



**HAL**  
open science

## Verification of Ensemble-Based Uncertainty Circles around Tropical Cyclone Track Forecasts

Thierry Dupont, Matthieu Plu, Philippe Caroff, Ghislain Faure

► **To cite this version:**

Thierry Dupont, Matthieu Plu, Philippe Caroff, Ghislain Faure. Verification of Ensemble-Based Uncertainty Circles around Tropical Cyclone Track Forecasts. *Weather and Forecasting*, 2011, 26 (5), pp.664–676. 10.1175/WAF-D-11-00007.1 . hal-00961377

**HAL Id: hal-00961377**

**<https://hal.science/hal-00961377>**

Submitted on 3 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## Verification of Ensemble-Based Uncertainty Circles around Tropical Cyclone Track Forecasts

THIERRY DUPONT

*RSMC La Réunion, Météo-France, Sainte Clotilde, France*

MATTHIEU PLU

*Laboratoire de l'Atmosphère et des Cyclones, Météo-France, and Unité Mixte 8105  
CNRS/Météo-France/Université de La Réunion, Sainte Clotilde, France*

PHILIPPE CAROFF

*RSMC La Réunion, Météo-France, Sainte Clotilde, France*

GHISLAIN FAURE

*Laboratoire de l'Atmosphère et des Cyclones, Météo-France, and Unité Mixte 8105  
CNRS/Météo-France/Université de La Réunion, Sainte Clotilde, France*

(Manuscript received 8 January 2011, in final form 15 March 2011)

### ABSTRACT

Several tropical cyclone forecasting centers issue uncertainty information with regard to their official track forecasts, generally using the climatological distribution of position error. However, such methods are not able to convey information that depends on the situation. The purpose of the present study is to assess the skill of the Ensemble Prediction System (EPS) from the European Centre for Medium-Range Weather Forecasts (ECMWF) at measuring the uncertainty of up to 3-day track forecasts issued by the Regional Specialized Meteorological Centre (RSMC) La Réunion in the southwestern Indian Ocean. The dispersion of cyclone positions in the EPS is extracted and translated at the RSMC forecast position. The verification relies on existing methods for probabilistic forecasts that are presently adapted to a cyclone-position metric. First, the probability distribution of forecast positions is compared to the climatological distribution using Brier scores. The probabilistic forecasts have better scores than the climatology, particularly after applying a simple calibration scheme. Second, uncertainty circles are built by fixing the probability at 75%. Their skill at detecting small and large error values is assessed. The circles have some skill for large errors up to the 3-day forecast (and maybe after); but the detection of small radii is skillful only up to 2-day forecasts. The applied methodology may be used to assess and to compare the skill of different probabilistic forecasting systems of cyclone position.

### 1. Introduction

Although tropical cyclone (TC) track forecasts have been steadily improving for several decades (Avila et al. 2006), some uncertainty still remains. A part of this uncertainty is due to an inherent predictability bound (Fraedrich and Leslie 1989; Plu 2011) that future

improvements in numerical models and in forecasting techniques will not be able to overcome. End users of TC forecasts, such as risk managers and public agencies, need both reliable track forecasts and an estimation of the forecast uncertainty. In the southwestern Indian Ocean (SWIO), the Regional Specialized Meteorological Centre of La Réunion (Météo-France) issues TC forecasts and warnings to the countries in this area for up to a 3-day lead time. RSMC La Réunion has developed a new technique to measure and to display the uncertainty of its official track forecast. The main purpose of the present article is a presentation and verification of this technique.

---

*Corresponding author address:* Thierry Dupont, Direction Interrégionale de La Réunion, Météo-France, BP 4, 97491 Sainte Clotilde CEDEX 09, France.  
E-mail: thierry.dupont@meteo.fr

Most RSMCs (Miami, Florida; Tokyo, Japan; and Hawaii) and Tropical Cyclone Warning Centers currently display uncertainty “cones” around their official track forecasts, using a climatological method based on their area of responsibility. For each forecast lead time, an uncertainty circle is built whose radius is taken as a fixed quantile (e.g., 67% for Miami, 70% for Tokyo) of the distribution of direct position error (DPE) computed over several previous seasons. The Joint Typhoon Warning Center (JTWC) in Hawaii produces uncertainty cones whose radii are the sum of the climatological average DPE and the forecasted 34-kt wind radius.

However, a fixed climatological error is not a sufficient measure of the uncertainty of a TC track since such tracks are expected to be sensitive to internal and external factors. Tropical cyclone motion is indeed driven by the evolution of its intensity, its structure, and its environment; and by the complex interactions between these factors. Since the first studies on TC motion (as compiled by Emanuel 2003), many mechanisms have been highlighted by researchers: advection by the environmental steering flow (Chan and Gray 1982), drifting by the planetary  $\beta$  (Rossby 1948) and similarly by the environmental gradient of vorticity (Shapiro 1992), asymmetries in the inner TC structure (Shapiro and Franklin 1998), and complex interactions between the TC structure and its environment (Chan et al. 2002) that are sensitive to many physical processes (Chan 2005). It is therefore expected that these complex interactions result in a high dependency of the TC track predictability on the cyclone features and on its environment. Forecasting a situation-dependent uncertainty should thus be of significant added value.

The purpose of the present article is to demonstrate the skill of the uncertainty circles built around the official RSMC La Réunion TC track forecast, computed from the Ensemble Prediction System (EPS) of the European Centre for Medium-Range Weather Forecasts (ECMWF). The area swept out by the circles at successive forecast lead times forms an uncertainty cone. The main reason why the EPS of ECMWF has been chosen is that it is recognized as one of the best among global ensemble systems (Buizza et al. 2005; Bourke et al. 2005). In addition, some specifics are dedicated to TC forecasts: the singular vectors used for computing the initial perturbations are targeted on a TC (Puri et al. 2001) and they are computed along a tangent-linear model using diabatic physics (Barkmeijer et al. 2001).

Very few studies have been dedicated to the verification of ensemble track forecasts. Recently, Yamaguchi et al. (2009) demonstrated that the spread of the Japan Meteorological Agency (JMA) ensemble in the northwest Pacific may be used as an indicator of the DPE of the

ensemble mean. A confidence level of the forecast position is deduced from the ensemble spread. A clear spread–skill relationship appears between the ensemble spread and the error of the ensemble mean. Majumdar and Finocchio (2010) found a similar relationship for ensembles in the North Atlantic. Furthermore, they showed some first probabilistic results on uncertainty circles deduced from ensemble forecasts. Some verification is performed using as a reference, the analyzed position of the TC a posteriori; this is usually called the best-track analysis. The measure of how often the best-track analysis is inside the uncertainty circle suggests that the EPS is able to discriminate between large and small error values.

From these first promising results obtained in different TC basins, the present article aims at going further. The uncertainty circles built from the EPS will be assessed according to a metric relevant to TC position. The probabilistic verification method will aim at testing whether these uncertainty circles better describe the distribution of the RSMC forecast than the climatological circles. The method will show

- some probabilistic scores associated with the distribution of position error deduced from the EPS and
- whether the uncertainty circles are able to discriminate between large and small forecast errors.

The outline of the article follows the presentation of these issues. After an overview on the data, the methodology, and the calibration, section 2 presents some scores associated with the probability distribution of forecast positions. Section 3 describes the construction of uncertainty circles at a fixed probability value and their skill at estimating the DPE. The article ends with a brief conclusion.

