Assessment of Atmospheric Reanalyses with Independent Observations in the Weddell Sea, Antarctica

M. O. Jonassen[1,2], I. Välisuo[3], T. Vihma[3] , P. Uotila[4] , A. P. Makshtas[5], and J. Launiainen[4]

[1]Department of Arctic Geophysics, The University Centre in Svalbard, Longyearbyen, Norway.

[2]Geophysical Institute, The University of Bergen, Bergen, Norway.

[3]Finnish Meteorological Institute, Helsinki, Finland.

[4]Institute for Atmospheric and Earth System Research, University of Helsinki, Helsinki, Finland.

[5]Arctic and Antarctic Research Institute, St. Petersburg, Russia.

Corresponding author: Marius O. Jonassen (marius.jonassen@unis.no)

Key Points:

- All four reanalyses assessed are generally too warm in the boundary layer.
- These and other biases are consistent with results from previous studies using more recent observations.
- As the amount of observations has varied in time and space, the consistency of the ABL warm bias calls for improving ABL parameterisations.

Abstract

Surface layer and upper-air in situ observations from two research vessel cruises and an ice station in the Weddell Sea from 1992 and 1996 are used to validate the four current atmospheric reanalysis products ERA-Interim, CFSR, JRA-55 and MERRA-2. Three of the observation datasets are independent, providing a rare opportunity to validate the reanalyses in the otherwise datasparse region of Antarctica against independent data. All four reanalyses reproduce 2 m temperatures that are warmer than the observations, with biases varying approximately between +2 K (CFSR) and +2.8 K (MERRA-2). Cloud fractions are also relatively poorly reproduced by the reanalyses, MERRA-2 and JRA-55 having the strongest positive and negative biases of about +30 % and -17 %. These biases contribute respectively to a negative (positive) downwelling shortwave radiation bias and a positive (negative) downwelling longwave radiation bias in these products. All four reanalyses are generally too warm in the atmospheric boundary layer (ABL), with biases up to +1.4 K (ERA-Interim). Skill scores of the error statistics reveal that ERA-Interim compares generally the most favorably against both the surface layer and the upper-air observations. CFSR compares the second best and JRA-55 and MERRA-2 have the

least favorable scores. The ABL warm bias is consistent with previous evaluation studies in high latitudes, where more recent observations have been applied. As the amount of observations has varied depending on the decade, season and region, the consistency of the warm bias suggests a need to improve ABL and surface parameterisations.

Plain Language Summary

Surface layer and upper-air in situ observations from two research vessel cruises and an ice station in the Weddell Sea from 1992 and 1996 are used to validate four atmospheric reanalyses products. Three of the observation datasets are independent, meaning that they have not been used in compiling the reanalyses. This provides a rare opportunity to validate the reanalyses in the otherwise datasparse region of Antarctica against independent data. The reanalyses differ in performance. However, all four reanalyses have large errors in the cloud cover and they also display too high temperatures in the lowermost part of the atmosphere. The latter finding is consistent with previous validation studies in polar regions, in which more recent observations have been applied. As the amount of observations has varied depending on the decade, season and region, the consistency of the warm bias suggests a need to improve the representation of physical processes in the lowest parts of the atmosphere in the reanalyses investigated.

1 Introduction

Reanalyses combine observations and a numerical prediction model providing four-dimensional gridded and dynamically coherent data that are used for a wide range of applications. Their usefulness is particularly high in the Arctic and Antarctic, where they can help fill in gaps and be used for attaining a more complete physical picture of climate change in a region where observational data are otherwise sparse and uneven in their distribution (Bromwich et al. 2013). Atmospheric reanalyses are also used for reconstructing near-surface temperature (Steig et al. 2009; Nicolas and Bromwich 2014), evaluating climate models (Rinke et al. 2006; Perez et al. 2014), and providing boundary conditions for land surface models, ice-ocean models and limited area atmospheric models (Assmann et al. 2013; Dutrieux et al. 2014; Lindsay et al. 2014).

Early reanalyses are known to have larger errors in the Arctic and Antarctic with respect to e.g. wind speed and direction, humidity, temperature, cloud cover and radiative fluxes (Screen and Simmonds 2011; Sturaro 2003; Bromwich et al. 2007; Walsh and Chapman 1998; Vancoppenolle et al. 2011). Due to these shortcomings, extensive work has been carried out in producing new reanalyses, incorporating among others more sophisticated assimilation methods, better representation of sea-ice and land-surface processes and better horizontal and vertical grid resolution. Examples of these current reanalyses are ERA-Interim (Dee et al. 2011) from the European Centre for Medium-Range Weather Forecasts (ECMWF), the Climate Forecast System Reanalysis (CFSR) (S. Saha et al. 2006) from the National Centers for Environmental Prediction (NCEP), the Japanese 55-year reanalysis (JRA-55) (Kobayashi et al. 2015) from the Japan Meteorological Agency (JMA), and MERRA version 2 (MERRA-2) (Gelaro et al. 2017) from the National Aeronautics and Space Administration (NASA). Even though substantial progress has been made in these products with respect to their predecessors (Tastula et al. 2013; Nygård et al. 2016), also these have their deficiencies. For example, spurious warming trends have been

identified in many parts of East Antarctica (Y. Wang et al. 2016) and near-surface cold biases have been found along the Antarctic coast (Bracegirdle and Marshall 2012; P. D. Jones and Lister 2014; R. W. Jones et al. 2016b).

The applications of reanalysis data, e.g. the provision of input data to models that are highly sensitive to the forcing conditions, make it important to validate reanalyses against available independent observational data which are often rare, particularly in the polar regions.

In this paper we validate the ERA-Interim, CFSR, JRA-55 and MERRA-2 reanalyses against a range of in situ observational data from the Weddell Sea. These data include vertical profiles and surface layer time series of temperature, humidity and wind speed as well as radiative and turbulent surface fluxes and cloud fraction. One main data source is the Ice Station Weddell (ISW), which was a U.S.-Russian campaign conducted from February to June in 1992, providing all of the above-mentioned types of data (Gordon and Ice Station Weddell Group of Principal Investigators and Chief Scientists 1993). The ISW is the hitherto longest lasting drift station of its kind and only two other ice stations have been performed in the Antarctic sea ice; Ice station Polarstern (ISPOL), which lasted for five weeks in December January 2004 and 2005 (Vihma et al. 2009; Bareiss and Görgen 2008) and the Sea Ice Mass Balance in the Antarctic (SIMBA), which took place over two weeks in September - October in 2007 (Vancoppenolle et al. 2011). The other data sources that we use in this study are radiosounding datasets originating from two summer and autumn-time research vessel cruises in the Weddell Sea, the first taking place in 1992 with *Akademik Fedorov* (hereafter '*Fedorov*') and the second taking place in 1996 with *Aranda*. All datasets except the *Fedorov* dataset were withheld from assimilation and are thus independent data. This study is made particularly relevant by the ongoing Year of Polar Prediction (YOPP) lasting from May 2017 to June 2019, which has a main focus on improving environmental prediction capabilities in polar regions (Jung et al. 2016; Goessling et al. 2016).

2 Datasets and Methodology

2.1 Observations

We use observational data from three different sources in this study. The first is the ISW, which was the first, and thus far the longest drifting sea ice station in the Antarctic, lasting from late February to late May 1992. Much of our knowledge of the atmospheric boundary layer over the Antarctic sea ice zone comes from this campaign (Edgar L. Andreas and Claffey 1995; Edgar L. Andreas 1995; Edgar L. Andreas, Claffy, and Makshtas 2000; Edgar L. Andreas 2002; Edgar L. Andreas, Jordan, and Makshtas 2004, 2005; Tastula et al. 2013; Tastula, Vihma, and Andreas 2012).

From ISW, we use cloud fraction, surface skin temperature and near-surface values of temperature, humidity and wind speed data as well as radiative and turbulent surface heat fluxes. Near-surface temperature and humidity were observed at heights of 0.1 m and 5 m above the ground level (AGL), while wind speed was observed at 5 m AGL. The turbulent sensible and latent heat fluxes were obtained using a sonic anemometer/thermometer and a hygrometer, both mounted at 4.65 m AGL above the surface. All these data are available as hourly averages from

the ISW dataset and we extracted these data for the time period between 25 February 18:00 and 29 May 1992 18:00. More details about these observations and the post-processing can be found in (Edgar L. Andreas, Jordan, and Makshtas 2004) and (Edgar L. Andreas, Jordan, and Makshtas 2005).

In addition, 40 airsonde soundings and 128 tethersondes soundings are available from the ISW (Claffey, Edgar, and Makshtas 1994), the first of which we found to provide reliable data only of temperature and the latter of which provided reliable temperature, humidity and wind data. For the tethersonde data, the maximum height reached by each sounding varied substantially (from 92 to 1350 m AGL) and the average height was 613 m AGL. In this study we only use soundings that reached at least 600 m AGL. Some time periods during ISW had more intensive sounding activity than others. In order not to weigh these periods excessively in the error statistics calculations, we selected profiles with a minimum temporal spacing of 12 hours for comparison against the reanalyses. In total we utilize 40 and 56 individual soundings from the ISW airsonde and tethersonde datasets, respectively.

Two other upper-air datasets used in this study consist of radiosonde soundings launched from two different research vessel cruises. The first is from the research vessel *Fedorov* in February 1992, which was used to supply the ISW. The second is from a cruise with the Finnish research vessel *Aranda*, which took place from the end of January to mid February, 1996 in the Weddell Sea (Vihma et al. 1997). A total of 77 soundings are available from *Aranda* and 40 from *Fedorov*. Of these, we excluded 17 from further analysis from *Aranda* and 11 from *Fedorov* due to poor data quality. Furthermore, like with the ISW soundings, we selected soundings with a minimum time difference of 12 hours, leading to 24 and 34 available soundings from *Aranda* and *Fedorov,* respectively. A map showing the locations from which the datasets used in this study were obtained can be seen in Figure 1.

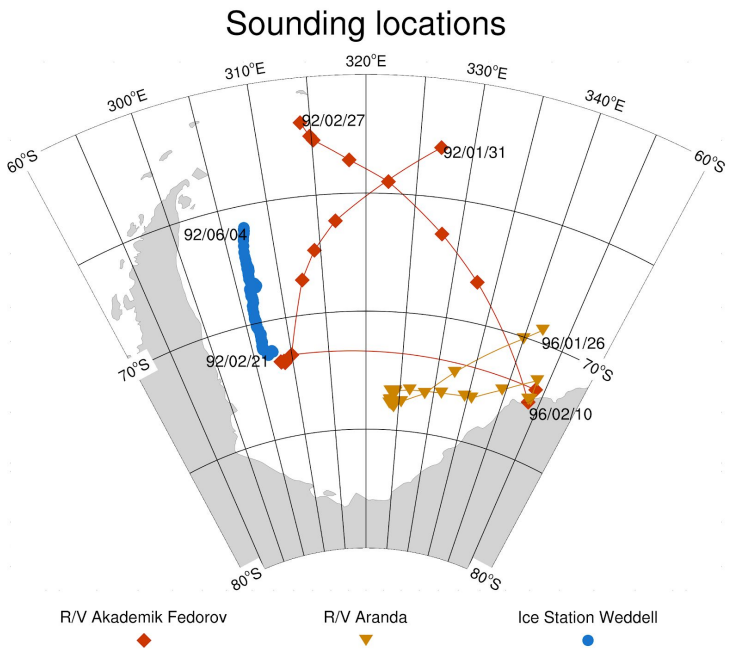

Figure 1. Tracks of *R/V Akademik Fedorov* (red rectangles and red, solid line) and Ice Station Weddell (blue circles and blue, solid line) from 1992 and of *R/V Aranda* (yellow triangles and yellow, solid line) 1996.

Amongst all observational datasets, the *Fedorov* dataset is the only that was sent to the Global Telecommunication System (GTS) and thereby the only dataset made available for assimilation for the reanalysis products considered herein. It is not clear, however, to what degree these observations were actually used for assimilation in the different products and or how much they were weighted in any assimilation. As part of our data post-processing, we removed unphysical values from the observations before we used them for validation.

2.2 Reanalyses

A wide range of observations have been assimilated in the reanalyses validated in this study, either using 3-dimensional variational data assimilation (3D-Var) (CFSR and MERRA-2) or 4D-Var (ERA-Interim and JRA-55). The reanalyses are available at horizontal resolutions in the range of 0.5-0.75° in a regular latitude-longitude grid (ERA-Interim, CFSR and MERRA-2) and 1.25° (JRA-55). We note that JRA-55 is run at a higher native resolution (about 0.5°), but the pressure level data (about 17 levels for all four products) that we use in this study are only available at this coarser resolution. The considered reanalysis variables are from the analysis fields. An exception is the near-surface temperature, humidity and wind in CFSR, which are only available as forecast fields. Another exception is the radiative and turbulent surface heat fluxes, which we obtained and or calculated as 6-hourly averages accumulated over the model forecasts.

2.3 Methodology

For comparison of the near-surface data from ISW, we extracted 6-hourly data from the hourly observations in order to match the time resolution of the reanalysis output. In the case of the temperature, humidity, wind speed and cloud fraction we did this by extracting every 6th datapoint from the respective time series. For the radiative and turbulent surface heat fluxes, however, we averaged the hourly observations into 6-hourly periods. Within each of these 6-hourly periods, some had less than 4 data points (more than 2 missing values) and we flagged these periods as missing. In the resulting observational 6-hourly datasets, about half of the data are missing in the case of the turbulent surface heat fluxes. For the surface radiative data, however, there are only 2% missing and for the cloud fraction data about 5% is missing. For the temperature, humidity and wind data about 15% are missing. We ignored all these data gaps in our error statistics calculations.

