Evaluating mesoscale model predictions and parameterisations against SGP ARM data on a seasonal time scale

Françoise M. Guichard, D. Parsons, J. Dudhia, J. Bresh

Evaluating Mesoscale Model Predictions of Clouds and Radiation with SGP ARM Data over a Seasonal Timescale

FRANÇOISE GUICHARD

CNRM-GAME (CNRS and Météo-France), Toulouse, France

DAVID B. PARSONS, JIMY DUDHIA, AND JAMES BRESCH

National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 16 July 2001, in final form 31 July 2002)

ABSTRACT

This study evaluates the predictions of radiative and cloud-related processes of the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (PSU–NCAR) Mesoscale Model (MM5). It is based on extensive comparison of three-dimensional forecast runs with local data from the Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) site collected at the Central Facility in Lamont, Oklahoma, over a seasonal timescale. Time series are built from simulations performed every day from 15 April to 23 June 1998 with a 10-km horizontal resolution. For the one single column centered on this site, a reasonable agreement is found between observed and simulated precipitation and surface fields time series. Indeed, the model is able to reproduce the timing and vertical extent of most major cloudy events, as revealed by radiative flux measurements, radar, and lidar data. The model encounters more difficulty with the prediction of cirrus and shallow clouds whereas deeper and long-lasting systems are much better captured. Day-to-day fluctuations of surface radiative fluxes, mostly explained by cloud cover changes, are similar in simulations and observations. Nevertheless, systematic differences have been identified. The downward longwave flux is overestimated under moist clear sky conditions. It is shown that the bias disappears with more sophisticated parameterizations such as Rapid Radiative Transfer Model (RRTM) and Community Climate Model, version 2 (CCM2) radiation schemes. The radiative impact of aerosols, not taken into account by the model, explains some of the discrepancies found under clear sky conditions. The differences, small compared to the short timescale variability, can reach up to 30 W m^{-2} on a 24-h timescale.

Overall, these results contribute to strengthen confidence in the realism of mesoscale forecast simulations. They also point out model weaknesses that may affect regional climate simulations: representation of low clouds, cirrus, and aerosols. Yet, the results suggest that these finescale simulations are appropriate for investigating parameterizations of cloud microphysics and radiative properties, as cloud timing and vertical extension are both reasonably captured.

1. Introduction

Evaluation and validation of atmospheric models coincided with and contributed to the emergence of these numerical tools; they are indeed as old as models themselves and of critical importance. With time, this task has become more and more complex. Numerical models have been continuously improved to reach a greater degree of realism, the latter being required in order to be able, via a modeling approach, to successfully address a large number of operational and research questions raised within the atmospheric sciences. The evaluation and validation similarly require more sophisticated observational approaches. In particular, the evaluation of model-simulated cloud and radiative processes requires observations, including accurate cloud data and radiation budgets, which are not provided by conventional data utilized for numerical weather prediction. Cloud processes also occur on a subgrid scale with respect to the resolution (both horizontal and vertical) of large-scale models. Cloud–radiation interactions depend on cloud height and thickness, cloud water content, but also microphysical characteristics of cloud such as the size and state of hydrometeors. Thus, large uncertainties affect the prediction of cloud processes by numerical models, including their interaction with radiation via water vapor transports and cloud cover as well as the formation of precipitation. Some aspects of numerical simulations are still difficult to evaluate from operational data alone.

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Dr. Francoise Guichard, CNRM-GAME, 42 av Coriolis, 31057 Toulouse Cedex France.
E-mail: francoise.guichard@meteo.fr

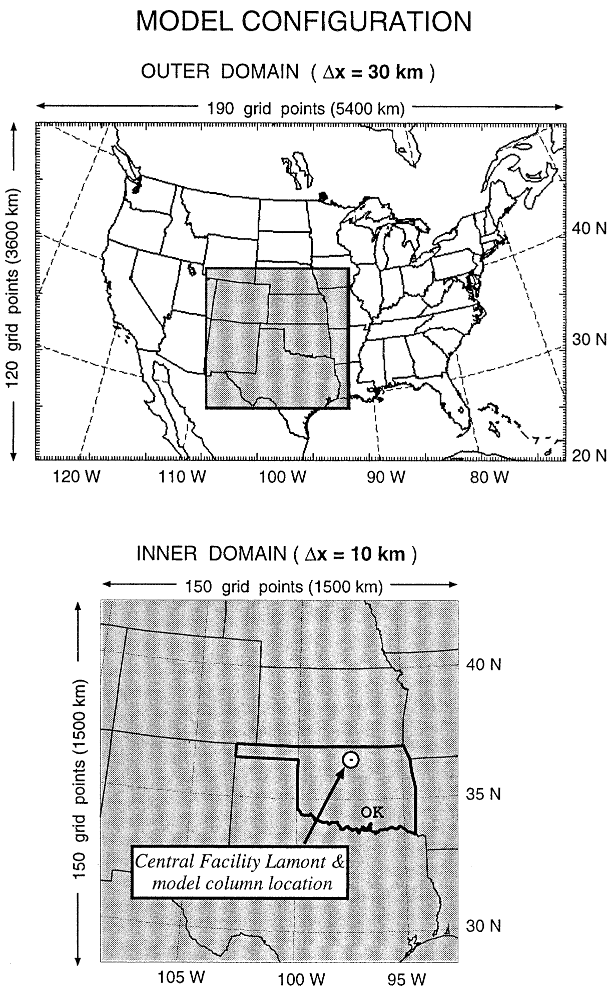

FIG. 1. Physical domain simulated with MM5; the grid point corresponding to the Central Facility site is marked with a black dot within a white disk.

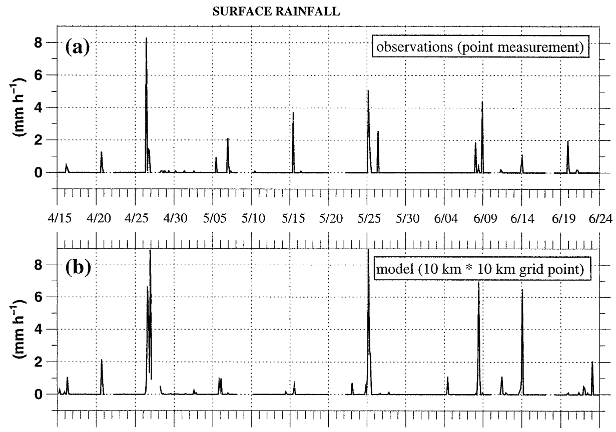

FIG. 2. Time series of surface rainfall (a) measured and (b) simulated at the Central Facility (3-h-mean values).

In the past decades, radiative fluxes and cloud cover diagnosed from satellite data have been widely used and proved to be very useful for the assessment of clouds simulated by weather forecast models (e.g., Morcrette 1991; Jakob 1999; Yang et al. 1999). These evaluations though mostly concern relatively large time- and space scales. Most of them did not directly document the accuracy of simulated cloud vertical structure and surface radiative fluxes, nor their small-scale variability, which are equally essential to assess. A further step can be achieved with observations documenting the evolution of cloud vertical structures and surface radiative fluxes on shorter time- and space scales. With the development of dedicated instruments (e.g., Clothiaux et al. 1999; Hogan et al. 2001), this type of investigation is beginning to be more systematically done, for instance by Bretherton et al. (2003), Duynkerke and Teixeira (2001), or Morcrette (2002). In this respect, the very large cloud and radiation dataset collected by the Atmospheric Radiation Measurement (ARM) experiment (Stokes and Schwartz 1994) provides a unique opportunity to evaluate and improve cloud and radiation parameterizations commonly used in atmospheric models.

A strategy often adopted for the purpose of evaluating model-simulated cloud and radiative processes is to utilize single-column models (SCMs) driven by appropriate boundary conditions derived from observations. The main advantage of this framework consists in isolating the parameterizations from the large-scale flow and, if needed, from surface processes (Randall et al. 1996). This method, largely developed within the Global Energy and Water Cycle Experiment (GEWEX) Cloud System Study (GCSS; Browning et al. 1993), also provides a common framework to perform both SCM and cloud-resolving model (CRM) simulations of given situations using the same boundary conditions. CRMs in turn, once validated with observations, provide synthetic datasets of cloud systems that can further help in testing and improving parameterizations of cloud-related processes (e.g., Gregory and Guichard 2002). In

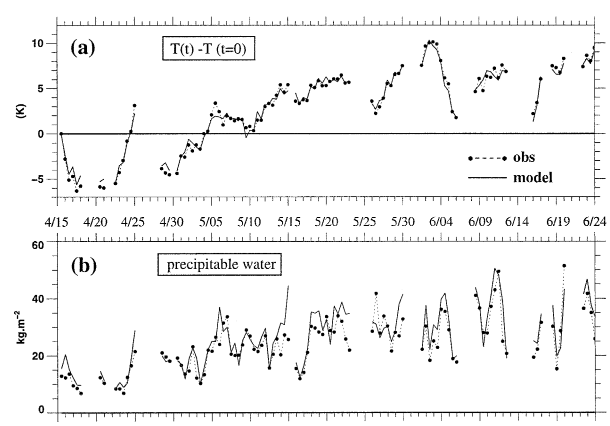

FIG. 3. Time series of observed and simulated (a) temperature and (b) moisture (precipitable water) integrated in the vertical from the lowest model level up to 100 mb using the 101 soundings available at 0000 and 1200 UTC at the Central Facility.

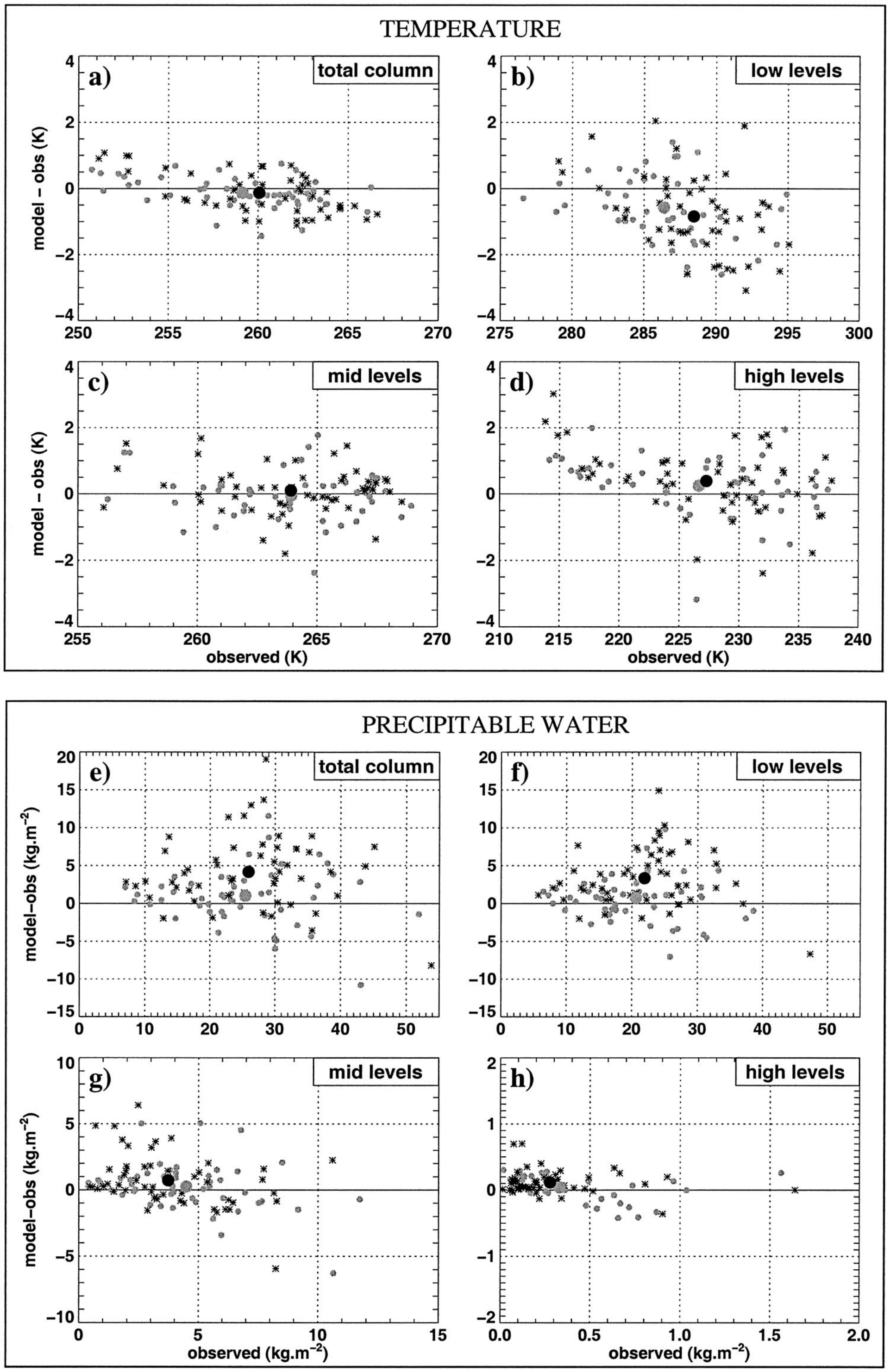

FIG. 4. Scatterplots of observed values versus model bias for temperature: (a) total column (surface to 100 mb), (b) low levels (surface to 680 mb), (c) midlevels (680 to 400 mb) and (d) high levels (400 to 100 mb) averages, (e)–(h) Same as in (a)–(d) except for precipitable water. (i)–(1) Same as (a)–(d) except for relative humidity—values at 1200 UTC (after 12 h of model run) and 0000 UTC (after 24 h of model run) correspond to small gray circles and black stars, respectively; the large gray and black circles indicate mean values for 1200 and 0000 UTC, respectively.

