Probabilistic Forecasting of (Severe) Thunderstorms in the Netherlands Using Model Output Statistics

MAURICE J. SCHMEITS, KEES J. KOK, AND DAAN H. P. VOGELEZANG

Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands

(Manuscript received 29 April 2004, in final form 7 September 2004)

ABSTRACT

The derivation and verification of logistic regression equations for the (conditional) probability of (severe) thunderstorms in the warm half-year (from mid-April to mid-October) in the Netherlands is described. For 12 regions of about 90 km × 80 km each, and for projections out to 48 h in advance (with 6-h periods), these equations have been derived using model output statistics (MOS). As a source for the predictands, lightning data from the Surveillance et d'Alerte Foudre par Interférométrie Radioélectrique (SAFIR) network have been used. The potential predictor dataset mainly consisted of the combined (postprocessed) output from two numerical weather prediction (NWP) models. It contained 15 traditional thunderstorm indices, computed from the High-Resolution Limited-Area Model (HIRLAM), and (postprocessed) output from the European Centre for Medium-Range Weather Forecasts (ECMWF) model. The most important predictor in the thunderstorm forecast system is the square root of the ECMWF 6-h convective precipitation sum, and the most important predictor in the *severe* thunderstorm forecast system is the HIRLAM Boyden index. The success of the square root of the ECMWF 6-h convective precipitation sum as a thunderstorm predictor indicates that there is a strong relation between the forecast convective precipitation by the ECMWF model and the occurrence of thunderstorms, at least in the Netherlands up to 3 days in advance. The overall verification results for the 0000, 0600, 1200, and 1800 UTC runs of the MOS (severe) thunderstorm forecast system are good, and, therefore, the system was made operational at the Royal Netherlands Meteorological Institute (KNMI) in April 2004.

1. Introduction

The Royal Netherlands Meteorological Institute (KNMI) is responsible for issuing warnings to the general public in the case of severe weather. If severe weather is expected to occur on the scale of a province in the Netherlands (Fig. 1) within the next 12 h, a so-called weather alarm is issued. Thunderstorms are one of the most damaging weather phenomena, especially if they are accompanied by large hail, (flash) flooding, wind gusts, or tornadoes. In the Netherlands, thunderstorms occur quite frequently during late spring, summer, and early autumn. Because thunderstorms are inherently stochastic in nature, this uncertainty is best accounted for by using probabilities in forecasting thunderstorms.

Since the early 1980s automated probability forecasts, based on model output statistics (MOS; Glahn and Lowry 1972; Wilks 1995), have been produced for the occurrence of a thunderstorm at minimally 1 out of 10 (predetermined) stations in the Netherlands during the period of 0000–2400 UTC (Lemcke and Kruizinga 1988). The MOS technique consists of determining a statistical relationship between a predictand (i.e., the occurrence of a thunderstorm in this case) and predictors from numerical weather prediction (NWP) model forecasts. Because society is getting more and more vulnerable to severe weather, there is a need to increase the temporal and spatial resolution of the thunderstorm forecasts. Because the resolution of operational NWP models has increased, with an accompanying improvement in forecast skill, and remote sensor observations of lightning have been used operationally for a number of years, it is now possible to make thunderstorm probability forecasts for regions of about 90 km × 80 km (Fig. 1; see section 2) out to several days. The temporal resolution is increased to periods of 6 h.

For projections out to 48 h in advance, we have developed an automated forecast system based on combined (postprocessed) output from the High-Resolution Limited-Area Model (HIRLAM; Undén et al. 2002) and the European Centre for Medium-Range Weather Forecasts (ECMWF) model. As a source for the predictands, we have used lightning data from the Surveillance et d'Alerte Foudre par Interférométrie

Corresponding author address: Dr. M. J. Schmeits, Royal Netherlands Meteorological Institute (KNMI), P.O. Box 201, 3730 AE De Bilt, Netherlands.
E-mail: schmeits@knmi.nl

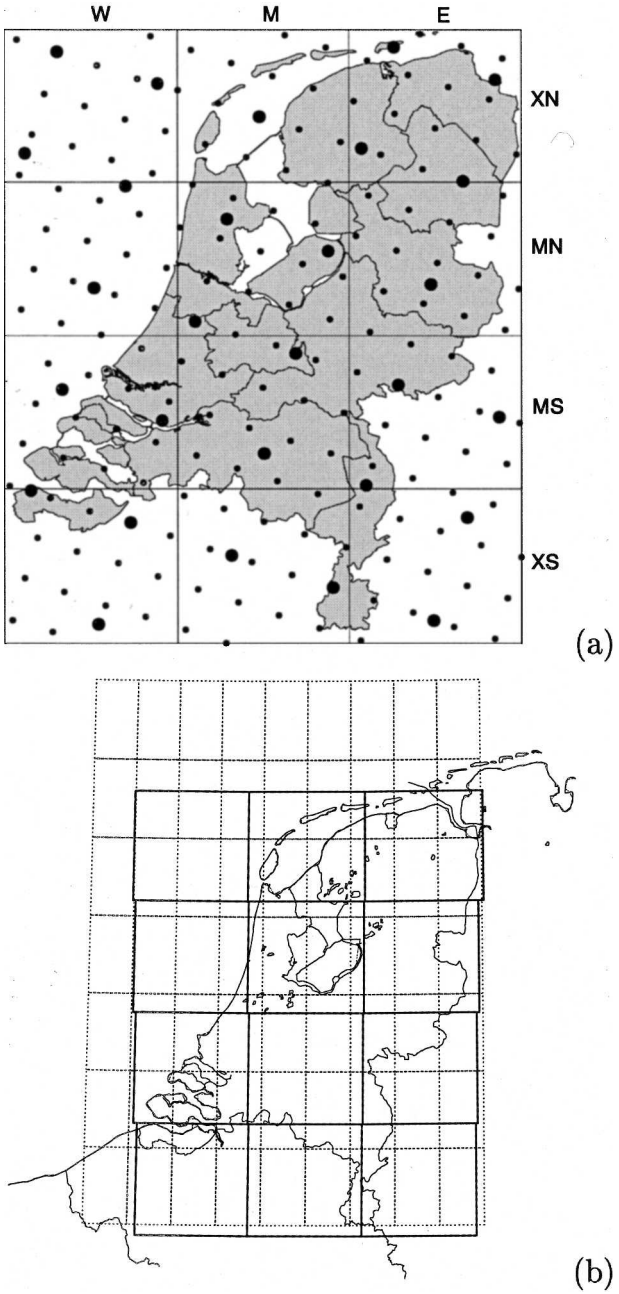

FIG. 1. (a) Geography of the Netherlands (gray shaded) and surroundings, subdivided in 12 regions (W: west, M: middle, E: east, N: north, S: south, and X: extreme). The province boundaries of the Netherlands are also indicated. The large (small) black circles indicate the HIRLAM grid points with 55-km (22 km) horizontal resolution. (b) Same subdivision as in (a) of the Netherlands and surroundings in 12 regions (solid rectangles), but in a different coordinate system; the dotted rectangles show the ECMWF grid at a horizontal resolution of ½°.

Radioélectrique (SAFIR) network (Wessels 1998). Two predictands are defined for each region and time period—the first is the probability of a thunderstorm (≥2 lightning discharges), and the second is the conditional probability of a *severe* thunderstorm (≥500 discharges) under the condition that ≥2 discharges will be detected. The criterion for severe thunderstorms used in this paper is different from the current weather alarm criterion for severe thunderstorms. The latter criterion is that at least 15 lightning discharges per minute are expected to occur within a radius of 15 km, possibly accompanied by cloud bursts, large hail, and/or very strong wind gusts (>102 km h^{-1}). Because this criterion is met only a few times per year in the Netherlands, standard statistical techniques (like logistic regression; Brelsford and Jones 1967; Wilks 1995) are not expected to lead to skillful forecasts, and, therefore, we have used another criterion for severe thunderstorms in this paper. In a subsequent study we will investigate whether it is possible to devise a MOS system that can be used as a tool to decide whether a weather alarm for severe thunderstorms should be issued.

The potential predictor set contains a set of traditional thunderstorm indices, computed from the HIRLAM forecasts, (postprocessed) output from the ECMWF model, the (co)sine of the day of the year, and the so-called P27 scores (Kruizinga 1979). The latter scores represent an objective classification of daily 500-hPa patterns (see section 2).

There are many papers describing the predictive potential of the various thunderstorm indices in a number of countries in Europe (e.g., Andersson et al. 1989; Jacovides and Yonetani 1990; Collier and Lilley 1994; Huntrieser et al. 1997), but to date only one study (Haklander and Van Delden 2003) systematically investigated the performance of a number of these indices in the Netherlands. However, our study is very different from that of Haklander and Van Delden (2003). For example, they investigated the predictive potential of 32 traditional thunderstorm indices, derived from rawinsonde observations of De Bilt (in the center of the Netherlands), instead of NWP model forecasts, which limited the projection of their method to 6 h. In addition, they assessed these 32 indices individually. Instead, we objectively select combinations of thunderstorm predictors, because "each of these indices has strengths and weaknesses, and no single index can be thought to provide a complete characterization of the state of the atmosphere" (Blanchard 1998). Haklander and Van Delden (2003) also used data from a different lightning detection system, they did not discriminate between "ordinary" and severe thunderstorms, and they limited their study to an area within 100 km from De Bilt. Their main conclusion was that the lifted index (Galway 1956) and the Boyden index (Boyden 1963) were good dichotomous thunderstorm predictors.

