

Evaluation of the performance of tropical cyclone track prediction techniques

JOHNNY C.L. CHAN

Department of Physics & Materials Science, City University of Hong Kong

सारा — इस शोध पत्र में उन विधियों की समीक्षा की गई है जिनके द्वारा उष्णकटिबंधीय चक्रवात (टी.सी.) की गति का पूर्वानुमान लगाने के लिए तकनीकों का मूल्यांकन किया जाता है। विभिन्न त्रुटि मापों (पूर्वानुमान त्रुटि, व्यवस्था से संबंधित त्रुटि तथा क्रॉस ट्रैक और अलॉंग ट्रैक त्रुटियाँ) की विस्तार से चर्चा की गई है। इस प्रकार के उदाहरण भी दिए गए हैं कि अक्षांश, देशांतर, तीव्रता, तीव्रता में परिवर्तन, आकार और पहले की गति सहित उपकटिबंधीय चक्रवात से संबंधित कारकों पर आधारित पूर्वानुमान को स्तरित करके किस तरह से इन तकनीकों का आगे और मूल्यांकन किया जा सकता है। उष्णकटिबंधीय चक्रवात से संबंधित पर्यावरणीय प्रवाह को प्रस्तुत करने के लिए आनुभविक - लॉबिक - प्रकार्य (ई. ओ. एफ.) विधि के अनुप्रयोग का भी प्रस्ताव रखा गया है। चूंकि ई. ओ. एफ. गुणांक विभिन्न प्रकार के प्रवाह क्षेत्रों का प्रतिनिधित्व करते हैं, इसलिए इन गुणांकों के परिमाण को पूर्वानुमान स्तरित करने के लिए उपयोग में लाया जा सकता है। इस प्रकार से पूर्वानुमान तकनीक का पूर्ण मूल्यांकन चक्रवात से संबद्ध कारकों तथा ई. ओ. एफ. गुणांक दोनों पर आधारित विभिन्न त्रुटि मापों के विश्लेषण का संयोजन है।

ABSTRACT. This paper reviews the methods by which techniques for predicting tropical cyclone (TC) motion can be evaluated. Different error measures (forecast error, systematic error, and cross-track and along-track errors) are described in detail. Examples are then given to show how these techniques can be further evaluated by stratifying the forecasts based on factors related to the TC, including latitude, longitude, intensity change, size and past movement. Application of the Empirical-Orthogonal-Function (EOF) approach to represent the environmental flow associated with the TCs is also proposed. The magnitudes of the EOF coefficients can then be used to stratify the forecasts since these coefficients represent different types of flow fields. A complete evaluation of a forecast technique then consists of a combination of analyzing the different error measures based on both the storm-related factors and the EOF coefficients.

Key words — Climatology Persistence (CLIPER), Forecast Error (FE), Tropical Cyclone (TC), Cross-Track (CT), Along Track (AT), Empirical Orthogonal Function (EOF).

1. Introduction

During the past several decades, numerous techniques have been developed for the prediction of tropical cyclone (TC) intensity and movement, with the latter forming the overwhelming majority. This is understandable since a perfect intensity forecast is useless unless the track prediction is also accurate. For this reason, this paper will only focus on the problem of TC track forecasts.

Techniques for the prediction of TC movement range from simple persistence and climatology, to statistical models using observed data and numerically-predicted parameters, to full global numerical-weather-prediction (NWP) models (see review in Elsberry 1995). Nowadays, each TC forecast centre generally has an entire suite of such forecast techniques for TC track prediction. However, in many cases, these techniques can produce very different forecasts. A typical example of this is

