

Thunderstorm Strike Probability Nowcasting, a New Algorithm

S. Dance^a, B.Ebert^a and D.Scurrah^a

^aCentre for Australian Weather and Climate Research, GPO Box 1289K, VIC, 3001, Australia
(S.Dance@bom.gov.au)

Abstract: Thunderstorm nowcasting is an important function for the severe weather forecaster. The usual method is for radar to detect thunderstorms, and for downstream systems to generate 'threat areas' for issuing warnings. However, these threat areas are difficult for other systems to work with as they are not mathematically well defined. In this paper we develop an algorithm, THESPA (THunderstorm Environment Strike Probability Algorithm), to translate a radar thunderstorm detection into a 'strike probability', a new concept for thunderstorm nowcasting. This represents the probability a given point will be affected by a thunderstorm in a given period. Being well defined, this can be combined with other strike probabilities or can be utilized by other systems or forecasters to produce a more valuable warning product. The paper includes studies that justify assumptions made about the statistics of thunderstorm motion, and results demonstrating how effective the algorithm is at thunderstorm nowcasting.

Keywords: thunderstorm; strike probability; radar; nowcasting;

1 INTRODUCTION

Nowcasting¹ thunderstorms is an important societal and economic activity. The Australian Bureau of Meteorology has a number of nowcast systems in place to detect thunderstorms (TSs), determine their speed, direction and severity, and make predictions about their location in the future. However, the current systems do not make probabilistic nowcasts, rather they delimit a homogeneous threat area without distinguishing high or low threat gradations. One such system, TIFS (Thunderstorm

¹A nowcast is a short term weather forecast.

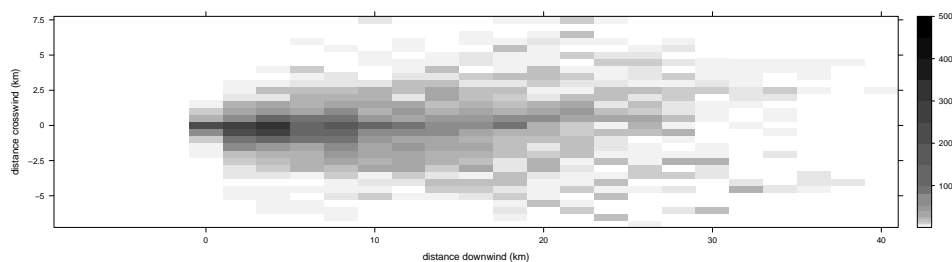


Figure 1: Histogram of detected TS positions compared with their nowcasts, which have been remapped to a nominal direction and speed of 90 degrees and 50 km h^{-1} , with detections remapped accordingly. Approximate counts for a given bin are indicated by the scale on the right.

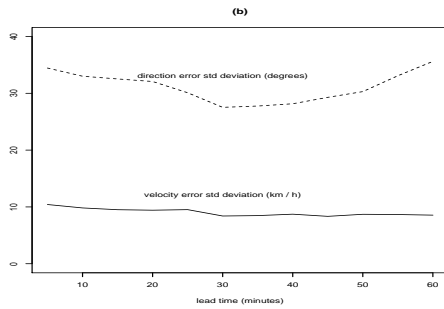
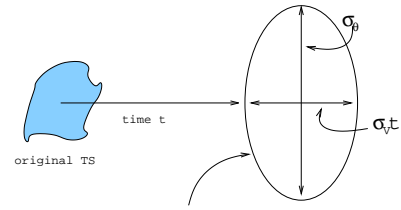


Figure 2(a) *TS nowcast verification statistics by lead time t (minutes) showing standard deviations of errors in V speed (km h^{-1}) (line) and θ direction (degrees) (dotted line)*



1 std deviation region for future TS CoG in polar coordinate system

Figure 2(b) *Shows the original TS and the 1 standard deviation ellipse for its CoG after time t given speed V , and uncertainty in velocity and direction of σ_V , σ_θ .*