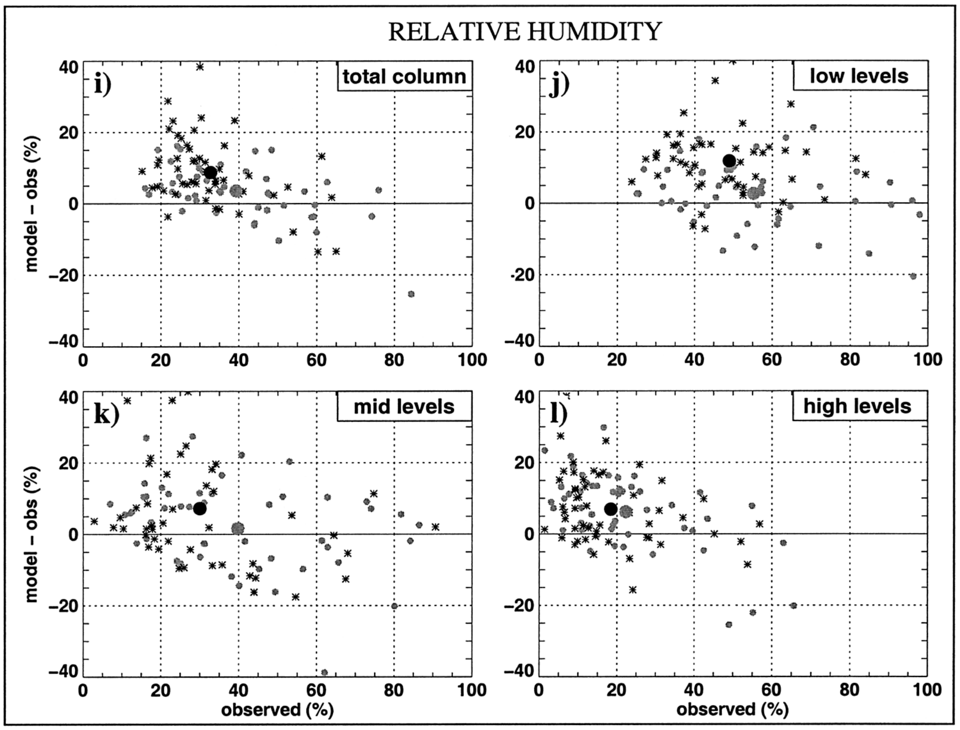

FIG. 4. (*Continued*)

practice, however, the level of accuracy required for boundary conditions is difficult to achieve from observations alone (e.g., Mace and Ackerman 1996; Parsons and Dudhia 1997; Zhang and Lin 1997; Guichard et al. 2000b), requiring additional measurements not routinely provided by ARM. Therefore the SCM approach is limited to relatively short (i.e., one week to one month) intensive observing periods (IOPs), leaving the overwhelming bulk of the ARM observations virtually untouched. As a result, observations from only a few ARM IOPs have been used so far to evaluate parameterizations within this SCM framework (Ghan et al. 2000; Xu et al. 2002; Xie et al. 2002).

An alternative and complementary approach is adopted hereafter, which makes use of the continuous data stream provided by ARM for directly evaluating the predictions and parameterizations of the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (PSU–NCAR) Mesoscale Model (MM5) at the location where observed ARM data are available, as in Morcrette (2002). The analyses were performed for a long series of daily forecast runs over a seasonal timescale, that is, a time period long enough to be relevant to climate goals. This study aims first at evaluating the accuracy of the simulated energy budget and cloud field, how they relate to each other, and also at investigating the major reasons for the differences between model and observations. The questions addressed by this study are: How well are the rainfall and surface radiative budget simulated? How closely are they connected to each other? What are the major differences between observed and simulated temperature and moisture profiles? How important are the errors related to differences in the cloud cover, that is, no cloud versus cloud? Are the model errors in surface radiative fluxes explained by weaknesses in the parameterization of cloud optical properties, or by a lack of aerosols in radiative calculations or by any other likely reason?

In this approach, the boundary conditions, including surface and large-scale forcing, are not prescribed, as is done for SCM runs, but calculated by the model (lateral boundary conditions are prescribed far upstream of the area of interest). As a result, within this ‘‘less controlled’’ framework, differences between the model and observations will also be related to inaccurate surface and/or large-scale forcing. However, because the subgrid (parameterized) processes and the resolved motions are tightly coupled, these errors also help identify major weaknesses of the parameterizations. In fact, large-scale advections prescribed in SCMs, derived from an observed network by methods such as the objective analysis, are also affected by significant uncertainties, as pointed out by Zhang et al. (2001). In the same way as relaxation techniques are applied to long SCM runs, the daily reinitiation contributes to the reduction of ‘‘model-generated’’ errors here. For instance, the consequences of a systematic overestimation of surface moisture fluxes on the simulated cloud field will be more limited than they would have been with multiday simulations. Finally, since the sounding data represents a series of point

TABLE 1. Summary of the data from the ARM SGP Central Facility used in this study.

Instrument observational system	Acronym	Measurement/retrieval used	Original sampling
Surface meteorological observation system	SMOS	Air temperature and precipitation	1800 s
Balloon-borne sounding system	BBSS	Vertical profiles of temperature and relative humidity	2 daily
Millimeter wave cloud radar	MMCR	Vertical profiles of reflectivity	10 s
Micropulse lidar	MPL	Cloud-base height	60 s
Belfort laser ceilometer	BLC	Cloud-base height	30 s
Microwave radiometer	MWR	Vertically integrated liquid water path	60 s
Solar-infrared radiation stations	SIRS	Downwelling surface longwave and shortwave radiative fluxes	60 s

measurements, a direct comparison is, of course, more straightforward for higher-resolution models.

Furthermore, such an analysis provides some guidance on the relevance of using outputs from mesoscale models to generate large-scale boundary conditions for SCMs in data-rich environments like the ARM Southern Great Plains (SGP) site of the United States. For instance, large-scale advection of ice anvils is typically ignored in the SCM approach, because of a lack of observations, but some studies have shown that this process is not negligible and should be included in the forcing on the same basis as are large-scale temperature and moisture advection (Petch and Dudhia 1998). In this respect, a mesoscale simulation integrating a large amount of observations in a consistent and physical way (as presented by Guo et al. 2000) could prove to be very helpful for improving the quality of these forcings. Finally, such evaluations are particularly relevant at the present time, with the emergence of mesoscale models for operational numerical weather forecasting (e.g., the development of the Weather Research and Forecasting Modeling System; Michalakes et al. 2000) and for regional climate studies (e.g., Giorgi et al. 1994).

The paper is organized as follows. The observations, model, and method are presented in section 2. Section 3 focuses on the validation of the simulations with various measurements characterizing the surface, the atmospheric column, and clouds at the ARM Central Facility. In section 4, we discuss some specific weaknesses of the model associated with the radiative parameterization and how the simulations are improved with a more sophisticated scheme. Finally, a summary is given in section 5.

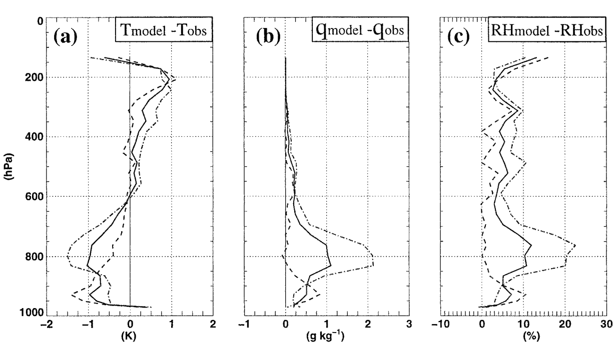

FIG. 5. Time-mean difference between model and observed (a) temperature, (b) specific humidity and (c) relative humidity profiles; solid, dashed and dashed-dotted lines correspond to averages over the 101 soundings, the 1200 UTC soundings and the 0000 UTC soundings, respectively.

2. Data, simulations, and method

Various sources of data are available at the ARM site. A large set of independent measurements characterizing the surface budget and the cloud fields has been used in the course of this study. It includes, in particular, surface temperature and precipitation data obtained from the Surface Meteorological Observation System, surface sensible and latent heat fluxes as well as longwave and shortwave radiation fluxes derived from the Energy Balance Bowen Ratio and Baseline Surface Radiation Network, reflectivities from a cloud radar and cloud-base heights detected by a micropulse lidar (Clothiaux et al. 1999), cloud water contents estimated by a microwave radiometer (Liljegren and Lesht 1996) and sounding data from the Balloon-Borne Sounding System (Lesht 1995). A summary of the data used in this study is given in Table 1.

The model evaluation focuses on real-time forecast runs obtained with the nonhydrostatic version of MM5 (Dudhia 1993; Grell et al. 1995). We take advantage of a series of high-resolution simulations with a two-way nested domain that were performed daily for the Storm and Mesoscale Ensemble Experiment in Spring 1998 (SAMEX ’98). MM5 version 2 (release 2-8) was utilized.

The horizontal grid mesh was 30 km in width for the outer domain, which covers the contiguous U.S. with 190 $\times$ 120 grid points; and 10 km for the inner domain, which includes the SGP ARM site (Fig. 1), with 151 $\times$ 151 grid points. The model uses a terrain-following sigma coordinate in the vertical, with 27 levels. The runs were performed with full physics. The time steps were 90 and 30 s for the outer and the inner domain, respectively. The parameterizations included the radiation scheme of Dudhia (1989), the National Centers for Environmental Prediction (NCEP) Medium-Range

Forecast (MRF) Model's planetary boundary scheme (Hong and Pan 1996), an explicit moisture scheme with ice physics (from Dudhia 1989), and the Grell (1993) cumulus parameterization. An upper-radiative boundary condition (Grell et al. 1995) was used to allow gravity waves to radiate through the top of the model without being reflected.

The simulated period extends over 70 days, from 15 April to 23 June 1998. For each day, a 27-h run was performed, except on 21 and 27 April. Initial and lateral boundary conditions were generated by interpolation of NCEP's ''early-Eta'' model analysis to the model grid. The model was initialized from the 0000 UTC analysis, with the fields being further reanalyzed (Manning and Haagenson 1992) with a multiquadric interpolation (Nuss and Titley 1994) of available surface and upper air data. A linear interpolation of the Eta forecast fields at 6, 12, 18, 24, and 30 h provided the time-varying lateral boundary conditions for the outer domain.