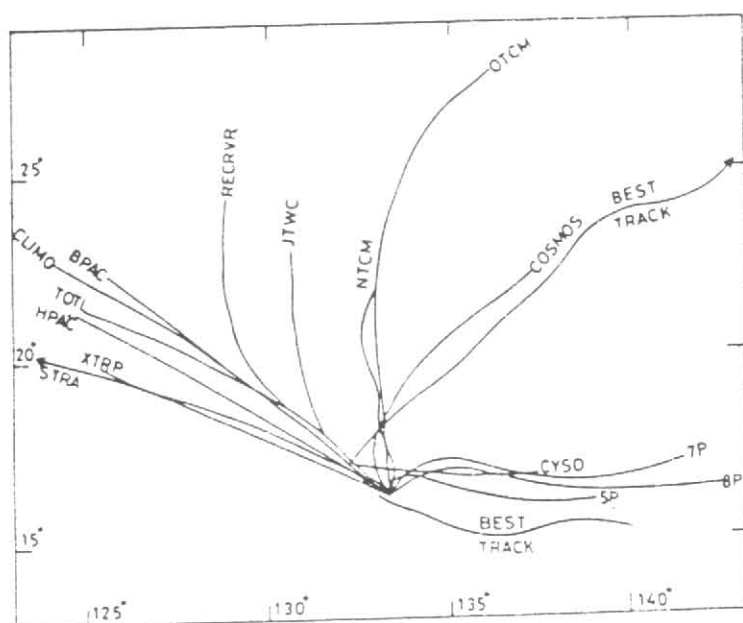


Fig. 1. An example of the array of forecasts available to a forecast centre. This one was to predict the movement of Tropical Storm Dom issued at 1200 UTC on 19 August 1983 (adapted from Tsui and Miller 1988).

shown in Fig. 1. Predictions near the point of recurvature of Tropical Storm Dom varied between a continuous northwestward motion to a complete directional reversal. Among these highly divergent forecast tracks, a forecaster must choose the one that he/she considers to be the most appropriate. The problem is : on what basis should the choice be made?

The main objective of this paper is, therefore, (i) to review previous studies on the establishment of such a basis and (ii) to propose an additional approach to evaluate the performance of a TC track forecast technique. The different error measures that have been employed to gauge the accuracy of a forecast are first discussed in section 2.

Methods for determining the performance of a forecast technique under different conditions are then presented in section 3. These methods generally consider parameters related to the TC itself. Since the movement of a TC is largely governed by its environmental flow (e.g., Chan and Gray 1982), the performance of those forecast techniques that incorporate information of such a flow should also be related to their ability to identify or predict this flow. A new approach is, therefore, presented in section 4 to address this issue. Section 5 then gives a summary of how a complete evaluation of the performance of a forecast technique should be carried out.

2. Error measures

(a) Forecast error

The most common measure of track forecast accuracy is the forecast error (FE) which is defined as the great-circle distance between the observed and the forecast positions (Fig. 2). Although the magnitude of the FE provides an indication on how far the predicted position is from the observed position, the accuracy of the forecast technique is generally gauged by comparing this error with some standard "no-skill" forecast. The generally-accepted standard is the statistical CLImatology-PERsistence (CLIPER) technique developed by Neumann (1972). This technique is considered as no-skill since it does not incorporate any synoptic data and is based purely on the information related to the past movement of the TC and its climatological characteristics. The skill S of a particular technique is then defined as,

$$S = \frac{E - E_c}{E_c} \times 100\%$$

where, E is the mean FE of the technique and E_c that of CLIPER. The more negative the value of S is, the more skillful will be the technique. A technique with $S > 0$ is considered to be not useful. Note that in calculating the mean FE for the technique and that of CLIPER, the same cases must be used because each individual forecast has its own characteristics, or "forecast difficulty" as pointed out