The observed cloud fraction was reported on a scale from 0 to 10 with intervals of 1, which we converted to % with intervals of 10 % for comparison with the reanalyses, whose cloud fractions we also rounded off to the nearest 10 %. We calculated 2 m values for the observed temperature and humidity and 10 m values for the wind speed using an iterative algorithm provided by (Launiainen and Vihma 1990). Surface pressure, which is essential for calculating the 2 m values, is not available in the ISW dataset and we therefore estimated it using the ensemble average of surface pressure in all four reanalysis products included in this study.

For comparison of the sounding data, we first linearly interpolated both the reanalysis and observation profile data to a common, vertical grid with 100 m intervals between 200 and 4000 m AGL. Thereafter, we linearly interpolated the reanalyses horizontally and in time to each timestamp and location of the observations. Naturally, in some cases the lowest pressure level (1000 hPa), and even the second lowest (975 hPa), in each reanalysis profile considered are located below the surface. To minimize the influence of this problem on the error statistics, we only focus on data down to 200 m AGL.

The error statistics that we apply in this study are bias, root mean square error (RMSE) and the correlation coefficient (r). In addition, for the surface layer data and cloud fraction data, we also consider slopes of linear regression lines and ratios of standard deviations (standard deviation of the reanalyses divided by that observed). We estimated the statistical significance at the 95 % level for the bias and r using the student-t test. The values of r are largely significant throughout the results, and we will explicitly state in the text if the values are not significant.

3 Results

3.1 Comparison of Surface Layer Data

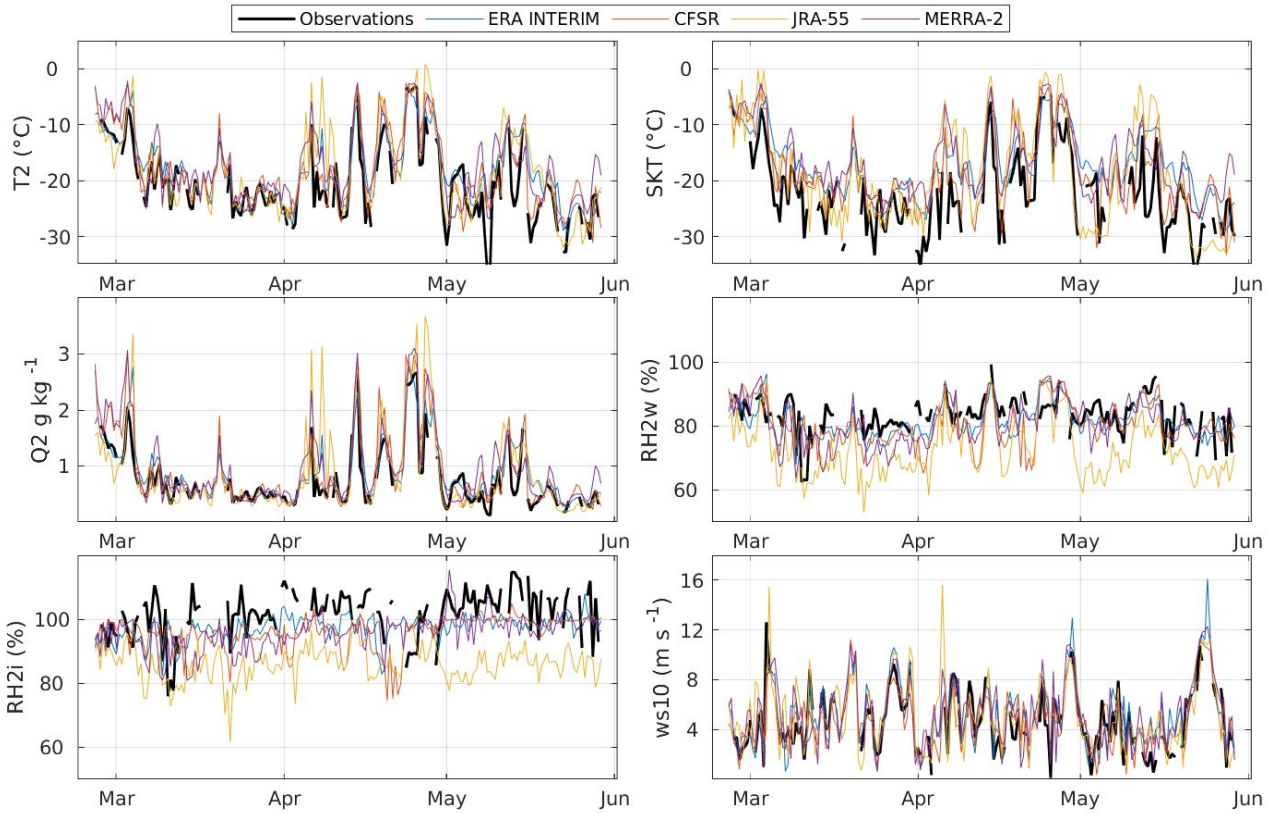

Figure 2
Time series of ISW surface layer observations and reanalysis data from ERA-Interim, CFSR, JRA-55 and MERRA-2 for the period from 25 February to 29 May, 1992. The variables presented are the 2 m temperature (T2), skin temperature (SKT), 2 m specific humidity (Q2), 2 m relative humidity with respect to water (RH2w), 2 m relative humidity with respect to ice (RH2i), and 10 m wind speed (U10). The time series are reduced from the original 6-hourly resolution to a 12-hourly resolution in this figure for improved clarity.

Figure 2 presents time series of the observed and modelled surface layer data of temperature (skin and at 2 m), specific humidity at 2 m, relative humidity with respect to water and ice at 2 m and wind speed at 10 m from the ISW period. We see from the observations that most of March was dominated by relatively cold and dry conditions, followed by a more variable, but generally warmer and moister weather type in April. Thereafter, in May, the weather turned cooler and drier again, still however, featuring a rather strong temporal variability, especially in temperature. The wind speed follows no such clear temporal pattern as the temperature and

humidity. The reanalyses generally capture the described, lower frequency variability in the variables investigated fairly well. However, on the shorter time-scales, they do feature obvious biases. We see a clear tendency for the reanalyses to overestimate the surface and near-surface temperatures and they also have a corresponding moist bias with regards to specific humidity and an underestimation of relative humidity. The latter is particularly evident with respect to ice, especially in the case of JRA-55.

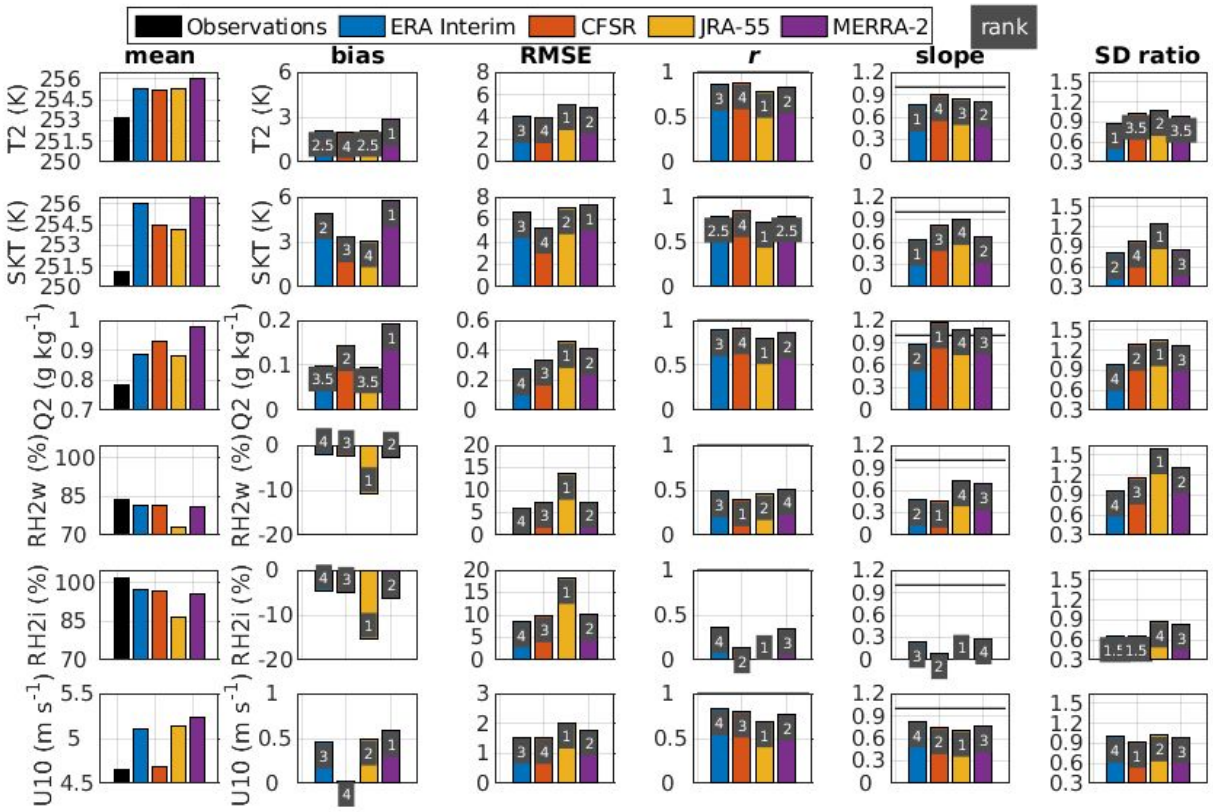

Figure 3. Mean bias, root mean square error (RMSE), correlation coefficient (*r*), slope of the linear regression line (slope) and the ratio of the standard deviation of the reanalyses divided by that observed (SD ratio) for the four reanalyses using Ice Station Weddell observations as reference. A positive bias indicates that the reanalysis product has a higher value than the observations. The statistics are presented for the 2 m temperature (T2), skin temperature (SKT), 2 m specific humidity (Q2), 2 m relative humidity with respect to water (RH2w), 2 m relative humidity with respect to ice (RH2i), and 10 m wind speed (U10). In the ranking indicated on each bar the best reanalysis is given 4 points and the worst 1 point. Mean values of each observed and modelled variable are given in the leftmost column of panels.

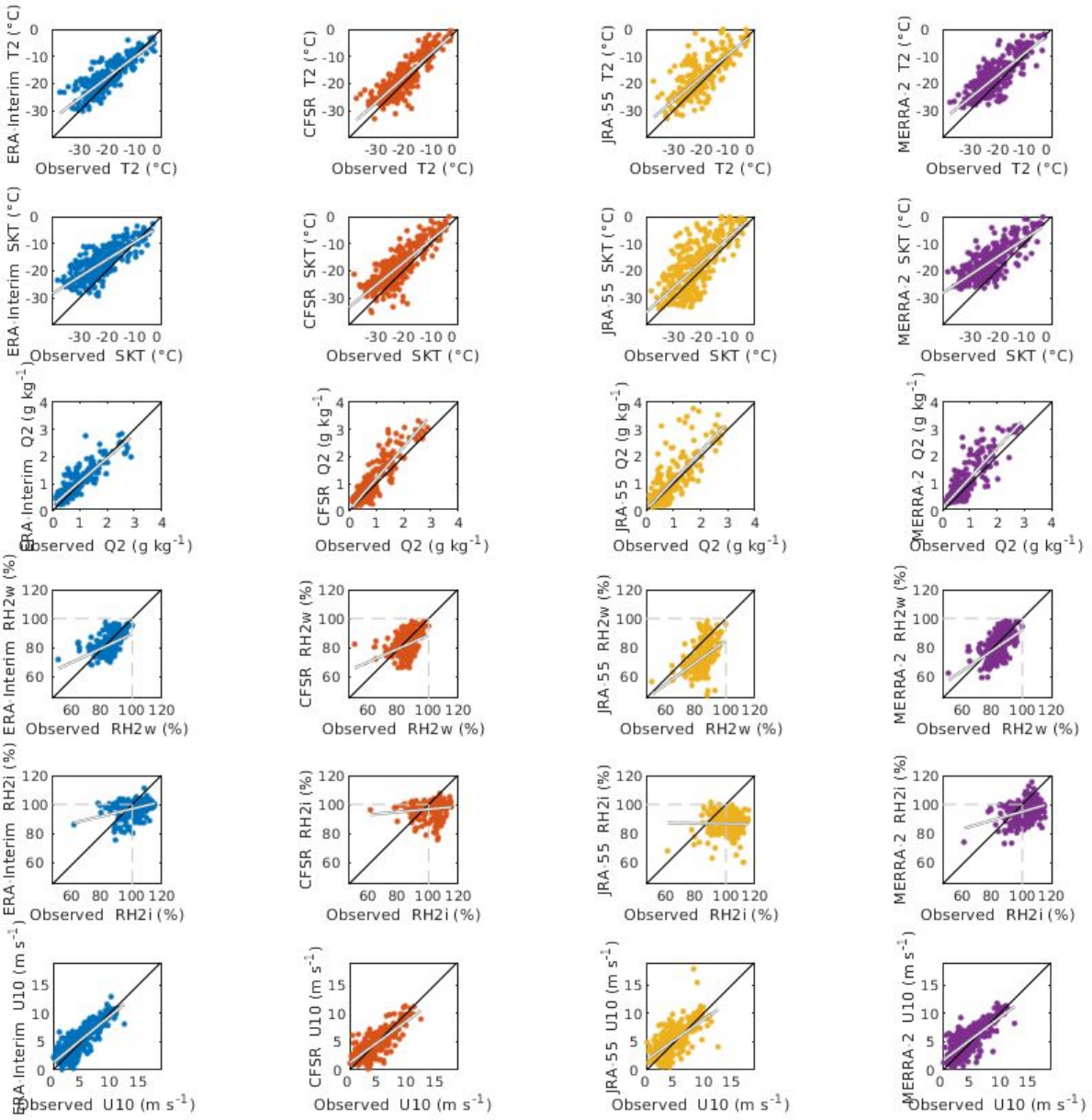

Figure 4. Scatterplots showing the observed 2 m temperature (T2), surface skin temperature (SKT), 2 m specific humidity (Q2), relative humidity with respect to water (RH2w), relative humidity with respect to ice (RH2i), and 10 m wind speed (U10) against the same variables in the four reanalysis products.

