Probabilistic Forecasts of (Severe) Thunderstorms for the Purpose of Issuing a Weather Alarm in the Netherlands

MAURICE J. SCHMEITS, KEES J. KOK, DAAN H. P. VOGELEZANG, AND RUDOLF M. VAN WESTRHENEN

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

(Manuscript received 12 December 2007, in final form 20 May 2008)

ABSTRACT

The development and verification of a new model output statistics (MOS) system is described; this system is intended to help forecasters decide whether a weather alarm for severe thunderstorms, based on high total lightning intensity, should be issued in the Netherlands. The system consists of logistic regression equations for both the probability of thunderstorms and the conditional probability of severe thunderstorms in the warm half-year (from mid-April to mid-October). These equations have been derived for 12 regions of about 90 km × 80 km each and for projections out to 12 h in advance (with 6-h periods). As a source for the predictands, reprocessed total lightning data from the Surveillance et d'Alerte Foudre par Interférométrie Radioélectrique (SAFIR) network have been used. The potential predictor dataset not only consisted of the combined postprocessed output from two numerical weather prediction (NWP) models, as in previous work by the first three authors, but it also contained an ensemble of advected radar and lightning data for the 0–6-h projections. The NWP model output dataset contained 17 traditional thunderstorm indices, computed from a reforecasting experiment with the High-Resolution Limited-Area Model (HIRLAM) and postprocessed output from the European Centre for Medium-Range Weather Forecasts (ECMWF) model. Brier skill scores and attributes diagrams show that the skill of the MOS thunderstorm forecast system is good and that the *severe* thunderstorm forecast system generally is also skillful, compared to the 2000–04 climatology, and therefore, the preoperational system was made operational at the Royal Netherlands Meteorological Institute (KNMI) in 2008.

1. Introduction

Thunderstorms are one of the least predictable weather phenomena, especially if they are severe. They may cause damage to property and electric utilities, and endanger humans and livestock. In the Netherlands, thunderstorms occur quite frequently during late spring, summer, and early autumn, whereas severe thunderstorms are quite rare. Unlike the United States, where the severe thunderstorm criterion is based on the occurrence of large hail (>0.75 in.), convective winds (>50 kt), and/or tornadoes, the Dutch criterion for severe thunderstorms is based on high total lightning intensity. It is described in more detail below, but first the connection between total lightning intensity and the U.S. criterion for severe thunderstorms is discussed.

Williams et al. (1999) showed (their Fig. 3) that, to a large extent, Florida thunderstorms can be subdivided into nonsevere and severe categories (according to the U.S. definition) on the basis of maximum total lightning intensity alone. To be more precise: no severe cases were found with a maximum total lightning intensity <60 discharges per minute, and for higher lightning intensities the majority of the cases were identified as severe. If these numbers can be transferred, the majority of the Dutch cases with maximum total lightning intensities above 60 discharges per minute would also be severe according to the U.S. definition. It is outside the scope of this study to investigate if that is indeed the case. Of course, there are differences in the severe thunderstorm climatologies of the Netherlands and Florida. In the Netherlands, tornadoes are very rare, but large hail and especially severe gusts occur regularly during severe thunderstorms.

In the case of severe weather, the Royal Netherlands Meteorological Institute (KNMI) is responsible for issuing warnings to the Dutch society. If the (subjective)

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

DOI: 10.1175/2008WAF2007102.1

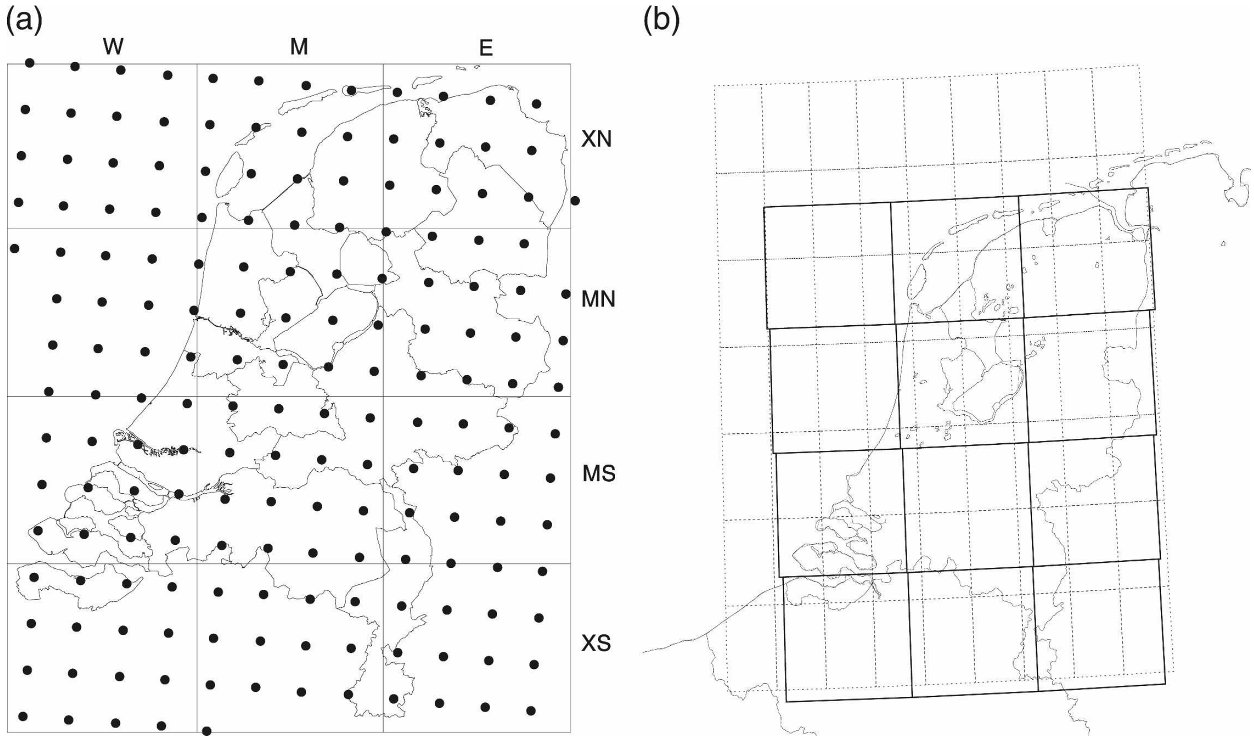


FIG. 1. (a) Geography of the Netherlands and surroundings, subdivided into 12 regions [west (W), middle (M), east (E), north (N), south (S), and extreme (X)]. The province boundaries of the Netherlands are also indicated. The black circles indicate the HIRLAM grid points with 22-km horizontal resolution. (b) Subdivisions are as in (a) (solid rectangles) but in a different coordinate system, and the dotted rectangles show the ECMWF grid at a horizontal resolution of 1/2° (adapted from Schmeits et al. 2005).

forecast probability of severe weather is ≥90% within a square of 50 km × 50 km anywhere in the Netherlands within the next 12 h, a so-called weather alarm is issued. The current weather alarm criterion for severe thunderstorms is a maximum total lightning intensity of at least 500 discharges per 5 minutes within a square of 50 km × 50 km. As this criterion is met only 2 days a year on average in the Netherlands, standard statistical techniques (like logistic regression; Brelsford and Jones 1967; Wilks 2006) are not expected to lead to skillful forecasts, and therefore we have used another (but relatively close) criterion for severe thunderstorms in this study.

