Cloud Predictions Diagnosed from Mesoscale Weather Model Forecasts

DONALD C. NORQUIST

Air Force Research Laboratory, Hanscom Air Force Base, Massachusetts

(Manuscript received 9 June 1998, in final form 28 September 1998)

ABSTRACT

Observed cloud characteristics, such as cloud cover, type, and base and top altitude, are of interest to the U.S. Air Force operational community for mission support. Predictions of such cloud characteristics are useful in support of a variety of mission activities. In this paper, a model output statistics approach to diagnosing these cloud characteristics from a forecast field generated by a mesoscale numerical weather prediction model is presented. Cloud characteristics information obtained from the air force RTNEPH cloud analysis supplied the cloud predictands, and forecast fields from the MM5 mesoscale numerical weather prediction model provided the weather variable predictors. Multiple linear regression (MLR) and multiple discriminant analysis (MDA) were used to develop the predictand–predictor relationships using 10 days of twice-daily cloud analyses and corresponding forecasts over a theater-scale grid. The consequent relationships were then applied to subsequent gridded forecast fields to obtain estimates of the distribution of the cloud characteristics at the forecast times. The methods used the most recent 10 days of cloud analyses and weather forecasts to develop the relationship for each successive application day.

The gridded cloud characteristics were diagnosed for 10 days in each of January and July of 1992 over a theater-scale region in southern Europe. The resulting diagnosed cloud predictions were verified against the RTNEPH analyses for forecast durations of 6–36 h at 6-h intervals. It is found that both the MLR and the MDA methods produced small mean errors in all the cloud variables. When compared with persistence, MLR showed skill in rmse in January, while MDA did not. On the other hand, MDA obtained a better score than MLR in percent diagnosed in the correct cloud amount category. Furthermore, the category selection method used with the MDA scheme effectively reproduced the cloud variables' category frequency distribution compared with that of the verification data, while MLR did not. In July, both methods showed skill with respect to persistence in cloud amount. Verification results for cloud type, base altitude, and thickness did not show appreciable skill with respect to persistence. Cloud-ceiling altitude diagnoses showed consistent skill compared to persistence for both methods in both months.

Visual depictions of the total cloud amount distribution as diagnosed by the methods showed that the MDA algorithm is capable of generating useful cloud prediction products. The images produced by the MLR scheme had unrealistically flat gradients of total cloud amount and too many occurrences of partly cloudy skies. The multiple discriminant analysis scheme is considered to be a useful short-term solution to the U.S. Air Force need for predictions of cloud characteristics in theater-scale areas.

1. Introduction

U.S. Air Force mission planners can benefit greatly from a prediction of the future state of cloud distribution in a locale of interest. Clouds impact choices of weapons systems, air combat strategies, ground surveillance opportunities, and aviation operations requiring visual line-of-sight. For these reasons and more, the U.S. Air Force considers cloud prediction a high priority in mission support requirements.

The cloud variables of highest interest to the air force are total cloud amount, cloud-ceiling altitude, and the amount, type, and base and top altitudes of individual cloud layers. Requirements call for the development of methods used to produce predictions of these quantities, and estimates of the accuracies of these predictions, over what are referred to in military terms as ''theater-scale'' areas (500–2000 km on a side). The predictions are required on a horizontal grid in theater, on grid spacings equivalent to those of the predicted standard meteorological variables (temperature, wind, humidity, pressure) as produced by the theater numerical weather prediction (NWP) model. Cloud forecasts are required out to 36 h beyond the current (most recent observation) time, with a stated degree of accuracy.

This paper describes the methods developed to provide the required cloud predictions and presents an assessment of their predictive skill. Because the cloud variables required are not explicitly predicted by any known mesoscale NWP model, a statistical forecast approach was taken in this study. Some current models produce explicit predictions of the cloud liquid and ice

Corresponding author address: Mr. Donald C. Norquist, % AFWA/DNX, 106 Peacekeeper Dr., Offutt AFB, NE 68113-4039.

water content. The vertical discretization of the model may give some indication of the vertical location of the cloud matter. However, there are no known physical relationships that can be used to derive the amount and type from the water content. The model cannot specify the vertical distribution of the cloud matter within the smallest vertical increment (layer) of the model, so base and top altitudes are ambiguous. Future efforts may seek to develop mesoscale NWP model capabilities to produce explicit predictions of all required cloud variables, obviating a need for a statistical forecast methodology.

The methods described here derive statistical relationships between NWP model-generated meteorological variables and observed or analyzed cloud variables, and then apply these relationships to later NWP model forecasts to ‘‘diagnose’’ the cloud variables from the predicted states. This technique is known as model output statistics (MOS). The U.S. National Weather Service has had a long history of the use of MOS in weather forecasting at specific sites. Glahn and Lowry (1972) applied the MOS technique to the prediction of a number of surface weather conditions as well as cloud amount. They developed and applied the MOS equations for individual locations using extensive samples of NWP variables and observed variables. They concluded that the MOS predictions can provide useful forecast guidance. Carter and Glahn (1976) used predictions from NWP models and surface observations of cloud amount to develop and apply MOS equations to forecast cloud amount at more than 200 U.S. stations. Verification scores indicated that the forecast method demonstrated skill comparable to subjective forecasts. Sky cover was one of several variables predicted by MOS equations developed by the Canadian Meteorological Centre (Brunet et al. 1988). They found that MOS forecasts had higher skill scores and were more reliable than ‘‘perfect prog’’ forecasts at longer projection times.

The MOS approach to cloud amount prediction has been attempted by several investigators in global-scale contexts. Trapnell (1992) applied the procedure of Mitchell and Hahn (1989) to the Air Force Global Weather Central (AFGWC) global spectral NWP model forecasts (Stobie 1986) and AFGWC gridded cloud analyses (RTNEPH; Hamill et al. 1992). The method relied primarily on statistical relationships between model-predicted relative humidity (RH) and the cloud analyses. Cianciolo (1993) applied an MOS approach to a 1-yr sample of AFGWC global spectral NWP model forecasts and RTNEPH cloud analyses. Total cloud amount was the sole predictand, but a large number of predicted and derived variables were used as predictors. Norquist et al. (1994, 1997) used a similar methodology with forecasts from the Phillips Laboratory global spectral model (Norquist et al. 1992; Norquist and Chang 1994) and RTNEPH cloud analyses. In this case, relationships between cloud amount and predictors were developed and applied for total cloud and for clouds in three discrete vertical segments of the troposphere, called ‘‘decks.’’ Finally, Nehrkorn and Zivkovic (1996) compared several methods of diagnosing cloud amount using the Phillips Laboratory model and RTNEPH analyses. A consistent finding among these studies was that the schemes with the lowest root-mean-square errors (rmses) had unrealistic frequency distributions, while those that sought to retain the proper frequency distribution of cloud amount had higher rmses.

In the present study, a multiple discriminant analysis (MDA) statistical method is adapted to relate each cloud variable to relevant predictors. Thus it was possible to accomplish a close match between the frequency distribution of the predicted cloud variables and the verifying analyses, while minimizing the loss of rmse skill. In addition, the resulting cloud distributions were more realistic than those produced by multiple linear regression (MLR), in that horizontal gradients were sharper and the overabundance of small (but nonzero) cloud amounts was eliminated. The cost of these improvements was a slight increase of rmse (what MLR is designed to minimize) and categorical, rather than continuous, predictands.

The purpose of this study was to determine the degree of skill that can be obtained in applying a statistical forecast technique to mesoscale cloud prediction. The method developed provides an immediate response to stated U.S. Air Force requirements for the predictions of the aforementioned cloud variables. Assuming that, to some degree, the state of cloudiness depends on the state of noncloud weather variables, the skill of this method should improve with the improvement of skill in mesoscale NWP forecasts (especially cloud-related variables) and the improvement in depiction of gridded cloud analyses. Finally, the skill of the diagnostic method can serve as a benchmark against which to assess the skill of prognostic cloud prediction techniques as they evolve.

2. Diagnostic methods

In this study, the general statistical forecasting method of MOS is used. This method determines an empirical relationship between predictors, which in the MOS method are output fields from an NWP model, and observations of the predictand. In the traditional MOS method, a time series of duration sufficient to account for all of the weather types that can occur at a location is collected from both the NWP model for the predictors and observations for the predictand. A regression analysis is applied to the predictor–predictand time series to arrive at an empirical relationship between the predictand and the selected combination of predictors. Then the relationship is applied to predictor values extracted from subsequent NWP model forecasts at that location, to ‘‘diagnose’’ the value of the predictand at the forecast time.

The earlier study (Norquist et al. 1994) describes the method used to select the predictor variables for the

TABLE 1. List and description of predictors available to theater cloud diagnosis. Predictor values at forecast time 3h before development forecast time are denoted ‘‘$t - 3$,’’ at development forecast times ‘‘$t - 0$.’’

1	Vorticity, predictand deck ave., $t - 3$
2	Divergence, predictand deck ave., $t - 3$
3	Temperature, predictand deck ave., $t - 3$
4	Precipitable water, predictand deck ave., $t - 3$
5	Total cloud water, predictand deck ave., $t - 3$
6	Total falling water, predictand deck ave., $t - 3$
7	RH, predictand deck ave., $t - 3$
8	Vertical velocity, predictand deck ave., $t - 3$
9	d(theta)/d(z), predictand deck ave., $t - 3$
10	Wind speed, predictand deck ave., $t - 3$
11	Wind shear, predictand deck ave., $t - 3$
12	Vorticity advection, predictand deck ave., $t - 3$
13	Temperature advection, predictand deck ave., $t - 3$
14	3D humidity divergence, predictand deck ave., $t - 3$
15	Condens. press. deficit, predictand deck ave., $t - 3$
16	d(theta − e)/d(z), predictand deck ave., $t - 3$
17	West wind component, predictand deck ave., $t - 3$
18	South wind component, predictand deck ave., $t - 3$
19	Maximum RH within predictand deck, $t - 3$
20	RH at layer above maximum RH (see #19), $t - 3$
21	Temperature at maximum RH (see #19), $t - 3$
22	d(theta)/d(z) at maximum RH (see #19), $t - 3$
23	Mean sea level pressure, $t - 3$
24	Ground temperature, $t - 3$
25	3-h stratiform surface precipitation, $t - 3$
26	3-h convective surface precipitation, $t - 3$
27	Surface-layer wind speed, $t - 3$
28	Vorticity, high deck ave., $t - 0$
29	Vorticity, middle deck ave., $t - 0$
30	Vorticity, low deck ave., $t - 0$
31	Divergence, high deck ave., $t - 0$
32	Divergence, middle deck ave., $t - 0$
33	Divergence, low deck ave., $t - 0$
34	RH, high deck ave., $t - 0$
35	RH, middle deck ave., $t - 0$
36	RH, low deck ave., $t - 0$
37	Vertical velocity, high deck ave., $t - 0$
38	Vertical velocity, middle deck ave., $t - 0$
39	Vertical velocity, low deck ave., $t - 0$
40	d(theta)/d(z), high deck ave., $t - 0$
41	d(theta)/d(z), middle deck ave., $t - 0$
42	d(theta)/d(z), low deck ave., $t - 0$
43	Wind speed, high deck ave., $t - 0$
44	Wind speed, middle deck ave., $t - 0$
45	Wind speed, low deck ave., $t - 0$
46	Wind shear, high deck ave., $t - 0$
47	Wind shear, middle deck ave., $t - 0$
48	Wind shear, low deck ave., $t - 0$
49	Maximum RH within high deck, $t - 0$
50	Maximum RH within middle deck, $t - 0$
51	Maximum RH within low deck, $t - 0$
52	Temperature, predictand deck ave., $t - 0$
53	Precipitable water, predictand deck ave., $t - 0$
54	Total cloud water, predictand deck ave., $t - 0$
55	Total falling water, predictand deck ave., $t - 0$
56	Vorticity advection, predictand deck ave., $t - 0$
57	Temperature advection, predictand deck ave., $t - 0$
58	3D humidity divergence, predictand deck ave., $t - 0$
59	Condens. press. deficit, predictand deck ave., $t - 0$
60	d(theta − e)/d(z), predictand deck ave., $t - 0$
61	West wind component, predictand deck ave., $t - 0$
62	South wind component, predictand deck ave., $t - 0$
63	RH for level above RH max, predictand deck, $t - 0$
64	Temperature at maximum RH level, $t - 0$
65	d(theta)/d(z) at maximum RH level, $t - 0$
66	Mean sea level pressure, $t - 0$
67	Ground temperature, $t - 0$
68	3-h stratiform surface precipitation rate, $t - 0$
69	3-h convective surface precipitation rate, $t - 0$
70	Surface-layer wind speed, $t - 0$
71	Maximum RH, squared within predictand deck, $t - 0$
72	Maximum RH, fourth within predictand deck, $t - 0$
73	RH (ice) at RH maximum, predictand deck, $t - 0$
74	Lifted-cond. − dist. at RH maximum, pred. deck, $t - 0$
75	Ln(Ri − No.), pred. deck (max RH deck for total), $t - 0$
76	Sine of latitude
77	Cosine of latitude
78	Sine of longitude
79	Cosine of longitude
80	Hours of sunshine before $t - 0$
81	Hours of darkness before $t - 0$
82	Surface terrain height (9-pt ave., 1/8 mesh data)
83	Standard deviation of surface terrain height
84	Percent of surface that is water
85	Eastward gradient of terrain height
86	Northward gradient of terrain height
87	3 × 3 × 3 minimum of Ln(Ri. − Number), $t - 0$
88	3 × 3 × 3 minimum of d(theta)/d(z), deck ave., $t - 0$
89	3 × 3 × 3 maximum of vertical shear, deck ave., $t - 0$
90	3 × 3 × 3 maximum of wind speed, deck ave., $t - 0$
91	3 × 3 maximum of 3-h convective precip. rate, $t - 0$
92	3 × 3 maximum of surface layer wind speed, $t - 0$
93	3 × 3 maximum of surface speed × terrain − var., $t - 0$
94	Surface wind times terrain var., wind/stability height, $t - 0$
95	Minimum of terrain var., wind/stability height, $t - 0$
96	RH squared, predictand deck ave., $t - 0$
97	RH fourth, predictand deck ave., $t - 0$
98	RH squared, predictand deck ave., $t - 3$
99	RH fourth, predicand deck ave., $t - 3$
100	Diagnosed total cloud amount, $t - 0$ (decks only)
101	Diagnosed total no. cloud layers, $t - 0$ (decks only)
102	Predictand, RTNEPH cloud quantity

