TECHNICAL MEMORANDUM

526

Evaluating cloud occurrence in the ECMWF Integrated Forecast System and three other operational forecast models using long-term ground-based radar and lidar measurements

D. Bouniol[1], A. Protat[2], J. Delanoë[3], J. Pelon[4], D.P. Donovan[5], J.-M. Piriou[1], F. Bouyssel[1], A.M. Tompkins[6], D.R. Wilson[7], Y. Morille[8], M. Haeffelin[9], E.J. O'Connor[3], R.J. Hogan[3] and A.J. Illingworth[3]

Research Department

[1] Centre National de la Recherche Météorologique. Météo-France, Toulouse, France
[2] Centre d'étude des Environnements Terrestre et Planétaires, CNRS, Vélizy, France
[3] Meteorology Department, University of Reading, Reading, UK
[4] Service d'aéronomie, CNRS, Paris, France
[5] Royal Netherlands Meteorological Institute, De Bilt, The Netherlands
[6] European Centre for Medium-range Weather Forecasts, Reading, UK
[7] Met Office, Exeter, UK
[8] Laboratoire de Météorologie Dynamique, CNRS, Palaiseau, France
[9] Institut Pierre-Simon Laplace, Palaiseau, France

December 2007

This paper has not been published and should be regarded as an Internal Report from ECMWF.
Permission to quote from it should be obtained from the ECMWF.

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

©Copyright 2008

European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, England

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.

Abstract

This paper assesses the ability of four operational weather forecast models (ECMWF, ARPEGE, RACMO and Met Office) to generate a cloud at the right location and time (*i.e.* the cloud occurrence) using a two year time series of observations collected by profiling ground-based active sensors (cloud radar and lidar) located at three different sites in western Europe (Cabauw, NL, Chilbolton, UK, Palaiseau, F.). Particular attention is given to potential biases that may arise from instrumentation differences (especially sensitivity) from one site to another and intermittent sampling.

It is shown that all models tend to overestimate the occurrence of high-level clouds, while for lower levels the errors are contrasting. The data set is then divided into seasons to evaluate the potential of the models to generate different cloud situations. Strong variations in cloud occurrence are found in the observations from one season to the same season the following year as well as between the seasons of a given year. Overall, the model biases observed using the whole data set are still found at seasonal scale, but the models generally manage to capture the seasonal variation in cloud occurrence well. This study demonstrates the usefulness of long time series in order to monitor and guide changes to a given parametrisation.

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 twenty 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 short wave/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 the parametrisation of clouds has 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.

Comparing clouds simulated by models with observations can be achieved using different approaches. A first approach is to build up regional or global climatologies of the cloud properties. One of the first that quantified the vertical cloud distribution was compiled by London (1957) using visual observations from the surface. 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, potentially allowing a global climatology of cloud properties to be constructed (*e.g.* Rossow and Schiffer 1991). The drawback is that information concerning detailed cloud vertical structure is usually lacking (Illingworth *et al.* 2007).

More recently the emergence of long time series of clouds properties, including their vertical structure in different synoptic situations, in particular through the Atmospheric Radiation Measurement (ARM) program (Stokes and Schwartz 1994), has opened a new perspective for model/observations comparisons. The principle is to implement a complete set of instruments (including cloud radar and lidar) in order to be able to document the cloud properties as fully as possible for a set of given operational sites. The use of long time series allows the cloud generation processes in models to be tested in a wide range of situations. This approach 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 a given parameter, such as cloud fraction (Hogan *et al.* 2001). Jakob (2003), in a Global Circulation Model (GCM) performance evaluation suggested (following Norris and Weaver 2001 and Tsedioudis and Jakob 2002) that the comparison should be sorted by dynamical regimes that can be defined by the model data.

Following the example of the ARM sites a network of cloud observing stations, all equipped with active sensors such as lidars and Dopplerised mm-wave radars, has been developed in Europe in the framework of the Cloudnet project. The objectives of this project and the cloud products derived at these sites are fully described in Illingworth *et al.* (2007). The three main objectives of Cloudnet are :

(i) to collect remotely sensed cloud data quasi-continuously from a network of three cloud remote sensing stations (Cabauw (NL), Chilbolton (UK) and Palaiseau (F)) and simultaneously gather hourly vertical profiles from four major operational models (ECMWF, ARPEGE, RACMO, and Met Office model) over the stations during two years (starting 1^{st} October 2002, ending 30 September 2004).

(ii) to develop and validate state-of-the art techniques for deriving the microphysical cloud properties from cloud radar, lidar and microwave radiometer observations.

(iii) to develop regional climatologies from observations to evaluate the representation of clouds in operational weather forecast models, with a focus on the representation of the fraction of the grid box that is filled with clouds (the so-called cloud fraction) and the ice water content.

The aim of the present two-part 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. After examining the mean model errors for the entire period, the model capability 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 rather more affected by convective systems. Such a seasonal study at midlatitude should therefore shed some light on the capability of the models to reproduce these very different types of cloud-producing situations.

It should be noted that the instrumentation implemented at the three Cloudnet sites is not strictly the same. All sites operate at least a radar (94 GHz for Chilbolton and Palaiseau, 35 GHz for Cabauw) and a ceilometer (355 nm for Cabauw and Chilbolton, 855 nm for Palaiseau) or a lidar (532 nm for Palaiseau). A full description of the instrumental set-up of the three sites is provided in Illingworth *et al.* (2007). Some differences in the regional characteristics can be expected from these differences in instrumentation, and especially the differences in instrument sensitivity. As a result we have chosen not to investigate in the present paper the regional variability of cloud properties and the capability of the models to reproduce it.

In the following section the cloud scheme of the four models (ECMWF, ARPEGE, RACMO and Met Office) involved in this evaluation are briefly described. This description is necessary in order to clearly specify which version of the models is evaluated in the present papers. Indeed, since operational centres are continuously improving their models, it must be noted that the conclusions drawn from our analysis apply to a specific set of model versions. The proposed methodology can be applied to later versions though, as well as to other models and other sites, as long as the observations are available. In Part I of the study (the present paper), we describe the methodology and evaluate if the models are able to simulate a cloud when observed, using the frequency of cloud occurrence as the evaluation parameter. In a second part (Bouniol *et al.* 2007) the observed cloud variables involved in the model cloud parametrisations are then analysed and compared to their representation in the models, namely the cloud fraction and the ice water content (the liquid water content is not investigated).

