Global verification of cloud fraction in models using CloudSat and CALIPSO data

Robin J Hogan[1,2], Andrew I Barrett[2,3], Thorwald H M Stein[2], Hélène Garçon[1,2*], Julien Delanoë[4], Richard M Forbes[1] and Alejandro Bodas-Salcedo[5]

In this paper, the CloudSat radar and CALIPSO lidar are used to evaluate the global distribution of vertically resolved cloud fraction in two forecast models. A new method is used to estimate the location of the base of liquid clouds from the lidar, exploiting the fact that multiple scattering leads to a monotonic relationship between the integrated backscatter and the cloud optical depth. Modeled mean cloud fraction is found to be about right in the upper troposphere, but at most latitudes mid-level cloud fraction is half that observed. We then use forecast verification measures to perform the first global evaluation of the extent to which the full profile of clouds are forecast at the right time and place. We find that winter-hemisphere upper-tropospheric clouds are most accurately forecast, while tropical and sub-tropical boundary-layer clouds are predicted with very little skill.

1. Introduction

Clouds have long been implicated as one of the largest sources of uncertainty in climate prediction [*Wielicki et al.*, 1995; *Boucher et al.*, 2013], but spaceborne cloud radar and lidar [*Stephens et al.*, 2002] are revolutionizing our ability to test the representation of clouds in models, and hence potentially to narrow this uncertainty. Global cloud distributions in models have been tested both using radar-lidar retrievals of quantities such as ice water content [e.g. *Delanoë et al.*, 2011], and by 'forward modeling' the observations from the model variables and comparing the distribution of modeled and actual observations [e.g. *Bodas-Salcedo et al.*, 2008].

However, such studies typically only evaluate the mean or the distribution of cloud properties, but not whether clouds are forecast at the right time. The problem of evaluating the association of forecast and observed cloud fields we refer to as *cloud verification,* and has previously been carried out using ground-based radar and lidar data [e.g. *Mace et al.*, 1998; *Jakob et al.*, 2004; *Hogan et al.*, 2009]. Cloud verification is important because, assuming that the large-scale dynamics is approximately correct in a short-range forecast, we can test whether the model produces clouds correctly in response to particular dynamical structures. Global cloud verification was demonstrated by *Wilkinson et al.* [2008], but their use only of spaceborne lidar, which is subject to significant attenuation, led to ambiguity in the common situation that clouds at one level (in either the model or reality) obscured clouds below.

In this paper we demonstrate how CloudSat and CALIPSO data may be used to evaluate the cloud fraction predicted by global forecast models, where cloud fraction is simply the fraction of each model gridbox that is filled with cloud. This is done in terms of both the monthly mean cloud fraction, and the extent to which clouds were forecast at the right time and location. We use forecasts by the Met Office and the European Centre for Medium Range Weather Forecasts (ECMWF), and while the duration of the comparison is limited to 1–2 months, this is sufficient to obtain robust signals and therefore to demonstrate the potential of the technique.

In section 2, we outline the method to estimate cloud fraction from CloudSat and CALIPSO, including a new method to estimate the thickness of liquid clouds. In section 3 we describe the model data and the verification measures used. In section 4, the cloud fraction climatology is evaluated, and in section 5, several verification measures are used to evaluate the extent to which the models forecast cloud at the right time.

2. Observations

2.1. The DARDAR product

We use as a starting point the 'DARDAR' CloudSat-CALIPSO cloud classification of *Delanoë et al.* [2010], in which the radar and lidar data are composited on to the same grid and each pixel is assigned a category such as clear sky, ice cloud, liquid cloud, rain, aerosol, mixtures of ice and supercooled liquid droplets, and mixtures of rain and liquid cloud droplets. The complementary sensitivities of spaceborne radar and lidar give us confidence that virtually all ice clouds are detected [e.g. *Stein et al.*, 2011]. We will not repeat the description provided by *Delanoë et al.* [2010], but highlight some properties of the observations that should be borne in mind:

- No attempt is made to distinguish 'ice cloud' from 'snow'. *Hogan et al.* [2001] argued that the two form a continuum and there is no sensible way to separate them in either ground-based or spaceborne observations. Hence all ice is considered as cloud in the observations. This is the same assumption as made in the Met Office model but it is different from the ECMWF model at the time of the comparison, which separated the ice cloud from snow with snow not interacting with radiation.

- As in models, rain in the observations is not treated as cloud. This means that when a melting layer is present, the observed cloud fraction can jump from 1 in the ice above to 0 in the rain below.

- Liquid clouds are identified from high-intensity lidar returns, but in optically thick clouds the lidar signal is strongly attenuated and affected by multiple scattering, presenting a challenge to locating cloud base. The radar is rarely useful for liquid cloud base because in warm clouds, the radar return tends to be dominated by drizzle which we do not classify as cloud. We overcome this problem by exploiting the lidar multiple scattering: in section 2.2 it is explained how this enables cloud optical depth to be estimated, and then in section 2.3 how cloud thickness can be estimated enabling us to modify the DARDAR cloud product.

[1]European Centre for Medium Range Weather Forecasts, Reading, United Kingdom.

[2]Department of Meteorology, University of Reading, Reading, United Kingdom.

[3]NERC National Centre for Earth Observation at University of Reading, Reading, United Kingdom.

