Evaluating forecasts of the evolution of the cloudy boundary layer using diurnal composites of radar and lidar observations

Andrew I. Barrett,[1] Robin J. Hogan,[1] and Ewan J. O'Connor[1]

[1]   Observations of boundary-layer cloud have been made using radar and lidar at Chilbolton, Hampshire, UK. These have been compared with output from 7 different global and regional models. Fifty-five cloudy days have been composited to reveal the mean diurnal variation of cloud top and base heights, cloud thickness and liquid water path of the clouds. To enable like-for-like comparison between model and observations, the observations have been averaged on to the grid of each model. The composites show a distinct diurnal cycle in observed cloud; the cloud height exhibits a sinusoidal variation throughout the day with a maximum at around 1600 and a minimum at around 0700 UTC. This diurnal cycle is captured by six of the seven models analysed, although the models generally under-predict both cloud top and cloud base heights throughout the day. The two worst performing models in terms of cloud boundaries also have biases of around a factor of two in liquid water path; these were the only two models that did not include an explicit formulation for cloud-top entrainment. **Citation:** Barrett, A. I., R. J. Hogan, and E. J. O'Connor (2009), Evaluating forecasts of the evolution of the cloudy boundary layer using diurnal composites of radar and lidar observations, *Geophys. Res. Lett.*, *36*, L17811, doi:10.1029/2009GL038919.

1.   Introduction

[2]   Low-level clouds and their location, height, thickness, persistence, and even existence depend on a number of subtle factors, which include surface fluxes, drizzle formation, entrainment at the boundary layer top and large scale subsidence. As such these clouds are a particular challenge to forecast accurately in numerical weather prediction models and even more so in lower resolution climate models [*Xu and Randall*, 1996]. The physical processes within the boundary layer that lead to the formation of these clouds are often dominated by turbulence and are therefore impossible to model explicitly; instead the important processes are parameterized [e.g., *Tiedtke*, 1993; *Lock*, 1998].

[3]   Previous studies investigating boundary-layer structure and stratocumulus cloud have shown that most models simulate boundary layer depths shallower than observed and liquid water paths that are too low. *Stevens et al.* [2007] showed that boundary layer depth is underestimated by the European Centre for Medium Range Weather Forecasts (ECMWF) model and reanalyses (ERA15 and ERA40), while *Hannay et al.* [2009] reported that the ECMWF model and three climate GCMs underpredict boundary-

layer depth and stratocumulus liquid water path in the South-East Pacific, in agreement with *Bretherton et al.* [2004] over the East Pacific. *Sengupta et al.* [2004] found that the ECMWF model boundary-layer cloud is slightly too high over the Southern Great Plains, USA, but *Palm et al.* [2005] used satellite-based lidar observations to show globally that the ECMWF model placed the maritime boundary layer top too low by 200–500 m.

[4]   To understand why there are model errors we need to look at the response of the models to changed forcings, such as over the diurnal cycle. Compositing is a way of revealing these errors and here boundary-layer cloud observations, made by radar and lidar, are composited over the diurnal cycle. Seven models from European centres are compared with observed cloud derived from radar, lidar and microwave radiometer observations made at Chilbolton, Hampshire, UK. The success of each model is judged based on hourly comparisons with observations of the presence, vertical location, thickness and water content of their boundary layer clouds.

2.   Method

[5]   We use cloud observations made continuously by the vertical pointing radar and lidar at Chilbolton, Hampshire, UK. Using the two wavelengths in tandem allows clouds to be located at a temporal resolution of 30 s, with a vertical resolution of 60 m. To allow a direct comparison to be made with the model data, the observations are transformed on to the grid of each model. The models are listed in Table 1 together with their horizontal and vertical resolution and information about the model formulation and parameterizations; see *Illingworth et al.* [2007] for further details. All forecast lead times are less than 36 hours, although RACMO and SMHI-RCA are both run using only ECMWF boundary conditions and without data assimilation. At each hour, data from the model grid boxes over the observation site have been recorded (except UKMO-Global, for which data are available only every 3 hours).

