Skill of a Ceiling and Visibility Local Ensemble Prediction System (LEPS) according to Fog-Type Prediction at Paris-Charles de Gaulle Airport

STEVIE ROQUELAURE,* ROBERT TARDIF,+ SAMUEL REMY, AND THIERRY BERGOT

GAME/CNRM, Météo-France, CNRS, Toulouse, France

(Manuscript received 30 September 2008, in final form 18 June 2009)

ABSTRACT

A specific event, called a low-visibility procedure (LVP), has been defined when visibility is under 600 m and/or the ceiling is under 60 m at Paris-Charles de Gaulle Airport, Paris, France, to ensure air traffic safety and to reduce the economic issues related to poor visibility conditions. The Local Ensemble Prediction System (LEPS) has been designed to estimate LVP likelihood in order to help forecasters in their tasks. This work evaluates the skill of LEPS for each type of LVP that takes place at the airport area during five winter seasons from 2002 to 2007. An event-based classification reveals that stratus base lowering, advection, and radiation fogs make up for 78% of the LVP cases that occurred near the airport during this period. This study also demonstrates that LEPS is skillful on these types of event for short-term forecasts. When the ensemble runs start with initialized LVP events, the prediction of advection fogs is as skillful as the prediction of radiation fog events and stratus base lowering. At 3 and 6 h before the runs where LVP events were initialized, LEPS still shows positive skill for radiation fog events and stratus base lowering cases.

1. Introduction

During the winter, airport authorities at Paris-Charles de Gaulle Airport, in Paris, France, are concerned with major management and safety issues caused by the occurrence of fog and low-ceiling events. Under these low-visibility conditions, low-visibility procedures (LVPs) are taken to safely manage the airport activity when visibility is less than 600 m and/or a ceiling is less than 60 m. Unfortunately, these procedures reduce by a factor of 2 the airport's efficiency for takeoffs and landings, causing flight delays or cancellations.

Most of the current operational systems for forecasting the fog and ceiling at an airport rely on statistical methods like model output statistics (MOS; Glahn and Lowry 1972; Koziara et al. 1983) or artificial neural networks (ANNs; Fabbian et al. 2007; Bremnes and Michaelides 2007; Marzban et al. 2007). These methods use statistical postprocessing of numerical model outputs to improve forecast quality, especially when model outputs are combined with surface observations. Their main advantage, compared to LEPS, is that they can provide spatial low-visibility forecasts. However, the lack of information on the physical processes beneath the fog formation and development is one of their undesirable characteristics because the link between the predictors and the predictand is purely statistical. Therefore, these statistical methods act like "black boxes" and the fitted statistical coefficients are physically difficult to interpret. Recently, the National Centers for Environmental Prediction (NCEP) have developed a fog ensemble prediction product based on the postprocessing of the Short-Range Ensemble Forecast (SREF) system. This system predicts the fog occurrence probability by diagnosing the fog conditions from the SREF ensemble members over the United States (Zhou et al. 2007).

As a part of the forecasting effort undertaken to facilitate the decision-making process at Paris-Charles de Gaulle Airport (Richardson 2000), a local ensemble prediction approach has been adopted. The Local Ensemble Prediction System (LEPS; Roquelaure and Bergot 2007, 2008, 2009) was designed to satisfy the

* Current affiliation: Atmospheric Science Laboratory, Graduate School of Science, Tohoku University, Sendai, Miyagi, Japan.

+ Current affiliation: Meteorological Research Division, Environment Canada, Dorval, Quebec, Canada.

Corresponding author address: Stevie Roquelaure, Atmospheric Science Laboratory, Graduate School of Science, Tohoku University, 6-3 Aoba, Aramaki, Aoba, Sendai, Miyagi 980-8578, Japan. E-mail: stevie@wind.geophys.tohoku.ac.jp

DOI: 10.1175/2009WAF2222213.1

specific requirements of predicting low-visibility and low-ceiling events. The local ensemble is generated from perturbations on initial conditions (atmosphere, soil, and low-cloud initialization), and mesoscale forcings are required to run the 1D numerical model. Perturbations are computed from past data; from on-site observations for initial conditions and from the 3D NWP version of the Aire Limitée Adaptation Dynamique Développement International (Aladin) model for mesoscale forcings. These perturbation computations have been described previously (Roquelaure and Bergot 2007). This local ensemble strategy provides the likelihood of LVP but it also has the potential to give insights into the physical processes during low-visibility events that are not accessible with MOS or ANN schemes. Moreover, LEPS is calibrated by the Bayesian averaging method (BMA), which highlights the main sources of uncertainty in the system by computing specific weights for each ensemble member. By these means, LEPS provides the likelihood of LVP over the airport for the next 12 h. These forecasts are necessary to maintain a high level of efficiency on the airport as well as reinforcing airport activity safety in poor visibility conditions.

In Roquelaure and Bergot (2008), the BMA calibration has been validated for the prediction of LVP variable over the Paris airport. The system provided reliable very short-term forecasts (12 h) of the LVP likelihood, along with interesting relative operating characteristics of LVP forecasts during the winter of 2004/05. However, some questions still remain as to how the ensemble is able to discriminate the type of LVP events. In fact, the LVP category includes various types of low-visibility events, which are controlled by different physical processes including radiation fog events, which are piloted by radiative cooling over the land; advection fog driven by advection, which induces the mixing of air parcels of contrasting temperatures; and lowering stratus events, which are induced by the moistening and/or the cooling of the layer below the boundary layer stratiform cloud, etc. This work tries to highlight the strengths and weaknesses of LEPS for each LVP type. It is reasonable to question the usefulness of a local ensemble prediction system for forecasting advection fog events since this type of fog event is driven by three-dimensional advections. In this study, the ensemble skill will be also assessed by comparing the ensemble mean with the reference deterministic forecast to analyze the mean ensemble LVP forecast's possible usefulness in the forecasting process.

This present paper focuses on the discrimination potential of LEPS related to each type of low-visibility event, by analyzing the system's performance in several areas (e.g., hit rates and false alarm rates, relative errors on formation–dissipation time). LVP events observed at the Paris-Charles de Gaulle Airport during the winter seasons of 2002–07 are used for this purpose. To achieve the identification of LVP types, a classification based on Tardif and Rasmussen's (2007) algorithm is performed over five winter seasons from 2002 to 2007 at Charles de Gaulle Airport to understand its low-visibility event climatology. Then, from this classification, LEPS performance levels are assessed to analyze the local ensemble forecasting behavior patterns for each type of event.

As a consequence, section 2 explores the climatology of LVP events over Charles de Gaulle Airport. Section 3 briefly presents LEPS and the BMA calibration applied to calibrate the ensemble. The BMA calibration method is applied over the full training dataset (2002–07) and compared to the previous computation of BMA weights in Roquelaure and Bergot (2008), which used a smaller training data sample composed by two winter seasons (2002–04). Section 4 describes the methodology adopted to analyze the LEPS discrimination potential and section 5 gives the results. Finally, section 6 summarizes the results and presents the main conclusions.