We have investigated what the maximum intensity threshold is that renders a skillful statistical forecast system. This has led to the following predictand definition for severe thunderstorms: the conditional probability of a severe thunderstorm (maximum lightning intensity ≥50, and for some 6-h periods also ≥100 and ≥200 discharges per 5 minutes) in a 90 km × 80 km region (Fig. 1) in a 6-h period, under the condition that ≥2 discharges will be detected in the same region in the same 6-h period (see section 2). This new criterion is different from the criterion for severe thunderstorms used by Schmeits et al. (2005) (≥500 discharges in a 6-h period), and is closer to the weather alarm criterion, in the sense that it is also based on 5-min lightning intensity thresholds. Although the predictand for severe thunderstorms is necessarily less extreme than the weather alarm criterion, it is expected that a newly developed forecast system, based on model output statistics (MOS), can be used as a tool to decide whether a weather alarm for severe thunderstorms should be issued. The MOS technique (Glahn and Lowry 1972; Wilks 2006) 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.

The predictand definition for thunderstorms is the same as in Schmeits et al. (2005, hereafter referred to as SKV05) and is defined as the probability of a thunderstorm (≥2 lightning discharges) in a 90 km × 80 km region (Fig. 1) in a 6-h period. As a source for the predictands, we have used reprocessed total lightning data from the Surveillance et Alerte Foudre par Interférométrie Radioélectrique (SAFIR) network (Noteboom 2006; Wessels 1998).

Since 2004, automated (severe) thunderstorm probability forecasts based on MOS have been produced during the warm half year (defined as the period from mid-April to mid-October) for the 90 km × 80 km regions in the Netherlands, out to 48 h in advance with 6-h periods (SKV05). This operational forecast system runs 4 times a day and is based on combined postprocessed output from the High-Resolution Limited Area Model (HIRLAM; Undén and Coauthors 2002) and the European Centre for Medium-Range Weather Forecasts (ECMWF) model.

In this paper we describe the development and verification of a new MOS system for both the probability of thunderstorms and the conditional probability of severe thunderstorms in the warm half-year in the Netherlands, out to 12 h in advance with 6-h periods. As in the development of the SKV05 system, the potential predictor set contains a set of traditional thunderstorm indices, computed from the HIRLAM reforecasting dataset, postprocessed output from the ECMWF model, the sine and cosine 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 3). A new feature is an ensemble of 18 members of advected radar and lightning data as potential predictor sources for the 0–6-h projections (see section 3). Because of the use of these advected observations as potential predictors, the run frequency can be higher than that of the operational system of SKV05 and is currently 8 times per day. The predictand definition for severe thunderstorms has also changed, as mentioned above.

In the late 1970s, an early generation MOS system for severe local storms was developed in the United States, based on manually digitized radar data, surface observations, and output from the Limited-Area Fine Mesh (LFM) model (Charba 1979). Later, a more-advanced MOS system was developed, based on radar reflectivity, lightning data, surface observations, and output from the Eta Model (Kitzmiller et al. 2002). Our study differs from the latter study in that we have combined the output from two NWP models, and we have used an ensemble of advected radar and lightning data instead of deterministically advected radar and lightning data. In addition, we have used a different definition for severe thunderstorms, based on total lightning intensity instead of the U.S. definition; we have not used surface observations; 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. To our knowledge, our study is the first that addresses exceedance probabilities of (on the order of minutes^{-1}) lightning intensity.

In section 2 the logistic regression method, the predictands, and their climatology are described, and in section 3 the potential and selected predictors are addressed. In section 4 we present an example of the performance of the objective thunderstorm forecast system during a day with severe weather, and in section 5 some of the verification results for the independent period from 1 July 2005 to 31 July 2007 are shown (warm half-years only). Finally, in section 6 we summarize and discuss the results.

2. Statistical method, predictands, and climatology

a. Logistic regression

The derivation of the MOS equations has been performed using the method of logistic regression (Brelsford and Jones 1967; Wilks 2006). According to this method, the probability Pr that an event y occurs is

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

The predictors x_i ($i = 1, 2, \ldots, n$) are selected via a so-called forward stepwise selection method (Wilks 2006). At each step, a predictor is chosen that produces the best regression in conjunction with the predictors chosen on previous steps; herewith, 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, an iterative method that maximizes the product of all computed probabilities of the (non)occurrence of the event in the dependent dataset. For an illustration of the logistic regression method, the interested reader is referred to Figs. 3 and 4 of SKV05. The datasets used in this study are (reprocessed) total lightning data from the SAFIR network (section 2b; Wessels 1998; Noteboom 2006), combined data from the two radars in the Netherlands (section 2b; Holleman 2007), and postprocessed output data from a reforecasting experiment with HIRLAM (Undén and Coauthors 2002) and from the ECMWF model (section 3a).

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 (0000–0600, 0300–0900, 0600–1200, 0900–1500, 1200–1800, 1500–2100, 1800–0000, or 2100–0300 UTC) in a region (Fig. 1). Herewith, both horizontal and vertical lightning discharges are taken

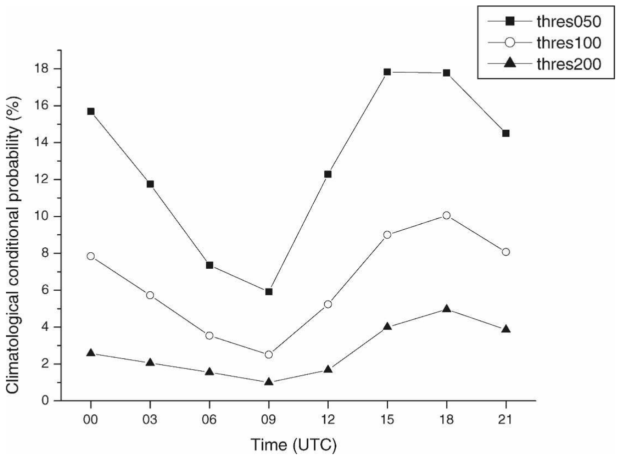


FIG. 2. Climatological conditional probability of severe thunderstorms (%) as a function of the central time of the corresponding 6-h period for the three different thresholds (50, 100, and 200 discharges per 5 minutes). All regions have been pooled. The climatologies have been computed using SAFIR reprocessed total lightning data (Noteboom 2006) from the warm half-years of 2000–04.

into account. The definition of a thunderstorm event is equal to the one used in SKV05. The predictand for thunderstorms is defined as the probability of a thunderstorm event.

An event is called a severe thunderstorm event if ≥50 lightning discharges per 5 minutes are detected by the SAFIR network in a 6-h time period in a region. Higher lightning intensities occur predominantly during the afternoon and evening (Fig. 2). Therefore, higher thresholds could also be handled statistically for the 1200–1800, 1500–2100, and 1800–0000 UTC periods (local summertime is UTC + 2 h), namely 100 and 200 discharges per 5 minutes. These criteria are different from the criterion used in SKV05, as outlined in the introduction. The predictand for severe thunderstorms is defined as the conditional probability of a severe thunderstorm event under the condition of a thunderstorm event.

Not only are the intensity thresholds in the MOS system lower than the weather alarm threshold, the areal size is also a factor of 3 larger (section 1). The areal size is larger because the thunderstorm frequency would otherwise become so low in the night and morning that there would be too few thunderstorm cases left to derive statistical equations. On the other hand, the 90×80 km^2 regions in the MOS system are specified a priori, while the weather alarm criterion is defined for *any* area of 50×50 km^2.

The climatological thunderstorm probabilities range from 4%–9% in the 0300–0900 UTC period to 6%–22% in the 1200–1800 UTC period, dependent on the region. During the afternoon and evening, the probabilities increase southward as a result of the climatological 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. These probabilities are not shown here because they are similar to the ones shown in Fig. 2a of SKV05.

As the predictand definition for severe thunderstorms is different from the one used in SKV05, we show the climatological probabilities of severe thunderstorms in Fig. 2. In this figure, a pronounced daily cycle can be seen with the highest climatological probabilities in the afternoon and evening periods. As the severe thunderstorm sample size is relatively small for each region separately, the regions have been pooled. Because the climatological *absolute* probabilities of severe thunderstorms are very small, for example, only between 0.2% and 4% for the threshold of 50 discharges per 5 minutes (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, for example, between 6% and 18% for the threshold of 50 discharges per 5 minutes (Fig. 2), and therefore conditional probabilities are used for severe thunderstorms.

