Contributions from a Local Ensemble Prediction System (LEPS) for Improving Fog and Low Cloud Forecasts at Airports

STEVIE ROQUELAURE AND THIERRY BERGOT

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

(Manuscript received 7 February 2008, in final form 19 June 2008)

ABSTRACT

At Paris' international airport, named Roissy Charles de Gaulle (CdG), air traffic safety and management as well as economic issues related to poor visibility conditions are crucial. Meteorologists face the challenge of supplying airport authorities with accurate forecasts of fog and cloud ceiling. A specific event, which is called a low visibility procedure (LVP), has been defined for a visibility under 600 m and/or a ceiling under 60 m. Forecasters have to provide two LVP human predictions at 0600 and 0900 local time, providing estimates of the LVP occurrence on the airport area for the next 3 h. This estimation has a probabilistic nature since the forecasters have to classify their forecasts into the following four forecast categories: "certain," "likely," "unlikely," and "excluded." A Local Ensemble Prediction System (LEPS) has been recently designed around the Code de Brouillard á l'Echelle Locale–Interactions between Soil, Biosphere, and Atmosphere (COBEL-ISBA) numerical model and has been tested to assess the predictability of LVP events and estimate their likelihood. This work compares the operational human LVP forecasts with LEPS LVP forecasts during the winter season 2004–05. This study shows that the use of LEPS for LVP prediction can significantly improve the current design of the operational LVP forecast by providing reliable forecasts up to 12 h ahead of time. Moreover, the system can be easily run on a personal computer without high computational resources.

1. Introduction

Over main international airports, forecasters have to deal with the prediction of infrequent events like fog and the life cycles of low clouds. At Paris' Charles de Gaulle (CdG) international airport, adverse ceiling and visibility conditions (visibility under 600 m and/or ceiling below 60 m) lead to the application of low visibility procedures (LVPs). The application of LVPs reduces airport efficiency for takeoffs–landings by a factor of two, causing aircraft delays or cancellations.

Since 2000 at Paris CdG, forecasters provide LVP bulletins at 0600 and 0900 local time to help airport authorities with their decision making process. The effort is made in order to more efficiently manage and secure the air traffic. The LVP bulletins provide very short-term forecasts of fog conditions for the next 3 h at the airport. With the use of all available information (local observations at the airport, 3D NWP forecasts, etc.), forecasters have to assess the LVP risk by providing a forecast of the following categories: excluded, unlikely, likely, and certain. As a consequence, the LVP human bulletins can be seen as probabilistic forecasts.

As part of the operational process, a local 1D approach has been implemented in 2005 at CdG airport to provide short-term forecasts of fog and low cloud evolution (Bergot et al. 2005; Bergot 2007). The 1D model, Code de Brouillard ál'Echelle Locale (local scale fog code; COBEL) (Bergot 1993), coupled with Interactions between Soil, Biosphere, and Atmosphere (ISBA) (Boone et al. 2000; Boone 2000), is used together with a 1D variational data assimilation (1DVAR) approach based on dedicated on-site local observations. A Local Ensemble Prediction System (LEPS) (Roquelaure and Bergot 2008) has been developed around COBEL-ISBA to explicitly estimate on a systematic basis the LVP likelihood at CdG airport. The system is built under the "perfect model hypothesis"; initial conditions and mesoscale forcings are then perturbed to produce the ensemble. A Bayesian model averaging (BMA) method has been applied for calibration of the local ensemble. Whereas the previous paper by Roquelaure and Bergot

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

DOI: 10.1175/2008WAF2222124.1

(2008) was on the validation of LEPS, including the implementation of the BMA for LVP prediction and the economic value aspect with a cost–loss study, this paper focuses on the comparison between human operational LVP forecasts and LEPS forecasts on the winter season 2004–05 at Roissy airport in order to show the added value of LEPS for LVP forecasting. LEPS has been run with a 3-h data assimilation frequency during the winter season 2004–05 and results are compared to the operational LVP bulletins. On-site observations during this period are used for validation of both LEPS and the operational bulletins.

As a consequence, the work presented in this paper aims to compare the operational LVP forecasts with those from the LEPS. In section 2 we describe the human operational LVP bulletin production process and characteristics. Section 3 presents LEPS. Section 4 compares the operational LVP bulletins to LEPS LVP forecasts over the winter season 2004–05. Section 5 exposes a case study occurring at the airport during the winter, on 17 January 2005. And finally, section 6 summarizes the results and concludes with the advantages of using LEPS for very short-term forecasts at CdG airport.

2. The operational LVP bulletins at Roissy Charles de Gaulle airport

Since 2000, LVP bulletins are issued at 0600 and 0900 LT by the Val d'Oise regional prediction center of Météo-France to support the morning air traffic control at the Paris CdG international airport. LVP conditions are defined by a visibility under 600 m and/or a cloud ceiling below 60 m. They are used to manage and secure the airport area and the air traffic by leading to actions like the delay or cancellation of flights, initiating specific procedures for protection, calling up extra staff, etc.

These bulletins deal with a rare event characterized by a very low climatological frequency (generally lower than 10%). Fogs are local phenomena, which result from the interaction between multiple physical processes (radiation, turbulence, advections, and microphysics). As a consequence the predictability of LVP events is difficult to estimate even for short-term predictions (Bergot 2007).

The production process of LVP bulletins is composed of two steps. First, the forecaster in charge produces a deterministic forecast based on his expertise using all available information like on-site observations and 3D forecasts from operational mesoscale numerical models. Then in the second step, he has to estimate the risk of LVP occurrence by classifying his forecast into the following categories: certain, likely, unlikely, or excluded. At the end of the process, the deterministic forecasts acquire a probabilistic nature and some probabilities are associated with each forecast category (see Table 1). The forecasters produce 3-h forecasts for the following forecast time periods:

- T1: between 0 and 30 min (0600–0630 or 0900–0930)
- T2: between 30 min and 1 h (0630–0700 or 0930–1000)
- T3: between 1 and 2 h (0700–0800 or 1000–1100)
- T4: between 2 and 3 h (0800–0900 or 1100–1200).

A forecasted LVP event is validated if the observed visibility goes below 600 m and/or the observed ceiling is less than 60 m (200 feet) during at least 6 min in one of the 4 forecast time periods listed above.

The evaluation of the LVP likelihood is subjective, since it relies on the forecaster's experience, which varies depending on the forecaster in charge of the bulletin production. Subjective forecasts usually lead to an overestimation of the risk of occurrence and to high false alarm rates (Murphy 1991). Moreover, the forecaster judgment is often influenced by the user's needs. On the one hand, the forecaster can be encouraged to take more "risk" in his forecast to avoid missed cases, if the user can tolerate high false alarm rates. On the other hand, he must be very careful if the user cannot tolerate the losses related to high false alarm rates.

3. The Local Ensemble Prediction System

a. LEPS construction

Ensemble prediction techniques are designed to estimate the level of confidence of a forecast. Theoretically, the goal is to make an explicit computation (through the Liouville equations) of the probability density function (pdf) of a forecast out of the pdf of the initial state (Ehrendorfer 1994). Multiple perturbed initial states, derived from the reference initial state, compose a sampling of the initial state pdf. However, even in a local approach, multiple model integrations of these perturbed states are costly and become rapidly prohibitive if one wants to obtain the whole true pdf. Therefore, the pdf has to be approximated using a finite sample of forecast scenarios. LEPS is built under the perfect model assumption and the model physic is not perturbed. The sampling strategy is based on perturbations of initial conditions and mesoscale forcings. The computation of the uncertainties of COBEL-ISBA inputs and the LVP forecast sensitivity to these uncertainties has been described in Roquelaure and Bergot (2007) and the validation of LEPS is described in Roquelaure and Bergot (2008).

