GEOPHYSICAL RESEARCH LETTERS, VOL. 25, NO.10, PAGES 1645-1648, MAY 15, 1998

# Validation of hydrometeor occurrence predicted by the ECMWF model using millimeter wave radar data

Gerald G. Mace
Department of Meteorology, University of Utah, Salt Lake City

Christian Jakob
European Center for Medium-Range Weather Forecasts

Kenneth P. Moran
NOAA Environmental Technology Laboratory, Boulder, CO

**Abstract.** Validation of hydrometeor prediction by global models is an important issue as it pertains to the accuracy of climate predictions. In this study we use data from a continuously operating millimeter wave radar at a research site in north central Oklahoma, USA to validate output from the operational ECMWF forecast model. We demonstrate that the ECMWF model shows good overall skill at predicting the vertical distribution of the clouds and precipitation that occurred over this site during winter 1997. However, we also show that the model tended to predict the onset of deep cloud events too soon, made the layers too deep and predicted dissipation somewhat later than observed.

## 1. Introduction

Clouds are known to play a substantial role in the global climate system. However, the ambiguity present in cloud climatologies derived from surface observer records or satellite observations precludes a thorough validation of cloud prediction schemes in general circulation models (GCMs). Since information regarding layer boundaries is often uncertain, causal mechanisms leading to agreement or disagreement of the model prediction with data is difficult to ascertain. Any model adjustments that bring compared parameters into better agreement are not guaranteed to improve the overall model skill in predicting the parameters that were not compared.

In this note, we explore an approach to model validation that uses a nearly continuous record of radar reflectivity observations collected at the Atmospheric Radiation Measurement Program (ARM; Stokes and Schwartz, 1994) Southern Great Plains (SGP) site from December 1996 through February, 1997. Even though this study considers only a single location, the comparison, nonetheless, allows a thorough evaluation of the model's skill at predicting the occurrence of clouds and precipitation in the vicinity of the SGP site during this particular winter season.

## 2. The Cloud Radar and Preliminary Data Processing

The millimeter cloud radar (MMCR; Moran et al., 1998) at the SGP site operates at a frequency of 35 GHz and is designed to map the distribution of clouds and precipitation in the vertical column above the instrument (Kropfli et al., 1995). While the radar uses a relatively low peak power transmitter (100 w), the high duty cycle (25%) and large antenna (57.2 dB gain) allow for quite high sensitivity (near -50 $dBZ_e$ at 5 km). This sensitivity is also attributable to pulse compression technology, the short wavelength (~8 mm) and long dwell times (~1s). Comparison of 1 year of MMCR data with lidar observations suggest that the MMCR observes approximately 95% of all cloud layers (E. Clothiaux, Personal Communication, 1998)

We combine the standard operational modes of the MMCR to generate a binary description of significant echo return in the time-range domain of the data using a statistical masking algorithm similar to that described by Clothiaux et al. (1995). The masking algorithm is applied to each mode sequentially and the masks are then combined. Additionally, the sensitivity of the radar in the lower range gates often leads to significant return associated with non-hydrometeor targets such as insects. We, therefore, combine the radar mask with observations from a co-located Belfort laser ceilometer to discriminate between hydrometeor and other targets.

## 3. The ECMWF Model

The operational ECMWF model is a global spectral model using triangular truncation at wave number 213, 31 levels in the vertical and mean orography. The corresponding model grid is designed to maintain a grid point spacing of about 60 km over the whole globe. The version of the model used for this study is, apart from the horizontal resolution, similar to that used in the ECMWF reanalysis project and is described in detail in Gibson et al (1997). The data assimilation system used is a three dimensional variational system as described in Courtier et al. (1998).

Of special interest for our work is the cloud parameterization (Tiedtke, 1993; Jakob, 1994). The scheme predicts both cloud fraction and cloud condensate using prognostic equations. The prognostic treatment of cloud fraction and the tight coupling of a variety of physical processes to cloud production and destruction are unique in an operational NWP model. It is therefore of substantial general interest to assess its performance.

The data necessary for the comparison to the radar at the SGP site is extracted from ECMWF's ARM data set which provides hourly model output averaged over the four model grid points that fall in an area from 97 to 98.3 W and 36 to 37 N. The data set is derived by combining consecutive 12 to 35 hour forecasts into a long time-series and includes the vertical distribution of cloud condensate and precipitation. The use of short-range

Paper number 98GL00845.
0094-8534/98/98GL-00845$05.00

forecasts ensures a reasonable representation of the large-scale circulation and therefore provides an ideal testbed for the model's parameterization. Hydrometeor boundaries are derived hourly from the model by identifying the occurrence of cloud and/or precipitation water or ice in each model sigma level if the value of any of the four condensate species at that level is greater than zero. Sensitivity tests using water content thresholds greater than zero for the occurrence of condensate in the model have been carried out and do not lead to significant changes in the results.