In this paper we describe the derivation and verification of MOS equations for the (conditional) probability of (severe) thunderstorms in the warm half-year (from mid-April to mid-October) in the Netherlands. In the early 1990s a similar MOS system was developed in the United States, based on output from the Nested

Grid Model (NGM; Reap 1994), and later, more skilful MOS systems were developed, based on output from the Global Spectral Model (GSM; Hughes 2001) and from the mesoscale Eta Model (Hughes 2002). Our study differs from those studies in that we have developed separate thunderstorm forecast equations for each region (instead of pooling all regions)—possibly leading to better verification scores—and we have combined the output from two NWP models. Also, we have used a different definition for severe thunderstorms, based on the lightning dataset instead of storm reports, and we have used another statistical method, namely, logistic regression instead of linear regression. According to Applequist et al. (2002), logistic regression is the preferred regression method in probabilistic forecasting.

In section 2, this method, the predictands, and the predictors are described. In section 3 we present an example of the performance of the objective thunderstorm forecast system during a day with severe weather and some of the verification results for the (independent) warm season of 2003. Finally, in section 4 we summarize and discuss the results.

2. Statistical method, predictands, and predictors

a. Logistic regression

The derivation of the MOS equations has been done using the method of logistic regression (Brelsford and Jones 1967; Wilks 1995). According to this method the probability Pr that an event y occurs is

$$\Pr\{y\} = \frac{1}{1 + \exp(a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n)}. \quad (1)$$

The predictors $x_i (i = 1, 2, \ldots, n)$ are selected via a so-called forward stepwise selection method (Wilks 1995). At each step, a predictor is chosen that produces the best regression in conjunction with the predictors chosen on previous steps; hereby, a significance threshold of 0.05 is specified. Each chosen predictor is kept in the equation unless the specified significance threshold of 0.10 is exceeded at a following step. The regression coefficients a_i are determined using the maximum likelihood method, which is an iterative method that maximizes the product of all computed probabilities of the (non) occurrence of the event in the dependent dataset. The datasets used in this study are lightning data from the SAFIR network (section 2b; Wessels 1998) and (postprocessed) output data from the HIRLAM (Undén et al. 2002) and ECMWF models (section 2c).

b. Predictand definitions and climatology

We have defined two events—one is called the thunderstorm event and the other the *severe* thunderstorm event. An event is called a thunderstorm event if ≥2 lightning discharges are detected by the SAFIR network in a 6-h time period (0300–0900, 0900–1500, 1500–2100 or 2100–0300 UTC) in a region (Fig. 1). The predictand for thunderstorms is defined as the probability of a thunderstorm event. An event is called a severe thunderstorm event if ≥500 lightning discharges are detected by the SAFIR network in a 6-h time period in a region. The predictand for severe thunderstorms is defined as the conditional probability of a severe thunderstorm event under the condition of a thunderstorm event. We have chosen a time period of 6 h because the HIRLAM and ECMWF model outputs have been archived with a time resolution of 6 h (section 2c). We preferred to use one region size for each projection. Therefore, a region size of about 90 km × 80 km has been chosen because it was estimated to be a spatial resolution for which skillful forecasts of thunderstorms are possible out to 48 h in advance.

The climatological thunderstorm probability charts are shown in Fig. 2a. The climatological thunderstorm probability is highest during the evening (1500–2100 UTC; local summer time is UTC + 2 h) in most regions (except for the three most northern regions) and lowest during the morning (0300–0900 UTC) in most regions, as a result of the diurnal cycle of convection. However, the climatological thunderstorm probability shows a different diurnal cycle in some of the coastal regions, for instance, in the region west (W)–extreme north (XN) (Fig. 1a) the climatological probability attains its maximum during the night (2100–0300 UTC) and its minimum during the afternoon (0900–1500 UTC), that is, it has a phase lag of 6 h. This phase lag is due to a different diurnal cycle of convection over sea compared to that over land. During the afternoon and evening the probabilities increase southward as a result of the climatologically higher maximum temperatures in the south, and during the night and morning the probabilities increase westward due to the closer proximity of the relatively warm seawater.

Because the climatological *absolute* probabilities of severe thunderstorms are only between 0% and 6% (not shown), the use of logistic regression (section 2a) would lead to optimized forecasts in the lower probability range but not in the higher-probability range. The climatological *conditional* probabilities of severe thunderstorms are higher (Fig. 2b), and, therefore, conditional probabilities are used for severe thunderstorms. The charts in Fig. 2b show a similar diurnal cycle as in Fig. 2a.

Because the severe thunderstorm sample size is relatively small for each region separately, the regions have been pooled. Because of the geographical differences in the climatological thunderstorm probability between land and coast, both the six land regions [west (W)–extreme south (XS), middle (M)–XS, east (E)–XS, M–midsouth (MS), E–MS, and E–midnorth (MN)] and the six so-called coastal regions [W–MS, W–MN, M–MN, W–extreme north (XN), M–XN, and E-XN] have been pooled to derive the severe thunderstorm

FIG. 2. (a) Climatological thunderstorm probability charts (%). (b) Climatological conditional probability charts of severe thunderstorms (%). For the periods of 2100–0300 and 0300–0900 UTC, all regions have been pooled, and for the periods of 0900–1500 and 1500–2100 UTC both the six land regions (W–XS, M–XS, E–XS, M–MS, E–MS, and E–MN) and the six coastal regions (W–MS, W–MN, M–MN, W–XN, M–XN, and E–XN) have been pooled. The climatologies have been computed using SAFIR data from the warm half-years of 1999–2001.

forecast equations for the afternoon and evening periods. For the night and morning periods all 12 regions have been pooled because of the even smaller sample size for each region separately during these times. Separate severe thunderstorm forecast equations have been derived for each projection and run time, resulting in a total of 42 forecast equations. No pooling has been performed in the derivation of the equations for the probability of ≥2 discharges, except for the initial predictor selection. This initial predictor selection has been performed for both the six pooled land regions and the six pooled coastal regions. Using this initial predictor selection as the potential predictor set for each region, separate thunderstorm forecast equations have been derived for each region, projection, and run time, resulting in a total of 348 forecast equations. The climatological probability charts of Figs. 2a and 2b are used as the reference forecasts in the verification of the regression equations (section 3).

c. Potential predictors

The HIRLAM output was used every 6 h from 6 to 48 h in advance from the 0000, 0600, 1200, and 1800 UTC cycles. We have computed a set of traditional thunderstorm indices from the HIRLAM forecasts. Subsequently, we have calculated the minimum, maximum, and average value of each index using all grid points in each of the 12 regions of Fig. 1a. These minimum, maximum, and average values of the indices are then used as potential predictors.

The ECMWF model output was used every 6 h from 18 to 72 h in advance from the 1200 UTC cycle. For several (postprocessed) output variables from the ECMWF forecasts we have also calculated the maximum values in each of the regions (Fig. 1b). In addition, the so-called P27 scores (Kruizinga 1979) have been used as potential predictors, as has the (co)sine of the day of the year. The three P27 scores, computed from the ECMWF model forecasts, are objective measures of the degree of zonality, meridionality, and cyclonality of the 500-hPa flow over western Europe. These scores verify at 0000 UTC. The other predictors verify at 0000, 0600, 1200, or 1800 UTC. For more information on the P27 classification, the reader is referred to Kruizinga (1979). We have not included (derived) output from the ECMWF Ensemble Prediction System (EPS; e.g., Molteni et al. 1996) as potential predictors, because we expected these to have no significant (additive) predictive potential (with respect to the other predictors) for lead times out to 72 h.

The HIRLAM output has a horizontal resolution of 55 (22) km from 1999 to 2001 (from 2002 to present) and the ECMWF model's horizontal resolution is T319 (T511) from 1999 to 2000 (from 2001 to present), but the ECMWF output is used at a resolution of ½° (Figs. 1a and 1b). HIRLAM has 31 nonequidistant levels in the vertical direction and the ECMWF model has 60 vertical levels. The change in horizontal resolution has

TABLE 1. Overview of derived HIRLAM predictors that are included in at least one (severe) thunderstorm forecast equation. Here, z is the (geopotential) height, T is the temperature (°C), T_d is the dewpoint temperature (°C), g is the acceleration due to gravity, LFC is the level of free convection, T_v is the virtual temperature, env. is the environment, ρ is the density, u is the zonal wind component, v is the meridional wind component, and q is the specific humidity.

