



HAL
open science

Probabilistic Forecasts of Tropical Cyclone Tracks and Intensities in the Southwest Indian Ocean Basin

François Bonnardot, Hubert Quetelard, Guillaume Jumaux, Marie-Dominique Leroux, Miloud Bessafi

► **To cite this version:**

François Bonnardot, Hubert Quetelard, Guillaume Jumaux, Marie-Dominique Leroux, Miloud Bessafi. Probabilistic Forecasts of Tropical Cyclone Tracks and Intensities in the Southwest Indian Ocean Basin. Quarterly Journal of the Royal Meteorological Society, 2018, 145 (719), pp.675-686. 10.1002/qj.3459 . hal-01968096

HAL Id: hal-01968096

<https://hal.science/hal-01968096>

Submitted on 18 Jul 2019

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.

Probabilistic Forecasts of Tropical Cyclone Tracks and Intensities in the Southwest Indian Ocean Basin

François Bonnardot, Hubert Quetelard, Guillaume Jumaux, Marie-Dominique Leroux, Miloud Bessafi

► **To cite this version:**

François Bonnardot, Hubert Quetelard, Guillaume Jumaux, Marie-Dominique Leroux, Miloud Bessafi. Probabilistic Forecasts of Tropical Cyclone Tracks and Intensities in the Southwest Indian Ocean Basin. Quarterly Journal of the Royal Meteorological Society, Wiley, 2018, 145 (719), pp.675-686. 10.1002/qj.3459 . hal-01968096

HAL Id: hal-01968096


<https://hal.archives-ouvertes.fr/hal-01968096>

Submitted on 18 Jul 2019

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.

Probabilistic forecasts of tropical cyclone tracks and intensities in the southwest Indian Ocean basin

Francois Bonnardot¹  | Hubert Quetelard¹ | Guillaume Jumaux¹ | Marie-Dominique Leroux¹ | Miloud Bessafi²

¹ Météo-France, Direction Interrégionale pour l'Océan Indien, Saint-Denis de La Réunion, France

²Laboratoire d'Electronique, d'Énergétique et des Procédés, Université de la Réunion, Réunion, France

Correspondence

Francois Bonnardot, Service Etudes et Climatologie, Météo-France DIROI, 50 boulevard du chaudron, 97490 Sainte-Clotilde, La Réunion, France.

Email: francois.bonnardot@meteo.fr

Funding information

Agence Nationale de la Recherche, ANR-14-CB03-0013 SPICY. French National Research Agency (ANR), ANR-14-CB03-0013 SPICY.

Under the responsibility of the Regional Specialized Meteorological Centre (RSMC) of La Réunion, the southwest Indian Ocean (SWIO) has tropical cyclone activity close to that of the North Atlantic. Like most territories of the SWIO basin, La Réunion island is highly vulnerable to cyclone-induced hazards and the potential impact of nearby storms is closely related to their track and intensity evolution. Although storm track and intensity forecasts have been steadily improving in the last decades, a great amount of uncertainty remains. Operational centres have therefore developed probabilistic products such as uncertainty cones for the prediction of storm track and intensity over the different cyclone basins. Unfortunately, the cone approach does not fully match the end-user needs for efficient decision-making. This article provides a method to generate alternate probabilistic scenarios of tropical system track and intensity forecasts around an official forecast. The method has been calibrated and evaluated to answer the needs of the Système de Prévision des Inondations en contexte Cyclonique (SPICY) project that aimed to explore new probabilistic forecast products for tropical system induced hazards such as coastal inundations. A hybrid method has been developed to benefit from both climatological and dynamical existing approaches. A first set of climatology-built scenarios is generated using the statistical distribution of RSMC La Réunion forecast errors. This initial set is then modulated using real-time information provided by the ensemble prediction system of the European Centre for Medium-range Weather Forecasts (ECMWF). The final product is a set of about 20 scenarios that are built around the official deterministic forecast with some associated probabilities. Performance scores demonstrate the efficiency of the method against other evaluated systems especially in the first 2 days of the forecast. The reasonable number of defined scenarios is cost efficient and makes it possible to perform further impact-oriented applications such as wave and storm surge simulations.

KEYWORDS

ensemble prediction, probabilistic forecasts, southwest Indian Ocean, tropical cyclones

1 | INTRODUCTION

Being able to forecast a tropical system track and intensity a few days ahead is a main issue for territories subject to storm hazards and their induced impacts. Islands such as La Réunion are regularly impacted by storms during the southwest

Indian Ocean (SWIO) austral summer. Due to the small size and sharp topography of the island, the level and location of the impacts are highly dependent on the final track and intensity of the storm passing nearby. Although numerical models have made great progress in the last decades in forecasting the tracks and intensities of such phenomena, there is

still a great deal of uncertainty associated with deterministic forecasts (e.g. Plu, 2011; DeMaria *et al.*, 2014). Small forecast errors may lead to inadequate decisions and consequently to material or human losses.

The Regional Specialized Meteorological Centre (RSMC) of La Réunion issues track forecasts and warnings for countries in the SWIO area up to a 5-day lead time whenever a tropical storm (TS) or tropical cyclone (TC) affects this region. In this particular basin, a TS is defined as a tropical system with estimated maximum 10 min average wind speed (V_{max}) between 34 and 63 knots (kn, 0.514 m/s). A TC is defined as a tropical system with V_{max} greater than 64 kn. In 2011, RSMC La Réunion developed a technique to evaluate and display track forecast uncertainty (Dupont *et al.*, 2011). Uncertainty circles are built around the official track forecast and calculated using the Ensemble Prediction System (EPS) of the European Centre for Medium-range Weather Forecasts (ECMWF). The cone of uncertainty comprises circles associated with each forecast lead time and represents the dispersion of the EPS ensemble. The greater the uncertainty, the wider the cone. This cone thus reflects the degree of confidence that forecasters can have in their track forecasts.

The uncertainty cone approach has proven valuable during the last few cyclone seasons. However, it does not fully match the end-user concerns and needs for detailed information on the potential impacts of winds, precipitations and oceanic phenomena such as swell and surge, with their associated probabilities. Hence the idea of designing a method that would generate alternate scenarios to the RSMC official deterministic forecast and that would quantify their associated probabilities to be able to subsequently derive probabilistic forecasts of cyclone-induced hazards (devastating winds, torrential rain, marine and river inundations) as described for example in Davis *et al.* (2010). The Système de Prévision des Inondations en context Cyclonique (SPICy) project (<http://spicy.brgm.fr/en>) led by the French Geological Survey (BRGM), in which Météo-France was involved, was a good opportunity to explore a new probabilistic approach for TS or TC track and intensity forecasts. This project aims at improving the information provided to authorities and users when a tropical system threatens La Réunion Island in particular but can be extended to other regions if it proves to provide effective innovative products for decision-makers. One component of the project is to provide probabilistic track and intensity forecasts to force atmospheric, wave and surge models and derive probabilistic forecasts of oceanic and coastal conditions.

Since the 2006–2007 hurricane season, the National Hurricane Center (NHC) has implemented a probabilistic model that gives spatial estimation of the probability of winds exceeding certain thresholds such as 34, 50 and 64 kn on the expected passage of a tropical system evolving in the Atlantic or North Pacific Oceans (DeMaria *et al.*, 2009). A Monte-Carlo method is used to generate 1,000 alternative forecasts around the *a-priori*-considered-the-most-likely

official forecast by randomly sampling from the NHC track and intensity forecast error 5-year distribution computed from previous cyclone seasons. Track errors are broken down into “Along-Track” (AT) and “Cross-Track” (CT) errors to facilitate computations. Each generated position is calculated by vectorially adding AT and CT errors to the official expected position while the random intensity error is added to the official intensity forecast. The originality of the method is that it takes into account the serial correlation of both track and intensity errors between the various lead times using linear regressions. The authors indeed demonstrate the correlation between TC track and intensity forecasts for all 12 h periods from 12 to 120 h. This Monte-Carlo method was initially considered within the SPICy project but soon had to be abandoned due to the amount of computing resources required to run 1,000 simulations to generate associated surface wind, ocean wave and surge. Alternative methods have been developed for TC track forecasts. Qi *et al.* (2014) propose a selective ensemble mean technique based on various ensemble prediction systems while Zhang and Yu (2017) explore a probabilistic forecast scheme based on the selective consensus of ensemble prediction systems. Jun *et al.* (2017) recently introduced an unequally weighted technique to generate multi-model ensemble forecasts for western North Pacific TC tracks. For intensity forecasts, probabilistic approaches are explored using, for example, a climatology-based forecast scheme (Chen *et al.*, 2016) or an analogue ensemble technique (Alessandrini *et al.*, 2018).