## 4. Results

Since the radar probes the atmosphere at much higher temporal and vertical resolution than the model output are stored, it is necessary to average the radar data into appropriate vertical and temporal frameworks. Since the model variables are available at the top of each hour of model integration and the model timestep is 30 minutes, the 30 minute period that extends from ±15 minutes from the top of each hour of the radar mask are examined and the occurrence or non occurrence of any hydrometeor return at any time within this period is then mapped to the model sigma levels. Frequency distributions of hydrometeor occurrence predicted by the model and observed by the radar as a function of time and height are then compared. There is a problematic aspect to this comparison that cannot be avoided. We compare a composite of four model grid points valid for a thirty minute period and 1.3°×1.0° area to a thirty minute composite of vertical profiles centered within the four grid point array. There may be instances where the model correctly predicts a condensate layer either upstream or downstream of the radar site within the four grid point array. If the radar observes no hydrometeor return, the event will be mistakenly counted as an incorrect prediction. Given that we are examining a fairly small area during a time when winds are typically strong and the time step is rather long, we do not expect this problem to significantly bias our statistics. This issue is discussed further below.

The three-month vertical frequency distribution of hydrometeor occurrence derived from the radar and model are shown in Figure 1. The agreement between observed and predicted frequencies of occurrence is, for the most part, quite good, although the model tends to predict somewhat greater frequencies than observed. The exception to this occurs in the few levels below 1 km where the radar observes a somewhat greater frequency of occurrence. A prominent peak of observed and predicted clouds/precipitation can be seen in the upper troposphere and the radar data also reveal a peak due to boundary layer condensate in the lowest levels. The observations also tend to indicate a well-defined minimum of occurrence between 2 and 3 km. This minimum in hydrometeor occurrence was also identified by Uttal et al. (1995) in a 1-month study of radar data collected during the FIRE II field program held at Coffeyville, Kansas during November and early December 1991.

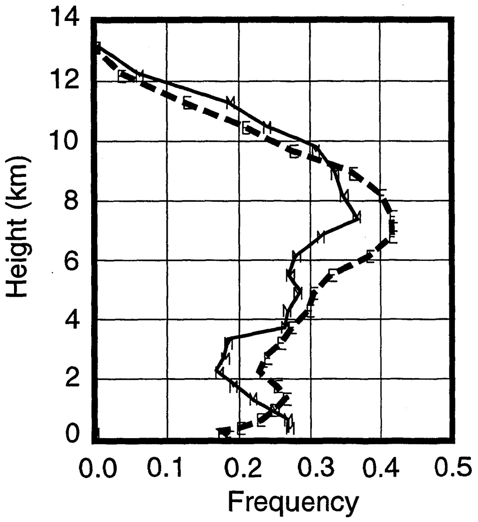

**Figure 1.** Comparison of the vertical frequency distribution of hydrometeor occurrence during the winter season 1996/1997. The ECMWF model is denoted by the dashed line with the symbol "E" and the radar by the solid line with the symbol "M".