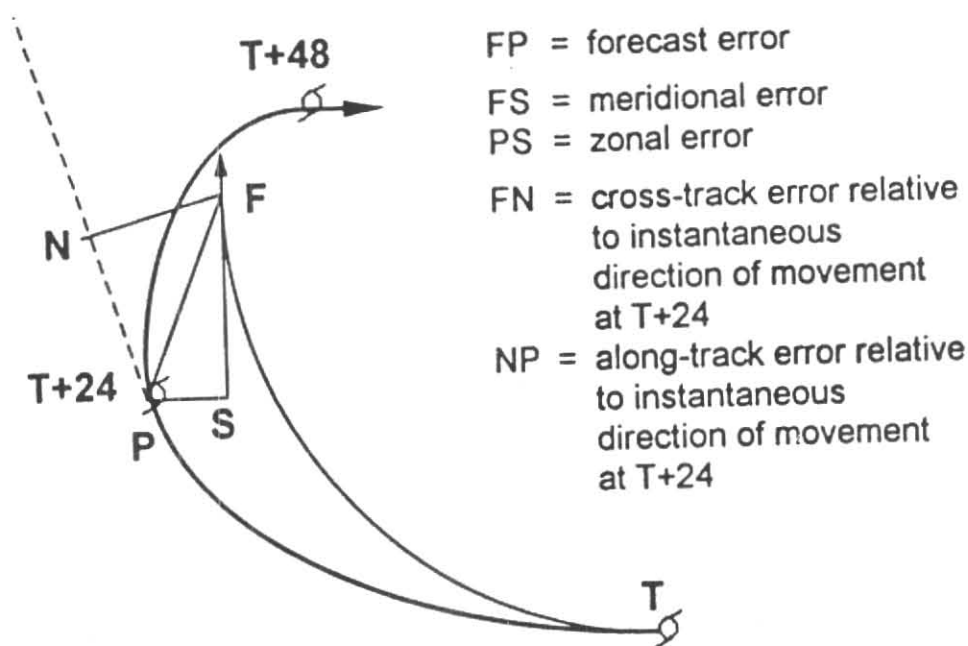


Fig. 2. Schematic showing how the forecast error, zonal/meridional error and the cross- and along-track errors relative to the instantaneous direction of movement at verifying time are calculated. T is the initial time of forecast, $T + 24$ (Point P) the best-track position 24 hours from T , $T + 48$ the best-track position 48 hours from T , and F the predicted position at $T + 24$. Heavy-solid line is the best track, the thin solid line the predicted track, and the dashed line the extrapolated line from the verifying position at $T + 24$.

by Pike and Neumann (1987). Only when the same cases are used will the calculation of the skill be meaningful. This is referred to as a homogeneous sample comparison. This concept must also be applied when comparing the accuracy of two or more forecast techniques.

(b) Systematic error

While the FE provides an idea of how far away the predicted position is from the observed (or best-track) position, and thus the accuracy of the forecast technique, it does not carry any information about the possible systematic bias in the latitudinal/longitudinal direction or along/normal to the direction of TC movement. Such information is useful not only in the understanding of the characteristics of the forecast technique (especially in the case of an NWP model), but also on how the technique might be improved, either through a modification of the technique, or a statistical correction to such a bias.

The systematic error in the zonal (meridional) direction can be calculated by taking the difference

(sign included) in longitudes (latitudes) between the observed and predicted positions and averaging the difference over a large sample (see Fig. 2). A monotonic increase or decrease of the zonal or meridional bias with time suggests that the bias may be systematic. For example, Elsberry and Frill (1980) found that the zonal (meridional) error of a regional NWP model for TC track prediction decreases (increases) with time. They then applied a backward extrapolation technique to reduce this bias and achieved better forecasts. Peak and Elsberry (1982) also applied a similar post-processing technique to another NWP TC prediction model to improve the forecasts. Since the purpose of this paper is not to discuss specific forecast techniques, these statistical correction methods will not be described in detail here. Interested readers can refer to the original papers.

It is worthwhile to point out that the existence of a systematic bias appears to be quite common in many NWP models whether or not they were designed specifically for the prediction of TC tracks. This was pointed out by Elsberry (1979, 1983) and later also found to be true for the UK Meteorological