TABLE 1. Table of correspondence between LEPS and the operational forecast categories.

Forecast categories of the local bulletins	Corresponding LEPS probability
Certain	$P > 90\%$
Likely	$50 \leq P < 90\%$
Unlikely	$0 < P < 50\%$
Excluded	$P = 0\%$

1) THE COBEL-ISBA NUMERICAL PREDICTION SCHEME

The local numerical prediction method currently used at Charles de Gaulle airport is based on the 1D high-resolution COBEL atmospheric model (Bergot 1993; Bergot and Guédalia 1994) coupled with the multilayer surface–vegetation–atmosphere transfer scheme ISBA (Boone et al. 2000; Boone 2000). COBEL-ISBA initial conditions are estimated using a 1D variational data assimilation system (Bergot et al. 2005). The system uses specific observations from a 30-m-high meteorological tower (atmospheric temperature and humidity, shortwave 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 (see http://www.cnrm.meteo.fr/aladin/).

LEPS is built around this local prediction scheme. Then COBEL-ISBA inputs are the atmospheric temperature and humidity profiles from the 1DVAR, the geostrophic wind profiles, the advection profiles, the cloud cover, and the soil temperature and 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.

2) ESTIMATION OF UNCERTAINTIES FOR INPUT PARAMETERS

Mesoscale forcing uncertainty computation 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 (see Roquelaure and Bergot 2007 for more details).

Initial condition uncertainties are estimated from errors on the observations for the soil and the lower part of the atmosphere where site observations are available (less than 30 m). Higher, NWP Aladin is used to provide both temperature and humidity profiles; as a consequence uncertainties are assessed with the spatiotemporal methodology described above for mesoscale forcings.

TABLE 2. The 54 ensemble members used in LEPS.

Perturbed parameter	Number of members
Reference member (unperturbed)	1
Advection	13
Cloud cover	11
Geostrophic wind	4
Atmospheric initial conditions	8
Soil initial conditions	8
Fog–stratus initialization	8

LEPS is composed by 54 members (Table 2) on the following parameters: atmospheric and soil initial conditions, fog–stratus initial conditions, geostrophic wind, cloud cover, and horizontal temperature and humidity advections (Roquelaure and Bergot 2008).

3) LEPS CALIBRATION

The calibration technique for LEPS follows the BMA method described in Raftery et al. (2005). The BMA is applied on a training dataset (winter seasons 2002–03 and 2003–04) to determine which members are the most efficient for the prediction of any quantity X (LVP in our case). Thanks to the apprenticeship, the BMA method assigns a weight to each member to improve the ensemble 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. The law of total probability states that the forecast probability density function, $p(X)$, is given by

$$p(X) = \sum_{k=1}^{K} p(X|M_k)p(M_k|X^T), \tag{1}$$

where $p(X|M_k)$ is the forecast pdf based on member M_k and $p(M_k|X^T)$ is the posterior probability of member M_k being correct on the training data. These posterior probabilities have to sum up to one, $\sum_{k=1}^{K} p(M_k|X^T) = 1$, and they can be interpreted as weights [$w_k = p(M_k|X^T)$]. The BMA weights w_k, $k = 1, \ldots, K$ and the variance σ^2 of the BMA pdf are estimated by maximum likelihood from the training data.

The strengths of the BMA method are the robustness of the weight computation algorithm and its simplicity in the case of binary forecasts, as is the case 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 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; Roquelaure and Bergot 2008). Overfitting occurs when the training data sample is too small and 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 improves the ensemble reliability (Roquelaure and Bergot 2008) because the sources of uncertainties (the soil and atmospheric initial conditions, the fog/stratus initial conditions, and the mesoscale forcings) are well sampled.

4) Verification scores

One of the most common measures of accuracy for verifying two-category probability forecasts is the Brier Score (BS) (Brier 1950). The Brier Score is used to evaluate ensemble skills. It is defined as the mean-square error of the probability forecast:

$$\mathrm{BS} = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2, \tag{2}$$

where N is the number of forecasts, p_i is the forecast probability, and o_i is the verifying observation (1 if LVP occurs, 0 if it does not). BS can be decomposed into three components; reliability, resolution, and uncertainty (Wilks 2006):

$$\mathrm{BS} = \mathrm{BS}_{\mathrm{rel}} - \mathrm{BS}_{\mathrm{resol}} + \mathrm{BS}_{\mathrm{unc}}, \tag{3}$$

where

$$\mathrm{BS}_{\mathrm{rel}} = \frac{1}{N}\sum_{k=1}^{T} n_k(p_k - \bar{o}_k)^2, \tag{4}$$

$$\mathrm{BS}_{\mathrm{resol}} = \frac{1}{N}\sum_{k=1}^{T} n_k(\bar{o}_k - \bar{o})^2, \tag{5}$$

$$\mathrm{BS}_{\mathrm{unc}} = \bar{o}(1 - \bar{o}). \tag{6}$$

When a sample of N forecasts has been divided in T categories, each comprises n_k forecasts of probability p_k; o_k is the observed frequency of LVP when the forecast was lying in that category, and $\bar{o}$ is the observed frequency of LVP in the whole sample.

Reliability ($\mathrm{BS}_{\mathrm{rel}}$) indicates the ability of the system to forecast accurate probabilities. Forecasted probabilities have to match observed frequencies. The reliability is negatively oriented, the smaller the better.

Resolution ($\mathrm{BS}_{\mathrm{resol}}$) reflects the ability of the system to provide the entire range of probabilities. The resolution is positively oriented, the higher the better.

Uncertainty ($\mathrm{BS}_{\mathrm{unc}}$) is the variance of observations. It indicates the intrinsic difficulty of forecasting the event and does not depend on the forecast system. Uncertainty is also the probability score of the sample climatology forecast.

The pseudo–relative operating characteristics (ROCs) curves also provide an efficient way of representing the quality of dichotomous, categorical, and also probabilistic forecasts. The method is based on ratios that measure the proportions of LVP events and nonevents for which warnings were provided. It evaluates the skill of the forecast system by comparing the hit rate (HR) and the pseudo–false alarm ratio (pseudo-FAR) of LVP events for different probability thresholds. Here in the case of rare events prediction, the pseudo-FAR is very convenient since it is computed as the ratio of forecasted and unobserved cases over LVP forecasted cases. This calculation removes the impact of the "no-no good forecasts" (no LVP forecasted and no LVP observed), which mostly dominates the data sample for rare events and hides the true skill of the LVP forecast system. If a is the number of observed and forecasted events, b is the number of not observed and forecasted events, and c is the number of observed and not forecasted events; HR and pseudo-FAR are defined by Eqs. (7) and (8):

$$\mathrm{HR} = \frac{a}{a + c}; \tag{7}$$

$$\text{pseudo-FAR} = \frac{b}{a + b}. \tag{8}$$

For skillful forecast systems, the pseudo-ROC curve bends toward the top left, where the HR is larger than the pseudo-FAR. The bottom left corner of ROC graphs is essentially dedicated to probabilistic information rather than the deterministic forecast (reference or mean); warnings are issued only when high percentages of members simulate LVP. Toward the top right corner, the criterion to issue warnings is relaxed, so they are issued more frequently. As a consequence, the HR increases significantly, but so does the pseudo-FAR.

