JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 114, D13104, doi:10.1029/2008JD010790, 2009

The performance of a global and mesoscale model over the central Arctic Ocean during late summer

C. E. Birch,[1] I. M. Brooks,[1] M. Tjernström,[2] S. F. Milton,[3] P. Earnshaw,[3] S. Söderberg,[4] and P. Ola G. Persson[5]

Received 17 July 2008; revised 22 April 2009; accepted 12 May 2009; published 10 July 2009.

[1] Measurements of turbulent fluxes, clouds, radiation, and profiles of mean meteorological parameters, obtained over an ice floe in the central Arctic Ocean during the Arctic Ocean Experiment 2001, are used to evaluate the performance of U.K. Met Office Unified Model (MetUM) and Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS) in the lower atmosphere during late summer. Both the latest version of the MetUM and the version operational in 2001 are used in the comparison to gain an insight as to whether updates to the model have improved its performance over the Arctic region. As with previous model evaluations over the Arctic, the pressure, humidity, and wind fields are satisfactorily represented in all three models. The older version of the MetUM underpredicts the occurrence of low-level Arctic clouds, and the liquid and ice cloud water partitioning is inaccurate compared to observations made during SHEBA. In the newer version, simulated ice and liquid water paths are improved, but the occurrence of low-level clouds are overpredicted. Both versions overestimate the amount of radiative heat absorbed at the surface, leading to a significant feedback of errors involving the surface albedo, which causes a large positive bias the surface temperature. Cloud forcing in COAMPS produces similar biases in the downwelling shortwave and longwave radiation fluxes to those produced by UM(G25). The surface albedo parameterization is, however, more realistic, and thus, the total heat flux and surface temperature are more accurate for the majority of the observation period.

Citation: Birch, C. E., I. M. Brooks, M. Tjernström, S. F. Milton, P. Earnshaw, S. Söderberg, and P. O. G. Persson (2009), The performance of a global and mesoscale model over the central Arctic Ocean during late summer, *J. Geophys. Res.*, *114*, D13104, doi:10.1029/2008JD010790.

1. Introduction

[2] Recent evidence has shown that temperatures in the Arctic are rising at almost twice the rate of the global average [*Solomon et al.*, 2007] and that this increase corresponds to a decrease in both sea ice thickness and extent [*Parkinson et al.*, 1999; *Nghiem et al.*, 2007; *Comiso et al.*, 2008]. This trend is predicted to continue and probably increase in the future [*Holland et al.*, 2006], and is partly due to processes such as the ice albedo feedback [*Curry et al.*, 1996]. Processes that occur in the Arctic are linked to both global ocean and atmospheric circulation [*Graversen*, 2006] and thus changes to the climate system over the central Arctic Ocean are expected to have a major impact elsewhere. For example, *Chapman and Walsh* [2007] suggest that a decrease in sea level pressure over the Bering Strait could cause a northward shift in the Pacific storm track, impacting the nearby coastal areas. It is therefore essential that both the present and future climate in the Arctic and its effect on global circulation can be simulated accurately by global and regional scale models.

[3] Multimodel averages currently produce the most confident next-century predictions of Arctic climate; however, there are large differences between individual model predictions, especially those related to the magnitude and spatial patterns of the warming [*Holland and Bitz*, 2003; *Serreze and Francis*, 2006] and to the extent and timing of the reduction in sea ice [*Arzel et al.*, 2006]. It has been suggested that this warming, along with the ice albedo feedback could produce abrupt reductions in summer Arctic sea ice [*Holland et al.*, 2006]. There is disagreement between models as to the timing of these events which is at least partly due to the varying progression of the warming in each model [*Serreze and Francis*, 2006].

[4] The same model inaccuracies reoccur in many of the model intercomparisons and evaluation studies of present-day conditions over the Arctic Ocean. Generally, the basic meteorological fields (pressure, temperature and winds) are

[1]Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds, UK.
[2]Department of Meteorology, Stockholm University, Stockholm, Sweden.
[3]Global Model Development and Diagnostics, Met Office, Exeter, Devon, UK.
[4]WeatherTech Scandinavia, Uppsala, Sweden.
[5]Cooperative Institute for Research in Environmental Sciences, University of Colorado/NOAA/Earth System Research Laboratory, Boulder, Colorado, USA.

0148-0227/09/2008JD010790$09.00

the most satisfactorily represented [*Tjernström et al.*, 2005; *Rinke et al.*, 2004], although even these variables often show some bias [e.g., *Chapman and Walsh*, 2007]. Models have been found to perform especially badly during the summer melt season [*Randall et al.*, 1998; *Makshtas et al.*, 2007], where surface heat fluxes show very little correlation to observations and the latent heat fluxes in particular are overestimated in most models [*Brunke et al.*, 2006; *Tjernström et al.*, 2005]. The other major issue with both regional and global models is their representation of clouds. There are problems with both simulated cloud occurrence and extent and with cloud optical and microphysical properties [*Tjernström et al.*, 2008]. This has consequences for other model-produced variables, most notably the surface heat fluxes and the radiation balance [*Tjernström et al.*, 2005; *Randall et al.*, 1998; *Walsh et al.*, 2002]. Since the existence of sea ice depends significantly on heat exchange between the surface and the atmosphere, it is vital to accurately represent these smaller scale processes to accurately predict future atmospheric and sea ice changes.

[5] The central Arctic Ocean is a unique environment, with a surface consisting of sea ice and open leads and which experiences near constant daylight during the summer months and darkness during the winter. In situ observations of the Arctic boundary layer were made during the Arctic Ocean Experiment (AOE) 2001 [*Tjernström et al.*, 2004a] and the Surface Heat Budget of the Arctic Ocean Experiment (SHEBA [*Uttal et al.*, 2002]). The summer boundary layer was found to be sometimes weakly stable [*Persson et al.*, 2002] but often well-mixed through its upper part and cloud layer, with a shallow stable surface layer [*Tjernström et al.*, 2004a]. Near-surface temperatures were relatively constant, between -1.7 and 0°C, due in large part to latent heat processes that act as a buffer against energy entering or leaving the surface. The near-surface humidity is high and always near ice saturation due to the high emission rate of water vapor from open leads compared with the low rate of removal by the ice surface [*Andreas et al.*, 2002]. The lower atmosphere is therefore most often cloudy, with a stratus cloud base commonly at around 100 m [*Tjernström et al.*, 2004a]. Strong capping inversions sometimes occur due to the advection of warm and relatively humid air aloft. Contrary to behavior at lower latitudes, it is possible that this also contributes to the high near-surface humidity and to cloud development and persistence in the boundary layer because entrainment will act as a source of boundary layer moisture [*Pinto*, 1998]. Multiple cloud layers with a temperature inversion associated with each of them, are also common [*Intrieri et al.*, 2002; *Tjernström et al.*, 2004a]. Cloud top was often found within the inversion, rather than below it, which is in contrast to low-latitude marine stratocumulus, where cloud top sits at the base of the inversion [*Tjernström*, 2005].

[6] This study uses surface observations and some surface-based remote sensing observations from AOE 2001 to evaluate the lower atmosphere simulated in a global and mesoscale model during the late summer melt/early freezeup period over the central Arctic Ocean. It aims to identify problems that occur in each of the models, especially those relating to their representation of the surface heat and radiative fluxes and clouds. Section 2 introduces the observational data set and section 3 describes the global and mesoscale models. There is a comparison of model diagnostics with the observations in section 4, including evaluations of the basic meteorological fields, surface turbulent and radiative fluxes and cloud occurrence. A case study in section 4.6 is used to investigate the errors found in cloud radiative forcing in more detail.

2. Observations

[7] The Arctic Ocean Experiment (AOE) 2001 [*Tjernström et al.*, 2004a] took place in a region of drifting pack ice between 88 and 89°N, 2–21 August 2001, on the Swedish icebreaker Oden. An 18 m meteorological mast was positioned on a large floe in the pack ice (1.5 × 3 km), approximately 300 m from the ship and 500 m from the nearest open leads. The micrometeorological data set includes mean profile measurements of wind speed at 5 levels (1.7, 3.4, 7.1, 12.9 and 17.3 m), humidity and air temperature at 2 levels (3.6 and 14.5 m) and wind direction at one level (18 m). High frequency measurements of the turbulent wind components and temperature were made using Gill sonic anemometers at heights of 4.7 and 15.4 m and of water vapor using Krypton hygrometers at heights of 3.6 and 14.5 m.

[8] Longwave and shortwave upwelling and downwelling radiation fluxes were measured at two sites during the field campaign. The first set of observations were made using Eppley pyranometers and pyrgeometers, which were situated on the ice near the meteorological mast and made measurements for the duration of the field campaign. A second set of shortwave radiation measurements, using Kipp and Zonen CM11 pyranometers were made periodically over an undisturbed snow surface on the pack ice halfway between the ship and the meteorological mast. All radiation measurements presented here, apart from the albedo observations and the upwelling shortwave radiation flux (discussed below) were made using the first set of sensors.

[9] The sensor measuring upwelling shortwave radiation from the first set of instruments did not work for the entire campaign so a polynomial function was fitted to a time series of surface albedo derived from the second set of shortwave radiation measurements (Figure 1). The upwelling shortwave radiation flux was then computed from this and the continuous downwelling shortwave radiation flux observations from the first set of instruments. Such an estimate is a potential source of error in the value of the observed upwelling shortwave radiation flux, SW_{up} and also in the net shortwave radiation, SW_{net} and net radiation, Rad_{net} fluxes. This is discussed further in section 4.5.

[10] The turbulence data sets from the meteorological mast are limited due to instrument problems during the field campaign. The turbulent winds were the least affected but the sonic temperature measurements suffered from contamination most likely caused by water droplets formed by condensation on the transducer heads. Water vapor measurements also suffered from condensation on the optical hygrometers. Rigorous checks were made to ensure data were used only from periods where there is high confidence it is uncontaminated. Firstly, a visual check of the time series was made and obvious periods of instrument failure and any erroneous, single outlying points were

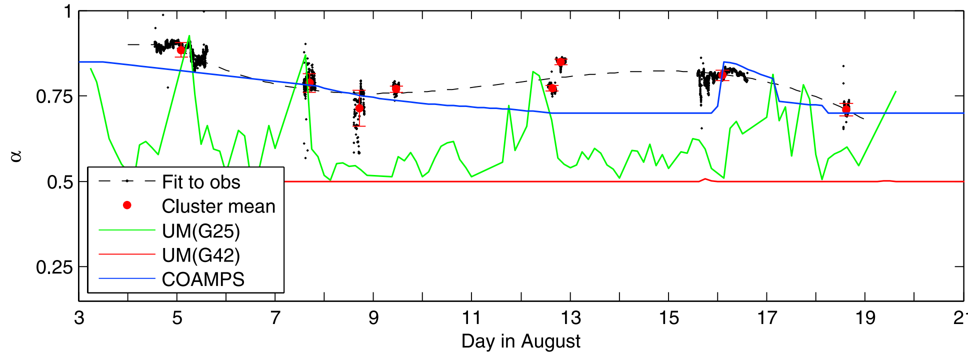

Figure 1. One minute averaged measurements of surface albedo made over an undisturbed snow surface and three hourly averaged model diagnostics. The measurements were made using the second set of upward and downward facing pyranometers, and a polynomial function is fitted to the time series. The mean of each cluster of data points and an error bar ±1 standard deviation about each mean is also shown.

removed. Corrections for crosswind contamination of the sonic temperature were made following *Schotanus et al.* [1983] and oxygen corrections to the water vapor measurements following *van Dijk et al.* [2003]. Eddy covariance fluxes of sensible heat, H, latent heat, E and the friction velocity, u_* were then estimated with a 30 minute averaging period. Measurements that were made when the instruments were downwind of the mast were removed from the data set. A flux footprint model [*Horst and Weil*, 1992] was used to determine that over 90% of the total flux is representative of a region of the ice surface that is less than 300 m from the mast. This suggests that the ship and open leads should have a very limited impact on the flux data set and it should therefore represent a surface covered almost entirely by pack ice.

[11] Additional measurements included surface pressure and 6 hourly radiosonde measurements of water vapor, pressure, temperature, and wind velocity up to 12 km. Three ISFF (Integrated Surface Flux Facility) stations were deployed on the ice, which made additional turbulence measurements at 3 m and air temperature, wind speed, humidity and pressure measurements at 2 m. Two of these (ISFF 1 and ISFF 2) were located on separate ice floes to the main ice camp, approximately 7 and 9 km from the ship, forming a rough triangle with Oden. ISFF 3 was located 1.5 km from the ice camp, near an open lead. The measurements made by the CSI ultrasonic anemometers and Krypton hygrometers suffered far less from the problems experienced by those on the main meteorological mast. The turbulent fluxes were computed in the same way as described above. There was an array of remote sensing instruments making continuous measurements, including a sodar to measure wind speed, direction and boundary layer structure, a ceilometer to measure cloud base and an S-band Doppler radar to observe clouds and precipitation (see *Tjernström et al.* [2004b] for further details).

