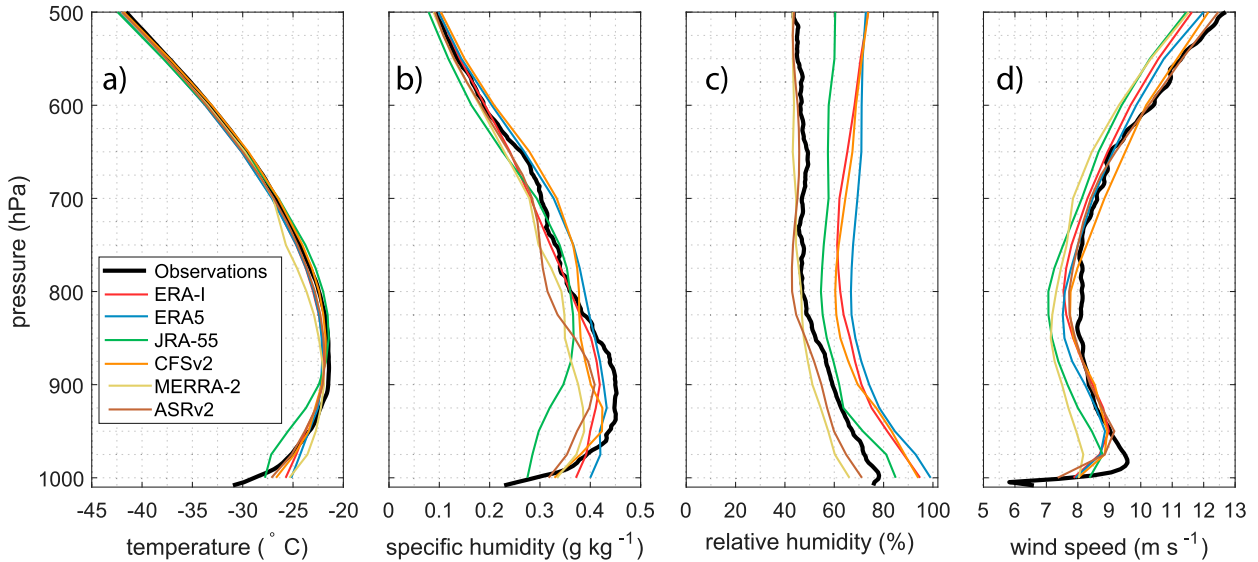


FIG. 4. (a) Mean temperature, (b) specific humidity, (c) relative humidity, and (d) wind speed profiles from N-ICE2015 radiosondes (black) launched when surface temperature was below  $-25^{\circ}\text{C}$  (60 profiles), compared with profiles from reanalyses (colored).

bias. In addition, all reanalyses have a small ( $<1^{\circ}\text{C}$ ) cold bias aloft, between 950 and 850 hPa (Fig. 4a). JRA-55 has a significantly larger cold bias from 900 to 975 hPa compared with the other reanalyses (Fig. 4a). Hence, despite having the best 2-m air temperature distribution for winter and smallest warm bias, JRA-55 does not simulate near-surface temperature profiles more accurately than other products. ASRv2 simulates the most representative temperature profiles during cold and stable winter periods (Fig. 4a).

The strength of the surface specific humidity inversion is substantially underestimated by all reanalyses (Fig. 4b). Each reanalysis exhibits a moist bias near the surface and dry bias from 950 to 850 hPa, where the maximum specific humidity is observed. These results are consistent with findings that ERA-I and JRA-55 underestimate the strength of specific humidity inversions observed at coastal meteorological stations in the Arctic (Naakka et al. 2018). JRA-55 has the largest dry specific humidity bias among all of the reanalyses, which explains the significant negative bias for total column water vapor during winter (Figs. 3f and 4b; Table 1). There is a large spread among the reanalyses for relative humidity (Fig. 4c). ERA5, ERA-I, and CFSv2 have large moist biases, of up to 20%, throughout the troposphere. In contrast, ASRv2 and MERRA-2 have small dry biases, with ASRv2 capturing the mean observed profile most accurately (Fig. 4c).

Most reanalyses slightly underestimate ( $<1\text{ m s}^{-1}$ ) wind speeds aloft during cold stable periods (Fig. 4d). In particular, all reanalyses underestimate the wind speed at 975 hPa, where there is a near-surface wind maximum.

Overall, ASRv2 has the most accurate wind profile for these conditions (Fig. 4d).

## 2) FORECAST FIELDS: SURFACE HEAT FLUXES AND ENERGY BUDGET

Overall, the forecast variables in the reanalyses are simulated less well than the analysis variables for the N-ICE2015 period (Table 1). Nonetheless, the wintertime downwelling longwave flux is captured remarkably well by all reanalyses (Figs. 2e and 5a,b; Table 1). The assimilation of temperature and humidity profiles from radiosondes likely improves the accuracy of these downward longwave fluxes. Correlation coefficients between the observations and reanalyses range from 0.92 in MERRA-2 to 0.96 in CFSv2, and RMSEs ranged from 20 to  $28\text{ W m}^{-2}$ . Four products have a positive bias (i.e., higher downward directed longwave flux), ranging from  $+4\text{ W m}^{-2}$  in ERA-I to  $+13\text{ W m}^{-2}$  in MERRA-2. In contrast, ASRv2 and JRA-55, have negative biases of  $-6$  and  $-13\text{ W m}^{-2}$ , respectively. We note that JRA-55 and ASRv2 have the smallest near-surface warm biases, and largest dry specific humidity biases of the six reanalyses (Figs. 3 and 4; Table 1). Interestingly, ERA5 has a larger positive downward longwave bias and larger RMSE than ERA-I (Table 1).

Correlation coefficients for the net longwave flux are lower than the downward longwave flux, in all reanalyses (Table 1). Correlation coefficients range from 0.65 in MERRA-2 to 0.84 in CFSv2. All reanalyses exhibit negative biases (i.e., upward) for the net longwave flux, which range from  $-3\text{ W m}^{-2}$  in MERRA-2 to  $-19\text{ W m}^{-2}$  in ASRv2 (Figs. 5c,d). These negative biases

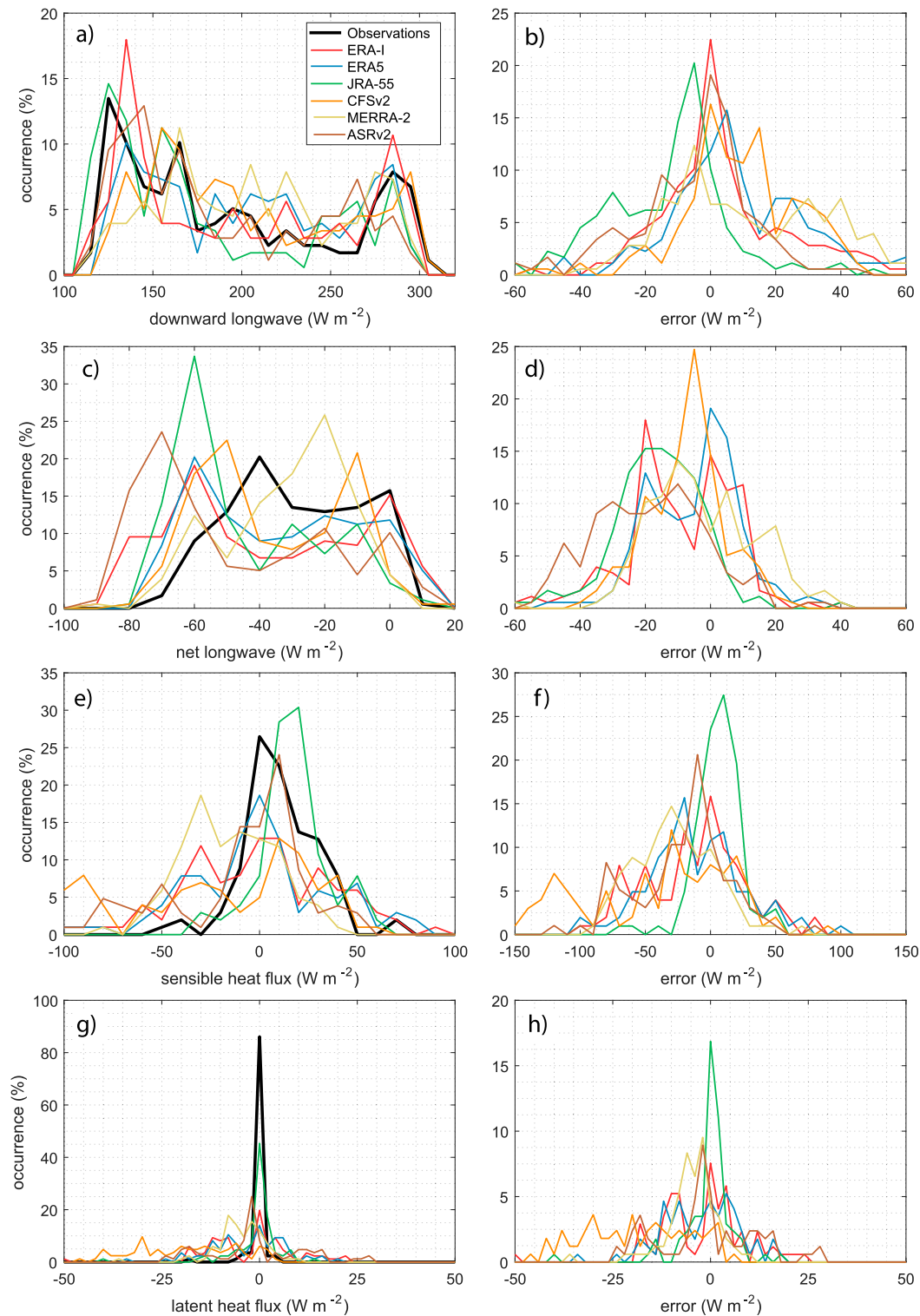


FIG. 5. Frequency distributions of N-ICE2015 observations (black) during winter compared with reanalyses (colors), and associated errors. (a) Downward longwave radiative flux at surface ( $10 \text{ W m}^{-2}$  bins), (b) downward longwave errors (reanalyses–observations) ( $5 \text{ W m}^{-2}$  bins), (c) net longwave radiative flux at surface ( $10 \text{ W m}^{-2}$  bins), (d) net longwave errors (reanalyses–observations) ( $5 \text{ W m}^{-2}$  bins), (e) sensible heat flux at surface ( $10 \text{ W m}^{-2}$  bins), (f) sensible heat flux errors ( $10 \text{ W m}^{-2}$  bins), (g) latent heat flux at surface ( $2 \text{ W m}^{-2}$  bins), and (h) latent heat flux errors ( $2 \text{ W m}^{-2}$  bins).