b. LEPS and the production of the LVP likelihood forecast

LEPS has been run during 5 months in the winter season 2004–05 at Charles de Gaulle airport. Twelve-hour runs have been performed with a 3-h data assimilation frequency (about 1200 runs per winter). Observations have been collected during the same period. Data from two winters (2002–04) were used for BMA weights computation and the winter 2004–05 is kept for validation in order to preserve the independence between the training and the verification datasets. As for the operational bulletin, the LVP events forecasted by LEPS are validated when the observed visibility goes down below 600 m and/or the observed ceiling is under 60 m (200 feet) at least 6 min during one of the previously defined 4 forecast time periods (0–30 min, 30 min–1 h, 1–2 h, and 2–3 h). In its operational configuration, LEPS is run with a 3-h data assimilation frequency and produces 12-h LVP airport forecasts at 0000, 0300, 0600, 0900, 1200, 1500, 1800, and 2100 UTC. Consequently, it is difficult to directly compare the forecast times between LEPS and the operational bulletin that are in local time (made at 0600 and 0900 LT). Then, there is a time lag between the operational bulletin forecast times and LEPS forecast times. In the winter season in Paris, the local time equals UTC + 1. LEPS initialized at 0300 UTC corresponds to 0400 LT (LEPS at 0600 UTC corresponds to 0700 LT). According to LEPS initialization time, the forecast times used for comparison with the operational bulletins are not equivalent. Table 3 summarizes the forecast time lags for the LVP bulletin at 0600 LT. For local bulletins at 0600, LEPS initialized at 0300 UTC compares longer forecast times with the operational bulletin whereas LEPS initialized at 0600 UTC compares shorter forecast times.

4. Comparison between the operational LVP bulletin and LEPS LVP forecast at Charles de Gaulle airport

a. *The operational human bulletin skill*

In the aeronautic prediction sector, it is important to avoid LVP event misses in order to anticipate and manage accurately the air traffic and the airport activity. This constraint affects the forecaster evaluation of the LVP risk. Forecasters can decide to either be careful or audacious by taking more or less risk in their prediction according to the situation. However, all tendencies are moderated by the user, which certainly requires accurate LVP forecasts. As a consequence, forecasters face the difficult challenge of providing accurate LVP forecasts and avoiding misses. This task seems contradictory since forecasters have to take the risk to overpredict LVP if they want to avoid misses. By doing that, they will produce a significant number of false alarms.

Figure 1, showing the number of cases forecasted in each forecast time bin, reveals this dilemma (black column). Actually, forecasters do not want to miss LVP events. As soon as they feel a weak risk of LVP, they classify it in the unlikely bin whatever the forecast time period (T1, T2, T3, and T4). Clearly, fewer cases are forecasted in the other categories. The forecasts classified as certain are essentially issued for forecast times up to 1 h (T1 and T2) and few likely forecasts are issued in all forecast time periods. So, the unlikely forecasts dominated at all forecast times.

Figure 2 shows the reliability diagrams for the four

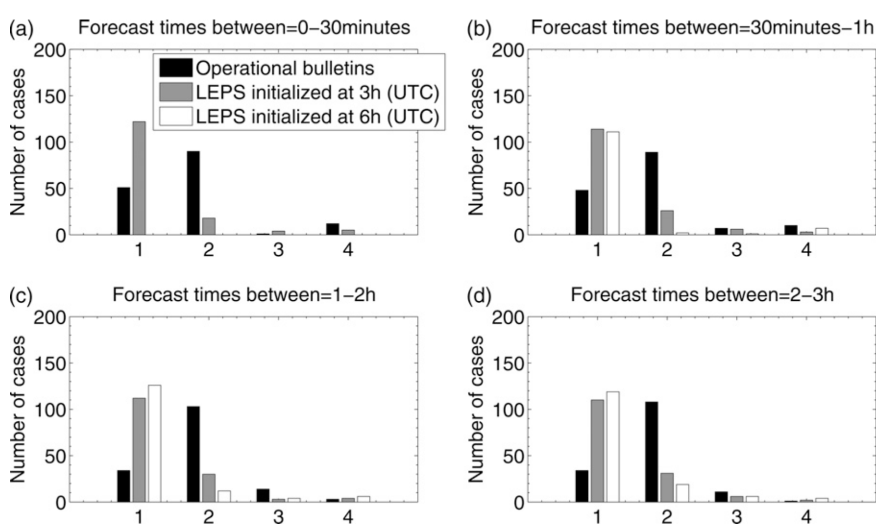

FIG. 1. The 0600 LT local operational and LEPS sharpness histograms for the defined LVP forecast time categories: (a) 0–30 min, (b) 30 min–1 h, (c) 1–2 h, and (d) 2–3 h. The forecasts are performed for the winter season 2004–05. Categories of forecasts are: 1 (excluded), 2 (unlikely), 3 (likely), and 4 (certain).

TABLE 3. Time lags between the operational human bulletins and LEPS forecasts in function of the LEPS initialization time for comparison with the 0600 LT bulletin.

Forecast time		
Local bulletin at 0600 LT	LEPS initialized at 0300 UTC	LEPS initialized at 0600 UTC
T1: 30 min	0230	—
T2: 1 h	0300	0000–initialization
T3: 2 h	0400	0100
T4: 3 h	0500	0200

probability forecast categories. Since these categories are defined for ranges of probabilities (Table 1), the correspondence with the observed frequency on the reliability diagrams in Fig. 2 is done using just single point values as follows: certain with the observed frequency of 100%, excluded with 0%, likely with 70%, and unlikely with 25%. The certain, excluded, and likely forecast categories are reliable for the longer-term forecasts in Figs. 2c,d (straight line). For shorter-range forecasts (Figs. 2a,b), the likely forecast category is less reliable, but it relies on few cases (see Fig. 1). Almost all the forecast uncertainty is contained in the unlikely class that is not reliable whatever the forecast time period. As a matter of fact, forecaster subjectivity is in the unlikely class where all potentially LVP cases are classified to try to avoid misses.

The pseudo-ROC curve (Fig. 3) confirms that the unlikely class leads to high false alarm rates. For this study, four thresholds have been chosen for the LEPS forecasts: $P > 90\%$ (lowest point on the LEPS pseudo-ROC curve), $P > 50\%$, $P > 20\%$, and $P > 0\%$ (highest point on the pseudo-ROC curve); and three thresholds for the operational bulletin: certain ($P > 90\%$, lowest point on the straight line), likely ($P > 50\%$), and unlikely ($P > 0\%$, highest point on the straight line) categories. For a forecast of 3 h, the forecasters forecast all the cases since the HR is 100%, but the pseudo-FAR varies between 60% and 80% according to the forecast time period. The operational forecast skill is good for T1 and T2 with high HR for the certain class,