HIRLAM predictors		
Predictor	Definition	Reference
Boyden index	$0.1\ (z_{700} - z_{1000}) - T_{700} - 200$	Boyden (1963)
Wet-bulb potential temperature at 500 hPa	θ_{w500}	
Bradbury index	$\theta_{w500} - \theta_{w850}$	Bradbury (1977)
Wet-bulb pseudopotential temperature at 500 hPa	θ_{ws500}	
Showalter index	$\theta_{ws500} - \theta_{w850}$	Showalter (1953)
Rackliff index	$\theta_{w925} - T_{500}$	Rackliff (1962)
Jefferson index	$1.6 \times \theta_{w925} - T_{500} - 11$	Jefferson (1963a,b)
Modified Jefferson index	$1.6 \times \theta_{w925} - T_{500} - 0.5\ (T - T_d)_{700} - 8$	Jefferson (1966)
Total totals index	$T_{850} + T_{d850} - 2T_{500}$	Miller (1967)
Cross totals index	$T_{d850} - T_{500}$	Miller (1967)
Level of neutral buoyancy (LNB)		
Lowest-level CAPE or lowest 50-/100-hPa CAPE (CAPE/CAPE50/CAPE100)	$g\int_{\mathrm{LFC}}^{LNB} \frac{T_v(\mathrm{parcel}) - T_v(\mathrm{env.})}{T_v(\mathrm{env.})}\,\mathrm{d}z$	Moncrieff and Miller (1976); Craven et al. (2002)
Moisture convergence	$\frac{\partial \rho u q}{\partial x} + \frac{\partial \rho v q}{\partial y}$	Van Delden (2001)
Shear	$\sqrt{(u_{500} - u_{500\mathrm{m}})^2 + (v_{500} - v_{500\mathrm{m}})^2}$	
Cosine of the wind direction at 850 hPa		

an effect on the statistics of the predictors and, therefore, on the forecast probabilities, which will be discussed in section 3.

The data from the first and third decades (10-day periods) of the warm months (from mid-April to mid-October) of 1999–2002 are used as the initial dependent set and the data from the second decades are used as the initial independent set. Both sets are used in the predictor selection process. Data from the second decades were used as the independent set instead of data from a whole season, because the horizontal HIRLAM resolution was changed in the beginning of 2002, and in this way the higher-resolution data are both in the dependent and independent datasets. After having selected the predictors for the MOS equations, the regression coefficients are updated using all of the data from the warm half-years of 1999–2002, that is, the eventual dependent (developmental) set. The warm season of 2003 is finally used as the eventual independent (verification) set. The verification results in this paper are all based on the latter dataset.

Most potential predictors contain one or more of the following ingredients that, together, are necessary and sufficient for the development of deep moist convection (and, hence, a thunderstorm): low static stability, moisture, and ascent of parcels to their level of free convection (LFC) by some lifting mechanism (Doswell 1987).

d. Selected predictors

In Tables 1 and 2 all of the predictors are shown that are selected in at least one (severe) thunderstorm forecast equation. This set of selected predictors is only a subset of the total (potential) predictor set that was investigated. Table 1 also includes the definitions of some traditional thunderstorm indices. The potential predictor set did not contain the following thunderstorm indices: the bulk Richardson number, convective inhibition, the K index, the lid strength index, the lifted index, the storm relative helicity, the severe weather threat (SWEAT) index, and the vertical totals index. These indices were not included, because they appeared to have no (additive) predictive potential (with respect to the other predictors) in a previous perfect prog (Wilks 1995) experiment (not shown). This does not mean that there could not have been useful information in these predictors in the MOS approach, but we chose to cull potential predictors in that way. The (severe) thunderstorm forecast equations contain at least 1 and at most 5 predictors. The maximum number of predictors has been set to 5, because more than 5 predictors often appeared to result in overfitting.

The square root of the ECMWF 6-h convective precipitation sum from 0000–0600/0600–1200/1200–1800 UTC turns out to be the most selected predictor in our thunderstorm forecast equations; it is included 286 times in a total of 348 forecast equations. The corresponding *large-scale* precipitation sum appeared to have no (additive) predictive potential (with respect to the other predictors) in the previous perfect prog experiment. Although the convection parameterization is changed quite often in NWP models, it was decided to include the ECMWF convective precipitation predictor, because it is the most important predictor. The last time the convection parameterization was changed in the ECMWF model was in January 2003 (Bechtold et al. 2003). As the warm season of 2003 is taken as the

TABLE 2. Overview of (derived) ECMWF and remaining predictors that are included in at least one (severe) thunderstorm forecast equation. Here u and v are the zonal and meridional wind components, respectively.

ECMWF predictors
Square root of 6-h convective precipitation sum from 0000–0600/0600–1200/1200–1800 UTC
Relative humidity at 850 hPa
Power-transformed relative humidity at 850 hPa
Meridional wind component at 850/300 hPa
Anomalous temperature at 1000 hPa
Anomalous geopotential height at 500 hPa
Wind speed at 1000/300 hPa
Advective heat 1000–850 hPa ($u_{850} \times v_{1000} - v_{850} \times u_{1000}$)
Curvature vorticity at 500/300 hPa
Shear vorticity at 300 hPa
Equivalent 500–850-hPa thickness
Temperature advection 500–1000 hPa
P27 score 2 (meridionality of the 500-hPa flow; Kruizinga 1979)
P27 score 3 (cyclonality of the 500-hPa flow; Kruizinga 1979)
Remaining predictors
(Co)sine of the day of the year

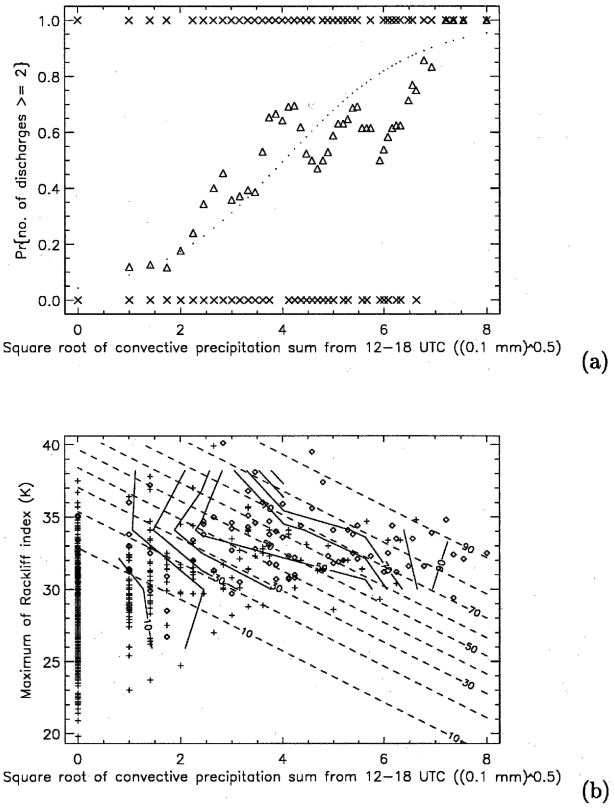

FIG. 3. (a) (Non) occurrence of thunderstorms in the period of 0900–1500 UTC (crosses) as a function of the first selected predictor, that is, the square root of the maximum 6-h convective precipitation sum from 1200 to 1800 UTC in the region E–MN, computed from the +30 h forecasts from the 1200 UTC ECMWF run. Additionally, the thunderstorm probabilities are indicated, as determined by either the running conditional means (triangles) or the fitted logistic regression function (dotted curve). (b) (Non) occurrence of thunderstorms in the period of 0900–1500 UTC, indicated by diamonds (pluses), as a function of the two selected predictors, that is, the square root of the maximum ECMWF 6-h convective precipitation sum from 1200 to 1800 UTC and the maximum of the HIRLAM Rackliff index in the region E–MN. The latter predictor has been computed from the +12 h forecasts from the 0000 UTC HIRLAM run. Also, the thunderstorm probabilities are indicated, as determined by either the conditional means (solid contours) or the fitted logistic regression function (dashed contours). All available data from the warm half-years of 1999–2002 are used in both (a) and (b).

independent verification set, we are able to check whether that change has a negative impact on the verification scores (see section 3). The strong relation between observed (convective) precipitation and observed lightning frequency was identified quite some time ago (e.g., Shackford 1960). Our study demonstrates that there is also a strong relation between the forecast convective precipitation by the ECMWF model and observed lightning, at least in the Netherlands, up to 3 days in advance.

Other predictors that are often selected in our thunderstorm forecast equations are the following thunderstorm indices, computed from the HIRLAM forecasts: the Jefferson index (included in 232 forecast equations), the level of neutral buoyancy (included in 99 forecast equations), and the Boyden index (included in 58 forecast equations). The Jefferson index assesses the latent instability of an air parcel at 925 hPa. The level of neutral buoyancy, or the limit of convection, is the level where a rising cloud parcel is about to get cooler than its environment and is used in the calculation of the convective available potential energy (CAPE; Table 1). The Boyden index accounts for pure conditional instability of a certain atmospheric layer and does not include moisture. However, because the Boyden index is combined with the convective precipitation sum in most equations and with other moisture-containing predictors in the remaining equations, the effect of moisture is included in the equations.

As an illustration of the logistic regression method, Fig. 3a shows the (non) occurrence of thunderstorms (i.e., the binary predictand indicated by crosses) in the period 0900–1500 UTC in the region E–MN, as a function of the first predictor (of two in total) that was selected by the forward stepwise selection method, using data from the warm half-years of 1999–2002. This first predictor is the square root of the maximum ECMWF 6-h convective precipitation sum from 1200 to 1800 UTC in the region E–MN. To visualize the relation between the binary predictand and the first predictor more clearly, running conditional means are shown as well (indicated by triangles in Fig. 3a), which can be interpreted as thunderstorm probabilities. For each of the predictor values the conditional mean has been determined by calculating the mean value of the binary predictand in the bin, centered around that predictor value. The bin size has been chosen to be one-tenth of the total range of predictor values. Addition-

ally, Fig. 3a shows the logistic regression curve, which is a smooth fit to the binary predictand and has a minimum of 4% for a convective precipitation sum equal to zero and a value of 95% for a convective precipitation sum equal to 6.4 mm.

