Combined Three-Stage 7-Day Weighted Analog Intensity Prediction Technique for Western North Pacific Tropical Cyclones: Demonstration of Optimum Performance

HSIAO-CHUNG TSAI

Department of Water Resources and Environmental Engineering, Tamkang University,
New Taipei City, Taiwan

RUSSELL L. ELSBERRY

Trauma, Health, and Hazards Center, University of Colorado, Colorado Springs, Colorado and
Naval Postgraduate School, Monterey, California

(Manuscript received 1 July 2019, in final form 30 September 2019)

ABSTRACT

The original 7-day weighted analog intensity Pacific (WAIP) prediction technique is improved by developing a new version for the preformation stage and combining it with the bifurcation version that is especially for the intensification stage and the ending-storm stage WAIP. This combined three-stage WAIP includes a calibrated intensity spread designed to include 68% of the verifying intensities. In this demonstration of the optimum performance of the combined WAIP, the Joint Typhoon Warning Center (JTWC) best tracks, the time to formation (T2F), and ending-storm time are utilized as inputs along with the initial intensity [i.e., either 15 or 20 kt ($1\,\text{kt} \approx 0.51\,\text{m s}^{-1}$)]. In the new "Before Formation" (defined as 25 kt) stage of the WAIP, the intensity evolution is represented by a square function between the initial intensity and the T2F (25 kt), and very small mean absolute errors (MAEs) and intensity spreads are achieved. In the bifurcation version, correct selections between two cluster WAIP intensity evolutions lead to MAEs that slowly increase to 17 kt at 144 h, and the intensity spreads are relatively small as well. Because the ending-storm time constrains the analog selection in the WAIP, the MAEs begin to decrease after 72 h and are only 10 kt at 156 h. A case study is presented as to how an ensemble storm-track forecast along with the T2F could provide WAIP 7-day intensity predictions beginning in the preformation stage, which indicates the potential for earlier guidance for the JTWC intensity forecasts of western North Pacific tropical cyclones.

1. Introduction

As DeMaria et al. (2014) have well documented in the Atlantic and in the western North Pacific, tropical cyclone (TC) intensity forecasts have not been improved as much as TC track forecasts. The forecasters have many skillful guidance products for track forecasting, and they first devote much effort to providing the most accurate track forecast in each situation. However, the global numerical weather prediction models that provide the primary guidance for the official track forecasts are generally not used for the intensity guidance due to their coarse resolution not being sufficient to resolve the inner-core convective processes that are critical to intensity changes. Since the forecaster's objective is to provide the intensity evolution that is most likely to occur given the official track forecast, the widely used Statistical Hurricane Intensity Prediction System (SHIPS; DeMaria and Kaplan 1994; DeMaria and Kaplan 1999), or the Statistical Typhoon Intensity Prediction System (STIPS; Knaff et al. 2005) in the western North Pacific, generates intensity forecasts by extracting global model-predicted variables related to TC intensity changes along that official track forecast.

Elsberry and Tsai (2014) was the first development of an analog technique for intensity and intensity spread predictions of western North Pacific TCs that are based on the hypothesis that the TC track is the predominant factor in the intensity forecast beyond 72 h. The 10 best historical track analogs were matched with the target TC track [for the development sample, the Joint Typhoon Warning Center (JTWC) best track file was used], and a

Corresponding author: R. L. Elsberry, relsberr@uccs.edu

simple arithmetic average intensity each 12 h out to 120 h corresponding to these 10 tracks was called the Situation-Dependent Intensity Prediction (SDIP). Tsai and Elsberry (2014a) then tested the SDIP by matching the 10 historical track analogs with the JTWC official track forecasts rather than best tracks. More weight was given to the 72–120-h portion of the track, and a new weighted intensity spread guidance product was provided that was calibrated to include about 68% of the verifying intensities at all forecast intervals. This new weighted analog intensity technique was 5 kt (1 kt $\approx$ 0.51 m s^{-1}) (20%) more accurate at 120 h than the JTWC official intensity forecasts.

Given this success with the 5-day forecasts, Tsai and Elsberry (2015) extended the forecasts to 7 days again with a perfect-prog approach that utilized the JTWC best track files from 1945 to 2009 to select the 10 best historical track analogs to the target TC track. While Tsai and Elsberry (2015) utilized a development sample from the 2000–09 seasons and an independent sample from the 2010–14 seasons, in a similar development of a 7-day weighted analog intensity technique for Atlantic hurricanes Tsai and Elsberry (2017a) found that it was better to randomly select 70% of the TCs in the entire sample to be the training set, and use the remaining 30% as the independent set. Specifically, this random sampling approach was more successful in obtaining training and independent sets with similar intensity biases so that the bias correction procedure had consistent performance for the independent set. Similarly, this approach improved the procedure for calibration of the "raw" intensity spreads each 12 h during the 7-day forecast interval to ensure that 68% of the WAIP intensities will verify within the calibrated intensity spread.

The "original 7-day WAIP" that will be the standard for comparison with the combined three-stage WAIP in this study is Tsai and Elsberry (2015) that has been redeveloped with the 70% training and 30% independent set approach. Since the early development studies were in a journal that is not widely available, a summary of the basic features and procedures for calculating the WAIP intensity and intensity spread is given in the appendix. It is noteworthy that because the searching for analogs can be done quickly on a desktop computer, the 7-day WAIP forecasts can be produced in about 1 min.

The first indication that a single version of the 7-day WAIP weighted analog technique could not be used for all stages of the TC life cycle was for the intensification stage when large intensity spreads occurred among the 10 historical analog intensity evolutions. Such bimodal or bifurcation situations may arise due to uncertainty in the timing of formation, timing and magnitude of rapid intensification periods, or track forecast uncertainty

leading to landfall versus nonlandfall. Tsai and Elsberry (2015) provided examples of 7-day WAIP forecasts of rapid intensification (their Figs. 8 and 9), rapid decay (their Fig. 10), and cyclones with extended periods of nondevelopment (their Fig. 11). Even though a minority of the 10 analogs may have indicated rapid intensification or rapid decay, the majority of the analogs generally do not indicate these tendencies. With these bimodal intensity evolutions, the weighted-mean WAIP intensity predictions will be "down the middle."

Since the Tsai and Elsberry (2014b, 2018) articles that describe the modifications of the 5-day and the 7-day WAIP to address these bifurcation situations are also in that journal that is not widely available, a detailed summary is given in the appendix. An objective technique is provided to detect these intensity bifurcation situations based on the magnitude of the "raw intensity spreads." Then a hierarchial cluster analysis (Wilks 2011) is applied to separate the analogs (in these 7-day bifurcation studies, 16 analogs were utilized) into two WAIP cluster intensity evolutions. Thus the bifurcation version WAIP outputs are the Cluster 1 intensity evolution with the larger maximum intensity and the Cluster 2 intensity evolution with the smaller maximum intensity, and separate intensity spreads are provided about each cluster intensity evolution. Tsai and Elsberry (2018) demonstrated that if a correct selection of the Cluster 1 or Cluster 2 WAIP forecast for each bifurcation situation was made, a substantial improvement in the intensity mean absolute errors (MAEs) was achieved relative to the original WAIP forecasts based on all 16 of the best analogs. Therefore, the Tsai and Elsberry (2018) bifurcation version of the WAIP will be utilized in the combined three-stage WAIP, and the optimum performance during the intensification stage from the correct selection of the Cluster 1 or the Cluster 2 WAIP intensity evolution will be demonstrated.

The second indication that a single version of the 7-day WAIP could not be used for all TC stages was that Tsai and Elsberry (2015) found an increasingly large overforecast intensity bias in the 5–7-day interval that they attributed to "ending storms" due to landfall, extratropical transition, or to nondevelopment within the 7-day forecast interval. Following Tsai and Elsberry (2017a) who had developed an ending-storm version for the Atlantic, Tsai and Elsberry (2017b) developed an ending-storm version of the 7-day WAIP with an additional constraint in the selection of the 10 best historical analogs that the intensity at the last matching point with the target TC track cannot exceed 50 kt. A separate calibration of the intensity spreads for the training set to ensure that 68% of the verifying intensities will be within the 12-h WAIP intensity spread values resulted in

smaller spreads (or higher confidence) for ending storms in the 5–7-day forecast intervals. Thus, some extra effort by the forecaster to identify ending-storm events of landfall, extratropical transition, or nondevelopment within 7 days will provide improved intensity and intensity spread guidance. Consequently, the second modification of the Tsai and Elsberry (2015) 7-day WAIP is to include the ending-storm stage WAIP version of Tsai and Elsberry (2017b).

A new effort has been to explore provision of intensity forecasts beginning in the preformation stage (for JTWC, formation is an intensity ≥ 25 kt) because western North Pacific tropical depressions can intensify to a typhoon so rapidly that earlier warnings are needed. While JTWC provides probabilistic alerts of formation in 24, 48, and 72 h, JTWC does not issue intensity forecasts for preformation circulations with intensities of 15 or 20 kt. Thus, the third stage in the new combined three-stage WAIP version is this preformation stage. As will be described in section 2a, a different approach than in Tsai and Elsberry 2015 will be utilized in the preformation stage that requires the forecaster to specify the time-to-formation (T2F).

The objective of this study is to demonstrate the optimum performance of a combined three-stage, 7-day WAIP version in which the preformation stage is combined with the bifurcation version for the intensification stage, and then addresses the ending-storm stage with an additional constraint on the historical analog selection. As described above, the original 7-day WAIP, the bifurcation version, and the ending-storm version were all developed with the JTWC best track files, which is the optimum performance for the WAIP as it assumes the JTWC official track forecast will have no error, and it separates the intensity error from the track forecast error contribution to the intensity error. Moreover, the combined WAIP incorporates the intensity evolutions of up to 16 historical track analogs in the weighted-mean intensity forecast and in the calibrated intensity spread, which provides some measure of the intensity uncertainty that might be attributed to track forecast errors. In this demonstration of the optimum performance of the combined WAIP, the TC track, the initial intensity, the T2F (25 kt), and the ending-storm time will be determined from best track files. The methodology for developing the preformation stage and for combining the three stages of the 7-day WAIP is described in section 2. The performance of the combined 7-day WAIP with the three stages is documented in section 3, and an example of the performance is provided when the preformation TC track is from an ensemble model rather than the JTWC best track file. A summary and some discussion of an operational test during the 2019 season are presented in section 4.

2. Methodology for preformation stage and combined three-stage WAIP

a. WAIP preformation stage

Searching for the 16 historical analogs to form a weighted average of those intensity evolutions during the preformation stage would lead to a large positive bias because the JTWC best track file including the preformation circulations during 2000–15 has 85% developing TCs to at least 35 kt. Thus a different approach than in Tsai and Elsberry (2015) was necessary to also provide 7-day WAIP intensity forecasts during the preformation stage.

In this optimum performance demonstration, the initial intensity (typically 15 or 20 kt) and the T2F of the pre-TC circulations in the western North Pacific are from JTWC best track files. Three functions (linear, exponential, and squared) were tested to describe the intensity evolutions between these two initial intensities and 25 kt, which JTWC designates as the T2F, but may also be 35 kt if desired. The squared function intensity evolution had somewhat smaller sample-mean biases and MAEs than the linear or exponential function intensity evolutions (not shown). The squared-function intensity evolutions for the first entries for all storms during 1985–2015 in the JTWC best track files that had an initial intensity of either 15 or 20 kt are shown in Fig. 1a. Note that a substantial fraction of these first entries at either 15 or 20 kt achieved an intensity of 25 kt (i.e., formation as a tropical depression) very quickly. Only seven (one) of the pre-TC circulations with a first-entry intensity of 15 kt (20 kt) took longer than 120 h to achieve formation. As will be demonstrated in section 3e, an ensemble model guidance product can provide the target TC track forecast for the combined WAIP analog selection in predicting the formations out to perhaps 120 h.

If all (i.e., not just the first entry) JTWC best track pre-TC circulations with initial intensities of 15 or 20 kt are considered, and at least 10 cases are required at each at 6-h forecast interval, the sample size is ~550 cases at time 0–12 h. However, the sample size then exponentially decreases to ~170 at 48 h and to 10 at 120 h (not shown). This rapid decrease in the sample size is consistent with the first-entry plot in Fig. 1a and confirms that a substantial fraction of the western North Pacific pre-TC circulations achieve formation as a tropical depression within 48 h.

