Track forecast: operational capability and new techniques - summary from the Tenth International Workshop on Tropical Cyclones (IWTC-10)

Article

Published Version

Conroy, A., Titley, H., Rivett, R., Feng, X. ORCID: https://orcid.org/0000-0003-4143-107X, Methven, J. ORCID: https://orcid.org/0000-0002-7636-6872, Hodges, K. ORCID: https://orcid.org/0000-0003-0894-229X, Brammer, A., Burton, A., Chakraborty, P., Chen, G., Cowan, L., Dunion, J. and Sarkar, A. (2023) Track forecast: operational capability and new techniques - summary from the Tenth International Workshop on Tropical Cyclones (IWTC-10). Tropical Cyclone Research and Review, 12 (1). pp. 64-80. ISSN 2225-6032 doi: https://doi.org/10.1016/j.tcrr.2023.05.002 Available at https://centaur.reading.ac.uk/111928/

To link to this article DOI: http://dx.doi.org/10.1016/j.tcrr.2023.05.002

Publisher: Elsevier

ESCAP/WMO
Typhoon Committee

Available online at www.sciencedirect.com

ScienceDirect

Tropical Cyclone Research and Review 12 (2023) 64–80

Tropical Cyclone Research and Review

www.keaipublishing.com/tcrr

Track forecast: Operational capability and new techniques - Summary from the Tenth International Workshop on Tropical Cyclones (IWTC-10)

Adam Conroy [a,*], Helen Titley [b], Rabi Rivett [a], Xiangbo Feng [c], John Methven [d], Kevin Hodges [d], Alan Brammer [e], Andrew Burton [a], Paromita Chakraborty [f], Guomin Chen [g], Levi Cowan [h], Jason Dunion [i], Abhijit Sarkar [f]

[a] *Bureau of Meteorology, West Perth, Australia*
[b] *Met Office, Exeter, UK*
[c] *National Centre for Atmospheric Science and Department of Meteorology, University of Reading, Reading, UK*
[d] *Department of Meteorology, University of Reading, Reading, UK*
[e] *Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, USA*
[f] *National Centre for Medium Range Weather Forecasting, Noida, India*
[g] *Shanghai Typhoon Institute of China Meteorological Administration, Shanghai, China*
[h] *Joint Typhoon Warning Center, Honolulu, USA*
[i] *Cooperative Institute for Marine and Atmospheric Studies, University of Miami, and Hurricane Research Division, NOAA/Atlantic Oceanographic and Meteorological Laboratory, Miami, USA*

Available online 9 May 2023

Abstract

In this paper, we summarize findings from the Tenth International Workshop on Tropical Cyclones (IWTC-10) subgroup on operational track forecasting techniques and capability.

The rate of improvement in the accuracy of official forecast tracks (OFTs) appears to be slowing down, at least for shorter lead times, where we may be approaching theoretical limits. Operational agencies continue to use consensus methods to produce the OFT with most continuing to rely on an unweighted consensus of four to nine NWP models. There continues to be limited use of weighted consensus techniques, which is likely a result of the skills and additional maintenance needed to support this approach. Improvements in the accuracy of ensemble mean tracks is leading to increased use of ensemble means in consensus tracks.

Operational agencies are increasingly producing situation-dependent depictions of track uncertainty, rather than relying on a static depiction of track forecast certainty based on accuracy statistics from the preceding 5 years. This trend has been facilitated by the greater availability of ensemble NWP guidance, particularly vortex parameter files, and improved spread in ensembles. Despite improving spread-skill relationships, most ensemble NWP systems remain under spread. Hence many operational centers are looking to leverage "super-ensembles" (ensembles of ensembles) to ensure the full spread of location probability is captured. This is an important area of service development for multi-hazard impact-based warnings as it supports better decision making by emergency managers and the community in the face of uncertainty.

Keywords: Tropical cyclones; Track forecasting; Ensembles; Track uncertainty

* Corresponding author.
E-mail address: Adam.Conroy@bom.gov.au (A. Conroy).
Peer review under responsibility of Shanghai Typhoon Institute of China Meteorological Administration.

ELSEVIER Production and Hosting by Elsevier on behalf of KeAi

https://doi.org/10.1016/j.tcrr.2023.05.002

1. Introduction

Forecasts of tropical cyclone impacts from storm surge, freshwater flooding, and extreme wind are subject to uncertainties that vary significantly from case to case and forecast run to forecast run, and to which stakeholders are sensitive in varied ways. Regional Specialized Meteorological Centres (RSMCs) value probabilistic guidance for TCs from ensemble forecast systems, but the "pull-through" of probabilistic information to operational warnings using those forecasts is slow (Titley et al., 2019). RSMCs are taking different approaches to producing probabilistic TC forecast products and using differing methods, and there is an opportunity for better coordination across RSMCs and forecast centers to develop best practices. The 2018 International Working Group on Tropical Cyclones (IWTC-9) recognized the need for improved probabilistic guidance for TCs globally and proposed several recommendations to streamline the use of ensemble forecast guidance and uncertainty information in operational forecast warnings and products (Titley et al. 2019), including the importance of a coordinated interdisciplinary approach in developing best practice.

In this paper we will give an overview over current capability in track forecasting at selected forecast centers. Some of the new techniques developed since IWTC-9 include greater use of Ensemble Prediction Systems (EPS) to express track uncertainty. Discussion will then move towards how we can better harness the available probabilistic guidance in tropical cyclone forecasting, which includes findings from Phase 1 of the Tropical Cyclone-Probabilistic Forecast Products (TC-PFP) project.

2. Methods of producing official forecast tracks

Operational warning agencies typically produce an Official Forecast Track (OFT) based on a consensus of model tracks, most commonly a combination of deterministic tracks and mean tracks from ensemble guidance. Improvements in skill from OFTs are decreasing, and in the future there is likely to be greater focus on probabilistic expression of TC impacts rather than the track itself. Nonetheless, an OFT is likely to remain a component of the tropical cyclone forecast service for some time to come. There may however be an argument to transition towards creating OFTs purely from ensemble data, particularly when this is the same data used to generate probabilistic forecasts of TC hazards.

2.1. Unweighted consensus

Most centers generate a forecast track based on an unweighted consensus of model guidance, usually a combination of deterministic models, with a growing number of centers also including some ensemble mean tracks as part of the consensus. Often there is a standard selection of consensus members selected, although sometimes the forecaster may exclude certain members, or occasionally use multiple runs of the same model (especially for those centers producing forecast tracks prior to TC formation, when not every model will have a track available).

Tropical cyclone forecasts at MetService (New Zealand) are prepared using an in-house tool called CyTrack, which allows the forecaster to create tropical cyclone tracks with a variety of model guidance and satellite imagery layered underneath. After the initial analysis, the forecast track is based on a consensus of the UK, EC and GFS deterministic forecast positions. The forecast positions from the previous forecast track are also included in the consensus, so that changes to forecast policy are made incrementally rather than a large jump when new model data is available. If the model analyses are different from the observed position, then an adjustment may also be made to the short-term track so there is no discontinuity along the track from analysis to forecast.

The Japan Meteorological Agency (JMA) mainly employs a consensus method for TC track forecasts. This approach involves taking the mean of predicted TC positions from multiple deterministic global models, including the Global Spectral Model (the JMA global deterministic model), European Centre for Medium-Range Weather Forecasts (ECMWF), National Centers for Environmental Prediction (NCEP) and United Kingdom Met Office (UKMO).

In a similar fashion, the China Meteorological Administration (CMA) typically produce a consensus track based on ECMWF, NCEP, and the CMA global and regional deterministic models. One of the component CMA models used in the consensus is a deep learning ensemble correction algorithm called TYTEC-AI (TYphoon Track Ensemble Correction-AI prediction technique). TYTEC-AI is a weighted consensus on multiple EPS. The weighting coefficient is applied to each component model (CMA, ECMWF, NCEP and UK ensembles) and is changing from year to year. TYTEC-AI has been found to outperform deterministic models when verified over the 2012–2019 period.

The RSMC La Reunion "official" track forecast is a consensus forecast, which relies on the outputs of all available deterministic numerical models and ensemble systems provided by various modeling centers (e.g. ECMWF, Meteo-France, the UK met Office, or NCEP), as well as knowledge on recent performances of each individual model.

The Joint Typhoon Warning Center (JTWC) has leveraged the consensus approach to tropical cyclone (TC) track forecasting since the late 1990s (Sampson and Schrader, 2000). An unweighted consensus tracker (CONW) consisting of nine members is currently used,[1] made up of global deterministic models and the means of global ensemble system forecasts. The JTWC consensus requires a minimum of two of the nine members be present in order to generate the CONW tracker. The Naval Research Laboratory (NRL) and JTWC annually review the performance and reliability of various models to assess the sensitivity of CONW accuracy to each member and to optimize overall accuracy of the consensus. For example, in 2019, the UK Met Office global ensemble mean (UEMN) was added to the track forecast consensus, and NCEP's HWRF model and the U.S.

[1] The nine members of CONW are GFS – NCEP, ECMWF, UKMET, GALWEM – U.S. Air Force, NAVGEM – U.S. Navy, JMA Global Spectral Model, ECMWF Ensemble Mean, GFS Ensemble Mean and UKMET Ensemble Mean.