The second and also last predictor that was selected in the regression equation under consideration is the maximum of the HIRLAM Rackliff index in the region E–MN. Like the Jefferson index, the Rackliff index assesses the latent instability of an air parcel at 925 hPa. Figure 3b shows again the (non) occurrence of thunderstorms, indicated by diamonds (pluses), the conditional means (solid contours), and the fitted logistic regression function (dashed contours), but now as a function of both predictors. It is evident that the inclusion of the second predictor in the equation is highly relevant. The minimum of the logistic regression function has decreased to 0%. High probabilities are forecast if both the Rackliff index and the convective precipitation sum are relatively high, as expected. However, there appear to be "forbidden" areas in the two-dimensional predictor space. One forbidden area consists of the combination of low Rackliff index values and high convective precipitation sums. Because low Rackliff index values correspond to nonconvective situations, it is logical that this area is not occupied. It is not clear, however, why the other area, that is, the combination of high Rackliff index values and high convective precipitation sums, is not occupied.

The predictor that has been selected most often in the *severe* thunderstorm forecast equations is the Boyden index, computed from the HIRLAM forecasts; it is included in 20 of the 42 forecast equations. Other important predictors are, again, the square root of the ECMWF 6-h convective precipitation sum from 0600–1200/1200–1800 UTC (included 14 times in the forecast equations), the ECMWF equivalent 500–850-hPa thickness (included in nine forecast equations), and the Bradbury index, computed from the HIRLAM forecasts (also included in nine forecast equations). Because the predictors in the severe thunderstorm forecast system are generally different from those in the thunderstorm forecast system for the same region and the same projection, the two systems contain independent information. Therefore, the performance of the combination of both systems is likely to be better than that of a system that would directly compute the absolute probabilities of severe thunderstorms.

It may seem surprising that the CAPE has not often been selected. This means that it does not have sufficient additive predictive potential with respect to the other predictors. However, quantities that are strongly related to the CAPE have often been selected in the equations, namely, the convective precipitation predictor and the level of neutral buoyancy. Still, as a single predictor CAPE appears to be a worse discriminator between the nonsevere thunderstorm cases on the one hand and the severe thunderstorm cases on the other hand than other predictors (e.g., Fig. 4). The Boyden index appears to be a good discriminator between those cases (e.g., Fig. 4a), despite the fact that it does not include moisture. Because the probability of a severe thunderstorm is conditional on the occurrence of a thunderstorm of any severity, it is not too surprising that a stability index without moisture can do a good job in forecasting the conditional probability of a severe thunderstorm. For a complete overview of all selected predictors, the reader is referred to Tables 1 and 2.

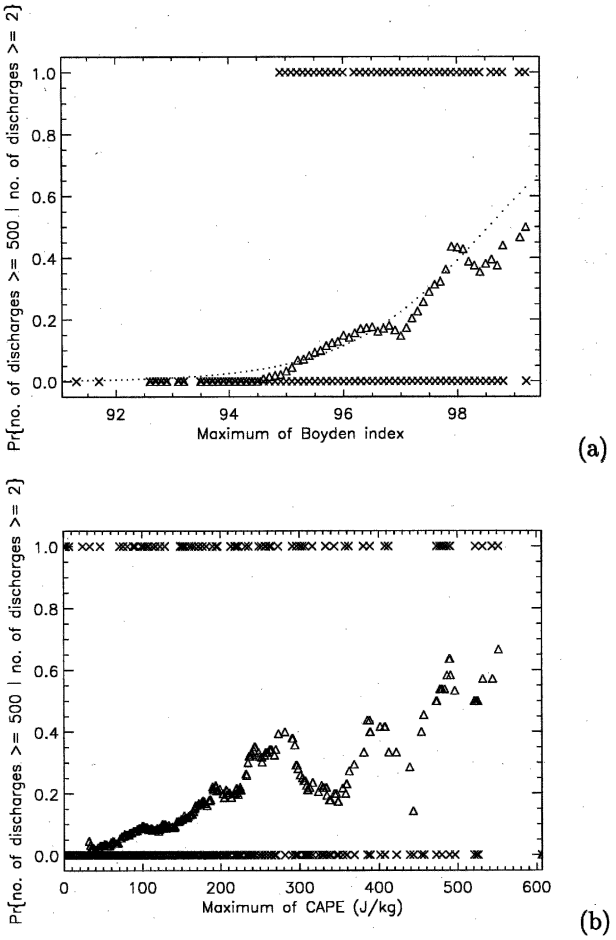

FIG. 4. (a) Occurrence of (non) severe thunderstorms in the period of 0900–1500 UTC (crosses) as a function of the first selected predictor, that is, the maximum Boyden index in each of the six land regions, computed from the +12 h forecasts from the 0000 UTC HIRLAM run. Additionally, the conditional probabilities of severe thunderstorms are indicated, as determined by either the running conditional means (triangles) or the fitted logistic regression function (dotted curve). (b) Same as in (a), but as a function of a nonselected predictor, namely, the maximum CAPE in each of the six land regions. All available data from the warm half-years of 1999–2002 are used in both (a) and (b).

3. Example of a (severe) thunderstorm forecast and objective verification results

In this section we present one case to demonstrate the MOS system for (severe) thunderstorms, and we

show some verification results of that system for the (independent) warm season of 2003.

a. Example of a (severe) thunderstorm forecast

The case that we show is from 2 June 2003. The +12 h and +18 h thunderstorm probability forecasts, computed by the 0000 UTC run of the MOS system, are shown in Figs. 5a and 5b, respectively. The +12 h and +18 h conditional probability forecasts of severe thunderstorms, computed by the same run, are shown in Figs. 5c and 5d, respectively. The highest thunderstorm probabilities have been forecast in the central and southern part of the Netherlands for the period of 0900–1500 UTC, and they have shifted to the east for the period of 1500–2100 UTC. The highest conditional probabilities of *severe* thunderstorms have been forecast in the southern part of the Netherlands for the period of 0900–1500 UTC, and these have also shifted to the east for the period of 1500–2100 UTC. To assess whether the forecast probabilities are higher than normal, they can be compared with the climatological probabilities (Fig. 2). In this case, the (conditional) probabilities of (severe) thunderstorms are higher than normal in almost all regions. The *absolute* probabilities of severe thunderstorms can be calculated by multiplying the thunderstorm probabilities (e.g., Fig. 5a) with the conditional probabilities of severe thunderstorms (Fig. 5c).

The locations of all lightning discharges, as detected by the SAFIR network, from 0900–1500 and 1500–2100 UTC are shown in Figs. 5e and 5f, respectively. In all but the most southwestern region (W–XS) lightning discharges have been observed in the period of 0900–1500 UTC, and in nine regions there have been at least 500 discharges detected by the SAFIR network. The observed pattern of lightning flashes has also shifted to the east in the period of 1500–2100 UTC, with 10 regions exceeding one flash and five regions (M–XN, E–XN, E–MN, E–MS, and M–MS) exceeding 500 flashes. Of course, (probability) forecasts cannot be verified using only one case, so we present now objective verification results for the (independent) warm season of 2003.

b. Verification results of the MOS thunderstorm forecast system

In this section some verification results of the MOS thunderstorm forecast system are presented for the (independent) warm season of 2003—to be more precise, the period from 19 May to 7 October 2003. As scalar verification scores, we have used the bias, the Brier score, two of its decomposition terms (reliability and resolution), and the Brier skill score (BSS). The definitions of these scores are given in the appendix. We also show attributes diagrams (Hsu and Murphy 1986). These are much more informative representations of forecast performance than scalar scores, because they are compact displays of the full distributions of forecasts and observations (Wilks 1995). Reliability, which measures the correspondence between the forecast probabilities and the event frequencies in each subsample, is an extremely important aspect of probability forecasts. If there are systematic deficiencies in the probability forecasts apparent from the attributes diagram, the forecast probabilities can, in principle, be remapped to their reliable verification frequencies (i.e., calibration). The second important decomposition term of the Brier score is the resolution. It measures the ability of different forecast probabilities to distinguish between different event frequencies. The last decomposition term of the Brier score, that is, the uncertainty, depends only on the variability of the observations and cannot be influenced by the forecast system.

In Figs. 6a–d we show attributes diagrams of the +6 to +24 h forecasts for the 0000 UTC run of the MOS thunderstorm forecast system for the region E–MN (Fig. 1a). Figure 6a shows the attributes diagram of the +6 h forecasts. Because the (sample) climatological probability is quite low in the period 0300–0900 UTC, the higher forecast probabilities are quite rare (see the histogram on the right of the figure). However, the Brier skill score is quite high (BSS = 39.7%). The bias is positive (3.7%), showing some overforecasting. The attributes diagram of the +12 h forecasts (Fig. 6b) shows a good skill and the resolution is quite good as well. This is the verification result for the regression equation that is visualized in Fig. 3b. The (sample) climatological probability is much higher in the period 0900–1500 UTC than in the 0300–0900 UTC period. The forecasts are quite reliable, and there is again a minor overforecasting bias. The +18 h forecasts (Fig. 6c) show good resolution, but at the price of degraded reliability, with substantial overforecasting, especially of the lower probabilities. The Brier skill score is 29.2%. The +24 h forecasts (Fig. 6d) also show an overforecasting bias. Because the resolution term is larger than the reliability term, the Brier score is smaller than the uncertainty term [see Eq. (A3) in the appendix], resulting again in a positive Brier skill score. For the +30, +42, and +48 h forecasts, the Brier (skill) score has increased (decreased) with respect to the +6, +18, and +24 h forecasts, respectively (Fig. 7a), as a result of a lower resolution, despite the increase in reliability (not shown).