The model time series was created from the last 24 h of each run, allowing for a relatively short 3-h spinup period. The evaluation was performed for a single 10 km $\times$ 10 km width vertical column of the model corresponding to the location of the ARM SGP Central Facility in Oklahoma. In Fig. 1, the tiny black dot surrounded by the white circle indicates the small size of the horizontal area covered by this simulated column, which is located in the 10-km-resolution inner domain. This obviously stresses the challenge encountered by the model for this very ''strict'' evaluation, requiring a large dataset in order to be helpful. For instance, it may happen that a mature cloud passing across the Central Facility was at first correctly initiated in the model (right timing and location). However, a small error in the direction of propagation, or a rapid dissipation of the cloud as it passed through a simulated area drier than observed (or for any other likely reason), could result in a cloud that never reached the Central Facility. Similarly, if the propagation speed of the cloud is too fast or too slow, it will reach the site too early or too late. In all these cases, the model will be in error. As noted earlier though, this finescale ''gridpoint'' comparison with the large observational dataset available at the Central Facility is more accurate and results in a more direct and meaningful evaluation of the day-to-day model behavior than if the same data were used to validate a wider simulated area, as would necessarily be the case for a model with a coarser grid, as in Mace et al. (1998) for instance, because of the very large temporal and spatial variability of cloud and cloud-related processes.

Observations are often sampled on a much finer timescale (on the order of 1 min) than usually needed for our analysis (on the order of 30 min or more). They have been averaged or sampled over an appropriate timescale. In what follows, we frequently use 30-min averaging, consistent with radiative calculations in the model, updated every 30 min. Broadly coherent with the column width, this 30-min time averaging implies an advective scale on the order of 10 to 20 km for wind speed values ranging from 5 to 10 m s^{-1}. For rainfall time series, we use averages over 3 h, which somehow minimize measurement uncertainties.[1] The advective scale increases accordingly to 50 to 200 km for this wind speed range. For radar reflectivity data, a 3-h sampling has been extracted from the high-frequency 10-s time–height series for comparison with model instantaneous hydrometeor profiles available every 3 h. Finally, cloud water content and cloud-base height measurements have not been averaged, in order to avoid mixing clear and cloudy measurements when scattered clouds prevail.

3. Evaluation of surface and cloud parameters

In this section, we evaluate surface and cloud parameters simulated by MM5. The former provide an insight on the accuracy of the simulated surface water and energy budget, but also on the relevance of the simulated cloud field, through its large signature on radiative fluxes. The accuracy of the cloud field is also investigated more directly with cloud data.

a. Precipitation

Time series of precipitation measured at the Central Facility with a tipping-bucket precipitation gauge is shown in Fig. 2a for the 70-day period. Precipitation mostly occurs in the form of a few strong rainy events lasting from a few hours up to more than one day. The model captures most of the rainy events (Fig. 2b), especially the stronger ones. The timing is also reasonably well reproduced except for a few cases (e.g., 5 and 15 May). In terms of intensity range, the agreement for each individual event appears reasonable, given the framework of this evaluation. The 70-day cumulative forecast rainfall is, however, 53% larger than observed, with a total of 198 mm to be compared to 129 mm from observations, largely because of a systematic overestimation of the rainfall rate for most strong convective events (e.g., on 26 April, 25 May, 13–14 June). Possible reasons explaining this discrepancy are numerous and include errors in forcing and/or microphysical and convective parameterizations. In addition, during heavy rain or strong, gusty winds, the collection efficiency of the rain gauge is reduced, but there is no specification of the expected accuracy for these conditions. The apparent overestimation (by 10 to 20 W m^{-2}) of the simulated surface latent heat flux (not shown) could play some role in this result, if it concerns a wider area than this single grid point, but its role was minimized with a model start at 0000 UTC each day. In effect, it corre-

[1] The instrument, a tipping-bucket rain gauge, produces a pulse output every 30 min. The uncertainty is, therefore, a minimum of one full bucket or 0.254 mm per period of integration. Values and timing of low rain rates should therefore be considered with caution.

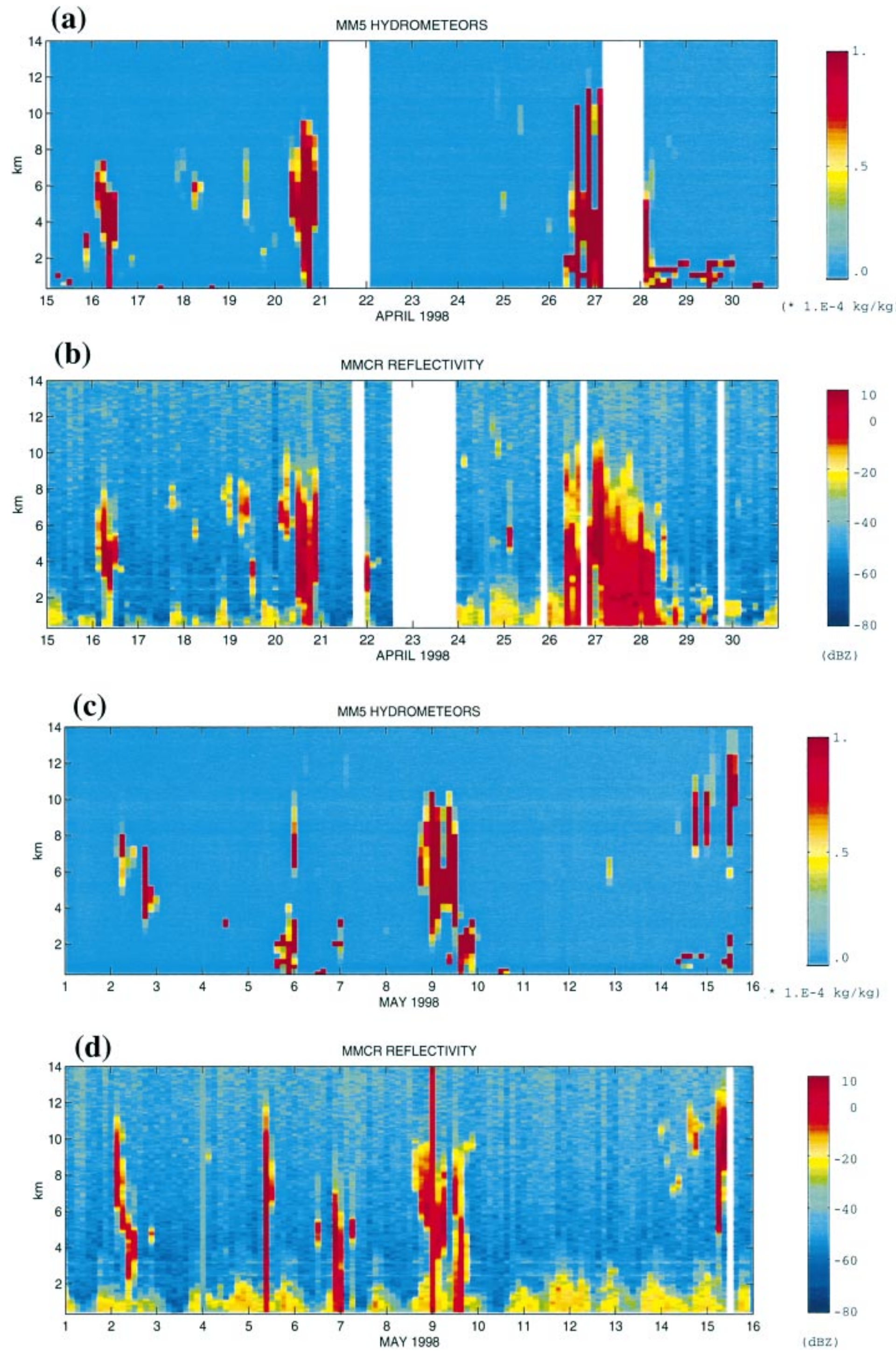

FIG. 6. Time–height cross section of (a) simulated hydrometeor mixing ratio and (b) radar reflectivity measured by the millimeter wave cloud radar (best reflectivity estimate) for the period 15–30 Apr; (c), (d), (e), (f), and (g), (h) same as (a), (b) except for the periods 1–15 May, 16–31 May, and 1–16 Jun 1998, respectively (3-h sampled instantaneous values).

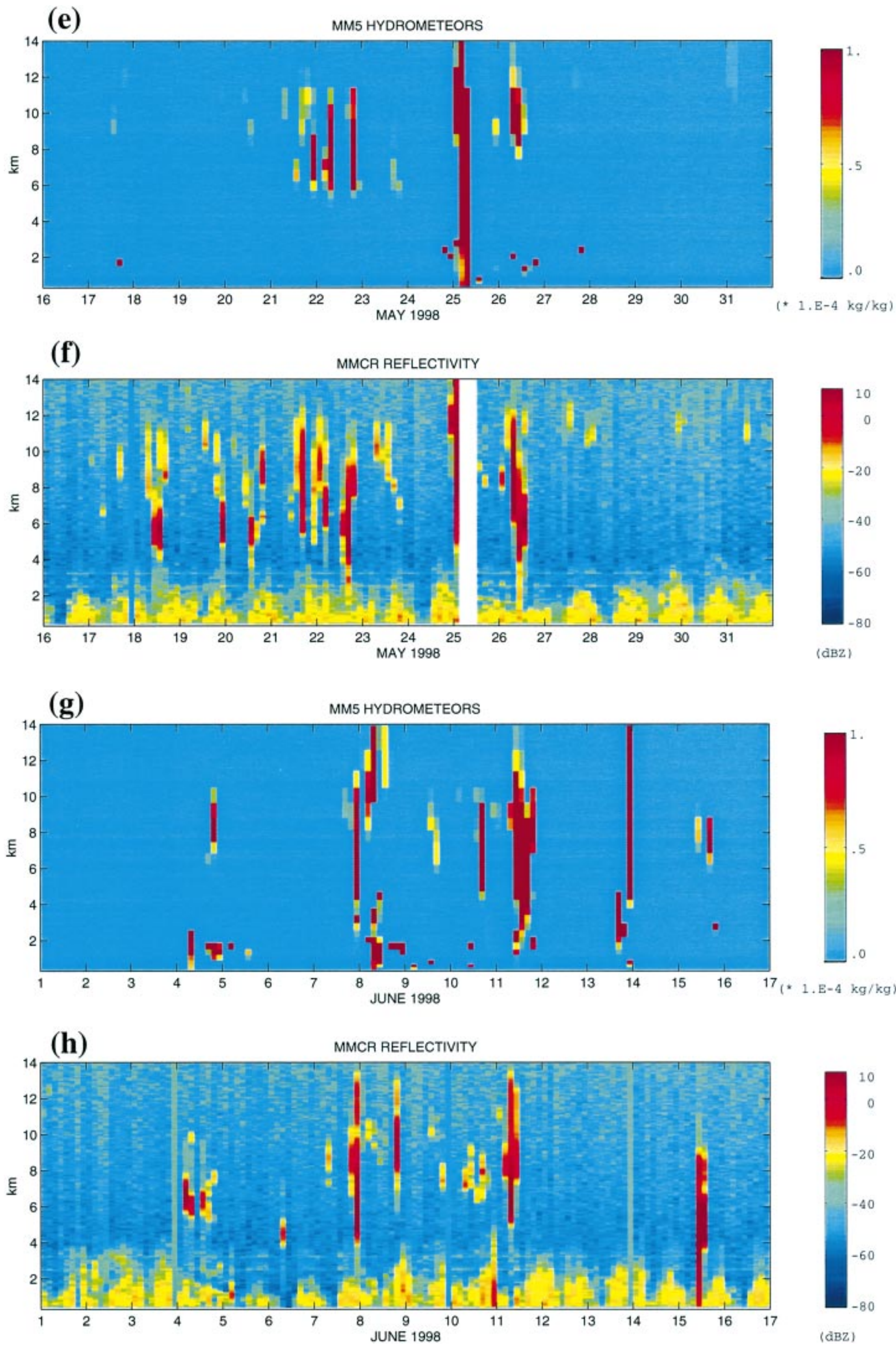

FIG. 6. (*Continued*)

sponds to a start around 1800 LST, so that most of the surface evaporation occurs during the second part of the simulation.

