Using continuous ground-based radar and lidar measurements for evaluating the representation of clouds in four operational models

Dominique Bouniol, Alain Protat, Julien Delanoë, Jacques Pelon, Jean-Marcel Piriou, François Bouyssel, Adrian M. Tompkins, Damian R. Wilson, Yohann Morille, Martial Haeffelin, et al.

▶ **To cite this version:**

Dominique Bouniol, Alain Protat, Julien Delanoë, Jacques Pelon, Jean-Marcel Piriou, et al.. Using continuous ground-based radar and lidar measurements for evaluating the representation of clouds in four operational models. Journal of Applied Meteorology and Climatology, 2010, 49 (9), pp.1971-1991. 10.1175/2010JAMC2333.1 . hal-00477076

HAL Id: hal-00477076
https://hal.science/hal-00477076

Submitted on 20 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Using Continuous Ground-Based Radar and Lidar Measurements for Evaluating the Representation of Clouds in Four Operational Models

DOMINIQUE BOUNIOL,[a] ALAIN PROTAT,[b,c] JULIEN DELANOË,[d] JACQUES PELON,[e] JEAN-MARCEL PIRIOU,[a] FRANÇOIS BOUYSSEL,[a] ADRIAN M. TOMPKINS,[f,*] DAMIAN R. WILSON,[g] YOHANN MORILLE,[h] MARTIAL HAEFFELIN,[i] EWAN J. O'CONNOR,[d,j] ROBIN J. HOGAN,[d] ANTHONY J. ILLINGWORTH,[d] DAVID P. DONOVAN,[k] AND HENK-KLEIN BALTINK[k]

[a] *GAME/CNRM, CNRS/Météo-France, Toulouse, France*
[b] *Centre for Australian Weather and Climate Research, Melbourne, Australia*
[c] *LATMOS-IPSL, CNRS/INSU, Vélizy, France*
[d] *Meteorology Department, University of Reading, Reading, United Kingdom*
[e] *LATMOS-IPSL, CNRS/INSU, Paris, France*
[f] *European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom*
[g] *Met Office, Exeter, United Kingdom*
[h] *Laboratoire de Météorologie Dynamique, CNRS, Palaiseau, France*
[i] *Institut Pierre-Simon Laplace, CNRS, Palaiseau, France*
[j] *Finnish Meteorological Institute, Helsinki, Finland*
[k] *Royal Netherlands Meteorological Institute, De Bilt, Netherlands*

(Manuscript received 25 June 2009, in final form 1 April 2010)

ABSTRACT

The ability of four operational weather forecast models [ECMWF, Action de Recherche Petite Echelle Grande Echelle model (ARPEGE), Regional Atmospheric Climate Model (RACMO), and Met Office] to generate a cloud at the right location and time (the cloud frequency of occurrence) is assessed in the present paper using a two-year time series of observations collected by profiling ground-based active remote sensors (cloud radar and lidar) located at three different sites in western Europe (Cabauw, Netherlands; Chilbolton, United Kingdom; and Palaiseau, France). Particular attention is given to potential biases that may arise from instrumentation differences (especially sensitivity) from one site to another and intermittent sampling. In a second step the statistical properties of the cloud variables involved in most advanced cloud schemes of numerical weather forecast models (ice water content and cloud fraction) are characterized and compared with their counterparts in the models. The two years of observations are first considered as a whole in order to evaluate the accuracy of the statistical representation of the cloud variables in each model. It is shown that all models tend to produce too many high-level clouds, with too-high cloud fraction and ice water content. The midlevel and low-level cloud occurrence is also generally overestimated, with too-low cloud fraction but a correct ice water content. The dataset is then divided into seasons to evaluate the potential of the models to generate different cloud situations in response to different large-scale forcings. Strong variations in cloud occurrence are found in the observations from one season to the same season the following year as well as in the seasonal cycle. Overall, the model biases observed using the whole dataset are still found at seasonal scale, but the models generally manage to well reproduce the observed seasonal variations in cloud occurrence. Overall, models do not generate the same cloud fraction distributions and these distributions do not agree with the observations. Another general conclusion is that the use of continuous ground-based radar and lidar observations is definitely a powerful tool for evaluating model cloud schemes and for a responsive assessment of the benefit achieved by changing or tuning a model cloud parameterization.

* Current affiliation: International Centre for Theoretical Physics, Trieste, Italy.

Corresponding author address: Dominique Bouniol, MOANA/GMME/CNRM, 42 avenue Gaspard Coriolis, 31057 Toulouse CEDEX, France.
E-mail: dominique.bouniol@meteo.fr

DOI: 10.1175/2010JAMC2333.1

© 2010 American Meteorological Society

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

1. Introduction

The improvement of the representation of clouds in models has been a major issue in climate and numerical weather prediction (NWP) over the last 20 years. For NWP the need of increasingly more accurate forecasts not only in cloud cover but in other variables modulated by cloud properties, such as surface precipitation, temperature or shortwave–UV radiation, highlights the need for an accurate prediction of the vertical and horizontal distributions of cloud ice and liquid water content (Illingworth et al. 2007).

Efforts to improve cloud parameterizations have led to an increase in the degree of realism of the simulated clouds (Jakob 2003). The degree of realism for these models is such that it becomes relevant to make direct comparisons with observations. As stressed by Siebesma et al. (2004), model improvement necessarily begins with an assessment of current model performance and the identification of model shortcomings.

In the past decades, the clouds simulated by weather forecast models have been assessed using radiative fluxes and cloud cover diagnosed from satellite data (Morcrette 1991; Jakob 1999; Yang et al. 1999). The strength of satellite observations lies in their monitoring capability, allowing for global climatologies of cloud properties to be constructed (e.g., Rossow and Schiffer 1983). The drawback is that information on the detailed vertical structure of clouds is usually lacking (Illingworth et al. 2007). A new opportunity for model/observations comparisons has been opened by the collection of long time series of cloud properties at given points, in particular through the Atmospheric Radiation Measurement Program (ARM) (Stokes and Schwartz 1994). The principle is to implement (at each operational site) a complete set of instruments (including cloud radar and lidar) in order to document the column cloud properties as fully as possible in different synoptic situations. Since April 2006, the spaceborne *CloudSat* (Stephens et al. 2002)/Clouds Aerosol lidar and Infrared Pathfinder Satellite Observations (CALIPSO; Winker et al. 2007) tandem mission provides a replica of these ground-based observations. However, because of the sun-synchronous orbit, the same point is always seen at the same local time, resulting in a poor sampling of the diurnal cycle. Liu et al. (2008) suggest that it is therefore important to interpret the day and night A-Train samples as independent samples. However, these instruments are most suited to a global evaluation of cloud properties (see Bodas-Salcedo et al. 2008). The ground-based approach is particularly suitable if one wants to accurately characterize the statistical properties of the cloud cover at a particular location, its vertical distribution, and its diurnal cycle. Ground-based instrumentation like that implemented by ARM can provide vertical profiles of cloud properties every 10 s with a vertical resolution between 45 to 90 m. The use of long time series allows the cloud-generation processes in models to be tested in a wide range of situations. The continuous ground-based sampling of cloud properties has already been successfully used in several studies either on a limited time period, for instance a season (Morcrette 2002; Guichard et al. 2003; Mathieu et al. 2006), or for given parameters, such as cloud fraction (Hogan et al. 2001), cloud occurrence, the influence of clouds on the radiation budget of the surface, the vertical atmospheric column, and the top of the atmosphere (Mace and Benson 2008).

Following the example of the ARM sites, a network of cloud-observing stations, all equipped with active sensors such as lidars and Doppler millimeter-wavelength radars, has been developed at three sites in Europe (Cabauw in the Netherlands, Chilbolton in the United Kingdom, and Palaiseau in France) in the framework of the Cloudnet project. It is noteworthy that another site, the Lindenberg observatory in Germany, now also has similar cloud remote sensing capabilities in western Europe. The objectives of this project, the instrumental setup and the cloud products derived at these sites are fully described in Illingworth et al. (2007). Some differences in cloud property characteristics may be interpreted as regional cloud properties. However, differences may result from differences in instrumentation implemented from one site to another. The regional variability of cloud properties is therefore not investigated in this paper.

The aim of the present paper is to compare the coincident observed and simulated clouds at the three stations during the two years of the project, and for the four models initially involved in Cloudnet [model from the European Center for Medium-Range Weather Forecasts (ECMWF), the Action de Recherche Petite Echelle Grande Echelle model from Météo-France (ARPEGE), the Regional Atmospheric Climate Model from the Royal Netherlands Meteorological Institute (KNMI; RACMO), and the Limited Area Model of the UK Met Office, referred to as Met Office]. After examining the mean model-observations differences for the entire period, the capability of the models to reproduce the seasonal variability is assessed. The winter season at midlatitudes is mainly affected by frontal systems, and therefore mainly stratiform precipitation, while summer is characterized by deep convective storms and associated cloud systems. Such a seasonal study at midlatitudes should therefore shed some light on the capability of the models to reproduce these very different types of cloud-producing large-scale situations. This methodology is applied to three cloud parameters: the cloud frequency

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

TABLE 1. Summary of the model setup and model cloud schemes involved in the Cloudnet project.

Model	ECMWF	Arpege1	Arpege2	RACMO	Met Office
Cloud scheme	Tiedtke (1993)		Xu and Randall (1996) for cloud fraction	Lenderink et al. (2003)	Smith (1990); Wilson and Ballard (1999)
	Global	Global	Global	Regional	Regional
Horizontal resolution (km)	39	24	24	54	12
No. of vertical levels	60	41	41	40	38
Cloud fraction	Prognostic	Diagnostic from water vapor excess	Diagnostic from IWC and LWC	Prognostic	Diagnostic for ice from the IWC
LWC	Prognostic from total water	Diagnostic from water vapor excess	Diagnostic	Prognostic	Prognostic with water vapor
IWC	Prognostic from total water	Diagnostic from water vapor excess	Diagnostic	Prognostic from total water	Prognostic

of occurrence (referred as cloud occurrence in the following), the cloud fraction, and the ice water content (IWC). The liquid water content is not evaluated in the present paper because of the difficulty in accurately retrieving this quantity with the present instrumental network.

In the following section the cloud schemes of the four models are briefly described. The cloud occurrence in the model at the four sites is then compared to the observations in section 3. In sections 4 and 5, the observed cloud variables involved in the weather forecast model cloud parameterizations (the cloud fraction and the IWC) are then analyzed and compared with the same quantities in the models. Concluding remarks are given in the final section.

2. Description of the models

A brief comparison of the models and of their cloud scheme is given in this section in order to clearly specify which version of each model is evaluated in the present paper. Operational centers are continuously upgrading their models and the conclusions drawn from our analysis apply to a specific set of model versions.

The four models involved in this evaluation are NWP models. ECMWF and ARPEGE are global NWP models, while RACMO and Met Office models involved in this study are regional models. All models use different horizontal and vertical resolutions (see Table 1). The profiles are extracted from the operational forecasts over the three observational sites (nearest grid point) every hour and correspond to $T + 12$ to $T + 36$ forecasts for ECMWF, ARPEGE, and RACMO models, and $T + 6$ to $T + 12$ forecasts 4 times per day for the Met Office model. All the comparisons proposed in the remaining part of this paper have also been performed as a function of the forecast times and all the obtained conclusions remain valid.

Table 1 provides a summary of the cloud schemes. While the emphasis is placed on the differences in the cloud schemes between the models, it is obvious that clouds are the result of complex interacting processes, and thus errors in cloud variables may result from deficiencies in other model components, such as the vertical diffusion or convection parameterizations for instance, or in the representation of the interaction between these model components. Moreover, since the cloud variables are from short-range operational forecasts, there will be a strong dependency on the data assimilation system. For example, during the study period ECMWF used a 4D variational assimilation system, which contrasts to the 3D system in place at the Met Office at the time. In particular, moisture analysis systems and humidity data usage and bias correction are particularly diverse (Anderson et al. 2005), which is relevant because relative humidity is the main predictor of cloud properties in both prognostic and diagnostic cloud schemes. Caution has therefore been exercised in this paper in order not to overstate the results of this study directly in terms of the differences in the cloud parameterizations themselves.