A detailed comparison of the surface layer data in terms of error statistics is given in Figure 3 and scatterplots between the different observed and modelled variables are shown in Figure 4. Considering the 2 m temperature, the mean observed value is 253.2 K and the reanalyses are all biased warm compared to this (Figure 3). MERRA-2 has the highest bias (+2.81 K) and CFSR

the lowest (+1.99 K). Furthermore, we see that the largest warm biases are found for the lowest observed temperatures in all four reanalyses (Figure 4). Correspondingly, the slope of all four linear regression lines is below 1. CFSR has the least scatter among the products and thus has the most favorable RMSE and *r* (3.99 K and 0.87), while JRA-55 has the most scatter and the least favorable RMSE and *r* (5.15 K and 0.78).

The mean observed skin temperature is 251.1 K and all four reanalyses feature warm biases, which are larger than for the 2 m temperature. JRA-55 has the lowest bias (3.03 K), whereas MERRA-2 has the highest (5.73 K). As for the 2 m temperature warm biases, the strongest warm biases in skin temperature are found for the lowest observed temperature (Figure 3) and the linear regression line slopes are all below 1. CFSR has the lowest RMSE (5.25 K) and highest *r* (0.84), while MERRA-2 has the highest RMSE (7.28 K) and JRA-55 the lowest *r* (0.78).

Considering the 2 m specific humidity, the mean observed value is 0.79 g kg^{-1}. All four reanalyses are significantly moister than this, corresponding to the warm biases found in the 2 m temperature. The highest moist bias is found in MERRA-2 at +0.19 g kg^{-1} (Figure 2). A large portion of the humidity values are clustered below 1 g kg^{-1} in all four products and the slopes of the regression lines are all close to one (Figure 3). Thus, the model moisture biases are not strongly affected by the observed humidity, though CFSR, JRA-55 and MERRA-2 have a slight tendency towards more positive humidity biases for higher observed humidity values. In terms of RMSE, ERA-Interim has the lowest value of 0.27 g kg^{-1}, and considering *r,* CFSR has the highest value of 0.91, both having relatively small scatter. JRA-55 has the worst values for both RMSE and *r* (0.45 g kg^{-1} and 0.8), and these are reflected in relatively large scatter. Also, JRA-55 features a larger range of humidity than what is observed, which leads to a larger standard deviation than observed (Figure 2).

Considering the 2 m relative humidity with respect to water, the mean observed value is close to 85 % and all four reanalyses have rather small negative biases within -2 to -3 %. An exception is JRA-55, where it is almost -11 %. The mean observed 2m relative humidity with respect to ice indicates super saturation (105 %), while the reanalyses all have values below 100 %. For JRA-55 the bias is beyond -15 %. Similar conclusions as for the biases can be drawn for the values of RMSE, with worse values with respect to ice than for water and they are particularly bad with respect to ice for JRA-55 with values in the vicinity of 20 %. The correlation coefficients, *r*, are around 0.5 for all reanalyses with respect to water. With respect to ice, however, they are considerably worse and vary between 0 (JRA-55, not significant) to around 0.3 (ERA-Interim and CFSR, significant). The scatter plots of observed and modelled 2m relative humidity (Figure 3) reveal why the values of *r* are this poor. We can clearly see that the reanalyses struggle in reproducing the highest observed values of relative humidity, which with respect to water implies slopes of the linear regression lines that are between 0.4 (CFSR) and 0.7 (JRA-55). With respect to ice, this slope is even down to 0 in JRA-55. A closer look at all data points of relative humidity with respect to ice reveals that none of these are above 100 % for JRA-55, while the other products do reproduce some values above 100 %.

For the 10 m wind speed, all four reanalyses feature positive biases and the highest is found in MERRA-2 (+0.61 m s^{-1}) (Figure 2). All these are statistically significant, except for CFSR,

which also has the lowest bias (+0.05 m s^{-1}). These positive biases are dominated by a large portion of relatively low observed wind speeds, and for higher observed wind speeds, all four products on average underestimate the wind speed (Figure 3). This is reflected by the slopes of the linear regression lines being below 1 for all four reanalyses. CFSR has the best RMSE (1.51 m s^{-1}) and ERA-Interim has the best *r* (0.83), and both products have relatively little scatter. The worst RMSE and *r* are found in JRA-55 (2.02 m s^{-1} and 0.69).

3.1.1 Summary of Comparison of Surface Layer Data and Reanalysis Performance with Respect to Relative Humidity

To summarize and make an overall assessment of the reanalyses' near-surface performance, we applied a ranking system to the error statistics of temperature, specific humidity and wind speed (labels on bars in Figure 2). The ranking gives 4 points to the best product and 1 point to the worst product for each of the error metrics and atmospheric parameters considered. In case of a tied ranking between two products, e.g. for the first place, both products receive 3.5 points. When all points are summed up for each reanalysis, ERA-Interim has the highest ranking (89 points), closely followed by CFSR (85 points). MERRA-2 and JRA-55 receive the fewest points (respectively 69 and 56 points).

Our comparison of the near-surface data reveals that all four renalyses struggle in reproducing the 2 m relative humidity, in particular with respect to ice (Figures 2, 3 and 4). Based on among other the near-surface observations from ISW, (E. L. Andreas et al. 2002) concluded that water vapor over polar sea ice is nearly always near saturation and often at supersaturation with respect to ice for temperatures between 0 and -25 ºC. For temperatures below -25 ºC, their results were less robust, due in part to the impaired reliability of the humidity sensors applied for such low temperatures, and this is also a limitation in the ISW observations. We investigate here in further detail how the 2 m relative humidity with respect to ice is reproduced by the reanalyses for temperatures in the range of -35 and 0 ºC. The results (Figure 5) reveal that all reanalyses do reproduce conditions close to saturation for the entire temperature range, all having relative humidity values of about 95 % or higher. JRA-55 is an exception to this, which consistently has values below 95 % for any temperature, and for temperatures between -20 and -30 ºC, even below 85 %. Neither of the products reproduces the observed onset of supersaturation at temperatures below -15 ºC, but for temperatures between -35 and -30 ºC, CFSR and ERA-Interim do show supersaturation. Though, as we comment on above, and as stated by (E. L. Andreas et al. 2002) (E. L. Andreas et al. 2002)the observational evidence for supersation based on the ISW for temperatures below -25 ºC is questionable dataset due to instrument limitations.

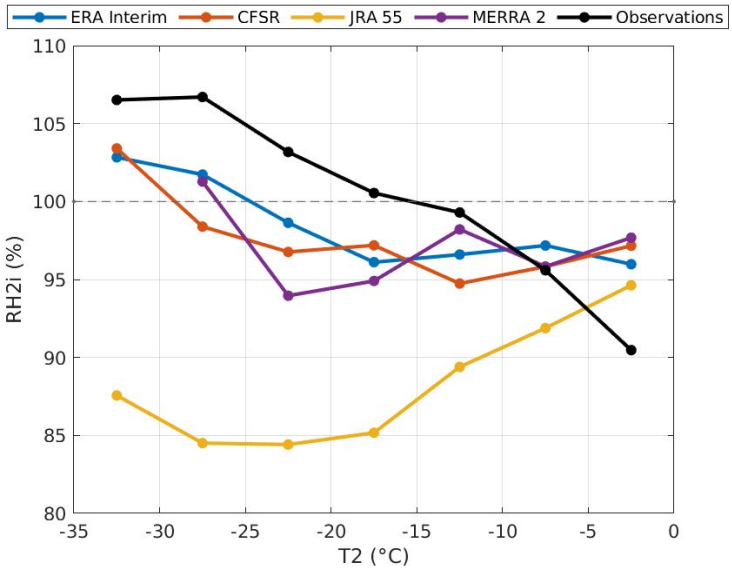

Figure 5. The relative humidity at 2 m with respect to ice (RH2i) from the Ice Stations Weddell observations and the four reanalyses averaged in bins of 5 ºC based on the observed and modelled 2 m temperature (T2).

3.2 Comparison with Radiative and Turbulent Surface Heat Fluxes and Cloud Fraction

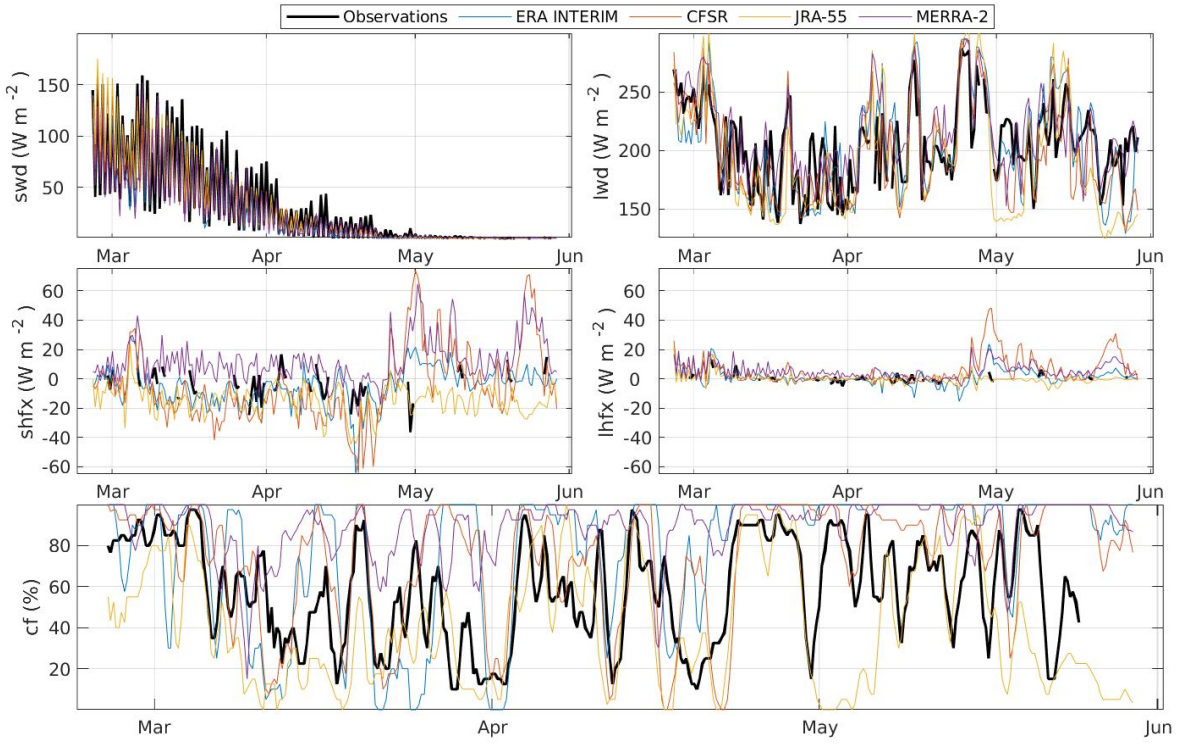

Figure 6

Time series of radiative and turbulent surface heat fluxes and cloud fraction from observations and reanalysis data from ERA-Interim, CFSR, JRA-55 and MERRA-2 for the ISW period from 25 February to 29 May, 1992. The variables presented are the downwelling short- and longwave radiation (swd and lwd), turbulent sensible and latent heat flux (shfx and lhfx) and the cloud fraction (cf). Except for the cloud fraction, the time series are reduced from the original 6-hourly resolution to a 12-hourly resolution in this figure for improved clarity. For the cloud fraction, a 2-day running average is applied for the same reason.

Figure 6 presents the radiative (downwelling short-and longwave) and turbulent surface heat fluxes (sensible and latent) and cloud fraction for the ISW period. As could be expected, the shortwave radiation diminishes rapidly through the time period of ISW and only in March and partly in April it features a clear diurnal signal. As could be expected, the downwelling longwave radiation follows a similar temporal development to the 2 m temperature and specific humidity data; it is fairly low in March, then it picks up again April, however, with quite some temporal variability. May has again lower, but still variable values. It is hard to spot a clear temporal signal in the observed sensible heat flux and the same goes for the latent heat flux. From the

beginning of May and onwards, the sensible and latent heat flux observations are unfortunately very sparse.