## 2. Probabilistic forecast of TC positions

### a. Data

The data sample is composed of the tracks of TCs of all intensities over the southwest Indian Ocean (between the equator and 40°S, from the African coast to 90°E), during the seasons 2007–08 and 2008–09. A total of 225 forecast tracks are available until the 3-day lead time. The observed TC positions are from the official RSMC best-track dataset. The RSMC forecast tracks result from a manually analyzed forecasting process, which relies on output from several numerical weather prediction models (from, e.g., Météo-France, ECMWF, the Met Office, and from different U.S. agencies) and some ensemble means. Consensus forecasts (Goerss 2000) that gather all or part of the numerical forecast tracks are used and they generally improve the RSMC forecast.

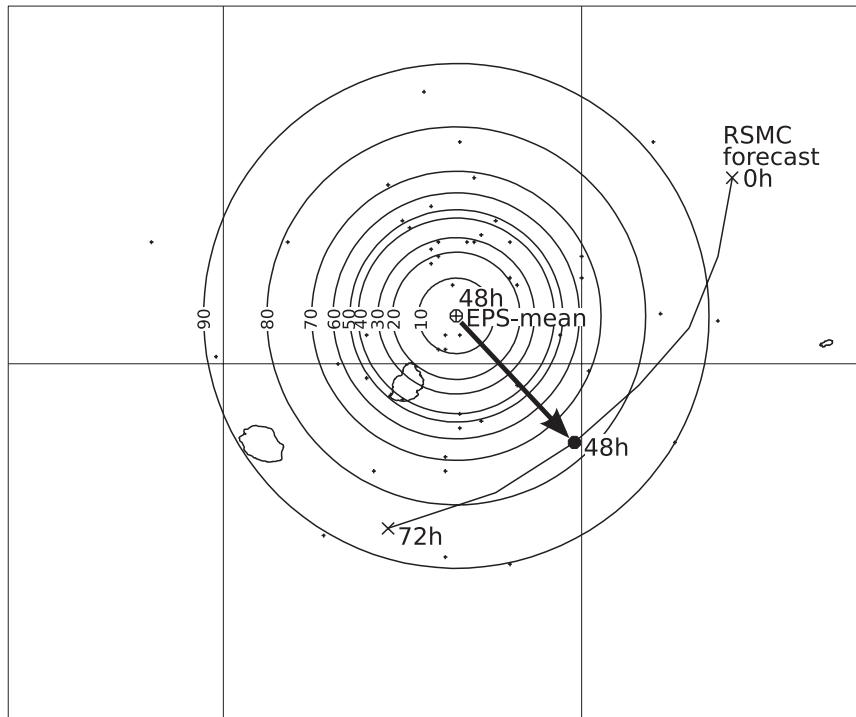


FIG. 1. Example of the construction of the distribution of forecast position probabilities at a given forecast lead time (here 48 h). The official RSMC forecast track is plotted in black, with the forecast position at 48 h indicated by a boldface black dot. The radii of the circles of different probabilities (%) are deduced from the  $n$  EPS positions (small crosses) by interpolation of the quantile  $p$  among the distribution of geodesic distance of the  $n$  members from the  $n$ -mean position. The circles describing the probability distribution are obtained after translation of their centers at the official RSMC forecast position, represented by the black arrow.

The TC positions forecasted by the EPS members are obtained from the ECMWF's tracking algorithm (described by van der Grijn 2002). The EPS was modified during the sample period; however, these changes are expected to have little impact on the spread of TC track forecasts. In particular its resolution remained the same (T399 until the 10-day lead time), and the method for computing the targeted TC initial perturbations (Puri et al. 2001) did not change.

The distribution of DPE, sometimes and hereafter called the distribution of climatological error, is measured by the distance between the RSMC forecast and the RSMC best-track position at the corresponding lead time. This distribution of DPE is computed over the 2007–08 and 2008–09 seasons.

*b. Definition of a probability distribution of forecast position*

The output from an ensemble includes the probability distribution of the forecast, usually displayed as the probability of a given meteorological parameter at a given point (rainfall, wind, etc). Concerning the position

of a TC, an ensemble forecast provides the successive positions of the TC for each of the members that predict this TC. To be mathematically and visually manageable, this ensemble of TC positions at a given forecast lead time is converted into a probability distribution of forecast position  $p(y)$ , defined by a set of concentric circles of different radii (Fig. 1). The radii are obtained from the ensemble forecast positions and then the circles are translated to the RSMC forecast position. If  $n$  (among the total ensemble size 51) EPS members have detected a TC, the radius of the circle of probability  $p(y)$  is obtained by interpolation of the quantile  $p(y)$  among the distribution of the geodesic distances of the  $n$  members from the  $n$ -mean position.

To summarize, the distribution of probability for a forecast lead time is represented by concentric circles centered on the RSMC forecast position with radii deduced from the ensemble spread (Fig. 1). As emphasized by Majumdar and Finocchio (2010), the use of circles leads to an isotropic distribution. The probability distribution is centered on the RSMC forecast position, which is assumed to be the most probable one. This is consistent with the fact that the RSMC forecasts perform globally