The ARPEGE model is a particular case in this study, since in the course of Cloudnet a totally different cloud parameterization [based on Xu and Randall (1996) for the formulation of cloud fraction] has replaced the original one on 14 April 2003. In both schemes though, the cloud variables are still diagnosed (no use of prognostic equations). Because of this complete change of philosophy during Cloudnet, the ARPEGE dataset has been split in two, labeled arpege1 and arpege2.

The ECMWF model profiles provided for this project come from a series of eight model releases, or cycles,

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

some of which included major changes to the model moist physics or data assimilation systems. The initial cycle in place at the project onset (1 October 2002) was 25R1, which remained operational until 14 January 2003. After this date, cycle 25r3 included a major upgrade to the cloud, convection schemes, and numerical solver methodology that led to a model climate with increased cloud ice amounts in both the tropics and midlatitudes, while cloud liquid water was reduced. Shortly before the end of the project cycle 28r3 made a major change to the numerics of the moist physics, leading to a substantially reduced global cloud liquid water amounts. However, the amount of data originated from this cycle does not affect the statistics reported here and the ECMWF time series is not split into the pre- and postcycle 25r3 periods.

The RACMO model, implemented at KNMI, uses the same dynamics as the High-Resolution Limited Area Model (HIRLAM) and is based on the ECMWF 23r4 cycle physics, which has been used for the 40-yr ECMWF Re-Analysis (ERA-40). It is expected that differences in cloud representation between ECMWF and RACMO would result mainly from differences in the dynamics, and data assimilation schemes, although the upgrades to the ECMWF system outlined above could also play some role in the observed differences.

The Met Office Unified Model used for these comparisons has one combined prognostic variable for water vapor and cloud liquid water and one prognostic variable for ice. In addition it must be noted that cloud fraction and IWC for all models except the Met Office include only nonprecipitating ice (see Illingworth et al. 2007).

3. Comparison of model cloud occurrence with observations

In this section the cloud occurrence in both the observations and in the four models is compared over the three Cloudnet sites. The model data are hourly snapshots over the Cloudnet sites. A preliminary work in these comparisons consists in ''degrading'' the observations (30-s time resolution, 30–60-m vertical resolution) to the scale of the model (1-h time resolution, 12–55-km horizontal resolution, coarser and varying vertical resolution). This is done following previous workers (e.g., Mace et al. 1998; Hogan et al. 2001) by assuming that the temporal sampling yields the equivalent of a two-dimensional slice through the three-dimensional grid box and averaging using the height levels of the models. Using the model wind speed as a function of height and the horizontal model grid-box size, the appropriate sampling time is calculated (this time is, however, constrained to lie between 10 and 60 min). It is assumed that in this time the cloud structure observed is predominantly due to the advection of structure within the grid box across the site, rather than evolution of the cloud during the period. These time series are two-dimensional while a model grid box represents a three-dimensional volume. It is assumed (Brooks et al. 2005; Hogan et al. 2001; Protat et al. 2010) that the fraction of the grid filled by clouds in the two-dimensional grid box represents the amount of cloud in the three-dimensional volume. Cloud fractions have been computed from the observed pixel subsample in each model grid box as the fraction of observed pixels that are categorized as cloud from the radar–lidar observations. When the instruments operate less than 20% of the averaging time required to compute cloud fraction in a model grid box, the obtained value was not included in the statistics.

A grid box (in model or observations) is considered cloudy when its cloud fraction is larger than 0.03. Different threshold values (lower than 0.1) have been tested and yielded very similar statistics [Hogan et al. (2001) reached the same conclusion]. This threshold is applied because most models (except Met Office) do not produce exact zero cloud fraction values. A given number of model grid box will be considered as cloud free (or with an underestimated cloud fraction) because only nonprecipitating hydrometeors are included in the cloud fraction calculation (except for Met Office). Hogan et al. (2001) modified the model cloud fractions by converting the snow fluxes in cloud fraction and they need to define thresholds on precipitation flux for this purpose. In this paper it has been chosen to keep the cloud fraction as it is in the model because these values are used for further calculations (radiative code for instance). However, it must be kept in mind that the amount of cloud (and cloud fraction) may be underestimated in particular at midlevels where precipitating ice may be encountered.

The cloud occurrence is defined as the ratio of cloudy hours to the total number of observational hours for a given height bin at the model resolution. In this section model cloud occurrence is compared with observations for the whole Cloudnet period in order to evaluate the ''climatological'' representation of clouds in the models. The seasonal variability of the model performances is then investigated in further detail.

a. Comparison for the whole Cloudnet period

1) Derivation of comparable observational and model profiles: role of intermittent radar and lidar sampling and instrumental effects

A first difficulty to overcome when comparing models and observations is to remove potential biases in the observations. Illingworth et al. (2007) detailed the instrumentation setup for each site as well the operation

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

issues that result in intermittent sampling. There are four: 1) The Palaiseau site 532-nm lidar system is not protected from precipitation damage and must be regularly checked by an operator; therefore its temporal sampling is mostly limited to daytime and to periods with low risk of precipitation. 2) The Chilbolton site radar only began operating continuously in April 2003 because of the previous failure of its tube (Hogan et al. 2003). 3) The Palaiseau site radar was only operated during daytime four days per week during the first year of the project. 4) Cabauw site radar operated only 30 days during autumn 2003.

Biases are also produced by the instrument sensitivities: for the lidar or ceilometer, the occurrence of a water cloud below an ice cloud will lead to total extinction of the lidar/ceilometer signal by the strong scattering by the water cloud droplets and any ice cloud above will not be detected. Similarly, cloud radars do not detect all thin high-altitude ice clouds because of their limited sensitivity (hereinafter the sensitivity effect). This sensitivity (generally given at 1 km) decreases as the inverse function of the squared distance from the radar. These clouds are generally considered as radiatively important when their optical depth is larger than 0.05 (Brown et al. 1995). A preliminary study of these biases has been conducted in Protat et al. (2006), focusing on the Palaiseau site and the ECMWF model.

A comparison of the cloud occurrence for the whole Cloudnet period at each site and for each model is shown in Fig. 1. To account for the continuous radar sensitivity loss (due to the power loss of the 95-GHz Klystron tubes; Hogan et al. 2003) at Palaiseau and Chilbolton, the sensitivity of the radars is computed at each time by interpolating from regular measurements of the transmitted power, instead of considering a mean value of sensitivity as in Protat et al. (2006). This radar sensitivity effect is then computed (as in Protat et al. 2006); that is, by converting the model IWC to a reflectivity value using the Liu and Illingworth (2000) IWC–Z relationships. These simulated reflectivity profiles suffer from two uncertainties: the accuracy of the IWC–Z relationship itself, and the fact that the model IWCs are used. Hence the magnitude of the instrumental effect correction is driven by the model IWC. This aspect has been taken into account in Fig. 1 by computing the cloud occurrence if the radar sensitivity was 3 dB higher/smaller. These (thin dotted) lines define an envelope around the occurrence with sensitivity effect taken into account giving a range of confidence in the final profiles. The comparison between model, after removal of the intermittent sampling effect (dashed line) and instrumental sensitivity effect, and observations should be made between the dotted line and the black line in Fig. 1.

The comparison of the different model (gray) lines in Fig. 1 indicates if the model subsample that includes sampling and instrumental effects (dotted line) is representative of the cloud properties derived from the total model sample (solid line). In agreement with Protat et al. (2006) it is found that the Palaiseau instrumental combination (95-GHz radar, 532-nm lidar) provides an unbiased estimate of cloud occurrence up to around 9.5 km of altitude (bottom line of Fig. 1). The same conclusion applies for the Cabauw site (35-GHz radar, ceilometer, first line of Fig. 1) except for the RACMO model, for which the account for the instrumental effect has a larger impact, probably owing to the fact that the RACMO model produces clouds significantly thinner (and thus not detectable with the cloud radar) than the others. For the Chilbolton site, the 94-GHz radar–ceilometer combination provides unbiased frequency of cloud occurrence up to around 7.5 km only. A visual check of the radar and ceilometer time series reveals that the ceilometer never detects ice clouds undetected by the radar, which is not the case with the much more powerful lidar at Palaiseau. Thus the instrumental complementarily observed at the Palaiseau site is not occurring at Chilbolton, resulting in biases starting as low as 7.5 km. The Cabauw site is also implementing a ceilometer, but the better sensitivity of the 35-GHz radar allows the maintenance of a representative sampling of high-level clouds up to 9.5-km height.

2) Comparison between models and observations

The observed cloud occurrence (thick black lines in Fig. 1) is now directly compared with the model occurrence including instrumental effects (dotted gray lines in Fig. 1). For further (and quantitative) comparison the clouds have also been binned in three categories: low level (between 0.5 and 3 km), midlevel (between 3 and 7 km), and high level (between 7 and 9.5 km). The upper altitude limit has been set to 9.5 km because it corresponds to an unbiased sample for comparison at Cabauw and Palaiseau (see previous section). A normalized difference in occurrence is computed at each height level for models and observations as $\Delta_{\text{norm}}\text{Occurrence} = (\text{Occurrence}_{\text{model}} - \text{Occurrence}_{\text{obs}})/\text{Occurrence}_{\text{obs}}$ are reported in Table 2 for each site (Cabauw–Chilbolton–Palaiseau), a 100% value meaning that the model occurrence is twice that observed. A normalized difference is also computed between 0.5 and 9.5 km of altitude (first line of Table 2).

The ECMWF model (first column in Fig. 1 and Table 2) exhibits a very good agreement for midlevel clouds (normalized differences smaller than 0.13) over the three