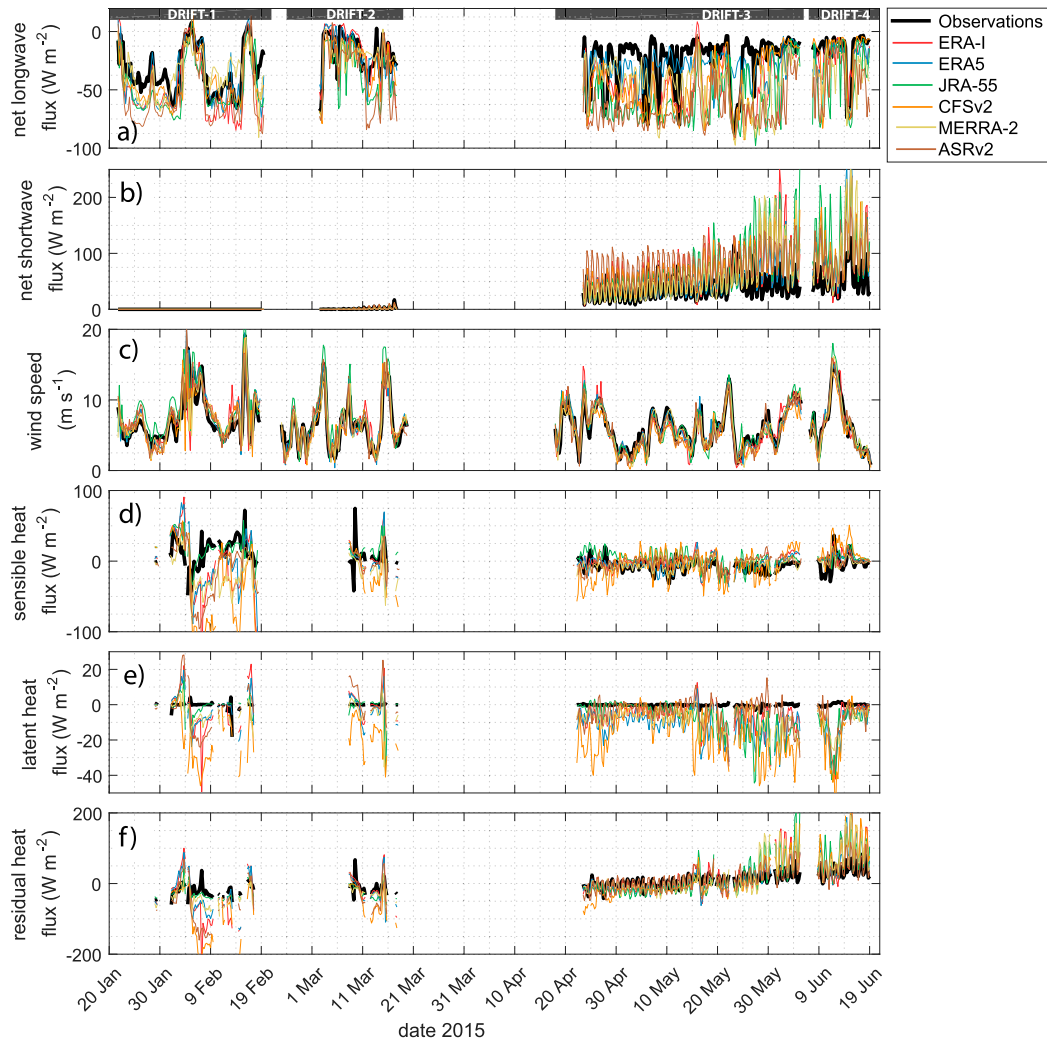


FIG. 6. Time series for observed (black) components of the surface energy balance during N-ICE2015, and those from the six reanalyses (colors). (a) Net longwave flux, (b) net shortwave flux, (c) 10-m wind speed, (d) sensible heat flux, (e) latent heat flux, and (f) residual heat flux at surface [sum of (a) + (b) + (d) + (e)]. All fluxes are defined as positive downward.

are largest during cold and stable periods (Fig. 6) and are foremost the result of an overly large upward longwave flux at the surface, resulting from the positive temperature bias.

The largest negative net longwave biases are found in ASRv2 and JRA-55 (Table 1). This is consistent with the negative downward longwave bias in these reanalyses, which compound the bias for the upward longwave flux. In contrast, the four remaining products exhibit positive biases for the downward longwave flux, which partially compensate the upward longwave flux bias. Nonetheless, the resultant net longwave flux bias remains negative. We note that while the net longwave bias in ERA5 is smaller in magnitude than ERA-I, this reflects larger compensating biases in downward longwave and upward

longwave radiation in ERA5 than ERA-I (Table 1). This highlights the importance of evaluating all terms of the energy budget independently (de Boer et al. 2014), rather than considering only net biases.

We next compare the sensible and latent heat fluxes in reanalyses with observed measurements of these turbulent heat fluxes over sea ice (Figs. 5 and 6). Observed latent heat fluxes over sea ice are near zero during the N-ICE2015 winter (Walden et al. 2017a) (Figs. 5g and 6e). These are consistent with observations from satellite data (Taylor et al. 2018). However, most reanalyses simulate large upward latent heat fluxes, with biases up to  $-22 \text{ W m}^{-2}$  (Figs. 5g,h and 6e; Table 2). Importantly, the range of values for latent heat fluxes simulated by the reanalyses is far larger than the observed values.

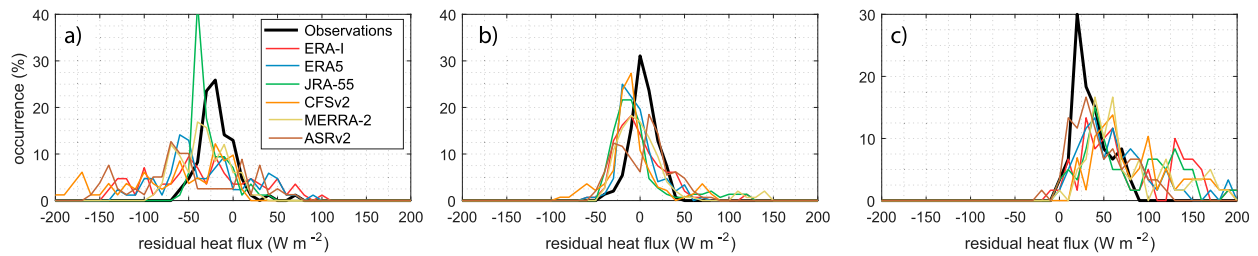


FIG. 7. Frequency distribution for the observed residual heat flux (black) with  $10 \text{ W m}^{-2}$  bins, compared with the six reanalyses (colors), for (a) winter (January–March), (b) spring (April–31 May), and (c) summer (after 1 Jun).

Over the winter drifts, less than 4% of the original 30-min average latent heat flux observations have a magnitude greater than  $5 \text{ W m}^{-2}$ . In contrast, simulated 6-h average latent heat fluxes frequently exceed  $10 \text{ W m}^{-2}$  in all reanalyses (Figs. 5g,h and 6e).

Sensible heat fluxes are typically of the correct order of magnitude in the reanalyses (Fig. 5e). However, the simulated fluxes are often in the opposite direction to the observations (Fig. 6d; Table 2). As a result, correlation coefficients are typically very low (0.11–0.74) (Table 1). The strong stable inversions observed during the N-ICE2015 winter result in a mean downward (positive) sensible heat flux of  $+14 \text{ W m}^{-2}$  over the sea ice (Table 2). However, JRA-55 is the only reanalyses that simulates a positive mean sensible heat flux (Table 2). Overall, JRA-55 performs best among all reanalyses for the sensible heat flux, with the highest correlation coefficient of 0.74, the smallest RMSE, and the smallest bias of  $+4 \text{ W m}^{-2}$  (Figs. 5e,f and 6d; Table 1). All other reanalyses have mean upward fluxes, and large negative biases that range from  $-14 \text{ W m}^{-2}$  in ERA5 to  $-46 \text{ W m}^{-2}$  in CFSv2. The negative sensible heat flux biases in these reanalyses are consistent with the reanalyses underestimating the strength of surface inversions in winter and having positive surface air temperature biases (Fig. 4a).

The poor performance of reanalyses for turbulent heat fluxes over sea ice is consistent with findings from the SHEBA campaign where observations were used to evaluate the ECMWF operational forecast model in 1997–98, with a lead time of 12–35 h (Beesley et al. 2000). Similarly, large errors in turbulent heat fluxes have been identified in several reanalyses over Antarctic sea ice (Tastula et al. 2013).

It is important to note that reanalyses provide grid cell average fluxes, in contrast to the point-based measurements that have a small footprint and were made over sea ice. The approximate area of a grid cell within the reanalyses ranges from  $225 \text{ km}^2$  in ERA5 to  $640 \text{ km}^2$  in ERA-I, and models typically only resolve features with length scales of 5–7 grid boxes (Skamarock 2004). JRA-55 is the only reanalysis with a mean ice fraction of

1.00 during the N-ICE2015 winter. It also has the smallest apparent sensible and latent heat flux biases (Table 2). With its dynamic sea ice model, CFSv2 has the largest mean open water fraction (0.07) during winter, among the different reanalyses. CFSv2 also suffers from the largest apparent sensible and latent heat flux biases (Table 2). To balance these apparent biases, CFSv2 would require a positive (i.e., upward) sensible and latent heat flux over the open water fraction of  $+640$  and  $+315 \text{ W m}^{-2}$ , respectively. ERA-I requires the smallest sensible ( $+410 \text{ W m}^{-2}$ ) and latent ( $+75 \text{ W m}^{-2}$ ) heat fluxes over the open water fraction to balance its apparent biases. There are no wintertime observations of sensible and latent heat fluxes over leads during N-ICE2015, but previous studies have estimated these could be on the order of  $+600$  and  $+150 \text{ W m}^{-2}$ , respectively (Maykut 1978; Marcq and Weiss 2012). Hence, the open water fraction of grid cells in reanalyses will be a major contributing factor to the apparent turbulent heat flux errors, and it is therefore not possible to say with certainty which reanalysis is most accurate. We also note that the open water fraction could contribute to an apparent bias in emitted longwave radiation; an open water fraction of 0.05 at the seawater freezing point of  $-1.8^\circ\text{C}$ , would produce an apparent bias of  $7 \text{ W m}^{-2}$  for a snow-surface temperature of  $-40^\circ\text{C}$  over the ice-covered fraction, or  $3.7 \text{ W m}^{-2}$  for  $-20^\circ\text{C}$ .