[4]LATMOS, Université de Versailles, St-Quentin-en-Yvelines, France.

[5]Met Office Hadley Centre, Exeter, United Kingdom.

[*]The fourth author was affiliated to the University of Reading during the time she contributed to this work, but is now affiliated to ECMWF.

Copyright 2016 by the American Geophysical Union.
0094-8276/16/$5.00

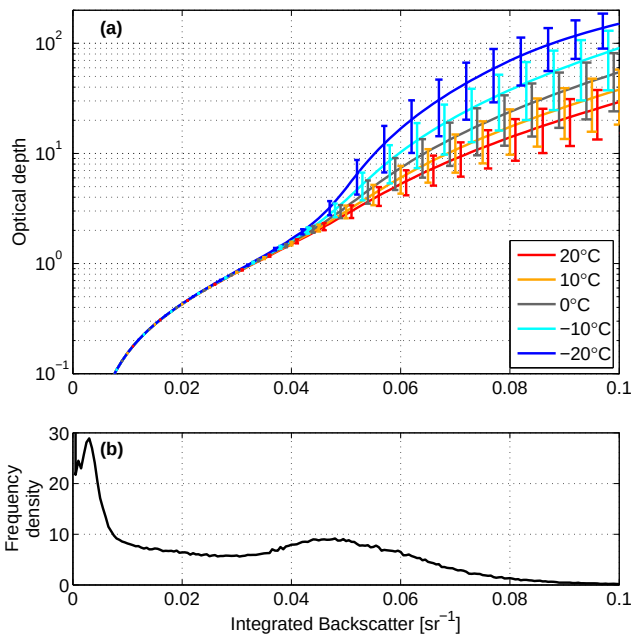

Figure 1. (a) Relationship between cloud optical depth and integrated backscatter for simulated lidar returns using *multiscatter*. Error bars show 1 standard deviation of the spread related to varying effective radius and adiabaticity assumptions. (b) Frequency density plot of integrated backscatter from five orbits of CALIPSO data.

The extent to which these issues affect the comparison of mean cloud fraction will be discussed in section 4, but the best verification measures used in section 5 are largely insensitive to whether there is a bias in the model or the observations.

2.2. Estimating liquid cloud optical depth from lidar integrated backscatter

As discussed above, the base of liquid clouds cannot be estimated accurately using either spaceborne radar or lidar. However, the presence of lidar multiple scattering from satellite platforms [e.g. *Wilkinson et al.*, 2008] means that the vertically-integrated attenuated backscatter through a liquid cloud is monotonically related to optical depth, which is in turn monotonically related to physical depth. Our method to estimate optical depth from integrated backscatter is now described.

To build a look-up table to capture this relationship, we simulated CALIPSO backscatter profiles for liquid clouds with different properties using *multiscatter* [*Hogan*, 2008; *Hogan and Battaglia*, 2008], a fast multiple-scattering algorithm, assuming a satellite altitude of 705 km and a receiver footprint on the cloud of diameter 91 m. The cloud properties were systematically varied through realistic ranges so that together with a best estimate of cloud fraction we also produce uncertainty ranges due to the assumptions made about cloud properties. The properties altered were cloud depth, cloud top temperature, adiabaticity and cloud-droplet effective radius.

Cloud adiabaticity is the ratio of the actual liquid water path (LWP) to that of an adiabatic cloud with the same depth and cloud base. Stratiform boundary-layer clouds are frequently close to adiabatic [*Wood*, 2012; *Albrecht et al.*, 1990], but can be sub-adiabatic if entrainment or drizzle are locally important [*Wood*, 2012]. In cumulus clouds, adiabaticity at cloud base is near 1, but falls to 0.2 near cloud top [*Warner*, 1955; *Albrecht et al.*, 1985; *Nicholls and Turton*, 1986]. *Zhou et al.* [2006] found that adiabaticity was typically around 0.79. Based on these values we choose a range of 0.8±0.2 for adiabaticity.

The cloud droplet effective radius strongly affects the scattering properties of the liquid clouds. *Miles et al.* [2000] reported typical values of 9.6±2.4 μm (1 standard deviation) for marine clouds and 5.4±2.1 μm for continental clouds. We choose a range of 10±5 μm for effective radius to cover these reported clouds. For all simulated clouds, the cloud top temperature was varied from −20 to 20°C in 10°C intervals and cloud optical depth was varied from 0 to 1000.

The geometric-optics extinction profile for these clouds was calculated and the total attenuated backscatter from each of the clouds was simulated using the multiscatter algorithm. Fig. 1a depicts the resulting lookup table for estimating optical depth (and its 1-sigma error) from integrated backscatter and temperature. It can be seen that there is good sensitivity to optical depth throughout the observed range of CALIPSO integrated backscatter (Fig. 1b). Note that for ground-based radar with a much smaller receiver footprint on the cloud, the relationship saturates at a particular value of integrated backscatter [e.g. *O'Connor et al.*, 2004].

