446

Performance and application of ECMWF EPS forecasts in the prediction of heavy rain and high winds in Hong Kong

Linus H.Y. Yeung, Edwin S.T. Lai, Queenie C.C. Lam, Philip K.Y. Chan, and P. Cheung

Operations Department

Hong Kong Observatory, 134A Nathan Road, Kowloon, Hong Kong

Abstract

ECMWF EPS meteograms for grid points in the vicinity of Hong Kong have been made available to the Hong Kong Observatory since January 2003 for evaluation purpose. The 6-hourly rainfall forecasts are verified against the analysed rainfall over Hong Kong using standard verification measures and skill scores as applied to the five EPS parameters of an ensemble distribution (i.e., the ensemble minimum, the 25% quartile, the median, the 75% quartile and the maximum). EPS rainfall prediction, trends and patterns on a day-to-day basis are also evaluated for ten selected heavy rain events, as well as case studies of EPS wind speed forecasts in relation to the prediction of high winds in tropical cyclone situations.

Through the study, usefulness of EPS forecasts in terms of accuracy, bias and relative value for a small region like Hong Kong is better understood. EPS rainfall forecasts probably should not be applied directly to operational forecasting due to problems of bias and large RMS errors; calibration and other post-processing will be required for both deterministic and probabilistic use. The EPS wind speed forecasts based on extreme members, meanwhile, are more promising for high-wind predictions in Hong Kong. For both extreme rainfall and high-wind forecasts, it is the extreme members in the ensemble (typically the ensemble MAX) that offer the most valuable information. But given the current limitations in information availability, forecasters would still have to rely on alternative guidance to first work out the most likely weather scenario, and then to refer to the EPS meteograms for the probable rainfall or winds that can be expected arising from this adopted scenario.

1 Introduction

By the end of 2003, the Hong Kong Observatory has extended the public deterministic weather forecasts from 5 days to 7 days. The extension of forecast range has created interest and prompted discussion on the merit of introducing a probabilistic element as an alternative approach in formulating the forecasts. The Observatory also has the responsibility of issuing warnings against severe weather that may affect Hong Kong, including rainstorms and tropical cyclones. Again, a deterministic approach may not be the best option, even in the very short range, given the highly volatile and unpredictable nature of such severe weather systems. To tackle the various operational forecasting and warning issues involved, use of probability forecast products from an ensemble prediction system (EPS) is explored.

By courtesy of the European Centre for Medium-range Weather Forecasts (ECMWF), EPS meteograms for two model grid points in the vicinity of Hong Kong have been made available to the Observatory in image format since January 2003 for evaluation purpose. More information on ECMWF EPS can be found at ECMWF's website [ECMWF(a)]. In the sample of EPS meteograms shown in Figure 1, probability forecasts for four weather elements are displayed in time series of distribution symbol up to 10 days ahead: (a) total cloud cover, (b) 6-hourly accumulated rainfall, (c) 10-metre wind speed and (d) 2-metre temperature. Each distribution symbol contains five parameters of the forecast ensemble, namely the ensemble minimum (or MIN), the 25% quartile, the median, the 75% quartile and the maximum (or MAX), hereafter referred to as the five EPS parameters. Also included for reference in the EPS meteograms are the TL255 control forecast (red dotted line) and the TL511 deterministic forecast (blue solid line).

To facilitate analysis and verification, the five EPS parameters for rainfall are manually extracted from the precipitation meteograms. As the EPS probability refers to the probability of rainfall being confined within a certain value, "forecast probability" of a rain event exceeding a certain rainfall threshold, which is our main interest in the discussion of heavy rain occurrence, is then simply 100% minus the EPS probability (e.g. if the EPS parameter of 75% quartile corresponds to a rainfall value of 10 mm, then the "forecast probability" of a rain event exceeding the 10-mm threshold would be 25%).
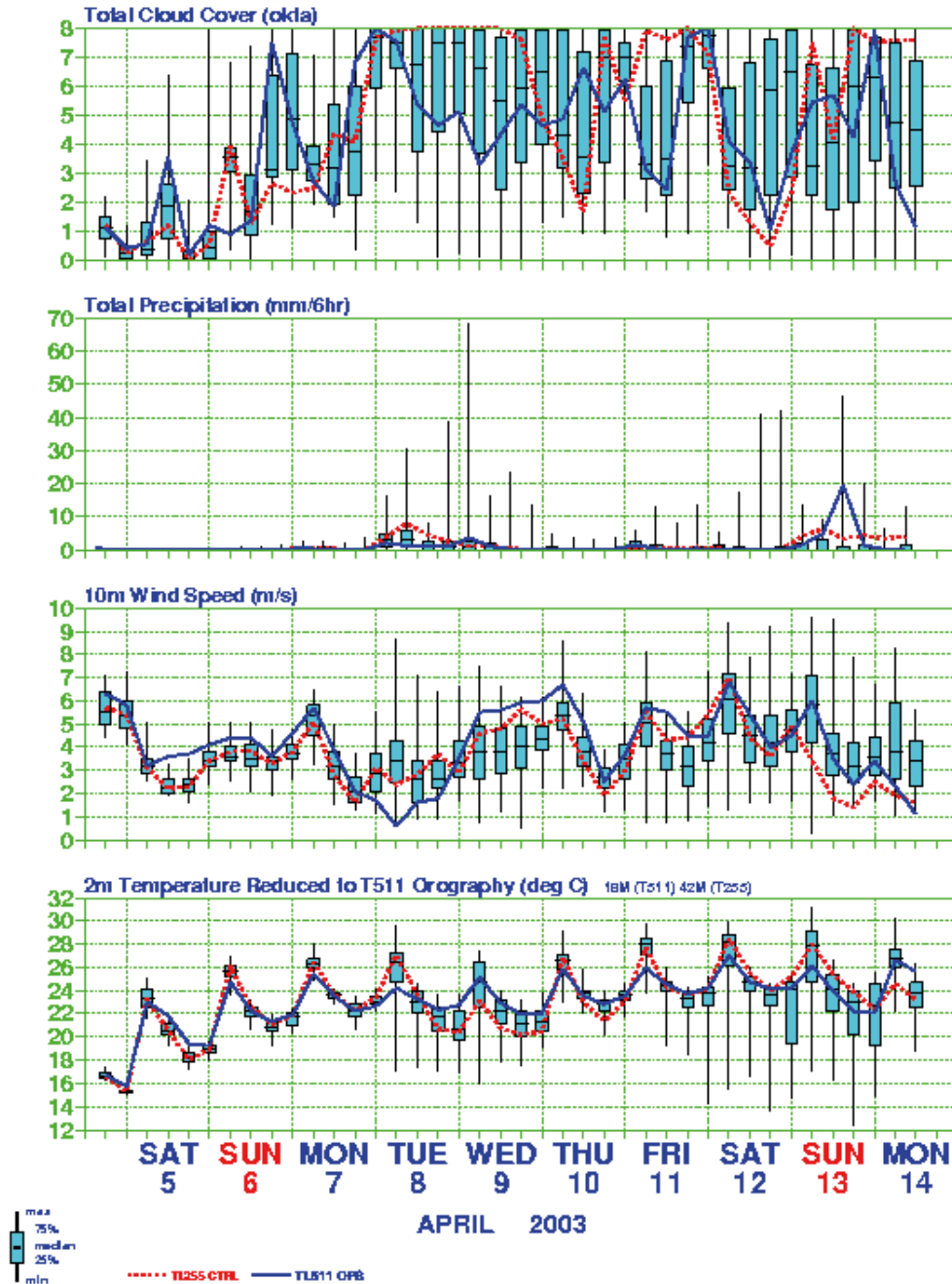
Figure 1 A sample EPS meteogram of ECMWF. As explained in the legend, the top spike, the bottom spike and the rectangle represent respectively the highest 25%, the lowest 25% and the 25-75% range of the ensemble values. The ensemble median is indicated by a dash in the rectangle. The blue solid line and the red broken line represent the TL511 deterministic and TL255 control forecasts respectively.

The two available model grid points are located at 22.8°N, 113.8°E over land (with fraction of land = 0.7 in its grid box, i.e. land point) and at 22.1°N, 114.5°E over sea (with fraction of land = 0.1 in its grid box, i.e. sea point). Neither is co-located exactly over the Hong Kong territory (see Figure 2). In this study, we shall

concentrate on forecasts at the land point in view of its higher land-sea mask value and hence more readily available observations and data from land stations for verification purpose.
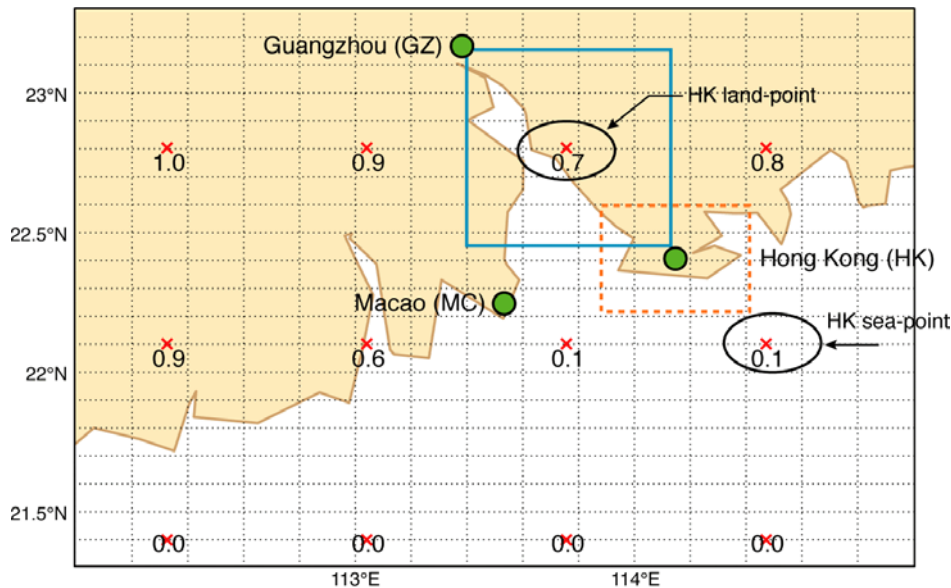

Figure 2 Land-sea mask for the ECMWF EPS meteograms. The red crosses are the exact positions of the model grid points. The values shown at each grid point is the land-sea mask which is defined as the fraction of land area in a grid box. Only points with values above 0.5 are considered as land points. As shown by the blue and the orange boxes respectively, the forecast area and the observation site only have a small overlapping area and their sizes are also different. Also shown in the figure are the locations (green dots) of Hong Kong, Guangzhou and Macao.

With emphasis on heavy rain and high wind forecasts in Hong Kong, the validation exercise will mainly focus on three aspects, results of which are given in Sections 2–4. Section 2 will be a discussion of the 6-hourly rainfall prediction as verified against the analysed rainfall over Hong Kong using standard verification measures and skill scores as applied to the five EPS parameters. Section 3 will evaluate the EPS rainfall prediction, trends and patterns on a day-to-day basis for ten selected heavy rain events. Case studies of EPS wind speed forecasts in relation to prediction of high winds due to tropical cyclones are presented in Section 4. Concluding remarks and other observations are summarized in Section 5.