[6]   The averaging process involves calculating the fraction of observed pixels that are cloudy in each model grid box. This is done for a time inversely proportional to the wind speed at that height, such that the product of the averaging time and the wind speed gives a distance the same as the width of the grid-box at that height. Figure 1 illustrates the results of this process on a single day of data. Averaging time is always constrained to be between 10 and 60 minutes, to ensure that there is no development of the cloud within the period. Further details about this process, performed as part of the Cloudnet project (www.cloudnet.org), are given by *Illingworth et al.* [2007].

---

[1]Department of Meteorology, University of Reading, Reading, UK.

Table 1. Summary of Each Model Used, Performance Statistics for Cloud Occurrence and Observed Biases in Models[a]

	UKMO-Meso	UKMO-Global	ECMWF	Météo-France	RACMO	SMHI-RCA-3.0	SMHI-RCA-3.5
Horizontal Resolution (km)	12	60	39	23.4	18	44	44
Number of Levels below 3 km	12	12	16	15	16	16	16
Grid-box depth at 1 km (m)	277	277	235	227	235	230	230
Mixing Scheme[b]	L/NL/EE	L/NL/EE	L/NL/EE	L	L/NL/EE	TKE	TKE/EE
Prognostic Cloud Variables[c]	q_t	q_t	q_l, A	none	q_l, A	none	none
A B	473 158	160 75	620 110	408 319	612 93	595 138	623 121
C D	121 424	32 162	231 344	79 499	259 295	348 224	387 174
Log odds ratio, ln θ	2.35	2.38	2.13	2.09	2.01	1.02	0.84
Cloud Top Error (m)	-226 ± 352	-274 ± 369	-126 ± 378	-567 ± 415	$+63 \pm 432$	$+39 \pm 594$	-70 ± 750
Cloud Base Error (m)	-183 ± 323	-212 ± 401	-84 ± 354	-326 ± 366	-115 ± 357	-225 ± 435	-412 ± 463
Cloud Thickness Error (m)	-43 ± 418	-62 ± 439	-42 ± 416	-241 ± 443	$+178 \pm 495$	$+263 \pm 624$	$+341 \pm 779$
100 m Temperature Bias[d] (K)	-0.39	-0.65	-0.71	$+0.99$	$+0.54$	n/a	$+0.22$
100 m Dewpoint Bias[d] (K)	-1.00	-0.10	-0.56	$+1.17$	-0.46	n/a	$+0.95$
Sensible Heat Bias[d] (W m^{-2})	n/a	-8.13	-0.21	$+12.90$	$+1.32$	n/a	-5.89
Latent Heat Bias[d] (W m^{-2})	n/a	$+1.55$	-1.86	$+3.69$	-11.53	n/a	$+8.13$
Subsidence at 3 km (cm s^{-1})	0.46	0.41	0.78	0.69	0.75	n/a	n/a

[a]The models are Met Office Mesoscale (UKMO-Meso), Met Office Global (UKMO-Global), European Centre for Medium Range Weather Forecasts (ECMWF), Météo-France ARPEGE, Royal Netherlands Meteorological Institute's Regional Atmospheric Climate Model (RACMO) and two versions of Swedish Meteorological and Hydrological Institute Rossby Centre Regional Atmosphere Model (SMHI-RCA). A, B, C, D refer to the values used in equation (1), with A being the number of data used for cloud height comparison. Positive errors indicate cloud is higher in the model. The variation is given by one standard deviation. See section 3 for details about mixing schemes.

[b]L, Local mixing; NL, Non-local mixing; EE, Explicit Entrainment; TKE, Prognostic Turbulent Kinetic Energy.

[c]Prognostic cloud variables are q_t – total water mixing ratio, q_l – liquid water mixing ratio and A – cloud fraction.

[d]Bias is difference from mean of available model data; n/a, not available.

