Validation of atmospheric reanalyses over the central Arctic Ocean

Erko Jakobson[1,2], Timo Vihma[3], Timo Palo[4], Liisi Jakobson[4], Hannes Keernik[1,2], Jaak Jaagus[4]

[1]Tartu Observatory, 61602 Tõravere, Estonia

[2]Department of Physics, University of Tartu, 50090, Tartu, Estonia

[3]Meteorological Research, Finnish Meteorological Institute, 00101Helsinki, Finland

[4]Department of Geography, University of Tartu, 50090, Tartu, Estonia

Correspondence should be addressed to Erko Jakobson, erko.jakobson@ut.ee

Abstract

Atmospheric reanalyses were validated against tethersonde sounding data on air temperature, air humidity and wind speed, collected during the drifting ice station Tara in the central Arctic in April – August 2007. The data were not assimilated into the reanalyses, providing a rare possibility for their independent validation, which was here made for the lowermost 890 m layer. The following reanalyses were included in the study: the European ERA-Interim, the Japanese JCDAS, and the U.S. NCEP-CFSR, NCEP-DOE, and NASA-MERRA. All reanalyses included large errors. ERA-Interim was ranked first; it outperformed the other reanalyses in the bias and root-mean-square-error (RMSE) for air temperature as well as in the bias, RMSE, and correlation coefficient for the wind speed. ERA-Interim suffered, however, from a warm bias of up to 2°C in the lowermost 400 m layer and a moist bias of 0.3 to 0.5 g kg^{-1} throughout the 890 m layer. The NCEP-CFSR, NCEP-DOE, and NASA-MERRA reanalyses outperformed the other reanalyses with respect to 2-m air temperature and specific humidity and 10-m wind speed, which makes them, especially NCEP-CSFR, better in providing turbulent flux forcing for sea ice models. Considering the whole vertical

profile, however, the older NCEP-DOE got the second highest overall ranking, being better than the new NCEP-CFSR. Considering the whole group of reanalyses, the largest air temperature errors surprisingly occurred during higher-than-average wind speeds. The observed biases in temperature, humidity, and wind speed were in many cases comparable or even larger than the climatological trends during the latest decades.

1. Introduction

Atmospheric reanalyses are widely applied in Arctic research (a) to study climate variability and trends [*Serreze et al.,* 2006; 2007; *Jakobson and Vihma,* 2010; *Cullather and Bosilovich,* 2011], (b) to better understand large-scale circulation and teleconnection patterns [*Thompson and Wallace,* 1998], (c) to validate climate models [*Rinke et al.,* 2006], and (d) to provide boundary conditions for ocean, sea ice, land-surface, and limited-area atmospheric models. In data sparse areas, such as the Arctic, reanalyses are arguably the best available source of integrated information on the four-dimensional structure of the atmosphere [*Screen and Simmonds*, 2011], although this is not necessarily true for all variables. The first global reanalyses included the National Center of Environmental Predictions / National Center for Atmospheric Research (NCEP/NCAR; *Kalnay et al.* [1996]) reanalysis, its improved version by NCEP and the Department of Energy (NCEP-DOE; *Kanamitsu et al.* [2002]), as well as the ERA-15 [*Gibson et al.,* 1997] and ERA-40 [*Uppala et al.,* 2005] reanalyses of the European Centre for Medium-Range Weather Forecasts. Comparisons against observations from the Arctic have, however, revealed major problems in each reanalysis: large errors exist in many variables, including near-surface air temperature, specific humidity, wind speed and direction [*Walsh and Chapman,* 1998; *Tjernström and Graversen,* 2009; *Screen and Simmonds,* 2011].