Since in the optimum performance demonstration the T2F is known, assuming a simple squared-function intensity evolution from initial intensities of 15 or 20–25 kt at the T2F as in Fig. 1a will lead to sample-mean intensity biases that are less than ~2 kt (Fig. 1b). These mean
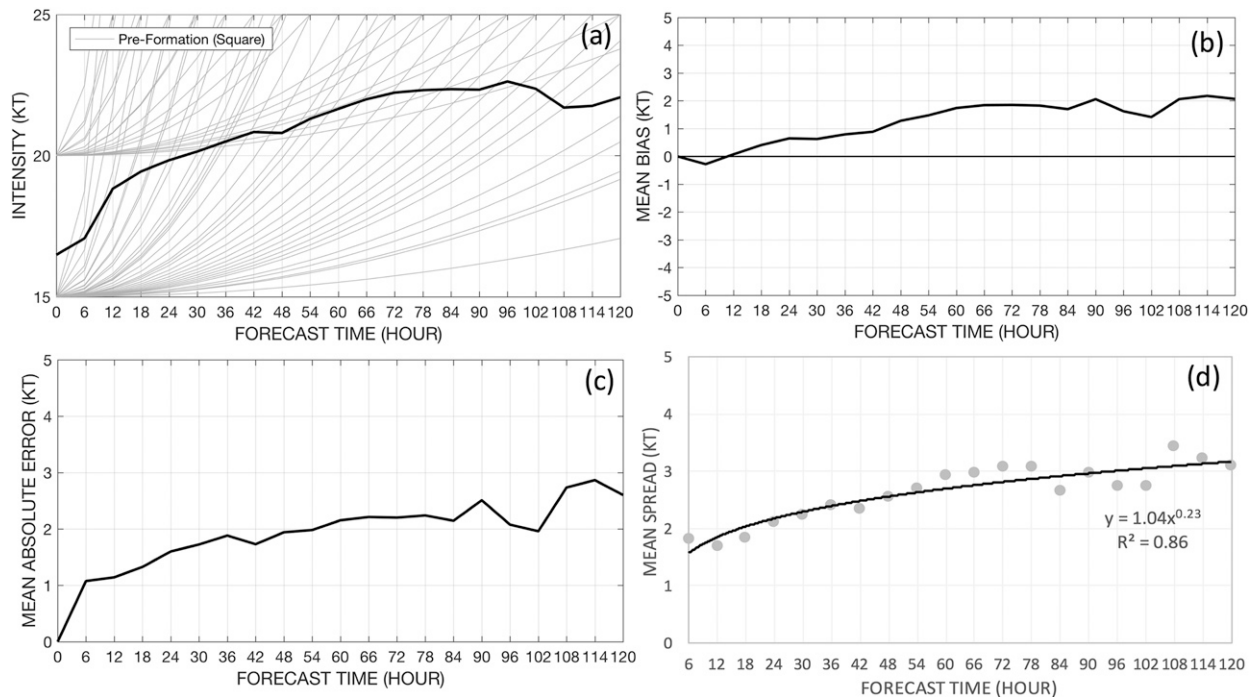
FIG. 1. (a) Intensity (kt) evolutions during the preformation stage represented by a squared function between an initial intensity of either 15 or 20 kt and an ending time to formation with an intensity of 25 kt. This sample is for the first entries for all storms in the JTWC best track files during 1985–2015 that had an initial intensity of either 15 or 20 kt. The heavy solid line is just the sample-mean intensities of the best track intensities at that forecast interval. (b) Mean intensity bias (kt) for the squared-function intensity evolutions as in (a) for all (not just the first entries) storms with initial intensities of either 15 or 20 kt in the JTWC best track files during 1985–2015. (c) As in (b), but for MAEs (kt). (d) As in (a), but for sample-mean intensity spreads (kt).

biases are positive, which means that the squared-function intensity forecasts in the preformation period will be too high, but a bias of 2 kt is small because TC intensity is only estimated to the nearest 5 kt. As shown in Fig. 1c, the MAEs are also ~2 kt, which is likely related to the positive bias in Fig. 1b. In conclusion, the sample-mean MAEs will be less than the observational intensity uncertainty of 5 kt over the entire forecast interval of 120 h from a simple-square function intensity evolution, *if* the T2F is accurately known.

A sensitivity test to the specification of the T2F for just the preformation period was carried out that assumed no error within 24 h, adding random errors at ±6 h to the T2F between 30 and 48 h, ±12 h between 54 and 72 h, ±18 h between 78 and 96 h, and ±24 h for greater than 96 h (not shown). The WAIP MAEs after adding these random errors were almost the same, which may be attributed to the smaller sample sizes of T2F at the longer forecast intervals (see Fig. 1a). Furthermore, the time intervals between these later T2Fs and the end of the forecast at 168 h (during which the intensity differences might be expected to grow) is becoming smaller and smaller. Consequently, the WAIP sensitivity to errors in T2F is not as large as might have been anticipated.

To get an estimate of the intensity spread about the squared function intensity evolutions during the preformation period, the root-mean square errors were estimated for the same 1985–2015 period as in Fig. 1a and a curve was fit to represent the mean spread values at each forecast interval (Fig. 1d). The fitted curve for the intensity spread y is $y = 1.04 x^{0.23}$, where x is the T2F time, and this curve explains 86% of the variance. As the fitted curve is leveling off at a 3-kt intensity spread at 120 h, the curve will be extended to longer forecast intervals as needed to provide the intensity spread for any T2F. Note that this preformation stage of the combined WAIP intensity only depends on the initial intensity and the T2F with an intensity of 25 kt and does not depend on any analog selection, bias correction, or calibration of the intensity spread.

b. Combined three-stage WAIP

The combination of the WAIP preformation stage described in section 2a with the WAIP bifurcation version (Tsai and Elsberry 2018) starting at T2F along the TC track at an intensity of 25 kt, but then modified if necessary with the WAIP ending-storm constraint on analog selection (Tsai and Elsberry 2017b), is the
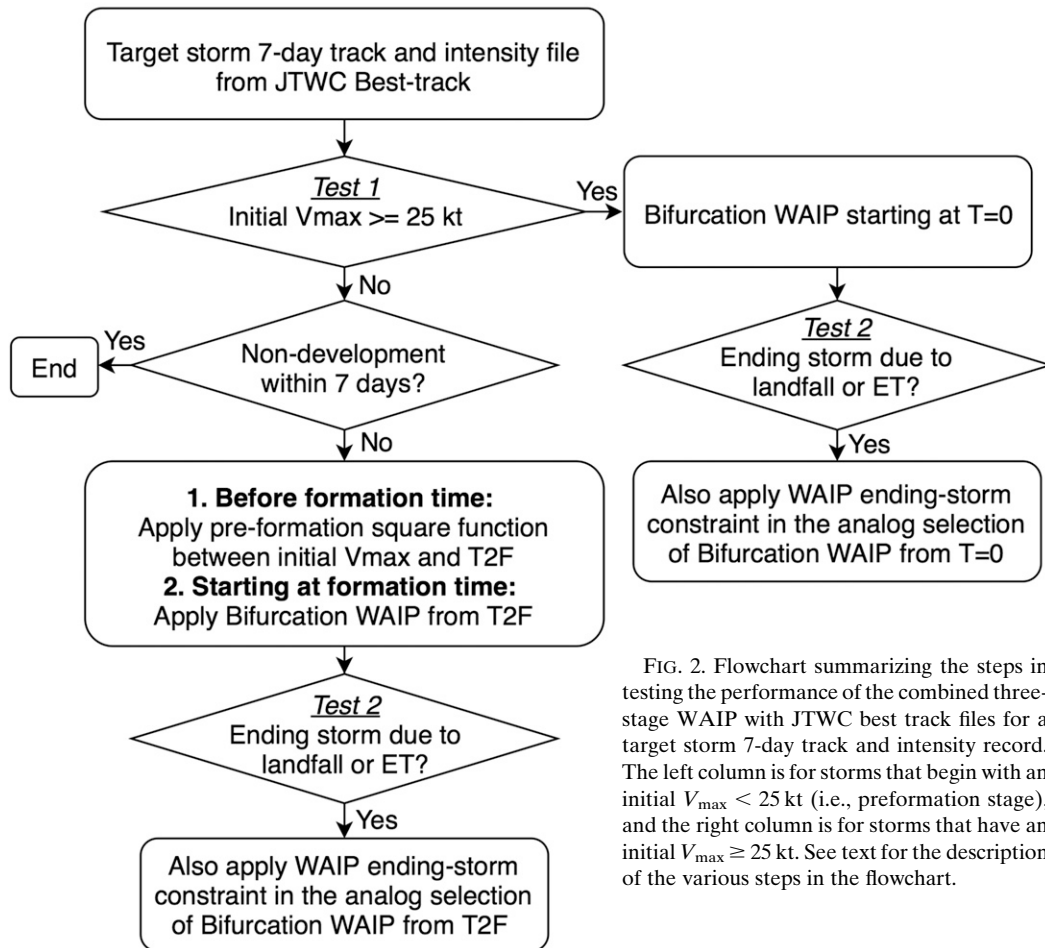
COMBINED 3-STAGE WAIP TEST WITH BEST-TRACKS

FIG. 2. Flowchart summarizing the steps in testing the performance of the combined three-stage WAIP with JTWC best track files for a target storm 7-day track and intensity record. The left column is for storms that begin with an initial $V_{max} < 25$ kt (i.e., preformation stage), and the right column is for storms that have an initial $V_{max} \geq 25$ kt. See text for the description of the various steps in the flowchart.

combined three-stage WAIP. The flowchart in Fig. 2 summarizes the steps in testing the performance of the combined three-stage WAIP with JTWC best track files. As indicated above, the original 7-day WAIP (Tsai and Elsberry 2015) was redeveloped with a training set of 70% randomly selected storms of the 2000–15 JTWC best track files and verified with an independent set of 30% of the storms. Rather than selecting only 10 historical analogs as in the Tsai and Elsberry (2015) WAIP version, a total of 16 analogs were selected as required for the bifurcation version (see the appendix for a description of this calculation). As indicated in Fig. 2, the bifurcation version of the WAIP is the central feature in the combined three-stage WAIP.

This optimum performance demonstration of the combined three-stage WAIP begins with a target storm 7-day track and intensity record from the JTWC best track file, and Test 1 is whether the initial intensity V_{max} is greater or less than 25 kt (Fig. 2). If the target storm is

at least a TC, the WAIP bifurcation version is initiated from time $T = 0$, with inputs of the 7-day track and the initial V_{max}. However, Test 2 is also applied to determine whether there is an ending-storm event due to a landfall or an extratropical transition along the 7-day target storm track. If there is not predicted to be an ending-storm event, the WAIP bifurcation version prediction to 168 h becomes the final intensity forecast and intensity spread guidance (Fig. 2, pathway on right side). If there is an ending-storm event, the historical analog selection is constrained such that the intensity at the last matching point with the target TC track cannot exceed 50 kt. Since the WAIP prediction is then constrained at $T = 0$ by the initial V_{max} and at the ending time by the 50-kt value, and also by the basic requirements to have a similar track and be within ± 30 days, the selected analog intensity evolutions tend to be quite similar so that final WAIP intensity forecast tends to be more accurate and have a smaller intensity spread (Tsai and Elsberry 2017b).

The more interesting and challenging pathway in Fig. 2 is when the initial intensity is less than 25 kt, that the target storm is in the preformation stage with an intensity of either 15 or 20 kt. The question is then whether this pre-TC circulation will not develop to at least 25 kt, which is very rare as indicated in Fig. 1a because almost all of the storms in the JTWC best track file during the 2000–15 study period attain at least 35 kt at some time during the life cycle. When the WAIP is moved to operational testing and applied to invests, the answer to the nondevelopment question will more frequently be "yes" and this will end consideration of such a pre-TC circulation.