2.3. Modifying the DARDAR cloud mask

Using the lookup table and the observed integrated backscatter in each profile (integrated from 500 m above liquid cloud top to 1,500 m below), we estimate the optical depth, τ, of each cloud layer, and its uncertainty. This is converted to an estimate of physical depth, H, in meters, using the relation from *Chiu et al.* [2014]:

$$H = \begin{cases} (513 \pm 51)\tau^{0.16\pm0.03} & \text{for drizzling clouds,} \\ (308 \pm 15)\tau^{0.25\pm0.01} & \text{for non-drizzling clouds.} \end{cases} \quad (1)$$

A cloud is determined to be drizzling if the peak radar reflectivity in the cloud layer exceeds −15 dBZ, consistent with the *Chiu et al.* [2014] formulation. The Chiu formulation was derived from clouds over land, so we also estimate cloud depth separately using a relation based on clouds over the ocean [*Minnis et al.*, 1992; *Chiu et al.*, 2014],

$$H = 58\tau^{0.56}, \quad (2)$$

which gives smaller geometric cloud depths and does not discriminate between drizzling and non-drizzling clouds. The uncertainty range for cloud depth is determined using (2) as the lower bound, and the largest values from (1) together with the upper uncertainty from the adiabaticity and effective radius ranges as the upper bound.

The DARDAR cloud mask is modified by adding or removing liquid cloud pixels at the base of the highest liquid layer identified in each profile to match the estimated H. The same procedure is applied to any layers lower in the profile, but because their lidar return may be significantly attenuated, if the retrieved H of such layers is smaller than the mean H of all unattenuated clouds at the same height in the two-month dataset used in this paper, then that mean cloud depth value is used instead. The presence of ice cloud is not modified.

3. Models and verification measures

3.1. ECMWF and Met Office global forecast models

We use the same Met Office and ECMWF model data as *Delanoë et al.* [2011]. The data were extracted under CloudSat and CALIPSO, and the pixels from the combined radar-lidar grid were associated with the nearest model gridbox. The model data available are the last three weeks of July 2006 (both models) and the entire month of February 2007 (ECMWF only). The Met Office data are from the forecast version of the Unified Model, cycle C40, with a horizontal resolution of around 40 km and 50 vertical levels. Model data are extracted in a similar way to *Bodas-Salcedo et al.* [2008]: 3-hourly snapshots are constructed from the 6-hourly analyses and intervening 3-hour forecasts, and the closest snapshot in

time is used for the comparison such that it is always within 1.5 h of the CloudSat overpass time. Cloud fraction in this model is diagnosed from the liquid and ice mixing ratios using a scheme based on Smith (1990).

The ECMWF data are from model cycle 32r3 with a horizontal resolution of around 40 km and 91 vertical levels. Three-hourly snapshots are extracted from a series of 12–33 h forecasts along the CloudSat track, such that the snapshot is again always within 1.5 h of the overpass time. In comparing the skill of the two models it should be borne in mind that the oldest ECMWF forecast used is 33 h old, while only 3 h for the Met Office. Cloud fraction in the ECMWF model is a prognostic variable as described by *Tiedtke* [1993], although the scheme has evolved considerably since his original paper.

3.2. Verification measures

The verification of cloud fraction forecasts is carried out in a similar way to *Hogan et al.* [2009]: each forecast cloud fraction is deemed to be a 'hit' if it and the corresponding observation both exceed the threshold of 0.1, a 'false alarm' if it is less than or equal to the threshold while the observed fraction exceeds the threshold, and likewise for 'misses' and 'correct negatives'. Then for each 4° latitude range and 1-km height range, the numbers of hits, false alarms, misses and correct negatives are summed and denoted a, b, c and d, respectively. The total number of forecasts in each latitude-height bin is $n = a + b + c + d$.

Four verification measures are computed from these variables. The first is the Gilbert Skill Score $\mathrm{GSS} = (a - a_r) \,/\, (a + b + c - a_r)$ [*Gilbert*, 1884], where $a_r = (a + b)(a + c)/n$ is the expected number of 'hits' by a random forecasting system. This measure is widely used in routine forecast verification under the name *Equitable Threat Score*. However, it is only strictly equitable (i.e. random forecasts yield a fixed expected score of zero) in the limit of large n, which led *Hogan et al.* [2010] to describe it as 'asymptotically equitable'. The second measure is the Log of Odds Ratio $\mathrm{LOR} = \ln(ad/bc)$ [*Stephenson*, 2000], and is also asymptotically equitable. *Stephenson et al.* [2008] showed that both GSS and LOR are degenerate for rare events, that is they tend to a meaningless limit (zero for GSS, infinity for LOR) as the base rate $s = (a+c)/n$ tends to zero. The first measure proposed that is non-degenerate for rare events while retaining asymptotic equitability is the Symmetric Extreme Dependency Score $\mathrm{SEDS} = \ln(a_r/a)/\ln(a/n)$ [*Hogan et al.*, 2009], which constitutes our third measure. The fourth is the Symmetric Extremal Dependence Index of *Ferro and Stephenson* [2011]:

$$\mathrm{SEDI} = \frac{\ln\left[F(1-H)\right] - \ln\left[H(1-F)\right]}{\ln\left[F(1-H)H(1-F)\right]}, \qquad (3)$$

where $H = a/(a + c)$ is the hit rate and $F = b/(b + d)$ is the false-alarm rate. Like SEDS, SEDI is asymptotically equitable and non-degenerate for rate events, but it is also base-rate independent and non-degenerate for overwhelmingly common events. Indeed, it was shown by *Hogan and Mason* [2011] to be the measure that combined the greatest number of desirable properties of any yet devised. In section 5, these measures are compared in the verification of global cloud forecasts, particularly in regard to their degeneracy in regions where cloud occurrence is very low.