To improve the situation, extensive work has recently been carried out in producing new reanalyses, such as the ECMWF ERA-Interim reanalysis [*Dee et al.,* 2011], the Japanese Meteorological Agency Climate Data Assimilation System (JCDAS, which is continuation of JRA-

25; *Onogi et al.* [2007]), the NCEP Climate Forecast System Reanalysis (CFSR; *Saha et al.* [2010]), and the NASA Modern Era Retrospective-Analysis for Research and Applications (MERRA; *Cullather and Bosilovich* [2011]). In general, these new reanalyses apply better horizontal and vertical resolution, better sea-ice and land-surface schemes, more extensive assimilation of satellite data, and more sophisticated assimilation methods. Although several recent studies have evaluated these new reanalyses in the Arctic (*Lüpkes et al.*, 2010; *Screen and Simmonds,* 2011; *Cullather and Bosilovich*, 2011; *Cuzzone and Vavrus*, 2011; Wilson et al., 2011), we are not aware of any study applying independent in situ data for the validation of the new reanalyses (for the older ones, see *Francis* (2002) and *Bromwich and Wang* (2005)). Hence, there is a strong need for such a study.

During the drift of the French schooner Tara in the central Arctic Ocean in 2007 the observations included tethersonde soundings (*Vihma et al.,* 2008), which have not been assimilated in any reanalysis. In this manuscript we apply these sounding data to validate the NCEP-DOE, NCEP-CFSR, ERA-Interim, JCDAS, and MERRA reanalyses. The somewhat older NCEP-DOE reanalysis is included in the study to quantify the progress made by NCEP-CFSR. We demonstrate large discrepancies between the five reanalyses, which all deviate a lot from observations, and discuss the reasons for and consequences of the errors and the remaining challenges.

2. Observations and data processing

In 2006–2007, meteorological, oceanographic, and sea ice measurements were made in the Arctic Ocean at the drifting ice station Tara [*Gascard et al*., 2008; *Vihma et al*., 2008] (Figure 1), which was a part of the European project DAMOCLES (Developing Arctic Modeling and Observation Capabilities for Long-Term Environmental Studies). From 25 April to 31 August 2007, a total of 95 tethersonde soundings during 39 sounding days were made. A Vaisala DigiCORA Tethersonde System was used to measure the vertical profiles of the air temperature, relative humidity, wind speed, and wind direction up to the height of 2 km. The measurement system

consisted of three tethersondes at approximately 15 m intervals in the vertical, attached to a tethered balloon. The balloon was ascended and descended with a constant speed of approximately $1 m\ s^{-1}$. The tethersonde system was only operated in non-precipitating conditions (except of very light snow fall) with wind speeds less than $15\ m\ s^{-1}$, and it was not ascended into thick clouds. The balloon was always lifted as high as the cloud conditions, wind speed, and the buoyancy of the $7\ m^3$ balloon allowed. The average of ascent and descent profiles of temperature, humidity, and wind direction were analysed. The data were averaged over the three tethersondes using a 20 m averaging interval. The average top height of the soundings was 1240 m. In this study, 29 profiles up to 890 m were used; to avoid giving excessive weight to data from days with a high sounding activity, we selected from each day only the profile that reached the highest altitude. Ten sounding days were omitted as no profiles reached the altitude of 890 m or the data quality was poor. The profiles selected for this study were measured between 09 and 18 UTC. The closest reanalysis output time was always selected for validation.

The reanalysis products validated were assimilated fields, i.e. the model first-guess fields corrected by assimilation of observations (when available). In MERRA, these are called as "the assimilated state". The other reanalysis archives include analysis and forecast fields; we validated the former, which are assimilated analyses as in the case of MERRA. All reanalysis products were horizontally linearly interpolated to Tara sounding sites. In the vertical, the reanalysis results were linearly interpolated from the reanalysis output levels to the sounding levels. In addition, the diagnostic reanalysis products for 2 m temperature and humidity and 10 m wind speed were validated. For all variables, the bias, root mean square error (RMSE) and correlation coefficient against observations were calculated, as well as the statistical significance of the bias and correlation in 95% confidence level. Correlation coefficients between observed and modeled air temperature and specific humidity were high, often exceeding 0.9, but these were due to the strong seasonal change from spring to summer, which was naturally captured by the reanalyses. Hence, for temperature and specific humidity, we only report correlations calculated using the 19 summer

4

soundings. We define summer as the period with the Tara 2 m air temperature above –1°C: from 9 June to 31 August (*Vihma et al.,* 2008).