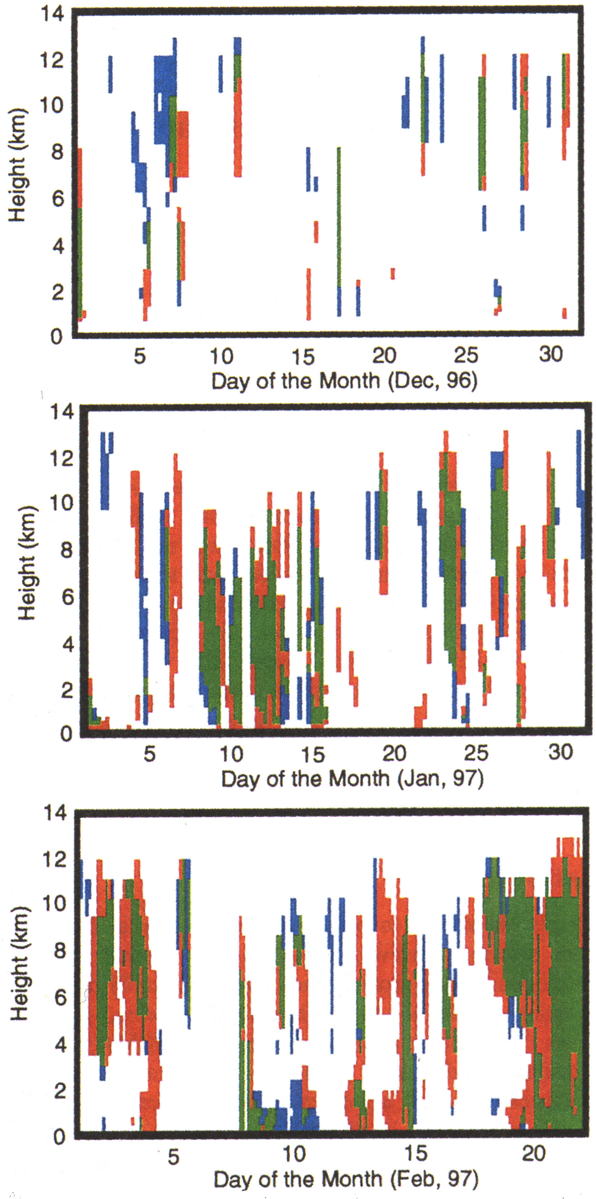

**Figure 2.** Day to day comparison of the ECMWF model output and the cloud radar-derived hydrometeor mask. The color code is shown in Table 1. The radar was not operational from 11-15 Dec. and 6-7 Jan. Data were only available from 00 UTC-12 UTC from 15-31 Dec.

In order to examine under what situations the model and radar agree or disagree, we present the hourly time-height comparisons

**Table 1.** Summary of Model-Radar Comparison Conditions*

| | Yes (Radar) | No (Radar) |
|---|---|---|
| Yes (Model) | *A*, Hit, Green | *B*, False Alarm, Red |
| No (Model) | *C*, Miss, Blue | *D*, Hit, White |

*A "Yes" status indicates the model (radar) predicted (observed) hydrometeor occurrence and a "No" indicates the model (radar) did not predict (observe) hydrometeor occurrence. The italicized letters denote the contingency table used in the quantitative analysis of model performance, "Hit", "Miss" and "False Alarm" denote how the comparison categories are referred to in the text and the colors refer to the contingency status in the time-height comparison in Figure 2.

for each month of the study period in Figure 2. The four possible outcomes of the comparison as well as the color code used in Figure 2 are summarized in Table 1. The major events evident in Figure 2 are well captured by the model. The model shows a tendency to predict hydrometeor onset too early, predict too great a layer depth and predict the dissipation of the layer too late. However, most of the false alarms occur in the vicinity (in either space or time) of observed events, pointing to possible timing and vertical misplacement problems in the model. There are only few false alarms, like that on the 13/14 February that occur in isolation. It also appears that the model tends to often not predict isolated cirrus events like that on 2 and 3 January.

In order to quantify the model performance, several objective measures have been calculated from the statistics gathered over the three months. The measures used are hit rate (*HR*), threat score (*TS*), probability of detection (*POD*) and false alarm rate (*FAR*) (Wilks, 1995):

$$HR=\frac{(A+D)}{(A+B+C+D)},\quad TS=\frac{(A)}{(A+B+C)},\quad POD=\frac{(A)}{(A+C)},$$

$$FAR=\frac{(B)}{(A+B)},$$

where *A*, *B*, *C*, and *D*, are defined in Table 1. Although these measures themselves give some indication of the performance of the model forecast, it is difficult to assess the model quality based only on their absolute values. In order to gauge the model performance, persistence and climatology forecasts have been created from the model output and radar data in the following way. One day of model output (00 UTC to 23 UTC) is gathered from a forecast initialized at 12 UTC of the previous day (see section 2). The persistence forecast is, therefore, created by using the radar observations at each level at 12 UTC as the forecast for the entire next day. To create a climatology forecast we use the three-month mean frequency of occurrence from the radar in each level as displayed in Figure 1. For each forecast time and level, a random number between 0 and 1 is computed and then compared to the mean frequency of occurrence. Neglecting any vertical correlation of clouds or precipitation, Hydrometer occurrence is predicted by the climatology forecast in cases where the random result is equal to or less than the mean frequency of occurrence, otherwise no cloud is predicted.