The confidence interval in SEDI may be estimated using the formula given by *Hogan and Mason* [2011]. We account approximately for serial correlation in the time-series by assuming that the number of independent samples per month in each latitude band is 1000: there are around 500 orbits in a month, each involving two over-passes of a particular latitude band. Since the return period is 16 days, much longer than the decorrelation time of the clouds, the orbits may be considered uncorrelated. The contingency table cell counts are scaled such that $a + b + c + d = 1000$ before applying the error formula of *Hogan and Mason* [2011].

4. Comparison of mean cloud fraction

In this section the mean cloud fraction is compared between models and observations. Figures 2a–2c depict mean cloud fraction from the observations and the two models in July 2006, while Figs. 2f and 2g depict the same but for the observations and ECMWF model in February 2007. Several features may be observed:

- Cloud fraction in the upper few kilometers of the troposphere was generally well captured by both models;
- Tropical upper-tropospheric cirrus was well captured by both models in July 2006, with models and observations all giving a peak cloud fraction of around 0.35 at an altitude of around 13 km;
- Mid-level cloud fraction (3–8 km) was underestimated in both models at all latitudes, missing as much as half of the total cloud fraction;
- Boundary-layer cloud fraction was predicted within the observational uncertainty, except in the Southern Ocean where both models underestimate cloud fraction by one-third, and both appear to overestimate the cloud fraction associated with Arctic summertime stratus;
- The one-third underestimate of Southern Ocean cloud fraction extended up to 6 km during both months;
- In July 2006, observed cloud fraction included polar stratospheric clouds south of 70°S up to 20 km, although their absence from the models is unsurprising given their very different physics.

To investigate the impact of the method to estimate cloud base height described in section 2, Fig. 3 shows profiles of observed and model cloud fraction at five latitudes, together with the range of uncertainty due only to the observational uncertainty in determining liquid cloud depth. The observed cloud fraction after modifying cloud depth is systematically less than the raw DARDAR cloud fraction, particularly in the tropical and subtropical boundary layer. This is because lidar multiple scattering has a 'pulse stretching' effect on the observed backscatter profile. For physically thin liquid clouds this tends to cause apparent signals below the true cloud base, which DARDAR interprets as cloud. Figures 3b and 3c show that after these modifications, the modeled boundary-layer cloud fraction in the tropics tends to lie within the observed uncertainty range. This range is almost entirely a result of the difference between the *Minnis et al.* [1992] and *Chiu et al.* [2014] relationships between optical depth and geometric depth, which may originate from their studies having been over ocean and land, respectively. Our perturbations of the cloud microphysical and macrophysical properties made much less difference to the retrieved cloud fraction, because the liquid cloud layers were mostly too thin to be sensitive to these assumptions (Fig 1b).

The underestimate of mid-level cloud fraction in mid-latitudes has been reported for a number of models [*Hogan et al.*, 2001; *Zhang et al.*, 2005; *Illingworth et al.*, 2007]. *Hogan et al.* [2001] found that the mid-level bias in the ECMWF model could be improved if 'snow' in the model, which at the time was treated diagnostically, was added to the model cloud fraction using suitable assumptions on fall speed and vertical overlap of precipitation. However, snow cannot be the whole story since the Met Office model, in which snow contributes to cloud fraction, exhibits the same underestimate in Fig. 2c, a result also found by *Bodas-Salcedo et al.* [2008] using a radar forward-modeling approach. Another likely cause of the mid-level difference is the tendency of many models to underestimate the occurrence of mixed-phase layer clouds [e.g. *Xie et al.*, 2008]. The Supporting Information includes plots of cloud fraction split into frequency of occurrence and mean amount when present.

5. Cloud placement verification

The verification measures introduced in section 3.2 are now used to evaluate the skill with which the models predict clouds in the