The reanalyses capture the downwelling shortwave radiation fairly well, and also the downwelling longwave radiation has no striking bias. When it comes to the sensible heat flux, there are clearer biases in the reanalyses. In March and April, MERRA-2 is generally biased high and the other datasets are biased somewhat low. For the periods around the beginning and end of May, CFSR and MERRA-2 stick out by having relatively high values of up to +80 W m^{-2}. May coincides with the time when the shortwave radiation is barely present anymore and the other heat flux components can be expected to dominate the surface energy balance. For the same times, ERA-Interim keeps between 0 and +20 W m^{-2} and JRA-55 is mostly on the negative side. Unfortunately, the observations are so sparse in these time periods that we are not able to make a robust judgement of these model estimates. The same general picture can be drawn for the latent heat flux, which is fairly stable between the beginning of the ISW period and until the end of April. After this time point, CFSR, and partly also MERRA-2, have relatively high fluxes of up to +70 W m^{-2} in the beginning and end of May. ERA-Interim has lower values and JRA-55 has values around 0 W m^{-2}. The cloud cover has no clear temporal trend and it is rather poorly represented in the reanalyses, with the largest biases found in MERRA-2, which generally overestimates it and in JRA-55, which generally underestimates it.

A detailed comparison of the surface layer data in terms of error statistics is given in Figure 7 and scatterplots between the different observed and modelled variables are shown in Figure 8.

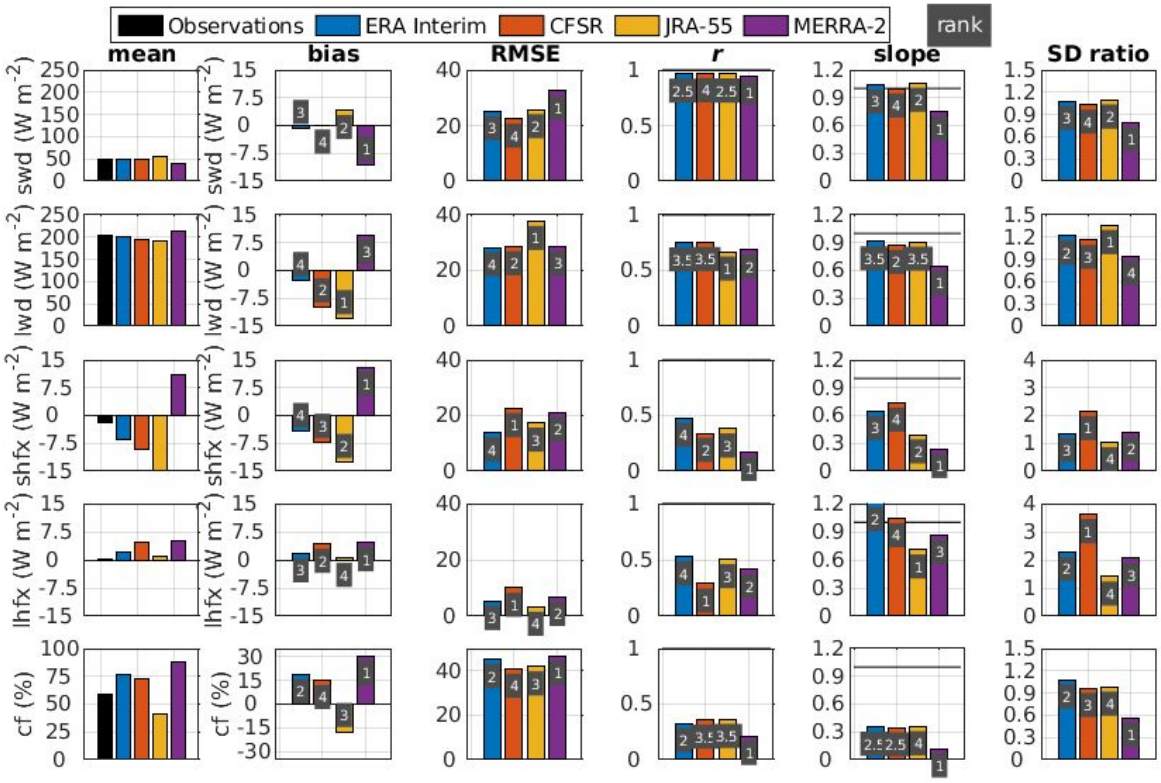

Figure 7. Mean bias, root mean square error (RMSE), correlation coefficient (*r*), slope of the linear regression line (slope) and the ratio of the standard deviation of the reanalyses divided by that observed (SD ratio) for the four reanalyses using Ice Station Weddell observations as reference. A positive bias indicates that the reanalysis product has a higher value than the observations. The statistics are presented for the downwelling shortwave radiation (swd), downwelling longwave radiation (lwd), turbulent sensible heat flux (shfx), turbulent latent heat flux (lhfx) and cloud fraction (cf). In the ranking indicated on each bar the best reanalysis is given 4 points and the worst 1 point. Mean values of each observed and modelled variable are given in the leftmost column of panels.

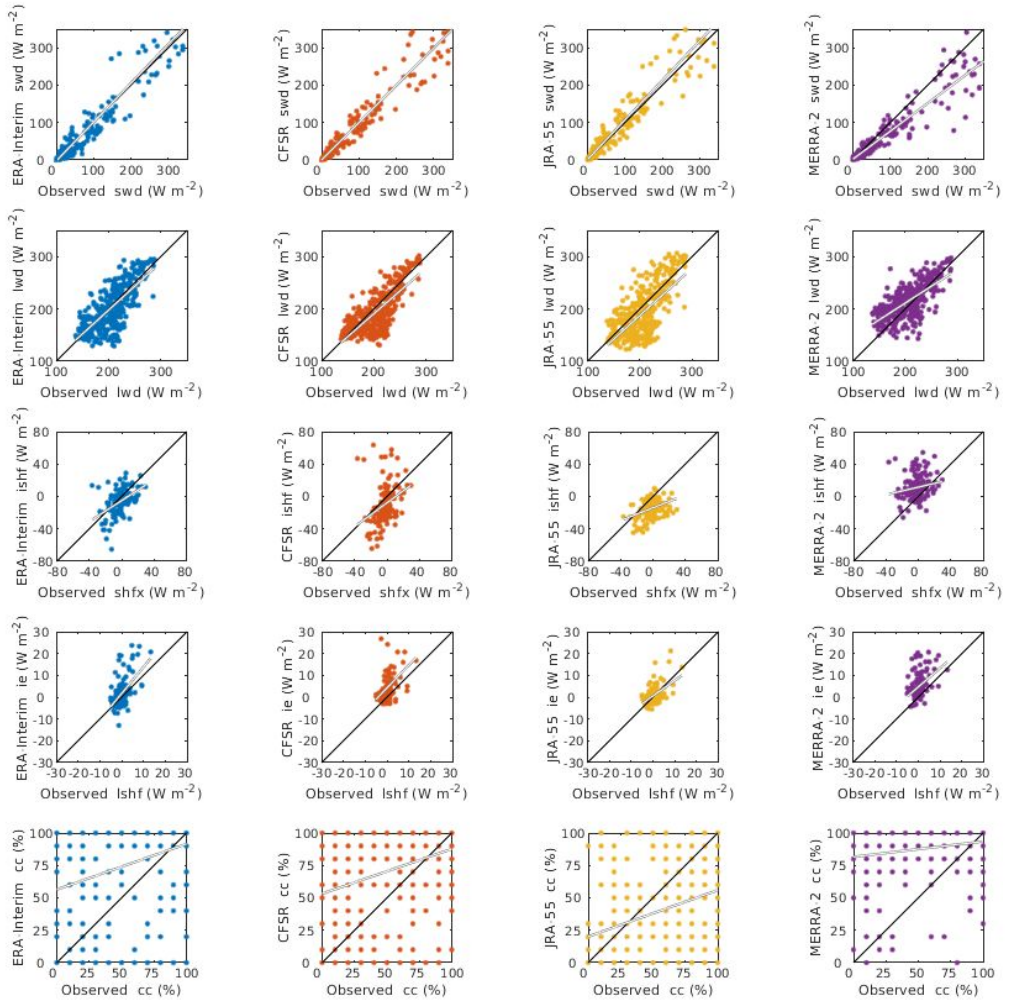

Figure 8. Scatterplots showing the observed downwelling shortwave radiation (swd), downwelling longwave radiation (lwd), turbulent sensible heat flux (shfx), turbulent latent heat flux (lhfx) and cloud fraction (cf) against the same variables in the four reanalysis products.

Considering the downwelling shortwave radiation, the mean observed value is 50.38 W m^{-2} and the reanalyses come rather close to this, and only MERRA-2 and JRA-55 feature significant biases of -10.71 W m^{-2} and +4.10 W m^{-2}, respectively (Figure 7). All four reanalyses have linear regression lines with slopes near 1, except MERRA-2, where it is 0.8. The RMSE is the most favourable in CFSR 22.36 W m^{-2} and the worst in MERRA-2 32.45 W m^{-2}. The values of *r* are very high across all four reanalyses, with values between 0.95 (MERRA-2) and 0.97 CFSR).

Turning to the downwelling longwave radiation, the mean observed value is 204.54 m^{-2}. The largest bias is found in JRA-55 (-13.03 W m^{-2}) and the smallest in ERA-Interim (-2.74 W m^{-2}, not significant). MERRA-2 is the only product with a positive bias (+9.56 W m^{-2}). The RMSE lies approximately within 20 and 40 W m^{-2} for all four products, which is similar to the values for

the downwelling shortwave radiation, JRA-55 having the least favourable value of 37.72 W m^{-2} and the three other products all having values around 28 W m^{-2}. The values of *r,* on the other hand, are lower for the downwelling longwave radiation than for the downwelling shortwave radiation and they are the worst for JRA-55 and MERRA-2 (about 0.68) and best for ERA-Interim and CFSR (0.75). As indicated by the slopes of the linear regression lines all being below 1, all four reanalyses have a tendency to underestimate higher observed values of the lwd (Figure 8). This tendency is the strongest in MERRA-2 with a slope of only about 0.6.

Regarding the turbulent surface flux of sensible heat, the mean observed value is -2.02 W m^{-2}. The reanalyses also have negative mean values, however they are larger, ERA-Interim having the largest bias of -4.36 W m^{-2}. MERRA-2 is an exception, with a positive bias of +13.07 W m^{-2}. The values of RMSE are approximately between 10 and 25 W m^{-2}. The slopes of the linear regression lines are all markedly below 1, with the lowest value of only 0.2 in MERRA-2 and the highest in CFSR with 0.7. Correspondingly, the correlation coefficients, *r*, are rather poor with values below 0.5 for all products and MERRA-2 having the worst (0.17), indicating a poor match between the data pairs in the linear sense. This poor match is also evident from looking at the scatterplots in Figure 8.

The mean observed turbulent surface flux of latent heat is only 0.27 W m^{-2}. The latent heat flux is on average slightly positively biased in all four reanalyses, with values between +0.5 W m^{-2} (JRA-55) and +4.2 W m^{-2} (CFSR). All these biases are significant, except for the bias in JRA-55. The RMSE values are also fairly low, between 3.30 W m^{-2} in JRA-55 and 10.03 W m^{-2} in CFSR. *r* is rather poor in all reanalyses, with values between 0.29 (CFSR) and 0.53 (ERA-Interim). The slopes of the linear regression lines are both above and below 1 in the respective products, with the strongest deviations from 1 found in ERA-Interim (1.2) and JRA-55 (0.7). All products feature a stronger variability in the latent heat flux than observed. This is seen in the standard deviation ratios, that are all positive, ranging between 1.4 in JRA-55 all the way up to 3.6 in CFSR.

Considering the cloud fraction, we see that the mean observed value is 58.4 %. ERA-Interim, CFSR and MERRA-2 all have positive biases, with the latter having the largest of +30.2 %. JRA-55, on the other hand, has a negative bias of -17.3 %. The RMSE is fairly large in all products, with values between 40.8 % (CFSR) and 46.6 % (MERRA-2). The mean *r* for all reanalyses (0.32) is the lowest for all variables considered and for MERRA-2 it is particularly low (0.21). The relatively high RMSE and *r* values are reflected in the scatter plots (Figure 8), which show a wide spread between the observed and simulated cloud fraction.

3.2.1 Summary of Comparison of Radiative and Turbulent Surface Heat Flux and Cloud fraction

We here summarize the reanalyses' performance with respect to the radiative and turbulent surface heat fluxes and cloud fraction applying the same ranking system as for the near-surface temperature, humidity and wind (see Section 3.1.1). When all points are summed up for each reanalysis, ERA-Interim has the highest ranking (74.5 points), followed by CFSR (68.5 points), JRA-55 (66 points) and lastly MERRA-2 (41 points).

3.2.2 the Prominent Cloud Fraction Biases

We find some of the most prominent biases in the reanalyses in their cloud fractions. Here, we investigate in further detail the impacts of these biases on the biases in downwelling long- and shortwave radiation. Since the solar radiation is almost gone by the 1 May, we set this as the last time point for the analysis. Before going into the biases themselves, we first establish the physical relationships between the cloud fraction and the other parameters mentioned above by calculating the correlation coefficients between them. The results can be seen in Figure 9.