If development within 7 days is an option, the first step (Fig. 2, middle column) is to apply the preformation squared function for the intensity evolution between the target storm initial time and the T2F, which is accurately known in this optimum performance demonstration because best track intensities are available (see section 2a). The second step is then to apply the WAIP bifurcation version starting at T2F with the inputs of the 7-day track, and the initial intensity of 25 kt at the T2F time (rather than at $T = 0$). As indicated above, Test 2 is applied before the bifurcation version begins to determine whether there is an ending-storm event due to a landfall or an extratropical transition along the 7-day target storm track. If no ending-storm event is predicted, the combination of the preformation intensity evolution plus the WAIP bifurcation version intensity forecast and intensity spread guidance becomes the final forecast. If an ending-storm event is predicted, the ending-storm constraint on the historical analog selection is applied so that the intensity at the last matching point with the target track cannot exceed 50 kt. Constraints at both the initial and ending times lead to more accurate WAIP intensity forecasts with smaller intensity spreads (Tsai and Elsberry 2017b).

c. Example of combined WAIP intensification stage and ending-storm prediction

An illustration of the original WAIP is given for pre-Typhoon Matsa (09W) at 1800 UTC 31 July 2005 (Fig. 3). The JTWC best track is indicated in Fig. 3a by the red circles each 12 h, and the 16 analog tracks in the original WAIP that start within ±30 days and have a similar track and initial intensity with this target storm are indicated with colored lines. Although Typhoon (TY) Matsu made landfall on the central China coast around 29°N, the original WAIP technique does not take this landfall into account in the selection of the 16 analogs. The 16 analog intensity evolutions in the original WAIP corresponding to the 16 tracks are shown in Fig. 3b. Note that about one-half of these intensity evolutions end between
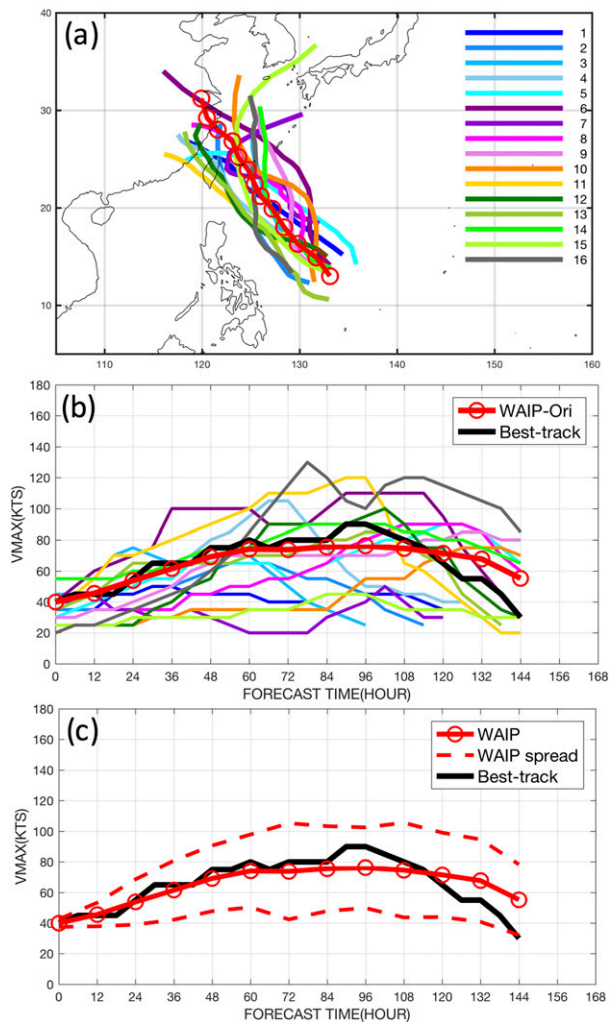

FIG. 3. Example for pre-Typhoon Matsu (09W) at 1800 UTC 31 Jul 2005 of a 7-day original WAIP forecast (without bifurcation version and without ending-storm constraint) with (a) JTWC best track (red circles and line) and 16 best historical analog tracks with colors according to rankings from 1st to 16th best (inset), (b) corresponding intensity (kt) evolutions for the 16 analogs [same colors as in (a)], the original WAIP intensity forecast (red circles and line), and the verifying intensity evolution (black line), and (c) repeat of the original WAIP forecast and verifying intensity evolutions from (b) plus the calibrated intensity spreads (kt; red dashed lines) relative to the WAIP forecast.

96 and 144 h, which corresponds to those analog tracks that also made landfall to the south of TY Matsu. However, the other half of the intensity evolutions have intensities ranging from 60 to 80 kt at 144 h, and these correspond to tracks in Fig. 3a that did not make landfall. The original WAIP (red circles in Fig. 3b) tends to "go down the middle" of the 16 analog intensity evolutions and has a weighted-mean V_{max} slightly smaller than 60 kt at 144 h when the verifying intensity is 30 kt (solid black line in Fig. 3b). The original WAIP and the

verifying intensity evolutions are repeated in Fig. 3c along with the intensity spreads each 12 h (dashed lines) that are calibrated to include 68% of the verifying intensities. Because the intensity spread among the 16 analogs in Fig. 3b is quite large (especially at 72 h with intensities ranging from 20 to 120 kt), the original WAIP intensity spread is also large (± 35 kt). Note that this original WAIP forecast is excellent until the last 24 h as Matsu was making landfall, and certainly the verifying intensities were well within the original WAIP intensity spread.

As indicated on the right side of Fig. 2, the Test 2 in the combined WAIP bifurcation approach is the check for an ending-storm event, which in this case was the landfall of Matsu. Consequently, the selection of the 16 analogs for the combined WAIP must also meet the ending-storm constraint that the intensity at the last matching time of the analog must be ≤ 50 kt. While a few of the analog landfalling tracks in Fig. 3a are again selected with this ending-storm constraint, additional analog tracks with landfalls replace nonlandfall tracks, and some new recurving storms are selected (Fig. 4a). More of the corresponding analog intensity evolutions (Fig. 4b) have larger V_{max} values, but note that every one of these analog intensity evolutions has an intensity less than 40 kt at 144 h. Consequently, the combined WAIP weighted-mean intensity at 144 h is just below 40 kt (Fig. 4b), which agrees much better with the verifying of 30 kt than in the original WAIP that did not have the ending-storm constraint (Fig. 3b). As indicated in Fig. 4c, the combined WAIP with the ending-storm constraint provides an excellent intensity forecast within about 5 kt through 144 h. However, the combined WAIP intensity spread at 72 h (Fig. 4c) is now even larger than for the original WAIP (Fig. 3c) with a range from 50 to 125 kt.

Such a large intensity spread will be automatically detected in the bifurcation version section of the combined WAIP. Two cluster intensity evolutions will be calculated with Cluster 1 (Fig. 5a) having the larger peak V_{max} and Cluster 2 (Fig. 5b) having the smaller peak V_{max}. Although Cluster 1 has overforecast the intensity of TY Matsu by about 15 kt at 72 h, overall this is an excellent forecast including a near-perfect intensity forecast at landfall at 144 h because of the additional ending-storm constraint. Note that the intensity spread about the Cluster 1 intensity evolution (Fig. 5a) is much smaller than in the original WAIP (Fig. 3c) or the combined WAIP with the ending-storm constraint but without the bifurcation application (Fig. 4c). By contrast, the Cluster 2 intensity forecast (Fig. 5b) has much smaller values, and the intensity spread is too small between 72 and 114 h as the verifying intensity (solid black
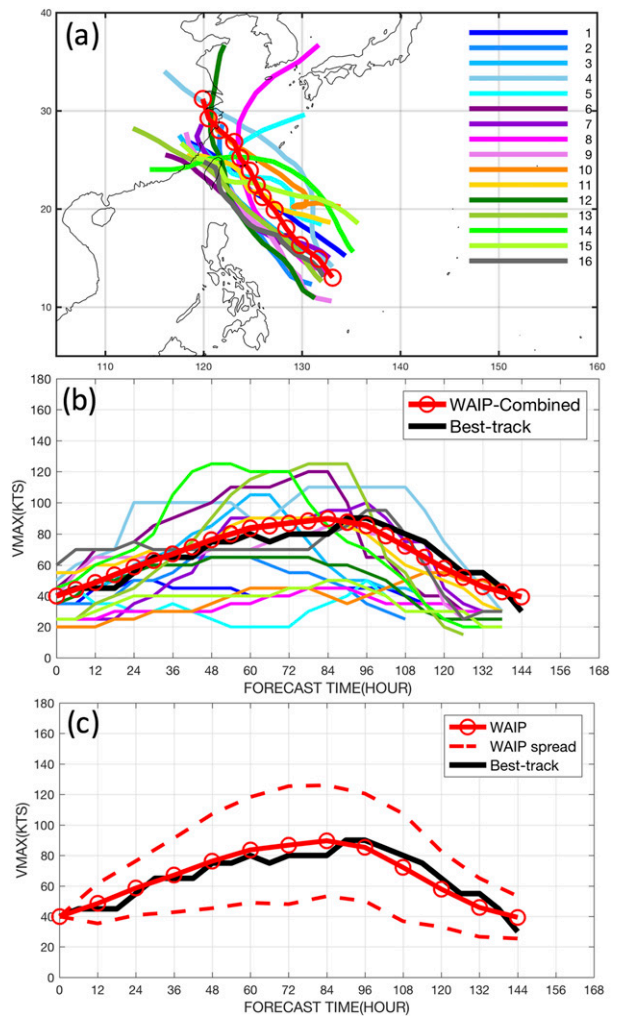

FIG. 4. Combined three-stage WAIP forecast as in Fig. 3, except here activating the ending-storm constraint on the selection of the 16 best historical analogs that must also have analog intensities < 50 kt at 144 h because Typhoon Matsu made landfall at that time. Note that some of the 16 analogs are different from Fig. 3, and the intensity spreads among these analogs satisfy the bifurcation situation condition (see text for description).

line) falls outside the intensity spread. Given the larger fraction of analog intensity evolutions with larger peak intensities (which will be grouped in Cluster 1) in Fig. 4b compared with the five analog nondevelopers, and the path of Matsu over a warm ocean region, Cluster 1 is clearly the more likely scenario. Whereas in these optimum performance examples the correct Cluster WAIP intensity is always selected, in operations the forecaster will need to make the selection based on other guidance and the recent performance of the technique.

The optimum performance of the combined three-stage WAIP technique will be described in section 3 with the qualifying statement that the initial intensity,
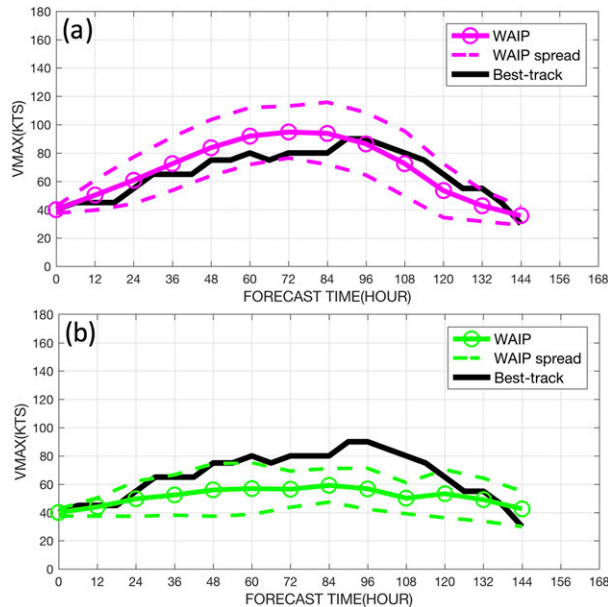
FIG. 5. Two cluster WAIP intensity (kt) evolutions (pink circles and lines) for the bifurcation situation among the 16 analogs in Fig. 4b vs the verifying intensity evolution (black lines) plus the calibrated intensity spreads (dashed lines) for (a) Cluster 1 evolution that has the larger peak V_{max} and (b) Cluster 2 evolution that has the smaller peak V_{max}.

the T2F, the correct bifurcation selections, and the ending-storm times have been specified from the JTWC best track files rather than operational conditions. While a preliminary sensitivity to the T2F test suggests little impact on the MAEs, the use of best track files that only contain storms that did intensify does not allow a test for Invests that may not intensify.