Given the actual state of the art, a 50% overestimation of rainfall over such a small (100 km^2) area is in fact relatively encouraging; at least it is smaller or within the range of reported results, which unfortunately concern much wider areas and/or shorter timescales than shown here. For instance, with NWP models, overestimation of almost 100% has been reported for daily mean precipitation rates (Higgins et al. 1996), as well as 40% overestimation of precipitation at basin and monthly scales (Betts and Viterbo 2000). Indeed, the correct timing and location of rain with global and mesoscale models is already a challenging issue. A commonly found situation is that a model has been able to predict a rainfall event, but the simulated rainfall accumulation is shifted by a few hundreds of kilometers from observations, and/or the amount of rain is much weaker/larger, or the affected area is twice broader/narrower (e.g., Pereira Fo et al. 1999, among many others). Therefore, quantitative precipitation forecasting (QPF) is an extremely difficult task.

In summary, the overall ''reasonable'' simulation of surface rainfall suggests a good timing of clouds that are precipitating, but does not guarantee the accurate simulation of the whole variety of clouds that play a role in the radiative budget at the ARM SGP Central Facility site.

b. Temperature and moisture fields

A correct simulation of clouds also requires that temperature and moisture fields in the model be close to those observed. By design, this will be the case, to a certain extent, as the model is reinitialized every day, in contrast to multiday simulations performed with SCMs or CRMs (e.g., Ghan et al. 2000; Xu and Randall (2000); Emanuel and Živković-Rothman 1999; Guichard et al. 2000b). The reasonable simulation of surface precipitation (Fig. 2) already points to this statement. Therefore, we expect relatively weak departures from observations—departures that can reflect model weaknesses though, and help to identify some parameterization problems.

Simulated thermodynamic fields are evaluated with sounding data (Lesht 1995). The 101 soundings available at 0000 and 1200 UTC for this period at the Central Facility are compared with model instantaneous fields at the same time. (0000 UTC in the model corresponds to simulated profiles after 24 h of simulation, not at the beginning of the simulation, when these (initial) fields agree much better with observed.) Figure 3 shows times series of temperature and moisture vertically integrated through the depth of the troposphere. During this springtime period, the atmosphere evolves from colder and drier to warmer and more moist conditions. The period also includes nonrainy periods lasting several days and showing moderate warming trends, as in May, as well as shorter periods characterized by sharper variations corresponding to different weather regimes (e.g., 22–25 April, 3–6 June). Day–night fluctuations are also partly captured by this 1200 UTC (early morning) to 0000 UTC (evening) time sampling. These various patterns are fairly well reproduced by the model (Fig. 3a). Indeed, most of the time, the column-mean temperature remains within 1 K of observed. The largest departure is related to a precipitating event that was not reproduced by the model (5 May). A careful examination shows that model errors tend to compensate each other during the first 20 days or so. This is not the case, however, later on—the model being 0.25 K colder than observed on average over the last 35 days. This is particularly obvious in June.

A correct simulation of water vapor is difficult because the observed structure of moisture in the atmosphere is strongly linked to processes that are subgrid scale in the model: convection and turbulence, surface fluxes and microphysics. A validation with data from only one given site is also delicate because of the patchy structure of the moisture field. Moreover, sounding data at this site show a dry bias similar to the one reported in Guichard et al. (2000a), a bias which is maximum during spring and summer, when relative humidity is high (Lesht 1999; Richardson et al. 2000). For these conditions, the instrumental bias can lead to errors in the precipitable water (PW) estimations reaching several kilograms per square meter. Despite these limitations, one can notice that time series of precipitable water exhibit several large fluctuations (10 to 20 kg m^{-2}) occurring within a day or so (Fig. 3b). The magnitude and time variations of PW are very similar in the model, though simulated PW is very frequently higher than observed, with both higher maxima and minima. Departures from observations can sometimes be partly explained by differences in the timing of rainy events with 3-h-frequency sounding data available in late April and part of May. For instance, for the period 5–7 May, soundings data show two PW peaks centered on the two rainy events that occurred on 5 May and end of 6/beginning of 7 May, respectively. The 5 May event was delayed by several hours in the model and the second event generated almost no precipitation in the model. Consistent with this behavior, simulated PW shows a first delayed peak and a second weak peak (not shown but partly seen by the 12-h sampling in Fig. 3b). Overall, simulated and observed time variations of PW are close, as expected from daily reinitialization.

In Fig. 4, model biases are presented for the column-mean and for three atmospheric layers (low level: surface to 680 mb; midlevel: 680 to 400 mb; and high level: 400 to 100 mb) as a function of observed value, distinguishing between early morning (1200 UTC) and evening (0000 UTC) biases. Column-mean temperatures almost never depart from more than 1 K from observed, but a weak trend can be seen in Fig. 4a, from a warm bias under colder conditions to a cold bias for warmer ones, with more scatter as observed temperature increases, for both morning and evening model biases. These two features are also characteristic of each of the three layers (Figs. 4b,c,d). Departures from the observed tend to compensate each other in the column, with, on average, a cold bias at low levels and a warm bias at high levels. On average, the bias is also larger in the evening than in the morning, but this average evening–morning difference of the bias (less than 1 K) is much smaller than the difference of the bias among individual soundings (up to a few Kelvin).

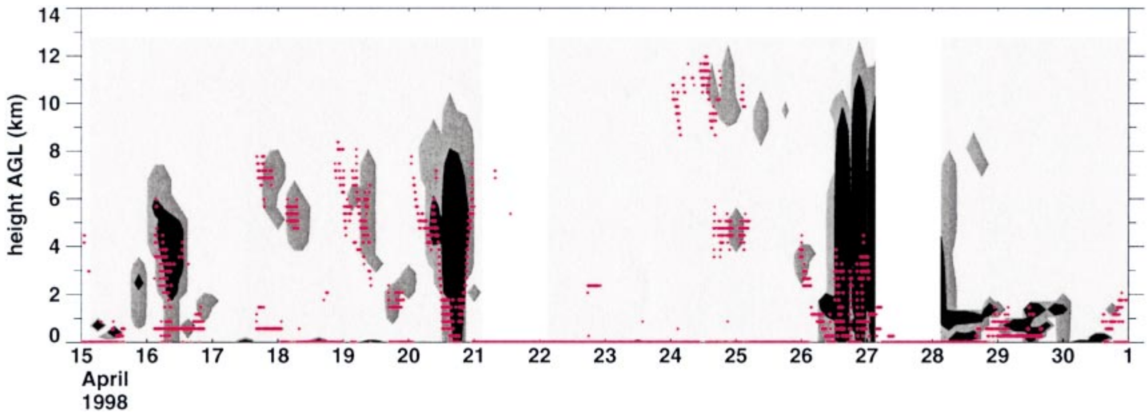

FIG. 7. Time–height cross section of simulated hydrometeor mixing ratio for the period 15–30 Apr 1998 (3-h sampling) and cloud-base heights (small red dots) measured by the micropulse lidar at the Central Facility (1-min sampling, data missing for part, of 15, 17, 29, and 30 Apr); light gray, gray, and black shadings correspond to hydrometeor mixing ratio values lower than 10^{-6}, between 10^{-6} and 10^{-4}, and higher than 10^{-4} kg kg^{-1}, respectively.

On average, the model is also too moist for each of the three layers, the time mean bias being always larger in the evening than in the morning (Figs. 4e–4h for precipitable water and Figs. 4i–4l for relative humidity). This result could be broadly consistent with the simulated surface moisture fluxes being too strong, thus providing too much moisture during daytime (if this problem is affecting a wider area than this grid point only). Despite the large scatter in the results, it is noticeable that dry biases are more commonly found with morning than with evening soundings, both at low and high levels (Figs. 4f and 4h). Most of the total PW bias (2.57 kg m^{-2}) is due to model errors in low levels (cf. Figs. 4e and 4f). Again, in Figs. 4e and 4f, a trend can be seen towards larger positive model biases when the atmosphere moistens. The range of fluctuation of instantaneous PW biases is comparable to the one reported in Morcrette (2002) for the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS).

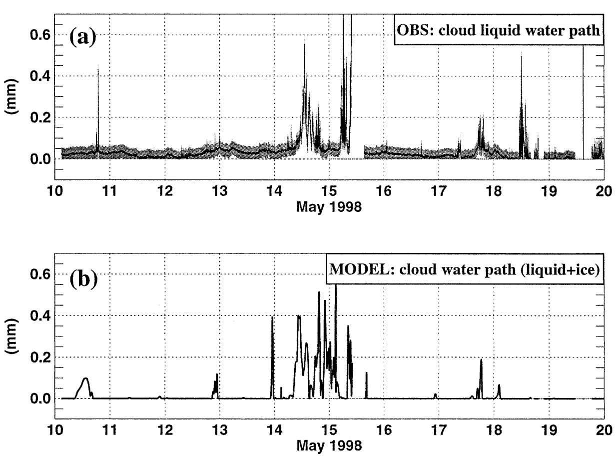

FIG. 8. Time series of 1-min-mean of (a) cloud liquid water path measured by the microwave water radiometer at the Central Facility and of (b) cloud liquid water path, including ice cloud water, simulated by MM5 for the period 10–20 May 1998.

Once averaged over the 101 profiles available at 1200 and 0000 UTC, simulated temperature departs from observations by less than 1 K at any level (Fig. 5a). The model is colder than observed below 600 hPa, except for the lowest level of the model, which is warmer by 0.4 K. The structure of the model bias evolves with time: it is already negative below 600 mb at the end of the night (1200 UTC) with a minimum located below 900 mb, this minimum reaches 800 mb in the evening (0000 UTC). The model is also slightly too warm above 600 hPa—this temperature bias concerns a relatively thin layer, located around 200 mb, at the end of the night, but it propagates to lower levels, down to 500 mb at the end of the day. (We investigate further the source of the temperature bias in section 4.)

The mean simulated moisture profile is within 1 g kg^{-1} of the measured (Fig. 5b, solid line), but the model systematically overestimates moisture, with a positive bias reaching 2 g kg^{-1} around 800 mb in the evening, despite the overestimation of simulated rainfall. This comparison may partly overestimate model errors because of the dry bias in measurements noted above, but the vertical structure of this model bias very likely reflects a model weakness. The model bias increases in height and intensity from the end of the night (1200 UTC) to the end of the afternoon (0000 UTC) and is strongly correlated with the time–height variation of the model cold bias. The model bias for relative humidity is less than +12% on average (Fig. 5c, solid line), but it significantly increases during daytime. It is mostly explained by specific humidity errors, rather than temperature biases, in particular in the upper troposphere though the model is warmer than observed at this height. Its vertical structure and time evolution are similar to the ones noted for the specific humidity bias.