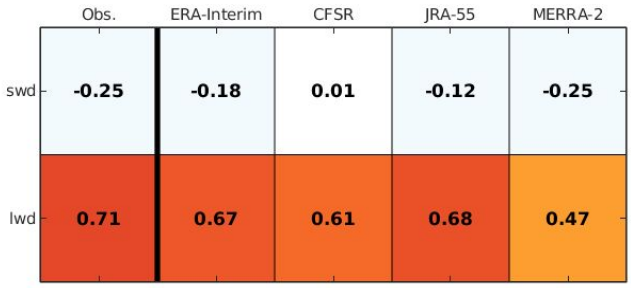

Figure 9
Correlation coefficients between the cloud fraction and the downwelling short- and longwave radiation (swd and lwd) from the observations (obs.) and the four reanalyses for the ISW period.

As could be expected, the cloud fraction correlates positively with the downwelling longwave radiation (r =0.70) and negatively with the shortwave radiation (r = -0.25) in the observations. The reanalyses capture these physical relationships fairly well, i.e. the correlations between the cloud fraction and the other variables are similar in the reanalyses to the observations. However, there are some deviations, the most prominent being the correlation between cloud fraction and shortwave radiation in CFSR, which is 0.01. Having established these physical relationships, we proceed by analysing the relationships between the biases in cloud fraction and the biases in the mentioned, correlated variables. We do so by calculating the correlation coefficients between the cloud fraction biases and the biases in the downwelling short- and longwave radiation. The results are presented in Figure 10.

Figure 10

Correlation coefficients between the biases in cloud fraction and the downwelling short- and longwave radiation and the surface skin temperature for the four reanalyses for the ISW period.

In the ensuing analysis, we will mainly focus on the products with the largest cloud fraction biases, i.e. MERRA-2 and JRA-55 with biases of +30 % and -17 %, respectively. The cited biases are from the whole duration of ISW (Figure 7), but these and the biases for the other parameters are very similar to those from the time period investigated (not shown).

The correlation coefficient between the biases in cloud fraction and shortwave radiation is -0.32 in the case of MERRA-2 and -0.25 in the case of JRA-55 (Figure 10). Even though correlation does not necessarily imply causality, these are indications that the positive (negative) cloud fraction bias in MERRA-2 (JRA-55) contributes to a negative (positive) shortwave radiation bias. Correspondingly, we see that for the downwelling longwave radiation the corresponding numbers are 0.55 (MERRA-2) and 0.50 (JRA-55). Thereby, the positive (negative) cloud fraction bias in MERRA-2 (JRA-55) contributes to a positive (negative) bias in the downwelling longwave radiation.

Regarding ERA-Interim and CFSR, both products have positive cloud fraction biases, but simultaneously they also have negative downwelling longwave radiation biases. This underlines the fact that downwelling longwave radiation is not strictly a function of cloud cover, as also seen by the non-perfect correlations in Figures 9. Indeed downwelling longwave radiation also depends on other factors such as cloud water vapor content and aerosols, as pointed out by e.g. (Walsh, Chapman, and Portis 2009; Zib et al. 2012) in their study of cloud fraction biases and effects on biases in radiative fluxes in reanalyses, which yielded qualitatively similar results to ours.

Furthermore, we should mention that (Tastula et al. 2013) found similar biases in the cloud fractions in ERA-Interim and CFSR using the ISW dataset as reference. For specific time periods of ISW, they related these positive biases to positive downwelling longwave radiation and argued that these also contributed to warm biases in the near surface temperature (see their Figure 8). Other studies have found similar surface layer warm biases in ERA-Interim in Antarctica and attributed them to too strong turbulent heat fluxes during very strong temperature stratifications (Fréville et al. 2014). Unfortunately, observations of turbulent surface fluxes of sensible and latent heat are rather sparse in the ISW dataset, in particular for periods with very strong temperature stratifications, and we therefore chose to leave them out of this correlation analysis.

3.3 Comparison With Upper-Air Observations

In the following we evaluate the reanalyses' upper-air performance between 200 and 4000 m AGL with respect to potential temperature, specific humidity and wind speed. We pay particular attention to differences in performance with respect to height AGL, geographical location and dataset independence. Regarding locations, we note that the ISW airsonde and tethersonde soundings were obtained solely over the Antarctic sea ice-pack in autumn and winter, whereas the *Aranda* and *Fedorov* soundings were ship-based and were thus made partly over the open

ocean and partly over the sea ice (Figure 1) in summer and autumn. The error statistics results for the *Aranda* and *Fedorov* datasets and the ISW airsondes and ISW tethersondes datasets are presented in Figure 11 and Figure 12, respectively.

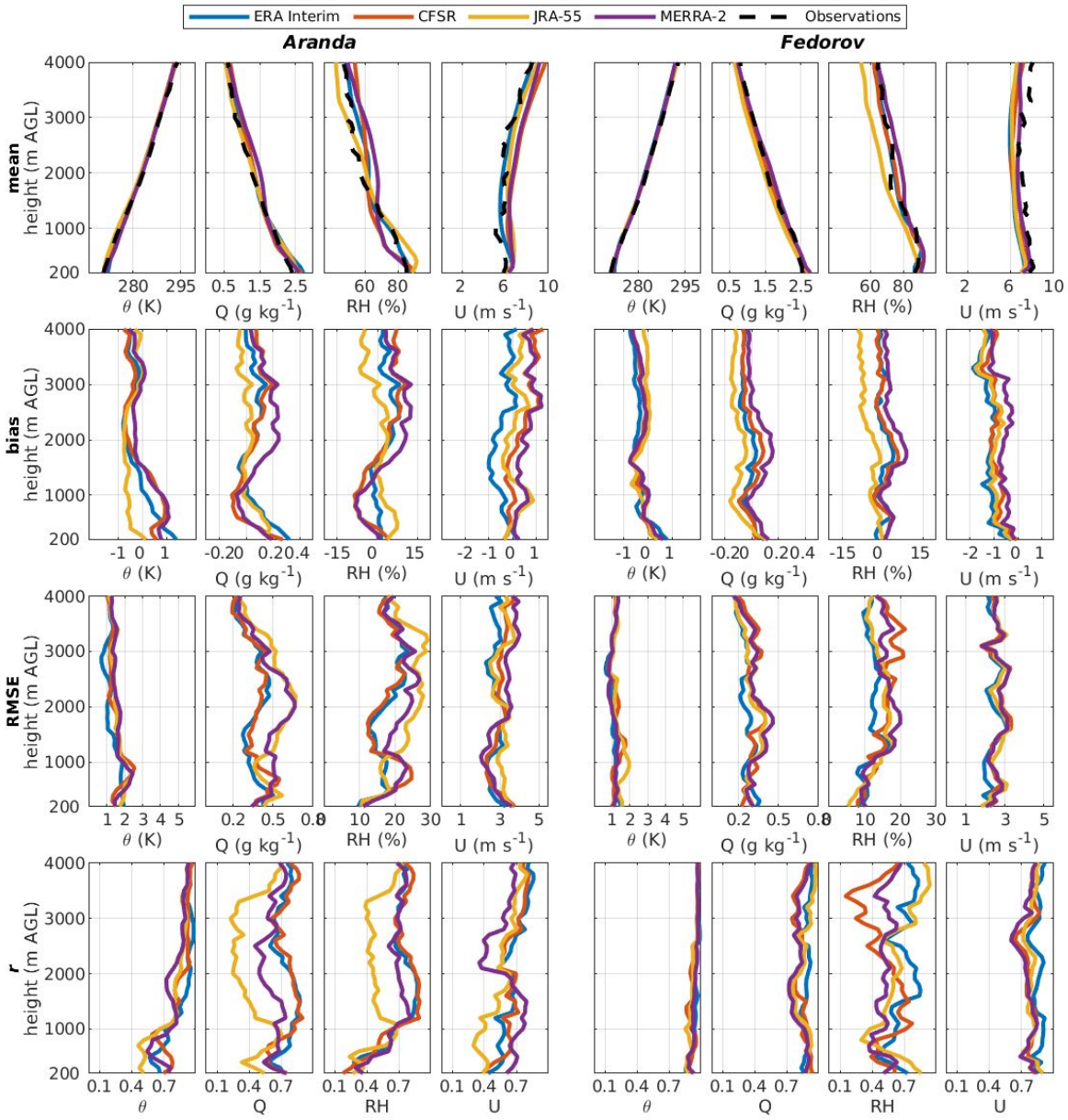

Figure 11. Profiles of mean, bias, root mean square error (RMSE) and correlation coefficient (*r*) for potential temperature (θ), specific humidity (Q) and wind speed (U) for the reanalyses using the *Aranda* and *Fedorov* profile datasets as reference.

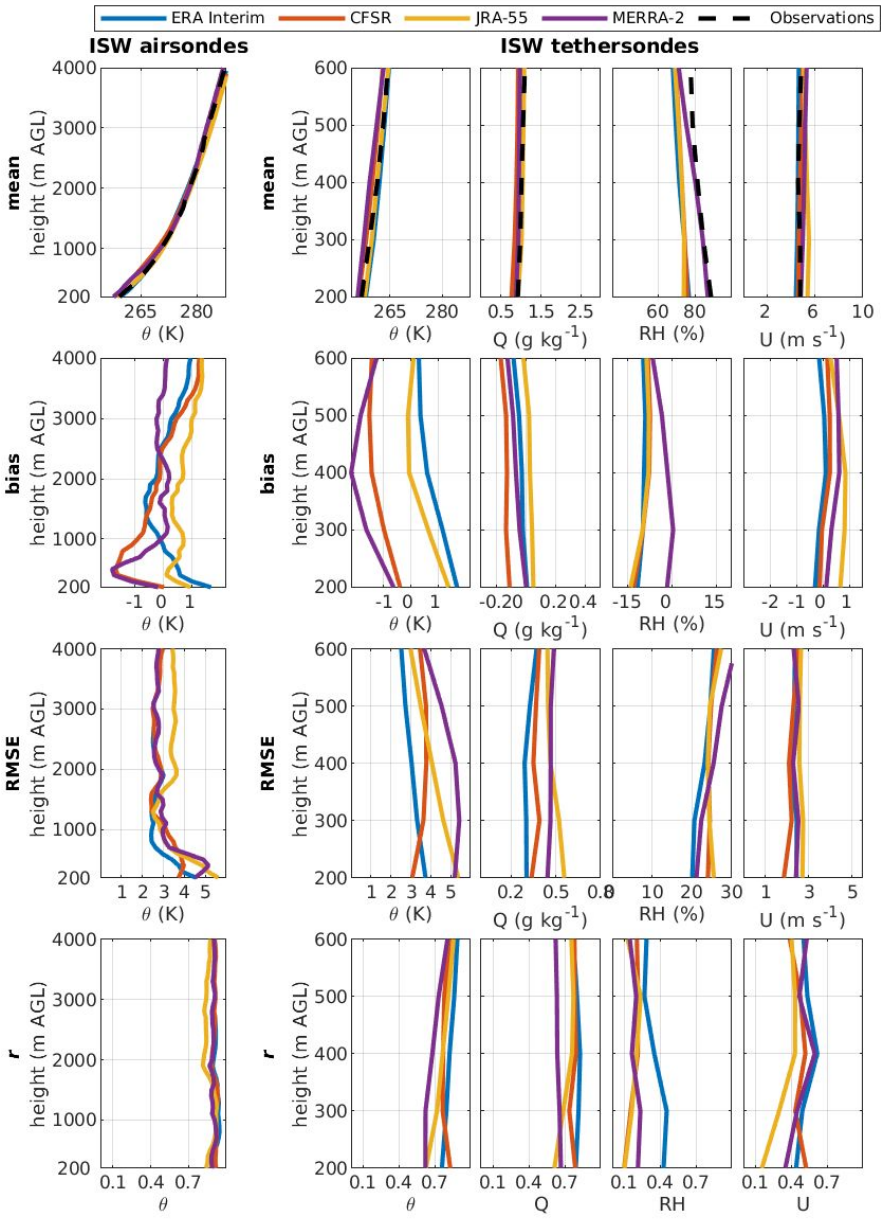

Figure 12. Same as in Figure 11, but for the ISW airsondes and ISW tethersondes datasets.

3.2.1 Temperature

The mean potential temperature profiles from the ship-based *Aranda* and *Fedorov* datasets (Figure 11) closely resemble each other, both in terms of their values at the lowest levels (about 270 K at 200 m) and their shape (variation with height). The mean ISW profiles (Figure 12), all observed over a compact sea ice field, deviate from these by being substantially colder (~255 K at 200 m) and more statically stably stratified in the lower hundreds of meters (~14 K km^{-1} vs ~7

K km^{-1}). The mean reanalyses profiles capture the general shapes of these observed profiles quite well. Though, there are some biases in the reanalyses. When compared against the *Aranda* and *Fedorov* datasets, the sign of the reanalysis biases depends on the altitude. In the lowest few hundred meters, there are warm biases of +0.1 to 1.4 K across all four reanalyses. Compared against the *Aranda* dataset, these warm biases are significant in all four reanalyses, except JRA-55, and they cover a relatively deep layer of up to 1500 m AGL. Compared against the *Fedorov* dataset, the low-level warm biases are generally smaller and occur in a shallower layer, and none of them are significant. Farther aloft, cold biases down to about -1 K dominate in both datasets. The spread in the biases between the reanalyses is generally larger for the independent *Aranda* and ISW sounding datasets than for the *Fedorov* dataset, especially at the lower levels. In the ISW soundings, the biases also differ in sign from the other two datasets; while there on average is a warm bias in ERA-Interim and JRA-55 in the lower hundreds of meters, there is a cold bias in MERRA-2 and CFSR for the same layer. We note that only the cold biases are statistically significant.