Separate severe thunderstorm forecast equations have been derived for each projection and run time, resulting in a total of 28 forecast equations. No pooling has been performed in the derivation of the equations for the probability of ≥2 discharges, except for the initial reduction in the number of potential predictors to assure some spatial consistency. Separate thunderstorm forecast equations have been derived for each region, projection, and run time, resulting in a total of 192 forecast equations. The 2000–04 climatological probabilities (e.g., Fig. 2) are used as the reference forecasts in the verification of the regression equations (section 5).

3. Predictors

a. Potential predictors

The HIRLAM output was used every 3 h from 6 to 15 h in advance from the 0000, 0600, 1200, and 1800 UTC cycles. We have computed a set of traditional thunderstorm indices from the HIRLAM reforecasting dataset. Subsequently, we have calculated the minimum, maximum, and average value of each index using all grid points in each of the 12 regions in Fig. 1a. These minimum, maximum, and average values of the indices are then used as potential predictors.

The ECMWF model output was used every 3 h from

12 to 39 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). Additionally, three of the so-called P27 scores (Kruizinga 1979) have been used as potential predictors, as have the sine and cosine of the day of the year. The 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, 0300, 0600, 0900, 1200, 1500, 1800, or 2100 UTC. For more information on the P27 classification, the reader is referred to Kruizinga (1979). As in SKV05, 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 39 h.

The HIRLAM output had a horizontal resolution of 22 km until October 2006 (Fig. 1a) and has operated at 11-km resolution from October 2006 to the present. The ECMWF model's horizontal resolution was T511 until February 2006 and has been T799 from February 2006 to the present, but the ECMWF output is used at a resolution of 1/2° (Fig. 1b). HIRLAM had 40 nonequidistant levels in the vertical direction and has had 60 levels since October 2006, while the ECMWF model had 60 vertical levels and has had 91 levels since February 2006. The resolution changes have an effect on the statistics of the predictors and therefore on the forecast probabilities, which will be discussed in section 5.

A new feature, compared with SKV05, is an ensemble of 18 members of advected radar and lightning data as potential predictor sources for the 0–6-h projections. The 5-min lightning and radar images at 0240, 0540, 0840, 1140, 1440, 1740, 2040 or 2340 UTC have been used as initial conditions. Subsequently, both the lightning and the radar image are advected for the following 6-h period using both the HIRLAM 700-hPa wind vectors and vectors computed from previous radar images. In addition to these basic vectors, vectors with magnitudes $\pm25\%$ and directions $\pm10°$ have been used. This leads to a total number of 18 ensemble members for both the advected lightning and radar data. Apart from potential predictors that are based on individual ensemble members, potential predictors representing characteristics of the total ensemble have also been used.

The rationale for this ensemble approach is now described. As is well known (e.g., Golding 2000), advecting cells with heavy precipitation and/or lightning for more than a couple of hours does not have much predictive value in a deterministic sense. This is due to the limited life cycle of individual convective cells or showers. In a probabilistic sense, however, there may be information in advecting them for 6 h or even longer. The reason is that this information does not originate from the individual cells but from the apparent—since observed—capability for developing thunderstorms that is advected to the area at risk. The observed rainfall and lightning features are manifestations of that capability, the extent of which may be much larger than the area of the observed features. Varying the length and direction of the advection vectors is a means of capturing the larger structure of that area. This type of predictor does not take into account the evolution of the system during advection. To some extent, this should be captured by the NWP predictors.

The data from the first and third 10-day series of the warm months (mid-April to mid-October) of 1 July 2002–30 June 2005 are used as the initial dependent set and the data from the second 10-day series serve as the initial independent set. Only the dependent set has been used in the predictor selection process. In this way the stability of the predictor selection could be tested on the independent set. After having selected the predictors for the MOS equations, the regression coefficients are updated using all data from the warm half-years of 1 July 2002–30 June 2005, that is, the eventual dependent (developmental) set. The period from 1 July 2005–31 July 2007 (warm half-years only) is finally used as the eventual independent (verification) set. The verification results in this paper are all based on the latter dataset.

b. *Selected predictors*

In Tables 1–3, all predictors are shown that are selected in at least one (severe) thunderstorm forecast equation. Table 2 also includes the definitions of some traditional thunderstorm indices. The (severe) thunderstorm forecast equations contain at least two and at most five predictors. The maximum number of predictors has been set to five, because more than five predictors often resulted in overfitting.

The most commonly selected predictor in the thunderstorm forecast system for the 0–6-h projections turned out to be the percentage of the total number of advection ensemble members with $\geq$4 discharges. It is included 72 times in a total of 96 forecast equations and occurs in all runs, except for the 2100 UTC run. As thunderstorms are often decaying around that time, it is not surprising that this predictor does not appear in the forecast equations of the 2100 UTC run. The prominence of this predictor illustrates that the ad hoc

TABLE 1. Overview of lightning and radar ensemble predictors that are included in at least one (severe) thunderstorm forecast equation. Here, radar advection ensemble indicates that part of the ensemble using only (perturbed) vectors from the radar images, and HIRLAM advection ensemble indicates that part of the ensemble using only (perturbed) HIRLAM 700-hPa wind vectors.

Lightning ensemble predictors
Percentage of total advection ensemble with ≥4 lightning discharges in 6 hours
Percentage of *radar* advection ensemble with ≥4 discharges in 6 hours
Binary predictor indicating whether at least one advection ensemble member shows ≥4 discharges in 6 hours
Binary predictor indicating whether a particular HIRLAM advection ensemble member shows ≥2 discharges in 5 minutes
Temporal maximum (average) 5-min lightning intensity from a particular radar advection ensemble member
Maximum 5-min lightning intensity from the total advection ensemble
Maximum 5-min lightning intensity from the HIRLAM advection ensemble
Radar ensemble predictors
Temporal maximum percentage of the region occupied by ≥10 and ≥30 mm h^{-1} radar pixels (Kitzmiller et al. 2002), respectively, from a particular radar advection ensemble member
Maximum percentage of the region occupied by ≥10 and ≥30 mm h^{-1} radar pixels, respectively, from total advection ensemble

choices that were made to construct the ensemble (section 3a) were quite good. However, it cannot be ruled out that other choices might have resulted in an even better predictor. As in SKV05, the square root of the ECMWF 6-h convective precipitation sum turns out to be an often-selected predictor as well. In fact, it is the most frequently selected predictor in the thunderstorm forecast system for the 6–12-h projections and is included 127 times in a total of 192 forecast equations. Therefore, it was decided to include the ECMWF convective precipitation predictor, despite the fact that the convection parameterization is changed quite often in NWP models.

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 97 forecast equations), the most unstable convective available potential energy (CAPE; included in 87 forecast equations), and the Boyden index (included in 55 forecast equations). The Jefferson index assesses the latent instability of an air parcel at 925 hPa. The CAPE (Table 2) represents the vertically integrated positive buoyancy of a parcel experiencing adiabatic ascent. In this case the parcel is lifted from the most unstable level in the lowest 400-hPa layer. This turned out to be a better predictor than CAPE based on a parcel lifted from the surface, or from the lowest 50- or 100-hPa layer. The Boyden index accounts for the pure conditional instability of a certain atmospheric layer and does not include moisture. However, as the Boyden index is combined with the convective precipitation sum in most equations and with

TABLE 2. Overview of the 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), θ_w is the wet-bulb potential temperature, θ_{ws} is the wet-bulb pseudopotential temperature, g is the acceleration due to gravity, LFC is the level of free convection, T_v is the virtual temperature, and env indicates environment.