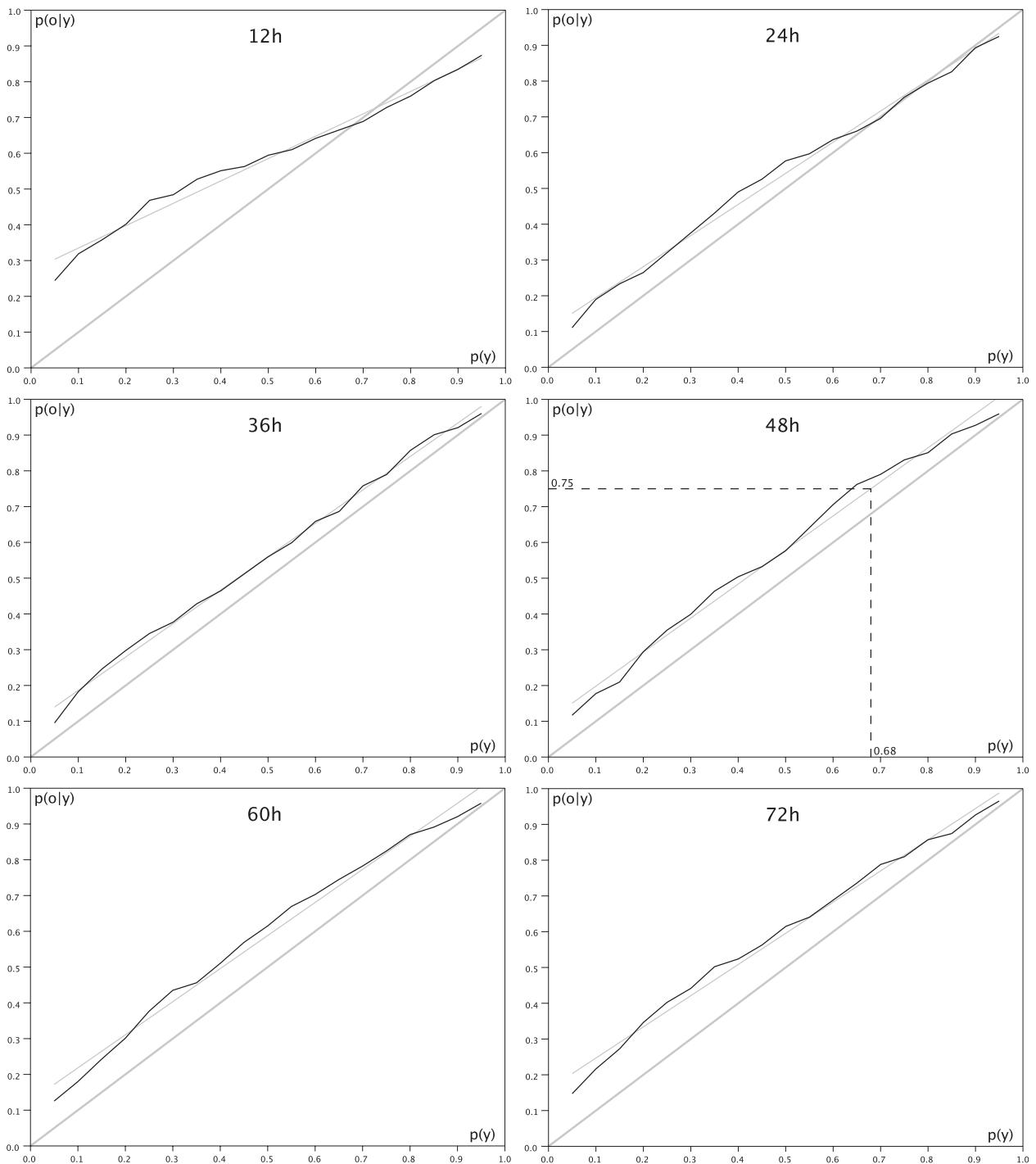


FIG. 2. Reliability diagrams for (top left) 12-h up to (bottom right) 72-h lead times showing the relationship between  $p(o|y)$  and  $p(y)$  (black curve), the linear regression used for calibration (gray thin line), and the diagonal calibrated curve (gray thick line). An example of calibration for 48 h is shown for a 75% probability; the radius is such that the circle contains 68% of the EPS members.

better than any individual numerical model and their consensus. Since the skill of the ensemble forecasts is sensitive to the position error of the circle center (Majumdar and Finocchio 2010), it is also important to ensure that the

forecast position used as the circle center has a good level of skill.

The probability of forecast position may be handled mathematically to compute objective probabilistic scores.

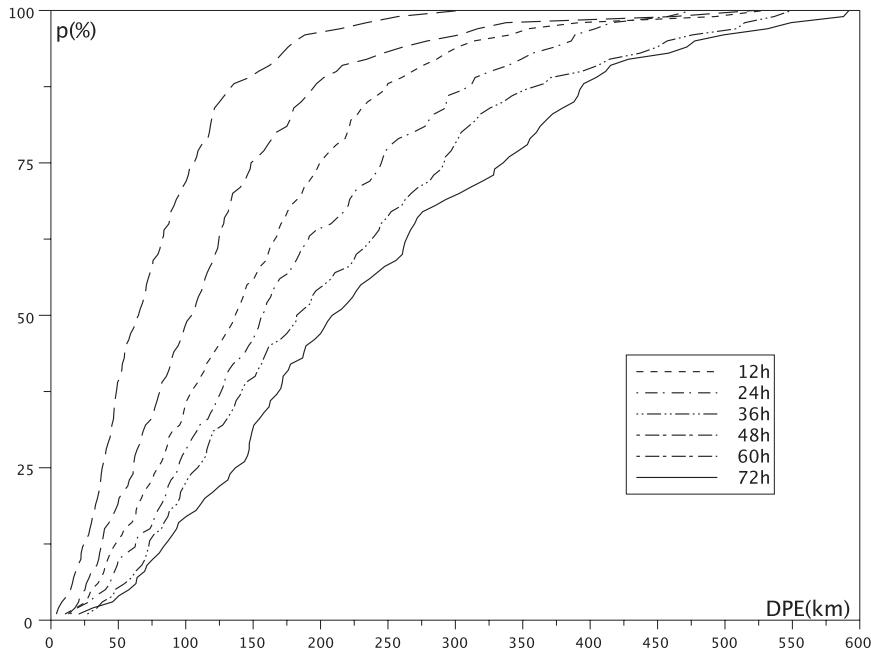


FIG. 3. Cumulated frequency of DPE representing the climatological distribution of DPE for 12 to 72 h lead times.

For a given circle of radius  $r$  around the RSMC forecast position, the probability that the forecast is inside this circle is noted  $p(y)$ , where  $y$  refers to the event where the forecast position is inside the circle. Verification is achieved a posteriori by looking at whether the observed best-track position is inside the circle, defined by the event  $o$ . For a large sample of forecasts and the associated circles of probability  $p(y)$ , the number of times the verifying analysis is inside this circle  $p(y)$  yields an observed conditional probability noted as  $p(o|y)$ .

The definition of the probability distributions  $p(y)$  and  $p(o|y)$  enables us to apply the existing verification methods of probabilistic forecasts (Wilks 2006). The following sections will deal with the verification of the probabilistic forecasts of TC positions.

### c. Reliability and calibration of the probabilistic forecasts

One of the first properties that is expected from a probabilistic forecast is that the forecast probability equals the observed probability; that is, when the forecast predicts an event with probability  $p\%$ , it actually happens  $p\%$  of the time. This property is measured by the reliability of the forecast (Wilks 2006). At every forecast lead time, the observed probability  $p(o|y)$  is computed for a given forecast probability  $p(y)$ , and the plots of  $p(o|y)$  as a function of  $p(y)$ , called reliability diagrams (Fig. 2), are used to diagnose the biases of the probabilistic forecast. A perfectly reliable forecast would imply the reliability

diagram follows the diagonal line  $p(o|y) = p(y)$ . When  $p(o|y) > p(y)$  [ $p(o|y) < p(y)$ ], the spread in the ensemble is smaller [larger] than that in the observations.

The reliability diagrams are in general quite close to the diagonal lines (Fig. 2). The most striking exception is for the 12-h lead time, for which there is a large bias for the small frequencies: the EPS dispersion for the smaller radii is too low. This flaw vanishes rapidly at longer lead times. For a 24-h lead time and after, the forecast probability is always lower than the observed probability. Overall, the curves are quite close to the diagonal lines, which suggests that the EPS dispersion is linked to the RSMC error distribution. The direct shifting of the EPS probability circles to the RSMC forecast position is thus a relevant method.

If the reliability of the forecast is not perfect, then a calibration should be applied. Since all of the reliability diagrams follow approximately a line (Fig. 2), a simple two-step calibration method may be applied to them:

- each curve on the reliability diagrams is approximated by a linear regression function  $p(o|y) = ap(y) + b$  (Fig. 2) and
- this function is used for calibrating the probability  $p(y)$  associated with radius  $r$ , that is, the calibrated probability associated with radius  $r$ , is taken to be  $p'(y) = [ap(y) + b]$ .

The operational implementation of such a calibration would be straightforward. For instance, for 48-h forecasts, a circle associated with 75% probability should be the smallest circle containing 68% of the EPS members (Fig. 2).

