

# Nowcast Guidance of Afternoon Convection Initiation for Taiwan

HUI-LING CHANG

*Research and Development Center, Central Weather Bureau, Taipei, Taiwan*

BARBARA G. BROWN

*National Center for Atmospheric Research, Boulder, Colorado*

PAO-SHIN CHU

*Department of Atmospheric Sciences, School of Ocean and Earth Science and Technology, University of Hawai'i at Mānoa, Honolulu, Hawaii*

YU-CHIENG LIOU

*Department of Atmospheric Sciences, National Central University, Zhong-Li, Taiwan*

WEN-HO WANG

*Meteorological Satellite Center, Central Weather Bureau, Taipei, Taiwan*

(Manuscript received 29 December 2016, in final form 25 July 2017)

## ABSTRACT

Focusing on afternoon thunderstorms in Taiwan during the warm season (May–October) under weak synoptic forcing, this study applied the Taiwan Auto-NowCaster (TANC) to produce 1-h likelihood nowcasts of afternoon convection initiation (ACI) using a fuzzy logic approach. The primary objective is to design more useful forecast products with uncertainty regions of predicted thunderstorms to provide nowcast guidance of ACI for forecasters. Four sensitivity tests on forecast performance were conducted to improve the usefulness of nowcasts for forecasters. The optimal likelihood threshold ( $L_t$ ) for ACIs, which is the likelihood value that best corresponds to the observed ACIs, was determined to be 0.6. Because of the high uncertainty on the exact location or timing of ACIs in nowcasts, location displacement and temporal shifting of ACIs should be considered in operational applications. When a spatial window of 5 km and a temporal window of 18 min are applied, the TANC displays moderate accuracy and satisfactory discrimination with an acceptable degree of overforecasting. The non-parametric Mann–Whitney test indicated that the performance of the TANC substantially surpasses the competing Space and Time Multiscale Analysis System–Weather Research and Forecasting Model, which serves as a pertinent reference for short-range (0–6 h) forecasts at the Central Weather Bureau in Taiwan.

## 1. Introduction

For short-range (0–6 h) forecasts, one of the most challenging tasks is to predict whether a convective storm will occur, as well as when and where it will happen. Relative to other convective systems, such as stationary fronts and typhoons, forecasting afternoon convective storms is more difficult because of their small spatial scale and very short lifetime. Sea breezes and anabatic winds play an important role in the initiation of

afternoon thunderstorms by moistening the boundary layer (Lin et al. 2011). The convection is often enhanced and new storms are generated when the outflows interact with terrain, sea breezes, and other outflows from adjacent storms (Szoke et al. 1985; Jou 1994). The afternoon convective activity in Taiwan peaks along the lower slope of the mountains rather than at higher elevations farther inland (Johnson and Bresch 1991; Lin and Kuo 1996; Chang 1997; Chen et al. 2001).

Wilson and Schreiber (1986) showed that the majority of thunderstorms in the Denver, Colorado, area formed along radar-detected convergence lines. As noted in

---

*Corresponding author:* Pao-Shin Chu, chu@hawaii.edu

DOI: 10.1175/WAF-D-16-0224.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](http://www.ametsoc.org/PUBSReuseLicenses) ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

previous studies (Mueller and Wilson 1989; Wilson and Mueller 1993), forecasters can often anticipate thunderstorm initiation by monitoring radar-detected convergence lines (boundaries) together with cloud development in the vicinity of the convergence line. However, forecasters often have difficulty in knowing the precise timing and location of storm initiation. Another issue is that not all convergence lines initiate storms, even when they collide in conditionally unstable environments (Stensrud and Maddox 1988). Wilson et al. (1998) showed that the initiation of convective storms is often controlled by boundary layer convergence features, environmental vertical wind shear, and buoyancy. Accordingly, successful forecasts of storm initiation depend on the accurate specification of the initial thermodynamic and kinematic fields with particular attention paid to convergence lines. Wilson et al. (2004) demonstrated that convective initiation (CI) can be predicted up to 1 h in advance by adopting a set of predictor fields and manually entering the location of boundary layer convergence lines in NCAR's Auto-Nowcaster (ANC) (Mueller et al. 2003).

Rapidly intensifying afternoon thunderstorms can lead to lightning strikes and heavy downpours, which may cause problems such as power failures, traffic jams, flooding, and aviation hazards. Accurate nowcasts provide disaster management agencies with valuable additional lead time to implement appropriate preventive actions against severe weather. Improving nowcasts of afternoon thunderstorms is one of the research priorities of the Central Weather Bureau (CWB) in Taiwan.

Currently, nowcasts of afternoon convective storms using numerical models are challenging. One reason for this is the crude representations of the model physics and convective schemes (Roberts et al. 2012). Another reason is that crucial characteristics of mesoscale boundaries, such as the frontal edges of land or sea breezes and anabatic or katabatic winds, cannot be adequately resolved by operationally available radar observations (when the boundaries are too far away from the radar, or too shallow for the radar to detect) or surface observations (generally sparsely spaced) used to initialize model fields in Taiwan. Therefore, mesoscale boundary information is unavailable in model initial fields (Benjamin et al. 2004; Stensrud et al. 2009). However, such information is critical for producing accurate forecasts of afternoon convection initiations (ACIs) using dynamical models. One way to mitigate this problem is to apply a statistical forecasting technique such as a fuzzy logic algorithm to mesoscale predictors that can be observed or forecast.

To provide objective guidance for afternoon thunderstorm predictions in northern Taiwan, Lin et al. (2012) developed a fuzzy logic algorithm using 277 cases

during the warm season (May–October) from 2005 to 2008 in the presence of dominant thermal forcing. Lin et al. (2012) found that the best predictors of afternoon thunderstorms were vapor pressure, humidity, wind direction, and wind speed of the boundary layer in the morning, as well as CAPE, dewpoint depression ( $T - T_d$ ), wind direction, and wind speed in the lower–middle layer of the troposphere (1000–500 hPa) from sounding data at 0800 local standard time (LST).

The use of high-resolution ensemble forecasts to predict CI is currently still being tested. In the Spring Forecasting Experiment in 2011 (SFE 2011), the potential utility of high-resolution (4 km) ensemble forecasts in providing guidance for CI forecasts was examined (Kain et al. 2013). The evaluation results indicate that when CI occurred in both models and radar observations, there was no systematic ensemble bias but considerable variance in the timing of the CI events. In addition, the default CI algorithms often overpredicted the frequency of CI events and sometimes totally missed convective events (Kain et al. 2013).

Focusing on ACIs in Taiwan under weak synoptic forcing, we applied the Taiwan Auto-NowCaster (TANC) to produce 1-h likelihood nowcasts of CI based on a fuzzy logic approach. Eight predictors were used in the study, and two of them were based on the radar climatology constructed by Lin et al. (2012). In this study we evaluate the forecast performance of the TANC in order to establish a reference for its future development and improvement. The ultimate goal is to provide forecasters with more useful nowcast products for guidance on ACIs in Taiwan.

This paper is organized as follows. The TANC and study data are introduced in section 2. The verification and analysis methodology are presented in section 3. Section 4 describes the sensitivity experiments on verification scores, including the sensitivity of scores to likelihood thresholds, spatial and temporal windows, as well as different combinations of spatial and temporal windows. In section 5, the TANC is compared with a high-resolution hot-start numerical weather prediction model, the Space and Time Multiscale Analysis System (Xie et al. 2011)–Weather Research and Forecasting (STMAS–WRF) Model, to validate the benefits of the TANC over other methods. A summary of the findings and suggestions for future research is provided in section 6.

## 2. TANC and study data

The TANC was introduced to the CWB by the National Center for Atmospheric Research (NCAR) (Mueller et al. 2003; Roberts and Rutledge 2003; Saxen et al. 2008) and was created specifically to predict convective storms

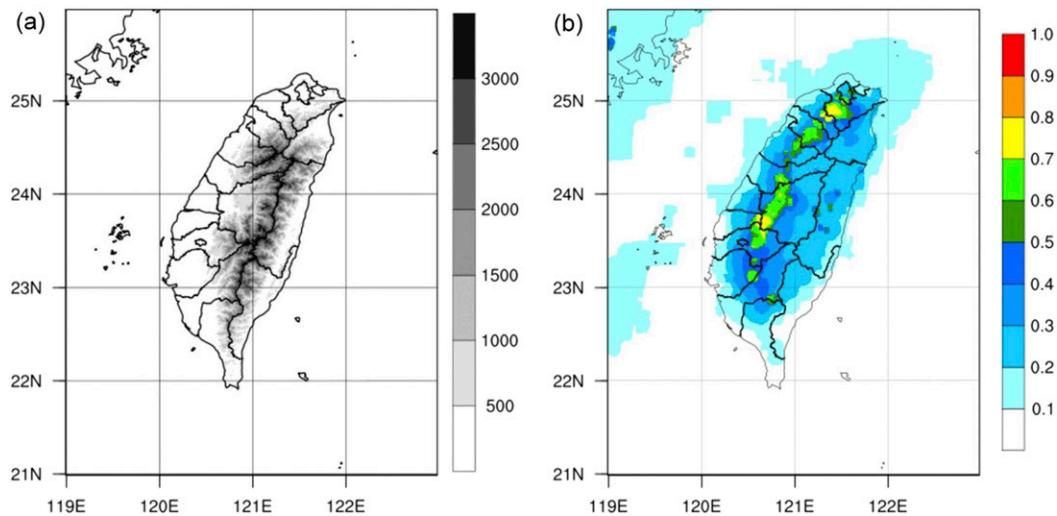


FIG. 1. (a) Terrain map of Taiwan. (b) TANC original nowcast product, showing the 1-h nowcast likelihoods of ACIs in Taiwan and its adjacent seas. The nowcast is issued at 0800 UTC 14 Jun 2015.

on a subtropical island with high mountains and complex terrain. The TANC covers Taiwan and its adjacent seas (Fig. 1a) with a  $0.01^\circ$  horizontal resolution; the system estimates the 1-h likelihood of CI every 6 min operationally. CI is defined as new convection with reflectivity  $\geq 35$  dBZ. The value of 35 dBZ is also the operational threshold for convection at the CWB.

The TANC nowcasts the likelihood of CI using eight predictors (Table 1) based on a fuzzy logic approach (Berenguer et al. 2006; Lin et al. 2012). The eight predictors include radar climatology frequency, radar climatology frequency trend, CAPE, CIN, average RH in

the 850–500-hPa layer, surface divergence, storm initiation locations, and radar-based cumulus cloudiness. Similar types of predictors are combined into one predictor group in the upstream modules (second row in Fig. 2), and the sum of the predictor weights for each upstream module is 1. For example, the CAPE and CIN predictors are combined into the stability upstream module, and the likelihood value of that module is obtained by summing the weighted likelihood values of the CAPE and CIN predictors ( $L_{\text{stability}} = W_{\text{CAPE}} \times L_{\text{CAPE}} + W_{\text{CIN}} \times L_{\text{CIN}}$ ;  $W_{\text{CAPE}} = 0.4$  and  $W_{\text{CIN}} = 0.6$ ). Four upstream modules (entrainment, small-scale convergence, CI proximity, and cumulus)

TABLE 1. TANC predictors with a brief description of each predictor, its use, and references.