diagnosis of cloud amount. Briefly, we used four different approaches in that study. The actual choices were based in part on the study by Cianciolo (1993) and in part on experience in NWP and local forecasting. The first approach was to consider the humidity–cloud relationship; thus, a number of humidity variables (relative and absolute) were included. Recognizing the fact that NWP model dynamic variables influence the distribution of atmospheric moisture, we included a number of mass and motion variables. Third, we observe that cloud structure exhibits a high degree of complexity that would not be spatially resolved by the model, yet might be related to NWP model variables that might indicate the amplitude of such unresolved turbulence. Finally, we acknowledge that cloud distribution may be dependent upon geography and time of day, regardless of the atmospheric state. This motivated the inclusion of geographic predictors. With only minor changes, the predictors used in the present study for the diagnosis of all predictands are taken from the predictors used in the earlier study for just cloud amount. Differences between the two predictor sets were dictated by the variables available from the two different NWP models and by the fact that the present study is for a theater-scale area as opposed to an entire hemisphere.

A variation of the MOS method described above is used, in which a smaller duration time series of a two-dimensional grid of NWP model predictors makes up the predictor value pool. Each two-dimensional grid is composed of the model gridpoint forecast values either averaged vertically within, or selected as the maximum or minimum value from, all the model layers within a discrete deck (vertical subdomain) of the model troposphere. A two-dimensional grid is formed for all the predictor variables in each deck, as listed in Table 1. The corresponding predictand values on this two-dimensional grid are taken from an analysis of the predictand, formed independently of the NWP model. The time duration of this predictand–predictors spatial–temporal data series should be extensive enough to include all the types of weather one would expect to encounter in the domain of the two-dimensional grid over the time that the resulting relationships between the two are applied. The use of a spatial–temporal series loses the geographic uniqueness of the individual location, and how its local conditions may influence the predictand–predictor relationship. However, many geographic predictors were included in the predictor pool to account for location influences. The primary motivation for the approach was to allow quick adaptation of the diagnosis method to changes to the forecast model or predictand analysis method.

In this study, a 10-day period for the development of the predictand–predictor relationship was used. At the end of each day, the process ingests the analyses of the predictand over the theater-scale region that were constructed for that day. For each of the analysis times, predictor values are extracted from all NWP model forecasts with verification times corresponding to the predictand analysis time. Only one-fourth of the available grid points were considered for inclusion in the predictor value pool, and these grid points were selected using a random number generator. These predictor values are combined with the relevant geographic values for the selected NWP model grid points, and with the predictand analysis values at those grid points. The predictor–predictand sets are formed and maintained separately by NWP model forecast duration and vertical deck. Once 10 days of these predictand–predictor sets of grid values are obtained for a given forecast duration and deck, they are subjected to the statistical regression process to arrive at a relationship between the predictand and combination of predictors. Because the process of forming the predictand–predictor sets ‘‘looks back’’ to get forecast predictors, which have valid times on the present day, the relationships for the longer duration forecast times are going to be based on forecasts initialized on earlier dates than for the shorter duration forecasts.

The first statistical method used in this study was MLR. MLR minimizes the mean-squared difference between a linear combination of selected predictors and the predictand in deriving the coefficients of the relationship between them. The IMSL STAT/LIBRARY software routine called DRSTEP with a forward stepwise selection method was used. This routine first finds the predictor from the pool that, based on the development dataset values, is most highly correlated with the predictand. Then it finds the predictor that, in linear combination with the most highly correlated predictor, is most highly correlated with the predictand. The process continues to select predictors in this manner until the increase in reduction in variance with each predictor added, which is the square of the total correlation (i.e., the correlation of the linear combination of chosen predictors with the predictand), becomes very small. In an earlier study of large-scale cloud systems (Norquist et al. 1994), we found that virtually no increase in reduction of variance occurs beyond 20 predictors, so we limited the number of selected predictors to 20 in this study. In some cases, the forward stepwise regression process did not find as many as 20 predictors before zero increase in reduction of variance was reached. Once the selection process is finished, DRSTEP computes the coefficients of the linear combination of these predictors with the predictand that minimizes the mean-squared difference between them. These coefficients are later used in application to subsequent forecasts of the selected predictors to diagnose values of the predictand.

Upon application on the MLR coefficients in our earlier study (Norquist et al. 1994), we found that the frequency distribution of the predictand (in this case, cloud amount) diagnosed did not match well with the frequency distribution of the predictand verification. In particular, the MLR produced too many cases of diagnosed values in the center of the spectrum of possible values and not enough in the extremes of the spectrum. This is a result of the MLR's tendency to minimize mean-squared error. We alleviated this problem somewhat by multiplying the regression slope (the dot product of the vector of predictor values and the corresponding coefficients) by a factor of 2 to ‘‘sharpen’’ (increase the number of extreme cases) the resulting frequency distribution. This step represents a compromise between minimized mean-square error (using the computed slope) and preservation of the predictand frequency distribution (using the augmented slope).

One of the principle findings of Norquist et al. (1994) was a lack of sharpness of the MLR-diagnosed cloud amounts relative to that of the cloud amount analyses. Sharpness is defined here as the percentage of grid points in the sample with cloud amount no greater than 20% or no less than 80%. The cloud analyses used in the study tended to have a maxima of frequency of occurrence at 0% and 100% cloud amount in a sort of ‘‘U-shaped’’ distribution. The MLR coefficients, even when augmented by the slope factor of 2, tended to skew the frequency distribution of cloud amount toward a distribution with a maximum at about 50% cloud cover.

As a way of addressing this problem, a later study (Norquist et al. 1997) experimented with an alternative to MLR known as multiple discriminant analysis (MDA). When applied to the same cloud amount datasets and large-scale NWP forecast predictor sets, we found that the MDA was capable of better preserving the cloud amount frequency distribution. Thus, MDA was used as a second statistical method in the current study.

MDA operates on common datasets of predictors and the corresponding predictands as does MLR, except that the predictand must be specified in categorical form. Therefore, a continuous predictand (like cloud amount) must be assigned to discrete categories, and the category number is used as the predictand. In MDA, the predictor values are used to discriminate between categories of the predictand. Predictor values are grouped according to the predictand category at the same grid point, and the mean and variance of the predictors in each group are determined. MDA performs best when the difference between group means is greater and the in-group variance is smaller. Thus, an appropriate predictor selection method for MDA would involve a screening of predictors on the basis of their collectively maximizing the ratio of these two quantities. MDA generates coefficients of a discriminant function for each category. When the coefficients are applied to the predictor values, a probability for each category is obtained. The actual category value assigned to the diagnosis must then be chosen on the basis of the diagnosed probabilities.

The multiple discriminant analysis algorithm DSCRM from the IMSL STAT/LIBRARY is used in this study. The algorithm DSCRM does not have the facility to select the most discriminating predictors from among a longer list of available predictors. The predictors selected in the MLR forward stepwise selection process are used as input to the MDA method. Each predictand was categorized before joining the corresponding predictor values as input to MDA. The MDA process commonly could not use all of the predictors submitted to it when processing the data in a 10-day development period for a given predictand. In this case, the DSCRM algorithm would fail because of a singularity found in calculating the determinant of the predictor matrix. When this happened, the number of predictors considered were reduced by one, and iteratively called DSCRM with one less predictor, eliminating successively more highly correlated predictors one at a time as ordered by the forward stepwise selection process. If the matrix still could not be solved, the least highly correlated predictor was removed, then the iterative process was repeated. If a solution could not be found with three or more predictors, the attempt was terminated and no MDA coefficients were obtained. This was a rare event and appeared to occur only for predictands that were less likely to be physically related to the weather variables from the NWP model.

Upon application of the calculated MDA coefficients, the probability of each category is diagnosed at each application grid point. Once this was done over the entire domain grid, it was necessary to select the predictand category to be assigned to each grid point based on these category probabilities. Norquist et al. (1997) tested five different category selection methods using global-scale cloud amount predictands and corresponding global-scale NWP forecast model predictors. The method that we found that produced the best combination of cloud amount diagnosis accuracy and preservation of the predictand frequency distribution was a method we called the iterative maximum probability method. The following paragraph describes this method.

In the iterative maximum probability method, we first compute the frequency of occurrence of the predictand over the last 2 days of the 10-day development period for each predictand category. Next, the frequency of occurrence proportion of the number of domain grid points in each category is computed (for each category, NTOT = frequency of occurrence times the number of domain points). Then the categories for each grid point are ordered from highest probability to lowest probability. On the first pass through the grid points, we identify the highest probability category for each grid point, compute the difference between that probability and the next highest probability, sort these probability differences over all grid points that have the same highest probability category, and select for that category the NTOT grid points that have the largest differences. Once this is done for all categories, we take a second pass through the unassigned grid points, and considering only those categories that are not full (i.e., do not have NTOT grid points assigned to them), identify the grid points that have an unfilled category as their second highest probability. Within these unfilled categories, we compute the difference between that probability and the next largest category's probability. We sort these differences and select for that category up to NTOT grid points with the highest differences. We repeat this process until at least all but one category is filled, then assign any unassigned grid points to the remaining unfilled category.