To get an idea of the verification scores for the other regions, Fig. 7a (7b) shows the Brier skill scores (biases) for the 0000 UTC run of the MOS thunderstorm forecast system for all regions. Although there is a large variation in Brier skill scores between the regions, the average BSSs for both the six land regions (indicated by the solid line in Fig. 7a) and the six coastal regions (dotted line) are positive, except for the +30 and +48 h forecasts for the coastal regions. The (average) BSS shows a large diurnal cycle as the most prominent feature. Apart from this diurnal cycle there is generally a (slight) decrease in the (average) BSS from one projec-

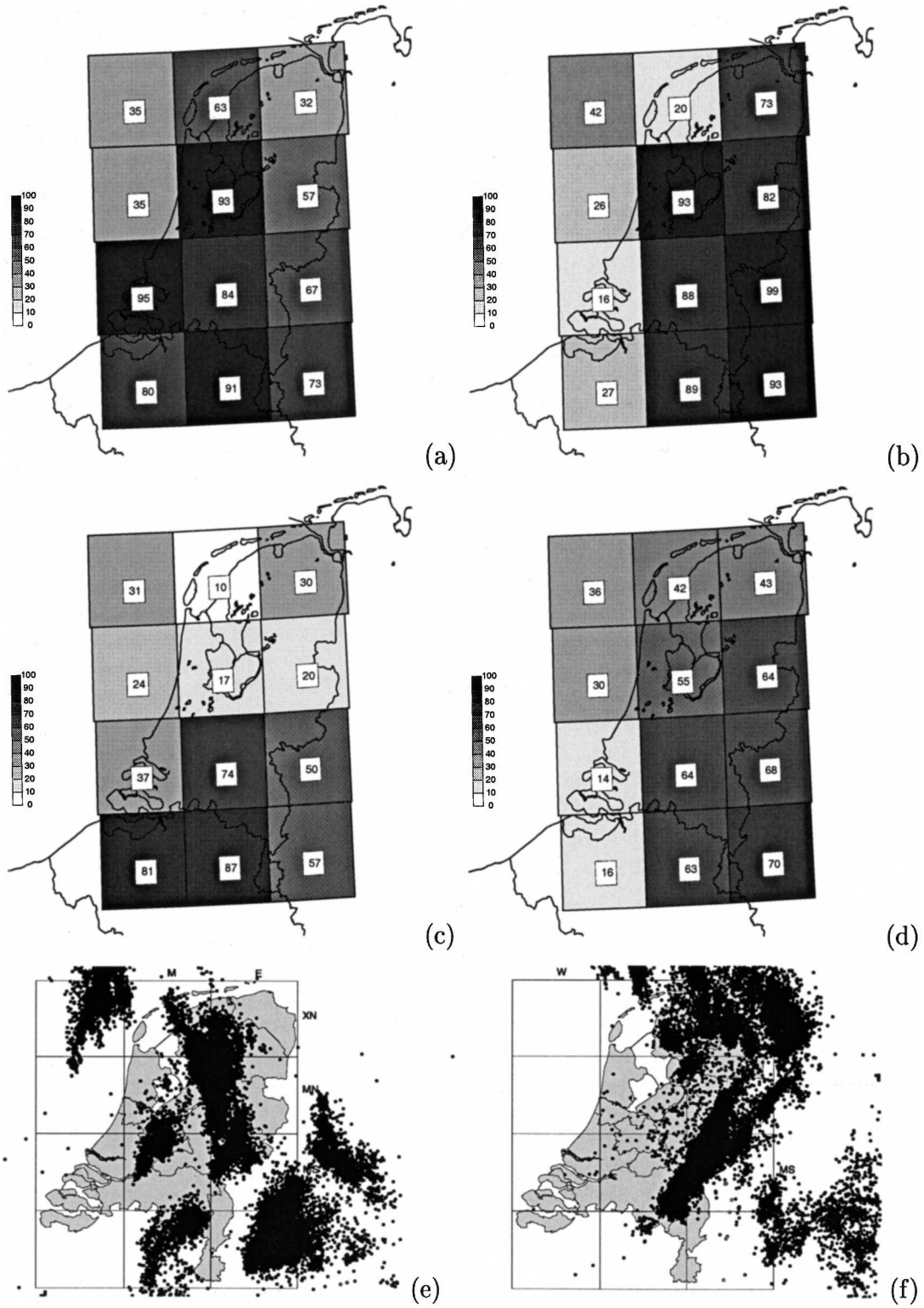

FIG. 5. Thunderstorm probability forecasts (%) of (a) +12 and (b) +18 h projections for (a) 0900–1500 and (b) 1500–2100 UTC on 2 Jun 2003. Conditional probability forecasts of severe thunderstorms (%) of (c) +12 and (d) +18 h projections for (c) 0900–1500 and (d) 1500–2100 UTC on the same day. These forecasts are computed by the 0000 UTC run of the MOS (severe) thunderstorm forecast system. The location of all lightning discharges (indicated by the black dots), as detected by the SAFIR network, is shown from (e) 0900–1500 and (f) 1500–2100 UTC on the same day.

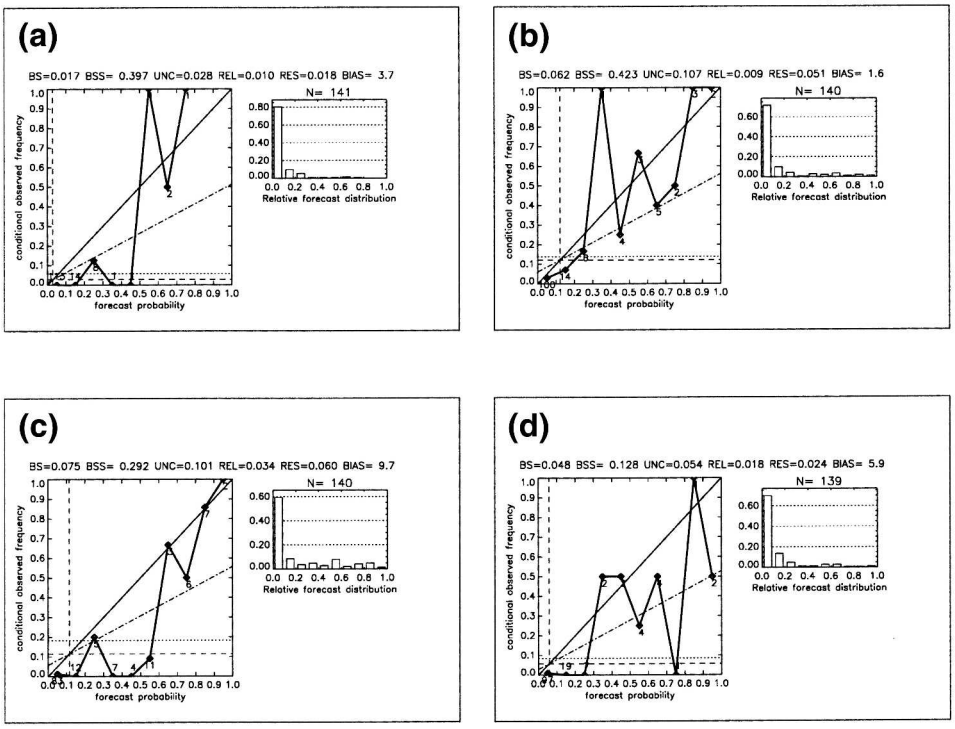

FIG. 6. Attributes diagrams of (a) +6, (b) +12, (c) +18, and (d) +24 h forecasts, as computed by the 0000 UTC run of the MOS thunderstorm forecast system for the E–MN region. The verification period is from 19 May to 7 Oct 2003. In these diagrams the observed frequencies of thunderstorm occurrence are shown, conditional on each of the 10 possible forecast probabilities (indicated by diamonds). For perfectly reliable forecasts these paired quantities are equal, yielding all points in the diagram falling on the diagonal line. The dotted line indicates the 1999–2001 climatology and the dashed line is the sample climatology. The dash-dotted line indicates the "no skill" line. The histogram on the right portrays the relative frequency of use of the forecasts. Here, UNC is uncertainty, REL is reliability, and RES is resolution [see Eq. (A3)], and N is the total number of cases.

tion to the projection 24 h later, as there should be. The only projections at which more than one region has a negative BSS are the +6, +30, and +48 h forecasts, verifying around 0600, 0600, and 0000 UTC, respectively. It is not surprising that these projections show the largest number of negative BSSs, because forecasting thunderstorms is most difficult for the night and especially the morning, when the relationship between the predictors and the occurrence of lightning is weakest, particularly in the coastal regions.