To meet the project goal, a method was built to generate a reasonable number of scenarios that modulates both track and intensity with a distinct probability assigned to each scenario. The limited number of scenarios fulfils operational computing time constraints while covering a large spectrum of possibilities.

Section 2 provides statistics of RSMC La Réunion TS and TC official forecast errors with respect to numerical models from the ECMWF. Section 3 describes the scenario generation method while section 4 analyses its performance. Section 5 is a summary of the article.

2 | STATISTICS ON RSMC LA RÉUNION OFFICIAL TS AND TC FORECAST ERRORS (TRACK AND INTENSITY)

The RSMC official (deterministic) forecast is chosen as the starting point of our ensemble generation process for a better consistency of the suite of forecasting products delivered in terms of track, intensity, and associated impacts. To validate this fundamental choice, statistics are computed over five recent cyclone seasons from 2011/2012 to 2015/2016 and over the RSMC area of responsibility (40°S–0°S/30°E–90°E) to evaluate the RSMC official forecast performance. For the purpose of the SPICy project, we focus on lead times 0 h (analysis), 12, 24, 48, 60 and 72 h.

TABLE 1 Intensity and position forecast errors at 0, 12, 24, 36, 48, 60 and 72 h lead times for three forecasting systems (RSMC, IFS and EPSm) during five TC seasons (2011/2012 to 2015/2016)

Lead time (h)		0	12	24	36	48	60	72	
DPE (km)	RSMC	28	63	96	130	168	206	253	
	IFS	53	74	100	133	168	207	249	
	EPSm	54	80	111	146	192	239	289	
DP-RMSE (km)	RSMC	39	76	119	161	206	253	312	
	IFS	65	93	125	164	202	249	303	
	EPSm	69	104	142	185	238	298	361	
CTE (km)	RSMC	0	-8	-15	-19	-22	-21	-19	
	IFS	-15	-18	-20	-18	-21	-25	-24	
	EPSm	-15	-20	-24	-26	-34	-37	-40	
ATE (km)	RSMC	-2	-14	-23	-34	-46	-61	-83	
	IFS	3	-5	-17	-32	-44	-64	-93	
	EPSm	-3	-17	-36	-54	-75	-93	-127	
Np	RSMC	1,183	1,178	1,166	1,134	1,083	1,014	941	
	IFS	581	569	543	505	473	432	396	
	EPSm	587	576	554	525	488	456	422	
DVmax (kn)	<i>Observed Vmax >=48 kn lead time</i>	RSMC	0	-2	-4	-4	-4	-3	-1
		IFS	-8	-7	-6	-6	-3	0	0
		EPSm	-14	-14	-15	-15	-14	-13	-14
Vmax-RMSE (kn)	RSMC	6	13	18	21	23	23	22	
	IFS	23	24	24	24	22	23	22	
	EPSm	25	26	27	27	26	26	26	
Nv	RSMC	498	508	509	489	450	396	347	
	IFS	249	249	231	209	180	157	130	
	EPSm	244	244	228	206	178	155	129	

Direct position mean (DPE) and root-mean-square (DP-RMSE) errors (km), as well as cross-track error (CTE, km), along-track error (ATE, km), maximum wind error (DVmax, kn) and maximum wind root-mean-square error (Vmax-RMSE, kn). The best score out of the three evaluated systems is highlighted in boldface. For maximum wind errors, cases when the observed wind is lower than 48 kn at the corresponding lead time are not considered (minimum stage of severe tropical storm). For each evaluated system, the size of the considered sample is indicated (Np for position and Nv for Vmax).

The “official” RSMC track is a consensus forecast resulting from a manually analysed forecasting process, which relies on output from available deterministic numerical models and ensemble systems provided by various modelling centres (e.g. ECMWF, Météo-France, the Met Office or National Centers for Environmental Prediction (NCEP)), as well as knowledge on individual model historical performances. Table 1 shows statistical scores of RSMC official forecast (RSMC), ECMWF deterministic model forecast (IFS) and ECMWF Ensemble Prediction System mean scenario from individual members (EPSm). Scores are computed at various lead times using RSMC La Réunion best-track (BT) dataset as a reference for observed positions and intensities. Leroux *et al.* (2018) provides detailed information on the techniques used at RSMC La Réunion to generate BT data (which eventually feed International Best Track Archive for Climate Stewardship (IBTrACS) global dataset). Data from the ECMWF models were obtained through the THORPEX Interactive Grand Global Ensemble (TIGGE)-portal (Swinbank *et al.*, 2016). Track forecasts are evaluated by the direct position error (DPE), cross-track error (CTE) and along-track error (ATE). TC intensities are defined by the maximum wind speed Vmax.

Vmax from BT data is an estimated maximum 10 min average wind speed whereas for the IFS and EPS models, Vmax is directly provided by ECMWF’s storm tracker. Intensity errors are obtained by subtracting the BT intensity from the predicted intensity Vmax.

Scores indicate that RSMC operational forecasts perform better than the IFS and EPS numerical systems at short term (0–36 h) both in position and intensity, with a mean DPE of 96 km at 24 h lead time (Table 1). However, at longer ranges (48–72 h), the RSMC and IFS systems perform equally with no significant intensity bias and a maximum wind root-mean-square error around 20 kn. Note that all forecasting systems generate an overall negative CTE bias that reaches some 20 km beyond 24 h lead time, highlighting a more or less systematic deviation of the forecast to the right of the observed track.

Table 1 also shows that scores of the mean EPS (EPSm) are worse than those of the RSMC and IFS systems for all parameters and all lead times. For the maximum wind parameter, a systematic negative bias is visible which is not surprising considering the coarser resolution of this ensemble system with respect to that of the IFS.

In conclusion, although the performance of RSMC predictions have not been compared with those of other numerical models but considering that ECMWF models are currently among the most accurate for TC forecasting over the southwest Indian Ocean basin (Yamaguchi *et al.*, 2015), it is assumed that the appraised operational forecast made by RSMC La Réunion is overall more skilful than numerical models taken individually (especially in the first 36 h of the forecast) and can thus be considered as a robust reference for our method which aims to generate alternate scenarios (from the official one) with associated probabilities.

3 | METHOD

As mentioned in section 2, the starting point of our method is the official forecast provided by RSMC La Réunion every 6 h whenever a TC is monitored within the RSMC area of responsibility.

3.1 | General overview of the method

Initially, two different existing techniques were considered. The first one, hereinafter referred to as “CLIM”, only considers the climatological errors (statistical ATE-CTE-DVmax error distribution over 5 years). The second one, hereinafter referred to as “EPS” considers a purely dynamical approach and uses the ECMWF ensemble system. In fact, further investigations showed that a hybrid approach (called “CLIM-EPS”) is generally preferable to the two individual ones.

The hybrid method is mainly a two-step approach. The first step consists in generating a first set of scenarios using the climatological technique (CLIM); the second step consists in modulating this first set of scenarios by integrating some information on the dispersion using the EPS system. In that manner, relevant real-time information is added to the predictability of the TC event. Basically, the original “climatological” probabilities applied to each scenario (issued from the statistical error distribution in the first step) are weighted by the EPS members’ distribution.

Since the final goal is to provide information on local impacts (waves, storm surge, and wind), the whole method is calibrated to end up with some 15–30 TC scenarios in order to be able to perform further applications on each scenario within reasonable computing time response to match operational constraints.

3.2 | Step 1: Build clusters from climatological errors

3.2.1 | Input data

To build our climatological scenarios, we use a sample of ATE-CTE-DVmax triplets of errors computed over the five TC seasons from 2011/2012 to 2015/2016. In contrast to section 2, errors are now obtained by subtracting the RSMC

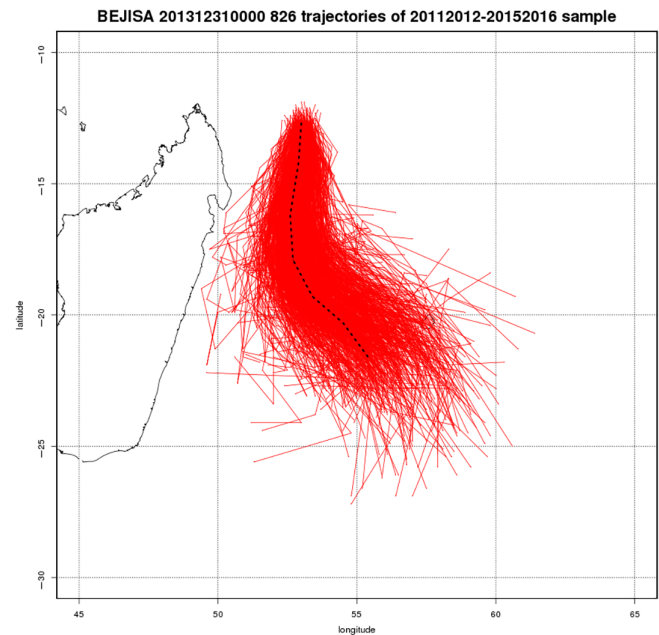


FIGURE 1 The 826 reconstructed tracks (in red), from 0 to 72 h lead time, around the RSMC official forecast (black dotted line) of TC *Bejisa* from base time 0000 UTC 31 December 2013 using the error sample from five TC seasons (2011/2012 to 2015/2016) [Colour figure can be viewed at wileyonlinelibrary.com].

forecast from the respective BT data for each of the three parameters. Therefore, DVmax is positive when the predicted intensity is underestimated, ATE and CTE are positive if the observed BT position is ahead and to the left of the predicted track for an anticlockwise rotation, respectively. The sample is composed of 826 forecasts. Each forecast is composed of seven triplets (one triplet per lead time from 0 to 72 h). Each element of the forecast error sample is therefore a vector of 3×7 variables.

