QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY
*Q. J. R. Meteorol. Soc.* **135**: 2046–2066 (2009)
Published online 21 July 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/qj.444

# A comparison of aircraft-based surface-layer observations over Denmark Strait and the Irminger Sea with meteorological analyses and QuikSCAT winds

I. A. Renfrew,[a]* G. N. Petersen,[a] D. A. J. Sproson,[a] G. W. K. Moore,[b] H. Adiwidjaja,[b]
S. Zhang[b] and R. North[c†]

[a]*School of Environmental Sciences, University of East Anglia, UK*
[b]*Department of Physics, University of Toronto, Canada*
[c]*Met Office, Exeter, UK*

**ABSTRACT:** A compilation of aircraft observations of the atmospheric surface layer is compared with several meteorological analyses and QuikSCAT wind products. The observations are taken during the Greenland Flow Distortion Experiment, in February and March 2007, during cold-air outbreak conditions and moderate to high wind speeds. About 150 data points spread over six days are used, with each data point derived from a 2-min run (equivalent to a 12 km spatial average). The observations were taken 30–50 m above the sea surface and are adjusted to standard heights. Surface-layer temperature, humidity and wind, as well as sea-surface temperature (SST) and surface turbulent fluxes are compared against co-located data from the ECMWF operational analyses, NCEP Global Reanalyses, NCEP North American Regional Reanalyses (NARR), Met Office North Atlantic European (NAE) operational analyses, two MM5 hindcasts, and two QuikSCAT products.

In general, the limited-area models are better at capturing the mesoscale high wind speed features and their associated structure; often the models underestimate the highest wind speeds and gradients. The most significant discrepancies are: a poor simulation of relative humidity by the NCEP global and MM5 models, a cold bias in 2 m air temperature near the sea-ice edge in the NAE model, and an overestimation of wind speed above 20 m $s^{-1}$ in the QuikSCAT wind products. In addition, the NCEP global, NARR and MM5 models all have significant discrepancies associated with the parametrisation of surface turbulent heat fluxes. A high-resolution prescription of the SST field is crucial in this region, although these were not generally used at this time. 

KEY WORDS surface-layer meteorology; surface fluxes; air–sea interaction; GFDex; Nordic Seas

*Received 7 November 2008; Revised 1 April 2009; Accepted 27 April 2009*

## 1. Introduction

The subpolar seas of the North Atlantic Ocean are a current focus of attention due to their key role within the climate system and, for example, through international programmes such as the Arctic–Subarctic Ocean Fluxes (ASOF) initiative (Dickson *et al.*, 2008a). These subpolar seas are where warm saline Atlantic inflow waters are cooled, mixed and overturned, through various mechanisms, eventually resulting in North Atlantic Deep Water (e.g. Dickson and Brown, 1994). These mechanisms are a key component of the ocean's thermohaline circulation. They include: open-ocean convection in the Greenland, Iceland, Norwegian, Labrador and Irminger Seas (e.g. LabSea Group, 1998; Marshall and Schott, 1999; Pickart *et al.* 2003); water mass mixing and modification on the continental shelves (e.g. Bacon *et al.*, 2002; Pickart *et al.*, 2005); and deep-water outflows through Denmark Strait and the Faroes–Scotland Channel (e.g. Bacon, 1998; Østerhus *et al.*, 2008; Dickson *et al.*, 2008b). These ocean processes are directly, or indirectly, forced by fluxes of heat, moisture and momentum at the ocean interface, making this region one where the 'communication' between the atmosphere and ocean is fundamentally important. These fluxes are primarily dictated by the transient weather systems of the region, so to provide accurate air–sea fluxes one must be able to capture the associated synoptic- and meso-scale weather systems. The regional-scale meteorology and climatology of the subpolar seas are rather complex: the North Atlantic stormtrack crosses the region, polar mesoscale cyclones are common, the sea ice is highly variable, and there are some major topographic obstacles such as Greenland and Iceland. Although numerical weather prediction (NWP) models now simulate synoptic-scale weather systems very well, there are still deficiencies at scales below about 1000 km,

*Correspondence to: Dr I. A. Renfrew, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.
E-mail: i.renfrew@uea.ac.uk
†The contribution of R. North was written in the course of her employment at the Met Office, UK and is published with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

i.e. much mesoscale variability is poorly represented or absent from NWP-based analyses (e.g. Chelton and Freilich, 2005; Chelton *et al.*, 2006; Condron *et al.*, 2006).

The objective of this study is to assess the quality of a selection of such NWP models and analyses during high wind speed wintertime conditions using a compilation of surface-layer observations from an aircraft-based field campaign. The aims are

- to quantify typical errors in the NWP analyses;
- to uncover any systematic model biases;
- to ascertain the resolution required to model the mesoscale features; and
- to evaluate the QuikSCAT wind products for this region.

High-quality surface-layer meteorological observations for the subpolar seas are relatively rare, and so despite the importance of validation for ocean studies, there have been relatively few such studies in this region. Renfrew *et al.* (2002) used 40 days of observations from a Labrador Sea Deep Convection Experiment cruise on board the *R/V Knorr* in a comparison with European Centre for Medium-Range Weather Forecasts (ECMWF) analyses and National Centers for Environmental Prediction (NCEP) reanalyses; while Josey (2001), Moore and Renfrew (2002), Josey *et al.* (2002), Sun *et al.* (2003) and Yu *et al.* (2004) used buoy and ship data in comparisons with NWP analyses for the Atlantic Ocean, and Smith *et al.* (2001) carried out an NCEP comparison over all the main oceans. With the exception of the recent comparison of Moore *et al.* (2008), there does not seem to have been any focus on the eastern subpolar seas. Although ship and buoy-based observations are the primary source of observations for such comparison studies, this type of observing platform does have disadvantages. For example, on synoptic time-scales they are essentially time series and so provide limited spatial information.

Using aircraft observations brings advantages and disadvantages. On the plus side, the aircraft is independent of the sea surface, and thus of surface waves, so wind observations should not be influenced by any sort of sheltering effects (e.g. Large *et al.*, 1995) which may affect buoys. Furthermore, due to the short sampling times, if the runs are carefully chosen, then flow distortion effects (a bane for ship-based observations) will not be important. Daily maintenance and expensive instrumentation mean that the quality of the observations should be high. Finally, using an aircraft allows spatial gradients to be sampled and compared to those in the NWP analyses. On the minus side, due to the expense and logistical constraints, relatively few independent data points are generated from any one field campaign and observations are typically spread over only a few days and a relatively small area.

The observations used in this comparison are compiled from the low-level components of six flights from the Greenland Flow Distortion experiment (GFDex). GFDex centered around an aircraft field campaign, based out of Keflavik in Iceland, during February and March 2007. The objectives of the field campaign were to obtain comprehensive observations of a number of mesoscale weather systems associated with the impact of the synoptic-scale atmospheric flow on the high topography of Greenland – such as tip jets, barrier flows and lee cyclones – as well as including a targeted observing component aimed at targeting sensitive area predictions. The six flights compiled here were flown to observe an easterly tip jet case (B268), a polar mesoscale cyclone (B271) and various barrier flow events (B274, B276, B277 & B278). Further details of individual flights can be found in the GFDex overview article by Renfrew *et al.* (2008). The meteorological conditions during the six flights can be categorised simply as 'cold-air outbreaks' associated with moderate to high wind speeds in northerly or north-easterly flows. In general the atmospheric surface layer was slightly unstable and the atmospheric boundary layer (ABL) close to moist neutral, i.e. conditions typical of such cold-air outbreaks (e.g. Brümmer, 1997; Renfrew and Moore, 1999).

The rest of this paper is organised as follows. Section 2 describes the observational data and the various model and satellite-derived products used in the comparison. Section 3 details the methodology used for extracting the comparison time series. Sections 4 and 5 describe the results for the surface-layer meteorology and the surface turbulent fluxes respectively. Section 6 summarises the various model performances, while section 7 concludes the paper.

## 2. Datasets

### 2.1. Observations

Flight-level measurements from the Facility for Airborne Atmospheric Measurement's (FAAM's) BAE-146 have been used to derive an observational 'database' for this comparison study. The database comprises numerous variables at approximately 150 separate times and locations, spread over six days (Figures 1 and 2; Table I). The flight-level observations have been divided into 2-min ($\sim$12 km) runs, and run averages are used for the comparison. The choice of 2 minutes is motivated by the sampling scales required for the turbulent flux analysis described in Petersen and Renfrew (2009). Here we have continued with this run length as 12 km is of a similar size to the grid boxes of the highest-resolution regional models. This suggests that, for these comparisons, each data point represents an independent verification point, i.e. each observational data point is likely to be in a different model grid box. For the lower-resolution products, there may be several observational data points within one grid box. The low-level legs which make up the database were all flown at 'minimum safe altitude' – between 32 and 51 m above the sea surface. The mean run altitude was 39 m, with a mean standard deviation of 3.5 m.

A brief description of the aircraft and the 'core' instrumentation in place during GFDex is provided in Renfrew *et al.* (2008). A more detailed description of the key flight-level measurements, their calibration



DOI: 10.1002/qj

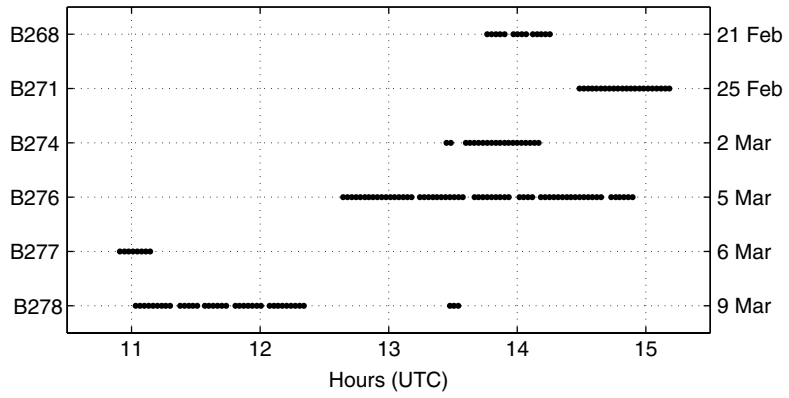


Figure 1. Times and dates of the GFDex observational database used in the comparison.

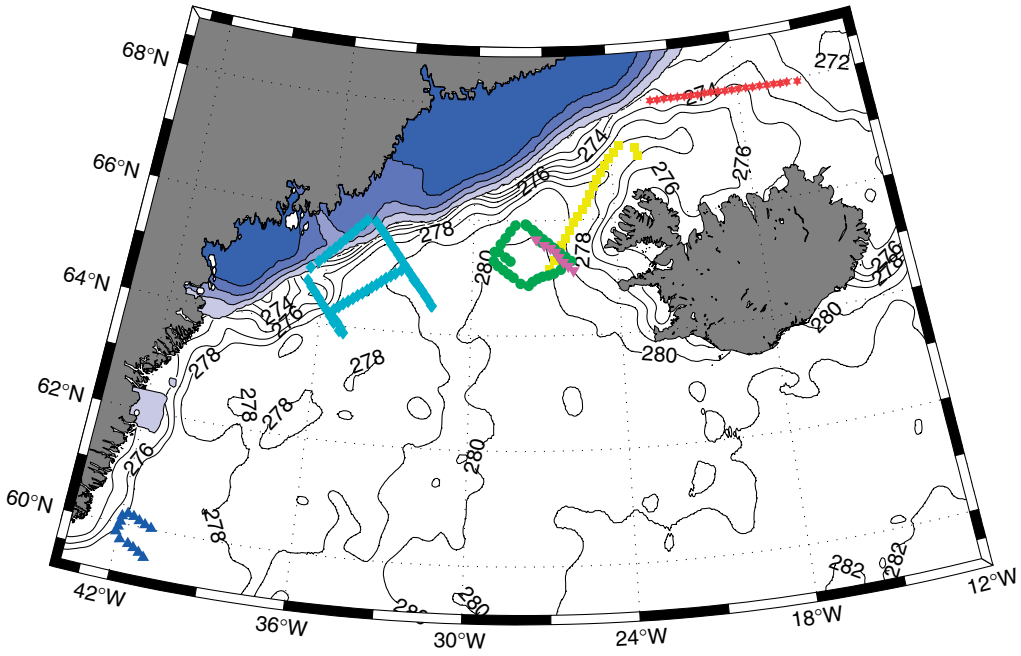


Figure 2. Locations of the observational database points. Data are from the low-level legs of flights B268 (triangles), B271 (stars), B274 (squares), B276 (diamonds), B277 (upturned triangles) and B278 (circles). SST (contours every 1K) and sea-ice concentration (shading every 20%) from the OSTIA dataset on the 5 March 2007 is also shown. This figure is available in colour online at www.interscience.wiley.com/journal/qj

procedures and their accuracies is given in Petersen and Renfrew (2009). A brief summary is reproduced here:

- A five-port turbulence probe on the nose of the aircraft, in conjunction with data from the Inertial Navigation Unit, the static pressure ports and other navigational aids provides three-dimensional wind velocities with an overall uncertainty of less than $\pm 0.5$ m s$^{-1}$.
- Flight-level static pressure is recorded with an uncertainty of order 0.5 hPa.
- Air temperature is measured with Rosemount temperature sensors (non de-iced and de-iced), which have an overall uncertainty of $\pm 0.3$ K, at 95% confidence for typical clear-air conditions, and relative errors $<0.01$ K. The non de-iced sensor is used for all flights except B274 where (due to icing) the de-iced sensor is used. Note that a small bias (determined from the other flights to be 0.48 K) is subtracted from the de-iced temperature to compensate for the warming induced by heating the de-iced sensor.
- Sea-surface temperature (SST) is derived from a downward-looking Heimann radiometer which measures in the range $8 - 14$ µm. Below cloud base, and if calibrated at the start of the leg, this can provide a good estimate of the SST – to a specified accuracy of ~0.7 K. Note that the Heimann estimates a 'skin' SST which, given the location and generally windy conditions of the observations, should be representative of a 'bulk' SST. Unfortunately the Heimann radiometer did not record, or was not properly calibrated, on three flights (B268, B271 and B274). For these days a substitute SST was calculated for use in the bulk flux algorithms, based on a high-resolution (~6 km) satellite-derived dataset from the Met Office – the Operational Sea surface Temperature and sea Ice Analysis (OSTIA) – Stark *et al.* (2007). These



DOI: 10.1002/qj

Table I. Summary of low-level missions during the GFDex field campaign which have been compiled into the observational database. The dropsonde column indicates firstly the number of dropsondes released and secondly the numbers that made the *qg* and *qu* operational data assimilation databases, where the *qg* cycle is the short-cut-off dataset used for the Met Office operational global forecast and the *qu* cycle is the longer-cut-off dataset used for the Met Office operational NAE forecast, as well as for an update for the global forecast. This would also be the number of dropsondes available for the ECMWF global and NCEP reanalyses models. Technical problems forced a delay in take-off for flight B276, explaining why none of these dropsondes made the forecast cycles.

| Flight | Date | Drop-sondes *qg/qu* | Time (mins) | No. of runs | Comments |
|---|---|---|---|---|---|
| B268 | 21 February 2007 | 10 *0/10* | 31 | 14 | Calibrated OSTIA SST used. Due to temperature sensor wetting all $T_{2m}$ and $q_{2m}$ points discarded and no turbulent heat fluxes available. $U_{10}$ calculated assuming neutral conditions. |
| B271 | 25 February 2007 | 16 *15/15* | 46 | 22 | Calibrated OSTIA SST used. |
| B274 | 2 March 2007 | 9 *8/8* | 43 | 20 | Calibrated OSTIA SST used. Due to icing, corrected de-iced temperature sensor used and INU-derived horizontal winds used. |
| B276 | 5 March 2007 | 8 *0/0* | 138 | 62 | The 14 runs over sea ice not used in this study. |
| B277 | 6 March 2007 | 17 *0/11* | 17 | 8 | Due to presence of liquid water, three $T_{2m}$, $q_{2m}$ and $U_{10}$ points discarded. |
| B278 | 9 March 2007 | 6 *0/6* | 90 | 39 | Due to presence of liquid water, two $T_{2m}$, $q_{2m}$ and $U_{10}$ points discarded. |

OSTIA SSTs were compared to the Heiman SSTs where possible and a systematic bias of +1.5 K was found for this region (Petersen and Renfrew, 2009). This bias was corrected for when the OSTIA data were used to determine the bulk fluxes. Note that only the aircraft-observed SSTs are included in the statistical comparisons in this study.