3. Models

3.1. Met Office Unified Model

[12] The Met Office Unified Model (MetUM) is a fully coupled ocean-atmosphere numerical model that supports both global and regional domains [*Davies et al.*, 2005; *Staniforth et al.*, 2006]. It can be run on many temporal scales, making it suitable for both numerical weather prediction (NWP) and climate modeling. Although it is arguably more important to simulate the Arctic region accurately on a climate timescale, the NWP version of the MetUM is used in this study since there are a number of advantages in using this framework to infer systematic errors in the parameterizations of climate models [*Phillips et al.*, 2004]. Firstly, the NWP short-range (12–36 hour) forecasts are run from initial states generated with state-of-the-art variational data assimilation [e.g., *Lorenc et al.*, 2000]. There are very few observations available for assimilation over the Arctic region but those that do exist minimize errors in the large-scale synoptic flow. In addition, there are no large biases in the circulation due to remote forcing effects (e.g., tropical/extratropical/polar interactions). Such remotely forced biases in the circulation of a climate model make it difficult to ascribe errors to specific parameterized physical processes. While ascribing errors is still nontrivial in NWP models, detailed observational data sets from field campaigns, such as in this study, can be used to evaluate the physical processes at the scale of individual weather systems. Data from radiosondes launched from Oden were assimilated into the MetUM forecasts via the Global Telecommunications System. The result of this is that the validation data set is not independent of the forecast diagnostics but it does however minimize errors in global circulation, allowing the focus of the model evaluation to be the parameterized processes. The MetUM is well placed to take advantage of this approach as the climate model (HadGEM1) and the global NWP version (G42) have a very similar dynamical and physical formulation [*Martin et al.*, 2006].

[13] Outputs from both the latest version (G42) of the global NWP model, and the version operational during 2001 (G25) are used to help determine whether updates to the model physics since 2001 have improved the simulation of the Arctic region. UM(G25) has a dynamical core based on the Eulerian and hydrostatic formulation described by *Cullen and Davies* [1991] and the physical formulation was similar to the HadAM3 climate version [*Pope et al.*, 2000]. Data sets from this version of the model are comprised of 12-hour operational forecasts, initialized from 00 UTC and 12 UTC analyses, sampled at 3-hour forecast intervals (t + 3, 6, 9, 12 hours) and cover the entire August

ice drift observation period. 3 hourly diagnostic data from every 12 hour forecast are concatenated to produce a continuous data series from 3 to 20 August.

[14] Since 2001 both the NWP and climate versions of the MetUM have undergone a large number of developments up until NWP model cycle G42 discussed in this study. The Eulerian/hydrostatic dynamical core has been replaced by a semi-Lagrangian, semi-implicit and nonhydrostatic formulation [*Davies et al.*, 2005] and many of the physical parameterizations have been updated [*Allan et al.*, 2007]. In addition, the 3D-Var (three-dimensional variational) data assimilation system [*Lorenc et al.*, 2000] has been replaced by a 4D-Var system [*Rawlins et al.*, 2007]. The operational global NWP horizontal resolution for the UM(G42) version is 0.375° latitude by 0.5625° longitude, but was run here at the same horizontal resolution as UM(G25), 0.56° by 0.83° to simplify the comparison. UM(G42) was run for the 2001 observation period, with initial conditions provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) 40 year reanalyses (ERA-40). The model forecast fields are output at 15 minute intervals out to 4 days. The second day of each forecast has been assembled in a similar way to the data in UM(G25) to obtain a continuous data set for the observation period. Using the second day of each UM(G42) forecast allows time for the necessary spinup after model initialization but keeps accumulated model errors to a minimum, allowing for optimum comparison with the older model version. In contrast, the UM(G25) data sets are from operational forecasts for which no spinup time is required due to the ongoing nature of the forecast and data assimilation cycle.

[15] Although the same horizontal grid resolution is used with both versions of the model, the vertical resolution in UM(G42) is much greater: 12 vertical levels in the lowest 3 km of the atmosphere, where the first 3 are at 10, 50 and 130 m, compared to UM(G25), which has 6 levels. Vertical grid box height is defined in pressure levels in this version of the model, the first 3 roughly translate to a few meters above the surface, 330 and 530 m. Observations from all over the globe were assimilated into the ERA-40 and 2001 MetUM analyses used to initialize the forecasts. Because there are only a very limited number of observation sites in the Arctic region, radiosondes from the AOE 2001 field campaign were submitted to the Global Telecommunications System during the field campaign and were thus utilized in the ERA-40 and 2001 MetUM analyses.

[16] The radiation scheme used in UM(G25) is described by *Slingo and Wilderspin* [1986] and *Slingo* [1989]. The cloud scheme uses a prognostic method, where both cloud ice and water contents are diagnosed from the relative humidity [*Smith*, 1990]. An improved radiation scheme based on the two-stream equations in both the longwave and shortwave spectral regions was introduced into UM(G42) following *Edwards and Slingo* [1996]. This allows for consistency in physical processes that are important in both spectral regions, such as overlapping cloud layers. It includes the treatment of the effects of nonspherical ice particles and allows multiple scattering between cloud layers. The cloud scheme in UM(G42) remains based on that by *Smith* [1990] but a cloud/precipitation microphysical scheme with prognostic ice was introduced [*Wilson and Ballard*, 1999], based on that by *Rutledge and Hobbs* [1983]. Cloud ice water content is now advected, although cloud water content is still determined from a diagnostic relationship with relative humidity.

[17] Both versions of the MetUM use a boundary layer scheme based on Monin-Obukhov similarity theory and surface fluxes computed following *Louis* [1979]. The surface roughness length of momentum, z_m, is set at a constant value of 0.003 m and it is assumed the surface roughness lengths of heat, z_h and humidity, z_q, are equal to $z_m/10$. The surface albedo in both versions of the model depends on the surface temperature. When the ice surface temperature is at its maximum (273.15 K) the albedo is 50% and this increases to a maximum of 80% as the surface temperature decreases to 263 K. Although the MetUM is a fully coupled ocean-atmosphere model, both NWP versions used here have fixed sea ice fractions over each forecast period. This far north both versions of the model assume 100% sea ice cover. It is only in the marginal ice zone that an open lead fraction is simulated. Sea ice thickness is also constant, at 2 m.

3.2. COAMPS

[18] The Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS) was developed by the Naval Research Laboratory, USA [*Hodur*, 1997; *NRL*, 2003]. It was run here with an outer domain covering the whole pan-Arctic region, including the marginal ice zone and some open water and land. The outer domain had a resolution of 54 km while two inner domains were nested at 18 and 6 km resolution respectively. The innermost domain was centered around the AOE 2001 observation locations. All domains had the same vertical grid; 45 vertical model levels in the lowest 3 km of the atmosphere, with the first three levels at 3, 10 and 17 m.

[19] The fluxes at the surface were modeled with a surface energy balance model adapted to sea ice conditions. It is based on a simple force-restore concept with a fixed ice thickness of 2 m using a "deep layer temperature" fixed at the freezing point of seawater, −1.7°C. Ice cannot melt or accumulate in the model but ice extent and fraction were updated every 24 hours during the model run from satellite observations. In the grid boxes as far north as the observation site, the surface is completely covered in ice, with no open lead fraction. The boundary layer turbulence scheme is based on the work of *Mellor and Yamada* [1974] and the surface turbulent fluxes are computed using a bulk Richardson number, based on the formulations presented by *Louis* [1979]. z_m is set at a constant value of 1.4×10^{-5} m and like the MetUM, it is assumed $z_h = z_q = z_m/10$.

[20] At the surface a simplified snow model is applied, with a skin-surface temperature parameterization. A fraction of any melted snow water is retained as liquid inside the snow layer and is allowed to refreeze if the bulk snow temperature sinks below 0°C. Snow albedo is set with a base value of 70% and a top value of 85%. At each new snowfall, the surface albedo is reset to the top value and is then relaxed back to the base value with a relaxation time of a few days during the melt conditions. Each grid point is either ice covered or open water, which was specified using Special Sensor Microwave Imager (SSM/I) satellite data.

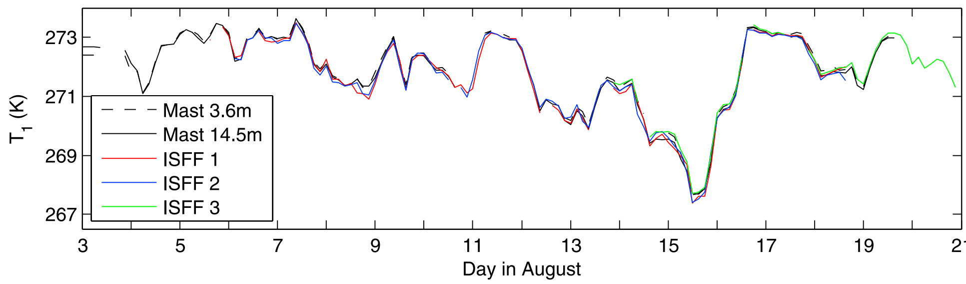

Figure 2. Three hourly averages of near-surface air temperature observations from the meteorological mast and the three ISFF stations.

[21] The moist microphysics scheme is based on one developed by *Rutledge and Hobbs* [1983] and consists of a bulk cloud microphysical model [*Lin et al.*, 1983] and a single-moment prediction of mixing ratio for 5 microphysical variables (vapor, pristine ice, snow, rain and cloud water). The size distribution of *Marshall and Palmer* [1948] is used, along with Kessler autoconversion [*Kessler*, 1969] and the Fletcher formulation for nucleation of pristine ice [*Fletcher*, 1962]. The radiation scheme performs both longwave and shortwave transfer calculations, based on the work of *Harshvardhan et al.* [1987].

[22] The outermost domain was forced by ERA-40 reanalysis data, which has a resolution of 1.5° latitude and 1.5° longitude. In contrast to the MetUM model runs, COAMPS was run in a "climate mode". The simulation, covering the entire AOE 2001 ice drift period, was run without any constraints from assimilation of observational data, except for that contained in the ERA-40 data used at the outermost boundary; it should be noted that the ERA-40 data does include the assimilation of the AOE 2001 radiosonde observations. With an outer domain covering the entire Arctic Ocean it is expected that the exact development of the atmospheric circulation will deviate more from the observations than those in the MetUM simulations. Systematic model errors present in all models are here allowed to fully develop over time and the chaotic nature of the atmospheric system ensures conditions well away from the lateral boundaries of the outermost domain deviate from reality. It is important to realize that such differences need not be erroneous in a physical sense but are an expression of the stochastic nature of the atmosphere. Due to these differences, the relative success of how the MetUM and COAMPS capture individual events cannot be assessed with confidence. Statistical comparisons however are useful, since biases over a longer period of time indicate fundamental differences in the model climates. It is more informative to compare the versions of the MetUM since these data sets were produced in a much more similar way and a comparison will give insight into whether the recent updates to the MetUM have increased its accuracy in the Arctic region.

4. Model Evaluation

4.1. Introduction to Evaluation

[23] When evaluating either global or regional scale models against observations a comparison of single point observations must be made with grid box averaged model diagnostics. Some care must be taken interpreting such comparisons since, for at least some of the variables, the two may represent rather different physical properties. The main meteorological mast was located on a large ice floe, 300 m from the open water around Oden and a significantly larger distance from open leads in all other directions. All observations discussed here were made either on or near the mast, apart from measurements made by the ceilometer and S-band radar, which were located on board Oden and by the ISFF stations, which were made on separate ice floes. The observations will represent conditions over the local pack ice, rather than conditions averaged over a region the size of a grid box, which will in reality contain a fraction of open leads. Compared to the pack ice, open leads can be a significant source of moisture, meaning conditions in their immediate vicinity can be quite different to those over the ice. Having said that, the ice and lead temperatures during August are much more similar than at other times of the year and the Arctic sea ice is relatively homogeneous compared to land surface types at lower latitudes. Figure 2 shows near-surface air temperature, T_1, measurements from the main meteorological mast and the three ISFF stations, one of which was located next to an open lead. This shows air temperature did not vary significantly over small distances on the main ice floe or between the middle and the edges of the ice floe. In addition to this, none of the three models include an open lead fraction in grid boxes this far north. Consequently, providing these issues are appreciated, it seems adequate to compare the observations and models in this way.

[24] Figure 2 also shows there is negligible difference between the air temperature measurements at 3.6 and 14.5 m on the mast. Since the measurements made at the upper mast height are more continuous, they are used in comparisons with modeled T_1, which refers to 1.5 m and 3 m above the surface in the MetUM and COAMPS respectively. The above is also true of the humidity measurements (not shown). Observed 10 m wind speed was derived by interpolation of the wind speed at the 7.1 and 12.9 m measurement levels for comparison with the models' wind speed at this height.