2   EPS 6-hourly Rainfall Prediction

With emphasis on EPS capability in the prediction of heavy precipitation, the verification data set has a purposely-designed bias towards rainy situations through the adoption of the following 2-step event selection strategy. The first step is to identify "rainy" days with issuance of thunderstorm or rainstorm warnings in Hong Kong from January to October in 2003. In the second step, the periods of 9 days before and 8 days following such "rainy" days (26 days in total) are also included for verification to cater for the "less rainy" situations. After rejecting days of overlapping among the rainy episodes, the final verification set contains a total of 191 model runs initialized at 12 UTC. Although results presented in this paper are essentially based on the 12-UTC set of EPS meteograms, the conclusions drawn are just as applicable to a similar verification set of 00-UTC EPS meteograms (see also Section 2.3 and Figure 13).

To obtain the "ground truth" at a sufficiently high temporal resolution for validating the 6-hourly rainfall prediction, Barnes analysis is applied to observed rainfall data, quality-checked against radar observation, collected by a dense raingauge network (more than 130 land-based raingauges) within Hong Kong over a

grid of 48 × 36 grid points at a spacing of about 1.5 km. The "observed" rainfall used for validation purpose is taken to be the value averaged over all the land points in the analysis grid (covering about one-third of the domain), so as to exclude the peripheral sea areas where the analysed values may be less reliable. The constraints imposed by surface rainfall data availability means that the EPS forecast area is located to the northwest of the verification area with only a relatively small overlapping region between the two (see Figure 2). The verification area of around 1,200 km2 is also smaller in size than the EPS forecast grid box, which typically covers an average area of about 6,400 km^2.

It turns out that the lowest 6-hourly "observed" rainfall as objectively analyzed over Hong Kong (see previous paragraph) on the "rainy" days is around 2.7 mm, hence giving us the baseline for defining "significant rainfall". At first glance, an event with 6-hourly rainfall of less than 10 mm, say 2.7 mm in 6 hours as in our definition of significant rainfall, may not seem too dramatic. But if we take into consideration the "sub-grid" variability of the actual rainfall distribution, a wide spectrum of severe rain events at smaller scales is actually possible. Localized rainstorms can in general be much more intense than the area-averaged values may otherwise suggest. In fact, the rain intensity can be up to an order of magnitude different.

Stratified into six categories of rainfall intensity, or "Ω" (in units of mm/6 hours): (a) $0 \leq \Omega < 1$, (b) $1 \leq \Omega < 5$, (c) $5 \leq \Omega < 10$, (d) $10 \leq \Omega < 20$, (e) $20 \leq \Omega < 30$ and (f) $\Omega \geq 30$, the frequency distributions of observed rainfall and EPS forecasts (MAX and median) are plotted in Figure 3, in which EPS forecasts at all forecast ranges (i.e. D+1 to D+10), i.e. a total of 7,640 rainfall forecasts, are included in compiling the histogram. It can be seen that about 80% of the observed rain events fall into category (a), with exponentially decreasing frequency for higher categories. There are only 11 cases (i.e., less than 2%) with "observed" rainfall equal to or greater than 30 mm/6 hr.
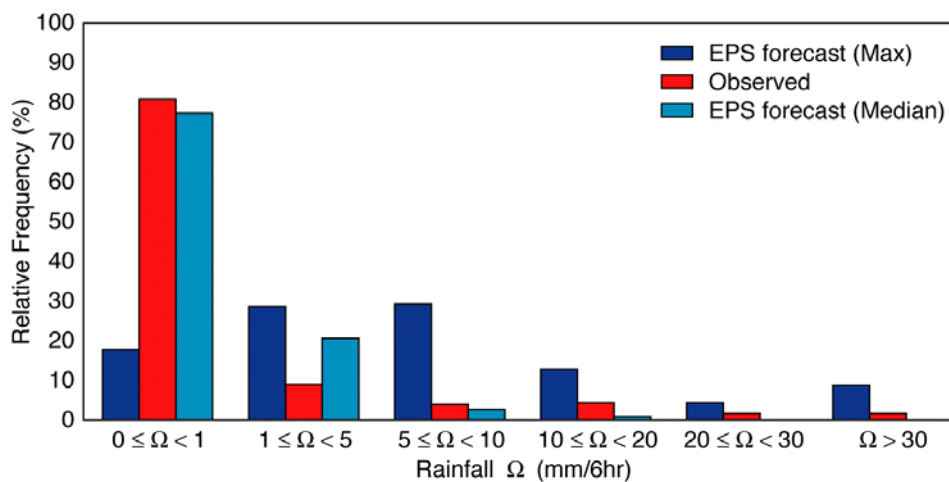

Figure 3 Frequency distributions of observed (red) rain events binned into six categories as shown by the labelling on the x-axis. Also shown in the histogram are the frequency distributions of EPS median (cyan) and EPS MAX (blue) forecast rainfall.

The six possible scenarios of distribution symbols with respect to the five EPS parameters are schematically illustrated in Figure 4, with actual occurrences for the six scenarios among the 7,640 EPS rainfall forecasts tabulated underneath. More than half of the forecasts have none of the distribution symbol or only the top spike showing. Less than 3% have a non-zero MIN; and even for this 3% of cases, the MAX can be approximated to the spread (i.e. MAX – MIN). But when rain is actually observed, the distribution symbol will show up 95% of the time, either in parts or in whole. On the other hand, when part or the whole of the

distribution symbol appears, rain is observed only 47% of the time. More discussion about the distributions of EPS forecast parameters are presented in Section 2.4.

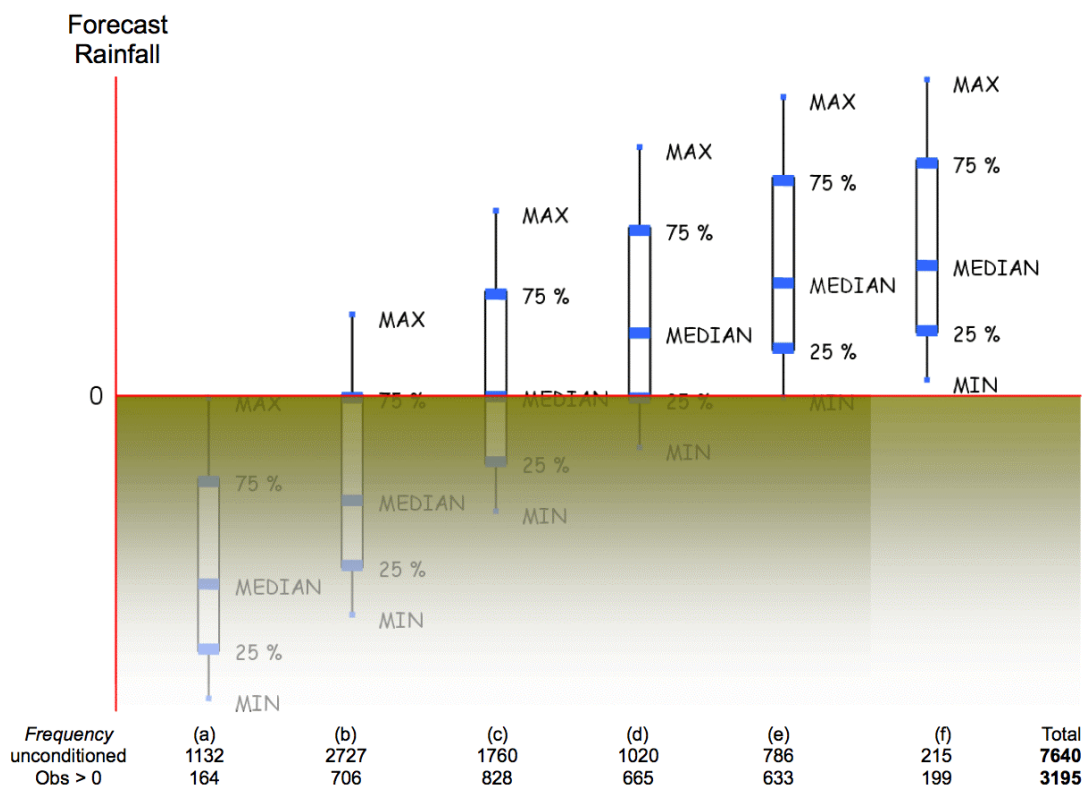| Frequency	(a)	(b)	(c)	(d)	(e)	(f)	Total
unconditioned	1132	2727	1760	1020	786	215	**7640**
Obs > 0	164	706	828	665	633	199	**3195**

Figure 4 Due to lack of detailed information on the ensemble distribution, the EPS rainfall forecasts can only be partitioned into the above six categories: (a) the entire ensemble forecast being zero; (b) the bottom 75 % of the ensemble are zero (when only thin spikes are seen); (c) the bottom 50 % are zero (when only the maximum and the 75-% quartile are seen as non-zero); (d) the bottom 25 % are zero (when only the maximum, the 75-% quartile and the median are seen as non-zero); (e) only the minimum is zero (when the maximum, the 75-% quartile, the median and the 25-% quartile are seen as non-zero); and (f) the entire ensemble are non-zero. The two numbers underneath each category indicate respectively the corresponding frequencies of occurrence for all conditions and on the condition that observed rainfall being greater zero (Obs > 0).