- Humidity is measured using a Lyman-Alpha absorption hygrometer, which measures total water content with an uncertainty of $\pm 0.15$ g kg$^{-1}$. The response of this instrument changes with time, which means it has to be carefully calibrated using a General Eastern cooled-mirror hygrometer. In addition, as the Lyman-Alpha measures total water content, periods when there is liquid water present have to be removed to generate a water vapour measurement. This has been carried out, as described in Petersen and Renfrew (2009), resulting in the elimination of four humidity points from the database.
- Flight-level altitude is measured by a radar altimeter. This has an uncertainty of $\pm 2\%$ below 760 m, so for an altitude of 40 m, the uncertainty is less than $\pm 1$ m.

The flight-level meteorological observations have been adjusted to 'standard' levels by using well-established stability-dependent surface-layer similarity theory (e.g. Smith, 1988; Renfrew *et al.*, 2002; Fairall *et al.*, 2003). In particular, the flight-level wind speed has been adjusted to a 10 m wind speed ($U_{10m}$) for comparison with the models and to a 10 m neutral wind speed ($U_{10N}$) for comparison with the QuikSCAT data; and the flight-level temperature and humidity data have been adjusted to 2 m temperature ($T_{2m}$), specific humidity ($q_{2m}$) and relative humidity ($RH_{2m}$). The latter is calculated from $q_{2m}$ and the saturated specific humidity at 2 m (calculated from $T_{2m}$) and, for consistency, the same procedure has been used to calculate model values of $RH_{2m}$. The flight-level static pressure has been adjusted to mean-sea-level pressure (MSLP) using hydrostatic balance with the run-mean altitude and density. The flight-level wind direction has not been adjusted to 10 m, but is assumed to be the same as observed at flight level.

We have used the COARE 3.0 bulk flux algorithm (Fairall *et al.*, 2003) for the above stability-dependent adjustments. The algorithm uses flight-level data to calculate the scaling parameters for wind (i.e. the friction velocity), temperature and humidity and the surface roughness lengths for wind, temperature and humidity. A sensitivity test using the eddy-correlation-derived scaling parameters and roughness lengths to carry out the same adjustment showed that there was no significant difference between the two adjustment methods: the mean differences for $U_{10m}$, $T_{2m}$ and $q_{2m}$ were 0.08 m s$^{-1}$, 0.08 K and 0.04 g kg$^{-1}$, respectively, with standard deviations of these differences of 0.22 m s$^{-1}$, 0.11 K and 0.04 g kg$^{-1}$. It was decided to use the COARE scaling parameters and roughness lengths so as to avoid the occasional outliers that are calculated using the covariance method and which are associated with the relatively large flux-sampling errors inherent in turbulent observations.

In addition to the surface-layer meteorological variables, comparisons against surface turbulent flux estimates are carried out. Direct calculations of the turbulent fluxes of momentum, sensible heat and latent heat using the covariance (eddy correlation) method are available



DOI: 10.1002/qj

for flights B268 (momentum only), B271, B276, B277 and B278 (Petersen and Renfrew, 2009). However these are *not* used for a direct comparison to the model data, because of the large variability in these turbulent quantities associated with flux-sampling errors (e.g. Fairall *et al.*, 2003; Petersen and Renfrew, 2009). Instead comparisons are made to flux estimates calculated using two commonly used bulk flux algorithms: the COARE 3.0 algorithm (as detailed in Fairall *et al.*, 2003) and the Smith (1988) algorithm, which uses constant values of heat and moisture exchange coefficients, in this case set to $C_{\mathrm{HN}} = 1.14 \times 10^{-3}$ and $C_{\mathrm{EN}} = 1.2 \times 10^{-3}$, based on the DeCosmo *et al.* (1996) study with adjustments for salinity effects (Fairall *et al.*, 2003). The COARE 3.0 algorithm is an updated and improved version of the previous COARE algorithm (Fairall *et al.*, 1996), in particular extending the wind speed range for which the algorithm is appropriate by the inclusion of a Charnock coefficient that depends on $U_{10\mathrm{N}}$ and by validation against observations in high wind speed conditions (e.g. Persson *et al.*, 2005). Variants of the Smith (1988) algorithm have been shown to be appropriate for high wind speed conditions in previous studies, e.g. Renfrew *et al.*, (2002).

Of a total of 165 points in the database, 14 are neglected as the aircraft was over sea ice or the marginal ice zone. Note that there is a slight mismatch between the observed sea-ice distribution and the sea-ice fields prescribed in the NWP models. This mismatch is a function of the model resolution. To allow for this, it was decided to neglect an additional five data points which were close to the ice edge (in reality) and had interpolated SSTs <271.5 K in the majority of the models. Hence the maximum size of the comparison dataset is 146 points. This is the number of data points used in the comparisons of MSLP and wind direction. Following the quality control procedures described above, this number is further reduced to 138 for $U_{10\mathrm{m}}$ and $U_{10\mathrm{N}}$; to 127 for $T_{2\mathrm{m}}$, $q_{2\mathrm{m}}$, $RH_{2\mathrm{m}}$ and the bulk fluxes and to 91 for SST. Note that due to the limited number of observations available, no statistical comparisons over the sea ice have been carried out.

### 2.2. ECMWF analyses

ECMWF operational global analyses are available every 6 hours. During 2007 the operational deterministic model at the ECMWF was running with a spectral truncation of T799 (equivalent to a grid spacing of ∼25 km). We extracted ECMWF time series at two resolutions. Firstly at a truncation of T511, or from an N400 reduced Gaussian grid for the surface variables, which is equivalent to a resolution of ∼40 km. This was the highest resolution data to which we had access. Secondly at a resolution of 1.125 degrees (from truncation T159 in the atmosphere and an N80 reduced Gaussian grid for surface variables) so equivalent to a resolution of ∼125 km. The latter is probably the most commonly used ECMWF analysis resolution and is the same as that used in the ERA-40 reanalyses (Uppala *et al.*, 2005). The surface-layer variables (pressure, temperature, humidity and wind) are available from the analyses fields, while the surface fluxes are 3-hour averages from the forecast initialised at 1200 UTC. The surface-layer variables are calculated by stability-dependent interpolation from the lowest model level (∼10 m). Note that although these variables are from model analyses, they are heavily influenced by the short-term forecasts that are used as part of the analysis procedure.

At this time the ECMWF used 0.5° resolution SST and sea-ice fields based on those provided by NCEP. The SST field was optimally interpolated from satellite, buoy and ship measurements. The sea-ice field was based on SSM/I passive microwave satellite-based measurements. ECMWF use a fractional sea-ice concentration with concentrations of <20% set to zero. Note that from 30 September 2008 the ECMWF system changed to use the OSTIA SST and sea-ice analysis. Note also that from June 1998 the ECMWF model has employed a two-way coupled wave model which can modify the near-surface winds. A recent assessment of the ECMWF's operational forecasting system for the Arctic region is provided by Jung and Leutbecher (2007).

### 2.3. NCEP reanalyses

NCEP reanalysis data are also available at two resolutions. Firstly from the commonly used NCEP/NCAR Global Reanalysis (Kalnay *et al.*, 1996; Kistler *et al.*, 2001); and secondly from the NCEP North American Regional Reanalysis (NARR; Mesinger *et al.*, 2006). The NCEP Global Reanalyses are available every 6 hours with the model run at a truncation of T62. The surface layer fields are available on a 2.5° × 2.5° latitude–longitude grid (equivalent to a horizontal resolution of ∼250 km). Clearly this global product is at a rather coarse resolution for this comparison, but as it is commonly used for forcing ocean models (either directly or via the CORE product; Large and Yeager, 2004), it is include here for completeness.

The NARR are available every 3 hours over much of the Northern Hemisphere from the central Pacific to the central Atlantic. Our domain of interest is close to the eastern boundary of the NARR domain. The NARR is based on the NCEP operational ETA model and is run with horizontal resolution of 32 km and with 45 vertical levels (Mesinger *et al.*, 2006). It has a significantly improved land surface model, when compared to the global model (e.g. Janjić, 1994; Chen *et al.*, 1997; Ek *et al.*, 2003), which has been thoroughly evaluated over land (Chen *et al.*, 1997; Ek *et al.*, 2003; Berbery *et al.*, 2003), although perhaps with less attention paid to the surface layer over the sea. The NARR uses a 1° Reynolds-based SST field, while in both reanalysis products a 100% or 0% sea-ice concentration is available. Unfortunately, due to a technical mistake, surface momentum flux data are not available from the NARR (F. Mesinger, personal communication, 2006).



DOI: 10.1002/qj

### 2.4. Met Office NAE analyses

Operational analysis and forecast data from the Met Office's NAE (North Atlantic and European) regional model have been obtained. The NAE is a limited-area version of the Met Office's Unified Model. At the time of GFDex, the NAE used version 6.1 of the Unified Model, with a horizontal resolution of 0.11° (∼12 km), using 600 × 360 grid points and 38 vertical levels. The NAE uses boundary conditions from the Met Office's global forecast, but is otherwise run as a stand-alone forecast model. Its domain covers most of the North Atlantic and Europe and a rotated pole configuration is used to allow a more even grid spacing across the domain. The Met Office Unified Model has a relatively sophisticated boundary-layer parametrisation scheme, based on diagnosing boundary-layer types and then applying local or non-local schemes, dependent upon the type (Lock *et al.*, 2000; Martin *et al.*, 2000). A few upgrades to the ABL scheme, as described in Brown *et al.* (2008), became operational in the NAE model on 26 September 2006. At the time of GFDex an SST analysis was performed daily combining *in situ* observations and either the NESDIS (National Environmental Satellite, Data, and Information Service) 50 × 50 km resolution product or, if not available, the 100 × 100 km resolution product. The NESDIS data were bias-corrected before being used. The background field was the previous day's analysis. The sea-ice field was derived from an NCEP product based on SSM/I passive microwave radiances and was available at 1/12°. Sea-ice concentrations less than 50% were set to zero. Note that in November 2007 (after GFDex) the SST boundary conditions for the NAE were changed to the high-resolution OSTIA analysis mentioned earlier. The sea-ice analysis was changed to that from the OSTIA system in July 2008.

It should be noted that the operational NAE had a minor upgrade on 6 March 2007, halfway through the GFDex campaign, so affecting observations from B277 and B278 in this study. In brief, these upgrades included (i) a change in satellite data usage in the data assimilation; (ii) a change to a dependence on $w$ (vertical velocity) instead of RH in the convection closure; (iii) a revision to the treatment of Exner pressure in the lateral boundary condition zone; and (iv) a revision to the aerosol/radiation coupling. It is not thought that these changes will affect the results of this study.

### 2.5. MM5 hindcast

The Pennsylvania State University–NCAR fifth-generation Mesoscale Model (MM5) version 3.7 has been employed to carry out a series of 36-hour hindcasts. For the 3-week GFDex period, the model was re-initialized at 0000 UTC every day and integrated for 36 hours. To allow for model spin-up, the first 12 hours of each integration was discarded. A single domain was used, centred at 67°N, 30°W and extending from about 60°W to 0°W and 50°N to 80°N. A grid size of 12 km was used, with 340 × 320 grid points in the horizontal. The MM5 configuration used 25 unevenly spaced levels in the vertical, with 10 levels in the lowest 1 km, the lowest at about 10 m. The model top is set at 50 hPa with a radiative upper boundary condition to minimize the reflection of internal gravity waves.

MM5 was configured using parametrisations that have been tried and tested for the polar regions i.e. the Polar MM5 configuration. These parametrisations include: (i) The Kain-Fritsch convective parametrisation with the shallow convective effects (Kain and Fritsch, 1993); (ii) The Goddard microphysics explicit moisture scheme with graupel/hail as an additional variable (Tao *et al.*, 1989; Tao and Simpson, 1993); (iii) A non-local planetary boundary layer scheme (Hong and Pan, 1996); (iv) the Dudhia radiation scheme; (v) A multi-layer soil model to predict land surface temperatures (Dudhia, 1996); and (vi) a modification to allow for the sea-ice fraction (Bromwich *et al.*, 2001; Cassano *et al.*, 2001).

The MM5 hindcasts were initialised and had boundary conditions provided from the NCEP final analyses (FNL) data, which is available 6-hourly at a resolution of 1° × 1°. Two sets of hindcasts were carried out. In the first (MM5-FNL) the NCEP 1° × 1° SST data and a 25 km National Snow and Ice Data Center (NSIDC) sea-ice and snow extent dataset were used as boundary conditions. In the second (MM5-OSTIA) the high-resolution (∼6 km) OSTIA SST and sea-ice data were used as boundary conditions (Stark *et al.*, 2007).