2. The event-based climatology of LVP over Paris-Charles de Gaulle Airport

a. The classification methodology: Fog types

The comprehensive fog-type classification, described by Tardif and Rasmussen (2007), is applied to identify the types of LVP events that commonly take place over Charles de Gaulle Airport. A specific LVP climatology for this airport location is then derived from the classification methodology. Five types of LVP events are discriminated through the algorithm: precipitation fog (PCP), radiation fog (RAD), advection fog (ADV), evaporation fog (EVP), and lowering of the cloud base (STL, low ceiling). Fog events that do not meet the criteria used to classify events into the five previous types, or those for which a condition has insufficient data for validation, characterizing the few hours prior to fog onset, are classified as unknown (UKN). The classification is based on the hypothesis that the formation of LVP events is driven by different physical mechanisms. As a consequence, the algorithm's goal is to identify the primary mechanisms responsible for the formation of each type of LVP event, via the use of simple conceptual models of fog formation applied to local surface observations at the airport. PCP fogs are mostly driven by the thermodynamical influence of evaporating precipitation (Petterssen 1969). RAD fogs develop under strong surface radiative cooling conditions (Meyer and Lala 1990; Baker et al. 2002). ADV fogs result from the

TABLE 1. Number of LVP events per winter and per type.

Winter season	No. of LVP events per winter per type						
	STL	RAD	ADV	UKN	PCP	EVP	Total LVP
2002/03	12	4	1	0	1	0	18
2003/04	20	18	2	4	7	1	52
2004/05	14	6	3	3	1	0	27
2005/06	10	8	2	6	1	0	27
2006/07	11	7	3	7	3	0	31
Total (2002–07)	67	43	11	20	13	1	155

mixing of air parcels of contrasting temperatures as moist and warm air flows over a colder surface (Baars et al. 2003). STL situations mostly occur under conditions of moistening and/or cooling of the layer below stratiform clouds and/or prolonged subsidence (Baker et al. 2002). EVP fogs are induced by the evaporation of surface water and mixing in the surface layer (Arya 2001). For more details on the classification and the simple conceptual models used, see Tardif and Rasmussen (2007). Compared to the hourly surface observations used in Tardif and Rasmussen's study, here 6-min local surface observations of visibility, temperature, wind, precipitation, cloud cover, and ceiling height are used for the classification. This high frequency of observations is clearly a great advantage in classifying the fog types, especially for short events, which can be totally missed with hourly data.

b. *The classification results over Charles de Gaulle Airport*

Table 1 shows the results of the LVP-type classification for Paris-Charles de Gaulle Airport. All winter seasons are composed of 5 months of data. However, during the first year of a data record, the observation measurements were started in December and as a consequence winter 2002/03 is composed of 5 months from December to April. Whereas, the other winters (2003–07) used data from October to February. Over the five winter seasons of the dataset, stratus base lowering events represent 43% of the LVP cases, radiation fog events count for 28%, and advection fog events for 7%. Precipitation and evaporation fog events are observed for, respectively, 8.4% and 0.6% of the LVP cases. Thirteen percent of the LVP cases cannot be discriminated by the algorithm and are classified as unknown.

Table 1 reveals the annual variability in observed LVP cases by looking at the results year by year. While the winter seasons from 2004 to 2007 have a total of about 30 LVP cases during each winter, with roughly the same frequency of LVP-type occurrences, 52 LVP cases are observed during winter season of 2003/04, which also has a twofold increase in the frequency of observations for STL, RAD, and PCP. The main differences between the winter season of 2002/03 and the other winters are due to the fact that the 5 months used from winter 2002/03 are from December to April, compared to October until February in the 2003–07 data. By starting the observation in December, the October and September fog cases are missed in the 2002/03 winter.

3. The Local Ensemble Prediction System

a. *LEPS methodology*

LEPS is built around the Code de Brouillard à l'Echelle Locale (COBEL, local-scale fog code) and Interaction Soil–Biosphere–Atmosphere (ISBA) local prediction scheme. The COBEL–ISBA initial conditions and mesoscale forcings are perturbed to obtain a 54-member ensemble (Roquelaure and Bergot 2007, 2008). The local numerical prediction method is currently used at Charles de Gaulle Airport to produce deterministic forecasts of LVP conditions over the airport. The 1D high-resolution COBEL atmospheric model (Bergot 1993; Bergot and Guédalia 1994) is coupled with the multilayer surface–vegetation–atmosphere transfer scheme of ISBA (Boone et al. 2000). The COBEL–ISBA initial conditions are estimated using a 1D variational data assimilation system (1DVAR; Bergot et al. 2005). The system uses specific observations from a 30-m-high meteorological tower (atmospheric temperature and humidity, and short- and longwave radiation fluxes) and soil measurements. The mesoscale influences are treated by external forcings. The mesoscale forcings (mesoscale advection, geostrophic wind, and cloud cover) are evaluated from the Météo-France operational NWP model Aladin (information online at http://www.cnrm.meteo.fr/aladin/).

The COBEL–ISBA inputs are the atmospheric temperature and humidity profiles from the 1DVAR system, the geostrophic wind profiles, the advection profiles, the cloud cover, the soil temperature, and the soil water content profiles. The model computes the following outputs within the boundary layer: the atmospheric temperature and humidity profiles, the wind profiles, the turbulent kinetic energy profiles, and the atmospheric cloud liquid water content from which the visibility is diagnosed.

b. *LEPS perturbations: Computation of the uncertainties*

The computation of the mesoscale forcing uncertainties is based on a spatiotemporal strategy, under the hypothesis that uncertainty is correlated with the ''intrinsic'' variability of the 3D NWP model Aladin. The

model variability is assessed in both space and time. The spatial variability is evaluated by comparing the forecast over an area of 3×3 grid points. This area is representative of the homogeneous surface conditions around the study area. The temporal variability is evaluated by comparing four Aladin runs (0000, 0600, 1200, and 1800 UTC) for the same verification time. At the end of this two-step procedure, the variability in both space and time is used to estimate the distribution of uncertainties on mesoscale forcings. Following this method of computing the uncertainties on mesoscale forcings, the horizontal temperature and humidity advections and the geostrophic wind are perturbed to produce the ensemble. The cloud cover is also perturbed by including members following three different scenarios: no cloud cover at all, cloud cover from the NWP model Aladin, and the persistence of the observed cloud cover. Initial condition uncertainties are estimated from observation errors in the soil and the lower part of the atmosphere where site observations are available (below 30 m). At higher elevations, the output from NWP model Aladin are used to provide both temperature and humidity profiles. As a consequence, uncertainties can be assessed with the spatiotemporal methodology described above for mesoscale forcings. The perturbed initial conditions in the ensemble are for the temperature and humidity profiles in both the atmosphere and the soil as well as the cloud initialization. For the cloud initialization, uncertainties are included on the cloud liquid water content and the cloud base or cloud top. For more details, see Roquelaure and Bergot (2008).