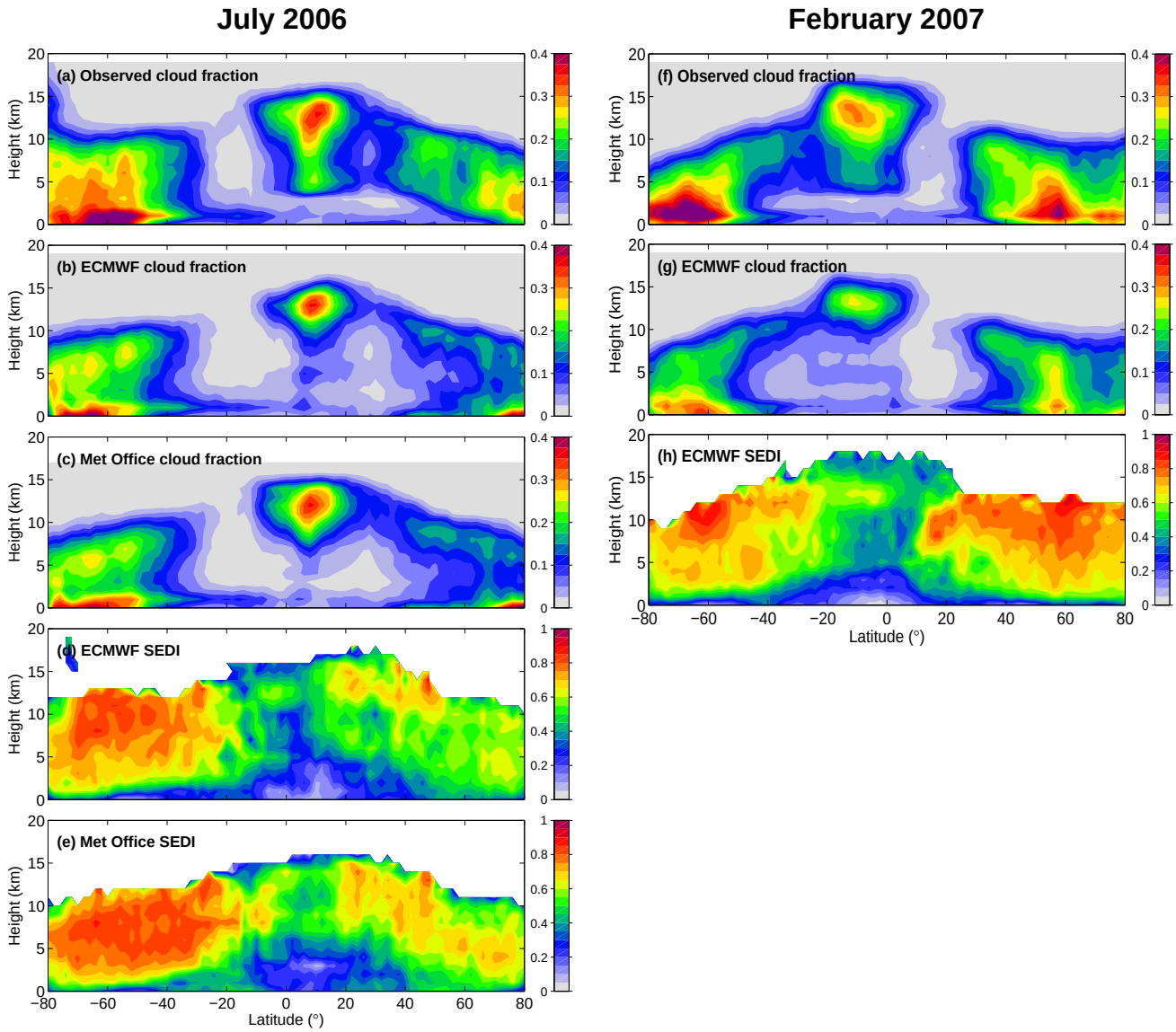

Figure 2. Evaluation of the ECMWF and Met Office models versus latitude and height. Results for July 2006 are shown in the left panels: (a) observed mean cloud fraction, (b) mean cloud fraction in the ECMWF model, (c) mean cloud fraction in the Met Office model, (d) Symmetric Extremal Dependence Index (SEDI) for the ECMWF model, and (e) SEDI for the Met Office model. Results for February 2007 (ECMWF model only) are shown in the right panels: (f) observed mean cloud fraction, (g) modeled mean cloud fraction, and (h) SEDI.

right place and right time. We first compare the behavior of the four measures: Fig. 4 shows these for the ECMWF model in July 2006. Although the color scales are different, each measure awards zero to random forecasts, and all have an upper bound of 1 corresponding to a perfect forecast except for LOR which is unbounded. The most obvious differences in behavior occur where cloud fraction in either or both model and observations tends to zero, which (as can be seen in Fig. 2) occurs near the tropopause, in the southern hemisphere mid- and upper-troposphere subtropics, and in the Antarctic stratosphere. In these regions the degeneracy of the first two measures to meaningless limits is apparent: GSS tends to zero and LOR to large values, as predicted by *Stephenson et al.* [2008]. The behavior of SEDS and SEDI, both designed to be non-degenerate, is much more satisfactory. In particular, in the winter mid-latitudes at a height of around 12 km, where mean cloud fraction is very low, we expect the clouds to be associated with the same large-scale weather systems that generate the clouds a few kilometers below. Since the score awarded should be a measure of the skill with which these weather systems were predicted, a good measure should report an approximately constant score with height in this region. SEDI achieves this, while SEDS has a modest increase of skill with height towards the tropopause. This behavior, together with the theoretical advantages of SEDI [*Hogan and Mason*, 2011], lead us to use SEDI for the remainder of this paper.

Figures 2d and 2e depict SEDI for the ECMWF and Met Office forecasts in July 2006. The patterns are very similar, although the Met Office appears to score consistently higher than ECMWF. This is most likely to be simply due to the fact that the Met Office forecasts are up to 30 h more recent than the ECMWF ones. However, some of the difference at mid-levels may be due to the Met Office's prognostic treatment of ice and snow, and indeed *Hogan et al.* [2009] found that in 2003–2004 the Met Office model predicted mid-level clouds with the greatest skill of the six models they evaluated.