This method has the property of imposing the frequency distribution of the predictand categories from the last 2 days of the 10-day development period upon the application diagnoses of the predictand. We found in our earlier study that this dramatically improves the match between the frequency distribution of the diagnosed predictand and the frequency distribution of the verification of the diagnoses, when compared to simply selecting the category with the highest diagnosed probability. This comes at the cost of only slightly higher rmses for the iterative maximum probability selection method. However, forcing the diagnosed predictands to have the development period's frequency distribution

TABLE 2. Cloud variable category values. Category indices are used as predictands in the MDA diagnostic scheme. RTNEPH and MLR values are converted to category indices for verification. Cloud amounts are rounded to the nearest 5% before conversion.

Variable	1	2	3	4	5	6
Amount (%)	0	5–20	25–40	45–60	65–80	85–100
Type						
Low	Cb	St	Sc	Cu		
Middle	As	Ns	Ac			
High	Cs	Cc	Ci			
Base altitude (m AGL)						
Low	0–396	397–792	793–1188	1189–1584	1585–1980	
Middle	1980–2590	2591–3199	3200–3809	3810–4418	4419–5028	
High	5028–6022	6023–7017	7018–8011	8012–9006	9007–10 000	
Thickness (m)						
Low	0–1600	1601–3200	3201–4800	4801–6400	6401–8000	
Middle	0–800	801–1600	1601–2400	2401–3200	3201–4000	
High	0–800	801–1600	1601–2400	2401–3200	3201–4000	
Number of layers	0	1	2	3	4	
Ceiling altitude (m)	0–2160	2161–4320	4321–6480	6481–8640	8641–10 800	

can run the risk of creating error in the diagnoses if the actual frequency distribution on the application day is dramatically different than it is on the last 2 days of the development period.

3. Data

The Air Force RTNEPH cloud analysis dataset (Hamill et al. 1992) was used as the source for the cloud variable predictands. The primary reasons for this choice were the availability of all required cloud variables, the relatively fine spatial resolution (47.625 km at 60° lat on a polar stereographic grid, known as one-eighth mesh), and its ready availability. The RTNEPH is known to have limitations in the accuracy of its specification of several of the cloud variables. For example, cloud-base altitudes deduced from satellite imagery are derived by subtracting a default thickness (based on cloud type) from the estimated cloud-top altitude. These realism limitations reduce the degree of physical relationship that the RTNEPH cloud data would otherwise have with noncloud atmospheric variables, and thus would be expected to influence the accuracy of the diagnostic method. If the diagnosis truly does have skill in relating realistic cloud and noncloud variables, one would expect the diagnosis skill to improve with increasing realism of the predictand.

At each RTNEPH grid point, the following data are included: total cloud amount, number of cloud layers, and, for up to four such layers, the amount, type, base altitude, and top altitude of each cloud layer. An additional piece of information used in this study was the time of the cloud measurement with respect to the analysis time. In this study, we used only data at grid points when this time differential was no more than 2 h.

The layer clouds in the RTNEPH database are assigned to cloud type categories according to their cloud-base altitude (USAFETAC 1991). Layer clouds with bases 0–1980 m above ground level (AGL) are considered low clouds and are assigned one of four cloud types: cumulonimbus, stratus, stratocumulus, or cumulus. Layer clouds with bases 1981–5028 m AGL are classified as middle clouds and are designated as altostratus, nimbostratus, or altocumulus. High clouds are those with bases above 5028 m AGL and are typed as cirrostratus, cirrocumulus, or cirrus.

With the exception of cloud type and number of cloud layers, the rest of the cloud variables are continuous, or noncategorical in nature. The range of values of each of these variables must be segmented into categories to be used in the MDA diagnosis method. The categories to which all cloud variables used in this study were assigned are listed in Table 2. The six cloud amount categories used were determined (Norquist et al. 1997) to be the best compromise between minimizing the range covered by each category and maximizing the skill of diagnosis. Base altitude and thickness categories were specified by dividing the full range of possible values into an arbitrarily chosen five categories.

The RTNEPH datasets for January and July 1992 in the Northern Hemisphere were provided by the Air Force Combat Climatology Center in Asheville, North Carolina. A midlatitude theater centered on northern Italy was extracted. The chosen theater consisted of 40 RTNEPH grid points on a side. The 40 × 40 gridpoint set was extracted from the RTNEPH dataset at 0000, 0600, 1200, and 1800 UTC for each day of both months, with ''missing'' flags set on for untimely (more than 2 h old) grid points. These extracted datasets were used for development and verification of the methods described in the previous section. They will be referred to hereafter as the theater RTNEPH cloud data.

Among the accuracy limitations of the RTNEPH is the inexactitude of the cloud-base altitude specification of the layer clouds. For this reason, each reported layer cloud at a given RTNEPH grid point was assigned to a

Table 3. RTNEPH frequency of occurrence (percent) of number of cloud layers with bases in each deck and in all decks for Jan and Jul 1992.

	Jan					Jul				
Deck	0	1	2	3	4	0	1	2	3	4
High	88.1	11.8	0.1	0.0	0.0	83.7	16.1	0.2	0.0	0.0
Middle	76.5	22.2	1.3	0.0	0.0	77.7	21.0	1.3	0.0	0.0
Low	47.5	45.8	6.2	0.4	0.0	56.3	37.7	5.5	0.5	0.0
All	34.9	39.9	19.6	5.0	0.6	44.6	29.2	18.8	6.2	1.2

deck (low, middle, or high) according to its cloud-base altitude. The fixed altitude (AGL) boundaries of the decks were identical to those for the cloud type categories. The one-eighth mesh terrain elevation datasets created by Schaaf et al. (1990) were used to convert reported cloud-base altitude (above mean sea level) to base altitude AGL.

Once all layer clouds were assigned to decks, the frequency distribution was computed by month for January and July 1992 theater RTNEPH cloud datasets. The frequency distribution of the number of cloud layers with bases in each deck is given for both months in Table 3. From the results, it was clear that more than two layers in a deck is a rare occurrence and that two layers in the same deck occurs never more than about 6% of the cases. Both months have more than 1% occurrence of two layers in both middle and low cloud decks.

Based on these results, the forecast diagnosis procedure was designed to develop and apply predictand–predictor relationships for two cloud layers in the middle and low decks, and just one layer in the high deck. In the middle and low decks, all reported cloud layers in each deck are considered, and the cloud layer with the greatest cloud amount (or, if two layers have the same amount, the one with the higher base altitude) is assigned to the ‘‘primary’’ layer for that deck. The layer with the second greatest cloud amount was assigned to the ‘‘secondary’’ cloud layer. When only one layer was reported, a value of zero cloud amount was assigned to the secondary layer. Since the presence of two or more layers in any of the three decks is rare, the secondary cloud layer is dominated by cases of zero cloud amount. However, it was necessary to carry a second layer in the lower two decks to account for their occurrence in the predictand database and to allow for infrequent occurrences of a total of four (or even possibly five) cloud layers.

We used the Pennsylvania State University–National Center for Atmospheric Research (NCAR) MM5 mesoscale numerical weather prediction model (Grell et al. 1994) to generate the weather variable predictors. MM5 has been chosen as the operational theater weather prediction model by the U.S. Air Force. We set up the model to run in hydrostatic mode, with an inner nest positioned on the one-eighth mesh theater grid points. The outer nest has a grid spacing three times that of the inner nest. We set the model's vertical coordinate system at 23 sigma levels, where sigma is defined by $(p - \mathrm{ptop})/(\mathrm{psfc} - \mathrm{ptop})$. Here the surface pressure, psfc, was derived from the analysis of pressure levels using the gridpoint terrain heights, and ptop = 50 hPa. The sigma levels were set at 0.05 intervals from sigma = 0.0 to sigma = 0.95, then sigma = 0.975, 0.990, and 1.0. The version of the MM5 used to generate the weather variable data for this study had the following configuration: mixed-phase ice processes, explicit moisture, cloud–radiation scheme, moist vertical diffusion in clouds, upper-radiative boundary layer, water vapor plus explicit cloud plus rain prediction, Grell (1993) cumulus parameterization scheme, surface heat and moisture fluxes, ground temperature variation, 13 land-use categories, cloud effects on surface radiation, stability-dependent drag coefficients, snow cover effects, evaporative cooling, shallow convection, and two-way nesting. Initial and boundary conditions for the MM5 forecast integrations were extracted from the global medium-range forecast (MRF) model of the National Centers for Environmental Prediction, acquired from NCAR on 2.5° lat × 2.5° long grids. Rawinsonde and surface observation data were used in a successive corrections analysis procedure to create gridded initial conditions, where MRF forecasts served as the ‘‘first-guess’’ for the analyses. Initialization consisted of removal of the vertically integrated mean divergence before forecast integration began. We used 5-min time steps in the forecast integration.

The NWP model-based predictors were derived for each grid point according to the following steps. We determined the top and bottom sigma layers for each deck by first computing the geopotential height and temperature of the model sigma levels. These were then used hydrostatically to compute the pressure corresponding to the cloud deck altitude (where the one-eighth mesh terrain elevations were added to AGL deck base altitudes) for each deck. This was then converted to sigma value. Stepping down from the highest sigma layer, the first sigma layer midpoint pressure found to be greater than the deck base sigma was considered the topmost sigma layer in the deck below, and the sigma layer above this is taken as the lowest sigma layer in the deck above. The topmost sigma layer used in the high cloud deck had a sigma level top of 0.20, 0.25, and 0.30 for latitudes 0°–20°, 20°–65°, and 65°–90°, respectively. We used the top of the lowest sigma layer as the low deck base at all grid points. For both the forecast duration being considered for the predictand–predictor relationship and 3 h prior, the sigma layer prognostic variables (except the moisture variables) were averaged (weighted by sigma layer depth) over each deck. The respective precipitable water values were computed for the prognostic moisture variables. In the latter computations, rain and snow water were combined (''total falling water'') and liquid and ice cloud water were combined (''total cloud water''). In addition to the model's prognostic variables, a number of derived variables were computed for each sigma layer and then averaged over the respective cloud deck sigma layers. Finally, several predictors were individual sigma layer values, selected because they represented the maximum value of the sigma layer values for their respective deck.

The forward stepwise regression selected the leading predictors (at most, 20) in each 10-day development period separately for each predictand in each deck (two predictands for each cloud mission impact variable in middle and low decks). These were provided to MLR and MDA for computation of the coefficients of the linear combination of selected predictors. One set of coefficients is generated for each predictand in MLR, while a set of coefficients for each predictand category is computed in MDA. These coefficients are then used in linear combination with subsequent forecast gridpoint values of the deck predictors to compute each forecast state predictand value directly in MLR, and the probability of each category of the predictand in MDA. The previously described category selection method, known as iterative maximum probability, is then used to determine the category selected for each grid point.

In preliminary results, some performance characteristics that could be improved by simple modifications to the methods were found. First, it was found that total number of layers was much better diagnosed directly by MDA (operating within the MLR algorithm) than by summing the number of cloudy layers diagnosed by MLR. The latter had a significant positive bias, reflecting the tendency of MLR to produce too many partial cloudy occurrences. Therefore, the MLR algorithm was modified to reduce the number of diagnosed cloud layers to the number directly diagnosed by MDA. The cloud-layer ''thinning'' was done considering the probability of nonzero cloud amount in each of the layers. They were ordered from least to most probable, based on the frequency of occurrence of cloud amount in the dependent predictand data over the 10-day development period. Taking them in order, the modified algorithm zeroed out the clouds in each layer until the number of layers diagnosed by MDA was achieved. Second, it was found that in some 10-day development periods, the MDA library software failed to find at least three satisfactory predictors (a minimum I set) out of the 20 offered to it by the forward stepwise selection method. In this case, no cloud variables were allowed to be diagnosed for those layers. When either cloud amount or base altitude could not be diagnosed by MDA due to too few satisfactory predictors found, the modified algorithm does not allow any cloud variables to be diagnosed for any layer in the deck in question—all variables are set to missing, and the cloud-base altitude is set to its midvalue for that deck to place the missing values in the correct deck in the output data. When thickness cannot be diagnosed but type can, the modified algorithm assigns the default thicknesses corresponding to the diagnosed type, as given by Hamill et al. (1992). When type cannot be diagnosed, it is set to missing.