Predictor, shorthand notation (units)	Predictor description and use	Reference
ClimoFreq (percent frequency)	Frequency of reflectivity $\geq 40$ dBZ based on four years' worth of radar data (2005–08)	Lin et al. (2011)
ClimoFreqTrend (percent frequency difference)	Based on the same data as above, this is the climatological trend in the frequency of reflectivity $\geq 40$ dBZ over the specific nowcast period	Lin et al. (2011)
CAPE ( $\text{J kg}^{-1}$ )	Convective available potential energy obtained from the latest CWB–WRF analysis fields; it is a measure of atmospheric stability	Lin et al. (2012); Trier et al. (2011)
CIN ( $\text{J kg}^{-1}$ )	Convective inhibition obtained from the latest CWB–WRF analysis field; it is also a measure of atmospheric stability	Lin et al. (2012); Trier et al. (2011)
RH <sub>avg</sub> (RH percentage)	Average relative humidity in the 850–500-hPa layer obtained from the latest CWB–WRF analysis fields; it is a measure of the amount of dry air that may be entrained into convective updrafts	Trier et al. (2011)
SurfDiv ( $\text{s}^{-1}$ )	Surface divergence obtained from surface station winds; it is used to help identify localized regions of updraft	Mueller et al. (2003)
StormInitLoc (dimensionless number)	Based on radar data, it is a measure of the closeness of storm initiation locations to a given grid point	Wilson and Roberts (2006)
RadarCu (dBZ)	Based on radar reflectivity data, it identifies regions of early storm growth aloft associated with cumulus clouds	Wilson and Mueller (1993)

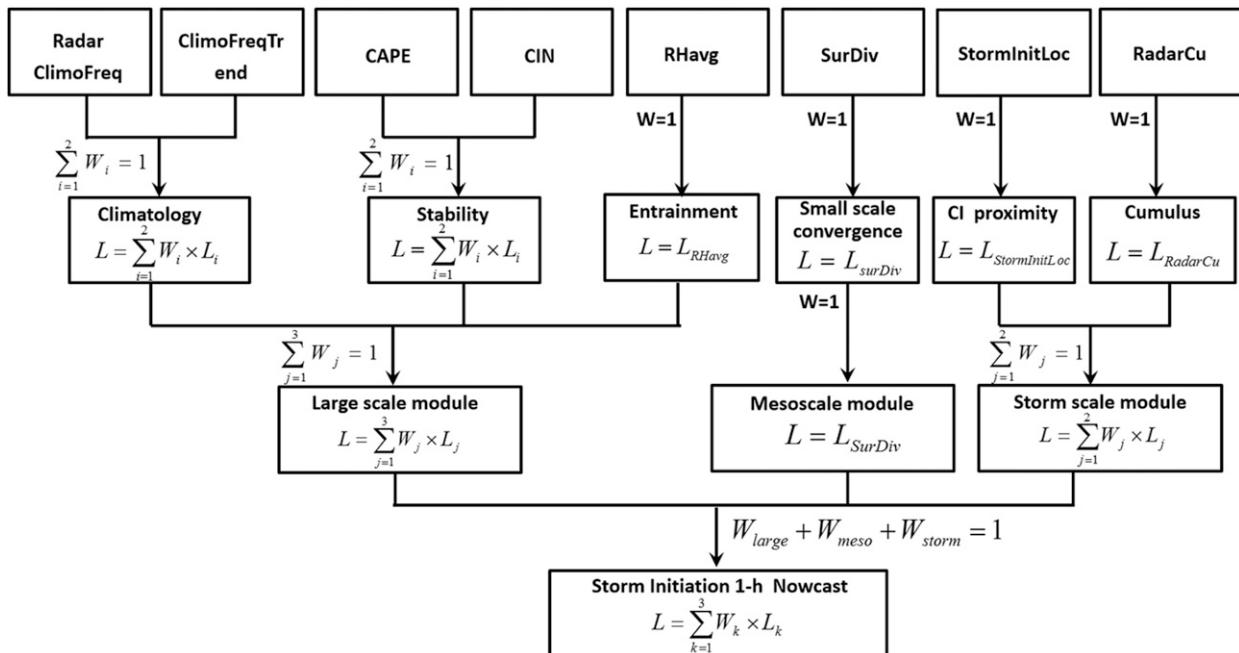


FIG. 2. Flowchart for TANC 1-h nowcasts of ACI. (first row) Eight predictors, as defined in Table 1, are combined into the (second row) upstream modules, and then the upstream modules are combined into the (third row) primary modules. Finally, the three primary modules are combined to produce the (bottom row) nowcast. The  $L_i$  and  $W_i$  are the likelihood and corresponding weight, respectively. See section 2 for further explanation.

have only one predictor; therefore, their likelihood value is that of the single predictor. The outputs from the upstream modules are then ingested into three primary modules (third row in Fig. 2) that represent forcing on the synoptic, mesoscale, and storm scales. Similarly, the sum of the upstream module weights for each primary module is 1. Finally, the three primary modules are combined to produce the likelihood of CI. The weights of the three modules are assigned based on their relative contributions to forcing new convection and the sum of the three weights is 1.

The TANC uses membership functions to determine the degree of association between various predictors and CI. Specifically, the predictor values are converted into likelihood values through fuzzy membership functions, which are derived from the statistics of prestorm environmental characteristics, climatology of radar reflectivity, and so on. “Fuzzy” indicates that the likelihood values range from  $-1$  to  $1$ . Higher positive values indicate an increased likelihood of CI in a region, lower negative values indicate a decreased likelihood, and  $0$  indicates a neutral likelihood (Mueller et al. 2003).

The conceptual models of TANC are based on determining the overlap of regions with a high climatological frequency and trend of convective storms, high instability, surface convergence, and other favorable conditions for triggering convection. The overlapping

regions of the various predictors are also the expected regions of CI (Mueller et al. 2003).

This study focuses on well-organized afternoon convective storms under weak synoptic forcing in warm seasons (May–October). Well-organized convection cases are defined as storms with 1) radar reflectivity  $\geq 35$  dBZ over areas  $\geq 300$  km<sup>2</sup> (i.e., at least over 1/120 Taiwan’s land area), which 2) lasted  $\geq 90$  min (the duration of a convective event, not the lifetime of individual convective cells) and 3) occurred between 0400 and 1200 UTC (i.e., between 1200 and 2000 LST). Nine days of afternoon convective storms (Table 2) that occurred in Taiwan from 2014 to 2015 were chosen for evaluation and a total of 312 nowcasts of 1 h were verified.

### 3. Verification methodology

#### a. Conversion from likelihood to Y/N forecasts

The TANC provides 1-h likelihood nowcasts of ACI (Fig. 1b), which indicate the uncertainty information associated with the forecasts. Note that a likelihood nowcast is different from a probabilistic forecast, even if both have a similar meaning: higher value represents higher possibility. The likelihood values from the TANC range from  $-1$  to  $1$  while the probability is bounded between  $0$  and  $1$ . Because the TANC likelihood nowcast

TABLE 2. Afternoon convection cases in 2014 and 2015.

Year	Event date (No. of validation times)	Time period (UTC)
2014	30 Jun (35)	0606–0930
	1 Jul (24)	0730–0948
	27 Aug (33)	0618–0930
	29 Aug (33)	0718–1030
	9 Sep (21)	0730–0930
2015	2 Jun (30)	0624–0918
	14 Jun (39)	0630–1018
	28 Jul (45)	0536–1000
	17 Sep (52)	0554–1100
Total: 312 nowcasts of 1 h		

is not a probabilistic forecast, forecasters need to know how to best use the TANC nowcasts. Another problem is that the TANC frequently shows large areas of low likelihood values ( $<0.3$ ) for CI (e.g., Fig. 1b), which results in forecasters mistakenly believing that CI may occur everywhere. To provide guidance on the most likely region for CI, we attempt to determine an optimal likelihood threshold (Lt) that best corresponds to the observed CI. Therefore, the likelihood forecasts are converted into yes/no (Y/N) forecasts. Moreover, the forecast uncertainty information is incorporated into the final nowcast products using the relaxation method described later. The conversion from likelihood to Y/N forecasts is performed by first selecting a relevant Lt. For example, if the Lt is 0.8, then if the likelihood exceeds this threshold it means that the TANC predicts there will be new convection in the next hour. Otherwise, the prediction is classified as a nonevent.

Currently, there is no direct observation that can unambiguously indicate whether new convection has initiated within the past hour. However, we need such information to determine whether the TANC nowcasts are correct. Here, we adopt the same approach as Lakshmanan et al. (2012). Two radar images 1 h apart were examined to find where new convection has occurred. The past observation was warped to best align it with the current observation using a cross-correlation optical flow method (Barron et al. 1994). This involves finding a smooth motion field based on the two images and then advecting the corresponding grid in the second image backward to align it with the first one. Once the two images have been aligned, a  $5 \times 5$  neighborhood ( $\sim 5 \text{ km} \times 5 \text{ km}$ ) of each pixel was searched to determine the convective state within the past hour. Each pixel of the radar image was then classified into one of four categories: new, ongoing, decaying, and no convection.

By using the aforementioned conversion, each grid point was classified into one of four possible conditions

TABLE 3. The  $2 \times 2$  contingency table.

		Forecast	
		Yes	No
Observation	Yes	Hit ( $h$ )	Miss ( $m$ )
	No	False alarm ( $f$ )	Correct rejection ( $c$ )

in a  $2 \times 2$  contingency table (Table 3), consequently enabling computation of the threat score (TS; [Bermowitz and Zurndorfer 1979](#)), bias ratio (BIAS), probability of detection (POD), false alarm ratio (FAR), Kuiper score (KS; [Murphy and Daan 1985](#)), and equitable threat score (ETS) for the forecast verification ([Wilks 2011](#); [Jolliffe and Stephenson 2012](#)). The aforementioned six verification scores can be found in [appendix A](#).

### b. Relaxation method

Compared with forecasting other weather systems, the uncertainty for ACI nowcasts is considerably higher. All kinds of uncertainties during the forecast process result in difficulty in predicting the exact location and timing of ACIs. For example, the predictors in the TANC, such as model-derived CAPE, CIN,  $\text{RH}_{\text{avg}}$ , and observed surface divergence, come with their own uncertainties, which contribute to the nowcast uncertainty represented by the fuzzy logic algorithm relying on a fixed number of predictors. That is, there is an inherent uncertainty in the statistical model even if the predictors were perfectly known. Therefore, the location displacement and temporal shift of the predicted storms should be accounted for in operational applications. What should the space–time tolerances be to achieve a level of accuracy that is considered acceptable? This question is addressed through the development of a relaxation method using a historical ACI dataset and taking into account the operational needs of forecasters in Taiwan. For example, a location displacement of 5 km is almost the maximum tolerable range for predicted storms given the small size of Taiwan.