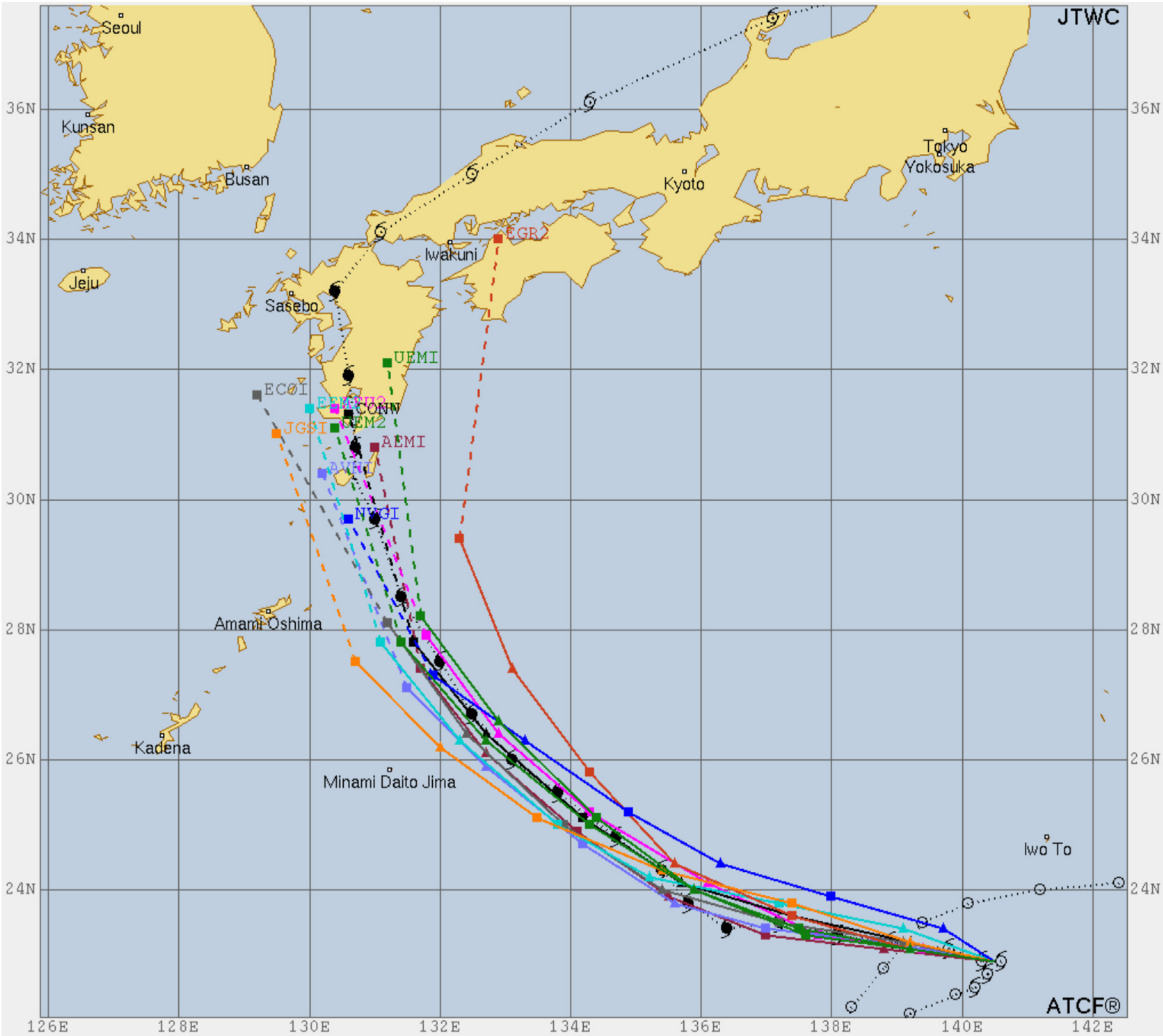

Fig. 1. Track forecasts from CONW members for Tropical Storm 16W initialized at 1200 UTC 14 September 2022 out to 72-h lead time, displayed in the Automated Tropical Cyclone Forecast (ATCF) system. The verifying track is plotted with black tropical cyclone symbols. (Image courtesy JTWC).

Navy's COAMPS-TC model were removed from the consensus. Interpolated versions of each model tracker are used when computing the consensus tracker, which shift the raw model track forecast in space and time to match the most recent best track position analyzed by JTWC. An example of the current CONW members as displayed in the Automated Tropical Cyclone Forecast (ATCF) system is shown in Fig. 1.

Forecasters at JTWC use the track forecast consensus (CONW) as a reasonable starting point for producing a TC forecast. This is done by displaying CONW and its members in a similar fashion to Fig. 1 within ATCF, allowing the forecast spread and the individual model predictions to be analyzed visually. After examining the guidance envelope, the forecaster then selects forecast positions from the map, taking into account known model biases and the previous forecast.

The Australian Bureau of Meteorology (ABoM) creates an unweighted consensus of a standard set of NWP models as a basis for the Official Forecast Track.[2] The consensus is created by a vector motion average of the component models at each timestep, which reduces discontinuities in the track as model availability changes. Once a consensus is created, the forecaster may make minor changes, particularly to bias the first 12 h of the OFT to recent motion.

Some customers have a requirement for an OFT prior to the formation of a tropical system. These are typically generated in a similar fashion to the normal OFT process, with the key differences being that the forecast starts at a future timestep

[2] The standard members of the ABoM consensus (subject to availability) are: ACCESS-G deterministic (Australian global model), ACCESS-G ensemble mean (2022 upgrade AG4 is the preferred model), ECMWF deterministic, EC ensemble mean, GFS deterministic, GFS ensemble mean, JMA deterministic, UK deterministic and UK ensemble mean.

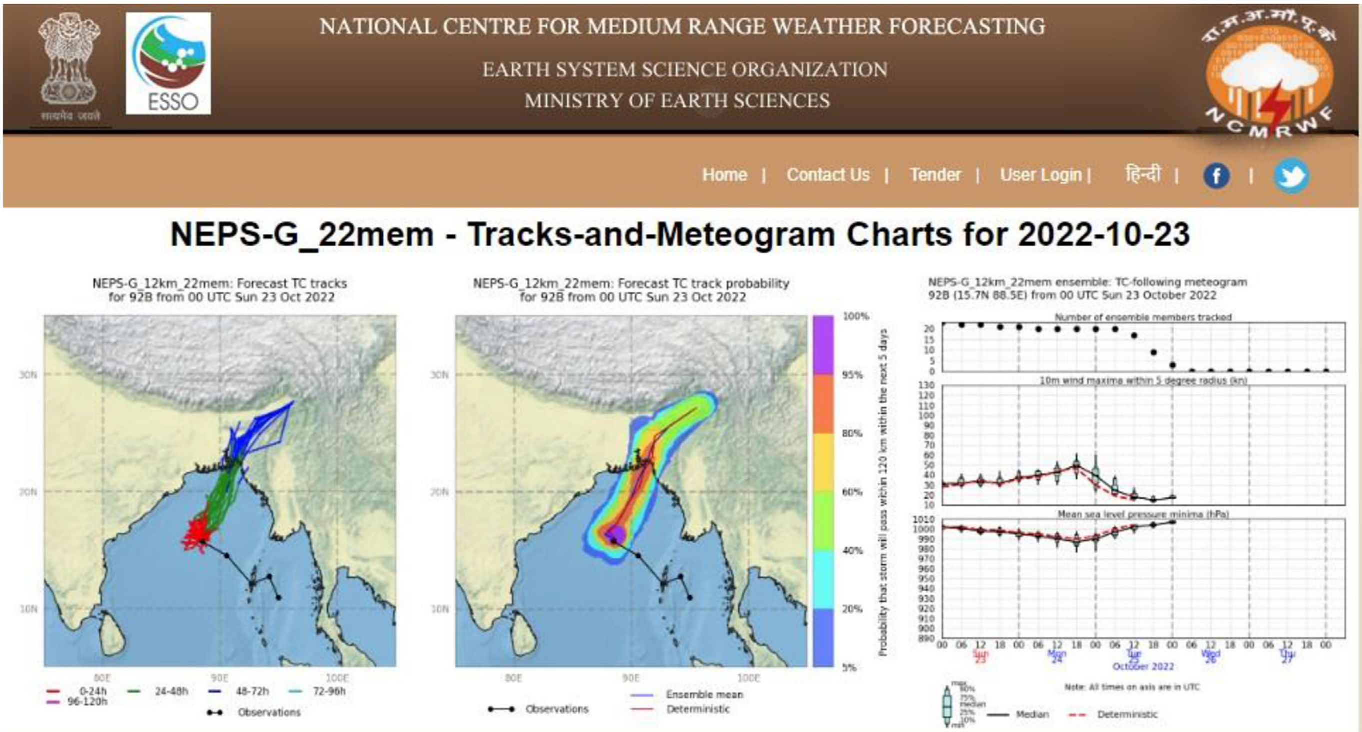

Fig. 2. Sample of a tropical cyclone forecast from NCMRWF (India).

(with no analysis position), and with less available guidance to generate the track from.

2.2. Weighted consensus

A weighted consensus track is one where a weighting coefficient is applied to each component model in the consensus, usually based on past performance. While this can increase the skill of the forecast track, it can require a large maintenance overhead, as whenever significant changes are made to the model, there needs to be a recalculation of the weighting coefficients. This usually needs to be done each season and can even be updated during the season. Because of this overhead, a weighted consensus is not feasible for some centers.

Deterministic track and intensity forecasts produced by the National Hurricane Center in the Atlantic and East Pacific basins come from an ensemble of inputs in a corrected consensus method. The Hurricane Forecast Improvement Program (HFIP) Corrected Consensus Approach (HCCA) technique produces skillful track and intensity forecasts utilizing up to 14 dynamical or statistical model forecasts to produce a consensus output (Simon et al. 2018). As part of this approach the models are weighted based on past performance and the error statistics are updated during the season, revealing expected performance of models as well as the uncertainty between skillful models.

2.3. Use of ensemble mean positions in track forecasting

While not common, there are forecast centers that present tropical cyclone forecasts which use purely ensemble data, including for the forecast track itself. The National Centre for Medium Range Weather Forecasting (NCMRWF – India) have moved towards displaying an ensemble average track for their tropical cyclone forecast tracks. The ensemble mean positions are calculated from the NEPS-G ensemble (the global ensemble prediction system of NCMRWF). This ensemble has a horizontal resolution of 12 km, and consists of 23 members (1 control, 11 perturbed on the most recent run, and 11 lagged members from the run 12 h previous). Systems are tracked using the bi-variate method adopted by the Met Office (UK). Fig. 2 shows a sample of a forecast issued on 23 October 2022. As well as the track, the forecast also shows a probability of location, and meteograms depicting the ensemble spread of 10 m wind and MSLP. The deterministic track shown on this graphic is NCUM, which is the global unified model of NCMRWF.

There remains debate as to whether a forecast track should even be produced as part of TC warning services, as it can draw focus to a single scenario, when the focus should be shifting to considering the range and likelihood of possible tracks. However, it is likely that forecast tracks will continue to be produced for some time yet, given how deeply embedded they are in TC forecast services. As more centers move towards using ensemble spread to generate a cone of uncertainty, or a depiction of forecast confidence, further consideration should be given to the suitability of using this ensemble guidance for generating the official forecast track. At ABoM, some comparisons have been made between super ensemble means and official forecast tracks (a “super ensemble” is defined as a combination of multiple ensemble runs, either of different ensemble models, different ensemble run times, or a combination of both. Investigations are ongoing but early results are promising. Further work should be done to compare super ensemble means to official forecast tracks on a normalized, global dataset, and with the inclusion of more ensemble prediction systems. An ensemble mean having comparable skill to existing methods would also bring with it the benefits of being more stable, and being consistent with probabilistic outputs from the same ensemble dataset.

3. Skill trends in official forecast tracks

Landsea and Cangialosi (2018) posed the question, “have we reached the limits of predictability for tropical cyclone track forecasting?” Errors in track position had decreased by two thirds in the Eastern North Pacific and North Atlantic basins during the period from 1990 to 2016. This was primarily due to “global modeling advances, data assimilation improvements, dramatic increases in observations primarily derived from satellite platforms, and use of ensemble forecast techniques”. It was pointed out that these improvements cannot continue indefinitely, and that there exists a “limit of predictability” in the atmosphere due to chaotic growth of errors with lead time.

Fig. 3 shows track errors in the North Atlantic and Eastern North Pacific relative to a climatology and persistence model (CLIPER5), which is used to normalize errors and account for track variability. This shows that overall there has been very little improvement in track forecasting skill from 2012 to 2016 in these basins. “The sample of 5 years is too short for meaningful statistical significance testing, but it does suggest there has been a slower rate of improvement or perhaps no additional advances” during this period.