Figure 1 shows, for TC *Bejisa* and for base time 0000 UTC 31 December 2013, the 826 reconstructed scenarios around the RSMC official forecast up to 72 h, using the whole error sample. For each element of the error sample (seasons 2011/2012 to 2015/2016), a new scenario can be built by adding the associated error vector to the RSMC initial scenario for each lead time up to 72 h.

Given the limited size of our error sample (826 elements and 5 seasons), the choice was made to consider all available forecasts to build the error statistical distribution. Nevertheless, Figure 2 shows some systematic dependencies of the forecast error on typical characteristics of the initial (observed) state of the forecast. Four criteria of the system are considered: initial longitude (LON), initial latitude (LAT), initial intensity (Vmax), and initial size (R34KT) as defined by the radius of gale-force wind (34 kn), following Leroux *et al.* (2018). Root-mean-square errors (RMSE) were calculated for direct position (DP) and maximum wind (Vmax). Results show a weak dependency of the position error on the four criteria. In contrast, intensity scores exhibit a strong dependency on initial size with larger errors for small systems (lower tercile of the sample) which confirms what Leroux

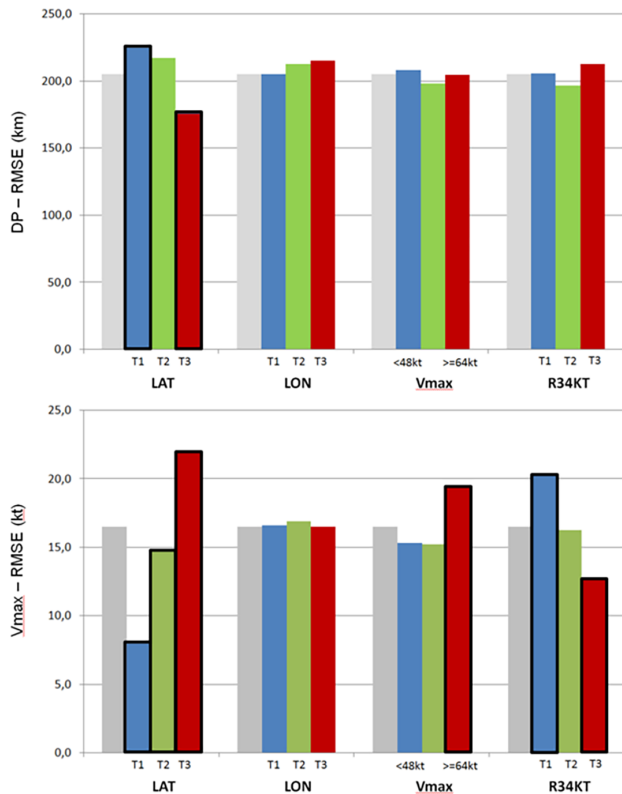


FIGURE 2 Mean DP and Vmax root-mean-square error (RMSE) of RSMC forecasts for lead times 24, 48 and 72 h during five TC seasons (2011/2012 to 2015/2016) considering four criteria: Initial latitude (LAT), longitude (LON), intensity (Vmax), and size as defined by the radius of gale-force wind (R34KT). In the LAT, LON and R34KT criteria, the initial sample (grey bar) is divided into three terciles: Lower (T1), median (T2) and upper (T3) terciles. For example, T1 in the R34KT criterion corresponds to small-size storms. In the Vmax criterion, the distribution is divided into three classes using the Vmax thresholds for TC classification (below 48 kn, between 48 and 64 kn, and above 64 kn) [Colour figure can be viewed at wileyonlinelibrary.com].

et al. (2018) found for mid-level systems. Intensity forecast errors seem bigger for TCs (initial Vmax greater than 64 kn at the initial forecast time). This result could be partly due to rapid intensification processes – which are still very difficult to forecast – knowing that there is a higher propensity for systems to intensify rapidly for an initial intensity between 65 and 75 kn (Leroux *et al.*, 2018). Finally, intensity forecast errors depend on storm latitude and decrease from the Equator (T3) to the Pole (T1, large negative values of LAT). With an increasing sample of forecast errors in the future, it might be valuable to consider these dependencies when building the climatological scenarios of an updated version of our method.

3.2.2 | Classification

As said before, operational constraints in terms of delivery time for final products limit the number of scenarios that we can consider. One efficient way to reduce the number of scenarios, without losing useful information within the spectrum of possibilities, is to partition the original error sample.

Clusters are built through an agglomerative hierarchical cluster analysis (HAC: Murtagh and Contreras, 2012) applied

on the 826 error vectors described above. The aggregation process is done following Ward's criterion (Ward, 1963). Prior to the HAC, each column of the initial error matrix (3×7 columns and 826 rows) is standardized and a principal component analysis (PCA) is applied to the 826 vectors of standardized error.

As often in classification operations, the critical part consists in choosing the most relevant number of classes. Many criteria can be considered but the important background constraint here is to end up with a limited ensemble of scenarios (maximum 30). As discussed later on, further tests showed that the number of climatological classes should not exceed 50 to comply with the operational constraint. On the other end, the number of climatological classes should be sufficient to end up with a final number of scenarios greater than 15. To objectively select the final number of classes, two indices are considered: the connectivity and the silhouette width (Handl *et al.*, 2005). The connectivity (to be minimized) indicates to what extent the neighbours of each element in the data space can be assigned to the same class. The silhouette width (to be maximized) analyses how well elements of the sample are clustered and brings some indication of the average distance between clusters. The evolutions of the connectivity index and of the silhouette width as a function of the number of classes varying from 30 to 60 indicate two privileged number of classes: 35 and 45 (Figure 3). Considering the two initial constraints in terms of the final number of scenarios (ensemble size), the best compromise was to keep 45 classes. With this choice, applying the whole method (further described), leads to a number of scenarios between 12 and 30 (81% of the cases having more than 15 scenarios). With 35 climatological clusters, 30% of the cases would have an ensemble size below 15.

3.2.3 | Climatological scenarios and their associated probability

A number of 45 distinct error vectors can be built by averaging the elements of each cluster of errors. Each vector is given a probability that is proportional to the number of elements that fall into that class. Then 45 climatology-built scenarios can be generated by applying the 45 error vectors (ATE-CTE-DVmax from 0 to 72 h lead time) to any official deterministic scenario issued by RSMC La Réunion. To illustrate this, Figure 4 shows the 45 CLIM scenarios associated with the TC *Bejisa* forecast issued on 0000 UTC 31 December 2013.

3.3 | Step 2: Consider “real-time” predictability

As demonstrated in Dupont *et al.* (2011), an uncertainty cone built using the distribution of EPS members is more relevant than a climatology-built cone. Step 2 of the method exposed here is meant to benefit from the valuable EPS inputs and add real-time uncertainty to the 45 CLIM scenarios built in step 1.