4. Verification methods

In a quantitative evaluation of the results, it was necessary to use predictand category as the basis for evaluation. The reason for this is that the MDA diagnosis method can specify only categories of the predictand, where each category contains a range of values (see Table 2). The category containing each MLR diagnosed value and RTNEPH specified value was identified, and that category value was assigned to the grid point before computing the verification statistics. Thus, all skill score results presented below are based on verification of predictand category values from MLR and MDA predictions against RTNEPH predictand categories (at the forecast valid time) as a reference. Scores over all twice-daily cloud predictions initialized on 13–22 January 1992 (referred to as the January period) and over the cloud predictions initialized 13–22 July 1992 (the July period) were computed. To provide a standard for skill comparison, a persistence forecast of predictand categories (from RTNEPH at the forecast initial time) was also verified against RTNEPH.

For cloud amount verification, the mean error (ME), rmse, and percent predicted correctly (PPC, percent diagnosed in the correct category) were computed. A fourth verification measure was the degree to which the categorical frequency distribution of the diagnosed clouds matched that of the verification. I call this measure the ''frequency distribution fit (FDF),'' given by the expression

$$\mathrm{FDF} = (1/N)\sum_{g=1}^{G} |(m_g/n_g) - 1|n_g,$$

where g is the category index of G categories, m_g is the number of grid points diagnosed to be in category g, n_g is the number of grid points observed in that category (the verification points), and N is the total number of verification points. An FDF value of zero represents the perfect representation of the verification's categorical frequency distribution.

The verification computations for cloud amount included all grid points where timely data were available in the RTNEPH. MLR and MDA cloud amounts were verified over the same set of grid points. Persistence (RTNEPH at the initial time) cloud amounts were verified over fewer points, because now both subject and reference had missing data. To see the effects of the different verification sample between the methods and RTNEPH, a verification of MLR and MDA was conducted over only those points where both the reference and initial time RTNEPH had nonmissing data. The results were not significantly different in skill measure, and the sample size was reduced significantly. Therefore, all available grid points in the verification of each method were used.

Objective verification of the other layer cloud variables was less straightforward. It is not clear how one can verify cloud type, base altitude, and thickness over all grid points (that have timely RTNEPH data) when some grid points are clear in a given cloud deck. In this study, I reasoned that the cloud amount verifications accounted for the degree of skill in clear versus cloud predictions, so that the accuracy of the cloud characteristics where they are present is what yet needs to be determined. Therefore, skill for cloud type, base altitude, and thickness was computed only over grid points where both the diagnosis to be verified (MLR, MDA, and persistence) and the verifying RTNEPH indicated nonzero cloud amount in a given cloud deck. This approach reduces the verification sample of grid points below the total number of timely grid points in the RTNEPH reference. In the results, the cloud type, base altitude, and thickness categorical skill score values are given only in decks where the number of grid points in the verification was at least 5% of the total number of timely RTNEPH grid points (i.e., the number of grid points used in the cloud amount verifications). This often excludes the high deck scores from being listed, because of the high frequency of occurrence of zero cloud amount in the high deck. In this paper, skill scores are shown only for the primary layers of the middle and low decks.

To the operational user of cloud predictions, it is vital that the predictions be displayed in some graphical manner that is easily understood and used. To date, we have experimented with graphical displays of total cloud amount in a two-dimensional visualization. For the purposes of this paper, such displays are useful for a subjective evaluation of the skill of the cloud diagnosis methods. We display total cloud amount predictions from MLR and MDA with forecast valid time RTNEPH as validation. We are aware of other studies (e.g., Albers et al. 1996) that have produced three-dimensional visualizations of gridded cloud analyses. The cloud predictions produced in this study would lend themselves well to such graphical rendering.

5. Results

The cloud diagnosis methodology described in section 2 was used to develop relationships between the cloud variables as predictands and the NWP model forecast variables and geographic variables as predictands. The 10-day development periods used in this study were 3–12 through 12–21 January and July 1992. The relationships developed were then applied to forecasts initialized on 13–22 January and July 1992 to produce the cloud ‘‘predictions’’ for these 10 days. Each 10-day development period was applied to the two (0000 and 1200 UTC initial times) forecasts initialized on the day after the development period. For the 6-h forecast cloud diagnosis, the development period initial times fell on the 10 days just prior to the application initial times. For the 12-, 18-, 24-, and 30-h forecasts, the development period initial times fell on the 10 days prior to the day before the day of the application initial times. For the 36-h forecast, there is a 2-day gap between the last initial time of the development forecasts and the first initial time of the application forecasts. For example, application times of 0000, 1200 UTC 13 January 1992 used relationships developed over 0000, 1200 UTC initial times on 3–12 January for the 6-h forecast diagnosis; 2–11 January for the 12-, 18-, 24-, and 30-h diagnoses; and 1–10 January for the 36-h diagnosis. This approach takes into account the fact that a particular forecast duration cannot be used for development until the forecast valid time RTNEPH cloud analysis is available to supply the predictands.

Once the cloud variables were diagnosed for each of the 10 days in the January and July verification periods, we verified the results using the methods described in section 4. The following sections describe the results of both the subjective and quantitative verifications for both months.

a. 13–22 January 1992

In Table 4, we list the leading predictors selected by the forward stepwise regression selection process for the ten 10-day development periods extending from 3–12 to 12–21 January 1992 for 6-h forecast cloud amount diagnosis. The predictors listed in the table as ‘‘leading’’ resulted from assigning points according to the order in which the predictors were selected in each development period, then adding these points over all 10 development periods. The resulting sums determined the rank as listed in the table. There was a significant variation in predictors selected during the course of this set of 10-day development periods. It was found that some predictors that figured prominently in the development relationships in the first few development periods would not appear in later development periods, while new predictors would emerge. Other predictors were strong contributors in all development periods. The predictors change because the cloud scene in a theater changes, and different cloud characteristics lead to the selection of different predictors in a theater-scale area. Cloud distribution in one 10-day training period can look significantly different from that of a later one. A gradual change was observed in predictors from development period to development period as the cloud characteristics changed with time in the theater.

TABLE 4. Leading cloud amount predictors for each cloud deck for the 10 10-day development periods from 3–12 to 12–21 Jan 1992, for the 6-h forecast ($t - 3$ is value 3 h before 6-h forecast time). Deck name is given in description when predictor selected is not from the same deck. Here RH = relative humidity, $d\theta/dz$ = vertical potential temperature gradient, ↑ = above, lvl = level, press. = pressure, cld = cloud, diag. = diagnotic, and sfc = surface.

Rank	High	Middle	Low	Total
1	Zonal wind	Max. RH, high	Diag. total cld. amt.	(Max. RH)^4
2	RH^2, $t - 3$	(Max. RH)^4	Max. RH, $t - 3$	Zonal wind
3	Meridional wind	Mean sea lvl. press.	Meridional wind, $t - 3$	$d\theta/dz$, $t - 3$
4	(Max. RH)^4	Diag. total cld. amt.	Zonal wind	RH^2
5	Wind shear, low	$d\theta/dz$, max. RH lvl.	RH, high	Sin (long)
6	Mean sea lvl. press.	Zonal wind, $t - 3$	RH ↑ max. RH lvl.	Max RH, high
7	Precipitable water	Max. RH, low	Max. RH, high	RH ↑ max. RH lvl.
8	Total cld. water, $t - 3$	RH, high	3 × 3 max sfc wind	Cos (lat)
9	Max. RH	Temperature	Max. RH, middle	Wind shear, low
10	RH	Min. sea lvl. press., $t - 3$	North elev gradient	RH, high

One should not try to attach physical meaning to derived statistical relationships. There are shortcomings in both the NWP model and the cloud analysis that will mask true physical relationships. For example, hours of sunlight before prediction time could be selected as a predictor simply because of the limitations of observing clouds at night. The statistical approach is an attempt to overcome these dataset limitations. In statistical forecasting, the regression methods determine the choice of the predictors used.

Table 5 lists the verification skill scores for cloud amount in the three decks, and total cloud. The total number of grid points possible for verification are the 39 × 39 theater grid points at each diagnosis time, times 20 diagnosis times, or 30 420 points. In the verification of MLR and MDA, in which a cloud amount is diag-

TABLE 5. Cloud amount category verification statistics for predictions initialized twice daily, 13–22 Jan 1992. Units for ME, rmse ~ 20% cloud amount.

	6-h	12-h	18-h	24-h	30-h	36-h	6-h	12-h	18-h	24-h	30-h	36-h
			ME						PPC			
			High						High			
MLR	−0.13	0.30	−0.15	−0.27	−0.16	−0.27	86.4	46.6	87.0	83.7	86.1	83.4
MDA	−0.02	0.04	−0.01	0.01	−0.02	0.06	85.9	77.6	85.5	79.1	84.7	76.1
Pers.	0.11	0.01	0.10	−0.01	0.08	−0.04	82.6	78.6	80.6	78.6	79.7	76.9
			Middle						Middle			
MLR	−0.22	0.20	−0.27	−0.36	−0.31	−0.42	60.8	37.7	58.1	67.7	55.3	68.1
MDA	−0.12	−0.02	−0.19	−0.13	−0.21	−0.13	61.7	69.5	60.1	70.2	57.5	69.0
Pers.	−0.31	−0.07	−0.35	−0.05	−0.41	−0.09	67.8	71.9	64.2	71.3	61.5	68.3
			Low						Low			
MLR	−0.11	0.42	−0.22	−0.22	−0.20	0.12	35.3	28.1	30.6	37.1	27.8	33.8
MDA	−0.24	0.16	−0.18	0.20	−0.13	0.29	43.3	47.6	41.5	44.5	36.1	40.8
Pers.	−0.47	−0.01	−0.46	0.60	−0.48	0.05	53.2	51.6	46.6	50.4	42.0	42.6
			Total						Total			
MLR	−0.12	−0.02	−0.03	−0.22	−0.09	0.02	34.7	37.3	26.8	30.5	24.6	29.1
MDA	−0.21	0.08	−0.23	0.14	−0.20	0.25	46.5	47.0	44.0	43.0	38.4	41.3
Pers.	−0.22	−0.05	−0.22	0.00	−0.27	−0.04	52.7	48.2	46.4	46.2	40.3	40.5
			Rmse						FDF			
			High						High			
MLR	0.89	1.21	0.90	1.10	0.92	1.12	0.09	0.87	0.09	0.13	0.09	0.15
MDA	1.09	1.43	1.13	1.40	1.16	1.51	0.02	0.03	0.02	0.02	0.02	0.05
Pers.	1.21	1.37	1.29	1.39	1.35	1.48	0.08	0.01	0.07	0.01	0.07	0.02
			Middle						Middle			
MLR	1.77	1.58	1.83	1.58	1.92	1.63	0.18	0.95	0.20	0.21	0.25	0.23
MDA	1.90	1.91	1.97	1.90	2.08	1.93	0.06	0.02	0.08	0.07	0.09	0.07
Pers.	1.83	1.78	2.03	1.82	2.11	1.94	0.19	0.03	0.21	0.02	0.25	0.05
			Low						Low			
MLR	2.29	2.27	2.46	2.27	2.57	2.29	0.45	0.73	0.57	0.43	0.56	0.50
MDA	2.44	2.43	2.57	2.59	2.79	2.69	0.12	0.08	0.11	0.09	0.08	0.14
Pers.	2.29	2.31	2.55	2.40	2.71	2.67	0.19	0.01	0.20	0.03	0.21	0.03
			Total						Total			
MLR	1.95	2.00	2.01	2.14	2.30	2.29	0.66	0.51	0.92	0.63	0.84	0.61
MDA	2.24	2.30	2.41	2.42	2.69	2.56	0.12	0.05	0.14	0.07	0.12	0.13
Pers.	2.10	2.15	2.38	2.26	2.59	2.50	0.16	0.03	0.17	0.02	0.20	0.02

nosed at every grid point, the number of verification grid points is limited only by the number of timely verifying RTNEPH grid points. The percentage of timely points was about 69% for 6-, 18-, and 30-h forecasts (0600 and 1800 UTC times) and 59% for 12-, 24-, and 36-h forecasts (0000 and 1200 UTC). In verification of persistence, only the timely points common to both the persisted and verifying RTNEPH could be used. The percentage of timely points common to both averaged about 43% over all six forecast times.