2 Description of the models

The four models involved in this evaluation are Numerical Weather Prediction (NWP) models and the cloud variables analysed in the following are operational forecast products. ECMWF and ARPEGE use global NWP

Table 1: Summary of the model cloud schemes involved in the Cloudnet project

	ECMWF	arpege1	arpege2	RACMO	Met Office
	global	global	global	regional	regional
Horizontal resolution [km]	39	24	24	18	12
Number of level	60	41	41	40	38
Cloud fraction	prognostic	diagnostic from water vapour excess	diagnostic from IWC and LWC	prognostic	diagnostic
LWC	prognostic from total cloud water	diagnostic from water vapour excess	diagnostic	prognostic	prognostic with water vapour
IWC	prognostic from total cloud water	diagnostic from water vapour excess	diagnostic	prognostic from total cloud water	prognostic

models, while RACMO and Met Office models are regional. The profiles are extracted 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 four times per day for the Met Office model. The models have different horizontal resolutions: 39, 24, 18 and 12 km, respectively for ECMWF, ARPEGE, RACMO and Met Office, and also a different vertical resolution: about 40 levels for ARPEGE, RACMO and Met Office and 60 levels for ECMWF, but all these levels stay in the range 0.1 to 1000 hPa.

In this section a brief overview of each model cloud scheme is given. It does not aim at giving a full description of each model cloud scheme but just at highlighting differences between models that can explain further differences observed in their cloud representation when compared to the observations. Table 1 provides a summary of the cloud scheme description performed in the following.

While the emphasis is placed on the differences in the cloud schemes between the models, it is obvious that clouds are the result of a complex array of interacting processes, and thus errors in cloud variables may result from deficiencies in other model components, such as the vertical diffusion or convection parametrisations for instance, or indeed in the representation of the interaction between these model components. Moreover, since the cloud variables are 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 UK 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 of relevance since relative humidity is the main predictor of cloud properties in both prognostic and diagnostic cloud schemes. It is thus necessary to guard against over-interpreting the results of this study directly in terms of the differences in the cloud parametrisations themselves.

2.1 ARPEGE model cloud scheme

The ARPEGE model is a particular case in this study, since it replaced its cloud parametrisation with a completely new scheme on 14 April 2003. In both schemes though, the cloud variables are diagnosed, there is no prognostic variable for clouds.

In the original diagnostic scheme, cloudiness has three sources: large scale supersaturation, sub-grid shallow convection and sub-grid deep convection supersaturation. For large scale and shallow convection the cloudiness and the condensed water contents are linked to the potential water vapour excess. For convection suspended water and cloudiness diagnostics are derived from the convective precipitation, which already accounts for all sub-grid information, including evaporation. The partition between solid and liquid phase is a function of temperature.

The main weaknesses of this scheme were to produce too much small cloud fraction and not enough ice water content. The cloud scheme used in the remaining time of the project, has diagnostic cloud variables (liquid/ice water content, rain/snow content, turbulent kinetic energy). A new formulation of cloudiness has been implemented, following Xu and Randall (1996) resulting in changes in the non-precipitating convection water vapour fluxes at the top of planetary boundary layer leading to a wetter boundary layer. Ice and liquid water content variables are used in order to compute cloudiness and both serve as inputs to the radiation scheme (short wave and long wave). In addition the cloud overlap assumption was switched from random overlap to maximum-random overlap.

Due to this complete change of philosophy during the Cloudnet project, the ARPEGE data set has been split in two, the first one labelled arpege1 corresponds to the diagnostic cloud scheme (up to 13 April 2003) and the second labelled arpege2 corresponding to the remaining time of the project.

2.2 ECMWF model cloud scheme

Cloud and large scale precipitation processes are described by prognostic equations for cloud liquid water/ice and cloud fraction, diagnostic relations for precipitation. The complete scheme is described in Tiedtke (1993). The partition between water and ice phase is made as a function of temperature, assuming that there is no ice supersaturation in the atmosphere.

The evolution of cloud water/ice and cloud fraction depends on transport of cloud water/ice and cloud area through the boundaries of the grid volume, formation of cloud water/ice and cloud area by convective processes, boundary layer turbulence and stratiform condensation processes, the rate of evaporation of cloud water/ice, the generation of precipitation from cloudwater, the dissipation of cloud/water by cloud top entrainment and the rate of decrease of cloud area due to evaporation.

The increase of cloud cover and cloud water in stratiform clouds (corresponding to non-convective processes such as large scale lifting of moist air and radiative cooling) is determined by how much of the cloud-free area exceeds saturation in one time step which in turn depends on the moisture distribution in the cloud-free area and how fast saturation is approached. New clouds are assumed to form when the grid average relative humidity exceeds a given threshold, assumed to be 80% throughout much of the atmosphere.

The ECMWF model profiles provided for this project derive from a series of eight model releases, or “cycles”, some of which included major changes to the model moist physics or data assimilation systems. The initial cycle in place at the project outset (1st October 2002) was 25R1, which remained operational until 14th January 2003. Compared to the ERA-40 reanalysis cycle 23R4, which was documented in Gregory *et al.*, (2000), this cycle includes numerous upgrades to the data assimilation scheme in addition to a revised short wave radiation scheme.

After 14th January 2003, cycle 25r3 included a major upgrade to the cloud and convection schemes. In particular, the changes to the cloud scheme physics and numerical solver methodology led to a model climate with increased cloud ice amounts in both the tropics and mid latitudes, while cloud liquid water was reduced.

Cycles 26r3 and 28r1 introduced in October 2003 and March 2004, respectively, both provided major upgrades to the analysis system and data usage. Cycle 28r3 made a major change to the numerics of the moist physics, leading to a substantially reduced global cloud liquid water amounts, however, this model version was introduced shortly before the end of the study period and thus will not affect the statistics reported here. Likewise, although a parallel treatment to the ARPEGE model should separate the ECMWF time series into the pre and post cycle 25r3 periods, the alterations to the cloud characteristics were more minor than that of ARPEGE, which introduced a completely new scheme, and thus the ECMWF data set is considered as a single unit.

2.3 KNMI Regional Atmospheric Climate MOdel (RACMO) cloud scheme

This model 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 ERA-40 reanalysis. A complete description of the RACMO cloud scheme is given in Lenderink *et al.* (2003). Therefore it is expected that differences in cloud representation between ECMWF and RACMO result mainly from differences in the dynamics, and data assimilation schemes, although the upgrades to the ECMWF system outlined above should also be taken into account.

2.4 Met Office model cloud scheme

The Met Office Unified Model used for these comparisons has one combined prognostic variable for water vapour and cloud liquid water and one prognostic variable for ice. Rain is diagnostic in the scheme. There are two sources of cloudiness : large scale and convective. At large scale, the split between water vapour and cloud liquid water is diagnosed using the Smith (1990) parametrisation scheme and ice phase and precipitation are described in Wilson and Ballard (1999).