The four measures of the three forecasts were calculated for each model level separately for the entire data set. The values for the entire data set are summarized in Table 2. The HR, which is the ratio of all correct forecasts to the total number of forecasts, has values above 0.6 for the three forecasts although the model is somewhat better. However, this measure can be misleading because it contains both correct forecasts of the occurrence and the non-occurrence of condensate. If the occurrence of the event is substantially less frequent than the non-occurrence and if forecasting the non-occurrence is 'trivial' then this measure will not be an appropriate gauge of the forecast performance. In our case, non-occurrence is the dominating event although in some levels the mean occurrence exceeds one third. Whether it is 'trivial' to forecast the non-occurrence of condensate can be a matter of debate but it is fair to say that it is the correct forecasting of the occurrence rather than the non-occurrence that is the target of the forecast.

A better measure in the case of dominance of the non-occurrence is the TS. Although the model clearly outperforms the idealized forecasts, the values of this score are lower. The values can be interpreted as the probability of the correctness of a cloud forecast. In that sense the values appear to be very low, but one must keep in mind that the results are produced by a global large-scale model with parameterized cloudiness and that we are asking the model to forecast clouds correctly at a certain level within the correct hour.

As the ratio of the correct forecasts of the event to the total number of observed events POD represents the likelihood of detecting the event to be forecasted. 1-POD is therefore an indicator for the number of misses. The model outperforms the other forecasts in this measure. Again, the interpretation of this measure in isolation can be misleading since a forecasting system that constantly predicts the event would be classified as perfect. It is necessary to evaluate the FAR as well. With FAR values between 0.4 and 0.5 the model is still superior to the persistence and climatology forecasts, however, close to 50% of all forecasts are not observed at the time they are predicted. Recall the issue raised earlier regarding the space-time ambiguity in the comparison of the grid point predictions to the vertical profile observations. Figure 2 clearly shows that many of the false alarms occur in the vicinity (either height or time) of observed events. If we consider this fact, we find that 41%, 52% and 60% of the false alarms occur within 1 vertical level and 1, 2, and 3 hours of an observed cloud or precipitation event, respectively. Further study is required to determine if these events were incorrectly predicted or were predicted to occur upstream or downstream of the radar.

## 5. Conclusions and Future Work

This work should be viewed primarily as a feasibility study for the use of long-term cloud radar data in the validation of cloud parameterizations in GCMs. We have shown that i) radar data are uniquely suited to validating modern parameterizations that predict the vertical distribution of clouds and precipitation and ii) that cloud prediction in the ECMWF forecast model is sufficiently skillful to allow for more in depth analyses of individual cases where the predicted macroscale and microphysical properties of the clouds can be evaluated against

**Table 2.** Model-Radar Comparison Results

| | *HR* | *TS* | *POD* | *FAR* |
|---|---|---|---|---|
| Model | 0.82 | 0.44 | 0.68 | 0.45 |
| Persistence | 0.76 | 0.23 | 0.37 | 0.61 |
| Climatology | 0.69 | 0.13 | 0.23 | 0.77 |

data and the coupling of the cloud prediction scheme with the resolved-scale dynamics can be assessed.

Regarding the first point, since the base and top of most layers can be identified unambiguously and compared directly to the model predictions, the use of millimeter-wave radar data in this type of validation exercise bypasses many of the ambiguities present in more conventional cloud climatologies. A notable exception to this is the ambiguity introduced by directly comparing a height-time section to predictions valid over a geographic area. In future studies where macroscale and microphysical properties are compared, the statistics of the properties can be compared instead of the occurrence. For instance, comparing frequency distributions of fractional cloud cover will reduce the ambiguity substantially. Also, it should be noted that we are able to perform detailed validation with radar observations at only a few gridpoints. The synergistic use of radar data for focused validation and satellite data for global validation will be required to thoroughly assess the accuracy of GCM cloud prediction schemes.

Regarding the second point, simply comparing the occurrence or nonoccurrence as we have done here is only the first step. Randall et al. (1996) outline a number of methods for parameterization validation. We have briefly examined the utility of one of these methods here. However, as discussed by Randall et al., the occurrence of clouds or precipitation at a model grid point is due to a complicated interaction between various elements in the model. Unraveling this interaction to diagnose why a prediction was successful or not is far from straightforward. For instance, it appears that isolated cirrus events are the cloud most likely missed by the model. This may be due to an incorrect prediction of the resolved-scale dynamics in the few cases we examined or to an incorrect formulation of the physics associated with cirrus formation and maintenance. We will evaluate these issues by examining composite statistics of specific cloud types and their relationship to the larger scale parameters (e.g. Mace et al., 1997) over a longer period of time. It will also be necessary to use the other evaluation methods discussed by Randall et al (1996).