At the airport location, the impacts of these uncertainties on COBEL–ISBA forecasts have been evaluated during the 2002/03 winter season (Roquelaure and Bergot 2007). The study has shown the time dependency of the forecast dispersion. The impacts of uncertainties on the initial conditions decrease during the first hours of the simulation (0–6 h), whereas the dispersion created by mesoscale forcings becomes more important in the second half of the simulation (6–12 h). The cloud radiative impacts on the dispersion are felt throughout the 12-h forecast period. The uncertainties on the low-cloud initializations also act upon the 12-h forecast period.

c. LEPS calibration

The calibration technique for LEPS follows the BMA method described by Raftery et al. (2005). The BMA approach is applied on a training dataset (winter seasons from 2002/03 to 2006/07) to determine which members are the most efficient for the prediction of any quantity X, which is the LVP in our case. Therefore, the BMA calibration is applied on the LVP variable without any consideration of the LVP type. Thanks to the learning over the training dataset, the BMA method assigns a weight to each member to improve the ensemble's reliability. As a consequence, each member is clearly identified and has its own characteristics. If K members are available in the training dataset X^T, BMA takes into account all members to learn about each member's efficiency in forecasting the variable X (see Roquelaure and Bergot 2008 for more details). The strengths of the BMA method are the robustness of the weight computation algorithm and its simplicity in cases of binary forecasts, as is the circumstance here. Actually, since we have to predict a binary variable, the LVP distribution is discrete (two values: 1 or 0) and there is no variance of the distribution to compute.

The weaknesses of the method are the possible overfitting and the colinearity between ensemble members over the training data (Wilson et al. 2007; Hamill 2007). Overfitting occurs when the training data sample is too small and the colinearity occurs when there is too much dependency between the ensemble members. Nevertheless, despite the overfitting due to the size of the training data sample and some colinearity between members, the BMA calibration has proven to be effective in LEPS and has improved the ensemble reliability (Roquelaure and Bergot 2008).

d. Validation of the BMA calibration over a longer training period

In Roquelaure and Bergot (2008), the training period was composed of two winter seasons, from 2002 to 2004, and the BMA weights were computed over this period. This present study takes advantage of the overall data sample (five winters) to compute the BMA weights, in order to give more robustness to the calibration by reducing the overfitting and colinearity between the members. The Brier score (BS; Brier 1950) and its decomposition into resolution, reliability, and uncertainty are going to be used for the verification.

Figure 1 shows that the calibration over five winter seasons gives similar Brier scores (Fig. 1a) and reliability scores (Fig. 1b) compared with the calibration over two winter seasons. In both cases, the BMA weights (not shown) are comparable and explain the same patterns of behavior observed in the reliability scores in Fig. 1b. Consequently, at least for the local area around Charles de Gaulle Airport, two winter seasons of data are sufficient to compute efficient BMA weights necessary to calibrate the local ensemble. This improvement improves the ensemble reliability by the BMA calibration and demonstrates adequate ensemble sampling. In actuality, an inappropriate ensemble sampling would not have improved the ensemble reliability. Thus, the BMA calibration is able to capture the main ensemble member

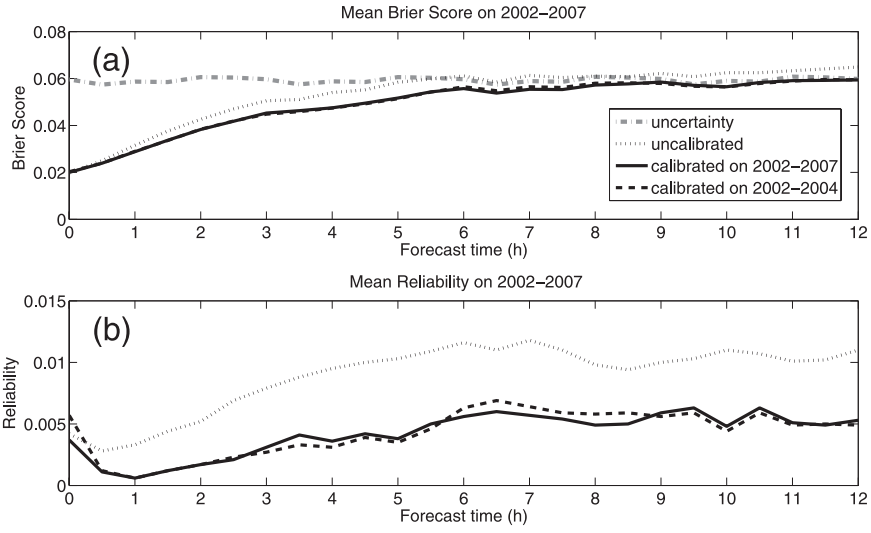

FIG. 1. (a) Brier score and (b) reliability for the five winter seasons from 2002 to 2007 for the uncalibrated ensemble (dotted line), LEPS calibrated for two winter seasons (2002–04) of data (dashed line), and LEPS calibrated for five winter seasons (2002–07) of data (solid line). In (a), the uncertainty (gray dashed–dotted line) is the probability score of the climatology.

contributions on the short 2-yr dataset as well as on a longer training dataset of five winter seasons. Moreover, cross-validated forecasts and verifications, using permutations of two-winter seasons to compute the BMA weights applied to a third independent winter, give similar results (not shown). As a consequence, the same 5-yr dataset is used for both calibration and validation in this study to take advantage of the full data training sample.

Table 2 presents the annual variability of the reliability and the resolution parts of the Brier score decomposition as well as the Brier score itself along with the uncertainty. In addition, as shown previously in Table 1, the winter seasons in the dataset do not have identical LVP event frequencies of occurrence. As a consequence, the mean results over the 12-h forecast period reveal the interannual variability of the Brier score, the resolution, the reliability, and the uncertainty (i.e., the probability score of the climatology). The percentages of variability of the BS, resolution, reliability, and uncertainty are 27%, 25%, 33%, and 23%, respectively (see Table 2).

4. Dataset and methodology

a. Methodology of analysis

Within an operational context, for most of the LVP cases, forecasters know which LVP type of event is going to occur but with some uncertainties. Forecasters may need to know the behavior of LEPS, or need to understand on which situations the system is doing well or poorly. Since LEPS is not able to discriminate LVP event types, the system only provides the overall LVP probability of occurrence (i.e., calibration performed for the LVP variable). The main goal of this study is to analyze the manner in which LEPS skill depends on the LVP event type in order to understand its predictive strengths. The analysis is performed only on observed LVP cases during the five winters of data to avoid the numerous numbers of days where no LVP occurred and good ''no–no'' forecasts were issued all day long (when no LVP is observed and no LVP is predicted). The only point of interest is the skill level of the local ensemble for these observed LVP cases (or dates); therefore, the skill level is conditional on the observed LVP type. As a result, the reason why incorrect forecasts are issued when LEPS predicts unobserved LVP events is not addressed in this present study; and should be addressed in the future.

TABLE 2. Variability of the scores over the five winter seasons from 2002 to 2007. The mean of the five winter seasons of the Brier score, the resolution, the reliability, and the uncertainty has been computed, as has their related standard deviations as a function of the forecast time and their mean results over the 12-h forecast period.