3. Results

3.1. Air temperature

Considering the mean profile averaged over the 29 soundings, the 10 m temperature was –4°C and a temperature inversion was based at 70 m with the temperature increasing from –4.2 to –2.8°C by the height of 890 m (Figure 2a). None of the reanalyses was successful in capturing the shape of the temperature profile. ERA-Interim and MERRA performed very well above 200 m, but had a significant warm bias of up to 2.0°C at lower levels. NCEP-CFSR was very good in the lowermost 200 m layer, but had a significant cold bias above 400 m, whereas NCEP-DOE yielded a strong surface-based inversion, with a large significant warm bias peaking at the height of 100 m. Upward of the lowest prognostic model level, JCDAS yielded a linear temperature gradient of –5°C km^{-1}, which strongly deviates from observations.

Considering RMSE (Figure 2b), ERA-Interim performed best with RMSE ranging from 1.9 to 3.0°C. MERRA and the NCEP reanalyses were approximately equally good, but for MERRA the RMSE did not depend much on the height, whereas the NCEP reanalyses had a clearly smallest RMSE close to the surface and peak values, up to 4.8°C for NCEP-CFSR, in the layer of 500-600 m. JCDAS had the largest RMSE in the whole profile, ranging from 4.5 to 5.9°C. Surprisingly, the RMSE did not decrease even at 500 m, where the observed and JCDAS mean profiles crossed.

The summertime correlation coefficient (r) between reanalysis and observed temperature above 400 m was from 0.6 to 0.9 for all reanalyses (not shown). Below 400 m r was smaller, mostly from 0.4 to 0.8. For summer soundings ($n = 19$), $r > 0.46$ is significant at the 95% confidence level.

Some profiles of JCDAS and NCEP-DOE differed from the observed profiles by even more than 10°C (not shown). The largest warm errors occurred in spring with observed 2-m temperatures close to –15°C. The largest cold errors occurred in summer during strong temperature inversions

that were not captured by the models, some 400 m above the inversion base. There was positive

correlation ($r = 0.24$ to 0.40) between the extreme temperature errors and the profile average wind

speeds in all models, though only in ERA-Interim the correlation was statistically significant ($p =$

0.03). Most of the temperature errors larger than 7.5°C (20 cases of 23) occurred when the wind

speed averaged over the profile was larger than 6 m/s; all three exceptions were from the JCDAS

model.

We defined the temperature inversion base height, depth, and strength as in *Kahl* [1990]

using a threshold of 0.3°C for the temperature increase with height [*Vihma et al.,* 2011]. From the

29 measured profiles, 23 included a temperature inversion. NCEP-DOE captured the occurrence of

inversions (but not their strength and depth) fairly well with only 4 missing inversions and 2 false

inversions. ERA-Interim followed with 5 missing inversions and 3 false inversions. NCEP-CFSR

missed 7 inversions and presented 3 false inversions. MERRA missed 9 inversions and presented 2

false inversions whereas JCDAS missed 11 inversions and gave 1 false inversion.

3.2. Air humidity

Air specific humidity, averaged over the 29 soundings, did not vary much with height

(Figure 2c), but the relative humidity decreased with height from 90% at the lowermost 100 m to

75% at 890 m (Figure 2e). Among the reanalyses, basically only ERA-Interim reproduced the shape

of the specific humidity profile, but with a significant moist bias of 0.3 to 0.5 g kg^{-1} throughout the

profile, and missing the humidity inversion in the lowermost 180 m. ERA-Interim relative humidity

had a significant moist bias of up to 9% in the whole profile. The observed mean specific humidity

profile was best captured by NCEP-CFSR, with mostly dry insignificant biases of up to 0.3 g kg^{-1}.