2.1   Contingency Table Statistics of EPS Forecasts

The distributions of the "hit", "miss", "false", and "correct reject" counts (denoted by H, M, F, and Z respectively; see Appendix A for definitions) of contingency tables are displayed as spider plots in Figure 5 for the five chosen rainfall thresholds of $\Omega \geq 1$, $\Omega \geq 5$, $\Omega \geq 10$, $\Omega \geq 20$ and $\Omega \geq 30$ (units again in mm/6 hours). Ideally, the webs should stretch towards both the H and Z corners and shrink away from the other two. In reality, they are far from ideal and look drastically different for different decision probabilities in different rainfall regimes. In the heavy rain regimes, most counts go to the "correct reject" corner no matter what the decision probability is. "Hit" counts even under loose decision criteria are still negligibly small. In effect, the analyses merely show that whatever the rainfall thresholds and whatever the decision criteria chosen, the EPS is good at forecasting no rain but is not particularly astute in picking out the rain events when they actually occur. The consequences of such highly asymmetrical contingency table statistics for extreme rainfall events will become apparent in the following discussion.
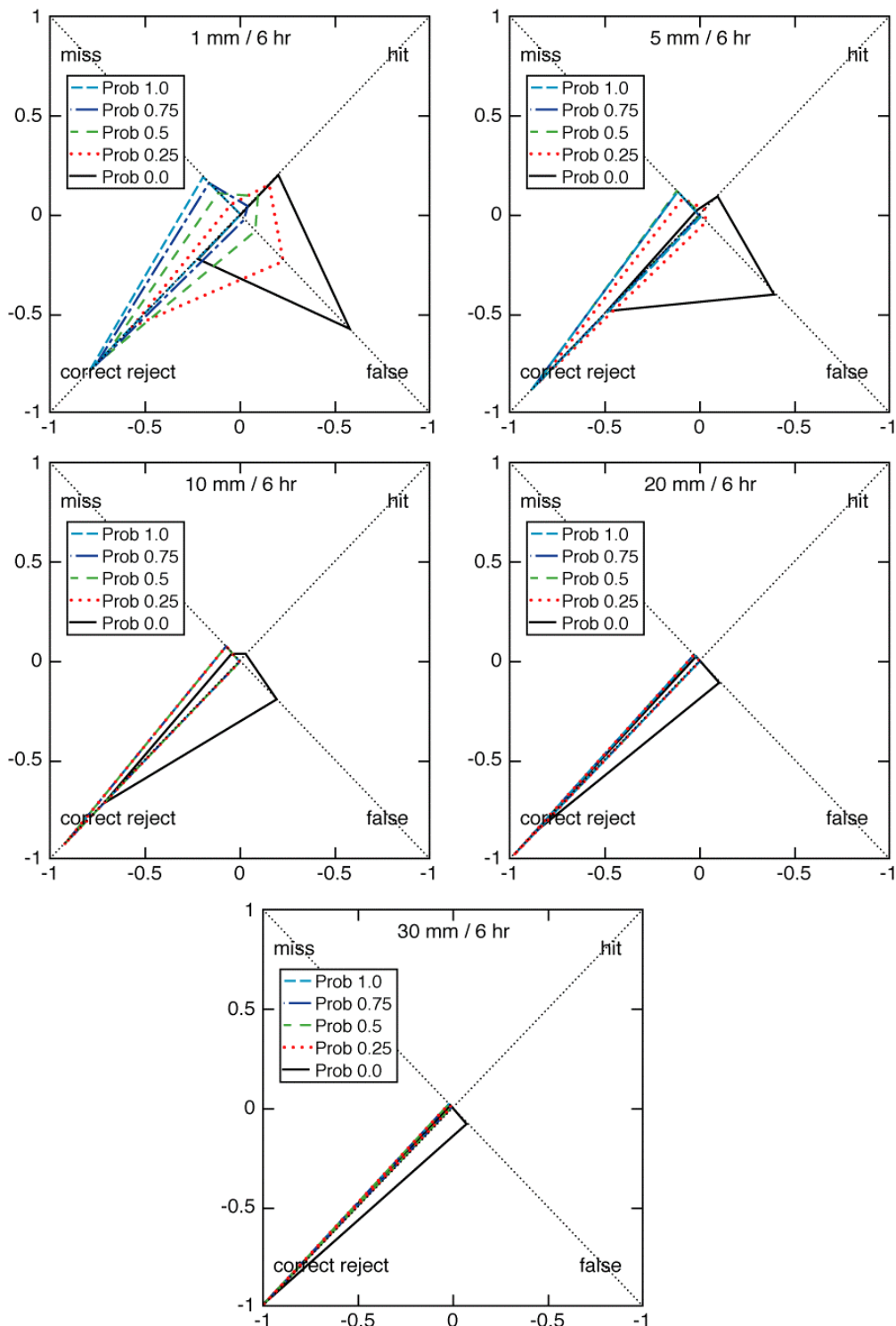
Figure 5 Spider plots showing the hit, miss, false, and correct-reject counts normalized by the total number of 7640 cases. Top-left to bottom-right panels display the distributions using threshold values of 1, 5, 10, 20 and 30 mm/6 hr respectively.

When plotted against forecast probability for the five rainfall thresholds, both the probability of detection, or POD defined as H/(H+M), and the probability of false detection, or PFD defined as F/(F+Z) [BOM; Jolliffe & Stephenson 2003], generally decrease as functions of decision probability (diagrams not shown). Such decays are particularly sharp for heavy rain events. They are related to the model bias towards low

probability as well as the intrinsic rarity of extreme events. In search of a reasonably high POD while maintaining a low PFD, Figure 6 plots the difference of POD–PFD against forecast probability. Under all thresholds, the useful probability range is skewed towards the lower end. For thresholds of 5 mm or smaller, the useful probability range is much wider; and 0.25 appears to be the optimal decision probability for such rain events in general. For thresholds of 10 mm or higher, the decision probability has to be smaller than 0.25.
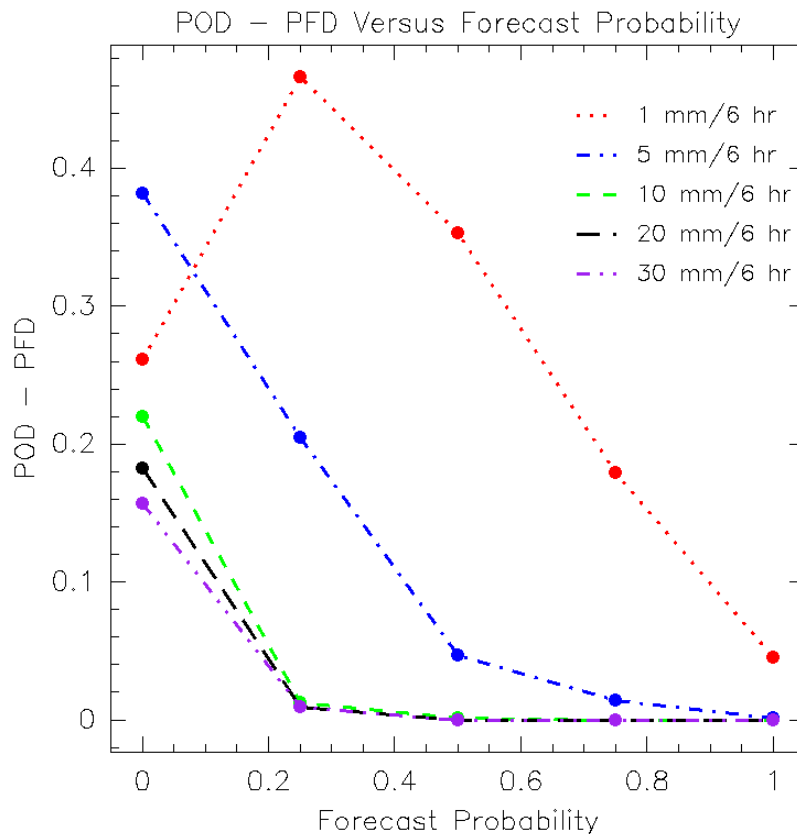

Figure 6 POD minus PFD versus forecast probability used for decision making under six different rainfall thresholds.

Relative Operating Characteristics (ROC) is basically a plot of POD against PFD [Mason & Graham 1999]. Figure 7 shows such ROC curves for various rainfall thresholds and decision probabilities. Skilful forecasts should score high in POD and low in PFD as much as possible, i.e., the ROC curves should bend towards the upper left corner of the plot. Put it another way, areas under the curves measure forecast skills. Random forecasts will lie on the diagonal line and have area of 0.5 unit. As shown in Figure 7, EPS forecasts in general subtend areas more than 0.5 unit, implying skilful forecasts under all rainfall thresholds. To delineate the change of skill levels more clearly, the ROC scores, defined as SROC = 2 ´ (area – 0.5) (see Appendix A for more details), is plotted in Figure 8 as a function of rainfall thresholds. It is evident that the skill level decays rapidly as rainfall threshold increases.
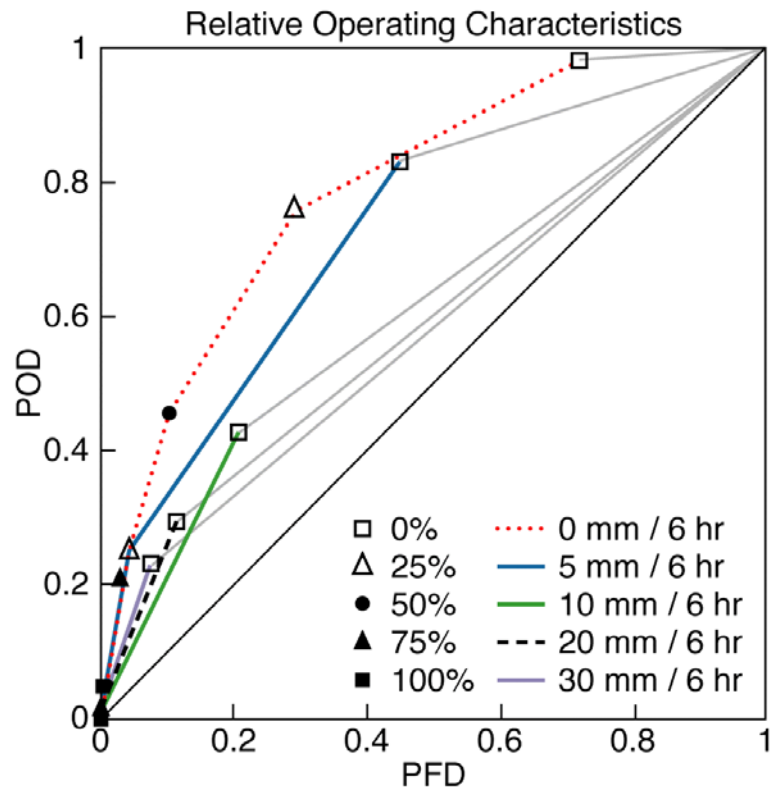
Figure 7 Relative Operating Characteristics (ROC). Areas under the ROC curves indicate forecast skills. An area of 1 means perfect forecast whereas 0.5 (the diagonal line) means forecast with no skill. Zero area implies perfectly bad forecast.
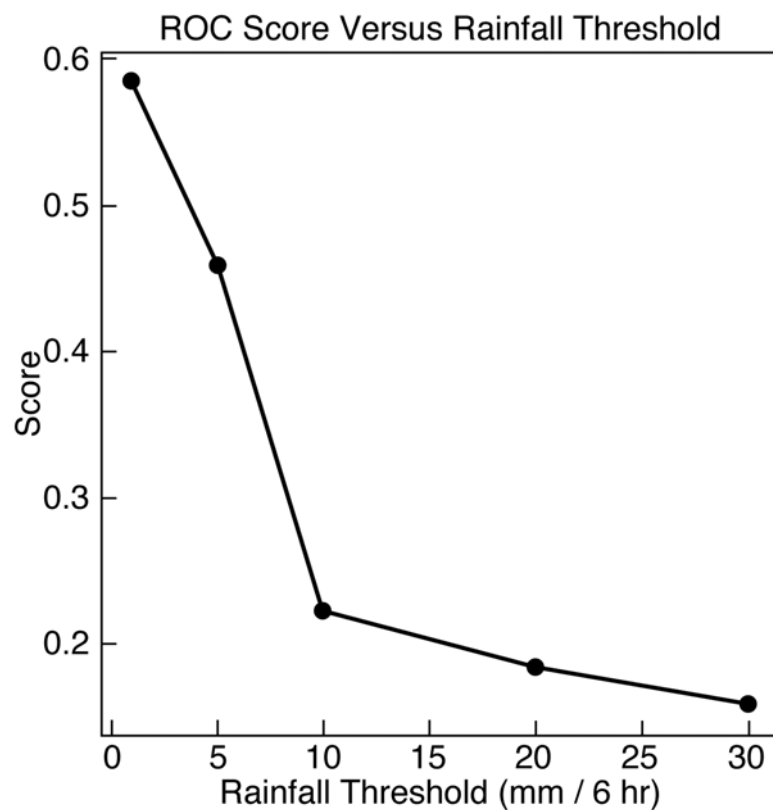

Figure 8 ROC scores versus rainfall threshold values.