	Mean scores over the 12-h forecast period		
Score	Mean score	Mean std dev of the score	Mean std dev of the score (%)
Brier score	0.049	0.013	27
Resolution	0.016	0.004	25
Reliability	0.006	0.002	33
Uncertainty	0.060	0.014	23

Thus, if a radiation fog event is observed at the airport during the night of 29–30 of December between 2000 and 0300 UTC, these will be the dates–times selected for the analysis. Within 24 h of the selected dates, errors in the forecasts occur when LEPS predicts the onset of the LVP event earlier than the observations or when the LVP event dissipates later than in the observations. Within this context of fog prediction, the limits of predictability are rapidly reached because of the small spatiotemporal scales involved in a fog event's life cycle. So, only very short-term or, at best, short-term predictions are conceivable for producing accurate and reliable fog forecasts. Since the methodology of the analysis focuses on the very short-term forecast, three initialization times are used to analyze the skill level for a LVP event:

- I0 is the initialization time when the LVP event is observed by the on-site measurements and initialized in the ensemble runs. The goal is to evaluate the prediction skill of initialized events as well as the LVP dissipation times for events whose life cycles are under 12 h.
- I3 covers the previous initialization times, respectively, 3 h before the I0 runs. The LVP is observed at the airport site during the first 3 h of the simulation. As a consequence, it is not initialized in LEPS and should be forecasted by the system at between 3 and 12 h of forecast time. This 3–12-h prediction interval is analyzed to assess the prediction skill of noninitialized events.
- I6 is the run that is 6 h before I0. The LVP event is observed and should be forecasted in the second half of the simulation (6–12 h). The focus is also on prediction skill of noninitialized events but for longer-term forecasts.

Actually, in these I3 and I6 runs, LVP should be predicted by the model after, respectively, 2–3 and 5–6 h of model integration. These runs evaluate the quality of the prediction of the fog formation time and the skill of the system in forecasting the development of LVP events. Our analysis will then focus on forecast times between 3 and 12 h in I3 and forecast times between 6 and 12 h in I6.

b. Scores used in the analysis

The skill levels of LEPS for the observed cases are evaluated by comparing the hit rates (HRs) and the false alarm rates (FARs) of the LVP events for the reference deterministic run, the ensemble mean solution, and the ensemble probabilistic forecasts. If we define a as the number of both observed and correctly forecasted events, b as the number of not observed but forecasted events, c as the number of observed but not forecasted events, and d is the number of not observed and not forecasted events, HR and FAR are given by

$$\mathrm{HR} = \frac{a}{a+c}; \tag{1}$$

$$\mathrm{FAR} = \frac{b}{b+d}. \tag{2}$$

The area under the ROC curve is also another measure of the system skill (Richardson 2000; Buizza 2001). As a consequence, it is convenient to compute a so-called ROC area score (ROCA), which is an integrated score used to compare the skill levels of the various LVP-type forecasts for I0, I3, and I6. The larger this area is, the better the forecast skill. A perfect system has a ROCA value of 1 while a system displaying no skill has a ROCA value of 0. To further evaluate the skill of the LVP system for fog-type forecasts in terms of a probabilistic context, reliability and sharpness diagrams are also used to assess the reliability and the distributions of the probability for each of the three initialization times.

c. Dataset

Thanks to the LVP event classification, all the observed LVP cases during winter seasons 2002–07 are listed and classified into one of the six types defined in section 2a: ADV, RAD, STL, PCP, EVP, and UKN. LEPS has been run for the five winter seasons at a 3-h data assimilation frequency, from December to April for winter 2002/03, and from October to February for the next four winters. Only the listed dates, identified by the classification, and the three 12-h forecasts at initialization times I0, I3, and I6 are conserved for the analysis. Forecasts and observations are compared on 1-h time intervals for the forecast validation. Thirty-six minutes of LVP conditions have to be observed during a 1-h window to define an observed LVP case.

5. Results

Precipitation fog results are not presented and not discussed because PCP events are synoptically related to cyclogenesis situations (Tardif and Rasmussen 2008), which are not well described in a 1D local numerical prediction approach. As a consequence, PCP fog events are not well forecasted by COBEL–ISBA and therefore not by LEPS. EVP cases, which are marginal LVP events with the lowest observed frequency (0.6%), are not discussed either. Results of UKN LVP events are shown in all the figures but are not discussed since these events cannot be clearly identified by the classification. 13% of the cases are categorized as unknown and should be driven by a combination of phenomena, revealing the

difficulties in forecasting some fog events. Three types of LVP events account for 78% of the LVP dataset: stratus lowering base (43%), followed by radiation (28%), and then advection fog events (7%). As a consequence, the discussion of the results is focused on these three types of LVP events.

a. LEPS from the deterministic point of view

In this section, instead of viewing the LVP variable as a disjointed event (i.e., fog and low ceiling) with yes–no binary forecasts deduced from crossing the LVP threshold, the LVP event is thought of as a continuous event by considering the evolution of the cloud-base height event. When the event cloud base is 0 m, a fog event occurs and when the cloud base goes under 60 m without going down to 0 m, a low-ceiling event occurs. By using this interpretation of LVP, the mean forecast of the ensemble can be defined as the mean cloud-base solution from the ensemble members. This interpretation of the LVP variable gives a deterministic solution for the mean solution of the ensemble, which can be directly compared with the cloud-base height of the reference run. Therefore, LVP events are predicted by the mean of the ensemble when the mean cloud base of the ensemble is less than 60 m. The type of LVP events for the comparison between the ensemble mean and the reference are also based on the classification described in section 2.

1) RELATIVE OPERATING CHARACTERISTICS

Deterministic forecasts clearly facilitate the decision-making process when these forecasts are highly accurate and reliable. However, because of model and analysis errors, deterministic forecasts are not infallible, especially in the case of predicting rare and sensitive events like fog. For this reason, probabilistic forecasts are increasingly used in meteorological fields. Nevertheless, it is interesting to examine the skill of the ensemble reference forecast and to compare it with the mean forecast skill of the ensemble. In the present section, the mean solution of the ensemble is compared with the cloud-base height of the reference forecast for the dates where both the reference and the mean of the ensemble solutions predict an LVP event for each type of LVP event identified in the classification. Thus, the comparison relies on the same number of cases because the reference solution discriminate more LVP cases than does the ensemble mean (not shown).

Table 3 shows the percentage of LVP events predicted by both the reference and the mean forecasts for each LVP type and all initialization times: I0, I3, and I6. Globally, between 8% and 75% of the cases are forecasted by both the reference and the mean. As expected, the highest percentages of forecasted events are obtained when LVP events are observed at the airport and then initialized in the ensemble for I0 with about 70% and the smallest percentages are for I6 with about 10%. ROC curves in Fig. 2 rely on these cases for I0. Figures 2a and 2b present the ROC curves and the ROC area scores for the reference and the ensemble mean. The ROC curves are a function of the forecast time. From right to left on each curve, the mean results are presented for the four following forecast time periods: 0–3 (far right points), 3–6, 6–9, and 9–12 h (far left points).

TABLE 3. Percentage of LVP events predicted by both the reference run and the ensemble mean forecasts per LVP types over the five winter seasons. All three Initialization times are shown (I0, I3, and I6).

	LVP events predicted (%)			
Initialization time	STL	RAD	ADV	UKN
I0	63	79	54	75
I3	52	32	9	35
I6	37	14	9	20

Figure 2 shows the ROC curves and ROCA scores for the various LVP types for the reference and the ensemble mean when LVP events are initialized (I0). First, Fig. 2 shows that the reference generally has a better HR that does the ensemble mean. However, what the reference gains on the HR is lost in terms of the FAR as compared to the ensemble mean. As a consequence, STL and RAD have similar ROCA scores for both the reference and the ensemble mean. For ADV, the ensemble mean improves the forecast skill by 27% as compared to the reference. For UKN, the ensemble reduces the forecast skill by 46%. For I3 and I6 relative operating characteristics, for both the ensemble mean and the reference forecasts, the conclusions are similar but the ROC curves are clearly less skillful as fewer cases were predicted (see percentages in Table 3).