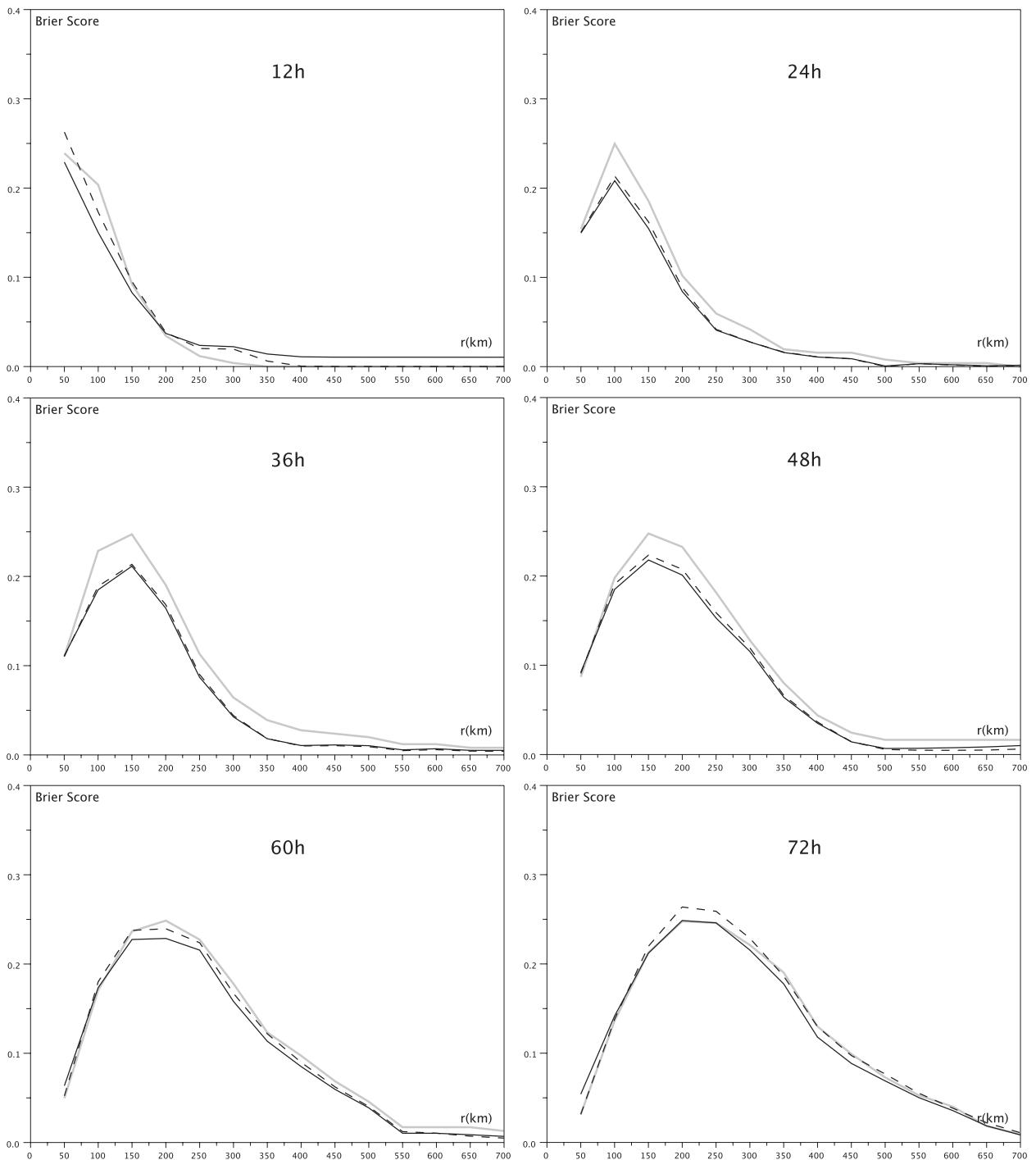


FIG. 4. Brier scores from (top left) 12-h up to (bottom right) 72-h lead times for the climatological forecast (thick gray curve), the uncalibrated EPS forecast (black dashed curve), and the calibrated EPS forecast (black solid curve).

#### d. Probabilistic scores

An objective evaluation of probabilistic forecasts may be achieved by probabilistic scores. Given a threshold radius  $r$ , the probability distributions  $p(y)$  and  $p(o|y)$  are

estimated for the ensemble forecast, with and without calibration. The Brier score measures the distance between the distribution of a probabilistic forecast and the distribution of the observations. Taking the  $K$  forecasts of the sample, it is expressed as