Model errors, increasing during daytime, are probably

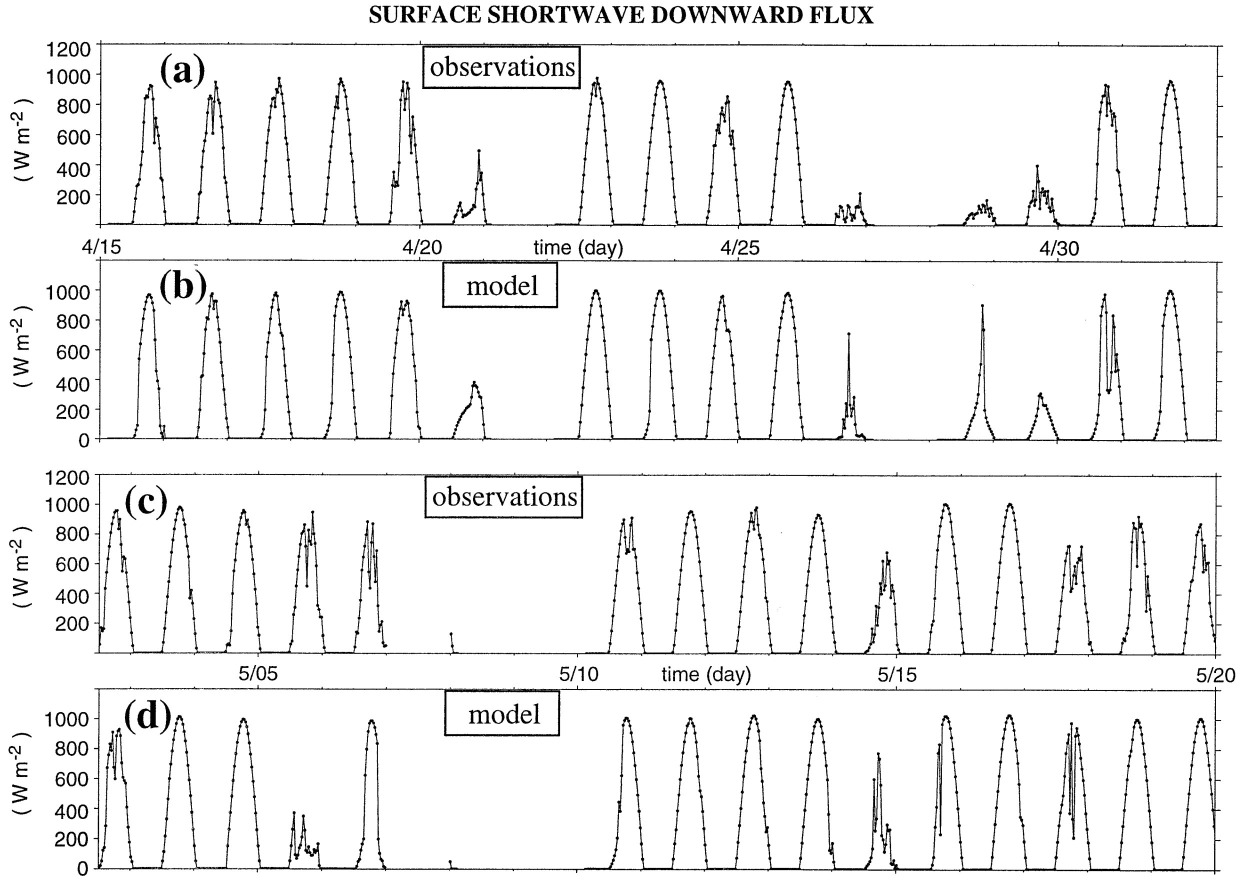

FIG. 9. Time series of (a), (c), (e), and (g) observed and (b), (d), (f), and (h) simulated surface shortwave downward flux (30-min average values) for period 15 Apr–24 Jun 1998.

partly explained by errors in surface heat fluxes, which mostly occur between 0000 and 1200 UTC. In addition, the coupled temperature–moisture bias suggests that the problem is also related to the representation of the boundary layer and shallow clouds in the model.

Thus, despite these discrepancies, indicative of model weaknesses, in particular in the surface and boundary layer parameterizations, time series of temperature and moisture show a reasonable agreement with measurements. The model thermodynamic state always stays relatively close to the real atmosphere, because of the short duration (27 h) of the set of runs, which helps to remove ''additional sources of errors'' (e.g., cumulative errors in the water budget) in the simulation of the cloud fields, a point investigated later.

c. Cloud occurrence

Figure 6 shows time–height series of observed radar reflectivity and simulated hydrometeor mixing ratio, with a time sampling of 3 h. The aim of this comparison is not to provide a quantitative evaluation of simulated cloud water content, but rather, to get a first-order evaluation of the accuracy of the simulated clouds in terms of timing and vertical extent. The radar reflectivities were measured by the millimeter wave cloud radar located at the Central Facility site. This instrument is pointing to zenith and operates at a frequency of 35 GHz. It is able to detect clouds up to 20 km, even thin clouds, due to its high sensitivity (on the order of -40 to -50 dBZ at 5-km AGL) with a 90-m vertical resolution (see Clothiaux et al. 1999 for a full description of the data provided by this instrument). In Figs. 6b,d,f,h, colors from yellow-green (~ -30 dBZ) to red (10 dBZ) indicate the presence of clouds, except in the lowest 2 km. In this layer, high reflectivity values, showing a diurnal cycle, are largely due to nonhydrometeor targets, for example, insects, so that detection of clouds in this layer with this instrument alone is not reliable. Also, the horizontal stripes and noisy pattern in the -40 to -60 dBZ range are related to the fact that these reflectivity data (referred to as ''best reflectivity estimate'') are based on four complementary operational modes. This comparison of radar reflectivity and simulated hydrometeor mixing ratio (Fig. 6) demonstrates that the model captures many aspects of the cloud fields passing above the radar.

First, the simulated and observed deep convective

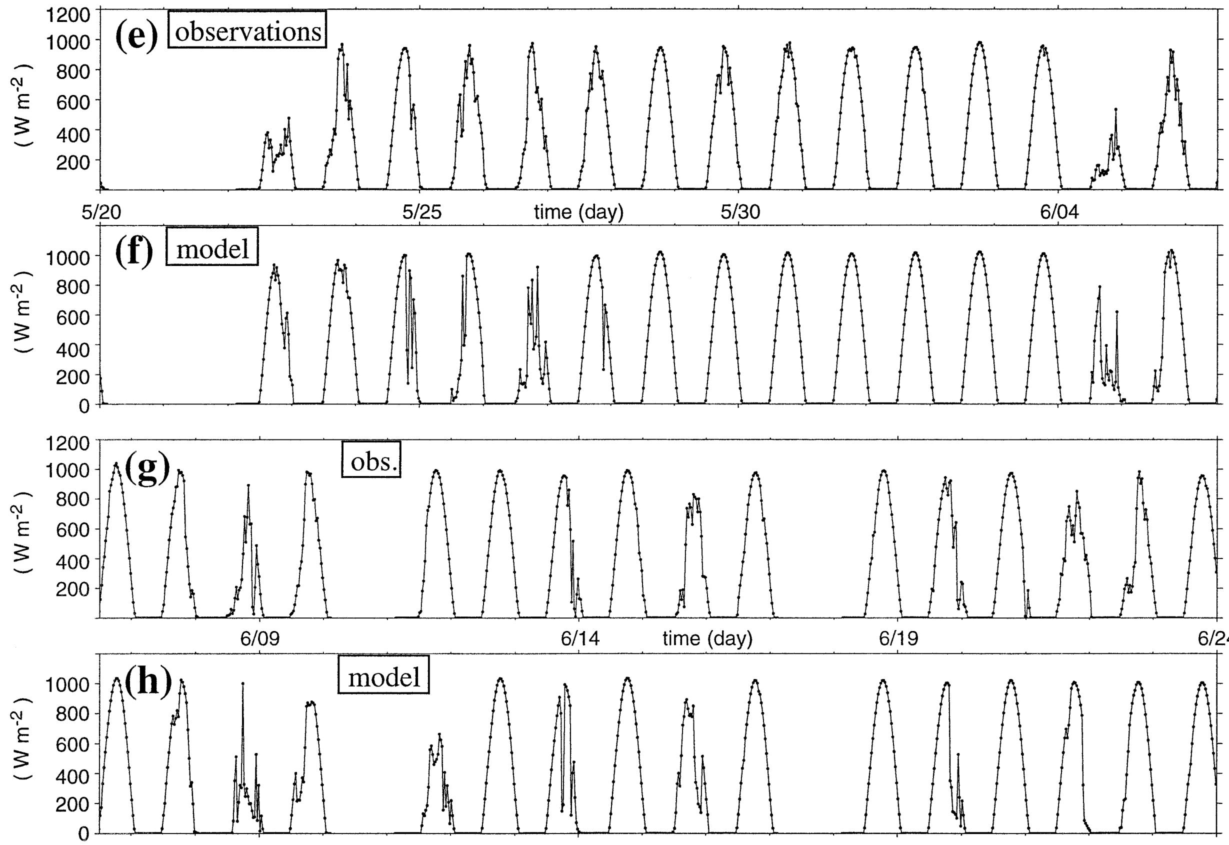

FIG. 9. (*Continued*)

clouds extending through the atmosphere up to 8 km or more share many common features in terms of timing, cloud-top height and lifetime, for instance on 16, 20, and 26–27 April, 9 and 25 May, and 8 and 11 June 1998. Also, on 26 May, when measurements indicate surface precipitation whereas MM5 does not (Fig. 2), the model, however, simulated a thick cloud. In contrast, on 11 June, MM5 does simulate precipitation when observations show none, but both the model and radar data indicate the occurrence of a thick, deep cloud. In fact, the model does not frequently simulate rainfall when the observations indicate clear sky conditions (and conversely). The radar data also show the frequent occurrence of midlevel clouds above 5 km, 2 to 4 km thick, for instance on 18 and 19 April, 2 and 17–24 May, and 5 June. These clouds are less well reproduced by MM5 than the deeper cloud systems. For instance, for the 17–24 May period, midlevel clouds are less numerous in the model than are detected by the radar. This conclusion is also valid for thin cirrus, as observed on the 27–31 May period (this point will be developed later). The characteristics of cloudy events also evolve with time. For instance, occurrences of thick raining clouds at the site tend to last longer during the first half of the period (Figs. 6b,d) than later on (Figs. 6f,h), as the nature of rainy events becomes more convective. Not surprisingly, it seems that the model performs better during the first half of the period.

The cloud-base height measured by the ground-based micropulse lidar provides additional elements for evaluating the simulated cloud field, especially in the lower atmosphere where the radar does not distinguish between hydrometeors and other targets. The vertical resolution of the lidar is 300 m, beginning at 120 m AGL and it can detect cloud base up to 15 km. First, our previous conclusions based on radar data are confirmed in Fig. 7, showing the consistency of the two instruments. One can also notice that the lidar frequently detects low-level clouds during daytime, for example, on 15, 16, 17, 19, and 30 April. This feature, typical of the whole 70-day period, is only partly captured by the model. Several of these low clouds probably correspond to broken cloud fields as suggested by the alternation of cloud/no cloud detection on a short timescale, resulting in the red line at 0 AGL in Fig. 7 for the days mentioned above. This result is also consistent with high-resolution satellite imagery (not shown).

An investigation of cloud water content is also possible with the microwave radiometer (MWR). This instrument provides an estimation of the cloud liquid water path in the atmospheric column above the radiometer, under nonrainy conditions. This estimation is compared to simulated cloud water path in Fig. 8 for a 10-day period. The gray shading in Fig. 8a indicates the range of uncertainty of the measurements. The observed 1-min average time series shows episodes lasting several hours during which the cloud liquid water path is large and highly variable (14–15 May). It also indicates the occurrence of clouds with small water content. Figure 8b shows that the model does not reproduce these thinner clouds. Instead, one can notice that the model generates shorter-duration clouds (lasting from one hour to a few hours, e.g., end of 13 and 14 May). The shorter-duration clouds in the simulation are more of a problem since the simulated values correspond to horizontal means over 100 km^2 whereas measurements are derived from local observations. This model weakness is probably partly related to the absence of a representation of the subgrid-scale nature of these clouds, including the cloud fraction and microphysical processes. A joint comparison of 1-min sampled time series of cloud-base height and downwelling shortwave radiative flux for the same period documents the occurrence of clouds when no firm conclusion can be drawn from the MWR data (for values within the range of uncertainty of the instrument). These data suggest that, except for 11 and 16 May, which were actually cloud-free, the model tends to underpredict low clouds.