The highest RMSE for all four reanalyses is found in the lowest few hundred meters when compared against the independent *Aranda* and ISW airsonde datasets. The latter dataset reveals the highest RMSE at 200-300 m AGL, with the best RMSE found in CFSR (3.8 K) and the worst in JRA-55 (5.5 K). The *Fedorov* dataset does not reveal a clear maximum in the RMSE of the reanalyses at the lower levels, and the spread in RMSE is lower than in the independent datasets.

The least favorable *r* is found for the *Aranda* dataset between 200 and 1000 m AGL, and JRA-55 has the lowest *r* of about 0.45. Farther aloft, *r* is around or higher than 0.7 in all reanalyses and datasets investigated. The overall highest values of *r* and the smallest spread between the reanalyses are found in *Fedorov* and the ISW airsondes.

3.2.2 Specific Humidity

The mean specific humidity profiles in the *Aranda* and *Fedorov* datasets have similar shapes and values, ranging from about 2.5 g kg^{-1} or slightly above at the lowest levels down to below 0.5 g kg^{-1} at the highest levels. In comparison, the ISW tethersonde profiles are substantially drier with values between 0.6 and 1 g kg^{-1} up to 600 m AGL, which corresponds to the lower temperatures in these profiles. The largest bias is found for ERA-Interim in the *Aranda* dataset at 200 m AGL (+0.32 g kg^{-1}). Further aloft, MERRA-2 has the highest positive bias with values up to +0.2 g kg^{-1} between 2000 and 3000 m AGL. Both the spread in bias across the reanalyses and their absolute values are lower in the *Fedorov* dataset than in the *Aranda* dataset. Also, the biases are not statistically significant for the *Fedorov* dataset, except in JRA-55, which has a significant dry bias at most height levels peaking at -0.16 g kg^{-1} at 900 m AGL. The ISW tethersonde dataset shows small biases in the reanalyses and only for CFSR it has a significant value at all height levels of down to -0.17 g kg^{-1}.

Both the largest RMSE (about 0.7 g kg^{-1} in JRA-55 and MERRA-2) and largest spread in RMSE between reanalyses are found at 2000 m AGL in the Aranda reanalyses profiles. This is where the highest RMSE is found for the *Fedorov* dataset too, but the absolute values and their variation with height are lower in all four reanalyses. In the ISW tethersonde dataset, there is

little variation with height in the RMSE across the reanalyses and ERA-Interim has the lowest values (below 0.3 g kg^{-1} for most height levels).

The correlation coefficient, *r,* features large variability with height for the *Aranda* dataset, in particular in JRA-55, ranging from only 0.25 (not significant) between 2000 and 3000 m AGL to 0.7 (significant) at 4000 m AGL. *r* is generally much higher (all above 0.7) for the *Fedorov* than for the *Aranda* dataset and the spread in *r* between the reanalyses is lower. Also in the ISW tethersonde dataset the spread in *r* is fairly low between the reanalyses and the values are all 0.65 or above.

3.2.3 Relative humidity

The mean relative humidity profiles in the *Aranda* and *Fedorov* datasets feature the same variability with height as the specific humidity, i.e. the highest values are found at the lowermost levels (about 85 %) and the lowest values at the uppermost levels (about 55 %). The ISW tethersonde profiles also have values around 85 % for the lower altitudes covered by that dataset (up to 600 m AGL). Looking at the reanalyses biases, these are the highest in the *Aranda* dataset between 2000 and 3000 m, MERRA-2 having highest values at close to +15 %. A similar, but somewhat smaller bias is seen in the *Fedorov* dataset in MERRA-2. The reanalyses display negative biases throughout with values down to almost -15 % at 200 m AGL. An exception is MERRA-2, which has a bias very close to 0 % for most altitudes.

As is generally the case for the other variables investigated, both the largest RMSE (up to 30 % in JRA-55 at 3-4000 m AGL) and the largest spread in RMSE between the reanalyses is found in the *Aranda* dataset. The ISW tethersonde dataset displays little variation with height in the RMSE across the reanalyses and ERA-Interim has the lowest values (below 25 % for most height levels).

The correlation coefficients, *r,* feature a very similar variability with height to the ones for specific humidity in the *Aranda* dataset. For the *Fedorov* dataset, however, there is a relatively much stronger variability in *r* with height for relative than for specific humidity, CFSR having the lowest values down to almost 0.1 at about 3500 m AGL. For the ISW tethersondes, the values of *r* are fairly poor with none of the reanalyses having values above 0.5 for any height level.

3.2.4 Wind Speed

While the mean wind speed generally increases with height in the *Aranda* dataset, there is an overall, slight decrease with height in the *Fedorov* profiles. This variation with height is seen by the reanalyses, but they do feature biases. We note that none of the biases in the *Aranda* dataset are significant. For the *Fedorov* dataset, however, ERA-Interim features significantly negative biases, peaking at about -1.8 m s^{-1} at 3300 m AGL. The absolute values and variability in the wind speed bias across the reanalyses are larger in the *Aranda* dataset than in the *Fedorov* dataset. The mean ISW tethersonde reanalyses profiles follow very closely the observed wind, and the agreement is particularly good for JRA-55.

The RMSE is roughly between 2 and 4 m s^{-1} in all three wind speed datasets. ERA-Interim has marginally the lowest RMSE values for the *Aranda* and *Fedorov* datasets, and the spread in RMSE between the reanalyses is somewhat smaller for the *Fedorov* than for the other datasets.

The highest *r* (at least 0.6) across all products and height levels are found for the *Fedorov* dataset. The reanalyses also feature the smallest variation in *r* for this dataset. The lowest, and also insignificant value of *r* (only 0.1) for any height level, is found at 200 m AGL in the tethersonde JRA-55 profiles.

3.2.5 Summary of Upper-Air Reanalysis Performance and the Influence of Temperature Inversions

Table 1 Vertically averaged values (Mean-columns) of bias, RMSE and correlation coefficient (*r*) for potential temperature (θ), specific humidity (Q) and wind speed (U) for the four reanalyses using *Aranda* radiosonde soundings as reference. Mean values for all four reanalyses of the vertical averages are given in the rightmost column (Mean all-column). Error statistics valid up to 1500 m AGL for potential temperature and specific humidity profiles with(without) temperature inversions are indicated as inv(noinv). The reanalyses are ranked (Rank-columns) for each error statistics according to their performance (except inversion and non-inversion statistics). In the ranking, the best reanalysis received 4 points and the worst 1 point. The last row sums the total ranking points.

	ERA-Interim		CFSR		JRA 55		MERRA 2		
	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank	Mean all
θ bias (K)	-0.18	2	-0.13	3	-0.44	1	0.09	4	-0.17
θ RMSE (K)	1.31	4	1.49	3	1.59	2	1.61	1	1.50
θ r	0.86	4	0.85	3	0.78	1	0.80	2	0.82
θ noinv bias (K)	0.38		0.33		-0.40		0.32		0.16
θ noinv RMSE (K)	1.46		1.54		1.47		1.47		1.48
θ noinv r	0.78		0.81		0.76		0.76		0.78
θ inv bias (K)	0.43		0.87		-0.61		1.07		0.44
θ inv RMSE (K)	1.97		2.09		2.32		2.12		2.12
θ inv r	0.63		0.61		0.49		0.64		0.59
Q bias (g kg^{-1})	0.07	2	0.06	3	0.02	4	0.11	1	0.07
Q RMSE (g kg^{-1})	0.35	4	0.37	3	0.48	1	0.47	2	0.42
Q r	0.75	3.5	0.75	3.5	0.43	1	0.63	2	0.64
Q noinv bias (g kg^{-1})	0.09		-0.01		-0.05		0.03		0.01
Q noinv RMSE (g kg^{-1})	0.31		0.34		0.34		0.42		0.35
Q noinv r	0.83		0.78		0.75		0.66		0.76
Q inv bias (g kg^{-1})	0.09		0.00		0.18		0.02		0.07
Q inv RMSE (g kg^{-1})	0.40		0.45		0.56		0.52		0.48
Q inv r	0.66		0.70		0.34		0.67		0.59
RH bias (RH %)	2.52	2	2.23	3	0.74	4	3.89	1	2.35
RH RMSE (RH %)	17.69	4	18.23	3	22.42	1	21.01	2	19.84
RH r	0.70	3	0.71	4	0.50	1	0.67	2	0.65
RH noinv bias (RH %)	0.34		-3.36		-0.64		-0.84		-1.12
RH noinv RMSE (RH %)	11.95		13.33		13.23		14.72		13.31
RH noinv r	0.66		0.68		0.69		0.68		0.68
RH inv bias (RH %)	-1.03		-5.48		7.87		-6.82		-1.37
RH inv RMSE (RH %)	18.24		21.33		21.76		22.48		20.95
RH inv r	0.47		0.41		0.24		0.45		0.39
U bias (m s^{-1})	-0.37	2	0.36	3	0.14	4	0.56	1	0.17
U RMSE (m s^{-1})	2.80	4	3.03	3	3.16	2	3.20	1	3.05
U r	0.67	4	0.66	3	0.56	1	0.63	2	0.63
Total rank points		38.5		37.5		23		21	

Table 2 Same as in Table 1, but for the ISW airsonde dataset and only for potential temperature.

	ERA-Interim		CFSR		JRA 55		MERRA 2		
	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank	Mean all
θ bias (K)	0.18	3	-0.09	4	0.76	1	-0.22	2	0.16
θ RMSE (K)	2.79	4	2.91	3	3.48	1	3.04	2	3.05
θ r	0.91	3.5	0.91	3.5	0.86	1	0.90	2	0.90
θ noinv bias (K)	0.66		0.74		1.32		1.72		1.11
θ noinv RMSE (K)	1.35		1.70		1.51		2.21		1.69
θ noinv r	0.89		0.87		0.95		0.82		0.88
θ inv bias (K)	0.04		-1.36		0.30		-1.04		-0.52
θ inv RMSE (K)	3.18		3.51		3.81		3.86		3.59
θ inv r	0.92		0.91		0.88		0.89		0.90
Total rank points		10.5		10.5		3		6	

The results of the ranking reveal, that for all of the four upper-air datasets, ERA-Interim has the highest ranking scores, except for the ISW airsonde dataset where it shares the first place with CFSR. JRA-55 and MERRA-2 have either the lowest or the second lowest ranking scores for all four datasets. An exception is the *Fedorov* dataset where JRA-55 shares second place with CFSR. In terms of the vertically averaged values of RMSE and *r,* ERA-Interim is best, with only a few exceptions, for all datasets and all parameters considered. Considering biases, however,

ERA-Interim has the worst or second-worst values for several datasets and parameters, such as for the potential temperature and wind speed for the *Aranda* and *Fedorov* datasets.

A closer inspection of the profile error statistics results presented Figures 7 and 8 and individual profiles (not shown) from the *Aranda* and ISW airsonde datasets, reveals that there are particularly large temperature errors in the lowermost 1500 m AGL of these datasets. This atmospheric layer coincides with the layer in which temperature inversions are known to be ubiquitous in the polar regions (Kahl 1990; Devasthale et al. 2010; Zhang and Seidel 2011). In order to investigate the impact of temperature inversions on the error statistics, we have calculated separate error statistics for the lowest 1500 m AGL for two subsets of profiles, with and without inversions in the lowermost 1500 m AGL. Following (R. W. Jones et al. 2016a) we define an inversion as an atmospheric layer in which the temperature increases with height by 2 K or more. We only use the *Aranda* and ISW airsonde datasets in the calculations of these statistics because of specific characteristics of the other datasets, outlined in the following. The *Fedorov* dataset is not independent and, compared to this dataset, the reanalyses do not have any significant temperature biases nor do they have substantially increased temperature RMSE in the lower layers. The ISW tethersonde dataset has a limited vertical range (up to 600 m AGL) and is therefore also excluded from this analysis.

In the *Aranda* dataset, we identified at least one temperature inversion in 16 out of 34 profiles, and in the ISW airsonde data the corresponding numbers were 32 out of 40 profiles. Among the latter, the temperature data was of poor quality in the lower 1500 m AGL in one profile and we therefore set the resulting number of ISW airsonde profiles containing no inversion to 7. The results from the *Aranda* reanalysis profile error statistics (Table 1) reveal that when temperature inversions are present, the mean potential temperature bias (mean of all reanalyses) in the lowermost 1500 m layer is +0.44 K, and for profiles without any inversion it is +0.16 K. For mean RMSE the values are 2.12 K (inversion) and 1.48 K (no inversion), and for the mean r they are 0.59 (inversion) and 0.78 (no inversion). Thus, there is on average a consistent degradation in model performance for potential temperature when the profiles include inversions and this result is also valid for when the statistics for each individual reanalysis are considered. The increased positive temperature biases with inversions present alludes to a general underestimation of the inversion strength. Furthermore, we also note that the error statistics for specific humidity are also generally worse for profiles with than without temperature inversions for the *Aranda* dataset and the same is seen for relative humidity.