Whilst it is difficult to determine when we will reach the limits of predictability (or indeed, whether we are already approaching it), knowledge of this will be critical to all users of tropical cyclone track forecast information. Regardless, as the rate of improvement in track forecasting slows, Landsea and Canglialosi suggest the emphasis should change to providing well-calibrated probabilistic forecasting information for tropical cyclone impacts. This includes finding new and improved techniques for characterizing track uncertainty (discussed in next section) and for generating probabilistic guidance for tropical cyclone impacts such as wind, storms surge, and rainfall. Ultimately, a full understanding of the spatial and temporal risk of these impacts will be of most importance to improving decision making.

A similar slowing improvement of track forecast performance since 2014 is evident in the western North Pacific. Fig. 4 shows trends in mean position error (PE) for five forecast centers in the western North Pacific (CMA, JMA, JTWC, KMA – Korea Meteorological Administration, and HKO – Hong Kong Observatory). Fig. 5 shows similar plots for global and regional NWP track forecasts. Both these figures show very little improvement in skill since 2014.

If the 96 and 120 h PEs are included (see Fig. 6), then it appears there are still some better skill improvements being made at longer lead times. Yu et al. (2021) suggest that this improving trend may continue for some time yet, particularly as improved analysis of position and structure of TCs result in better initialization in NWP guidance, as well as reduced initial position errors in forecast tracks feed into lower errors for the duration of the forecast track. However, it should be noted that verification scores routinely show ECMWF outperforming other global models for track despite not using synthetic observations, and hence having larger position errors at analysis time.

4. Characterizing track uncertainty

In the past, the uncertainty in tropical cyclone location has most commonly been based on fixed values such as the historic forecast error. However, a static area like this fails to take into account any situationally dependent information. With the increased availability and skill of ensemble data, it is possible to more accurately calculate the probability of tropical cyclone location through the forecast period (Titley et al., 2020). Being able to represent this probability accurately not only increases the value of the track forecast, but also can feed into other impact-based products that rely on the location of the tropical cyclone. This will ultimately enable users of the forecast products to make better decisions.

The area that depicts track uncertainty has typically been referred to as an uncertainty area or a cone of uncertainty. However, as there is a move towards using dynamical methods to determine track uncertainty, it makes sense for the name of the area representing this uncertainty to match the method. For a forecast timestep, there exists a probability distribution of where the system will be located at that time (a probability of location). The aim of using dynamic methods is to attempt to create an areal representation of this probability distribution at a pre-selected confidence level. For example, if you select a confidence level of 80%, then the system would be expected to be located within that area 80% of the time. A suitable name for this expression of uncertainty at the particular forecast time is the ‘Forecast Confidence Area’ (FCA). And if we sum together all the FCAs from the analysis time to that particular forecast time, the swath created would be the ‘Forecast Confidence Cone’ (FCC). Not only does this express the uncertainty with more positive language, the terminology is consistent with the method used to generate it.

4.1. Statistical methods

For centers that use a statistical method in generating a spatial depiction of the track uncertainty, the area is usually based on an average forecast error over a number of previous seasons of forecasts produced by the center. This has the advantage of being predictable and easy to calculate, but it misses out on capturing situationally dependent information. Many centers use this method, although some do allow for forecaster manipulation to take into account model guidance (discussed in ‘Statistical-dynamical methods').

The Joint Typhoon Warning Center (JTWC) takes a slightly different approach to other centers. Fig. 7 shows an

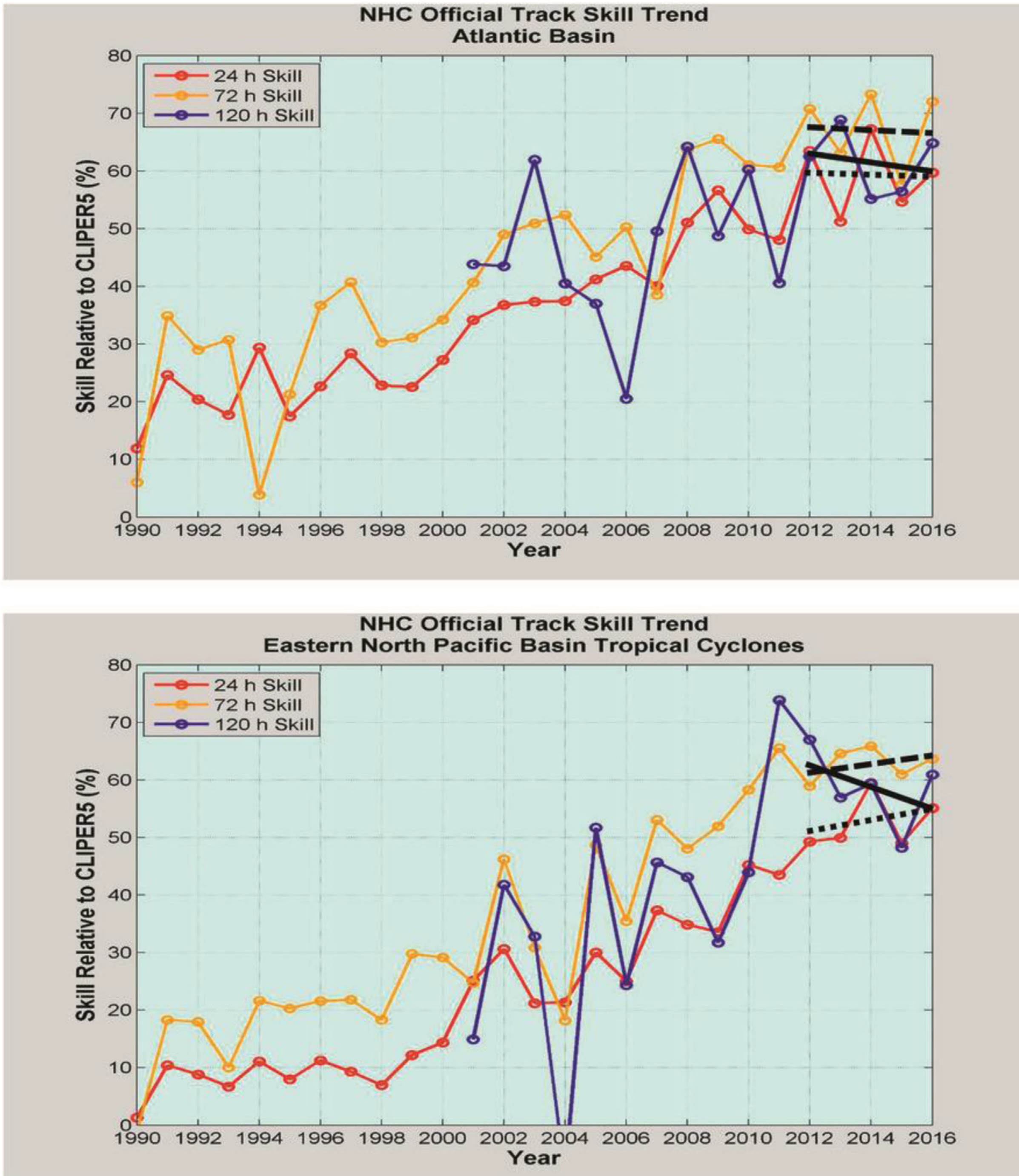

Fig. 3. 1990–2016 track forecast skill (improvements relative to CLIPER5) from the NHC for (top) the Atlantic basin and (bottom) the eastern North Pacific basin for the 24- (red), 72- (yellow), and 120-h (blue) predictions. The 2012–16 best fit linear trend is appended for each time series in black: 24 h, dotted; 72 h, dashed; and 120 h, solid. (Bulletin of the American Meteorological Society 99, 11; 10.1175/BAMS-D-17-0136.1).

example of a forecast graphic produced by the Joint Typhoon Warning Center – JTWC (USA). The TC forecast graphic includes a hatched region known as the "danger area" or "ship avoidance area." This is defined at each forecast point as a circle centered on that point with a radius equal to the JTWC 5-year mean track forecast error plus the maximum forecast radius of 34 kt winds at that point. This region functions as an error swath within which gale-force winds may occur, based on historical track error. It does not currently account for uncertainty or asymmetry in the wind radii forecast, or for asymmetry in track forecast uncertainty. As a result of adding the forecast gale radius to the climatological area, the swath appears larger than a climatologically based cone such as that used by the National Hurricane Center (NHC – USA), even though it is based on a similar level of uncertainty.

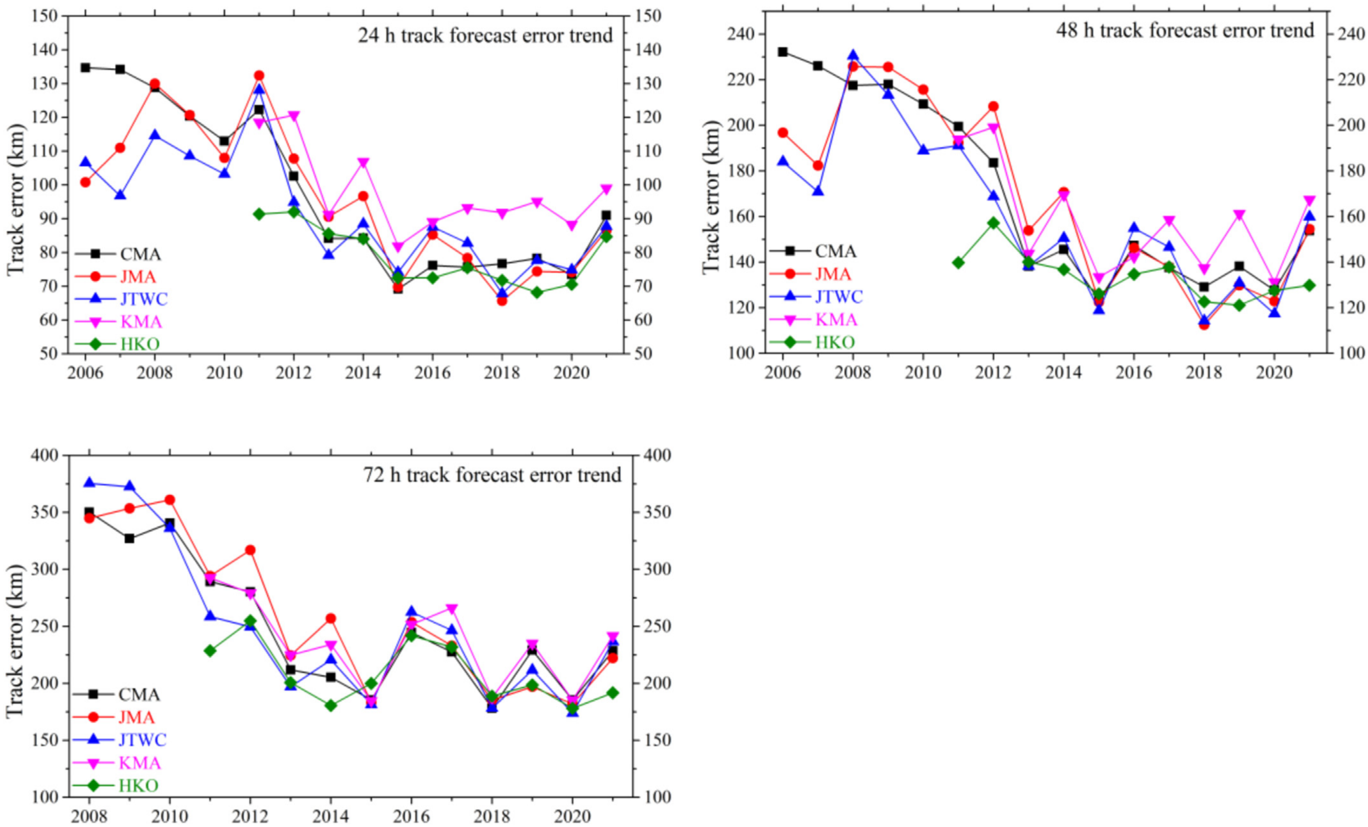

Fig. 4. Position error trends for each official typhoon prediction agency at the lead times of 24, 48, and 72 h (Figure courtesy of Chen and Yang (2022)).