#### 1) Spatial relaxation

For location displacement, we relax restrictions from a pixel-to-pixel verification to a verification of a circle with a radius of  $N$  grid points. This spatial relaxation method is similar to that of [Lakshmanan et al. \(2012\)](#), but with a modification to render uniform location displacement in all directions. In [Lakshmanan et al. \(2012\)](#), the pixel-to-pixel verification was relaxed to a verification of a square area of  $(2N + 1)(2N + 1)$  grid points (Fig. 3). As a result, the tolerable location displacement of a predicted storm is larger in the diagonal than in the other radial

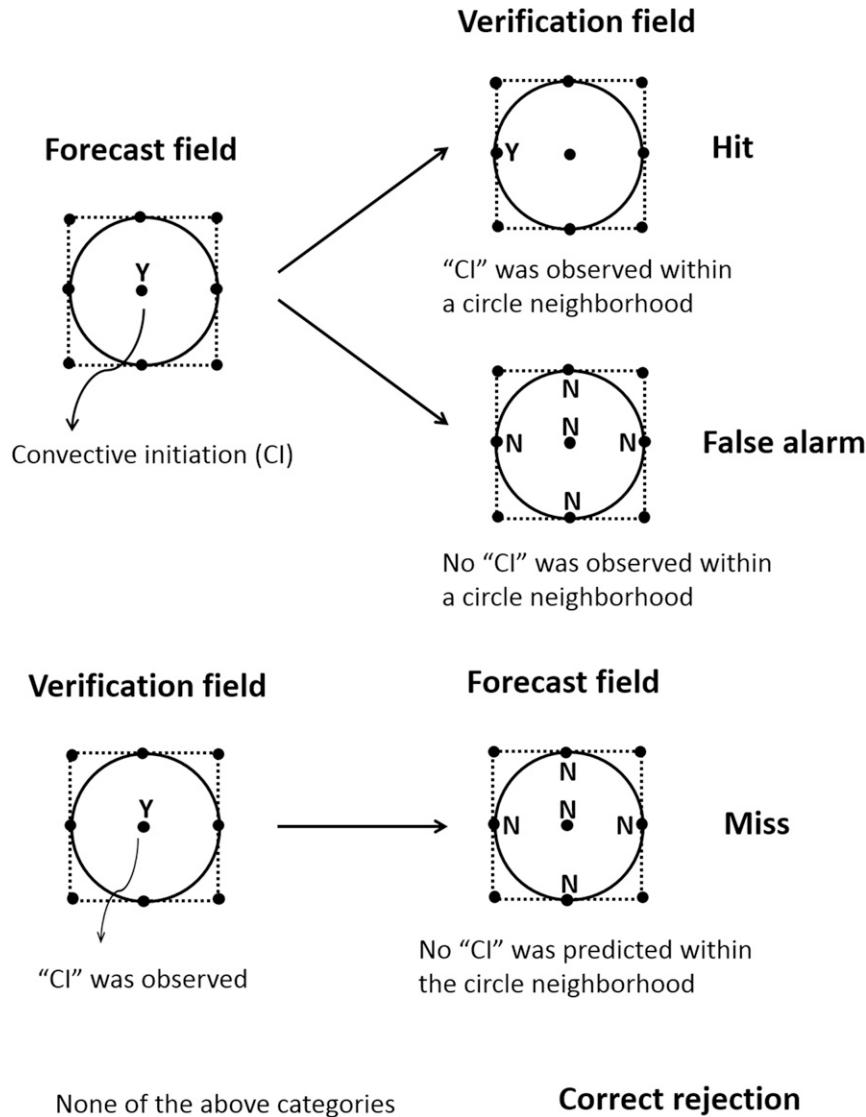


FIG. 3. Spatial relaxation method with  $N = 1$ , which means one grid of storm location displacement is allowed. The circle area is based on this study and the square one is from Lakshmanan et al. (2012).

directions. To overcome this issue, the area of tolerable location displacement was modified to a circle with radius of  $N$  grid points.

Figure 3 illustrates the spatial relaxation method with  $N = 1$ , which means only one grid point of location displacement is an acceptable tolerance for the CI nowcasts. Suppose that the TANC predicted a certain grid point as CI. If a pixel-to-pixel verification was applied, this grid point would be classified as a "hit"  $h$  only when new convection was observed at the same grid point in the verification field. However, if one grid point of location displacement was allowed ( $N = 1$ ), this verifying grid point would be

regarded as  $h$  when new convection was observed within a circle with radius of one grid point. Therefore, allowing for location displacement increases the frequency of  $h$ . A "false alarm"  $f$  requires the nowcast to predict CI but without new convection observed at the same grid point through a pixel-to-pixel verification; however, under the  $N = 1$  relaxation, this grid point would be classified as  $f$  if no new convection was observed within a circle with radius of one grid point. Therefore, allowing for location displacement reduces the frequency of  $f$ .

2) A "miss"  $m$  requires new convection to be observed but no CI to be predicted at the same grid point

using a pixel-to-pixel verification; however, under the  $N = 1$  relaxation, this grid point would be classified as  $m$  if no CI was predicted within a circle with radius of one grid point. Therefore, allowing for location displacement reduces the frequency of  $m$ . None of the aforementioned categories would be classified as a “correct rejection”  $c$  in the contingency table (Table 3). The spatial relaxation works favorably because the horizontal resolution of the TANC is high (approximately 1 km). The smoothed forecasts will generally have better verification scores than the unsmoothed forecasts.

Temporal relaxation is also accounted for in the verification of CI nowcasts. Suppose a temporal window ( $T = 12$  min) is considered, which means a temporal shift of less than 12 min is an acceptable tolerance for the CI nowcasts. That is, if the TANC 1-h nowcast predicts CI, the new convection is expected to occur in the next 48–72 min ( $1\text{ h} \pm 12\text{ min}$ ). If the initial time of the 1-h CI nowcast was 0630 UTC, the predicted new convection would likely occur between 0718 and 0742 UTC ( $0730\text{ UTC} \pm 12\text{ min}$ ) under the  $T = 12$  min relaxation. In other words, the new convection predicted by the TANC nowcasts between 0618 and 0642 UTC ( $0630\text{ UTC} \pm 12\text{ min}$ ) would probably occur at 0730 UTC. The temporal relaxation works well because the temporal resolution of the TANC is high (6 min).

Temporal relaxation is similar to a time-lagged ensemble TANC and the idea of temporal relaxation comes from time-lagged ensembles. For time-lagged ensembles, previous forecasts initialized at different times were used to construct members for the ensemble forecasts. Applying a similar concept to the TANC, we use the previous nowcasts initialized at different times to predict CI at a specific validation time. The TANC nowcasts with temporal relaxation can be regarded as time-lagged ensemble nowcasts. The difference between time-lagged ensembles and temporal relaxation is that time-lagged ensembles produce a mean forecast or a probabilistic forecast, but temporal relaxation produces a Y/N forecast. That is, as long as there is a previous nowcast predicting a grid point as CI, this grid will be regarded as CI for temporal relaxation.

The main reason to adopt the time-lagged ensemble concept in the TANC nowcast is that a short-range forecast generally possesses a relatively strong dependency on the initial conditions. Forecast errors in a very short range may be strongly correlated to uncertainties in the initial analysis (Lu et al. 2007). The time-lagged ensembles can be interpreted as the forecasts obtained from a set of perturbed initial conditions (Van den Dool and Rukhovets 1994).

In section 4, which focuses on sensitivity experiments, we apply the spatial and temporal relaxation methods to evaluate the forecast performance of the TANC under different spatial and temporal windows. Based on the evaluation results, we determine the most likely regions and the less likely, but still possible, areas for CI for TANC nowcast products.

### c. Confidence interval

Confidence intervals are used to describe the uncertainty associated with a sample estimate of a population parameter. To construct a confidence interval, a confidence level (e.g., 95%) must first be selected. The confidence level indicates the probability that the confidence interval captures the true population parameter given a distribution of samples. Therefore, confidence intervals provide more meaningful information about forecast performance, and enable credible comparisons of nowcast performance among different spatial or temporal window settings when sample size is limited (312 nowcasts in this study). In this study, we applied the bootstrap method (Chu 2002; Wilks 2011) to construct the 95% confidence interval of median TS, BIAS, KS, and ETS based on 10 000 bootstrap samples. There was very little autocorrelation in the verification statistics so they could be considered nearly independent.

## 4. Sensitivity experiments

In this section, we discuss sensitivity tests for various Lts that were conducted to determine an optimal Lt to provide guidance on the most likely region for CI. In addition, sensitivities of verification scores to different spatial ( $\pm 1$ –10 km) and temporal ( $\pm 6$ –30 min) windows were also investigated to determine acceptable spatial and temporal uncertainty ranges for the purpose of displaying the less likely, but still possible, regions for CI.

### a. Sensitivity of scores to different Lts

The forecast performance for no relaxation at different Lts (Fig. 4) shows that the median TS and FAR values do not greatly vary when Lt is between 0.3 and 0.6; however, the median BIAS and POD values exhibit a clear decrease with increasing Lt. The optimal Lt is selected using the following arguments: a lower Lt produces a higher POD, but also leads to overforecasting (i.e., a larger BIAS). Therefore, the POD alone should not be used for determining the optimal Lt. With a focus on the BIAS, the ratio is too large when Lt is between 0.3 and 0.5. When Lt equals 0.6, the TANC displays an acceptable degree of overforecasting. If Lt is increased to 0.7, the TANC exhibits underforecasting ( $\text{BIAS} < 1$ ) and the TS decreases