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)
Showalter index	$\theta_{ws500} - \theta_{w850}$	Showalter (1953)
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)
Vertical totals index	$T_{850} - T_{500}$	Miller (1967)
Level of neutral buoyancy (LNB)		
Lowest 100-hPa CAPE and most unstable CAPE	$g\int_{\mathrm{LFC}}^{\mathrm{LNB}} \frac{T_v(\mathrm{parcel}) - T_v(\mathrm{env})}{T_v(\mathrm{env})}\,dz$	Moncrieff and Miller (1976); Craven et al. (2002)
Square root of 3- and 6-h accumulated convective precipitation		

TABLE 3. Overview of the (postprocessed) ECMWF and cosine predictors that are included in at least one (severe) thunderstorm forecast equation.

ECMWF predictors
Square root of 3- and 6-h accumulated convective precipitation
Relative humidity at 850 hPa
Power-transformed relative humidity at 850 hPa
Meridional wind component at 850 and 300 hPa
Anomalous temperature at 1000 hPa
Wind speed at 1000 hPa
Equivalent 500–850-hPa thickness
Temperature advection between 500 and 1000 hPa
P27 score 3 [cyclonality of the 500-hPa flow; Kruizinga (1979)]
Cosine predictor
Cosine of the day of the year

other moisture containing predictors in most of the remaining equations, the effect of moisture is nearly always included.

Haklander and van Delden (2003) also found that the CAPE is a good predictor for the probability of thunderstorms, while the Jefferson and Boyden indices are less good. They investigated the separate 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.

The above-mentioned list of the most commonly selected predictors is actually quite similar to the corresponding list in SKV05, apart from the lightning ensemble predictor and the most unstable CAPE, which were not included in the potential predictor set of that study. However, a comparable predictor to the latter was often selected in the SKV05 study, namely the level of neutral buoyancy—based on a parcel lifted from the surface—which is used in the computation of CAPE (Table 2).

The predictor that has been selected most often in the *severe* thunderstorm forecast equations is the Bradbury index, computed from the HIRLAM forecasts. The Bradbury index (Table 2) assesses the potential instability between 850 and 500 hPa. Other important predictors are again the Jefferson index, computed from the HIRLAM forecasts; a number of predictors from the ensemble of advected lightning data (Table 1; only 0–6-h projections of all runs, again except for the 2100 UTC run); and the square root of the ECMWF 3-h convective precipitation sum. For the purpose of consistency, the same predictors are used in the logistic regression equations for the three different thresholds (50, 100, and 200 discharges per 5 minutes) but, of course, with different regression coefficients. As 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 (SKV05). A complete overview of all selected predictors is given in Tables 1–3.

4. Example of a (severe) thunderstorm forecast

In this section we present one case to demonstrate the MOS system for thunderstorms and severe thunderstorms. The case that we show is from 8 June 2007. The 6–12-h thunderstorm probability forecasts, computed by the 0900 UTC run of the MOS system, are shown in Fig. 3a. The 6–12-h conditional probability forecasts of severe thunderstorms (threshold of 200 discharges per 5 minutes), computed by the same run, are shown in Fig. 3b. The highest thunderstorm probabilities have been forecast in the central and western parts of the Netherlands for the period 1500–2100 UTC. The highest conditional probabilities of *severe* thunderstorms have been forecast in three southwestern regions for the same period. To assess whether the forecast probabilities are higher than normal, they can be compared with the climatological probabilities (e.g., Fig. 2). In this case most of the probabilities of thunderstorms and conditional probabilities of severe thunderstorms are much higher than normal. The *absolute* probabilities of severe thunderstorms can be calculated by multiplying the conditional probabilities of severe thunderstorms (Fig. 3b) by the thunderstorm probabilities (Fig. 3a).

The 0–6-h thunderstorm probability forecasts, computed by the 1500 UTC run of the MOS system, are shown in Fig. 3c. The 0–6-h conditional probability forecasts of severe thunderstorms (threshold of 200 discharges per 5 minutes), computed by the same run, are shown in Fig. 3d. Figure 3e shows the 5-min lightning intensity at 1440 UTC. This is the initial field, which was advected subsequently. In Figs. 3c and 3d the effects of the lightning advection predictors can be seen. The probabilities have increased in the regions where the thunderstorms are likely to be advected to.

The regional maximum 5-min lightning intensity, as detected by the SAFIR network, from 1500–2100 UTC is shown in Fig. 3f. In all but the most northeastern region, lightning discharges have been observed in the period 1500–2100 UTC and in five regions even the threshold of 200 discharges per 5 minutes has been exceeded. Moreover, a weather alarm for severe thunderstorms has been issued, but not before the criterion was reached for the first time. Of course, (probability) fore-