In both July 2006 and February 2007, the highest skill is in the winter-hemisphere mid-latitude mid- and upper troposphere. *Illingworth et al.* [2007] also found higher skill in winter, which

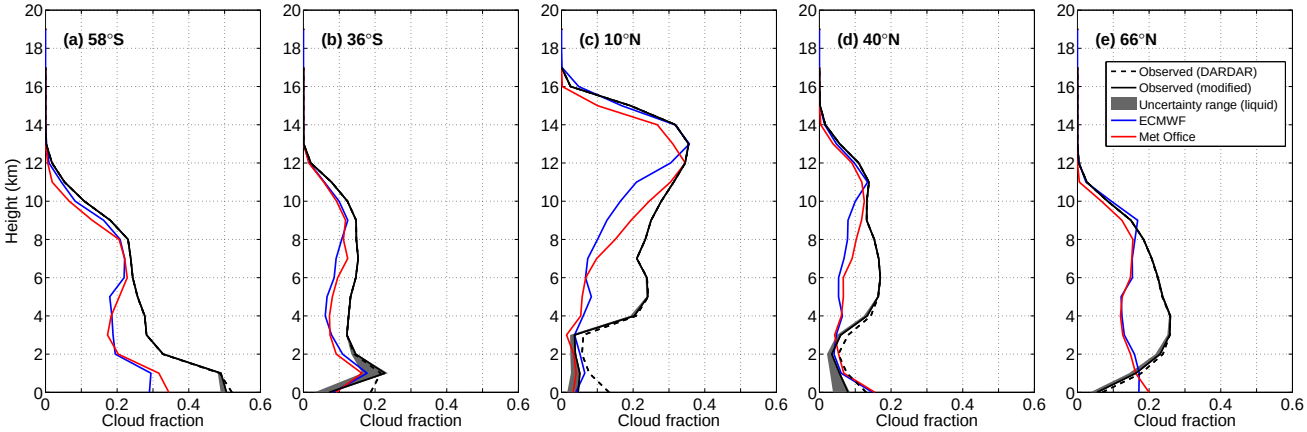

Figure 3. Mean cloud fraction profiles at five different latitudes (58°S, 36°S, 10°N, 40°N, 66°N) for observations and both models during July 2006. The shaded gray region shows the uncertainty in cloud fraction related only to uncertainty of liquid cloud depth.

they attributed to more predictable large-scale cyclonic systems in winter versus less predictable small-scale convective systems in summer. However, they used GSS, so their seasonal cycle in skill will have been amplified by the dependence of GSS on cloud occurrence, which is higher in winter. Interestingly, Figs. 2d and 2h are not mirror images of each other: mid-latitude clouds in the southern-hemisphere (SH) summer are significantly better predicted by ECMWF than those in the northern-hemisphere (NH) summer. Specifically, SH summer values of SEDI above the boundary layer between 50°S and 70°S typically lie in the range 0.65–0.85 (95% confidence intervals ± 0.07), while NH summer values in latitudes 50–70°N lie typically in the range 0.50–0.65 (± 0.08). This difference is likely due to the absence of major land masses in latitudes 50–70°S leading to a much reduced occurrence of summertime convection.

The tropical mid- and upper troposphere exhibits much less skill than in mid-latitudes, because tropical cloud systems are associated with convection, which is intrinsically less predictable than mid-latitude cyclones. To investigate the potential role of the diurnal cycle, SEDI was computed separately for daytime and nighttime (not shown), using A-Train equator crossing times of approximately 0130 and 1330 local time. The results in all parts of latitude-height space are very similar to those shown in Fig. 2. This implies that the reduced skill in the tropics is not so much due to a poor diurnal cycle in the model as to the generally poor placement of convective clouds in the model. Note that this is not simply a measure of the skill of the convection scheme in triggering a particular gridbox at exactly the right time (which would be expected to have virtually no skill at all), but more of the the skill of clouds originating from convective detrainment that may persist long after the convection was triggered.

Finally, we consider boundary-layer clouds, which are predicted with much less skill than those at higher altitudes. The poorest performance occurs in the tropics below 1.5 km; for example, in 2006 the ECMWF value of SEDI at 1-km above the equator was 0.10 ± 0.12, implying that the cloud fraction forecast is not significantly better than a random forecast. In mid-latitudes the skill is a little higher (and significantly better than random), and there is a tendency for lower skill in the SH in both winter and summer. These findings can be explained by the fact that cumulus clouds, which are ubiquitous over tropical oceans, are associated with small spatial and temporal scales that are inherently unpredictable. Another factor is that satellite sounders rarely provide useful information down into the boundary layer, so assimilation systems are more dependent on surface and radiosonde observations, which are much sparser in the tropics and SH.