### 2.6. QuikSCAT data

Satellite scatterometer winds from the Seawinds scatterometer on board the QuikSCAT satellite have provided, for the first time, twice-daily near-global measurements of near-surface wind speed and direction over the world's oceans. Two versions of Level 2B QuikSCAT wind products have been used in this comparison. The 0.25 degree pass data (Version 3 retrieval algorithm) available from Remote Sensing Systems (referred to here as QS-RSS), and the 25 km NASA DIRTH product available from NASA's Jet Propulsion Laboratory (referred to as QS-NASA-D). The NASA DIRTH product uses a more advanced wind direction ambiguity removal algorithm (JPL, 2006) which is generally more reliable than the 'standard' NASA product (indeed this was what we found). The DIRTH algorithm is used to generate the NASA L3 gridded wind products. Both QuikSCAT products have been processed and quality controlled along broadly similar lines, e.g. Ebuchi *et al.* (2002). The QS-RSS winds are retrieved using the 'Ku-2001' model (Wentz *et al.*, 2001; and http://www.remss.com). They are processed for rain flagging and sea-ice detection by making recourse to contemporaneous satellite microwave measurements. Wentz *et al.* (2001) quote typical accuracies of 1 m $s^{-1}$ and 15 degrees in 10 m wind speed and direction. In a comparison with over 30 000 buoy data, Ebuchi *et al.* (2002) find correlation coefficients of 0.925 and 0.959–0.977 for wind speed and direction and rms differences of 1.01 m $s^{-1}$ and 26.5–18.6 degrees. The QS-NASA-D winds are retrieved using a



DOI: 10.1002/qj

'Wind Vector Cell' retrieval algorithm (JPL, 2006). Similar rain flagging and sea-ice detection checks are implemented (e.g. Ebuchi *et al.*, 2002; Chelton and Freilich, 2005; JPL, 2006). In the same buoy comparison study, Ebuchi *et al.* (2002) find correlation coefficients of 0.925 and 0.948–0.973 for wind speed and direction, with rms differences of 1.01 m $s^{-1}$ and 29.6–19.5 degrees. Chelton and Freilich (2005) find similar statistics in a separate comparison against buoy data. They conclude that QuikSCAT winds have component errors of about 0.75 m $s^{-1}$ in the along-wind direction and 1.5 m $s^{-1}$ in the cross-wind direction.

It is perhaps worth noting that scatterometers do not actually measure winds, rather they measure radar backscatter cross-section, which is dependent on small-scale surface roughness, which depends on surface stress. Hence *surface stress* is the most direct geophysical variable that is calculated and wind speeds are derived assuming neutral stratification, i.e. the QuikSCAT wind product is equivalent to a neutral 10 m wind speed ($U_{10N}$). The geophysical retrieval models have been tested for wind speeds up to 65 m $s^{-1}$ (Yueh *et al.*, 2001; Fernandez *et al.*, 2006). Fernandez *et al.* (2006) present aircraft observations of 10 m winds measured indirectly using the SFMR (Simultaneous multi-Frequency Microwave Radiometer) which provides $U_{10N}$ winds equivalent to a 10 min average.[†] They examine aircraft-observed normalised radar cross-section and show a clear flattening of reponse with wind speed when $U_{10N} > 25$ m $s^{-1}$. Consequently wind retrievals for scatterometers in high wind speed conditions are more difficult. Fernandez *et al.* suggest modifications to the geophysical retrieval models for winds in the range 25–65 m $s^{-1}$, but as far as we are aware these have not yet been implemented.

There has been little *in situ* validation of QuikSCAT data for wind speeds greater than about 20 m $s^{-1}$ as buoys tend not to be located in such windy locations. The buoy observations used by Ebuchi *et al.* (2002) and Chelton and Freilich (2005) sampled almost exclusively wind speeds lower than this value, with the few comparisons available for winds greater than 20 m $s^{-1}$ having uncomfortably large residuals. Recently Moore *et al.* (2008) have presented a limited QuikSCAT comparison against buoy data from Cape Farewell, Greenland, the windiest location in the world ocean. Their comparison showed relatively large rms errors (2.5 m $s^{-1}$) at this location, and hinted at a systematic discrepancy at higher wind speeds (greater than ~17 m $s^{-1}$). However there must always be some concern about the accuracy of buoy observations during such conditions, where there is the possibility of sheltering effects due to the concomittant high waves (e.g. Large *et al.*, 1995).

[†]Note that in a comparison of SFMR winds against GPS dropsonde observations, the overall rms difference was 3.3 m $s^{-1}$; Uhlhorn and Black, 2003.

### 2.7. Discussion

It is worth noting that, during the GFDex field campaign, 144 dropsondes were released from the FAAM aircraft and a number of radiosonde stations in Greenland and Iceland, as well as E-ASAP (EU-Automated Shipboard Aerological Programme) vessels, were carrying out additional (0600 and 1800 UTC) soundings. All of this *additional* GFDex sounding data was transmitted onto the GTS in real time. The vast majority of the dropsonde data made the 1200 UTC forecast cycles at the meteorological agencies (Table I). This being the case, the ECMWF, NCEP and Met Office NWP products will all have been influenced by the GFDex dropsonde data on the day of that flight. Only the MM5 hindcasts do not include this data, as they are initialised from 0000 UTC. Due to this additional sounding data, one might anticipate that the NWP products would be better than average at this time for this region. In other words, this comparison probably represents an upper bound on analyses quality. Despite this fact, a comparison of surface-layer and surface flux variables is still instructive. It provides a rigorous test of whether the models are able to simulate and parametrise the atmospheric boundary layer and associated air–sea fluxes during these cold-air outbreak events, given a reasonably well-constrained set of initial conditions.

## 3. Methodology

The strategy employed in this study has been to assess each model or satellite-derived product in the best way for that product, i.e. to compare at the best temporal and spatial resolution available. In other words, we are assessing the quality of each product against the observational 'truth', rather than carrying out a strict intercomparison. The basic methodology has been to extract model and satellite-derived 'time series' by matching in time and space against the observational data. Figure 1 illustrates the temporal range of the observational data (betweeen 1100 and 1500 UTC over six days), while Figure 2 illustrates the spatial domain of the observations; the data points cover approximately 2000 km of the atmospheric surface layer.

A matching in *time* is carried out as follows:

- The ECMWF analyses and the NCEP Global Reanalysis data are available only every 6 hours and so the 1200 UTC model fields are used with no adjustment for temporal differences.
- The NCEP North American Regional Reanalysis data are available every 3 hours. Model data are extracted at one analysis time for each day, the closer of either the 1200 or 1500 UTC analysis.
- The Met Office NAE data are available every hour from forecasts run every 6 hours. Only data from the 1200 UTC run on the day of comparison were used. Data were extracted at the nearest hour to the observation time.
- The MM5 hindcast simulations have output fields every hour. The MM5 data are extracted at the



DOI: 10.1002/qj

nearest hour to the mean time of the observations on each day.

- QuikSCAT satellite passes are typically available in the early morning (0600–0900 UTC) and the early evening (1800–2100 UTC). The swath data have been linearly interpolated in time to the mean time of the observations on each day. Sensitivity tests using data from a continuously recording meteorological buoy (Moore *et al.*, 2008) suggest that, under these sort of meteorological conditions, the autocorrelation time-scale is such that this interpolation is a reasonable approach. The correlation between an interpolation over the 12-hour period and the truth is 0.85, suggesting such a comparison is worthwhile. Furthermore, the GFDex flights all targeted well-defined meteorological features (e.g. tip jets, barrier winds, mesoscale cyclones) which were all reasonably long-lived. A check of the morning and afternoon passes for each flight showed that any well-defined spatial gradients in wind speed were consistent during the day (not shown).

A matching in *space* is carried out using triangular (Delaunay-based) linear interpolation from the three nearest gridpoints. Triangular interpolation has an advantage over bi-linear interpolation for non-regular grids such as the swath grids of the QuikSCAT data (e.g. Chelton and Freilich, 2005).

## 4. Surface-layer data comparison

This section comprises a general comparison of surface-layer data and then a more quantitative comparison for selected variables via scatter plots and tabulated statistics. Figure 3 shows 'time series' style plots of the key surface-layer meteorological variables, although it is worth reiterating that these plots should be more accurately described as a compilation of six 'spatial series' from the six flights (only data from the NAE model are from different times). The aircraft observations are, in essence, snapshots of the atmosphere and so these plots illustrate how well the models are able to capture spatial gradients, as well as their general correspondence. In general the models do capture the day-to-day variability and the major spatial gradients seen in the observations. In other words, the models are generally able to capture the mesoscale gradients in the ABL associated with these six different synoptic-scale situations. However there is rather a large range in some variables (the temperatures, relative humidity and wind speed) and some obvious biases, suggesting a quantitative comparison is required. To this end, Table II notes means and standard deviations of all variables. A linear regression has been performed with the aircraft

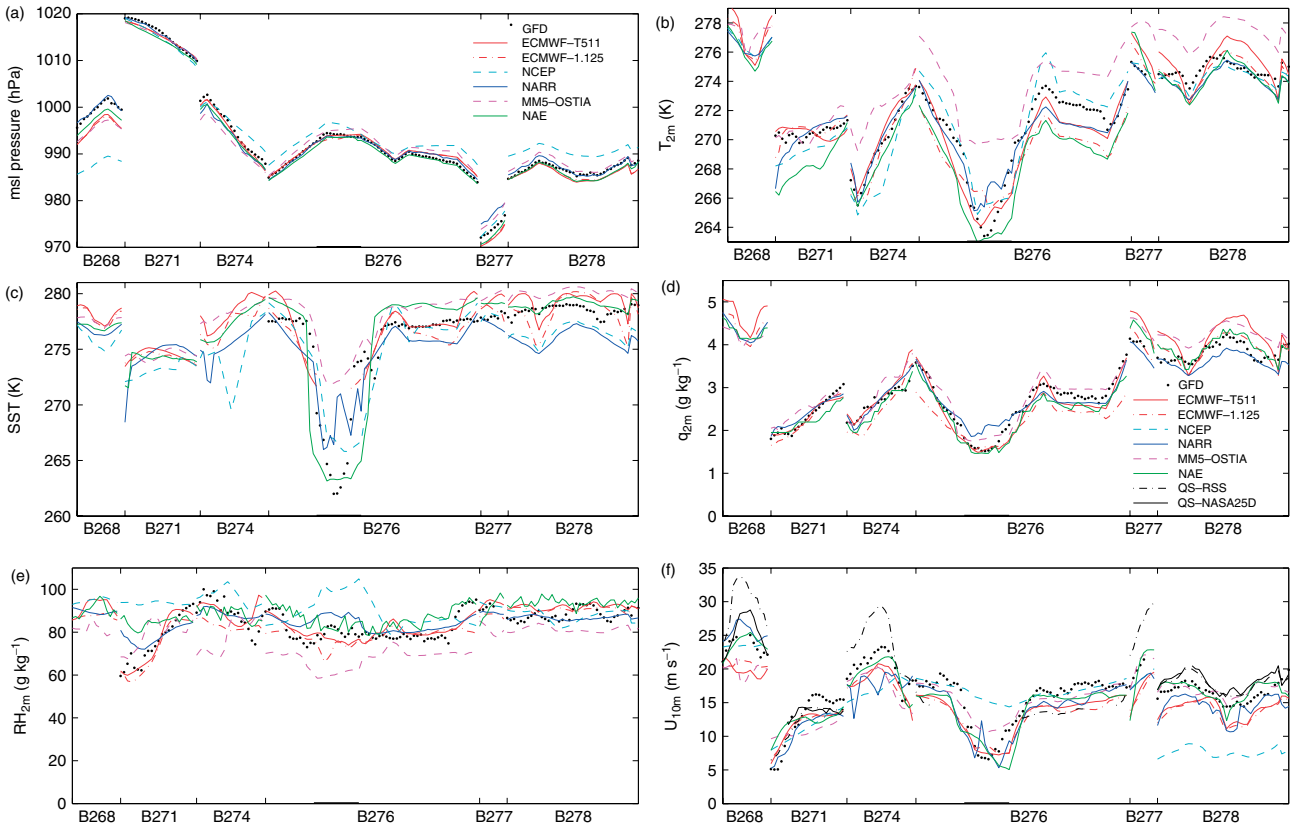


Figure 3. Spatial 'time series' plots showing observations (dots) and model or satellite products (see legend) for the 6 low-level flights. Each observation is a 2-minute (∼12 km) average. The variables shown are (a) mean-sea-level pressure, mslp, (b) 2 m air temperature, $T_{2m}$, (c) SST (d) 2 m specific humidity, $q_{2m}$, (e) 2 m relative humidity, $RH_{2m}$, and (f) 10 m wind speed, $U_{10m}$. A bold horizontal line on each panel marks where the data are taken over observed sea ice.