At large scale, the Smith (1990) scheme assumes a triangular distribution of the total water content inside the grid, with a width determined by a critical relative humidity defined as the minimum grid-box mean relative humidity at which clouds will start to form. Condensation and cloud occur for the fraction of the vapour plus liquid content that exceeds saturation. No supersaturation with respect to water is allowed. The Wilson and Ballard (1999) scheme is designed for model processes that occur in frontal systems. This scheme has a prognostic variable representing all ice particles and considered transfer processes which act between vapour, liquid water, ice and rain. The ice cloud fraction is deduced from the inversion of the implicit relationship that exists in the Smith (1990) scheme between cloud fraction and condensed water content. This provides a diagnostic relationship for ice cloud fraction as a function of ice water content, specific saturation humidity and critical relative humidity.

At convective scale, condensed water is based on a mass-flux convective scheme. These quantities are then used within a diagnostic relationship in order to compute a convective cloud fraction.

3 Comparison of model cloud occurrence with observations

In this section the frequency of cloud occurrence in observations and in the four models is compared over the three Cloudnet sites. The frequency of cloud occurrence is defined as the ratio of cloudy hours to the total number of observational hours for a given level. A grid box is considered cloudy when its cloud fraction (in model and in observations, computed from temporal averaging of Cloudnet cloud profiles in the equivalent time/size model grid box following Mace *et al.* 1998 or Hogan *et al.* 2001) is larger than 0.03. Different threshold values (lower than 0.1) have been tested leading to the same statistics. This threshold is applied because most models (except Met Office) do not produce exact zero cloud fraction values. In previous studies such as in Hogan *et al.* (2001) this threshold was set to 0.05 and they also stated that their results were found to be fairly insensitive to the exact value of this threshold.

In the following, model frequency of cloud occurrence is compared to 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.

3.1 Comparison for the whole Cloudnet period

3.1.1 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. During Cloudnet the cloud radars and lidars were not all originally designed for unattended operational use. For instance, the 532 nm lidar system at SIRTA (Haeffelin *et al.* 2005) is not protected from precipitation damage and must be regularly checked by an operator. As a result, the temporal sampling of Palaiseau lidar observations is mostly limited to daytime and to periods with low risk of precipitation. In the same way, intermittent problems with the different cloud radars have occurred during Cloudnet, which resulted in a partial sampling. For instance the radar of the Chilbolton site began to operate continuously in April 2003 because of the failure of its tube (Hogan *et al.* 2003) and during the first year of the project the radar of the Palaiseau site was only operated during daytime four days a week. 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 due to their limited sensitivity. Such 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. In the present paper, this study is carried out for the three sites and the four models, in order to quantify and take good account of these biases individually for each model and at each site.

A comparison of the frequency of cloud occurrence for the whole Cloudnet period at each site and for each model is shown in Figure 1. The lower left panel (Palaiseau and ECWMF) of this figure corresponds to a mix of Figure 1(a) and 2(a) of Protat *et al.* (2006). The solid grey line corresponds to the whole model sample (the two years of operations with a one hour resolution), the dashed line is the model sub-sample corresponding to the radar-lidar hours of operations, and the dotted line is the model sub-sample corresponding to the radar and lidar hours of operations and accounting for the above-mentioned instrumental radar-lidar effects. In order to compute the radar sensitivity effect, instead of considering a mean value of sensitivity as in Protat *et al.* (2006), we have refined the analysis by using the decrease of sensitivity due to power loss of the 95 GHz radar tubes at Palaiseau and Chilbolton interpolated from regular measurements during the Cloudnet period. This radar sensitivity effect is then computed as in Protat *et al.* (2006), *i.e.* by converting the model ice water content to a reflectivity value using the Liu and Illingworth (2000) IWC-Z relationships. Liquid water clouds were identified using the lidar as they often fell below the sensitivity limit of the radar. Full details of the procedure whereby the radar lidar returns were corrected for any attenuation and categorised as cloud can be found in Illingworth *et al.* (2007). The model-derived reflectivities below the true cloud radar sensitivity at each site are then removed from the model sub-sample corresponding to the dotted lines in Figure 1. The final comparison between model and observations should thus be made between the dotted line (after removal of the instrumental