3. Optimum performance verifications for combined three-stage WAIP

The combined WAIP will be compared with the original 7-day WAIP (Tsai and Elsberry 2015) by first summarizing the 30% independent sample verifications during 2000–15 for the three stages and then the "All Sample." That is, the "Before Formation" subsample will be discussed in section 3a. The "After Formation" subsample when intensity bifurcation situations were detected will be first summarized in section 3b, and then the After Formation subsample when ending storms were detected will be described in section 3c. The All Sample verification for the combined three-stage WAIP will then be summarized in section 3d. Finally, the opportunity for earlier preformation combined WAIP guidance that has been possible from the JTWC best track files will be illustrated by using an ensemble storm-track forecast as the target for analog selection.

a. Before formation verification

Recall that the sample mean bias and MAEs in Figs. 1b and 1c, respectively, were based on describing the intensity evolutions as a squared function from initial times to the T2Fs based on the 1985–2015 JTWC best track files. The verification sample here is for the combined WAIP Before Formation forecasts with a sample size of at least 10 cases that began from an initial intensity of either 15 or 20 kt. Consequently, the sample size decreases from ~310 forecasts at 12 h (i.e., likely 20-kt initial intensities rapidly achieving a formation defined as 25 kt) to only ~20 forecasts at a T2F of 60 h (Fig. 6a), which is the last T2F considered here as there were <10 forecasts with a T2F of 72 h.

The verification is then between the combined WAIP Before Formation dataset versus the original WAIP forecast, which simply averaged the intensities of the 10 best historical analogs in the JTWC best track file that matched the target storm track and initial intensity. As expected from the first 60 h of the forecast times in Fig. 1c, the combined WAIP intensity MAEs in this independent sample during 2000–15 also have values of 1–2 kt when the T2Fs are in the range of 12–60 h (Fig. 6b). By contrast, the MAEs for the original WAIP increase rapidly to 10 kt for T2F = 36 h and more than 20 kt for T2F = 60 h. The explanation is that the JTWC best track file from which the 10 best historical analogs are selected has a large majority of storms that do achieve tropical storm (34 kt) intensity, and thus for those original WAIP analogs with T2F = 36 h (60 h) the average intensity is biased high by 10 kt (20 kt). This large positive bias from the original WAIP analog selection was the motivation to constrain the WAIP intensities via the squared intensity evolution knowing the T2F when the intensity < 25 kt.

Another verification metric is the correlation coefficients of the combined WAIP intensity forecasts with the verifying intensities each 12 h (Fig. 6c). Because the Before Formation WAIP forecasts are constrained to be between the initial intensity of either 15 and 20 kt and an intensity of 25 kt at the T2F, the correlation coefficients for the combined WAIP are very high (0.8) even when the T2F is 60 h. By contrast, the correlation coefficients for the original WAIP forecasts decrease rapidly to 0.3 (9% explained variance) at 36 h, which may be attributed to the rapidly increasing MAEs in Fig. 6b.

b. After formation with bifurcation verification

Because the WAIP bifurcation version (Tsai and Elsberry 2018) is in a journal that is not widely available, the key steps in the development of this version of WAIP are described in the appendix. The flowchart in
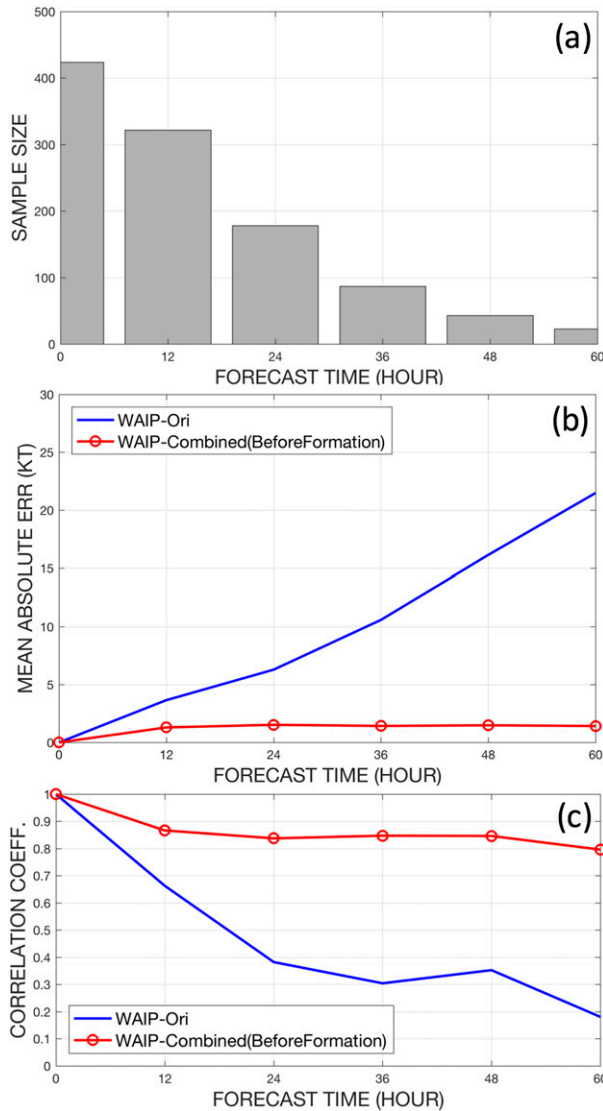
FIG. 6. Verification of the independent sample combined WAIP intensities (kt) for the Before Formation stage with (a) sample sizes when at least 10 cases are available for verification for the forecast time to formation (T2F) on the abscissa, (b) mean absolute errors (kt) for the combined WAIP (red circles and line) vs the original WAIP of Tsai and Elsberry (2015), and (c) correlation coefficients of the combined WAIP and original WAIP intensity forecasts with the verifying intensities.

Fig. A1 summarizes the objective detection of an intensity bifurcation situation, which is based on the weighted-mean intensity spread of 16 analogs [Eq. (A1)]. If that weighted-mean spread (WMS) for a WAIP forecast exceeds the overall sample WMS, a bifurcation intensity situation exists and a hierarchial cluster analysis is applied to the 16 analog intensity evolutions each 12–168 h to separate them into two clusters. The weighted-mean intensities and weighted-mean intensity

spreads are calculated with Cluster 1 having the larger peak $V_{\max}$ in Fig. 5a and the Cluster 2 with the smaller peak $V_{\max}$ in Fig. 5b. Note that the Cluster 1 selection for the large intensity spread case in Fig. 4b was made up of various peak $V_{\max}$ values occurring at various forecast intervals, so the weighted-mean intensity will not necessarily be a conservative estimate of the actual peak $V_{\max}$. Tsai and Elsberry (2018) provide some guidance-on-guidance for the cluster selection based on the numbers of the 16 analogs in each cluster. That is, the selection of Cluster 1 or of Cluster 2 is proposed to be the cluster that has a majority of the 16 analogs. If the forecaster is still uncertain after examining other intensity guidance products, the forecaster should select the original WAIP that is the mean of all 16 analogs as the MAEs will not be that much larger, and thus not risk an ''All Wrong'' cluster selection.

Since this is an optimal performance evaluation that is based on JTWC best track information, the verification here assumes a correct selection of either Cluster 1 or Cluster 2 in each bifurcation situation. A comparison will also be given with the original WAIP forecast that is simply a weighted-mean of all 16 analogs and thus tends to go down the middle of the Cluster 1 and Cluster 2 intensity evolutions. The sample sizes for the independent sample of combined WAIP forecasts with a bifurcation are shown in the inset of Fig. 7a. A total of ~525 cases are available from 12 to 72 h with a decrease to ~450 cases at 132 h and then a rapid decrease to ~215 cases at 168 h. The importance of an All Correct cluster selection is indicated by the MAEs in Fig. 7a since the 60 h intensity errors are only 15 kt and the MAEs remain below 17 kt through 144 h. By contrast, the All Wrong cluster selection has a MAE of 39 kt at 72 h, which may be attributed to every rapid intensification (rapid decay) situation being wrongly selected as a Cluster 2 (Cluster 1) intensity evolution with a smaller (larger) $V_{\max}$. Furthermore, the All Correct MAEs are 5–6 kt smaller in the 60–108-h forecast period than the original WAIP forecasts. Thus, bifurcation situations are opportunities for the forecaster to add value relative to a ''down the middle'' All 16 analog intensity evolution by a correct selection between the two cluster intensity evolutions provided in the combined WAIP.

The All Correct combined WAIP cluster intensity forecasts also have high correlation coefficients with the verifying intensities (Fig. 7b). Although the correlation coefficients decrease to 0.7 at 72 h, they then increase to between 0.75 and 0.79 for the 96–144 h forecast interval. By contrast, the All Wrong cluster intensity selection has correlation coefficients that decrease very rapidly to less than 0.4 before 48 h. Even though the MAEs at 72 and 84 h for the original WAIP do not seem that bad
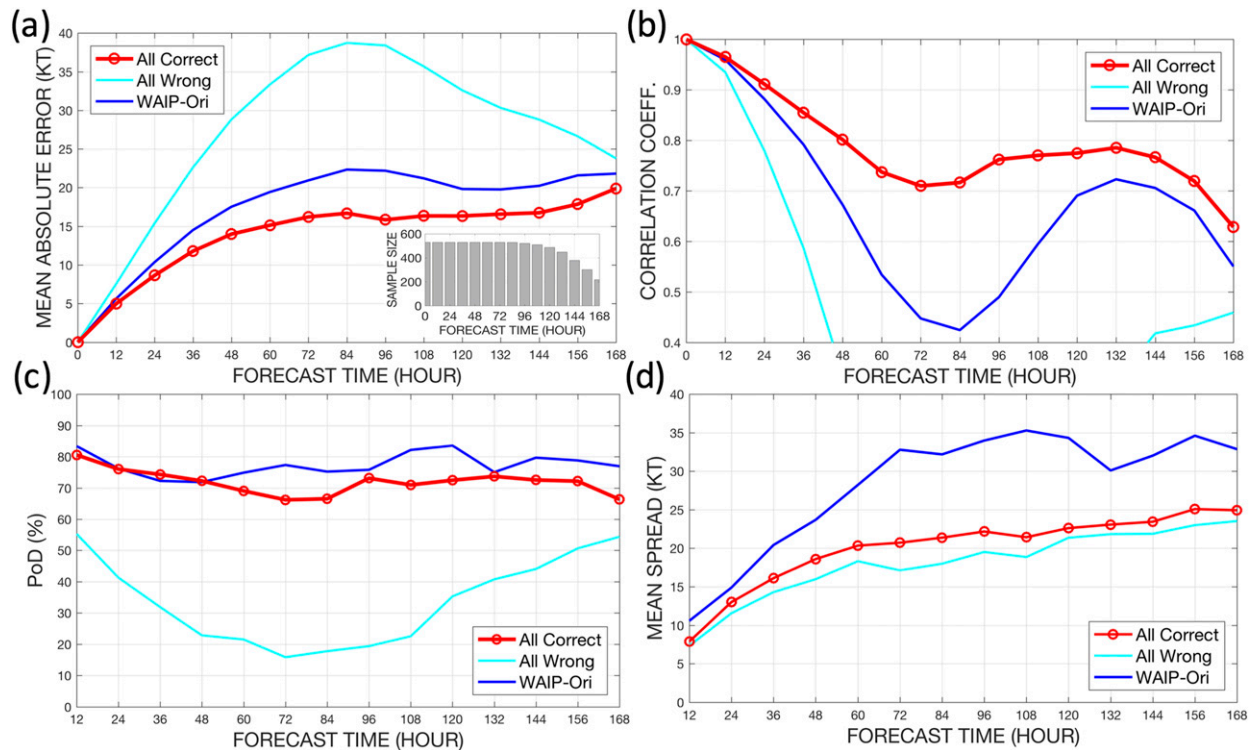
FIG. 7. Verification of the independent sample combined WAIP intensity (kt) forecasts for the After Formation period with bifurcation situations for the All Correct (red circles and lines) and the All Wrong (light blue lines) selections of the Cluster 1 and Cluster 2 intensity evolutions, and compared with the original WAIP intensity evolutions (dark blue lines) in terms of (a) MAEs (kt), (b) correlation coefficients with verifying intensities, (c) probability of detection, and (d) sample-mean intensity spreads (kt).

(5–6 kt larger) compared to the All Correct MAEs (Fig. 7a), the original WAIP intensity correlation coefficients have decreased to less than 0.45 at those time intervals (Fig. 7b). However, the original WAIP correlation coefficients recover to 0.7 at 132 h when the MAE difference is only 3 kt. Thus, the correct cluster intensity evolution selection is most important in the 48–120 h of the forecast and is likely related to correctly picking Cluster 1 with the larger peak V_{max} in rapid intensification cases.