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

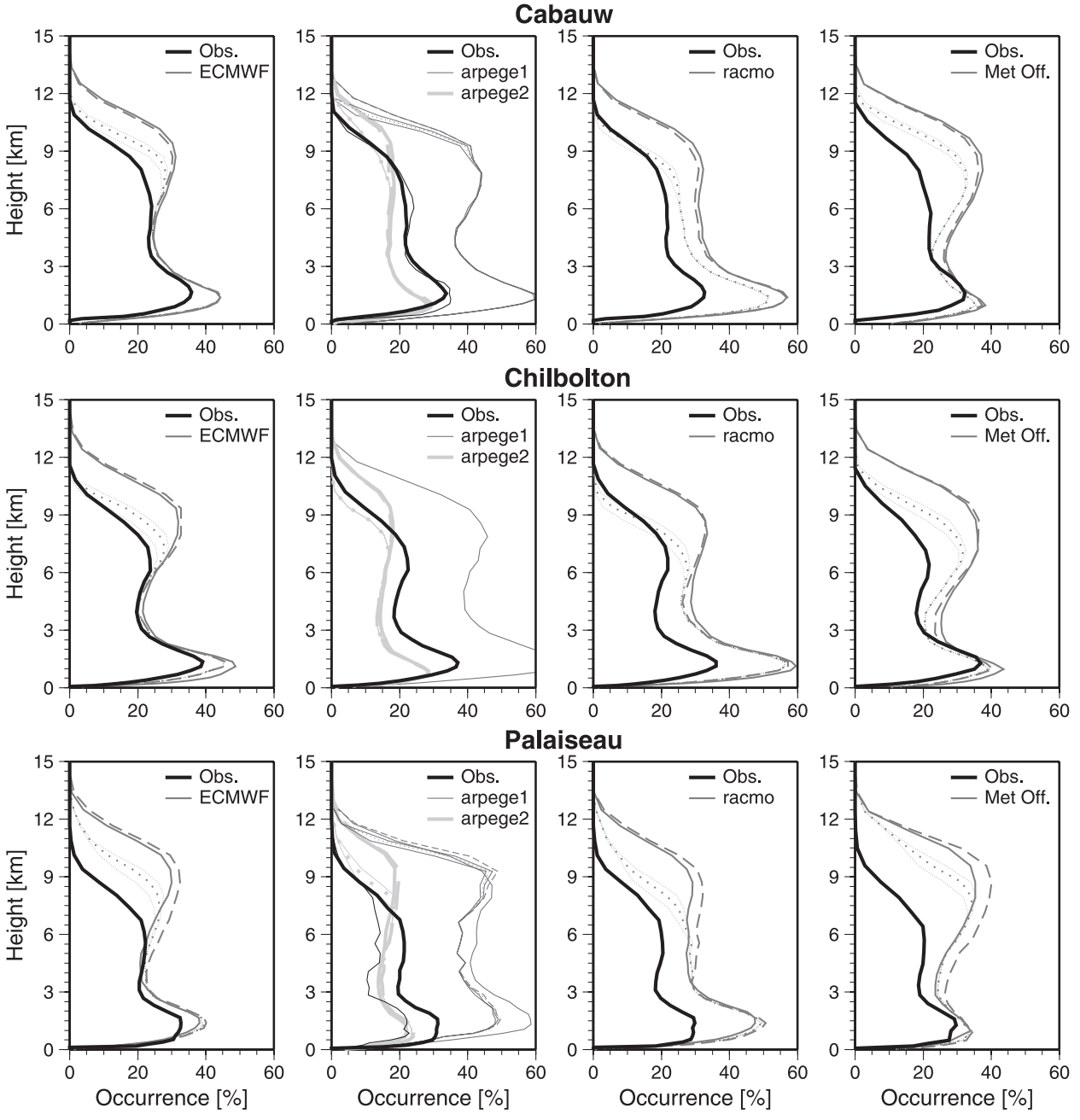

FIG. 1. Frequency of cloud occurrence for models and observations obtained at the three sites for the whole Cloudnet period (1 Oct 2002–30 Sep 2004). Black line shows the frequency of cloud occurrence obtained from the observations. For ARPEGE, the thinnest line corresponds to arpege1 period, the thickest line to arpege2. For models, the gray lines with different styles correspond to the model samples: solid line for whole model sample, dashed line for model subsample corresponding to instrument hours of operations, and dotted line for model subsample corresponding to instrument hours of operations and including instrumental effects. Thinnest dotted lines show the same occurrence including sensitivity effect but accounting for a 3-dB more–less sensitive radar. Each line corresponds to an observatory (top to bottom) Cabauw, Chilbolton, and Palaiseau. (left to right) comparison with ECMWF, ARPEGE (1 and 2), RACMO, and Met Office.

sites and an overestimation of the high-level and low-level cloud occurrences. Although the overestimation is about the same over all sites for the low-level cloud occurrence (0.11 to 0.14), the overestimation of the high-level cloud occurrence is much larger at Palaiseau (1.57) than at Chilbolton and Cabauw (0.40–046).

The ARPEGE model (second column in Fig. 1 and second and third in Table 2) exhibits a radically different

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

TABLE 2. Normalized differences in cloud occurrence defined as $\Delta_{norm}\text{Occurrence} = (\text{Occurrence}_{model} - \text{Occurrence}_{obs})/\text{Occurrence}_{obs}$ for all clouds between 0.5 and 9.5 km of height and for the three cloud categories (low, mid-, and high level); see text for definition. Three numbers are reported corresponding to the normalized differences at Cabauw/Chilbolton/Palaiseau.

	ECMWF	Arpege1	Arpege2	RACMO	Met Office
Δ(all level)	0.19/0.14/0.22	0.74/—/1.81	−0.18/−0.27/−0.13	0.39/0.45/0.56	0.19/0.17/0.41
Δ(low level)	0.11/0.14/0.15	0.61/—/1.30	−0.15/−0.28/−0.17	0.56/0.58/0.59	0.053/0.013/0.15
Δ(midlevel)	0.092/0.076/0.13	0.71/—/2.06	−0.23/−0.27/−0.15	0.25/0.36/0.47	0.25/0.20/0.31
Δ(high level)	0.46/0.40/1.57	1.56/—/4.31	−0.13/−0.26/0.38	0.12/−0.0003/0.93	0.82/0.74/2.21

behavior between the two cloud schemes used during the Cloudnet period. The arpege1 scheme (thin light gray lines in Fig. 1 to be compared with the thin black line) produces a very strong and systematic overestimation (larger than 0.70). It also generates large occurrences of high-level clouds that are classified as ''detectable'' by the instruments, since there is no difference between the total model profile and the profile with instrumental effects included. The arpege2 scheme (thick dark gray lines in Fig. 1 to be compared with the thick black line) significantly improves the frequency of occurrence but a systematic underestimation of about 0.13–0.18 appears up to 8 km of altitude. This scheme seems to produce the best overall estimate of cloud occurrence for all sites. This result is somewhat surprising, since ARPEGE is the only model that does not treat clouds with prognostic equations.

The RACMO model (third column in Fig. 1 and fourth in Table 2) includes a similar cloud scheme as ECMWF. It is nevertheless clearly characterized by a much larger and systematic overestimation of cloud occurrence for midlevel and low-level clouds, but a better high-level cloud occurrence than ECMWF. This result highlights the importance of accurately representing the large-scale forcing for an accurate representation of the clouds by the cloud scheme.

The Met Office model (fourth column in Fig. 1 and Table 2) produces the best low-level cloud occurrence of all models (fractional difference between 0.053 and 0.15). It tends, however, to produce overestimates of midlevel cloud occurrence similar to those observed in RACMO, and to produce the largest overestimations of all models for the high-level cloud occurrence (between 0.82 and 2.21).

It is seen that all models significantly overestimate the occurrence of thick clouds at high levels. Among the models, the Met Office, arpege1, and ECMWF produce the largest overestimations. The arpege2 cloud scheme seems to best match the observations overall, while the ECMWF and Met Office schemes are the most accurate to reproduce the frequency of cloud occurrence of midlevel and low-level clouds, respectively. It must also be noted that the occurrence computed from the model may be underestimated. Indeed this parameter is derived from the model cloud fraction that (except for the Met Office) is only computed from the nonprecipitating hydrometeors categories. Therefore the overestimation of high clouds may have been minimized in the present case and may be even larger.

b. Comparisons at seasonal scale

Cloudnet dataset is split into seasons with the objectives to document the seasonal (from one season to another during a year) and season-to-season (from one season a given year to the same season another year) variability of cloud occurrence in western Europe and to evaluate if the models are able to reproduce it. This is (especially at midlatitudes) an indirect way of investigating if the model cloud schemes produce the right differences in cloud properties in response to a different large-scale forcing.

When binning the dataset according to seasons, the intermittent sampling of instruments has to be carefully taken into account. During Cloudnet instrumentation at both the Chilbolton and Palaiseau sites suffered from technical difficulties. Apart from autumn 2003, where only 30 days of operations for this season exist, data from Cabauw are by far the most continuous of the three sites. Moreover, because the Cabauw radar did not experience significant power loss during the project (which has unfortunately been the case for the two 95-GHz radars), only the Cabauw dataset is used to investigate the season-to-season and seasonal variability.

1) SEASON-TO-SEASON VARIABILITY

Figure 2 shows the variations of cloud occurrence for each season (two seasons by panel) derived from the observations. The third panel shows the differences in occurrence between one season and the same season the next year. It appears that except for spring (comparison of solid and dashed gray lines in middle panel or dark thick line in right panel), where the curves are similar for the two years, strong differences occurred for a given season between the two years.

The large season-to-season variability observed in Fig. 2 also indicates that one should not extrapolate to other seasons of other years the properties of cloud occurrence derived from the two years of the Cloudnet project. This result is rather different from the same kind

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

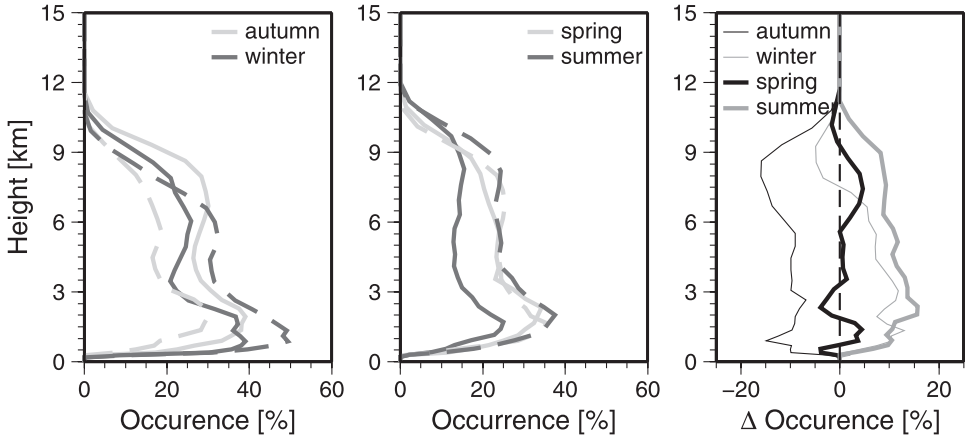

FIG. 2. Season-to-season comparison of frequency of cloud occurrence at the Cabauw site from the observations, for the first year (from autumn 2002 to summer 2003, solid lines) and second year (from autumn 2003 up to summer 2004, dashed lines) of the project: (left) autumn–winter, (center) spring–summer, and (right) difference (% $\text{season}_{\text{year2}}$ – % $\text{season}_{\text{year1}}$) in each season cloud frequency of occurrence.

of analysis conducted by Mace and Benson (2008) at the ARM Southern Great Plains site but using eight years of continuous observations. They found that each season has certain distinct characteristics and radiative influence. With these two years of observations, we can, however, evaluate if the models are able to reproduce the observed variability. For this purpose, the cloud occurrence has been calculated for each model and each season (not shown). It is found that all models are able to reproduce the fairly small variability of cloud occurrence during spring between the two years of the project. The observed large variability in summer (more clouds during the second year), in low- and midlevel clouds during winter and in low-level clouds during autumn are also well captured by the models. However, the models are not as successful to reproduce the decrease trend in high level cloud occurrence for the second year during the cold season.

In general, it appears that the models seem to reproduce reasonably well the season-to-season cloud variability observed during Cloudnet. This indicates that, despite their differences, the data assimilation schemes are reproducing the seasonal evolution of frontal activity and large-scale humidity structures well, and that the moist physics parameterizations are able to translate this into appropriate moist convective activity and associate convective or stratiform cloud properties.

2) SEASONAL CYCLE

Figure 3 shows the same curves as Fig. 2 but this time, sorted per year of the project in order to document the seasonal cycle of cloud occurrence. Figure 3 (left panel) shows the cloud occurrence seasonal cycle mean differences for the three cloud categories with respect to the spring season ($\Delta\text{occurrence} = \text{occurrence}_{\text{season}} - \text{occurrence}_{\text{spring}}$) as bar plot. The spring profile has been used as a reference here only because it is nearly identical for the two years of the projects (Fig. 2) for observations and for models. The large seasonal differences in cloud occurrence for the two years of the project appear immediately on these plots (comparison of top and bottom line). For the first year of the project (first line of the left panel), the main feature in the observations is that the summer season is characterized by a lower occurrence than for the other seasons. This is consistent with the blocking high pressure system that remained over Europe during much of the summer leading to record high temperatures over many regions of western Europe (Levinson and Waple 2004). The second year (bottom line of the left panel) shows a completely different behavior with nearly the same cloud occurrence for summer and spring, the smallest occurrences observed in autumn for all clouds, and the largest occurrences of low-level and midlevel clouds observed in winter.

Because of the change in the ARPEGE parameterization within the first year of the project, the seasonal cycle cannot be investigated for arpege1. The seasonal cycle of the first year is rather well reproduced by all models for low and midlevel clouds. In particular, the lower frequencies of occurrence observed throughout the troposphere for summer 2003 is present in all models. In fact, the models would have to be very poor to be unable to reproduce the large-scale blocking high pressure system in their analyses, and maintain this through the 36-h forecast range. It would be a much more stringent test of the models in general to see if they are able to maintain the ability to forecast the cloud variability in the medium (5 to 10 days) and extended range forecasts.