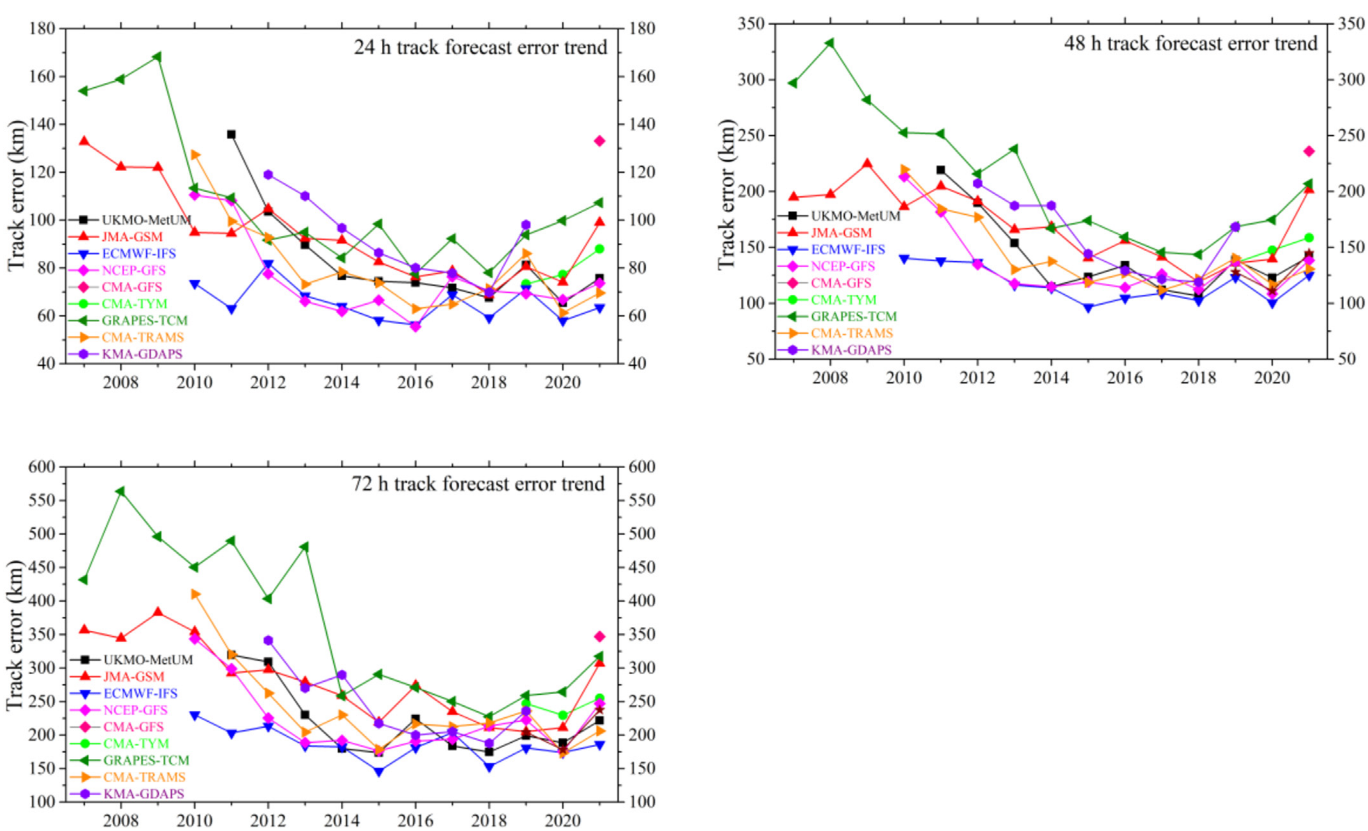

Fig. 5. Position error trends for global and regional NWP models from 2007 to 2021 at the lead times of 24, 48, and 72 h (Figure courtesy of Chen and Yang (2022)).

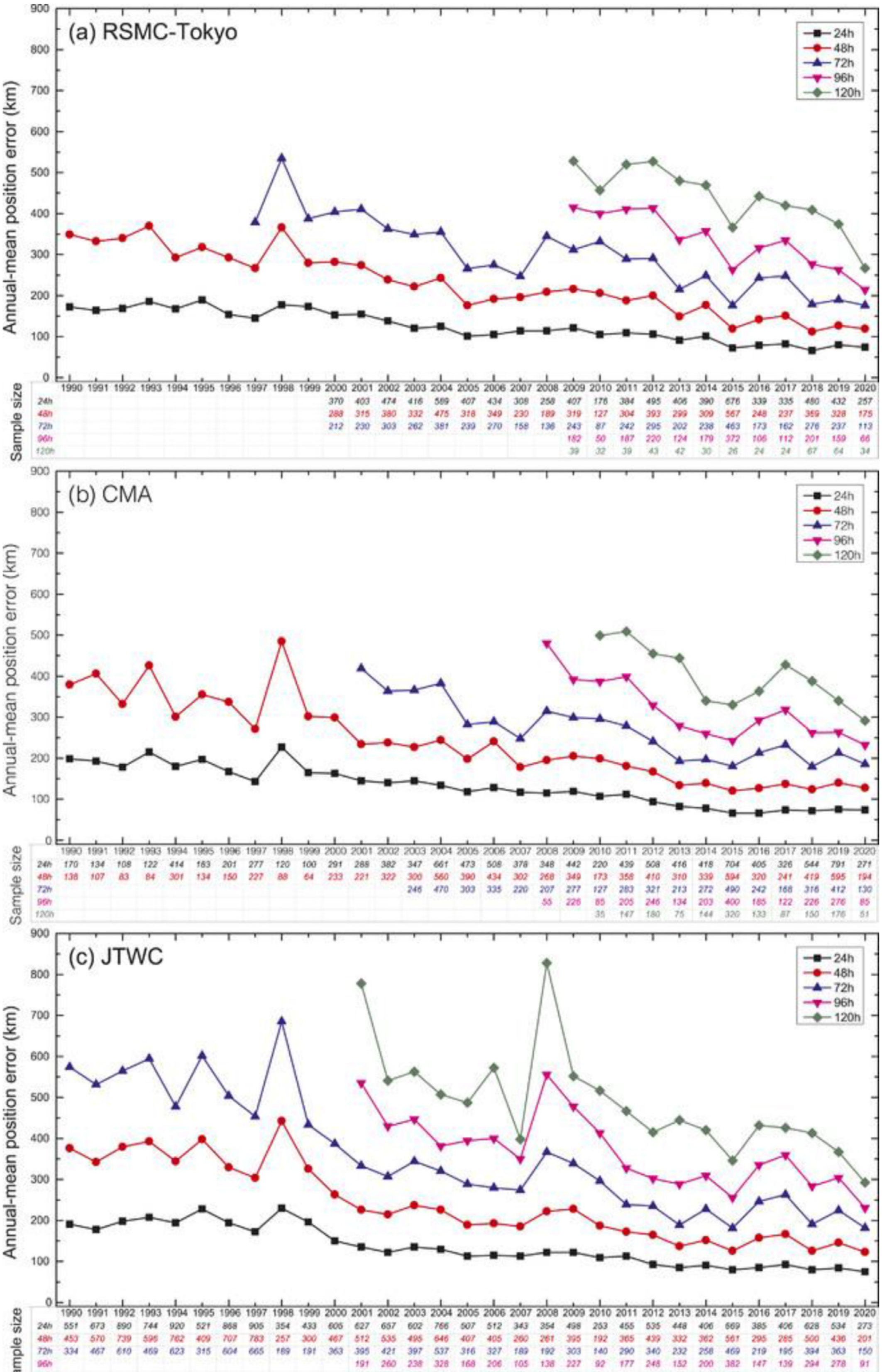

Fig. 6. The annual-mean position error (km) of TC track forecasts from 1990 to 2020 at (a) RSMC-Tokyo, (b) CMA, and (c) JTWC. The sample sizes are listed at the bottom of each chart. Note that the sample size is not available for RSMC-Tokyo before 2000. (Figure courtesy of Yu et al. (2021)).