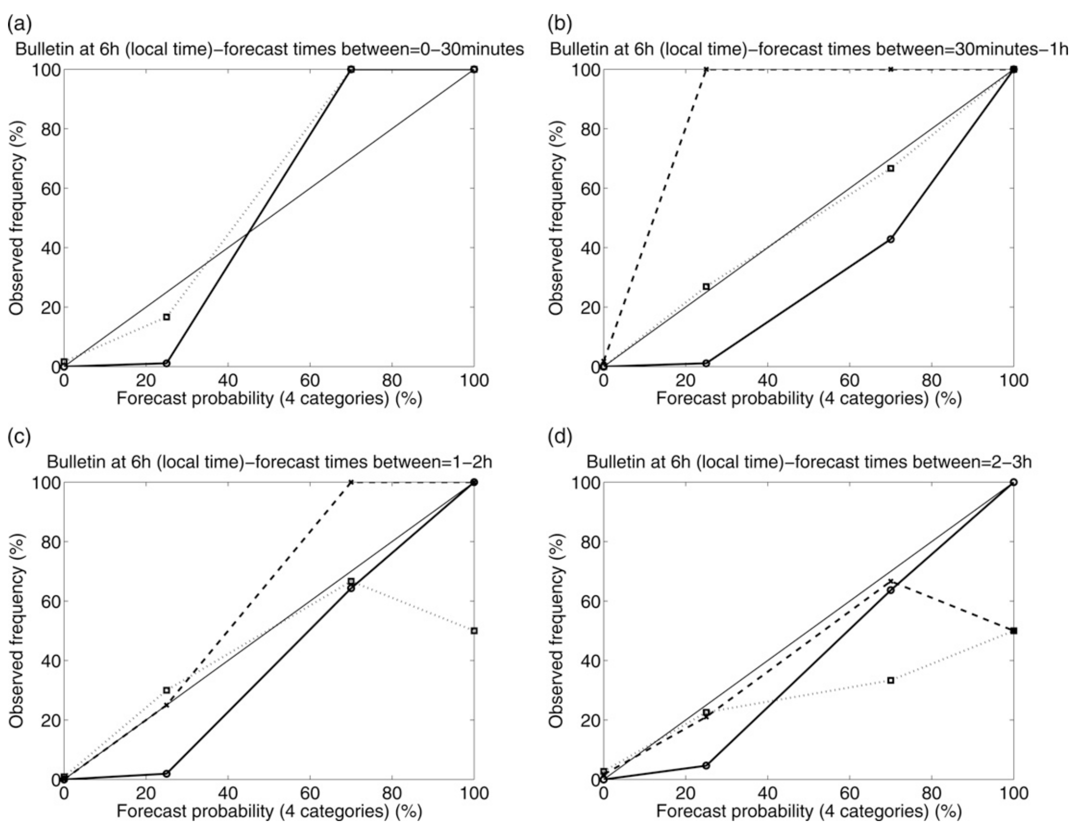

FIG. 2. The 0600 LT local operational and LEPS reliability diagrams for the LVP defined forecast time categories: (a) 0–30 min, (b) 30 min–1 h, (c) 1–2 h, and (d) 2–3 h. The forecasts are performed for winter season 2004–05: LEPS initialized at 0300 UTC (dotted line), LEPS initialized at 0600 UTC (dashed line), and operational forecasts at Roissy (straight line). The correspondence with the observed frequency is done with single points: certain with an observed frequency of 100%, excluded with 0%, likely with 70%, and unlikely with 25%.

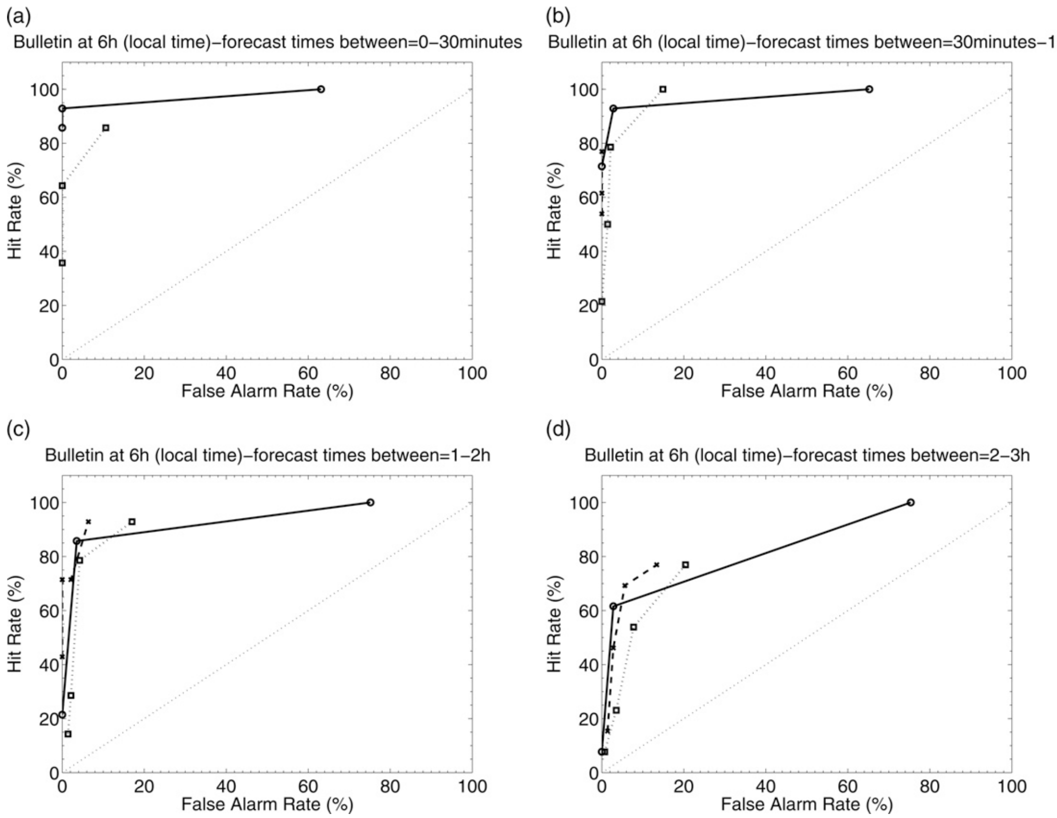

FIG. 3. The 0600 LT local operational and LEPS pseudo-ROC curves for the LVP defined forecast time categories: (a) 0–30 min, (b) 30 min–1 h, (c) 1–2 h, and (d) 2–3 h. The forecasts are performed for the winter season 2004–05: LEPS initialized at 0300 UTC (dotted line), LEPS initialized at 0600 UTC (dashed line), and operational forecasts at Roissy (straight line). Four thresholds have been chosen for the LEPS forecasts: $P > 90\%$ (lowest point on the LEPS pseudo-ROC curve), $P > 50\%$, $P > 20\%$, and $P > 0\%$ (highest point on the pseudo-ROC curve); and three thresholds for the operational bulletin: certain ($P > 90\%$, lowest point on the straight line), likely ($P > 50\%$), and unlikely ($P > 0\%$, highest point on the straight line).

respectively, 85% and 70% with almost no false alarms. In T3 and T4, the HR fails for the certain categories at, respectively, 20% and 8%. The forecast uncertainty is held in the unlikely categories, which lead to high false alarms (between 60% and 80% according to the forecast time period) and maximum detection rates (100%).

b. *LEPS LVP forecast skill*

LEPS computes explicitly the LVP probability density function. The system makes a systematic evaluation of the LVP likelihood, based on ensemble forecasts of the numerical model COBEL-ISBA. LEPS is able to provide objective probabilities of LVP conditions over the airport. Figure 1 (gray and white columns) and Figure 2 (dashed and dotted lines) show that LEPS is more reliable than the operational bulletin for the 3-h forecast and for both initialization times (0300 and 0600 UTC), except for the certain probability category at longer-range forecasts (1–2 and 2–3 h). The unlikely LEPS class is reliable compared to the operational bulletin. The LEPS unlikely forecast category contains one-third of the number of cases included in the unlikely class of the operational bulletin. The additional cases are found in the excluded class, which dominates the LEPS forecasts (Fig. 1). As a consequence, LEPS provides a more realistic histogram with an "L" shape than the operational bulletin having fewer observed cases. This L shape is expected for events with low climatological frequencies, such as fog. Distributions with this L shape are characteristic for rare events corresponding to much more excluded risk events than forecasted ones.