Interactive Forecast System, Bally [2004]), displays TSs detected by a radar processing system TITAN (Thunderstorm Identification, Tracking, Analysis And Nowcasting, Dixon and Wiener [1993]) in a graphical user interface (GUI) that allows forecasters to manipulate the nowcast threat area and issue warnings. Currently, this system initially generates an ad hoc threat area based on TS velocity and size. We wish to improve on this by producing a mathematically soundly based probabilistic threat area that allows the output to be combined with probabilistic products from other systems, and forecasters to issue more meaningful warnings. In this paper we present an algorithm, THESPA (THunderstorm Environment Strike Probability Algorithm), to generate such probabilistic nowcasts. In this, we borrow the concept of the 'strike probability' from cyclone forecasting, which gives the probability a given place will be affected by the phenomenon over the nominated period.

In Section 2 we talk about the antecedents of the term "strike probability". In Section 3 we describe TS motion studies which result in an empirical strike probability which we wish to capture with THESPA, and an analysis of how the standard deviation of thunderstorm motion measures vary through the lead time (how far in the future the nowcast is predicting). In Section 4 we describe the assumptions the algorithm depends upon. We then describe the reasoning behind the algorithm by dividing the problem into two stages: firstly in Section 5 we derive a probability density function giving the chance a TS centre of gravity (CoG) touches a point at any time over the nowcast period (ie, 30 or 60 minutes). Then in Section 6 we go on to establish the probability distribution that a point is affected by *any part* of the TS at any time in the nowcast period, ie, the strike probability. In Section 7 we show examples of the distribution, together with statistical results.

2 PREVIOUS WORK

The phrase 'strike probability' has been borrowed from tropical cyclone forecasting. Here the two main approaches are based on numerical weather prediction (NWP) which model the cyclones explicitly. One approach to introducing probability is to employ ensembles of numerical models, the weighted percentage of the ensemble that puts the cyclone within a set distance of a given point is the strike probability for that point [Weber, 2003].

Another approach [Templeton and Keenan, 1982] to the strike probability problem is to use a single NWP forecast given as a number of time steps. The strike probability is calculated by interpolating between the steps, assuming the cyclone path is given by a polynomial, the shape of the cyclone is circular, the probability distribution of each step is given by an *a priori* bivariate normal distribution, and there is statistical independence between steps.

Our approach is closer in spirit to the latter, where probabilities are introduced through an explicit

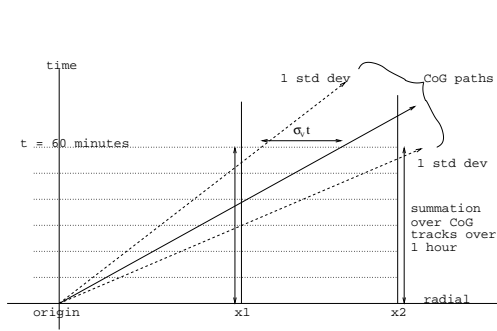


Figure 3(a) Time-radial slice with two points x_1 , x_2 and indicating the 1 standard deviation subset of CoG paths that intersect those points.

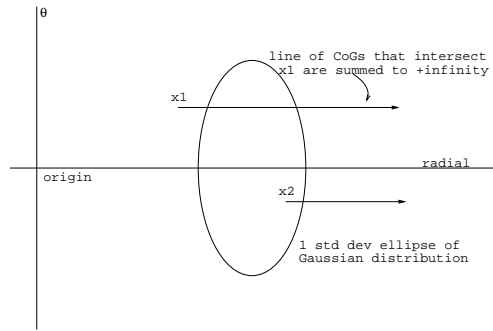


Figure 3(b) Final (ie, 60 minutes) 2D Gaussian showing lines of projected trajectories of the CoG that intersect the points x_1 and x_2 , in polar coordinates.

assumption of a bivariate normal distribution, although our approach assumes a linear path, a storm shape given by the initial detection, and complete dependence along potential storm tracks and mutual exclusivity between tracks.