Frequency Bias, or FB defined as the ratio (H+F)/(H+M), measures model bias in terms of event occurrence [Wilks 1995]. Figure 9 plots FB against forecast probability. Over-predictions, unbiased predictions and under-predictions have FB > 1, FB = 1, and FB < 1 respectively. The EPS rainfall forecasts are found to be strongly biased. They are over-predicting the occurrence at lower decision probabilities but under-predicting at higher probabilities. The cross-over, i.e. unbiased, region lies roughly around 0.45 for rain with thresholds ³ 1 mm/6 hr and in the range of 0.15–0.25 for heavier rain with thresholds ³ 5 mm/6 hr.
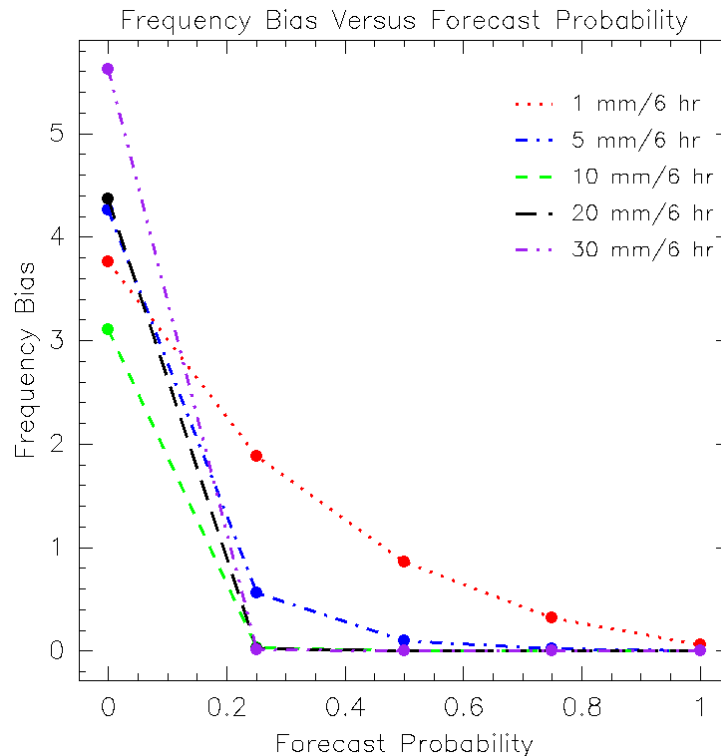

Figure 9 Frequency bias (FB) of EPS forecasts for various rainfall threshold values. FB > (<) 1 means over (under) forecast. FB of unity means unbiased forecasts.

2.2   Reliability, Sharpness and Spread

Reliability diagram plots the observed relative frequency of occurrence against the forecast probability of occurrence [ECMWF(b); Jolliffe & Stephenson 2003]. Figure 10 shows the reliability curves for the five chosen rainfall threshold values. For the rainfall threshold of 1 mm/6 hr, EPS systematically over-predicts the probability of occurrence. It is mostly unbiased for the 5 and 10 mm/6 hr thresholds below the forecast probability of 0.5. The large deviations from the diagonal line at the high end of forecast probability are mainly a reflection of small sample sizes. For higher rainfall thresholds, it is a common feature that EPS does not forecast high probability at all. At decision probability of 0.25, where sample sizes are still significantly large, the plots seem to signify a loss of reliability as the rainfall threshold increases.

The distribution curves in Figure 11 show the sharpness (see Appendix A for explanation) of EPS rainfall forecasts. It can be seen that the probability forecasts are not sharp for rainfall thresholds of 1 mm/6 hr. Sharpness increases for higher rain intensity but the curves only skew towards the low probability end, i.e., the non-occurrence of such events.
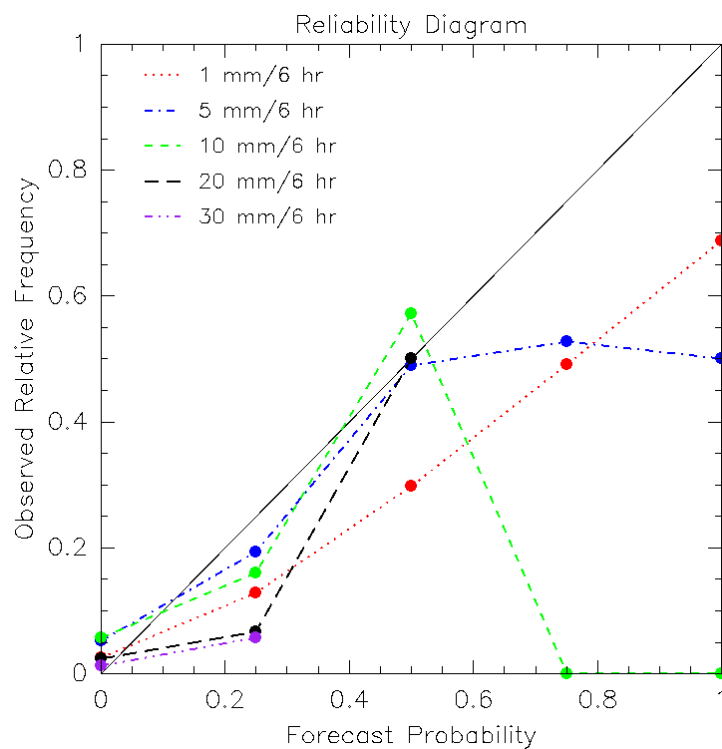
Figure 10 Reliability curves for six rainfall threshold values. Perfect forecasts will lie on the diagonal from (0,0) to (1,1). Reliability curves above (below) diagonal line indicate systematic under- (over-) forecast trends.
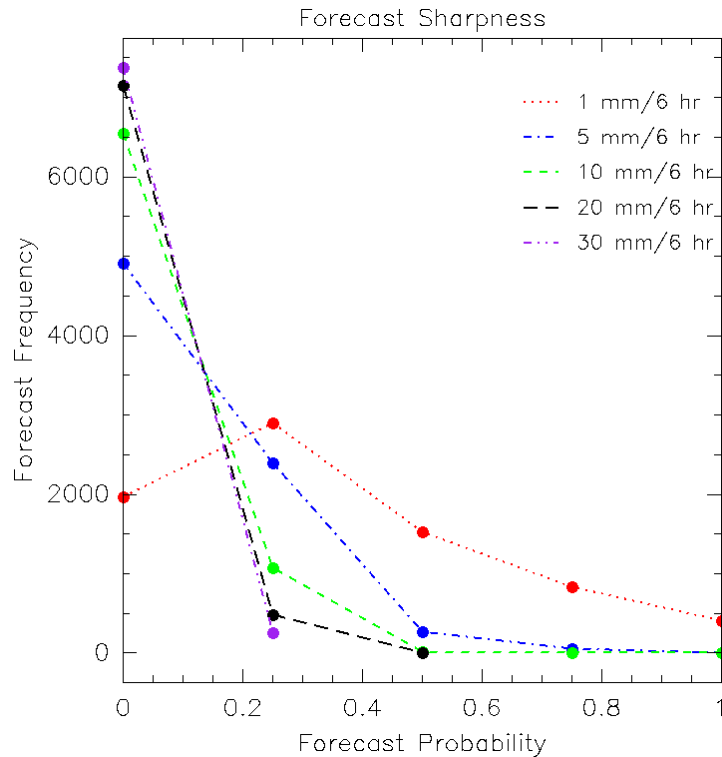
Figure 11 Plots showing the sharpness of the EPS forecast in resolving rain events. Sharp probability forecasts have frequency distributions skewed towards both low and high probabilities. In contrast, climatological forecasts will distribute mostly around 0.5 and have little frequency for the two probability extremes.

Talagrand diagram is a rank histogram for checking where the verifying observations fall with respect to the ensemble members arranged in some kind of ascending order [ECMWF(b)]. As explained in Appendix A, it is a useful tool for monitoring bias in the ensemble spread. Without detailed information from individual ensemble members, we can only generate a limited version of the Talagrand diagrams with only six bins based on the five EPS parameters. The 0th and 5th bins contain observed values below the forecast minima and above the forecast maxima respectively; the 1st bin contains the number of cases with observed values falling between MIN and the 25% quartile; the 2nd bin between 25% quartile and the median; the 3rd bin between the median and the 75% quartile; and the 4th bin between the 75% quartile and MAX. Referring to the six possible scenarios of distribution symbol in Figure 4, while zero observed rainfall cases will naturally fall to the 0th bin in scenario (f), they are all arbitrarily assigned to the 1st bin for scenarios (a), (b), (c), (d) and (e), resulting in the dominant peak seen in the unconditioned plot in the top panel of Figure 12. It also reflects that the EPS forecasts score very high in correct rejections (see Section 2.1). In the bottom panel of Figure 12, the Talagrand diagrams are conditioned on cases with non-zero observed rainfall to screen out the dominant peak in the 1st bin. The peak at the 4th bin results from the fact that in most rainy cases, it is only the top spike that can encompass actual rain intensity. There is also a substantial number of cases falling into the 5th bin, indicating that the ensemble spread does not extend quite far enough into the extreme rainfall regime.
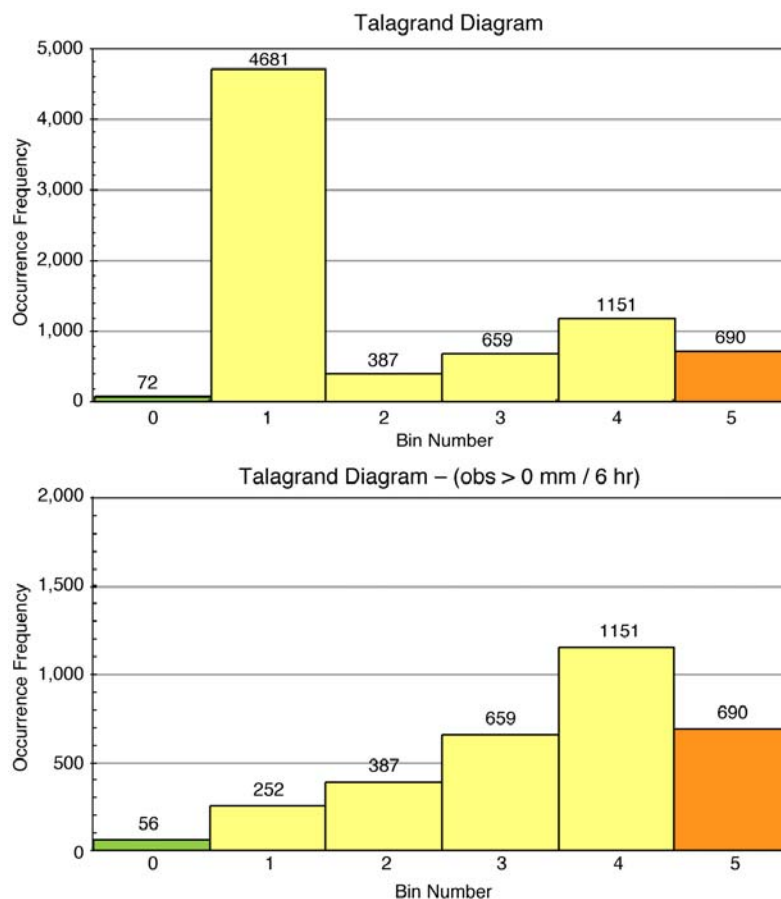
Figure 12 Talagrand Diagrams: unconditioned (top panel) and conditioned on cases with observed 6-hour rainfall greater than zero (bottom panel). The bin numbers from 0 to 5 refer to the following six ranges respectively for binning the observed rainfall Ω_{obs} : (0) $\Omega_{obs} < EPS(MIN)$, (1) $EPS(MIN) \leq \Omega_{obs} < EPS(25\%)$, (2) $EPS(25\%) \leq \Omega_{obs} < EPS(median)$, (3) $EPS(median) \leq \Omega_{obs} < EPS(75\%)$, (4) $EPS(75\%) \leq \Omega_{obs} \leq EPS(MAX)$, and (5) $\Omega_{obs} > EPS(MAX)$.