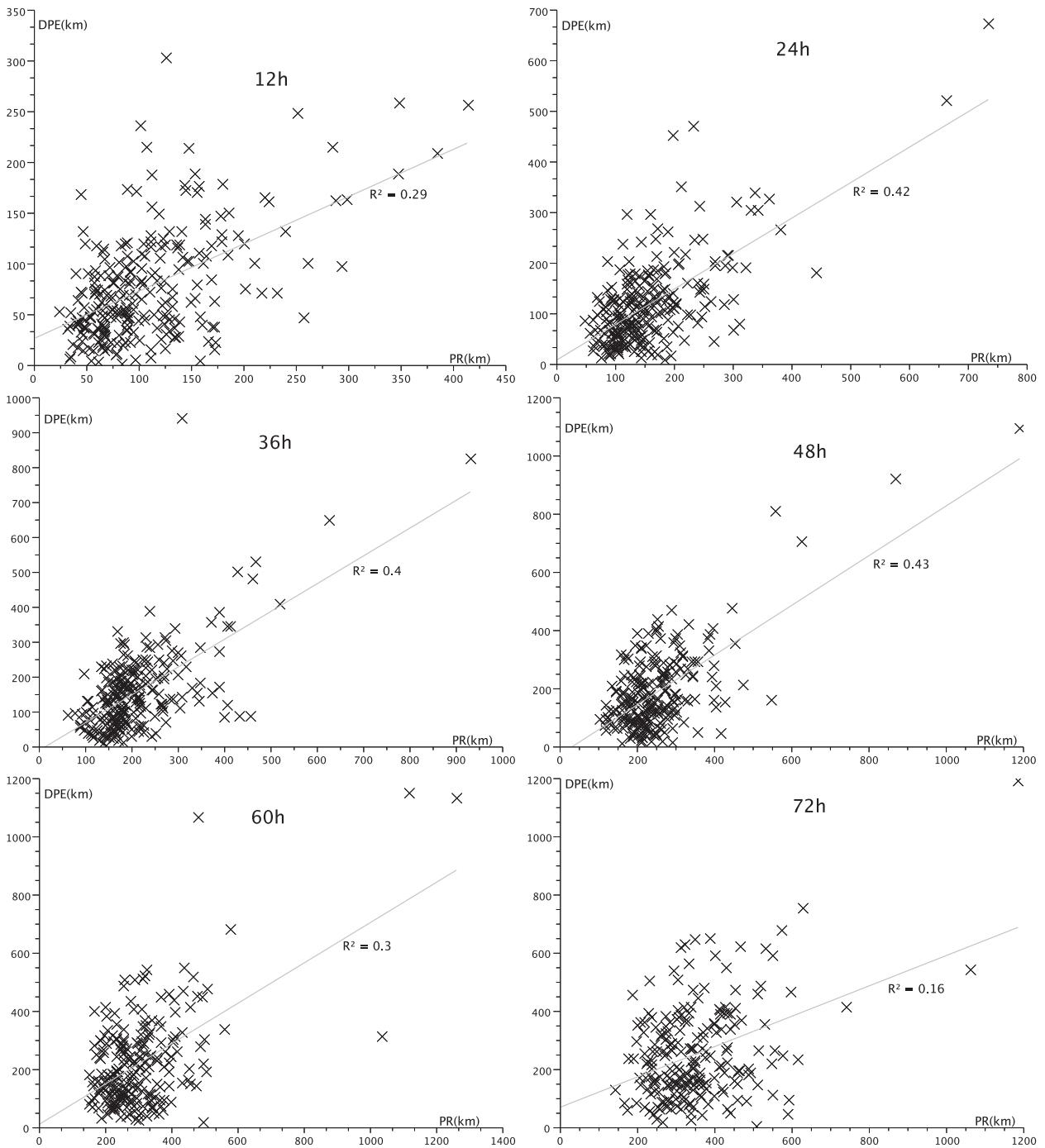


FIG. 5. Cloud points of the DPE as a function of the PR from (top left) 12-h up to (bottom right) 72-h lead times. The linear regressions (black line) and the correlation coefficients  $R^2$  are indicated.

$$BS(r) = \frac{1}{K} \sum_{k=1}^K [y_k(r) - o_k(r)]^2, \quad (1)$$

where  $y_k(r)$  is the forecast probability associated with radius  $r$  and  $o_k(r)$  is 1 (0) if the verifying best-track

analysis of the  $k$  forecast is inside (outside) the probability circle of radius  $r$ . For the purpose of comparison, the scores of the climatological forecasts are also computed, for which  $y_k(r)$  is the same for all  $k$  and it is deduced from the DPE distribution (Fig. 3) at each forecast lead time.

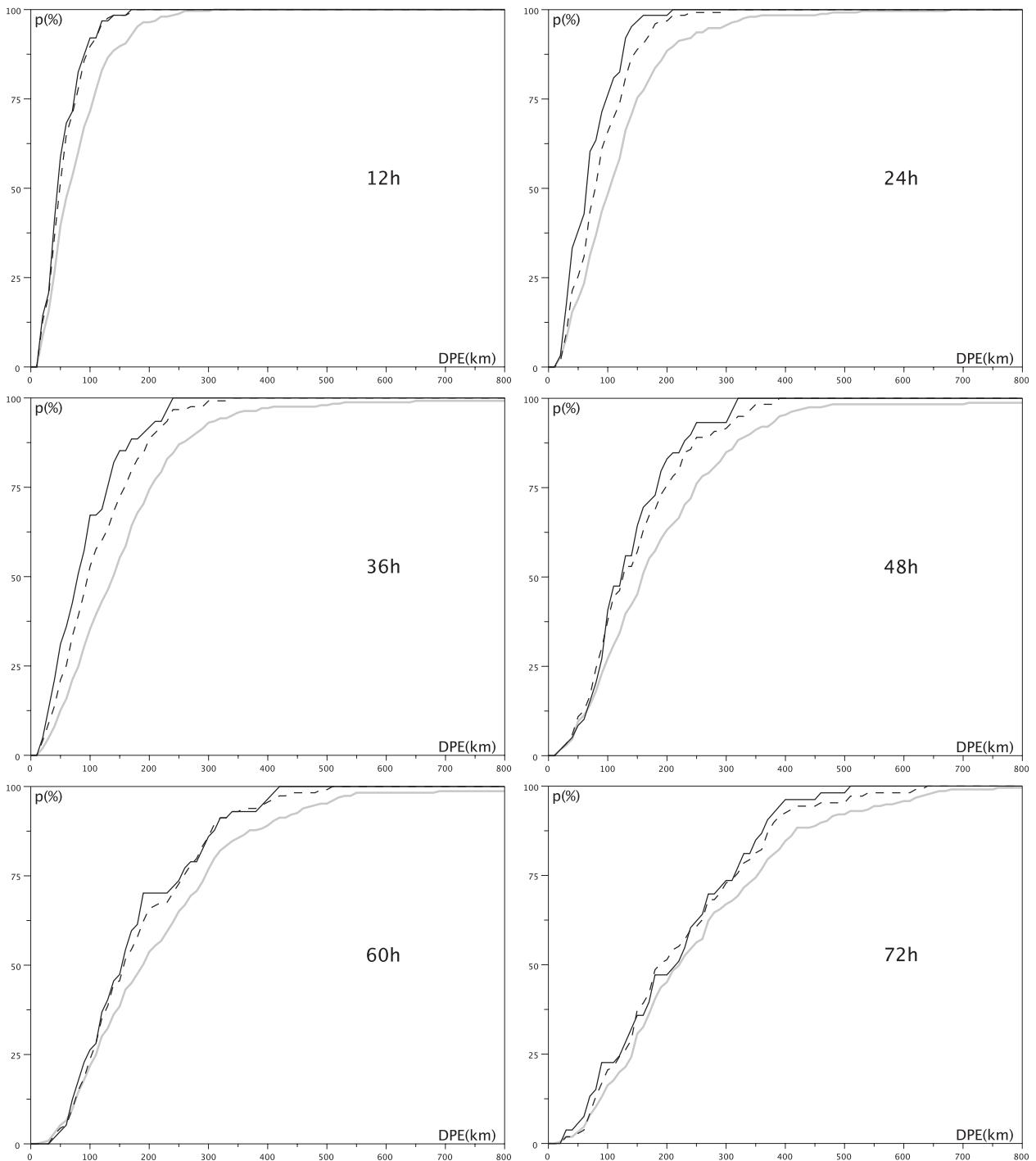


FIG. 6. Cumulated frequency of DPE for (top left) 12-h up to (bottom right) 72-h terms lead times conditioned on the predicted radius being lower than its first quartile [ $PR < Q_{PR}(0.25)$ , black solid line], being lower than its median [ $PR < Q_{PR}(0.5)$ , black dashed line], and unconditionally (gray thick line).

The Brier score encompasses some information about the reliability of the forecast (i.e., comparison of the observation and forecast distributions) and its resolution (i.e., how close each individual ensemble forecast is to

the observation). A Brier score approaching zero indicates a situation where the distance between the distribution of a probabilistic forecast and the distribution of the observations approaches zero, or where the probabilistic