2) RELATIVE MEAN ERRORS IN LVP FORMATION AND DISSIPATION TIMES

Figure 3 shows the relative mean errors on the formation and dissipation times of each type of LVP event predicted by both the reference (non boldface symbols) and the ensemble mean (boldface symbols) forecasts during the five winter seasons. For initialized RAD and ADV events, the dissipation time is well forecast with mean errors within a range of ± 1 h for both the reference and the ensemble mean. The dissipation of STL is too late by about 3 h for the ensemble mean and by 5 h for the reference. The ensemble mean dissipates all types of LVP events earlier than does the reference forecast, by about 2 h for STL, RAD, and UKN and by about 1 h for ADV.

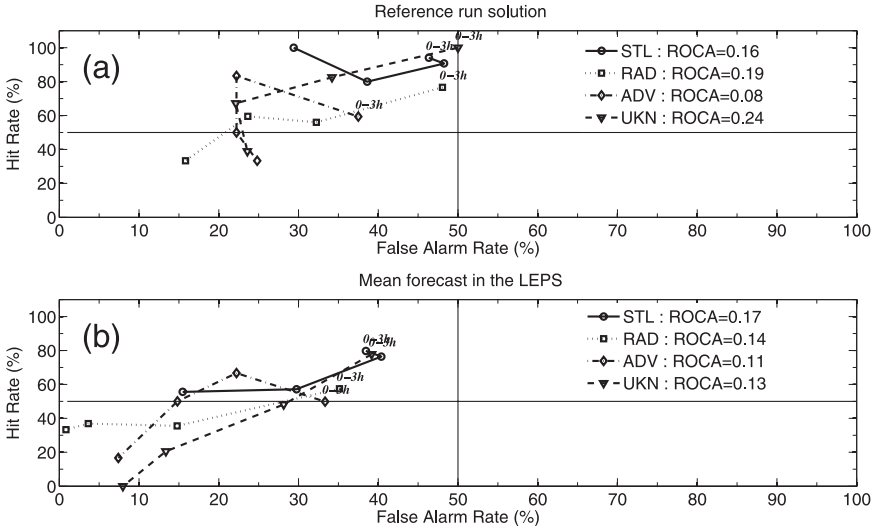

FIG. 2. ROC curves for the (a) reference LVP forecast and (b) mean LVP ensemble forecast over the five winter seasons for initialization time I0. The ROC curves are a function of the forecast time. From right to left on each curve, the mean results for the four following forecast time periods are shown: 0–3 h (indicated on the curves), then 3–6 h, 6–9 h, and 9–12 h (far left points).

For forecasts initialized 3 h prior to fog onset in Fig. 3b (I3), the errors in the formation time are less than 1 h for all of the events. The reference RAD dissipation is particularly good with unbiased results. The mean errors in the dissipation for STL also decreased as compared to I0. Once again, the ensemble mean dissipates earlier than the reference runs, moving the errors closer to unbiased predictions, with the exception of RAD fog events that dissipate 3 h too soon. Note that ADV statistics are not at all valuable since very few advective cases are forecasted and mean dissipation errors are over 6 h.

For forecasts initialized 6 h prior to fog onset (I6), LVPs form 2–3 h too early for both the reference and the ensemble mean. The mean errors in the dissipation time spread out with two distinct patterns of behavior for the reference and the ensemble mean forecasts, which

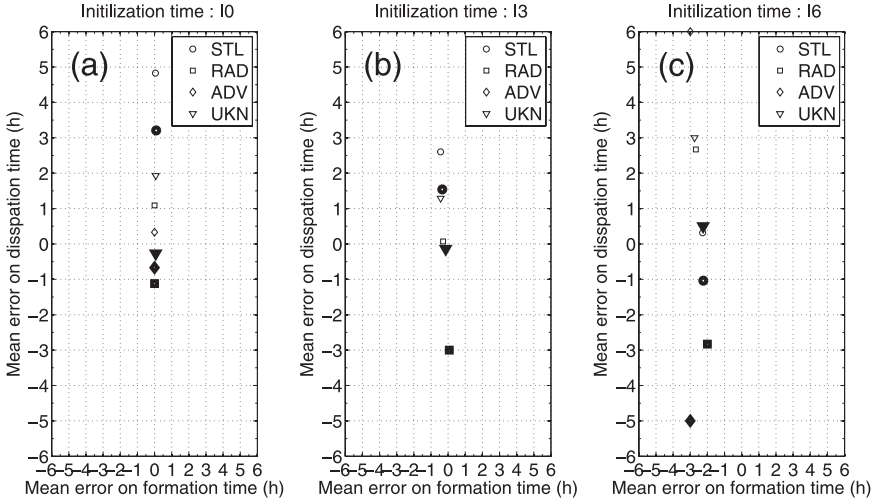

FIG. 3. Relative mean errors on the LVP formation and dissipation time as a function of the type of LVP events and for the initialization times (a) I0, (b) I3, and (c) I6 over the five winter seasons. Results are presented for all cases that are predicted by both the reference (non-bold signs) and the mean (bold signs) LVP forecasts. Positive values mean a later formation–dissipation and negative values mean an earlier formation–dissipation of the LVP event.

respectively dissipate too late and too early within the range of +3 h and −3 h. One can see that the dissipation time of STL is well forecast in both the reference and the ensemble mean, with less than 1 h of error. Again, ADV statistics are not at all useful since very few advective cases are forecasted and mean errors of the time of ADV dissipation are over 6 h. The uncertainties in these mean errors are independent of the type of LVP events and the differences among various forecasts shown in Fig. 3 are within the range of 1–3 h in the formation time and 3–5 h in the dissipation time.

3) SUMMARY OF DETERMINISTIC FORECASTS

Results show that the reference generally has a better HR than the ensemble mean. However, what the reference gains on the HR is lost in term of the FAR as compared to the ensemble mean. STL and RAD have similar ROCA scores for both the reference and the ensemble mean. For ADV, the ensemble mean improves the forecast skill by 27% as compared to the reference. For UKN the ensemble reduces the forecast skill by 46%. Results also show that radiation fog events are well forecasted by the reference run whatever the initialization time. This result is not surprising for two reasons. First, the COBEL model design takes advantage of a high vertical resolution grid, which is important for adequately forecasting radiation fog development and burnoff. Second, the model also includes a very detailed description of longwave radiation (Vehil et al. 1989), which has a 232 spectral intervals between 4 and 100 μm to evaluate the surface radiative cooling, and the hypothesis of horizontal homogeneity made in the 1D model design is more sustainable in the case of RAD fog events.

Long-range prediction of STL (I6) is also extremely skillful with less than 1 h of error in the burnoff time for both the reference and the ensemble mean forecasts.