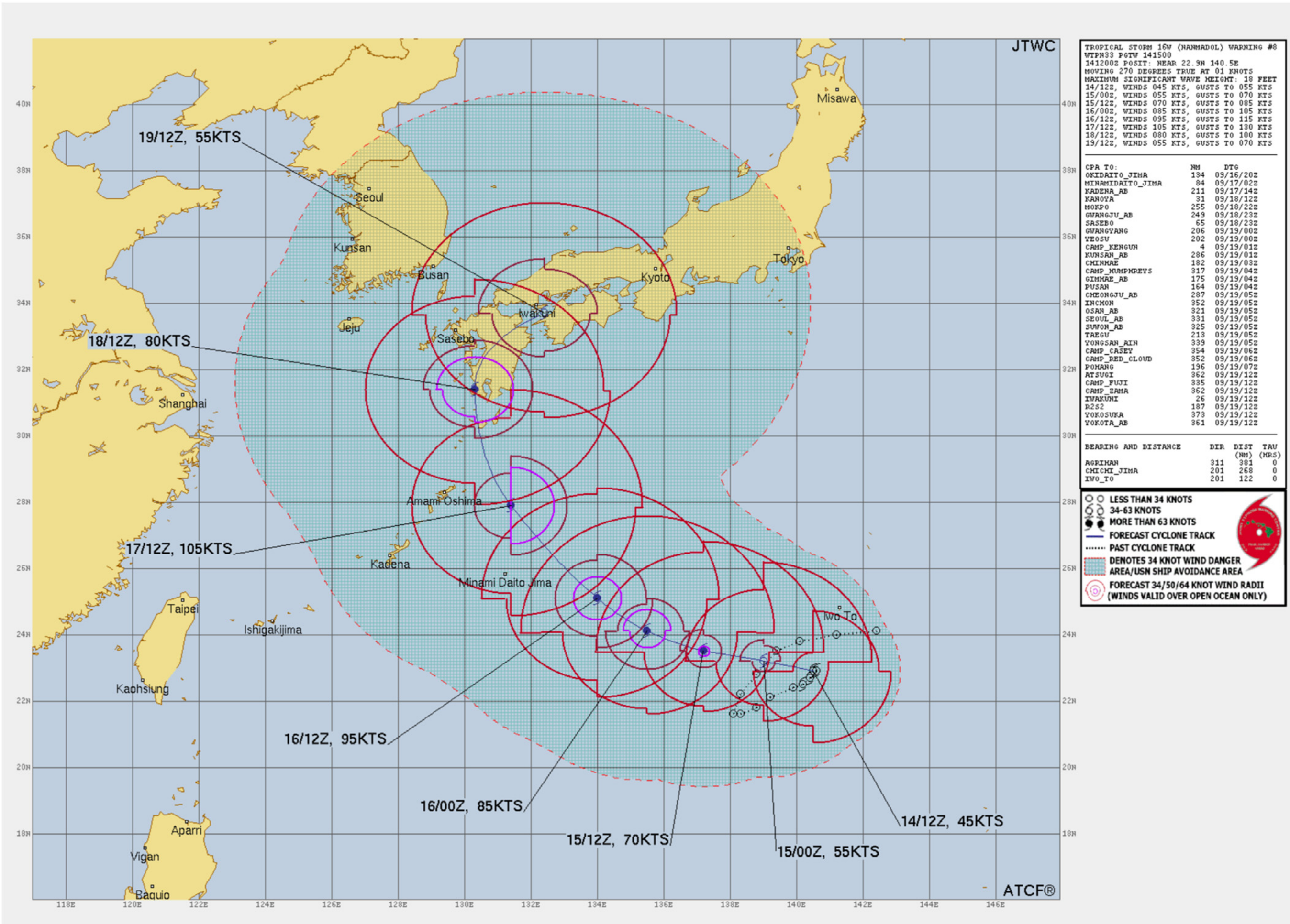

Fig. 7. An example of a JTWC forecast graphic for Tropical Storm 16W, issued at 1500 UTC 14 September 2022. The TC forecast positions are denoted by TC symbols along a solid blue curve. The red, dark red, and purple bounding areas centered on each forecast position denote the forecast radii of 34 kt, 50 kt, and 64 kt winds, respectively. Forecast intensity is annotated for each forecast point in text, and the ship avoidance area is indicated by the blue hatched region. (Image courtesy of JTWC).

4.2. Statistical-dynamical methods

In some cases, the statistically generated uncertainty is combined with information from NWP in order to improve the skill and account for situationally dependent information. This can simply be done by subjectively altering the area generated by climatological error to account for apparent spread of model guidance, or by more elaborate methods.

At some centers, dynamically information is incorporated into the track uncertainty in a subjective fashion. The Met-Service (NZ) estimate position uncertainty subjectively to encompass the spread of the UK, EC and GFS deterministic forecast positions, then the cone of uncertainty is adjusted to include the majority of the EC ensemble tracks. Other ensemble data from the MOGREPS and GEFS is also assessed, as well as deterministic guidance from NAVGEM, CMC, ACCESS and HWRF, to determine how the forecast policy matches the overall range of model outcomes. This context is useful for communicating track uncertainty in briefings to stakeholders, particularly at longer time frames, so a 'worst case' scenario can be assessed alongside the most likely.

The Australian Bureau of Meteorology (ABoM) used a similar process prior to the 2021/22 tropical cyclone season. The 'first guess' at position uncertainty was a cone generated based on the climatological error at each time step. The forecaster would then overlay deterministic and ensemble track guidance, and manually adjust the area to better reflect NWP guidance. This has since been updated and is detailed in the following section.

4.3. Predominantly dynamical methods

Some centers have already adopted processes for expressing forecast uncertainty by predominantly dynamical methods, and others have plans to make such changes in the future. Typically, this process will generate a Cone of Uncertainty/Forecast Confidence Area using ensemble forecast data as the basis, with some statistical input whenever there is insufficient ensemble data. There can also be bias applied early in track to the analysis position uncertainty.

Unlike RSMC New Delhi, which uses static cones of uncertainty, NCMRWF (India) forecast tracks currently show a color shaded probability of location on their TC forecasts which is based on the NCMRWF global ensemble NEPS-G (shown earlier in Fig. 2). This area is uncalibrated and based on only one model.

At RMSC Tokyo (JMA) the uncertainty at each forecast timestep is expressed via a probability circle, which shows the range into which the center of a TC is expected to move with 70% probability at each validation time. Starting in June 2019, the radius for all forecast times up to 120 h is determined by the multiple ensemble method, which is solely according to the confidence level based on the cumulative ensemble spread calculated using multiple ensemble prediction systems (EPSs) consisting of ECMWF, NCEP and UKMO global EPSs in addition to JMA GEPS (Fukuda and Yamaguchi 2019). The radii are based on the confidence level at each forecast time, which is classified as 'high', 'medium' or 'low', and is based on the degree of ensemble spread. When this method started in 2019, the confidence level was determined from calculating all ensemble spread values from the period 2016 to 2018 and dividing them based on degree of spread into three groups (high, medium, and low) such that the ratio is 4:4:2. A standard radius is used at each forecast timestep based on the confidence level assigned. The radii used for each confidence level is reviewed every few years, but at the time of writing the data from 2016 to 2018 continues to be used.

Since the 2011/12 tropical cyclone season, RSMC La Reunion have incorporated dynamical information into their Cone of Uncertainty (referred to by La Reunion as a 'forecast confidence area'). This has been built around the official RSMC La Reunion track forecast (75% confidence level circles for each lead time) and based on the spread of the ensemble forecast members of the ECMWF ensemble prediction system (EPS). A new version of the forecast confidence area was implemented in 2020 (for the cyclone season 2020–2021) based on SPICy scenarios, developed as part of the *Système de Prévision des Inondations en contexte Cyclonique* (SPICy) project.

In the framework of the SPICy project a methodology was developed at RSMC La Reunion (Meteo-France) in order to generate alternate probabilistic scenarios of tropical cyclones track and intensity forecasts around the official forecast.

Initially two different techniques were considered, one only using climatological errors, and the other purely dynamical using the ECMWF ensemble system (EPS) outputs. A hybrid approach was eventually implemented after being assessed as generally more preferable. A first set of climatology-built scenarios is generated using the statistical distribution of RSMC La Reunion forecast errors. This initial set is then modulated using real-time information provided by the ensemble prediction system (EPS) of the European Centre for Medium-range Weather Forecasts (ECMWF).

The method (described in Bonnardot et al., (2019)) was first developed using five TC seasons of climatological data, from 2011/12–2015/16, including 826 forecasts. These were classified into 45 classes from which alternative scenarios are built around the official RSMC La Reunion forecast. The real time ECMWF ensemble forecast members are then realigned on the RSMC forecast by first calculating differences between each member and the ensemble mean. Each of those scenarios is then assigned to the nearest climatological class and a probability of the previously built climatological scenarios is determined based on the number of realigned ECMWF ensemble members assigned to it. This leads to a decrease in the number of scenarios as climatological classes with zero realigned members assigned are given a probability of zero. In this way, the final ensemble is modulated using the real time ECMWF forecast information to produce a cost-efficient set of approximately 15–30 alternative scenarios (with associated probabilities) that can be delivered within the operational timeframes required.

However, since the article was written the method has slightly evolved. In its latest version, all climatological classes are kept but the initial probability (climatological one) is modulated with respect to the number of ECMWF members associated to each climatological class. Therefore, the final number of scenarios is unchanged at the end of the process (45 classes - but with more classes associated to a zero probability). Those weighted alternate scenarios are then used to determine the confidence circles for each lead time (with a calibrated 75% probability).

For the 2021/22 tropical cyclone season, ABoM changed its method for expressing track uncertainty from a climatologically based cone of uncertainty to a dynamically generated area. For each forecast timestep, an 80% FCA is produced based on a calibrated super ensemble of guidance. An 80% FCA represents the area within which the system will be located at a particular time, to a confidence level of 80%. In other words, it represents the region in which the tropical system center is expected to be located for 80% percent of cases. The 80% confidence level was chosen so that this new method produced similar sized areas when compared to the previous, more subjective, method. The individual FCAs are combined into swaths (FCCs). The FCCs are for the time range starting at analysis time and ending at forecast hour +24, +48, +72 and + 120. The FCCs are the swath of the 80% FCAs in the time range. Due to the skill of the analysis position and its influence on short term forecast uncertainty, the output is weighted to the OFT uncertainty at analysis time, changing to be all super ensemble guidance by forecast hour 36.

Last season, the FCA for super ensemble guidance was derived using an Adaptive Bandwidth Kernel Density Estimate (KDE) to generate a Probability of Location grid from all the guidance fixes, then finding the contour where the sum of the grid cells within the contour gave a FCA which had 80% of analyzed systems within it. This means the FCA's are calibrated.

Included in the forecast system is the option to weight towards the historical error method. This was used when ensemble guidance (either due to model performance or issues with the vortex tracker) was problematic or when the system was weakening and the number of guidance tracks decreased below a certain threshold.

Verification performed when developing this system indicated that a super ensemble of ECMWF EPS, MOGREPS and GEFS resulted in increased spread and once calibrated, was better performing than any single or two model combination. A lagged super ensemble (including the previous run in the super ensemble along with the current run) further increased the spread and, once calibrated, resulted little change in performance relative to the not lagged super ensemble. However, a lagged super ensemble has the benefits of maintaining some stability between forecasts and increasing the number of ensemble members. For the KDE method in particular, more members mean less uncertainty in the region of the FCA, with simulations indicating a lag of 6 h (2 runs of ECMWF, MOGREPS and GEFS) was optimum.

Forecaster and user feedback from the 2021/2022 season was that the FCA's from the KDE method had “precision exceeding accuracy”. The perception was that some FCAs had wavey edges, where they should be straighter or more uniformly curved. In response to this, for the 2022/23 season the FCAs are to be generated using a Gaussian Mixture Model (GMM) instead of the KDE.

The GMM is superior to the existing KDE method in a number of ways:

- GMM appears to outperform the KDE when the number of data points is in the range of the data being used. That is, it gives a more consistent and better estimation.
- GMM delivers shapes which are more regular, which is a desirable property from a customer point of view (See Fig. 8).

For the 2023/24 season, it is expected that forecasters will also have the ability to quality control the super ensemble track guidance prior to the generation of the FCAs. It is also intended that Confidence Areas will be generated for each system of interest at the 00Z and 12Z timesteps out to seven days, even prior to system genesis. This will form the basis of the Tropical Cyclone Outlook (TC genesis product) and will result in there being a smooth transition from outlook products to warning products as a system develops.