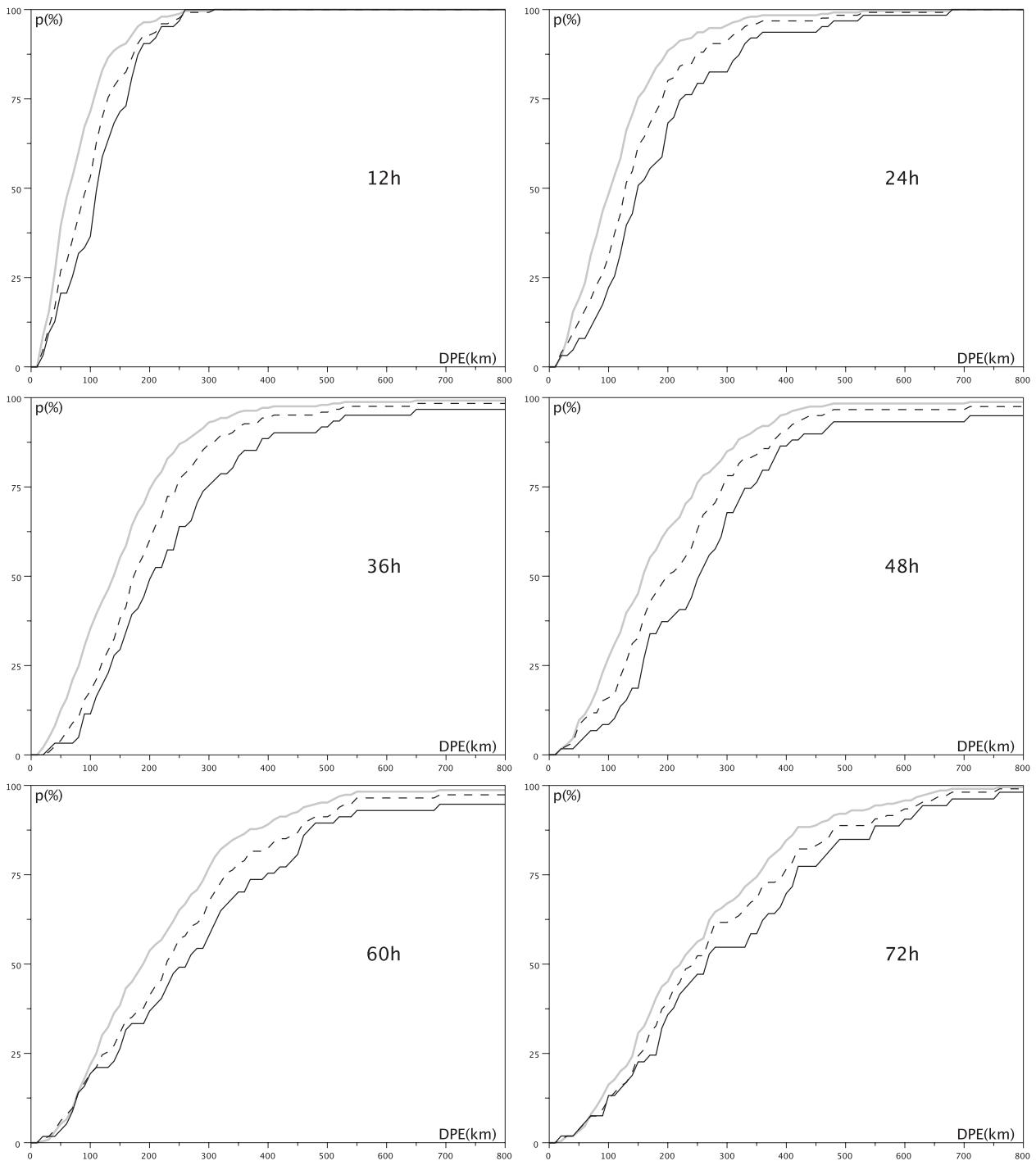


FIG. 7. As in Fig. 6, but for  $PR > Q_{PR}(0.75)$  (black solid line) and  $PR > Q_{PR}(0.5)$  (black dashed line).

forecast is perfectly reliable (it matches the observed probability) and it has a perfect resolution (it is able to adapt its spread to the situation).

For every lead time and for the probabilistic and the climatological forecasts, the Brier scores are plotted as a function of the radius  $r$  (Fig. 4). The calibrated probabilistic

forecast is better (e.g., lower Brier scores) than the climatology at almost every radii and at every range. The only exceptions are for the large radii at a lead time of 12 h and for small radii at lead times of 60 and 72 h. In these cases, the results may not be significant because the sample is small (their frequency is available in Fig. 3).

TABLE 1. Contingency table (%) for the detection of small errors [ $PR < Q_{PR}(0.5)$  with  $DPE < Q_{DPE}(0.5)$ ] for a 36-h lead time.

	$PR < Q_{PR}(0.5)$	$PR \geq Q_{PR}(0.5)$
$DPE < Q_{DPE}(0.5)$	34%	16%
$DPE \geq Q_{DPE}(0.5)$	16%	34%

When the uncalibrated forecast score is worse than the climatological forecast, the calibrated forecast is better or similar to the latter. The simple calibration method that is used is therefore relevant.

The probabilistic scores demonstrate that the EPS probabilistic forecast brings some valuable information about the distribution of the actual TC position versus climatology. The issues that will be addressed next are

- how to represent the uncertainty around the forecast position and
- how to measure the skill of this method at estimating the size of the DPE.

### 3. Construction and validation of uncertainty circles

#### a. A definition of uncertainty circles

To represent graphically the uncertainty of the forecast position, the probability will be set to a fixed value, similarly to the existing climatological methods used by operational centers. An uncertainty circle is then given by a single predicted radius (PR), associated with this calibrated probability, centered at the RSMC forecast position. A probability value of 75% has been chosen among other possibilities (e.g., 50% or 67%). One of the reasons for choosing this 75% value is that the EPS seems to be rather well calibrated at this point (Fig. 2). In addition, as suggested by the results of Majumdar and Finocchio (2010), only slight statistical differences are expected between the 50%, 67%, and 75% probability circles. Since the purpose here is to measure the uncertainty, the relevant information is the size of the circle rather than the probability value, provided it is fixed.

The validation processes of the uncertainty circles and of the PR should employ different methods from the above-mentioned probabilistic verification. By definition, what is expected from uncertainty circles is a measure of the uncertainty, that is when the PR is large (small), the error should be large (small). Like previous studies (Yamaguchi et al. 2009; Majumdar and Finocchio 2010), the link between the DPE and the PR will be assessed on a statistical basis to more precisely measure the skill at detecting large and small forecast errors.

TABLE 2. As in Table 1, but for the detection of very small errors [ $PR < Q_{PR}(0.25)$  with  $DPE < Q_{DPE}(0.25)$ ].

	$PR < Q_{PR}(0.25)$	$PR \geq Q_{PR}(0.25)$
$DPE < Q_{DPE}(0.25)$	13%	12%
$DPE \geq Q_{DPE}(0.25)$	12%	63%

#### b. Spread–skill relationship

A first step is to relate the PR to the DPE (Fig. 5) in order to find a relationship between the spread (PR) and skill (DPE). The cloud points are rather spread out, but some correlation may be found. The highest correlation values ( $R^2 \simeq 0.4$ ) exist between the 24- and 48-h lead times, and then correlation vanishes after 60 h. This result suggests a moderate relationship between the PR and the DPE until 60 h.

A more informative diagnostic is to determine whether predicting a large (or small) radius has an impact on the distribution of the forecast error. A measure for this is the conditional probability distribution of DPE for a PR being lower or higher than a fixed value. At every forecast lead time, such conditional distributions are compared to the unconditional DPE distributions. Let  $Q_f(x)$  be the quantile  $x$  of the distribution  $f$  ( $f$  is DPE or PR). Figure 6 shows the cumulated frequencies of DPE conditioned on  $PR < Q_{PR}(0.5)$  and to  $PR < Q_{PR}(0.25)$ , as well as the unconditional one. For every lead time, the conditional distributions appear to the left of the unconditional ones, which means that when the PR is small, the DPE actually tends to be small. The DPE conditioned on  $PR < Q_{PR}(0.25)$  generally indicates a smaller radii than the DPE conditioned on  $PR < Q_{PR}(0.5)$ , which means that there is some skill at discriminating between a small PR and a very small one. Similar conclusions apply to the cumulated frequencies of DPE conditioned on  $PR > Q_{PR}(0.5)$  and  $PR > Q_{PR}(0.75)$  (Fig. 7): there is some skill at detecting large DPE when the PR is large. In addition, large radii versus very large radii seem to be more effectively discriminated than small versus very small radii: the curves between the conditional distributions tend to be more separated for large radii (Fig. 7) than for small radii (Fig. 6).