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

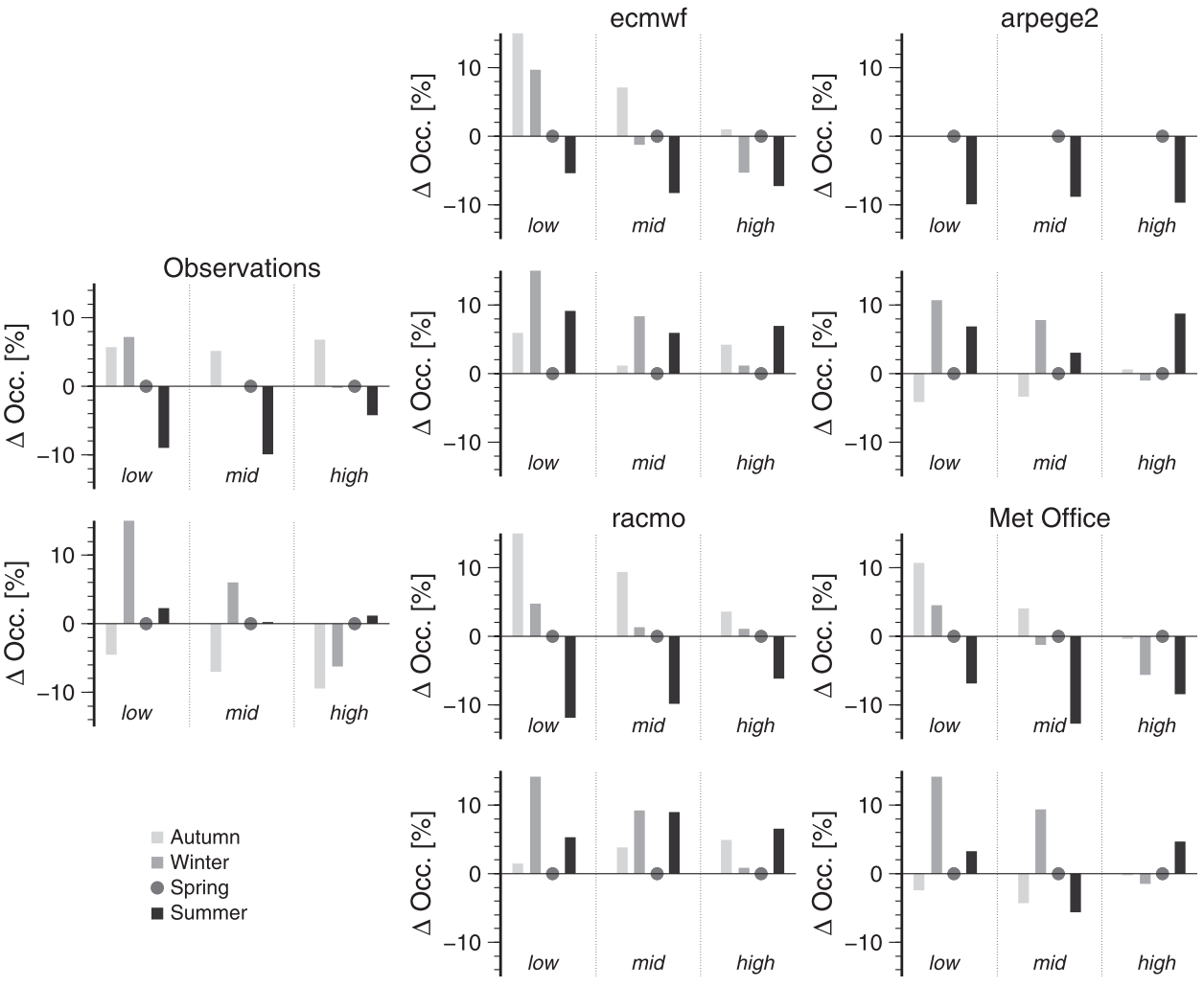

FIG. 3. Mean differences in seasonal variability of frequency of cloud occurrence with respect to spring (symbolized by the gray dot, Δoccurrence = occurrence$_{season}$ − occurrence$_{spring}$) for observations of (left) all the models, (center) ECMWF and RACMO, (right) arpege2 and Met Office, and the three cloud categories (low-, mid-, and high-level clouds, from left to right, separated by the thin lines) at the Cabauw site. For each dataset, the top part displays the seasonal cycle of the first year of project, the bottom part the seasonal cycle for the second year (beginning at autumn).

More interesting is that, despite their ability to forecast the reduction in cloud with the high pressure system, the ECMWF and RACMO models tend to generate too many low-level clouds (as much as in spring) and these results remain true the second year. This points to model problems in reproducing the low-level temperature inversions often associated with these systems, and indicates that the turbulence and/or shallow convection scheme suffer from shortcomings. For the second year these models are not able to reproduce the smaller cloud occurrence in autumn; however, the larger occurrence of low-level and midlevel clouds for winter 2004 is captured in ARPEGE and Met Office. In general, it is encouraging to see that the main seasonal signatures detected on the observations (maximum of occurrence of low-level clouds during winter and maximum of occurrence of high-level clouds during summer) are well captured by the models.

4. Comparison of model cloud fraction with observations

Cloud fraction is an important parameter since it is a crucial input to the radiation schemes. It is now a prognostic variable of the cloud schemes held in the ECMWF and RACMO models (see Table 1). In this section mean cloud fraction profiles and probability distribution functions (pdfs) of cloud fraction as a function of height are derived. The rationale for using pdfs in addition to the mean profiles considered in earlier studies (Mace et al.

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

1998; Hogan et al. 2001) is to investigate how the model is distributing the cloud fraction between 0 and 1, since a mean value of 0.5 can be the result of very different cloud fraction distributions. As in section 3, the statistics of cloud fraction are derived using the two years of Cloudnet observations and are first documented and compared with the four models. The dataset is then split into seasons in order to characterize the seasonal variability and evaluate if the models are able to reproduce it.

a. Comparison for the whole Cloudnet period

The mean profiles and pdfs of cloud fraction for models and observations using the whole Cloudnet period are displayed in the two first columns of Fig. 4 for the Cabauw site. Each row corresponds to a different model, with the first column showing the distributions obtained using the model subsample where models and observations agree on a cloud occurrence (a grid box being defined cloudy when the cloud fraction is larger than 0.03). The same distribution has been computed using the whole model dataset (not shown) resulting in similar pdfs except for the very high levels (between 11 and 13.5 km of height) where some clouds of rather small cloud fraction exist. The occurrence of such clouds is observed in the model profiles in Fig. 1 (solid gray line). However, when the criterion of occurrence agreement between model and observation is taken into account, these high altitude clouds are excluded from the statistics. The second column in Fig. 4 corresponds to the distribution obtained from observations (remapped at each model resolution). The solid black lines in the two first columns in Fig. 4 superimposed on each panel are the mean profiles of cloud fraction. This should be distinguished from the mean cloud fraction including occasions when no clouds were present shown in Fig. 6a of Illingworth et al. (2007), which results in much smaller values because it includes the cloud-free profiles. As clearly seen in Fig. 4, the mean profile does not correspond to the maximum of the pdf, because it is the average of a distribution, which has two maxima at the largest and smallest cloud fractions. If the mean cloud fraction profile is used and not the distribution in the cloud fraction evaluation, this could lead to questionable results, as it is quite possible for a model to have the correct mean cloud fraction value but resulting from the average of a completely wrong distribution. To illustrate this point, the third column compares the mean distribution of cloud fraction for each cloud category. The same distributions obtained for the two other sites (Chilbolton and Palaiseau) are not shown as they do not exhibit significant differences with the Cabauw data. Model skills are also found to be similar over the three sites.

The observed fraction of low-level clouds is characterized by a bimodal distribution, with about the same amount of clouds with low cloud fraction (less than 0.3) and high cloud fraction (larger than 0.85). Midlevel clouds are predominantly characterized by a monomodal distribution, peaking at large cloud fraction values (larger than 0.8). In contrast the high-altitude clouds are characterized by a bimodal distribution (very small and very large cloud fraction values).

The model cloud fraction pdfs can be directly compared to those observed (first two columns of Fig. 4 or gray and black lines in third column). The ECMWF and RACMO models share similar skills and discrepancies with respect to the observations. The core of small cloud fractions observed for low-level clouds extends too much upward (up to 8 km instead of 3 km in the observations) indicating that the ECMWF and RACMO models produce too many clouds with low cloud fraction at midlevels. This signature is also reflected by the smaller amount of clouds with large cloud fractions between 3 and 7 km (30% as compared with more than 40% observed, roughly) in these models. The ECMWF and RACMO models also clearly have difficulties in producing high-level clouds with small cloud fractions (high-level clouds are essentially characterized by very large cloud fractions in Fig. 4), which is not in agreement with the observations that clearly exhibit a bimodal distribution.

The Met Office model is characterized by a different cloud fraction distribution, with many more small cloud fractions overall (between 0 and 0.4), and a much more homogeneous distribution of cloud fractions between 0 and 1 than in the observations. The most striking feature is the lack of mid- and low-level clouds with a large cloud fraction.

The impact of the change in cloud scheme in the ARPEGE model is obvious. It is in particular clear from Fig. 4 (second and third line) that the cloud scheme in the first version of ARPEGE was unable to generate large cloud fractions at any height. The second scheme is clearly much better, with a good representation of the bimodal distribution of low-level clouds, and of the large cloud fractions in midlevel clouds. However, it is less accurate in describing the distribution of cloud fraction of high-level clouds, with too many high-level clouds characterized by a large cloud fraction, although there is some indication of a bimodal distribution, suggesting that further improvements or tuning of this scheme should be possible. The impact of the change in the ARPEGE cloud scheme on cloud fraction was also investigated by Illingworth et al. (2007) who noted that errors in cloud occurrence and cloud fraction in the previous scheme (arpege1) compensated to give a reasonably unbiased

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

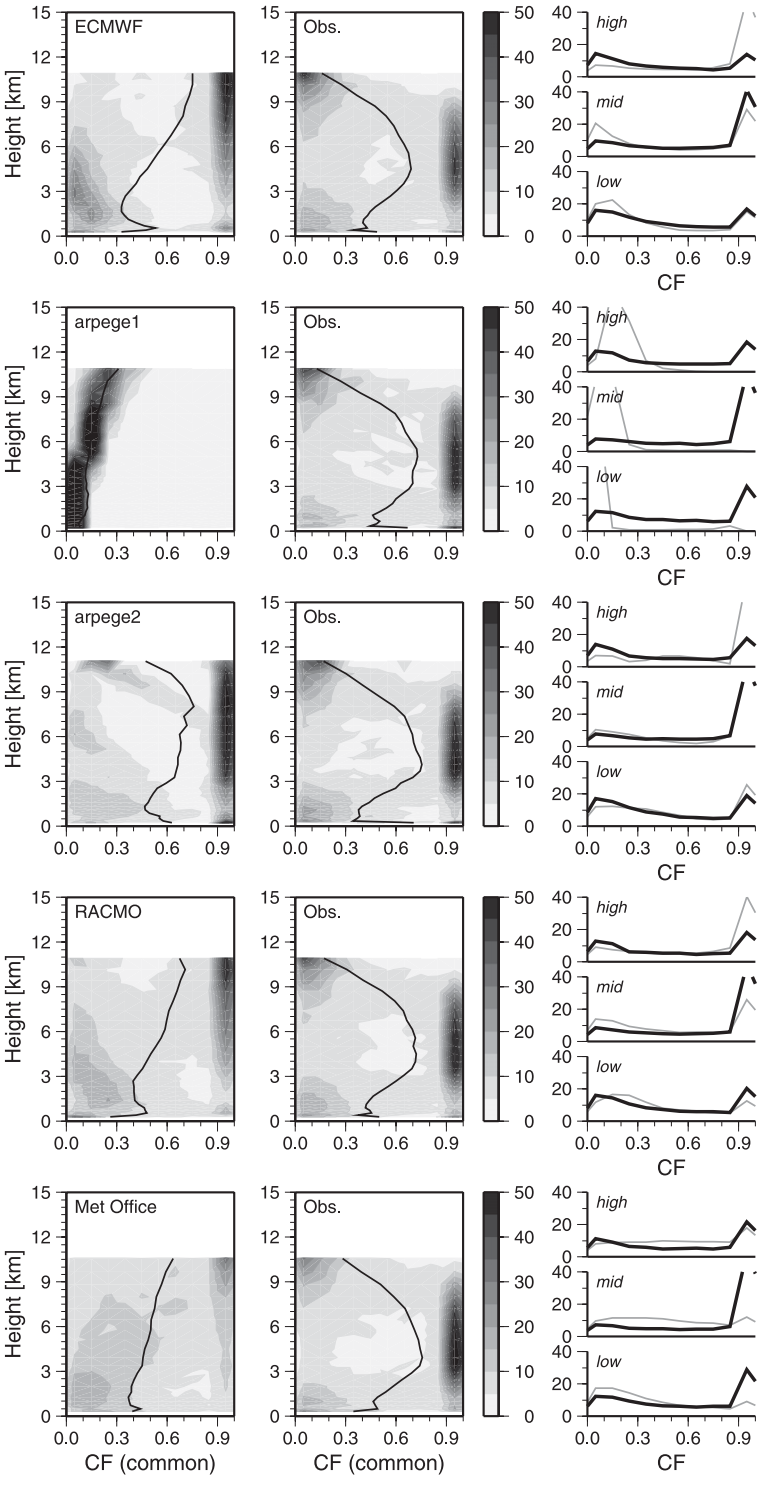

FIG. 4. Pdfs of cloud fractions when a cloud is present [in both (left) model and (center) observation] as a function of height in kilometers at the Cabauw site for the two-year period with the contour shading changing for every 5%. The black solid line shows the mean value of cloud fraction when observation and model agree on cloud occurrence. Each row corresponds to a model (or model version) (top to bottom) ECMWF, arpege1, arpege2, RACMO, and Met Office. (right) Mean pdfs of cloud fraction for model (gray line) and observation (black line) for the three cloud categories (high-, mid-, low-level from top to bottom).