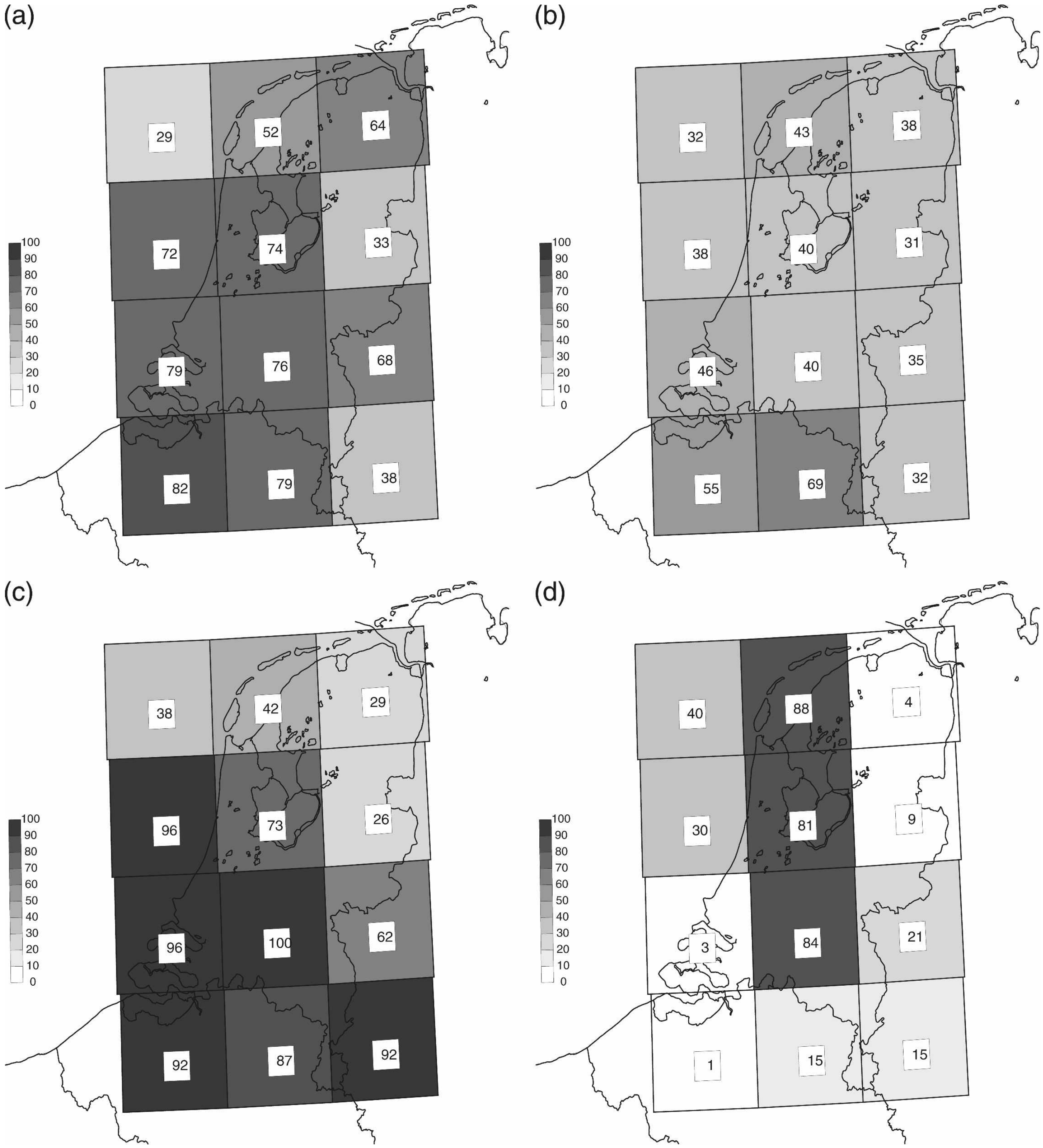


FIG. 3. The (a), (b) 6–12- and (c), (d) 0–6-h forecasts of the MOS (severe) thunderstorm forecast system, valid for 1500–2100 UTC 8 Jun 2007. Shown in (a) and (c) are thunderstorm probability forecasts (%), whereas (b) and (d) show conditional probability forecasts (%) of ≥200 discharges per 5 minutes. (e) Lightning intensity [(5 min)$^{-1}$ (50 km × 50 km)$^{-1}$] at 1440 UTC with the following color shading: lightning intensity <50 discharges per 5 minutes (gray), 50–100 discharges per 5 minutes (green), 100–200 discharges per 5 minutes (blue), 200–400 discharges per 5 minutes (red), and 400–500 discharges per 5 minutes (yellow). (f) Regional maximum 5-min lightning intensity, as detected by the SAFIR network, from 1500–2100 UTC 8 Jun 2007.

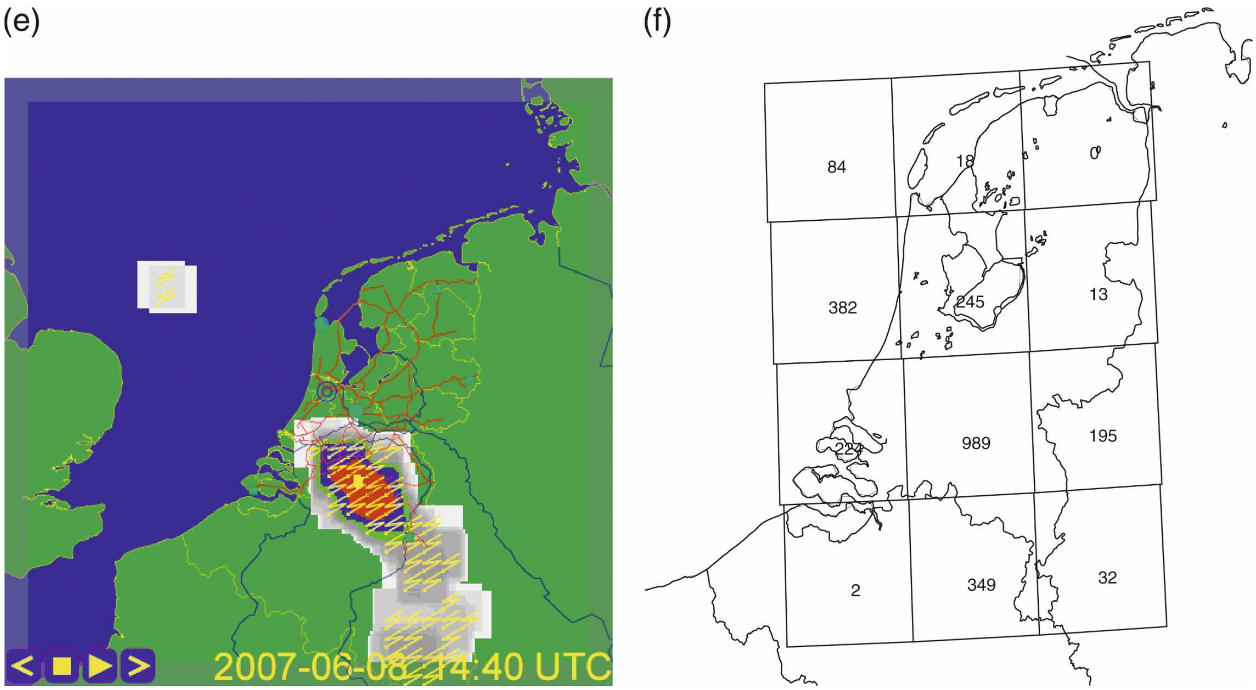


FIG. 3. *(Continued)*

casts cannot be verified using only one case, so we now present objective verification results.

5. Verification results

a. *Verification results of the MOS thunderstorm forecast system*

In this section some verification results of the MOS thunderstorm forecast system (≥2 discharges) are presented for the independent warm months of 1 July 2005–31 July 2007. As scalar verification scores, we have used the bias, the Brier score (BS), two of its decomposition terms (reliability and resolution), and the Brier skill score (BSS). The definitions of these scores are given in Wilks (2006). We also show attributes diagrams (Fig. 4; 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 2006).

In Figs. 4a–d we show attributes diagrams of the 0–6-h forecasts for four runs of the MOS thunderstorm forecast system for the region M-MS (Fig. 1a). Figure 4a shows the attributes diagram of the 0000 UTC run. As the climatological probability is quite low in the period 0000–0600 UTC, the higher forecast probabilities are quite rare (see the histogram on the right in Fig. 4a). Nevertheless, the BSS is clearly positive (BSS = 21.5%). The attributes diagram of the 0600 UTC run (Fig. 4b) shows high skill (BSS = 47.9%) because of the excellent reliability and the good resolution, and it shows hardly any bias. The forecasts of the 1200 UTC run (Fig. 4c) show even better resolution but slightly worse reliability, with some overforecasting. The forecasts of the 1800 UTC run (Fig. 4d) show high skill (BSS = 45.2%) as well and an overforecasting bias of the higher probabilities.

To get an idea of the verification scores for the other regions, Fig. 5a shows the Brier skill scores for the 0–6-h forecasts of the eight runs of the MOS thunderstorm forecast system for all regions, and Fig. 5b shows the Brier skill scores for the 6–12-h forecasts. 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 lines in Fig. 5) and the six coastal regions (dashed lines) are positive. The average BSS of the six land regions shows a clear diurnal cycle with the highest skill in the afternoon and evening (1200–1800 UTC), whereas the average BSS of the six coastal regions shows no diurnal cycle. Forecasting thunderstorms is most difficult during the nighttime (0000 UTC) and especially during the morning (0300 UTC), when the relationship between the predictors and the occurrence of lightning is weakest, particularly in the land regions. The Brier skill scores for the 6–12-h forecasts (Fig. 5b) are generally smaller than those for