We finally consider the overall surface energy budget over sea ice in the observations and the reanalyses. This budget is equal to the sum of the net radiative flux (longwave + shortwave) and the sensible and latent heat fluxes (Walden et al. 2017a). The resultant imbalance can be considered as a residual heat flux, which is balanced by an ocean heat flux through the sea ice and/or a change of energy storage in the snow layer adjacent to the atmosphere. We do not decompose these terms here. During winter, the observed residual heat flux is negative, with a mean value of  $-13 \text{ W m}^{-2}$  and modal value of  $-20 \text{ W m}^{-2}$  (Figs. 6f and 7a). The negative residual heat flux implies that the surface is losing energy, as we would expect in winter (Walden et al. 2017a). Individual terms of the surface energy budget reveal that the

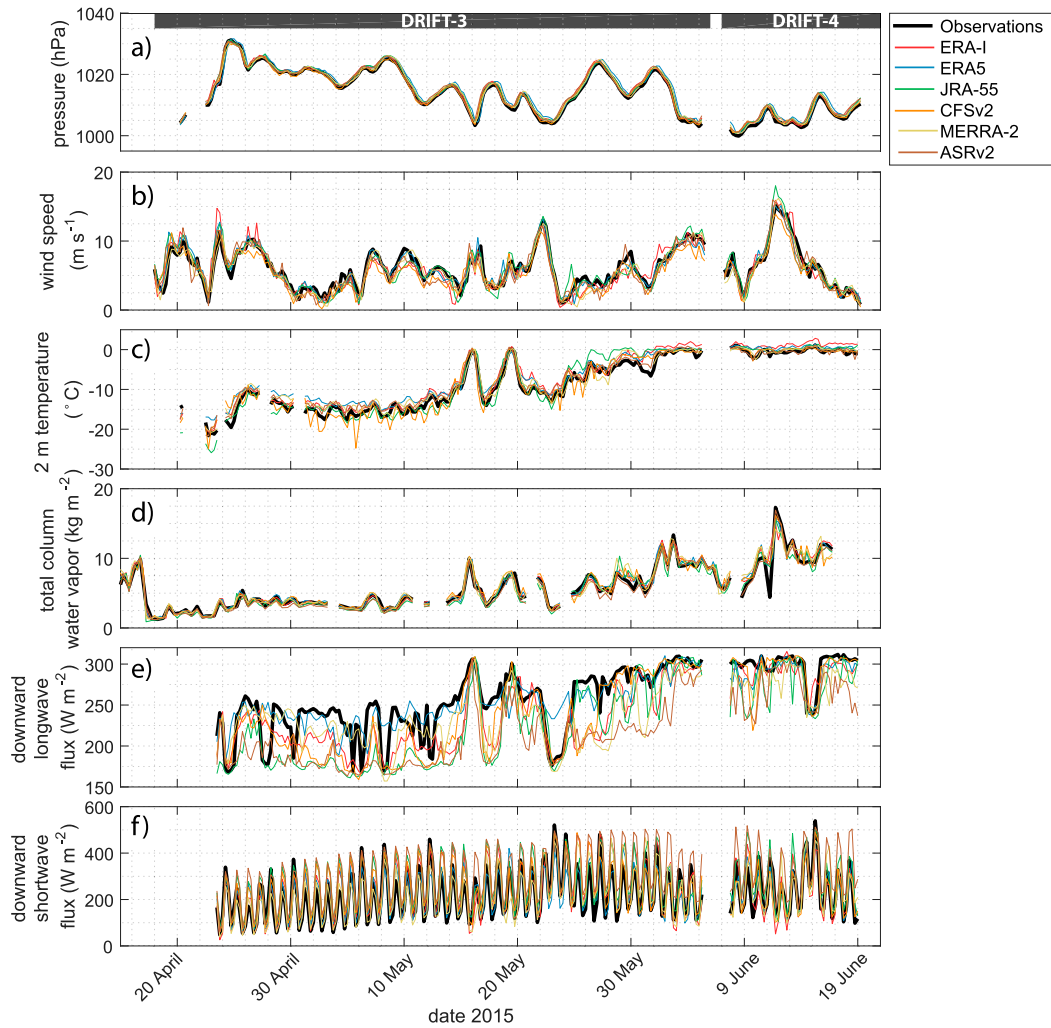


FIG. 8. Time series of N-ICE2015 observations (black) for Drifts 3 and 4, covering spring (April–May) and early summer (June), compared with values from the six reanalyses (colors). (a) Mean sea level pressure, (b) 10-m wind speed, (c) 2-m air temperature, (d) total column water vapor, (e) downward longwave radiative flux at surface, and (f) downward shortwave flux at surface.

radiative cooling is partially balanced by a downward sensible heat flux (Figs. 5 and 6; Table 2). In the reanalyses, the mean winter residual heat fluxes range from  $-26 \text{ W m}^{-2}$  in JRA-55 to  $-88 \text{ W m}^{-2}$  in CFSv2 (Figs. 6f and 7a; Table 2). Hence, all of the reanalyses have substantial negative biases. The overly negative energy budget in the reanalyses is caused by the near-surface winter warm bias, and thus overly strong radiative cooling. The bias is also further compounded by the large negative sensible and latent heat flux biases (Table 2).

### b. Spring and early summer

The spring and summer period of N-ICE2015 cover the third and fourth ice drifts from 18 April to 5 June and 7 June to 21 June 2015. These drifts are situated in closer proximity to the ice edge compared with the two winter

drifts (Fig. 1). With the exception of two warm events on 16 May and 19 May, associated with storms, near-surface temperatures in spring do not rise above  $-10^\circ\text{C}$  until 24 May (Cohen et al. 2017). Following this date, the near-surface air temperature, total column water vapor, and downward longwave flux increase progressively until 1 June, when the 2-m temperature reaches a near-constant  $0^\circ\text{C}$  (Fig. 8). We classify 1 June as the onset of summer (Cohen et al. 2017), although this timing is likely influenced by the ship's drift reaching close proximity to the ice edge as well as the seasonal progression (Fig. 1).

### 1) ANALYSIS VARIABLES: SURFACE METEOROLOGY AND VERTICAL PROFILES

Similar to the winter season, we find close agreement between the reanalyses analysis fields and observations

of mean sea level pressure, 2-m temperature, 10-m wind speed, and total column water vapor during spring and early summer (Fig. 8; Table 1).

Correlation coefficients between the reanalyses and observed 2-m temperature are high during spring, ranging from 0.93 to 0.98 (Table 1). After temperatures approach 0°C, during summer, there is less variability and so correlations are lower (0.57–0.81). CFSv2 has a nonsignificant cold bias during spring. However, all other reanalyses have warm biases in both spring and summer (Figs. 8c and 9a; Table 1). ERA5 (+1.7°C) has a larger warm bias than ERA-I (+1.3°C) during the cooler spring months, but during the summer period ERA5 (+0.8°C) has a smaller bias than ERA-I (+1.6°C) (Figs. 8c and 9a; Table 1). Near-surface air temperature biases and RMSEs are smaller during spring and summer compared with winter, in all reanalyses (Table 1). Observations from N-ICE2015 show that the surface layer was frequently unstable during spring (Walden et al. 2017a; Kayser et al. 2017). The smaller temperature biases during spring and summer, compared with winter, are therefore consistent with reanalyses having a temperature- and/or stability-dependent warm bias, with the largest biases during cold-stable conditions.

Correlation coefficients for the total column water vapor are high (0.94–0.99) during spring. As with the 2-m temperature, correlation coefficients are lower during summer (0.80–0.94), compared with the winter and spring seasons. Absolute biases and RMSEs are also larger during summer, compared with winter and spring, although this reflects higher background water vapor content and variability (Table 1; Figs. 2d, 3f, 8d, and 9c). JRA-55 and ASRv2 have dry biases in all seasons, ranging from  $-0.1$  to  $-0.4 \text{ kg m}^{-2}$  in spring and summer (Table 1; Fig. 9c). The other four reanalyses have moist biases in spring and summer, ranging from  $+0.1$  to  $+0.5 \text{ kg m}^{-2}$ , although the biases are often nonsignificant (Figs. 8d and 9c; Table 1).

Correlation coefficients between the reanalyses and observed 10-m wind speed increase from 0.85–0.91 in spring to 0.94–0.97 in summer (Fig. 8b; Table 1). RMSEs during the spring and summer are also smaller than winter values in all reanalyses. During winter, most reanalyses have a small positive 10-m wind speed bias, whereas in spring biases are predominantly negative (Figs. 3d and 9b; Table 1). During the summer period, three reanalyses have a positive wind speed bias and three have negative biases, and most of the biases are nonsignificant. ERA5 performs better than ERA-I for the wind speed during winter, spring, and summer, with higher correlation coefficients, smaller biases, and smaller RMSEs in each season (Table 1). Interestingly, despite the higher horizontal resolution and vertical resolution in ASRv2 than

most of the global reanalyses, it does not perform noticeably better for the 10-m wind speed (Table 1). This may reflect the fact that our observations are from the Arctic Ocean, far away from the complex topography that is better resolved by this regional reanalysis.

Previous studies have shown that atmospheric reanalyses have difficulties simulating realistic clouds, particularly during spring and summer months (Walsh et al. 2009; Lindsay et al. 2014; Wesslén et al. 2014). We therefore focus our analyses of radiosondes from the spring and summer months of N-ICE2015 on the presence of clouds. We choose three sets of conditions to study, with two examples from each case (Fig. 10). The first case corresponds to clear-sky conditions, which were observed on 8 and 23 May 2015. The second case is where thick clouds were observed down to the surface, such as on 25 May and 2 June. The final case corresponds to times when lifted cloud layers were present. Examples of these conditions occurred on 30 April and 6 May.

There were relatively few cloud-free days during the N-ICE2015 spring and summer (Cohen et al. 2017; Walden et al. 2017a). However, on these cloud-free days, most of the reanalyses simulate the shape of the moisture profiles relatively well, albeit with a tendency toward a positive relative humidity bias near the surface in many products (Figs. 10a,b). For both examples, ERA5 simulates a spurious thin cloud layer at 950–975 hPa (Figs. 10a,b). On 23 May, ASRv2 also has a distinct moist bias at 750 hPa.

The reanalyses mostly capture the general shape of moisture profiles at times when thick cloud layers extend close to the surface, below 900 hPa (Figs. 10c,d). However, the reanalyses often strongly underestimate the strength of the near-surface specific humidity inversions in these clouds. These inversions may also be simulated at the wrong height. As a result, the reanalyses often have a dry bias at the lower levels of these clouds (Figs. 10c,d). For example, on 25 May ERA-I, CFSv2, and MERRA-2 have large dry biases for both the specific and relative humidity below 850 hPa. Interestingly, on 25 May ASRv2 simulates the most accurate moisture profile, and on 2 June ASRv2 has the largest dry bias among all reanalyses.

The reanalyses typically perform worst at times when multiple cloud layers are observed (Figs. 10e,f). All of the reanalyses fail to capture the small-scale variability in the specific and relative humidity in these layers, and often the reanalyses underestimate the specific humidity within the cloud layers. As a result, the cloud layers in the reanalyses are either absent, too thin, or at the wrong height (Figs. 10e,f).