LEPS initialized at 0300 UTC uses longer forecast times than the 0600 UTC runs for the comparison with the operational bulletins. For the time periods T3 and T4, LEPS initialized at 0300 UTC has a less significant potential of discrimination of LVP events than LEPS initialized at 0600 UTC for the forecasts categories certain and likely, because the longer forecast times are used for the comparison. The pseudo-ROC curves

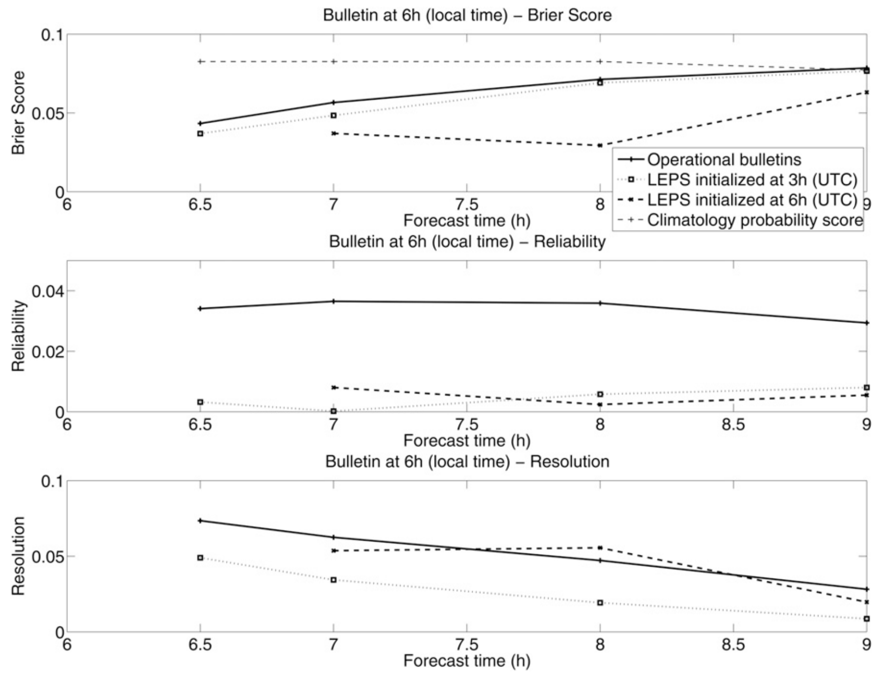

FIG. 4. (top) BS, (middle) reliability, and (bottom) resolution for the 0600 LT operational forecast and LEPS for the LVP defined forecast time categories: 0–30 min, 30 min–1 h, 1–2 h, and 2–3 h. The forecasts are performed for the winter season 2004–05: LEPS initialized at 0300 UTC (dotted line), LEPS initialized at 0600 UTC (dashed line), and operational forecasts at Roissy (straight line).

(Fig. 3) show that LEPS leads to significantly less false alarms than the operational bulletin. The unlikely forecasts issued by LEPS have a much lower pseudo-FAR, around 10%–20% instead of 60%–80% in the operational human bulletin, which is a positive consequence of the increased skill of LEPS forecasts. However, LEPS is not able to predict all the LVP cases during the winter since the ensemble is not able to capture all the LVP cases.

c. Summary: The Brier score and its decomposition

The Brier score and its decomposition into reliability, resolution, and uncertainty parts have been computed to assess LEPS skill and its contributions for the production of LVP bulletins. Figure 4 and Table 4 confirm the previous results and provide a summary of LEPS skill over the winter season.

- All three forecasts (operational, LEPS initialized at 0300 and 0600 UTC) give a better BS than the climatology for the four forecast time periods. The BS of LEPS initialized at 0300 UTC, which uses longer forecast times for comparison, is worse than the BS for LEPS initialized at 0600 UTC. Compared to the climatological uncertainty, LEPS initialized at 0300 and 0600 UTC improved the Brier Score on average by, respectively, 28% and 46%. Compared to the operational bulletins, LEPS initialized at 0300 and 0600 UTC improved the Brier Score by, respectively, 9% and 38% on average.
- LEPS major contribution is on the reliability of its forecasts. Both LEPS runs (initialized at 0300 and 0600 UTC) improve the reliability part of the BS by about 85% compared to the operational bulletins.
- The operational bulletin resolution is almost equal to LEPS resolution (initialized at 0600 UTC). But, resolution for LEPS initialized at 0300 UTC is less skillful because of longer forecast times used for the comparison with the operational bulletins (resolution is worsened by 52%).

5. A case study

a. The observed LVP case at CdG airport

The case presented in this section is not a classical nocturnal radiation fog. This is a case of stratus lowering occurring during the night of 17 January 2005 at Roissy Charles de Gaulle airport. Moreover, stratus-lowering events are interesting because they are more complex to forecast than classical radiation fog where

TABLE 4. Mean improvement–lowering of the BS and its decomposition computed by $(X_{\text{leps}} - X_{\text{baseline}})/X_{\text{baseline}}$, where X is either the BS, reliability, or resolution, and the baseline is either the uncertainty (clim) or the operational forecast bulletins (OP). A negative value means an improvement in the BS and the reliability but a lowering of the resolution.

	Percentage of improvement/damage on the score (%)			
	$\text{LEPS}_{0300\ \text{UTC}}$ vs clim	$\text{LEPS}_{0600\ \text{UTC}}$ vs clim	$\text{LEPS}_{0300\ \text{UTC}}$ vs OP	$\text{LEPS}_{0600\ \text{UTC}}$ vs OP
BS	−28	−46	−9	−38
Reliability	—	—	−87	−84
Resolution	—	—	−52	−9

the radiative cooling is clearly the main physical process responsible for the fog formation. In the case of stratus lowering, identifying the main physical process involved in the dynamical evolution of the cloud base is not at all straightforward because the cloud-base evolution is driven by a combination of different physical processes. It is interesting to see how LEPS handles stratus-lowering events. At 0000 UTC, the cloud base was close to 100 m (Fig. 5). The cloud base went down slowly between 0000 and 0300 UTC and reached the ground with a visibility close to 600 m between 0300 and 0600 UTC. As explained previously, the main physical processes involved in this LVP event are not obvious and clearly recognizable with available on-site observations. However, a cooling is observed at the ground during the night, and between 0000 and 0600 UTC the temperature decreases by almost 1.5°C, helping the surface condensation and the fog formation (Fig. 6). The upward longwave flux at the ground is about 5 W m^{-2} higher than the downward longwave flux (Fig. 7). Probably the cloud-base lowering is also caused by the cloud-base cooling by longwave emissions or advections because the radiative cooling is not strong enough to totally explain the evolution of the cloud base. The LVP event ends with sunrise; the surface was heated by shortwave radiation with the fog dissipating but the low cloud persisting with its base remaining under 60 m until 1200 UTC.

b. *LEPS LVP forecast*

1) LEPS PROBABILISTIC FORECASTS

In its operational configuration, LEPS is run with a 3-h data assimilation frequency and produces 12-h LVP