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

mean cloud cover. The same result is found in the present study if one goes back to Fig. 1 observing that arpege1 produces far too many clouds but with too small a cloud fraction (especially at midlevels, as shown in Fig. 4). This combined effect of two inaccurate representations of cloud fraction and cloud occurrence can result in unbiased total cloud coverage. This point clearly highlights the value of comparing both the occurrence and fraction, not only the total cloud cover.

b. Comparison at seasonal scale

As detailed in section 3 the Cabauw site provides the most continuous sampling, so in the following the seasonal cycle and the season-to-season variability of cloud fraction are documented only at this site.

1) SEASON-TO-SEASON VARIABILITY

Figure 5 (two first rows) shows the evolution of the distribution of cloud fraction for the observations as a function of the season and demonstrates that the season-to-season variability is fairly weak for all seasons. To facilitate the comparisons, the mean pdfs have been computed for each cloud category and season (third row, first year in gray, second year in black). For all seasons the general structure of the pdfs is similar to that already discussed for the two-year dataset (Fig. 4): a bimodal distribution for low-level clouds (small and large cloud fractions, but with an amplitude of each peak varying with the season), a monomodal distribution for the midlevel clouds (large cloud fraction values are predominant), and a bimodal distribution for the high-level clouds (very small and very large cloud fraction values). The largest variability (which is still relatively small) occurs in autumn for all cloud categories.

The same mean seasonal pdfs by cloud categories for all models are shown for the second year in the bottom row of Fig. 5 (with the observed mean distribution superimposed in dashed red line) similar results were found for the first year (not shown). As in observations, the same features as for the two-year dataset is found for all models: large cloud fraction for high-level clouds, bimodal distributions for midlevel and low-level clouds. However, the seasonal variability of these structures is in contrast larger and requires further analysis.

2) SEASONAL CYCLE

The purpose of this section is to study how the distribution changes from one season to another as a response to different large-scale forcings, and if the models reproduce these changes. The larger seasonal variability is observed for mid- and low-level clouds in both models and observations [comparison from one column to another of gray (black) lines for observations in the middle panel of Fig. 5 and gray and black lines for models in the bottom panel of Fig. 5]. A very small variability is found when comparing the spring season of one year to the following in both models and observations, this season is used as a reference to allow comparisons in relative amplitude changes [$\Delta \mathrm{pdf}_{\mathrm{season}}(\mathrm{CF}) = \mathrm{pdf}_{\mathrm{season(CF)}} - \mathrm{pdfspring}(\mathrm{CF})$]. Figure 6 shows these differences in mean distributions for the four models for the second year of the project. The observations are displayed in dashed red.

At low levels during winter (middle column of Fig. 6), all models tend to fairly well reproduce the observed larger amount of large cloud fraction and the smaller amount of small cloud fractions. However, the models fail to reproduce the observed similar distribution between spring and autumn. The main feature that models represent well for midlevel clouds is the smaller amount of large cloud fraction cloud for summer (except Met Office). For the other seasons, the agreement between models and observations is poorer (with even an anticorrelation for autumn). For high-level clouds, all models are unable to reproduce the main observed seasonal features.

Overall, the problems detected using the two-year dataset are still found at seasonal scale (in particular for high-level clouds). The ability of the models to reproduce the changes in the distributions for low- and mid-level clouds appears better for winter and summer. This clearly implies that the first-order focus for model improvement should be put on the improvement of the model representation of the statistical properties of clouds before worrying too much about more subtle responses of the cloud schemes to different large-scale forcings over different seasons.

5. Comparison of ice water content between models and observations

The second variable generally held in NWP model prognostic ice cloud schemes (IWC) is now evaluated in this section. The same methodology as that employed for cloud fraction is used (distributions and mean profiles).

There is no remote sensing instrumentation able to provide a direct measurement of the IWC profiles. The profiling instruments (radar and lidar) measure radar reflectivity, Doppler velocity, and lidar backscatter coefficient. Numerous methods with different degrees of complexity exist and have been developed in the Cloudnet project to derive IWC either from radar reflectivity only (Liu and Illingworth 2000; Protat et al. 2007) or using an additional constraint such as air temperature (Liu and Illingworth 2000; Hogan et al. 2006;

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

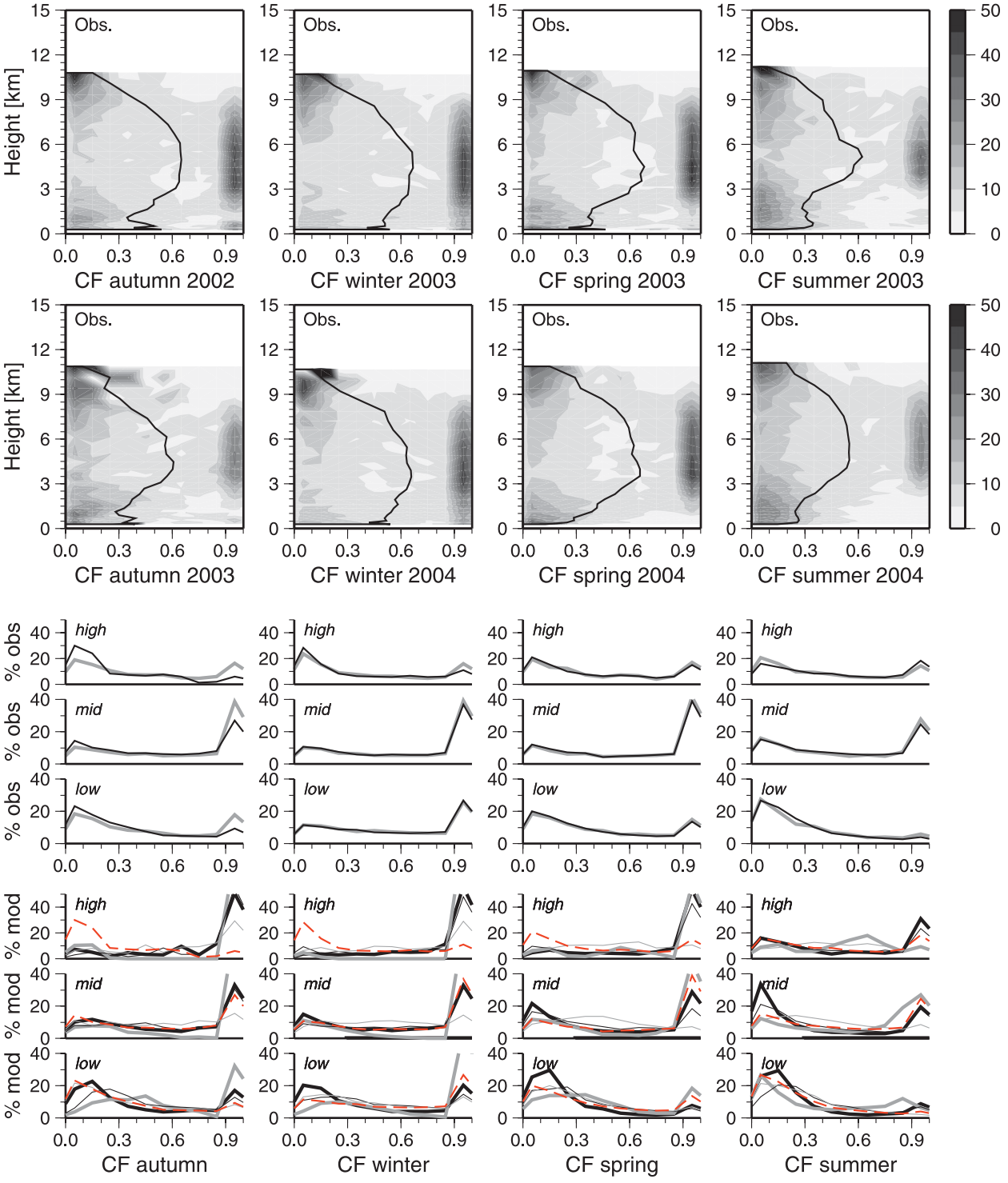

FIG. 5. Observed pdfs of cloud fraction for each season of the Cloudnet Project using the ECMWF model grid at the Cabauw site from autumn to summer for (top row) the first year and (second row) the second year. The black solid line shows the mean value of the distribution. (third–fifth rows) Comparison between first (gray solid line) and second year (black solid line) of the mean cloud fraction pdfs (top to bottom) for low-, mid-, and high-level clouds. (sixth–eight rows) The mean distribution for the three cloud categories for models (ECMWF: thick black; arpege2: thick gray; RACMO: thin black; Met Office: thin gray) for each season of the second year of the project; the superimposed dashed red line is the mean pdf from observations.

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

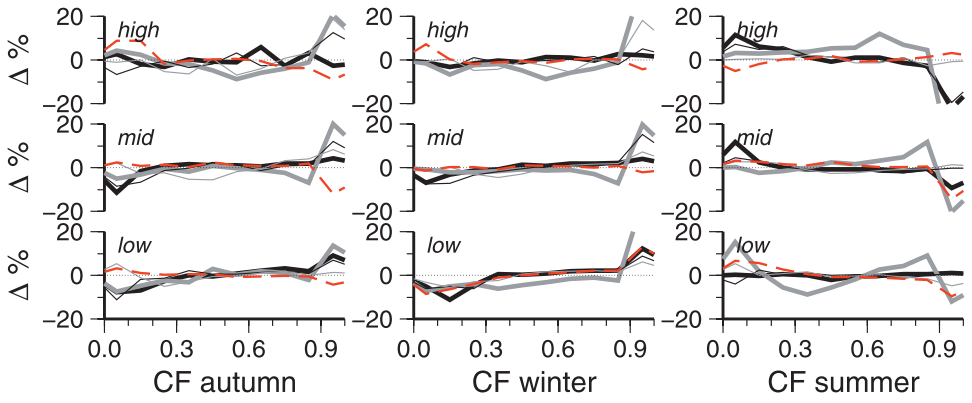

FIG. 6. Differences in mean cloud fraction distribution with respect to spring 2004 [$\Delta pdf_{season}(CF) = pdf_{season}(CF) - pdf_{spring}(CF)$] for models (ECMWF in thick black, arpege2 in thick gray, RACMO in thin black, Met Office in thin gray lines, and observations in dashed red line) for (left to right) autumn 2003, winter 2004, summer 2004 for (top to bottom) high-, mid-, and low-level clouds.