At this point, however, the results show that the model captures many features of the observed cloud field, including the observed timing, the cloud-base and -top heights and the cloud liquid water path of the deeper clouds. Thus, predictions of cloud occurrence in MM5 appear to be in the same range of accuracy as reported in Hogan et al. (2001) or Morcrette (2002) for the ECMWF IFS. We now investigate how this conclusion translates to the simulated downward shortwave radiative flux at the surface.

d. *Shortwave radiative flux*

The largest day-to-day fluctuations of the downward shortwave radiative flux are related to the variations in the cloud coverage as shown in Figs. 9a,c,e,g. Several of the cloud events are associated with rain. These fluctuations are also well reproduced by the model (Figs. 9b,d,f,h). A close day-to-day examination shows that the model captures the signature of several cloudy but nonrainy periods. Indeed, the 70-day-mean simulated shortwave downward flux departs from the measured flux by $+11.5$ W m^{-2}. Moreover, a typical problem with pyranometers used to measure this flux is a tendency to underestimate the actual flux by several watts per square meter, the error being larger under clear sky conditions (D. Slater, 1999, personal communication). This instrumental bias likely reaches -5 to -10 W m^{-2} for the present time period. Nighttime negative values have been set to 0 in the present analysis but no correction was applied to daytime measurements. Assuming the time-mean instrument error is the same for nighttime and daytime periods, the difference between the simulated and measured shortwave downwelling flux would decrease to approximately $+8$ W m^{-2}, suggesting that the model only slightly overestimates the shortwave downward flux at the surface.

On 30-min and even daily timescales, however, the error is much larger (Fig. 10). When the timing of rainy and cloudy events is not well reproduced, simulated 30-min-mean fluxes depart positively and negatively by several hundreds of watts per square meter from the observed. These large errors do not always significantly compensate each other, even on a 24-h timescale (e.g., 5 May). On a longer timescale also, three important factors have been identified that induce a systematic overestimation of the shortwave flux at the surface. These factors include an underestimation of cirrus and midlevel clouds, underestimation of low clouds, and the absence of aerosols in the model, as reported later.

When the radar, the lidar, the microwave radiometer and the model all indicate clear sky conditions, the simulated surface shortwave downward flux can still be larger than observed by a few tens of watts per square meter. In fact, under clear sky conditions, the observed shortwave downward flux shows much more day-to-day variation than simulated. This feature, especially pronounced in May, is at least partly due to a smoke pall advected from Central America fires into the ARM SGP site under favorable winds. This event was reported in Peppler et al. (2000), who found elevated levels of aerosol loading in May. Table 2 illustrates this point for two clear sky days, respectively preceding and following a frontal passage (on 15 May) which cleaned the atmosphere of aerosols. At local noon, the measured shortwave downward flux at the surface is higher by approximately 100 W m^{-2} on 16 May than on 11 May; the diffuse radiative flux being also greatly reduced, which is consistent with the low-aerosol optical thickness shown in Peppler et al. (2000) for this day. This leads to a 30 W m^{-2} difference for the 24-h average total flux. It is also striking to notice the good agreement between the simulated and observed flux for 16 May. These two contrasting days represent extreme cases. They suggest however that 24-h-mean differences on the order of 10 W m^{-2} are to be expected under clear sky conditions between the model and observations due to the absence of aerosols in the simulation. Thus, aerosols are likely to play a significant role in the model overestimation of the shortwave downward flux at the surface.

The impact of underestimating cirrus and low-level clouds is investigated for a particularly critical period, extending from 27 May to 2 June. For this period, the simulated column is almost completely clear, whereas cloud radar data indicate the frequent presence of thin (2-km width or less) high cirrus located between 10 and 14 km. These clouds are partly captured by the lidar (Fig. 11). Low clouds were also frequent during daytime (e.g., 27 and 29 May). The model predicts some of those

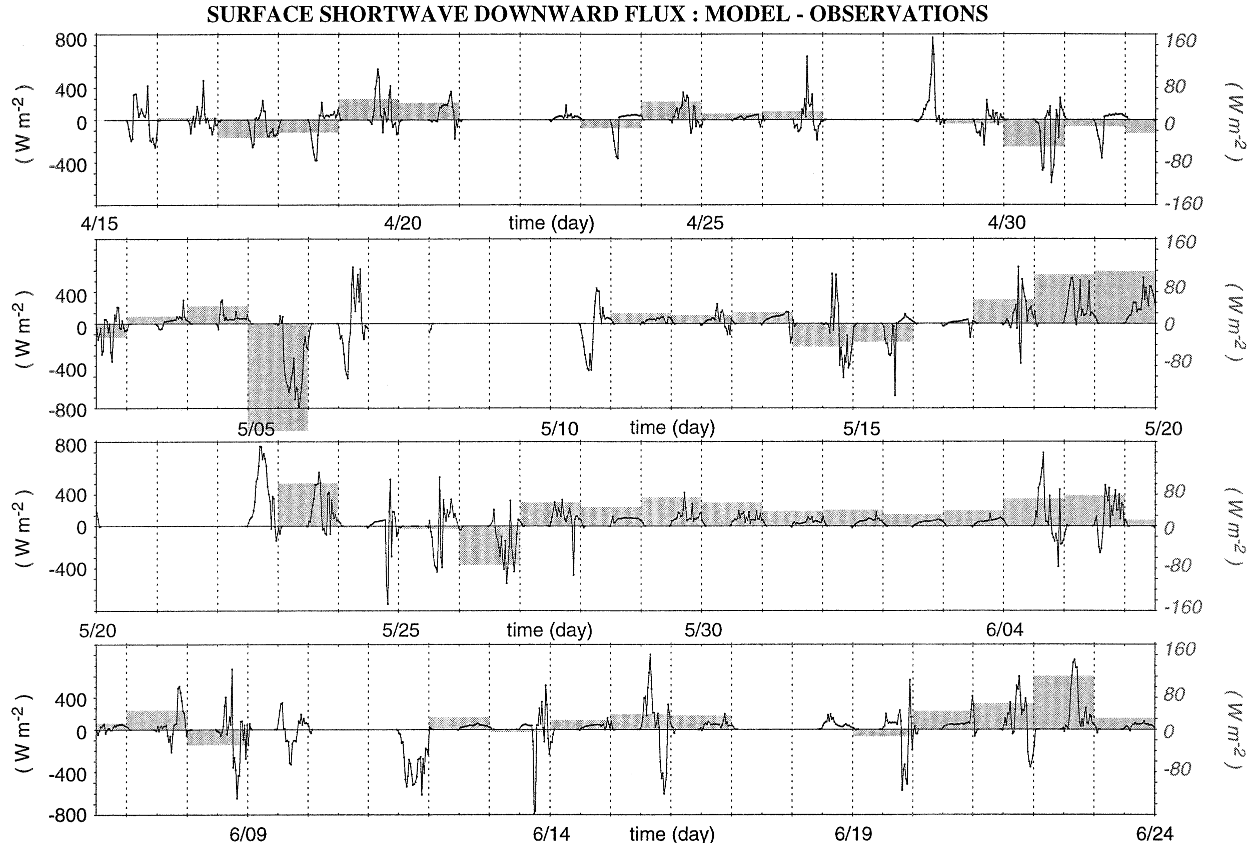

FIG. 10. Time series of differences between simulated and measured surface shortwave downward flux (30-min average values) for period 15 Apr–24 Jun 1998. Gray shading indicates 24-h average values, using the *y* axis on the right-hand side.

clouds (e.g., 27 and 30 May), but clearly underestimates the cloud occurrence for this period. As a result, the simulated shortwave downward surface flux is systematically overestimated (Table 3). Depending on the timing (local noon, evening, etc.) and duration of these low clouds and cirrus, the 24-h-mean cloud radiative forcing ranges from a few to more than 40 W m^{-2} at the Central Facility site.[2] Low clouds apparently have a larger impact than cirrus but both contribute significantly to this forcing. (The aerosols likely played a role during this period too, as the 24-h-mean simulated flux is larger than observed by 20 W m^{-2} on the 2 June under clear sky conditions.) Various factors may be involved in the underestimation of cirrus in the model including the microphysics (e.g., the ''ice settling'' terms are maybe too efficient and/or the ice initiation processes are underestimated at high altitude where the model is warmer than observed). In addition the methodology might play a role too. In effect, the simulated time series consists of a set of daily runs. Each day, the cloud cover is totally removed when the model is reinitialized from a cold start. Therefore, it is difficult to keep track of thin long-lived cirrus from one day to the other. Observations show that cirrus are underestimated overall (e.g., 18–21 and 27–31 May). In addition, when there is cloud ice in the column after 27 h of runtime, the amount of ice is often larger than the one computed in the next run for the same time (i.e., after 3 h of runtime). This is, in some ways, reminiscent of neglecting cloud ad-

[2] This rough estimation is based on a comparison of the first 6 days with the last clear sky day (2 June), neglecting variations of the solar and aerosol radiative forcing as well as temperature and moisture changes over the 7 days.

TABLE 2. The 24 h-mean surface shortwave downward flux at the surface—time average from 0000 LT to 0000 LT next day.

Day (1998)	Measured (W m^{-2}) Tot	Measured (W m^{-2}) Diffuse	Simulated (W m^{-2}) Tot	Simulated–measured (W m^{-2}) Tot
11 May	318	63	338	+20
16 May	348	30	353	+5
16 May–11 May	30	−33	15	–

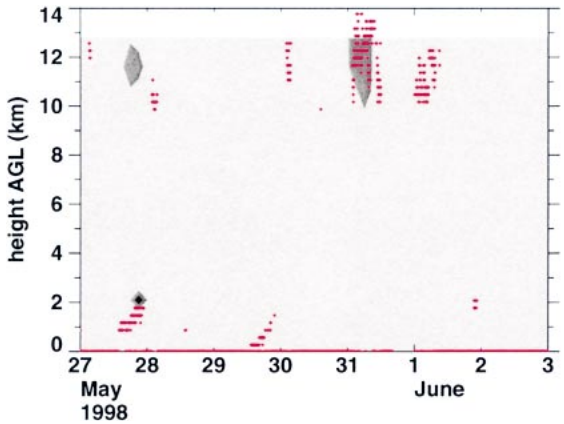

FIG. 11. Same as in Fig. 7 except for the period 27 May–3 Jun 1998.

vection in SCMs (Petch and Dudhia 1998), as the life cycle of these clouds is ignored in both cases. The utilization of the model in a three- or four-dimensional data assimilation (FDDA) mode (Parsons and Dudhia 1997; Guo et al. 2000) with prognostic cloud water at that scale or for initialization techniques that include water processes may contribute to improved simulation of thin cirrus. Finally, the small scale of cloud-related processes, particularly low-level cumulus, also revealed by the large temporal variability of the shortwave flux (Fig. 12), suggests that a treatment of the subgrid-scale nature of these processes should not be included, even at this finescale, as advocated by Pincus and Klein (2000).

4. Improvement of MM5 longwave radiation

As stressed by Chevallier and Morcrette (2000), many radiative transfer models suffer from a clear sky bias in the longwave at the surface. As reported next, we also found some specific problems with the simulation of longwave radiation in MM5. The impact of a better parameterization of the longwave radiative fluxes is also briefly presented.

Time series of simulated and observed longwave downward fluxes at the surface are shown in Fig. 13. The general trend over the season is well captured, beginning with values around 300 W m^{-2} in April, and then gradually increasing up to more than 400 W m^{-2} in late June. Sudden increases associated with the occurrence of clouds are also fairly well reproduced [clearly seen in daily plots (not shown) and still obvious on 8 June in Fig. 13]. The simulated flux agrees quite well with observations during the first 20 days. After that period, however, it is generally overestimated by values of the order of several tens of watts per square meter. A careful examination shows that this bias is not associated with clouds. In fact, under cloudy conditions, model and observations usually agree, for example, for 4, 9, and 14 June. At the same time, the bias is too large to be explained by errors in the simulated thermodynamic fields. A better understanding of this problem is achieved with offline tests. Fourteen fully clear sky days are extracted from the 70-day period. MM5 temperature and moisture profiles for these days are then used as input to three different radiation schemes: namely the MM5 (Dudhia 1989) and Community Climate Model, version 3 (CCM3; Kiehl et al. 1998) radiation schemes and the Rapid Radiative Transfer Model (RRTM; Mlawer et al. 1997), the latter being used in various atmospheric models (e.g., in the operational ECMWF IFS). Briefly, the MM5 radiation scheme is a very simple broadband scheme using a spectrally integrated emissivity function, involving temperature, moisture, and pressure; whereas the CCM3 scheme and RRTM are more sophisticated—both using several spectral intervals (6 and 16, respectively), consider trace gases, and take into account the water vapor continuum. These last two schemes also differ by their algorithms: the CCM3 radiation model employing a broadband emissivity and absorptivity parameterization over each interval and RRTM using a correlated-*k* method. Both the CCM3 radiation scheme and RRTM lead to a substantial improvement compared to the standard radiation scheme, with a systematic reduction of the downward longwave radiative flux at the surface. This decrease is largely explained by differences in the treatment of water vapor radiative properties in the longwave interval. This point is illustrated in Fig. 14, which shows that the difference between the fluxes computed with CCM3 and MM5 radiative schemes increases almost linearly with the precipitable water amount. The slope is on the order of 2

TABLE 3. The 24-h-mean surface shortwave downward flux for the period 27 May–2 Jun 1998 (time average from 0000 to 0000 LT next day). Cloud cover information is for daytime only; observations based on the radar, lidar, and radiometer datasets.