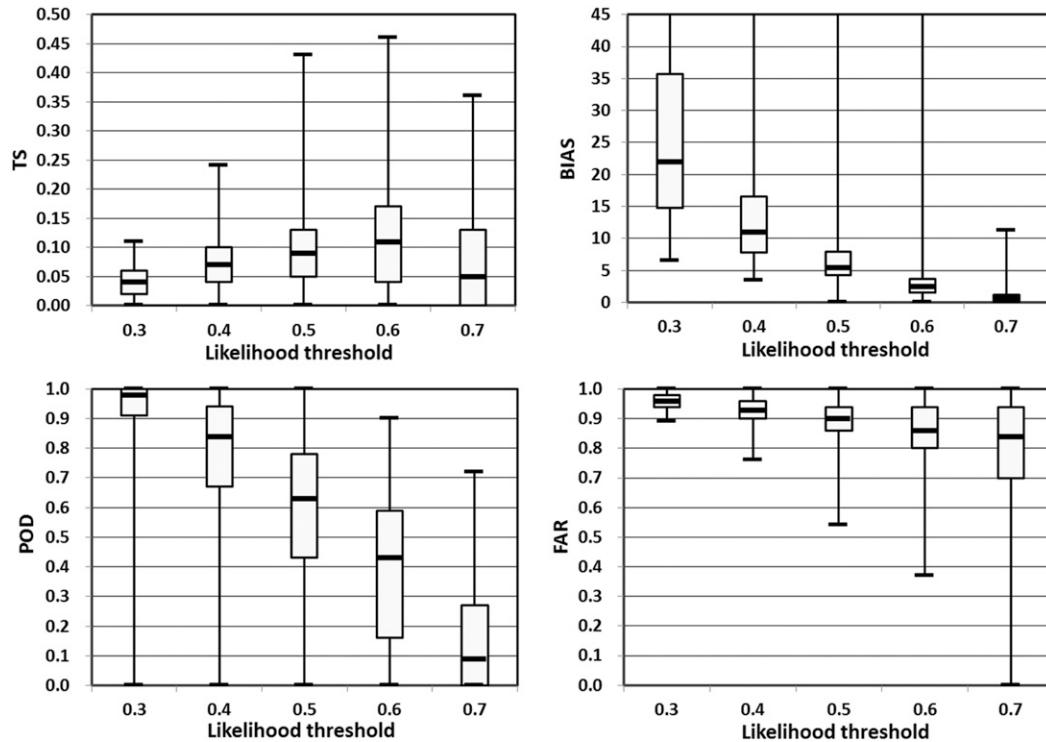


FIG. 4. Boxplots of verification scores at different Lts from TANC, including TS, BIAS, POD, and FAR.

considerably. Thus, an Lt of 0.6 is selected as the optimal value for 1-h nowcasts of ACIs.

#### b. Sensitivity of scores to different spatial windows

A boxplot analysis (Fig. 4) revealed that the distributions of verification scores are asymmetrical; therefore, we selected the median as the test statistic to represent the central tendency of forecast performance rather than the mean. Additionally, considering the uncertainty associated with sampling variability and the limitations in sample size, 95% confidence intervals of the score median were constructed through bootstrapping.

Figure 5 displays the score median with a 95% confidence interval associated with different spatial windows when the optimal Lt (0.6) was applied. When the spatial window  $N$  equals zero, the verification results were determined using a pixel-to-pixel verification. When  $N$  was extended out to five grid points, the median TS, BIAS, KS, and ETS values were approximately 0.33, 1.87, 0.84, and 0.33, respectively. In this case the TANC displayed moderate accuracy and satisfactory discrimination but also an acceptable degree of overforecasting. As mentioned in appendix A, the values of the ETS and TS are similar because ACIs can be regarded as “rare events” in the TANC domain and the chance of random hits is very low for the TS. For this reason we only show the

TS in the latter analysis. In addition, the KS approaches the POD for rare events; thus, we use the KS together with the BIAS to evaluate the forecast quality of TANC to ensure that the high value of KS does not result from serious overforecasting. The 95% confidence intervals were narrow, indicating that the uncertainty caused by sampling variability or limitations in sample size is very small.

Regarding the aforementioned two sensitivity tests, the CWB provides one formulation of the operational TANC nowcast product with an optimal Lt of 0.6 and a spatial window of five grid points (Fig. 6a). Figure 6b shows the same nowcast product using the spatial relaxation from Lakshmanan et al. (2012) for comparison. When adopting a square of  $(2N + 1) \times (2N + 1)$  grid points as the area of tolerable location displacement, the boundary of the uncertainty area of ACI will have a zigzag or square shape. In other words, the tolerable location displacement is not identical in all directions. Therefore, we opt to use a circle with a radius of  $N$  grid points as the area of tolerable location displacement.

Five grid points were selected as the spatial window for the TANC for two reasons: 1) the KS becomes saturated with this setting (Fig. 5), and 2) a location displacement of five grid points ( $\sim 5$  km) is the maximum tolerable range considered by forecasters

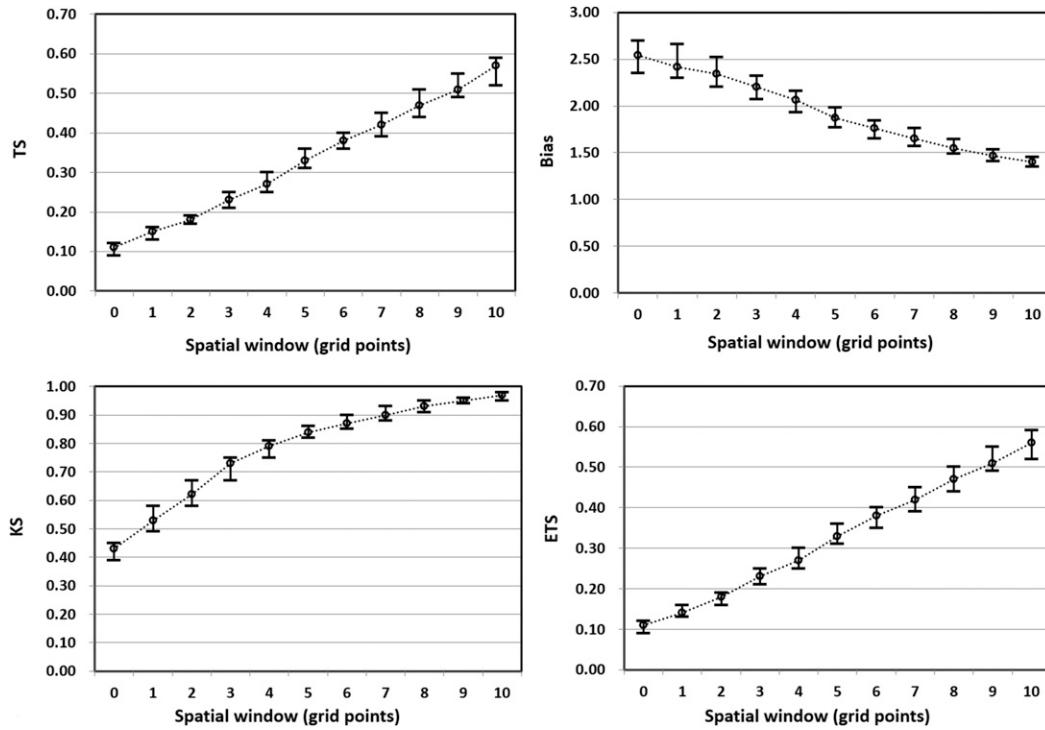


FIG. 5. Score median values and 95% confidence intervals for different spatial windows (grid points) from TANC, including TS, BIAS, KS, and ETS.

due to the small size of Taiwan. This nowcast product serves as a critical reference for ACI warnings at the CWB.

*c. Sensitivity of scores to different temporal windows*

Figure 7 is similar to Fig. 5, but for different temporal windows. For a temporal window of 0 min (i.e., a point-to-point verification), the forecast was verified with only the observation at the validation time. The verification scores improved substantially when the restrictions were

relaxed from point-to-point verification to verification over a temporal window  $T$  of  $\pm 6$  min. When  $T$  was extended to  $\pm 18$  min (when the KS tends to become saturated), the TS, BIAS, and KS values reached 0.36, 2.01, and 0.86, respectively. Similar to the results from spatial relaxation, the TANC displayed moderate accuracy and satisfactory discrimination but also an acceptable degree of overforecasting; we are confident in these verification results because of the narrow 95% confidence intervals.

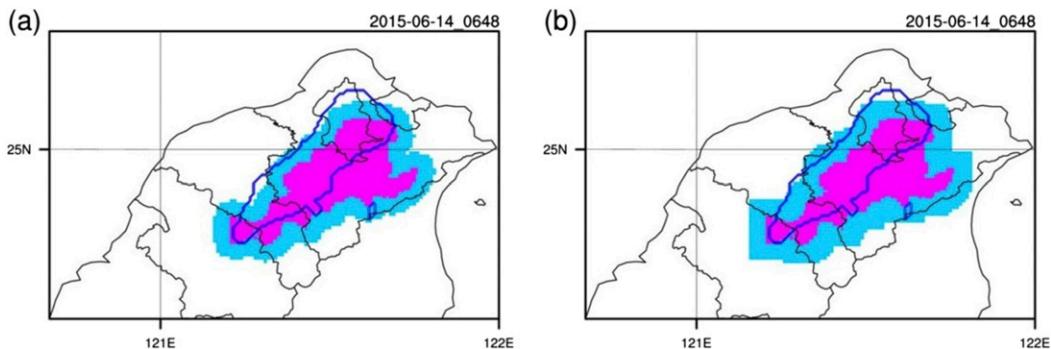


FIG. 6. TANC nowcast product design based on sensitivity tests of scores to different spatial windows using the spatial relaxation of (a) this study and (b) Lakshmanan et al. (2012). The pink shading shows the most likely regions for CI (i.e., the areas with likelihood  $\geq 0.6$ ). The blue shading shows the less likely but still possible areas of CI. The dark blue contours of observed CI are also overlaid for verification. The 1-h TANC nowcast for northern Taiwan was issued at 0648 UTC 14 Jun 2015.

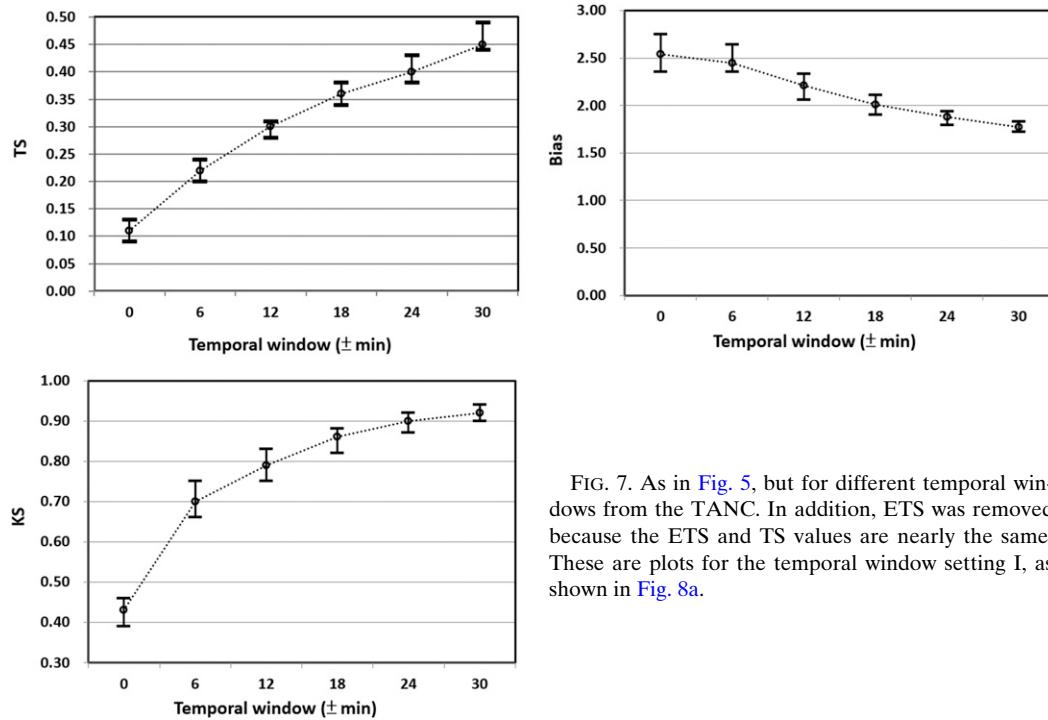


FIG. 7. As in Fig. 5, but for different temporal windows from the TANC. In addition, ETS was removed because the ETS and TS values are nearly the same. These are plots for the temporal window setting I, as shown in Fig. 8a.