Protat et al. 2007), Doppler radar velocity (Matrosov et al. 2002; Mace et al. 2002; Delanoë et al. 2007), or lidar backscatter (Donovan and van Lammeren 2001; Wang and Sassen 2002; Tinel et al. 2005). In this paper we adopt the approach that consists of evaluating the model cloud variables using the cloud variables retrieved from the radar–lidar observations. Another increasingly popular and complementary approach consists in computing pseudoradar and lidar observations from model outputs (using so-called radar and lidar simulators; e.g., Haynes et al. 2007; Chepfer et al. 2008; Bodas-Salcedo et al. 2008). The main reason for which the retrieval strategy is adopted in this study is because we evaluate operational models, and generally such simulators would have to be run online since they need input parameters which are not recorded operationally at a sufficiently high temporal resolution. Besides, in the case of retrieving IWC from measurements, similar types of hypotheses are included in these simulators (particle size distributions for instance, particle habits, etc.). As a result, the errors on the retrieved reflectivities are probably of the same order of magnitude as the error on the retrieved IWC.

Heymsfield et al. (2008) intercompared a half-dozen radar and radar–lidar retrieval methods using a test dataset. Several of these methods have been applied in the Cloudnet project. In this paper it is shown that the radar/lidar retrieval technique potentially provides the most accurate method of retrieving the IWC and with the lowest standard deviation of the error relative to radar-only techniques. However, as discussed in Illingworth et al. (2007), the radar–lidar technique can only be applied to a small fraction of all ice clouds, owing to the lack of sensitivity of the cloud radars to thin cirrus clouds and the rapid extinction of the lidar signal for clouds of optical thickness larger than 2–3. For the radar-only techniques, the Doppler velocity–reflectivity retrieval technique also did very well in this intercomparison, but a significant high bias was found for the IWC–Z–T techniques. Several retrieved IWC were available in the Cloudnet project either from radar–lidar techniques or from radar-only techniques. The different retrieval methods produce very similar pdfs (not shown): a skewed distribution with wider part of the distribution for small IWC. A quantitative comparison between radar–lidar technique and radar-only technique is difficult because the retrieval methods cannot be applied to the exactly the same dataset, radar–lidar techniques were suitable for small optical depth clouds. For given dataset Heymsfield et al. (2008) show that radar–lidar methods present a mean ratio to observations of 1.03 ± 00.66 and that for radar-only techniques is 1.00 ± 0.60. In view of these results, the Doppler velocity–reflectivity method named RadOn (Delanoë et al. 2007) has been used in the present paper, as it does represent a good tradeoff between accuracy of the IWC retrieval and the amount of clouds to which this method can be applied. Delanoë et al. (2007) estimates the bias on retrieved IWC at 0.4% with a standard deviation lower than 18%. In the following, comparisons are performed using model grid-box averaged observed IWC, the model values for not completely filled grid box being unchanged.

a. Comparison for the whole Cloudnet period

Figure 7 (two first columns) displays the pdfs of IWC obtained from the model time series and from the observations (remapped at the model resolutions) at the Cabauw site, when model and observations agree on a cloud occurrence. For each distribution the mean IWC profiles have also been computed (solid lines). These mean profiles are repeated and superimposed in the

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

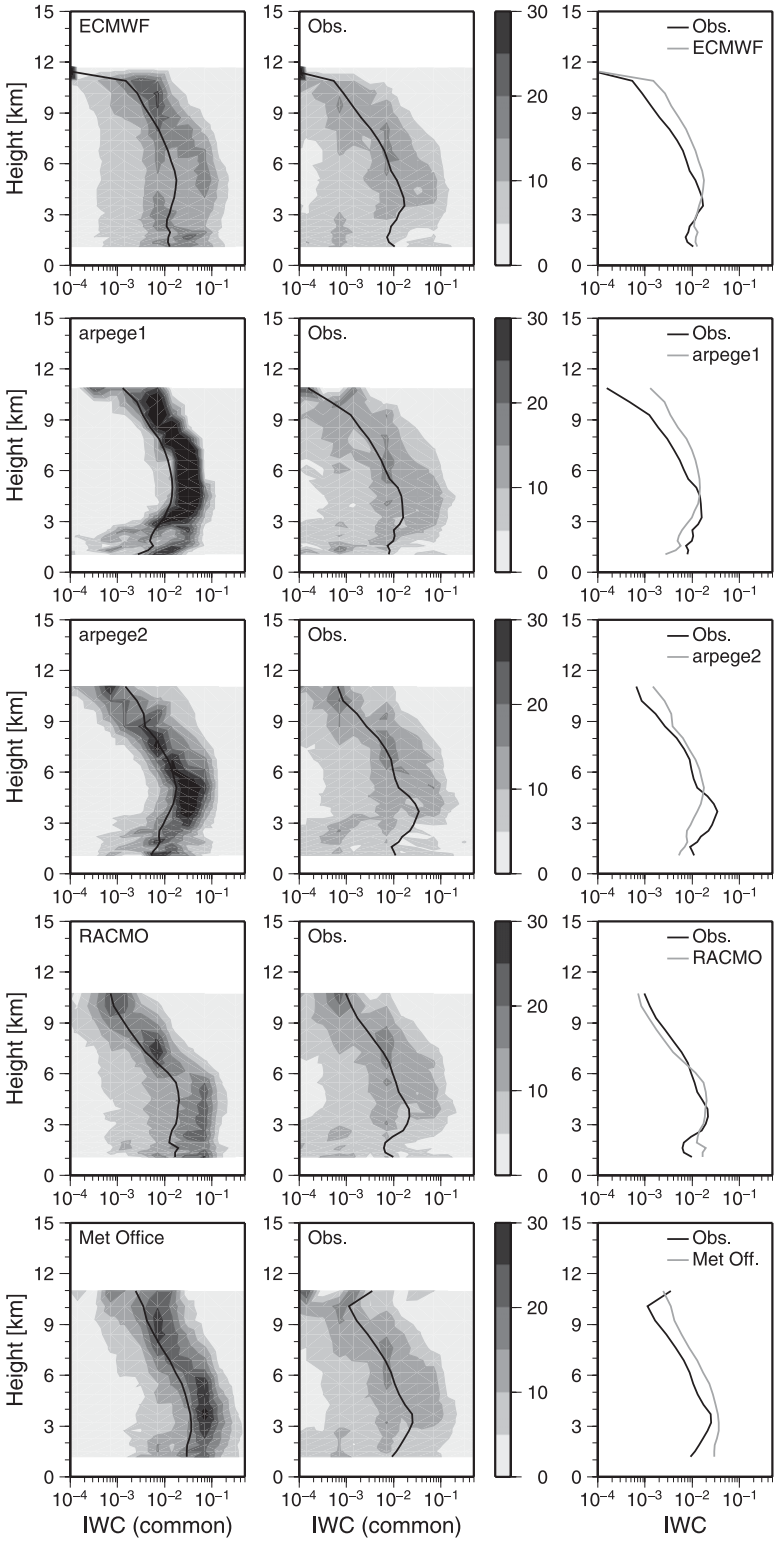

FIG. 7. Pdfs of IWC (g m^{-3}) for the two-year Cloudnet period at (left) the Cabauw site for models [one model or model version per line (top to bottom): ECMWF, arpege1, arpege2, RACMO and Met Office]; (center) observations. The black line shows the mean value. (right) The two previous mean value profiles for model (gray) and observations (black).

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

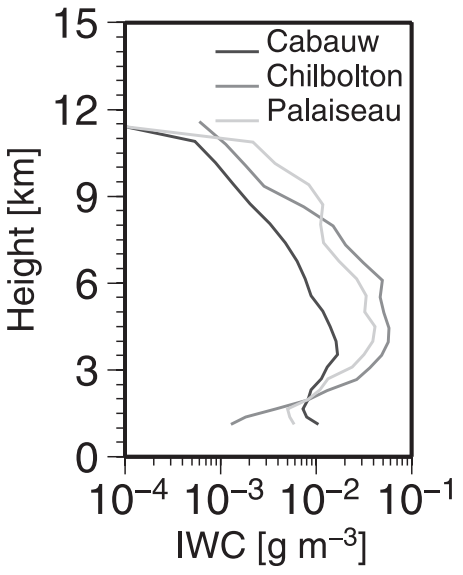

FIG. 8. Comparison of mean IWC profiles (g m^{-3}) (computed on the ECMWF grid) obtained at the three sites for the whole Cloudnet period.

third column in order to facilitate direct comparisons. The observed IWC distributions are skewed toward small IWCs, and with a peak IWC of about 10^{-2} g m^{-3} at 3.5–4.0 km of height and a decrease above and below. The comparison of the observed and model IWC distributions of the two first columns of Fig. 7 clearly shows that the models reproduce this skewed distribution rather well. The IWC distribution of the models is nevertheless generally much narrower than the distribution obtained from the observations. The ECMWF and RACMO models seem to best reproduce the observed width of the IWC distributions at all heights, but above 7 km of altitude the RACMO IWC distribution is skewed toward small IWC. This result explains the difference between the dotted and solid line in Fig. 1 for this model. Indeed, in Fig. 1 the solid dotted line does not include the cloud occurrence when the reflectivity values computed from the IWC were below the detection threshold of the radar. This event often occurs for the small IWC observed at these levels in this model. The Met Office model produces a way-too-narrow distribution, significantly skewed toward the largest IWCs at all heights. Bodas-Salcedo et al. (2008) found similar results when comparing this same model with *CloudSat* near-global observations. They explain this narrower distribution by the fact that nature presents a larger variability in particular for the particle size distribution leading to a larger range of IWC values. This variability in particle size distribution is not represented in the model. The comparison of the IWC distributions produced by the first and second parameterization for ARPEGE shows that IWCs are significantly smaller (about one order of magnitude) for clouds higher than 6 km when the second parameterization is used. This effect is not obvious from the mean values but can be observed on the contours where the distribution is centered on larger IWC. This feature is not in agreement with the observations, so it is solely due to the change in parameterization.

The impact of radar sensitivity on the results must be addressed, because all the retrieval algorithms use the radar reflectivity as an input. The 35-GHz radar at the Cabauw site is more sensitive (−55 dBZ at 1 km) than the other two 95-GHz radars at Palaiseau and Chilbolton (−50 and −45 dBZ at 1 km, respectively) and the 95-GHz cloud radars experienced a gradual sensitivity loss during the project (Hogan et al. 2003). Therefore, some lower reflectivity values are included at the Cabauw site corresponding to very small IWC, which would not be detected at the two other sites. This is confirmed in Fig. 8 where the mean profiles derived from the observations at the three sites are compared: the Chilbolton and Palaiseau IWC profiles are similar whereas the Cabauw IWC profile has smaller IWCs. This demonstrates that the IWCs missed by cloud radars from −50 to −45 dBZ sensitivity at 1 km do play a significant role in the mean IWC profile. Therefore in order to include the largest range of IWCs in the comparisons, we have chosen to use only the Cabauw observations to evaluate the model IWCs.