DOI: 10.1002/qj

Table II. Means and standard deviations of the surface-layer meteorological data over water. The comparison products are analyses from: the ECMWF at 1.125 deg. and T511/N400 resolutions; NCEP Global Reanalyses; NCEP North American Regional Reanalyses (NARR); Met Office operational analyses from the NAE; MM5 hindcasts using NFL and OSTIA surface boundary conditions; and QuikSCAT winds using the RSS and NASA-DIRTH retrieval algorithms.

| | MSLP (hPa) | $T_{2m}$ (K) | SST (K) | $q_{2m}$ (g kg$^{-1}$) | $RH_{2m}$ (%) | $U_{10m}$ (m s$^{-1}$) | $WD$ (deg.) | |
|---|---|---|---|---|---|---|---|---|
| Mean | **993.1** | **272.3** | **277.8** | **3.10** | **84** | **17.3** | **29** | **Aircraft** |
| | 992.5 | 271.6 | 277.6 | 2.92 | 83 | 14.8 | 33 | ECMWF-1deg |
| | 992.4 | 272.4 | 278.5 | 3.19 | 86 | 15.1 | 32 | ECMWF-T511 |
| | 993.8 | 272.2 | 276.8 | 3.22 | 89 | 14.2 | 34 | NCEP-reanalysis |
| | 993.2 | 271.9 | 276.1 | 3.00 | 84 | 15.8 | 22 | NCEP-NARR |
| | 992.1 | 271.1 | 278.9 | 3.02 | 89 | 16.6 | 32 | Met Office NAE |
| | 993.6 | 272.9 | 276.5 | 2.97 | 77 | 16.2 | 35 | UT-MM5-NFL |
| | 993.1 | 274.6 | 279.5 | 3.34 | 77 | 16.3 | 36 | UT-MM5-OSTIA |
| | – | – | – | – | – | 18.1 | 22 | QS-RSS |
| | – | – | – | – | – | 18.0 | 33 | QS-NASA-D |
| Std. dev. | **10.9** | **2.2** | **1.1** | **0.64** | **8** | **3.4** | **30** | **Aircraft** |
| | 10.8 | 2.6 | 1.6 | 0.79 | 8 | 2.8 | 26 | ECMWF-1deg |
| | 10.9 | 2.6 | 1.6 | 0.85 | 8 | 2.7 | 26 | ECMWF-T511 |
| | 9.7 | 2.8 | 1.3 | 0.51 | 5 | 5.0 | 39 | NCEP-reanalysis |
| | 10.5 | 2.1 | 0.9 | 0.56 | 4 | 3.8 | 29 | NCEP-NARR |
| | 10.7 | 2.9 | 0.6 | 0.76 | 5 | 3.2 | 28 | Met Office NAE |
| | 9.8 | 2.3 | 1.0 | 0.64 | 8 | 2.8 | 25 | UT-MM5-NFL |
| | 10.0 | 2.7 | 0.9 | 0.75 | 6 | 2.9 | 27 | UT-MM5-OSTIA |
| | – | – | – | – | – | 5.8 | 32 | QS-RSS |
| | – | – | – | – | – | 3.8 | 28 | QS-NASA-D |

observations as the independent variable and the model data the dependent variable. Table III notes the correlation coefficient, $r$, between the aircraft observations and each model time series, the slope of the linear regression line, the bias error (which is the mean difference between the time series), and the rms error. A correlation coefficient and slope close to 1 indicate the variable is well modelled.

Figure 3 illustrates that, as one might expect, the MSLP compares very well for most models ($r$ typically 0.99 and regression slopes close to 1 in Table III), the exception being the coarse-resolution NCEP Global Reanalyses which is noticeably too low on B268 and too high on B274, B276 and B278, suggesting the 250 km resolution of this model is not sufficient to capture these events. Excluding the NCEP Global, the largest MSLP ranges are for B268 and B277 – the events with the strongest winds. B277 was a strong barrier wind event observed as the flow exited Denmark Strait, where pronouced ageostrophic forcing is present (Petersen *et al.*, 2009). It is also the case that the dropsondes released during this flight were too late for the forecast cycles (Table I). Overall, both the ECMWF model and the NAE have a small negative bias in MSLP.

Figures 3(b) and (c) show $T_{2m}$ and SST. The range of model temperatures is surprisingly large, typically $\pm 2$ K for any one point, illustrating that the models' performances here are varied. In general, all the models do a reasonable job in capturing the spatial gradients, although there are exceptions. For B271 (from the wake of Cape Tobin, Greenland, into a polar low; Figure 2 and Renfrew *et al.*, 2008), the relatively small gradient in $T_{2m}$ is overestimated by the NARR and NAE, and underestimated by the ECMWF-T511. For B277, the $T_{2m}$ gradient is overestimated by the ECMWF models and the NAE. Inspection of the SST plot suggests that some of these difficulties, as well as the surprisingly large range of temperatures, can be attributed to this surface boundary condition. There is a close coupling between the SST and surface-layer air temperatures and so any significant discrepancies in SST will be reflected in $T_{2m}$. As illustrated in Figure 2, the observations are generally taken less than 200 km from the sea-ice edge and close to the continental shelf break, so in a region where the SST gradients are relatively large. Mismatches between the model's sea-ice and SST fields and reality are the cause of some of the abrupt discrepancies seen in Figure 3 (e.g. the spikes in ECMWF and NARR SSTs in B278). Comparing the various model SSTs against the observations, the broad-scale spatial gradients correspond, but the details do not.

Figures 4 and 5 show scatter plots of $T_{2m}$ and SST for each of the model comparisons. The correlations and slopes are generally good for $T_{2m}$ ($r = 0.90$–$0.95$, Table III), but there are clear biases of $-0.7$ K for the ECMWF-1deg; $-1.3$ K for the NAE; and 2.3 K for the MM5-OSTIA. Interestingly it seems the ECMWF-1deg bias is simply due to insufficient resolution in the archived data, because the ECMWF-T511 comparison has zero bias. The $T_{2m}$ cold bias in the NAE is despite a 1.1 K warm bias in the SST. Referring to Figure 3, it



DOI: 10.1002/qj

Table III. Statistical comparisons for the surface-layer meteorological data over water; see text for details. To avoid false statistics for this comparison, the wind direction (*WD*) data have been rotated to be between −180 and +180 degrees. The bias and RMS errors are dimensional with units as specified in Table II.

| | MSLP | $T_{2m}$ | SST | $q_{2m}$ | $RH_{2m}$ | $U_{10m}$ | *WD* | Aircraft versus |
|---|---|---|---|---|---|---|---|---|
| Correlation coefficient | 0.99 | 0.91 | 0.51 | 0.94 | 0.72 | 0.93 | 0.94 | ECMWF-1deg |
| | 0.99 | 0.92 | 0.62 | 0.96 | 0.83 | 0.92 | 0.94 | ECMWF-T511 |
| | 0.92 | 0.92 | 0.42 | 0.87 | 0.09 | 0.62 | 0.73 | NCEP-reanalysis |
| | 0.99 | 0.95 | 0.25 | 0.97 | 0.86 | 0.88 | 0.95 | NCEP-NARR |
| | 0.99 | 0.93 | 0.58 | 0.97 | 0.64 | 0.87 | 0.91 | Met Office NAE |
| | 0.99 | 0.90 | 0.32 | 0.94 | 0.32 | 0.83 | 0.90 | UT-MM5-NFL |
| | 0.99 | 0.90 | 0.90 | 0.97 | 0.20 | 0.81 | 0.90 | UT-MM5-OSTIA |
| | – | – | – | – | – | 0.88 | 0.81 | QS-RSS |
| | – | – | – | – | – | 0.90 | 0.92 | QS-NASA-D |
| Slope | 0.97 | 1.07 | 0.74 | 1.15 | 0.71 | 0.77 | 0.83 | ECMWF-1deg |
| | 0.99 | 1.09 | 0.90 | 1.27 | 0.85 | 0.73 | 0.81 | ECMWF-T511 |
| | 0.81 | 1.14 | 0.53 | 0.70 | 0.06 | 0.93 | 0.95 | NCEP-reanalysis |
| | 0.96 | 0.90 | 0.22 | 0.85 | 0.46 | 0.99 | 0.92 | NCEP-NARR |
| | 0.98 | 1.18 | 0.33 | 1.15 | 0.38 | 0.83 | 0.86 | Met Office NAE |
| | 0.88 | 0.91 | 0.31 | 0.94 | 0.32 | 0.70 | 0.77 | UT-MM5-NFL |
| | 0.90 | 1.08 | 0.78 | 1.13 | 0.16 | 0.69 | 0.82 | UT-MM5-OSTIA |
| | – | – | – | – | – | 1.39 | 0.80 | QS-RSS |
| | – | – | – | – | – | 1.06 | 0.82 | QS-NASA-D |
| Bias error | −0.7 | −0.7 | −0.2 | −0.18 | −1.7 | −2.5 | 4 | ECMWF-1deg |
| | −0.7 | 0.0 | 0.7 | 0.09 | 1.2 | −2.2 | 3 | ECMWF-T511 |
| | 0.7 | −0.2 | −1.0 | 0.19 | 5.1 | −3.1 | 5 | NCEP-reanalysis |
| | 0.1 | −0.4 | −1.7 | −0.10 | 0.0 | −1.5 | −7 | NCEP-NARR |
| | −1.0 | −1.3 | 1.1 | −0.08 | 4.9 | −0.7 | 3 | Met Office NAE |
| | 0.5 | 0.6 | −1.3 | −0.13 | −7.0 | −1.1 | 6 | UT-MM5-NFL |
| | 0.0 | 2.3 | 1.7 | 0.24 | −7.5 | −1.0 | 8 | UT-MM5-OSTIA |
| | – | – | – | – | – | 0.8 | −7 | QS-RSS |
| | – | – | – | – | – | 1.0 | 5 | QS-NASA-D |
| RMS error | 1.5 | 1.3 | 1.4 | 0.34 | 6.2 | 2.8 | 11 | ECMWF-1deg |
| | 1.5 | 1.0 | 1.4 | 0.31 | 4.9 | 2.6 | 11 | ECMWF-T511 |
| | 4.4 | 1.2 | 1.6 | 0.34 | 10.5 | 5.0 | 27 | NCEP-reanalysis |
| | 1.1 | 0.8 | 2.1 | 0.19 | 4.9 | 2.4 | 12 | NCEP-NARR |
| | 1.2 | 1.7 | 1.4 | 0.23 | 7.9 | 1.8 | 13 | Met Office NAE |
| | 2.2 | 1.2 | 1.8 | 0.25 | 11.7 | 2.2 | 14 | UT-MM5-NFL |
| | 2.0 | 2.5 | 1.7 | 0.32 | 11.9 | 2.2 | 15 | UT-MM5-OSTIA |
| | – | – | – | – | – | 3.3 | 21 | QS-RSS |
| | – | – | – | – | – | 1.9 | 13 | QS-NASA-D |

seems to be notable only during B271 and B276 – the flights closest to the sea ice. The NAE's wind direction corresponds well (Tables II, III) indicating no systematic advection errors, which suggests a more local problem with the model's ABL scheme over and immediately downwind of the model's sea ice during these sort of cold-air outbreak conditions. A comparison of NAE versus observed soundings generally shows a cold bias throughout the ABL over and adjacent to areas of sea ice during B271 and B276 (as discussed above) and also during flights B273 and B274. Away from the sea-ice edge, the NAE's temperature soundings generally compare very well. Hence we suspect that a parametrisation problem over model sea ice is the source of the problem. In the NAE operational model, sea ice is set to 2 m thick. In reality, sea ice in this region is highly heterogeneous, being a mixture of old (thick) ice advected out of the Arctic via the Fram Strait and new (thin) ice which has frozen over winter in the Greenland and Iceland Seas. It may be the case that a uniform setting of 2 m thickness provides too much insulation of the cold air from the warmer waters. In case-study simulations of the barrier flow cases, we have experimented with this and found that changing the ice thickness to 0.6 m warmed the marine ABL by 0.5–1 K adjacent to the sea ice.

The 2.3 K bias in the MM5-OSTIA data is concomitant with a 1.7 K bias in the SST. Indeed, when the MM5 model was run with the default NFL SST boundary condition (which had a SST bias of −1.3 K) the corresponding $T_{2m}$ bias was only 0.6 K – confirming the key role the SST plays in determining the ABL temperatures.



DOI: 10.1002/qj

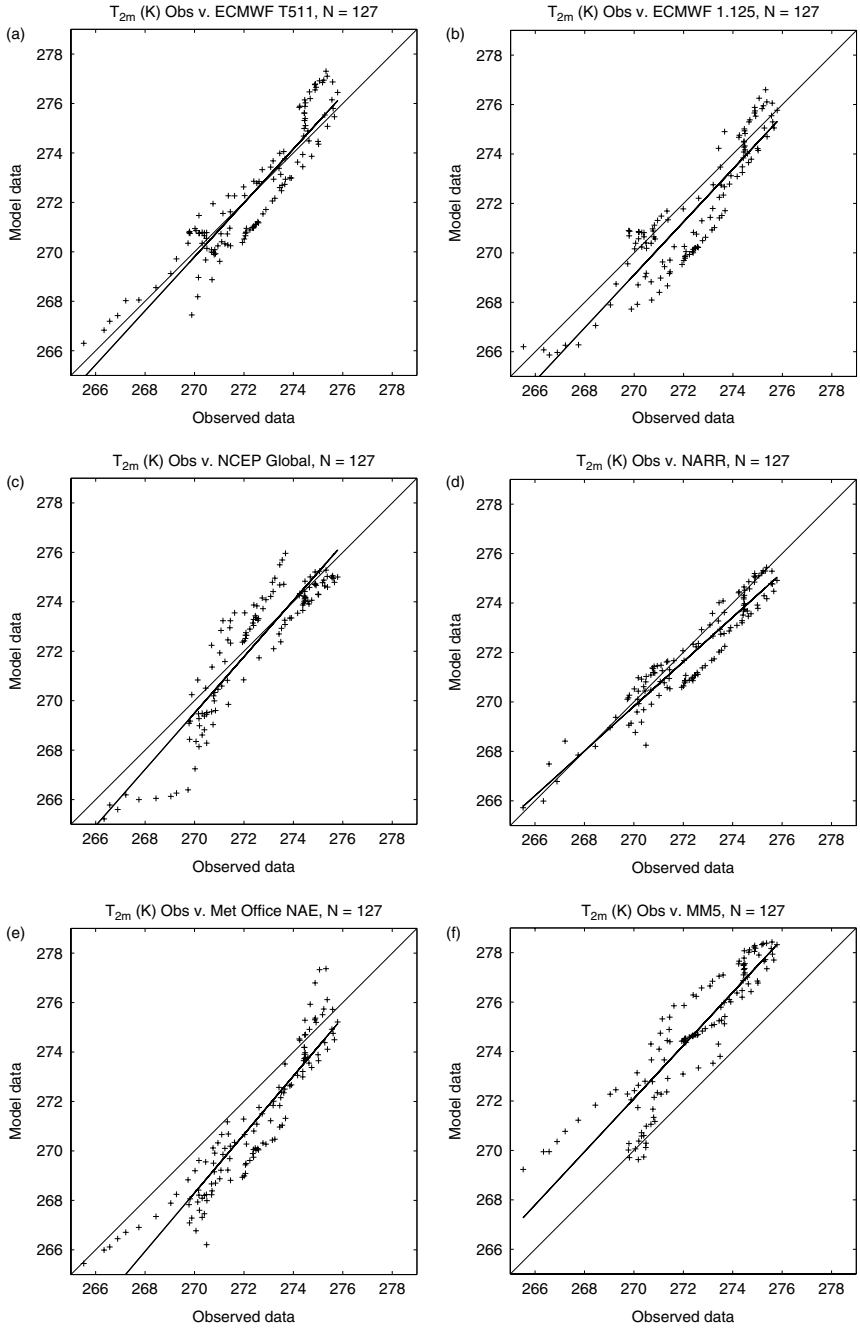


Figure 4. Scatter plots of aircraft observations versus model 2 m air temperature ($T_{2m}$) for models: ECMWF-T511, ECMWF-1deg, NCEP Global Reanalysis, NARR, Met Office NAE and MM5-OSTIA, as indicated on each panel. A linear regression line is shown, where the observed data are assumed to be independent and the model data dependent.