The bias (Fig. 7b) also shows a large variation between the regions, but for most regions it is positive, indicating overforecasting. One of the reasons for this overforecasting might be the increase in HIRLAM resolution, which, although present in about 30% of the development sample, has probably led to more extreme values of the thunderstorm indices in the verification sample and, therefore, to higher forecast probabilities than in the development sample. Another reason might be the change in the ECMWF model's convective parameterization, which leads to more convective precipitation in the presence of a stable surface layer (e.g., over land during the night; Bechtold et al. 2003) and, therefore, to higher forecast probabilities (in these cases) than in the development sample. However, Fig. 7b shows the highest biases in the evening, while the stable surface layer emerges only after sunset (i.e., late in the evening during the summer). Therefore, the increase in HIRLAM resolution might have a larger effect on the bias. Figure 7b shows also that the average bias is larger for the land regions than for the coastal regions. This implies that the skill of the MOS equations for the land regions could be further improved by calibration or by updating the regression coefficients when more data from the high-resolution HIRLAM are included, leading to smaller biases and, hence, to higher BSSs.

We can conclude that the overall skill of the 0000 UTC run of the MOS thunderstorm forecast system is good. Because of this good skill and the apparently minor effect on the bias, we can also conclude that there is no large systematic impact of the change in the ECMWF model's convective parameterization on the MOS system. The verification results for the 0600, 1200, and 1800 UTC runs are similar (not shown), showing overall a good skill as well.

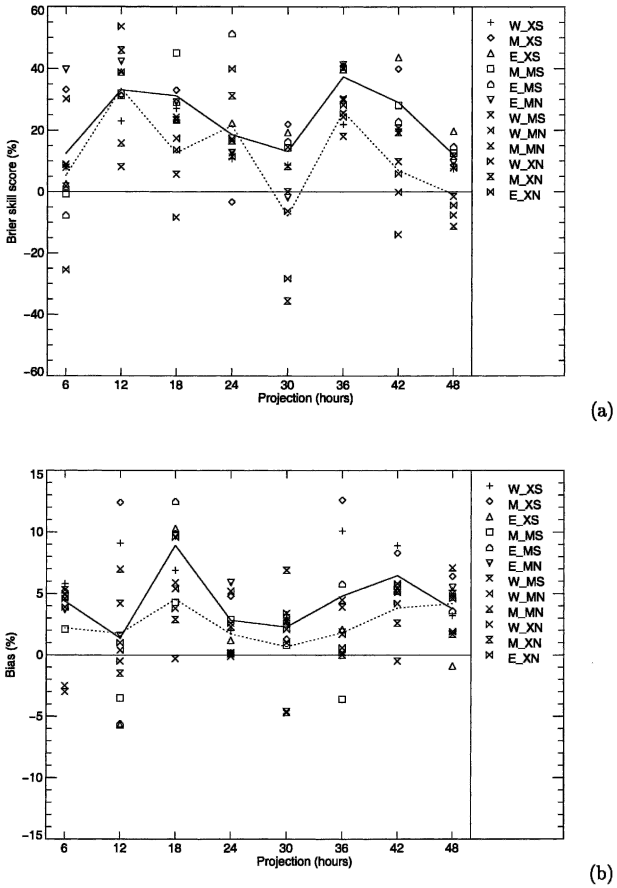

FIG. 7. (a) Brier skill score with respect to the 1999–2001 climatology (Fig. 2a) and (b) bias, as a function of projection, for the 0000 UTC run of the MOS thunderstorm forecast system for all 12 regions (indicated by the different symbols). The solid line represents the average (a) BSS or (b) bias for the six land regions, and the dotted line represents the average (a) BSS or (b) bias for the six coastal regions. The verification period is from 19 May to 7 Oct 2003.

c. Verification results of the MOS severe thunderstorm forecast system

Here some verification results of the MOS *severe* thunderstorm forecast system are presented for the same period as in the previous section. In Figs. 8a–f we show attributes diagrams of the +6 to +24 h forecasts for the 0000 UTC run of the MOS severe thunderstorm forecast system. Figure 8a shows the attributes diagram of the +6 h forecasts for all 12 pooled regions. Because the (sample) climatological probability is quite low in the period of 0300–0900 UTC, the higher forecast probabilities are quite rare. The Brier score is larger than the uncertainty term, mostly due to the bad resolution, resulting in a negative Brier skill score. The bias is positive (5.9%), showing overforecasting. The attributes diagram of the +12 h forecasts for the six pooled land regions (Fig. 8b) shows a good skill and the resolution is good as well. The sample climatological probability is much higher in the period of 0900–1500 UTC than in the 0300–0900 UTC period. The forecasts are quite reliable and have hardly any bias. Figure 8c displays the attributes diagram of the +12 h forecasts for the six pooled coastal regions. The lower probabilities are quite reliable, but the higher probabilities are unreliable and suffer from substantial overforecasting. This results in a positive bias and a negative Brier skill score, also because of the bad resolution. The +18 h forecasts for the six land and six coastal regions (Figs. 8d and 8e, respectively) have good skill, are quite reliable, and have a good resolution. Both have hardly any bias. The +24 h forecasts for all 12 pooled regions (Fig. 8f) are reliable and have a good resolution, resulting in a positive Brier skill score.

To get an idea of the verification scores for the other projections, Fig. 9a (9b) shows the Brier skill scores (biases) for the 0000 UTC run of the MOS severe thunderstorm forecast system for all projections. Because the regions have been pooled in this case (Fig. 9a), contrary to Fig. 7a, we have computed 90% bootstrap confidence intervals (Wilks 1995), in order to get an impression of the uncertainty associated with the BSSs. These confidence intervals are indicated by error bars in Fig. 9a. As was the case for the thunderstorm forecast system, the severe thunderstorm forecast system shows a large diurnal cycle in Brier skill scores (Fig. 9a). The only projections at which all regions (probably) have a negative BSS are the +6 and +30 h forecasts, verifying around 0600 UTC. It is not surprising that these projections show a negative BSS, because forecasting (severe) thunderstorms is most difficult for the morning, when the relationship between the predictors and the occurrence of a (severe) thunderstorm is weakest. More surprising is the negative (zero) BSS for the coastal regions at the +12 (+36) h projection, verifying around 1200 UTC. However, this result is in line with the results obtained during the derivation of the equations, in that the relationship between the predictors and the occurrence of a severe thunderstorm appeared to be rather weak during the afternoon. Figure 9b shows that the bias is positive, except for the +18 and +36 h forecasts, indicating overforecasting for most projections. We can conclude that the overall skill of the 0000 UTC run of the MOS severe thunderstorm forecast system is good. The verification results for the 0600, 1200, and 1800 UTC runs are similar (not shown), showing overall a good skill as well.

4. Summary and discussion

We have described the derivation and verification of MOS (Glahn and Lowry 1972; Wilks 1995) equations for the (conditional) probability of (severe) thunderstorms in the warm half-year (from mid-April to mid-October) in the Netherlands. For 12 regions of about 90 km × 80 km each (Fig. 1), and for projections out to