ADV dissipation is well forecast when fog events are observed at the airport and initialized in the ensemble runs (I0) with mean errors on burnoff time in the range of ±−1 h.

b. LEPS from the probabilistic point of view

The Local Ensemble Prediction System skill is now assessed with ROC curves and the computation of the ROC area, which give a score to compare the prediction skill of LVP types. The following probability thresholds are plotted: $P > 90\%$, $>70\%$, $>30\%$, $>10\%$, and $>0\%$. Since fog events have low climatological frequencies, sample size is a concern. As a result, reliability and sharpness diagrams are also presented to evaluate the sample sizes of the datasets for the different fog types.

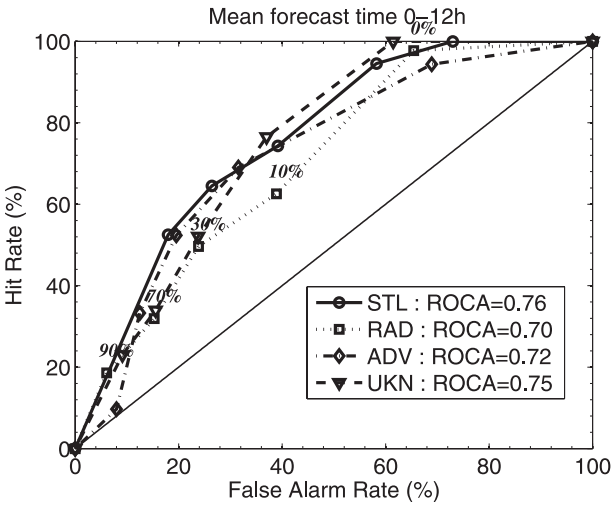

FIG. 4. ROC curves for the initialization time I0. The mean results over the 12-h forecast period of the simulation are shown. The points on the curves are for the following probability thresholds from left to right: $P > 90\%$, $>70\%$, $>30\%$, $>10\%$, and $>0\%$. The ROCA is also presented for each type of LVP event.

1) I0: LVP FORECAST SKILL WHEN EVENTS ARE OBSERVED ON THE AIRPORT AND INITIALIZED IN THE ENSEMBLE RUNS

For initialized LVP events, the ROCA scores reveal that the skill of the ensemble for STL, RAD, and ADV are all around 0.7 (Fig. 4). The skill, for STL, is 8% and 5% better than that of RAD and ADV, respectively. ADV events are well forecasted by LEPS when they are initialized and the skill is even 3% greater than for RAD over the 12-h forecast period. The reliability and sharpness diagrams in Fig. 5 show that the reliability is satisfactory for all types of LVP events, even if LEPS has been calibrated for the LVP variable.

2) I3 AND I6: LVP FORMATION–MATURATION FORECAST SKILL AND LONG-RANGE FORECAST OF NONINITIALIZED LVP EVENTS

With runs I3, 3 h before the initialized ones I0, LVP onsets should have been forecast after 2–3 h of simulation. Analyzing the forecast between 3 and 12 h provides an indication of LEPS's skill for the formation–maturation phase as well as the dissipation phase for events shorter than 9 h. Figure 6 shows that STL and RAD retain about the same level of skill as in I0 with ROCA scores of 0.7. For ADV, LEPS's ROCA score decreases by 43% and falls to 0.4. Figure 7 shows than the sample size becomes the main issue and that only STL still have enough cases to construct a good reliability diagram and LEPS tends to underpredict LVP events. ADV events are forecasted only with probabilities under 10%.

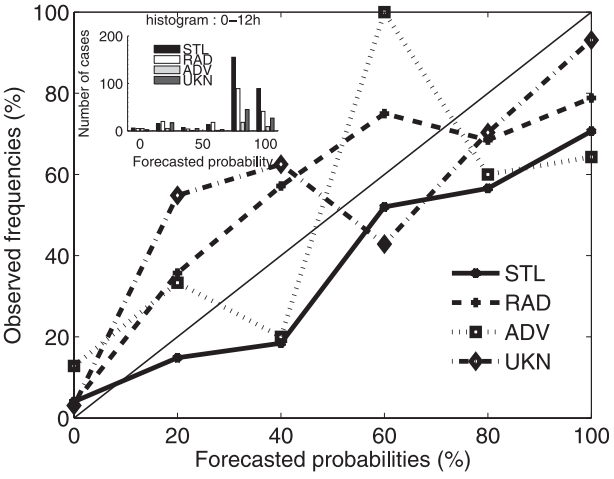

FIG. 5. Reliability and sharpness diagrams for the initialization time I0 (mean result over the 12-h forecast period). Results are shown for 20% probability bins.

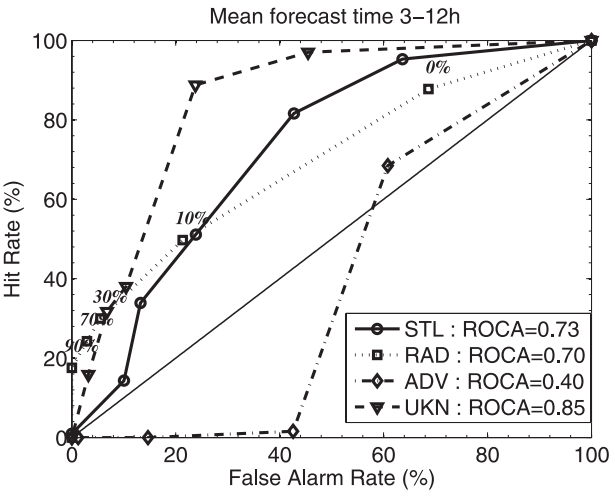

FIG. 6. As in Fig. 4, but for initialization time I3 with the mean result presented over 3–12 h.

Longer-range forecasts of noninitialized LVP events (I6) lead to similar conclusions to those with I3 except that the scores are lower (not shown). Longer-range forecasts for I6 yield smaller ROCA scores than I3. For STL and RAD events, the scores decrease by 10% and 23%, respectively, and ADV possesses almost the same level of skill as in I3 with a ROCA of 0.42. Reliability and sharpness diagrams reveal that fewer cases than I3 are forecast by LEPS 6 h before the onset of fog events. As a result, LEPS forecasts become bimodal with either probabilities under 40% or probabilities over 70% (also not shown). Low probabilities are much more frequent than high probabilities for long-range forecasts and low probabilities underestimate the observed frequencies, while high probabilities tend to overestimate the observed frequencies of LVP events.

3) SUMMARY ON PROBABILISTIC FORECASTS

The use of probabilistic forecasts imposes a decision-making procedure (actions) that is not an easy task to accomplish successfully. The user has to define a probability threshold that is linked to an HR and an FAR. However, ensembles are designed to assess predictability and one of the basic characteristics of an ensemble is that the ensemble predictions spread out as the lead time increases. As a consequence, high probabilities are only obtained for short-term forecasts. Therefore, the user has to define probability thresholds following the type of event to be predicted as well as the forecast ranges that are to be covered. The longer the forecast range is, the lower the probability threshold needs to be. Or, the higher the HR is required to be, the lower the probability threshold needs to be.