Pagowski and Moore (2001) show the importance of sea-ice concentration within model grid cells in simulating ABL temperatures both over the marginal ice zone and downstream as, with only a 0 or 100% ice concentration, differences of 2–3 K were found up to 300 km downwind of the sea-ice edge. This limitation may explain the overestimate in $T_{2m}$ gradients seen in both of the NCEP models, e.g. through the B271 comparison (Figure 3).

Only SST observations from flights B276, B277 and B278 are available and these cover a more restricted geographical area with only a few observations below ~276 K. Figure 2 shows it is the relatively warm SSTs overlying the deeper waters to the southwest of Iceland (between Greenland's continental shelf and the mid-Atlantic ridge) which are mainly sampled during these three flights. Despite these limitations, Figure 5 makes it clear that the scatter in the SST comparisons is large (e.g. relative to that of $T_{2m}$) particularly for the ECMWF, NCEP and NARR data. The correlations ($r = 0.25$–$0.62$) and the slopes (0.22–0.90) are generally poor, although it is worth noting the regression slopes are heavily influenced by only a few data points at lower temperatures. There are biases of −1.0 K for the NCEP Global, −1.7 K



DOI: 10.1002/qj

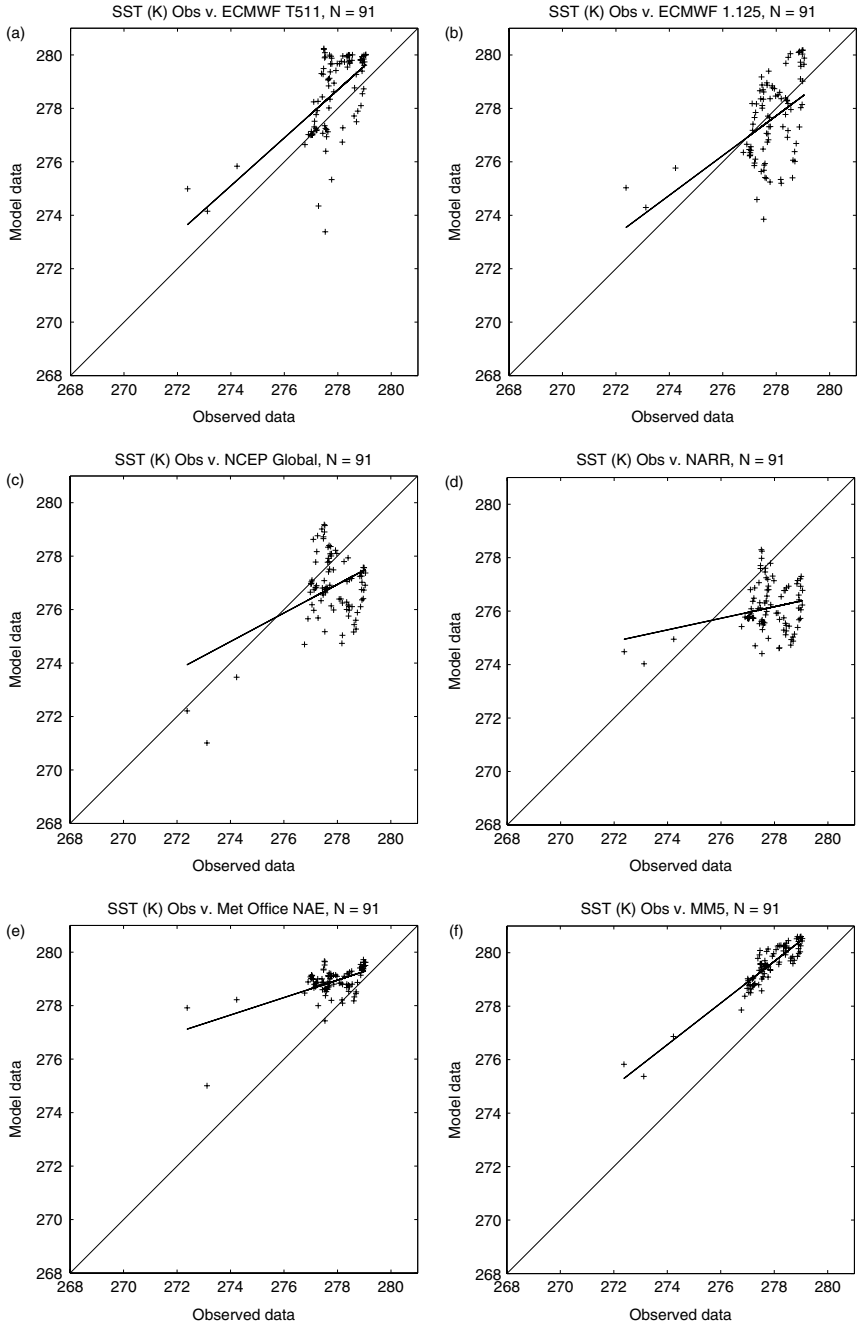


Figure 5. Scatter plots of aircraft observations versus model SST for models: ECMWF-T511, ECMWF-1deg, NCEP Global Reanalysis, NARR, Met Office NAE and MM5-OSTIA, as indicated.

for the NARR, 1.1 K for the NAE, and 1.7 K for the MM5-OSTIA. All of these biases are larger than the specified accuracy of the aircraft's radiometer ($\pm 0.7$ K). It is possible that the frequent wave breaking and white caps observed have affected these radiometer observations, although sensitivity analyses based on flux gradient calculations, suggest this is not the case (Petersen and Renfrew, 2009). Section 2 discusses the source of each model's SST field. It is clear that the coarse resolution of most of these ($0.5^\circ$ for the ECMWF, $1^\circ$ for the NCEP and NARR models) is one source of the large scatter. Indeed the scatter is much reduced for the higher-resolution MM5-OSTIA model, which uses the higher-resolution OSTIA SST data (6 km), and for the NAE although this must be an artefact of interpolation onto the 12 km model grid. Unfortunately these higher-resolution datasets also have substantial biases in this comparison. The OSTIA SST has a bias of 1.7 K and a rms error of 1.6 K compared to the aircraft observations – well above the rms error design specification of 0.8 K versus *in situ* observations (Stark *et al.*, 2007). Donlon *et al.* (2007) describe how such SST analyses employ multiple sources of satellite data along with *in situ* observations from ships and buoys, combined via a multi-scale optimal interpolation scheme. In the Irminger Sea and Denmark Strait area, *in situ* observations are rarely available; the proximity of the



DOI: 10.1002/qj

sea-ice zone leads to enhanced SST gradients and cloud cover often hampers infrared satellite observations. This means the quoted temporal and spatial resolution of e.g. the OSTIA data (i.e. daily and 6 km resolution) is probably not an accurate reflection of the product's actual resolution under such conditions. It seems that under these sort of conditions the SST fields prescribed to NWP models may be in error by up to $\pm 1.6$ K at any one location.

Figures 3(d) and (e) show the specific and relative humidities. For $q_{2\mathrm{m}}$ the general correspondence is good and the spatial gradients are well simulated. The more notable discrepancies can be traced back to poorly corresponding $T_{2\mathrm{m}}$, such as for the ECMWF-1deg data at the start of B276. Tables II and III show that the quantitative comparison for $q_{2\mathrm{m}}$ tends to follow that of $T_{2\mathrm{m}}$, with similar correlation coefficients, slopes and biases (the exception being the NCEP Global comparison). The fact that specific humidity is so strongly governed by temperature necessitates a comparison of $RH_{2\mathrm{m}}$ and here the discrepancies are more transparent. The NCEP Global Reanalysis data are almost uniformly too high (by 5% on average) and $r$ is only 0.09, implying there is very little skill in RH simulation in this model. The same result was found, and discussed in some detail, for similar meteorological conditions over the Labrador Sea by Renfrew *et al.* (2002). The NAE model's $RH_{2\mathrm{m}}$ is also generally too high (by 5% on average), but the correlation and slope are reasonable, suggesting the humidity is modelled with some skill but is biased too moist for these conditions. Note that comparing NAE and observed soundings shows that generally there is a moist bias throughout the ABL. All the cases are unstable boundary layers (mixing driven by a positive buoyancy source at the surface) and so it is expected that any bias generated at the surface would propagate upwards through the depth of the ABL. Conversely the MM5's $RH_{2\mathrm{m}}$ data are almost uniformly too low (by 7% on average), with a poor $r$ and slope, as well as the highest rms errors (up to 12%). This result is consistent for the MM5 model whatever the SST boundary conditions. In terms of RH, the ECMWF and NARR models perform reasonably well. The ECMWF comparisons show the best correlation coefficients and linear regression slope – suggesting a good ABL parametrisation – while the NARR also has excellent correlation coefficients and relatively small rms errors, but a slightly low slope. The reason is that the NARR (and the NAE) fail to capture the lower RH values observed close to the ice edge (at the start of B271 and in B276, Figure 3), where they are both also too cold. (This low RH slope was also found in a comparison of the NARR with the *R/V Knorr* observations of Renfrew *et al.* (2002) that will be documented in a later paper.)

Figure 3(f) shows 10 m wind speed ($U_{10\mathrm{m}}$) from the aircraft, models and the two QuikSCAT time series using the RSS retrieval and the NASA-DIRTH retrieval. In keeping with the MSLP comparison, the general correspondence is reasonable. However an inspection of the spatial gradients shows that many of the models fail to capture the sharp horizontal gradients associated with the mesoscale weather systems observed; for example, the gradients across the easterly tip jet of B268 are not reproduced by the ECMWF or NCEP models (nor by MM5), while the strong shear across the polar low of B271 is not reproduced by the NCEP or MM5 models. It is clear that, at a resolution of 2.5 degrees, the NCEP Global Reanalysis is simply too coarse to represent these sharp gradients in $U_{10\mathrm{m}}$. Figure 6 illustrates its rather poor overall correspondence, with $r = 0.62$, a bias of $-3.1$ m s$^{-1}$ and rms error of 5.0 m s$^{-1}$ (Table III). All the other models perform reasonably, with $r$ ranging from 0.92 at best (ECMWF) to 0.62 at worst (MM5). The NARR has the best slope (0.99), with the other models tending to have a slope that is too low (0.7–0.8), i.e. they are *not* always capturing the highest wind speeds. There are negative biases for *all* the models, ranging from $-0.7$ m s$^{-1}$ at best (NAE), through to $-2.2$ to $-2.5$ m s$^{-1}$ (ECMWF), with $-3.1$ m s$^{-1}$ at worst (NCEP). There is a rough correspondence between model grid size and the magnitude of this bias.

Recall that both ECMWF comparisons are from the same model forecast run at T799 ($\sim$25 km resolution), the difference being simply their archived resolution, i.e. T511 ($\sim$40 km) or T159 (1.125 degrees). One might have expected the T511 product to allow better representation of these mesoscale jets, but this is not the case. The $U_{10\mathrm{m}}$ statistics are almost the same, which suggests that there is something in the ECMWF model set-up which 'smooths out' these mesoscale features. This result is consistent with the comments of Chelton *et al.* (2006), who illustrate through spectral analysis that both the ECMWF and NCEP operational analyses underestimate the variance at scales less than about 1000 km as compared to QuikSCAT winds. Examining Figure 3 again, it seems that the higher-resolution limited-area models (NAE, MM5 and NARR) *are capable* of resolving the sharp gradients and high winds, but only the NAE does so consistently through the 6 days; whereas the global models (ECMWF and NCEP) are simply *not capable* of simulating the higher wind speed jets.

Figure 3 also illustrates the QuikSCAT winds time series and Figure 6 compares $U_{10\mathrm{N}}$ against that observed. Unfortunately there are some missing data, either close to the ice edge or from when the retrieval algorithms failed to determine valid winds for either the morning or evening pass on the comparison day. For the QS-RSS product, 126 from a possible 138 data points are available, however for the QS-NASA-D product only 64 data points are available spread over flights B268, B271 and B278. Data from the other days have been erroneously flagged as over sea ice by the NASA-D algorithm. Visual and flight-level aircraft observations have been carefully checked and no sea ice has been detected. Both the QuikSCAT comparisons are rather poor; inferior to the regional-scale NWP models. The slopes are too high – 1.39 for QS-RSS and 1.06 for QS-NASA-D. For the QS-RSS product, this illustrates a serious overestimation of wind speeds above about 18 m s$^{-1}$. The maximum QS-RSS wind is 33 m s$^{-1}$, which is 8 m s$^{-1}$ higher than the observed maximum. The rms errors are 3.3 and 1.9 m s$^{-1}$, along with 21 and 13 degrees for wind