These results confirm that the spread of the EPS has some skill at detecting the forecast uncertainty of the RSMC forecast. Majumdar and Finocchio (2010) suggested that the EPS spread could be useful in discriminating between large and small forecast errors of the

TABLE 3. As in Table 1, but for the detection of very large errors [ $PR > Q_{PR}(0.75)$  with  $DPE > Q_{DPE}(0.75)$ ].

	$PR > Q_{PR}(0.75)$	$PR \leq Q_{PR}(0.75)$
$DPE > Q_{DPE}(0.75)$	12%	13%
$DPE \leq Q_{DPE}(0.75)$	13%	62%

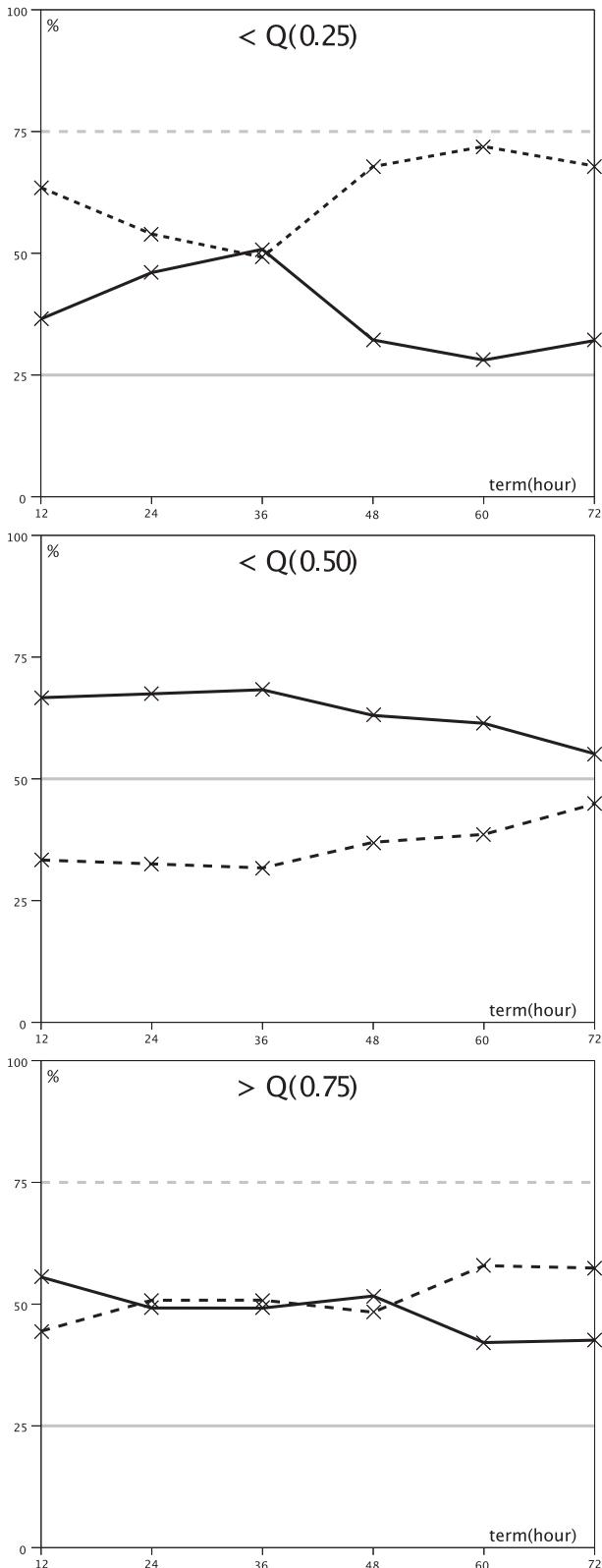


FIG. 8. POD (solid lines) and FAR (dashed lines) as a function of the forecast lead time for the EPS forecast (black lines) and a random forecast (gray lines) with no skill: (top)  $<Q(0.25)$ , (middle)  $<Q(0.5)$ , and (bottom)  $>Q(0.75)$ .

ensemble mean. The present results prove that the spread of the EPS is also skillful at detecting small and large errors in the RSMC forecast.

### c. Detection rate of the amplitude of DPE

The skill of the present method at detecting large and small errors may be quantitatively estimated. The following decision rule is applied to the size of the uncertainty circles: if the PR is lower than a given quantile, then it is decided that the forecast DPE will be lower than the corresponding quantile. The decision rule and the associated contingency tables are defined for three cases:  $PR < Q_{PR}(0.5)$  with  $DPE < Q_{DPE}(0.5)$  (detection of small error; Table 1),  $PR < Q_{PR}(0.25)$  with  $DPE < Q_{DPE}(0.25)$  (detection of very small error; Table 2), and  $PR > Q_{PR}(0.75)$  with  $DPE > Q_{DPE}(0.75)$  (detection of very large error; Table 3). They are shown for 36-h lead time, but they have been computed for all lead times. The contingency table for the detection of large errors [ $PR > Q_{PR}(0.5)$  with  $DPE > Q_{DPE}(0.5)$ ] is also given by Table 1.

These tables lead to input data for computing some probability scores, namely, the probability of detection (POD) and the false alarm rate (FAR) associated with the detection of large and small error values. For instance, the POD associated with the event  $<Q(0.25)$  measures how often the radius is predicted to be very small [ $PR < Q_{PR}(0.25)$ ] when the error is verified to be very small [ $DPE < Q_{DPE}(0.25)$ ]. Similarly, the FAR associated with the event  $<Q(0.25)$  measures how often the DPE is not very small [above  $Q_{DPE}(0.25)$ ] when the PR is very small [below  $Q_{PR}(0.25)$ ]. Among two forecast systems, the better one has the higher POD and the lower FAR. To demonstrate the skill of the EPS uncertainty circles, the POD and FAR are compared to a no-skill forecast. Since the climatological circles are case independent, they do not give any information about the detection of the size of the DPE. Therefore, the forecast skill is obtained by comparison with a random forecast, obtained by picking up a PR value among the climatological DPE distribution (Fig. 3). The POD and FAR values for the events  $<Q_{DPE}$  and  $>Q_{DPE}$  of such a random forecast may be easily computed. A skillful forecast should have a higher POD and a lower FAR than this random forecast.

The POD and FAR associated with the events  $<Q(0.5)$ ,  $<Q(0.25)$ , and  $>Q(0.75)$  for the ensemble method are plotted in Fig. 8, together with the scores of a random forecast. The scores for the event  $>Q(0.5)$  are the same as the ones of  $<Q(0.5)$ . To illustrate the meaning of the scores, take the event  $<Q(0.5)$  for a 36-h lead time. The POD value (65%, according to Fig. 8) implies that 65% of the small error was actually detected. The FAR (35%, according to Fig. 8) implies that 35% of the small error

forecasts were not observed. These scores are better than a random forecast ( $POD = 50\%$ ,  $FAR = 50\%$ ). Thus, the  $POD$  and  $FAR$  scores emphasize the skill of the uncertainty circles at estimating the DPE. A similar interpretation applies to the very large [ $>Q(0.75)$ ] and the very small [ $<Q(0.25)$ ] circles.