Difference vectors are computed with respect to the ensemble mean for each of the 51 individual EPS members;

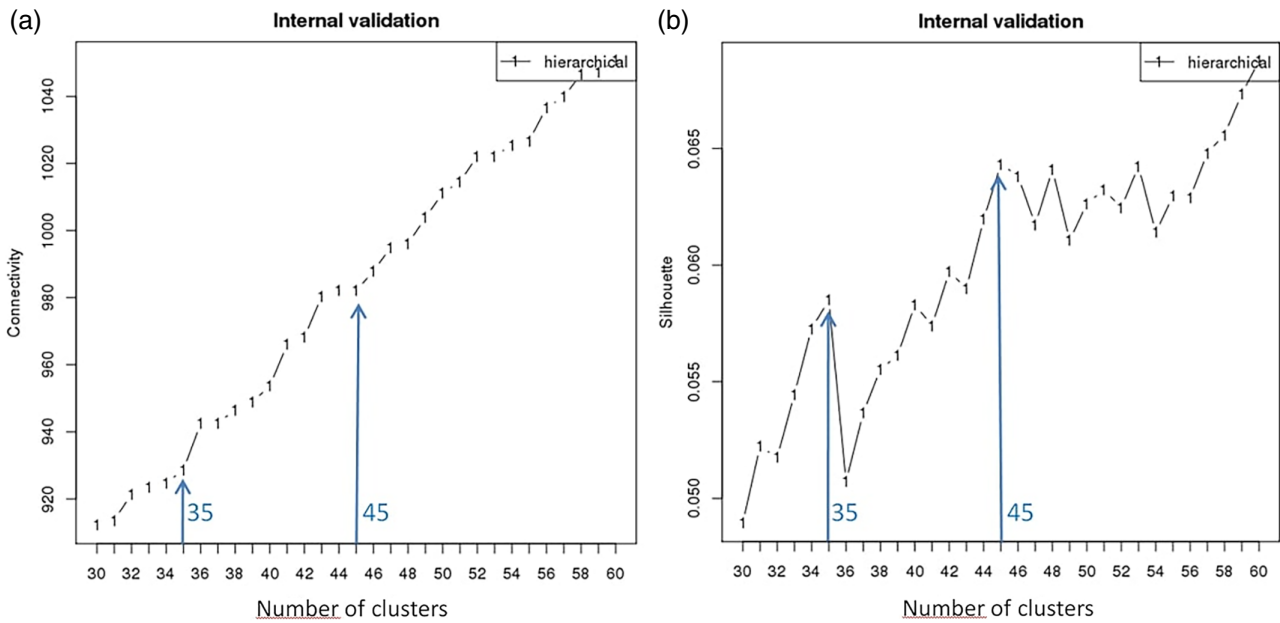


FIGURE 3 (a) Connectivity index and (b) silhouette width as a function of the number of clusters related to the HAC performed on the 826 error vectors defining the sample [Colour figure can be viewed at wileyonlinelibrary.com].

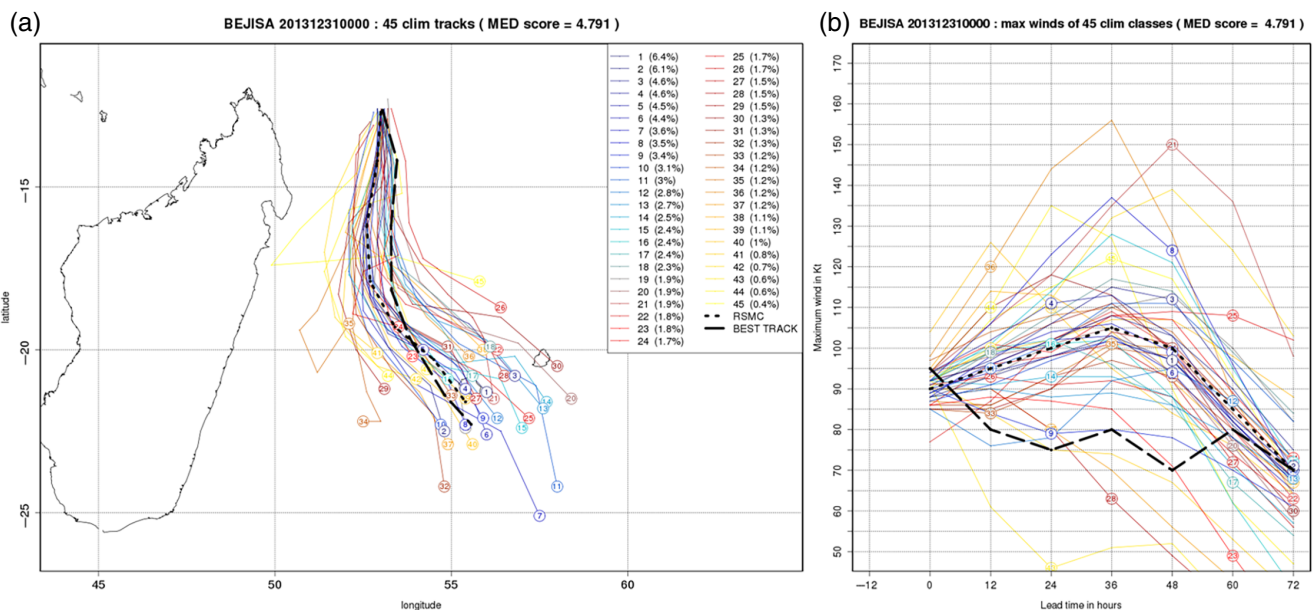


FIGURE 4 (a) Tracks and (b) maximum wind V_{max} (kn) up to 72 h lead time for 40 CLIM scenarios (coloured lines) around the RSMC official forecast (black dotted line) issued at 0000 UTC 31 December 2013 for TC *Bejisa*. Each scenario is referenced by a number indicated at the end of the track and carries a probability (%) indicated in the right panel legend. BT data are added for verification (black dashed line) [Colour figure can be viewed at wileyonlinelibrary.com].

these 51 ATE-CTE-DV $_{max}$ vectors are then added to the RSMC official track (e.g. Figure 5 for TC *Bejisa*) to build a new set of 51 tracks. In consistence with the CLIM scenarios built in step 1, the EPS ensemble is realigned (or centred) on the reference RSMC forecast (e.g. Figure 6 for TC *Bejisa*). This realigned EPS ensemble forecast is called EPS-REC in the rest of the article.

Each of the 51 EPS-REC scenarios is then assigned to the nearest climatological class, using a Euclidian distance criterion which leads to a modulation of the initial probability of each of the 45 CLIM scenarios previously built. The final

probability ($P_{weighted.cl}$) of a given scenario “ cl ” is that of the CLIM class (P_{cl}) weighted by the number of EPS-REC members ($nb.members.cl$) that fall into class cl :

$$P_{weighted.cl} = P_{cl} \times \left(\frac{nb.members.cl * \sum_{k=1}^{45} (nb.elements.k)}{\sum_{k=1}^{45} (nb.elements.k * nb.members.k)} \right),$$

where $nb.elements.k$ is the number of initial elements contained in the 45 CLIM clusters. Note that the probability of a climatological class decreases all the way to 0 when none of the 51 EPS-REC members are assigned to it. In that case,

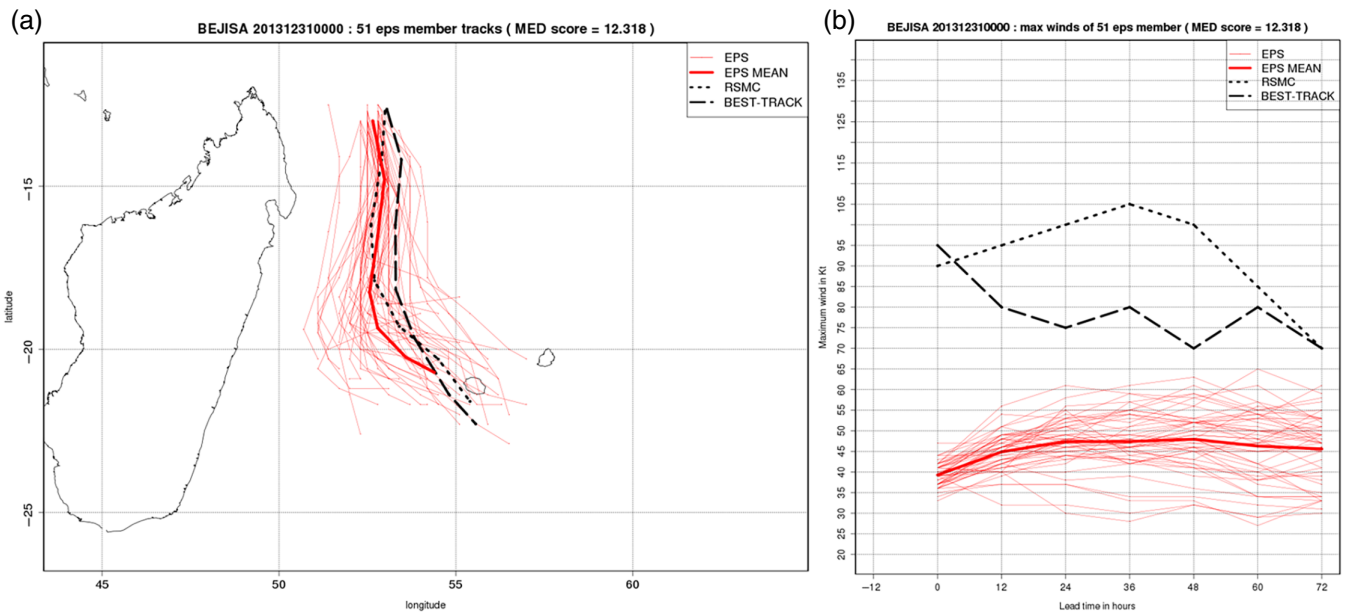


FIGURE 5 As in Figure 4 but for the 51 EPS scenarios (red lines) [Colour figure can be viewed at wileyonlinelibrary.com].