The rapid increase in the combined WAIP MAEs from 12 to 72 h and then the slow increase in the MAEs for the remainder of the 7-day forecast (Fig. 7a) may be an indication of a limit to predictability of TC intensity for this technique. Thus, it is important to provide an intensity uncertainty metric for a 7-day WAIP forecast. Tsai and Elsberry (2015) had provided a calibrated intensity spread that was designed such that the original WAIP intensity forecasts would lie within that calibrated intensity spread for 68% of the forecasts (see the appendix for summary of the steps in the intensity spread calibrations). If the independent sample of original WAIP forecasts had exactly the same characteristics as the training sample used in the calibration calculation, the

probability of detection (PoD) for the original WAIP in Fig. 7c (dark blue line) would have been equal to 68% during the entire 12–168-h period. However, the calibrated intensity spreads are overly large for the independent sample of original WAIP forecasts with values exceeding 68%, and thus they are "overdetermined."

In this combined WAIP study, only the All Sample intensity spread calibration was applied, rather than developing separate bias corrections for the bifurcation subsamples as had been done by Tsai and Elsberry (2015). As indicated in Fig. 7c (red line), the All Correct bifurcation WAIP forecasts also have slightly overdetermined PoDs except at 72–84 h and again at 168 h. Applying that All Sample original WAIP calibration of intensity spreads is not effective for the All Wrong bifurcation WAIP forecasts (Fig. 7c, light blue line) as the PoDs are well below 68%.

As indicated above, the bifurcation situations by definition have large intensity spreads among the analogs. Indeed, the sample-mean intensity spreads for the independent sample of original WAIP forecasts are very large (Fig. 7d, dark blue line), and this accounts for the high PoDs for the original WAIP in Fig. 7c. By contrast, the success of intensity spread calibration for the All
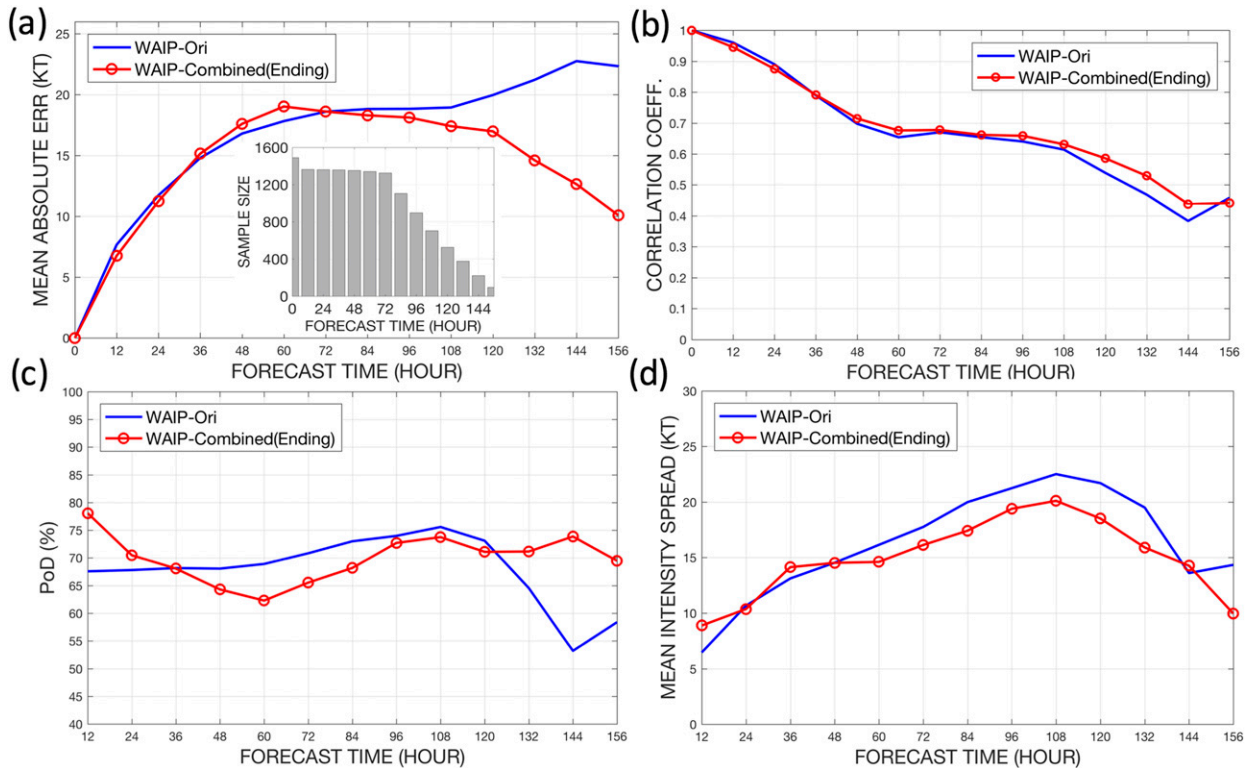
FIG. 8. Verification as in Fig. 7, except for combined WAIP intensity forecasts After Formation with ending-storm events.

Correct combined WAIP bifurcation forecasts leads to much smaller sample-mean intensity spreads (Fig. 7d, red line) that still ensure a PoD near the desired 68%. Even though the All Sample intensity spread calibration has been applied, an All Correct combined WAIP cluster intensity selection with small MAEs (Fig. 7a) after 48 h also has relatively small intensity spreads about those combined WAIP intensity forecasts (Fig. 7d). Applying the All Sample intensity spread calibration to the All Wrong bifurcation WAIP intensity forecasts led to small sample-mean intensity spreads (Fig. 7d, light blue line). However, these unrealistic small intensity spreads do not enclose the large MAEs of the All Wrong forecasts (Fig. 7a), so the PoDs for the All Wrong forecasts are very poor (Fig. 7c).

In summary, the WAIP bifurcation version of Tsai and Elsberry (2018) is an important component of the After Formation combined WAIP forecasts. In this optimum performance evaluation, the All Correct cluster intensity selection illustrates the potential benefits, but in operations the forecaster will need to make that cluster selection.

c. After formation with ending-storm verification

As indicated in the flowchart in Fig. 2, the application of the bifurcation WAIP either from $T = 0$ or $T =$ T2F first requires a test whether an ending-storm event

(landfall or extratropical transition) will occur during the 7-day forecast interval. The sample sizes for the independent sample of combined WAIP forecasts with ending-storm events are shown in the inset of Fig. 8a. Almost 1400 WAIP forecasts have an ending event within 72 h, and then the number of events decreases to ~900 at 96 h, ~500 at 120 h, and ~100 at 156 h. Recall that the historical analog selection constraint for these ending-storm events is that the analog intensity must be ≤50 kt at the ending time.

The verification for this independent sample of combined WAIP forecasts (Fig. 8a) is similar to the Tsai and Elsberry (2017b) test. That is, the most important result is that the MAEs begin to decrease with increasing forecast intervals greater than 72 h and decrease to only 10 kt at 156 h (Fig. 8a). By contrast, the original WAIP MAEs continue to increase with forecast intervals greater than 120 h and are 12 kt larger at 156 h than for the combined WAIP forecasts. While one might have also expected a substantial improvement in the combined WAIP intensity forecast correlation coefficients with the verifying intensities, the actual improvement is less than 0.05 and only in the 120–144-h intervals (Fig. 8b).

As was the case for the independent sample of original WAIP and combined WAIP bifurcation intensity

verifications in terms of PoD (Fig. 7c), the original WAIP has the better PoD verification (e.g., near 68%) for at least the first 72 h than does the combined WAIP with the ending-storm constraint (Fig. 8c). The explanation is again that the All Sample intensity spread calibration has been applied to both versions rather than calculating a separate intensity spread calibration just for the combined WAIP ending-storm forecasts. However, the larger original WAIP MAEs at 132–156 h (Fig. 8a) are associated with smaller PoDs than for the combined WAIP (Fig. 8c) that has smaller MAEs in this time interval. The sample-mean intensity spreads for the combined WAIP are also smaller than for the original WAIP from 60 to 132 h and are only 10 kt at 156 h (Fig. 8d). So in addition to the combined WAIP with the ending-storm constraint having a very small MAE = 10 kt at 156 h, the intensity uncertainty at 156 h is also very small (±10 kt).

The overall good performance for this independent sample of combined WAIP forecasts with ending-storm events demonstrates the advantage of constraining the historical analog selection for both the initial intensity and the final intensity, as well as requiring a similar TC track within ±30 days of the date of the target TC. Thus, including the ending-storm stage as an integral component of the combined three-stage WAIP is expected to have a substantial contribution to the overall success of the combined WAIP.

d. Combined three-stage WAIP verification

In addition to the After Formation with bifurcation situations (section 3b) and with ending-storm events (section 3c), there are other After Formation combined WAIP forecasts that extend to 168 h without involving either bifurcations or ending-storm events. Because these other After Formation forecasts will essentially be identical to the original WAIP, those forecasts will not be compared separately here. However, those other After Formation forecasts are included in the All Sample combined three-stage WAIP verification described in this section.

The independent All Sample of 30% randomly selected storms during 2000–15 contains ~2100 forecasts in the 12–72-h period, but then the sample decreases to ~600 forecasts at 168 h (Fig. 9a, inset). The improved MAEs for the combined WAIP intensity forecasts (Fig. 9a, red circles and line) in the first 72 h relative to the original WAIP (Fig. 9a, blue line) may be attributed to the very small intensity errors for the Before Formation period (Fig. 6b), even though that sample size is not that large (Fig. 6a). For example, the 90 forecasts with a T2F of 36 h would have essentially zero intensity errors for the first 36 h, and then the After Formation

WAIP intensity errors would only average ~10 kt after another 36 h according to the error growth rate in Fig. 6b. Not only would this contribute to an error reduction at 72 h, this improvement should be sustained beyond 72 h when the WAIP error growth rate is smaller.

Another contribution to the smaller combined WAIP intensity forecast errors in the 72–144-h forecast interval is the All Correct bifurcation WAIP Cluster intensity forecasts (Fig. 7a). Even though there are only ~525 such bifurcation cases (Fig. 7a, inset), the All Sample MAEs are smaller than the original WAIP MAEs in this forecast interval (Fig. 9a). Although the WAIP ending-storm stage has very small MAEs at 156 h (Fig. 8a), this small sample of ~50 forecasts is not sufficient to offset the increasing MAE trend for the All Sample combined WAIP that includes ~600 forecasts at 168 h (Fig. 9a). Nevertheless, the MAE reductions from the special treatment of the three stages in the TC life cycle in the combined WAIP are substantial relative to the original WAIP MAEs.

A homogeneous comparison with the JTWC intensity errors has not been attempted because the independent sample has been randomly drawn from the 2000–15 seasons. Rather, a comparison has been made with the JTWC 18-yr (2000–17) average intensity errors at 24, 48, 72, 96, and 120 h, which were 11.3, 16.8, 20.2, 22.5, and 24.2 kt (J. Darlow, JTWC, 2018, personal communication). Since these JTWC average intensity errors are quite similar to the original WAIP intensity errors in Fig. 9a, the improvement of the combined WAIP intensity errors relative to the original WAIP suggests that the combined WAIP technique may provide useful guidance for the JTWC forecasters. Again, the qualifying statement that this evaluation is for optimum performance of the combined WAIP in that best track inputs have been used, and an All Correct bifurcation cluster has been assumed.