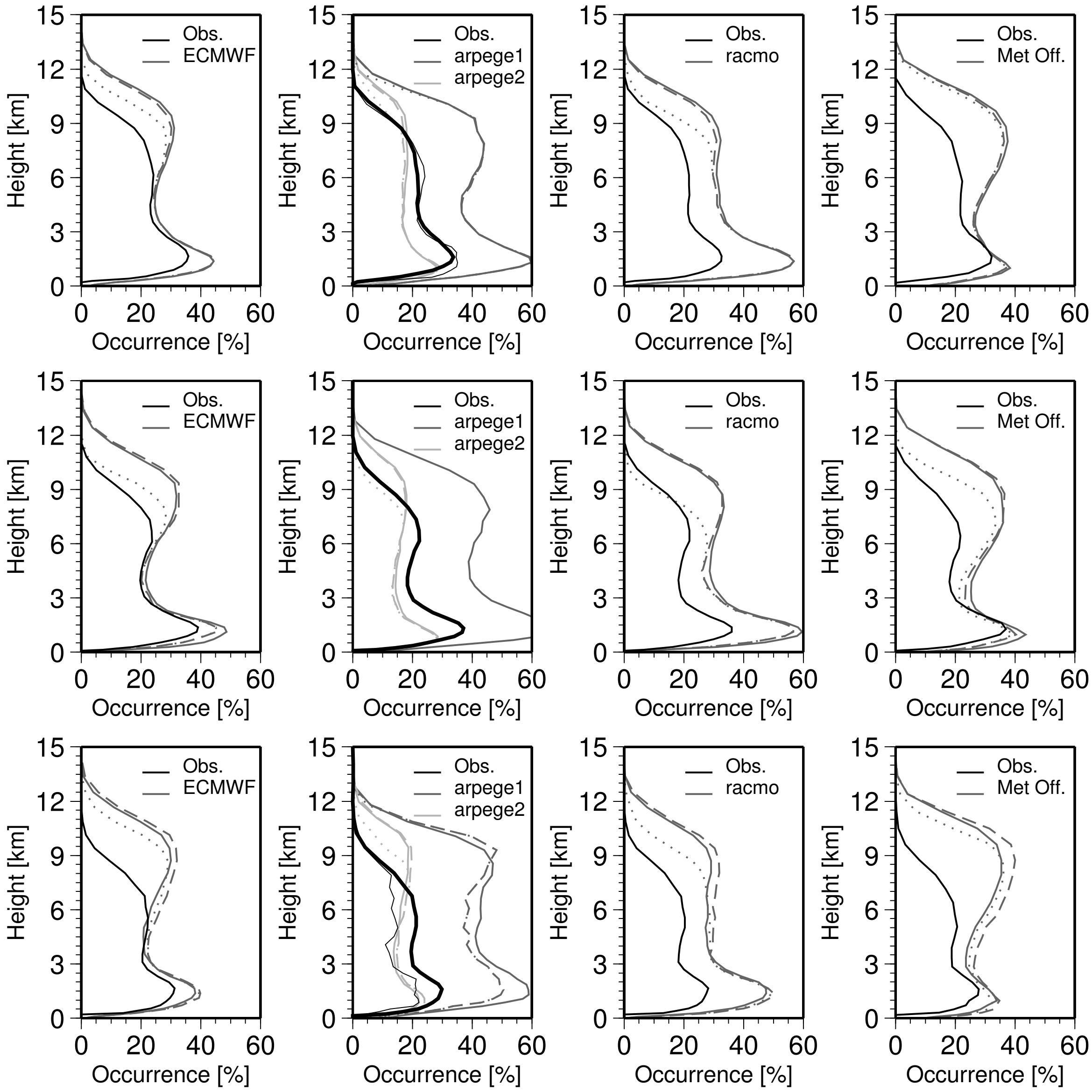

Figure 1: Frequency of cloud occurrence for models and observations obtained at the three sites for the whole Cloudnet period (01 October 2002 to 30 September 2004). Black line shows the frequency of cloud occurrence obtained from the observations. For ARPEGE, the thinnest line corresponds to arpege1 period, the boldest line to arpege2. The grey lines with different styles correspond to the model samples : solid line for whole model sample, dashed line for model sub-sample corresponding to instrument hours of operations and dotted line for model sub-sample corresponding to instrument hours of operations and including instrumental effects. Each line corresponds to an observatory from top to bottom: Cabauw, Chilbolton and Palaiseau. Each panel is dedicated to a comparison with a particular model : from left to right ECMWF, ARPEGE (1 and 2), RACMO and Met Office.

and sampling effects) and the black line (observations).

The comparison between the solid grey, dashed, and dotted lines in Figure 1 is interesting in several aspects. First it shows if the model sub-sample that includes sampling and instrumental effects is representative of the cloud properties derived from the total model sample. Even if it is anticipated that not all clouds are included in the comparisons, we can still compare model and observations. Protat *et al.* (2006) showed for the Palaiseau site (95 GHz radar, 532 nm lidar) and the ECMWF model that owing to the instrumental effects a radar-lidar combination could do a reasonable job providing unbiased frequency of cloud occurrence up to around 9 km altitude, which is confirmed by the lower left panel of Figure 1. Figure 1 shows that the same conclusion applies for the Cabauw site (35 GHz radar, ceilometer) as well as the other models except for the RACMO model, for which the account for the instrumental effect has a larger effect, probably owing to the fact that the RACMO model produces clouds significantly thinner (and thus not detectable with the cloud radar) than the others. Regarding the Chilbolton site, the 94 GHz radar-ceilometer combination provides unbiased frequency of cloud occurrence up to around 7.5 km only. The cloud radar performances are roughly similar to the performances as that of Palaiseau. A visual check on radar and ceilometer time series reveals that the ceilometer never detects an ice cloud that the radar does not detect, which was not the case with the high-output lidar at Palaiseau, which as discussed in Protat *et al.* (2006) significantly helps improve the frequency of cloud occurrence statistics above 8 km. Thus the compensating instrumental effect observed at the Palaiseau site is not occurring at Chilbolton, resulting in biases above 7.5 km. It is however occurring at the Cabauw site due to the increased sensitivity of the radar.

3.1.2 Comparison between models and observations

The frequency of occurrence obtained from the data (black lines in Figure 1) is now directly compared to the model frequency of occurrence including instrumental effects (grey dotted lines). It must be noted here that the model profiles corrected for instrumental effects using IWC-Z relationships (dotted line) suffer from two uncertainties: the accuracy of the IWC-Z relationship itself, and the fact that the model IWCs are used to translate the IWCs into a radar detection threshold. Hence the magnitude of the instrumental effect correction is driven by the model IWC.

The ECMWF model (first column in Figure 1) exhibits a very good agreement for mid-level clouds (between 3 and 7 km) over the three sites and an over-estimation of the high-level (above 7 km) and low-level cloud (below 3 km) occurrences. Although the overestimation is about the same over all sites for the low-level cloud occurrence (about 10%), the overestimation of the high-level cloud occurrence is much larger at Palaiseau (up to 20-25%) than at Chilbolton and Cabauw (5-10%).

The ARPEGE model (second column in Figure 1) exhibits a radically different behaviour between the two cloud schemes used during the Cloudnet period. The diagnostic cloud scheme labelled arpege1 produces a very strong and systematic overestimation. This scheme also produces large occurrences of high-level clouds which are classified as ”detectable” by the instruments, since there is no difference between the total profile and the profile with instrumental effects included. The modified ARPEGE diagnostic cloud scheme (labelled arpege2) significantly improves the frequency of occurrence but a systematic underestimation of about 5% appears up to 8 km altitude. It appears to produce the best overall estimate of cloud occurrence for all sites. This is somewhat surprising, since ARPEGE is the only model which does not treat clouds with prognostic equations.

The RACMO model (third column in Figure 1) includes a similar cloud scheme as ECMWF. It is nevertheless clearly characterised by a much larger and systematic over-estimation of cloud occurrence at all levels than ECMWF, except for high-level cloud occurrence (once the instrumental effects are included) for which it provides a better cloud occurrence than ECMWF overall.

The Met Office model (fourth column in Figure 1) produces the best cloud occurrences of all models for the low-level clouds, overestimations similar to those observed in RACMO of the mid-level cloud occurrence, and the largest overestimations among models for the high-level cloud occurrence.