The timing of the temporal window also needs to take into account the operational needs of the forecasters to have the latest information at the time that they must produce their forecasts. Therefore, the next topic addressed pertains to how to set the temporal window on the basis of the operational needs of the forecasters. As mentioned in section 3b, a temporal window of  $\pm T$  min means that the convective storms predicted by the nowcasts during the period of the issue time  $\pm T$  min are likely to occur 1 h after the issue time. An example of this is illustrated in Fig. 8a, where  $T = 18$ . Based on the viewpoint of operational applications, only

the forecasts that have already been generated (i.e., the forecasts issued earlier than 0930 LST) can be used to provide additional information for the latest 1-h nowcast (i.e., the forecast issued at 0930 LST). Therefore, it is important to consider whether the forecast performance will be severely affected if the nowcast is set at the ending point of the temporal window (Fig. 8b) given the same window size.

Figure 9 is similar to Fig. 7; however, in Fig. 7 the nowcast is set at the center of the temporal window, as shown in Fig. 8a (temporal window setting I). In contrast, in Fig. 9 the nowcast is set at the ending point of

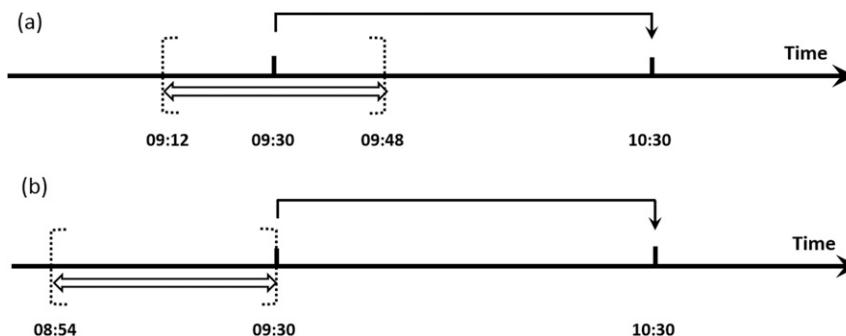


FIG. 8. Schematic diagram of the temporal forecasting window. (a) A temporal window of  $\pm 18$  min (i.e., 36 min). The nowcast is set at the center of the temporal window (temporal window setting I), and (b) the temporal window size (36 min) is the same as in (a), but the nowcast is set at the ending point of the temporal window (temporal window setting II).

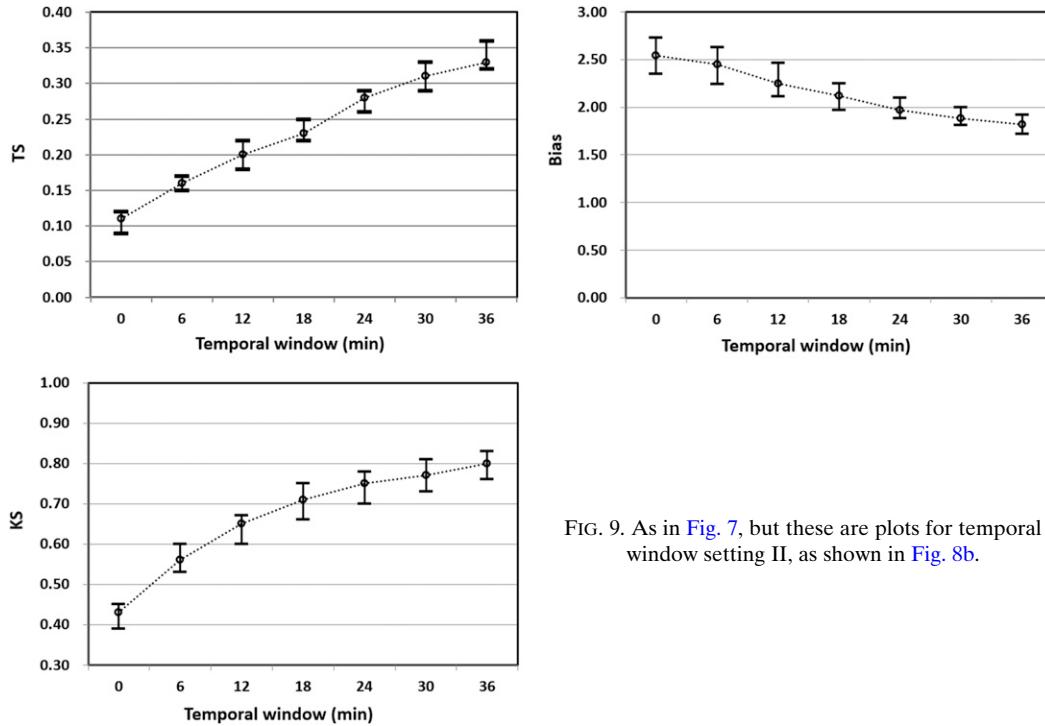


FIG. 9. As in Fig. 7, but these are plots for temporal window setting II, as shown in Fig. 8b.

the temporal window, as shown in Fig. 8b (temporal window setting II). If we compare Figs. 7 and 9, we see that the forecast performance is sensitive both to the window length and the ending point of the temporal window (Table 4). The forecast performance of setting I with a window of  $\pm 6/\pm 12$  min slightly exceeds that of setting II with a window of 12/24 min, but is comparable to that of setting II with a window of 18/30 min, except for the BIAS. Note that temporal window setting II possesses better BIAS given the same window size. We did not compare the larger temporal windows because a temporal shift of 30 min is the maximum tolerable range considered by forecasters at the CWB.

These results indicate that the forecast performance of TANC is more sensitive to the window length than to

the ending point of the temporal window. The forecast performance will not be severely affected if the nowcast is set at the ending point of the temporal window, given the same window size. Therefore, for operational considerations, we adopt the window setting II to determine the most adequate temporal shift of predicted storms. One advantage of setting II is that it has a more reasonable value for the BIAS given the same temporal window length.

*d. Sensitivity of scores to different combinations of spatial and temporal windows*

The results of sensitivity tests when spatial and temporal windows are combined (Fig. 10) showed that both expanding spatial windows and lengthening temporal

TABLE 4. Comparison of verification scores between temporal window settings I and II. The nowcast is set at the center of the temporal window in setting I, but at the ending point of the temporal window in setting II.

Temporal window (min)	Setting I	$\pm 6$		$\pm 12$	
	Setting II	12	18	24	30
TS	Setting I	0.22		0.30	
	Setting II	0.20	0.23	0.28	0.31
BIAS	Setting I	2.45		2.21	
	Setting II	2.25	2.12	1.97	1.88
KS	Setting I	0.70		0.79	
	Setting II	0.65	0.71	0.75	0.77
ETS	Setting I	0.22		0.30	
	Setting II	0.19	0.23	0.27	0.31

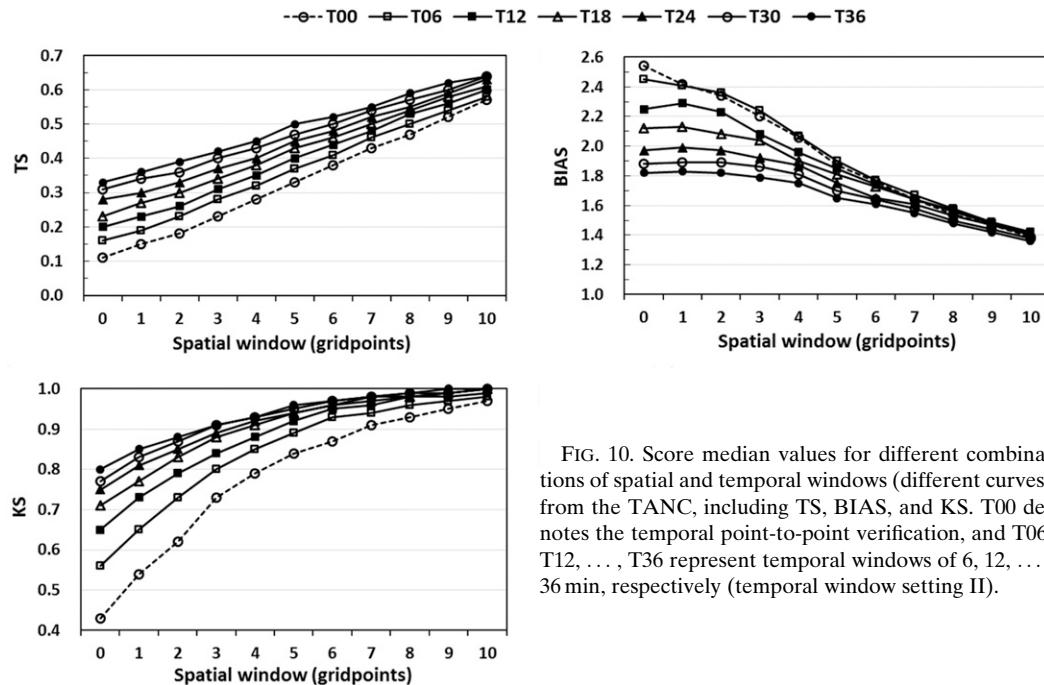


FIG. 10. Score median values for different combinations of spatial and temporal windows (different curves) from the TANC, including TS, BIAS, and KS. T00 denotes the temporal point-to-point verification, and T06, T12, . . . , T36 represent temporal windows of 6, 12, . . . , 36 min, respectively (temporal window setting II).

windows would yield better verification scores. Focusing on the KS, the saturation point is almost at the spatial window of five grid points and the temporal window of 18 min. Adopting this window setting as the forecast guidance for the ACIs, the TS, BIAS, and KS values are 0.43, 1.81, and 0.94, respectively. That is, the TANC displayed moderate accuracy and satisfactory discrimination, but also an acceptable degree of overforecasting. According to the results from the four aforementioned sensitivity tests, the CWB provides a TANC nowcast product for ACIs with an optimal  $L_t$  of 0.6, spatial window of five grid points, and temporal window of 18 min (Fig. 11).