The cold bias clearly dominated over the dry bias, seen as a mostly significant (i.e., significant at

most measurement heights) wet bias in the relative humidity of NCEP-CFSR. MERRA specific

humidity had significant dry bias above 150 m with the magnitude increasing with altitude to 0.5 g

kg^{-1}. MERRA relative humidity had a mostly significant dry bias of approximately 6% in the whole

profile. The mean specific humidity profiles of JCDAS and NCEP-DOE showed a mostly

significant moist bias in the lowermost 600 m and a strong decrease of air moisture upward of the

height of 100-150 m. JCDAS had a mostly significant positive bias in relative humidity but NCEP-

DOE values, vertically interpolated between the model levels, almost perfectly matched the

observations in the layer of 100 to 550 m. This was, however, due to the combined effects of warm

and moist (in the sense of specific humidity) bias.

Considering the specific humidity RMSE, NCEP-CFSR outperformed the other reanalyses

in the whole profile (Figure 2d), but for relative humidity, none of the reanalyses was clearly better

than the others (Figure 2f). Except of JCDAS specific humidity, the RMSE of specific and relative

humidity increased with height up to the altitude of at least 500 m.

The summertime correlation coefficient r for specific humidity (not shown) varied in NCEP-

CFSR and ERA-Interim from 0.4 to 0.7. JCDAS had a minimum r of 0.2 and MERRA even below

0.1. The worst model was NCEP-DOE with a negative r down to –0.5 above the height of 250 m.

Correlations for the relative humidity were generally smaller than for the specific humidity, r

typically ranging from 0.2 to 0.6.

In the individual profiles of specific humidity (not shown), ERA-Interim had the highest

positive errors, even exceeding 3 g kg^{-1}. In three cases the observed specific humidity was

approximately 2.5 g kg^{-1}, while the ERA-Interim specific humidity was approximately 6 g kg^{-1}. All

these cases were associated with a strong (8°C) temperature inversion. In these cases the

temperature profile was fairly well reproduced by ERA-Interim, but the reanalysis relative humidity

was excessive, close to 100% compared to the observed 35 to 60%. The highest negative errors in

specific humidity, less than –2 g kg^{-1}, occurred in NCEP reanalyses, associated with strong (8°C)

temperature inversions that were not captured by the reanalyses.

A layer with a specific humidity increase larger than 0.2 g kg^{-1} was considered as a

humidity inversion. The measured 29 profiles included 21 cases with a humidity inversion; the

maximum inversion strength was 2.6 g kg^{-1}. NCEP-DOE captured 15 inversions and 4 false

inversions. ERA-Interim captured 10 inversions and MERRA 9 inversions with no false inversions. NCEP-CFSR captured 7 inversions and 1 false inversion whereas JCDAS captured 4 inversions and 3 false inversions.

3.3 Wind speed

The observed wind speed averaged over the whole measurement period was 3.2 m s^{-1} at the height of 10 m, increasing to 5.6 m s^{-1} by the height of 130 m, and was gradually increasing further upwards to 6.2 m s^{-1} (Figure 2g). Considering the mean wind speed, averaged over the 29 soundings, ERA-Interim and JCDAS yielded statistically significantly too strong 10 m wind speed, whereas higher than 30 m MERRA and NCEP-CFSR had significantly too low wind speed. The mean wind speed profile was best captured by ERA-Interim and JCDAS with the magnitude of the negative bias smaller than 0.6 m s^{-1}. At all prognostic model levels, NCEP-DOE underestimated wind speed by 1 m s^{-1} and NCEP-CFSR and MERRA by 1.7–1.8 m s^{-1}. NCEP-CFSR and MERRA, however, outperformed the other reanalyses for the 10 m wind speed.