DOI: 10.1002/qj

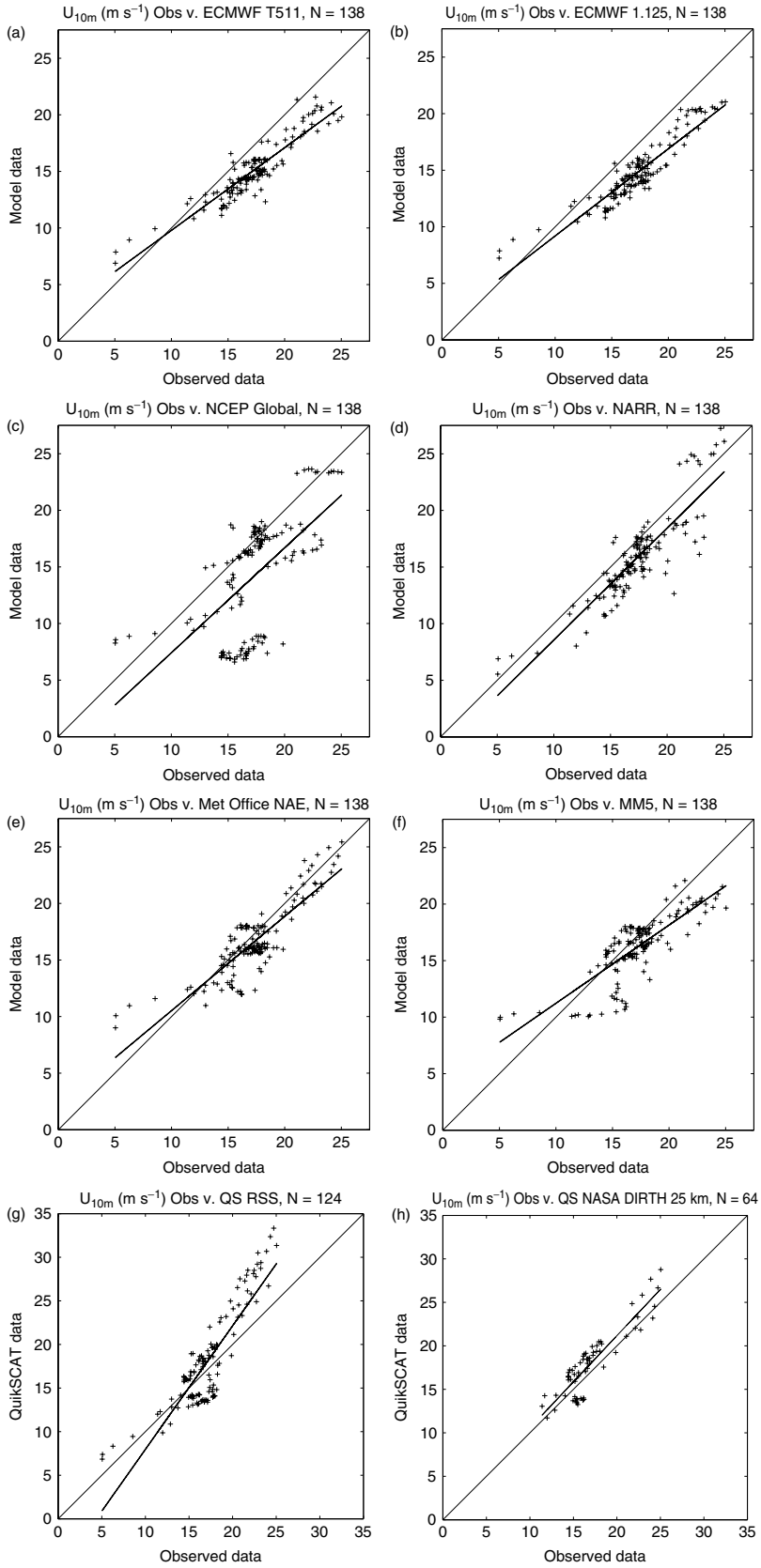


Figure 6. Scatter plots of aircraft observations versus model 10 m wind speed ($U_{10m}$) for models and satellite products: ECMWF-T511, ECMWF-1deg, NCEP Global Reanalyses, NARR, Met Office NAE, MM5-OSTIA, QuikSCAT RSS and QuikSCAT NASA-DIRTH, as indicated. Note that a larger range is used for the QuikSCAT plots.



DOI: 10.1002/qj

direction, for the QS-RSS and QS-NASA-D algorithms respectively (Table III). These correlations and rms errors are much worse than previous buoy comparison studies, where rms errors are typically 1 to 1.5 m $s^{-1}$ and ~20 degrees (Ebuchi *et al.*, 2002; Chelton and Freilich, 2005).

The performance of the QS-RSS retrieval is poor, particularly at high wind speeds. As mentioned in section 2, Ebuchi *et al.* (2002) also found large wind speed residuals above ~18 m $s^{-1}$, particularly for the RSS retrieval, in a comparison against thousands of co-located buoy observations (although their figures suggest only a handful of these observations were for $U_{10N} > 20$ m $s^{-1}$). Fernandez *et al.* (2006) present further evidence and also some potential retrieval-model modifications, although thus far we do not think these have been implemented. Our results, as well as a comparison against buoy observations from the Cape Farewell area by Moore *et al.* (2008), suggest there is still a large high bias with the QS-RSS geophysical retrieval model for winds >18 m $s^{-1}$, and also a slight high bias with the QS-NASA-D model, although both studies rely on a relatively small number of co-located observations. This conclusion is corroborated by a Met Office technical report (Keogh and Offiler, 2006) in which they describe a bias correction that is applied to the QuikSCAT operational winds during the data assimilation process in the operational forecasting suite. This bias correction is to account for QuikSCAT overestimating winds at high wind speeds, compared with fixed-platform observations and the NWP model.

The spatial 'time series' plots in Figure 3 provide a one-dimensional picture of the wind field and the other variables, however it can be useful to examine two-dimensional pictures. Figure 2 shows as background the OSTIA SST field on 5 March 2007 (B276) and is broadly representative of the SST field during the three-week period. For the sake of brevity, further spatial plots are not reproduced here. However spatial plots of, for example MSLP and $U_{10m}$, are reproduced in other GFDex studies: for 21 February (B268), Renfrew *et al.* (2009) show ECMWF and QuikSCAT maps; for 2–6 March (B274, B276 and B277), Petersen *et al.* (2009) show ECMWF maps; and for 9 March (B278), Haine *et al.* (2009) show NCEP Global and MM5 maps.

## 5. Surface turbulent flux comparison

A brief comparison of surface turbulent fluxes is provided here for two reasons: firstly as validation for possible ocean modelling studies and secondly as systematic errors in surface fluxes will feed back into the model simulations, for example, by fluxing too much heat into the model's lowest atmospheric level. Surface turbulent momentum, heat and moisture fluxes must be parametrised in NWP models, generally via a 'bulk flux algorithm' which requires surface-layer temperature, specific humidity and wind (plus SST) as input data and the prescription of non-dimensional exchange coefficients (or equivalently roughness lengths) which, in general, are a function of atmospheric stability and wind speed. Determining appropriate bulk flux algorithms, and in particular the exchange coefficients, requires an empirical approach, i.e. observations of the turbulent fluxes, which means the results rely on measuring turbulent quantities with large flux-sampling errors (e.g. Fairall *et al.*, 2003; Persson *et al.*, 2005; Petersen and Renfrew, 2009). This background explains why the various state-of-the-art NWP models in current use have quite different bulk flux algorithms, especially for high wind speeds.

For the momentum fluxes, most algorithms make use of a modified Charnock formula, with various values of the Charnock constant. For the heat and moisture fluxes, a wider variety of formulae are employed. The ECMWF algorithm uses a 'smooth flow' formulation, with scalar roughness lengths proportional to $1/u_*$ (details in e.g. Brunke *et al.*, 2003). The ECMWF algorithm has been found to be within the bounds of observational uncertainty in similar cold-air outbreak conditions (Renfrew *et al.*, 2002) and to be relatively unproblematic (Brunke *et al.*, 2002, 2003). The NCEP Global Reanalysis sets the scalar roughness lengths using an empirical quadratic expression of the roughness Reynolds number (e.g. Zeng *et al.*, 1998). Previous studies have established that the bulk flux algorithm used in the NCEP Global Reanalysis is inappropriate at high wind speeds, particularly during periods of high heat fluxes (Zeng *et al.*, 1998; Renfrew *et al.*, 2002). The NARR, based on the ETA model, employs a viscous sublayer based on the Liu *et al.* (1979) relationships and dependent on the roughness Reynolds number, underneath a standard Mellor and Yamada (1974) 'level 2' surface layer (details in e.g. Janjić, 1994). At the time of the GFDex field campaign, the Met Office NAE employed constant scalar roughness lengths, however this has now been changed so as to be more appropriate at higher winds (details in e.g. Edwards, 2007). The default MM5 algorithm sets the scalar roughness lengths equal to the roughness length. As discussed by Pagowski and Moore (2001) this is inappropriate over the ocean and leads to large overestimates in the surface turbulent heat fluxes.

Figure 7 and Table IV provide a comparison of observational estimates against the model surface fluxes of momentum ($\tau$), sensible heat (SH) and latent heat (LH). In Figure 7 two bulk flux estimates – described in section 2.1 – the adjusted Smith (1988) algorithm (circles) and the COARE 3.0 algorithm (dots) are plotted. As with the surface-layer variables (Figure 3) the general correspondence is reasonable, but on a close examination there are numerous discrepancies. For the momentum flux, the model differences are primarily a reflection of differences in the surface-layer winds, e.g. the NCEP Reanalysis winds are often too low and do not represent the mesoscale variability well. The other models do as well as could be expected given the $U_{10m}$ correspondence. Table IV provides comparison statistics for the COARE 3.0 fluxes (the Smith bulk flux statistics are similar). The $\tau$ correlation coefficients and slopes are similar to those of $U_{10m}$ and there is a negative bias of about 25% of the mean observed value for the ECMWF and NCEP Global models – explained by a bias in $U_{10m}$.



DOI: 10.1002/qj

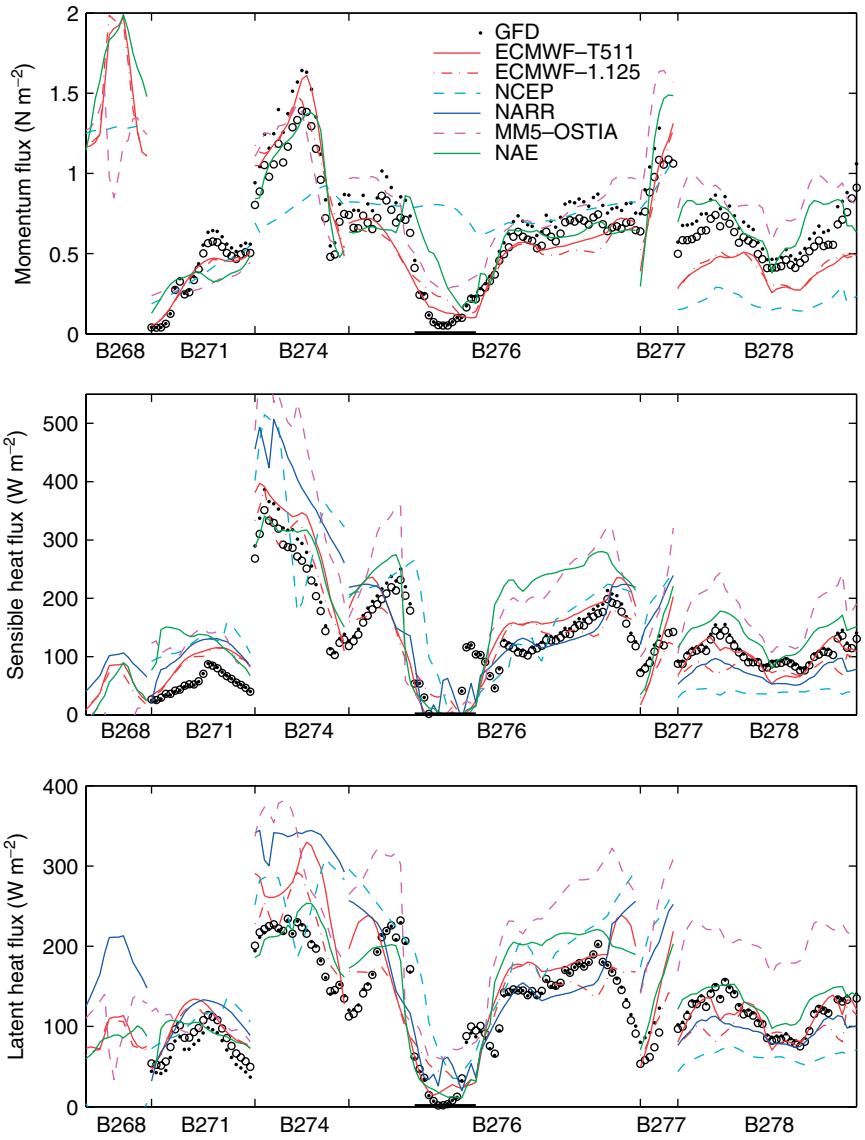


Figure 7. Spatial time series plots of surface turbulent fluxes: (a) surface momentum flux, (b) sensible heat flux and (c) latent heat flux. Observations are based on the COARE 3.0 (dots) and Smith (circles) bulk flux algorithms as described in section 2.1. Output model fluxes are shown in the legend. A bold horizontal line on each panel marks where the data are taken over observed sea ice.

None of the model's algorithms appear to be inconsistent with this comparison.

For the surface heat fluxes there are quite radical differences between the various model time series. The NCEP heat fluxes are overestimated when $U_{10m}$ corresponds well (e.g. during B276), in line with previous studies. The MM5 and NARR fluxes are generally much higher than the observations, especially during high wind speeds. The NAE fluxes are also generally too high. Figure 8 shows scatter plots for the LH fluxes and is also illustrative of the SH flux comparisons. In general the scatter is much larger than for the winds, temperatures, etc. The correlation coefficients for the NCEP Global and NARR comparisons are rather low and the slope for the MM5 model is rather high. There are bias errors of 21 and 31 W m$^{-2}$ for SH and LH in the NARR comparison and 88 and 91 W m$^{-2}$ in the MM5-OSTIA, compared to mean observed fluxes of 140 and 134 W m$^{-2}$ respectively. These positive biases are despite $\Delta T$ being too low (Table II). This suggests that the bulk flux algorithm used in the NARR model may not be the most appropriate for these high heat flux conditions. These results also corroborate the analysis of Pagowski and Moore (2001) regarding the inappropriate default scalar roughness lengths set in the MM5 model. The NAE heat fluxes appear well modelled, in that $r$ and the regression slopes are reasonable, but there is a considerable bias of 50 and 26 W m$^{-2}$ for SH and LH respectively, caused by $T_{2m}$ being too low and the SST being too high, so that $\Delta T$ and $\Delta q$ are both too large. The NAE's mean $\Delta T$ is 7.8 K, compared to an observed value of 5.5 K. The ECMWF heat fluxes are about the most consistent with regression slopes close to 1 and relatively small bias errors. Rather surprisingly the ECMWF-1deg. comparison has smaller bias errors and rms errors than the higher resolution ECMWF-T511 comparison, seemingly due to compensating errors: the



DOI: 10.1002/qj

Table IV. Statistical comparisons for the surface flux data over water. The observations used in this comparison are bulk flux estimates based on the COARE 3.0 algorithm (Fairall *et al.*, 2003). Similar correspondences were generally found when comparing against the Smith (1988) algorithm, as described in section 2.1. The bias and RMS errors are dimensional with units N m$^{-2}$, W m$^{-2}$ and W m$^{-2}$ respectively.

| | $\tau$ | SH | LH | |
|---|---|---|---|---|
| Correlation | 0.89 | 0.88 | 0.78 | ECMWF-1deg |
| coefficient | 0.90 | 0.90 | 0.79 | ECMWF-T511 |
| | 0.56 | 0.75 | 0.60 | NCEP-Reanalyses |
| | – | 0.84 | 0.67 | NCEP-NARR |
| | 0.87 | 0.85 | 0.87 | Met Office NAE |
| | 0.80 | 0.85 | 0.77 | UT-MM5-NFL |
| | 0.81 | 0.97 | 0.89 | UT-MM5-OSTIA |
| Slope | 0.86 | 0.85 | 0.83 | ECMWF-1deg |
| | 0.93 | 1.00 | 1.06 | ECMWF-T511 |
| | 0.50 | 1.10 | 1.04 | NCEP-Reanalyses |
| | – | 1.22 | 1.15 | NCEP-NARR |
| | 0.75 | 0.79 | 0.87 | Met Office NAE |
| | 0.78 | 1.18 | 0.82 | UT-MM5-NFL |
| | 0.80 | 1.51 | 1.36 | UT-MM5-OSTIA |
| Bias error | −0.18 | −4 | 4 | ECMWF-1deg |
| | −0.16 | 16 | 24 | ECMWF-T511 |
| | −0.20 | 15 | 27 | NCEP-Reanalyses |
| | – | 21 | 31 | NCEP-NARR |
| | −0.07 | 50 | 26 | Met Office NAE |
| | 0.02 | 39 | 33 | UT-MM5-NFL |
| | 0.04 | 88 | 91 | UT-MM5-OSTIA |
| RMS error | 0.22 | 37 | 34 | ECMWF-1deg |
| | 0.21 | 40 | 48 | ECMWF-T511 |
| | 0.34 | 79 | 73 | NCEP-Reanalyses |
| | – | 67 | 71 | NCEP-NARR |
| | 0.17 | 65 | 36 | Met Office NAE |
| | 0.18 | 70 | 48 | UT-MM5-NFL |
| | 0.19 | 102 | 98 | UT-MM5-OSTIA |

1.125-degree data $\Delta T$ is too large (as $T_{2m}$ is too low on average; Figure 4) and both ECMWF data have winds that are too low. The smallest rms errors for both SH and LH fluxes are about 25% of the mean values (as is the case for the $\tau$ fluxes), providing an indication of how difficult it is to consistently simulate well these flux quantities.