2.3   Skills and Values over Deterministic Forecasts

The Brier Scores (BS) and Brier Skill Scores (BSS) (see Appendix A for details) versus event thresholds and BS against forecast range are shown in Figure 13 and Figure 14 respectively. Figure 13 suggests that EPS rainfall forecasts are in general more skilful than their high-resolution deterministic counterparts for small amounts of precipitation up to the 10 mm/6 hr threshold, and that there are no significant differences in skill levels between the 00-UTC and 12-UTC sets of EPS meteograms. The apparently lower BS for both EPS and TL511 forecasts in the heavy rain regimes (threshold > 10 mm/hr) are due to the dominance of correct rejection as shown in the contingency table statistics (see Section 2.1) and may not be a true reflection of the model skill in such regimes. Throughout the forecast range up to 10 days (Figure 14), BS of EPS forecasts (dashed lines) are consistently lower than those of TL511 forecasts (solid lines). The differences are apparently more pronounced at lower rainfall thresholds for which the frequency of events is generally higher. BSS plots (diagram not shown) reveal that skill levels do not show significant degradations with increasing forecast range. The diurnally related periodic fluctuations observed in the BS plots in Figure 14 could well be related to both the time of occurrence of the rain events and the sample size in each category of forecasts. For example, out of the 11 cases with rainfall greater than or equal to 30 mm/6 hr, five of them occurred during 00–06 UTC and three during 12–18 UTC. As such, accuracy of forecasts validating at these two periods may be naturally lower.

Figure 15 displays the relative economic values [Richardson 2000], VR, of EPS rainfall forecasts against cost-loss ratio, C/L, for the lower rainfall threshold values. Also shown in the plots are VR of TL511 forecasts (blue solid line) and the optimal VR (grey solid line) obtained by taking the envelope of all VR curves. In general, the optimal VRs are higher than those of deterministic forecasts in all C/L regimes. VR diminishes as the rainfall threshold increases and its peak shifts towards the lower values of C/L. For thresholds of 20 or 30 mm/6 hr (diagrams not shown), the relative economic values of EPS forecasts become very small while those of TL511 appear to vanish. This is in accordance with the decreasing trend of climatological frequency of occurrence of significant rainfall.

2.4   Deterministic Use of EPS forecasts

The general impression among users is that the ensemble median usually under-predicts the actual rainfall, especially in the heavy rain regimes. In fact, as shown in Figure 3, while the ensemble median is reasonably close to observations for rainfall at 10 mm/6 hr or below, it dies out too quickly in higher rainfall regimes. The distribution of ensemble maxima suffers from the following two problems: (i) relative frequency too low under 5 mm/6 hr, and (ii) too high for more intense rain. None of these EPS parameters seems adequate to be used deterministically; the question is whether multiple, or all, EPS parameters taken together may be useful in developing some deterministic application of EPS rainfall forecasts.
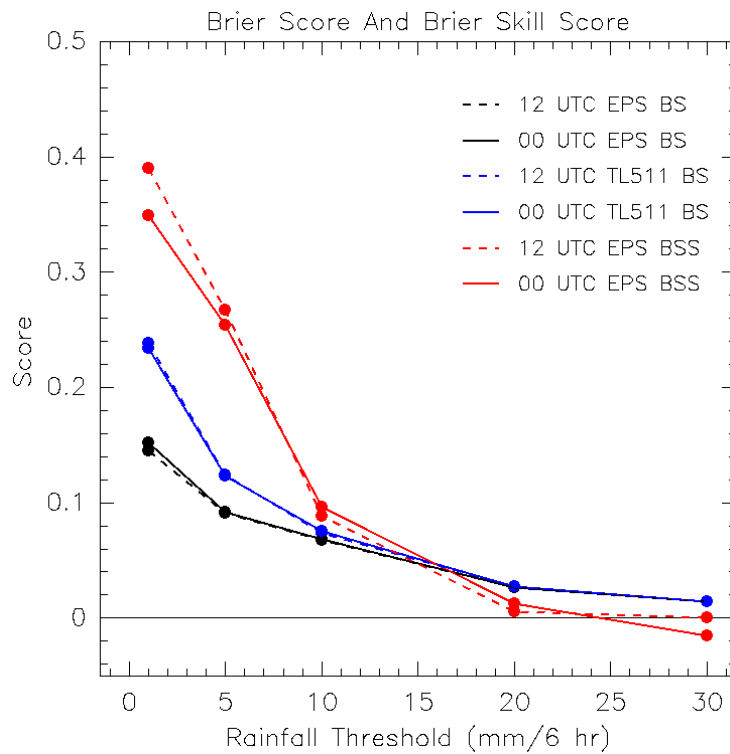
Figure 13 Comparison of Brier Scores (BS) for EPS (dark lines) and TL511 (blue lines) forecasts. The 00-UTC and 12-UTC ensembles are indicated by solid and dashed lines respectively. Brier Skill Scores (BSS) for EPS are also shown as red lines. (See Appendix A for calculation of EPS BSS).
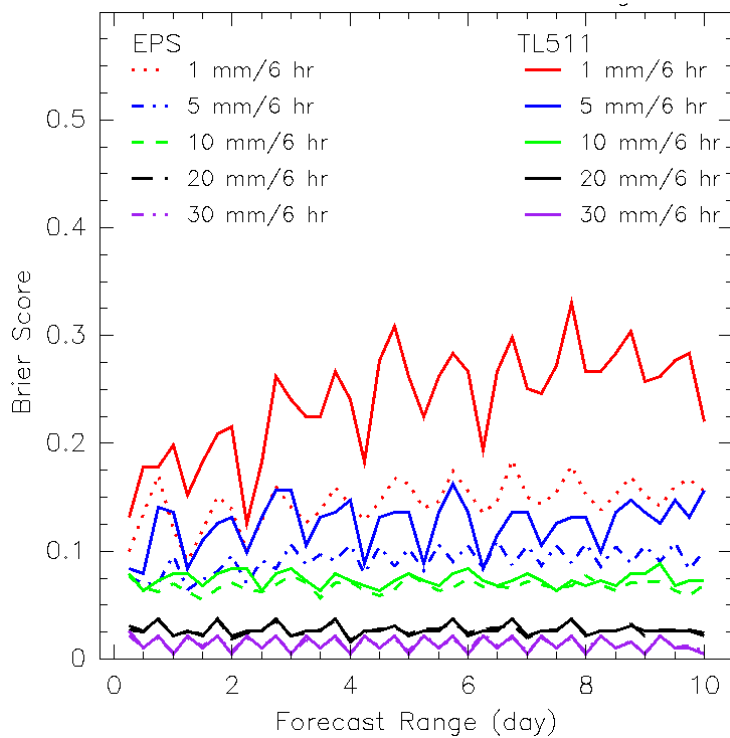

Figure 14 Brier Scores (BS) against forecast range. This plot compares the BS of EPS (broken lines) and TL511 deterministic forecast (solid lines).
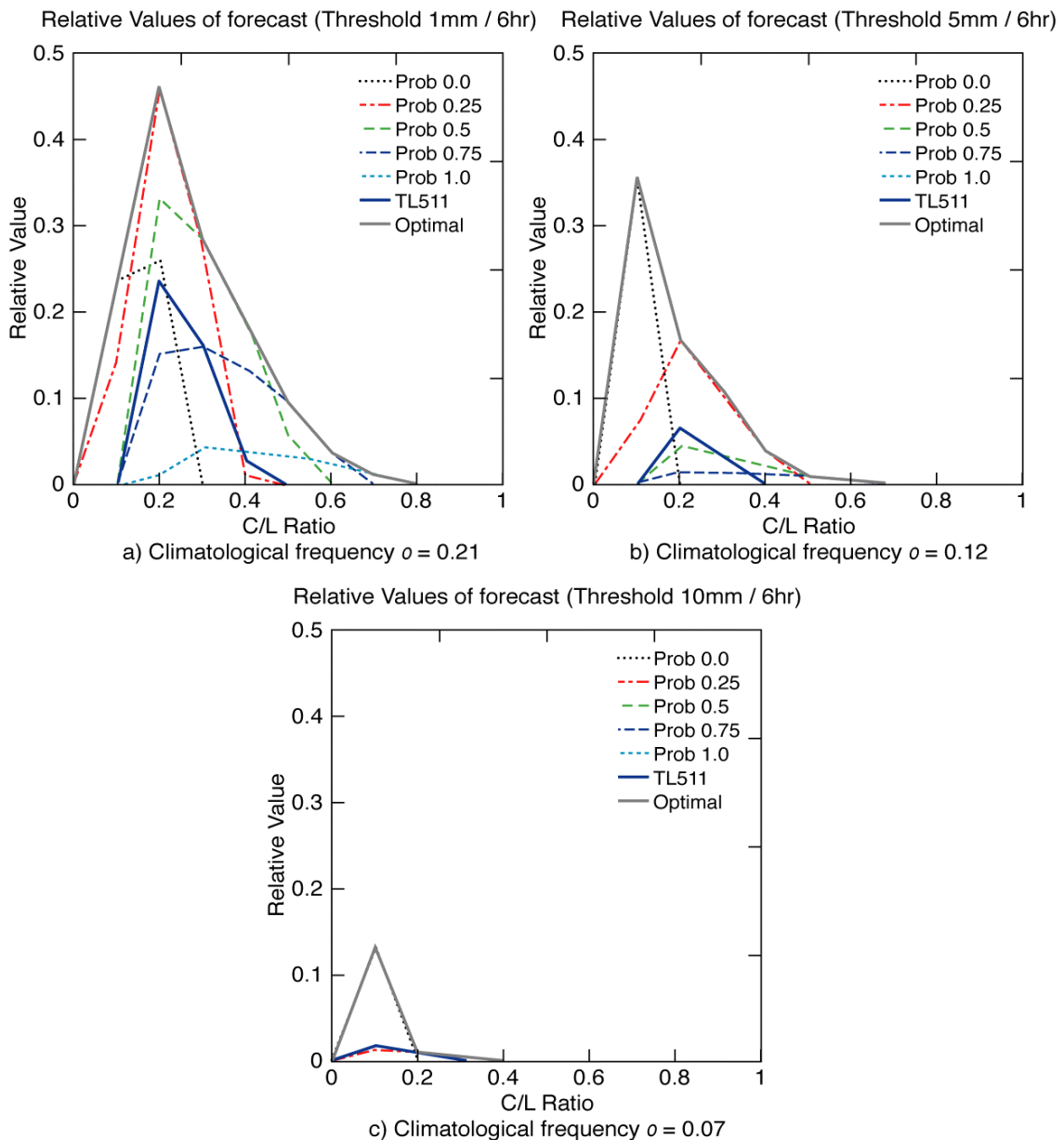
Figure 15 Panels (a), (b) and (c) show the relative values of EPS rainfall forecasts for the thresholds of 1, 5 and 10 mm/6 hr respectively. Grey solid lines are the optimal values obtained by taking the envelop of all component curves. Blue solid lines represent the relative values of the deterministic TL511 forecasts.