The reanalyses mostly simulate the shape of temperature and wind profiles well in spring and summer,

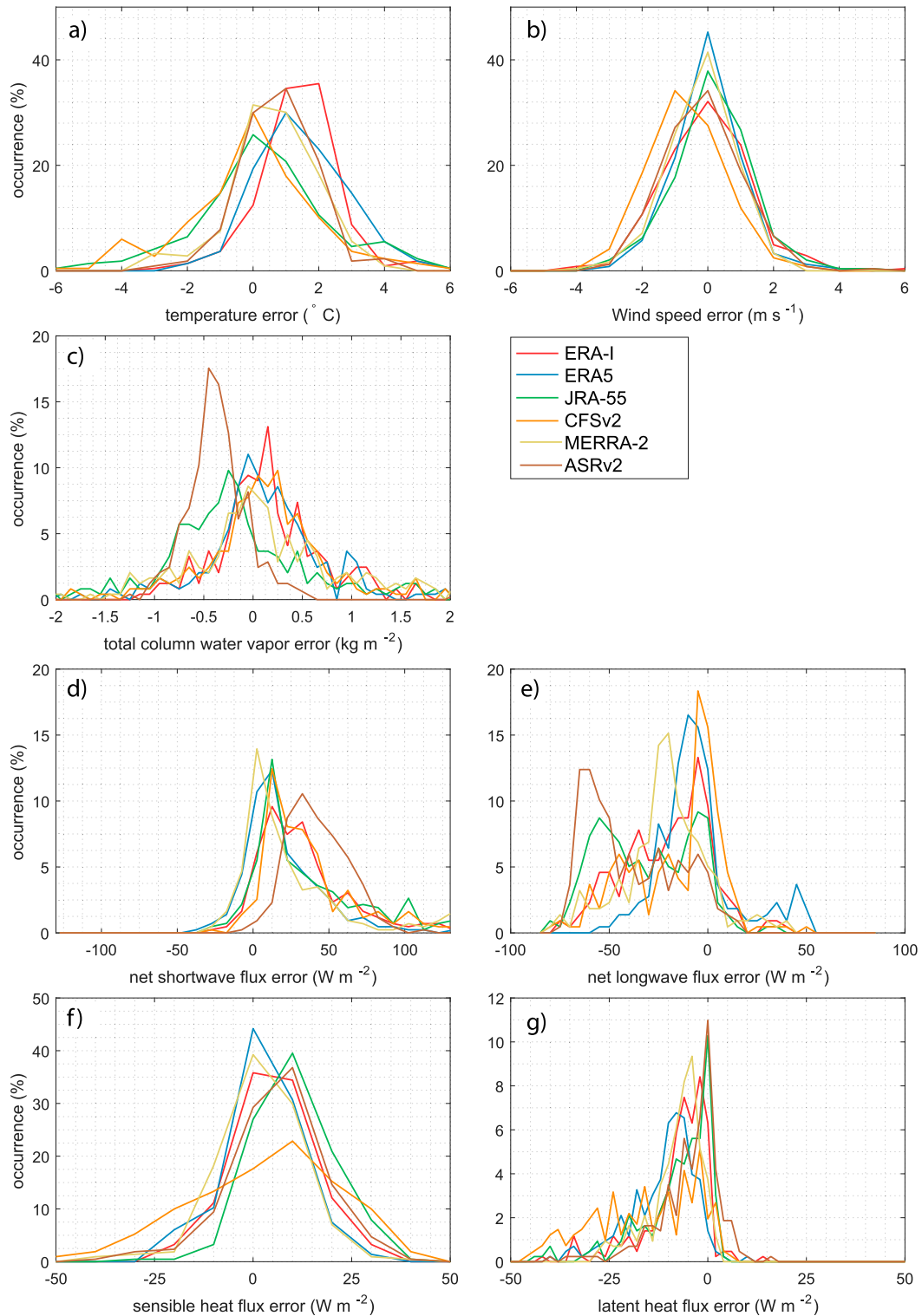


FIG. 9. Frequency distribution of errors (reanalyses – observations) during spring and summer for the six reanalyses, with respect to observations from N-ICE2015. (a) 2-m temperature ( $1^{\circ}\text{C}$  bins), (b) 10-m wind speed ( $1 \text{ m s}^{-1}$  bins), (c) total column water vapor ( $0.1 \text{ kg m}^{-2}$  bins), (d) net shortwave flux at surface ( $10 \text{ W m}^{-2}$  bins), (e) net longwave flux at surface ( $5 \text{ W m}^{-2}$  bins), (f) surface sensible heat flux ( $10 \text{ W m}^{-2}$  bins), and (g) surface latent heat flux ( $2 \text{ W m}^{-2}$  bins).

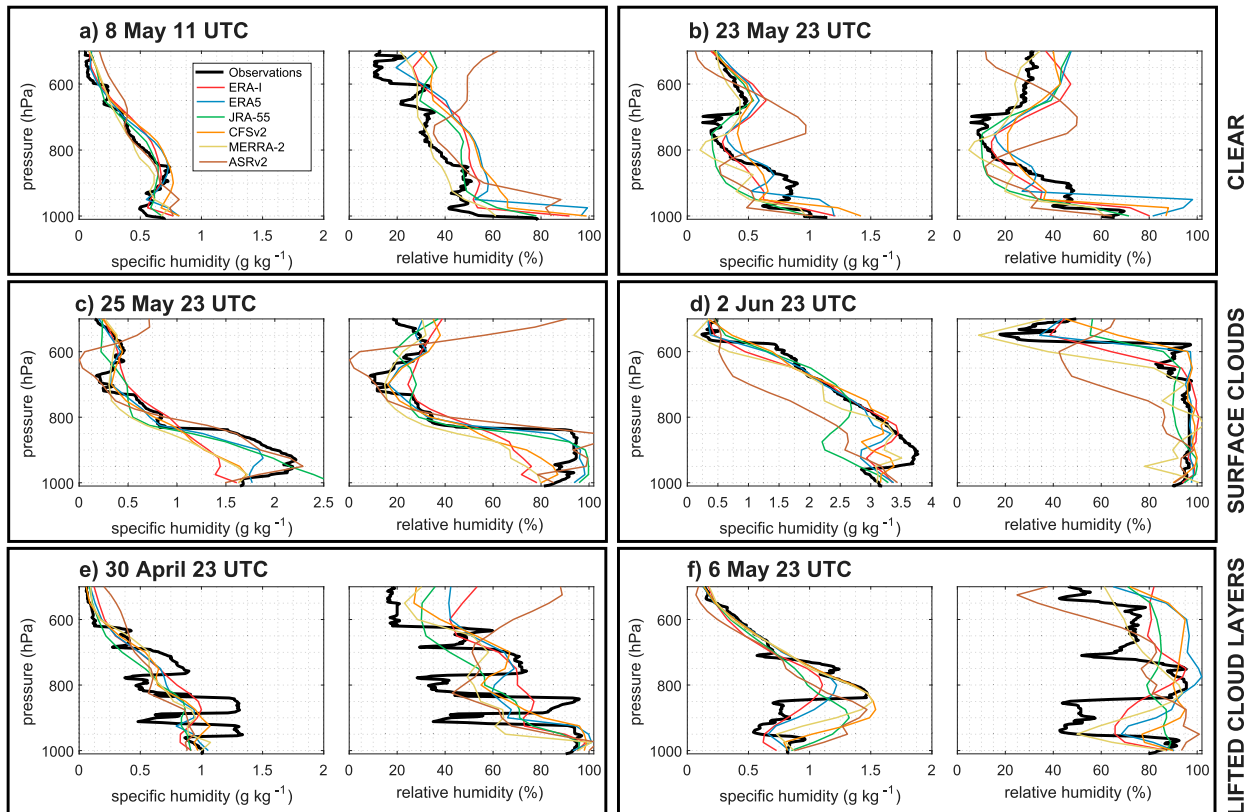


FIG. 10. Examples of vertical profiles of specific and relative humidity from N-ICE2015 radiosondes, compared with reanalyses for (a),(b) clear days, (c),(d) days with clouds close to surface, and (e),(f) lifted cloud layers.

including the six examples shown here (Fig. 11). However, the reanalyses frequently underestimate the strength of surface, elevated, and/or cloud-top temperature inversions. For example, on 25 May all of the reanalyses simulate the observed cloud-top inversion at 825 hPa (Fig. 11c). However, the strength of this inversion is substantially underestimated in all reanalyses. The strong inversion in the observations likely indicates the presence of a cloud-top liquid water layer. Such layers of cloud liquid water generate strong radiative cooling, leading to the formation of inversions (Morrison et al. 2011). Reanalyses are known to underestimate the concentration of liquid water in Arctic clouds (Pithan et al. 2016, 2014; Engström et al. 2014; Wesslén et al. 2014; de Boer et al. 2014). The absence of this cloud liquid water layer in the reanalyses, or presence of less liquid water, would result in less radiative cooling and thus a weaker cloud-top inversion, as we see (Fig. 11c). Small cloud-top inversions are also visible in the observations for the elevated cloud layers at 850 hPa on 30 April and 750 hPa on 6 May (Figs. 11e,f). However, temperature inversions are not visible at these heights in any reanalysis. This could indicate that the cloud layers are absent in the reanalyses, or that the vertical resolution

of the reanalyses is not sufficient to accurately resolve these features. Cloud liquid water and ice content measurements are not available for N-ICE2015, and so cannot be evaluated further here.

## 2) FORECAST VARIABLES: SURFACE HEAT FLUXES AND ENERGY BUDGET

The reanalyses perform significantly worse for the radiative fluxes during the spring and summer months of N-ICE2015, compared with winter (Figs. 2, 6, and 8; Table 1). While most reanalyses exhibit a small positive bias for the downward longwave radiative flux in winter, there are substantial negative biases in spring and summer (Figs. 2e and 8e; Table 1). With the exception of ERA5, biases in spring range from  $-19 \text{ W m}^{-2}$  in MERRA-2 to  $-46 \text{ W m}^{-2}$  in ASRv2. During summer, biases range from  $-3 \text{ W m}^{-2}$  in CFSv2 to  $-31 \text{ W m}^{-2}$  in ASRv2. Correlation coefficients for the downward longwave flux range from 0.92 to 0.95 during winter, but just from 0.38 to 0.80 in spring and summer. RMSEs are largest during the spring period and range from  $27 \text{ W m}^{-2}$  in ERA5 to  $54 \text{ W m}^{-2}$  in ASRv2. For comparison, RMSEs in winter range from  $20\text{--}28 \text{ W m}^{-2}$ , and  $17\text{--}40 \text{ W m}^{-2}$  in summer (Table 1). Interestingly, ERA5