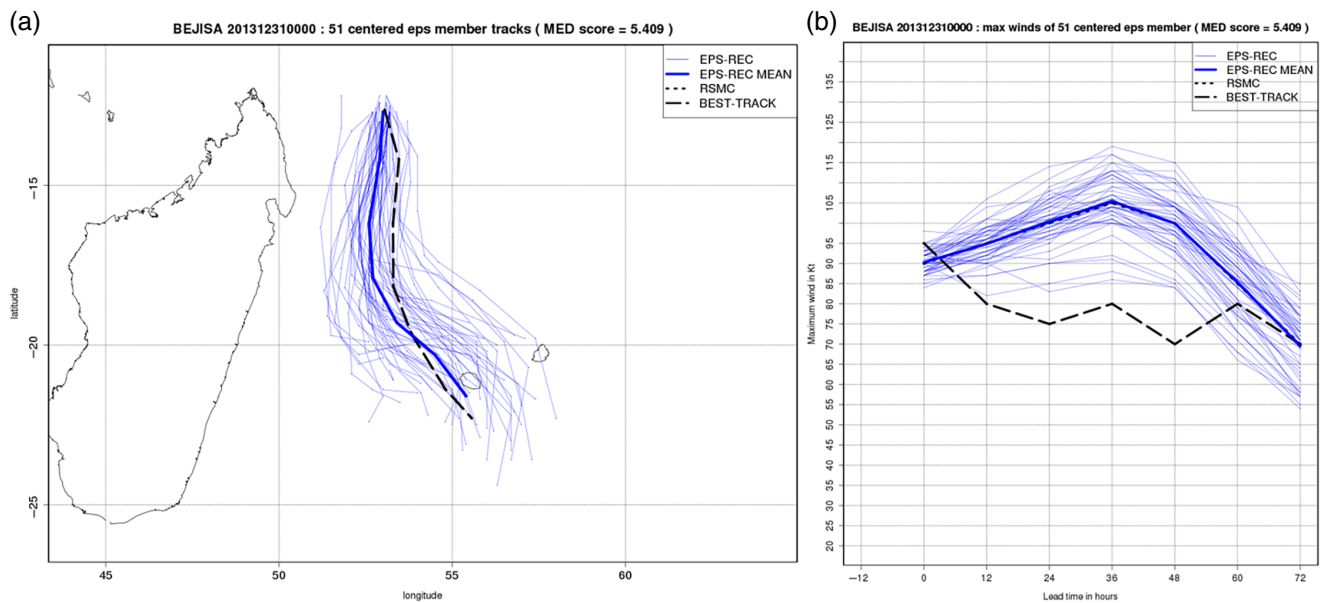


FIGURE 6 As in Figure 4 but for the 51 EPS-REC scenarios (blue lines), corresponding to the realignment of the 51 EPS scenarios displayed in Figure 5 on the RSMC official forecast (black dashed line) [Colour figure can be viewed at wileyonlinelibrary.com].

the climatological class disappears from the set of 45 original classes. This weighting process typically decreases the initial number of 45 CLIM classes to some 15–30 CLIM-EPS scenarios (e.g. Figure 7 for TC *Bejisa*).

4 | EVALUATION

The performance of a probabilistic forecasting system can be evaluated in many different ways depending on the targeted goal. A new probabilistic forecasting system is usually evaluated against a “reference” pre-existing system. In our case, the three methods (CLIM, CLIM-EPS and EPS-REC) are evaluated with respect to the ECMWF EPS system that provides 51 members that are supposedly equally probable.

The evaluation is performed for two distinct periods: seasons 2011/2012 to 2015/2016, which is the learning period used for calibrating the method, and seasons 2016/2017 to 2017/2018 that provide independent data.

4.1 | Mean Euclidian distance score

We can evaluate the similarity between a predicted scenario and an observed one (BT) by looking at the mean Euclidian distance (MED) over the seven considered lead times (0–72 h) for both track and intensity parameters. For a given ensemble forecast (among EPS, EPS-REC, CLIM or CLIM-EPS), in order to consider both the position (DPE) and intensity (DVmax) errors, the Euclidian distance (*deu*) is computed in

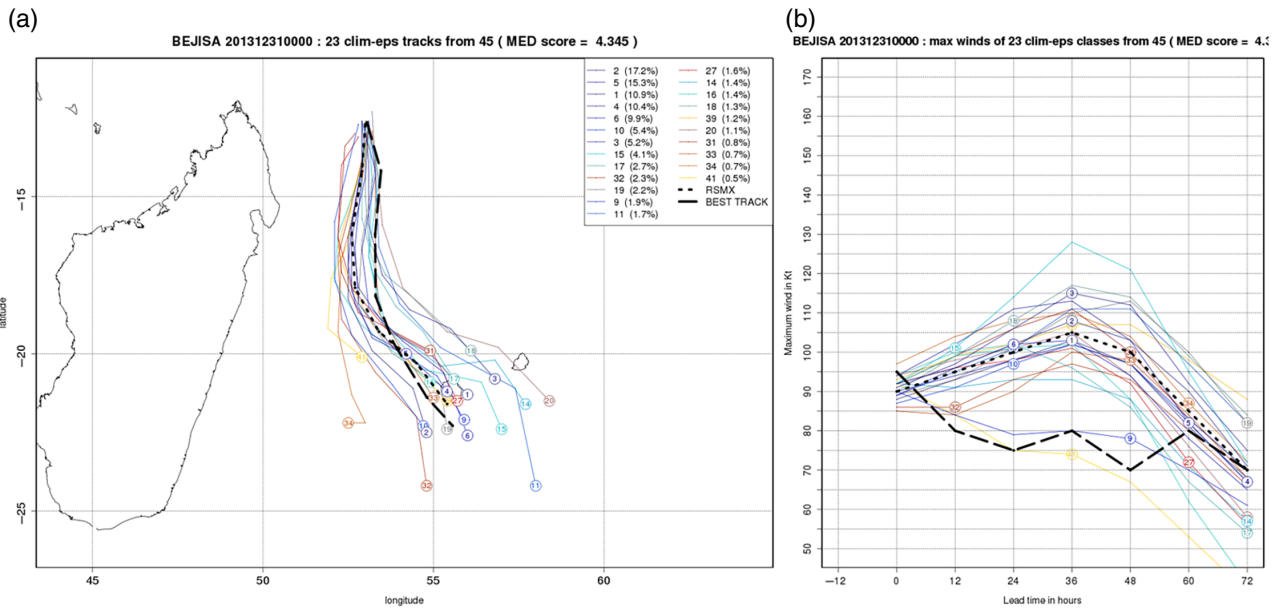


FIGURE 7 As in Figure 4 but for the 23 CLIM-EPS scenarios (coloured lines) [Colour figure can be viewed at wileyonlinelibrary.com].

the Principal Component Analysis space obtained from the classification process (cf. section 3.2.2). When considering an ensemble of forecasts, the MED is given by:

$$MED = \sum_{cl=1}^{nbcl} (deu * P.cl),$$

where $nbcl$ is the number of scenarios in the considered forecast (necessarily lower than or equal to 45 for CLIM and CLIM-EPS and equal to 51 for EPS and EPS-REC), deu the Euclidian distance for class cl and $P.cl$ the probability of scenario number cl (whatever method is considered among EPS, EPS-REC, CLIM or CLIM-EPS). A perfect forecast leads to a MED of 0. The lowest MED corresponds to the most accurate track and intensity forecast. Statistics in Table 2 show that the three evaluated methods (CLIM, CLIM-EPS and EPS-REC) give better scores than the reference system (EPS), the CLIM-EPS hybrid method performing the best whatever period is considered (the learning period 2011/2012 to 2015/2016 or the independent period 2016/2017 to 2017/2018). We can note that, overall, MED scores are better over the independent period when we would expect a decrease of scores due to potential over-fitting issues. This indicates better performances of both EPS system and RSMC forecasts during the last two cyclone seasons, with respect to the five previous ones. It also probably rules out a possible over-learning problem in the method design.

4.2 | Probability density function of forecast errors

Another way to evaluate the performance of the method is to look at the probability distribution as a function of the error. For a given lead time and error interval:

$$P = \frac{1}{nb.fcst} \sum_{i=1}^{nb.fcst} \left(\sum_{cl=1}^{nbcl} (P.cl * F.cl) \right),$$

where P is the mean probability between 0 and 1, $nb.fcst$ is the number of forecasts in the sample, $nbcl$ is the number of classes in each forecasting system, $P.cl$ is the probability of the considered member (cl) and $F.cl$ is assigned to 1 when a pre-defined condition is verified (e.g. $100 \text{ km} < \text{DPE} < 200 \text{ km}$) and 0 otherwise. The pre-defined conditions are relative to the DPE or DVmax parameters here.