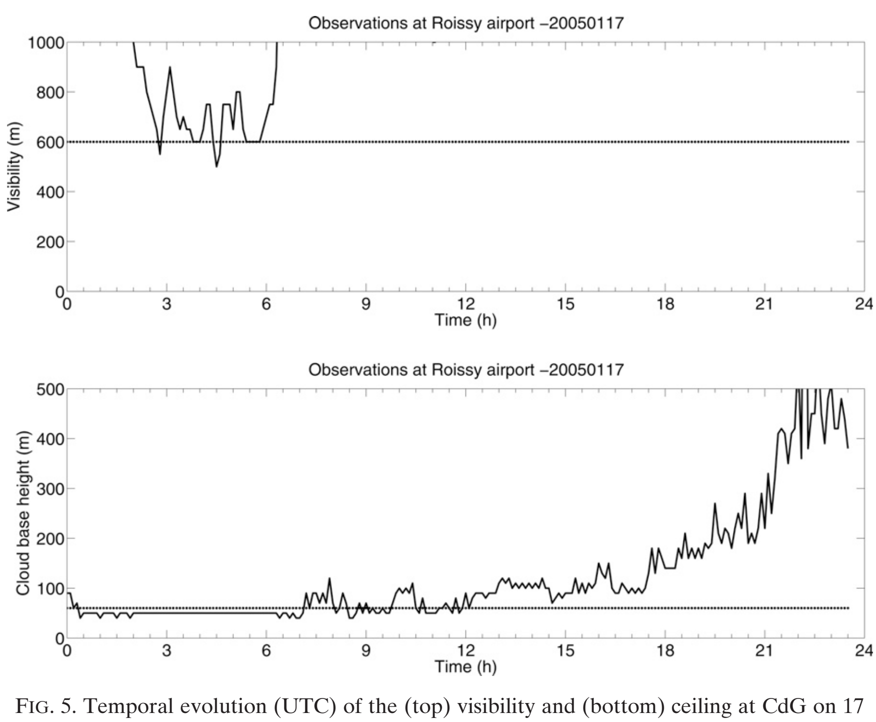

FIG. 5. Temporal evolution (UTC) of the (top) visibility and (bottom) ceiling at CdG on 17 Jan 2005.

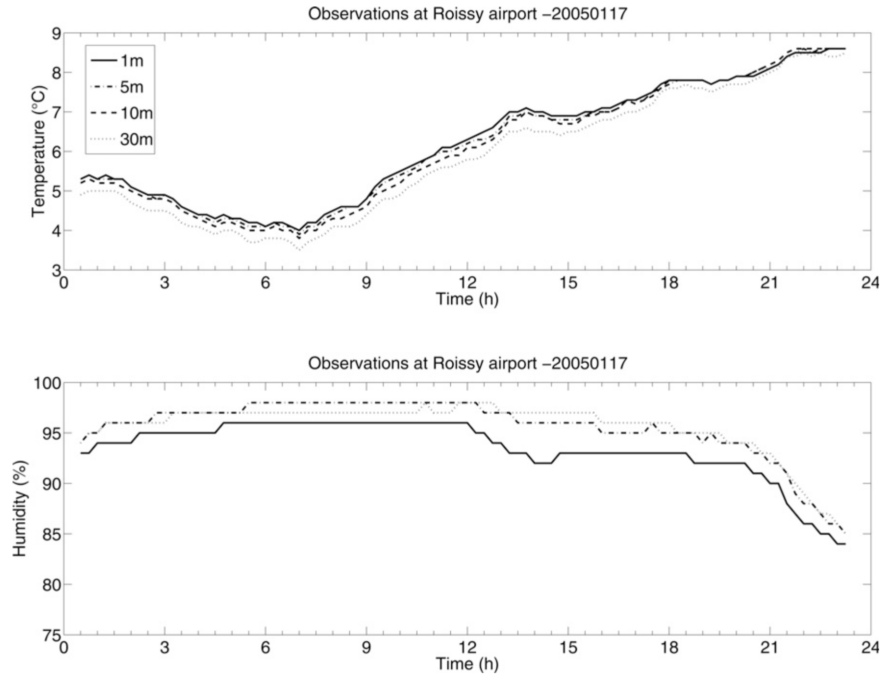

FIG. 6. Temporal evolution (UTC) of the (top) temperature and (bottom) humidity at CdG on 17 Jan 2005.

forecasts over the airport at 0000, 0300, 0600, 0900, 1200, 1500, 1800, and 2100 UTC. Then, it is possible to take into account the conditions leading to the LVP event and refine the forecast with each new run.

At 0000 UTC the low cloud is initialized with a cloud base at 100 m (Fig. 8). This figure presents spaghetti diagrams, which display the evolution of the cloud base for all raw ensemble members. The spaghetti

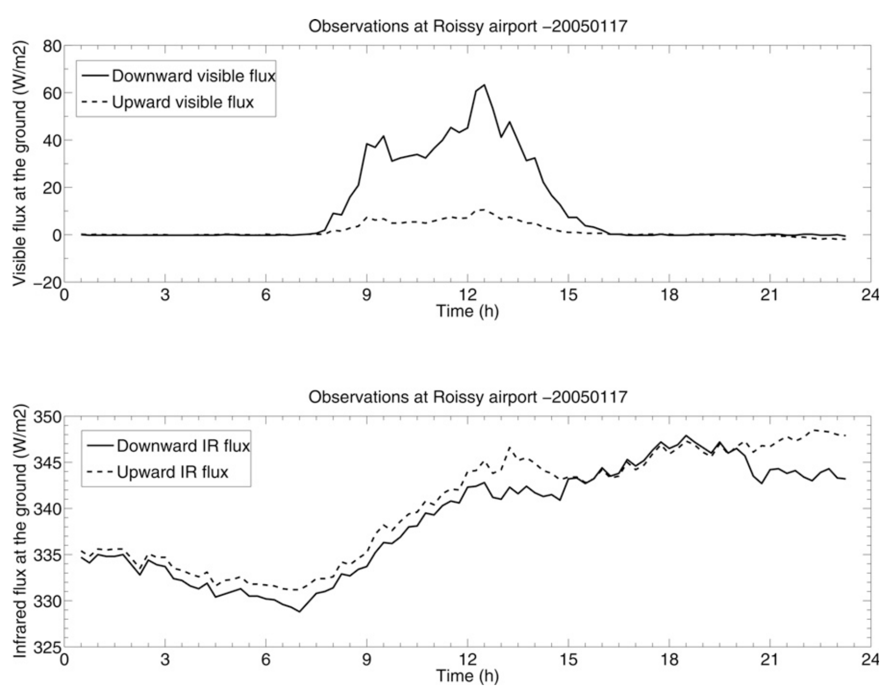

FIG. 7. Temporal evolution (UTC) of the (top) shortwave and (bottom) longwave radiation at CdG on 17 Jan 2005.