[25] The albedo and radiation measurements were made over undisturbed snow on the pack ice and therefore do not fully represent a region of sea ice on spatial scales the size of a model grid box, which includes a fraction of open leads and melt ponds. Without further radiation measurements over these various surface types, the effect of an open water fraction on the surface albedo is difficult to quantify. For this study however, both models assume the sea ice fraction is 100% at 88–89°N and thus evaluating the model data

Table 1. Statistics of Model Diagnostics Compared to Observations Using Three Hourly Averages[a]

			UM(G25)				UM(G42)				COAMPS			
	Unit	$\bar{x}_{obs}$	a.b.	σ	R	IoA	a.b.	σ	R	IoA	a.b.	σ	R	IoA
p	hPa	1004.16	0.24	1.25	0.98	0.99	0.52	1.70	0.97	0.98	−2.21	3.31	0.86	0.89
U_{10m}	m s^{-1}	4.39	−0.56	1.32	0.71	0.81	−0.16	1.62	0.67	0.80	0.60	2.58	0.37	0.58
u_*	W m^{-2}	0.19	0.05	0.08	0.70	0.78	0.02	0.05	0.84	0.90	−0.05	0.10	0.21	0.49
q_1	g kg^{-1}	3.38	0.06	0.26	0.69	0.73	0.17	0.26	0.71	0.69	−0.03	0.35	0.45	0.68
T_1	K	271.77	0.79	1.19	0.50	0.54	1.33	1.30	0.25	0.47	0.53	1.32	0.39	0.59
T_{ice}	K	271.72	0.99	1.26	0.28	0.47	1.38	1.30	0.53	0.44	1.03	1.49	−0.16	0.37
$cldfrac$	–	0.79	0.01	0.31	−0.01	0.34	0.19	0.27	0.15	0.46	–	–	–	–
LW_{dn}	W m^{-2}	296.50	−9.34	14.62	0.60	0.71	8.90	17.84	0.18	0.48	−11.02	27.68	0.00	0.27
LW_{up}	W m^{-2}	309.10	4.49	6.01	0.13	0.45	6.51	5.93	0.31	0.44	4.71	6.79	−0.16	0.37
SW_{dn}	W m^{-2}	135.23	16.15	51.63	0.63	0.74	−35.76	40.17	0.49	0.60	24.60	53.13	0.38	0.57
SW_{up}	W m^{-2}	107.58	−23.19	35.28	0.60	0.70	−57.83	31.81	0.49	0.48	12.80	41.50	0.46	0.65
LW_{net}	W m^{-2}	−12.60	−13.83	16.10	0.42	0.57	2.39	16.14	0.08	0.35	−15.73	25.98	−0.11	0.19
SW_{net}	W m^{-2}	27.66	39.34	27.46	0.40	0.30	22.07	16.96	0.33	0.40	11.80	14.87	0.19	0.41
Rad_{net}	W m^{-2}	15.06	25.51	21.69	−0.08	0.27	24.46	16.18	0.04	0.32	−3.93	19.11	0.11	0.37
H	W m^{-2}	−2.07	0.49	6.75	0.52	0.59	2.35	5.51	0.37	0.53	0.32	5.00	0.27	0.52
E	W m^{-2}	−5.09	−9.99	8.83	0.14	0.34	−4.00	7.85	0.07	0.37	2.97	5.20	0.06	0.40
tot_{hflx}	W m^{-2}	7.09	16.01	18.92	−0.07	0.34	22.81	17.61	0.22	0.42	−0.64	18.00	0.31	0.46

[a]The absolute bias (a.b.) is the mean difference between each observed and modeled parameter. A positive bias implies that, for a given parameter, the model produces a value of higher magnitude than that observed. The mean observation over the entire field campaign ($\bar{x}_{obs}$), the standard deviation (σ) of the difference between each three hourly averaged observation and modeled value, the correlation coefficient (R), and the "Index of Agreement" (IoA) are also given.

using radiation measurements over ice surfaces only is considered valid and adequate for the methods of analysis used here.

[26] The comparisons in this study are conducted using either time series or time-height cross sections of various variables. To complement this, a basic statistical analysis is also presented in Table 1, which compares 117 three-hourly mean model and observational data points. The absolute bias is the mean difference between each observed and modeled parameter. The mean observation is also given, along with the standard deviation, σ of the differences between each 3 hourly observation and modeled value. Models that reproduce the observations to a high degree of accuracy should have a low absolute bias and a low standard

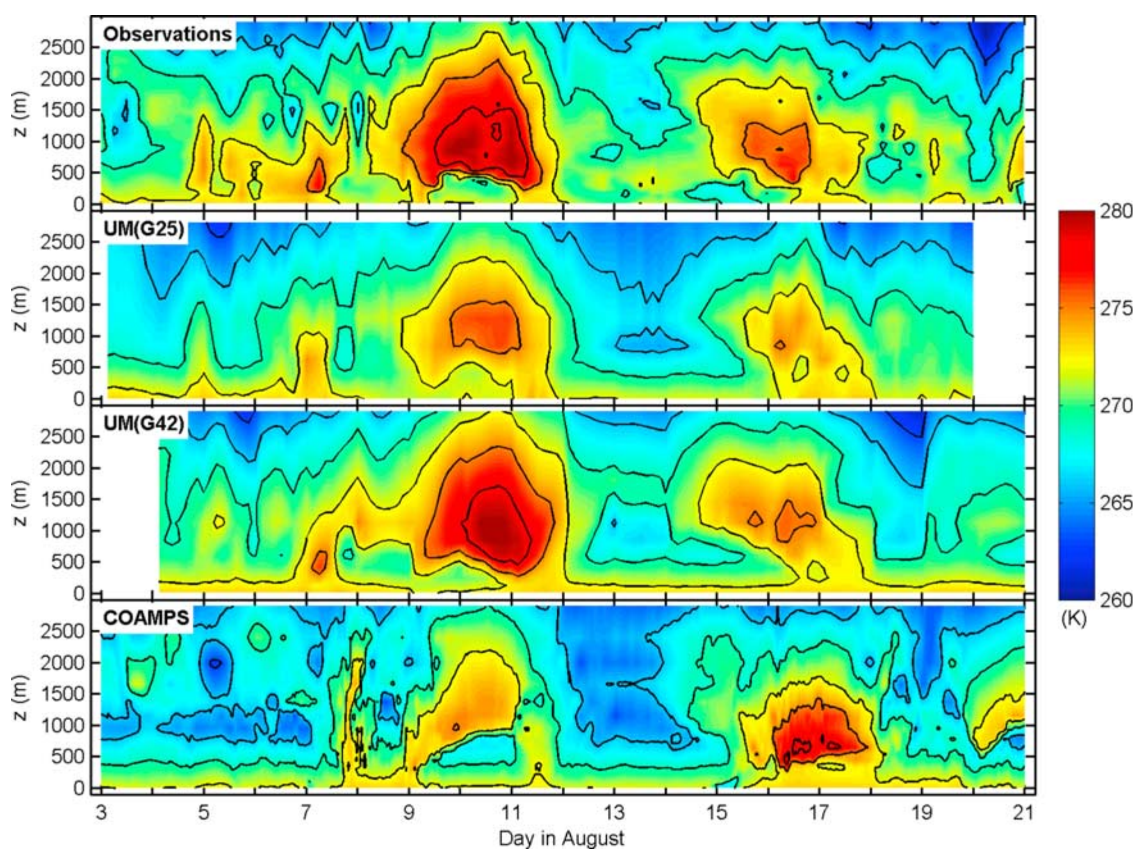

Figure 3. Air temperature measurements from the six hourly radiosondes compared to model diagnostics during the AOE 2001 observation period. Isopleths are at 3 K intervals.

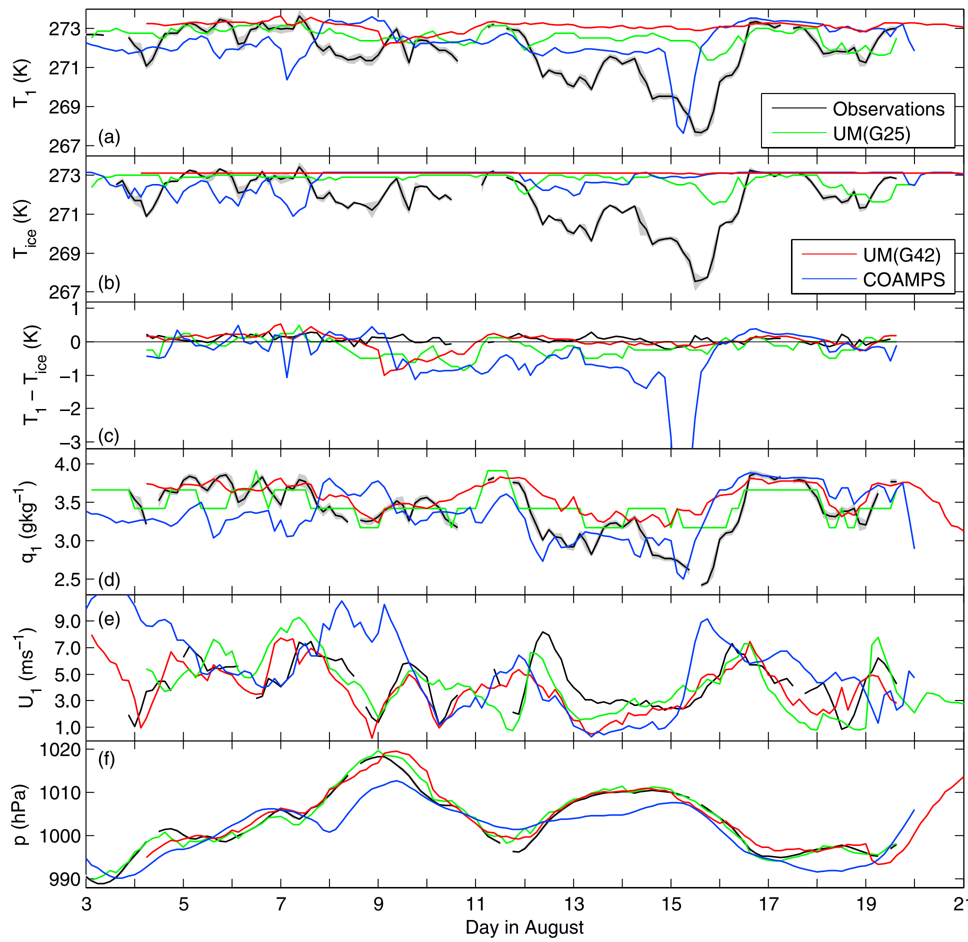

Figure 4. Three hourly mean observations and model comparisons during the AOE 2001 observation period. Three hourly mean (a) near-surface air temperature T_1, (b) ice surface temperature T_{ice}, (c) $T_1 - T_{ice}$, (d) near-surface specific humidity q_1, (e) 10-m wind speed U_{10m}, and (f) surface pressure p. All measurements, except for T_{ice}, were made on the meteorological mast. The gray area represents ±1 standard deviation about each 3-hour mean observation.

deviation. When testing for the degree of correlation, a model could produce the correct signal even if it is out of phase with the observations, returning a low or even negative correlation coefficient. For this reason both the correlation coefficient, R and the "Index of Agreement", IoA have been computed, since the IoA takes into account phase differences between two signals [*Tjernström et al.*, 2005].

4.2. Basic Meteorological Fields

[27] Figure 3 shows a time-height cross section comparing air temperature from radiosonde observations with that simulated in the models. There are two obvious warm periods above 500 m between 9–12 and 15–18 August, which all three models reproduce more or less accurately. Warmer air was also observed up to 1500 m between 4.5 and 7.5 August, which is less well represented by the two MetUM models and not at all by COAMPS. A cold period occurs throughout the lowest 3 km of the atmosphere between 12 and 16 August, with a distinct region of cold air in the lowest 400 m on 14–15 August. The MetUM simulates the cold air aloft with reasonable accuracy, with UM(G42) producing the best results. The cold air close to the surface however, is not reproduced at all in either version of the model. This is also illustrated in the Figure 4a, which shows T_1 over the entire ice drift period. The observed cold period on 15 August is not at all evident in UM(G42), which keeps the temperature fairly constant, very close to 273 K, the freezing point of fresh water and UM(G25) produces only a slight decrease in temperature. COAMPS produces a drop in temperature close to the surface on 15 August but for a much shorter duration than the observed cold event. All three models have a mean positive bias in T_1 (i.e. the models are too warm), with UM(G42) showing the largest discrepancy. None of the models are well correlated with the observations.

[28] Ice surface temperature, T_{ice}, measurements were derived from the surface longwave radiation flux following *Persson et al.* [2002]. Observed T_{ice} ranges between 273 and 267 K (Figure 4b) and T_1 follows a similar variation over time. All three models show a positive mean bias in T_{ice} of at least 1 K. UM(G42) performs the worst, where T_{ice} remains at 273.1 K for almost the duration of the observation period, except for a very small decrease on 15 August. COAMPS produces a similar magnitude of error in T_{ice} and none of the models are correlated well to the observations.

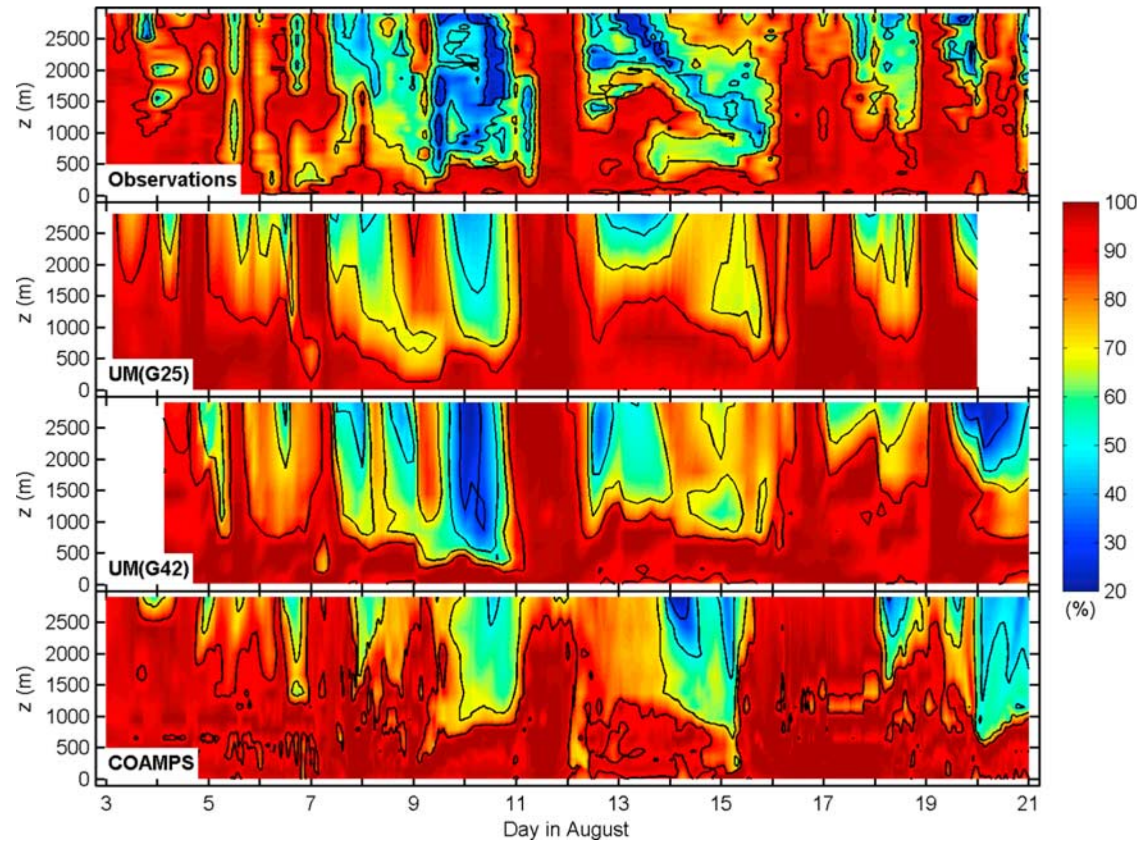

Figure 5. Same as Figure 3, but for relative humidity. Isopleths are at 20% intervals.