Figure 16 shows forecast errors versus observed and forecast rainfall by using the five EPS parameters as predictors. The forecast errors are averaged over all forecasts with the same valid time. As shown in the upper panel, the magnitude of forecast errors in general increase linearly as the observed rainfall increases. The linear pattern is particularly prominent for predictors other than the ensemble maximum. This confirms the impression that in most cases, the lower portion of the ensemble are forecasting very little or nil rainfall even in the heavy rain regimes. This also explains the systematic under-forecasting bias for observed rainfall above 10 mm/6 hr. For observed rainfall below 10 mm/6 hr, the errors of ensemble maximum behave rather differently from others. They scatter quite randomly and are mostly over-forecasting. The lower panel of Figure 16 displays the same forecast errors but with respect to the forecast values of the five predictors. The following points are noted: (a) data points are segregated into different groups according to forecast rainfall;

(b) only the ensemble maximum is capable of producing rainfall over 10 mm/6 hr; (c) the upper bound of the forecast errors is sharp and linearly increasing with forecast rainfall; (d) negative forecast errors are distributed rather randomly and do not seem to have a clear lower bound.
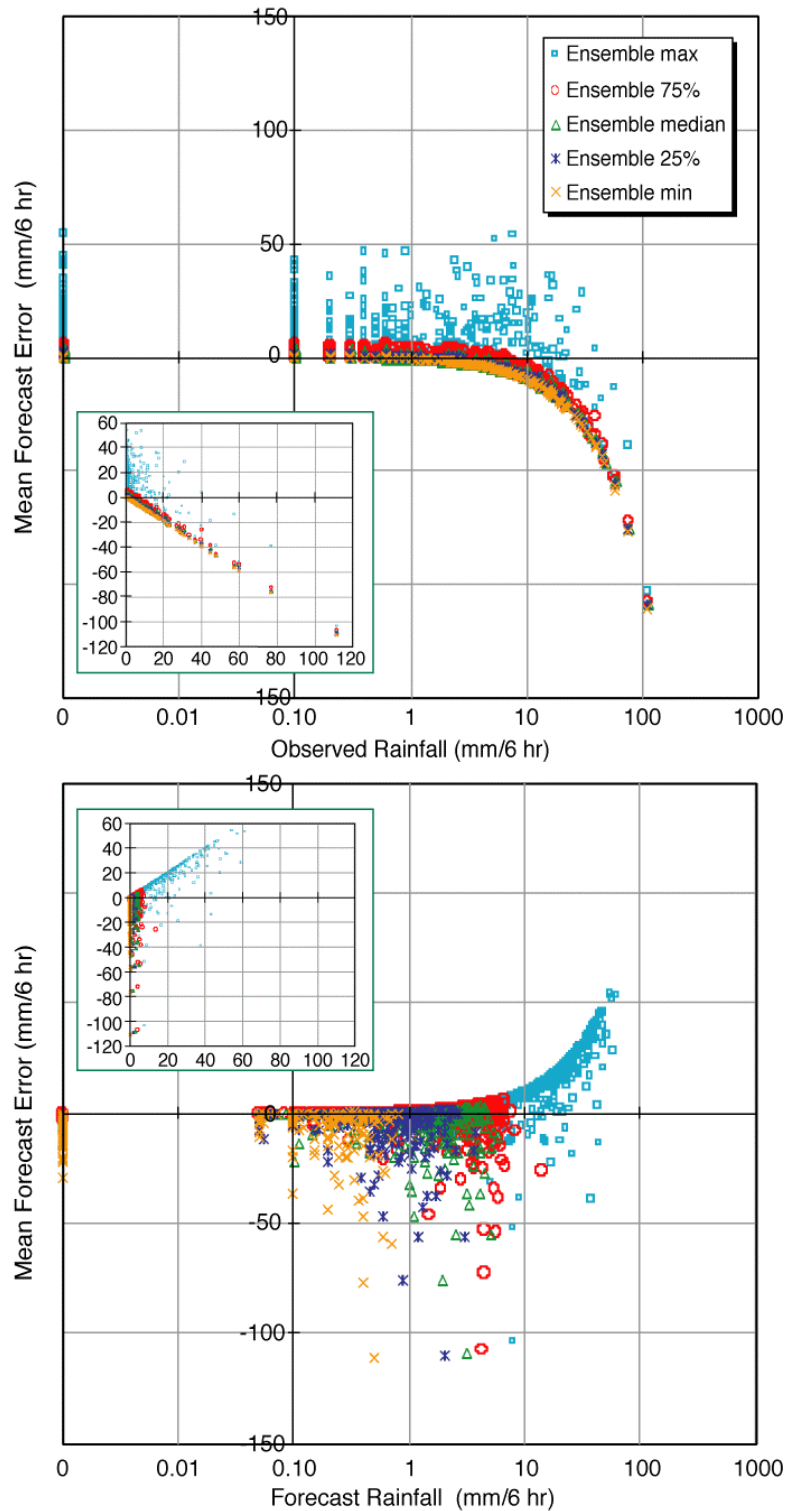

Figure 16 Mean forecast errors of ensemble parameters versus observed rainfall (upper) and the parameters themselves (lower). Insets show the corresponding plots in linear scale. Note: only cases with non-zero observed rainfall are included.

One immediate implication for using the ensemble parameters in a deterministic mode is that we have to employ different parameters for different forecast regimes. Specifically, for forecasting rain intensity over 10 mm/6 hr, the only useful predictor appears to be the ensemble maximum. The resulting error can be as big as the predicted value but the nice feature here is that the error growth is linear and predictable. Whether or not one should issue such a risky deterministic forecast will then depend on its relative value. As shown in Figure 15, the decision probability of zero is responsible for the majority of contribution in calculating the optimal relative value for the 10 mm/6 hr rainfall threshold. From a decision-maker point of view, it will still be worthwhile betting on extreme rainfall events as predicted by the ensemble maximum, provided that the cost-loss ratio is relatively low, say, below 0.2.

For predicting rain intensity less than 10 mm/6 hr, there are more choices for predictors, including the 75% quartile, median, or even the 25% quartile. The large negative errors associated with these predictors are associated with extreme rainfall predictions. Therefore, provided we do not use these three predictors for the purposes they are not supposed to serve, the forecast errors can be relatively small. As for the case of issuing extreme rainfall forecast discussed in the last paragraph, the choice of predictors among the three possible options may be facilitated by making reference to the corresponding relative economic value diagrams and the cost-loss ratio deemed affordable by the decision-maker.

3 EPS Precipitation Trends on a Day-to-day Basis in Significant Weather Events

While EPS 6-hourly quantitative precipitation forecasts may not be particularly successful in picking out the extreme rain events, an alternative approach is to assess its usefulness in terms of heavy rain likelihood on a day-to-day basis. As such, ten weather events are selected (Table 1) for verification purpose. The cases cover a wide spectrum of weather scenarios, from monsoon troughs in springtime, to tropical cyclones in summer, to depressions or mesocyclones in autumn, and to cold fronts in winter. Half of these cases have led to local rainstorm warnings being issued in Hong Kong; the other half are not as heavy but still produce rainfall values of appreciable amount. In all these cases, the weather systems involved are well anticipated based on either actual observations or numerical prognoses, or both. The remaining key question in operational forecasting is to assess on a day-to-day basis the chance of heavy rain accompanying such inclement weather systems and what is the heaviest rain that can be expected.

Case Date	Event Scenario
Case 1 - 8 Apr 03	Monsoon trough case. Rainstorm warnings issued in Hong Kong.
Case 2 - May 03	Monsoon trough case. Rainstorm warnings issued in Hong Kong.
Case 3 - 9-11 Jun 03	Monsoon trough case. Rainstorm warnings issued in Hong Kong.
Case 4 - 23-25 Jul 03	Tropical cyclone case.
Case 5 - 24-25 Aug 03	Tropical cyclone case. Rainstorm warnings issued in Hong Kong.
Case 6 - 2-3 Sep 03	Tropical cyclone case. Rainstorm warnings issued in Hong Kong.
Case 7 - 14-15 Sep 03	Case of lingering depression or meso-cyclone.
Case 8 - 10-11 Oct 03	Case of lingering depression or meso-cyclone.
Case 9 - 8 Nov 03	Cold front case.
Case 10 - 18 Jan 04	Cold front case.

Table 1 Summary of the ten selected weather events for EPS verification

To evaluate to what extent EPS products can help out in this problem, forecasters tend to look at the rainfall extremes indicated by the EPS meteograms and compare their trends or changes from one model run to the next in the hope of getting a feel on the likely confidence level. Without more specific information on the individual ensemble members, what the forecasters are effectively looking at are the magnitudes and spread of the top spikes in the precipitation distribution symbols. To mimic what the forecasters are doing in practice, we add up all the 6-hourly maximum rainfall values at the top of the spikes for the days identified under the ten weather events. As explained in Note 1 of Table 2, this is done for all forecast ranges from D+1 to D+7. The "envelope of extremes" so derived will no doubt be an over-estimation of even the worst EPS scenario predicted by the most pessimistic ensemble member. The rationale is that if the actual rainfall is even higher than these envelope figures, then we can justifiably conclude that the EPS prognoses have failed for the event.

Date	EPS(D+1)	HK	MC	GZ	AVE	MAX	SUM
8 Apr 03	47	65	33	6	35	65	104
5 May 03	24	141	142	16	100	142	299
9-11 Jun 03	230	309	250	100	220	309	659
23-25 Jul 03	365	41	15	32	29	41	88
24-25 Aug 03	135	138	56	39	78	138	233
2-3 Sep 03	434	103	94	115	104	103	312
14-15 Sep 03	227	186	138	84	136	186	408
10-11 Oct 03	62	49	42	0	30	49	91
8 Nov 03	39	33	0	1	11	33	34
18 Jan 04	41	15	14	30	20	30	59

Table 2 EPS(D+1) rainfall extremes tabulated against observed rainfall at Hong Kong (HK), Macao (MC) and Guangzhou (GZ) (see Figure 2 for locations).

Note 1: EPS value is the summation of all 6-hourly extreme forecasts (i.e. maximum value attained by the top spike in the distribution symbol) on the days concerned. EPS(D+1) means that the forecast is made on the day before the first day in the rain period; e.g. for the 3-day rain period of 9-11 Jun 03, the forecast is based on model run on 12 UTC 8 Jun 03 and covers twelve 6-hourly periods from 2 am local time 9 Jun 03 (18 UTC 8 Jun 03) to 2 am local time 12 Jun 03 (18 UTC 11 Jun). EPS(D+2) to EPS(D+7) in Table 3 carry similar meanings.

Note 2: HK, MC and GZ rainfall from mid-night to mid-night.

Note 3: AVE, MAX and SUM are respectively average, maximum and summation of HK, MC and GZ rainfall.

As shown in Figure 2, since the EPS forecast area for the land grid is slightly off to the northwest of Hong Kong (HK), we also try to incorporate rainfall data from the neighbouring cities of Macao (MC) and Guangzhou (GZ) in the verification exercises (Table 2). It should be noted that for this part of the verification, actual rainfall values from all three cities are single station data, including those for Hong Kong, which are measured at the Hong Kong Observatory headquarters (and hence not the same as the analyzed territory-wide "observed" rainfall used in Section 2). A direct comparison is then made with the EPS prediction, and the validation results and failure rates compiled. Table 3 is an example of EPS(D+1) forecasts as compared against HK rainfall. The same is applied to EPS forecasts up to the D+7 range, with

comparisons made against HK, AVE (average of HK, MC and GZ rainfall), MAX (maximum of HK, MC and GZ rainfall) and SUM (summation of HK, MC and GZ rainfall) at all forecast ranges. A summary of the failure rates is tabulated in Table 4.