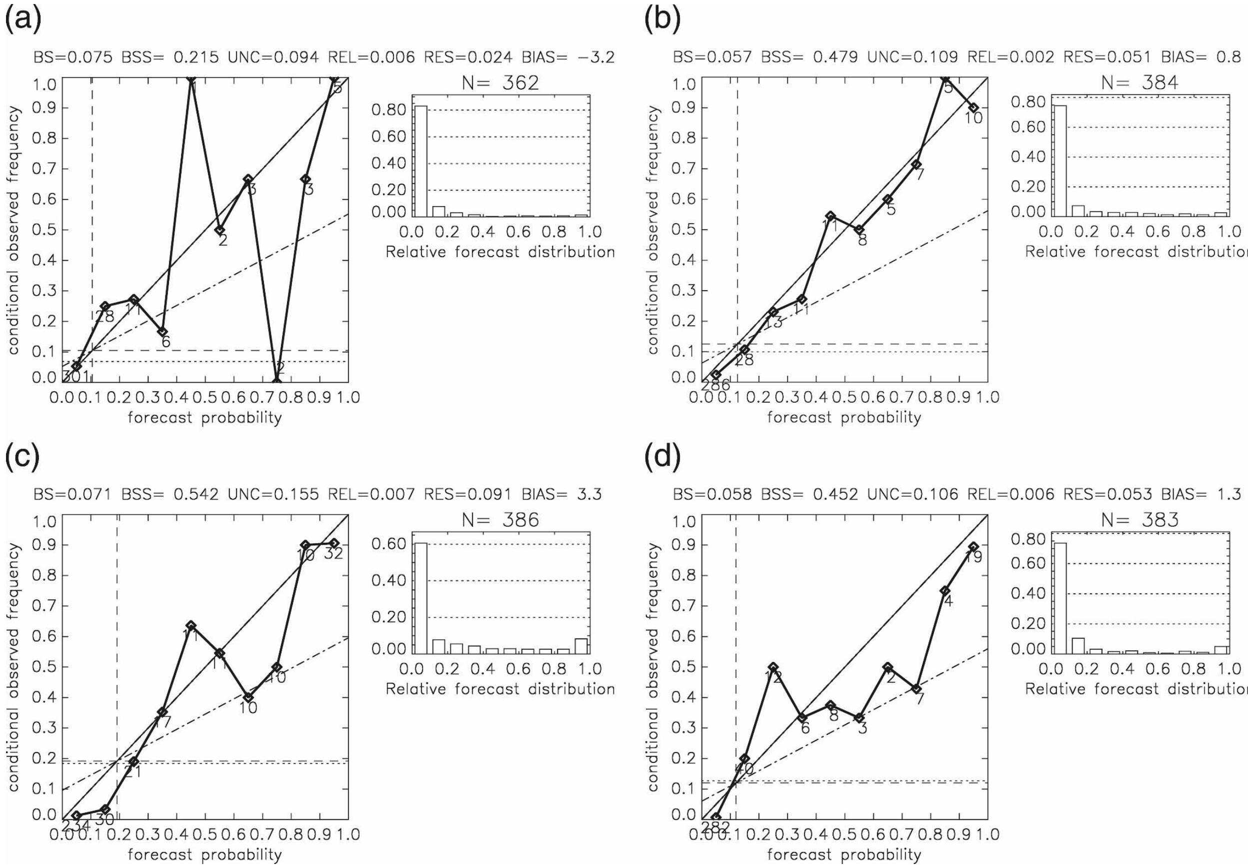


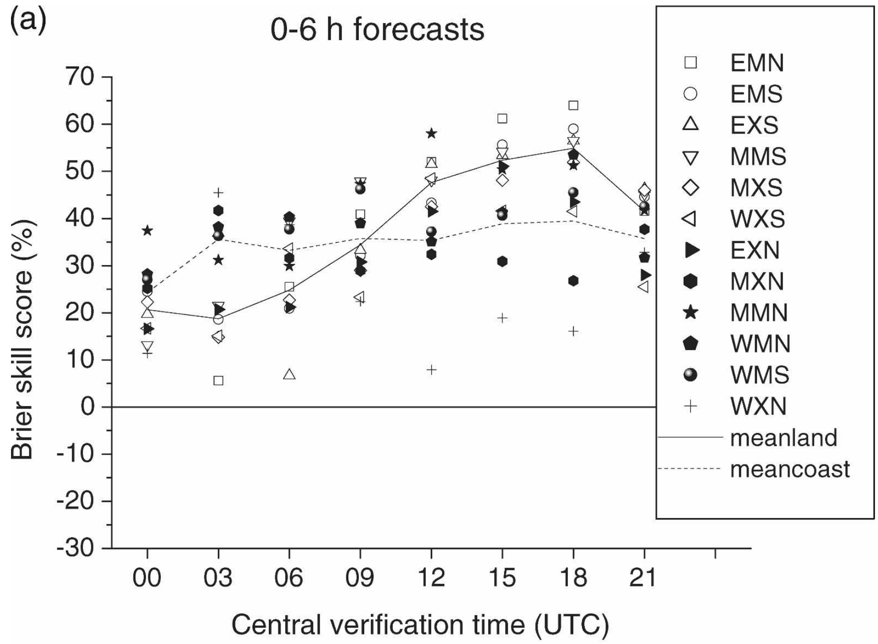
FIG. 4. Attributes diagrams of 0–6-h forecasts, as computed by the (a) 0000, (b) 0600, (c) 1200, and (d) 1800 UTC runs of the MOS thunderstorm forecast system for the region M-MS. The verification period is from 1 Jul 2005 to 31 Jul 2007 (warm half-years only). In these diagrams the observed frequencies 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 2000–04 climatology, and the dashed line the sample climatology. The dashed–dotted line indicates the "no skill" line (in terms of the BSS). The histogram on the right portrays the relative frequency of the use of the forecasts. Here, UNC is short for uncertainty, REL for reliability, and RES for resolution (see e.g., Wilks 2006), and *N* is the total number of cases.

the 0–6-h forecasts (Fig. 5a), as a result of a lower-resolution term and a slight decrease in reliability (not shown). Apart from the fact that the skill of a forecast system decreases with increasing forecast projections, the absence of the most important predictor for the 0–6-h projections (i.e., the percentage of the total number of advection ensemble members with $\geq$4 discharges) is expected to play a role as well. We conclude from these verification results that the overall skill of the MOS thunderstorm forecast system is good compared to the 2000–04 climatology. Finally, the Brier skill scores of this new thunderstorm forecast system are generally higher than those of the operational system of SKV05 (not shown).

The bias (not shown) also shows some variation between the regions, but for most regions it is slightly positive, indicating overforecasting. A plausible reason for this overforecasting is the increase in ECMWF and/or HIRLAM resolution, which has led to more extreme values of the predictors in the verification sample and therefore to higher forecast probabilities than in the development sample.

b. 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 subsection. In Figs. 6a–h we show attributes diagrams of the 0–6-h forecasts for the 0000, 0600, 1200, 1500, and 1800 UTC runs of the MOS severe thunderstorm forecast system for the thresholds of 50 and 100 discharges per 5 minutes. All 12 regions have been pooled. The severe thunderstorm



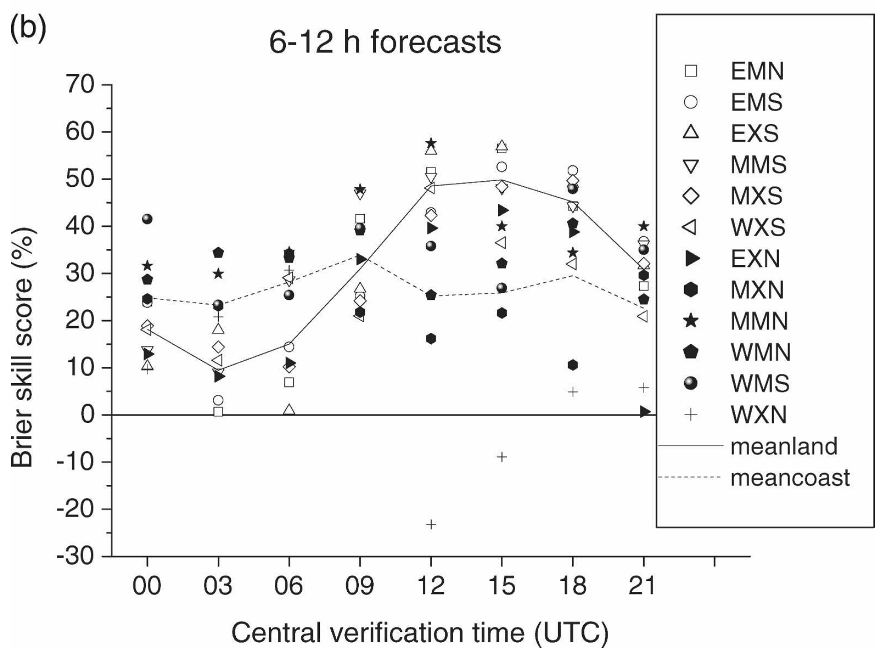


FIG. 5. BSS (%) with respect to the 2000–04 climatology as a function of central verification time for the eight runs of the MOS thunderstorm forecast system (≥2 discharges) for all 12 regions. The solid line represents the average BSS for the six land regions (E-MN, E-MS, E-XS, M-MS, M-XS, and W-XS; open symbols), and the dashed line represents the average BSS for the six coastal regions (E-XN, M-XN, M-MN, W-MN, W-MS, and W-XN; filled symbols). The verification period is from 1 Jul 2005 to 31 Jul 2007 (warm half-years only) for the (a) 0–6- and (b) 6–12-h forecasts.

forecast system shows the highest skill for the evening period (1500–2100 UTC; Figs. 6e and 6f) with the highest skill for the lowest threshold, as expected. Surprisingly, the system shows the lowest skill for the afternoon period (0900–1500 UTC; not shown), presumably because of the high bias in that forecast due to the increase in ECMWF resolution, which appears to be more apparent in this forecast than in most of the other ones. Figure 6c shows the attributes diagram of the 1200 UTC run for the threshold of 50 discharges per 5 minutes. Although the Brier skill score of these forecasts is 0, the diagram shows a positive-sloping reliability curve, which is a clear indication that the forecasts have more forecast value than do the climatological forecasts (section 6; Mason 2004). The 6–12-h forecasts have negative, zero, or only slightly positive Brier skill scores (not shown), when verified over the total verification period, but the forecasts for the 1200–1800, 1500–2100, and 1800–0000 UTC periods show again positive-sloping reliability curves.