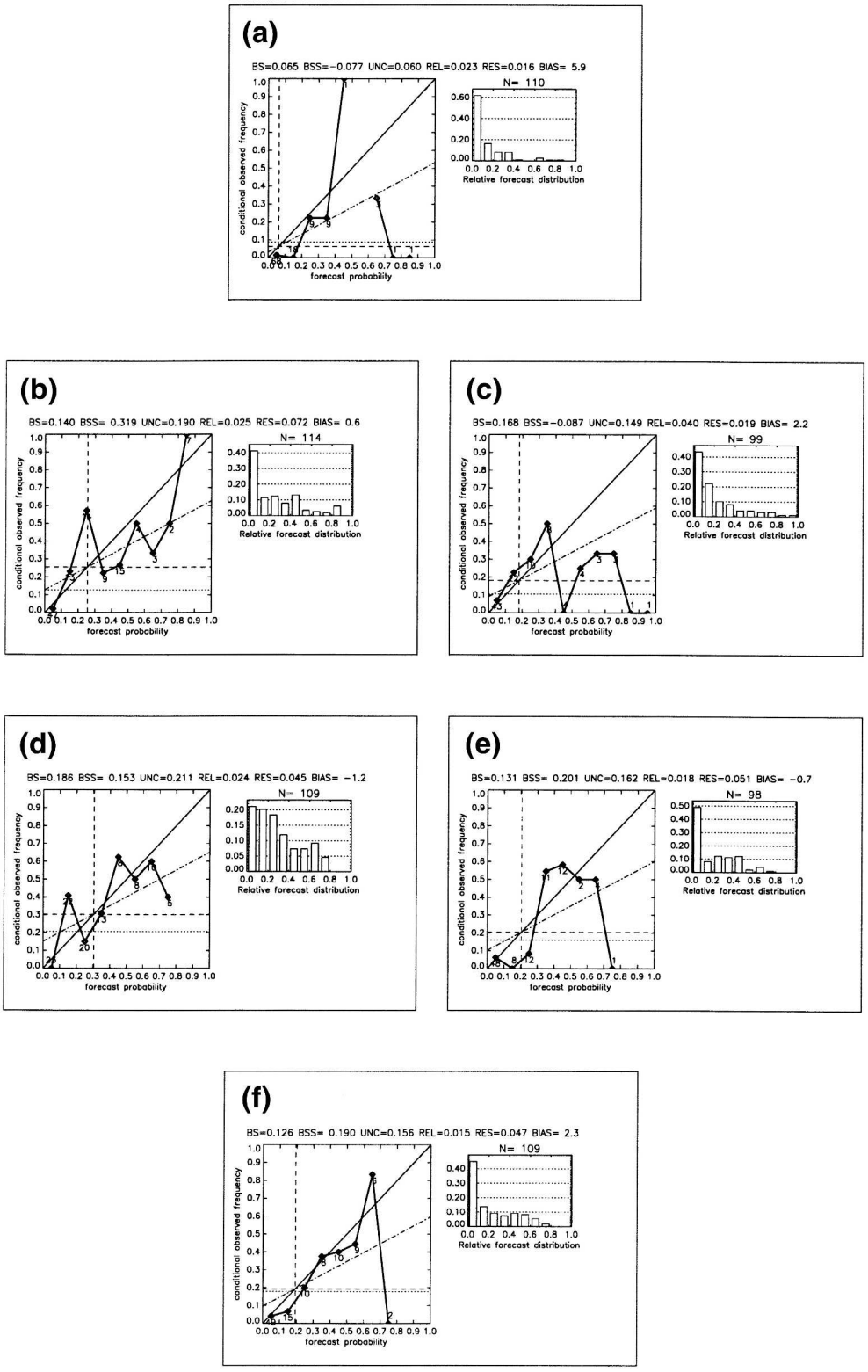

FIG. 8. Attributes diagrams of (a) +6 h forecasts for all regions, (b) +12 h forecasts for land regions, (c) +12 h forecasts for coastal regions, (d) +18 h forecasts for land regions, (e) +18 h forecasts for coastal regions, and (f) +24 h forecasts for all regions, as computed by the 0000 UTC run of the MOS severe thunderstorm forecast system. The verification period is from 19 May to 7 Oct 2003. See the caption of Fig. 6 for more information.

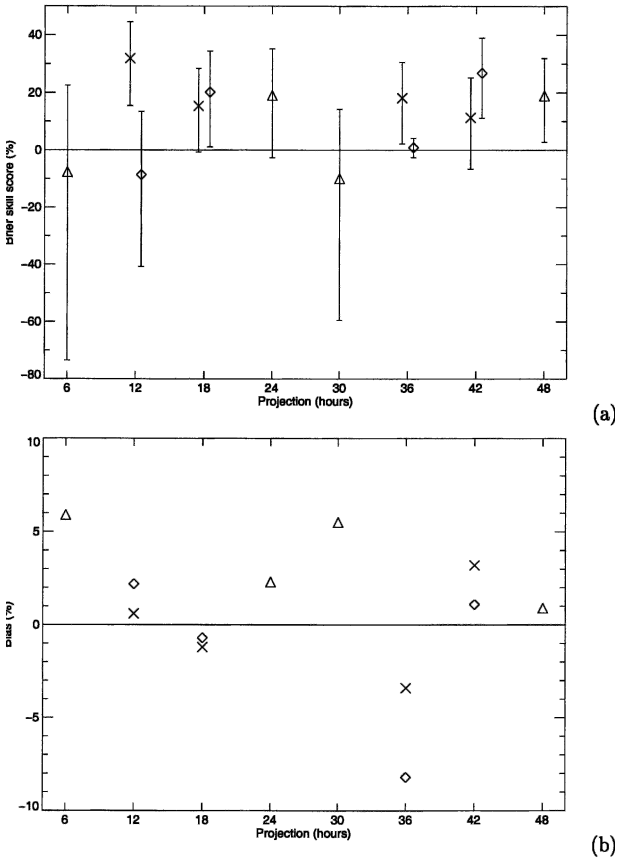

FIG. 9. (a) Brier skill score with respect to the 1999–2001 climatology (Fig. 2b) and (b) bias as a function of projection for the 0000 UTC run of the MOS severe thunderstorm forecast system for the six pooled land regions (crosses), for the six pooled coastal regions (diamonds), and for all 12 pooled regions (triangles). The error bars in (a) indicate the 90% bootstrap confidence intervals (Wilks 1995); the error bars belonging to the land and coastal regions are centered around the projection time for legibility. The verification period is from 19 May to 7 Oct 2003.

48 h in advance (with 6-h periods), we have developed these equations using combined (postprocessed) output from the HIRLAM (Undén et al. 2002) and ECMWF models as the potential predictor dataset. The predictands are derived from the SAFIR (Wessels 1998) lightning data, being either the probability of a thunderstorm (≥2 discharges) or the conditional probability of a *severe* thunderstorm (≥500 discharges) under the condition that ≥2 discharges will be detected. The equations have been derived using the method of logistic regression (Brelsford and Jones 1967; Wilks 1995), which is the preferred regression method in probabilistic forecasting (Applequist et al. 2002).

When using the MOS approach, the combination of output data from two different NWP models is an advantage and a disadvantage at the same time. The advantage is that the combination of output from two different NWP models usually leads to better forecasts (e.g., Thompson 1977; McCalla and Kalnay 1988), but the disadvantage is that the frequency of model changes is higher, potentially leading to less stable MOS equations. By including predictors in the forecast equations that are not too sensitive to model changes we hope to have minimized the latter effect.

The square root of the ECMWF 6-h convective precipitation sum is the most important predictor in the thunderstorm forecast system, and the Boyden index, computed from the HIRLAM forecasts, is the most important predictor in the *severe* thunderstorm forecast system (see Tables 1 and 2 for an overview of all selected predictors). The strong relation between observed (convective) precipitation and observed lightning frequency was identified quite some time ago (e.g., Shackford 1960). Our study demonstrates that there is also a strong relation between the forecast convective precipitation by the ECMWF model and observed lightning, at least in the Netherlands up to 3 days in advance (e.g., Fig. 3a). In examining several numerical forecasts of severe weather events in the United States, Kain et al. (2003) noted that convective rainfall coverage was often correctly predicted by the Eta Model when severe weather occurred, but small rainfall amounts generated by the convective parameterization scheme failed to indicate the severity of the convection. Here our MOS approach has a distinct advantage in that it objectively transforms the square root of the convective precipitation sum (together with other predictors) into the (conditional) probability of (severe) thunderstorms, as long as there is a distinct relation between the occurrence of (severe) thunderstorms and the square root of the convective precipitation sum.

Although the convection parameterization is changed quite often in NWP models, it was decided to include the ECMWF convective precipitation predictor, because it is the most important predictor. The last time the convection parameterization was changed in the ECMWF model was in January 2003 (Bechtold et al. 2003). Because the warm season of 2003 was taken as the independent verification set, it could be investigated whether there is a discernible (negative) effect of that change on the verification scores. From the verification results it is concluded that a negative effect cannot be discerned.

The square root of the HIRLAM 6-h convective precipitation sum has not been added as a potential predictor, because the convection parameterization is changed more often in HIRLAM than in the ECMWF model. To justify this decision, it has been tested for a few projections whether the HIRLAM convective precipitation predictor would have additive predictive potential with respect to the other predictors. It could be concluded that it generally had some additive predictive potential, but the increases in Brier skill scores were minor.

The Boyden index appears to be a good discriminator between the nonsevere thunderstorm cases on the one hand and the severe thunderstorm cases on the

other hand (e.g., Fig. 4a). Blanchard (1998) and Kain et al. (2003) also noted that the conditional probability of severe convection appears to increase as lapse rates in the lower to middle troposphere increase. Haklander and Van Delden (2003) did not discriminate between "ordinary" and severe thunderstorms, but they also concluded that the Boyden index is a good thunderstorm predictor.

Apart from the "saw tooth" pattern seen in several of our attributes diagrams (Figs. 6 and 8) due to the small sample size, the reliability curves generally have a positive slope. Mason (2004) showed that positive-sloping reliability curves correspond to positive values of the Brier skill score with random guessing as a strategy, which is intuitively appealing because of the implication that the conditional probability of the forecast event increases as the forecast probability increases. Moreover, our Brier skill scores are generally positive with respect to climatology, with the latter being a harsh standard. Mason (2004) showed that the Brier skill score is harsh because the expected value of this skill score is negative if nonclimatological forecast probabilities are issued.

We can conclude from the verification results (Figs. 6–9) that the overall skill of the 0000 UTC run of the MOS (severe) thunderstorm forecast system is good. Forecasting thunderstorms is most difficult for the night (2100–0300 UTC), and especially the morning (0300–0900 UTC), when the relationship between the predictors and the occurrence of lightning is weakest, particularly in the coastal regions. Forecasting *severe* thunderstorms is most difficult for the morning in all regions and for the afternoon (0900–1500 UTC) in the coastal regions. Because the overall verification results for the 0600, 1200, and 1800 UTC runs of the MOS (severe) thunderstorm forecast system are also good, the system was made operational at KNMI in April 2004. However, the system shows (slight) overforecasting, presumably as a result of the increase in HIRLAM resolution in the beginning of 2002. This can be eliminated by calibration or by updating the regression coefficients when more data from the high-resolution HIRLAM are included.

The Brier skill scores of the U.S. GSM-based MOS 6-h thunderstorm forecast system (from 12 to 48 h in advance) are between 6% and 15% (Hughes 2001), and the U.S. Eta Model–based MOS 6-h thunderstorm forecast system shows similar BSSs (Hughes 2002). Of course, a direct comparison cannot be made, but it is evident from Fig. 7a that our MOS thunderstorm forecast system shows higher BSSs, apart from the morning times. Because similar predictors were selected in the regression equations of Hughes (2001, 2002), this is not likely to explain the differences in BSSs. The higher BSSs of our MOS system may result from the fact that we have developed separate thunderstorm forecast equations for each region. Other reasons might be that we have combined output from two different NWP models, that we have used a different statistical method, or a combination of these.