The ME skill scores indicate that the MLR and MDA methods do not generally produce a significant bias in the diagnosed cloud amounts. In the three decks, MDA appears to produce a lower ME than MLR while MLR does better than MDA in total cloud amount. While MLR cloud amount diagnosis does in some cases generate a bias over 0.3 (about 6% cloud amount), there is no discernible trend of bias with forecast duration. Neither MLR nor MDA show nearly the variation of ME with forecast duration as is evident in the persistence scores.

We see that in the rmse scores, MLR consistently demonstrates the best skill of the three methods. We expect that MLR should produce better rmse skill than MDA, since it is designed to minimize mean-squared error in its estimate. There does not appear to be any noticeable growth in the rmse of MLR with forecast duration, at least for the three decks. There is such an increase in rmse with forecast duration in persistence, though it is not monotonic. The rmse skill of the MDA method does not beat persistence, but at best can be considered only competitive. The increase in rmse going from high cloud to low cloud in all three methods reflects the decrease in frequency of occurrence of zero cloud amount from high to low cloud. Thus, with a greater likelihood of zero cloud amount, a diagnosis of zero cloud amount at any grid point is more likely to be correct.

This same trend of decreasing skill with lower cloud deck is seen as also in the PPC scores. In this score, MLR appears to slightly outperform MDA in the high deck (except in 12-h forecasts), while MDA performs better than MLR in the other two decks and total cloud. In the high deck, persistence is slightly less skillful than MLR and MDA, but is good as or better than either method in middle, low, and total cloud. Persistence shows a noticeable decline in skill with forecast duration (except in the middle deck), but MLR and MDA do not.

In the measure of the agreement between the category frequency distribution between each method and the reference (FDF), MDA significantly betters MLR consistently in all decks and forecast durations. This is to be expected, since the category selection method used in association with MDA was designed to reproduce the frequency distribution of the last 2 days of the predictand data. Therefore, the MDA FDF values are a measure of the change of category frequency distribution between the last 2 days of the development period and the subsequent application day. The higher FDF values for MLR reflect the tendency of MLR to minimize mean-squared error at the expense of losing the sharpness of the frequency distribution. As will be seen in the subjective evaluation, MLR tends to produce too many occurrences of partial cloudiness and too few occurrences of clear or overcast.

The MLR skill scores for 12-h forecasts are worse than at other forecast durations in the high and middle decks. Upon investigation, it was found that the MDA diagnosis of total number of cloud layers, operating in conjunction with the MLR cloud diagnosis algorithms, failed to find a sufficient number of predictors in several of the 10-day development periods. Therefore, there was no governing value of total number of cloud layers to perform the ‘‘thinning’’ of cloud layers diagnosed by MLR. This resulted in superfluous cloud layers being left in the diagnosed cloud data, which resulted in worse scores in the deck cloud amounts. The fact that too much deck cloudiness was diagnosed at 12-h forecast times by MLR is reflected in the positive values of deck ME, in contrast to the negative values at all other forecast durations.

In an attempt to determine the statistical representativeness of the verification statistics shown, the development and application of the methods was continued for the next 8-day period (23–30 January 1992). The verification statistics for this period were very similar in trend and nature to those for the original 10-day period, with the exception of the ME, which were less negative. This similarity exists in spite of the fact that the cloud characteristics of the two periods were quite different. In Fig. 1 the 10 × 10 gridpoint average of RTNEPH total cloud amount for the 18-day period for four theater subareas is shown. Two of the subareas increase in clouds in the first 10 days then decrease in the last eight, while one subarea has just the opposite trend, and the fourth grows and stays high in cloudiness. The stability of the statistics between the two periods in light of the large differences in cloud characteristics suggest the Table 5 statistics are representative of a greater duration of verification.

Although similar skill scores were calculated for the diagnoses of cloud type, base altitude, and thickness, the scores are not shown. Neither the MLR nor MDA produced better cloud diagnoses than persistence for these variables, as measured by rmse and PPC. The MLR and MDA diagnoses had small values of ME, and the MDA diagnoses had consistently small values of FDF.

There are two ways to diagnose the number of cloud layers at a grid point using these methods. First, MDA diagnoses total number of cloud layers as a separate predictand, using the same predictors as for total cloud amount. Second, once either MLR or MDA diagnoses cloud amount for all decks over the entire theater, one can count the number of layers diagnosed at each grid point. As was mentioned previously, the direct diagnosis

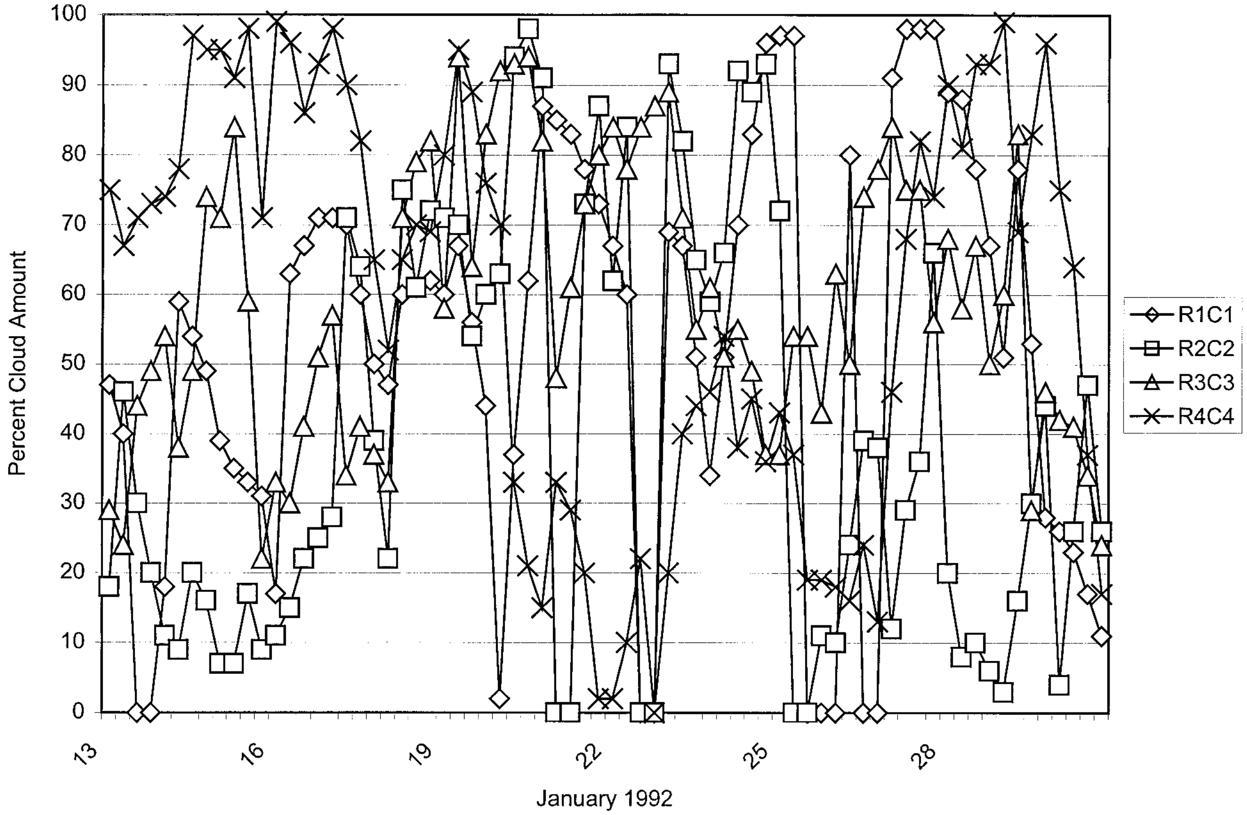

FIG. 1. RTNEPH total cloud amount (%) averaged over four 10 × 10 gridpoint theater subareas during the period 13–30 Jan 1992. Subarea indices refer to row (R, south–north) and column (C, west–east) of the sixteen 10 × 10 gridpoint subareas in the theater.

of number of cloud layers by MDA was imposed upon the MLR cloud amount diagnosis, because it was clearly superior in skill to the count of cloud layers diagnosed by MLR. In comparing the direct diagnosis of number of layers with the count of layers diagnosed by MDA, it was found that the verification scores were very similar. They were also competitive with the scores for the number of cloud layers resulting from the verification of persistence.

The algorithms also give two separate estimates of cloud-ceiling altitude. For both MLR and MDA, cloud-ceiling altitude (category for MDA) is diagnosed directly as a separate predictand, using the same predictors as for total cloud amount. In addition, for both methods, the ceiling altitude is taken to be the base altitude of the lowest diagnosed layer with greater than 50% cloud cover. Table 6 shows the category verification skill scores for all four methods and persistence as a reference. The directly diagnosed cloud-ceiling altitudes have a lower ME than those deduced from cloud-base

TABLE 6. Cloud-ceiling altitude category verification statistics for predictions initialized twice daily, 13–22 Jan 1992. Regular script MLR and MDA represent directly diagnosed ceiling altitude. Italicized MLR and MDA represents ceiling altitude assigned as base altitude of the lowest cloud layer with cloud amount greater than 50%. One unit for ME and rmse is ~2160 m.

	6-h	12-h	18-h	24-h	30-h	36-h	6-h	12-h	18-h	24-h	30-h	36-h
	ME						PPC					
MLR	−0.01	0.06	0.03	−0.01	−0.02	0.09	78.9	63.9	72.4	69.3	76.8	61.3
MLR	−0.12	−0.26	−0.07	−0.20	−0.11	−0.25	84.9	75.9	79.0	78.5	81.8	79.1
MDA	0.00	−0.01	−0.01	0.00	−0.03	0.02	79.8	67.4	78.0	66.5	78.2	65.9
MDA	−0.05	−0.17	−0.03	−0.12	−0.02	−0.09	81.8	72.3	78.6	69.9	77.7	69.9
Pers.	0.18	0.03	0.21	−0.03	0.19	−0.05	76.9	72.9	72.4	72.7	71.1	65.9
	Rmse						FDF					
MLR	0.56	0.99	0.62	0.85	0.59	0.94	0.01	0.16	0.08	0.14	0.02	0.29
MLR	0.50	0.88	0.55	0.80	0.55	0.81	0.17	0.33	0.09	0.22	0.16	0.24
MDA	0.65	1.09	0.67	1.09	0.66	1.12	0.03	0.05	0.04	0.04	0.06	0.07
MDA	0.59	0.93	0.66	1.02	0.69	1.01	0.08	0.20	0.10	0.17	0.09	0.13
Pers.	0.81	1.00	0.94	1.09	0.96	1.20	0.20	0.02	0.20	0.04	0.16	0.06

TABLE 7. Cloud amount and cloud-ceiling altitude category verification statistics for 18-h predictions initialized twice daily, 13–22 Jan 1992. Regular verification uses forecast valid time RTNEPH as reference. Random verifications randomly select forecast initial time and RTNEPH valid time from 10-day period.