3 THUNDERSTORM MOTION STUDIES

To determine the statistics of the thunderstorm motion a database of 1682 thunderstorm tracks² was collected from the Kurnell radar near Sydney, Australia, with 8302 TITAN nowcasts of up to one hour and verified against actual motion [Ebert et al., 2005]. It is likely that in places other than Sydney the results found below may vary. The data consists of a forecaster selected set of TITAN output.

We aggregated all TS nowcasts and subsequent detections into one composite histogram, as an empirical “strike probability” chart, where the velocity of the first member of each storm track was taken as the nowcast. The detected locations of the storms in each track, in storm-relative coordinates (nowcasts were remapped to a nominal 50 km h^{-1} and 90 degrees , with detections for each timestep along the TS track remapped accordingly), were recorded in a 25×30 spatial grid by incrementing a counter at the detected position of the TS along its track (each gridbox was $2 \text{ km} \times 0.5 \text{ km}$). This produced the 2D histogram of storm detections shown in Figure 1.

It can be seen that the detected TSs produced a histogram with an elliptical core and triangular envelope. Few TSs last the full hour (which would put the detections on the 50 km mark). In fact most TSs last only a short time or are slow compared with their initial speed, as evident from the large spike in the histogram less than 10 km from the origin. This may be related to the use of TITAN 35 dBZ data, which would include weak storms with a short life.

Another analysis (Figure 2(a)) showed that the standard deviation for speed (V) and direction (θ) errors between nowcast and detection (σ_V and σ_θ), are roughly constant (around 10 km h^{-1} and 30 degrees respectively) over all lead times. This is consistent with the findings of Bellon and Austin [1978] where velocity error and direction error is roughly constant over lead time. Thus we can use these standard deviations for predicting TS motion as THESPA requires (see Section 4).

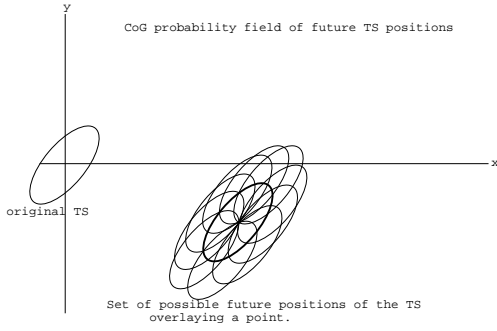


Figure 4(a) Schematic probability density field of possible positions of the CoG of the TS, with for one particular point, the set of possible future positions of the TS that overlay it (TS overlay set). Note that those shown (light ellipses) are examples of the extremes of the TS overlay set, with the particular point being right on their boundaries, and the collection of their CoGs laying on the dark solid ellipse (the TSA boundary). The TS overlay set also includes all those ellipses whose CoGs lie within the TSA boundary.

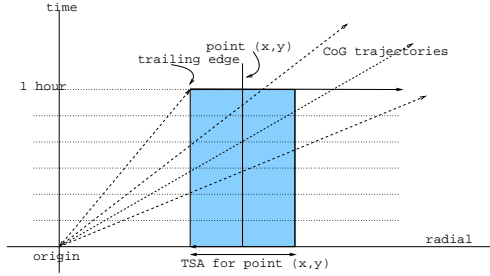


Figure 4(b) Time-radial slice through the set of potential TS CoG trajectories that intersect a point (x, y) (TSA) at any time during the nowcast period (say 30 or 60 minutes). The probability density associated with the trailing edge position already accounts for the contributions from all trajectories passing through it (ie, all trajectories below that point in the figure), and thus that point is the only point one needs to consider to compute the probability density for that radial through the TSA.

4 ASSUMPTIONS

In this paper we present a foundation for the calculation of the probability that a TS is over (affects) some point (x, y) at some time in the nowcast period, given an existing TS at time $t = 0$ defined by its shape and size (here given as an ellipse), speed V and direction θ . For mathematical simplicity it is assumed the TS moves at constant velocity, lasts for the full nowcast period, and does not change shape or size. Of course, these are not realistic assumptions, but are necessary for the development of the algorithm. For instance, it is not clear whether the TS will die early, or become stronger, split, or be replaced by a new TS. We also assume that the distribution of TS speed and direction errors (difference between actual speed and direction and nowcast speed and direction) is a 2 dimensional bivariate normal distribution [Brunk, 1965], and that the standard deviations of this distribution, σ_V and σ_θ , are constant over the nowcast period (as justified by the study of TS statistics in Section 3). The forecaster should be made aware of these assumptions when using the system, although the assumption of linearity corresponds closely to current practice.