For the ISW airsonde dataset, the situation is somewhat different with respect to the potential temperature profile error statistics presented in Table 2; while the mean RMSE for all reanalyses degrades from the non-inversion (1.69 K) to the inversion profiles (3.59 K), the bias and r improve from +1.11 to -0.52 K and 0.88 to 0.9, respectively. For ERA-Interim, the bias is reduced from 0.66 to 0.04 K when going from non-inversion to inversion profiles, and the corresponding change for JRA-55 is from +1.32 to +0.95 K. For CFSR and MERRA-2, on the other hand, there is a change in sign in the biases and they go from +0.74 to -1.36 K and +1.72 to -1.04 K, respectively, when the non-inversion profiles are compared with the inversion profiles. This would suggest a general overestimation of the inversion strengths by CFSR and MERRA-2, and an underestimation by ERA-Interim and JRA-55. This hypothesis is supported by the

comparison of the means of the non-inversion and inversion ISW airsonde temperature profiles, as presented in Figure 6. We note, however, that the number of non-inversion profiles is relatively low (7), limiting the statistical robustness of this result.

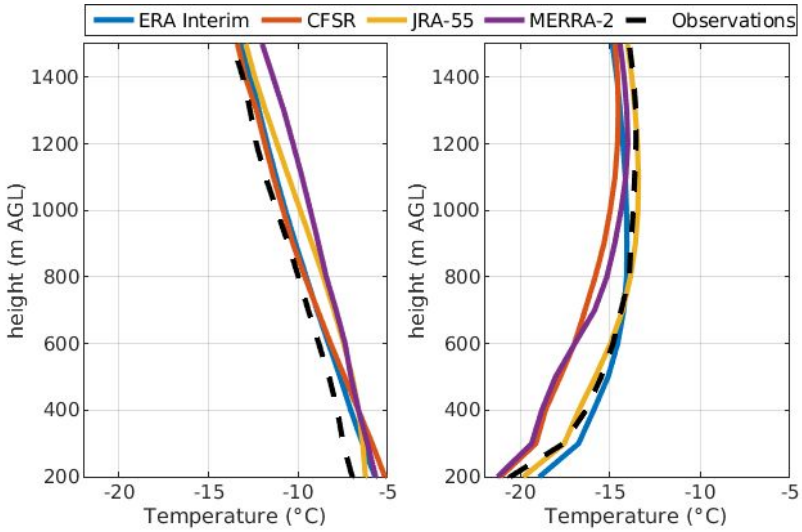

Figure 9. Mean of ISW airsonde temperature profiles split in two groups: (left panel) 7 profiles where no temperature inversion was observed and (right panel) 32 profiles where at least one temperature inversion was observed by the ISW airsonde.

4 Discussion and Conclusions

In this study, we used surface layer and upper-air data from two research cruises and one ice station in the Weddell Sea from 1992 and 1996 to validate the four current reanalyses ERA-Interim, CFSR, JRA-55 and MERRA-2.

In terms of surface layer performance, we find that CFSR and ERA-Interim perform the best when the ranking scores are summed, ERA-Interim having the overall highest score. We note that CFSR performs particularly well for the 2 m temperature, where it has the best bias, RMSE and *r,* and it also performs well for 10 m wind speed where it is the only product with a non-significant bias. We do not know why CFSR performs this well for these variables, but note that among the reanalyses evaluated here CFSR is the only product based on a coupled atmosphere - sea ice - ocean model (Suranjana Saha et al. 2010), which is expected to yield advantage close to the air - sea and air- ice interphases. ERA-Interim, on the other hand, performs very well when it comes to the relative humidity, both with respect to water and with respect to ice. It is also the product that in general comes the closest to reproducing the observed saturation and supersaturation with respect to ice for the temperature range investigated, as has previously been documented by (E. L. Andreas et al. 2002) for the ISW observations.

All four reanalyses feature warm biases in their 2 m temperature, ranging from +1.99 K in CFSR to +2.81 K in MERRA-2. There are few other evaluation studies of reanalyses covering the Antarctic sea ice zone, largely because few observations exist and those that do cover rather short time periods. This is particularly true for the current reanalyses such as evaluated herein. One study that does evaluate current reanalyses for the Antarctic sea ice, however, is that by (Jones et al. 2016). Based on data from three research vessel cruises, they documented near-surface cold biases in ERA-Interim, CFSR, JRA-55 and MERRA in the Amundsen Sea Embayment in West Antarctica. Though, these biases were dominated by strong negative values close to the coast (approaching -6 K). Farther offshore, they found data points with weaker and even positive temperature biases in all four reanalyses. Regarding older reanalyses, the NCEP-National Center for Atmospheric Research (NCAR) reanalysis (Kalnay et al. 1996), has been found to have a cold bias by (Vihma 2002) and (Vancoppenolle et al. 2011), who used buoy data and data from SIMBA and ISPOL in their evaluations. Using ERA-Interim data from 1979 to 2013, (Jones and Lister 2014), also found cold biases along the coast, whereas they also documented prominent warm biases in the Antarctic interior. Using satellite data, (Fréville et al. 2014) also found widespread warm biases in the Antarctic interior.

We find prominent biases in the reanalyses' cloud fraction data, and these are particularly strong in MERRA-2 (+30.2 %) and JRA-55 (-17.3 %). Our analysis indicates that these cloud fraction biases contribute to a positive downwelling longwave radiation bias in MERRA-2 and a negative downwelling longwave radiation bias in JRA-55. Similarly, they also contribute to a negative downwelling shortwave bias in MERRA-2 and a positive downwelling shortwave bias in JRA-55. Thus, these biases partly compensate each other in the surface energy balance. However, all variance in the radiative flux biases is not explained by the cloud fraction biases, and we find that both ERA-Interim and CFSR have negative downward longwave radiation

biases in spite of their positive cloud fraction biases. Radiative fluxes are indeed also functions of more complex cloud properties such as aerosols and water vapor content (Walsh et al. 2009; Zib et al. 2012; Vancoppenolle et al. 2011). Former studies have found that substantial cloud fraction biases are common in reanalyses in polar regions and, like us, that positive biases typically induce positive biases in downwelling longwave radiation and negative biases in shortwave radiation (Walsh, Chapman, and Portis 2009; Zib et al. 2012).

Reliable near-surface variables in reanalyses are important for climate research, where variables such as 2 m air temperature and 10 m wind speed often receive much attention, and for usage of reanalyses in driving ocean and sea-ice models. For example, too high near-surface wind speeds, as we document for ERA-Interim, JRA-55 and MERRA-2, and to some degree for CFSR, would lead to overestimation of the wind stress and its curl, with implications to e.g. ocean dynamics and transport, and sea-ice drift (Suranjana Saha et al. 2010; Uotila et al. 2014).

Regarding upper-air performance, ERA-Interim outperforms all the other reanalyses when the rank scores are summed up for all datasets, except the ISW airsondes where it shares the first place with CFSR. However, ERA-Interim does suffer from some prominent biases, including a significant warm bias of up to +1.4 K for the *Aranda* reanalyses soundings. This low-level warm bias in ERA-Interim is consistent with findings in more recent data from the eastern side of the Antarctic peninsula (Nygård et al. 2016) and also with data from the Arctic (Lüpkes et al. 2010; Jakobson et al. 2012; Wesslén et al. 2014; de Boer et al. 2014; Liu et al. 2008). Corresponding to this warm bias, there is a significant moist bias in the lowermost layers of the reanalyses, when compared against the *Aranda* soundings and, to a lesser degree, the *Fedorov* soundings. At higher levels (above 1500 m AGL), there are significant cold biases in both the *Fedorov* and *Aranda* profiles in all four reanalyses. Similar cold biases were found by (Nygård et al. 2016) for reanalysis data from the eastern side of the Antarctic Peninsula. They pointed out that such cold biases are consistent with biases found in satellite data (J. Wang et al. 2013; Boylan et al. 2015), and reasoned that this might be a source for these cold biases. However, none of the satellite datasets addressed in (J. Wang et al. 2013; Boylan et al. 2015) and (J. Wang et al. 2013; Boylan et al. 2015) were available in 1992 and 1996.

Regarding data independency, the *Fedorov* dataset was the only dataset made available for assimilation through the GTS. Assuming that the *Fedorov* data was used in data assimilation, this is probably a major reason why the error statistics are generally more favorable and there is mostly a smaller spread between the reanalyses for this dataset than for the other upper-air datasets. Still, significant biases remain, as for example upper-level cold biases down to -0.7 K in ERA-Interim. Hence, there are indications that the data assimilation system does not fully utilize available soundings, e.g. by giving them too little weight. This is consistent with previous studies, e.g. by (Lüpkes et al. 2010; Jakobson et al. 2012; Wesslén et al. 2014; de Boer et al. 2014; Liu et al. 2008), who found biases similar in nature to these in spite of the fact that the radiosonde data used for validation were sent to GTS and were thus available for assimilation.

Considering spatial variability in reanalysis performance, previous validation studies for the Antarctic have revealed substantial differences between different products

(R. W. Jones et al. 2016b; Nygård et al. 2016). In our study, we identified strong spatial differences with respect to lower-level potential temperature. In the strongly, statically stratified lower layers of the ISW soundings, solely performed over the Antarctic sea ice, JRA-55 and ERA-Interim both feature warm biases (though not significant), just like in the ship-based *Aranda* and *Fedorov* soundings. CFSR and MERRA-2, however, feature significant cold biases in the lower levels of the ISW soundings. Temperature inversions are known to be difficult to represent appropriately in models and several previous studies have found larger errors in connection to these (Lüpkes et al. 2010; Pavelsky et al. 2010; Harden, Renfrew, and Petersen 2011; R. W. Jones et al. 2016b). It is therefore not entirely surprising that such differences occur. We do see that the reanalyses error statistics for the lower 1500 m AGL degrade when going from the non-inversion profiles to profiles containing inversions for the *Aranda* dataset, and to some degree also for the ISW airsonde dataset.

Finally, it is worth noting that even though the observations considered in this study are older than in several recent evaluation studies (Nygård et al. 2016; R. W. Jones et al. 2016b; Lüpkes et al. 2010), the documented biases are largely similar in nature, including dominant warm biases in the ABL. As the amount of observations has varied depending on the decade, season and region, the consistency of the warm bias in the ABL suggests a need to improve ABL and surface parameterisations. In addition to reanalyses, a warm near-surface bias in conditions of a stable boundary layer is a common feature in numerical weather prediction, and often attributed to excessive heat and momentum fluxes in the stable boundary layer (Cuxart et al. 2006; Vihma et al. 2014)

Acknowledgments

We express our deep gratitude for the work of the late Edgar L. Andreas in Ice Station Weddell. This study was supported by the Academy of Finland (contract 304345) and the EC Marie Curie Support Action LAWINE (grant 707262). A. P. Makshtas was supported by the Ministry of Education and Science of the Russian Federation (Project RFMEFI61617X0076). The observational datasets used in this study are available at the websites indicated in the supporting information.

References

Andreas, Edgar L. 1995. “Air-Ice Drag Coefficients in the Western Weddell Sea: 2. A Model Based on Form Drag and Drifting Snow.” *Journal of Geophysical Research*, Bull., 100 (C3): 4833.

———. 2002. “Parameterizing Scalar Transfer over Snow and Ice: A Review.” *Journal of Hydrometeorology* 3 (4): 417–32.

Andreas, Edgar L., and Kerry J. Claffey. 1995. “Air-Ice Drag Coefficients in the Western Weddell Sea: 1. Values Deduced from Profile Measurements.” *Journal of Geophysical Research* 100 (C3): 4821.

Andreas, Edgar L., Kerry J. Claffy, and Aleksandr P. Makshtas. 2000. “Low-Level Atmospheric Jets And Inversions Over The Western Weddell Sea.” *Boundary-Layer Meteorology* 97 (3): 459–86.

Andreas, Edgar L., Rachel E. Jordan, and Aleksandar P. Makshtas. 2005. “Parameterizing Turbulent Exchange over Sea Ice: The Ice Station Weddell Results.” *Boundary-Layer Meteorology* 114 (2): 439–60.

Andreas, Edgar L., Rachel E. Jordan, and Aleksandr P. Makshtas. 2004. “Simulations of Snow, Ice, and

Near-Surface Atmospheric Processes on Ice Station Weddell." *Journal of Hydrometeorology* 5 (4): 611–24.

Andreas, E. L., S. P. Guest, P. O. G. Persson, C. W. Fairall, T. W. Horst, R. E. Moritz, and S. R. Semmer. 2002. "Near-Surface Water Vapor over Polar Sea Ice Is Always near Ice Saturation." *Journal of Geophysical Research* 107 (C10). https://doi.org/10.1029/2000jc000411.

Assmann, K. M., A. Jenkins, D. R. Shoosmith, D. P. Walker, S. S. Jacobs, and K. W. Nicholls. 2013. "Variability of Circumpolar Deep Water Transport onto the Amundsen Sea Continental Shelf through a Shelf Break Trough." *Journal of Geophysical Research, C: Oceans* 118 (12): 6603–20.

Boer, G. de, M. D. Shupe, P. M. Caldwell, S. E. Bauer, O. Persson, J. S. Boyle, M. Kelley, S. A. Klein, and M. Tjernström. 2014. "Near-Surface Meteorology during the Arctic Summer Cloud Ocean Study (ASCOS): Evaluation of Reanalyses and Global Climate Models." *Atmospheric Chemistry and Physics* 14 (1): 427–45.