Since it is the radiative and latent heat effects of clouds and precipitation that influence climate predictions most directly, the ultimate validation target should be the cloud microphysical parameters that influence radiation and latent heat release. In the near future, a suite of algorithms will be run operationally using the SGP data stream to retrieve the microphysical cloud properties of many cloud types (e.g. Mace et al., 1998).

Finally, this study highlights the potential utility of a space-borne cloud radar. As we have shown here, simply comparing hydrometeor occurrence at a particular point is a potentially powerful diagnostic exercise. Among many other scientific uses, cloud radar data collected from a polar orbiting satellite would allow extension of this study to a global scale. This will certainly move the validation of GCM parameterizations to a much more rigorous level and will ultimately improve the parameterization of clouds in GCMs.

**Acknowledgements.** This research was supported by the Environmental Science Division of the U. S. Department of Energy (grant DE-FG02-90ER61071 and contract number 350153-AQ5). This study benefited from several useful discussions with Dr. Steve Krueger and Dr. Eugene Clothiaux.

## References

Clothiaux, E. E., M. A. Miller, B. A. Albrecht, T. P. Ackerman, J. Verlinde, D. M. Babb, R. M. Peters, and W. J. Syrett, An evaluation of a 94-GHz radar for remote sensing of cloud properties, *J. Atmos. Oceanic Technol.*, *12*, 201-229, 1995.

Courtier, P. E. Anderson, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, and M. Fisher, The ECMWF implementation of three dimensional variational assimilation (3D-VAR). Part I: Formulation. Submitted *Quart. J. Roy. Met. Soc* (In Press), 1998.

Gibson, J. K., P. Kallberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano, ECMWF Re_Analysis Project Report Series. 1. ERA description. *ECMWF, July 1997*

Jakob, C., The Impact of the New Cloud Scheme on ECMWF's Integrated Forecasting System (IFS), *Proceedings of ECMWF/GEWEX Workshop on Modeling, Validation and Assimilation of Clouds*, ECMWF, November 1994, 1994.

Kropfli, R. A., S. Y. Matrosov, T. Uttal, B. W. Orr, A. S. Frisch, K. A. Clark, B. W. Bartram, R. F. Reinking, J. B. Snider, B. E. Martner, Cloud physics studies with 8mm wavelength radar, *Atmos. Res..*, *35*, 299-313, 1995.

Mace, G.G., T.P. Ackerman, E.E. Clothiaux, and B.A. Albrecht, A study of composite cirrus morphology and its relationship to large-scale meteorology. *J. Geophys. Res.*, 102, 13581-13593, 1997.

Mace, G. G., S. Kinne, T. P. Ackerman, An examination of cirrus cloud characteristics using data from a millimeter wave radar and lidar: The 24 April SUCCESS case study. *Geophys. Res Let.* (In Press), 1998.

Moran, K. P., B. E. Martner, M. J. Post, R. A. Krofli, D. C. Welsh, and K. B. Widener, An unattended cloud-profiling radar for use in climate research. *Bull. Am Meteorol. Soc.* (In Press), 1998.

Randall, D. A., K. Xu, R. J. C. Sommerville, and S. Iacobellis, Single column models and cloud ensemble models as links between observations and climate models. *J. Clim.* 9, 1683-1697, 1996.

Schmidt, G., Complementary code and digital filtering for detection of weak VHF radar signals from the mesosphere. *Geosci. Electronics*, GE-11, 154-161, 1979.

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

Uttal, T., E. E. Clothiaux, T. P. Ackerman, J. M. Intrieri, and W. L. Eberhard, Cloud boundary statistics during FIRE II, *J. Atmos Sci*, *52*, 4276-4284, 1995.

Wilks, D.S., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press, 238-242

---

G. Mace , Department of Meteorology, University of Utah, Salt Lake City, Utah, 84112 (e-mail: mace@atmos.met.utah.edu)

C. Jakob, European Center for Medium Range Weather Forecasts, Reading England (e-mail: paj@ecmwf.int)

K. Moran, NOAA, R/E/ET4 325 Broadway, Boulder CO 80303 (e-mail: kmoran@etl.noaa.gov)

(Received: October 21, 1997; Revised: January 9, 1998; Accepted: February 16, 1998.)