5 PROBABILITY DENSITY FUNCTION OF POINTS MEETING THE CENTRE OF THE TS

In this section we calculate the probability density function that a given point meets the CoG of the TS at some time in the nowcast period. This is based on the assumption (above) that the future positions of the TS CoG is given by a two dimensional Gaussian distribution [Brunk, 1965], with standard deviations σ_V , σ_θ (Figure 2(b)).

Below, without loss of generality, we assume the velocity is oriented along the x axis. In the implementation these results can be rotated to the actual direction of the thunderstorm. The formula for the above 2D Gaussian distribution of time-dependent joint speed and direction errors in (r, θ) polar space³ is:

$$f(r, \theta) = (1/2\pi\sigma_V t \sigma_\theta) e^{-(((r-Vt)/\sqrt{2}\sigma_V t)^2 + (\theta/\sqrt{2}\sigma_\theta)^2)} \quad (1)$$

where t is the lead time, and which is normalized so that integrating over all r, θ gives a probability

²A track records the position of a single identified storm for each time step during its life.

³To convert from Cartesian to polar coordinates, use $r = \sqrt{x^2 + y^2}$ and $\theta = \text{atan}(x/y)$.

of 1. Note that the dimension of the probability value this function returns is $angle^{-1}distance^{-1}$.

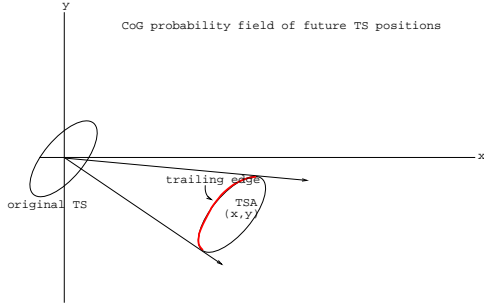


Figure 5: Shows a plan in Cartesian coordinates of the original TS, the probability field of places that meet the CoG of the TS, the TS area (TSA) for a position (x, y) , and the “trailing edge” that contributes all the probability to the sum for that position.

the probabilities of all the possible locations of the CoG that pass through that point up to time t . These CoG positions map onto a line on the 2D Gaussian shown in Figure 3(b).

The probabilities along these lines of CoG intersections are summed to positive infinity for positive radial distance and negative infinity for negative radial distance to produce the probability per degree of θ that the CoG of the thunderstorm met the point x_1 or x_2 at any time in the nowcast period. That is, for x_1 in Figure 3(b) one sums over most of the Gaussian, and for x_2 , which is further from the origin, only the tail end of the Gaussian distribution. Since each CoG trajectory is mutually exclusive of all others, it is possible to combine the probabilities additively:

$$r > 0 : probCoG(r, \theta) = \int_r^{+\infty} (1/2\pi\sigma_V t \sigma_\theta) e^{-((s-Vt)/\sqrt{2}\sigma_V t)^2 + (\theta/\sqrt{2}\sigma_\theta)^2} ds \quad (2)$$

After some manipulation,

$$r > 0 : probCoG(r, \theta) = (1/\sqrt{2\pi}\sigma_\theta) e^{-(\theta/\sqrt{2}\sigma_\theta)^2} (1 - erf((r - Vt)/\sqrt{2}\sigma_V t))/2 \quad (3)$$

and

$$r < 0 : probCoG(r, \theta) = \int_{-\infty}^r (1/\pi\sigma_V t \sigma_\theta) e^{-((s-Vt)/\sigma_V t)^2 + (\theta/\sigma_\theta)^2} ds = \\ (1/\sqrt{2\pi}\sigma_\theta) e^{-(\theta/\sqrt{2}\sigma_\theta)^2} (1 + erf((r - Vt)/\sqrt{2}\sigma_V t))/2 \quad (4)$$

where erf is the **error function** [Fristedt and Gray, 1997], s and ds are radial distance measures, and (r, θ) is a point in polar coordinates. Note that at $r = 0$ the values returned by Equations 3 and 4 are not equal. In fact, we never evaluate them at $r = 0$ as that occurs for points within the detected TS, where of course the probability that the point is affected by the TS is 1.