The improvement of the combined WAIP compared to the original WAIP in terms of correlation coefficients with the verifying intensities is particularly noteworthy at 72 h and beyond (Fig. 9b). Specifically, it was the After Formation with bifurcation situations subsample that had correlation coefficients that increased to around 0.8 between 96 and 144 h (Fig. 7b) similar to the All Sample (Fig. 9b). Although the Before Formation combined WAIP correlation coefficients were very high (Fig. 6c), the sample sizes were small (Fig. 6a). Furthermore, the original WAIP correlation coefficients were already high, so the Before Formation cases did not have a significant impact in the All Sample (Fig. 9b). The combined WAIP with ending-storm events subsample also cannot be used to explain the improved correlation coefficients for the All Sample as that subsample had decreasing correlation
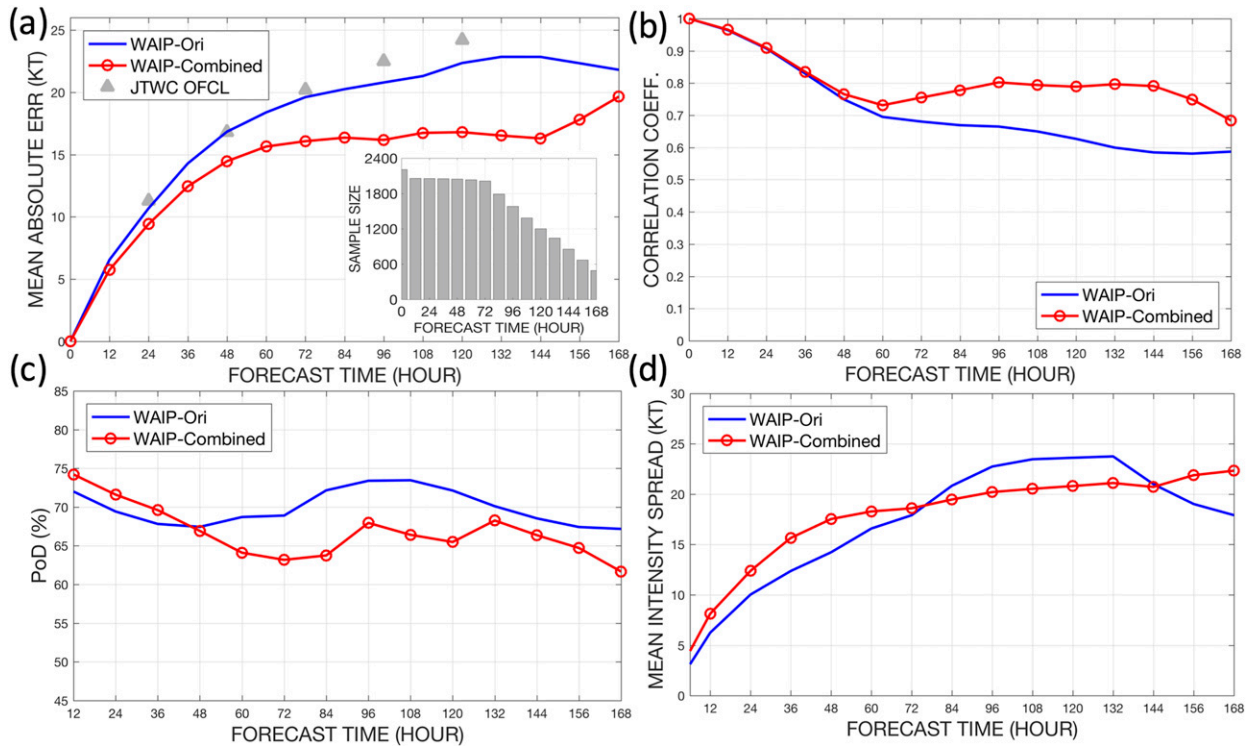
FIG. 9. Verification as in Fig. 7, except for All Sample combined WAIP intensity forecasts. The 2000–17 average JTWC official intensity forecast errors each 24 through 120 h are indicated as triangles in (a).

coefficients after 72 h (Fig. 8b). Thus, an opportunity exists for highly accurate (correlation coefficients of 0.80) intensity forecasts extending to 6 days if the forecaster always selects the correct WAIP cluster intensity evolution in bifurcation situations.

The PoD for the independent set of original WAIP forecasts (Fig. 9c, dark blue line) is close to the desired 68% over most of the 168 h forecast interval. However, the application of the All Sample calibration for the independent sample of combined WAIP forecasts is not as successful with higher PoDs at 12–24 h and too low PoDs in the 60–84-h forecast intervals (Fig. 9c, red circles and line). This need for a separate calibration of the intensity spreads for the combined WAIP forecasts also affects the sample-mean intensity spreads (Fig. 9d). Because the independent sample of the original WAIP intensity spreads is relatively better calibrated, the intensity spreads out to 72 h required to include 68% of the verifying intensities are smaller (Fig. 9d, dark blue lines) than for the less well calibrated combined WAIP intensity spreads (Fig. 9d, red circles and line). In the 84–132-h forecast interval where the original WAIP intensity spreads are larger, the corresponding PoDs are too high (Fig. 9c, dark blue line). Even though the combined WAIP intensity spreads are not well calibrated in terms of the PoDs not all being equal to 68%

(Fig. 9c, red circles and line), the small (~1 kt) increase in the combined WAIP intensity spread (Fig. 9d, red circles and line) from 84 to 144 h indicates nearly constant uncertainty about a nearly constant MAE (Fig. 9a) has been achieved by the combined WAIP. Thus, the JTWC will have a more accurate guidance product that will potentially allow them to extend their intensity forecasts to 7 days and provide a useful intensity uncertainty measure that their customers can use to evaluate their risk in terms of the TC intensity given an accurate track forecast beyond 72 h.

The areal distribution of the improvements in these combined three-stage WAIP intensity forecasts over the original WAIP forecasts are shown for four forecast intervals in Fig. 10. At 24 h (Fig. 10a), improvements are achieved over most of the western North Pacific and South China Sea. The exceptions are primarily over the 10° latitude × 10° longitude areas that include Taiwan and the central and northern Philippines, which may be attributed to WAIP forecasts initiated east of these islands since this technique does not account for landfall on islands with significant topography. That is, such island landfalls are not considered to be ending-storm events for which a constraint would be put on the historical best analogs to have intensities ≤ 50 kt. At 48 h (Fig. 10b), the largest improvements tend to be at low
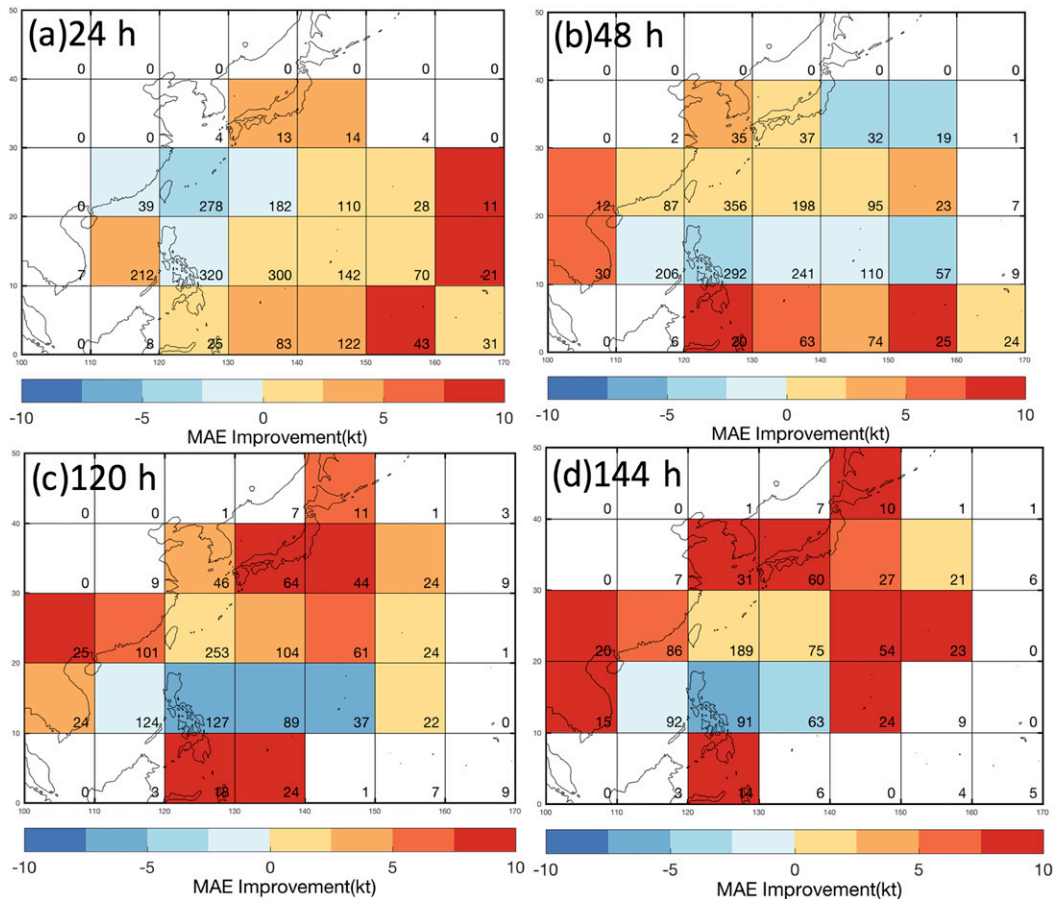
FIG. 10. Areal distribution of independent All Sample combined WAIP MAE (kt; color contour scale below) improvements relative to the original WAIP forecasts at (a) 24, (b) 48, (c) 120, and (d) 144 h. The sample size in each box is indicated, and color shading is only provided if the sample size is at least 10.

latitudes in the Philippine Sea, which may be attributed to the Before Formation component in the combined WAIP (not shown), and in landfall areas of East Asia and Southeast Asia, which is attributed to the ending-storm component. The most frequent areas of non-improved combined WAIP forecast are over and just to the east of the Philippines and over the South China Sea to the west. Again, this deficiency is attributed to westward-translating TCs with intensity changes related to passage over the Philippines that were not properly predicted by the present ending-storm component of WAIP. Similar improvements (with larger magnitudes) apply at 120 h (Fig. 10c) as at 48 h both for the landfalls along the East Asia and Southeast Asia coast and for the not well-predicted intensity changes with westward-translating TCs over the Philippines. The areal pattern of improvement and nonimprovement of combined WAIP intensity forecasts is continued at 144 h (Fig. 10d). The general conclusion is that the All Sample combined three-stage WAIP has improved intensity forecasts especially

over most of the western North Pacific and East Asia coasts, but the forecaster must make adjustments to account for the intensity changes associated with westward-translating TCs over the northern and central Philippines.

The areal distributions of the Before Formation stage WAIP improvements are between the equator and 20°N, and the primary time contribution is to the 12- and 24-h forecasts (not shown). The areal distributions of the improvements due to ending-storm events of landfall and extratropical transitions may be inferred in Fig. 10, as described above. Over the remainder of the western North Pacific, the combined WAIP improvements over the original WAIP are primarily due to the After Formation WAIP with bifurcations (Fig. 11). To again present the optimum performance, the All Correct selections of WAIP Cluster 1 or Cluster 2 intensity evolutions have been assumed in these bifurcation situations. Recall from Fig. 8a that the success of the WAIP forecasts with bifurcations was the improvement over the original
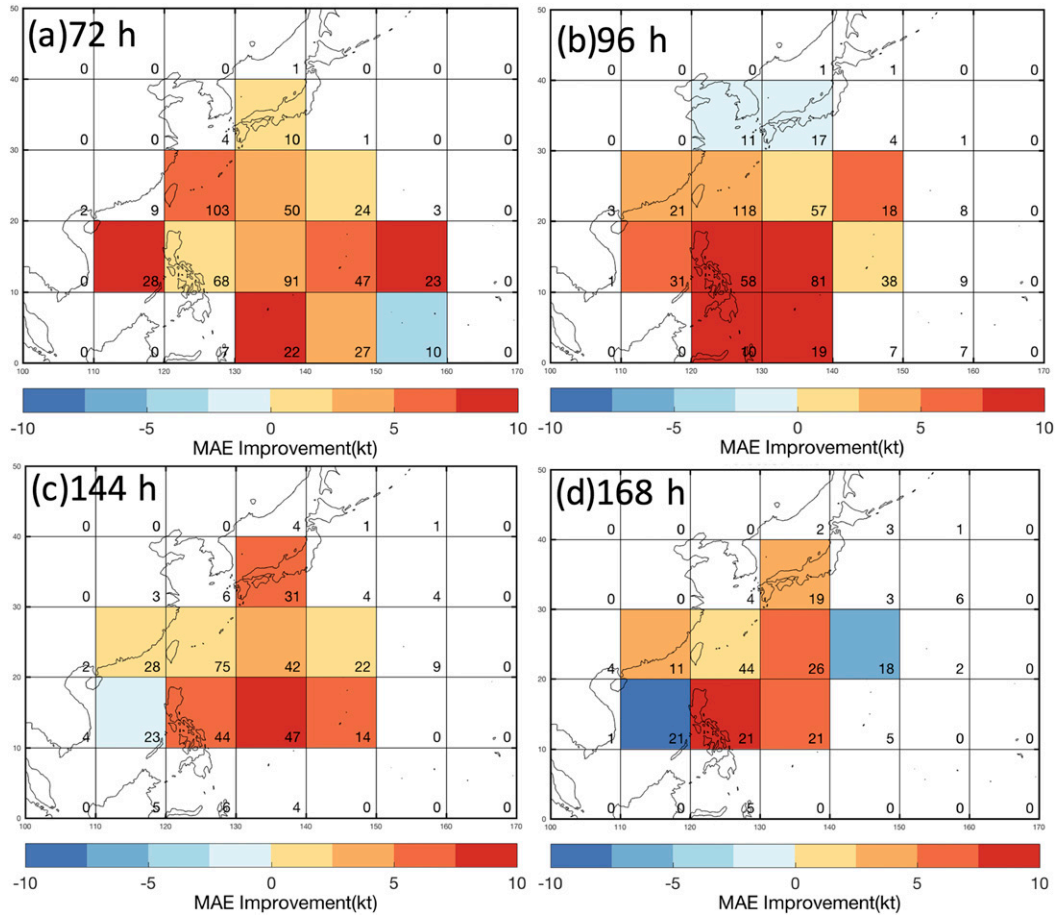
FIG. 11. Areal distribution of independent sample combined WAIP MAE (kt; color contour scale below) improvements as in Fig. 10, except just for After Formation bifurcations in which All Correct cluster intensity forecasts have been selected.