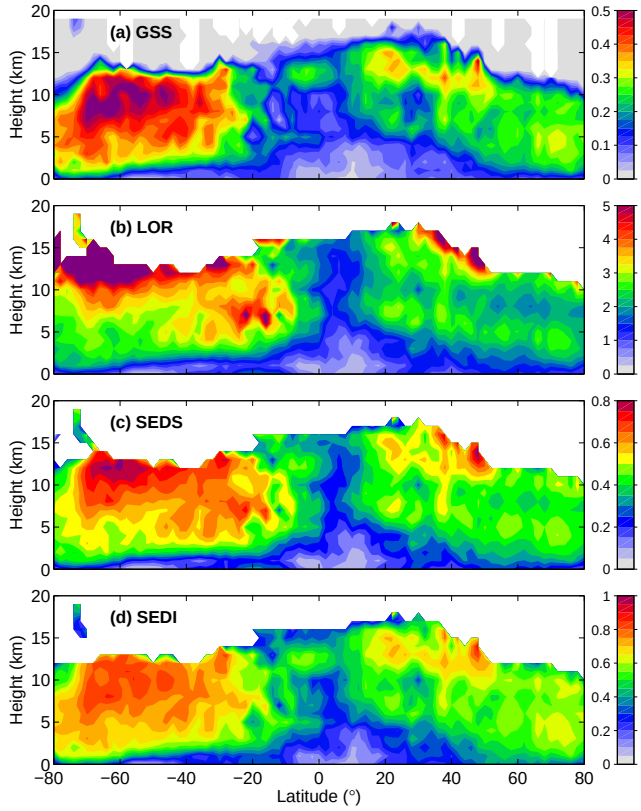

Figure 4. Four verification measures applied to ECMWF cloud fraction in July 2006: (a) Gilbert Skill Score, (b) Log of Odds Ratio, (c) Symmetric Extreme Dependency Score, and (d) Symmetric Extremal Dependence Index.

6. Conclusions

In this paper we have demonstrated how combined spaceborne radar and lidar may be used to evaluate cloud fraction predictions by weather forecast models, both in terms of bias and skill. While the two model versions used here are not particularly recent, they are sufficient to demonstrate the promise of the technique. We find that both models under-predict mid-level cloud fraction at all latitudes and boundary-layer cloud in the Southern Ocean, in agreement with previous studies. We then performed the first verification

of the full vertical profile of cloud forecasts on a global scale. A comparison of the behavior of different verification measures revealed the strong dependence of the Gilbert Skill Score, GSS (also known as the Equitable Threat Score), on the frequency of cloud occurrence. Its widespread use in operational verification is therefore rather concerning, since variations purely due to a correlation with frequency of occurrence may be being misinterpreted as a change in performance of the underlying forecast system. Much more reliable behavior was found with the Symmetric Extremal Dependence Index, SEDI.

We discovered a very wide range in the skill of forecasts of different global cloud types, from tropical boundary-layer clouds having $\mathrm{SEDI} < 0.2$ through tropical convection and cirrus with $0.4 < \mathrm{SEDI} < 0.6$ to mid-latitude winter-season cyclonic systems with $0.6 < \mathrm{SEDI} < 0.85$. These differences are partly because convective clouds are inherently less predictable than larger-scale weather systems, but they also suggest that the physics of boundary-layer clouds and deep convection has the greatest scope for improvement in global models. The approach to cloud verification presented in this paper could be used in an operational context within the model development cycle to objectively assess the impact of changes to model physics.

Acknowledgements

We thank Christine Chiu for valuable discussions. A. Bodas-Salcedo was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101).

References

Albrecht B A, Penc R S and Schubert W H 1985 An observational study of cloud-topped mixed layers *J. Atmos. Sci.* **42** 800–822

Albrecht B A, Fairall C W, Thomson D W, White A B, Snider J B and Schubert W H 1990 Surface-based remote sensing of the observed and the adiabatic liquid water content of stratocumulus clouds *Geophys. Res. Lett.* **17** 89–92

Bodas-Salcedo A, Webb M J, Brooks M E, Ringer M A, Williams K D, Milton S F and Wilson D R 2008 Evaluating cloud systems in the Met Office global forecast model using simulated CloudSat radar reflectivities *J. Geophys. Res.* **113** D00A13, doi:10.1029/2007JD009620

Boucher O, Randall D, Artaxo P, Bretherton C, Feingold G, Forster P, Kerminen V-M, Kondo Y, Liao H, Lohmann U, Rasch P, Satheesh S K, Sherwood S, Stevens B and Zhang X Y 2013 Clouds and Aerosols *In Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker T F, Qin D, Plattner G K, Tignor M, Allen S K, Boschung J, Nauels A, Xia Y, Bex V and Midgley P M (eds.)]. Cambridge.

Chiu J C, Holmes J A, Hogan R J and O'Connor E J 2015 The interdependence of continental warm cloud properties derived from unexploited solar background signals in ground-based lidar measurements *Atmos. Chem. Phys.* **14** 8389–8401

Delanoë J and Hogan R J 2010 Combined CloudSat-CALIPSO-MODIS retrievals of the properties of ice clouds *J. Geophys. Res.* **115** D00H29, doi:10.1029/2009JD012346

Delanoë J, Hogan R J, Forbes R M, Bodas-Salcedo A and Stein T H M 2011 Evaluation of ice cloud representation in the ECMWF and UK Met Office models using CloudSat and CALIPSO data *Q. J. R. Meteorol. Soc.* **137** 2064–2078

Ferro C A T and Stephenson D B 2011 Extremal dependence indices: improved verification measures for deterministic forecasts of rare binary events *Weath. Forecasting* **26** 699–713

Gilbert G K 1884 Finley's tornado predictions *Amer. Meteorol. J.* **1** 166–172

Hogan R J and Mason I B 2011 Deterministic forecasts of binary events *in Forecast verification: a practitioner's guide in atmospheric science* Jolliffe I T and Stephenson D B (eds.) 2nd Ed., Wiley