[29] The radiosonde observations show that relative humidity was constantly above 90% in the lowest 100 m of the atmosphere, which all three models simulate well (Figure 5). There are periods of high humidity throughout the lowest 3 km of the atmosphere on 3–7, 11, 16, and 19 August, which are also represented well in the models. The observations show two prolonged periods of low humidity aloft, occurring between 9–11 and 12.5–16 August and there are additional shorter low-humidity periods throughout the measurement period. UM(G25) and UM(G42) simulate most of the low-humidity events well (e.g., 10 August) but neither produce low enough humidities between 14–16 August. COAMPS generally represents the timing of low- and high-humidity periods accurately but away from the surface there is a general bias toward higher humidities than those observed. The near-surface specific humidity, q_1 (Figure 4d and Table 1) is positively biased in both versions of the MetUM and negatively biased in COAMPS but the bias is small and all three models show at least reasonable correlation to the observations.

[30] Observed wind speeds up to 3 km in altitude were often below 5 m s^{-1} but there are notable periods of stronger winds on 5–9, 12, and 16 August (Figure 6). Both versions of the MetUM capture the high wind events with reasonable accuracy, although there is a tendency to underestimate the speed. COAMPS reproduces the magnitude of the high wind events with some accuracy but these events are often phase shifted in time. This is not unexpected since COAMPS is free to develop without daily assimilated observations, apart from at the model boundaries. Both these points are highlighted in the 10 m wind speed, $U_{10\mathrm{m}}$ in Table 1 and Figure 4e; both versions of the MetUM show a negative bias, although it is much smaller in UM(G42). The wind speed in COAMPS is positively biased and has a lower correlation coefficient than the MetUM. Modeled surface pressure is by far the best simulated diagnostic (Figure 4f), where the bias is notably larger and the correlation notably less in COAMPS than in either version of the MetUM.

[31] The p, $U_{10\mathrm{m}}$ and q_1 fields and air temperatures away from the surface are represented reasonably well in all three models. This is not surprising since the AOE 2001 radiosonde observations were utilized in the UM(G25) forecasts and to produce the ERA-40 data used to initialize the UM(G42) forecasts. The models should therefore be expected to reproduce these basic meteorological fields with at least reasonable accuracy. COAMPS performs notably worse than the MetUM in these basic parameters because it was run without any constraints from assimilation of observational data, except for at the outermost boundaries. The fact that the difference between the correlation coefficient and the IoA for $U_{10\mathrm{m}}$ and q_1 is much greater in COAMPS than the MetUM indicates that the general signal is correct but it is out of phase with the observations. Errors in the surface flux and cloud diagnostics produced by inaccuracies in the larger scale circulation rather than in the physical parameterizations will occur in all three models but are likely to be more significant in COAMPS. It is therefore important to assess the success of a model compared to the observations based on mean values over

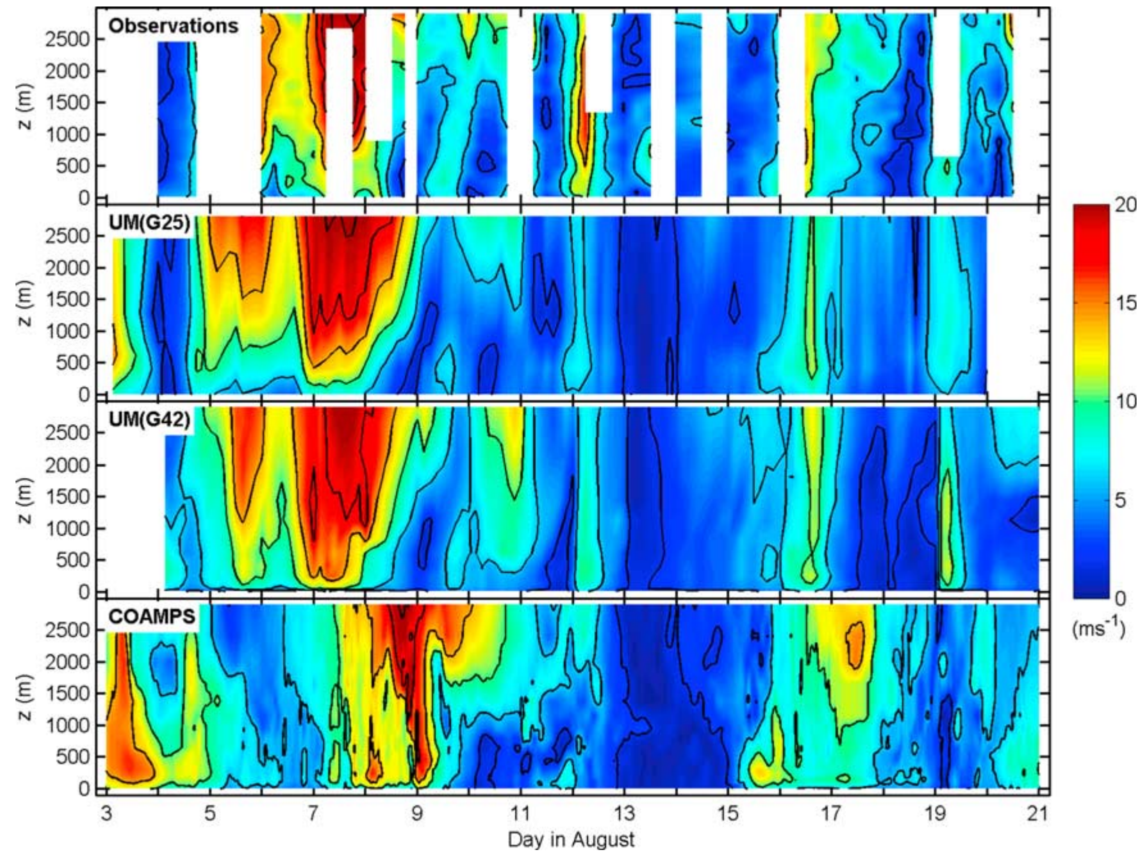

Figure 6. Same as Figure 3, but for wind speed. Isopleths are at 4 m s^{-1} intervals. Missing observations are due to instrument failure.

extended periods of time rather than on its representation of individual weather events and that MetUM-COAMPS comparisons should be made with caution due to the fundamental differences in their formulations.

4.3. Surface Turbulent Fluxes

[32] Observed and modeled friction velocity u_*, and the turbulent fluxes of sensible, H and latent, E heat are presented in Figure 7, along with a statistical analysis in Table 1. Throughout this paper, the surface radiative and turbulent fluxes are defined such that a positive flux represents a transfer of energy to the surface. There were significant problems with ice and condensation forming on the sensing heads of the sonic anemometers and Krypton hygrometers on the meteorological mast during AOE 2001, limiting the turbulent flux data set that is available for analysis. The measurements from the three ISFF stations are however more extensive and there is reasonably good agreement between these and the mast data, giving confidence that the measurements used from each location are representative of average conditions over the whole region.

[33] The values of u_* produced by UM(G25) and UM(G42) are well correlated to the observations, which is expected since the correlation between modeled and observed U_{10m} is also high. Both versions of the MetUM produce at a small positive bias in u_*, even though the wind speeds show a small negative bias. Figure 8 compares the value of u_* to the value of U_{10m}, where the gradient of each line is representative of the transfer coefficient at 10 m above the surface. The range of gradients produced by the observations is most likely indicative of the spatial variation in roughness length over the measurement sites. *Tjernström* [2005] estimated the mean value of z_m during the AOE campaign at 0.003 m. This is an order of magnitude higher than the value computed for SHEBA [*Persson et al.*, 2002], although that value represents average conditions over the entire 12 month campaign rather than over the summer months only. z_m is set to a constant value of 0.003 m in the MetUM, equal to that observed. The transfer coefficient produced by UM(G25) is too large, explaining the slight positive bias in u_*. Since the value of z_m is accurate in the model this bias could be explained by its representation of atmospheric stability. The transfer coefficient produced by UM(G42) is closer to the observations, accounting for the smaller bias in u_*.

[34] The correlation between observed u_* and that produced by COAMPS is poor compared to that between the MetUM and the observations. This is most likely due to the lower correlation between the modeled and observed wind speeds. COAMPS produces an overall negative bias in u_*, even though the overall bias in U_{10m} is positive. Figure 8 suggests this is due to an underestimation of the transfer coefficient, consistent with the low value of z_m used in COAMPS (1.4×10^{-5} m); two orders of magnitude lower than that observed during AOE 2001.

[35] All three models show good agreement in the sensible heat flux during many periods of the field campaign (Figure 7b), although there is a tendency towards magni-

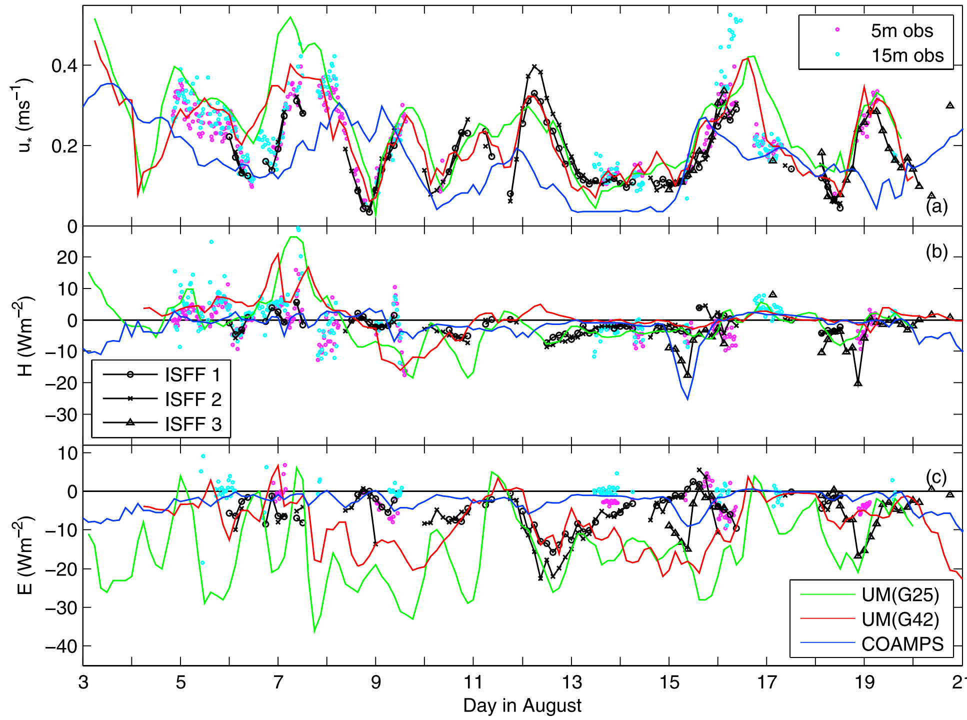

Figure 7. Surface flux observations and model diagnostics: (a) friction velocity, (b) sensible heat, (c) latent heat. Model diagnostics and measurements from the ISFF stations are presented as three hourly averages, and the measurements from the meteorological mast are half hourly averaged fluxes. A positive flux represents a transfer of energy to the surface.

tudes that are too low (Table 1). Correlation between the models and the observations is generally low and the standard deviation of the bias high. The correlation coefficient in UM(G42) is similar to that produced by the other two models but the mean absolute bias is much larger.

[36] Figure 8b shows H/U_{10m} plotted against $T_1 - T_{ice}$ for each model. Observations include measurements made from the meteorological mast only due to the lack of T_{ice} or upwelling longwave radiation flux measurements at the ISFF stations. In COAMPS $T_1 - T_{ice}$ (Figure 4c) is mostly too large in magnitude and on average over the entire observation period is the wrong sign compared to the observations. This should lead to an overestimation of the magnitude of H compared to the observations. However, the transfer coefficient is much smaller than that produced by the observations. This compensates for the overestimation of $T_1 - T_{ice}$. Both versions of the MetUM overestimate the transfer coefficient for H but the magnitude of H produced by the models is underestimated due to the low values of $T_1 - T_{ice}$.

[37] Model biases in the latent heat flux, E are much larger than in either H or u_* and the correlation between each model and the observations is very low. Both versions of the MetUM produce a negative bias in E (too much energy emitted from the surface), which is consistent with other modeled and observed latent heat flux comparisons such as in the works of *Brunke et al.* [2006] and *Tjernström et al.* [2005]. COAMPS however produces magnitudes of E that are lower than the observations, at least in part due to the low value of z_q.