Day	Model		Observations		Mod–Obs
(1998)	(W m^{-2})	Cloud cover	(W m^{-2})	Cloud cover	(W m^{-2})
27 May	333	A few low clouds and thin cirrus	291	Low clouds and cirrus	+42
28 May	355	Clear sky	326	Almost no clouds	+29
29 May	350	Clear sky	291	Mostly low clouds	+59
30 May	355	Very thin cirrus	313	Thin cirrus	+42
31 May	355	Clear sky	330	Very thin cirrus	+25
1 Jun	356	Clear sky	330	A few low clouds	+26
2 Jun	360	Clear sky	340	Clear sky	+20

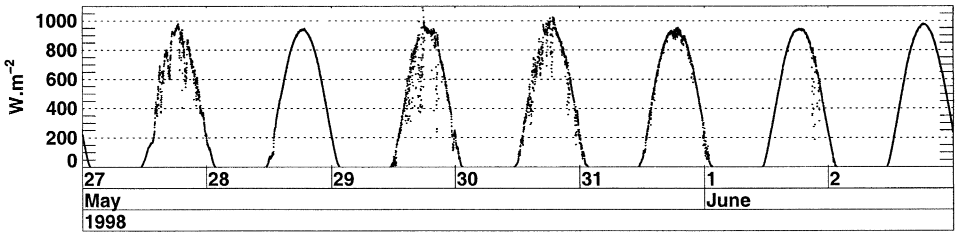

FIG. 12. Time series of observed surface shortwave downward flux at the Central Facility from 27 May to 3 Jun 1998 (1-min sampling).

W kg^{-1}. This explains why, in Fig. 13, the agreement of the simulated surface longwave downward flux with observations is quite good during the first 20 days which are relatively dry, but worse later on, as the total precipitable water increases significantly.

The existing radiation scheme (Dudhia 1989) was first designed to interact with clouds. This is a broadband emissivity scheme; at that time, it was necessary for the scheme to be efficient in terms of computational cost. However, it is likely that a more sophisticated spectral parameterization of radiation, as used in CCM2, CCM3 radiative schemes, and RRTM, becomes necessary for specific types of study, in particular, when the surface energy budget is involved, for instance for regional climate studies.

Offline radiation tests show that either CCM3 or RRTM longwave (LW) parameterizations improve the simulation of the downward LW flux at the surface. Here, the impact of implementing an alternative radiation scheme in the model is analyzed with two in-line tests. Results from these runs are presented for 29 May and compared to the standard run. This day, very moist and clear, corresponds to the most critical day throughout the 70-day period. A first run is performed with CCM2[3] radiation (CCM2 RAD; Hack et al. 1993) and a second one with RRTM longwave radiation scheme (plus the standard MM5 shortwave radiation scheme).

[3] CCM2 rather than CCM3 is used for in-line tests for practical reasons. In fact, they are very similar for clear sky LW calculations (the main difference in this case being the treatment of trace gas, which does not impact our results and conclusions).

Figure 15a clearly shows the large improvement achieved in this fully coupled run. The longwave flux is decreased by more than 50 W m^{-2}, the predicted and observed now agreeing to within 20 W m^{-2}. The spikes in the observed flux (e.g., at 12 and 17 h) are related to the presence of clouds. None of the simulations predicted cloud, so this feature is not reproduced. It is also worth noting that the two radiative schemes—RRTM and CCM2 RAD—lead to surface flux values that differ by up to 20 W m^{-2}. It is not possible from this result only to conclude that one scheme is better than the other: errors in the simulated temperature and moisture profiles can impact the results (by several watts per square meter) and measurements have their own uncertainties too (on the order of 4 W m^{-2}). Rather, at that point, one can put forward that both schemes improve the simulation in comparable ways. The differences between the simulations with the RRTM longwave scheme and CCM2 radiation, respectively, are also consistent with Iacono et al. (2000) radiative calculations for a typical midlatitude summer atmosphere.

A positive impact of the improvement of radiative fluxes occurs at the surface (Fig. 15b). In the standard runs, surface air frequently does not cool enough at night (not shown). For this particular case, temperature was too high especially at the end of the night by more than 4 K. This is not the case with RRTM and CCM2 RAD

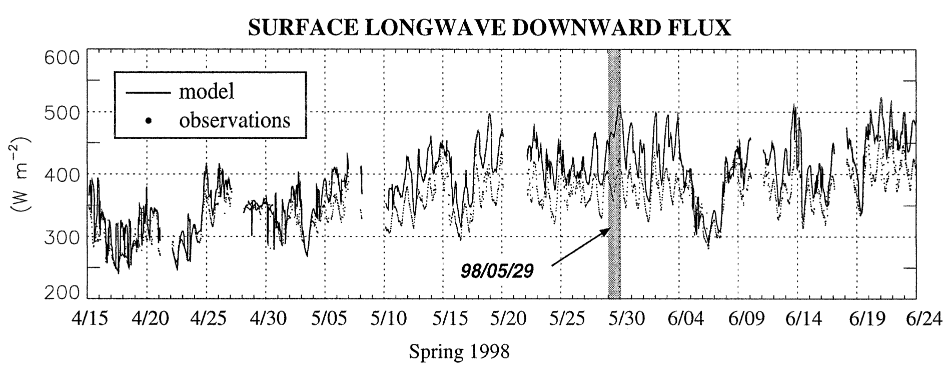

FIG. 13. Time series of simulated (solid line) and observed (dots) surface longwave downward flux (30-min-mean values). Gray shading indicates the day chosen for in-line sensitivity tests of the radiation scheme.

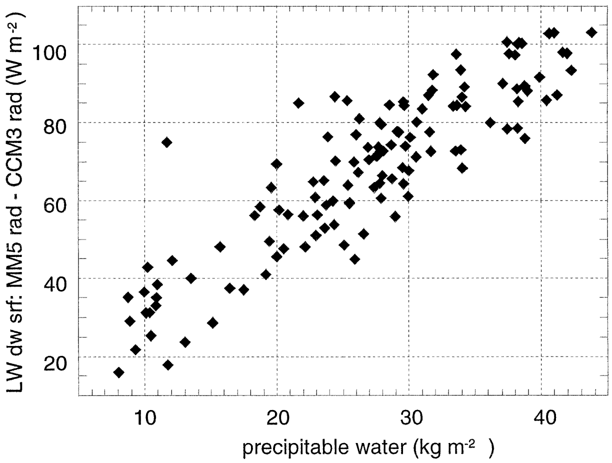

FIG. 14. Scatterplot of the differences between clear sky longwave flux at the surface simulated by the MM5 and CCM3 radiative schemes as a function of precipitable water (offline test for 14 clear sky days, 112 points).

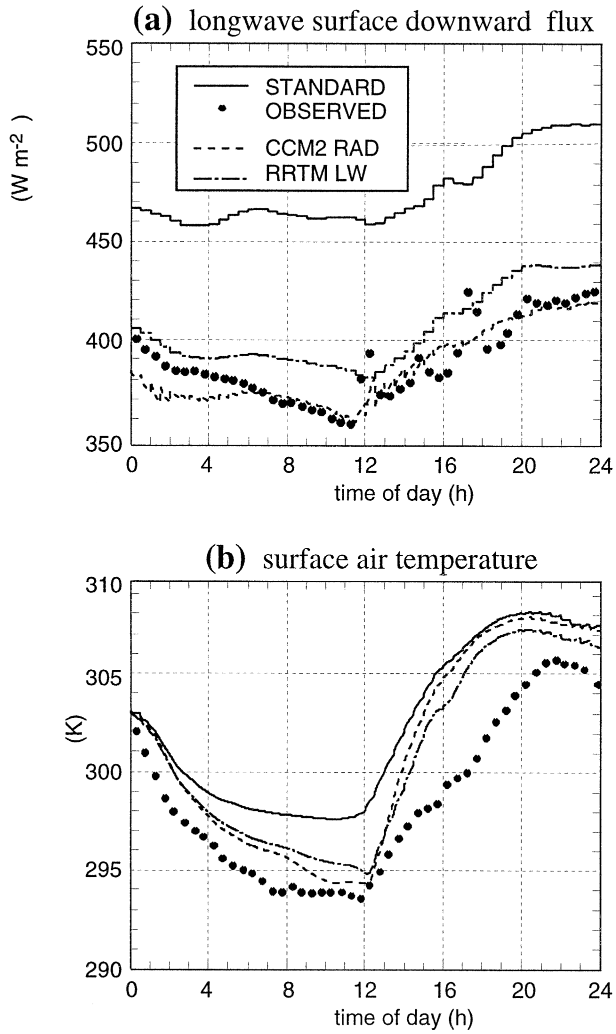

FIG. 15. Comparison with observations of (a) the longwave downward flux and (b) the air temperature at the surface simulated by MM5 for the 29 May 1998 with three different radiative schemes.

and the slope of the nighttime cooling is now very close to observed. Both schemes also act to reduce the overestimation of sensible heat flux that was found in the standard run under clear sky conditions, thus resulting in an improved surface energy budget.

Mean atmospheric cooling rates, Q_{rad}, are presented in Fig. 16. Above 800 mb, values of Q_{rad} are relatively close to each other for the three radiation schemes. Major differences occur below: the previously noted too-large downward longwave flux at the surface in standard runs is linked to a very strong cooling rate in the boundary layer, much stronger than with either CCM2 RAD or RRTM. For this extreme case (29 May), the difference is significant. It is probably balanced by turbulence, which propagates the bias to the surface. In effect, the difference between these radiative cooling rates alone would lead to larger differences in the simulated temperatures than obtained for the three simulations. Yet, these results suggest that the cold bias previously found below 800 hPa (Fig. 5a) could be reduced with a more sophisticated longwave radiation scheme producing weaker longwave radiative cooling rates in this layer.

5. Summary

This study has presented an evaluation of the surface energy budget and cloud field simulated by the mesoscale model MM5 at the ARM SGP Central Facility site in Lamont, Oklahoma. The analysis is performed on a seasonal timescale, with the help of the continuous flow of data provided by ARM, for a 70-day period extending from 15 April to 23 June 1998. Model outputs were generated from time series of daily real-time forecast runs performed over North America. The simulations included an inner two-way nested domain where the horizontal resolution was increased to 10 km, the SGP site being located in this inner domain. The comparison with observations regards one single vertical column only, the one centered on the Central Facility site where a large amount of data are available. Thus, this evaluation corresponds to a very strict and demanding test.