WAIP during the first 72 h, and then that improvement was sustained over the remainder of the forecast intervals. The areal distribution of that bifurcation combined WAIP improvement at 72 h is illustrated in Fig. 11a. Note that improvements are achieved for all areas (except for between 0°–10°N, 150°–160°E), and especially for the large samples in the 20°–30°N, 120°–130°E and 10°–20°N, 130°–140°E areas where rapid intensifications over high sea surface temperatures might be expected. Even the area over the northern and central Philippines has small improvement in these bifurcation situations. At 96 h (Fig. 11b), the combined WAIP forecasts involving bifurcations are improved over all areas south of 30°N, which includes sample-mean improvements of 7.5–10 kt over the area from 0°–20°N, 120°–140°E that includes all of the Philippines. These improvements continue at 144 h (Fig. 11c) over most of the area, except the South China Sea has a small (0–2.5 kt) degradation. Even though a larger (7.5–10 kt) degradation over the South China Sea and an area 20°–30°N,

140°–150°E occurs for the 168 h combined WAIP forecasts (Fig. 11d), the remainder of the western North Pacific with at least 10 bifurcation forecasts has improvements relative to the original WAIP. Therefore, the addition of the bifurcation version WAIP to the combined, three-stage WAIP has contributed to the largest areas with improvements in the western North Pacific.

e. Case study demonstration of preformation stage

While the improvement of the combined WAIP MAEs during the preformation stage relative to the original WAIP in Fig. 6b is very large after 36 h, the sample sizes for this optimum performance demonstration then become very small. That is, the JTWC best track file contains very few cases for which the first entry is more than 36 h prior to the T2F. Furthermore, the squared function representation of the intensity evolution from a well-known initial intensity to a specified T2F over less than 36 h is likely to be quite accurate (Fig. 1c) with a small mean bias (Fig. 1b) and with a small mean spread (Fig. 1d).
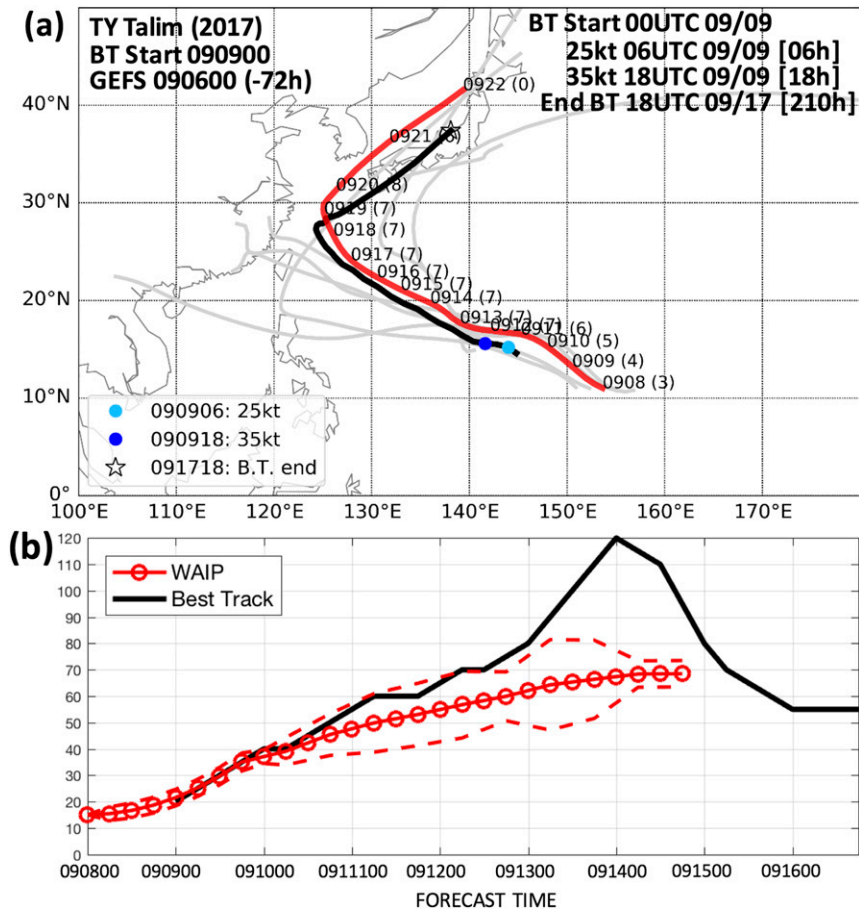
FIG. 12. (a) GEFS-based weighted mean vector motion (WMVM) ensemble storm-track forecast (red line) from 0000 UTC 6 Sep 2017 labeled with month/day and number of ensemble member tracks in gray that has been matched with the JTWC best track of TY Talim (2017, black line) that starts at 0600 UTC 9 Sep. (b) Combined WAIP intensity (red circles) and intensity spread (red dashed line) vs best track intensity (black line).

Elsberry et al. (2011) had demonstrated that the ECMWF ensemble weighted mean vector motion (WMVM) track forecasts in the western North Pacific during the 2009 season typically began the pre-TC circulations 2–3 days before the first entry in the JTWC best track file. A similar capability for early predictions of the beginning of pre-TC circulations by the NCEP Global Ensemble Forecast System (GEFS) was demonstrated during the 2015 season. Since the combined WAIP intensity technique after formation only requires a track forecast and an initial intensity, these WMVM track forecasts can provide the required WAIP input even before the JTWC has issued an invest.

A case study with a GEFS track forecast for TY Talim (2017) that according to JTWC started at 0000 UTC 9 September 2017 with an intensity of 25 kt (35 kt) just 6 h (18 h) later (black line, Fig. 12a) will be presented as an example of how combined WAIP predictions could provide intensity predictions for much of the 8.75 day life cycle of TY Talim starting from the preformation stage. Note that this GEFS forecast from 0000 UTC 6 September began the pre-Talim circulation at 0000 UTC 8 September, which provides a preformation period of 42 h before the T2F (here 35 kt is utilized) at 1800 UTC 9 September (i.e., day 3.75 in that GEFS forecast). Assuming an initial intensity of 15 kt at the 0000 UTC 8 September starting time in the GEFS forecast, and for this optimum performance demonstration the actual T2F (35 kt) is specified at 1800 UTC 9 September, the preformation stage of the combined WAIP intensity forecast (Fig. 12b) is simply a squared function connecting the 15-kt initial intensity with 35 kt at the T2F. While no JTWC intensity estimates are available for validation during the first 24 h, this preformation stage WAIP forecast is coincident with the JTWC best track intensities from their start at 0000 UTC

9 September to the T2F (35 kt) at 1800 UTC 9 September (Fig. 12b).

Although the GEFS WMVM track forecast has a maximum of only 7 (out of a possible 21) members that are widely spread (Fig. 12a, gray lines), the WMVM track forecast after the formation time has good agreement with the path of Talim, but has an increasingly large slow along-track bias due to the track spread. That is, this GEFS forecast has Talim recurving at ~0000 UTC 18 September, but Talim had recurved earlier and the JTWC best track actually ends over central Japan already at 1800 UTC 17 September (Fig. 12a, black line). In operations, the forecaster will need to account for a slow bias when the ensemble track spread is large.

The combined WAIP intensification stage forecast that starts at the T2F is also coincident with the JTWC best track intensities for the first 24 h. Whereas the WAIP does not intensify Talim as much as observed during the next 48 h, the JTWC best track intensities are within the 68% intensity spread (dashed red lines, Fig. 12b). However, this statistical–dynamic WAIP technique does not predict the rapid intensification from 80 kt at 0000 UTC 13 September to 120 kt at 0000 UTC 14 September, which are days 7 and 8 of this 0000 UTC 6 September GEFS forecast. In principle, a JTWC forecaster could have on 6 September issued an alert that a pre-TC circulation would start near 10°N, 154°E on 8 September, become a tropical storm around 10 September, and likely become a typhoon as early as 12 September (based on the upper bound of the intensity spread). In fact, Talim did become a typhoon at 0000 UTC 12 September (Fig. 12b, triangles).

The JTWC already has the GEFS WMVM track forecasts as in Fig. 12a. In such a situation with a track forecast that likely would threaten Japan, the JTWC would certainly want to issue an alert as soon as possible. The combined WAIP technique with its preformation stage will provide JTWC with an additional capability to predict the likely maximum intensity within the next 7 days. Since it would be easy for the forecaster to specify an initial intensity of the pre-TC circulation in Fig. 12a, and specifying an ending-storm time within 7 days is not required in this case, the key input to the combined WAIP forecast in Fig. 12b is the T2F. A technique to provide the JTWC forecaster an objective estimate of the T2F based on the GEFS forecast variables is in operational testing at JTWC and will be reported at the end of the 2019 season.

4. Summary and conclusions

The original 7-day WAIP intensity and intensity spread forecast technique (Tsai and Elsberry 2015) that

considered all stages of the TC life cycle was found to have large intensity spreads, especially during the intensification stage, and an increasingly large positive intensity bias with increasing forecast intervals. Thus, Tsai and Elsberry (2014b, 2018) developed a bifurcation version of WAIP that calculated two cluster forecast intensity evolutions with separate intensity spreads, and demonstrated that an All Correct selection of the cluster forecasts resulted in considerably smaller MAEs and corresponding intensity spreads. Tsai and Elsberry (2017b) developed a WAIP version for ending-storm events of landfall and extratropical transition, or simply nondevelopment during the 7-day forecast interval, which eliminated the increasing positive intensity bias with time. In the ending-storm WAIP version, the MAEs begin to decrease after 72 h and are only 10 kt at 156 h.

In this optimum performance evaluation in which the T2F is known, a special WAIP intensity and intensity spread forecast approach for the preformation stage is demonstrated to have very small MAEs and intensity spreads. Thus, a combined three-stage WAIP version is developed starting with the preformation stage and continuing with the Tsai and Elsberry (2018) bifurcation version for the intensification stage and then the Tsai and Elsberry (2017b) ending-storm version of WAIP. An optimum version of WAIP is evaluated as the JTWC best tracks and intensities are utilized as inputs, and a correct selection of the cluster WAIP intensity in bifurcation situations is assumed. With these qualifying conditions, a substantial reduction in MAEs has been demonstrated, and the intensity spreads are relatively small as well, during all three stages of the TC life cycle considered here. It is emphasized that these combined WAIP forecasts can be calculated on a desktop computer in about 1 min.