In conclusion, it is seen that the models significantly overestimate the occurrence of thick cloud at high-levels. Among the models, the Met Office, arpege1, and ECMWF cloud schemes produce the largest overestimations of occurrence of such clouds. The arpege2 cloud scheme seems to best match the observations overall, while

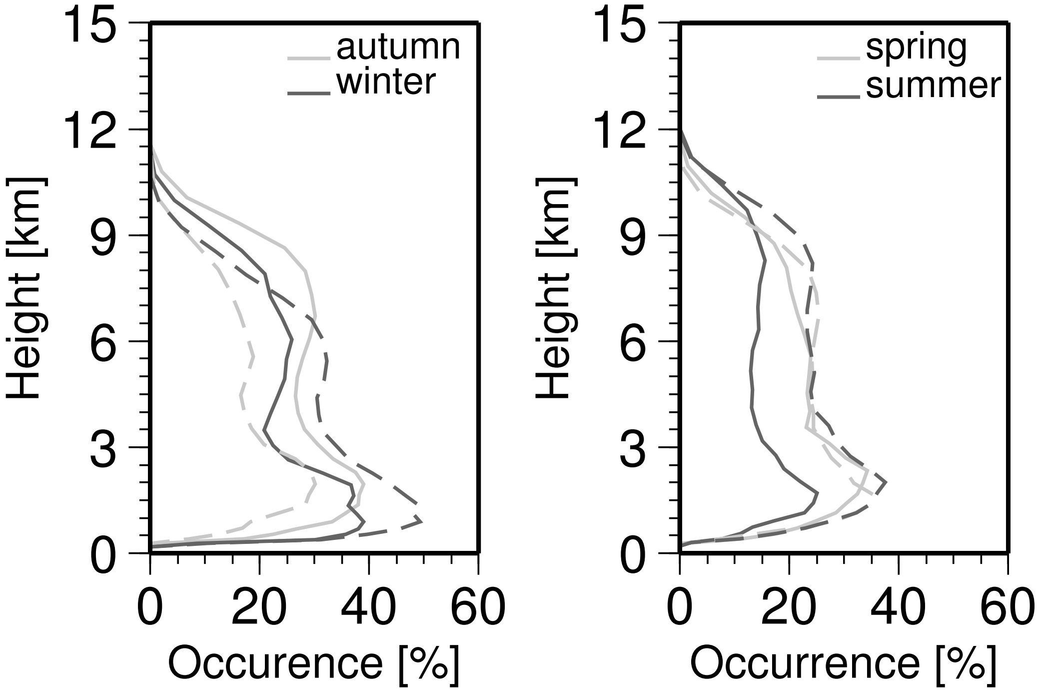

Figure 2: Season to season comparison of frequency of cloud occurrence at the Cabauw site from the observations, for the first year (autumn 2002 to summer 2003, solid lines) and second year (autumn 2003 up to summer 2004, dashed lines) of the project

the ECMWF and Met Office schemes are the most accurate to reproduce the frequency of cloud occurrence of mid-level and low-level clouds, respectively.

3.2 Comparisons at seasonal scale

In this section the Cloudnet data set is split into seasons (winter is from 22 December to 20 march, spring is from 21 march to 20 June, summer is from 21 June to 20 September and autumn from 21 September to 21 December). The objectives here are 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.

When splitting the dataset into 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 is the case for the two 95 GHz radars, Hogan *et al.* 2003), only the Cabauw dataset is used to investigate the season-to-season and seasonal variability.

3.2.1 Season-to-season variability

Figure 2 shows the variations in frequency of cloud occurrence for each season (two seasons by panel) derived from the observations at the Cabauw site. It appears that except for spring (comparison of solid and dashed grey lines in right panel), where the curves for the two years present roughly the same shape, strong differences

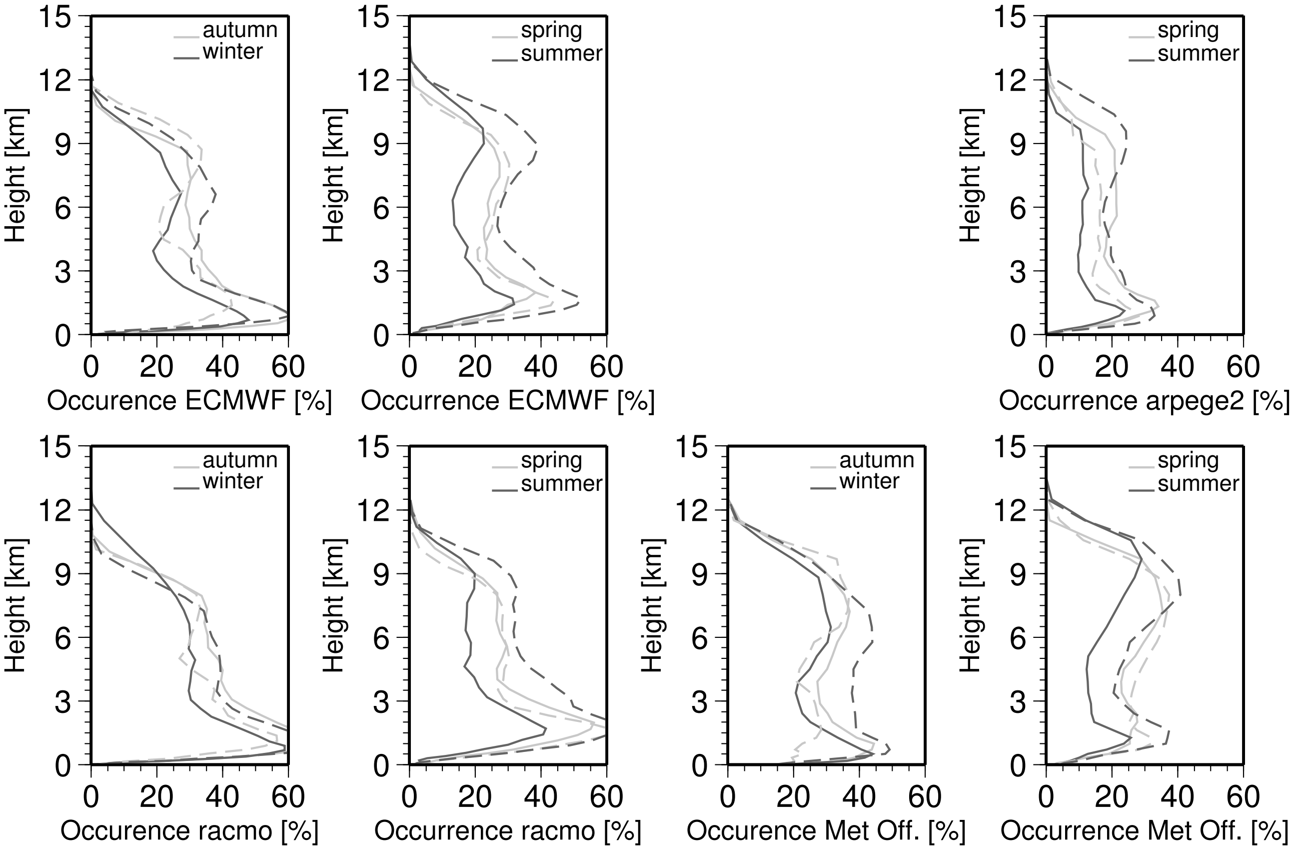

Figure 3: Season to season comparison of frequency of cloud occurrence obtained from the models : ECMWF and ARPEGE (first line), RACMO and Met Office (second line) at the Cabauw site, divided into the first (solid) and second (dashed) years of the project.