6 PROBABILITY DISTRIBUTION FOR POINTS AFFECTED BY THE TS

In the last section we found a probability density function for points meeting the CoG of the TS in the nowcast period. In this section, we derive a probability distribution that points will be *affected*

One problem here is that the 2D Gaussian distribution is not designed for a restricted space like the polar plane. We deal with this by remapping the polar plane so that polar angles $\theta > 90$ degrees are converted to $180 - \theta$, $\theta < -90$ are converted to $-180 - \theta$ and polar distance r is converted to $-r$ in both these cases. This allows us to use the same probability density function (PDF) in front of and behind the storm with respect to its velocity, with the caveat that there will be a discontinuity in the PDF around ± 90 degrees.

To compute the probability that any given point (r, θ) meets the CoG of the thunderstorm during time t , say 60 minutes, consider Figure 3(a), which shows time as the vertical axis, and the radial dimension as the horizontal axis.

For a given point x_1 or x_2 , one has to integrate

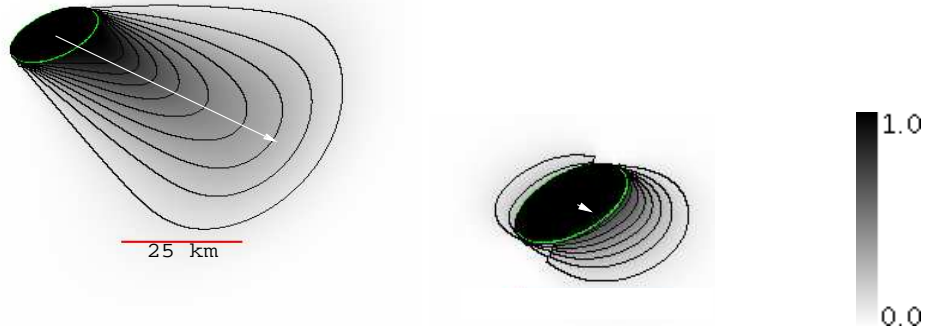


Figure 6: Examples of output from algorithm with typical values derived from TITAN storm tracking. On the left, $\theta = 115$ degrees, $\sigma_\theta = 30$ degrees, $V = 54 \text{ km h}^{-1}$, $\sigma_V = 10 \text{ km h}^{-1}$, on the right, $\theta = 115$ degrees, $\sigma_\theta = 30$ degrees, $V = 4 \text{ km h}^{-1}$, $\sigma_V = 10 \text{ km h}^{-1}$, with probability contours shown at 0.1 intervals. The original storms are shown as ellipses with nowcast velocity as an arrow. A scale marker shows 25 km. Black shading represents a strike probability of 1, white shading represents a strike probability of zero, as indicated by the greyscale bar on the right.

by any part of the TS over the nowcast period. It is assumed that the size and shape of the TS does not change through this period.

To compute the required probability given a specific shaped TS, we need for each point (x, y) to combine the probabilities of all possible future positions of the TS that overlay that point (dubbed TS overlay set, and whose set of CoGs is dubbed TS area or TSA), see Figure 4(a). Thus for each such TS position overlaying (x, y) , we find the probability at its CoG, and for the whole TSA, need to combine such probabilities for all ellipses within the TS overlay set. This requires an operation similar to convolution in image processing, in which the possible TS positions are “raster scanned” over the point.