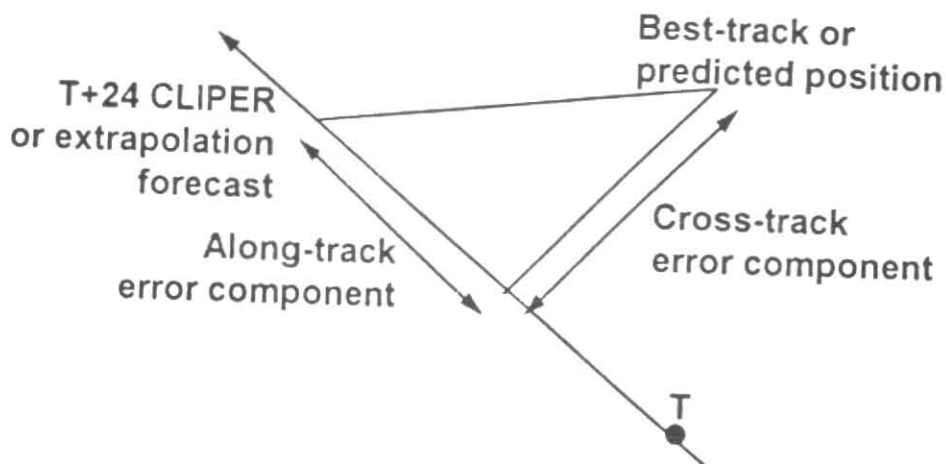


Fig. 3. Schematic showing how the cross- and along-track errors relative to either an extrapolation or CLIPER forecast are calculated. T is the initial time of forecast and $T + 24$ the forecast position at 24 h from either the extrapolation or CLIPER forecast.

Office global model (Chan and Kay 1993). Thus, in the evaluation of the performance of an NWP model for TC track prediction, possible zonal and meridional biases must be considered.

(c) Cross-and along-track error

Although systematic errors can indicate the overall bias of the technique in the zonal and/or meridional directions, the mixing of eastward- and westward-moving TCs makes the interpretation difficult. To relate to the direction of TC movement, cross-track (CT) and along-track (AT) errors can be computed. These can provide information on the direction and speed accuracies of the technique. The former is especially important in recurving situations.

The CT/AT errors are generally computed relative to the instantaneous direction of motion at the verification time (Shapiro and Neumann 1984, Tsui and Miller 1988). The CT error is the normal distance from the predicted position to the line extrapolated from the direction of motion at the verification time (see Fig. 2). The AT error is the distance between the verifying position and the point N in Fig. 2. It is obvious that the smaller the CT error, the better will be the direction forecast of the technique. However, the interpretation of the AT error is not as simple. Even if the forecast technique gives a perfect prediction of the TC speed, the AT error can be negative if the direction prediction is incorrect.

To provide an easier interpretation of the CT/AT errors, they can be calculated relative to some

standard "no-skill" forecast, such as, extrapolation (Peak and Elsberry 1986) or CLIPER (Chan *et al.* 1987). That is, the CT/AT components of both the predicted and the observed positions are calculated relative to a standard forecast (Fig. 3).

The CT components of the best track positions can then be categorized to be either to the left of, near or to the right of the standard track. A forecast position falling in the same category as that of the best track can be considered as a "correct" forecast while that with two categories off a "bad forecast". The number of cases can then be totalled to give a score. Similarly, the AT components can be stratified as faster than, near or slower than the prediction by the standard technique. For further details of this calculation, the reader can refer to Chan *et al.* (1987) and Chan and Kay (1993).

(d) Summary

Each measure discussed in this section evaluates a particular aspect of the performance of a forecast technique. To provide an overall picture of the usefulness of the technique, all these measures should, therefore, be calculated. However, because TCs can have very different characteristics and can be embedded in different types of environmental flow, the evaluation of a particular technique must also take these factors into consideration. That is, the forecasts must be stratified according to these types or characteristics. The results will then provide information as to the accuracy of the technique under a certain set of environmental or TC-related conditions. A forecaster can then use this information

to decide whether the technique can be used as guidance under a particular situation. The types of stratification will be discussed in the next two sections.

3. Stratifications based on storm-related parameters

The storm-related parameters that have been considered in the past include latitude, longitude, intensity, intensity change, size and past movement. It should be noted that these parameters are not necessarily independent of one another. For example, a TC can move poleward and westward as it intensifies. An increase in size can also occur even during the decaying stage of a TC (Merrill 1984, Weatherford and Gray 1988). Nevertheless, by stratifying the forecasts using each of these parameters, the performance of a forecast technique can be better defined. The stratification is usually done by breaking up the entire sample of forecasts into two or three categories based on the distribution of cases for a particular parameter, with about the same number of forecasts in each category. The error measures described in the last section are then applied to the forecasts in each category. Since NWP techniques are becoming more common, examples in the discussion here will be drawn from these techniques.