[7] From the available data at the time of the analysis (most days in the period April 2003–September 2004) suitable days have been selected by eye from the radar and lidar observations and observed cloud fraction. Days have been deemed suitable when: (1) Cloud is observed in the lowest 3 km in consecutive hours, at least once during the day, (2) No significant cloud is present above 3 km, (3) No local frontal influences are present. These conditions are chosen to isolate errors in the boundary layer schemes, without the potential influence of large-scale forcing. In practice these conditions restrict the days chosen to times when anticyclonic conditions are present. Of the 55 days chosen, 36 occur in the months August–November, with the remaining days being distributed within the remaining months.

[8] Data from seven European weather centre models have been used for comparison with observations. Suitable data from the model output have been selected to allow comparison with the observations from the same time. These are: (1) Cloud top height – defined as the height of the top of the highest grid-box below 3 km with cloud fraction greater than 0.1, (2) Cloud base height – the bottom of the lowest grid-box with cloud fraction greater than 0.1, (3) Cloud thickness – the difference in heights between the cloud top and cloud base. In all cases, only a single layer of cloud was present.

3. Results and Discussion

[9] We first assess the skill of model predictions of boundary-layer cloud occurrence. A model is deemed to have correctly predicted cloud presence if, for the hour in question, both model and observations have cloud fraction greater than the threshold value of 0.1 in any grid-box below 3 km. Equally, a correct forecast of a cloud-free boundary layer requires that all grid-boxes below 3 km have cloud fraction less than 0.1 when no cloud is observed. The skill of the prediction of cloud presence has then been quantified using the log odds ratio [*Stephenson*, 2000], defined as:

$$\ln \theta = \ln \frac{AD}{BC} \qquad (1)$$

where A is the number of correct cloud forecasts, D is the number of correct forecasts of clear sky. B and C are the number of false predictions of clear sky and cloud respectively. The log odds ratio was chosen for its properties, namely that a random forecast gives a score of
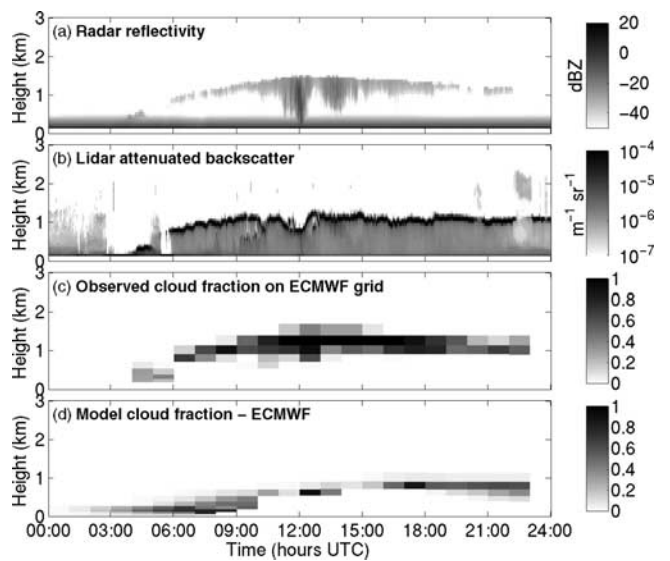
Figure 1. Observations and ECMWF model output for 18 September 2003 at Chilbolton: (a) radar return, (b) lidar return, (c) observed cloud fraction averaged to ECMWF model grid, and (d) ECMWF model forecast cloud fraction.
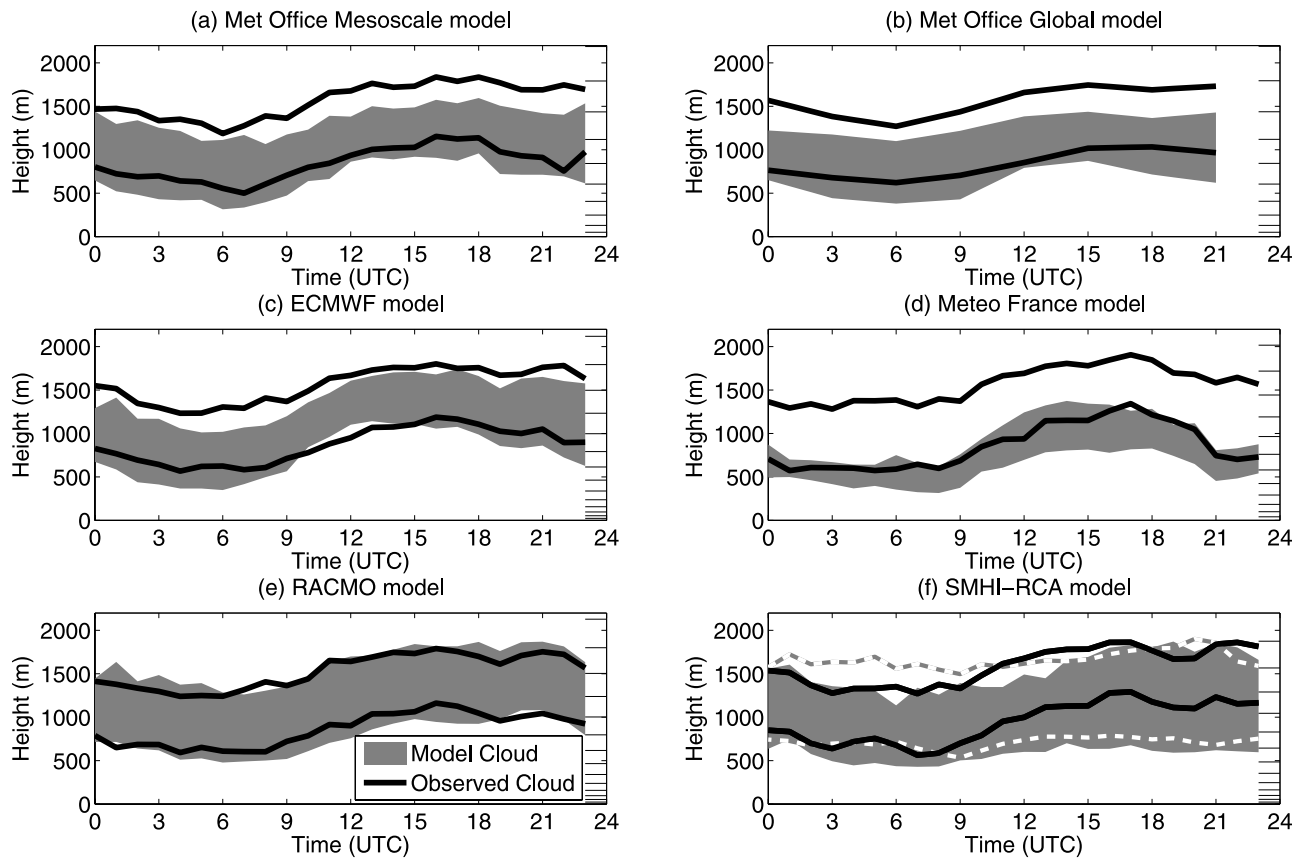
Figure 2. (a−e) A composite of the observed cloud boundaries (lines) and model cloud (shaded) throughout the day for the 7 models analysed. The vertical resolution of each model is shown by horizontal lines between 23−24 hours. (f) Two model versions, v3.0 is shown with dashed lines whilst v3.5 is shown with the solid fill. See text for model differences.

0 and it is not easily hedged. Detailed analysis of the merits of different skill scores for cloud verification is given by *Hogan et al.* [2009]. Model performance for the prediction of cloud presence can be seen in Table 1.

[10] The log odds ratio shows that the two Met Office models (UKMO-Meso, 2.35; UKMO-Global, 2.38) perform best, and are essentially unbiased. SMHI-RCA-3.0, SMHI-RCA-3.5, ECMWF and RACMO have the largest number of correct cloud forecasts, and therefore more data for the cloud height comparison, but all show a bias towards too much cloud and have lower scores (1.02, 0.84, 2.13 and 2.01 respectively). Conversely, Météo-France (2.09) shows a large bias towards too little cloud.