	High	Middle	Low	Total	Ceiling	High	Middle	Low	Total	Ceiling
			Rmse					PPC		
MLR reg.	0.90	1.83	2.46	2.01	0.55	87.0	58.1	30.6	26.8	79.0
MLR ran.	1.01	2.23	3.01	2.80	0.62	83.8	48.9	20.5	18.3	78.2
MDA reg.	1.13	1.97	2.57	2.41	0.66	85.5	60.1	41.5	44.0	78.6
MDA ran.	1.29	2.46	3.24	3.13	0.72	81.8	47.5	26.1	29.3	76.1

altitude in both MLR and MDA, although none of the error levels are excessive. On the other hand, the deduced cloud-ceiling altitudes have a lower rmse than their diagnosed counterparts. Deduced MLR cloud-ceiling altitudes are consistently the best of the five methods. Unlike the cloud-layer quantities, it appears that both types of MLR and MDA estimates of cloud-ceiling altitude show skill relative to persistence. MDA-diagnosed ceiling has a slightly better PPC score than MLR-diagnosed ceiling. MLR and MDA both show more PPC skill in ceilings deduced from cloud-base altitude than diagnosed directly. As in rmse, ceilings deduced from MLR cloud-base altitudes show the most skill consistently over all forecast durations. As expected, MDA-diagnosed ceilings have the best FDF values overall among the methods, and both MDA- and MLR-diagnosed ceilings have better FDF score than their deduced counterparts.

The performance of ceiling appears to be better in absolute score and with respect to persistence than that of cloud amount or base altitude. As can be seen in Table 2, the altitude categories for cloud ceiling are much coarser (have a greater depth) than for base altitude. The verification scores are by category, and it is easier to hit the correct category when the categories are wider. Thus, the absolute scores are better for ceiling than for base altitude. As for cloud amount, Table 5 indicates that there was a slight tendency on the part of MLR and MDA to overestimate the number of clear cases. These clear cases are not included in the verification of ceiling, because the verification sample is limited to grid points where both the method and RTNEPH had greater than 50% cloud amount. If we limited the cloud amount verification to cases where cloud amount for both the method and RTNEPH were greater than 50%, the skill levels would likely improve.

To provide further information on the level of skill demonstrated by the diagnostic methods, the following additional verification was performed. A random number generator was used to select both the diagnosed cloud forecast initial time and the reference (RTNEPH) valid time from the 20 twice-daily synoptic times being verified. The idea here is to determine how much greater skill is shown by verifying each forecast diagnosis against its correct valid time reference compared to a completely random match between forecast valid times and reference times. Table 7 shows the results of this random verification against the regular verification results for cloud amount and ceiling altitude as taken from Tables 5 and 6. As can be seen, the cloud amounts show significant skill with respect to the random verifications for the cloud amount in all four decks, with less significant skill compared to random verification for ceiling. The random verification was also conducted for cloud type, base altitude, and thickness, and the results were compared with the regular verification skill scores. The comparison showed that diagnosis skill for these variables was no greater than the skill scores calculated from the random verifications.

For the subjective evaluation of the diagnosed cloud forecasts, the diagnosed and RTNEPH total cloud amount for the two synoptic times in the 10-day verification period that had the fewest untimely points in the RTNEPH were displayed. A short duration forecast for one case and a longer duration forecast for the other were shown. The graphical-depictions of these two cases are shown in Fig. 2.

In Fig. 2a, we see that MLR and MDA 6-h forecast diagnoses placed the overcast areas in the same locations, in fairly good agreement with the RTNEPH. Both MLR and MDA fail to reproduce the small clear area in the midst of the overcast area in the northeastern portion of the theater. The generally clear area in the southwestern portion of the theater depicted in the RTNEPH image is too cloudy in the MLR diagnosis. The MDA diagnosis does a better job of representing the clear area but fails to pick up the small patches of cloud that break up the clear area in the RTNEPH. As compared with the RTNEPH, the MDA gradient is much more realistic than that of the MLR diagnosis. The flat gradient of the MLR is a symptom of MLR's tendency to skew the cloud amount frequency distribution toward partial cloudiness, to minimize mean-squared error.

In Fig. 2b, MLR and MDA 30-h forecast diagnoses do not capture enough of the overcast region dominating most of the RTNEPH image. As a result, the clear and scattered cloud areas in the northern portions of the two diagnoses are too extensive. Because MLR tends to underpredict the cloud amount category extremes, it produces less of the RTNEPH overcast than does MDA. The RTNEPH depicts sharp boundaries between the overcast and clear areas. This is captured better by the MDA than by MLR.

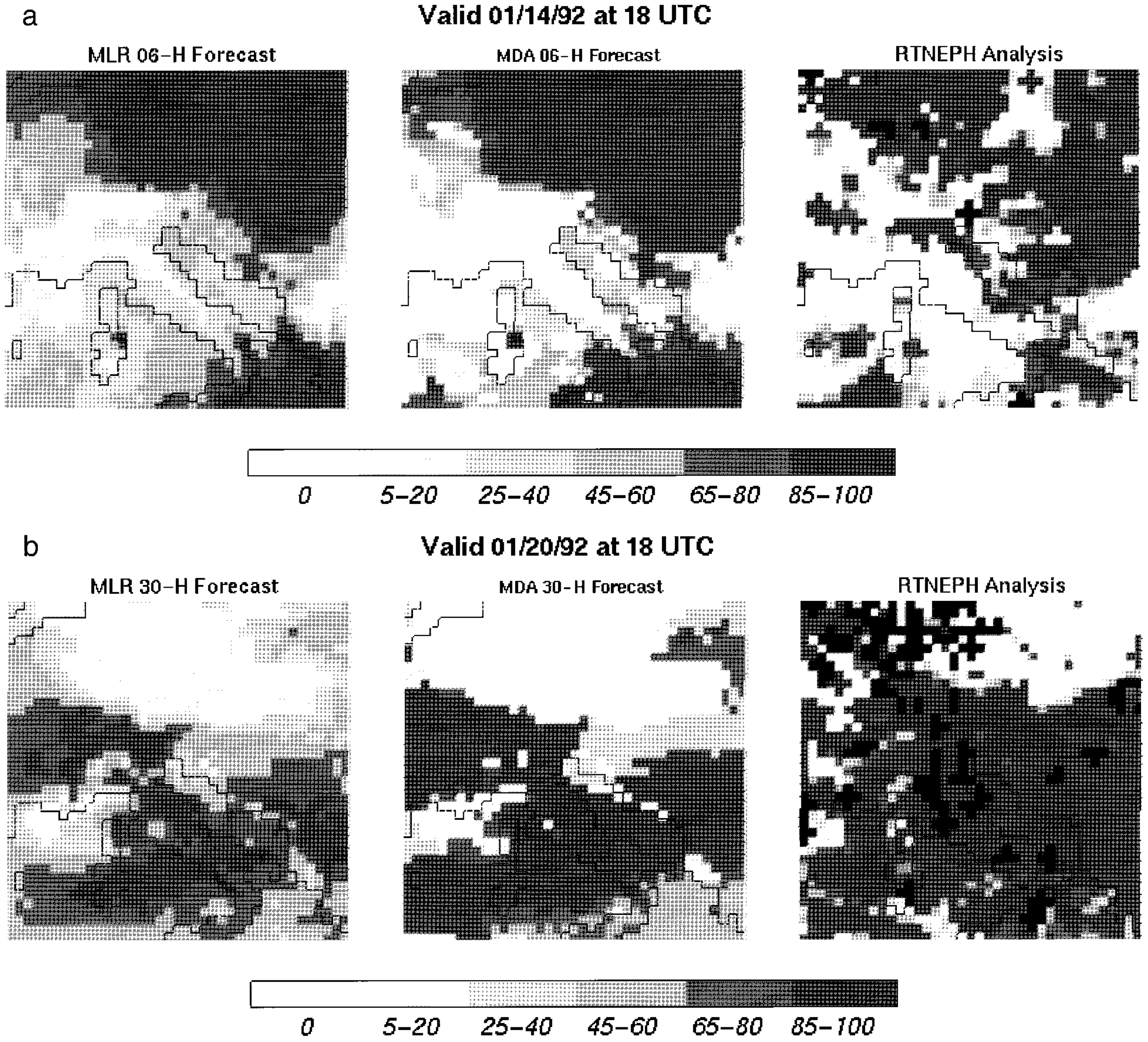

FIG. 2. Total cloud amount (%) over the Central Mediterranean theater for (a) 6-h MLR (left) and MDA (center) forecasts and RTNEPH analysis (right) valid at 1800 UTC 14 Jan 1992 and (b) 30-h MLR and MDA forecasts and RTNEPH analysis valid at 1800 UTC 20 Jan 1992. Black locations on RTNEPH images indicate untimely (more than 2 h old) data.

b. 13–22 July 1992

Table 8 lists the cloud amount verification statistics for the July 10-day verification period. The percentage of timely points was about 72% for 6-, 18-, and 30-h forecasts (0600 and 1800 UTC times) and 68% for 12-, 24-, and 36-h forecasts (0000 and 1200 UTC). In verification of persistence, only the timely points common to both the persisted and verifying RTNEPH could be used. The percentage of timely points common to both averaged about 52% over all six forecast times.

The ME skill scores indicate that, as in January, the MLR and MDA methods generally produce a small bias in the diagnosed cloud amounts. The MLR MEs are consistently negative. This is due to the ‘‘thinning’’ of the number of MLR-diagnosed cloudy layers using the MDA-diagnosed value of the predictand ‘‘total number of cloud layers.’’ By contrast, the MDA MEs alternate between more negative (at 0600 and 1800 UTC) and more positive (at 0000 and 1200 UTC) values. The larger positive MEs in the low and total cloud amounts at the 12-, 24-, and 36-h forecast durations in MDA were due to a lesser amount of cloud over the theater after 15 July than before that date. Thus, the frequency distribution imposed on the diagnosis represented a cloudier period than in the application period. The reason this effect is not seen in the 6-, 18-, and 30-h forecasts is that the difference in cloudiness before and after 15 July was greater at 1200 UTC (the time of day of greatest cloudiness) than at 0600 and 1800 UTC.

Table 8. Cloud amount category verification statistics for predictions initialized twice daily, 13–22 Jul 1992. Units for ME, rmse ~ 20% cloud amount.