The scores are always better for the ensemble method than for the random forecast, but they are sometimes very close. The discrimination of DPE apart from its median value  $Q(0.5)$  by the PR is valuable for all lead times, but it is close to the random value at 72 h. The  $POD$  and  $FAR$  associated with the detection of very large DPE [ $>Q(0.75)$ ] are always around 50%. The scores for the detection of very small DPE are better than the random forecast until 36 h, but then they jump to values that are close to the random forecast. Therefore, the uncertainty circles are valuable in detecting a large error at least until a 72 h lead time; and they are able to detect small uncertainty until a 48 h lead time.

#### 4. Conclusions

A method of construction of uncertainty circles around an official TC track forecast, using the ECMWF EPS, has been described and assessed. The circles are calibrated by using a simple scheme. A validation in two steps is performed, first by the computation of scores on the probability distribution of forecast position and second by measuring the skill at forecasting whether the position error will be large or small. The main conclusions of this work are that

- the Brier scores associated with the calibrated probability distribution given by the EPS are better than the ones from a climatological forecast, and calibration is useful and sometimes necessary to obtain valuable scores;
- the uncertainty circles, defined by the radius of a calibrated probability 75%, are well correlated with the DPE until 60 h;
- the conditional distributions of DPE to the radii of uncertainty circles demonstrate that the method is skillful at discriminating between large and small forecast errors; and
- the size of the PR is able to indicate a small or a large DPE at least until 72 h, but this skill becomes small from 48 h and beyond for the smallest values of DPE.

A result of particular interest is that the EPS is not only skillful at measuring the EPS ensemble mean error (Majumdar and Finocchio 2010), but also the RSMC position error. This suggests that the EPS spread is able to measure an intrinsic part of the forecast uncertainty, with acknowledgment that the RSMC track forecasts and the

EPS forecasts are not totally independent. The good results presented in this article will lead to the operational implementation in the near future of the uncertainty circles by RSMC La Réunion based on the ECMWF EPS. Track forecasts with uncertainty cones will also be extended until the 5-day lead time.

Although the EPS of ECMWF is dedicated to medium-range forecasts and has generally good scores for lead times of several days, the detection of small TC forecast position errors has little skill after a 48-h lead time. It is still an open question whether this flaw is due to the ECMWF EPS itself, to the method of building circles, or to an intrinsic property of TC track predictability. The comparison of different ensemble systems would provide some information to answer this question. The method proposed in the present article, particularly the Brier scores and the detection scores of large and small radii, draw a framework that we can use to quantify and to compare different probabilistic forecasting systems of TC position. The impacts of the future evolutions of the EPS on TC track forecasts may be measured, and EPS forecasts issued from different centers, including those from the recently improved system of Météo-France, may be compared. It would also be interesting to assess the value of a multi-ensemble system such as The Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble (TIGGE; Bougeault et al. 2010).

The method builds uncertainty circles, but other shapes may be more suitable and could result in better scores. For instance, an ellipsis with an axis along the track would discriminate among the along- and the cross-track errors. Another issue to consider is when ensemble track forecasts splits into several directions. In such a case, a single RSMC track forecast described by an uncertainty circle may not be relevant. The assessment of different uncertainty shapes would require more refined methods.

*Acknowledgments.* The support by ECMWF for providing skillful data for TC forecasting is gratefully acknowledged by RSMC La Réunion. This work was successfully initiated with the help of Samuel Westrelin and Olivier Cabanes. We would also like to thank Nicole Girardot and Sébastien Langlade for their suggestions, and the two reviewers for their relevant comments, which helped to improve the manuscript.

#### REFERENCES

- Avila, L. A., P. Caroff, J. Callaghan, J. Franklin, and M. DeMaria, 2006: Track forecasts. *Proc. Sixth Int. Workshop on Tropical Cyclones*, San José, Costa Rica, WMO, 12–18.
- Barkmeijer, J., R. Buizza, T. N. Palmer, K. Puri, and J.-F. Mahfouf, 2001: Tropical singular vectors computed with linearized diabatic physics. *Quart. J. Roy. Meteor. Soc.*, **127**, 685–708.

- Bougeault, P., and Coauthors, 2010: The THORPEX Interactive Grand Global Ensemble. *Bull. Amer. Meteor. Soc.*, **91**, 1059–1072.
- Bourke, W., R. Buizza, and M. Naughton, 2005: Performance of the ECMWF and the BoM ensemble systems in the Southern Hemisphere. *Mon. Wea. Rev.*, **132**, 2338–2357.
- Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076–1097.
- Chan, J. C., 2005: The physics of tropical cyclone motion. *Annu. Rev. Fluid Mech.*, **37**, 99–128.
- , and W. M. Gray, 1982: Tropical cyclone movement and surrounding flow relationships. *Mon. Wea. Rev.*, **110**, 1354–1374.
- , F. M. Ko, and Y. M. Lei, 2002: Relationship between potential vorticity tendency and tropical cyclone motion. *J. Atmos. Sci.*, **59**, 1317–1336.
- Emanuel, K., 2003: Tropical cyclones. *Annu. Rev. Earth Planet. Sci.*, **31**, 75–104.
- Fraedrich, K., and M. Leslie, 1989: Estimates of cyclone track predictability. I: Tropical cyclones in the Australian region. *Quart. J. Roy. Meteor. Soc.*, **115**, 79–92.
- Goerss, J. S., 2000: Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Wea. Rev.*, **128**, 1187–1193.
- Majumdar, S. J., and P. M. Finocchio, 2010: On the ability of global ensemble prediction systems to predict tropical cyclone track probabilities. *Wea. Forecasting*, **25**, 659–680.
- Plu, M., 2011: A new assessment of the predictability of tropical cyclone tracks. *Mon. Wea. Rev.*, in press.
- Puri, K., J. Barkmeijer, and T. N. Palmer, 2001: Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. *Quart. J. Roy. Meteor. Soc.*, **127**, 709–734.
- Rossby, C. G., 1948: On displacements and intensity changes of atmospheric vortices. *J. Mar. Res.*, **7**, 175–187.
- Shapiro, L. J., 1992: Hurricane vortex motion and evolution in a three-layer model. *J. Atmos. Sci.*, **49**, 140–154.
- , and J. L. Franklin, 1998: Potential vorticity asymmetries and tropical cyclone motion. *Mon. Wea. Rev.*, **127**, 124–131.
- van der Grijn, G., 2002: Tropical cyclone forecasting at ECMWF: New products and validation. ECMWF Tech. Memo. 386, 13 pp. [Available online at [http://www.ecmwf.int/publications/library/ecpublications/\\_pdf/tm/301-400/tm386.pdf](http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/301-400/tm386.pdf).]
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 627 pp.
- Yamaguchi, M., R. Sakai, M. Kyoda, T. Komori, and T. Kadowaki, 2009: Typhoon ensemble prediction system developed at the Japan Meteorological Agency. *Mon. Wea. Rev.*, **137**, 2592–2604.