## 6. A summary of model performance

### 6.1. ECMWF operational analyses

The ECMWF model does not capture the highest wind speeds observed, despite an operational horizontal resolution of T799 and archived data available at truncation T511/N400, equivalent to ∼40 km. This suggests mesoscale atmospheric flow features are being 'smoothed out' in some way, a result that is in line with the spectral analysis and conclusions of Chelton *et al.* (2006). At T511/N400, the model produces good estimates for the surface-layer temperature and humidities, despite a large scatter in the SST. But at lower archived resolution (1.125 deg) a bias of −0.7 K in $T_{2m}$ is introduced. The ECMWF surface turbulent fluxes correspond reasonably well with the observations; the statistical comparison is in line with previous studies (Josey, 2001; Josey *et al.*, 2002; Renfrew *et al.*, 2002). Overall the correspondence is comparable to that of Renfrew *et al.* (2002)[‡] for similar cold-air outbreak conditions over the Labrador Sea.

### 6.2. NCEP global reanalyses

The NCEP global reanalyses are simply too coarse to adequately resolve the mesoscale flow features observed in this dataset. In particular, the correspondence in $U_{10m}$ is very poor. The $T_{2m}$ and SST correspondences are reasonable for the model's resolution, but there is still a positive bias in RH as discussed in Renfrew *et al.* (2002). The flux correlations are poor. The inappropriate scalar roughness-length parametrisation that has been discussed in previous studies (Zeng *et al.*, 1998; Renfrew *et al.*, 2002) is evident on occasion, but the generally poor correspondence in other variables means it is less obvious in this study.

### 6.3. NCEP NARR

The NCEP NARR comparison is generally good. At this horizontal resolution (32 km) the highest wind speeds can be simulated, although overall there is a negative bias (−1.5 m s$^{-1}$) in $U_{10m}$. The correspondence in surface-layer temperature and humidities is relatively good, compared to the other models, suggesting that the ABL parametrisations are adequate, with the caveat that the slope in the $RH_{2m}$ comparison is too low (the model is too moist at lower RHs) and $\Delta T$ is rather low. The rms errors for $T_{2m}$ and $q_{2m}$ are similar to the comparison against buoy data of Moore *et al.* (2008). Given the above, the correspondence of the surface heat fluxes is disappointing, suggesting the bulk flux algorithm is *not* optimal. The NARR bulk flux algorithm employs a viscous sublayer, which becomes negligible for high wind speeds, whereupon flux transports are simply set by the ABL parametrisation, which is based on Mellor and Yamada (1974)'s level 2 scheme (Janjić, 1994). Evidence from this comparison suggests this set-up can, but does not always, lead to fluxes much larger than either the observations or the other models.

### 6.4. Met Office NAE

The NAE operational model does well at capturing the observed high winds associated with the barrier flows

[‡]Note that Renfrew *et al.* (2002) present bias, slope, random and total errors in their Tables II and IV. They state on p. 389 that the 'total error' is equal to the rms error, but unfortunately this is incorrect; they do not tabulate the rms errors.



DOI: 10.1002/qj

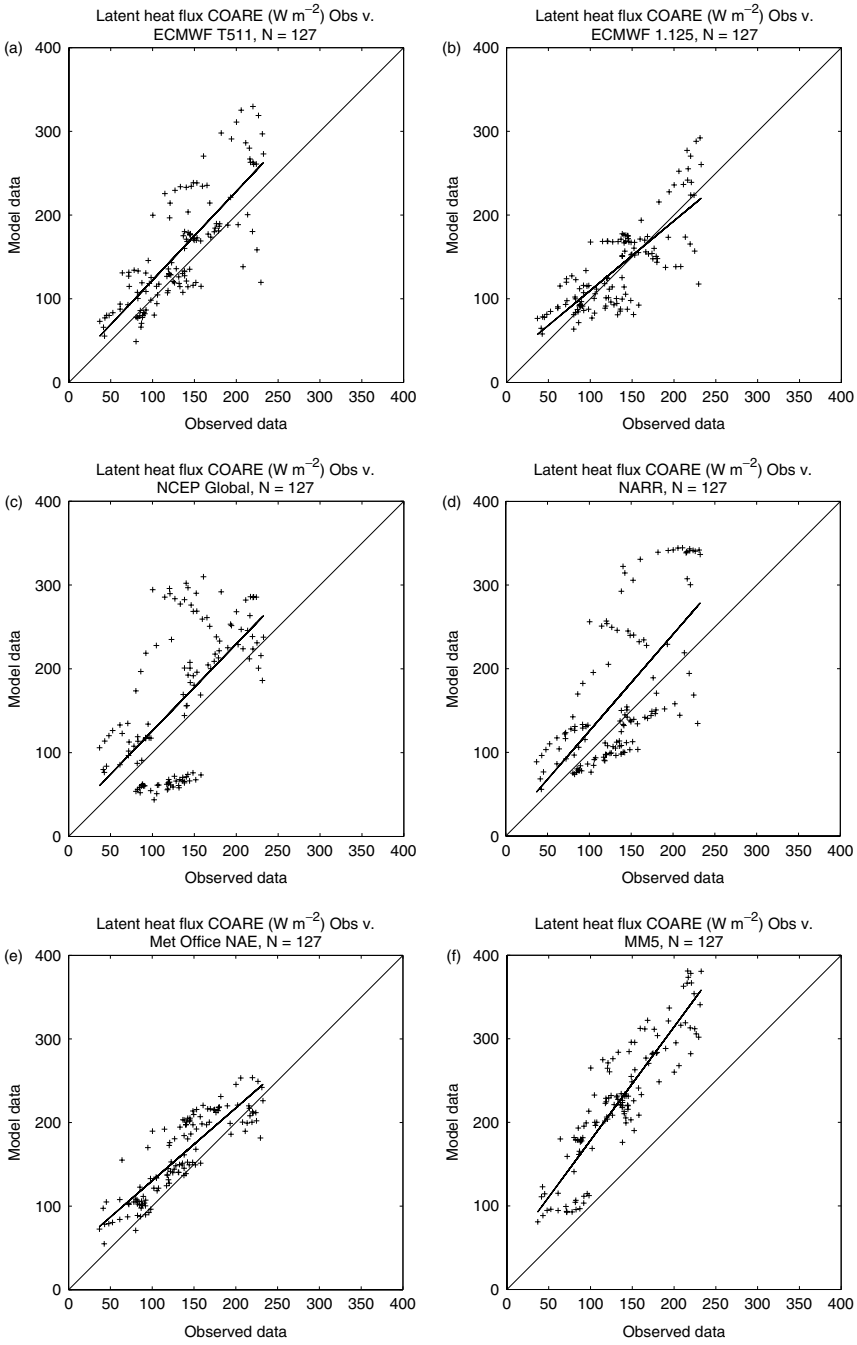


Figure 8. Scatter plots of aircraft observations versus model surface latent heat fluxes (LH) for models ECMWF-T511, ECMWF-1deg, NCEP Global Reanalyses, NARR, Met Office NAE and MM5-OSTIA, as indicated.

and jets; there is only a small bias ($-0.7$ m s$^{-1}$) in $U_{10\mathrm{m}}$. However, adjacent and over areas of sea ice there is a pronounced cold bias in $T_{2\mathrm{m}}$ ($-1.3$ K on average), which is *not* explained by a $+1.1$ K bias in the SST. This problem may be explained by the model's sea ice thickness being set at a uniform 2 m, which is perhaps too thick on average for this region. The NAE $RH_{2\mathrm{m}}$ has a similar low-slope problem to that of the NARR model. The surface turbulent fluxes are generally well-modelled, but the relatively large biases in $\Delta T$ and $\Delta q$ result in relatively large biases in the sensible and latent heat fluxes.

### 6.5. MM5 hindcasts

The MM5 simulations are able to capture the high wind speed jets, but sometimes do not. A bias of $+1.7$ K in the OSTIA SST used in the second hindcast leads to a positive bias in $T_{2\mathrm{m}}$ (2.3 K) and in $q_{2\mathrm{m}}$ (0.24 g kg$^{-1}$). The $RH_{2\mathrm{m}}$ corresponds poorly to the observations, suggesting little skill for this field during these conditions. The MM5 default bulk flux algorithm results in a poor regression slope and large bias, which leads to large overestimates in the sensible and latent heat fluxes. In line with Pagowski and Moore (2001), this suggests the algorithm is inappropriate for high heat flux conditions.



DOI: 10.1002/qj

### 6.6. QuikSCAT winds

The QuikSCAT wind comparisons are very poor and poor for the RSS and NASA-D retrieval algorithms, respectively. The RSS algorithm appears to be more problematic at high wind speeds: the regression slope is 1.39 and the rms error is 3.3 m $s^{-1}$, which is above the instrument's design specifications. The NASA-D algorithm has better error statistics: the regression slope is 1.06 and the rms error is 1.9 m $s^{-1}$, but there is less data available due to rather conservative sea-ice flagging. Combining these results with the recent buoy comparison of Moore *et al.* (2008) provides *in situ* evidence for problems in the retrieval algorithms at high wind speeds, particularly in the RSS algorithm, which should be urgently addressed.

## 7. Conclusions

To simulate the high winds associated with extratropical mesoscale weather systems – such as tip jets, barrier flows and polar lows – a model resolution of order 20 km is necessary, but is not sufficient, as appropriate ABL and surface flux parametrisations are also crucial. In regions of the subpolar and polar seas, relatively close to the sea-ice edge, the current generation of NWP models still have problems in accurately simulating ABL temperature and humidity, perhaps being unable to transit from stable to unstable conditions quickly enough. An accurate prescription of the SST is essential, but at the time of GFDex these were generally prescribed at a relatively coarse resolution compared to the atmospheric model grid. The operational use of a new generation of high-resolution SST products (e.g. Donlon *et al.*, 2007) will no doubt improve the quality of SST fields, but there are still likely to be relatively high discrepancies in cloudy areas of high SST gradients. The use of surface turbulent fluxes from NWP models is *not* recommended without an investigation of the surface flux algorithm used and validation against observations.

## Acknowledgements

We would like to thank all at the Facility for Airborne Atmospheric Measurement for enabling the GFDex field campaign. The UEA authors were funded by the Natural Environmental Research Council (grant NE/C003365/1) and the University of Toronto authors by the Canadian Foundation for Climate and Atmospheric Sciences (GR-641). The EUFAR (European Fleet for Airborne Research) and EUCOS (European Coordinated Observing System) schemes contributed additional resources to the field campaign. This study was largely carried out at the University of Toronto during a period of sabbatical leave for the lead author, for which he would like to thank the Dept. of Physics and NERC for financial contributions. We would like to thank F. Mesinger for discussion on the NARR dataset. We would also like to thank Sid Clough (Met Office) for early discussions on this work and for comments on the manuscript by Adrian Lock, Clive Jones and Adrian Hines at the Met Office and Thomas Jung at the ECMWF.

## References

Bacon S. 1998. Decadal variability in the outflow from the Nordic seas to the deep Atlantic Ocean. *Nature* **394**: 871–874.

Bacon S, Reverdin G, Rigor IG, Smith HM. 2002. A fresh water jet on the east Greenland shelf. *J. Geophys. Res.* **107**: 3068, DOI: 10.1029/2001JC000935.

Berbery EH, Luo Y, Mitchell KE, Betts AK. 2003. Eta model estimated land surface processes and the hydrologic cycle of the Mississippi basin. *J. Geophys. Res.* **108**(D22): 8852, DOI: 10.1029/2002JD003192, 2003.

Brümmer B. 1997. Boundary layer mass, water and heat budgets in wintertime cold-air outbreaks from the arctic sea ice. *Mon. Weather Rev.* **125**: 1824–1837.

Bromwich DH, Cassano J, Klein T, Heinemann G, Hines K, Steffen K, Box JE. 2001. Mesoscale modeling of katabatic winds over Greenland with the Polar MM5. *Mon. Weather Rev.* **129**: 2290–2309.

Brown AR, Beare RJ, Edwards JM, Lock AP, Keogh SJ, Milton F, Walters DN. 2008. Upgrades to the boundary-layer scheme in the Met Office numerical weather prediction model. *Boundary-layer Meteorol.* **128**: 117–132.

Brunke MA, Zeng X, Anderson S. 2002. Uncertainties in sea surface turbulent flux algorithms and data sets *J. Geophys. Res.* **107**(C10): 3141, DOI: 10.1029/2001JC000992.