Hogan R J, Jakob C and Illingworth A J 2001 Comparison of ECMWF winter-season cloud fraction with radar derived values *J. Appl. Meteorol.* **40** 513–525

Hogan R J 2008 Fast lidar and radar multiple-scattering models: Part 1: Small-angle scattering using the photon variance-covariance method *J. Atmos. Sci.* **65** 3621–3635

Hogan R J and Battaglia A 2008 Fast lidar and radar multiple-scattering models: Part 2: Wide-angle scattering using the time-dependent two-stream approximation *J. Atmos. Sci.* **65** 3636–3651

Hogan R J, O'Connor E J and Illingworth A J 2009 Verification of cloud-fraction forecasts *Q. J. R. Meteorol. Soc.* **135** 1494–1511

Hogan R J, Ferro C A T, Jolliffe I T and Stephenson D B 2010 Equitability revisited: why the 'equitable threat score' is not equitable *Weath. Forecasting* **25** 710–726

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** 883–898

Jakob C, Pincus R, Hannay C, and Xu K-M 2004 The use of cloud radar for model evaluation: A probabilistic approach *J. Geophys. Res.* **109** D03202, doi:10.1029/2003JD003473

Mace G G, Jakob C, and Moran K P 1998 Validation of hydrometeor occurrence predicted by the ECMWF model using millimeter wave radar data *Geophys. Res. Lett.* **25** 1645–1648

Mace GG, Zhang Q, Vaughan M, Marchand R, Stephens G, Trepte C, and Winker D 2009 A description of hydrometeor layer occurrence statistics derived from the first year of merged CloudSat nad CALIPSO data *J. Geophys. Res.* **114** D00A26 doi:10.1029/2007JD009755

Miles N L, Verlinde J and Clothiaux E E 2000 Cloud droplet size distributions in low-level stratiform clouds *J. Atmos. Sci.* **57** 295–311

Minnis P, Heck P W, Young D F, Fairall C W and Snider J B 1992 Stratocumulus cloud properties derived from simultaneous satellite and island-based instrumentation during FIRE *J. Appl. Meteorol.* **31** 317–339

Nicholls S and Turton J D 1986 An observational study of the structure of stratiform cloud sheets: Part II. Entrainment *Q. J. R. Meteorol. Soc.* **112** 461–480

O'Connor E J, Illingworth A J and Hogan R J 2004 A technique for autocalibration of cloud lidar *J. Atmos. Oceanic Tech.,* **21,** 777–786.

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

Stein T H M, Delanoë J and Hogan R J 2011 A comparison between four different retrieval methods for ice-cloud properties using data from the CloudSat, CALIPSO, and MODIS satellites *J. Appl. Meteorol. Climatology* **50** 1952-1969

Stephens G L et al. 2002 The CloudSat mission and the A-train *Bull. Am. Meteorol. Soc.* **83** 1771–1790

Stephenson D B 2000 Use of the odds ratio for diagnosing forecast skill *Weath. Forecasting* **15** 221–232

Stephenson D B, Casati B, Ferro C A T and Wilson C A 2008 The extreme dependency score: a non-vanishing measure for forecasts of rare events *Meteorol. Apps.* **15** 41–50

Tiedtke M 1993 Representation of clouds in large-scale models *Mon. Weath. Rev.* **121** 3040–3061

Warner, J 1955 The water content of cumuliform cloud *Tellus* **7** 449–457

Wilkinson J M, Hogan R J, Illingworth A J and Benedetti A 2008 Use of a lidar forward model for global comparisons of cloud fraction between the ICESat lidar and the ECMWF model *Mon. Weath. Rev.* **136** 3742–3759

Wielicki B A, Harrison E F, Cess R D, King M D and Randall D A 1995 Mission to Planet Earth: Role of clouds and radiation in climate *Bull. Am. Meteorol. Soc.* **76** 2125–2153

Wood R 2012 Stratocumulus clouds *Mon. Wea. Rev.* **140** 2373–2423

Xie S, Boyle J, Klein S A, Liu X and Ghan S 2008 Simulations of Arctic mixed-phase clouds in forecasts with CAM3 and AM2 for M-PACE *J. Geophys. Res.* **113** D04211, doi:10.1029/2007JD009225

Yang G, and Slingo J M 2001 The diurnal cycle in the tropics *Mon. Weath. Rev.* **129** 784–801

Zhang M H, Lin W Y, Klein S A, Bacmeister J T, Bony S, Cederwall R T, Del Genio A D, Hack J J, Loeb N G, Lohmann U, Minnis P, Musat I, Pincus R, Stier P, Suarez M J, Webb M J, Wu J B, Xie S C, Yao M-S and Zhang J H 2005 Comparing clouds and their seasonal variations in 10 atmospheric general circulation models with satellite measurements *J. Geophys. Res.* **110** D15S02, doi:10.1029/2004JD005021

Zhou M, Zeng X, Brunke M, Zhang Z and Fairall C 2006 An analysis of statistical characteristics of stratus and stratocumulus over eastern Pacific *Geophys. Res. Lett.* **33** L024796, doi:10.1029/2005GL024796

Corresponding author: R. J. Hogan, ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom (r.j.hogan@ecmwf.int)