Figure 8 shows the probability density function (PDF) of the DPE variable for lead times 36 and 72 h (season 2011/2012 to 2015/2016). The CLIM-EPS method (orange line) performs the best for both lead times. At 36 h lead time for example, the CLIM-EPS method yields higher cumulated probabilities at low DPEs ($< 100 \text{ km}$) and lower cumulated probabilities at high DPEs ($> 200 \text{ km}$). Scores on independent data (not shown) confirm the ones obtained within the learning period. A finer analysis can be obtained by splitting the sample into three categories according to the quality of the RSMC official forecast: “bad”, “normal” and “good”. Each forecast is assigned to a category using the three terciles of the MED “climatology” based on the RSMC forecasts sample. Figure 9 shows that the CLIM-EPS method is much more skilful when the RSMC forecast can be considered as “good” ($MED < 4,64$). In the “bad” forecast category ($MED > 5,92$), the skill of the three methods is rather equivalent, with a weak advantage for the CLIM-EPS method. Note that the reference method (EPS) is the least skilful of all, whatever the lead time and DPE.

When considering intensity errors (DVmax), the CLIM-EPS and EPS-REC methods perform equally for all lead times (Figure 10). The EPS distribution is skewed toward negative errors, indicating that the EPS method underestimates the maximum wind, most likely due to its coarser horizontal resolution. The three evaluated methods seem to globally correct that mean EPS bias since their error distribution is quite symmetric and centred on zero.

TABLE 2 Mean Euclidian distance (MED); Brier score and associated skill score in parentheses (using EPS as the reference method) for different lead times and different forecasting methods (CLIM, CLIM-EPS, EPS-REC and EPS)

Lead time (h)	MED	Brier score (Brier skill score)					
		12	24	36	48	60	72
2011/2012 to 2015/2016							
CLIM	5.8	0.27 (−0.17)	0.36 (−0.38)	0.4 (−0.33)	0.42 (−0.27)	0.42 (−0.24)	0.39 (−0.18)
CLIM-EPS	5.37	0.15 (0.35)	0.20 (0.23)	0.27 (0.1)	0.3 (0.09)	0.33 (0.03)	0.33 (0)
EPS-REC	6.1	0.14 (0.39)	0.22 (0.15)	0.28 (0.07)	0.33 (0)	0.35 (−0.03)	0.34 (−0.03)
EPS	7.76	0.23 (0)	0.26 (0)	0.3 (0)	0.33 (0)	0.34 (0)	0.33 (0)
2016/2017 to 2017/2018							
CLIM	5.25	0.28 (−0.22)	0.37 (−0.37)	0.43 (−0.34)	0.46 (−0.31)	0.43 (−0.23)	0.4 (−0.14)
CLIM-EPS	4.67	0.12 (0.48)	0.18 (0.33)	0.28 (0.13)	0.32 (0.09)	0.33 (0.06)	0.32 (0.09)
EPS-REC	5.52	0.13 (0.43)	0.22 (0.19)	0.31 (0.03)	0.33 (0.06)	0.34 (0.03)	0.34 (0.03)
EPS	6.84	0.23 (0)	0.27 (0)	0.32 (0)	0.35 (0)	0.35 (0)	0.35 (0)

Brier scores consider the binary event of a system at the minimum severe tropical storm (Vmax greater than 48 knots) stage passing within a radius of 100 km from the BT position. MED and Brier scores are calculated over two different periods: The learning period corresponding to seasons 2011/2012 to 2015/2016, and an independent period corresponding to seasons 2016/2017 to 2017/2018.

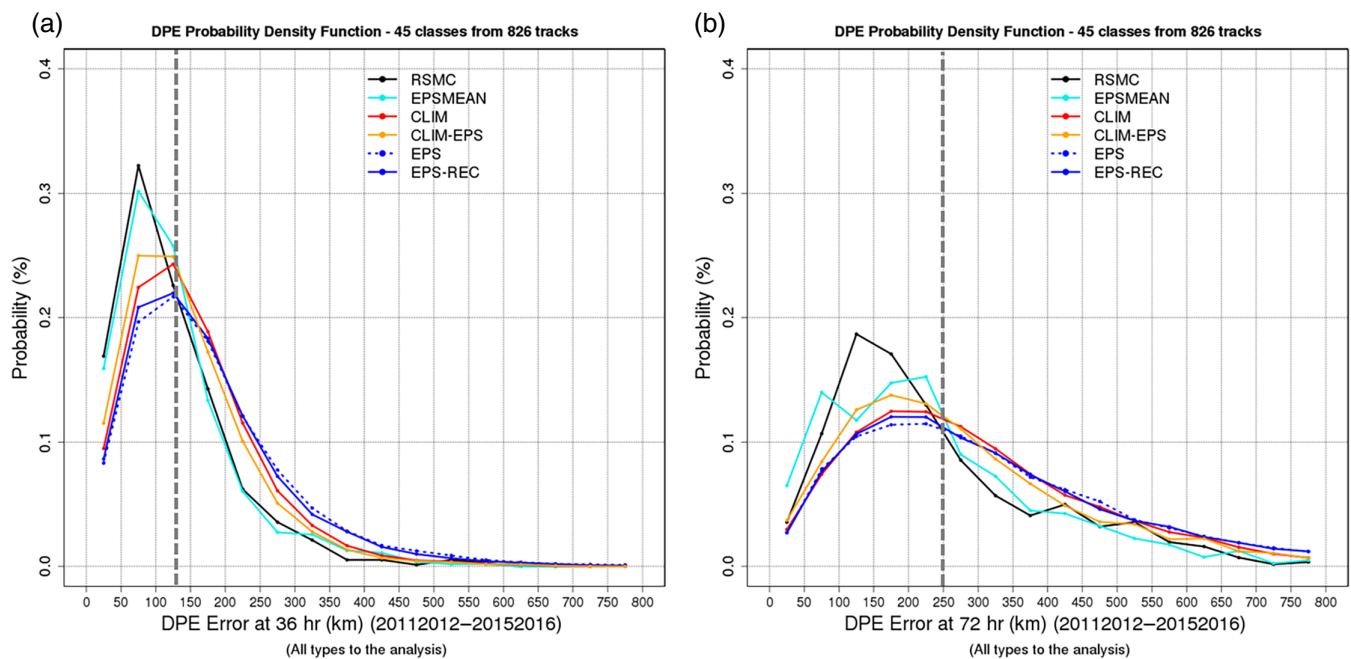


FIGURE 8 Direct position error (DPE) probability density functions at (a) 36 h and (b) 72 h lead times for the CLIM (red line), CLIM-EPS (orange line), EPS-REC (solid blue line) and EPS (dotted blue line) methods. RSMC (black line) and EPSMEAN (cyan line) DPE are displayed as they stand for two “reference” deterministic forecasts. The vertical grey line indicates the mean RSMC DPE value for each corresponding lead time [Colour figure can be viewed at wileyonlinelibrary.com].

4.3 | Skill scores

The capacity of a probabilistic forecasting system to predict the potential impacts of an event affecting inhabited land is another performance criterion. In the SWIO basin, it is commonly assumed that a tropical disturbance will highly impact the population – through strong winds, torrential rain and coastal submersion – when it passes within a radius of 100 km from the island at the minimum Severe Tropical Storm stage (STS). Note that an STS is a TS with Vmax between 48 and 63 knots. This criterion could be used to calculate Brier Skill Scores (BSSs) for the three methods. Unfortunately, the number of such “high impact” weather events in the evaluation period is too small for the scores to be

statistically significant (Leroux *et al.*, 2018). A way to bypass this limitation and increase the size of the “high impact” weather events is to consider that each location in the BT dataset that is associated with a minimum STS event during the 2011/2012–2015/2016 period is potentially inhabited.