Boylan, Patrick, Junhong Wang, Stephen A. Cohn, Eric Fetzer, Eric S. Maddy, and Sun Wong. 2015. "Validation of AIRS Version 6 Temperature Profiles and Surface-Based Inversions over Antarctica Using Concordiasi Dropsonde Data: AIRS v6 Antarctic Surface Inversions." *Journal of Geophysical Research, D: Atmospheres*, Ser. on Atmos. Oceanic and Planet. Phys, 120 (3): 992–1007.

Bracegirdle, Thomas J., and Gareth J. Marshall. 2012. "The Reliability of Antarctic Tropospheric Pressure and Temperature in the Latest Global Reanalyses." *Journal of Climate* 25 (20): 7138–46.

Bromwich, David H., Ryan L. Fogt, Kevin I. Hodges, and John E. Walsh. 2007. "A Tropospheric Assessment of the ERA-40, NCEP, and JRA-25 Global Reanalyses in the Polar Regions." *Journal of Geophysical Research* 112 (D10): 1147.

Bromwich, David H., Julien P. Nicolas, Andrew J. Monaghan, Matthew A. Lazzara, Linda M. Keller, George A. Weidner, and Aaron B. Wilson. 2013. "Central West Antarctica among the Most Rapidly Warming Regions on Earth." *Nature Geoscience* 6 (2): 139–45.

Claffey, K. J., L. A. Edgar, and A. Makshtas. 1994. "Upper-Air Data Collected on Ice Station Weddell." *United States. Army. Corps of Engineers & National Science Foundation (U.S.).* https://catalogue.nla.gov.au/Record/4106075.

Cuxart, J., A. A. M. Holtslag, R. J. Beare, E. Bazile, A. Beljaars, A. Cheng, L. Conangla, et al. 2006. "Single-Column Model Intercomparison for a Stably Stratified Atmospheric Boundary Layer." *Boundary-Layer Meteorology* 118 (2): 273–303.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, et al. 2011. "The ERA-Interim Reanalysis: Configuration and Performance of the Data Assimilation System." *Quarterly Journal of the Royal Meteorological Society* 137 (656): 553–97.

Devasthale, A., U. Willén, K-G Karlsson, and C. G. Jones. 2010. "Quantifying the Clear-Sky Temperature Inversion Frequency and Strength over the Arctic Ocean during Summer and Winter Seasons from AIRS Profiles." *Atmospheric Chemistry and Physics* 10 (12): 5565–72.

Dutrieux, P., J. De Rydt, A. Jenkins, P. R. Holland, H. K. Ha, S. H. Lee, E. J. Steig, Q. Ding, E. P. Abrahamsen, and M. Schroder. 2014. "Strong Sensitivity of Pine Island Ice-Shelf Melting to Climatic Variability." *Science* 343 (6167): 174–78.

Fréville, H., E. Brun, G. Picard, N. Tatarinova, L. Arnaud, C. Lanconelli, C. Reijmer, and M. van den Broeke. 2014. "Using MODIS Land Surface Temperatures and the Crocus Snow Model to Understand the Warm Bias of ERA-Interim Reanalyses at the Surface in Antarctica." *The Cryosphere Discussions* 8 (1): 55–84.

Gelaro, Ronald, Will McCarty, Max J. Suárez, Ricardo Todling, Andrea Molod, Lawrence Takacs, Cynthia A. Randles, et al. 2017. "The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)." *Journal of Climate* 30 (14): 5419–54.

Goessling, Helge F., Thomas Jung, Stefanie Klebe, Jenny Baeseman, Peter Bauer, Peter Chen, Matthieu Chevallier, et al. 2016. "Paving the Way for the Year of Polar Prediction." *Bulletin of the American*

Meteorological Society 97 (4): ES85–88.
Gordon, Arnold L., and Ice Station Weddell Group of Principal Investigators and Chief Scientists. 1993. "Weddell Sea Exploration from Ice Station." *Eos, Transactions American Geophysical Union* 74 (11): 121–26.
Harden, B. E., I. A. Renfrew, and G. N. Petersen. 2011. "A Climatology of Wintertime Barrier Winds off Southeast Greenland." *Journal of Climate* 24 (17): 4701–17.
Jakobson, Erko, Timo Vihma, Timo Palo, Liisi Jakobson, Hannes Keernik, and Jaak Jaagus. 2012. "Validation of Atmospheric Reanalyses over the Central Arctic Ocean." *Geophysical Research Letters* 39 (10). https://doi.org/10.1029/2012GL051591.
Jones, P. D., and D. H. Lister. 2014. "Antarctic near-Surface Air Temperatures Compared with ERA-Interim Values since 1979." *International Journal of Climatology* 35 (7): 1354–66.
Jones, R. W., I. A. Renfrew, A. Orr, B. G. M. Webber, D. M. Holland, and M. A. Lazzara. 2016a. "Evaluation of Four Global Reanalysis Products Using in Situ Observations in the Amundsen Sea Embayment, Antarctica." *Journal of Geophysical Research, D: Atmospheres* 121 (11): 6240–57.
———. 2016b. "Evaluation of Four Global Reanalysis Products Using in Situ Observations in the Amundsen Sea Embayment, Antarctica: Amundsen Sea Reanalyses Evaluation." *Journal of Geophysical Research, D: Atmospheres* 121 (11): 6240–57.
Jung, Thomas, Neil D. Gordon, Peter Bauer, David H. Bromwich, Matthieu Chevallier, Jonathan J. Day, Jackie Dawson, et al. 2016. "Advancing Polar Prediction Capabilities on Daily to Seasonal Time Scales." *Bulletin of the American Meteorological Society* 97 (9): 1631–47.
Kahl, Jonathan D. 1990. "Characteristics of the Low-Level Temperature Inversion along the Alaskan Arctic Coast." *International Journal of Climatology* 10 (5): 537–48.
Kobayashi, Shinya, O. T. A. Yukinari, Yayoi Harada, Ayataka Ebita, Masami Moriya, Hirokatsu Onoda, Kazutoshi Onogi, et al. 2015. "The JRA-55 Reanalysis: General Specifications and Basic Characteristics." *Journal of the Meteorological Society of Japan. Ser. II* 93 (1): 5–48.
Launiainen, J., and T. Vihma. 1990. "Derivation of Turbulent Surface Fluxes — An Iterative Flux-Profile Method Allowing Arbitrary Observing Heights." *Environmental Software* 5 (3): 113–24.
Lindsay, R., M. Wensnahan, A. Schweiger, and J. Zhang. 2014. "Evaluation of Seven Different Atmospheric Reanalysis Products in the Arctic*." *Journal of Climate* 27 (7): 2588–2606.
Liu, Jiping, Zhanhai Zhang, Yongyun Hu, Liqi Chen, Yongjiu Dai, and Xiaobo Ren. 2008. "Assessment of Surface Air Temperature over the Arctic Ocean in Reanalysis and IPCC AR4 Model Simulations with IABP/POLES Observations." *Journal of Geophysical Research* 113 (D10): 155.
Lüpkes, C., T. Vihma, E. Jakobson, G. König-Langlo, and A. Tetzlaff. 2010. "Meteorological Observations from Ship Cruises during Summer to the Central Arctic: A Comparison with Reanalysis Data." *Geophysical Research Letters* 37 (9). https://doi.org/10.1029/2010GL042724.
Nicolas, Julien P., and David H. Bromwich. 2014. "New Reconstruction of Antarctic Near-Surface Temperatures: Multidecadal Trends and Reliability of Global Reanalyses*." *Journal of Climate* 27 (21): 8070–93.
Nygård, Tiina, Timo Vihma, Gerit Birnbaum, Jörg Hartmann, John King, Tom Lachlan-Cope, Russell Ladkin, Christof Lüpkes, and Alexandra Weiss. 2016. "Validation of Eight Atmospheric Reanalyses in the Antarctic Peninsula Region." *Quarterly Journal of the Royal Meteorological Society* 142 (695): 684–92.
Pavelsky, Tamlin M., Julien Boé, Alex Hall, and Eric J. Fetzer. 2010. "Atmospheric Inversion Strength over Polar Oceans in Winter Regulated by Sea Ice." *Climate Dynamics* 36 (5-6): 945–55.
Perez, Jorge, Melisa Menendez, Fernando J. Mendez, and Inigo J. Losada. 2014. "Evaluating the Performance of CMIP3 and CMIP5 Global Climate Models over the North-East Atlantic Region." *Climate Dynamics* 43 (9-10): 2663–80.
Rinke, A., K. Dethloff, J. J. Cassano, J. H. Christensen, J. A. Curry, P. Du, E. Girard, et al. 2006.

"Evaluation of an Ensemble of Arctic Regional Climate Models: Spatiotemporal Fields during the SHEBA Year." *Climate Dynamics* 27 (4): 433–35.

Saha, S., S. Nadiga, C. Thiaw, J. Wang, W. Wang, Q. Zhang, H. M. Van den Dool, et al. 2006. "The NCEP Climate Forecast System." *Journal of Climate* 19 (15): 3483–3517.

Saha, Suranjana, Shrinivas Moorthi, Hua-Lu Pan, Xingren Wu, Jiande Wang, Sudhir Nadiga, Patrick Tripp, et al. 2010. "The NCEP Climate Forecast System Reanalysis." *Bulletin of the American Meteorological Society* 91 (8): 1015–58.

Screen, James A., and Ian Simmonds. 2011. "Erroneous Arctic Temperature Trends in the ERA-40 Reanalysis: A Closer Look." *Journal of Climate* 24 (10): 2620–27.

Steig, Eric J., David P. Schneider, Scott D. Rutherford, Michael E. Mann, Josefino C. Comiso, and Drew T. Shindell. 2009. "Warming of the Antarctic Ice-Sheet Surface since the 1957 International Geophysical Year." *Nature* 457 (7228): 459–62.

Sturaro, G. 2003. "A Closer Look at the Climatological Discontinuities Present in the NCEP/NCAR Reanalysis Temperature due to the Introduction of Satellite Data." *Climate Dynamics* 21 (3-4): 309–16.

Tastula, Esa-Matti, Timo Vihma, and Edgar L. Andreas. 2012. "Evaluation of Polar WRF from Modeling the Atmospheric Boundary Layer over Antarctic Sea Ice in Autumn and Winter." *Monthly Weather Review* 140 (12): 3919–35.

Tastula, Esa-Matti, Timo Vihma, Edgar L. Andreas, and Boris Galperin. 2013. "Validation of the Diurnal Cycles in Atmospheric Reanalyses over Antarctic Sea Ice." *Journal of Geophysical Research, D: Atmospheres* 118 (10): 4194–4204.

Uotila, P., P. R. Holland, T. Vihma, S. J. Marsland, and N. Kimura. 2014. "Is Realistic Antarctic Sea-Ice Extent in Climate Models the Result of Excessive Ice Drift?" *Ocean Modelling* 79 (July): 33–42.

Vancoppenolle, Martin, Ralph Timmermann, Stephen F. Ackley, Thierry Fichefet, Hugues Goosse, Petra Heil, Katherine C. Leonard, et al. 2011. "Assessment of Radiation Forcing Data Sets for Large-Scale Sea Ice Models in the Southern Ocean." *Deep-Sea Research. Part II, Topical Studies in Oceanography* 58 (9-10): 1237–49.

Vihma, T., J. Launiainen, J. Uotila, and A. Kotro. 1997. "FINNARP Air-Sea-Ice Interaction Experiment in the Weddell Sea in 1996-1997." *Antarctic Reports of Finland 7*, 30.

Vihma, T., R. Pirazzini, I. Fer, I. A. Renfrew, J. Sedlar, M. Tjernström, C. Lüpkes, et al. 2014. "Advances in Understanding and Parameterization of Small-Scale Physical Processes in the Marine Arctic Climate System: A Review." *Atmospheric Chemistry and Physics* 14 (17): 9403–50.

Walsh, John E., and William L. Chapman. 1998. "Arctic Cloud–Radiation–Temperature Associations in Observational Data and Atmospheric Reanalyses." *Journal of Climate* 11 (11): 3030–45.

Walsh, John E., William L. Chapman, and Diane H. Portis. 2009. "Arctic Cloud Fraction and Radiative Fluxes in Atmospheric Reanalyses." *Journal of Climate* 22 (9): 2316–34.

Wang, Junhong, Terry Hock, Stephen A. Cohn, Charlie Martin, Nick Potts, Tony Reale, Bomin Sun, and Frank Tilley. 2013. "Unprecedented Upper-Air Dropsonde Observations over Antarctica from the 2010 Concordiasi Experiment: Validation of Satellite-Retrieved Temperature Profiles." *Geophysical Research Letters* 40 (6): 1231–36.

Wang, Yang, Dong Zhou, Armin Bunde, and Shlomo Havlin. 2016. "Testing Reanalysis Data Sets in Antarctica: Trends, Persistence Properties, and Trend Significance." *Journal of Geophysical Research, D: Atmospheres* 121 (21): 12,839–12,855.

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." *Atmospheric Chemistry and Physics* 14 (5): 2605–24.

Zhang, Yehui, and Dian J. Seidel. 2011. "Challenges in Estimating Trends in Arctic Surface-Based Inversions from Radiosonde Data." *Geophysical Research Letters* 38 (17).

https://doi.org/10.1029/2011GL048728.

Zib, Behnjamin J., Xiquan Dong, Baike Xi, and Aaron Kennedy. 2012. "Evaluation and Intercomparison of Cloud Fraction and Radiative Fluxes in Recent Reanalyses over the Arctic Using BSRN Surface Observations." *Journal of Climate* 25 (7): 2291–2305.