The RMSE for the 10 m wind speed was approximately 1.5 m s^{-1} for all reanalyses (Figure 2h). At higher levels ERA-Interim was clearly the best, followed by NCEP-DOE and JCDAS, while the new NCEP-CFSR and MERRA reanalyses were clearly the worst. The correlation coefficient against the observed wind speed (not shown) was best for ERA-Interim (vertical average 0.7) and worst for MERRA (r was 0.3–0.5 below 400 m but dropped to 0.1–0.2 above 500 m). The largest errors in individual profiles, both negative and positive, exceeded 5 m s^{-1} in magnitude in MERRA and NCEP-CFSR. A large overestimation of wind speed concurred with calm weather. A large underestimation coincided above the inversion base of strong temperature inversions that were not captured by the models.

3.4 All variables studied

To summarize the results we present a ranking of the reanalyses, with the bias, RMSE and correlation of air temperature, specific and relative humidity, and wind speed, vertically averaged over the 890 m layer (Table 1).

The observations covered spring and summer seasons, which differ, among others, from the point of view of lower boundary conditions for the atmosphere. Considering the air temperature, the biases and RMSEs were better for summer (except for CFSR), but correlations were generally better for spring. These results may be related to the prevailing near-zero surface temperatures in summer, the more variable temperatures in spring, and the capability of reanalyses to reproduce the synoptic-scale variations. Considering the humidity variables, the results were generally better for spring, seen for most reanalyses as better correlations and RMSEs of specific humidity and smaller biases of relative humidity. Summer conditions with a lot of low clouds, fog, and melt ponds seem a challenge for humidity analyses. The correlation coefficients for wind speed were better in spring for all reanalyses. We note, however, that the number of spring soundings used in the validation was small (10), which prevents us from firm conclusions.

4. Discussion and conclusions

The tethersonde observations should not be considered as representing the climatology of the study period and region. This is because the balloon could not be operated during winds stronger than 15 m s^{-1} and the balloon was not ascended into clouds. In any case, the data are suitable for reanalysis validation in conditions with neither low clouds nor very strong winds, and according to our knowledge represent the longest independent (not assimilated to models) data set of vertical profiles of wind, air temperature, and air humidity over the Arctic Ocean.

All reanalyses suffered from large errors in the vertical profiles of air temperature and humidity, and two reanalyses also had large errors in the wind speed profile. Combining the validation results for temperature, humidity and wind, ERA-Interim got the highest overall ranking. ERA-Interim outperformed the other reanalyses in the bias and RMSE for air temperature as well as

in the bias, RMSE, and correlation for the wind speed (Table 1). ERA-Interim got also the second place in reproducing the observed temperature and humidity inversions. Except of the inversions, ERA-Interim was not particularly successful estimating the air humidity. The main challenge for ECMWF is to get rid of the warm and moist biases. These have been long-lasting problems: already in 1990s ECMWF operational analyses and ERA-40 included warm and moist near-surface biases in the Arctic and Antarctic [*Beesley et al.*, 2000; *Curry et al.,* 2002; *Vihma et al.,* 2002; *Tastula and Vihma,* 2011].

Our results on the warm bias of up to 2°C in ERA-Interim in the lowermost 400 m layer is well in agreement with *Lüpkes et al.* [2010], who observed a warm bias of up to 1.7°C in the lowermost 300 m in ERA-Interim on the basis of R/V Polarstern rawinsonde soundings within 500 km of Tara in August 2007. *Lüpkes et al.* [2010] also observed a moist bias throughout the lowermost 1 km, which agrees well with our results. It is noteworthy that the R/V Polarstern soundings were assimilated into ERA-Interim. This suggests that the data assimilation system does not work effectively; even when observations are available, they are not well utilized, probably getting too low weight compared to the first-guess field (compare to the conclusions of *Atlaskin and Vihma* [2012]).