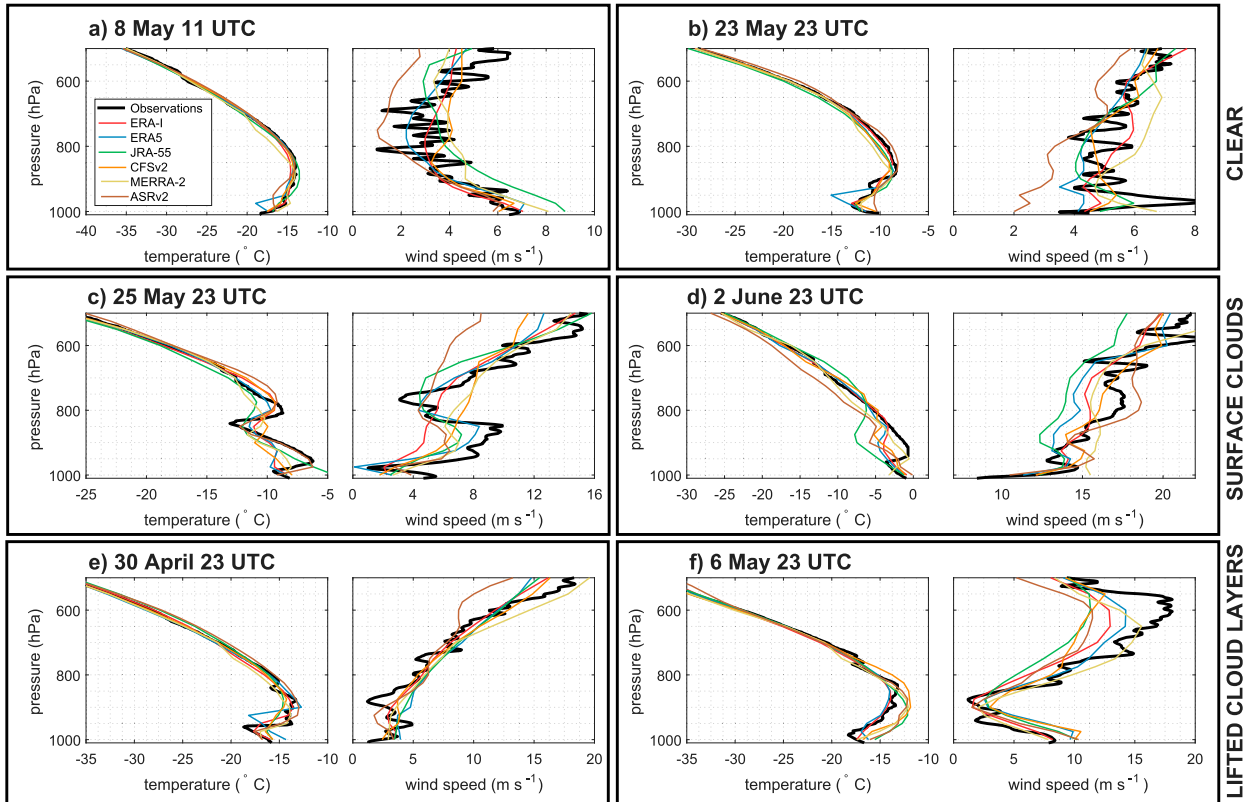


FIG. 11. Examples of vertical profiles of temperature and wind speed from N-ICE2015 radiosondes, compared with reanalyses for (a),(b) clear days, (c),(d) days with clouds close to surface, and (e),(f) lifted cloud layers.

is the only reanalysis with a positive and/or non-significant downward longwave bias during spring. The magnitudes of the downward longwave biases and RMSEs in ERA5 are substantially smaller than ERA-I and the other reanalyses during spring and summer (Fig. 8e; Table 1). ERA5 is also the only reanalyses to have a smaller downward longwave bias in spring compared with winter and summer. Nonetheless, the correlation coefficient in spring is lower than summer and winter (Table 1).

Correlation coefficients between the observed net longwave radiative fluxes and reanalyses are very low in spring and summer (Fig. 6a; Table 1). These values range from 0.15 (ERA5) to 0.41 (CFSv2 and ERA-I) during spring, and from 0.39 (ASRv2) to 0.80 (ERA5) in summer. The largest biases ( $-43 \text{ W m}^{-2}$ ) and RMSEs ( $49 \text{ W m}^{-2}$ ) are found in spring, rather than summer (Fig. 6a; Table 1). As with the winter season, all reanalyses have a negative (upward) net longwave bias during spring and summer (Fig. 9e; Table 1). During spring and summer, this negative bias is primarily driven by a negative bias in the downward longwave flux (Fig. 8e). In contrast, during winter the bias is the result of the warm bias at the surface and thus overly strong

upward longwave flux (Fig. 2). ERA5 has the smallest RMSEs among all reanalyses for the net longwave flux during spring and summer, and performs considerably better than ERA-I. In contrast, the regional reanalysis ASRv2 has the largest net longwave biases among all products in all seasons (Table 1).

Most reanalyses have positive (i.e., downward) biases for the surface net and downward shortwave fluxes during spring (Figs. 6b and 8f; Table 1). Spring biases for the net shortwave flux range from  $+18 \text{ W m}^{-2}$  in ERA5 to  $+38 \text{ W m}^{-2}$  in ASRv2 (Fig. 9d; Table 1). For the downward shortwave flux, spring biases range from  $-2 \text{ W m}^{-2}$  in MERRA-2 to  $+79 \text{ W m}^{-2}$  in ASRv2. MERRA-2 is the only reanalysis with a negative bias, and this is nonsignificant. During summer, four reanalyses have negative downward shortwave flux biases, three of which are nonsignificant (Table 1). Summer biases range from  $-15 \text{ W m}^{-2}$  in CFSv2 to  $+93 \text{ W m}^{-2}$  in ASRv2. Despite the negative downward shortwave flux biases, all reanalyses have positive biases for the net shortwave flux in summer, ranging from  $+41 \text{ W m}^{-2}$  in ASRv2 to  $+65 \text{ W m}^{-2}$  in JRA-55. Especially during summer months, the net shortwave bias is often more positive than the downward

shortwave bias (Table 1). This indicates that the surface albedo in the reanalyses is too low, compared with the observations. Reanalyses are known to treat the albedo of snow-covered sea ice crudely, resulting in substantial errors (de Boer et al. 2014; Wesslén et al. 2014). Moreover, we note again that the observations are point measurements made over snow-covered sea ice, while the reanalyses provide grid cell averages including an open water fraction with low albedo (Table 2). The lowest mean ice concentration during summer is 0.67 in ERA-I.

Low-level clouds were remarkably persistent throughout the spring months of N-ICE2015 (Walden et al. 2017a). However, it seems that with the exception of ERA5, these persistent clouds are not accurately simulated by the reanalyses (Fig. 6). As a result, we see negative downward longwave flux biases at the surface in most reanalyses, and overly strong radiative cooling. Typically, these biases are partially compensated positive downward and net shortwave flux biases (Figs. 6, 8, and 9; Table 1), which further suggest a lack of clouds and/or poorly simulated cloud properties and humidity profiles (Wyser et al. 2008). It is interesting that while optimized for the polar environment, ASRv2 has the largest radiative flux biases among all reanalyses (Figs. 8 and 9; Table 1), suggesting that this reanalysis does not have an improved representation of spring clouds in the Arctic.

ASRv2 and JRA-55 clearly suffer from a similar problem of absent clouds during summer, resulting in large positive biases for the downward shortwave flux and negative biases for the net longwave flux (Figs. 8e,f and 9d,e; Table 1). In contrast, the other reanalyses have negative biases for both the downward shortwave flux and downward longwave flux (Table 1). In these cases, it appears that clouds are likely present in the reanalyses and observations, but the cloud properties (e.g., phase, temperature, height, and liquid and/or ice water content) are simulated poorly by the reanalyses (Wyser et al. 2008). As a result, the simulated clouds reflect too much incoming shortwave radiation and emit too little longwave radiation downward toward the surface.

As with winter, the mean observed latent heat fluxes over sea ice are near zero throughout the spring and summer drifts (Fig. 6e). However, all reanalyses simulate large negative (upward) latent heat fluxes (Table 2; Figs. 6e and 9g). For example, all reanalyses simulate sustained large negative latent heat fluxes, ranging from  $-25$  to  $-60 \text{ W m}^{-2}$ , during a storm event on 11 June with mean wind speeds of  $15 \text{ m s}^{-1}$  (Fig. 6). However, the observed daily mean latent heat flux at this time is positive and near zero. Latent heat flux biases in spring and summer are mostly larger than the winter

season (Table 2). This may reflect the closer proximity to the ice edge of the winter drifts, and thus lower mean ice concentration within the reanalyses grid cells (Fig. 1; Table 2). However, JRA-55 has a mean ice fraction of 1.00 throughout the spring and summer drifts and has a substantial negative latent heat flux bias of  $-13 \text{ W m}^{-2}$  during the final weeks of the field campaign (Fig. 9g). Furthermore, in spring and summer, the temperature contrast between the ice and open water is much smaller than in winter, reducing the difference in turbulent fluxes over these two surfaces. It is thus clear that major errors exist in the reanalyses for surface latent heat fluxes, regardless of ice concentration.

The sensible heat fluxes observed over sea ice are of smaller magnitude in spring and summer compared with winter (Table 1; Fig. 6d). The average fluxes are also negative (i.e., upward) rather than positive (Table 2). This is consistent with the fact that unstable conditions were frequently measured at the surface during spring (Walden et al. 2017a). Typically, most reanalyses have positive sensible heat flux biases during spring and summer (Fig. 9f; Table 2). In spring, CFSv2 is the only reanalysis with a negative sensible heat flux bias of  $-1 \text{ W m}^{-2}$ . The other reanalyses have positive biases ranging from  $+1$  to  $+12 \text{ W m}^{-2}$ , although several of these are nonsignificant (Table 2). JRA-55 has the largest bias during spring and an overall mean positive flux, in contrast to the observed negative flux. During early summer the observed sensible heat flux is  $-3 \text{ W m}^{-2}$ , but all reanalyses have large positive biases ranging from  $+2$  to  $+19 \text{ W m}^{-2}$ . The bias in JRA-55 is of similar magnitude to the other reanalyses during summer, despite having no open water fraction (Fig. 9d; Tables 1 and 2).

In spring, the observed residual heat flux is near zero (Walden et al. 2017a) (Fig. 6f). At this time most reanalyses have a negative bias, albeit of smaller magnitude than winter (Fig. 7b). These biases ranged from  $0 \text{ W m}^{-2}$  in ASRv2 to  $-17 \text{ W m}^{-2}$  in MERRA-2 (Table 2). The negative biases are primarily caused by a combination of the near-surface warm bias and lack of clouds, which result in overly strong radiative cooling. This bias is compounded by the negative latent heat flux bias in all reanalyses (Table 2). However, the biases are partially compensated by the positive net shortwave bias and, in most cases, positive sensible heat flux bias (Fig. 6). Importantly, the small residual heat flux bias in ASRv2 reflects several large compensating biases, and not the accurate representation of the surface energy budget (Tables 1 and 2).