From this assumption, Brier Scores (BSs) and BSSs can be calculated for each of the seven lead times as follows:

$$BSS = 1 - \frac{BS}{BS_{Ref}} \text{ with } BS = \frac{1}{nb.fcst} \sum_{i=1}^{nb.fcst} (p_i - o_i)^2 \text{ and}$$

$$p_i = \sum_{cl=1}^{nb.cl} (P_{cl} * F_{cl}),$$

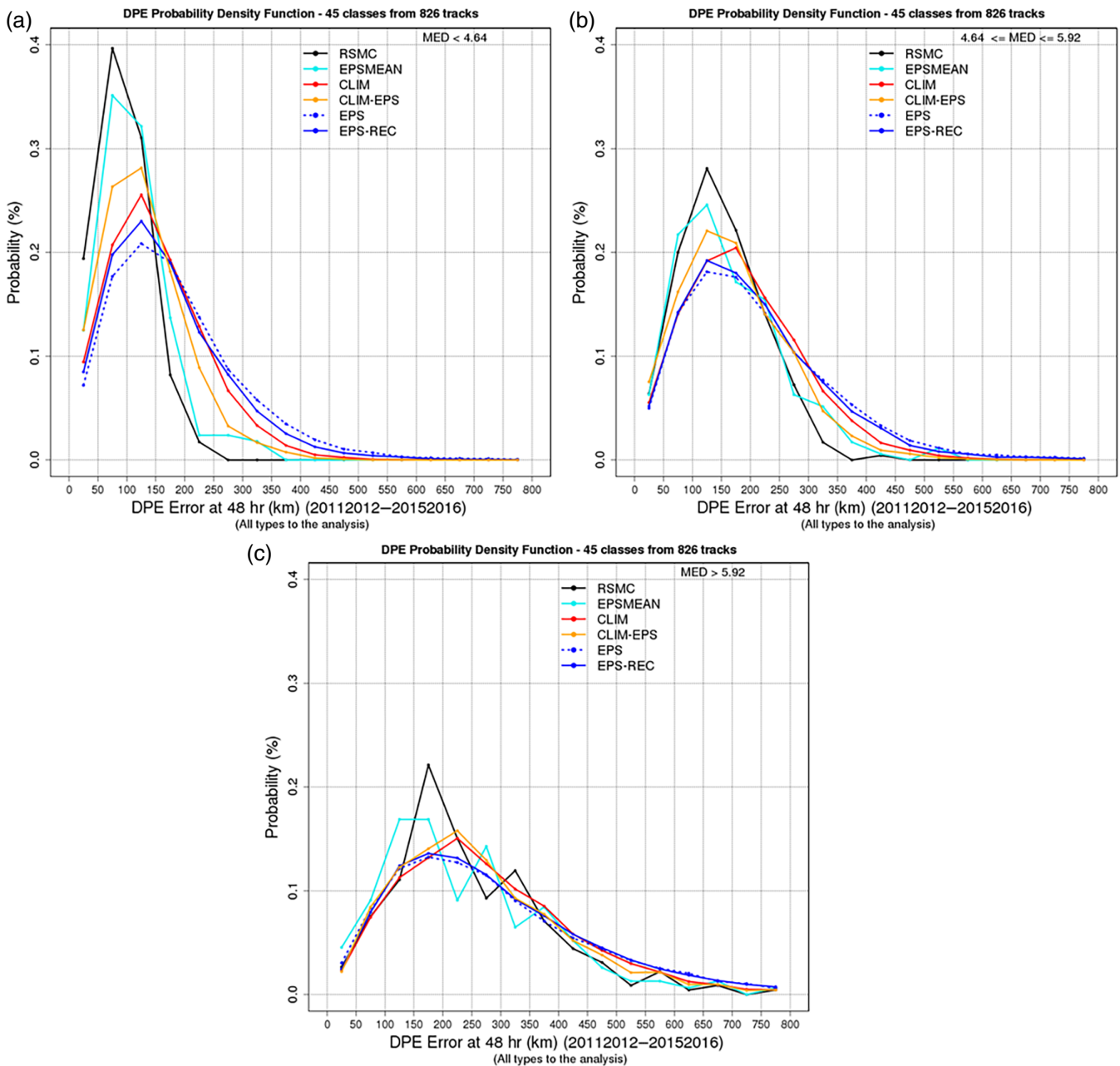


FIGURE 9 Direct position error (DPE) probability density functions at 48 h lead times for the CLIM (red line), CLIM-EPS (orange line), EPS-REC (solid blue line) and EPS (dotted blue line) methods, for three qualities of the RSMC official forecast: (a) “good”, (b) “medium” and (c) “bad”. The three categories are built considering the three terciles of the mean Euclidian distance in the RSMC forecast sample. RSMC (black line) and EPSMEAN (cyan line) DPE are displayed as they stand for two “reference” deterministic forecasts [Colour figure can be viewed at wileyonlinelibrary.com].

where the skill baseline is the EPS Brier score (BS_{Ref}), o_i is the observed event (1 if the system is at least an STS and 0 otherwise), p_i is the cumulative forecasted probability of having an STS within 100 km of the observed storm location (BT data) for a given lead time. Scores in Table 2 indicate that the CLIM-EPS hybrid method is more skilful than the others up to 48 h. Beyond that lead time, the CLIM-EPS and EPS-REC performances are very close to that of the original EPS ensemble. The CLIM method is the less skilful for all lead times, confirming the benefit of considering the real-time information provided by the EPS spread to modulate the initial CLIM probabilities. As for MED scores and PDF errors, impact-oriented skill evaluated through Brier scores indicate

that performances of all methods are comparable or better for the independent period.

5 | DISCUSSION AND CONCLUSIONS

A method has been developed to build alternate scenarios (with associated probabilities) from the official deterministic track and intensity forecast released by RSMC La Réunion from 0 to 72 h lead time. A first set of 40 scenarios (CLIM) is produced based on a 5-year climatology of track and intensity forecast errors from the operational RSMC forecast centre. Probabilities are derived from the statistical distribution

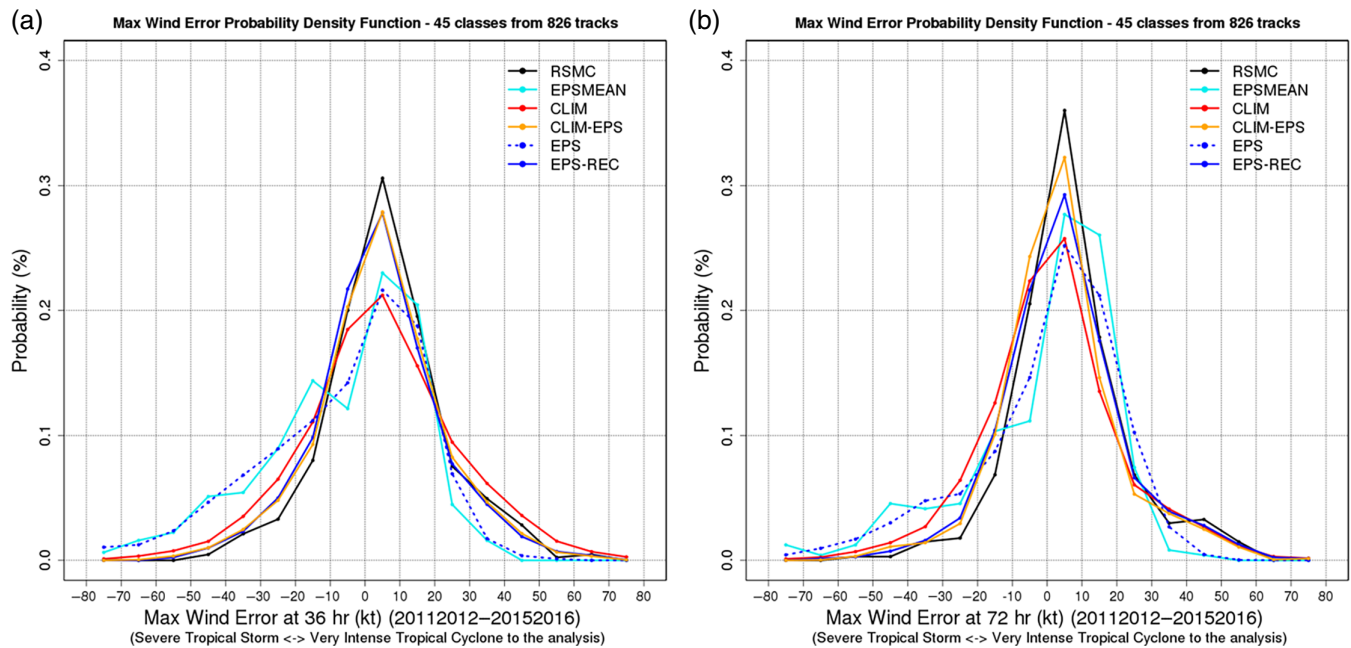


FIGURE 10 As in Figure 8 but for maximum wind error (DV_{max}) [Colour figure can be viewed at wileyonlinelibrary.com].

of errors. The CLIM probabilities are then weighted using the “realigned” ECMWF ensemble of scenarios (EPS-REC). This hybrid CLIM-EPS method yields some 15–30 final scenarios depending on the EPS ensemble spread.

The three methods (CLIM, EPS-REC and CLIM-EPS) are then evaluated using various scores. The evaluation is performed using more than 800 RSMC official forecasts performed since 2011.

All scores clearly indicate that the two methods (CLIM-EPS or EPS-REC) perform the best with respect to the original EPS or CLIM methods. This confirms the added value of:

- using the RSMC official deterministic forecast as the reference for the ensemble generation process,
- using the real-time EPS ensemble spread information to modulate the climatologically-built ensemble.