Date	EPS Rainfall Forecast Extremes (see Note 1 under Table 2)						
	D+1	D+2	D+3	D+4	D+5	D+6	D+7
8 Apr 03	x		x	x			x
5 May 03	x	x	x	x	x	x	x
9-11 Jun 03	x			x			
23-25 Jul 03							NA
24-25 Aug 03				NA			
2-3 Sep 03						x	x
14-15 Sep 03			x	x	x	x	x
10-11 Oct 03			x	x	x	x	x
8 Nov 03		x	x	x		x	
18 Jan 04			x	x	x		
Failure rates	30%	20%	60%	78%	40%	50%	56%

Table 3 EPS extreme rainfall forecasts verified against HK rainfall for the ten selected weather events. "x" means actual rainfall is at least 10% higher than EPS extremes; "NA" means forecast not available.

Rainfall Type	EPS Rainfall Forecast Extremes (see Note 1 under Table 2)						
	D+1	D+2	D+3	D+4	D+5	D+6	D+7
HK	30%	20%	60%	78%	40%	50%	56%
AVE	10%	10%	30%	56%	30%	40%	33%
MAX	30%	20%	60%	78%	40%	60%	56%
SUM	70%	70%	60%	78%	70%	80%	56%

Table 4 Failure rates of EPS (D+1) to (D+7) extreme rainfall forecasts verified against HK rainfall and AVE, MAX and SUM of HK-MC-GZ rainfall for the ten selected weather events

As can be expected, because the selection of cases is strongly biased towards the Hong Kong situation, the failure rates between HK and MAX are naturally very similar. Both HK and MAX, as well SUM which produces the worst results, turn out to be inferior to AVE in terms of failure rates, confirming the notion that the EPS forecast is probably more representative of the spatially averaged situation (over a grid box of approximately 80´80 km2 in area) rather than tied to a specific location or linked to a prediction of probable maximum rainfall within the area of coverage.

In terms of forecast range, D+2 prediction offers the best results in all aspects. This could be attributed to the fact that the ECMWF EPS is set up based on the growing modes of perturbation optimized for a 48-hour period [ECMWF(a)]. In fact, if one has to do a location-specific forecast, say for Hong Kong, the best option is to put more faith on the D+2 prediction. Even then, what the EPS(D+2) forecasts can offer at best is to impose a ceiling on the likely rainfall that can be expected in association with inclement weather systems that are forecast to affect Hong Kong. In other words, if a tropical cyclone is expected to pass by Hong Kong in the next day or two and forecaster have to assess how much rain it is going to bring, then they should be wary of going for a value that is higher than the envelope figure of rainfall extremes indicated by the EPS(D+2) forecast.

4 Case Studies on High Winds associated with Tropical Cyclones

In summer, occurrence of high winds in Hong Kong is mainly due to the southwest monsoon or approaching tropical cyclones. In the latter case, the major factors affecting the wind distributions over Hong Kong include: (i) cyclone track and proximity, (ii) cyclone size, (iii) cyclone intensity and structure, (iv) local terrain effects, and (v) combined effect with other prevailing weather systems. However, it is statistically difficult to make a probabilistic verification of the EPS wind speed forecasts for Hong Kong with the present data set. All that can be done is a case-by-case comparison based on the available tropical cyclone events.

In 2003, only three typhoons, namely Imbudo, Krovanh and Dujuan, came close enough to bring gale force or higher winds to Hong Kong. Severe Tropical Storm Koni also skirted past but only strong winds were reported locally. All four tropical cyclones tracked west-northwestward across the northern part of the South China Sea. Dujuan scored a direct hit over Hong Kong and hence induced the highest winds, while the other three eventually landed to the west of Hong Kong.Figure 17 shows their best tracks as analysed by the Hong Kong Observatory, overlaid with ECMWF TL511 forecast tracks.

The dates of closest approach and hence the highest wind recorded in Hong Kong during the passages of Koni, Imbudo, Krovanh and Dujuan are respectively 21 July, 24 July, 24 August and 2 September 2003. The maximum winds predicted by EPS meteograms for those dates at the land point are extracted at forecast range from D+2 to D+10. The extracted EPS maximum winds along with the maximum winds indicated by the TL511 deterministic forecasts are tabulated in Table 5. Actual wind distributions on those dates, presented as the lowest, median and highest of wind speed values, as measured from 24 automatic weather stations within Hong Kong, are also shown for easy comparison.

| Statistics	Maximum of 10-min Mean Wind Speed at 24 AWSs (m/s)				Maximum of EPS 10-metre Wind Speed Forecasts (m/s)				Maximum of TL511 10-metre Wind Speed Forecasts (m/s)			
	Koni	Imbudo	Krovanh	Dujuan	Koni	Imbudo	Krovanh	Dujuan	Koni	Imbudo	Krovanh	Dujuan
Highest	16.7	29.9	26.8	33.5	14.0	32.1	25.0	26.9	8.2	25.4	11.9	12.0
Median	8.4	15.4	12.9	18.5	10.9	18.9	17.6	13.5	3.8	7.0	7.6	7.7
Lowest	3.6	7.3	7.6	8.0	6.2	7.7	8.9	6.0	3.2	3.3	3.5	1.9
Forecast Range (Day) 2					11.7	28.6	17.5	26.9	8.2	13.8	9.6	12.0
3					11.3	32.1	21.5	16.3	7.4	22.8	11.9	8.1
4					11.7	25.5	17.7	10.3	7.9	25.4	9.0	9.0
5					-	20.8	-	14.5	-	7.5	-	11.4
6					9.8	8.4	22.0	18.1	3.2	6.6	10.1	7.7
7					6.2	7.7	15.0	7.3	3.7	3.5	3.7	4.0
8					9.1	-	10.7	6.0	3.8	-	4.2	2.2
9					10.6	17.0	8.9	13.5	3.4	3.3	3.5	1.9
10					14.0	16.5	25.0	9.8	3.8	3.5	6.2	2.6

Table 5 Maximum 10-minute mean wind speed recorded at 24 automatic weather stations (AWSs) in Hong Kong during the passages of TC Koni (21 July 2003), Imbudo (24 July 2003), Krovanh (24 August 2003) and Dujuan (2 September 2003) and the corresponding extreme forecasts by using the EPS 10-metre wind speed and the TL511 forecasts.

As shown in Table 5, EPS prognoses definitely have an advantage over their deterministic counterparts. In practice, as long as forecasters are confident about the forecast track and can pin down the date of closest approach, EPS would be able to offer useful guidance in the estimate of the probable maximum wind that is likely to be experienced in Hong Kong. It should also be noted that EPS maximum wind forecasts at longer range are not necessarily poorer. In fact, for Koni and Krovanh, the best EPS estimates as compared to observed maximum wind were actually made at the D+10 range! Only in the case of Imbudo was the tendency of more reliable prediction at the shorter range observed; while for the direct hit case of Dujuan, only the D+2 forecast came remotely close to the ground truth. In other words, through a continuous monitoring of the EPS maximum wind forecasts starting from the D+10 range, it is still possible for
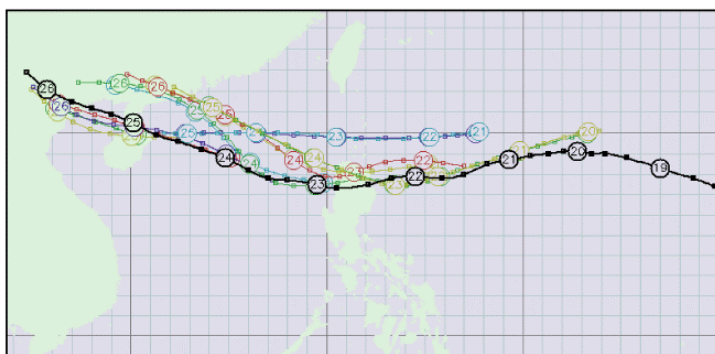
forecasters to make a reasonable guess at the longer range and to gradually refine their estimates closer to the event. Another interesting observation from Table 5 is that the EPS maximum winds over different forecast ranges appear to provide a good indication for the regional differences in wind distribution over Hong Kong.
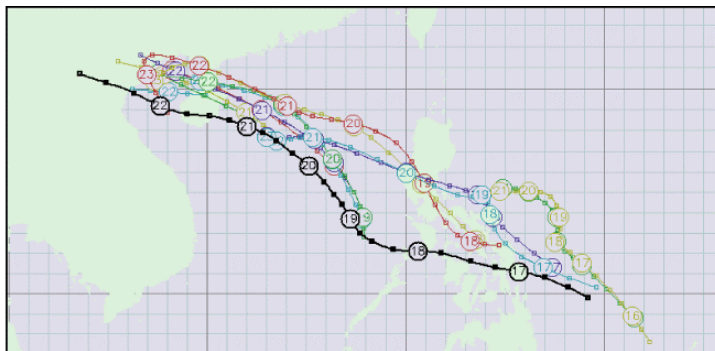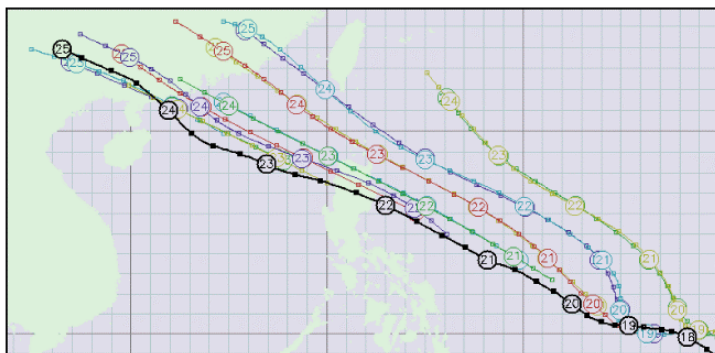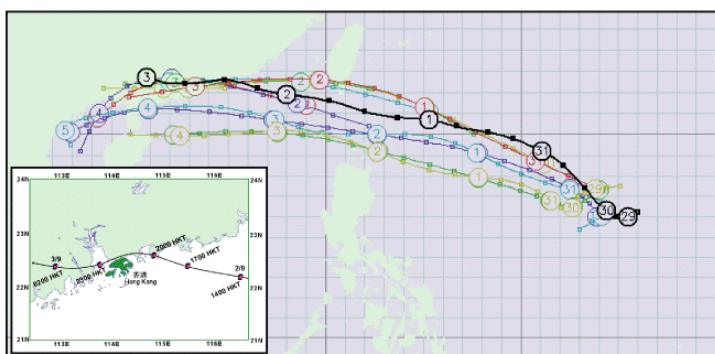

Figure 17 HKO best tracks (dark lines) and ECMWF TL511 forecast tracks (coloured lines) for tropical cyclones Koni, Imbudo, Krovanh and Dujuan in 2003. Numbers in circles indicate 00-UTC positions

While we should try not to over-generalize with the limited number of cases available, it is worth noting however that the EPS maximum wind forecasts are nearly 100% correct after categorizing the forecasts according to the following four thresholds: (i) strong winds or above (wind speed ³ 11.5 m/s), (ii) gale winds or above (wind speed ³ 17.5 m/s), (iii) storm winds or above (wind speed ³ 24.5 m/s), or (iv) hurricane force winds (wind speed ³ 33.0 m/s). As evident from Table 6, EPS high-wind predictions are correct under all criteria, except for the single missed case of hurricane force winds due to Dujuan. In comparison, the performance of deterministic forecasts as shown in Table 7 is notably inferior with many missed events.