[11] We now evaluate the location of cloud boundaries. The values in Table 1 show that all models simulate cloud base lower than is observed on average and only RACMO and SMHI-RCA-3.0 predict cloud top height too high. RACMO is the least biased model, predicting cloud top height too high by 63 m and cloud base height too low by 115 m on average. In general, with the exception of the Météo-France and SMHI-RCA models, the model biases are smaller than the model grid box depth at cloud level (see Table 1 for values at 1 km).

[12] More information is revealed by producing diurnal composites of cloud heights. Figure 2 shows the composited cloud boundaries for each of the models. The observed cloud (black outline) shows a clear diurnal cycle in height, with cloud lowest around 7 UTC and highest at 16 UTC.

The average cloud thickness stays near constant throughout the day and once averaged to the model grid is about 500 m thick. The model cloud (gray fill) exhibits a similar diurnal cycle for all models except SMHI-RCA. SMHI-RCA-3.0 has a poor representation of the diurnal cycle while SMHI-RCA-3.5 now shows a diurnal cycle in cloud top height but not cloud base. The model cloud remains too low throughout the day for many models.

[13] The RACMO model provides a remarkably good fit to the observations, following the mean cloud top and base heights very well. RACMO performs better than ECMWF despite having near identical vertical grids and model physics, although RACMO does have a better horizontal resolution and exhibits more variation in the error, shown by a larger standard deviation.

[14] The Liquid Water Path (LWP) of the modeled and observed cloud has also been composited. A diurnal cycle of LWP, similar to that of cloud height, is presented in Figure 3. The range of observations derived from dual-wavelength microwave radiometer is shown by the gray shaded region. A similar diurnal cycle is present in some models, although UKMO-Meso and UKMO-Global do not show this. The model biases in LWP are similar to those seen in cloud height and thickness, with SMHI-RCA-3.0 substantially overpredicting cloud thickness and LWP, Météo-France underpredicting both and the other models generally performing better.
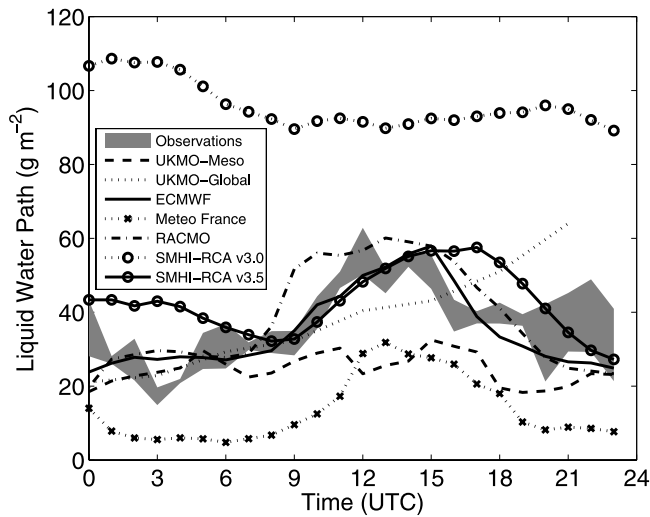
Figure 3. A composite of the observed (shaded) and model (lines) Liquid Water Path throughout the day. The observed LWPs for each model fall within the shaded region.

[15] To help understand what may be the cause of the model errors we must identify aspects of the model that are important in forecasting stratocumulus. Two such aspects are the surface scheme and the treatment of turbulent mixing. We note from Table 1 that we have three types of turbulent mixing scheme. Both Met Office models, the ECMWF and RACMO models employ a non-local mixing scheme in unstable conditions, and when the boundary layer is topped by stratocumulus, each uses an explicit formulation for the entrainment rate as a function of cloud-top radiative divergence following *Lock* [1998]. By contrast, the Météo-France model uses a local mixing scheme in unstable conditions following *Louis* [1979], with no explicit treatment of cloud-top entrainment. SMHI-RCA uses prognostic turbulent kinetic energy (TKE) which includes cloud-top entrainment in version 3.5, but not in version 3.0 (dashed lines in Figure 2f). Also shown in Table 1 are the differences from the multi-model mean for surface sensible and latent heat fluxes and boundary-layer temperature and dewpoint. We will now discuss possible explanations for the differences between model performance in terms of the different model formulations.