Brunke MA, Fairall CW, Zeng X, Eymard L, Curry JA. 2003. Which bulk aerodynamic algorithms are least problematic in computing ocean surface turbulent fluxes? *J. Climate* **16**: 619–635.

Cassano JJ, Box JE, Bromwich DH, Li L, Steffen K. 2001. Evaluation of polar MM5 simulations of Greenland's atmospheric circulation. *J. Geophys. Res.* **106**: 33867–33890.

Chen F, Janjić Z, Mitchell KE. 1997. Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP mesoscale Eta model. *Boundary-layer Meteorol.* **85**: 391–421.

Chelton DB, Freilich MH. 2005. Scatterometer-based assessment of 10-m wind analyses from the Operational ECMWF and NCEP numerical weather prediction models. *Mon. Weather Rev.* **133**: 409–429.

Chelton DB, Freilich MH, Sienkiewicz JM, Von Ahn JM. 2006. On the use of QuikSCAT scatterometer measurements of surface winds for marine weather prediction. *Mon. Weather Rev.* **134**: 2055–2071.

Condron A, Bigg GR, Renfrew IA. 2006. Polar mesoscale cyclones in the northeast Atlantic: Comparing climatologies from ERA-40 and satellite imagery. *Mon. Weather Rev.* **134**: 1518–1533.

DeCosmo J, Katsaros KB, Smith SD, Anderson RJ, Oost WA, Bumke K, Chadwick H. 1996. Air-sea exchange of water vapor and sensible heat: The Humidity Exchange Over the Sea (HEXOS) results. *J. Geophys. Res.* **101**(C5): 12001–12016.

Dickson RR, Brown J. 1994. The production of North Atlantic Deep Water. *J. Geophy. Res.* **99**(C6): 12319–12341.

Dickson RR, Meincke J, Rhines P. 2008a. *Arctic-Subarctic Ocean Fluxes: Defining the role of the Northern Seas in Climate.* Springer: Berlin.

Dickson RR, Dye S, Jonsson S, Kohl A, Macrander A, Marnela M, Meinke J, Olsen S, Rudels B, Valdimarsson H, Voet G. 2008b. The overflow flux west of Iceland: Variability, origins and forcing. Pp 443–474 in *Arctic-Subarctic Ocean Fluxes: Defining the role of the Northern Seas in Climate.* Dickson RR, Meincke J, Rhines P. (eds.) Springer: Berlin.

Donlon C, Robinson I, Casey KS, Vazquez-Cuervo J, Armstrong E, Arino O, Gentemann C, May D, LeBorgne P, Piollé J, Barton I, Beggs H, Poulter DJS, Merchant CJ, Bingham A, Heinz S, Harris A, Wick G, Emery B, Minnett P, Evans R, Llewellyn-Jones D, Mutlow C, Reynolds RW, Kawamura H, Rayner N. 2007. The Global Ocean Data Assimilation Experiment High-resolution Sea Surface Temperature Pilot Project. *Bull. Amer. Meteorol. Soc.* **88**: 1197–1213.

Dudhia J. 1996. 'A multi-layer soil temperature model for MM5'. Preprints for Sixth PSU/NCAR Mesoscale Model Users' Workshop, 22–24 July 1996, Boulder, Colorado, Pp 49–50.

Ebuchi N, Graber HC, Caruso MJ. 2002. Evaluation of wind vectors observed by QuikSCAT/SeaWinds using ocean buoy data. *J. Atmos. Oceanic Technol.* **19**: 2049–2062.

Edwards JM. 2007. Oceanic latent heat fluxes: Consistency with the atmospheric hydrological and energy cycles and general

circulation modeling. *J. Geophys. Res.* **112**: D06115, DOI: 10.1029/2006JD007324.
Ek MB, Mitchell KE, Lin Y, Rogers E, Grunmann P, Koren V, Gayno G, Tarpley JD. 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.* **108**: 8851, DOI: 10.1029/2002JD003296.
Fairall CW, Bradley EF, Rogers DP, Edson JB, Young GS. 1996. Bulk parameterization of air-sea fluxes for Tropical Ocean–Global Atmosphere Coupled-Ocean Atmosphere Response Experiment. *J. Geophys. Res.* **101**(C2): 3747–3764.
Fairall CW, Bradley EF, Hare JE, Grachev AA, Edson JB. 2003. Bulk parameterization of air-sea fluxes: Updates and verification for the COARE algorithm. *J. Climate* **16**: 571–591.
Fernandez DE, Carswell JR, Frasier S, Chang PS, Black PG, Marks FD. 2006. Dual-polarized C- and Ku-band ocean backscatter response to hurricane-force winds. *J. Geophys. Res.* **111**: C08013, DOI: 10.1029/2005JC003048.
Haine TWN, Zhang S, Moore GWK, Renfrew IA. 2009. On the impact of high-resolution, high-frequency meteorological forcing on Denmark Strait ocean circulation. *Q. J. R. Meteorol. Soc.* **135**: 2067–2085.
Hong S-Y, Pan H-L. 1996. Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Weather Rev.* **124**: 2322–2339.
Janjić ZI. 1994. The step-mountain Eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Weather Rev.* **122**: 927–945.
Josey S. 2001. A comparison of ECMWF, NCEP/NCAR and SOC surface heat fluxes with moored buoy measurements in the subduction region of the north-east Atlantic. *J. Climate* **14**: 1780–1789.
Josey SA, Kent EC, Taylor PK. 2002. Wind stress forcing of the ocean in the SOC climatology: Comparisons with the NCEP–NCAR, ECMWF, UWM/COADS, and Hellerman and Rosenstein datasets. *J. Phys. Oceanogr.* **32**: 1993–2019.
JPL. 2006. QuikSCAT Science Data Product: User's Manual, Version 3.0, D-18053 – Rev A. Overview and Geophysical Data Products. Jet Propulsion Laboratory: Pasadena.
Jung T, Leutbecher M. 2007. Performance of the ECMWF forecasting system in the Arctic during winter. *Q. J. R. Meteorol. Soc.* **133**: 1327–1340.
Kain JS, Fritsch JM. 1993. Convective parameterization for mesoscale models: The Kain–Fritsch scheme. Cumulus parameterization. *Meteorol. Monogr.* **46**: 165–170. Amer. Meteorol. Soc: Boston.
Kalnay E, Kanamitsu M, Kirtler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetma A, Reynolds R, Jenne R, Joseph D. 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteorol. Soc.* **77**: 437–471.
Keogh SJ, Offiler D. 2006. 'Seawinds revisited'. Forecasting Research Technical Report No. 472. Met Office: Exeter, UK.
Kistler R, Kalnay E, Collins W, Saha S, White G, Woollen J, Chelliah M, Ebisuzaki W, Kanamitsu M, Kousky V, van den Dool H, Jenne R, Fiorino M. 2001. The NCEP–NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteorol. Soc.* **82**: 247–267.
LabSea Group. 1998. The Labrador Sea Deep Convection Experiment. *Bull. Amer. Meteorol. Soc.* **79**: 2033–2058.
Large W, Yeager SG. 2004. 'Diurnal to decadal global forcing for ocean and sea-ice models: The datasets and flux climatologies'. Technical Note NCAR/TN-460+STR. NCAR: Boulder.
Large W, Morzel J, Crawford GB. 1995. Accounting for surface wave distortion of the marine wind profile in low-level ocean storms wind measurements. *J. Phys. Oceanogr.* **25**: 2959–2971.
Liu WT, Katsaros KB, Businger JB. 1979. Bulk parameterization of air-sea exchanges of heat and water vapor including the molecular constraints at the interface. *J. Atmos. Sci.* **36**: 1722–1735.
Lock AP, Brown AR, Bush MR, Martin GM, Smith RNB. 2000. A new boundary-layer mixing scheme. Part I: Scheme description and single-column model tests. *Mon. Weather Rev.* **128**: 3187–3199.
Marshall J, Schott F. 1999. Open ocean convection: Observations, theory and models, *Rev. Geophys.* **37**: 1–64.
Martin GM, Bush MR, Brown AR, Lock AP, Smith RNB. 2000. A new boundary-layer mixing scheme. Part II: Tests in climate and mesoscale models. *Mon. Weather Rev.* **128**: 3200–3217.
Mellor GL, Yamada T. 1974. A hierarchy of turbulence closure models for planetary boundary layers. *J. Atmos. Sci.* **31**: 1791–1806.
Mesinger F, DiMego G, Kalnay E, Mitchell K, Shafran PC, Ebisuzaki W, Jović D, Woollen J, Rogers E, Berbery EH, Ek MB, Fan Y, Grumbine R, Higgins W, Li H, Lin Y, Manikin G, Parrish D, Shi W. 2006. North American regional reanalysis. *Bull. Amer. Meteorol. Soc.* **87**: 343–360.
Moore GWK, Renfrew IA. 2002. An assessment of the surface turbulent heat fluxes from the NCEP reanalysis over western boundary currents. *J. Climate* **15**: 2020–2037.
Moore GWK, Pickart RS, Renfrew IA. 2008. Buoy observations from the windiest location in the world ocean, Cape Farewell, Greenland. *Geophys. Res. Lett.* **35**: L18802, DOI: 10.1029/2008GL034845.
Østerhus S, Sherwin T, Quadfasel D, Hansen B. 2008. The overflow transport east of Iceland. Pp 427–442 in *Arctic-Subarctic Ocean Fluxes: Defining the role of the Northern Seas in Climate.* Dickson RR, Meincke J, Rhines P. (eds.) Springer: Berlin.
Pagowski M, Moore GWK. 2001. A numerical study of an extreme cold-air outbreak over the Labrador Sea: Sea-ice air–sea interaction and the development of polar lows. *Mon. Weather Rev.* **129**: 47–72.
Persson POG, Hare JE, Fairall CW, Otto WD. 2005. Air–sea interaction processes in warm and cold sectors of extratropical cyclonic storms observed during FASTEX. *Q. J. R. Meteorol. Soc.* **131**: 877–912.
Petersen GN, Renfrew IA. 2009. Aircraft-based observations of air-sea fluxes over Denmark Strait and the Irminger Sea during high wind speed conditions. *Q. J. R. Meteorol. Soc.* **135**: 2030–2045.
Petersen GN, Renfrew IA, Moore GWK. 2009. An overview of barrier winds off southeastern Greenland during GFDex. *Q. J. R. Meteorol. Soc.* **135**: 1950–1967.
Pickart RS, Spall MA, Ribergaard MH, Moore GWK, Milliff RF. 2003. Deep convection in the Irminger Sea forced by the Greenland tip jet. *Nature* **424**: 152–156.
Pickart RS, Torres DJ, Fratantoni PS. 2005. The East Greenland spill jet. *J. Phys. Oceanogr.* **35**: 1037–1053.
Renfrew IA, Moore GWK. 1999. An extreme cold air outbreak over the Labrador Sea: Roll vortices and air–sea interaction *Mon. Weather Rev.* **127**: 2379–2394.
Renfrew IA, Moore GWK, Guest PS, Bumke K. 2002. A comparison of surface-layer and surface turbulent-flux observations over the Labrador Sea with ECMWF analyses and NCEP reanalyses. *J. Phys. Oceanogr.* **32**: 383–400.
Renfrew IA, Moore GWK, Kristjánsson JE, Ólafsson H, Gray SL, Petersen GN, Bovis K, Brown PRA, Føre I, Haine T, Hay C, Irvine EA, Lawrence A, Ohigashi T, Outten S, Pickart RS, Shapiro M, Sproson D, Swinbank R, Woolley A, Zhang S. 2008. The Greenland Flow Distortion experiment. *Bull. Amer. Meteorol. Soc.* **89**: 1307–1324.
Renfrew IA, Outten SD, Moore GWK. 2009. An easterly tip jet off Cape Farewell, Greenland. I: Aircraft observations. *Q. J. R. Meteorol. Soc.* **135**: 1919–1933.
Smith SD. 1988. Coefficients for sea surface wind stress, heat flux and wind profiles as a function of wind speed and temperature. *J. Geophys. Res.* **93**: 15467–15472.
Smith SR, Legler DM, Verzone KV. 2001. Quantifying uncertainties in NCEP reanalyses using high-quality research vessel observations. *J. Climate* **14**: 4062–4072.
Stark JD, Donlon CJ, Martin MJ, McCulloch ME. 2007. 'OSTIA: An operational, high-resolution, real-time, global sea surface temperature analysis system'. In: Proceedings of *OCEANS 2007 - Europe*. DOI: 10.1109/OCEANSE.2007.4302251.
Sun B, Yu L, Weller RA. 2003. Comparisons of surface meteorology and turbulent heat fluxes over the Atlantic: NWP model analyses versus moored buoy observations. *J. Climate* **16**: 679–695.
Tao WK, Simpson J. 1993. The Goddard cumulus ensemble model. Part I: Model description. *Terr. Atmos. Ocean. Sci.* **4**: 35–72.
Tao WK, Simpson J, McCumber M. 1989. An ice-water saturation adjustment. *Mon. Weather Rev.* **17**: 231–235.
Uhlhorn EW, Black PG. 2003. Verification of remotely sensed sea surface winds in hurricanes. *J. Atmos. Oceanic Technol.* **20**: 99–116.
Uppala SM, Kållberg PW, Simmons AJ, Andrae U, Da Costa Bechtold V, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, Van de Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosazac M, Fisher M, Fuentes M, Hagemann S, Hólm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP,

Mahfouf J-F, Morcrette J-J, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KA, Untch A, Vasiljevic D, Viterbo P, Woollen J. 2005. The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.* **131**: 2961–3012.

Wentz F, Smith DK, Mears CA, Gentemann CL. 2001. 'Advanced algorithms for QuikScat and SeaWinds/AMSR'. Proceedings of IGARSS 2001, Sydney, Australia, IEEE, 1079–1081.

Yu L, Weller RA, Sun B. 2004. Improving latent and sensible heat flux estimates for the Atlantic Ocean (1988–1999) by a synthesis approach. *J. Climate* **17**: 373–393.

Yueh SH, Stiles B, Tsai W-Y, Hu H, Liu WT. 2001. 'QuikSCAT geophysical model function for hurricane wind and rain'. Proceedings of IGARSS 2001, Sydney, Australia, IEEE, 1089–1091.

Zeng X, Zhao M, Dickinson RE. 1998. Intercomparison of bulk aerodynamic algorithms for the computation of sea surface fluxes using TOGA COARE and TAO data. *J. Climate* **11**: 2628–2644.




DOI: 10.1002/qj