Regarding LEPS skill as related to LVP types, Fig. 8 summarizes the ROC curve when an HR over 50% is required. The specific skill of each type of LVP is presented with the associated probability thresholds and FAR, which would allow for the detection of at least 50% of the events. STL achieves the highest level of skill among the main three types of LVPs for all initialization times. ADV and RAD have similar skill aptitudes for initialized cases. However, LEPS loses a significant amount of skill for the prediction of noninitialized ADV events. As mentioned previously, high probability thresholds are obtained for I0 because, for this initialization time, LVPs are observed at the airport and therefore are initialized in the ensemble members. For an HR threshold above 50%, the probability criterion should be

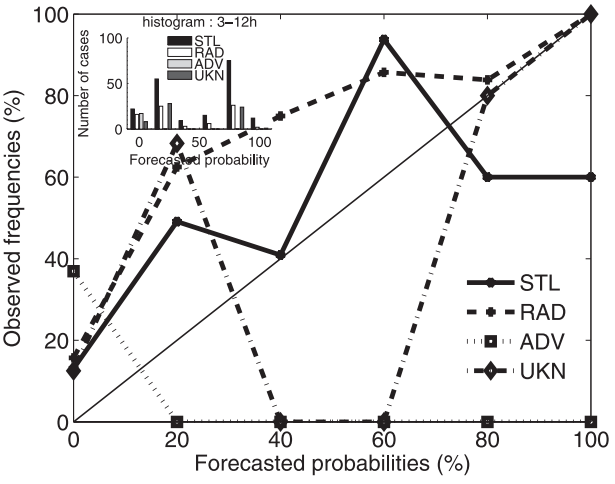

FIG. 7. As in Fig. 5, but for the initialization time I3 with the mean result over 3–12 h of the forecast.

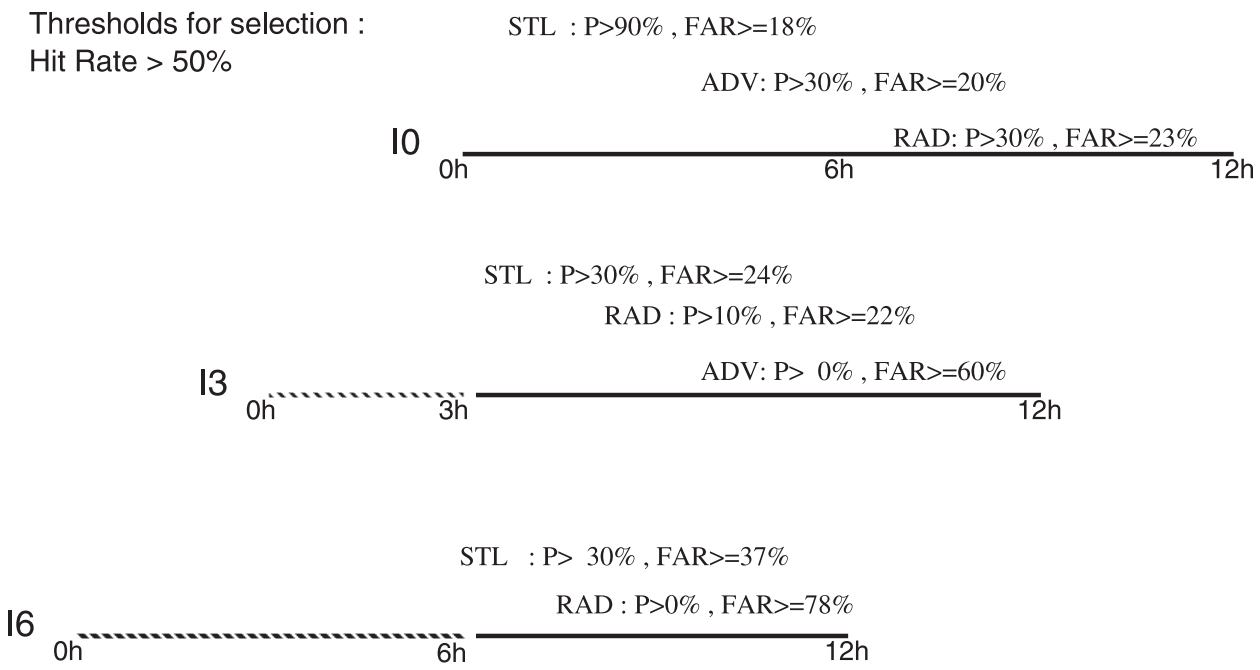

FIG. 8. Schematic summary of the ROC curves. By selecting a criterion of protection–prevention for HR > 50%. The skill levels of all types of LVP events are indicated for all initialized runs (I0, I3 and I6), as well as the corresponding FAR and probability thresholds.

relaxed for noninitialized runs I3 and I6; however, this will lead to an increase in the FAR.

Although, a 5-yr dataset has been used in the study, the results suffer from undersampled fog-type events. However, LEPS skill has been assessed for different LVP events that occurred at the Paris airport. One interesting result is that this local approach is able to provide valuable forecasts of advection fog events when the ensemble runs start with the initialization of the advection fog events. Even if advections are three-dimensional processes, the local influences of these mesoscale advections are treated by the local ensemble when ADV events are initialized. Therefore, we are able to obtain satisfactory results for the dissipation times of ADV events.

6. Conclusions

The LVP-type dependent skill of the Local Ensemble Prediction System has been assessed in this study to determine the skill of the system for various types of LVPs at Paris-Charles de Gaulle Airport. During the forecasting process, forecasters can identify the types of events occurring at the airport, which is not possible in LEPS. So, this study gives some insights into the system's performance for each type of LVP.

An interesting result of the LEPS calibration process (in section 3) demonstrates that the calibration is as efficient with two winters of data as with 5 winters of data. Thus, LEPS can be calibrated on short training samples and can still produce skillful forecasts, which is not the case in traditional 3D ensembles that require decades of data. The BMA calibration is very convenient in this local approach as it is applied to a single prediction threshold. As a result, each forecast solution is binary (1 if LVP is predicted and 0 if LVP is not predicted). Thus, the BMA is able to learn the main characteristics of the ensemble members and to assign to each member a specific weight that defines its ''identity.'' Therefore, when the ensemble is correctly sampled (i.e., the main sources of uncertainty are correctly described), the BMA improves the ensemble reliability, as has been shown in Roquelaure and Bergot (2008). Also shown in a previous LEPS paper by the present authors (2007), the first half of the run (0–6 h) is dominated by the uncertainty of the initial conditions and the second half (6–12 h) of the run by the uncertainty of the mesoscale forcings (e.g., advection, cloud cover).

Ensembles produce very large datasets that can be used not only as forecast tools but also for diagnostic purposes. Here, the local ensemble is used in a very basic way, in the sense that its forecasting skill is assessed through the reference, the ensemble mean (the mean cloud-base solution of the ensemble), and the ensemble probabilistic forecasts for each type of LVP identified at the airport site. The LVP-type identification has been carried out by using the classification algorithm of Tardif and Rasmussen (2007). The results show that 78% of the LVP dataset is dominated by three types of LVP events:

stratus lowering base (43%), radiation (28%), and advections fog events (7%).

The reference and the ensemble mean forecast results show the following:

1) Among all LVP type events, STL forecast skill is the highest for both the reference and the ensemble mean and ADV events are the most difficult to forecast.
2) The reference is very skillful for RAD fog prediction, which is not a surprise because COBEL has been designed to focus on radiation processes.
3) The relative mean forecast error of the LVP event formation time is under 2 h, which is in agreement with results of Bergot (2007).
4) The relative mean forecast error for LVP burnoff is generally under 3 h, which is also in agreement with the results in Bergot (2007).