The comparisons between the model and observation are plotted in the third column of Fig. 7 and they clearly show that the shape of the mean IWC profiles is rather well reproduced by all models. However, observed differences appear larger than uncertainty on the retrieved IWC. The ECMWF and arpege1 cloud schemes both tend to overestimate IWC above 4-km height, and slightly underestimate below. In the case of ECMWF, this overestimation is the result of a too-strong production of large IWCs, while in the case of arpege1 it is merely the result of a pdf that is way too narrow (not including enough low IWC values). The second ARPEGE cloud scheme reproduces the observed IWC profile more faithfully than the first cloud scheme above 4-km height, but strongly underestimates IWC below 4-km height. The Met Office cloud scheme produces a systematic overestimation of IWC at all heights, but captures the shape of the profile very well. Finally, the RACMO model reproduces almost perfectly the observed IWC profile, as a result of a relatively good representation of the observed IWC distribution.

b. Comparison at seasonal scale

The seasonal cycle (evolution from one season to another during a year) and season-to-season (from one

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

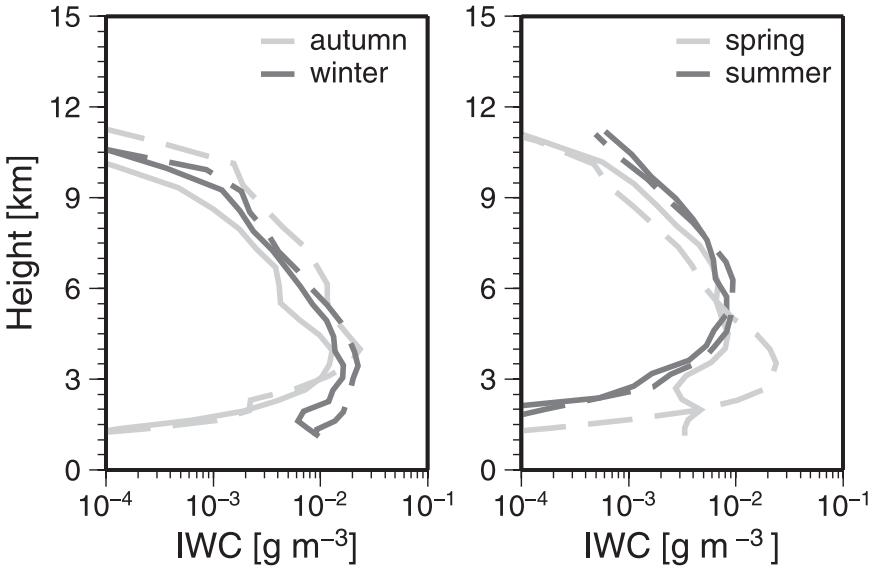

FIG. 9. Seasonal evolution of IWC at the Cabauw site: (left) autumn–winter and (right) spring–summer Solid lines correspond to the first year of the project (from autumn 2002 up to summer 2003); dashed line are for the second year (from autumn 2003 up to summer 2004).

season a given year to the same season another year) variabilities are investigated in this section for IWC using the Cabauw observations. A clear result is that the seasonal variability of the IWC distributions in both the observations and the model remain fairly similar (not shown). Below, we therefore concentrate on the comparison of mean IWC profiles with that produced in models.

1) SEASON-TO-SEASON VARIABILITY

Figure 9 shows the mean profiles of observed IWC for autumn and winter on the left panel and spring and summer on the right panel. There is some variability between the two years of the project (comparison of dashed and solid lines for a given season) in autumn and spring, while it is fairly small in winter and summer. To characterize this variability and evaluate the ability of the model to reproduce it (and also because the logarithmic display of IWC may be misleading), a normalized difference is computed for low-, mid-, and high-level clouds. The normalized difference is computed for a given cloud category as $\Delta_{\text{norm}}\text{IWC} = (\text{IWC}_{\text{season year2}} - \text{IWC}_{\text{season year1}})/\text{IWC}_{\text{season year1}}$.

Table 3 shows these normalized differences for each cloud category. As observed in Fig. 9, the smaller variability is observed for winter and summer. One may also observe that the second year of the project exhibits rather larger IWCs. However, even if some models tend to capture this increase (e.g., Met Office), the magnitudes do not really match. This finding can be generalized to all models and cloud categories: even if the models get the right trend, they fail to reproduce the magnitude.

2) SEASONAL CYCLE

A significant seasonal variability may be expected, simply owing to the fact that ice is present in the troposphere at higher levels during the hottest seasons and extends to lower altitudes during the coldest season. To highlight the seasonal cycle, the summer seasons of both years have been chosen as a reference and the normalized differences [$\Delta_{\text{norm}}\text{IWC} = (\text{IWC}_{\text{season}} - \text{IWC}_{\text{summer}})/\text{IWC}_{\text{summer}}$] are computed for each season and for each

TABLE 3. Normalized mean differences of IWC [$\Delta_{\text{norm}}\text{IWC} = (\text{IWC}_{\text{season year2}} - \text{IWC}_{\text{season year1}})/\text{IWC}_{\text{season year1}}$] for models and observations between first and second year of the project each time for low-/mid-/high-level clouds at the Cabauw site.

	Obs	ECMWF	ARPEGE	RACMO	Met Office
Δ_{norm}(autumn)	0.05/1.01/1.92	1.61/1.21/1.47	—	0.43/0.75/−0.15	0.43/0.03/−0.03
Δ_{norm}(winter)	−0.09/0.22/−0.19	−0.04/−0.12/0.13	—	2.68/0.22/−0.67	−0.35/0.04/0.001
Δ_{norm}(spring)	0.56/1.25/−0.35	−0.20/0.25/0.09	−0.52/0.07/0.29	0.003/−0.18/-0.19	0.14/−0.13/−0.25
Δ_{norm}(summer)	0.72/0.20/0.05	−0.02/−0.07/0.20	0.08/0.53/0.37	0.22/0.32/1.03	−0.35/0.02/0.47

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

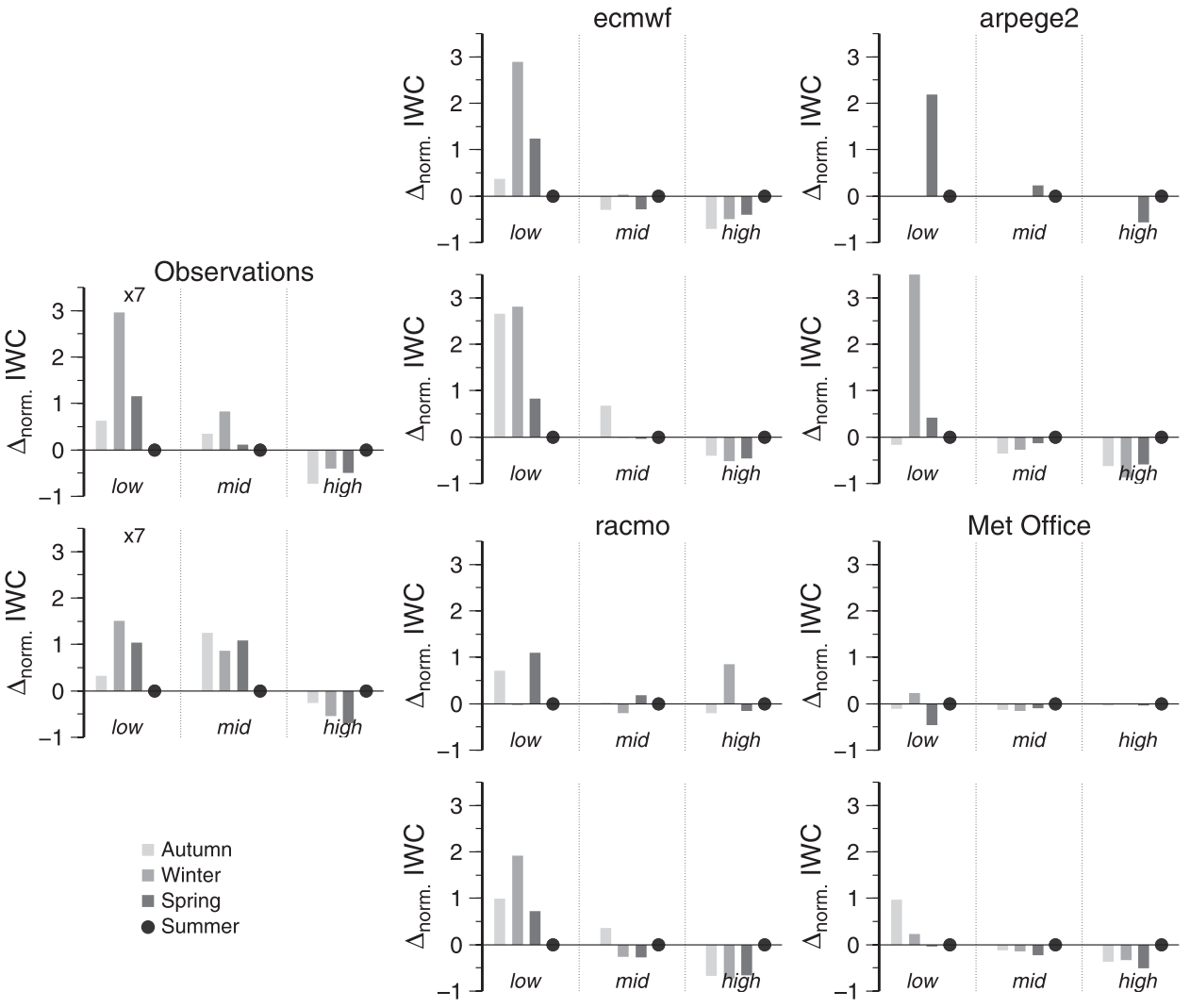

FIG. 10. Seasonal cycle of normalized differences of mean IWC with respect to spring [symbolized by the black dot, $\Delta_{norm}IWC = (IWC_{season} - IWC_{summer})/IWC_{summer}$] for (left) observations and (center) ECMWF and RACMO, (right) arpege2 and Met Office at the Cabauw site for (left to right in each dataset) the low-, mid-, and high-level clouds. For each dataset, the top part displayed the seasons of the first year, the bottom the seasons for the second year (beginning at autumn). Note that the bars for observed low-level clouds are shown at one-seventh of their true height.

cloud category. The choice of summer as a reference comes from the previous subsection where the smaller season-to-season variability is observed for this season. The results are displayed in Fig. 10 with the first column showing the results from the observations. The observed minimum in IWC for low-level clouds during summer season appears clearly in this figure. Below 3 km of height, the most prominent feature is a clear IWC maximum in winter (with a normalized difference reaching 20 and 10 for winter 2003 and 2004, respectively, out of the range in Fig. 10), as expected, and a gradually decreasing mean IWC from the warmest to the coldest season (with spring and autumn presenting rather similar values). In contrast, small changes from one season to another appear for midlevel clouds with larger IWC relative to summer. Last, high-level clouds are characterized by a systematic reduction of IWC with respect to summer.

The same mean normalized differences have been built for the two years of all models and are displayed in Fig. 10 (second and third columns with two models by column). The reference season is still summer, although this season does not exhibit a reduced variability between the two years for the models. It may be observed that ECMWF reproduces rather well the observed seasonal cycle of IWC for low-level clouds (in particular the peak value for winter), while the Met Office tends to

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

produce the same profile for all seasons (with very few variations observed for all cloud categories). The weak changes in mean IWC for midlevel clouds are well reproduced by all models (even if it is with the wrong sign) as well as the reduction for high-level clouds with respect to summer.

6. Conclusions

This paper describes an evaluation of the representation of clouds in four numerical weather prediction models (ECMWF, ARPEGE, RACMO, Met Office) by making use of long time series of cloud parameters collected by radar and lidar at three different sites in western Europe. The philosophy used consists of analyzing the two-year dataset of observations and models, to produce profiles of cloud occurrence, pdfs of cloud fraction, and IWC, and to compare them with those deduced from models. Clouds have then been subdivided into three categories: high-, mid-, and low-level clouds. The dataset has also been further analyzed on a seasonal basis.

Particular attention has been given to minimize the effects of potential biases linked to the instrumental combination (95- or 35-GHz radar in combination with a lidar or ceilometer) and especially the difference in instrument sensitivity. These instrumental effects have been taken into account, showing that either a radar combined with a lidar or a radar with a higher sensitivity combined with a lidar ceilometer provide similar results, with an unbiased cloud frequency of occurrence sampling up to 9 km of altitude.

For the high-level clouds all the models tend to overestimate the cloud occurrence and all schemes except the first version of the ARPEGE model fail to produce the small cloud fraction values observed at this height. All models but ARPEGE indeed exhibit mainly large cloud fraction values, while a bimodal distribution is found in the observations. The IWC is generally overestimated, except by RACMO. So the picture is as follows: there are too many high-level clouds in the models, and with too large cloud fraction and IWC. These clouds are considered radiatively important because of their feedback on weather and climate and therefore their relatively inaccurate representation in models may be significant when computing fluxes with the radiation scheme.

The occurrence of midlevel is found to be largely overestimated by all models but the second version of the ARPEGE scheme. The models produce midlevel clouds with low cloud fraction too frequently, while a monomodal distribution peaking at large cloud fraction is found in the observations. In contrast, the IWC of midlevel clouds is generally well reproduced by the models.