Stratifying by latitude generally reflects the ability of the model in predicting the movement of TCs in the easterlies or westerlies. For example, a recent study by Chan and Kay (1993) indicates that the UK Meteorological Office (UKMO) global NWP model has good skill for TCs in the Western North Pacific (WNP) north of $\sim 20^{\circ}\text{N}$ but performs worse than CLIPER for TCs south of this latitude in most cases. They attributed this result to the fact that the UKMO model predicts the midlatitude (westerly) flow better than the tropical (easterly) flow. Thus, a forecaster should not use the UKMO model forecast as guidance for TCs south of 20°N . Note, however, that this statement may not apply to the most recent version of the model because it has been modified since Chan and Kay (1993) study (see Radford 1994 and Heming *et al.* 1995).

Longitude stratifications give a different perspective on the performance of a forecast technique. In their evaluation of the performance of the Nested Tropical Cyclone Model, Chan *et al.* (1987) found that the model had good skill for TCs

in the western part of the WNP and attributed this result to the relative abundance of data in the region. However, Chan and Kay (1993) found that the UKMO predictions were quite poor for these TCs. They suggested that data quality apparently could not improve upon the problems associated with the interaction between the TC circulation and topography. Thus, such a stratification provides insight into how a model may be improved.

In general, most NWP techniques perform poorer for weak TCs (Chan *et al.* 1987, Aberson *et al.* 1993, Chan and Kay 1993). This is probably related to the ability of the model analysis in defining the TC circulation when the TC is relatively weak. Recent theoretical studies have shown that a proper representation of this circulation is crucial in predicting the TC movement through its interaction with the environmental flow (see review in Elsberry 1995). Therefore, without a good definition of the vortex circulation, the prediction may then become poor. This, in fact, has been proposed as a main source of error in some operational NWP TC track forecasts (Chan and Lam 1987, Serrano and Undén 1994, Chan 1995, Heming *et al.* 1995).

Since TCs tend to intensify at low latitudes, results from the stratifications for intensity change generally follow those of the latitude categories. This is the conclusion of Chan and Kay (1993) in their study of the UKMO forecasts.

No unique definition of TC size exists. Frank and Gray (1980) and Chan *et al.* (1987) used the radius of 15 m s^{-1} surface winds as a proxy of size while Merrill (1984) employed the radius of outermost-closed surface isobars. According to Merrill (1984), a TC generally increases in size as it moves poleward and has the largest extent shortly before its demise. Therefore, results for the size stratifications should be consistent with these of latitude (Chan *et al.* 1987).

Elsberry and Peak (1986) stratified the forecasts based on the past 12-h TC motion. They found that depending on the characteristics of the technique and its ability to represent the synoptic-scale flow (which governs to a large extent the past 12-h motion), the forecasts from the technique can have very different skills.

To summarize, stratifications of forecasts based on storm-related parameters enable a detailed

evaluation of a particular technique. A forecaster can, therefore, decide whether to use the forecast from the technique when the current TC has a certain set of characteristics. Further, developers of the technique can use this type of evaluation to improve the performance of the technique.

4. Stratification based on environmental factors

While results from the stratifications based on storm-related factors can be attributed to differences in the environmental flow (*e.g.*, westerly vs. easterly), the entire flow pattern cannot be described. In this section, it is proposed that stratifications based on different types of environmental flow may give further insights into the performance of a TC track forecast technique.

The large-scale environmental flow associated with a TC can be described by either a geopotential height field or a wind field. The best way to represent such a field is through the use of a set of empirical orthogonal functions (EOFs). Shaffer and Elsberry (1982) applied the EOF approach to project the geopotential height fields associated with a large number of TCs onto a set of eigenvectors. The EOF coefficients of the first ten eigenvectors (out of a total 120) are found to explain about 85% of the total variance of the fields. Furthermore, the values of the EOF coefficients corresponding to some of these eigenvectors show discriminatory power on the direction of TC movement. Therefore, it should be possible to stratify the forecasts from a particular technique based on the value of the coefficient associated with the eigenvectors.

Since the geopotential height gradients are generally weak in the tropics, Peak *et al.* (1986) and Schott *et al.* (1987) applied the EOF technique on the zonal and meridional winds of the large-scale environments associated with TCs. They found that the EOF coefficients can actually be used to develop statistical techniques for predicting TC movement.