During early summer, the residual heat flux is positive with a mean value of  $32 \text{ W m}^{-2}$  (Walden et al. 2017a). All of the reanalyses have a positive residual heat flux at

this time and moreover, accurately captured the timing of the transition toward a positive energy budget (Fig. 6f). Nonetheless, all reanalyses have large positive biases ranging from  $+9 \text{ W m}^{-2}$  in ASRv2 to  $+51 \text{ W m}^{-2}$  in CFSv2 (Fig. 7c; Table 2). The primary source of this bias is the positive net shortwave flux at the surface, caused by the low surface albedo and in some cases poorly simulated clouds. The low albedo in the reanalyses is likely due to a combination of the open water fraction of the grid cell and simplistic treatment of snow on sea ice; we note neither the net shortwave biases nor the residual heat flux biases are negatively correlated with mean ice concentration in the reanalyses (Tables 1 and 2). The positive net shortwave bias is partially compensated by stronger radiative cooling and an overly strong upward latent heat flux from the surface, while a positive sensible heat flux compounds the bias in all reanalyses (Figs. 9d–g; Table 2). Thus, while large, the net bias for the residual heat flux masks several larger compensating biases in the individual components of the energy budget. Often these biases are of similar or greater magnitude than the observed flux (Tables 1 and 2).

#### 4. Discussion

A winter warm bias in atmospheric reanalyses, over Arctic sea ice, has been reported by many earlier studies (Beesley et al. 2000; Makshtas et al. 2007; Liu et al. 2008; Tjernström and Graversen 2009; Lindsay et al. 2014; Cullather et al. 2016; Graham et al. 2017a). It is therefore not surprising that we identify a similar bias here in the latest generation of reanalyses. Interestingly, despite the higher vertical resolution of the newly released ERA5, the winter warm bias is larger than that in ERA-I. More reassuringly, the Arctic regional reanalyses, ASRv2 simulates stable surface temperature inversions more accurately than any of the other reanalyses, and has the highest combined score for correlation, bias and RMSE (Figs. 2c, 3a,b, and 4a; Table 1). This can likely be attributed to the land surface model used in ASRv2 being optimized for polar environments (Hines et al. 2015).

Major improvements have been made in the treatment of atmospheric moisture, clouds, and precipitation in atmospheric reanalyses over the last two decades (Walsh and Chapman 1998; Cullather et al. 2000; Engström et al. 2014; Sotiropoulou et al. 2015; Bromwich et al. 2016; Boisvert et al. 2018). Nonetheless, problems with Arctic clouds continue to afflict reanalyses in all seasons (Cullather et al. 2016). Among reanalyses analyzed previously, products from the ECMWF have often been found to perform better than those by other groups with respect to Arctic clouds (Walsh et al. 2009; Lindsay et al.

2014; Wesslén et al. 2014; de Boer et al. 2014). Consistent with this pattern, we find that ERA5 has the smallest radiative flux biases during the spring and summer periods of N-ICE2015 (Figs. 6a,b, 8e,f, and 9d,e; Table 1). In contrast, ASRv2 has some of the largest biases related to clouds during spring and summer. Results from the Arctic Summer Cloud–Ocean Study (ASCOS) campaign during autumn 2008, also indicated that ASRv1 and ASRv2 performs less well than ERA-I with respect to Arctic clouds (Wesslén et al. 2014). Nonetheless, we urge caution in making broad statements about how representative the N-ICE2015 biases are for the wider Arctic and specific seasons. It is likely that the accurate simulation of clouds varies by location, season and weather pattern in all reanalyses, including ERA5.

All of the reanalyses show low skill simulating turbulent heat fluxes over sea ice (Figs. 5, 6, and 9; Tables 1 and 2). The simulated fluxes are often of the wrong magnitude and/or direction. Notably, the apparent sensible heat flux biases for N-ICE2015 are of larger magnitude than those identified in earlier reanalyses and models for the SHEBA campaign (Cullather and Bosilovich 2012; Beesley et al. 2000). The larger sensible heat flux biases in reanalyses for N-ICE2015, compared with SHEBA, likely reflect the closer proximity of the ice edge and thinner, younger sea ice during N-ICE2015. This results in a higher open water fraction in the reanalysis grid cell. During the winter season, large localized turbulent heat fluxes occur over open water areas and leads (Maykut 1978; Marcq and Weiss 2012). These large open water fluxes are included in the reanalyses grid cell average fluxes, but are not reflected in the point measurements made over sea ice. This disparity can result in large apparent biases in the turbulent heat fluxes simulated by reanalyses.

Turbulent heat fluxes are notoriously difficult to model, and depend on the accurate simulation of multiple variables including the near-surface temperature and humidity profiles, wind speed and direction, and surface radiative fluxes. While reanalyses typically capture the general evolution of these fields, at any given time errors may exist that propagate through into the calculation of the turbulent heat fluxes. A recent study demonstrated that by swapping input data from ERA-I to the Atmospheric Infrared Sounder (AIRS) while using the same flux calculation scheme, differences in the simulated latent heat flux could reach  $40 \text{ W m}^{-2}$  over the Beaufort Sea (Boisvert et al. 2015). Impressively, satellite-derived sensible and latent heat fluxes indicate relatively high skill for capturing the observed N-ICE2015 turbulent heat fluxes (Taylor et al. 2018). Sensible and latent heat fluxes derived from AIRS had RMSEs of just 5 and  $1 \text{ W m}^{-2}$ , respectively (Taylor et al.

TABLE 3. Table highlights the strengths and weaknesses of each reanalysis. For the strengths we list the variables with the highest score in a given season, and weaknesses show the lowest scores. Scores are provided in Table 1.

Reanalyses	Strengths	Weaknesses
ERA-I	Downward longwave radiation in winter Total column water vapor in spring	2-m temperature in summer
ERA5	10-m wind speed in winter and spring Surface turbulent and radiative fluxes in spring and summer	2-m temperature in winter and spring
JRA-55	Turbulent heat fluxes in winter	Water vapor path in winter (cold and dry bias aloft). 10-m wind speed in winter and summer
CFSv2	Captures longwave radiative fluxes well in all seasons.	10-m wind speed in spring Turbulent heat fluxes in winter and summer
MERRA-2	2-m temperature in summer 10-m wind speed in summer	Total column water vapor in spring and summer Downward longwave radiation in winter
ASRv2	2-m temperature in winter and spring Total column water vapor in winter and summer	Radiative fluxes in all seasons Turbulent heat fluxes in spring

2018). In contrast, RMSEs in the six reanalyses are up to  $70 \text{ W m}^{-2}$  (Table 1). Our understanding of turbulent heat fluxes in the Arctic is severely hampered by a lack of in situ observations and the short spatial and temporal scales of these fluxes. The strong performance of satellite data for measuring these parameters is therefore encouraging (Taylor et al. 2018).

We find substantial negative residual heat flux biases (up to  $-76 \text{ W m}^{-2}$ ) in all reanalyses during winter (Figs. 6f and 7a; Table 2). In CFSv2, the bias is 5 times larger than the observed flux. Likewise, during the summer period all reanalyses have large positive residual heat flux biases (Fig. 7). The smallest residual heat flux biases are found during spring. However, this is the result of large compensating biases in the individual terms of the surface energy budget (Fig. 9). These large apparent biases in the surface energy budget over sea ice must be taken into consideration if using these products to force sea ice models (Tables 1 and 2).

## 5. Summary

In this study, we evaluate the performance of six atmospheric reanalyses (ERA-I, ERA5, JRA-55, CFSv2, MERRA-2, and ASRv2) over Arctic sea ice from winter until early summer. The reanalyses are evaluated against a comprehensive suite of observations from the N-ICE2015 field campaign, which consists of a rare 5-month ice drift in pack ice north of Svalbard from January–June 2015.

Overall, the reanalyses perform remarkably well for the winter season (January–March). We find high correlation coefficients ( $>0.90$ ) between the reanalyses and observations for most of the surface meteorology parameters, as well as the downwelling longwave radiative flux. Nonetheless, all reanalyses have a positive winter 2-m temperature bias that ranges from  $+1.1^\circ\text{C}$  in JRA-55 to  $+3.8^\circ\text{C}$  in CFSv2. This winter warm bias is

associated primarily with poorly resolved (too weak) surface inversions during cold-stable periods. While JRA-55 has the best near-surface temperature distribution and smallest warm bias during winter, it suffers from a large cold and dry bias aloft. ASRv2 simulates surface inversions most accurately. Interestingly, the winter warm bias is larger in the newly released ERA5 than ERA-I. In all reanalyses, the winter warm bias results in an excessive upward and net longwave flux from the surface. Mean winter net longwave biases range from  $-3 \text{ W m}^{-2}$  in MERRA-2 to  $-19 \text{ W m}^{-2}$  in ASRv2.

The representation of radiative fluxes in reanalyses is found to be worse during spring (April–June), compared with winter and summer. Correlation coefficients for the net longwave radiative flux in spring range from 0.15 in ERA5 to 0.41 in CFSv2. All reanalyses fail to simulate accurately the persistent clouds observed during spring. This results in a pattern of negative net longwave biases at the surface and positive shortwave biases. Notably, ERA5 performs better than ERA-I, and all other reanalyses, in its simulation of surface radiative fluxes during spring and summer. In contrast, ASRv2 has the largest radiative flux biases in spring, with  $+38 \text{ W m}^{-2}$  for the net shortwave flux and  $-43 \text{ W m}^{-2}$  for the net longwave flux.

Our analyses demonstrate that reanalyses have major difficulties resolving individual components of the surface energy budget over sea ice. All reanalyses show poor skill in simulating surface turbulent heat fluxes over sea ice. We find low correlation coefficients (0.02), large apparent biases ( $46 \text{ W m}^{-2}$ ), and large RMSEs ( $70 \text{ W m}^{-2}$ ) for the sensible and latent heat fluxes during winter, spring, and summer. These apparent errors can partially be explained by the difference between the point observations made over sea ice and grid cell average values outputted by the reanalyses, which include a small but important open water fraction (Table 2). In winter, negative biases in the turbulent

heat fluxes compound the negative net longwave radiative biases to generate large negative residual heat flux biases in all reanalyses. These residual heat flux biases range from  $-13 \text{ W m}^{-2}$  in JRA-55 to  $-76 \text{ W m}^{-2}$  in CFSv2. In summer, we find large positive residual heat flux biases in all reanalyses, ranging from  $+9 \text{ W m}^{-2}$  in ASRv2 to  $+51 \text{ W m}^{-2}$  in CFSv2, which are the result of several large compensating biases in the energy budget.

We conclude that all of the reanalyses products considered in this study show high skill in the simulation of analysis fields during the N-ICE2015 period (Table 1). However, some large errors exist in the simulation of radiative and turbulent heat fluxes. Therefore the representation of the surface energy budget over sea ice is often relatively poor. No single reanalysis product is superior overall. Instead, each reanalysis has strengths and weaknesses for different variables, which vary by season, as we summarize in Table 3. Finally, we urge caution in making broad statements about how representative the seasonal N-ICE2015 biases are for the wider Arctic.