Both NCEP reanalyses and MERRA outperformed the other reanalyses with respect to 2-m air temperature and specific humidity and 10-m wind speed. This is an important result for those who apply reanalyses to provide atmospheric forcing for sea ice models in retrospective simulations. If one reanalysis should be selected, NCEP-CFSR is recommended on the basis of this study; it was among the best for all near-surface variables validated here (Figure 2). It should be remembered, however, that also radiative fluxes and precipitation, not validated in this study, are essential in the atmospheric forcing for sea ice. As the near-surface variables depend on a complex interaction of various processes, it is very difficult to evaluate what is the reason for the success of NCEP-CFSR. We only note that this reanalysis includes a comparably sophisticated treatment of sea ice, including its fractional coverage and prognostic ice and snow thickness [*Saha et al.*, 2010].

The difficulties in improving reanalyses are demonstrated by the fact that at the prognostic model levels, the older NCEP-DOE got the second highest overall ranking (Table 1) and outperformed the new NCEP-CFSR for the wind speed and, close to the height of 800 m, also for the air temperature. NCEP-DOE was the best reanalysis capturing both temperature and humidity inversions, though the model had the most sparse vertical resolution. NCEP-CFSR had a cold bias up to 2°C but was clearly the best reanalysis for specific humidity. NCEP-CFSR also had a large bias, -1.7 m s^{-1}, in the wind speed. Further, the new MERRA reanalysis suffered from serious problems in air humidity and wind speed, being the driest reanalysis with weakest winds. Moreover, MERRA had clearly lowest ($r = 0.1–0.2$) correlations with the observed wind speed; these occurred from 500 m upwards, where the other reanalyses had correlation coefficients of 0.35 to 0.7.

JCDAS reanalysis suffered from poor temperature and humidity profiles. JCDAS was the weakest model in capturing temperature and humidity inversions. The average temperature profile was close to moist-adiabatic, which suggests that the boundary layer scheme yields too much mixing. JCDAS results for the wind speed were, however, almost as good as those of ERA-Interim.

An interesting aspect in the validation results was that the largest air temperature errors did not occur in conditions of very stable stratification, which is usually the case [*Atlaskin and Vihma*, 2012], but in conditions of higher-than-average wind speeds. This may be related to the large role of lateral advection in controlling the air temperature variability over the Arctic Ocean, especially in spring and summer 2007 [*Graversen et al.*, 2011].

The observed biases in temperature, humidity, and wind speed are in many cases comparable or even larger than the climatological trends during the latest decades [*Serreze et al.*, 2009]. This calls for caution when applying reanalysis data in climatological studies. A good aspect in reanalysis is that the model and the data assimilation systems keeps the same throughout the reanalysis period, but changes occur in the availability of observations, and these may result in wrong conclusions on the trends, as demonstrated e.g. by *Screen and Simmonds* [2011].

References

Atlaskin, E., and T. Vihma (2012), Evaluation of NWP results for wintertime nocturnal boundary layer temperatures over Europe and Finland, *Q. J. R. Meteorol. Soc.*, doi:10.1002/qj.1885.

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.

Bromwich, D. H., and S.-H. Wang (2005), Evaluation of the NCEP-NCAR and ECMWF 15- and 40-yr reanalyses using rawinsonde data from two independent Arctic field experiments, *Mon. Weather Rev.*, 133, 3562–3578.

Cullather, Richard I., and M. G. Bosilovich (2011), The Moisture Budget of the Polar Atmosphere in MERRA, *J. Climate*, 24, 2861–2879, doi: http://dx.doi.org/10.1175/2010JCLI4090.1.

Curry, J. A., J. L. Schramm, A. Alam, R. Reeder, T. E. Arbetter, and P. Guest (2002), Evaluation of data sets used to force sea ice models in the Arctic Ocean, *J. Geophys. Res.*, 107, NO. C10, 8027, 10.1029/2000JC000466.