Finally, future developments in our MOS system may be an increase in the temporal resolution from 6 to 3 h, an increase in the spatial resolution (dependent on projection), and the inclusion of previously detected discharges in surrounding regions as a potential predictor set for the 0–3-h projections.

Acknowledgments. The authors thank the following people at KNMI for their help in the Indices for Extreme Convective Situations (INDECS) project, of which the most important results are described in this paper. We thank Rudolf van Westrhenen, Erik van Meijgaard, and Cisco de Bruijn for programming part of the software to compute the thunderstorm indices; Iwan Holleman for general information about the SAFIR data and for supplying the software to make transformations between coordinate systems; and Frans van der Wel for a short introduction to Geographic Information Systems (GIS). Seijo Kruizinga, Jeanette Onvlee, and the anonymous reviewers are thanked for their comments on earlier versions of the manuscript.

APPENDIX

A Definitions of Some Verification Scores

In this appendix the definitions of the bias and Brier (skill) score are given. The bias is defined as the difference between the mean forecast probability and the observed frequency,

$$\text{Bias} = \overline{p^f} - \overline{p^o}, \tag{A1}$$

where p^f is the forecast probability and p^o is the observation of the event, being either 0 or 1.

The Brier score (BS) is defined as the mean-squared error of the probability forecasts,

$$\text{BS} = \frac{1}{N}\sum_{j=1}^{N}(p_j^f - p_j^o)^2. \tag{A2}$$

The Brier score is a strictly proper score (Wilks 1995) and is negatively oriented, with perfect forecasts exhibiting BS=0. The Brier score in Eq. (A2) can be decomposed into three terms (Murphy 1973; Wilks 1995), schematically written as

$$\text{BS} = \text{reliability} - \text{resolution} + \text{uncertainty}. \tag{A3}$$

In this equation the reliability term has a negative orientation and consists of a weighted average of the squared differences between the forecast probabilities and the relative frequencies of the forecast event in each subsample (with a total of 10 subsamples in our case). The resolution term in Eq. (A3) has a positive orientation and summarizes the ability of the forecasts

to discern subsample forecast periods with different relative frequencies of the event. The uncertainty term in Eq. (A3) is related to the sample climatology. Therefore, it depends only on the variability of the observations and cannot be influenced by the forecast system. For more information on the Brier score decomposition, the reader is referred to Wilks (1995).

The Brier score of a forecast system (BS) is often compared to that of a reference system (BS_{ref}) and expressed in the form of a skill score, the Brier skill score (BSS),

$$BSS = \frac{BS - BS_{ref}}{BS_{perf} - BS_{ref}} = 1 - \frac{BS}{BS_{ref}}, \qquad (A4)$$

because the Brier score of perfect forecasts (BS_{perf}) equals 0. As usual, the climatological relative frequencies are used as the reference forecasts in this study.

REFERENCES

Andersson, T., M. Andersson, C. Jacobsson, and S. Nilsson, 1989: Thermodynamic indices for forecasting thunderstorms in southern Sweden. *Meteor. Mag.,* **116,** 141–146.

Applequist, S., G. E. Gahrs, R. L. Pfeffer, and X.-F. Niu, 2002: Comparison of methodologies for probabilistic quantitative precipitation forecasting. *Wea. Forecasting,* **17,** 783–799.

Bechtold, P., F. Lalaurette, A. Ghelli, and M. Miller, 2003: Forecasts of severe convection. *ECMWF Newsletter*, No. 98, ECMWF, Reading, United Kingdom, 8–16.

Blanchard, D. O., 1998: Assessing the vertical distribution of convective available potential energy. *Wea. Forecasting,* **13,** 870–877.

Boyden, C. J., 1963: A simple instability index for use as a synoptic parameter. *Meteor. Mag.,* **92,** 198–210.

Bradbury, T. A. M., 1977: The use of wet-bulb potential temperature charts. *Meteor. Mag.,* **106,** 233–251.

Brelsford, W. M., and R. H. Jones, 1967: Estimating probabilities. *Mon. Wea. Rev.,* **95,** 570–576.

Collier, C. G., and R. B. E. Lilley, 1994: Forecasting thunderstorm initiation in north-west Europe using thermodynamic indices, satellite and radar data. *Meteor. Appl.,* **1,** 75–84.

Craven, J. P., R. E. Jewell, and H. E. Brooks, 2002: Comparison between observed convective cloud-base heights and lifting condensation level for two different lifted parcels. *Wea. Forecasting,* **17,** 885–890.

Doswell, C. A., III, 1987: The distinction between large-scale and mesoscale contribution to severe convection: A case study example. *Wea. Forecasting,* **2,** 3–16.

Galway, J. G., 1956: The lifted index as a predictor of latent instability. *Bull. Amer. Meteor. Soc.,* **37,** 528–529.

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

Haklander, A. J., and A. Van Delden, 2003: Thunderstorm predictors and their forecast skill for the Netherlands. *Atmos. Res.,* **67–68,** 273–299.

Hsu, W.-R., and A. H. Murphy, 1986: The attributes diagram: A geometrical framework for assessing the quality of probability forecasts. *Int. J. Forecasting,* **2,** 285–293.

Hughes, K. K., 2001: Development of MOS thunderstorm and severe thunderstorm forecast equations with multiple data sources. Preprints, *18th Conf. on Weather Analysis and Forecasting,* Fort Lauderdale, FL, Amer. Meteor. Soc., 191–195.

——, 2002: Automated gridded forecast guidance for thunderstorms and severe local storms based on the Eta model. Preprints, *21st Conf. on Severe Local Storms,* San Antonio, TX, Amer. Meteor. Soc., J19–J22.

Huntrieser, H., H. H. Schiesser, W. Schmid, and A. Waldvogel, 1997: Comparison of traditional and newly developed thunderstorm indices for Switzerland. *Wea. Forecasting,* **12,** 108–125.

Jacovides, C. P., and T. Yonetani, 1990: An evaluation of stability indices for thunderstorm prediction in Greater Cyprus. *Wea. Forecasting,* **5,** 559–569.

Jefferson, G. J., 1963a: A further development of the instability index. *Meteor. Mag.,* **92,** 313–316.

——, 1963b: A modified instability index. *Meteor. Mag.,* **92,** 92–96.

——, 1966: Letter to the editor. *Meteor. Mag.,* **95,** 381–382.

Kain, J. S., M. E. Baldwin, and S. J. Weiss, 2003: Parameterized updraft mass flux as a predictor of convective intensity. *Wea. Forecasting,* **18,** 106–116.

Kruizinga, S., 1979: Objective classification of daily 500 mbar patterns. Preprints, *Sixth Conf. on Probability and Statistics in the Atmospheric Sciences,* Banff, AB, Canada, Amer. Meteor. Soc., 126–129.

Lemcke, C., and S. Kruizinga, 1988: Model output statistics forecasts: Three years of operational experience in the Netherlands. *Mon. Wea. Rev.,* **116,** 1077–1090.

Mason, S. J., 2004: On using "climatology" as a reference strategy in the Brier and ranked probability skill scores. *Mon. Wea. Rev.,* **132,** 1891–1895.

McCalla, C., and E. Kalnay, 1988: Short and medium range forecast skill and the agreement between operational models. Preprints, *Eighth Conf. on Numerical Weather Prediction,* Baltimore, MD, Amer. Meteor. Soc., 634–640.

Miller, R. C., 1967: Notes on analysis and severe storm forecasting procedures of the Military Weather Warning Center. AWS, U.S. Air Force Tech. Rep. 200, 94 pp.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The new ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.,* **122,** 73–119.

Moncrieff, M. W., and M. J. Miller, 1976: The dynamics and simulation of tropical cumulonimbus and squall lines. *Quart. J. Roy. Meteor. Soc.,* **102,** 373–394.

Murphy, A. H., 1973: A new vector partition of the probability score. *J. Appl. Meteor.,* **12,** 595–600.

Rackliff, P. G., 1962: Application of an instability index to regional forecasting. *Meteor. Mag.,* **91,** 113–120.

Reap, R. M., 1994: 24-h NGM based probability and categorical forecasts of thunderstorms and severe local storms for the contiguous U.S. NWS Technical Procedures Bulletin, No. 419, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 14 pp.

Shackford, C. R., 1960: Radar indications of a precipitation-lightning relationship in New England thunderstorms. *J. Atmos. Sci.,* **17,** 15–19.

Showalter, A. K., 1953: A stability index for thunderstorm forecasting. *Bull. Amer. Meteor. Soc.,* **34,** 250–252.

Thompson, P. D., 1977: How to improve accuracy by combining independent forecasts. *Mon. Wea. Rev.,* **105,** 228–229.

Undén, P., and Coauthors, 2002: HIRLAM-5 scientific documentation. SMHI, 144 pp.

Van Delden, A. J., 2001: The synoptic setting of thunderstorms in Western Europe. *Atmos. Res.,* **56,** 89–110.

Wessels, H. R. A., 1998: Evaluation of a radio interferometry lightning positioning system. KNMI Scientific Rep. WR-98-04, 26 pp.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction.* Academic Press, 467 pp.