However, we cannot simply add the probabilities, as they are correlated. That is, some potential positions of the CoG entail other positions, namely those of TS trajectories that lie in the same direction but are faster (which have already passed over the same point earlier in the nowcast period), and which now (at the end of the nowcast period) lie further away from the origin, see Figure 4(b). These “downstream” points should not be included in the probability sum. Since the probability density of its CoG position is in units of angle^{-1} , we can find the total probability that the point (x, y) was affected by the TS ($\text{strikeProb}(x, y)$) by integrating along the “trailing edge” of the TSA (points of the TSA closest to the origin, Figure 5) for that point. Let TE_{xy} be a function from polar angle to radial distance defining the shape of the trailing edge for the given TSA, ie, $TE_{xy}(\theta) \rightarrow r$, with domain θ_1 to θ_2 being the angular limits of the trailing edge, then replacing r in $\text{probCoG}(r, \theta)$ (see Equations 3 and 4) with $TE_{xy}(\theta)$ gives:

$$\text{strikeProb}(x, y) = \int_{\theta_1}^{\theta_2} \text{probCoG}(TE_{xy}(\theta), \theta) d\theta \quad (5)$$

Note that this equation is far cheaper to evaluate than a full convolution due to the much smaller number of points along the trailing edge compared with the full area of the TSA. One then repeats this operation for each point in the geographic region to generate the complete probability field. Note, this is not a probability **density** function, each point now has an associated **probability**. To deal with the case of multiple TSs in the same region, the strike probability distribution can be calculated for each TS separately, and then pointwise combined with the usual rules of probability (assuming independence).

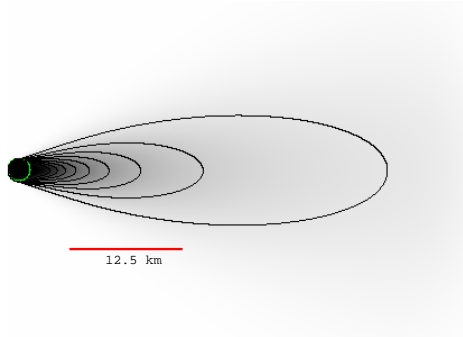


Figure 7(a) Example of algorithm output with values chosen to emulate data in the histogram in Figure 1 ($\theta = 90$ degrees, $\sigma_\theta = 30$ degrees, $V = 50 \text{ km h}^{-1}$, $\sigma_V = 10 \text{ km h}^{-1}$), with probability contours shown at 0.1 intervals, and a scale marker showing 12.5 km. Black shading represents a strike probability of 1, white shading represents a strike probability of zero, as in Figure 6.

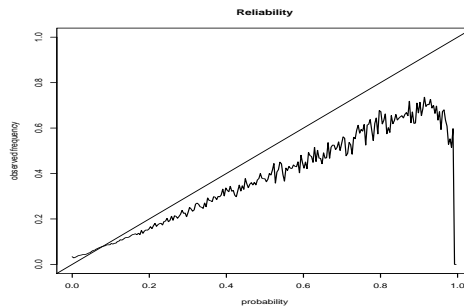


Figure 7(b) Reliability chart for thunderstorm strike probability versus observations: for each probability value the observed frequency of thunderstorms at pixels with that nowcast probability is plotted, compared with the ideal curve, the straight diagonal line.

7 RESULTS

The THESPA algorithm described above was coded in Java. The results for some typical thunderstorm parameters are shown in Figure 6. The original TS shape is shown as an ellipse, whose interior of course has a probability of 1 (black) that the point is affected by the TS. For high speed relative to speed standard deviation (σ_V), the probability field shows a smoothed “triangular” envelope centered around the velocity vector (left side of Figure 6). For slower moving TSs there is a significant chance the TS will move opposite to its initial velocity vector, which shows up as finite probability on the back side of the TS (with respect to velocity). This is clear in the contoured output shown to the right in Figure 6. One can also see in this instance, around the ‘ends’ of the ellipse (at 90 degrees from the velocity vector), the discontinuity due to the remapping of the polar coordinate system described in Section 5.