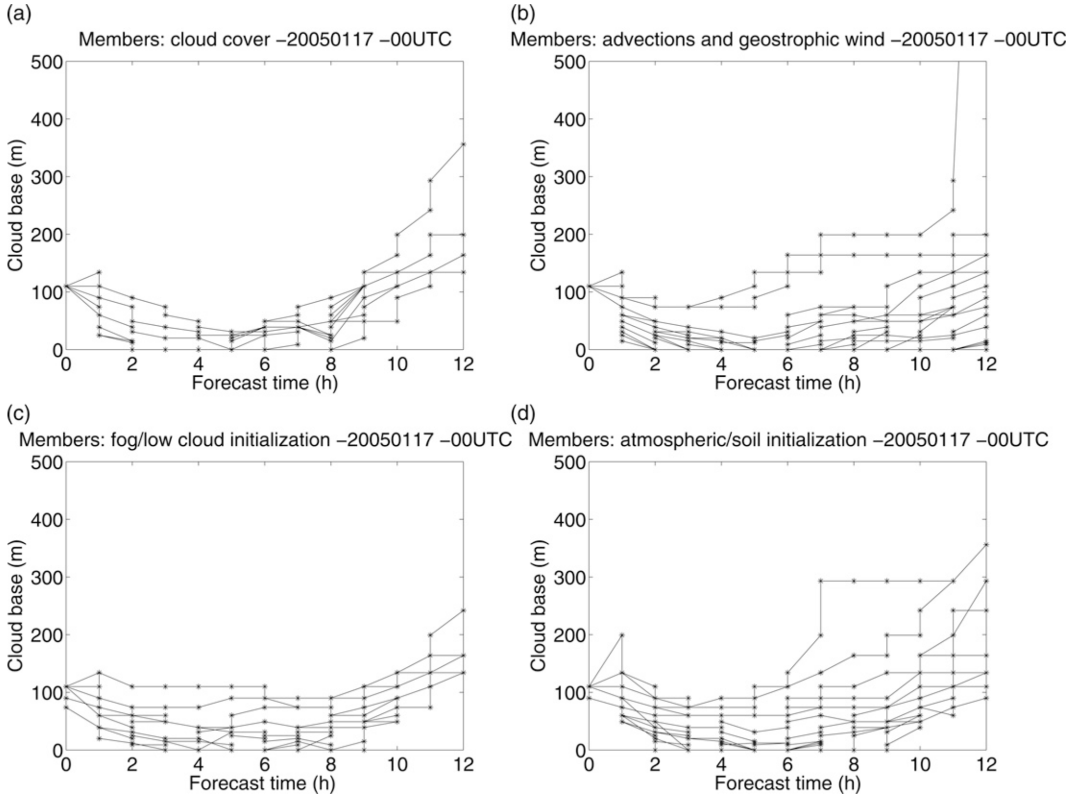

FIG. 8. The 54 rough ensemble members that were initialized for the 17 Jan 2005 0000 UTC forecast: (a) cloud cover, (b) advection and geostrophic wind, (c) fog–stratus, and (d) atmospheric–soil uncalibrated ensemble members.

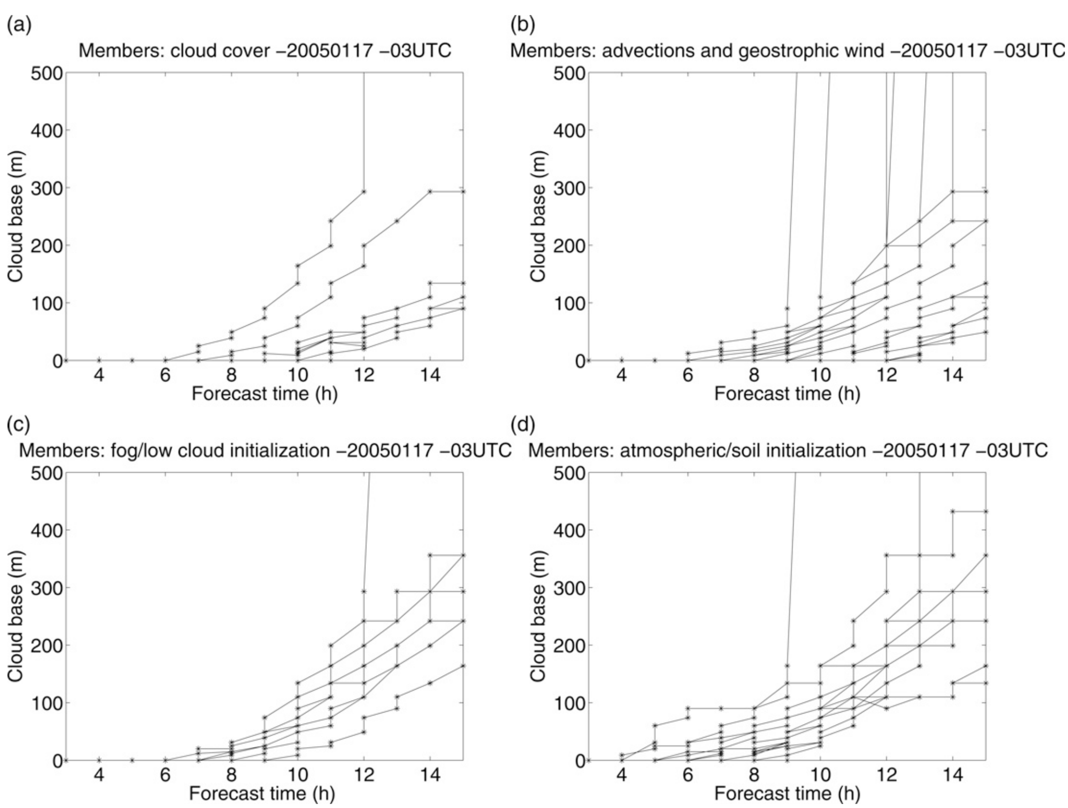

FIG. 9. Same as in Fig. 8, but for 0300 UTC.

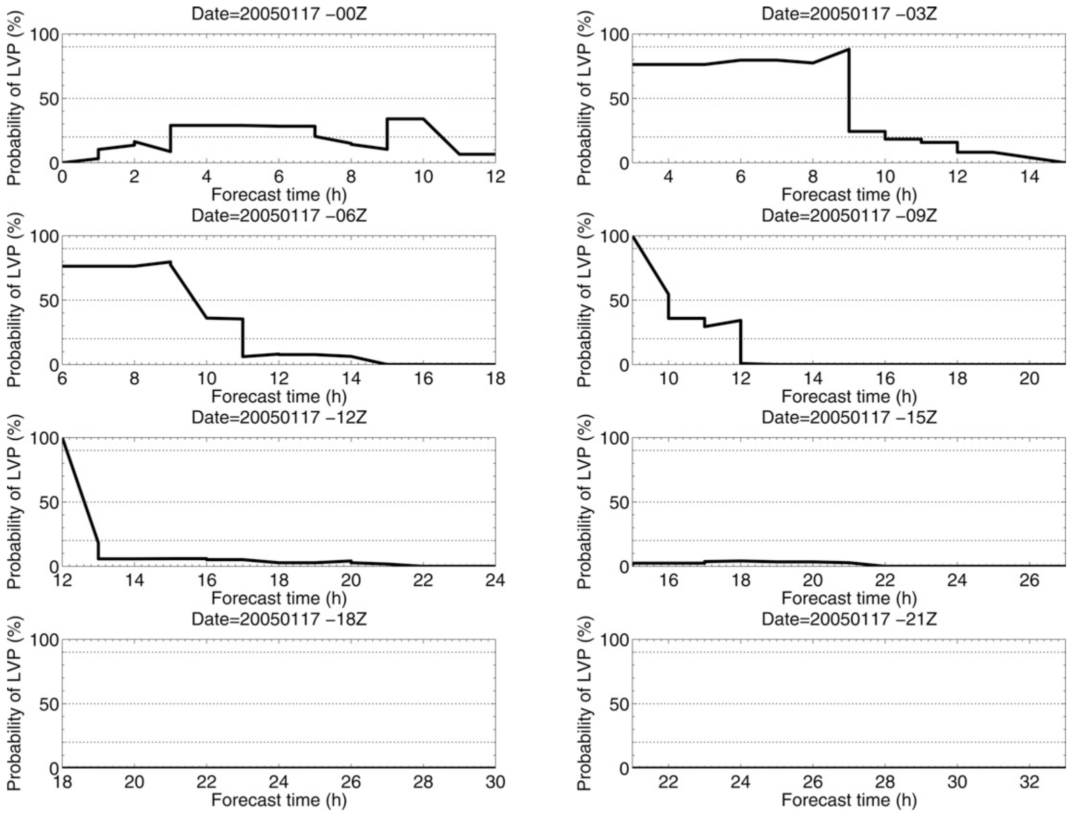

FIG. 10. The LVP probabilistic forecasts for the case study, 17 Jan 2005. The 8 daily runs were initialized at 0000, 0300, 0600, 0900, 1200, 1500, 1800, and 2100 UTC.

diagram shows that almost all the uncalibrated ensemble members forecast the cloud-base height decrease after 2–3 h of simulation. The fog formation is well captured as well as the fog dissipation after sunrise. At 0000 UTC, LEPS forecasts an unlikely risk of LVP during the 12-h forecast period. However, between 0300 and 0800 UTC the LVP risk increases and reaches 30% at 0900 UTC.