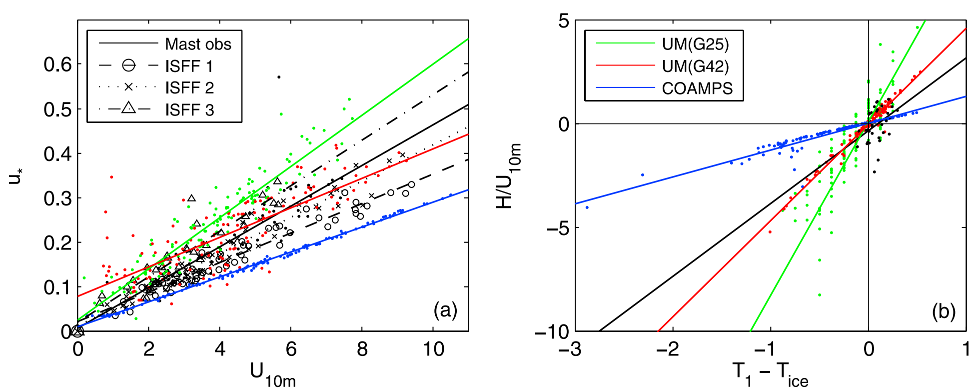

Figure 8. (a) Three hourly averaged u_* against U_{10m}. (b) Three hourly averaged H/U against $T_1 - T_{ice}$.

[38] Another potential source of error in both modeled *H* and *E* is the representation of snow and ice in the models. In reality the surface temperature of sea ice adjusts very rapidly to changes in atmospheric forcing caused by, for example, variations in the radiative fluxes due to changing cloud conditions. Since neither the MetUM nor COAMPS incorporate a fully coupled ice model, the force-restore method used within them requires a relatively thick layer of ice at the surface to change temperature. This process may not occur quickly enough in the models, meaning the surface temperature reacts too slowly to changes in surface forcing and thus potentially causes errors in the modeled surface turbulent fluxes, which are forced by processes on synoptic or shorter timescales.

4.4. Cloud Occurrence

[39] Cloud fraction is a difficult quantity to measure and represent accurately. Observations were derived from ceilometer measurements, which retrieved cloud base height at a single point in the sky at a frequency of 4 samples per minute. A cloud fraction parameter was then computed from this by taking a time average of the measurements over a 3 hour period. The cloud fraction variable determined by the MetUM is a parameterized spatial average, where cloud fraction on each model level in a grid box is used to compute a total fraction assuming maximum overlap (this type of cloud field is unavailable from COAMPS). A comparison of modeled and observed cloud fraction is, however still worthwhile since a temporal average of clouds moving over a single point in the sky should have a quantitative relationship to a spatially averaged model parameter. Figure 9a and Table 1 show these quantities. UM(G42) generally overpredicts cloud fraction, keeping it at 100% for the majority of the time but it does reproduce some periods of decreased cloud fraction found in the observations, such as on 18–19 August. This is in agreement with the findings of *Tjernström et al.* [2008], who found regional scale models produce clear conditions less frequently than what was observed during SHEBA. Over the whole observation period UM(G25) produces a lower absolute bias than UM(G42), although it shows less correlation with the observations.

[40] Success in the representation of cloud occurrence cannot be assessed using only cloud fraction, since in theory a model could generate a perfect annual cycle of cloud fraction but still produce cloud at incorrect heights and with the wrong radiative properties. A more informative way of assessing modeled cloud is through the cloud ice and liquid water concentrations. Figures 9b and 9c show time series of ice water path (IWP) and liquid water path (LWP) for each of the three models and Table 2 presents the mean modeled IWP, LWP and total cloud water over the entire period. Since observations of these variables are not available from AOE 2001, mean values observed when clouds were present during August at SHEBA [*Shupe and Matrosov*, 2006] are used as representative values for comparison. Additionally, mean cloud base measurements from the ceilometer and backscatter from the S-band cloud and precipitation radar can be compared to the model time-height cross sections of total cloud water concentrations (Figures 9d–9f). Although patchy, the S-band radar shows several periods in which cloud extends to above 3 km, for example, on 11 August. These deeper clouds are associated with the passage of synoptic scale frontal systems, which included some precipitation. Low-level clouds or fog, which are too close to the surface for the S-band radar to observe, are indicated by the ceilometer cloud base measurements; cloud base was typically between 100 and 200 m.

[41] Both versions of the MetUM show distinct periods during which cloud extends up to approximately 7 km (e.g., 11, 16.5–17.0, and 19.0–19.5 August) and where radar data is available, the timing of these events is correct. The most obvious difference in cloud between the two MetUM models is the near persistent cloud layer below 1 km in UM(G42) (e.g., 12–15 August). In general, UM(G25) underpredicts low cloud and UM(G42) produces a layer of low-level cloud which occurs too frequently compared to the observations and is not necessarily correct in its altitude, thickness or radiative properties. During the periods with deeper clouds both models produce peaks in IWP and LWP, although the magnitude of the LWP (IWP) peaks are significantly larger (smaller) in UM(G42). Furthermore, the LWP is between 25 and 100 g m^{-2} in UM(G42) and near zero in UM(G25) during the low-cloud periods such as 12–16 August. The partition between ice and liquid cloud water in UM(G42) is consistent with the SHEBA data (Table 2). UM(G25) however, underestimates the value of liquid water and overestimates the value of ice water.

[42] COAMPS produces high concentrations of cloud water at single grid points and zero cloud water at others, producing a sharp gradient between grid boxes containing high and zero cloud water concentrations, which accounts for the peculiar-looking profiles. The model produces cloudy skies for the greater part of the observation period, with cloud up to 10 km for the majority of the time. There is a distinct segregation of ice and liquid cloud water, where cloud water below 5 km is liquid and water above 5 km is ice (not shown). The IWP is similar to the observations during SHEBA, though the mean LWP is significantly lower.

4.5. Radiation and Total Heat Flux

[43] To produce accurate climate predictions it is critical that the surface energy budget, including the radiative fluxes, are modeled correctly. Cloud fraction, thickness, and optical and microphysical properties all significantly influence the radiation balance at the surface. An evaluation of the modeled surface radiation budget, while important in its own right, will also give further insight into the success of cloud representation in the models.

[44] As noted in section 2, the sensor measuring SW_{up} at the mast site failed during the field campaign. Albedo is calculated from a second set of SW_{dn} and SW_{up} measurements, that were made periodically during the campaign. From this data, the albedo of the surface is estimated using a polynomial fit to the data clusters (Figure 1). To avoid unrealistic values produced by an extension of the polynomial to times before the first albedo observations were made, a constant value of 0.9 (the mean of the first observation cluster) is used for the previous day (4 August). The albedo is then used to calculate SW_{up} using SW_{dn} measurements from the first set of sensors. This process introduces some uncertainty in the radiation flux estimates. To assess the extent of this error the mean and standard

deviation of each cluster of albedo data points is computed. The mean albedo measurement ±1 mean standard deviation is 0.796 ± 0.02. This is then used to calculate the error range in the values of mean observed SW_{up}, SW_{net} and Rad_{net}, which are 107.58 ± 2.7, 27.66 ± 2.7 and 15.06 ± 2.7 W m^{-2} respectively. The error is relatively small and even the uncertainty in the radiation fluxes calculated by the standard deviation of the cluster means (0.06, producing an

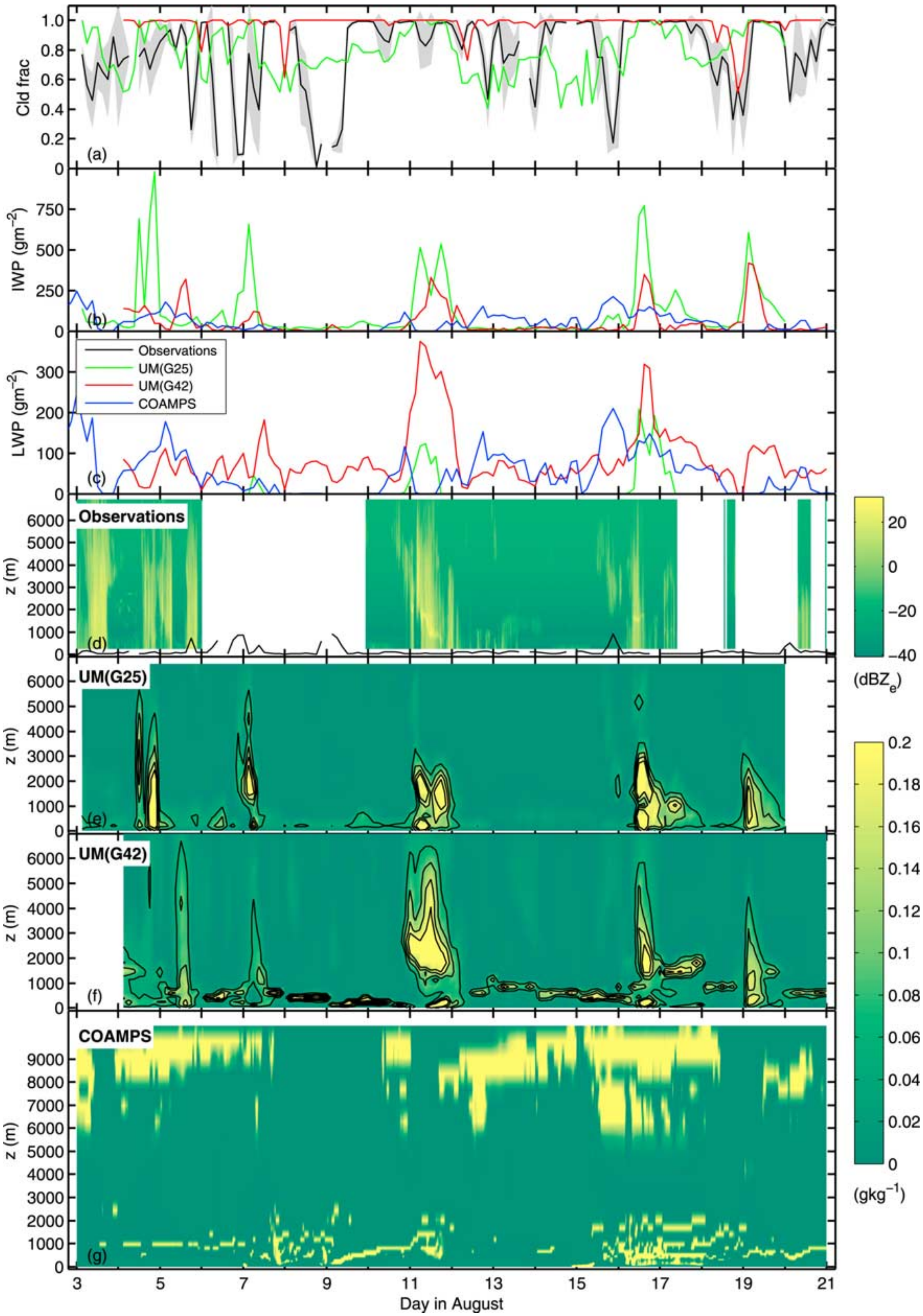

Figure 9

Table 2. Mean Modeled Liquid and Ice Water Paths (g m^{-2}) Compared to Mean Observations During Periods Where Clouds Were Present for the Month of August From SHEBA [*Shupe and Matrosov*, 2006]

	Observations	G25	G42	COAMPS
IWP	50–60	116	51	51
LWP	70–90	11	83	15
IWP + LWP	120–150	127	134	66

uncertainty of ±8.2 W m^{-2}) is not significant enough to change the general relationship between each model and the observations.

[45] Table 1 lists the mean absolute biases in the radiation components and Figure 10 shows scatterplots of the modeled and observed individual component and net surface radiative fluxes. The net surface radiative fluxes are defined such that a positive flux represents a transfer of energy to the surface. An important result from both the statistics and Figure 10 is the lack of correlation with the observations in all three models. The correlation is generally better in the separate upwelling and downwelling long and shortwave radiation components than in the net radiation fluxes, where the accumulation of errors in the separate components produces large biases. Since the downwelling radiation fluxes, LW_{dn} and SW_{dn}, are the important fluxes when considering the effects of cloud on the radiation balance, these are considered first.

[46] Both UM(G25) and COAMPS overestimate SW_{dn} and underestimate LW_{dn} (Table 1). *Shupe and Intrieri* [2004] have found that the radiative properties of clouds with LWP values that are less than 20–50 g m^{-2} depend strongly on the value of the LWP, whereas clouds with larger LWPs behave almost as black bodies and thus the absolute value of the LWP is of less importance. In both UM(G25) and COAMPS the mean LWP is less than 20 g m^{-2} and much lower than expected based on the SHEBA data. This is the most likely cause of overestimated SW_{dn} and underestimated LW_{dn}. UM(G42) overestimates LW_{dn} and underestimates SW_{dn}; mean IWP and LWPs are much closer to the expected values and therefore the positive bias in cloud fraction is a more likely cause for the biases in downwelling radiation.

[47] LW_{up} is dependent on the temperature of the surface and is overestimated in all three models due to the positive bias in T_{ice}. These errors are however, small compared to those in SW_{dn} and LW_{dn} due to the relatively small temporal variation in T_{ice} during August. The value of modeled SW_{up} depends on the magnitude of SW_{dn} and the albedo of the surface. Figure 1 shows surface albedo observations made over the duration of the field campaign. The albedo over sea ice in the MetUM can vary between a minimum of 50% and a maximum of 80%, depending on the temperature of the surface. Due to the overestimation of T_{ice} the albedo produced by both versions of the MetUM is too small and the error in UM(G42) is especially prominent; its almost constant value of 0.5 is obviously unrealistic. For UM(G25), the overestimation of SW_{dn} partially compensates for the underestimation of albedo, leading to a smaller underestimation of SW_{up}. The surface albedo in COAMPS is based on the amount of time elapsed since the last snowfall, rather than T_{ice} and produces the highest and most realistic values for albedo of all the models and therefore values of SW_{up} with the smallest bias.