5. WMO TC-PFP Tropical Cyclone-Probabilistic Forecast Products (TC-PFP) seamless GDPFS Pilot Project

The Tropical Cyclone-Probabilistic Forecast Products (TC-PFP) effort is a WMO Seamless GDPFS Pilot Project established in response to recommendations from the 2018 IWTC-9 in Honolulu, Hawaii. The main goal of TC-PFP is to coordinate across RSMCs, TCWCs, and other forecast centers to help identify best practice guidance for probabilistic TC forecasts incorporating a value cycle approach. TC-PFP is being implemented in 3 phases: Phase 1 (TC formation and position) began in 2020; Phase 2 (TC intensity and structure) will begin in 2023; and Phase 3 (rainfall and storm surge) will start in 2024. Phase 1 efforts are described in detail by Dunion et al. (2023; anticipated TCRR submission). They surveyed numerous forecast centers around the world and found that many are developing their own probabilistic forecast

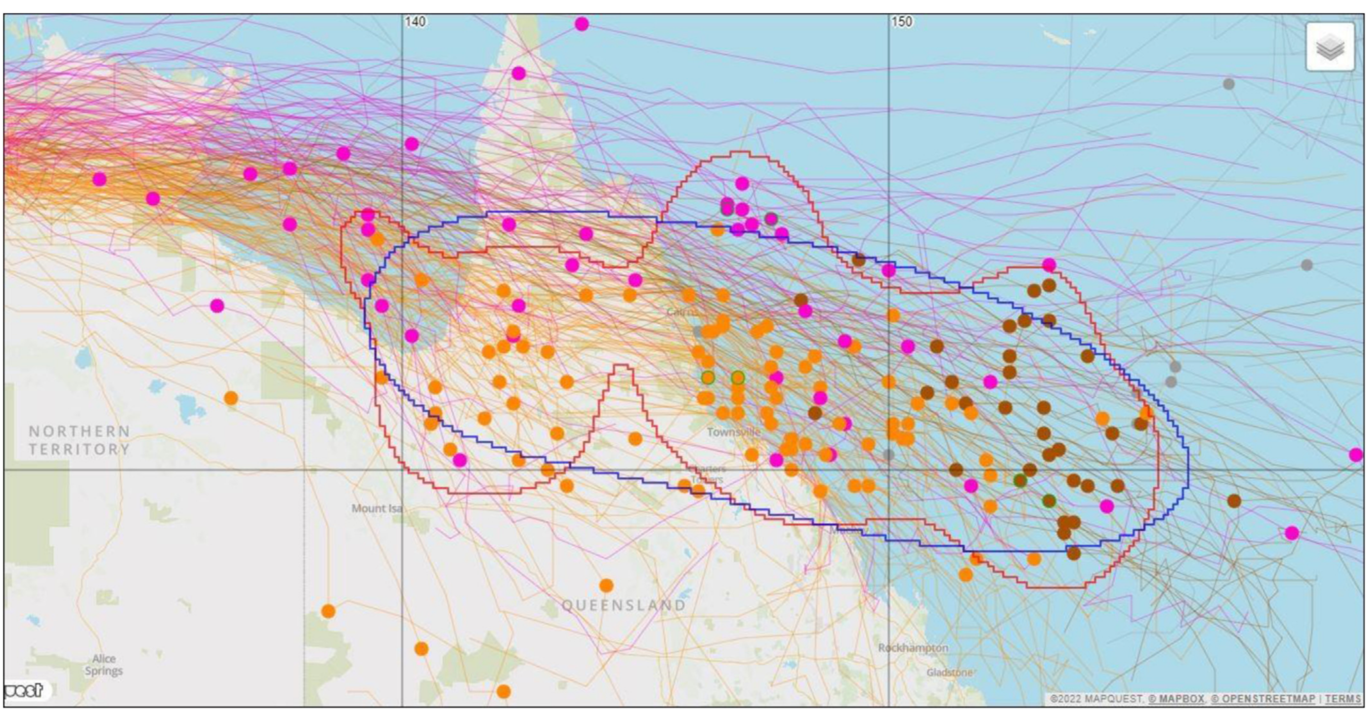

Fig. 8. A comparison of Forecast Confidence Area generation methods. The red line is an uncalibrated KDE calculation, the blue line is an uncalibrated GMM calculation. Note how the GMM produces a more desirable shape whilst still covering a similar sized area. This image is for the system that would eventually develop into Tropical Cyclone Seth. The displayed tracks are from a lagged super ensemble of ECMWF, MOGREPS and GEFS, with the runs from 00Z 26 December 2021, and 18Z 25 December 2021. The time shown is 00Z 31 December 2021 (i.e. 120 h from the most recent ensemble run) (Image courtesy of Australian Bureau of Meteorology).

techniques for predicting TC track and intensity. Several common challenges that many centres face when developing probabilistic TC forecasts were also identified by these communications. Dunion and coauthors, (2023) present recommendations related to probabilistic forecasts of TC formation and position that included enhancing communication between centres, improving accessibility and standardization of probabilistic forecast data, and optimizing product design and communication to end users through a value-cycle approach. It is anticipated that TC-PFP Phase 1–3 efforts should be closely coordinated with IWTC-10 and future IWTC workshops and will be an ongoing process that WMO could help steward.

6. Importance of vortex tracking and clustering algorithms and use of super ensembles

6.1. Vortex trackers

For tropical cyclone guidance, vortex trackers have a strong influence on the characteristics and useability of the data for operations as well as verification. Inconsistency between how systems are tracked between different ensemble prediction systems is an issue that needs to be overcome when developing probabilistic guidance that incorporates multiple ensemble guidance sources.

Whether the tracker is “pre-formation”, or “post-formation” has an influence on data availability. Post-formation means vortices are only tracked if initialized with an initial position, which is usually a manual analysis. Pre-formation means vortices are found and tracked in the grids without needing a manual analysis position to initiate the tracker. Pre-formation trackers produce data prior to the formation of a system and therefore have a more uniform amount of data across the 10 days of the forecast (the common length of guidance). Pre-formation trackers can be used to produce pre-formation forecast tracks, genesis, and outlook products. Post-formation trackers only produce data once the system has formed and produce much less data towards the end of the 10 days due to the average life of tropical cyclones meaning that many no longer exist towards the end of the forecast period. Several trackers combine the tracking of both pre-formation and post-formation tracks, with the option of applying different thresholds for each, and are the preferred type.

The threshold at which vortices are tracked also influences data availability and suitability for use in a super ensemble. For example, the GFDL tracker will track very weak systems, and this makes difficult to use pre-formation as it can be hard to distinguish between weak, shallow systems and ones that have the potential to develop. The MOGREPS tracker (Heming, 2017) has a higher threshold for tracking vortices, which means less overall tracks. Fig. 9 compares the data availability from EC (ECWMF EPS) and UK (MOGREPS) and the US (GEFS) over a three-and-a-half-year period. The EC and UK are pre-formation tracks, and the US is post-formation. A data point refers to a unique combination of tropical system, base time and forecast time. The US has less datapoints at forecast hour 12 because it is post-formation; it needs an analyzed system to initiate the tracker. Due to the typical life span of a system at tropical cyclone intensity, this also explains the bigger decrease in data points for the US as the forecast hour increases. The UK

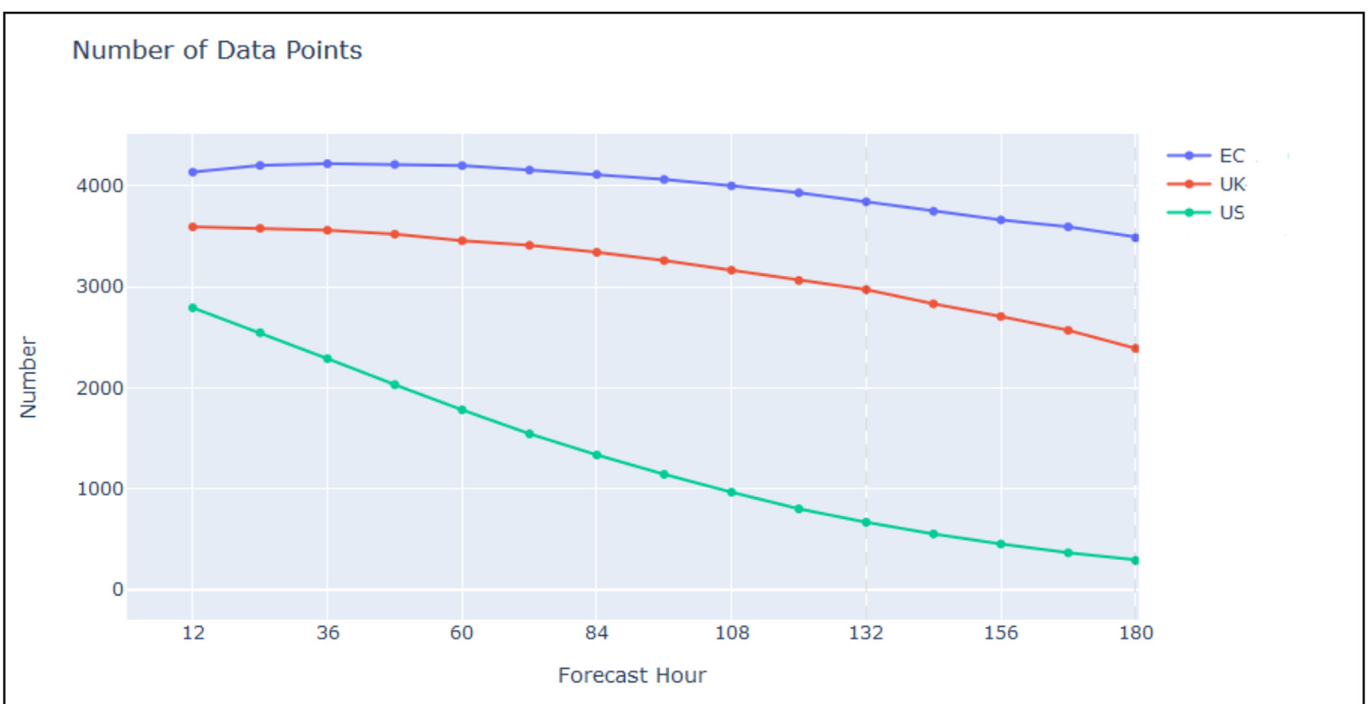

Fig. 9. Chart showing the number of data points available from each ensemble. The data is from January 2019 to September 2022 inclusive; the base times are 00Z and 12Z and the UK data includes the previous base time (18Z for the 00Z run, and 06Z for the 12Z run). The data is for analyzed tropical systems that appear in the track data and the IBTrACS dataset.

(MOGREPS) has less data points than the EC due to the tracker having a higher threshold to track a vortex and also having less members (36 vs 51). Less members means less overall data points that have more than 10 available fixes.

A study by Xiangbo Feng, John Methven and Kevin Hodges from the University of Reading aimed to compare four different tracking methods. This was performed to support the TC-PFP project. The study looked at western North Pacific tropical cyclones in the 2020 typhoon season and applied four different trackers to the same ECMWF EPS data.

The tracking methods compared were:

- Met Office (MO) operational tracker (Heming, 2017). Tracks provided by Helen Titley (Met Office).
- Geophysical Fluid Dynamics Laboratory (GFDL) operational tracker (Marchok, 2021). Tracks provided by Tim Marchok.
- ECMWF (EC) operational tracker (Vitart et al., 2012; Magnusson et al., 2021). Tracks are downloaded from the TIGGE TC database, and are for named systems only.
- Reading (RD) tracker (Hodges et al., 2017).

The study found that the estimate of ensemble spread is relatively insensitive to the tracker used (less than 10 km difference between trackers in the first 6 days of the forecast).