(that can be as large as 10%) occurred for a given season between the two years. Overall, these differences appear to be systematic throughout the vertical column (with the exception of the high clouds in winter).

The large season-to-season variability observed in Figure 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. Longer time series are required to derive robust statistical properties at seasonal scale which could be used for statistical evaluation of model or spaceborne cloud retrievals. However, with these two years of observations, we can evaluate if the models are able to reproduce the observed variability. For this purpose, Figure 3 shows the same curves as depicted in Figure 2 but for each model. Due to the change in the ARPEGE parametrisation after winter 2003, autumn and winter seasons cannot be compared between the two years of the project. Therefore the third panel of the first line in Figure 3 is left blank. It is first observed from Figure 3 that all models are able to reproduce the fairly small variability of frequency of cloud occurrence during spring between the two years of the project. The observed large variability in winter and summer (more clouds during the second year) are also well captured by the models. The models are not as successful to reproduce the large variability observed in autumn.

In conclusion 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 seasonal evolution of cyclone activity and large-scale humidity structures well, and that the moist physics parametrisations are able to translate this into appropriate moist convective activity and associate convective or stratiform cloud properties.

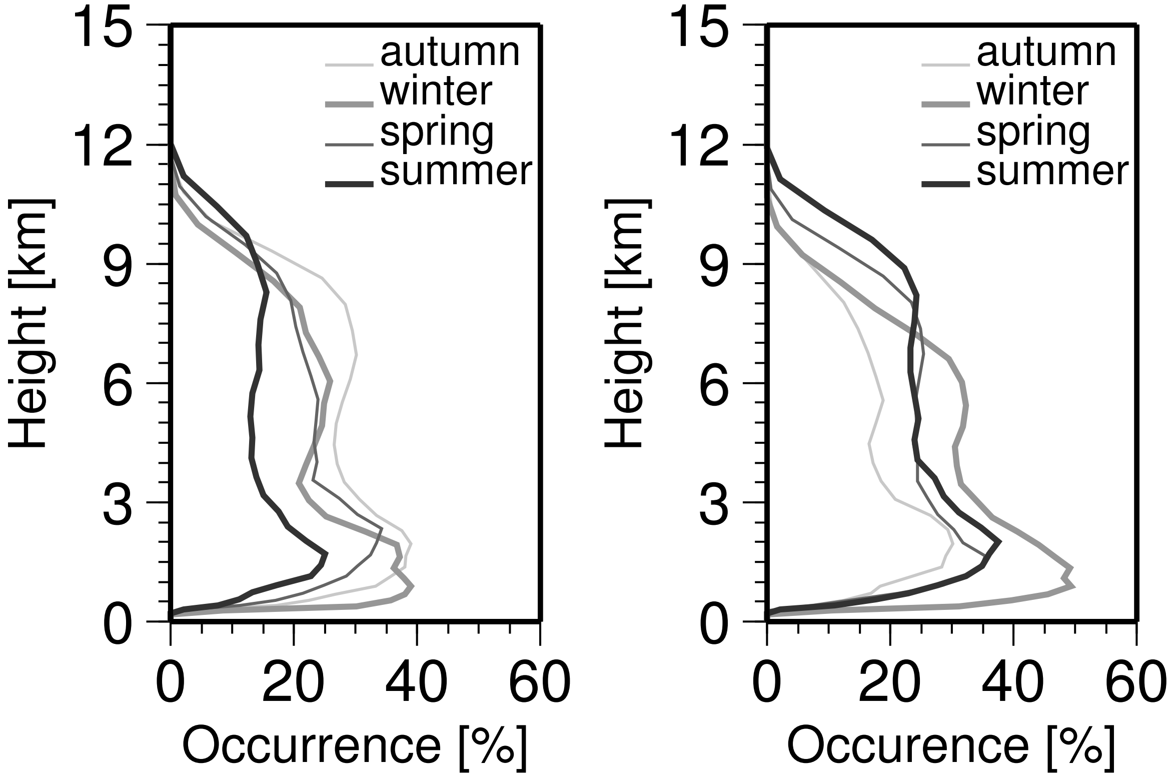

Figure 4: Seasonal variability of frequency of cloud occurrence deduced from observations at the Cabauw site for the first (left panel) and second (right panel) year of the project.

3.2.2 Seasonal variability

Figure 4 shows the same curves as Figure 2 but this time, sorted per year of the project in order to document the seasonal variability of cloud occurrence and evaluate if the models are able to reproduce it. If one wants to compare the evolution of cloud occurrence between the two years the curves (thin solid grey line) for spring may be used as a reference since it has been observed in the previous section that the profile of cloud occurrence is similar during the two years.

For the first year of the project (Figure 4, left panel) the main feature in the observations is that the summer season is characterised by a smaller frequency of 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 (Figure 4, right panel) shows a completely different behaviour with the same frequency of occurrence for summer and spring, the smallest occurrences are observed in autumn for all clouds, and the largest occurrences of low-level and mid-level clouds are observed in winter. During the first year the same occurrence of high-level clouds (above 9 km) was observed for all seasons, while during the second year there was a large variability between the seasons.