[48] Rad_{net} is overestimated in UM(G25) and UM(G42) by 25.5 and 24.5 W m^{-2} respectively. The error in LW_{net}, and more specifically in LW_{dn} dominates in COAMPS and is reflected by an underestimation in Rad_{net} of 3.9 W m^{-2}. The bias in LW_{dn} and SW_{dn} in COAMPS are similar to those in UM(G25), indicating both models produce a similar magnitude of error from cloud forcing. This suggests the cause of the large bias found in Rad_{net} in UM(G25) and thus in UM(G42) is dominated by the unrealistic surface albedo parameterization in the MetUM, with errors in cloud radiative forcing having a smaller but still important effect.

[49] Table 1 shows the mean observed net heat flux at the surface is +7.1 W m^{-2}. This is about half the value observed at SHEBA during the month of August (+15–19 W m^{-2} [*Persson et al.*, 2002]). SHEBA measurements were made at a lower latitude and thus experience slightly higher insulation. UM(G25) and UM(G42) overestimate the observed value by +16.0 and +22.8 W m^{-2} respectively, even though the biases in the latent heat flux compensate for the errors in Rad_{net} to some extent. T_{ice} is calculated iteratively in the models from the turbulent heat fluxes and radiative terms in the surface energy budget. If any of these terms cause too much energy to be absorbed by the surface, modeled T_{ice} is overestimated. Since the albedo of the surface in the MetUM is based on T_{ice}, this causes an important feedback at the surface of the model. Errors in the model radiative fluxes cause an overestimation of the total heat flux; T_{ice} is positively biased and the albedo is underestimated. This underestimation caused too much SW_{dn} to be absorbed at the surface, further increasing the error in the total heat flux and T_{ice}. In UM(G42) this locks the albedo at its lowest value of 50%. In UM(G25) the same feedback occurs but to a lesser extent due to the smaller bias in the total heat flux.

[50] The temperature of the ice surface in the MetUM cannot increase above 0°C. In reality, when an ice surface is at 0°C additional heat input would melt the ice. In these models, where ice extent is prescribed and ice thickness is constant, an imbalance in the heat flux cannot cause the ice to melt and disappear. However, in a version of the model with a fully coupled ice model where ice extent and thickness are explicitly simulated, this imbalance could

Figure 9. Cloud observations and model diagnostics. (a) Three hourly averaged cloud fraction. The gray area represents ±1 standard deviation about each 3-hour mean observation. (b) Modeled ice water path (no observations available). (c) Modeled liquid water path (no observations available). (d) Radar backscatter from the S-band cloud and precipitation radar. Backscatter is proportional to the amount of condensate in the atmospheric column, wherein the threshold is at approximately 0 to +5 dBZ_e. The black line shows three hourly averaged mean cloud base measurements derived from the ceilometer. (e) UM(G25) profile of modeled total frozen plus liquid cloud water concentration. Isopleths are at 0.05 g kg^{-1} intervals. (f) Same as Figure 9e, but for UM(G42). (g) Same as Figure 9e, but for COAMPS.

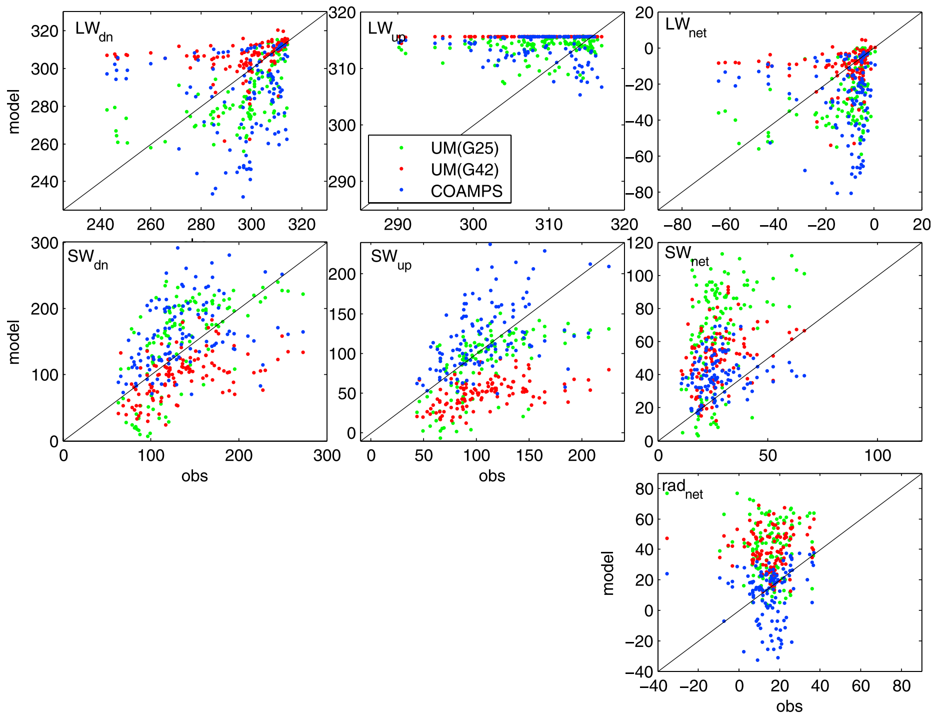

Figure 10. Comparison of three hourly averaged modeled and observed radiative fluxes. A positive flux in LW_{net}, SW_{net} and Rad_{net} represents a transfer of energy to the surface.

cause excess ice melt over the course of the summer season producing inaccuracies in future predictions of sea ice extent and other variables. In the climate version of the MetUM (Hadley Centre Global Environment Model, HadGEM1) the albedo parameterization is less simplistic. It is more dependent on snow depth and unlike the NWP version of the MetUM, includes the effects of melt ponds and an open lead fraction even at such high latitudes. The erroneous albedo feedback therefore does not occur in HadGEM1 and the bias in the surface total heat flux should be less extreme. This does not suggest however that the heat budget at the surface is error free, since errors in the radiation fluxes described here are likely to also apply in the climate version of the model.

[51] In COAMPS the underestimation of Rad_{net} is offset largely by biases in the turbulent heat fluxes, producing only a small underestimation of the total heat flux. There is however a large positive bias in T_{ice} and T_1, a result which is not expected. This is discussed in more detail in the following section.

4.6. Case Study

[52] Here we examine a period of relatively cooler temperatures observed in the lowest 3 km of the atmosphere between 12.0 and 16.0 August (Figure 3). *Tjernström et al.* [2004a] show that during the summer months, the near-surface air temperature is most frequently at 0°C or −1.7°C, the melting points of fresh and seawater respectively. This indicates strong control of the near-surface air temperature by a surface consisting of snow, ice, open leads and melt ponds. If colder air is advected over a sea ice surface, the surface warms the atmosphere through the release of sensible heat and then through latent heat as melt pond and seawater begin to freeze. For the regional average air temperature to drop below −1.7°C for a significant amount of time a layer of ice must form on top of a sufficient fraction of melt pond and open lead surfaces, significantly reducing the magnitude of the heat fluxes. Formation of a thin layer of ice on top of melt ponds and open leads was observed visually during this period.

[53] Figure 11 shows 5 day back trajectories ending at the observation site together with a plot of sea ice extent from the UM(G25) analyses, in which sea ice fraction is diagnosed from the assimilation of satellite data. The start of the observed temperature decrease (11.75 August) coincides with a change in air mass origin, from air originating over warmer, open ocean, to air that has spent at least 5 days over the pack ice. This suggests that the cold air results from advection from another region of the Arctic rather than local cooling; this is supported by the fact colder temperatures were observed up to 3 km, rather than only at the surface. If this temperature decrease was caused by local radiative cooling at the surface, the observed heat fluxes would be positive (downward). Over the entire cold period the observed sensible heat flux is negative, only returning to positive once the air temperature recovered on 16 August (Figure 7) and the observed total heat flux remains positive, at 2.47 W m^{-2} (Table 3) even though T_{ice} decreases significantly, which is contrary to what is expected. This disparity is most likely due to uncertainties in the observed values that make up the surface energy budget. The maximum uncertainty in the net radiation measurements is 8.2 W m^{-2}. This, along with a typical uncertainty of 20% in the eddy-covariance measurements of sensible and latent heat [e.g., *DeCosmo et al.*, 1996] results in a potential total heat flux down to −6.63 W m^{-2} for the cold period, which could

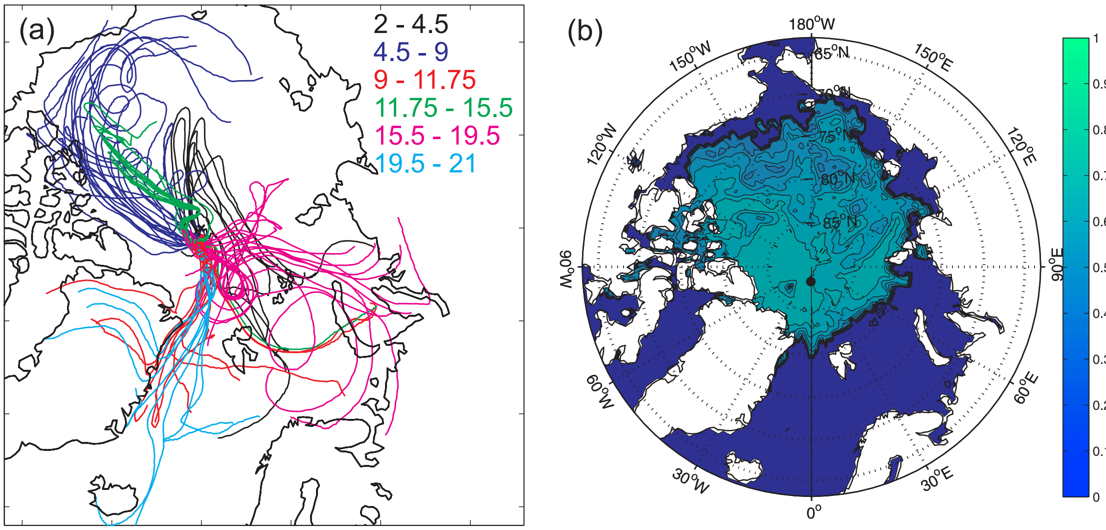

Figure 11. (a) 5-day back trajectories ending at the observation site during AOE 2001, calculated with the *McGrath* [1989] three-dimensional trajectory model utilizing ECMWF analyses. The time of arrival of the air masses at the observation site are in decimal days in August. (b) Sea ice fraction from UM(G25), which is diagnosed from satellite observations. The black dot marks the location of the AOE 2001 observation site.

easily have caused the decrease in observed surface temperature.

[54] During the periods 12–14 and 14–15 August cloud was observed up to 2000 m and 400 m respectively (Figure 9d). The ceilometer observations show a near constant layer of low-level cloud, apart from a period with decreased cloud cover during the second half of 15 August, coinciding with the coldest T_{ice} and T_1 observations (Figures 9d, 4a, and 4b). During this decrease in cloud cover, Rad_{net} decreases and becomes negative for a short time (Figure 12c), indicating radiative cooling of the surface, thus further enforcing the cold period.

[55] Figure 3 shows decreased air temperatures above the surface in all three models during the cold period, indicating that they have to some extent reproduced the advection of cold air over the observation site. This cold period is not seen in modeled T_1 and T_{ice}, except briefly in COAMPS, due to errors in the surface energy budget, where the representation of clouds play a significant role. The observations show that low-level clouds prevail during the cold period. The properties of these clouds and their impact on the radiation budget during the cold period are now assessed using a comparison with periods where different cloud conditions are prevalent. For this we use a number of periods when the passage of synoptic scale frontal systems produced cloud that extended to above 3 km (11.0–12.0, 16.5–17.0, and 19.0–19.5 August). The absolute model biases for T_{ice}, T_1 and the radiation and turbulent fluxes, computed in the same way as those in Table 1 are presented for the “cold period” (12–15.5 August) and the “deep cloud” periods in Table 3.