Up to 48 h lead time, the CLIM-EPS method is more skilful than the EPS-REC method, especially when looking at performances with an impact-oriented view (Brier Skill Scores). The advantage of the CLIM-EPS method is more obvious for track forecasts than for intensity forecasts for which the EPS-REC method can be considered to perform as well as the CLIM-EPS. The most interesting advantage of the CLIM-EPS method is for cases where the RSMC official forecast is skilful. In these cases, probabilities are more concentrated around the official forecast than in other methods.

Although the importance of using the RSMC official forecast against the original EPS ensemble scenarios seems clear, the choice between CLIM-EPS and EPS-REC could be argued. Without considering scores, the first and main advantage of the CLIM-EPS method is that it allows the number of scenarios to be reduced by a factor of two or three (51

for EPS-REC, 15–30 for CLIM-EPS). This is valuable considering the limited available resources to develop further operational impact-oriented applications (wind, rain, wave or surge simulations) that could ideally be run on each scenario. A second reason for choosing the CLIM-EPS method is that initial perturbations applied to the EPS ensemble members are mainly calibrated for medium-range lead times (beyond 48–72 h). Therefore, the spread of the EPS ensemble may be artificially too high in the first 24–48 h of the forecast which could explain an overall better performance of the CLIM-EPS method for short-term forecasts.

As discussed previously, the limited number of alternative scenarios generated throughout the method (15–30) was defined in order to fit the objectives of the SPICy project. However, this reduced ensemble approach may lead to some underestimation of the plume’s dispersion especially when the predictability is low. In other words, one may ask the question: how often does the true track or intensity fall outside the range of the proposed scenarios? In order to answer this question, “miss” rates have been calculated for both track and intensity. For each forecast and lead time from 0 to 72 h, we consider the forecast as a “miss” when more than half of the observed positions (BT) corresponding to each lead time are outside the cloud of points formed by the scenarios at this specific lead time. When considering tracks, the “miss” rate is 5% for the 550 cases from 2011/2012 to 2015/2016. For intensity, this rate rises to 13% which could be considered as a limitation of the method. However, when calculating the same scores for the most recent two-year period (2016/2017 to 2017/2018), the miss rate for intensity drops to 2%. This may be partially attributed to the EPS horizontal resolution improvement in 2015 that may have improved the spread skill of the EPS intensity forecast with beneficial feedback on our CLIM-EPS forecast.

Future developments should lead to improvements and should overpass some limitations of the CLIM-EPS method. We mention here three main evolutions that should be considered:

- The first one consists in improving the probability weighting module that uses the EPS ensemble spread to modulate the CLIM ensemble. The technique used and described in section 3.3 has the advantage of substantially reducing the size of the ensemble. This behaviour is quite convenient when the predictability is high. On the other hand, the final CLIM-EPS ensemble may be too narrow whenever the predictability drops or the dispersion of the EPS ensemble increases. The weighting technique could be revised to increase the spectrum of possibilities (granularity) in situations that are less predictable.
- The method does not consider modulations of the storm structure within the produced ensemble. This is a main limitation considering that our final objective is to evaluate a storm “impact”. TC-induced hazards (rain, wind, waves and surge) on a small mountainous island like La Réunion are highly dependent on the size of the storm that threatens it.
- As discussed in section 3.2.1, forecast errors show some systematic dependencies with typical characteristics of the initial (observed) state of the forecast. Therefore, the forecast error classification process could consider various learning subsamples in order to modulate the weight of the initial CLIM scenarios.

The method has been implemented quasi-operationally at RSMC La Réunion during the 2016/2017 and 2017/2018 TC seasons. Its performances are to be evaluated regularly with respect to ECMWF EPS current and future evolutions.

ACKNOWLEDGEMENTS

We acknowledge the French National Research Agency (ANR) for funding this research (ANR-14-CB03-0013 SPICY). We also warmly thank Sophie Lecacheux from BRGM for an efficient coordination of this project and her constant support and suggestions on this study.

ORCID

Francois Bonnardot  <https://orcid.org/0000-0002-8729-2020>

REFERENCES

- Alessandrini, S., Delle Monache, L., Rozoff, C.M. and Lewis, W.E. (2018) Probabilistic prediction of tropical cyclone intensity with an analog ensemble. *Monthly Weather Review*, 146, 1723–1744. <https://doi.org/10.1175/MWR-D-17-0314.1>.
- Chen, P.Y., Yu, H., Brown, B., Chen, G.M., and Wan, R.J. (2016) A probabilistic climatology-based analogue intensity forecast scheme for tropical cyclones. *Quarterly Journal of the Royal Meteorological Society*, 142(669), 2386–2397.
- Davis, J.R., Paramygin, V.A., Forrest, D. and Sheng, Y.P. (2010) Toward the probabilistic simulation of storm surge and inundation in a limited-resource environment. *Monthly Weather Review*, 138, 2953–2974. <https://doi.org/10.1175/2010MWR3136.1>.
- DeMaria, M., Knaff, J.A., Knabb, R., Lauer, C., Sampson, C.R. and DeMaria, R.T. (2009) A new method for estimating tropical cyclone wind speed probabilities. *Weather and Forecasting*, 24, 1573–1591. <https://doi.org/10.1175/2009WAF222286.1>.
- DeMaria, M., Sampson, C.R., Knaff, J.A. and Musgrave, K.D. (2014) Is tropical cyclone intensity guidance improving? *Bulletin of the American Meteorological Society*, 95, 387–398. <https://doi.org/10.1175/BAMS-D-12-00240.1>.
- Dupont, T., Plu, M., Caroff, P. and Faure, G. (2011) Verification of ensemble-based uncertainty circles around tropical cyclone track forecasts. *Weather and Forecasting*, 26, 664–676. <https://doi.org/10.1175/WAF-D-11-00007.1>.
- Handl, J., Knowles, J. and Kell, D.B. (2005) Computational cluster validation in post-genomic data analysis. *Bioinformatics*, 21(15), 3201–3212.
- Jun, S., Kang, N.-Y., Lee, W. and Chun, Y. (2017) An alternative multi-model ensemble forecast for tropical cyclone tracks in the western North Pacific. *Atmosphere*, 8(9), 174. <https://doi.org/10.3390/atmos8090174>.
- Leroux, M., Meister, J., Mekies, D., Dorla, A. and Caroff, P. (2018) A climatology of southwest Indian Ocean tropical systems: their number, tracks, impacts, sizes, empirical maximum potential intensity, and intensity changes. *Journal of Applied Meteorology and Climatology*, 57, 1021–1041. <https://doi.org/10.1175/JAMC-D-17-0094.1>.
- Murtagh, F. and Contreras, P. (2012) Algorithms for hierarchical clustering: an overview. *WIREs Data Mining and Knowledge Discovery*, 2, 86–97. <https://doi.org/10.1002/widm.53>.
- Plu, M. (2011) A new assessment of the predictability of tropical cyclone tracks. *Monthly Weather Review*, 139, 3600–3608. <https://doi.org/10.1175/2011MWR3627.1>.
- Qi, L.B., Yu, H. and Chen, P.Y. (2014) Selective ensemble mean technique for tropical cyclone track forecast by using ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 140, 805–813.
- Swinbank, R., Kyouda, M., Buchanan, P., Froude, L., Hamill, T.M., Hewson, T.D., Keller, J.H., Matsueda, M., Methven, J., Pappenberger, F., Scheuerer, M., Tittley, H.A., Wilson, L. and Yamaguchi, M. (2016) The TIGGE project and its achievements. *Bulletin of the American Meteorological Society*, 97, 49–67. <https://doi.org/10.1175/BAMS-D-13-00191.1>.
- Ward Jr., J.H. (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236–244. <https://doi.org/10.1080/01621459.1963.10500845>.
- Yamaguchi, M., Vitart, F., Lang, S.T., Magnusson, L., Elsberry, R.L., Elliott, G., Kyouda, M. and Nakazawa, T. (2015) Global distribution of the skill of tropical cyclone activity forecasts on short- to medium-range time scales. *Weather and Forecasting*, 30, 1695–1709. <https://doi.org/10.1175/WAF-D-14-00136.1>.
- Zhang, X.P. and Yu, H. (2017) A probabilistic tropical cyclone track forecast scheme based on the selective consensus of ensemble prediction systems. *Weather and Forecasting*, 32(12), 2143–2157. <https://doi.org/10.1175/WAF-D-17-0071.1>.

How to cite this article: Bonnardot F, Quetelard H, Jumaux G, Leroux M-D, Bessafi M. Probabilistic forecasts of tropical cyclone tracks and intensities in the southwest Indian Ocean basin. *Q J R Meteorol Soc* 2019;1–12. <https://doi.org/10.1002/qj.3459>