When the individual years are considered, the attributes diagrams for the periods July–mid-October 2005 and mid-April–July 2007 show better overall skill than the diagrams for the warm half-year of 2006 (not shown). Factors that may have contributed to the lesser skill in the warm half-year of 2006 are sampling effects and/or the low frequencies of severe thunderstorms during that half-year. Nevertheless, the evaluation of the system by forecasters during the period 24 May–15 October 2006 was positive, as was evident from the results of a 6-h Web-based questionnaire during this period (Schmeits et al. 2007). Apparently, a forecast system can be much more useful in practice than the (harsh) objective verification results might suggest.

A similar conclusion can be drawn from Kain et al. (2006), who reported results from the Storm Prediction Center/National Severe Storms Laboratory (SPC/NSSL) 2004 Spring Program. During that program two different probabilistic forecasts of severe weather were prepared. The first was a control human forecast, with data access restricted to operational data streams. The second was an experimental human forecast, with access to high-resolution output. While the differences in objective skill scores between these experimental and control human forecasts were very small, the difference in mean subjective rating was relatively large.

We conclude from the verification results shown here (Fig. 6) that the severe thunderstorm forecast system is generally skillful, compared to the 2000–04 climatology, at least for the thresholds of 50 and 100 discharges per 5 minutes. Verification results for the higher threshold of 200 discharges per 5 minutes (not shown) are less conclusive, mainly because of the low observed frequency. Finally, we would like to stress that the Brier skill scores for the initial independent dataset (section 3a) were clearly positive for all thresholds (not shown).

6. Summary and discussion

We have described the derivation and verification of new MOS equations for both the probability of thunderstorms and the conditional probability of severe thunderstorms during the warm half-year (i.e., mid-April to mid-October) in the Netherlands. We have developed these equations for 12 regions of about 90 × 80 km^2 each (Fig. 1) and for projections out to 12 h in advance (with 6-h periods). As the potential predictor dataset we have not only used combined postprocessed output from the HIRLAM and ECMWF models as in SKV05, but also an ensemble of advected radar and lightning data for the 0–6-h projections. The predictands are derived from SAFIR (reprocessed) total lightning data, being either the probability of a thunderstorm (≥2 discharges) or the conditional probability of a *severe* thunderstorm (maximum lightning intensity ≥50, and for some 6-h periods also ≥100 and ≥200 discharges per 5 minutes) under the condition that ≥2

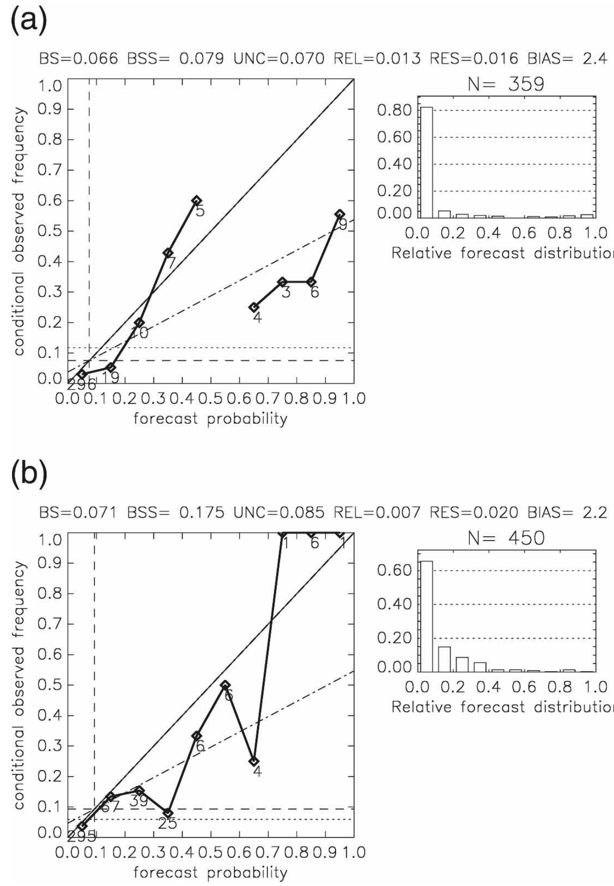


FIG. 6. Attributes diagrams of 0–6-h forecasts, as computed by five different runs of the MOS severe thunderstorm forecast system: (a) 0000, (b) 0600, (c), (d) 1200, (e), (f) 1500, and (g), (h) 1800 UTC runs. Here, (a)–(c), (e), and (g) show attributes diagrams for the threshold of 50 discharges per 5 minutes, and (d), (f), and (h) show those for the threshold of 100 discharges per 5 minutes. The verification period is from 1 Jul 2005 to 31 Jul 2007 (warm half-years only). See the caption of Fig. 4 for more information.