We now analyse the ability of the models to reproduce the observed seasonal variability. Figure 5 shows the same curves as Figure 4 but for the models. Due to the change in the ARPEGE parametrisation within the first year of the project, the inter-seasonal variability cannot be investigated for this model during the first year of the project. The smaller 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

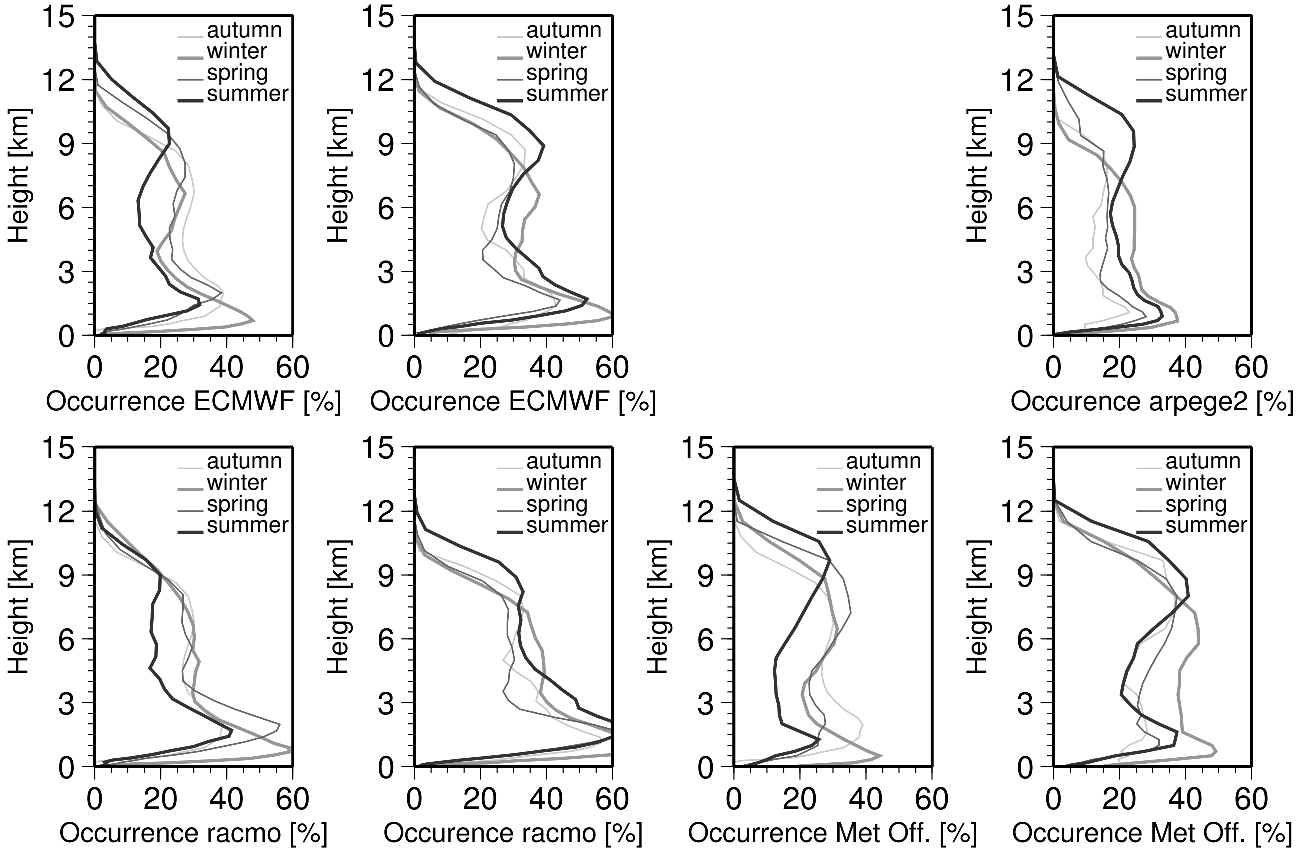

Figure 5: Seasonal variability of frequency of cloud occurrence from the models : (ECMWF, ARPEGE on the first line, RACMO and Met Office on the second line) at the Cabauw site. For a given model, the left-hand side panel displays the seasons of the first year of project, the right-hand side panel the seasons for the second year.

blocking high pressure system in their analyses, and maintain this through the 36 hour 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 for forecast the cloud variability in the medium (5 to 10 day) and extended range forecasts.

More interestingly, 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). This points to model problems with 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 the models are not able to reproduce the smaller cloud occurrence in autumn, however the larger frequency of occurrence of low-level and mid-level clouds for winter 2004 is captured in ARPEGE and Met Office. Finally, only RACMO seems to reproduce well the different behaviour of high cloud occurrence between the two years, with the same occurrence of high-level clouds (above 9 km) for all seasons, and a large variability between the seasons for the second year. In general, it is also 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 Conclusions

This paper presents an evaluation of the cloud representation in four numerical weather prediction models by making use of long time series of cloud parameters collected at three different sites in western Europe. The ability of the models to reproduce the distribution of cloud observed at a given location has been evaluated. The companion paper (Bouniol *et al.* 2007) is dedicated to the evaluation of the parameters involved in the state-of-the-art cloud schemes (IWC and cloud fraction).

The cloud occurrence in the models has been evaluated with the whole Cloudnet data set and on a seasonal basis in order to highlight potential to generate different weather situations. Particular attention has been given to potential biases linked to the instrumental combination (94 GHz or 35 GHz radar in combination with a lidar or ceilometer) and especially difference in instrumental sensitivity. These instrumental effects have been taken into account showing that either a radar combined either a lidar or radar with a high sensitivity and a ceilometer provide similar results, with an unbiased cloud occurrence sampling up to 9 km altitude. These results should be considered with some caution though, since some errors are introduced in the sensitivity threshold applied to the model reflectivities arising from the use of an IWC-Z relationship to translate the model IWC into reflectivity.

To summarise the main results of this paper, the comparison of the frequency of cloud occurrence derived from observations with model statistics show that all the models significantly overestimate thick cloud occurrences at high-levels. Among the models, the Met Office, arpege1, and ECMWF cloud schemes produce the largest overestimations of occurrence of such clouds. 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 mid-level and low-level clouds, respectively.

The objective of a season to season comparison is to determine if short validation periods can be used for tuning of a cloud scheme, and it is indeed shown that biases observed over the 2-year comparison are seen in individual seasons. However, the models struggled to reproduce all aspects of inter and intra seasonal variability indicating that a long time series than 2 years is required for a statistically robust evaluation.

It has also been demonstrated in this paper, as in the overview of the Cloudnet project of Illingworth *et al.* (2007), that it was possible to follow a model evolution in cloud parametrisation using the monitoring by ground-based remote-sensing stations. In this paper the case of the ARPEGE model has been particularly illustrative. However the benefit from a change in parametrisation cannot be evaluated with a single variable like the cloud occurrence, which is mostly related to a grid-box threshold (deciding if cloud occurred or not) and therefore to the thermodynamical state of the model. The next step is to consider the cloud condensate amounts and cover; the aim of the companion paper (Bouniol *et al.* 2007).

Finally it would be interesting to repeat the same exercise but using the spaceborne CloudSat radar and CALIPSO lidar measurements available since April (2006) as part of the A-Train mission. Indeed 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 liquid clouds below, and the cloud radar would be of similar sensitivity as the ground-based radars.