High Wind Threshold (m/s)		S.T.S.Koni	T.Imbudo	T.Krovanh	T.Dujuan
Strong	11.5	H	H	H	H
Gale	17.5	Z	H	H	H
Storm	24.5	Z	H	H	H
Hurricane	33.0	Z	Z	Z	M

Table 6 Hit (H), miss (M) and correct rejection (Z) of EPS high wind forecasts for TC Koni (21 July 2003), Imbudo (24 July 2003), Krovanh (24 August 2003) and Dujuan (2 September 2003).

High Wind Threshold (m/s)		S.T.S.Koni	T.Imbudo	T.Krovanh	T.Dujuan
Strong	11.5	M	H	H	H
Gale	17.5	Z	H	M	M
Storm	24.5	Z	H	M	M
Hurricane	33.0	Z	Z	Z	M

Table 7 Hit (H), miss (M) and correct rejection (Z) of TL511 deterministic high wind forecasts for TC Koni (21 July 2003), Imbudo (24 July 2003), Krovanh (24 August 2003) and Dujuan (2 September 2003)

5   Concluding Remarks

In this study, the application of ECMWF EPS meteograms in the operational forecasting of rain and high winds is reviewed in the context of Hong Kong's experience. Given that the EPS forecasts are made for grid points not exactly co-located with Hong Kong and that detailed probability distribution among ensemble members and other related information such as clustering are not available, only some preliminary conclusions can be drawn as follows:

(a) the EPS guidance is in general more skilful and useful than the higher-resolution deterministic forecast;

(b) for rain forecast, EPS is essentially good at forecasting "no rain" but tends to under-predict the rainfall amount in significant rain events;

(c) the intrinsic limitations in model capability to forecast heavy rain on the finer temporal and spatial scales seem to be reflected also in the diurnal fluctuations of the skill scores and poorer EPS performances in monsoon trough scenarios which typically produce the heaviest rain;

(d) for POD against PFD consideration in most rain events, the forecast probability of 25% (i.e. EPS 75% quartile) appears in general to offer the optimal decision probability; and for deterministic prediction of heavier rain using the EPS parameters, the ensemble maximum is the only predictor of some use;

(e) for rainfall trend forecasts on a day-to-day basis, D+2 forecast of the highest probable rainfall in the course of an anticipated rain event appears to be the most reliable;

(f) for high wind assessment in tropical cyclone situations, a continuous monitoring of EPS maximum wind forecast on a day-to-day basis starting from the D+10 forecast is useful for estimating the probable maximum winds that are likely to hit Hong Kong.

Probability forecasts based on EPS data is a relative new experience to the forecasters. Through the present study, usefulness of EPS forecasts in terms of accuracy, bias and relative value for a small region like Hong Kong is better understood. Judging from the verification results alone, the rainfall information contained in an EPS forecast probably should not be applied directly to operational forecasting due to problems of bias and large RMS errors. Calibration and other post-processing will be required for both deterministic and probabilistic use. The EPS wind speed forecasts based on extreme members, meanwhile, are more promising for high-wind predictions in Hong Kong. The EPS far exceeds its deterministic counterparts in terms of performance in validation results.

As a general observation, for both extreme rainfall and high-wind forecasts, it is the extreme members in the ensemble (typically the ensemble MAX) that offer the most valuable information. Other less extreme members, perhaps the TL511 deterministic forecast included most of the time, are obviously reflecting the corresponding scenarios in less acute weather regimes. This means there is no point in applying, say, the ensemble median to predict potential torrential rain or destructive winds associated with an anticipated direct-hit typhoon. At the moment, there is not sufficient information from the EPS meteograms to enable the forecasters to meaningfully assess the probability or relative likelihood of different weather scenarios. A 5% chance as against a 10% chance for a certain extreme weather event may lead to very different warning and emergency response strategies. Until such time when more detailed EPS information or full ensemble members' data are made available, forecasters would still have to rely on alternative guidance, e.g. other NWP models, expert systems, synoptic and mesoscale analyses, as well as experience, to first work out the most likely weather scenario, and then to refer to the EPS meteograms for the probable rainfall or winds that can be expected arising from this adopted scenario.

Acknowledgement

References

Published

Jolliffe, I.T. and D.B. Stephenson (Ed.), 2003: Economic Value and Skill, Forecast Verification – A Practitioner's Guide in Atmospheric Science, Wiley.

Mason, S. J. and N. E. Graham, 1999: Conditional Probabilities, Relative Operating Characteristics, and Relative Operating Levels, Weather and Forecasting, 14, 713–725.

Richardson, D., 2000: Skill and Economic Value of the ECMWF Ensemble Prediction System, Quarterly Journal of the Royal Meteorological Society, 126, 649-668.

Wilks, D.S., 1995: Statistical Methods in the Atmospheric Sciences, Academic Press.

Online

ECMWF(a):

http://www.ecmwf.int/products/forecasts/guide/The_Ensemble_Prediction_System_EPS.html

ECMWF (b):

http://www.ecmwf.int/products/forecasts/guide/Verification_of_probabilistic_forecasts.html

BOM:

http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_pages.html

Appendix: Calculation of Verification Measures

For reader's easy reference, this appendix summarizes all the formula used in this report. It is essentially an extract from various reference listed in the last section.

Brier Score

Brier Score has the following definition:

$$BS = \left\langle \left[p(x) - \delta(x) \right]^2 \right\rangle$$

Like its deterministic counterpart, BS measures the mean square "error" of a probability forecast, $p(x)$, for an event x. For deterministic forecast, $p(x)$ is either 1 (x is predicted to occur) or zero (x not occurring). $\delta(x)$ keeps account of the actual occurrence of the event and takes on boolean values:

$$\delta(x) = \begin{cases} 1 \text{ if } x \text{ occurs} \\ 0 \text{ otherwise} \end{cases}$$

For perfect probability forecast, $BS = BS_{perfect} = 0$.

Brier Skill Score

Definition:

$$BSS = \frac{BS - BS_{ref}}{BS_{perfect} - BS_{ref}} = 1 - \frac{BS}{BS_{ref}}$$

where BSref refers to the BS of a reference forecast and BSperfect is zero. For the present verification, we chose the output of the high-resolution deterministic T511 model as reference forecast.

Contingency Table

	Observations		
Forecasts	Observed, 'Y'	Not observed, 'N'	Total
Forecast, 'Y'	H	F	W
Not Forecast, 'N'	M	Z	W'
Total	O	O'	N

where

$$O = H + M$$
$$O' = F + Z$$
$$W = H + F$$
$$W' = M + Z$$

and

$$N = O + O' = W + W' = H + F + M + Z$$

An event is said to occur (tagged 'Y') if the forecast or observed value is greater than or equal to a threshold value η and vice versa. H, F, M and Z are respectively the hit, false, miss and correct rejection counts.

Probability of Detection

Definition:

$$POD = \frac{H}{H + M}$$

Probability of False Detection

Definition:

$$PFD = \frac{F}{F + Z}$$

Note that *PFD* is sometimes called false alarm rate (*FAR*) in the literature. To avoid confusion with the other more common definition for *FAR*, i.e.

$$FAR = \frac{F}{H + F}$$

PFD is used throughout the article. As shown in Figure 5, both H and F can be simultaneously zero for extreme events, rendering FAR undefined. In contrast, PFD, which is conditioned on observations, is well defined and thus a preferred measure than FAR in extreme event forecasts.

Frequency Bias

Definition:

$$FB = \frac{H + F}{H + M}$$

Talagrand Diagram

Talagrand Diagram is a rank histogram for checking where the verifying observations fall with respect to the ensemble forecasts arranged in ascending order. By original construction, all the N member forecasts are ranked from the lowest to the highest predicted rainfall values and the observed values are binned into the $N - 1$ intervals between the two forecast extrema. For observations falling beyond the forecast extrema, they are put into either the 0th or the Nth bin, depending on whether they are below the forecast minimum or above the forecast maximum. There are totally $N + 1$ bins for building up the rank histogram.

For a perfect ensemble, observations are equally likely to fall into any one of the $N + 1$ bins and a flat diagram is thus expected. While a "U" shaped distribution implies that the ensemble spread is too small, a dome-shaped distribution means too large an ensemble spread is. Asymmetric distributions, such as "L" or "J"-shaped diagrams, imply that the ensemble contains bias.

To construct a Talagrand Diagram out of EPS meteograms, in which detailed information about individual member is lacking, we can only have six bins as determined by the five EPS parameters. The 0th and 5th bins contain observed values below the forecast minima and above the forecast maxima respectively; the 1st bin contains the number of cases with observed values falling between MIN and the 25% quartile; the 2nd bin between 25% quartile and the median; the 3rd bin between the median and the 75% quartile; and the 4th bin between the 75% quartile and MAX.

Relative Operating Characteristics (ROC)

For a given rainfall threshold, the POD is plotted against PFD using different probability values as the forecast criteria. For forecasts with no skill, the resulting plot will be the (0,0)-(1,1) diagonal line. ROC curves for skillful and poor forecasts will bend towards the (1,0) and (0,1) points respectively. The area under the curve is a measure of the forecast skill. Perfectly good, random, and perfectly bad forecast will have areas of 1, 0.5 and zero unit respectively. The ROC Score is formally defined as

$$S_{ROC} = 2 \times (area - 0.5)$$
$$= \begin{cases} 1 \text{ for perfectly good forecast} \\ 0 \text{ for random forecast} \\ -1 \text{ for perfectly bad forecast} \end{cases}$$

Reliability Diagram

Reliability diagram checks the agreement between forecast probability and mean observed frequency. It plots the latter against the former quantity with the range of forecast probabilities divided into K bins. For perfectly reliable forecasts, the data points should lie in the proximity of the diagonal line. Deviations from the diagonal gives the conditional bias: (a) below — overforecasting; (b) above — underforecasting.

The flatter the reliability curve, the less resolution (ability of the forecast to resolve the set of sample events into subsets with characteristically different outcomes) it has. A forecast of climatology does not discriminate at all between events and non-events, and thus has no resolution. Points between the "no skill" line and the diagonal contribute positively to the Brier skill score. The frequency distribution of forecasts in each probability bin shows the sharpness (tendency to forecast probabilities near 0 or 1, as opposed to values clustered around the mean) of the forecast. The reliability diagram is conditioned on the forecasts and is a good partner to the ROC, which is conditioned on the observations.

In practice, the following steps are taken to generate a reliability plot for a given rainfall threshold:

1. read the corresponding forecast probabilities from EPS meteograms and sort into 5 bins between 0 and 100%, namely 0-12.4%, 12.5-37.4%, 37.5-62.4%, 62.5-87.4% and 87.5-100%. For plotting purpose, "0", "0.25", "0.5", "0.75 and "1" are used respectively to label these bins.

2. denote the number of forecasts in each bin by π_i

3. for a given probability bin, count the number of cases, denoted by υ_i, with actual rainfall greater than or equal to the threshold value;.

4. calculate the observed frequency according to $OBSFREQ = \dfrac{\upsilon_i}{\pi_i}$