One of the motivations for this optimum performance evaluation of the combined three-stage WAIP is to demonstrate to the forecasters the value that they can add relative to the original 7-day WAIP (Tsai and Elsberry 2015) by (i) an accurate specification of the T2F; (ii) a correct selection of the Cluster 1 or Cluster 2 WAIP intensity evolutions in bifurcation situations; and (iii) specification of the ending-storm time along the JTWC official track forecast. As indicated in Fig. 1a, a large majority of western North Pacific pre-TC circulations identified by JTWC achieve 25-kt intensities within 72 h, and JTWC issues TC Formation Alerts that indicate the likelihood of a formation within 72 h. An objective technique has been developed and is in operational testing to identify a T2F (defined as 35 kt rather than 25 kt as in this study) utilizing the lower-tropospheric and upper-tropospheric warm core magnitudes (WCM) along

ensemble storm-track forecasts as in Fig. 12a. The time series of these WCMs also provides an indication of nondevelopment of a predicted ensemble storm, and also an indication of the timing of an extratropical transition, which along with the predicted landfall time along the official track forecast as the ending-storm condition, are the required inputs to WAIP. As described in the appendix and in section 3b, objective guidance is provided as to when the WAIP intensity spread indicates a bifurcation situation exists, and two WAIP cluster intensity evolutions with separate intensity spreads are provided. Because these bifurcation situations are primarily during the intensification phase, and the Cluster 1 intensity evolution with the larger peak $V_{\max}$ would be an indicator of more rapid intensification, we have confidence that an experienced forecaster will almost always make the correct selection of Cluster 1. When the forecaster is uncertain as to the selection, the All 16 analog intensity evolution will still be a reasonable alternative. Thus, the combined WAIP is in operational testing during the typhoon season in the western North Pacific by the JTWC, and a report of the performance will be provided at the end of the season.

APPENDIX

Bifurcation Version of WAIP

Some key features of the original 7-day WAIP of Tsai and Elsberry (2015) and the WAIP bifurcation version of Tsai and Elsberry (2018) are summarized here as background for this combined WAIP version. The 10 best historical track analogs occurring within ±30 days are ranked according to an average track distance d_T between the JTWC best track in the development sample and the target storm track, with a higher weight being given to the 72–168-h interval. The analogs are also ranked according to the initial intensity difference d_V. In WAIP, the final ranking ($\text{Rank}_{\text{analog}}$) of the candidate analogs are based on weighting factors of 0.8 for the track ranking and 0.2 for the intensity ranking, and then the $\text{Rank}_{\text{analog}}$ is sorted in ascending order to select

the 10 best analogs. A weighted mean intensity V_w at each time is

$$V_w = \sum_{i=1}^{n} (w_i V_i) \bigg/ \sum_{i=1}^{n} w_i,$$

where V_i is the intensity of the ith track analog, and

$$w_i = (d/d_{\tau,i}) \bigg/ \sum_{i=1}^{n} (1/d_{\tau,i}).$$

Because the "raw" intensity spread σ at each 12 h among the 10 best historical analogs is not adequate to represent the full probability density function, an intensity spread calibration is applied to ensure that 68.26% of the verifying intensities will likely be within the intensity spread. The calibrated intensity spread σ' at each time $t = t_i$ is

$$\sigma' = |a\sigma + b|,$$

where a and b are the calibration factors, and the raw intensity spread is weighted as above:

$$\sigma = \sqrt{\sum_{i=1}^{n} [w_i (V_i - V_w)^2] \bigg/ \sum_{i=1}^{n} w_i},$$

where $i = 1 - n$ analogs.

Tsai and Elsberry (2014b) developed a procedure for detecting an intensity bifurcation situation based on the magnitudes of the 5-day WAIP calibrated intensity spreads, and then calculated two cluster intensity evolutions with separate intensity spreads that were considerably smaller than the original WAIP intensity spread that was utilized to detect the bifurcation situations. By definition, the intensity cluster 1 (C1) is always that cluster with the larger maximum intensity, and intensity cluster 2 (C2) is the alternate solution with a lower maximum intensity.

The flowchart in Fig. A1 summarizes the objective detection procedure for an intensity bifurcation situation utilizing N analog intensities, and Tsai and Elsberry (2018) found increasing the number of analogs from 10 to 16 resulted in larger sample sizes and better 7-day WAIP performance in bifurcation situations. First, the weighted-mean spread (WMS_i) is calculated for each individual case:

$$\text{WMS}_i = \sum_{t=0\,\text{h}}^{120\,\text{h}} (w_i \sigma_{i,t}) \bigg/ \sum_{t=0\,\text{h}}^{120\,\text{h}} w_t, \qquad \text{(A1)}$$

where $i = 1 - n$ samples, $\sigma_{i,t}$ is the intensity spread each 12 h, and w_t is the same weighting factor as for the weighted-mean intensities in WAIP. That is, increasing weight is given to the intensities from the initial time to
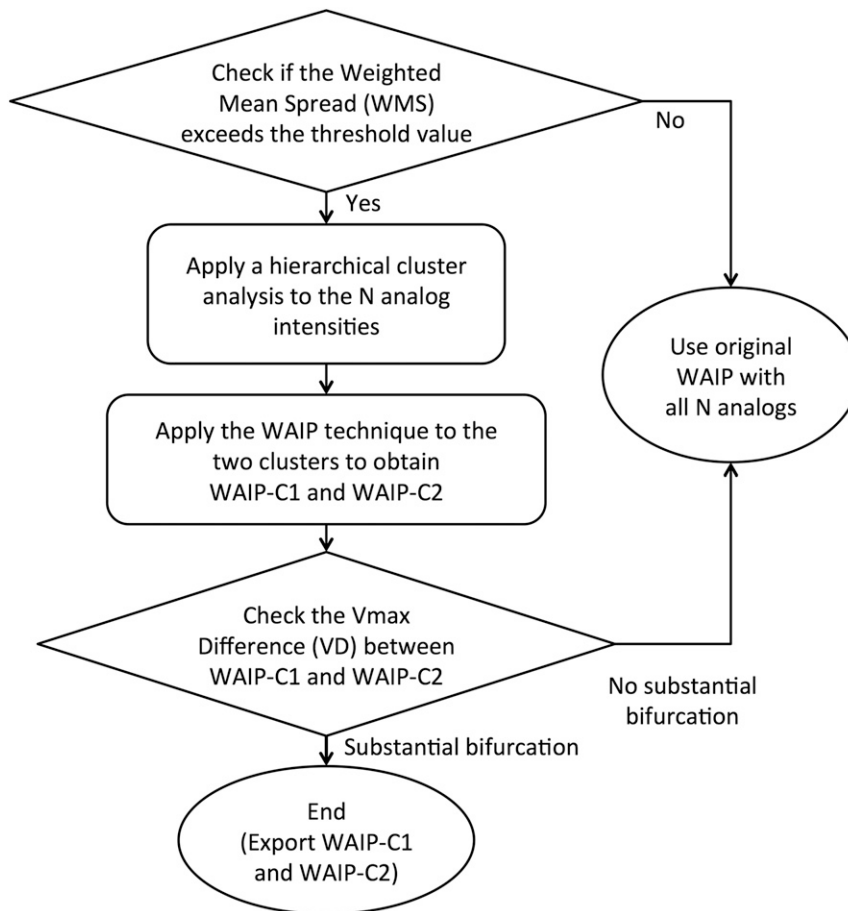
FIG. A1. Flowchart adapted from Tsai and Elsberry (2014b) of the decisions to objectively detect the existence of a WAIP intensity bifurcation situation and thereby use a hierarchical cluster analysis of the N best analog intensities to determine two clusters for which the WAIP intensities have differences that meet the threshold condition for a substantial intensity bifurcation (see text for definition).

72 h, and from 72 to 120 h the weighting factor is twice the value at $t = 0$ h. The WMS for the overall sample is

$$\text{WMS} = \sum_{i=1}^{n} \text{WMS}_i / n. \tag{A2}$$

Note that the threshold value for detecting an intensity bifurcation at the beginning of the flowchart in Fig. A1 is set as this overall sample mean WMS in Eq. (A2).

If the WMS_i now over the 7-day (168 h) forecast period for an individual case of N best-analog intensities exceeds the threshold WMS value, the second step in the flowchart (Fig. A1) is to apply a hierarchical cluster analysis (Wilks 2011) to the N analog intensities to separate them into two clusters. If there are at least three analogs in each cluster, the WAIP technique is separately applied to the two clusters to produce weighted-mean intensities and weighted-mean intensity

spreads each 12–168 h. Whereas Tsai and Elsberry (2014b) had utilized the same bias correction and intensity spread calibration for the two clusters as for the 10-analog WAIP forecasts, Tsai and Elsberry (2018) found it was necessary to derive new bias corrections and intensity spread calibrations for the two clusters in the 7-day WAIP forecasts.

The third step in the flowchart (Fig. A1) is a requirement that the cluster 1 WAIP and cluster 2 WAIP intensities must be substantially different in bifurcation situations. The motivation for this condition is simply that some meaningful intensity difference should exist between the cluster intensity forecasts that would justify having the JTWC forecaster search for alternate scenarios. Testing for this substantial bifurcation condition is based on the intensity difference:

$$\text{VD} = |\text{WAIP\_C1} - \text{WAIP\_C2}|, \tag{A3}$$

at each 12-h forecast interval, and after some testing Tsai and Elsberry (2014b) had specified the threshold VD = 15 kt. The two cluster WAIP forecasts are considered to represent a substantial intensity bifurcation if this VD threshold is exceeded for at least 25% of the 12-h forecast intervals within the 168-h forecast period. By contrast, if this VD threshold is not satisfied for 25% of the 12-h forecast intervals, it is considered that no substantial intensity bifurcation exists and the original WAIP based on all of the available analogs should be used (Fig. A1, right side pathway).

REFERENCES

DeMaria, M., and J. Kaplan, 1994: A Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic basin. *Wea. Forecasting*, **9**, 209–220, https://doi.org/10.1175/1520-0434(1994)009<0209:ASHIPS>2.0.CO;2.

——, and ——, 1999: An updated Statistical Hurricane Prediction Scheme (SHIPS) for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **14**, 326–337, https://doi.org/10.1175/1520-0434(1999)014<0326:AUSHIP>2.0.CO;2.

——, C. R. Sampson, J. A. Knaff, and K. D. Musgrave, 2014: Is tropical cyclone intensity guidance improving? *Bull. Amer. Meteor. Soc.*, **95**, 387–398, https://doi.org/10.1175/BAMS-D-12-00240.1.

Elsberry, R. L., and H.-C. Tsai, 2014: Situation-dependent intensity skill metric and intensity spread guidance for western North Pacific tropical cyclones. *Asia-Pac. J. Atmos. Sci.*, **50**, 297–306, https://doi.org/10.1007/s13143-014-0018-5.

——, M. S. Jordan, and F. Vitart, 2011: Evolution of the ECMWF 32-day ensemble predictions during 2009 season of western North Pacific tropical cyclone events on intraseasonal time-scales. *Asia-Pac. J. Atmos. Sci.*, **47**, 305–318, https://doi.org/10.1007/s13143-011-0017-8.

Knaff, J. A., C. R. Sampson, and M. DeMaria, 2005: An operational statistical typhoon intensity prediction scheme for the western North Pacific. *Wea. Forecasting*, **20**, 688–699, https://doi.org/10.1175/WAF863.1.

Tsai, H.-C., and R. L. Elsberry, 2014a: Applications of situation-dependent intensity and intensity spread predictions based on a weighted analog technique. *Asia-Pac. J. Atmos. Sci.*, **50**, 507–518, https://doi.org/10.1007/s13143-014-0040-7.

——, and ——, 2014b: Improved tropical cyclone intensity and intensity spread predictions in bifurcation situations. *Asia-Pac. J. Atmos. Sci.*, **50**, 117–128, https://doi.org/10.1007/s13143-014-0054-1.

——, and ——, 2015: Seven-day intensity and intensity spread predictions for western North Pacific tropical cyclones. *Asia-Pac. J. Atmos. Sci.*, **51**, 331–342, https://doi.org/10.1007/s13143-015-0082-5.

——, and ——, 2017a: Seven-day intensity and intensity spread predictions for Atlantic tropical cyclones. *Wea. Forecasting*, **32**, 141–147, https://doi.org/10.1175/WAF-D-16-0165.1.

——, and ——, 2017b: Ending storm version of the seven-day weighted analog intensity prediction technique for western North Pacific tropical cyclones. *Wea. Forecasting*, **32**, 2229–2235, https://doi.org/10.1175/WAF-D-17-0151.1.

——, and ——, 2018: Seven-day intensity and intensity spread predictions in bifurcation situations with guidance-on-guidance for western North Pacific tropical cyclones. *Asia-Pac. J. Atmos. Sci.*, **54**, 421–430, https://doi.org/10.1007/s13143-018-0008-0.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.