Low-level cloud occurrence is generally overestimated by the models (except for arpege2) with a different magnitude from one model to another. They also appear to be characterized by too small a cloud fraction as compared with observations, owing to an underrepresentation of large cloud fraction values of low-level clouds. The models show reasonable skills for IWC, because they seem to produce IWC within the same order of magnitude as the observations. The exception is the Met Office model, which produces a systematic and large overestimation of IWC.

A general comment that applies to all cloud categories is that the distribution of IWC is too narrow in the models. This highlights the inability of the models to reproduce the observed variability of this parameter whatever the degree of sophistication of the model cloud scheme. Bodas-Salcedo et al. (2008) found similar results in the Met Office model. They suggest that two-moment schemes (using IWC and ice particle concentration) should be used to mitigate this problem. However, such a solution may be very time consuming for NWP models. From this overall comparison it is also interesting to see that the arpege2 model, although it is still using a diagnostic cloud scheme and a coarser resolution than Met Office, produces overall the ''best'' representation of clouds and questions the real value of the fully prognostic schemes.

The seasonal cycle study aims at investigating if the models are able to reproduce the differences in the observations; that is, how well the parameterizations reproduce differences in statistical cloud properties in response to different large-scale forcings encountered all along the year at midlatitudes. Regarding the cloud frequency of occurrence, it is found in this paper that the seasonal cycle is generally well captured by the models (less accurate results in autumn, though) at all levels. In contrast, the cloud fraction variations are generally not well reproduced, in particular at high levels. For IWC the seasonal variability found in the models do not fully agree with that observed either, especially for midlevel clouds. However, the variations are reasonably well captured for ECMWF, arpege2, and RACMO.

The same exercise would be interesting to repeat but using the spaceborne *CloudSat* radar and CALIPSO lidar measurements available since June 2006 as part of the A-Train mission. In this case the three ground-based stations are overflown using exactly the same instrumentation. The geographical differences in cloud occurrence can then be investigated and possibly related to synoptic conditions and/or cloud regimes. This spaceborne cloud radar/lidar tandem is also of particular interest for the evaluation of occurrence of high-level clouds, since the lidar would not be attenuated by the

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

liquid clouds below, and the cloud radar would be of similar sensitivity as the ground-based radars.

Last, it has been shown that the continuous operation of cloud profiling stations do have the potential to provide an efficient way of rapidly evaluating the impact of any change in the parameterization schemes of operational weather forecast models. For example, it was possible to demonstrate that the new cloud scheme in the ARPEGE model and its subsequent tuning was having a positive impact in the representation of clouds. The A-Train satellite now provides cloud profiles on a global scale. These will also be very valuable to investigate the cloud properties in regions where ground-based observations are generally sparse but where clouds have a large influence on the climate system such as in the tropical belt (including Africa and South America). However, because of the sun-synchronous orbit, the temporal and diurnal sampling at these latitudes is limited. Therefore, adapted strategy of the present methodology is under investigation in order to evaluate the cloud parameterizations of global models (such as ECMWF or ARPEGE model) in these very different meteorological conditions.

Acknowledgments. This research received funding from the European Union Cloudnet project (Grant EVK2-CT-2000-00065) and NERC Grant NER/Z/S/2003/00643. We acknowledged the Experimental Site for Atmospheric Research (CESAR), the Chilbolton Facility for Atmospheric and Radio Research, part of Rutherford Appleton Laboratory, and the SIRTA Observatory of the Institut Pierre-Simon Laplace and the support by their staff for provision of the original radar and lidar datasets used in this study.

REFERENCES

Anderson, E., and Coauthors, 2005: Assimilation and modeling of the atmospheric hydrological cycle in the ECMWF forecasting system. *Bull. Amer. Meteor. Soc.,* **86,** 387–402.

Bodas-Salcedo, A., M. J. Webb, M. E. Brooks, M. A. Ringer, K. D. Williams, S. F. Milton, and D. R. Wilson, 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.

Brooks, M. E., R. J. Hogan, and A. J. Illingworth, 2005: Parameterizing the difference in cloud fraction defined by area and by volume as observed with radar and lidar. *J. Atmos. Sci.,* **62,** 2248–2260.

Brown, P. R. A., A. J. Illingworth, A. J. Heymsfield, G. M. McFarqhar, K. A. Browning, and M. Gosset, 1995: The role of spaceborne millimeter-wave radar in the global monitoring of ice cloud. *J. Appl. Meteor.,* **34,** 2346–2366.

Chepfer, H., S. Bony, D. Winker, M. Chiriaco, J.-L. Dufresne, and G. Seze, 2008: Use of CALIPSO lidar observations to evaluate the cloudiness simulated by a climate model. *Geophys. Res. Lett.,* **35,** L15704, doi:10.1029/2008GL034207.

Delanoë, J., A. Protat, D. Bouniol, A. Heymsfield, A. Bansemer, and P. Brown, 2007: The characterization of ice clouds properties from Doppler radar measurements. *J. Appl. Meteor. Climatol.,* **46,** 1682–1698.

Donovan, D. P., and C. A. P. van Lammeren, 2001: Cloud effective particle size and water content profile retrievals using combined lidar and radar observations. Part 1: Theory and examples. *J. Geophys. Res.,* **106** (D21), 27 425–27 448.

Guichard, F., D. B. Parsons, J. Dudhia, and J. Bresch, 2003: Evaluating mesoscale model predictions of clouds and radiations with SGP ARM data over a seasonal timescale. *Mon. Wea. Rev.,* **131,** 926–944.

Haynes, J. M., R. T. Marchand, Z. Luo, A. Bodas-Salcedo, and G. L. Stephens, 2007: A multipurpose radar simulation package: QuickBeam. *Bull. Amer. Meteor. Soc.,* **88,** 1723–1727.

Heymsfield, A. J., and Coauthors, 2008: Testing IWC retrieval methods using radar and ancillary measurements with in situ data. *J. Appl. Meteor. Climatol.,* **47,** 135–163.

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

——, D. Bouniol, D. N. Ladd, E. J. O'Connor, and A. J. Illingworth, 2003: Absolute calibration of 94/95 GHz radar using rain. *J. Atmos. Oceanic Technol.,* **20,** 572–580.

——, M. P. Mittermaier, and A. J. Illingworth, 2006: The retrieval of ice water content from reflectivity factor and temperature and its use in evaluating a mesoscale model. *J. Appl. Meteor. Climatol.,* **45,** 301–317.

Illingworth, A. J., and Coauthors, 2007: Cloudnet—Continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bull. Amer. Meteor. Soc.,* **88,** 883–898.

Jakob, C., 1999: Cloud cover in the ECMWF reanalysis. *J. Climate,* **12,** 947–959.

——, 2003: An improved strategy for the evaluation of cloud parameterizations in GCMs. *Bull. Amer. Meteor. Soc.,* **84,** 1387–1401.

Lenderink, G., B. van den Hurk, E. van Meijgaard, A. P. van Ulden, and J. W. M. Cuijpers, 2003: Simulation of present-day climate in RACMO2: First results and model developments. Tech. Rep., 252, KNMI, De Bilt, Netherlands.

Levinson, D. H., and A. M. Waple, 2004: State of the climate in 2003. *Bull. Amer. Meteor. Soc.,* **85,** S1–S72.

Liu, C.-L., and A. J. Illingworth, 2000: Toward more accurate retrievals of ice water content from radar measurements of clouds. *J. Appl. Meteor.,* **39,** 1130–1146.

Liu, C., E. J. Zipser, G. Mace, and S. Benson, 2008: Implications of the differences between daytime and nighttime *CloudSat* observations over the tropics. *J. Geophys. Res.,* **113,** D00A04, doi:10.1029/2008JD009783.

Mace, G. G., and S. Benson, 2008: The vertical structure of cloud occurrence and radiative forcing at the SGP ARM site as revealed by 8 years of continuous data. *J. Climate,* **21,** 2591–2610.

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

——, A. Heymsfield, and M. Poellot, 2002: On retrieving the microphysical properties of cirrus clouds using the moments of the millimeter-wavelength Doppler spectrum. *J. Geophys. Res.,* **107,** 4815, doi:10.1029/2001JD001308.

Mathieu, A., J.-M. Piriou, M. Haeffelin, P. Drobinsky, F. Vinit, and F. Bouyssel, 2006: Identification of error sources in planetary boundary layer cloud forecast using SIRTA observations. *Geophys. Res. Lett.,* **33,** L19812, doi:10.1029/2006GL026001.

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020

Matrosov, S. Y., A. V. Korolev, and A. J. Heymsfield, 2002: Profiling cloud ice mass and particle characteristic size from Doppler radar measurements. *J. Atmos. Oceanic Technol.,* **19,** 1003–1018.

Morcrette, J.-J., 1991: Radiation and cloud radiative properties in the ECMWF operational forecast model. *J. Geophys. Res.,* **96,** 9121–9132.

——, 2002: Assessment of the ECMWF model cloudiness and surface radiation fields at the ARM SGP site. *Mon. Wea. Rev.,* **130,** 257–277.

Protat, A., A. Armstrong, M. Haeffelin, Y. Morille, J. Pelon, J. Delanoë, and D. Bouniol, 2006: Impact of conditional sampling and instrumental limitations on the statistics of cloud properties derived from cloud radar and lidar at SIRTA. *Geophys. Res. Lett.,* **33,** L11805, doi:10.1029/2005GL025340.

——, J. Delanoë, D. Bouniol, A. J. Heymsfield, A. Bansemer, and P. Brown, 2007: Evaluation of ice water content retrievals from cloud radar reflectivity and temperature using a large airborne in-situ microphysical database. *J. Appl. Meteor. Climatol.,* **46,** 557–572.

——, ——, A. Plana-Fattori, P. T. May, and E. O'Connor, 2010: The statistical properties of tropical ice clouds generated by the West African and Australian monsoons, from ground-based radar–lidar observations. *Quart. J. Roy. Meteor. Soc.,* **136,** 345–363.

Rossow, W. B., and R. A. Schiffer, 1983: The International Satellite Cloud Climatology Project (ISCCP): The first project of the World Climate Research Program. *Bull. Amer. Meteor. Soc.,* **64,** 779–784.

Siebesma, A. P., and Coauthors, 2004: Cloud representation in general-circulation models over the northern Pacific Ocean: A EUROCS intercomparison study. *Quart. J. Roy. Meteor. Soc.,* **130,** 3245–3268.

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

Stephens, G. L., and Coauthors, 2002: The *CloudSat* mission and the A-Train. *Bull. Amer. Meteor. Soc.,* **83,** 1771–1790.

Stokes, G. M., and S. E. Schwartz, 1994: The Atmospheric Radiation Measurement (ARM) Program: Programmatic background and design of the Cloud and Radiation Test Bed. *Bull. Amer. Meteor. Soc.,* **75,** 1201–1221.

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

Tinel, C., J. Testud, J. Pelon, R. J. Hogan, A. Protat, A. Delanoë, and D. Bouniol, 2005: The retrieval of ice cloud properties from cloud radar and lidar synergy. *J. Appl. Meteor.,* **44,** 860–875.

Wang, Z., and K. Sassen, 2002: Cirrus cloud microphysical property retrieval using lidar and radar measurements. Part I: Algorithm description and comparison with in situ data. *J. Appl. Meteor.,* **41,** 218–229.

Wilson, D. R., and S. P. Ballard, 1999: A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Quart. J. Roy. Meteor. Soc.,* **125,** 1607–1636.

Winker, D. M., W. H. Hunt, and M. J. McGill, 2007: Initial performance assessment of CALIOP. *Geophys. Res. Lett.,* **34,** LI9803, doi:10.1029/2007GL030135.

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

Yang, S. K., Y.-T. Hou, A. J. Miller, and K. A. Campana, 1999: Evaluation of the Earth Radiation Budget in NCEP–NCAR reanalysis with ERBE. *J. Climate,* **12,** 477–493.

Downloaded from http://journals.ametsoc.org/jamc/article-pdf/49/9/1971/3557234/2010jamc2333_1.pdf by guest on 20 November 2020