### 5. Comparison with high-resolution hot-start model STMAS–WRF

The next topic we address is whether the TANC can surpass other short-range forecast models currently available. We compared the TANC with a high-resolution hot-start numerical weather prediction model, the STMAS–WRF, to validate its value. The STMAS–WRF, which is also called variational Local Analysis and Prediction System–WRF (LAPS–WRF), applies diabatic data assimilation to mitigate the “spinup” problem and, thus, can produce more accurate forecasts during the early prediction stage (0–6 h) (Chang et al. 2012, 2015). Currently, the STMAS–WRF serves as a crucial reference for short-range (0–6 h) forecasts at the CWB in Taiwan.

The STMAS–WRF domain covers the island of Taiwan and its nearby sea areas with a horizontal resolution of 3 km. It runs hourly with a forecast length of 12 h. Accordingly, we compared only the hourly TANC nowcasts with the corresponding STMAS–WRF 1-h forecasts. To make a fair comparison, all the results below are from pixel-to-pixel verification without any spatial and temporal relaxation for the TANC and STMAS–WRF. For the TANC, we considered 0.6 as the  $L_t$ , which means the TANC is considered to be “predicting CI” if its likelihood value exceeds 0.6. For the STMAS–WRF, we chose 35 dBZ as the threshold for the vertical maximum reflectivity to match with the TANC’s operational definition of convection, which means the STMAS–WRF is regarded as “predicting a convection” if its vertical maximum reflectivity exceeds 35 dBZ in the 1-h forecast.

Because the input data for the STMAS–WRF, such as model background, radar reflectivity, and surface observation data, were not archived in 2014, we could not rerun the STMAS–WRF for the five cases in 2014 in Table 2. Therefore, in addition to the four cases in 2015, we chose four additional cases in 2016 as verification cases to compare the forecast performance of these two systems. A total of 26 forecasts of 1 h from the eight cases were verified.

Here, we use the performance diagram (Roebber 2009) to discern differences in the forecast performance of the TANC and STMAS–WRF (Fig. 12). The

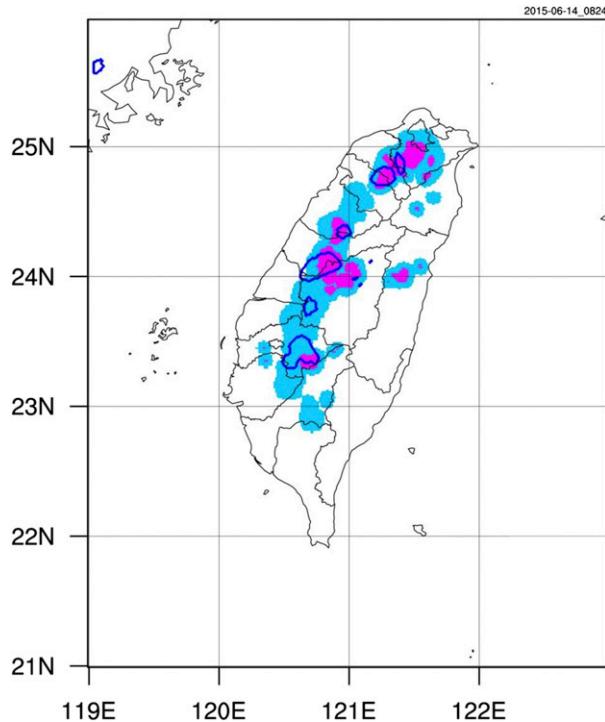


FIG. 11. TANC 1-h nowcast guidance for ACI issued at 0824 UTC 14 Jun 2015. The pink shading shows the most likely regions for ACI (i.e., the areas with likelihood  $\geq 0.6$ ). The blue shading shows the tolerable areas of forecast errors (i.e., storm displacement or time shifting). The dark blue contours of observed ACI are also overlaid for verification.

performance diagram is based on the fact that TS, BIAS, POD, and the success ratio (SR;  $SR = 1 - FAR$ ) are functionally related to each other. The point representing the scores for a perfect forecast (POD, SR, bias score, and TS equal unity) lies in the top-right corner of the diagram and deviations in a particular direction indicate the relative differences in the POD and SR and, consequently, the BIAS and TS. Figure 12 shows that, overall, the TANC had higher prediction accuracy than the STMAS-WRF did for the TS. A total of 17 of the 26 runs from the TANC had TS values greater than 0.1, whereas the STMAS-WRF had only two runs with TS higher than 0.1. Both systems displayed overforecasting but the STMAS-WRF displayed more pronounced overforecasting. The TANC had most parts of runs with BIAS between 1 and 3; by contrast, most of the STMAS-WRF runs had BIAS greater than 3. Generally, the SR values of the TANC were higher than those of the STMAS-WRF. According to these samples of afternoon convective storms, the TANC outperformed the STMAS-WRF. These results also exemplify the difficulties in predicting ACIs by using short-range numerical models, even with a hot-start model.

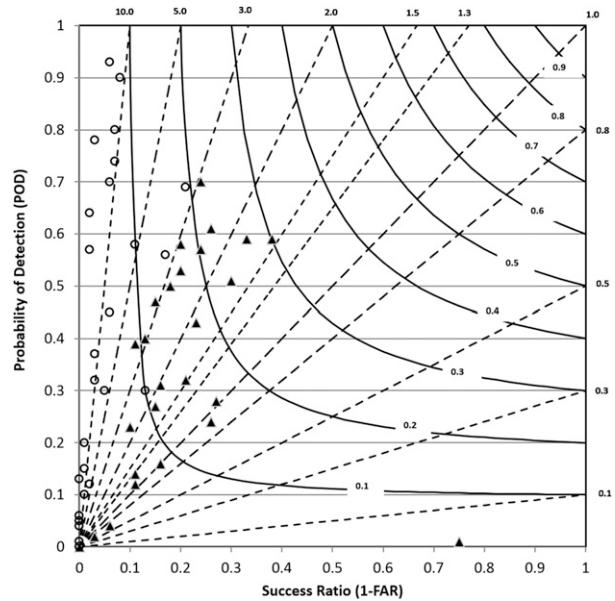


FIG. 12. Performance diagram showing the forecast verification results for the TANC (red dots) and STMAS-WRF (blue dots) based on eight afternoon thunderstorm cases in 2015 and 2016. Curved lines show TS and straight dotted lines represent BIAS.

The sample median TS of the TANC was markedly higher than that of the STMAS-WRF (Fig. 13); however, the sampling variability of the TANC was larger (i.e., the 95% confidence interval was broader). In addition, the median TS difference (paired results) between these two systems showed that the TANC outperformed the STMAS-WRF when applying the median TS as an index of forecast ability.

To test for significant differences in the median TS values between both systems, a Mann-Whitney test (also referred to as a two-sample rank test) was used for hypothesis testing. Although the Mann-Whitney test does not require normally distributed data, a key assumption underlying the test is that the individual samples are independent. A check of the autocorrelation functions of the TS for both systems exhibited small values at various lags, reflecting a weak serial dependence (i.e., nearly independent) in the sample data. Subsequently, the Mann-Whitney test (Mann 1945; Larson 1982) was conducted at a specified level of significance ( $\alpha = 0.01$ ) with the following hypotheses:

$$H_0, \text{ median } TS_1 = \text{median } TS_2; \text{ and} \\ H_A, \text{ median } TS_1 \neq \text{median } TS_2,$$

where the subscripts 1 and 2 for TS refer to the STMAS-WRF and TANC, respectively. Based on the eight cases during 2015 and 2016, the  $p$  value is 0.00008 for a

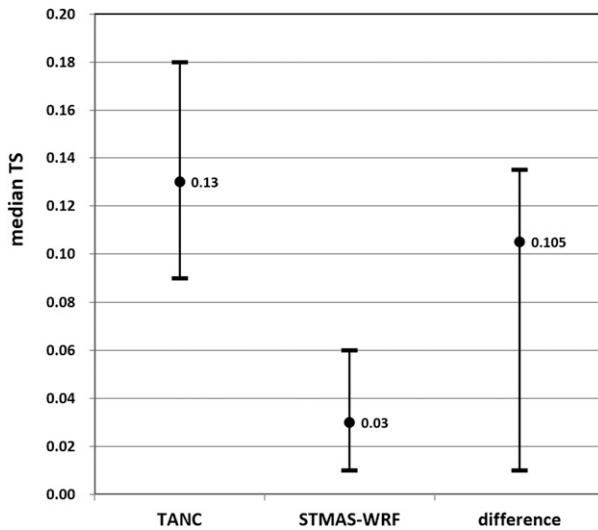


FIG. 13. Sample median TS for TANC and STMAS-WRF, as well as the median TS difference between these two systems with a 95% confidence interval on the basis of a 10 000 bootstrap resampling process.

two-sided test ( $\pm Z_w$ ) and a normal approximation. This result indicates that the probability of no differences existing in forecast ability between both systems was 0.008%. In other words, a statistically significant difference exists between both systems. Please refer to [appendix B](#) for a more detailed description of the Mann-Whitney test.

Strictly speaking, the STMAS-WRF forecasts should be regarded as convective storm nowcasts, which involve nowcasts of storm initiation, growth, and dissipation. Among these three attributes, storm initiation is the most difficult to predict in practice. Although the TANC focuses only on storm initiation, it offers better forecast performance than the STMAS-WRF for these cases.

## 6. Conclusions and future work

Focusing on nine days' worth of afternoon thunderstorms under weak synoptic forcing in 2014 and 2015, we apply the TANC to generate 1-h likelihood nowcasts of ACIs based on a fuzzy logic approach. The primary purpose is to provide more useful nowcast guidance of ACIs for forecasters.

Sensitivity experiments for various  $L_t$ s were conducted to determine the optimal  $L_t$  for indicating ACI. The criterion of threshold selection is optimized to balance the hits against false alarms (or POD against BIAS) in the ACI forecasts. A higher POD, which indicates a greater chance of ACI being detected, is not necessarily the best choice. The optimal value of  $L_t$  for ACI is suggested to be 0.6. The sensitivity

experiments on spatial and temporal windows showed that a combination of a spatial window of 5 km and a temporal window of 18 min is preferred as the acceptable uncertainty range of forecast errors when operational needs are taken into account. Under these conditions, the TANC displays moderate accuracy and satisfactory discrimination with an acceptable degree of overforecasting.

Based on the results from sensitivity experiments, we designed a new nowcast product that only displays the most likely regions (Fig. 11, in pink) for ACIs (i.e., the areas with  $L_t \geq 0.6$ ) instead of likelihood contours. In addition, the tolerable areas of storm displacement (5 km) and temporal shift (18 min) are also shown to indicate the less likely, but still possible, areas of ACIs (Fig. 11, in blue). The TANC was also compared with an operational high-resolution hot-start model (STMAS-WRF) to validate its nowcast value. Verification results and a Mann-Whitney test suggest that the TANC significantly surpasses the competing STMAS-WRF model with a very low  $p$  value.