	6-h	12-h	18-h	24-h	30-h	36-h	6-h	12-h	18-h	24-h	30-h	36-h
			ME						PPC			
			High						High			
MLR	−0.21	−0.23	−0.18	−0.21	−0.15	−0.18	80.9	74.4	78.5	76.7	78.3	75.9
MDA	−0.11	0.03	−0.04	0.05	−0.06	0.04	81.7	74.1	80.9	74.6	81.9	76.4
Pers.	−0.04	0.02	0.00	0.03	−0.02	0.03	77.9	72.6	73.8	74.1	75.5	74.8
			Middle						Middle			
MLR	−0.13	−0.17	−0.15	−0.17	−0.10	−0.15	72.1	69.0	72.4	67.4	70.3	67.4
MDA	−0.08	0.08	0.02	0.03	−0.03	0.08	75.5	69.8	72.9	71.2	74.8	70.6
Pers.	−0.8	0.04	−0.03	0.02	−0.06	0.04	72.6	69.9	68.7	71.7	69.1	68.9
			Low						Low			
MLR	−0.18	−0.07	−0.20	−0.12	−0.18	−0.09	54.3	56.0	54.0	55.8	54.8	55.6
MDA	−0.18	0.23	0.00	0.24	−0.05	0.34	58.2	48.3	56.6	48.2	57.8	46.5
Pers.	−0.22	0.02	−0.11	0.08	−0.23	0.13	55.8	38.4	43.5	54.9	48.7	35.7
			Total						Total			
MLR	−0.30	−0.29	−0.36	−0.32	−0.26	−0.31	51.4	49.4	53.3	46.9	50.3	48.1
MDA	−0.17	0.39	0.05	0.37	−0.05	0.53	55.1	40.2	52.5	39.2	52.9	38.0
Pers.	0.00	0.06	0.15	0.10	0.02	0.17	49.6	33.0	38.3	41.1	39.8	29.7
			Rmse						FDF			
			High						High			
MLR	1.02	1.10	1.02	1.03	1.00	1.02	0.14	0.13	0.15	0.15	0.12	0.15
MDA	1.07	1.33	1.12	1.30	1.09	1.29	0.09	0.02	0.03	0.03	0.06	0.04
Pers.	1.17	1.36	1.35	1.33	1.30	1.31	0.03	0.01	0.01	0.02	0.02	0.03
			Middle						Middle			
MLR	1.17	1.20	1.17	1.18	1.19	1.14	0.08	0.18	0.09	0.22	0.11	0.24
MDA	1.27	1.52	1.37	1.45	1.30	1.47	0.08	0.05	0.01	0.02	0.03	0.04
Pers.	1.38	1.52	1.53	1.47	1.55	1.55	0.07	0.03	0.05	0.02	0.07	0.03
			Low						Low			
MLR	1.51	1.30	1.50	1.29	1.50	1.34	0.19	0.12	0.24	0.21	0.24	0.16
MDA	1.66	1.74	1.71	1.75	1.68	1.83	0.11	0.12	0.04	0.13	0.03	0.18
Pers.	1.73	1.97	2.08	1.70	2.07	2.11	0.16	0.04	0.14	0.05	0.12	0.07
			Total						Total			
MLR	1.56	1.46	1.54	1.46	1.59	1.53	0.15	0.19	0.19	0.19	0.21	0.21
MDA	1.74	1.95	1.80	1.93	1.77	2.02	0.11	0.19	0.04	0.18	0.04	0.26
Pers.	1.71	2.02	2.13	1.91	2.18	2.21	0.14	0.04	0.24	0.05	0.12	0.08

Because the diagnostic methods were developed and diagnosed at the two different sets of times, it is hard to compare their performance between 6-, 18-, and 30-h forecasts on one hand (developed and diagnosed at 0600 and 1800 UTC), and 12-, 24-, and 36-h forecasts on the other (developed and diagnosed at 0000 and 1200 UTC). If we would have initialized the forecasts four times per day, the development and diagnosis times would have been the same for all forecast durations. In this case, the verification scores would have been more comparable between the forecast durations.

In the July rmse and PPC statistics, we see a major difference from the January results. Here, MLR and MDA skill is more consistently improved over persistence, particularly in low and total, than in January. This is likely due at least in part to the fact that summertime cloudiness is less persistent than wintertime cloudiness. There is a greater degree of temporal variability, making a persistence forecast less likely to be successful. Even so, there does seem to be some indication of real skill in the diagnostic methods. As usual, MLR beats MDA in rmse, while the reverse is generally true for PPC. The exceptions are at the 12-, 24-, and 36-h times, when MDA imposed too much cloudiness, putting too many grid points in the greater cloud amount categories. Persistence appears to be most competitive with MLR and MDA at 24 h, suggesting some modest skill in ‘‘diurnal persistence’’ in the summertime. Finally, the FDF scores show once again that MDA generally reproduces the RTNEPH frequency distribution best, except for the overdiagnosed cloudiness at the 12-, 24-, and 36-h durations in low total cloud.

Due to space limitations, the verification statistics for cloud type, base altitude, and thickness for the July cloud forecast diagnoses are not included. It was found that for cloud type, the skill of MLR and MDA with respect to persistence was greater in July than in January. At some forecast durations, MLR and MDA cloud type skills, as measured by PPC, were greater than that for persistence while for other durations they were less. For cloud-base altitude thickness, rmse scores for MLR and MDA were better than persistence at some forecast durations, and worse at others. Cloud-base altitude and thickness PPC scores for persistence were as good or better than their MLR and MDA counterparts. Cloud-base altitude biases for MLR and MDA were consistently small, while for persistence they were large and positive at 6-, 18-, and 30-h forecast durations. Cloud

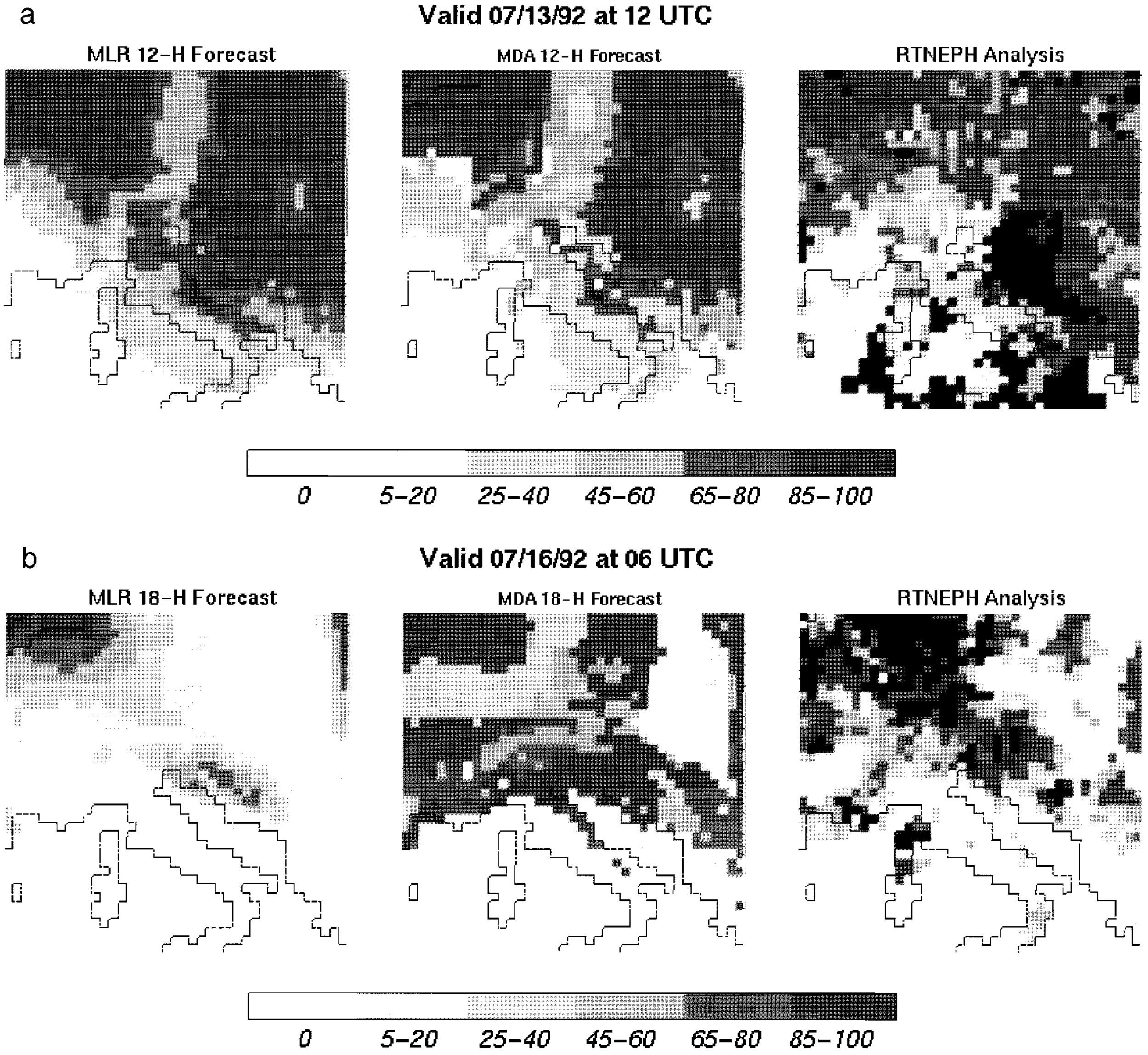

FIG. 3. Same as in Fig. 1 except for (a) 12-h MLR and MDA forecasts and RTNEPH analysis valid at 1200 UTC 13 Jul 1992 and (b) 18-h MLR and MDA forecasts and RTNEPH analysis valid at 0600 UTC 16 Jul 1992.

thickness biases were consistently small for MDA and persistence, but were larger in magnitude for MLR at some forecast durations. MDA almost always produced the best FDF scores for type, base altitude, and thickness.

The July cloud-ceiling altitude verification scores (not shown) echoed the January trends. Using the base altitude of the lowest cloud layer with greater than 50% cloud amount yielded better rmse and PPC scores in MLR and MDA than directly diagnosing ceiling altitude from both methods. The former approach resulted in somewhat larger biases and for MDA degraded the FDF. The most important result was that in July, the rmse and PPC cloud-ceiling altitude verification scores for MLR and MDA were better than persistence, regardless of whether they were diagnosed directly or were inferred from the base altitude of the lowest layer with greater than 50% cloud amount.

Figure 3 shows the depictions of diagnosed and RTNEPH total cloud amount for two times during the July verification period. In Fig. 3a, the MLR and MDA 12-h forecast diagnoses valid at 1200 UTC 13 July 1992 show overcast areas in the northwest and northeast portions of the theater, with lesser cloud amounts in the north central portion. This appears to be a fairly good representation of the total cloud conditions over the northern part of the plot as shown in the RTNEPH depiction. Both MLR and MDA understate the cloud amount in the west central part of the theater, and MLR produces a larger area of scattered clouds west of Italy than is displayed by RTNEPH. Neither diagnosis has the extent of spatial variability of cloud amount that is seen in the RTNEPH depiction. In Fig. 3b, the RTNEPH analysis at 0600 UTC 16 July 1992 shows a band of broken to overcast cloud amounts covering the northwestern portion of the theater, extending down into the central portion. The spatial extent of the coverage is excessive in the MDA 18-h diagnosis valid at this same time. The pattern of the total cloudiness diagnosed by the MLR matches fairly well with the major features of the RTNEPH cloudiness, but the amounts appear to be understated. The clear slot in the northeast is located well in the MLR, but the MLR lacks the cloud appendages in the west and east central portions of the theater that are evident in the RTNEPH image.

6. Discussion

The methods used in this study to diagnose cloud variables from mesoscale numerical weather prediction model forecasts involved the development of statistical linear relationships between selected predictors and the cloud variable predictands. Separate relationships were developed for each forecast duration. This approach has the advantage of being able to account for the change in forecast error with increasing duration. It has the potential disadvantage of yielding a sequence of cloud diagnoses that are unrealistically discontinuous in time when displayed in animation. This latter issue was not considered in this study. While such a time discontinuity might be avoided by using one duration's relationship to diagnose all durations' cloud variables, the risk of introducing a greater error due to failing to account for forecast error growth looms with this approach. It was found that there was no discernible decrease of cloud amount diagnosis skill with increasing forecast duration in either January or July diagnoses.

A noticeable trend in the verification statistics was the similarity of the skill scores at 6-, 18-, and 30-h forecast durations, and at 12-, 24-, and 36-h forecast durations. We attribute the trend toward distinctions between the two sets of scores to the differences in availability of timely data between 0600 and 1800 UTC, the times used to develop and verify the former set of forecast durations, and 0000 and 1200 UTC, the times used to develop and verify the latter. Different amounts, ages, and sources of timely (no more than 2 h old) RTNEPH cloud data were used in the analysis at the two different sets of times. This resulted in two different sets of verification scores. Thus, it was possible to compare skill scores only for the 6-, 18-, and 30-h verifications with each other and to compare the 12-, 24-, and 36-h verifications with each other. This would have been overcome if forecasts were initialized four times per day, rather than just two as in this study.