[16] The Météo-France model exhibits a strong diurnal cycle in cloud height, but the cloud is far too low at all times during the day. Yet this model has the largest sensible heat flux and has the warmest boundary-layer of all models. The fact that modeled cloud is too low is likely a result of the local mixing scheme used in this model in unstable conditions (following *Louis* [1979]), with no explicit treatment of cloud-top entrainment. It was found by *Beljaars and Betts* [1992] that a lack of explicit entrainment resulted in cloudy boundary layers that tended to be too shallow, cold and moist, due to too little growth into the (potentially) warmer and drier free troposphere above. *Holtslag and Boville* [1993] found that using local instead of non-local mixing has similar consequences. This could explain the shallower cloudy boundary layer in the Météo-France model and its lower cloud base. The warm rather than cold

bias in this model is presumably due to the higher surface sensible heat flux. The bias towards too little cloud in the model is also related to the mixing scheme, as the model boundary layer fails to grow to sufficient depth to allow air at the top to reach saturation, hence observed boundary-layer cloud may be above the model boundary-layer top, whilst the model boundary-layer is cloud-free.

[17] Further evidence for the importance of including explicit entrainment is found when considering the performance of the SMHI-RCA model, for which we had versions with and without this parameterization. The turbulence scheme in SMHI-RCA-3.0 is based on prognostic TKE combined with a diagnostic length scale [*Cuxart et al.*, 2000]. Although the scheme includes diffusion of moist conserved variables, the length scales do not include cloud condensation effects and so entrainment at cloud top tends to be underestimated (U. Willén, personal communication). The lack of diurnal cycle in boundary-layer cloud height could also be related to issues with the cloud microphysics leading to overestimated water content (also reported by *van Lipzig et al.* [2006] and *Illingworth et al.* [2007]).

[18] The diurnal cycle exhibited in the SMHI-RCA-3.5 model is an improvement on version 3.0 (shown by dashed lines in Figure 2f). Version 3.5 includes moist turbulence including entrainment and improved cloud microphysics. The addition of explicit entrainment appears to have much improved the diurnal cycle in cloud top height. Although SMHI-RCA-3.5 is one of the warmer models, it has the smallest diurnal range of temperature. This, caused by a weak sensible heat flux, combined with a large latent heat flux keeps the boundary layer air near saturation throughout the day, hence a low cloud base with little diurnal change in height. The same bias in fluxes probably results in the worsening of the cloud base height and cloud thickness and a greater bias towards too much cloud. The liquid water path now also shows a diurnal cycle and the bias significantly decreases. Other changes in the new version are the convection scheme and land surface physiographic database. The affect of vertical resolution on boundary-layer cloud simulation was also investigated using the SMHI-RCA model. Runs with 40 vertical levels, 8 below 3 km, of both model versions showed a slight worsening in performance in cloud presence and location; however, the difference was not dramatic.

4. Summary

[19] Seven models have been compared with boundary layer cloud observations made using radar and lidar. Diurnal compositing of the observed and modeled cloud has allowed us to quantify how well the models simulate the evolution of the cloudy boundary layer. The averaging of observations to the model grids has been key in allowing suitable comparisons between observations and model data but the averaging process has not significantly altered the data.

[20] The analysis has shown a large variation between the performance of the models. UKMO-Meso and UKMO-Global demonstrate the best forecast skill, based on the log odds ratio, when predicting the presence of boundary-layer cloud, but they place the cloud 200–300 m too low on average (around one model grid level). ECMWF and

RACMO show less bias in cloud height, but do exhibit a bias towards predicting cloud when none is observed. The Météo-France model has clear sky too frequently and when it does simulate cloud it is typically too low, too thin and contains too little water. SMHI-RCA produces cloud too frequently, has cloud base too low, but showed a large improvement in LWP and cloud top height between versions 3.0 and 3.5.