[56] At times when deep clouds were observed, the biases for almost all variables in all three models are smaller than during periods where only low-level cloud was present. This is because all three models simulate the passage of the frontal systems and the occurrence and radiative properties of the associated deep clouds with reasonable accuracy and the radiative fluxes are less sensitive to the precise values of

Table 3. Mean Observational Values and Absolute Biases of Temperature, Radiation and Heat Flux Diagnostics During the Cold Period and During Periods With Deep Cloud Cover

		Cold Period				Deep Cloud Periods			
	Unit	$\bar{x}_{obs}$	G25	G42	COAMPS	$\bar{x}_{obs}$	G25	G42	COAMPS
T_1	K	270.20	2.17	2.88	1.29	272.77	−0.32	0.43	0.10
T_{ice}	K	270.16	2.47	2.93	2.52	272.70	−0.26	0.40	0.41
$cldfrac$	–	0.80	−0.15	0.17	–	0.91	0.06	0.06	–
LW_{dn}	W m^{-2}	288.61	−12.52	13.12	−25.22	309.56	−4.94	1.12	−9.81
LW_{up}	W m^{-2}	302.04	11.57	13.54	11.47	313.61	−1.05	2.01	1.87
SW_{dn}	W m^{-2}	141.40	39.13	−36.81	11.26	86.70	−43.01	−34.91	30.66
SW_{up}	W m^{-2}	115.01	−10.93	−62.70	−7.98	65.74	−48.94	−39.85	21.03
LW_{net}	W m^{-2}	−13.43	−24.09	−0.42	−36.69	−4.05	−3.89	−0.89	−11.67
SW_{net}	W m^{-2}	26.39	50.06	25.88	19.24	20.96	5.92	4.93	9.63
Rad_{net}	W m^{-2}	12.96	25.98	25.46	−17.45	16.91	2.03	4.04	−2.05
H	W m^{-2}	−3.50	−0.67	2.92	−0.96	−0.37	−0.12	0.86	−1.76
E	W m^{-2}	−6.99	−10.32	−6.20	3.83	−4.58	−6.58	−2.75	−2.01
tot_{hflx}	W m^{-2}	2.47	14.99	22.18	−14.58	11.96	−4.67	2.15	−5.82

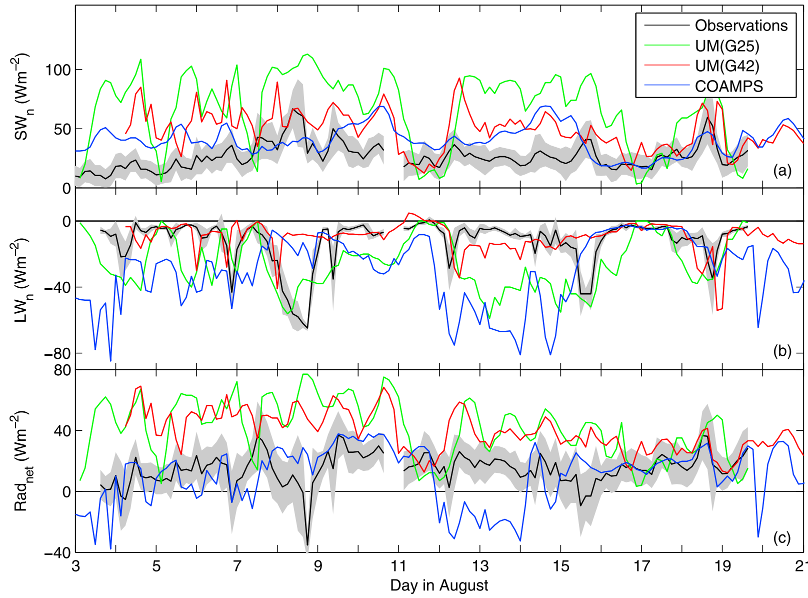

Figure 12. Three hourly averaged surface radiative flux observations and model comparisons showing (a) net shortwave, (b) net longwave, and (c) net radiation. A positive flux represents a transfer of energy to the surface. The gray area represents ±1 standard deviation about each 3-hour mean observation and for SW_{net} and Rad_{net}, includes an estimate of the error produced in the computation of SW_{up} of 8.2 W m^{-2}.

LWP and IWP when their magnitudes are large. It must be noted however, that although the simulated cloud fractions are accurate during these periods, the absolute biases in SW_{up} and SW_{dn} are still large and it is the result of the difference in these errors that produces the small error in SW_{net}. The direction of the biases in SW_{up} and SW_{dn} in COAMPS are also of the opposite sign to those in the MetUM. The resulting values of modeled Rad_{net} are all within 4.1 W m^{-2} of that observed. The predominately negative biases in the sensible and latent heat fluxes lead to a small negative bias in the total heat flux in all three models and simulated $T_{\rm ice}$ and T_1 are within 0.5 K of the observed values during these periods.

[57] During the cold period, UM(G25) produces unrealistic clear conditions, seen in Figure 9e and in the cloud fraction bias in Table 3. Over the entire observation period, incorrect partitioning of mean ice and liquid cloud water also prevails. A combination of these factors causes an underestimation of LW_{dn} and an overestimation of SW_{dn} by the model. UM(G42) produces a near constant layer of low-level cloud during the cold period, which perhaps looks fairly realistic, although the biases in Table 3 show the cloud fraction in this version of the model is overestimated. Since the partitioning of ice and liquid water is approximately correct, the cause of the overestimated LW_{dn} and underestimated SW_{dn} is the overprediction of low-level clouds. Biases in SW_{dn} and LW_{dn} due to errors in cloud occurrence and cloud radiative forcing, coupled with a large negative bias in SW_{up} caused by errors in the parameterization of the surface albedo produces a positive bias in Rad_{net} of 25 W m^{-2} in both versions of the model. Errors in H and E act to compensate for these errors to some extent but a positive bias remains in the total heat flux of 15.0 and 22.2 W m^{-2} in UM(G25) and UM(G42) respectively. These large errors account for the large biases in $T_{\rm ice}$ and T_1.

[58] The errors produced by COAMPS during the cold period are large, seem unphysical and have a large effect on the mean statistics for the model over the month of August. The separation of statistics for this model into the "deep cloud" and "cold" periods in the same way as the MetUM and the production of a set of statistics for all times other than the cold period is therefore especially helpful. During the periods with deep clouds the biases in COAMPS are similar to those in the MetUM, causing a small negative bias in the total heat flux and fairly accurate $T_{\rm ice}$ and T_1. The same can be said during periods of the field campaign other than during the cold period (Table 4). This shows that the representation of cloud forcing, surface albedo and the turbulent fluxes in the model are generally reasonable enough to produce $T_{\rm ice}$ and T_1 with only a small positive bias.

[59] During the cold period, errors in the up and downwelling radiation components are generally smaller in COAMPS than those produced by the MetUM, apart from

Table 4. Mean Observational Values and Absolute Biases of Temperature, Radiation and Heat Flux Diagnostics for COAMPS at All Times Except the Cold Period

	Unit	$\bar{x}_{obs}$	COAMPS
T_1	K	273.44	0.12
$T_{\rm ice}$	K	272.38	0.37
LW_{dn}	W m^{-2}	299.19	−6.35
LW_{up}	W m^{-2}	312.05	1.75
SW_{dn}	W m^{-2}	135.87	29.90
SW_{up}	W m^{-2}	106.95	20.46
LW_{net}	W m^{-2}	−12.86	−8.11
SW_{net}	W m^{-2}	28.91	9.44
Rad_{net}	W m^{-2}	16.05	1.33
H	W m^{-2}	−1.39	0.93
E	W m^{-2}	−3.99	2.47
tot_{hflx}	W m^{-2}	10.67	4.73

the significant underestimation of LW_{dn}. This error is most likely caused by the relatively small amount of warm, low-level cloud produced by the model during this period (Figure 9g) or too low LWP (Table 2), and results in a value of mean Rad_{net} that is of the wrong sign. This is offset to some extent in the total heat flux by the bias in E, producing a total heat flux that is both too large in magnitude and of the wrong sign; a large amount of heat is emitted from the surface by the model compared to a small amount of heat absorbed at the surface in the observations. Large negative biases in T_{ice} and T_1 would therefore be expected, but this is not the case.

[60] In COAMPS, grid boxes containing sea ice can consist of a fraction of bare and snow covered ice. The model computes T_{ice} using a weighted average of the snow and bare ice surface temperatures. When the total heat flux becomes large and negative on 11 August, T_{ice} and T_1 begin to decrease as expected. At the start of 13 August there is a decrease in the fraction of the surface that is covered in snow. This alters the weighting in the computation of T_{ice} and since the sea ice surface temperature in the model is higher than the snow surface temperature this decrease in snow cover increases T_{ice} to values above what would be expected due to the changes radiative fluxes alone. This overestimation of T_{ice} during 13 and 14 August keeps the decrease in T_1 moderate until 15 August, when a pool of very cold air is advected over the observation site in the model. This is visible in plots of near-surface air temperature fields over the Arctic region (not shown) and in the large negative sensible and latent heat fluxes produced by the model. The observed decrease in T_1 on 15 August is not accompanied by a decrease in T_{ice} (Figure 4). This is due to the fact an increase in Rad_{net} of approximately 20 W m^{-2} occurs on 15th August, offsetting the loss of energy from the surface through the turbulent heat fluxes.

5. Summary

[61] AOE 2001 field observations made over the Arctic pack ice during August 2001 are used to evaluate two versions of the global NWP MetUM and the mesoscale model, COAMPS. The UM(G25) data set is comprised of forecasts from the U.K. Met Office archives, produced by the version of the model that was in operation in 2001. UM(G42) is the latest version of the model, which contains a large number of developments to its formulation and physical parameterizations. Daily forecasts were produced for August 2001 using initial conditions from ERA-40 data. COAMPS was run with an outer domain covering the whole pan-Arctic region and contained two nested inner domains, the smallest of which was centered around the AOE 2001 observation site. The outermost domain was forced by ERA-40 data and in contrast to the MetUM model data, COAMPS was run in a "climate mode" for the entire AOE 2001 ice drift period, without any constraints except those at the outermost boundaries.

[62] The wind speed, surface pressure and relative humidity fields are at least reasonably represented in all three models. This is expected since the radiosonde observations made during AOE 2001 were assimilated into the UM(G25) forecasts and into the ERA-40 data used to initialize UM(G42) and as boundary conditions in COAMPS. Biases in these fields are larger and the correlation with the observations is worse in COAMPS and events are often phase shifted in time. This is due to the reduced constraints used in this model run. The air temperature in all three models away from the surface is represented with reasonable accuracy but close to the surface there are large positive biases. UM(G42) shows the largest bias, where T_1 and T_{ice} remain close to 273 K for the duration of the observation period.

[63] u_* is represented reasonably well in all three models, though with some explainable errors. The observed surface sensible and latent turbulent heat fluxes are negative (heat emitted from the surface) but small in magnitude. The MetUM underestimates the magnitude of the sensible heat flux, likely due to biases in T_1 and T_{ice}, and the bias in the latent heat flux is large in both versions of the MetUM. The direction of the sensible and latent heat fluxes in COAMPS are correct but the magnitudes of both are underestimated, which is most likely due to the small roughness lengths used in the parameterizations compared with the MetUM and those suggested by the observations.

[64] The MetUM computes the surface albedo as a function of T_{ice}. When the ice surface temperature is at its maximum (273.1 K) the albedo is 50% and this increases to a maximum of 80% with decreasing T_{ice}. The albedo in both versions of the model is underestimated due to the positive bias in T_{ice}. This affects the value of modeled SW_{up} and thus the entire radiation balance, creating an important feedback of errors. The climate version of the MetUM (HadGEM1) uses a more sophisticated albedo scheme, which computes the surface albedo based on surface temperature, snow depth, open lead and melt pond fraction and therefore does not suffer from this feedback. It is recommended that in future NWP versions of the MetUM the albedo over sea ice is less dependent on surface temperature and like HadGEM1 and COAMPS, is controlled by the amount of snow, ice and liquid water present at the surface, as was observed by *Perovich et al.* [2002] and *Persson et al.* [2002].

[65] All three models reproduce the occurrence and radiative properties of deep cloud, associated with synoptic scale frontal events with reasonable accuracy. During periods where only low-level cloud was observed, UM(G25) underpredicts cloud fraction and both it and COAMPS produce too little cloud liquid water compared to that observed during the SHEBA experiment. This causes an underestimation of LW_{dn} and an overestimation of SW_{dn}. The partitioning of ice and liquid cloud water in UM(G42) is more representative of typical conditions and unlike UM(G25), the newer version of the model produces a layer of low-level cloud for the majority of the observation period, possibly due to the increased vertical grid resolution in this version of the model. Although it "looks" as though it reproduces the observations with greater accuracy, cloud fraction is overpredicted leading to an overestimation of LW_{dn} and a underestimation of SW_{dn}. Similar biases in SW_{dn} and LW_{dn} produced by UM(G25) and COAMPS suggest errors in cloud forcing are similar in both models. The larger bias in Rad_{net} in UM(G25) and UM(G42) compared to COAMPS is therefore most likely dominated by the surface albedo parameterization rather than cloud forcing. The bias in the surface turbulent heat fluxes act to offset the overes-

timation of Rad_{net} to some extent but the total heat flux in the MetUM remains overestimated in both versions of the model.

[66] The changes in model formulation between versions G25 and G42 of the MetUM have made little difference to the accuracy of modeled surface pressure, relative humidity, wind speed fields and air temperature away from the surface, since these diagnostics were already reproduced with high accuracy. The production of more low-level clouds in UM(G42), although seemingly more accurate has lead to increased biases in the surface radiation balance and thus in T_{ice} and T_1. The bias in H has increased, most likely due to the increased errors in T_{ice} and T_1 but the bias in E has decreased by approximately 50%.

[67] Although there are significant errors in both SW_{up} and SW_{dn} in COAMPS, at all times other than during the cold period COAMPS produces only a small bias in the net radiation flux. This and the small biases in H and E lead to only a small errors in the average total heat flux and thus T_{ice} and T_1 are reproduced reasonably accurately. During the cold period, errors originating from cloud representation and in the reproduction of surface snow and liquid water processes at the surface produce a large positive bias in T_{ice} and T_1 during this period. This bias has a large effect on the statistics for COAMPS for the whole month of August.

[68] In all three models errors in the turbulent heat fluxes compensate for errors in the net radiation flux in the total heat flux, therefore improving one aspect of the model will not necessarily improve overall model performance. Since accurate representation of all components of the surface energy budget is central to accurate climate predictions, it is imperative to improve model parameterizations of the surface heat fluxes and of cloud properties. Improvements to simulated Arctic cloud occurrence and radiative properties in regional and global scale models generally is challenging. Progress in this area has been limited by a lack of *in situ* observational data. The processes that cause the formation and persistence of summer low-level Arctic clouds are not well understood and therefore polar specific parameterizations have not been fully developed. The recent Arctic Summer Ocean Cloud Study (ASCOS), the latest experiment in the AOE series, was conducted during August 2008 with the aim of solving some of these issues. The extensive cloud, radiation and turbulent flux data sets gained from this campaign will assist in solving these problems in the coming years.