The analysis shows that the model captures fairly well the timing of most precipitating events, though it usually overestimates cumulative rainfall. The model thermodynamic state always stays close to the real atmosphere, as expected with daily reinitialization. Temperature and mixing ratio depart by less than 1 K and 1 g kg^{-1}, respectively from sounding data on average over the period. Larger differences are located below 800 hPa where the model is colder and moister than observed. This reasonable agreement contributes to the quality of the simulated cloud cover. Indeed, the cloud cover compares favorably with both millimeter wave cloud radar data and cloud-base height retrieved from the micropulse lidar for long-lived deep clouds and to a lesser

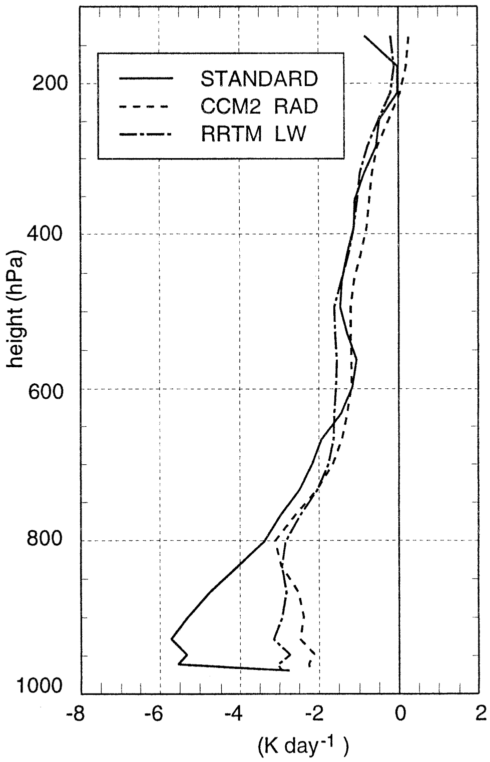

FIG. 16. 24-h-mean radiative heating rate for MM5 simulations of 29 May 1998.

extend for midlevel clouds; low-level clouds and thin cirrus are frequently underestimated though. The 70-day average shortwave downward flux is larger than observed by 5 to 10 W m^{-2}, though 30-min- and 24-h-mean differences are frequently much larger due to the incorrect prediction of cloud occurrence (cloud vs no cloud). The underestimation of cirrus and low clouds and the absence of aerosol radiative effects in the simulation partly explain the model overestimation of the shortwave downward flux. The impact of an advected smoke pall associated with distant fires was noted in the observed aerosol fields. An overestimation of the downward longwave flux at the surface was also found under clear sky, for moist atmospheric conditions. This problem disappears with the implementation in MM5 of more sophisticated radiation schemes (RRTM and CCM2 radiation schemes).

The present study shows that this type of approach is indeed suitable for the evaluation of a mesoscale model performance. The availability of a data stream over long timescales proved to be very valuable. For instance, the nature of errors from the LW radiation scheme were clearly evident through the seasonal cycle in precipitable water. The level of accuracy of the cloud cover suggests that this type of finescale simulation is an appropriate test bed for investigating cloud microphysical and radiative parameterizations. Overall, the quality of the results also indicates that forecast runs at this scale, performed in a three- or four-dimensional data assimilation mode in data-rich environments such as the SGP site, could, as a complement to ''purely observed datasets,'' represent an interesting tool for evaluating and improving parameterizations that are used in large-scale models. The approach is not limited to IOPs and thus can take better advantage of the time series of ARM data.

Acknowledgments. This research was supported by DOE ARM Grant DE-AI0297ER62359 and begun while the lead author was a visitor to the Atmospheric Technology Division of NCAR. We are grateful to the ARM instrument mentors for providing quality measurements and useful advice. W. O. J. Brown, and C. Martin are acknowledged for their assistance with the software, as well as C. Zender, who provided the CCM3 radiative code. Two anonymous reviewers are thanked for their extensive comments, which helped to improved the initial manuscript.

REFERENCES

Betts, A. K., and P. Viterbo, 2000: Hydrological budgets and surface energy balance of seven subbasins of the Mackenzie River from the ECMWF model. *J. Hydrometeor.,* **1,** 47–60.

Bretherton, C. S., S. R. de Roode, C. Jakob, E. L Andreas, J. Intrieri, R. E. Moritz, and P. O. G. Persson, 2003: A comparison of the ECMWF forecast model with observations over the annual cycle at SHEBA. *J. Geophys. Res.,* in press.

Browning, K. A., and GEWEX Cloud System Science Team, 1993: The GEWEX Cloud System Study (GCSS). *Bull. Amer. Meteor. Soc.,* **74,** 387–400.

Chevallier, F., and J.-J. Morcrette, 2000: Comparison of model fluxes with surface and top-of-the-atmosphere observations. *Mon. Wea. Rev.,* **128,** 3839–3852.

Clothiaux, E. E., and Coauthors, 1999: The Atmospheric Radiation Measurement Program cloud radars: Operational modes. *J. Atmos. Oceanic Technol.,* **16,** 819–827.

Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.,* **46,** 3077–3107.

——, 1993: A nonhydrostatic version of the Penn State–NCAR mesoscale model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.,* **121,** 1493–1513.

Duynkerke, P. G., and J. Teixeira, 2001: Comparison of the ECMWF reanalysis with FIRE I observations: Diurnal variation of marine stratocumulus. *J. Climate,* **14,** 1466–1478.

Emanuel, K. A., and M. Živković-Rothman, 1999: Development and evaluation of a convection scheme for use in climate models. *J. Atmos. Sci.,* **56,** 1766–1782.

Ghan, S., and Coauthors, 2000: A comparison of single column model simulations of summertime midlatitude continental convection. *J. Geophys. Res.,* **105** (D2), 2091–2124.

Giorgi, F., C. Shields Brodeur, and G. T. Bates, 1994: Regional climate change scenarios over the United States produced with a nested regional climate model. *J. Climate,* **7,** 375–399.

Gregory, D., and F. Guichard, 2002: Aspects of the parametrization of organized convection: Contrasting cloud resolving model and single column realizations. *Quart. J. Roy. Meteor. Soc.,* **128,** 625–646.

Grell, G. A., 1993: Prognostic evaluation of assumptions used by cumulus parameterizations. *Mon. Wea. Rev.,* **121,** 764–787.

——, J. Dudhia, and D. Stauffer, 1995: A description of the fifth-generation Penn State/NCAR mesoscale model (MM5). Tech.

Note NCAR/TN-398+STR, 138 pp. [Available from National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307.]

Guichard, F., D. B. Parsons, and E. Miller, 2000a: Thermodynamical and radiative impact of the correction of sounding humidity bias in the Tropics. *J. Climate,* **13,** 3611–3624.

——, J.-L. Redelsperger, and J.-P. Lafore, 2000b: Cloud-resolving simulation of convective activity during TOGA-COARE: Sensitivity to external sources of uncertainties. *Quart. J. Roy. Meteor. Soc.,* **126,** 3067–3096.

Guo, Y.-R., Y.-H. Kuo, J. Dudhia, D. Parsons, and C. Rocken, 2000: Four-dimensional variational data assimilation of heterogeneous mesoscale observations for a strong convective case. *Mon. Wea. Rev.,* **128,** 619–643.

Hack, J. J., B. A. Boville, B. P. Brieglieb, P. J. Rasch, and D. L. Williamson, 1993: Description of the NCAR community climate model. Tech. Note NCAR/TN-382+IA, 97 pp. [Available from National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307.]

Higgins, R. W., K. C. Mo, and S. D. Schubert, 1996: The moisture budget of the central United States in spring as evaluated in the NCEP/NCAR and the NASA/DAO reanalyses. *Mon. Wea. Rev.,* **124,** 939–963.

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.

Hong, S.-Y., and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.,* **124,** 2322–2339.

Iacono, M. J., E. J. Mlawer, S. A. Clough, and J.-J. Morcrette, 2000: Impact of an improved longwave radiation model, RRTM, on the energy budget and thermodynamic properties of the NCAR community climate model, CCM3. *J. Geophys. Res.,* **105,** 14 873–14 890.

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

Kiehl, J. T., J. J. Hack, G. B. Bonan, B. A. Boville, D. L. Williamson, and P. J. Rasch, 1998: The National Center for Atmospheric Research Community Climate Model: CCM3. *J. Climate,* **11,** 1131–1149.

Lesht, B. M., 1995: An evaluation of ARM radiosonde operational performance. Preprints, *Ninth Symp. on Meteorological Observations and Instrumentation*, Charlotte, NC, Amer. Meteor. Soc., 6–10.

——, 1999: Reanalysis of radiosonde data from the 1996 and 1997 water vapor intensive observation periods: Application of the Vaisala RS-80H contamination correction algorithm to dual-sonde soundings. *Proc. Ninth ARM Science Team Meeting*, San Antonio, TX, Dept. of Energy.

Liljegren, J. C., and B. M. Lesht, 1996: Measurements of integrated water vapor and cloud liquid water from microwave radiometers at the DOE ARM cloud and radiation test bed in the U.S. Southern Great Plains. *Proc. IEEE Int. Geosciences and Remote Sensing Symp.,* Lincoln, NE, IEEE, 1675–1677.

Mace, G. G., and T. P. Ackerman, 1996: Assessment of error in synoptic-scale diagnostics derived from wind profiler and radiosonde network data. *Mon. Wea. Rev.,* **124,** 1521–1534.

——, 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.

Manning, K. W., and P. L. Haagenson, 1992: Data ingest and objective analysis for the PSU/NCAR modeling system: Programs DATAGRID and RAWINS. NCAR Tech. Note 376+IA, 209 pp.

Michalakes, J., S. Chen, J. Dudhia, L. Hart, J. Klemp, J. Middlecoff, and W. Skamarock, 2000: Development of a next-generation regional weather research and forecast model. *Proc. Ninth ECMWF Workshop on the Use of Parallel Processors in Meteorology*, Reading, UK, ECMWF, 269–276.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.,* **102** (D14), 16 663–16 682.

Morcrette, J.-J., 1991: Evaluation of model-generated cloudiness: Satellite-observed and model-generated diurnal variability of brightness temperature. *Mon. Wea. Rev.,* **119,** 1205–1224.

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

Nuss, W. A., and D. W. Titley, 1994: Use of multiquadric interpolation for meteorological objective analysis. *Mon. Wea. Rev.,* **122,** 1611–1631.

Parsons, D. B., and J. Dudhia, 1997: Observing system simulation experiments and objective analysis tests in support of the goals of the atmospheric radiation measurement program. *Mon. Wea. Rev.,* **125,** 2353–2381.

Peppler, R. A., and Coauthors, 2000: ARM Southern Great Plains site observations of the smoke pall associated with the 1998 Central American fires. *Bull. Amer. Meteor. Soc.,* **81,** 2563–2592.

Pereira Fo, A. J., K. C. Crawford, and D. J. Stensrud, 1999: Mesoscale precipitation fields. Part II: Hydrometeorologic modeling. *J. Appl. Meteor.,* **38,** 102–125.

Petch, J. C., and J. Dudhia, 1998: The importance of the horizontal advection of hydrometeors in a single-column model. *J. Climate,* **11,** 2437–2452.

Pincus, R., and S. A. Klein, 2000: Unresolved spatial variability and process rates in large scale models. *J. Geophys. Res.,* **105,** 27 059–27 065.

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

Richardson, S. J., F. Guichard, and B. M. Lesht, 2000: The radiative impact of the radiosonde relative humidity bias. *Proc. 10th ARM Science Team Meeting*, San Antonio, TX, Dept. of Energy.

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.

Xie, S., and Coauthors, 2002: Intercomparison and evaluation of cumulus parametrizations under summertime midlatitude continental conditions. *Quart. J. Roy. Meteor. Soc.,* **128,** 1095–1136.

Xu, K.-M., and D. A. Randall, 2000: Explicit simulation of midlatitude cumulus ensembles: Comparison with ARM data. *J. Atmos. Sci.,* **57,** 2839–2858.

——, and Coauthors, 2002: An intercomparison of cloud-resolving models with the ARM summer 1997 IOP data. *Quart. J. Roy. Meteor. Soc.,* **128,** 593–624.

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.

Zhang, M. H., and J. L. Lin, 1997: Constrained variational analysis of sounding data based on column-integrated budgets of mass, heat, moisture, and momentum: Approach and application to ARM measurements. *J. Atmos. Sci.,* **54,** 1503–1524.

——, ——, R. T. Cederwall, J. J. Yio, and S. C. Xie, 2001: Objective analysis of ARM IOP data: Method and sensitivity. *Mon. Wea. Rev.,* **129,** 295–311.