Cuzzone, J., and S. Vavrus (2011), The relationships between Arctic sea ice and cloud-related variables in the ERA-Interim reanalysis and CCSM3, *Environ. Res. Lett.*, 6, 014016, doi:10.1088/1748-9326/6/1/014016.

Dee, D. P., *et al.* (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, 137: 553–597.

Francis, J. A. (2002), Validation of reanalysis upper-level winds in the Arctic with independent rawinsonde data, *Geophys. Res. Lett.*, 29(9), 1315, doi:10.1029/2001GL014578.

Gascard, J.-C., *et al*. (2008), Exploring Arctic transpolar drift during dramatic sea ice retreat, *Eos Trans. AGU*, 89, doi:10.1029/2008EO030001.

Gibson, J. K., P. Kållberg, S. Uppala, A. Nomura, A. Hernandez, and E. Serrano (1997), *ERA Description, ECMWF ERA-15 Project Report Series*, NO. 1., ECMWF, Reading, UK (available from www.ecmwf.int/publications).

Graversen, R. G., T. Mauritsen, S. Drijfhout, M. Tjernström, and S. Mårtensson (2011), Warm winds from the Pacific caused extensive Arctic sea-ice melt in summer 2007, *Clim. Dyn.*, 36, 2103–2112, DOI 10.1007/s00382-010-0809-z.

Jakobson, E., and T. Vihma (2010), Atmospheric moisture budget over the Arctic on the basis of the ERA-40 reanalysis, *Int. J. Climatol.*, 30, 2175–2194, DOI: 10.1002/joc.2039.

Kalnay, E., *et al*. (1996), The NCEP/NCAR 40-Year Reanalysis Project, *Bulletin of the American Meteorological Society*, 77 (3), 437–471.

Kahl, J.D. (1990), Characteristics of the low-level temperature inversion along the Alaskan Arctic coast, *Int. J. Climatology*, 10, 537–548.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter (2002), NCEP-DOE AMIP-II Reanalysis (R-2), *Bull. Amer. Met. Soc.*, 83, 1631–1643.

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, *Geophys. Res. Lett.*, 37, L09810, doi:10.1029/2010GL042724.

Onogi, K., *et al*. (2007), The JRA-25 Reanalysis, *J. Meteor. Soc. Japan*, 85, 369–432.

Rinke A., W. Maslowski, K. Dethloff, and J. Clement (2006), Influence of sea ice on the atmosphere: a study with an Arctic atmospheric regional climate model, *J. Geophys. Res.*, 111, D16103.

Saha, *et al*. (2010), The NCEP Climate Forecast System Reanalysis, *Bull. Amer. Meteor. Soc.*, 91, 8, 1015–1057.

Screen, J. A., and I. Simmonds (2011), Erroneous Arctic Temperature Trends in the ERA-40 Reanalysis: A Closer Look, *J. Clim.*, 24, 2620–2627.

Serreze, M. C., A. P. Barrett, A. G. Slater, R. A. Woodgate, K. Aagaard, R. B. Lammers, M. Steele, R. Moritz, M. Meredith, and C. M. Lee (2006), The large-scale freshwater cycle of the Arctic, *J. Geophys. Res.*, 111, C11010, doi:10.1029/2005JC003424.

Serreze, M. C., A. P. Barrett, A. G. Slater, M. Steele, J. Zhang, and K. E. Trenberth (2007), The large-scale energy budget of the Arctic, *J. Geophys. Res*., 112, D11122, doi:10.1029/2006JD008230.

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

Tastula, E.-M., and T. Vihma (2011), WRF model experiments on the Antarctic atmosphere in winter, *Mon. Wea. Rev*., 139, 1279–1291, doi:10.1175/2010MWR3478.1.

Thompson, D. W. J., and J. M. Wallace (1998), The Arctic Oscillation signature in the wintertime geopotential height and temperature fields, *Geophys. Res. Lett*., 25, 1297–1300.