[69] **Acknowledgments.** This work was funded by U.K. Natural Environmental Research Council (studentship and grant number NE/E010008/1), and U.K. Met Office. AOE 2001 was a multinational expedition. Logistics were funded by Swedish Secretariat for Polar Research and partly by the Knut and Alice Wallenberg Foundation. We thank the AOE 2001 participants for sharing their data with us. The S-band radar was operated by Scott Abbott, and we thank Allen White for providing the data from it. The ISFF data sets were provided by John Militzer and Steve Oncley and the albedo measurements by Bertil Larsson and Maria Lundin. We also thank the two anonymous reviewers for their helpful comments and suggestions.

References

Allan, R., A. Slingo, S. Milton, and M. Brooks (2007), Evaluation of the Met Office global forecast model using geostationary earth radiation budget (GERB) data, *Q. J. R. Meteorol. Soc.*, *133*, 1993–2010.

Andreas, E. L., P. S. 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, *J. Geophys. Res.*, *107*(C10), 8033, doi:10.1029/2000JC000411.

Arzel, O., T. Fichefet, and H. Goosse (2006), Sea ice evolution over the 20th and 21st centuries as simulated by current AOGCMs, *Ocean Model.*, *12*(3–4), 401–415.

Brunke, M., J. Zhou, X. Zeng, and E. Andreas (2006), An intercomparision of bulk aerodynamic algorithms used over sea ice with data from the Surface Heat Budget of the Arctic Ocean (SHEBA) experiment, *J. Geophys. Res.*, *111*, C09001, doi:10.1029/2005JC002907.

Chapman, W. L., and J. E. Walsh (2007), Simulations of Arctic temperature and pressure by global coupled models, *J. Clim.*, *20*(4), 609–632.

Comiso, J. C., C. L. Parkinson, R. Gersten, and L. Stock (2008), Accelerated decline in the Arctic sea ice cover, *Geophys. Res. Lett.*, *35*, L01703, doi:10.1029/2007GL031972.

Cullen, M. J. P., and T. Davies (1991), A conservative split-explicit integration scheme with fourth order horizontal advection, *Q. J. R. Meteorol. Soc.*, *117*, 993–1002.

Curry, J., W. Rossow, D. Randall, and J. Schramm (1996), Overview of Arctic cloud and radiation characteristics, *J. Clim.*, *9*(8), 1731–1764.

Davies, T., M. J. P. Cullen, A. J. Malcolm, M. H. Mawson, A. Staniforth, A. A. White, and S. Wood (2005), A new dynamical core for the Met Office's global and regional modeling of the atmosphere, *Q. J. R. Meteorol. Soc.*, *131*(609), 1759–1782.

DeCosmo, J., K. B. Katsaros, S. D. Smith, R. J. Anderson, W. A. Oost, K. Bumke, and H. Chadwick (1996), Air-sea exchange over water vapor and sensible heat: The humidity exchange over the sea (HEXOS) results, *J. Geophys. Res.*, *101*(C5), 12,001–12,016.

Edwards, J. M., and A. Slingo (1996), Studies with a flexible new radiation code: I. Choosing a configuration for a large-scale model, *Q. J. R. Meteorol. Soc.*, *122*(531), 689–719.

Fletcher, N. H. (1962), *The Physics of Rainclouds*, Cambridge Univ. Press, Cambridge, U. K.

Graversen, R. (2006), Do changes in the midlatitude circulation have any impact on the Arctic surface air temperature trend?, *J. Clim.*, *19*(20), 5422–5438.

Harshvardhan, R. Davies, T. Randall, and D. Corsetti (1987), A fast radiation parameterization for atmospheric circulation models, *J. Geophys. Res.*, *92*(D1), 1009–1060.

Hodur, R. M. (1997), The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS), *Mon. Weather Rev.*, *125*(7), 1414–1430.

Holland, M. M., and C. M. Bitz (2003), Polar amplification of climate change in coupled models, *Clim. Dyn.*, *21*(3–4), 221–232.

Holland, M. M., C. M. Bitz, and L. Tremblay (2006), Future abrupt reductions in the summer Arctic sea ice, *Geophys. Res. Lett.*, *22*, L23503, doi:10.1029/2006GL028024.

Horst, T. W., and J. C. Weil (1992), Footprint estimation for scalar flux measurements in the atmospheric surface layer, *Boundary Layer Meteorol.*, *59*(3), 279–296.

Intrieri, J., C. W. Fairall, M. Shupe, P. Persson, E. Andreas, P. Guest, and R. Moritz (2002), An annual cycle of Arctic surface cloud forcing at SHEBA, *J. Geophys. Res.*, *107*(C10), 8039, doi:10.1029/2000JC000439.

Kessler, E. (1969), On the distribution and continuity of water substance in atmospheric circulations, *Meteorol. Monogr.*, *Am. Meteorol. Soc.*, *32*(10), 82–84.

Lin, Y. L., R. D. Farley, and H. D. Orville (1983), Bulk parameterization of the snow field in a cloud model, *J. Clim. Appl. Meteorol.*, *22*(6), 1065–1092.

Lorenc, A. C., et al. (2000), The Met Office global three-dimensional variational data assimilation scheme, *Q. J. R. Meteorol. Soc.*, *126*(570), 2991–3012.

Louis, J. F. (1979), A parametric model of vertical eddy fluxes in the atmosphere, *Boundary Layer Meteorol.*, *17*(2), 187–202.

Makshtas, A., D. Atkinson, M. S. S. Kulakov, R. Krishfield, and A. Proshutinsky (2007), Atmospheric forcing validation for modeling the central Arctic, *Geophys. Res. Lett.*, *34*, L20706, doi:10.1029/2007GL031378.

Marshall, J. S., and W. M. Palmer (1948), The distribution of raindrops with size, *J. Meteorol.*, *5*, 165–166.

Martin, G. M., M. A. Ringer, V. D. Pope, A. Jones, C. Dearden, and T. J. Hinton (2006), The physical properties of the atmosphere in the new Hadley Centre Global Environment Model, HadGEM1: Part I. Model description and global climatology, *J. Clim.*, *19*, 1274–1301.

McGrath, R. (1989), Trajectory models and their use in the Irish Meteorological Service, in *Memorandum No. 112/89*, p. 12, Irish Meteorol. Serv., Dublin, Ireland.

Mellor, G. L., and T. Yamada (1974), Development of a turbulence closure for geophysical problems, *Geophys. Space Phys.*, *20*, 851–875.

Nghiem, S. V., I. G. Rigor, D. K. Perovich, P. Clemente-Colón, and J. W. Weatherly (2007), Rapid reduction of Arctic perennial sea ice, *Geophys. Res. Lett.*, *34*, L19504, doi:10.1029/2007GL031138.

NRL (2003), *COAMPS: Version 3 model description. General Theory and Equations*, *NRL/PU/7500-03-448*, Naval Research Laboratory, Marine Meteorology Division, Monterey, Calif.

Parkinson, C. L., D. J. Cavalieri, P. Gloerson, H. J. Zwally, and J. C. Comiso (1999), Arctic sea ice extents, areas and trends, *J. Geophys. Res.*, *104*(C9), 20,837–20,856.

Perovich, D. K., T. C. Grenfell, B. Light, and P. V. Hobbs (2002), Seasonal evolution of the albedo of multiyear Arctic sea ice, *J. Geophys. Res.*, *107*(C10), 8044, doi:10.1029/2000JC000438.

Persson, P., C. W. Fairall, E. Andreas, P. Guest, and D. Perovich (2002), Measurements near the atmospheric surface flux group tower at SHEBA: Near-surface conditions and surface energy budget, *J. Geophys. Res.*, *107*(C10), 8045, doi:10.1029/2000JC000705.

Phillips, T. J., et al. (2004), Evaluating parameterizations in general circulation models, *Bull. Am. Meteorol. Soc.*, *85*(12), 1903–1915.

Pinto, J. O. (1998), Autumnal mixed-phase cloudy boundary-layers in the Arctic, *J. Atmos. Sci.*, *55*(11), 2016–2038.

Pope, V. D., M. L. Gallani, P. R. Rowntree, and R. A. Stratton (2000), The impact of new physical parameterizations in the Hadley Centre climate model-HadAM3, *Clim. Dyn.*, *16*, 123–146.

Randall, D., et al. (1998), Status of and outlook for large-scale modeling of atmosphere-ice-ocean interactions in the Arctic, *Bull. Am. Meteorol. Soc.*, *79*(2), 197–219.

Rawlins, F., S. P. Ballard, K. J. Bovis, A. M. Clayton, D. Li, G. W. Inverarity, A. C. Lorenc, and T. J. Payne (2007), The Met Office global 4-dimensional variational data assimilation scheme, *Q. J. R. Meteorol. Soc.*, *133*, 347–362.

Rinke, A., P. Marbaix, and K. Dethloff (2004), Internal variability in Arctic regional climate simulations: Case study for the SHEBA year, *Clim. Res.*, *27*(3), 197–209.

Rutledge, S. A., and P. V. Hobbs (1983), The mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones: VIII. A model for the "seeder-feeder" process in warm-frontal rain bands, *J. Atmos. Sci.*, *40*(5), 1185–1206.

Schotanus, P., F. T. M. Nieuwstadt, and H. De Bruin (1983), Temperature measurements with a sonic anemometer and its application to heat and moisture fluxes, *Boundary Layer Meteorol.*, *26*(1), 81–93.

Serreze, M. C., and J. Francis (2006), The Arctic amplification debate, *Clim. Change*, *76*, 241–264.

Shupe, M. D., and J. M. Intrieri (2004), Cloud radiative forcing of the Arctic surface: The influence of cloud properties, surface albedo, and solar zenith angle, *J. Clim.*, *17*(3), 616–628.

Shupe, M. D., and S. Y. Matrosov (2006), Arctic mixed-phase cloud properties derived from surface-based sensors ar SHEBA, *J. Atmos. Sci.*, *63*(2), 697–711.

Slingo, A. (1989), A GCM parametrization for the shortwave radiation properties of water clouds, *J. Atmos. Sci.*, *46*, 1419–1427.

Slingo, A., and R. Wilderspin (1986), Development of a revised long-wave radiation scheme for an atmospheric general circulation model, *Q. J. R. Meteorol. Soc.*, *112*, 371–386.

Smith, R. N. B. (1990), A scheme for predicting layer clouds and their water content in a general circulation model, *Q. J. R. Meteorol. Soc.*, *116*, 435–460.

Solomon, S., D. Qin, M. Manning, K. Averyt, M. M. B. Tignor, H. L. Miller, and Z. Chen (Eds.) (2007), *Climate Change 2007: The Physical Science Basis, IPCC*, Cambridge Univ. Press, Cambridge, U. K.

Staniforth, A., A. White, N. Wood, J. Thuburn, M. Zerroukat, E. Cordero, and T. Davies (2006), *Joy of UM 6.3—Model Formulation*, Met Office, U. K.

Tjernström, M. (2005), The summer Arctic boundary layer during the Arctic Ocean Experiment 2001 (AOE-2001), *Boundary Layer Meteorol.*, *117*(1), 5–36.

Tjernström, M., C. Leck, P. Persson, M. Jensen, S. Oncley, and A. Targino (2004a), The summertime Arctic atmosphere. Meteorological measurements during the Arctic Ocean Experiment 2001, *Bull. Am. Meteorol. Soc.*, *84*(9), 1305–1321.

Tjernström, M., C. Leck, P. Persson, M. J. Jensen, S. Oncley, and A. Targino (2004b), The summertime Arctic atmosphere. Meteorological measurements during the Arctic Ocean Experiment 2001. supplement—Experimental equipment, *Bull. Am. Meteorol. Soc.*, *84*(9), ES14–ES18.

Tjernström, M., et al. (2005), Modeling the Arctic boundary layer: An evaluation of six ARCMIP regional-scale models using data from the SHEBA project, *Boundary Layer Meteorol.*, *117*(2), 337–381.

Tjernström, M., J. Sedlar, and M. D. Shupe (2008), How well do regional climate models reproduce radiation and clouds in the Arctic?, *J. Appl. Meteorol. Climatol.*, *47*(9), 2405–2422, doi:10.1175/2008JAMC1845.1.

Uttal, T., et al. (2002), Surface heat budget of the Arctic Ocean, *Bull. Am. Meteorol. Soc.*, *83*(2), 255–275.

van Dijk, A., W. Kohsiek, and H. de Bruin (2003), Oxygen sensitivity of krypton and Lyman-α hygrometers, *J. Atmos. Ocean. Technol.*, *20*(1), 143–151.

Walsh, J., V. Kattsov, W. Chapman, V. Govorkova, and T. Pavlova (2002), Comparison of Arctic climate simulations by uncoupled and coupled global models, *J. Clim.*, *15*(12), 1429–1446.

Wilson, D. R., and S. P. Ballard (1999), A microphysically based precipitation scheme for the U. K. Meteorological Office Unified Model, *Q. J. R. Meteorol. Soc.*, *125*, 1607–1636.

C. E. Birch and I. M. Brooks, Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds, Environment Building, Leeds LS2 9JT, UK. (c.birch@see.leeds.ac.uk)

P. Earnshaw and S. F. Milton, Global Model Development and Diagnostics, Met Office, Fitzroy Road, Exeter, Devon EX1 3PB, UK.

P. O. G. Persson, Cooperative Institute for Research in Environmental Sciences, University of Colorado/NOAA/Earth System Research Laboratory, 325 Broadway, Boulder, CO 80305, USA.

S. Söderberg, WeatherTech Scandinavia, Odinslund 2, Dekanhuset, SE-753 10 Uppsala, Sweden.

M. Tjernström, Department of Meteorology, Stockholm University, SE-10691 Stockholm, Sweden.