To compare the output of the algorithm with actual storm behaviour, we ran our algorithm using parameters chosen to match the histogram shown in Figure 1, producing the probability field shown in Figure 7(a). It can be seen that there is indeed some similarity between the two figures, suggesting that our algorithm is generating plausible strike probabilities.

To further compare THESPA with actual storm behaviour, we processed the TITAN database (see Section 3) so that for each storm track, we generated a strike probability nowcast from the first detection that included a nowcast velocity (this is the second detection, as TITAN requires 2 detections to compute velocity), provided the storm was strong enough (ie, VIL, or vertically integrated liquid, above 30 kg m^{-2}) to last a reasonable time. For each increment in predicted strike probability we counted the number of observed TS occurrences and non-occurrences to calculate an observed frequency. We ignored locations where there were no nowcasts. The resultant reliability chart (in which the frequency of a phenomenon is plotted against its nowcast probability [Jolliffe and Stephenson, 2003]) is shown in Figure 7(b). Overall, higher nowcast strike probabilities are associated with higher observed frequency of TS detection. However, at very high probabilities (over 0.9) the low number of cases gives erratic results. In general, THESPA has a tendency to overestimate the probabilities. This is likely to be related to the assumption that storms last for the full nowcast period and do not experience growth or decay.

8 CONCLUSION

The THESPA algorithm presented here has been designed for mathematical tractability and algorithmic simplicity. However, algorithmic simplicity cannot capture the variety of real phenomena. For instance, If a storm track lasts less than the nowcast period, it is either a brief storm, it split or merged, or TITAN lost track of it, whereas our algorithm assumes each storm lasts, and is tracked, for the full nowcast period. Secondly, the observations take place at most every 5 minutes, whereas our algorithm assumes continuous observation. Despite these caveats, THESPA performed reasonably well with good reliability.

Future developments include dealing with storm lifetimes of less than the nowcast period, and performing the calculations within Cartesian coordinates, thus removing the discontinuities resulting from the use of polar coordinates. In this paper we have adapted the concept of strike probability to thunderstorms, and described an algorithm for consistently and defensibly generating them. The algorithm uses a number of simplifying assumptions to allow it to be run in an interactive package (TIFS) in real time. Our claim is that the resultant distribution is nonetheless useful, especially in light of the results given in Section 7.

ACKNOWLEDGEMENTS

The authors wish to thank the anonymous reviewers of this paper who made many useful suggestions.

REFERENCES

- Bally, J. The thunderstorm interactive forecast system: Turning automated thunderstorm tracks into severe weather warnings. *Weather Forecasting*, 19:64–72, 2004.
- Bellon, A. and G. L. Austin. The evaluation of two years of real-time operation of a short-term precipitation forecasting procedure (SHARP). *Journal of Applied Meteorology*, 17(12):1778–1787, 1978.
- Brunk, H. D. *An Introduction to Mathematical Statistics*. Blaisdell Publishing Company, 1965.
- Dixon, M. and G. Wiener. TITAN: thunderstorm identification, tracking, analysis and nowcasting - a radar based methodology. *J. Atmos. Oceanic Tech.*, 10(6):785–797, 1993.
- Ebert, E., T. Keenan, J. Bally, and S. Dance. Verification of operational thunderstorm nowcasts. In *World Weather Research Program Symposium on Nowcasting and Very Short Range Forecasting*, page 8.12, Toulouse, France, 2005.
- Fristedt, B. and L. Gray. *A Modern Approach to Probability Theory*. Birkhaeuser Boston, 1997.
- Jolliffe, I. T. and D. B. Stephenson. *Forecast Verification*. John Wiley and Sons, 2003.
- Templeton, J. I. and T. D. Keenan. Tropical cyclone strike probability forecasting in the Australian region. Technical Report 49, Bureau of Meteorology, GPO Box 1289K, Melbourne, VIC, 3001, Australia, 1982.
- Weber, H. C. Hurricane track prediction using a statistical ensemble of numerical models. *Monthly Weather Review*, 131(5):749–770, 2003.