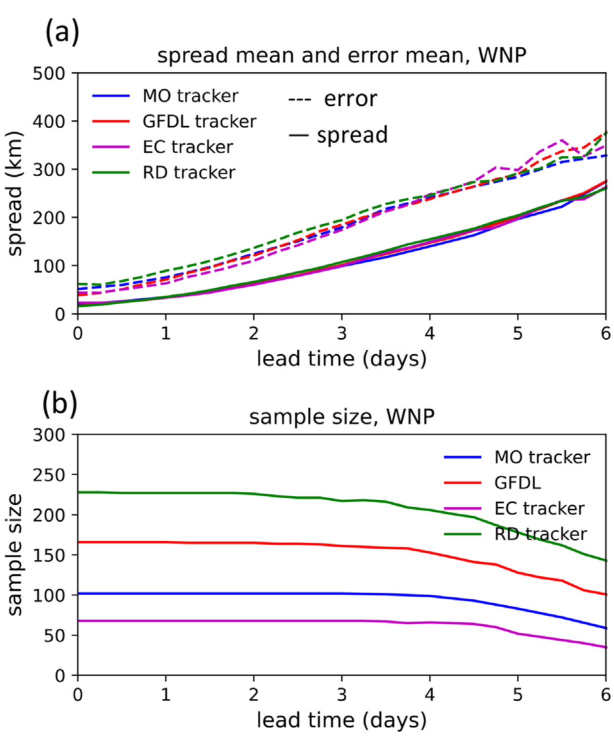

Fig. 10. (a) Ensemble spread (solid lines) and position error for the ensemble mean (dashed lines) averaged over TC forecasts for the western North Pacific basin in 2020 (b) The number of forecasts (i.e., different start dates). in (a). The calculation only includes the forecasts, in which i) ensemble members are matched to observations within the first three days of the forecast and with the mean of four consecutive track points ≤4° away from observation, and ii) the minimum number of matched members over the ensemble forecast members is 30 of 51.

However, the difference in error of the ensemble mean was more pronounced, varying by up to 20–25 km between trackers across the first 6 days. Fig. 10 shows this data. The error in ensemble mean track is also significantly greater than spread showing that the EC ensemble is under-dispersive. In these calculations only forecasts with named storms at t = 0 were included, enabling unambiguous comparison of features identified by different trackers. In each forecast, the error is measured using the great circle distance of ensemble mean position points to the corresponding observed track points (from IBTrACS). The ensemble spread is calculated by calculating the square root of the average of the squared distances (i.e., standard deviation) between track points in the ensemble members to the ensemble mean positions, and then the statistics are composited over all forecasts during the season containing named storms.

Due to the differences in feature identification (both in variables and thresholds used), there was a significant difference in the number of tracks identified by each tracker, even for named storms. The comparison between trackers has been made without normalization of this data (i.e. not every timestep of every storm compared is identified in all four track datasets). This is due to the fact there would not be a suitable sample size for some trackers, and also the assumption that the cases where all four trackers have a position would mostly be confined to mature storms (where the differences in trackers would be less pronounced). By not normalizing the tracks, the data is more comparable to what is currently available in an operational environment.

The differences in mean error between trackers may be related to differences in how the trackers calculate position, and also the sample size differences between trackers (the trackers that have a lower threshold for storm intensity will track weaker systems, which would be likely to introduce larger errors to the ensemble mean).

6.2. Clustering algorithms

When using multiple ensemble sources, or when using pre-formation tracks, there is a need to ensure that tracks from different ensemble members are each associated with a particular storm. The group of tracks from different ensemble methods diagnosed to represent the same storm are referred to as a cluster. At the Australian Bureau of Meteorology (ABoM), the clustering algorithm applied to ensembles is built upon the method described in Aijaz et al. (2019). This method consists of looping through all ensemble members for each forecast time. When the start of a track from a particular member is found, it is compared to existing clusters, and if it close enough to one then it is added to that cluster. If it is too far away from an existing cluster, then a new cluster is created. At the end of this looping process, a set of clusters will have been created for both existing systems and for systems that have yet to form.

As an overview, the clustering algorithm:

- Aims to minimize the total (sum) distance of a member's tracks to the clusters.

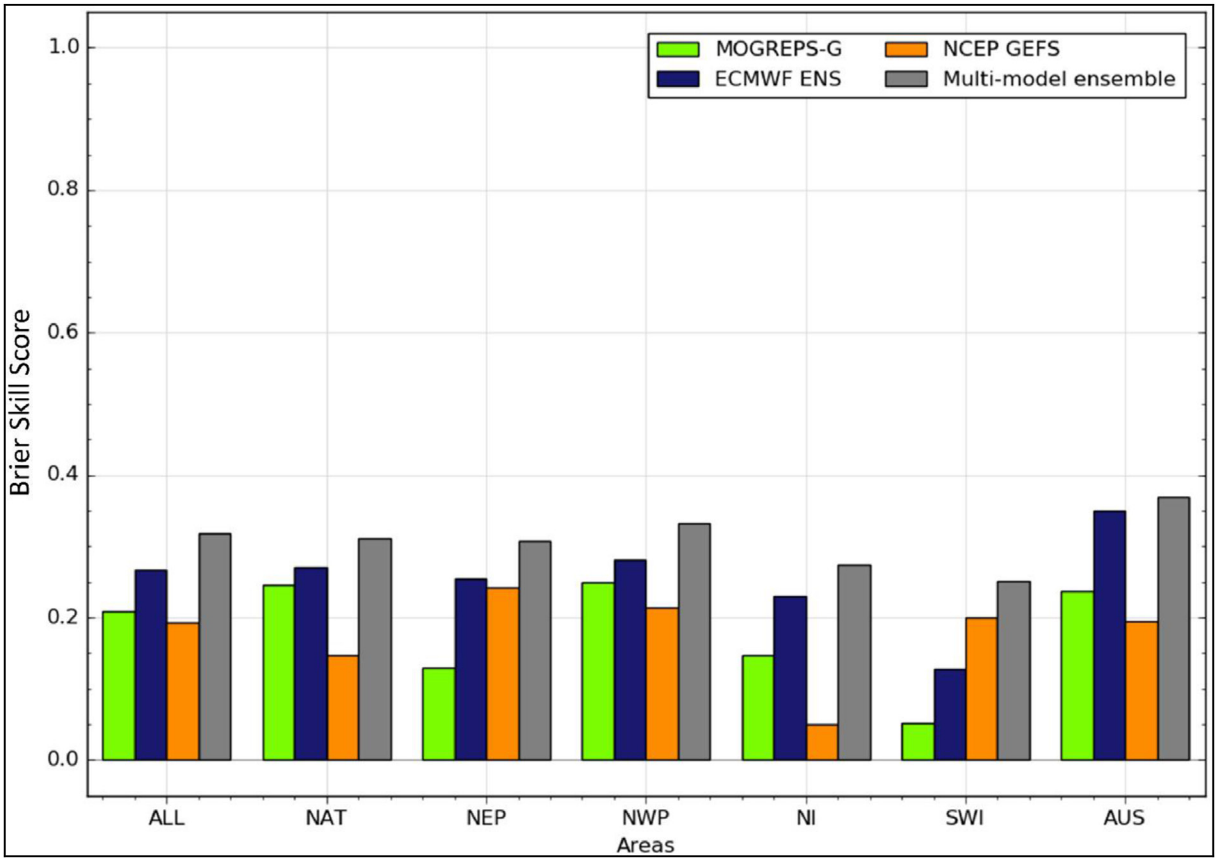

Fig. 11. BSS for the overall track probability forecasts for all named tropical cyclones (left) and for named tropical cyclones in each of the individual six basins, for each individual ensemble and the multi-model ensemble. The reference forecast used in the skill score calculation is the consensus (mean) track of the three deterministic forecasts NAT = North Atlantic, NEP = eastern North Pacific, NWP = western North Pacific, NI = Northern Indian Ocean, SWI = SW Indian Ocean, AUS = Australian region) (image courtesy of Titley et al. (2020)).

- Has a distance which is trajectory based, using a version of a Fréchet Distance, which weights towards the start of the track. The spatial distance is between a track and the mean of the cluster's tracks. The actual distance is a function of the spatial distance and the inverse of the number of unique ensemble members in the cluster, which preferences clusters with more tracks.
- Ensures a member can only have one track per cluster, unless they do not “conflict” (i.e. do not overlap with multiple positions at the same forecast timestep).
- Uses a large distance threshold when determining if a track is “close enough”. Typically, it is the “one track per cluster” rule which creates new clusters.

The method can also be used to “cluster” different models and base times together by clustering the mean tracks of each ensemble. This allows for combination into a super ensemble, even prior to system formation, and also for continuity across different forecast issue times. A clustering method that works on systems prior to formation allows for the equivalent of post-formation information to be made available in a pre-formation environment, which allows for system-based forecast products to be issued prior to tropical cyclone formation. For users of the forecast products, this can allow for longer range responses to tropical cyclone events, which in turn can improve decision making.

6.3. Evaluation of multi-model ensemble track probability forecasts

Ensemble forecast systems provide a useful way to assess situationally dependent forecast uncertainty. However, single-model ensembles tend to be under-spread and can fail to provide a range of solutions that covers the full probability distribution. A multi-model ensemble approach can help solve the problem of under-spread, with a set of independent and skillful ensemble providing a more complete representation of the uncertainty (in a similar fashion to how a consensus of deterministic models shows greater skill than each individual member of the consensus). Titley et al. (2020) evaluated the performance of track probability forecasts from a multi-model ensemble of MOGREPS-G, ECMWF ENS and NCEP GEFS across the period from January 2017 to December 2018, and found that this multi-model ensemble increased the skill and

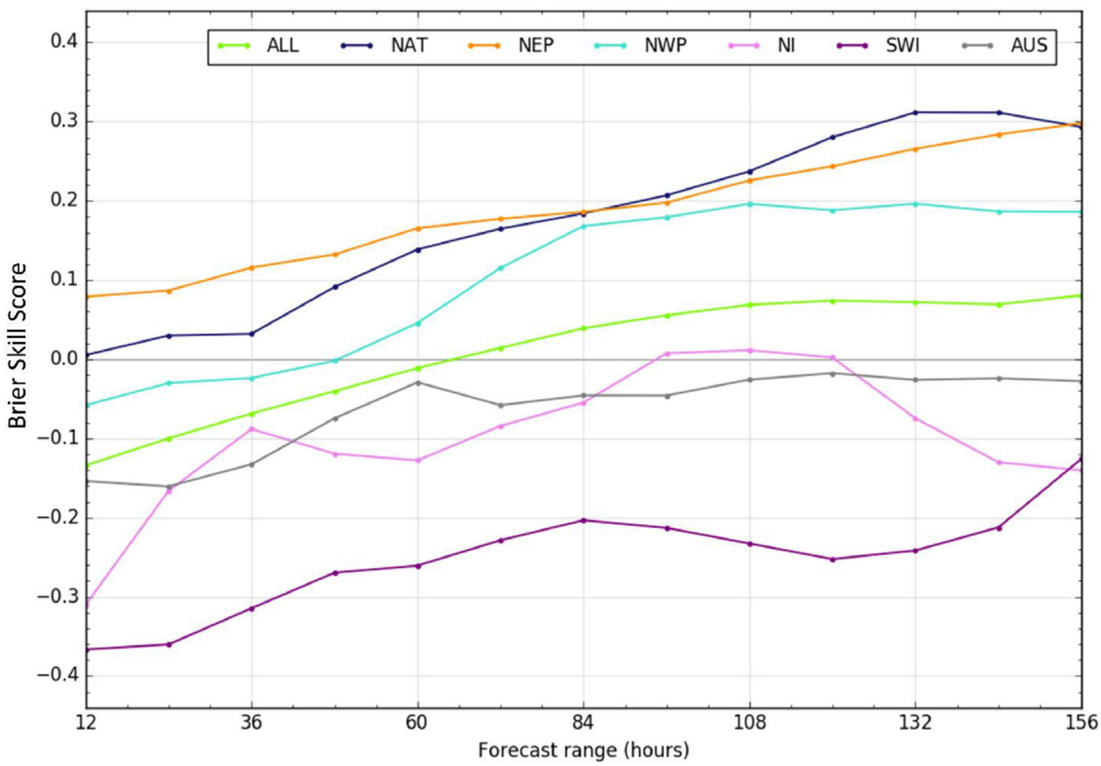

Fig. 12. Brier skill score (BSS) for the MOGREPS-G/ECMWF ENS/NCEP GEFS multi-model ensemble, for all storms (in green) and split into the six tropical cyclone basins, for the 24-h track probability forecasts centered around each forecast range. The reference forecast is the multi-model consensus of the parent deterministic models (image courtesy of Titley et al. (2020)).

value of probabilistic forecasts over the individual ensemble models. This was found to be applicable over all basins, as can be seen in Fig. 11.