[21] The four models which include non-local mixing in unstable conditions and explicit entrainment at cloud-top demonstrate the best simulation of the diurnal cycle of stratocumulus. The local mixing scheme in Météo-France results in a boundary layer which is too shallow and too moist. Even a large sensible heat flux, making this the warmest model, does not allow the boundary layer to grow to sufficient depth to match observations. Changes to the SMHI-RCA model, which included an explicit entrainment formulation, resulted in a better simulation of the diurnal cycle in cloud top height.

References

Beljaars, A. C. M., and A. K. Betts (1992), Validation of the boundary layer representation in the ECMWF model, in *Validation of Models Over Europe*, vol. 2, pp. 159–195, Eur. Cent. for Medium-Range Weather Forecasts, Reading, UK.

Bretherton, C. S., T. Uttal, C. W. Fairall, S. Yuter, R. Weller, D. Baumgardner, K. Comstock, R. Wood, and G. Raga (2004), The EPIC 2001 stratocumulus study, *Bull. Am. Meteorol. Soc.*, *85*, 967–977.

Cuxart, J., P. Bougeault, and J.-L. Redelsperger (2000), A turbulence scheme allowing for mesoscale and large-eddy simulations, *Q. J. R. Meteorol. Soc.*, *126*, 1–30.

Hannay, C., D. L. Williamson, J. J. Hack, J. T. Kiehl, J. G. Olson, S. A. Klein, C. S. Bretherton, and M. Köhler (2009), Evaluation of forecasted southeast Pacific stratocumulus in the NCAR, GFDL, and ECMWF models, *J. Clim.*, *22*, 2871–2889.

Hogan, R. J., E. J. O'Connor, and A. J. Illingworth (2009), Verification of cloud-fraction forecasts, *Q. J. R. Meteorol. Soc.*, in press.

Holtslag, A. A. M., and B. A. Boville (1993), Local versus nonlocal boundary-layer diffusion in a global climate model, *J. Clim.*, *6*, 1825–1842.

Illingworth, A. J., et al. (2007), Cloudnet—Continuous evaluation of cloud profiles in seven operational models using ground-based observations, *Bull. Am. Meteorol. Soc.*, *88*, 885–898.

Lock, A. P. (1998), The parametrization of entrainment in cloudy boundary layers, *Q. J. R. Meteorol. Soc.*, *124*, 2729–2753.

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

Palm, S. P., A. Benedetti, and J. Spinhirne (2005), Validation of ECMWF global forecast model parameters using GLAS atmospheric channel measurements, *Geophys. Res. Lett.*, *32*, L22S09, doi:10.1029/2005GL023535.

Sengupta, M., E. E. Clothiaux, and T. P. Ackerman (2004), Climatology of warm boundary layer clouds at the ARM SGP site and their comparison to models, *J. Clim.*, *17*, 4760–4782.

Stephenson, D. B. (2000), Use of the "odds ratio" for diagnosing forecast skill, *Weather Forecast.*, *15*, 221–232.

Stevens, B., A. Beljaars, S. Bordoni, C. Holloway, M. Köhler, S. Krueger, V. Savic-Jovcic, and Y. Zhang (2007), On the structure of the lower troposphere in the summertime stratocumulus regime of the northeast Pacific, *Mon. Weather Rev.*, *135*, 985–1005.

Tiedtke, M. (1993), Representation of clouds in large-scale models, *Mon. Weather Rev.*, *121*, 3040–3061.

van Lipzig, N. P. M., et al. (2006), Model predicted low-level cloud parameters: 1. Comparison with observations from the BALTEX Bridge Campaigns, *Atmos. Res.*, *82*, 55–82.

Xu, K. M., and D. A. Randall (1996), Explicit simulation of cumulus ensembles with the GATE Phase III data: Comparison with observations, *J. Atmos. Sci.*, *53*, 3710–3736.

───────────────
A. I. Barrett, R. J. Hogan, and E. J. O'Connor, Department of Meteorology, University of Reading, Earley Gate, PO Box 243, Reading RG6 6BB, UK. (a.i.barrett@reading.ac.uk)