*Acknowledgments.* We wish to acknowledge three anonymous reviewers for their constructive comments. We thank all members of the N-ICE2015 team and the crew from R/V *Lance* for their help collecting and producing the N-ICE2015 datasets. We thank Lesheng Bai of the Polar Meteorology Group at The Ohio State University for supplying the ASRv2 output, interpolated on to pressure levels, over the N-ICE2015 domain. This work has been supported by the Norwegian Polar Institute's Centre for Ice, Climate and Ecosystems (ICE) through the N-ICE project. Further support was provided to R.M.G., L.C., B.S., A.R., and S.R.H. by the German Academic Exchange Service (DAAD) and PPP Norway. L.C. was supported by the Arktis 2030 program of the Ministries of Foreign Affairs and Climate and Environment of Norway, through the project ID Arctic. N.R. acknowledges support from the Erasmus+ programme for traineeships for a 3-month internship at the Norwegian Polar Institute. B.S. and A.R. acknowledge support by the SFB/TR 172 "Arctic Amplification: Climate Relevant Atmospheric and Surface Processes, and Feedback Mechanisms (AC)<sup>3</sup>" funded by the Deutsche Forschungsgemeinschaft (DFG). V.P.W. acknowledges support from the U.S.–Norway Fulbright Distinguished Arctic Chair Program and Washington State University. N-ICE2015 datasets cited in this study are available through Norwegian Polar Institute's data archive ([data.npolar.no](http://data.npolar.no)). ERA-I, ERA5, JRA-55, CFSv2, and ASRv2, were downloaded through the National Center for Atmospheric Research's research data archive at the University of Colorado

([rda.ucar.edu](http://rda.ucar.edu)). MERRA-2 files were downloaded from NASA's Earth Data archive ([earthdata.nasa.gov](http://earthdata.nasa.gov)).

## REFERENCES

- Beesley, J. A., C. S. Bretherton, C. Jakob, E. L. Andreas, J. M. Intrieri, and T. A. Uttal, 2000: A comparison of cloud and boundary layer variables in the ECMWF forecast model with observations at Surface Heat Budget of the Arctic Ocean (SHEBA) ice camp. *J. Geophys. Res.*, **105**, 12 337–12 349, <https://doi.org/10.1029/2000JD900079>.
- Bekryaev, R. V., I. V. Polyakov, and V. A. Alexeev, 2010: Role of polar amplification in long-term surface air temperature variations and modern Arctic warming. *J. Climate*, **23**, 3888–3906, <https://doi.org/10.1175/2010JCLI3297.1>.
- Boisvert, L. N., and J. C. Stroeve, 2015: The Arctic is becoming warmer and wetter as revealed by the Atmospheric Infrared Sounder. *Geophys. Res. Lett.*, **42**, 4439–4446, <https://doi.org/10.1002/2015GL063775>.
- , D. L. Wu, T. Vihma, and J. Susskind, 2015: Verification of air/surface humidity differences from AIRS and ERA-Interim in support of turbulent flux estimation in the Arctic. *J. Geophys. Res. Atmos.*, **120**, 945–963, <https://doi.org/10.1002/2014JD021666>.
- , M. A. Webster, A. A. Petty, T. Markus, D. H. Bromwich, and R. I. Cullather, 2018: Intercomparison of precipitation estimates over the Arctic Ocean and its peripheral seas from reanalyses. *J. Climate*, **31**, 8441–8462, <https://doi.org/10.1175/JCLI-D-18-0125.1>.
- Bosilovich, M., and Coauthors, 2015: MERRA-2: Initial evaluation of the climate. NASA/TM-2015-104606, Vol. 43, 139 pp., <https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich803.pdf>.
- Bromwich, D. H., K. M. Hines, and L.-S. Bai, 2009: Development and testing of Polar Weather Research and Forecasting model: 2. Arctic Ocean. *J. Geophys. Res.*, **114**, D08122, <https://doi.org/10.1029/2008JD010300>.
- , A. B. Wilson, L.-S. Bai, G. W. K. Moore, and P. Bauer, 2014: Contrasting the regional Arctic System Reanalysis with the global ERA-Interim Reanalysis for the Arctic. *Quart. J. Roy. Meteor. Soc.*, **142**, 644–658, <https://doi.org/10.1002/qj.2527>.
- , —, L. S. Bai, G. W. K. Moore, and P. Bauer, 2016: A comparison of the regional Arctic System Reanalysis and the global ERA-Interim Reanalysis for the Arctic. *Quart. J. Roy. Meteor. Soc.*, **142**, 644–658, <https://doi.org/10.1002/qj.2527>.
- , and Coauthors, 2018: The Arctic System Reanalysis, version 2. *Bull. Amer. Meteor. Soc.*, **99**, 805–828, <https://doi.org/10.1175/BAMS-D-16-0215.1>.
- Cohen, L., S. R. Hudson, V. P. Walden, and M. A. Granskog, 2017: Meteorological conditions in a thinner Arctic sea ice regime from winter through summer during the Norwegian Young Sea Ice expedition (N-ICE2015). *J. Geophys. Res. Atmos.*, **122**, 7235–7259, <https://doi.org/10.1002/2016JD026034>.
- Cullather, R. I., and M. G. Bosilovich, 2012: The energy budget of the polar atmosphere in MERRA. *J. Climate*, **25**, 5–24, <https://doi.org/10.1175/2011JCLI4138.1>.
- , D. H. Bromwich, and M. C. Serreze, 2000: The atmospheric hydrology cycle over the Arctic basin from reanalyses. Part I: Comparison with observations and previous studies. *J. Climate*, **13**, 923–937, [https://doi.org/10.1175/1520-0442\(2000\)013<0923:TAHCOT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<0923:TAHCOT>2.0.CO;2).
- , T. M. Hamill, D. Bromwich, X. Wu, and P. Taylor, 2016: Systematic Improvements of Reanalyses in the Arctic

- (SIRTA): A white paper. 51 pp., <http://www.iarppcollaborations.org/uploads/cms/documents/sirta-white-paper-final.pdf>.
- de Boer, G., and Coauthors, 2014: Near-surface meteorology during the Arctic Summer Cloud Ocean Study (ASCOS): Evaluation of reanalyses and global climate models. *Atmos. Chem. Phys.*, **14**, 427–445, <https://doi.org/10.5194/acp-14-427-2014>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Dorn, W., K. Dethloff, and A. Rinke, 2009: Improved simulation of feedbacks between atmosphere and sea ice over the Arctic Ocean in a coupled regional climate model. *Ocean Modell.*, **29**, 103–114, <https://doi.org/10.1016/j.ocemod.2009.03.010>.
- Engström, A., J. Karlsson, and G. Svensson, 2014: The importance of representing mixed-phase clouds for simulating distinctive atmospheric states in the Arctic. *J. Climate*, **27**, 265–272, <https://doi.org/10.1175/JCLI-D-13-00271.1>.
- Fetterer, F., K. Knowles, W. Meier, M. Savoie, and A. K. Windnagel, 2017: Sea Ice Index, version 3 [subset: N\_201502\_extent\_v3.0.tif]. National Snow and Ice Data Center, accessed 15 August 2018, <https://doi.org/10.7265/N5K072F8>.
- Finkelstein, P. L., and P. F. Sims, 2001: Sampling error in eddy correlation flux measurements. *J. Geophys. Res.*, **106**, 3503–3509, <https://doi.org/10.1029/2000JD900731>.
- Graham, R. M., and Coauthors, 2017a: A comparison of the two Arctic atmospheric winter states observed during N-ICE2015 and SHEBA. *J. Geophys. Res. Atmos.*, **122**, 5716–5737, <https://doi.org/10.1002/2016JD025475>.
- , L. Cohen, A. A. Petty, L. N. Boisvert, A. Rinke, S. R. Hudson, M. Nicolaus, and M. A. Granskog, 2017b: Increasing frequency and duration of Arctic winter warming events. *Geophys. Res. Lett.*, **44**, 6974–6983, <https://doi.org/10.1002/2017GL073395>.
- Granskog, M. A., P. Assmy, S. Gerland, G. Spreen, H. Steen, and L. H. Smedsrud, 2016: Arctic research on thin ice: Consequences of Arctic sea ice loss. *Eos, Trans. Amer. Geophys. Union*, **97**, 22–26, <https://doi.org/10.1029/2016EO044097>.
- , I. Fer, A. Rinke, and H. Steen, 2018: Atmosphere–ice–ocean–ecosystem processes in a thinner Arctic sea ice regime: The Norwegian Young Sea ICE (N-ICE2015) expedition. *J. Geophys. Res. Oceans*, **123**, 1586–1594, <https://doi.org/10.1002/2017JC013328>.
- Harada, Y., and Coauthors, 2016: The JRA-55 Reanalysis: Representation of atmospheric circulation and climate variability. *J. Meteor. Soc. Japan*, **94**, 269–302, <https://doi.org/10.2151/jmsj.2016-015>.
- Hines, K. M., and D. H. Bromwich, 2008: Development and testing of Polar Weather Research and Forecasting (WRF) Model. Part I: Greenland Ice Sheet meteorology. *Mon. Wea. Rev.*, **136**, 1971–1989, <https://doi.org/10.1175/2007MWR2112.1>.
- , —, L. Bai, C. M. Bitz, J. G. Powers, and K. W. Manning, 2015: Sea ice enhancements to Polar WRF. *Mon. Wea. Rev.*, **143**, 2363–2385, <https://doi.org/10.1175/MWR-D-14-00344.1>.
- Hudson, S. R., L. Cohen, and V. Walden, 2015: N-ICE2015 surface meteorology [data set]. Norwegian Polar Institute, accessed 15 January 2018, <https://doi.org/10.21334/npolar.2015.056a61d1>.
- , —, and —, 2016: N-ICE2015 surface broadband radiation data [data set]. Norwegian Polar Institute, accessed 20 April 2017, <https://doi.org/10.21334/npolar.2016.a89cb766>.
- , —, M. Kayser, M. Maturilli, J. Kim, S. Park, W. Moon, and M. A. Granskog, 2017: N-ICE2015 atmospheric profiles from radiosondes [data set]. Norwegian Polar Institute, accessed 5 March 2019, <https://doi.org/10.21334/npolar.2017.216df9b3>.
- Hyland, R. W., and A. Wexler, 1983: Formulations for the thermodynamic properties of the saturated phases of H<sub>2</sub>O from 173.15 K to 473.15 K. *ASHRAE Trans.*, **89**, 500–519.
- Jakobson, E., T. Vihma, T. Palo, L. Jakobson, H. Keernik, and J. Jaagus, 2012: Validation of atmospheric reanalyses over the central Arctic Ocean. *Geophys. Res. Lett.*, **39**, L10802, <https://doi.org/10.1029/2012GL051591>.
- Kapsch, M. L., N. Skific, R. G. Graversen, M. Tjernström, and J. A. Francis, 2019: Summers with low Arctic sea ice linked to persistence of spring atmospheric circulation patterns. *Climate Dyn.*, **52**, 2497–2512, <https://doi.org/10.1007/s00382-018-4279-z>.
- Kayser, M., and Coauthors, 2017: Vertical thermodynamic structure of the troposphere during the Norwegian young sea ICE expedition (N-ICE2015). *J. Geophys. Res. Atmos.*, **122**, 10855–10872, <https://doi.org/10.1002/2016JD026089>.
- King, J., G. Spreen, S. Gerland, C. Haas, S. Hendricks, L. Kaleschke, and C. Wang, 2017: Sea-ice thickness from field measurements in the northwestern Barents Sea. *J. Geophys. Res. Oceans*, **122**, 1497–1512, <https://doi.org/10.1002/2016JC012199>.
- Klaus, D., K. Dethloff, W. Dorn, A. Rinke, and D. L. Wu, 2016: New insight of Arctic cloud parameterization from regional climate model simulations, satellite-based, and drifting station data. *Geophys. Res. Lett.*, **43**, 5450–5459, <https://doi.org/10.1002/2015GL067530>.
- Kobayashi, S., and Coauthors, 2015: The JRA-55 Reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, <https://doi.org/10.2151/jmsj.2015-001>.
- Künsch, H. R., 1989: The jackknife and the bootstrap for general stationary observations. *Ann. Stat.*, **17**, 1217–1241, <https://doi.org/10.1214/aos/1176347265>.
- Lindsay, R., and A. Schweiger, 2015: Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations. *Cryosphere*, **9**, 269–283, <https://doi.org/10.5194/tcd-8-4545-2014>.
- , M. Wensnahan, A. Schweiger, and J. Zhang, 2014: Evaluation of seven different atmospheric reanalysis products in the Arctic. *J. Climate*, **27**, 2588–2606, <https://doi.org/10.1175/JCLI-D-13-00014.1>.
- Liu, J., Z. Zhang, Y. Y. Hu, L. Chen, Y. Dai, and X. Ren, 2008: Assessment of surface air temperature over the Arctic Ocean in reanalysis and IPCC AR4 model simulations with IABP/POLES observations. *J. Geophys. Res.*, **113**, D10105, <https://doi.org/10.1029/2007JD009380>.
- Makshtas, A., D. Atkinson, M. Kulakov, S. Shutilin, R. Krishfield, and A. Proshutinsky, 2007: Atmospheric forcing validation for modeling the central Arctic. *Geophys. Res. Lett.*, **34**, L20706, <https://doi.org/10.1029/2007GL031378>.
- Marq, S., and J. Weiss, 2012: Influence of sea ice lead-width distribution on turbulent heat transfer between the ocean and the atmosphere. *Cryosphere*, **6**, 143–156, <https://doi.org/10.5194/tc-6-143-2012>.
- Maykut, G. A., 1978: Energy exchange over young sea ice in the central Arctic. *J. Geophys. Res. Oceans*, **83**, 3646–3658, <https://doi.org/10.1029/JC083IC07P03646>.
- Morrison, H., G. de Boer, G. Feingold, J. Harrington, M. D. Shupe, and K. Sulia, 2011: Resilience of persistent Arctic mixed-phase clouds. *Nat. Geosci.*, **5**, 11–17, <https://doi.org/10.1038/ngeo1332>.
- Mortin, J., G. Svensson, R. G. Graversen, M. L. Kapsch, J. C. Stroeve, and L. N. Boisvert, 2016: Melt onset over Arctic sea ice controlled by atmospheric moisture transport. *Geophys. Res. Lett.*, **43**, 6636–6642, <https://doi.org/10.1002/2016GL069330>.