At 0300 UTC, the fog is detected on the site and the fog layer is initialized in all the ensemble members (Fig. 9). Between 0300 and 0900 UTC, the probability is 79% (likely risk). After 0900 UTC the probability decreases rapidly to 22%, then becomes 0% between 0900 and 1500 UTC where LVP is excluded by LEPS. All runs between 0600 and 2100 UTC confirm that the LVP risk is very low after 1300 UTC and can be considered as excluded after 1300 UTC. Figure 10 shows the LVP forecast of LEPS and the calibrated probabilities issued by the system for the lowering stratus case described previously.

2) CONTRIBUTION OF LEPS IN THE PRODUCTION OF THE LVP BULLETINS

Tables 5 and 6 present LEPS LVP forecasts by comparing them with the operational bulletins and the observations for the 0600- and 0900 LT local bulletins at the airport for the lowering stratus case study. On the four defined forecast time periods, LEPS LVP forecasts are sharper than the LVP bulletin. LEPS gives higher probabilities than the operation bulletins. LEPS could have helped in the decision making process for this case. Moreover, LEPS can provide longer forecasts, as the 6-h forecast remains accurate on the lowering stratus case. The LVP case is well forecast by LEPS, especially the dissipation phase (0900 UTC run; Table 6).

6. Conclusions

Currently at Roissy Charles de Gaulle airport, the LVP operational forecasts are composed of 2 3-h forecast bulletins at 0600 and 0900 LT. Forecasters used their expertise to estimate the LVP risk by classifying theirs forecasts into one the following categories: certain, likely, unlikely, and excluded. With this procedure they give a level of confidence on the LVP forecast. Thus, the LVP bulletins have a subjective probabilistic nature, which leads to biases on the forecasts and too many false alarms. Actually, the forecast categories that represent the prediction uncertainty (likely and unlikely)

TABLE 5. LEPS and the operational LVP bulletin at 0600 LT for the case study of 17 Jan 2005. C, L, U, and E, respectively, represent LVP forecast categories of certain, likely, unlikely, and excluded.

LT	0630	0700	0800	0900	1000	1100	1200
Observations	LVP	LVP	LVP	LVP	LVP	LVP	LVP
Operational bulletin	L	L	U	U	—	—	—
LEPS initialized at 0300 UTC	L	L	L	L	U	U	U
LEPS initialized at 0600 UTC	—	L	L	L	U	U	U

overestimated the number of LVP cases during the test winter season 2004–05. The unlikely forecast category is populated by too many cases, since forecasters use this category to predict all the borderline cases to avoid misses.

The Local Ensemble Prediction System (LEPS) has been designed around the COBEL-ISBA local prediction system integrated at Charles de Gaulle airport for the purpose of providing accurate forecasts of low visibility events over the airport. This 1D ensemble prediction strategy does not require high computational resources and can be run in real time on a personal computer. It is also a flexible strategy since it can be applied to other airports; however, a recalibration would be necessary, as each calibration is specific to the local climatology. LEPS compared to the operational bulletin gives reliable forecast probabilistic categories for the very short forecast range prediction period. The ensemble and its BMA calibration procedure proceed to an objective evaluation of the LVP risk, relying on the runs from the one-dimensional numerical model COBEL-ISBA. This systematic procedure of evaluation of the LVP risk produces reliable forecasts for the four defined probabilistic forecast categories (certain, likely, unlikely, and excluded). Then the forecast distribution provided by LEPS corresponds to a rare event distribution with the ''excluded'' forecast category holding the majority of the cases. LEPS reduces the number of false alarms by 50% or 60% according to the forecast time compared to the human oprational bulletin. LEPS has also been proven to be reliable on the 12-h forecast time period (Roquelaure and Bergot 2007); as a consequence the system can be used to provide longer forecast times than 3 h. It can also be used to refine the four LVP forecast categories (certain, likely, unlikely, and excluded) and define more appropriate ones. The unlikely forecast category ($0 < P < 50\%$) should be more restricted; another category should be defined for situations when the LVP risk is under 10%.

LEPS has been able to provide valuable forecasts for the case study, which was a lowering stratus during the night of 17 January 2005 at Roissy Charles de Gaulle airport. The 3-h data frequency has been useful in capturing the event and forecasting the time of the transition from an LVP to non-LVP event. In the case study, LEPS has been more efficient than the operational bulletin since it provided clearer insight about the evolving conditions. This system provides a potential for improving the LVP bulletin at the Roissy Charles de Gaulle airport by giving the possibility of increasing the forecast time period and refining the forecast categories. This case has highlighted LEPS' potential and its contributions for the prediction of LVP situations. Spaghetti diagrams and the ensemble mean are also interesting and can be provided to the forecasters from LEPS.

In conclusion, LEPS provides more reliable forecasts than the operational procedure used for the bulletins. LEPS reduces significantly the number of false alarms for the cost of a few missed cases. The current form on the LVP bulletin at Roissy CdG airport can then be improved by the use of this local ensemble prediction system. Forecasters will be able to produce more accurate LVP bulletins by integrating the additional information provided by LEPS: ensemble probability, the ensemble mean, and spaghetti diagrams. The mean of the ensemble is generally a more reliable forecast than the reference deterministic run, since the ensemble mean eliminates the unpredictable ensemble components and preserves the predictable ones. As the consequence,

TABLE 6. Same as in Table 5, but for LVP bulletin at 0900 LT.

LT	0930	1000	1100	1200	1300	1400	1500
Observations	LVP	LVP	LVP	LVP	LVP	no	no
Operational bulletin	U	U	U	U	—	—	—
LEPS initialized at 0600 UTC	L	L	U	U	U	U	U
LEPS initialized at 0900 UTC	—	C	L	U	U	E	E

forecasters can use the ensemble mean as a synthesis of the ensemble behavior. The reliability of forecasts will be significantly improved and the number of false alarms reduced as shown in this study.

REFERENCES

Bergot, T., 1993: Modélisation du brouillard à l'aide d'un modèle 1D forcé par des champs mésoéchelle: Application à la prévision. Ph.D. thesis, Université Paul Sabatier, 192 pp. [Available from CNRM, Météo-France, 42 Avenue Coriolis, 31057 Toulouse CEDEX, France.]

——, 2007: Quality assessment of the Cobel-Isba numerical forecast system of fog and low clowds. *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., 2000: Modélisation des processus hydrologiques dans le schéma de surface ISBA: Inclusion d'un réservoir hydrologique, du gel et la modélisation de la neige. Ph.D. thesis, Université Paul Sabatier, 236 pp. [Available from CNRM, Météo-France, 42 Avenue Coriolis, 31 057 Toulouse CEDEX, France.]

——, 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.,* **39,** 1544–1569.

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

Ehrendorfer, M., 1994: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: Theory. *Mon. Wea. Rev.,* **122,** 703–713.

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.

Murphy, A. H., 1991: Probabilities, odds and forecasts of rare events. *Wea. Forecasting,* **6,** 302–307.

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.

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 (LEPS) for fog and low clouds: Construction, Bayesian model averaging calibration, and validation. *J. Appl. Meteor. Climatol.,* **47,** 3072–3088.

Wilks, D. S., 2006: *Statistical Methods in Atmospheric Sciences.* 2nd ed. International Geophysical Series, Vol. 59, Academic Press, 627 pp.

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.