Based on the 2014 and 2015 cases in this study, about 90% of the samples had likelihood values between 0 and 0.2, few had negative likelihood values, and no samples had a likelihood value greater than 0.95. In the beginning, we disregarded the few negative likelihood values, reset the negative likelihood values to zero, and treated likelihood nowcasts as probabilistic forecasts. Analysis of the reliability diagram showed that the TANC was obviously overforecasting and the likelihoods were not statistically reliable. We also tried to incorporate spatial and temporal windows into the likelihood nowcasts by averaging the likelihood values or taking the median values within the spatial and temporal windows to make them more statistically reliable. Improvement is very limited, however, because the TANC likelihood nowcasts were seriously overforecasting with poor resolution. We think the fundamental reason for the statistical unreliability, even though spatial and temporal windows were incorporated into the likelihood nowcasts, lies in the fact that the likelihood forecasts should not be regarded as probabilistic forecasts.

To produce more accurate ACI nowcasts, some predictors of the TANC should be changed. Among the eight predictors, three predictors (CAPE, CIN, and  $RH_{avg}$ ) come from the analysis field of the regional CWB-WRF model, which is updated every 6 h with a horizontal resolution of 15 km. Additionally, the surface divergence (SurfDiv) predictor is calculated from wind observations at stations with a spacing of approximately 9–10 km. The temporal and spatial resolutions of these four predictors are too low to resolve the atmospheric

characteristics required for ACI nowcasts. In the future, analysis fields of the STMAS–WRF model should be considered as replacements for their equivalents in the CWB–WRF model to provide higher temporal and spatial resolution information on wind, thermal, and humidity fields.

Apart from the predictors listed in Table 1, we will also take into account other predictors. For example, satellite cumulus cloud IR temperature change (Roberts and Rutledge 2003) and onset time of the sea breeze from the coast or the time of sea-breeze passage at key surface stations (Lin et al. 2012) are relevant variables to be considered. Information on the previous day's convective storms would also be beneficial for forecasting the current day's event because of the nature of the persistence of thunderstorm activity. In addition, we plan to track convective cells to determine their movement speed and direction in the future. This capability would provide information on the propagation of predicted storms. By doing so, the regions of spatial forecast uncertainty could be narrowed down by simply performing spatial relaxation along the propagation direction of a storm, instead of searching in all directions in our present study.

*Acknowledgments.* The authors thank the Central Weather Bureau for providing the TANC dataset and computer resources, as well as the NCAR and NOAA/MDL for technical support. The methodology of clustering together similar types of predictors into upstream and primary modules and weighting them, as shown in Fig. 2, was originally developed by Dr. James Wilson of NCAR as part of the TANC software transferred to the CWB. We also thank Dr. LingYan Xin and Mr. John Crockett at the Meteorological Development Laboratory, NOAA/NWS, as well as Dr. Wei-Peng Huang and Miss Yu-Shuang Tang at the CWB, for helpful discussions. Thanks to May Izumi at the University of Hawai'i at Mānoa for editing this manuscript. HLC is sponsored by the Ministry of Science and Technology, Republic of China, under Grants MOST-104-2625-M-052-004, MOST-105-2625-M-052-001, and MOST-106-2625-M-052-001. This is SOEST Contribution Number 10002. NCAR is funded by the U.S. National Science Foundation. PSC was funded by the Honolulu Board of Water Supply under Award C13548001.

## APPENDIX A

### Dichotomous Verification Scores

The six verification scores used in this study are defined as follows:

$$TS = \frac{h}{h + f + m}, \quad (\text{A1})$$

$$BIAS = \frac{h + f}{h + m}, \quad (\text{A2})$$

$$POD = \frac{h}{h + m}, \quad (\text{A3})$$

$$FAR = \frac{f}{h + f}, \quad (\text{A4})$$

$$KS = \text{hit rate } H - \text{false alarm rate } F$$

$$= \frac{h}{h + m} - \frac{f}{f + c}, \quad \text{and} \quad (\text{A5})$$

$$ETS = \frac{h - h_{\text{rdm}}}{h + f + m - h_{\text{rdm}}},$$

$$h_{\text{rdm}} = \frac{(h + m)(h + f)}{h + m + f + c}, \quad (\text{A6})$$

where  $h$ ,  $m$ ,  $f$ , and  $c$  are the four elements in a  $2 \times 2$  contingency table (Table 3).

The TS is also named the critical success index (CSI; Donaldson et al. 1975) or the Gilbert score. We adopted the TS as the index of forecast ability in this study because it was commonly used in various previous studies in assessing the accuracy of thunderstorm forecasts (Huntrieser et al. 1997; Mitchell et al. 1998; Mueller et al. 2003; Mazur et al. 2009), despite having some shortcomings. For example, the TS depends on the climatological frequency of events because some hits can occur purely as a result of random chance. That is, the TS is prone to lower values for rare events and higher values for frequent events. Therefore, a higher TS does not necessarily indicate more accurate forecasting ability. For the evaluation of ACIs in this study, however, the TS value approximated the ETS value because afternoon thunderstorms can be regarded as rare events in the TANC domain.

The ETS is also called the Gilbert skill score (GSS; Gilbert 1884), which is constructed using the Gilbert score (i.e., the TS) as the accuracy measure in the definition of a skill score. That is,  $GSS = (TS - TS_{\text{ref}})/(TS_{\text{perf}} - TS_{\text{ref}})$ , where  $TS_{\text{ref}}$  is the TS value from reference forecasts (i.e., random forecasts) and  $TS_{\text{perf}} = 1$  is the TS value from perfect forecasts. If one needs to compare forecast performance between different samples, the ETS is a better choice than the TS because the ETS could adjust for the effects of differences in the climatological frequencies of the event between samples.

The KS is also known as the Hanssen–Kuipers discriminant score (HKS; Hanssen and Kuipers 1965; Woodcock 1976) or the true skill statistic (TSS; Flueck 1987) and is equal to the hit rate  $H$  minus the false alarm rate  $F$ . The  $H$  (= POD) and  $F$  values are conditioned on

observations. The KS describes how effectively the forecast separates events from nonevents. In other words, the KS measures the discrimination ability of a forecast system. The KS ranges from  $-1$  to  $1$ . A positive KS indicates a forecast with positive discrimination ability, with the perfect score of  $1$ . The KS approaches the POD whenever forecasting is dominated by correct forecasts of nonoccurrence (e.g., forecasts of rare events like severe local storms). A forecaster can maximize the KS by overforecasting rare events; therefore, using the KS alone to evaluate the forecast success of rare events is not strictly proper (Doswell et al. 1990). To avoid this, we adopted other scores (TS, BIAS, and FAR) together with the KS to give a more complete picture of the forecast quality of afternoon convection, which can be regarded as a rare event in the TANC domain. Details for the KS can be found in Gandin and Murphy (1992).

## APPENDIX B

### The Mann–Whitney Test

The Mann–Whitney test (Mann 1945; Larson 1982) determines  $p$  values to assess the significance of differences between two samples by using a normal approximation, which is calculated as follows:

$$Z_w = \frac{\left| W - \frac{n(m+n+1)}{2} \right| - 0.5}{\sqrt{\frac{mn}{12} \left[ (m+n+1) - \frac{\sum_{i=1}^K (t_i^3 - t_i)}{(m+n)(m+n-1)} \right]}}$$

where  $W$  is the test statistics;  $n$  and  $m$  are the size of sample 1 (STMAS–WRF) and sample 2 (TANC), respectively;  $K$  is the number of sets of ties (i.e., pairs with the same values); and  $t_i$  is the number of tied values in the  $i$ th set of ties. The  $W$  value can be calculated using the following four steps: 1) pool the data from the two samples into one batch, 2) rank all the TS values in ascending order, 3) calculate and assign the average rank for the samples that are tied, and 4) calculate the sum of the ranks of the first sample (i.e.,  $W$  value). After the value of  $Z_w$  is calculated, it can be translated into a  $p$  value using a normal approximation and a two-sided test in this study.

## REFERENCES

- Barron, J., D. Fleet, and S. Beauchemin, 1994: Performance of optical flow techniques. *Int. J. Comput. Vis.*, **12**, 43–77, doi:10.1007/BF01420984.
- Benjamin, S., and Coauthors, 2004: An hourly assimilation–forecast cycle: The RUC. *Mon. Wea. Rev.*, **132**, 495–518, doi:10.1175/1520-0493(2004)132<0495:AHACTR>2.0.CO;2.
- Berenguer, M., D. Sempere-Torres, C. Corral, and R. Sánchez-Diezma, 2006: A fuzzy logic technique for identifying non-precipitating echoes in radar scans. *J. Atmos. Oceanic Technol.*, **23**, 1157–1180, doi:10.1175/JTECH1914.1.
- Bermowitz, R. J., and E. A. Zurndorfer, 1979: Automated guidance for predicting quantitative precipitation. *Mon. Wea. Rev.*, **107**, 122–128, doi:10.1175/1520-0493(1979)107<0122:AGFPQP>2.0.CO;2.
- Chang, H. L., 1997: The study on the afternoon convection in Taiwan area. M.S. thesis, Dept. of Atmospheric Sciences, National Central University, Zhongli, Taiwan, 200 pp.
- , H. Yuan, and P. L. Lin, 2012: Short-range (0–12h) PQPFs from time-lagged multimodel ensembles using LAPS. *Mon. Wea. Rev.*, **140**, 1496–1516, doi:10.1175/MWR-D-11-00085.1.
- , S.-C. Yang, H. Yuan, P. L. Lin, and Y. C. Liou, 2015: Analysis of relative operating characteristic and economic value using the LAPS ensemble prediction system in Taiwan area. *Mon. Wea. Rev.*, **143**, 1833–1848, doi:10.1175/MWR-D-14-00189.1.
- Chen, G. T.-J., H. C. Chou, T. C. Chang, and C. S. Liu, 2001: Frontal and non-frontal convection over northern Taiwan in mei-yu season (in Chinese with English abstract). *Atmos. Sci.*, **29**, 37–52.
- Chu, P.-S., 2002: Large-scale circulation features associated with decadal variations of tropical cyclone activity over the central North Pacific. *J. Climate*, **15**, 2678–2689, doi:10.1175/1520-0442(2002)015<2678:LSCFAW>2.0.CO;2.
- Donaldson, R. J., R. M. Dyer, and M. J. Krauss, 1975: An objective evaluator of techniques for predicting severe weather events. Preprints, *Ninth Conf. on Severe Local Storms*, Norman, OK, Amer. Meteor. Soc., 321–326.
- Doswell, C. A., III, R. Davies-Jones, and D. L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting*, **5**, 576–585, doi:10.1175/1520-0434(1990)005<0576:OSMOSI>2.0.CO;2.
- Flueck, J. A., 1987: A study of some measures of forecast verification. Preprints, *10th Conf. on Probability and Statistics in Atmospheric Sciences*, Edmonton, AB, Canada, Amer. Meteor. Soc., 69–73.
- Gandin, L. S., and A. H. Murphy, 1992: Equitable skill scores for categorical forecasts. *Mon. Wea. Rev.*, **120**, 361–370, doi:10.1175/1520-0493(1992)120<0361:ESSFCF>2.0.CO;2.
- Gilbert, G. K., 1884: Finley’s tornado predictions. *Amer. Meteor. J.*, **1**, 166–172.
- Hanssen, A. W., and W. J. A. Kuipers, 1965: On the relationship between the frequency of rain and various meteorological parameters. *Meded. Verh.*, **81**, 2–15.
- Huntrieser, H., H. H. Schiesser, W. Schmid, and A. Waldvogel, 1997: Comparison of traditional and newly developed thunderstorm indices for Switzerland. *Wea. Forecasting*, **12**, 108–125, doi:10.1175/1520-0434(1997)012<0108:COTAND>2.0.CO;2.
- Johnson, R. H., and J. F. Bresch, 1991: Diagnosed characteristics of precipitation systems over Taiwan during the May–June 1987 TAMEX. *Mon. Wea. Rev.*, **119**, 2540–2557, doi:10.1175/1520-0493(1991)119<2540:DCOPSO>2.0.CO;2.
- Jolliffe, I. T., and D. B. Stephenson, 2012: *Forecast Verification: A Practitioner’s Guide in Atmospheric Science*. 2nd ed. John Wiley and Sons, 274 pp.
- Jou, B. J.-D., 1994: Mountain-originated mesoscale precipitation system in northern Taiwan: A case study 21 June 1991. *Terr. Atmos. Oceanic Sci.*, **5**, 169–197, doi:10.3319/TAO.1994.5.2.169(TAMEX).