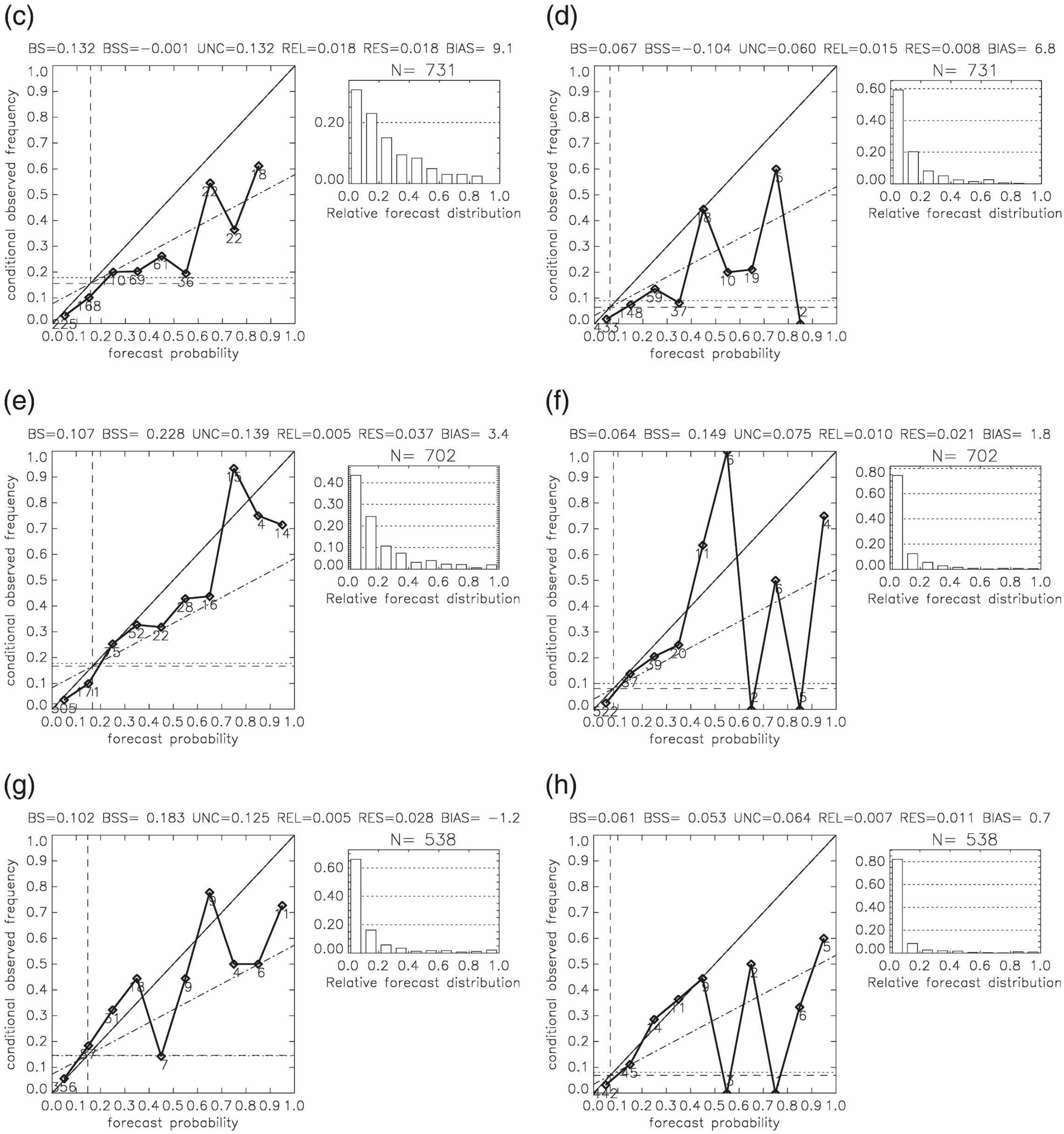


FIG. 6. *(Continued)*

discharges will be detected. The equations have been derived using the method of logistic regression, which is the preferred regression method in probabilistic forecasting (Applequist et al. 2002). The MOS system was made preoperational at KNMI in the spring of 2006. It is expected that this new MOS system will help the forecasters to decide whether a weather alarm for severe thunderstorms should be issued.

When using the MOS approach, the combination of output data from two different NWP models has both advantages and disadvantages (SKV05). Among the advantages 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 a disadvantage is that the frequency of the 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. However, the equations should be updated shortly, as is apparent from the high biases in some of the forecasts, because of the higher resolution of both the HIRLAM and ECMWF models.

The ensemble approach used in the advection of the lightning and radar data works well, as is evident from the prominence of the associated predictors (Table 1). In all 0–6-h forecast equations, apart from those of the 2100 UTC run, lightning and/or radar ensemble predictors have been selected. In these equations the lightning ensemble predictors have been selected more often than the radar ensemble predictors.

For the 0–6-h projections the most frequently selected predictor in the thunderstorm forecast system turned out to be the percentage of the total number of advection ensemble members with $\geq$4 discharges. Overall, the square root of the ECMWF 6-h convective precipitation sum is the most often selected predictor in the thunderstorm forecast system, as in SKV05. The most often selected predictor in the *severe* thunderstorm forecast system is the Bradbury index, computed from the HIRLAM forecasts (see Tables 1–3 for an overview of all selected predictors).

Contrary to SKV05, the square roots of the HIRLAM 3- and 6-h convective precipitation sums have now been added as potential predictors. These predictors were selected a few times, so in those cases they had some additive predictive potential with respect to the other predictors, but the square root of the ECMWF 6-h convective precipitation sum is a much better predictor for probabilistic thunderstorm forecasts.

The Bradbury index appears to be a good discriminator between nonsevere and severe thunderstorm cases. Charba (1979) and Kitzmiller et al. (2002) found that a similar index, namely the total totals index (Miller 1967), is an important predictor for (conditional) probabilistic forecasts of severe weather, whereas that index has not been selected in our study. On the other hand, the vertical totals index (Table 2) has been selected, but only once. 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.

Apart from the "sawtooth" pattern seen in several of our attributes diagrams (Figs. 4 and 6) due to the low frequency of the higher forecast probabilities, 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 observed frequency increases as the forecast probability increases. Moreover, our Brier skill scores are often positive with respect to climatology, 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 that the MOS thunderstorm forecast system has a good skill (Figs. 4 and 5), and that the *severe* thunderstorm forecast system generally is also skillful (Fig. 6), compared to the 2000–04 climatology. However, the system shows (slight) overforecasting, plausibly as a result of the increase in ECMWF and/or HIRLAM resolution in 2006. This can be eliminated by updating the regression coefficients when data from the high-resolution models are included.

Recently, the preoperational system has been extended to include the 12–48-h projections. This system is intended to replace the operational (severe) thunderstorm forecast system of SKV05, if its skill is satisfactory. Verification results for the warm-half year of 2007 indicate the system to be skillful out to +48 h (not shown). Finally, future developments in this MOS system may consist of even more extreme criteria for severe thunderstorms and the inclusion of advected Meteosat Second Generation (MSG) data as a potential predictor source for the 0–6-h projections.

Acknowledgments. The authors thank Iwan Holleman (KNMI) for supplying part of the software to process the SAFIR data, Ernst de Vreede (KNMI) for providing Fig. 3e, and the anonymous reviewers for their comments on an earlier version of the manuscript.

REFERENCES

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.

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.

Charba, J. P., 1979: Two to six hour severe local storm probabilities: An operational forecast system. *Mon. Wea. Rev.,* **107,** 1263–1274.

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.

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.

Golding, B. W., 2000: Quantitative precipitation forecasting in the UK. *J. Hydrol.,* **239,** 286–305.

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

Holleman, I., 2007: Bias adjustment and long-term verification of radar-based precipitation estimates. *Meteor. Appl.,* **14,** 195–203.

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.

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.

——, S. J. Weiss, J. J. Levit, M. E. Baldwin, and D. R. Bright, 2006: Examination of convection-allowing configurations of the WRF model for the prediction of severe convective weather: The SPC/NSSL Spring Program 2004. *Wea. Forecasting,* **21,** 167–181.

Kitzmiller, D. H., F. G. Samplatsky, C. Mello, and J. Dai, 2002: Probabilistic forecasts of severe local storms in the 0–3 hour timeframe from an advective-statistical technique. Preprints, *21st Conf. on Severe Local Storms/19th Conf. on Weather Analysis and Forecasting/15th Conf. on Numerical Weather Prediction/,* San Antonio, TX, Amer. Meteor. Soc., JP1.10. [Available online at http://ams.confex.com/ams/pdfpapers/46765.pdf.]

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.

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 Tech. Rep. 200, U.S. Air Force, Scott AFB, IL, 94 pp. [Available from Headquarters, AWS, Scott AFB, IL 62225.]

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.

Noteboom, S., 2006: Processing, validatie, en analyse van bliksemdata uit het SAFIR/FLITS systeem (in Dutch). KNMI Internal Rep. IR-2006-01, 71 pp. [Available from KNMI, P.O. Box 201, 3730 AE De Bilt, Netherlands.]

Schmeits, M. J., C. J. Kok, and D. H. P. Vogelezang, 2005: Probabilistic forecasting of (severe) thunderstorms in the Netherlands using model output statistics. *Wea. Forecasting,* **20,** 134–148.

——, ——, ——, and R. M. van Westrhenen, 2007: Kansverwachtingen voor onweer ten behoeve van uitgifte weeralarm (KOUW). KNMI Internal Rep. IR-2007-03, 22 pp. [Available online at http://www.knmi.nl/publications/fulltexts/kouw_ir.pdf.]

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, Norrköping, Sweden, 144 pp.

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

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* 2nd ed. Academic Press, 627 pp.

Williams, E., and Coauthors, 1999: The behavior of total lightning activity in severe Florida thunderstorms. *Atmos. Res.,* **51,** 245–265.