Fig. 12 shows how the BSS varies over time compared to the reference multi-model consensus of the parent deterministic models across all basins and on the global dataset. It shows that the comparative skill of the multi-model ensemble becomes greater as the forecast lead time increases. On a global scale, the multi-model ensemble shows positive skill over the multi-model consensus beyond 60 h lead time. It needs to be noted that for some basins, the BSS remained mostly negative throughout the forecast period, even though for these basins the BSS was positive when considered across the entire 7 day forecast period. This is a reflection of the 7-day verification being more forgiving of along track errors (i.e. the track position may be quite accurate, but the timing of movement along the track has more error). The sample sizes were smallest in these basins, and a case-by-case assessment for the worst performing basin (the SW Indian Ocean) found that the deterministic errors were relatively low for the period studied when compared to previous seasons. This would make it more difficult for the multi-model ensemble to outperform the reference consensus.

7. Summary and conclusions

The generation of a forecast track via a consensus of model guidance is a well-established process used by all agencies surveyed for this report. The accuracy of official forecast tracks constructed in this manner continues to improve, though the rate of improvement has slowed in the last four years and it is possible we are reaching predictability limits, at least at shorter lead times. However, the established consensus methodology only depicts a single scenario. There is a growing need to have an accurate understanding of the full range of possible scenarios, especially when considering impact-based forecasts and warnings.

Recent advances and new techniques in track forecasting have centered around improving the depiction of location probability for a system. The best way to do this comes from harnessing the information from skillful ensemble prediction systems. Individual ensemble systems tend to be under spread, but the use of an ensemble of ensembles (a "super ensemble") can overcome this limitation. Challenges remain in developing a semi-standardized set of best practices for publicly-available probabilistic TC genesis and TC track forecasts and how to tailor both the products and associated communication to suit the requirements of the users.

As agencies move towards the inclusion of dynamic location probability in TC forecasts, consideration should be given to reducing the prominence of the forecast track itself (and even moving away from the concept of a forecast track and towards a true probabilistic representation of potential TC impacts).

Vortex parameter files are an important vehicle for sharing NWP data, allowing forecast centers to implement a super ensemble approach without needing the technical capability to run vortex tracking algorithms on NWP grids. The content of a vortex parameter file is dependent on the algorithm used. Many tracking algorithms are sensitive to model resolution, and some perform better than others with respect to tracking weak circulations. Further work is needed to support optimization within a standardized framework so that all operational centers are able to deliver to their communities the full value of a super-ensemble-based probabilistic approach to track forecasting.

Acknowledgments

The authors would like to thank their respective affiliate organizations for supporting their participation in IWTC-10. They would also like to thank the WMO World Weather Research Programme and WMO Tropical Cyclone Programme for jointly organizing IWTC-10 and co-chairs J. Courtney and R. Rogers for their leadership and direction.

Additional thanks go out to Philippe Caroff for information on SpiCy, and Takuya Hosomi for information on the JMA forecast process.

Acronyms used in the report

ABoM Australian Bureau of Meteorology
ACCESS-G Australian Community Climate and Earth-System Simulator - Global (ABoM)
ATCF Automated Tropical Cyclone Forecast
AUS Australian
BSS Brier Skill Score
CLIPER Climatology and Persistence Model
CMA China Meteorological Administration
CMC Canadian Meteorological Centre
COAMPS Coupled Ocean/ Atmosphere Mesoscale Prediction System
CONW JTWC track forecast consensus
COU Cone of Uncertainty
EC/ECMWF European Centre for Medium-Range Weather Forecasts
ENS Ensemble
EPS Esnsemble Prediction System
FCA Forecast Confidence Area
FCC Forecast Confidence Cone
GALWEM Global Air-Land Weather Exploitation Model
GDPFS Global Data-Processing and Forecasting System
GEFS/GEPS Global Ensemble Forecast System (USA)
GEPS Global Ensemble Prediction System
GFDL Geophysical Fluid Dynamics Laboratory
GFS/US Global Forecast System
GMM Gaussian Mixture Model
GPCE Goerss probability consensus estimate
HCCA Hurricane Forecast Improvement Program (HFIP) Corrected Consensus Approach
HCCA HFIP Corrected Consensus Approach
HFIP Hurricane Forecast Improvement Program
HFIP Hurricane Forecast Improvement Program
HKO HKO – Hong Kong Observatory
HWRF Hurricane Weather Research and Forecasting
IBTrACS International Best Track Archive for Climate Stewardship
JMA Japan Meteorological Agency
JTWC Joint Typhoon Warning Center
KDE Kernel Density Estimate
km kilometres
KMA Korea Meteorological Administration
kt knots
MOGREPS Met Office Global and Regional Ensemble Prediction System (UK)
NAT North Atlantic
NAVGEM Navy Global Environmental Model (US)
NCEP National Center for Environmental Prediction
NCEP National Centers for Environmental Prediction
NCMRWF National Centre for Medium Range Weather Forecasting
NCMRWF National Centre for Medium Range Weather Forecasting (India)
NCUM NCMRWF Global Unified Model
NEP Eastern North Pacific
NEPS-G NCMRWF Global Ensemble Prediction System
NHC National Hurricane Centre
NI Northern Indian Ocean
NWP Numerical Weather Prediction
NWP Western North Pacific
NZ New Zealand
OFT Official Forecast Track
PE Position Error
RD Reading
RSMC Regional Specialized Meteorological Centre
SPICy Système de Prévision des Inondations en contexte Cyclonique
SWI Southwest Indian Ocean
TC Tropical cyclone
TC-PFP Tropical Cyclone Probabilistic Forecast Products (project)
TCWC Tropical Cyclone Warning Centre
UK/UKMET/UKMO/MO United Kingdom Met Office
USA/US United States of America
WMO World Meteorological Organisation

References

Aijaz, S., Kepert, J.D., Ye, H., Huang, Z., Hawksford, A., 2019. Bias correction of tropical cyclone parameters in the ECMWF ensemble prediction system in Australia. Mon. Wea. Rev. 147 (11), 4261–4285. https://journals.ametsoc.org/view/journals/mwre/147/11/mwr-d-18-0377.1.xml.

Bonnardot, F., Quetelard, H., Jumaux, G., Leroux, M.-D., Bessafi, M., 2019. Probabilistic forecasts of tropical cyclone tracks and intensities in the southwest Indian Ocean basin. Q. J. R. Meteorol. Soc. 145, 675–686. https://doi.org/10.1002/qj.3459.

Chen, G.M., Yang, M.Q., 2022. Verification of tropical cyclone operational forecast in 2021. Tech. Rep. TLFDP-WMO. http://www.tlfdp.net/#/main/verification-reports.

Dunion, J., coauthors, 2023. Recommendations for improved tropical cyclone formation and position probabilistic forecast products (anticipated submission). tropical cyclone Research and review.

Fukuda, J., Yamaguchi, M., 2019. Determining Probability-Circle Radii of Tropical Cyclone Track Forecasts with Multiple Ensembles. RSMC Tokyo – Typhoon Center. *Technical Review No. 21* (April 2019).

Heming, J.T., 2017. Tropical cyclone tracking and verification techniques for Met Office numerical weather prediction models. Met. Apps 24, 1–8. https://doi.org/10.1002/met.1599.

Landsea, C.W., Cangialosi, J.P., 2018. Have we reached the limits of predictability for tropical cyclone track forecasting? Bull. Amer. Meteorol. Soc. 99, 2237–2243. https://doi.org/10.1175/BAMS-D-17-0136.1.

Magnusson, L., Majumdar, S., Emerton, R., Richardson, D., Alonso-Balmaseda, M., Baugh, C., Zsoter, E., 2021. Tropical Cyclone Activities at ECMWF. European Centre for Medium Range Weather Forecasts. *Technical Memo* 888.

Marchok, T., 2021. Important factors in the tracking of tropical cyclones in operational models. J. Appl. Meteorol. Climatol. 60, 1265–1284. https://doi.org/10.1175/JAMC-D-20-0175.1.

Sampson, C.R., Schrader, A.J., 2000. The automated tropical cyclone forecasting system (version 3.2). Bull. Amer. Meteorol. Soc. 81, 1131–1240. https://doi.org/10.1175/1520-0477(2000)081<1231:TATCFS>2.3.CO;2.

Simon, A., Penny, A.B., DeMaria, M., Franklin, J.L., Pasch, R.J., Rappaport, E.N., Zelinsky, D.A., 2018. A description of the real-time HFIP corrected consensus approach (HCCA) for tropical cyclone track and intensity guidance. Wea. Forecast. 33, 37–57. https://doi.org/10.1175/WAF-D-17-0068.1.

Titley, H.A., Bowyer, R.L., Cloke, H.L., 2020. A global evaluation of multi-model ensemble tropical cyclone track probability forecasts. Quart. J. Roy. Meteorol. Soc. 146 (726), 531–545.

Titley, H.A., Yamaguchi, M., Magnusson, L., 2019. Current and potential use of ensemble forecasts in operational TC forecasting: results from a global forecaster survey. Trop. Cyclone Res. Rev. 8 (3), 166–180. https://doi.org/10.1016/j.tcrr.2019.10.005. ISSN 2225-6032.

Vitart, F., Prates, F., Bonet, A., Sahin, C., 2012. New tropical cyclone products on the web. ECMWF Newsl. 130, 17–23.

Yu, H., Chen, G.-M., Zhou, C., Wong, W.K., Yang, M.-Q., Xu, Y.-L., Chen, P.-Y., Wan, R.-J., Hu, X.-R., 2021. Are we reaching the limit of tropical cyclone track predictability in the western North Pacific? Bull. Amer. Meteorol. Soc. 103, E410–E428. https://doi.org/10.1175/BAMS-D-20-0308.1.