- Kain, J. S., and Coauthors, 2013: A feasibility study for probabilistic convection initiation forecasts based on explicit numerical guidance. *Bull. Amer. Meteor. Soc.*, **94**, 1213–1225, doi:[10.1175/BAMS-D-11-00264.1](https://doi.org/10.1175/BAMS-D-11-00264.1).
- Lakshmanan, V., J. Crockett, K. Sperow, M. Ba, and L. Xin, 2012: Tuning AutoNowcaster automatically. *Wea. Forecasting*, **27**, 1568–1579, doi:[10.1175/WAF-D-11-00141.1](https://doi.org/10.1175/WAF-D-11-00141.1).
- Larson, H. J., 1982: *Introduction to Probability Theory and Statistical Inference*. 3rd ed. John Wiley and Sons, 638 pp.
- Lin, P. F., P. L. Chang, B. J.-D. Jou, J. W. Wilson, and R. D. Roberts, 2011: Warm season afternoon thunderstorm characteristics under weak synoptic-scale forcing over Taiwan Island. *Wea. Forecasting*, **26**, 44–60, doi:[10.1175/2010WAF2222386.1](https://doi.org/10.1175/2010WAF2222386.1).
- , —, —, —, and —, 2012: Objective prediction of warm season afternoon thunderstorms in northern Taiwan using a fuzzy logic approach. *Wea. Forecasting*, **27**, 1178–1197, doi:[10.1175/WAF-D-11-00105.1](https://doi.org/10.1175/WAF-D-11-00105.1).
- Lin, S. M., and H. C. Kuo, 1996: A study of summertime afternoon convection in southern Taiwan during 1994 (in Chinese with English abstract). *Atmos. Sci.*, **24**, 249–280.
- Lu, C., H. Yuan, B. Schwartz, and S. Benjamin, 2007: Short-range forecast using time-lagged ensembles. *Wea. Forecasting*, **22**, 580–595, doi:[10.1175/WAF999.1](https://doi.org/10.1175/WAF999.1).
- Mann, H. B., 1945: Non-parametric test against trends. *Econometrica*, **13**, 245–259, doi:[10.2307/1907187](https://doi.org/10.2307/1907187).
- Mazur, R. J., J. F. Weaver, and T. H. Vonder Haar, 2009: A preliminary statistical study of correlations between inflow feeder clouds, supercell or multicell thunderstorms, and severe weather. *Wea. Forecasting*, **24**, 921–934, doi:[10.1175/2009WAF2222149.1](https://doi.org/10.1175/2009WAF2222149.1).
- Mitchell, E. D., S. V. Vasiloff, G. J. Stumpf, A. Witt, M. D. Eilts, J. T. Johnson, and K. W. Thomas, 1998: The National Severe Storms Laboratory tornado detection algorithm. *Wea. Forecasting*, **13**, 352–366, doi:[10.1175/1520-0434\(1998\)013<0352:TNSSLT>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<0352:TNSSLT>2.0.CO;2).
- Mueller, C. K., and J. W. Wilson, 1989: Evaluation of the TDWR nowcasting experiment. Preprints, *24th Conf. on Radar Meteorology*, Tallahassee, FL, Amer. Meteor. Soc., 224–227.
- , T. Saxen, R. Roberts, J. Wilson, T. Betancourt, S. Dettling, N. Oien, and H. Yee, 2003: NCAR Auto-Nowcast system. *Wea. Forecasting*, **18**, 545–561, doi:[10.1175/1520-0434\(2003\)018<0545:NAS>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0545:NAS>2.0.CO;2).
- Murphy, A. H., and H. Daan, 1985: Forecast evaluation. *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, A. H. Murphy and R. W. Katz, Eds., Westview Press, 379–437.
- Roberts, R. D., and S. Rutledge, 2003: Nowcasting storm initiation and growth using GOES-8 and WSR-88D data. *Wea. Forecasting*, **18**, 562–584, doi:[10.1175/1520-0434\(2003\)018<0562:NSIAGU>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0562:NSIAGU>2.0.CO;2).
- , A. R. Anderson, E. Nelson, B. G. Brown, J. W. Wilson, M. Pocerich, and T. Saxen, 2012: Impacts of forecaster involvement on convective storm initiation and evolution nowcasting. *Wea. Forecasting*, **27**, 1061–1089, doi:[10.1175/WAF-D-11-00087.1](https://doi.org/10.1175/WAF-D-11-00087.1).
- Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, doi:[10.1175/2008WAF2222159.1](https://doi.org/10.1175/2008WAF2222159.1).
- Saxen, T. R., and Coauthors, 2008: The operational mesogamma-scale analysis and forecast system of the U.S. Army Test and Evaluation Command. Part IV: The White Sands Missile Range Auto-Nowcast system. *J. Appl. Meteor. Climatol.*, **47**, 1123–1139, doi:[10.1175/2007JAMC1656.1](https://doi.org/10.1175/2007JAMC1656.1).
- Stensrud, D. J., and R. A. Maddox, 1988: Opposing mesoscale circulations: A case study. *Wea. Forecasting*, **3**, 189–204, doi:[10.1175/1520-0434\(1988\)003<0189:OMCACS>2.0.CO;2](https://doi.org/10.1175/1520-0434(1988)003<0189:OMCACS>2.0.CO;2).
- , and Coauthors, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499, doi:[10.1175/2009BAMS2795.1](https://doi.org/10.1175/2009BAMS2795.1).
- Szoke, E. J., C. K. Mueller, J. W. Wilson, and E. J. Zipser, 1985: Development of convection and severe weather along outflow boundaries in northeast Colorado: Diagnosing the above-surface environment with a mobile sounding system. Preprints, *14th Conf. on Severe Local Storms*, Indianapolis, IN, Amer. Meteor. Soc., 386–389.
- Trier, S. B., M. A. LeMone, F. Chen, and K. W. Manning, 2011: Effects of surface heat and moisture exchange on ARW-WRF warm-season precipitation forecasts over the central United States. *Wea. Forecasting*, **26**, 3–25, doi:[10.1175/2010WAF2222426.1](https://doi.org/10.1175/2010WAF2222426.1).
- Van den Dool, H. M., and L. Rukhovets, 1994: On the weights for an ensemble-averaged 6–10-day forecast. *Wea. Forecasting*, **9**, 457–465, doi:[10.1175/1520-0434\(1994\)009<0457:OTWFAE>2.0.CO;2](https://doi.org/10.1175/1520-0434(1994)009<0457:OTWFAE>2.0.CO;2).
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, 676 pp.
- Wilson, J. W., and W. E. Schreiber, 1986: Initiation of convective storms by radar-observed boundary layer convergent lines. *Mon. Wea. Rev.*, **114**, 2516–2536, doi:[10.1175/1520-0493\(1986\)114<2516:IOCSAR>2.0.CO;2](https://doi.org/10.1175/1520-0493(1986)114<2516:IOCSAR>2.0.CO;2).
- , and C. K. Mueller, 1993: Nowcasts of thunderstorm initiation and evolution. *Wea. Forecasting*, **8**, 113–131, doi:[10.1175/1520-0434\(1993\)008<0113:NOTIAE>2.0.CO;2](https://doi.org/10.1175/1520-0434(1993)008<0113:NOTIAE>2.0.CO;2).
- , and R. D. Roberts, 2006: Summary of convective storm initiation and evolution during IHOP: Observational and modeling perspective. *Mon. Wea. Rev.*, **134**, 23–47, doi:[10.1175/MWR3069.1](https://doi.org/10.1175/MWR3069.1).
- , N. Crook, C. K. Mueller, J. Sun, and M. Dixon, 1998: Nowcasting thunderstorms: A status report. *Bull. Amer. Meteor. Soc.*, **79**, 2079–2099, doi:[10.1175/1520-0477\(1998\)079<2079:NTASR>2.0.CO;2](https://doi.org/10.1175/1520-0477(1998)079<2079:NTASR>2.0.CO;2).
- , E. E. Ebert, T. R. Saxen, R. D. Roberts, C. K. Mueller, M. Sleigh, C. E. Pierce, and A. Seed, 2004: Sydney 2000 Forecast Demonstration Project: Convective storm nowcasting. *Wea. Forecasting*, **19**, 131–150, doi:[10.1175/1520-0434\(2004\)019<0131:SFDPCS>2.0.CO;2](https://doi.org/10.1175/1520-0434(2004)019<0131:SFDPCS>2.0.CO;2).
- Woodcock, F., 1976: The evaluation of yes/no forecasts for scientific and administrative purposes. *Mon. Wea. Rev.*, **104**, 1209–1214, doi:[10.1175/1520-0493\(1976\)104<1209:TEOYFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1976)104<1209:TEOYFF>2.0.CO;2).
- Xie, Y., S. Koch, J. McGinley, S. Albers, P. E. Bieringer, M. Wolfson, and M. Chan, 2011: A space-time multiscale analysis system: A sequential variational analysis approach. *Mon. Wea. Rev.*, **139**, 1224–1240, doi:[10.1175/2010MWR3338.1](https://doi.org/10.1175/2010MWR3338.1).