LEPS probabilistic forecast results show the following:

1) The highest skill level is found for the prediction of RAD fog. This result can be explained by the design of COBEL, which focuses on representation of radiation processes.
2) STL events are also particularly well forecasted, especially for long-range forecasts; this is largely because of the low cloud initialization procedure, which initializes the clouds in all of the ensemble members. When the various scenarios spread out, the ensemble manages to encompass reality.
3) ADV fog prediction shows considerable skill but only for the initialized runs. This can be explained by the fact that perturbed advection members in the ensemble have significant weights in the ensemble calibration, which has a significant impact on the initialized fog events (Roquelaure and Bergot 2008). However, advections are three-dimensional phenomena and noninitialized ADV events are not well forecasted in the 1D modeling approach.

It is always necessary to keep in mind that probabilistic forecasts are user dependent. The choice of the criterion of protection–prevention actions depends entirely on the user's needs. One of the simple rules for understanding LEPS may be that the longer the forecast range is, the lower the probability threshold needs to be. Or, the higher the HR is required to be, the lower the probability threshold needs to be.

In conclusion, the LVP climatology, as well as the frequency of occurrence of each LVP event type, has been defined at Charles de Gaulle Airport. From 6 h before the initialization of LVP through COBEL–ISBA assimilation procedure, LEPS is able to produce skillful forecasts for stratus base lowering cases, and advection and radiation fog events, which represent 80% of the LVP cases in the five-winter training dataset. Via this study, forecasters can understand how LEPS will perform with respect of the type of LVP event to be forecast. This information is useful for forecasters who will be able to adequately use LEPS following its strengths.

Acknowledgments. The authors wish to thank the personnel of Météo-France involved in this research study for their help and fruitful comments, in particular the division of aeronautic prediction.

REFERENCES

Arya, P. A., 2001: *Introduction to Micrometeorology*. 2nd ed. Academic Press, 420 pp.

Baars, J. A., M. Witiw, and A. Al-Habash, 2003: Determining fog type in the Los Angeles basin using historic surface observation data. Preprints, *16th Conf. on Probability and Statistics in the Atmospheric Sciences/13th Symp. on Global Change and Climate Variations,* Long Beach, CA, Amer. Meteor. Soc., J3.8. [Available online at http://ams.confex.com/ams/pdfpapers/29133.pdf.]

Baker, R., J. Cramer, and J. Peters, 2002: Radiation fog: UPS Airlines conceptual models and forecast methods. Preprints, *10th Conf. on Aviation, Range, and Aerospace Meteorology,* Portland, OR, Amer. Meteor. Soc., 154–159.

Bergot, T., 1993: Modélisation du brouillard à l'aide d'un modèle 1D forcé par des champs mésoéchelle: Application à la prévision (Fog modelization using 1D model forced by mesoscale forcings: Forecasting applications) Ph.D. thesis 1546, Université Paul Sabatier, Toulouse, France, 192 pp. [Available from CNRM, Météo-France, 42 Ave. Coriolis, 31057 Toulouse CEDEX, France.]

——, 2007: Quality assessment of the Cobel-Isba numerical forecast system of fog and low clouds. *J. Pure Appl. Geophys.,* **164,** 1265–1282.

——, and D. Guédalia, 1994: Numerical forecasting of radiation fog. Part I: Numerical model and sensitivity tests. *Mon. Wea. Rev.,* **122,** 1218–1230.

——, D. Carrer, J. Noilhan, and P. Bougeault, 2005: Improved site-specific numerical prediction of fog and low clouds: A feasibility study. *Wea. Forecasting,* **20,** 627–646.

Boone, A., V. Masson, T. Meyers, and J. Noilhan, 2000: The influence of the inclusion of soil freezing on simulations by a soil–vegetation–atmosphere transfer scheme. *J. Appl. Meteor.,* **9,** 1544–1569.

Bremnes, J. B., and S. C. Michaelides, 2007: Probabilistic visibility forecasting using neural networks. *Pure Appl. Geophys.,* **164,** 1365–1381.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.,* **78,** 1–3.

Buizza, R., 2001: Accuracy and potential economic value of categorical and probabilistic forecasts of discrete events. *Mon. Wea. Rev.,* **129,** 2329–2345.

Fabbian, D., R. De Dear, and S. Lellyett, 2007: Application of artificial neural network forecasts to predict fog at Canberra International Airport. *Wea. Forecasting,* **22,** 372–381.

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.

Hamill, T. M., 2007: Comments on ''Calibrated surface temperature forecasts from the Canadian Ensemble Prediction System using Bayesian model averaging.'' *Mon. Wea. Rev.,* **135,** 4226–4230.

Koziara, M. C., J. R. Robert, and W. J. Thompson, 1983: Estimating marine fog probability using a model output statistics scheme. *Mon. Wea. Rev.,* **111,** 2333–2340.

Marzban, C., S. Leyton, and B. Colman, 2007: Ceiling and visibility via neural networks. *Wea. Forecasting,* **22,** 466–479.

Meyer, M. B., and G. G. Lala, 1990: Climatological aspects of radiation fog occurrence at Albany, New York. *J. Climate,* **3,** 577–586.

Petterssen, S., 1969: *Introduction to Meteorology*. 3rd ed. McGraw-Hill, 333 pp.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.,* **133,** 1155–1174.

Richardson, D. S., 2000: Skill and economic value of the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.,* **126,** 649–668.

Roquelaure, S., and T. Bergot, 2007: Seasonal sensitivity on COBEL-ISBA local forecast system for fog and low clouds. *J. Pure Appl. Geophys.,* **164,** 1283–1301.

——, and ——, 2008: A Local Ensemble Prediction System (L-EPS) for fog and low clouds: Construction, Bayesian model averaging calibration, and validation. *J. Appl. Meteor. Climatol.,* **47,** 3072–3088.

——, and ——, 2009: Contributions from a Local Ensemble Prediction System (LEPS) at improving fog and low clouds forecasts on airports. *Wea. Forecasting,* **24,** 39–52.

Tardif, R., and R. M. Rasmussen, 2007: Event-based climatology and typology of fog in the New York region. *J. Appl. Meteor. Climatol.,* **46,** 1141–1168.

——, and ——, 2008: Process-oriented analysis of environmental conditions associated with precipitation fog events in the New York city region. *J. Appl. Meteor. Climatol.,* **47,** 1681–1703.

Vehil, R., J. Monneris, D. Guédalia, and P. Sarthou, 1989: Study of radiative effects (long-wave and short-wave) within a fog layer. *Atmos. Res.,* **23,** 179–194.

Wilson, L. J., S. Beauregard, A. E. Raftery, and R. Verret, 2007: Calibrated surface temperature forecasts from the Canadian Ensemble Prediction System using Bayesian model averaging. *Mon. Wea. Rev.,* **135,** 1364–1385.

Zhou, B., J. Du, B. Ferrier, J. McQueen, and G. Dimego, 2007: Numerical forecast of fog - Central solutions. Preprints, *18th Conf. on Numerical Weather Prediction,* Park City, UT, Amer. Meteor. Soc., 8A.6. [Available online at http://ams.confex.com/ams/pdfpapers/123669.pdf.]