Because the zonal and meridional winds are projected separately onto different EOF-spaces, it is not possible to recombine the projected zonal and meridional winds to study the vector wind field. Therefore, Ford *et al.* (1993) used relative vorticity as a representation of the environmental flow and projected the vorticity values onto an EOF space. The EOF coefficients are then used to derive

prediction equations for recurvature. More recently, Ng and Chan (1996) attempted to project the actual wind vectors onto a complex EOF space so that a reconstruction of the projected wind vectors becomes possible. They found that the first eigenvector alone can explain about 60% of the variance of the entire wind field. Furthermore, as in the case of Shaffer and Elsberry (1982), the magnitudes of the complex EOF coefficient corresponding to some of the eigenvectors relate to the direction and/or speed of TC movement.

Thus, it appears that the magnitudes of the EOF coefficients (real or complex) can be used to represent different synoptic flow patterns so that they can act as parameters for the stratifications of the forecasts. Results from this type of stratification should provide further information as to the environmental conditions under which a particular forecast technique is skillful.

5. Summary and conclusion

In operational forecasting of tropical cyclone (TC) movement, a forecaster is often presented with an entire suite of forecasts provided by a whole range of techniques, ranging from CLIPER-type to global NWP models. The forecaster must then decide on which of these forecasts can be used as a guidance under the current situation. This paper reviews the methods by which a particular forecast technique can be evaluated. Different error measures can be computed, each providing a different piece of information. The forecast error shows the absolute accuracy of the technique while the zonal and meridional error (or systematic error) gives hints of the existence of a systematic bias. Cross-track (CT) and Along-track (AT) errors calculated relative to the instantaneous direction of motion at verifying time indicates the accuracy of the technique in terms of its directionality. Computing the CT/AT errors relative to some standard technique, such as extrapolation or CLIPER, allows an assessment of the ability of the forecast technique in predicting turning and speed changes.

To determine the usefulness of a particular technique under a certain set of conditions, these error measures have to be stratified according to different storm-related factors and environmental flow patterns. The former can include latitude, longitude, intensity, intensity change, size and past TC

movement. It is also proposed that the environmental flow patterns be represented by a set of empirical orthogonal functions (EOFs) so that stratifications can be made using the magnitudes of the coefficients corresponding to these EOFs.

It is suggested that the methodology outlined in this paper should be adopted by operational centres in their evaluation of the techniques available to them. Such an evaluation provides objective criteria for the forecaster to choose the technique that has the best performance under a certain set of storm-related factors and environmental flow conditions. The weaknesses of a certain technique identified through such an evaluation can also provide input for further improvements of the technique.

Acknowledgements

The author would like to thank the Research Grants Council of the Universities Grants Committee of Hong Kong, and the US office of Naval Research for their support of the author's research in the evaluation of the performance of tropical cyclone track prediction techniques.