Acknowledgements 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, Bauer P, Beljaars A, Chevallier F, Holm E, Janisková M, Kallberg P, Kelly G, Lopez P, Mc Nally A, Moreau E, Simmons A, Thepaut J-N, Tompkins AM. 2005. Assimilation and modeling of the atmospheric hydrological cycle in the ECMWF forecasting system. *Bull. Am. Meteorol. Soc.***86**: 387–402.

Bouniol D, Protat A, Delanoë J, Pelon J, Donovan DP, Piriou J-M, Bouyssel F, Tompkins AM, Wilson D, Morille Y, Haeffelin M, O'Connor E, Hogan RJ, Illingworth AJ. 2007. Using continuous ground-based radar and lidar measurements for evaluating the representation of clouds in four operational models. Part II: Cloud fraction and ice water content. *Quart. J. Roy. Meteor. Soc.*, submitted to Quart. J. Roy. Meteor. Soc.

Brown PRA, Illingworth AJ, Heymsfield AJ, Mc Farqhar GM, Browning KA, Gosset M. 1995. The role of Spaceborne Millimetr-Wave Radar in the Global Monitoring of Ice Cloud. *J. Appl. Meteor.***34**: 2346–2366.

Haeffelin, M, Barths L, Bock 0, Boitel C, Bony S, Bouniol D, Chepfer H, Chiriaco M, Cuesta J, Delanoë J, Drobinski P, Dufresne J-L, Flamant C., Grall M, Hodzic A, Hourdin F, Lapouge F, Lemaître Y, Mathieu A, Morille Y, Naud C, Noël V, O'Hirok B, Pelon J, Pietras C, Protat A, Romand B, Scialom G, Vautard R. 2005. SIRTA, a ground-based atmospheric observatory for cloud and aerosol research. *An. Geophys.* **23**: 253–275.

Gregory D, Morcrette J-J, Jakob C, Beljaars ACM, Stockdale T. 2000. Revision of convection, radiation and cloud schemes in the ECMWF integrated forecasting system. *Quart. J. R. Meteorol. Soc.***126**: 1685–1710.

Guichard F, Parsons DB, Dudhia J, Bresch J. 2003. Evaluating Mesoscale Model Predictions of Clouds and Radiations with SGP ARM Data over a seasonal Timescale. *Mon. Weather Rev.* **131**: 926–944.

Hogan RJ, Bouniol D, Ladd DN, O'Connor EJ, Illingworth AJ. 2003. Absolute calibration of 9495 GHz radar using rain. *J. Atmos. Oceanic Technol.***20**(4): 572–580.

Hogan RJ, Jakob C, Illingworth AJ. 2001. Comparison of ECMWF Winter-Season Cloud Fraction with Radar-Derived Values. *J. Appl. Meteor.***40**: 513–525.

Illingworth AJ, Hogan RJ, O'Connor EJ, Bouniol D, Brooks ME, Delanoë J, Donovan DP, Gaussiat N, Goddard JWF, Haeffelin M, Klein Baltink H, Krasnov OA, Pelon J, Piriou J-M, Protat A, Russchenberg HWJ, Seifert A, Tompkins AM, van Zadelhoff G-J, Vinit F, Willen U, Wilson DR, Wrench CL. 2007. Cloudnet - continuous evaluation of cloud profiles in seven operational modeles using ground-based observations. *Bull. Am. Meteorol. Soc.***88**-6, 883–898.

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

Jakob C. 2003. An improved strategy for the evaluation of cloud parametrizations in GCMs. *Bull. Am. Meteorol. Soc.* 1387–1401.

Lenderink G, van den Hurk B, van Meijgaard E, van Ulden AP, Cuijpers JWM. 2003. Simulation of present-day climate in RACMO2: first results and model developments. *Tech. Rep.***252**, KNMI, de Bilt, The Netherlands

Levinson DH, Waple AM. 2004. State of the climate in 2003. *Bull. Am. Meteorol. Soc.***85**: S1–S72.

Liu CL, Illingworth AJ. 2000. Toward more accurate retrievals of ice water content from radar measurements of clouds. *J. Appl. Meteor.***39**: 1130–1146.

London J. 1957. A study of the atmospheric heat balance. Final Report Contract AF19(122)-165, College of Engineering, New York University, 99pp.

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

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

Morcrette J-J. 1991. Radiation and cloud radiative properties in the ECMWF prognostic cloud scheme using LITE. *J. Geophys. Res.* **26**: 1417–1420.

Morcrette J-J. 2002. Assessment of the ECMWF Model Cloudiness and Surface Radiation Fields at the ARM SGP site. *Mon. Weather Rev.* **130**: 257–277.

Norris J, Weaver C. 2001. Improved techniques for evaluating GCM cloudiness applied to the NCAR CCM3. *J. Climate* **14**: 2540–2550.

Protat A, Armstrong A, Haeffelin M, Morille Y, Pelon J, Delanoë J, Bouniol D. 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.

Rossow WB, Schiffer RA. 1983. The International Satellite Cloud Climatology Project (ISCCP): The first project of the World Climate Research Program. *Bull. Am. Meteorol. Soc.* **64**: 779–784.

Siebesma AP, jakob C, Lenderink G, Neggers RAJ, Teixeira J, van Meijgarrd E, Calco J, Chlond A, Grenier H, Jones C, Kölher M, Kitagawa H, Marquet P, Lock AP, Müller F, Olmeda D, Severijns C. 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 RNB. 1990. A scheme for predicting layer clouds and their water content in a general circulation model. *Quart. J. R. Meteor. Soc.* **116**: 435–460.

Stockes GM, Schwartz SE. 1994. The Atmospheric Radiation Measurement (ARM) Program: Programmatic background and design of the Cloud and Radiation Test Bed. *Bull. Am. Meteorol. Soc.* **75**: 1201–1221.

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

Tsedioudis G, Jakob C. 2002. Evaluation of midlatitude cloudproperties in a weather and a climate model: Dependence on dynamic regime and spatial resolution. *J. Geophys. Res.* **107**: 4781, doi:10.1029/2002JD002259.

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

Xu K-M, Randall DA. 1996. Explicit Simulation of Cumulus Ensemble with the GATE Phase III Data: Comparison with Observations. *J. Atmos. Sci.***53**: 3710–3736.

Yang S-K, Hou Y-T, Miller AJ, Campana KA. 1999. Evaluatioin of the Earth Radiation Budget in NCEP-NCAR reanalysis with ERBE. *J. Climate* **12**(2): 477–493.