Tjernström, M., and R. G. Graversen (2009), The vertical structure of the lower Arctic troposphere analysed from observations and ERA-40 reanalysis, *Q. J. Roy. Meteor. Soc*., 135, 431–443.

Uppala, S.M., *et al*. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc*., 131, 2961–3012.

Vihma, T., J. Uotila, B. Cheng, and J. Launiainen (2002), Surface heat budget over the Weddell Sea: buoy results and comparisons with large-scale models, *J. Geophys. Res.,* 107 (C2), 3013, doi: 10.1029/2000JC00037.

Vihma, T., J. Jaagus, E. Jakobson, and T. Palo (2008), Meteorological conditions in the Arctic Ocean in spring and summer 2007 as recorded on the drifting ice station Tara, *Gephys. Res. Lett.*, 35, L18706, doi: 10.1029/2008GL034681.

Vihma, T., T. Kilpeläinen, M. Manninen, A. Sjöblom, E. Jakobson, T. Palo, J. Jaagus, and M. Maturilli (2011), Characteristics of temperature and humidity inversions and low-level jets over Svalbard fjords in spring, *Advances in Meteorology*, Volume 2011, Article ID 486807, 14 p., doi:10.1155/2011/486807.

Walsh, J. E., and W. L. Chapman (1998), Arctic cloud–radiation–temperature associations in observational data and atmospheric reanalyses, *J. Climate,* 11, 3030–3045.

Wilson, A. B., D. H. Bromwich, and K. M. Hines (2011), Evaluation of Polar WRF forecasts on the Arctic System Reanalysis domain: Surface and upper air analysis, *J. Geophys. Res*., 116, D11112, doi:10.1029/2010JD015013.

Figures

Figure 1. Drift track of TARA during the period of tethersonde soundings, from 25 April to 31 August, 2007. The red line shows the September minimum sea ice extent.

Figure 2. From 29 single profiles calculated: a) average temperature, b) RMSE (root mean square error) of temperature, c) average specific humidity, d) RMSE of specific humidity, e) average relative humidity, f) RMSE of relative humidity, g) average wind speed, h) RMSE of wind speed.

Table 1. Vertically averaged values of the magnitude of bias, RMSE, and correlation coefficient of air temperature (*Ta*), specific humidity (*Qa*), relative humidity (*RH*), and wind speed (*V*). Correlation coefficients for air temperature and specific humidity were calculated on the basis of 19 soundings in summer, but for all other parameters on the basis of the whole period of 29 soundings. In the ranking the best reanalyses is indicated by 5 points and the worst by 1 point.

	ERA-Interim		NCEP-DOE		NCEP-CFSR		MERRA		JCDAS	
	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank
Ta |bias|	0.51	5	1.17	3	1.36	2	0.63	4	1.42	1
Ta RMSE	2.61	5	3.31	3	3.53	2	3.15	4	5.30	1
Ta Correl	0.74	4	0.79	5	0.70	2	0.74	3	0.63	1
Qa |bias|	0.40	1	0.20	4	0.14	5	0.33	2	0.25	3
Qa RMSE	0.75	4	0.81	2	0.54	5	0.75	3	0.81	1
Qa Correl	0.56	4	-0.17	1	0.58	5	0.33	2	0.47	3
RH |bias|	6.89	2	2.35	5	5.82	3	5.26	4	8.68	1
RH RMSE	15.7	3	15.9	2	15.3	5	15.4	4	16.8	1
RH Correl	0.41	3	0.48	4	0.29	2	0.52	5	0.23	1
V |bias|	0.43	5	0.90	3	1.69	2	1.85	1	0.47	4
V RMSE	1.80	5	2.03	4	2.70	2	2.91	1	2.20	3
V Correl	0.71	5	0.59	4	0.44	2	0.28	1	0.52	3
Total points		46		40		37		34		23