References

- Aberson, S. D., DeMaria, M. and Kohler, R.H., 1993, "A four-year (1989-1992) sample of a nested barotropic hurricane track forecast model (VICBAR)", *Preprints, 20th Conf. Hurr. Trop. Meteor., Amer. Meteor. Soc., Boston, MA*, 292-295.
- Chan, J. C. L., 1995, "Performance of global and regional NWP models in their prediction of Typhoon Nat (1991)", *Wea. Forecasting*, **10**, 400-410.
- Chan, J. C. L. and Gray W. M., 1982, "Tropical cyclone movement and surrounding flow relationships", *Mon. Wea. Rev.*, **110**, 1354-1374.
- Chan, J. C. L., Williams, B.J. and Elsberry, R.L., 1987, "Performance of the Nested Tropical Cyclone Model as a function of five storm-related parameters", *Mon. Wea. Rev.*, **115**, 1238-1252.
- Chan, J. C. L. and Lam, H., 1989, "Performance of the ECMWF model in predicting the movement of Typhoon Wayne (1986)", *Wea. Forecasting*, **4**, 234-245.
- Chan, J. C. L. and Kay, W. K., 1993, "Performance of the UK Meteorological Office Global Model in predicting the movement of tropical cyclones", *Wea. Forecasting*, **8**, 326-336.
- Elsberry, R. L., 1979, "Application of tropical cyclone models", *Bull. Amer. Meteor. Soc.*, **60**, 750-762.
- Elsberry, R. L., 1983, "Recent developments in tropical cyclone track forecasting", *Pap. Meteor. Res.*, **6**, 117-128.
- Elsberry, R. L., 1995, "Tropical cyclone motion. Chapter 4 in *Global Perspectives on Tropical Cyclones*", Report No. TCP-38, WMO/TD-No. 693, World Meteorological Organization, Geneva, Switzerland, 106-197.
- Elsberry, R. L. and Frill, D. R., 1980, "Statistical post-processing of dynamical tropical cyclone model track forecasts", *Mon. Wea. Rev.*, **108**, 1219-1225.
- Elsberry, R. L. and Peak, J. E., 1986, "An evaluation of tropical cyclone aids based on cross-track and along-track components", *Mon. Wea. Rev.*, **114**, 147-155.
- Ford, D. M., Elsberry, R. L., Harr, P. A. and Dobos, P.H., 1993, "Forecasting tropical cyclone recurvature. Part II: An objective technique using an empirical orthogonal function representation of vorticity fields", *Mon. Wea. Rev.*, **121**, 1279-1290.
- Frank, W. M. and Gray, W. M., 1980, "Radius and frequency of 15 m s^{-1} (30 kt) winds around tropical cyclones", *Mon. Wea. Rev.*, **108**, 219-223.
- Heming, J. T., Chan, J. C. L. and Radford A. M., 1995, "A new scheme for the initialisation of tropical cyclones in the UK Meteorological Office Global Model", *Meteor. Appl.*, **2**, 171-184.
- Merrill, R. T., 1984, "A comparison of large and small tropical cyclones", *Mon. Wea. Rev.*, **112**, 1408-1418.
- Neumann, C. J., 1972, "An alternative to the HURRAN tropical cyclone forecast system", *Mon. Wea. Rev.*, **100**, 245-255.
- Ng, W.Y. and Chan J.C.L., 1996, "Representation of the environmental wind fields associated with tropical cyclone motion using complex EOFs", *Research Report AP-96-03*. Dept. of Physics & Mat. Sci, City University of Hong Kong, 40 pp.
- Peak, J. E. and Elsberry, R. L., 1982, "A simplified statistical post-processing technique for adjusting tropical cyclone tracks," *Pap. Meteor. Res.*, **5**, 1-13.
- Peak, J.E. and Elsberry, R.L., 1986, "Prediction of tropical cyclone turning and acceleration using empirical orthogonal function representations", *Mon. Wea. Rev.*, **114**, 156-164.
- Peak, J.E., Wilson, W.E., Elsberry, R. L. and Chan, J.C.L., 1986, "Forecasting tropical cyclone motion using empirical orthogonal function representations of the environmental wind fields", *Mon. Wea. Rev.*, **114**, 2466-2477.
- Pike, A.C. and Neumann, C.J., 1987, "The variation of track forecast difficulty among tropical cyclone-basins", *Wea. Forecasting*, **2**, 237-241.
- Radford, A.M., 1994, "Forecasting the movement of tropical cyclones at the Met. Office," *Meteor. Appl.*, **1**, 355-363.
- Schott, T.B., Chan, J.C.L. and Elsberry, R.L., 1987, "Further applications of empirical orthogonal functions of wind fields for tropical cyclone motion studies", *Mon. Wea. Rev.*, **115**, 1225-1237.
- Serrano, E. and Uden, P., 1994, "Evaluation of a tropical cyclone bogussing method in data assimilation and forecasting", *Mon. Wea. Rev.*, **122**, 1523-1547.

- Shaffer, A.R. and Elsberry, R.L., 1982, "A statistical-climatological tropical cyclone track prediction technique using an EOF representation of synoptic forcing", *Mon. Wea. Rev.*, **110**, 1945-1954.
- Shapiro, L. J. and Neumann, C.J., 1984, "On the orientation of grid systems for the statistical prediction of tropical cyclone motion", *Mon. Wea. Rev.*, **112**, 188-199.
- Tsui, T.L. and Miller, R. J., 1988, "Evaluation of western North Pacific tropical cyclone objective forecast aids", *Wea. Forecasting*, **3**, 76-85.
- Weatherford, C. L. and Gray, W. M., 1988, "Typhoon structure as revealed by aircraft reconnaissance, Part II; Structural variability", *Mon. Wea. Rev.*, **116**, 1044-1056.
-