It was found that there can be a change from day to day in the predictors that are used when developing predictor–cloud amount relationships over a theater-sized area. This is no doubt due to the change in cloud distribution over such a small area. The theater-scale area is on the same scale of a single high or low pressure system. As the system moves across the theater, the character of the clouds change with the movement, and thus the weather associated with the changing clouds. For example, in the period 1–23 July 1992, there were five relative maxima in the time series of theater-averaged total cloud amount.

The two diagnosis methods considered in this study generally were able to avoid creating a biased cloud amount distribution. This is because the methods were able to account for any systematic errors in the NWP model when the cloud–predictor relationships were developed. The category selection method associated with the MDA was designed to diagnose the same frequency distribution of the cloud variable as was present in the dependent data. Thus, the MDA scheme may produce nontrivial biases when the application period has a much different amount of cloudiness in the theater than was present during the development period.

The degree to which the MLR scheme was able to replicate the frequency distribution of cloud amounts in the RTNEPH depended on the frequency of occurrence of the extreme cloud amount categories. In July, the MLR did a good job in reproducing the cloud amount distribution. This is because the extreme categories were only somewhat more likely than the midrange categories; that is, the probability of partly cloudy skies was high. In January, MLR could not produce a high enough frequency of occurrence of 85%–100% cloud amount. We consider this to be a major limitation of the MLR scheme in its application to diagnosing cloud amounts over a theater-scale area. It is the reason that we designed and employed the MDA scheme. Nehrkorn and Zivkovic (1996) also found that the MLR scheme was unable to reproduce the degree of cloud amount variance present in the cloud analysis when it was applied to forecasts from a global NWP model. The MDA scheme, with its frequency distribution–preserving category selection method, performed in a superior manner in reproducing the frequency of occurrence of cloud amount as seen in the verifying RTNEPH.

The price that is paid by MDA in preserving the frequency distribution of the predictand is the loss of rmse skill as compared with MLR. The MLR is designed to minimize mean-squared error and suffers skewing of the frequency distribution of the predictand as a result. This is particularly acute in the case of cloud amount, which, in our study, consistently showed a ''U-shaped'' frequency distribution. Even though MDA did not score as well as MLR in rmse, in some cases the MDA scores were still better than persistence. The most consistent and clear case of this was for July. This is the case in which MLR and MDA showed the greatest skill with respect to persistence. While MLR excelled over persistence in rmse for January, MDA did not score as well in rmse as persistence in January.

The score ‘‘percent predicted correctly’’ (PPC) combines features of skill with frequency distribution reproduction. PPC scores can be quite high in cases where a single category of cloud amount is quite likely. On the other hand, if this is one of the extreme categories, MLR will generally score quite low in PPC. As with rmse, MLR and MDA enjoyed their greatest PPC advantage over persistence for July. In January, persistence was as good as or better than the diagnosis method in PPC.

The MLR and MDA diagnosis schemes were also applied to the cloud-layer predictands: cloud type, base altitude, and thickness. These variables as specified in the RTNEPH cloud analyses are considered to have less reliability than the cloud amount. For this reason, we expect them to have a lower correlation to weather variables than does cloud amount. This would in turn lead to poorer performance upon diagnosis. In general, we found that only in certain instances did type, base altitude, and thickness diagnosis by either scheme excel over a persistence forecast. For example, cloud type skill was greater for July than January. As in cloud amount, biases for the cloud-layer predictands from both schemes were small, and frequency distribution reproduction by MDA was consistently good.

Cloud-ceiling altitude was diagnosed directly by both methods and was also deduced from the lowest cloud-base altitude with a cloud amount of greater than 50%. We found that both diagnosed and deduced cloud-ceiling altitudes showed skill with respect to persistence. Deduced cloud altitudes had better rmse and PPC skill scores than their diagnosed counterparts for both MLR and MDA. The diagnosed ceilings had slightly smaller MEs than the deduced values, and MDA-diagnosed ceilings reproduced the frequency distribution better than MDA-deduced ceilings.

We conclude from these results that the MLR and MDA cloud amount diagnosis procedures are unable to consistently outperform persistence in quantitative skill in the theater-scale area used in this study. The degree of skill shown with respect to persistence varies with season in this midlatitude locale, in which improvement over persistence was evident in the summer. This may be due to the greater temporal variability of cloudiness over this theater in summer than in winter. Because the MLR scheme is prone to misrepresenting the frequency distribution of cloud amount, especially when it is characterized by a high probability of one of the cloud amount extremes, the MDA is clearly the more risk-free choice of the schemes.

The diagnoses of cloud type, base altitude, and thickness also do not show consistent skill with respect to persistence. Given the inability of the cloud amount diagnoses to do so, we expect this to be true for the less reliably specified cloud-layer variables. However, cloud-ceiling altitudes produced by the diagnostic schemes, especially when deduced from the lowest cloud-layer base altitude, can excel over persistence. We conclude that there is useful skill in deducing cloud-ceiling altitudes from the MDA cloud amount and cloud-base altitude diagnoses.

The subjective evaluation of the cloud diagnosis schemes showed that the MDA is able to generate realistic total cloud amount patterns. The gradients of total cloud amount generated by MDA are much more like those depicted in the verifying RTNEPH than those produced by MLR, which tends to produce too many grid points with midrange cloud amounts. The resulting gradients are not sharp enough when compared with RTNEPH. In our view, the MDA-generated total cloud amount depictions show enough realism to be useful as forecast guidance tools. As for their skill or usefulness compared to prediction products of cloud trajectory models, it is an unanswered question until such a comparison takes place.

There are inherent limitations to the capabilities of statistically based cloud diagnosis schemes when applied to the forecasts from numerical weather prediction models. The schemes can perform only as well as the model predicts the weather associated with the cloudiness. Also, there are no explicit physical relationships involved in the diagnosis products, but simply a statistical association of the predictand with the selected predictor. Finally, the reliability of the predictand is a major limiting factor in the realism of the diagnosed cloud predictions, because the predictand is used to develop the diagnosis schemes.

With such limitations, diagnostic cloud schemes can be seen only as a short-term solution to the need for cloud prediction. Yet a short-term solution is needed at this time. Trajectory-based cloud schemes have even greater limitations, such as no ability to represent weather-related cloud growth or decay. Prognostic schemes have been developed on very limited samples of quantitative cloud measurements, predict water concentration and have no proven methods of translating that into observed cloud variables, and have received very little evaluation of the cloud predictions themselves. The primary contribution of this study is to provide and evaluate a method that is ready to be used now to give predictions of future cloud states. Not only can it predict cloud amounts in layers but also types, base altitudes, and thicknesses. The MDA-based cloud diagnosis procedure described in this paper can provide the basis for a short-term theater-scale cloud prediction capability. The long-term solution is the prognostic cloud scheme. But much work must be done to mature the prognostic method into one that is reliable and useful for observed cloud variable predictions. In the mean time, the MDA-based cloud diagnosis procedure described herein can serve as a useful tool to satisfy the need for theater-scale predictions of observed cloud variables.

Acknowledgments. I appreciate the efforts of Don Aiken and Dan DeBenedictus in executing the MM5 model and processing the RTNEPH datasets. Don Chisholm’s interest and encouragement was helpful in the development of the diagnostic methods. This work was funded through the U.S. Air Force 63707F Program.

REFERENCES

Albers, S. C., J. A. McGinley, D. L. Birkenheuer, and J. R. Smart, 1996: The local analysis and prediction system (LAPS): Analysis of clouds, precipitation, and temperature. *Wea. Forecasting,* **11,** 273–287.

Brunet, N., R. Verret, and N. Yacowar, 1988: An objective comparison of model output statistics and perfect prog systems in producing numerical weather element forecasts. *Wea. Forecasting,* **3,** 273–283.

Carter, G. M., and H. R. Glahn, 1976: Objective prediction of cloud amount based on model output statistics. *Mon. Wea. Rev.,* **104,** 1565–1572.

Cianciolo, M. E., 1993: Short-range cloud amount forecasting with model output statistics. PL-TR-93-2205, Phillips Laboratory (AFMC), Hanscom AFB, MA, 166 pp. [NTIS ADA 274769.]

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statisics (MOS) in objective weather forecasting. *J. Appl. Meteor.,* **11,** 1203–1211.

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

——, J. Dudhia, and D. R. Stauffer, 1994: A description of the fifth-generation Penn State/NCAR mesoscale model (MM5). NCAR/TN-398 + STR, National Center for Atmospheric Research Boulder, CO, 121 pp. [Available online at http://www.mmm.ucar.edu/mm5.]

Hamill, T. M., R. P. d'Entremont, and J. T. Bunting, 1992: A description of the Air Force real-time nephanalysis model. *Wea. Forecasting,* **7,** 288–306.

Mitchell, K. E., and D. C. Hahn, 1989: Objective development of diagnostic cloud forecast schemes in global and regional models. Preprints, *Seventh Conf. on Atmospheric Radiation,* San Francisco, CA, Amer. Meteor. Soc., J138–J145.

Nehrkorn, T., and M. Zivkovic, 1996: A comparison of diagnostic cloud cover schemes. *Mon. Wea. Rev.,* **124,** 1732–1745.

Norquist, D. C., and S. Chang, 1994: Diagnosis and correction of systematic humidity error in a global numerical weather prediction model. *Mon. Wea. Rev.,* **122,** 2442–2460.

——, C.-H. Yang, S. Chang, and D. C. Hahn, 1992: Phillips Laboratory global spectral numerical weather prediction model. PL-TR-92-2225, Phillips Laboratory (AFMC), Hanscom AFB, MA, 166 pp. [NTIS ADA 267293.]

——, H. S. Muench, D. L. Aiken, and D. C. Hahn, 1994: Diagnosing cloudiness from global numerical weather prediction model forecasts. PL-TR-94-2211, Phillips Laboratory (AFMC), Hanscom AFB, MA, 152 pp. [NTIS ADA 289456.]

——, D. L. Aiken, and D. A. DeBenedictus, 1997: Cloud cover predictions diagnosed from global numerical weather prediction model forecasts. PL-TR-97-2015, Phillips Laboratory (AFMC), Hanscom AFB, MA, 150 pp. [NTIS ADA 331790.]

Schaaf, C. B., J. M. Ward, H. S. Muench, R. P. d'Entremont, M. K. Griffin, G. B. Gustafson, and J. T. Bunting, 1990: The hemispheric eighth mesh terrain elevation and geography data sets. GL-TR-90-0223, Geophysics Laboratory (AFSC), Hanscom AFB, MA, 16 pp. [Available from AFRL/VSBE, 29 Randolph Rd., Hanscom AFB, MA 01731-3010.]

Stobie, J. G., 1986: AFGWC's advanced weather analysis and prediction system (AWAPS). AFGWC/TN-86-001, Air Force Global Weather Central, Offutt AFB, NE, 98 pp. [Available from AFWA, 106 Peacekeeper Dr., Offutt AFB, NE 68113-4039.]

Trapnell, R. N., 1992: Cloud Curve algorithm test program. PL-TR-92-2052, Phillips Laboratory (AFMC), Hanscom AFB, MA, 170 pp. [NTIS ADA 253918.]

USAFETAC, 1991: RTNEPH USAFETAC climatic database users handbook no. 1. USAFETAC/UH—86/001, USAFETAC, Asheville, NC, 18 pp. [Available from AFCCC, 151 Patton Ave., Rm 120, Asheville, NC 28806-5002.]