- Naakka, T., T. Nygård, and T. Vihma, 2018: Arctic humidity inversions: Climatology and processes. *J. Climate*, **31**, 3765–3787, <https://doi.org/10.1175/JCLI-D-17-0497.1>.
- Overland, J. E., and M. Wang, 2016: Recent extreme Arctic temperatures are due to a split polar vortex. *J. Climate*, **29**, 5609–5616, <https://doi.org/10.1175/JCLI-D-16-0320.1>.
- Park, D.-S. R., S. Lee, and S. B. Feldstein, 2015: Attribution of the recent winter sea-ice decline over the Atlantic sector of the Arctic Ocean. *J. Climate*, **28**, 4027–4033, <https://doi.org/10.1175/JCLI-D-15-0042.1>.
- Patton, A., D. N. Politis, and H. White, 2009: Correction to “Automatic block-length selection for the dependent bootstrap” by D. Politis and H. White. *Econometric Rev.*, **28**, 372–375, <https://doi.org/10.1080/07474930802459016>.
- Pithan, F., B. Medeiros, and T. Mauritsen, 2014: Mixed-phase clouds cause climate model biases in Arctic wintertime temperature inversions. *Climate Dyn.*, **43**, 289–303, <https://doi.org/10.1007/s00382-013-1964-9>.
- , and Coauthors, 2016: Select strengths and biases of models in representing the Arctic winter boundary layer over sea ice: The Larcform1 single column model intercomparison. *J. Adv. Model. Earth Syst.*, **8**, 1345–1357, <https://doi.org/10.1002/2016MS000630>.
- Rinke, A., K. Dethloff, W. Dorn, D. Handorf, and J. C. Moore, 2013: Simulated Arctic atmospheric feedbacks associated with late summer sea ice anomalies. *J. Geophys. Res. Atmos.*, **118**, 7698–7714, <https://doi.org/10.1002/jgrd.50584>.
- , M. Maturilli, R. M. Graham, H. Matthes, D. Handorf, L. Cohen, S. R. Hudson, and J. C. Moore, 2017: Extreme cyclone events in the Arctic: Wintertime variability and trends. *Environ. Res. Lett.*, **12**, 094006, <https://doi.org/10.1088/1748-9326/aa7def>.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, <https://doi.org/10.1175/2010BAMS3001.1>.
- , and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, <https://doi.org/10.1175/JCLI-D-12-00823.1>.
- Schweiger, A., R. Lindsay, J. Zhang, M. Steele, H. Stern, and R. Kwok, 2011: Uncertainty in modeled Arctic sea ice volume. *J. Geophys. Res.*, **116**, C00D06, <https://doi.org/10.1029/2011JC007084>.
- Screen, J. A., and I. Simmonds, 2012: Declining summer snowfall in the Arctic: Causes, impacts and feedbacks. *Climate Dyn.*, **38**, 2243–2256, <https://doi.org/10.1007/s00382-011-1105-2>.
- , —, C. Deser, and R. Tomas, 2013: The atmospheric response to three decades of observed Arctic sea ice loss. *J. Climate*, **26**, 1230–1248, <https://doi.org/10.1175/JCLI-D-12-00063.1>.
- Serreze, M. C., and J. A. Francis, 2006: The Arctic amplification debate. *Climatic Change*, **76**, 241–264, <https://doi.org/10.1007/s10584-005-9017-y>.
- , A. P. Barrett, and J. Stroeve, 2012: Recent changes in tropospheric water vapor over the Arctic as assessed from radiosondes and atmospheric reanalyses. *J. Geophys. Res.*, **117**, D10104, <https://doi.org/10.1029/2011JD017421>.
- Simmons, A. J., and P. Poli, 2014: Arctic warming in ERA-Interim and other analyses. *Quart. J. Roy. Meteor. Soc.*, **141**, 1147–1162, <https://doi.org/10.1002/qj.2422>.
- Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032, <https://doi.org/10.1175/MWR2830.1>.
- Solomon, A., M. D. Shupe, O. Persson, H. Morrison, T. Yamaguchi, P. M. Caldwell, and G. de Boer, 2014: The sensitivity of springtime Arctic mixed-phase stratocumulus clouds to surface-layer and cloud-top inversion-layer moisture sources. *J. Atmos. Sci.*, **71**, 574–595, <https://doi.org/10.1175/JAS-D-13-0179.1>.
- Sotiropoulou, G., J. Sedlar, R. Forbes, and M. Tjernström, 2015: Summer Arctic clouds in the ECMWF forecast model: An evaluation of cloud parametrization schemes. *Quart. J. Roy. Meteor. Soc.*, **142**, 387–400, <https://doi.org/10.1002/qj.2658>.
- Stroeve, J., and D. Notz, 2018: Changing state of Arctic sea ice across all seasons. *Environ. Res. Lett.*, **13**, 103001, <https://doi.org/10.1088/1748-9326/aade56>.
- , M. C. Serreze, M. M. Holland, J. E. Kay, J. Malanik, and A. P. Barrett, 2012: The Arctic’s rapidly shrinking sea ice cover: A research synthesis. *Climatic Change*, **110**, 1005–1027, <https://doi.org/10.1007/s10584-011-0101-1>.
- Tastula, E. M., T. Vihma, E. L. Andreas, and B. Galperin, 2013: Validation of the diurnal cycles in atmospheric reanalyses over Antarctic sea ice. *J. Geophys. Res.*, **118**, 4194–4204, <https://doi.org/10.1002/jgrd.50336>.
- Taylor, P., B. Hegyi, R. Boeke, and L. Boisvert, 2018: On the increasing importance of air–sea exchanges in a thawing Arctic: A review. *Atmosphere*, **9**, 41, <https://doi.org/10.3390/atmos9020041>.
- Tjernström, M., and R. G. Gravensén, 2009: The vertical structure of the lower Arctic troposphere analysed from observations and the ERA-40 reanalysis. *Quart. J. Roy. Meteor. Soc.*, **135**, 431–443, <https://doi.org/10.1002/qj.380>.
- Walden, V. P., S. R. Hudson, L. Cohen, S. Y. Murphy, and M. A. Granskog, 2017a: Atmospheric components of the surface energy budget over young sea ice: Results from the N-ICE2015 campaign. *J. Geophys. Res. Atmos.*, **122**, 8427–8446, <https://doi.org/10.1002/2016JD026091>.
- , S. Murphy, S. R. Hudson, and L. Cohen, 2017b: N-ICE2015 atmospheric turbulent fluxes [data set]. Norwegian Polar Institute, accessed 2 August 2018, <https://doi.org/10.21334/npolar.2017.298013b7>.
- Walsh, J. E., and W. L. Chapman, 1998: Arctic cloud–radiation–temperature associations in observational data and atmospheric reanalyses. *J. Climate*, **11**, 3030–3045, [https://doi.org/10.1175/1520-0442\(1998\)011<3030:ACRTAI>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<3030:ACRTAI>2.0.CO;2).
- , —, and D. H. Portis, 2009: Arctic cloud fraction and radiative fluxes in atmospheric reanalyses. *J. Climate*, **22**, 2316–2334, <https://doi.org/10.1175/2008JCLI2213.1>.
- Wesslén, C., M. Tjernström, D. H. Bromwich, G. de Boer, A. M. L. Ekman, L. S. Bai, and S. H. Wang, 2014: The Arctic summer atmosphere: An evaluation of reanalyses using ASCOS data. *Atmos. Chem. Phys.*, **14**, 2605–2624, <https://doi.org/10.5194/acp-14-2605-2014>.
- Wyser, K., and Coauthors, 2008: An evaluation of Arctic cloud and radiation processes during the SHEBA year: Simulation results from eight Arctic regional climate models. *Climate Dyn.*, **30**, 203–223, <https://doi.org/10.1007/s00382-007-0286-1>.