# The Utility of Atmospheric Sounding Stability Indices for the Evaluation Thunderstorm Potential in Belize

Dennis S. Gonguez
Meteorologist
National Meteorological Service of Belize

ABSTRACT

Stability indices derived from the early morning atmospheric profile have long been considered individually as indicators of thunderstorm potential. This study uses a combination of the more commonly used indices along with moisture and buoyancy elements to develop a thunderstorm threat or risk level within the period of 12 hours from observation time. Verification measures such as false alarm rates and ratios, proportion of perfect forecasts, probability of detection along with the critical success index and bias were used to statistically assess the value of the categorical forecasts. The results were compared with success measures from a single-valued index, the K-index, which is in common use as an indicator for thunderstorm potential.

## 1. INTRODUCTION

The potential of the environment to support thunderstorms is routinely assessed on a daily basis and forecast products defining the likelihood of their occurrence are disseminated in accordance. In Belize a large portion of this assessment is subjective. This research aims at removing some of that subjectivity by injecting a more scientific and empirical method of evaluating thunderstorm potential.

It is a long time tradition that weather forecasters use individual stability indices to assess the likelihood of thunderstorm during the day. Most of these indices specifically evaluate the convective and severe weather potential of the atmosphere using measures of thermal and moisture properties. Each index has its own strengths and weaknesses and no single index can be considered to completely describe the state of the atmosphere A dilemma then arises when these indices, considered individually, provide conflicting or contradictory information on the stability of the atmosphere. This study reveals that the use of a combination of indices, buoyant energy measure and moisture variables provide a better indication of the possibility of thunderstorms as compared with a single value index.

The first section covers data sources along with a description of the different stability indices. This is followed by the section dealing with the method used to assign classes of risk/threat levels to each selected index.

The third section comprises the results of performance measures of forecast quality of the combined indices along with a comparison with an individual index. Included in this section are select examples of successes and failures of using a single valued index as compared with an aggregate. The final section involves the summary and discussions along with future avenues for research in severe thunderstorm forecasting in Belize.

*Corresponding author address*: Dennis S. Gonguez, National Meteorological Service of Belize, Philip Goldson International Airport, Ladyville, Belize, Central America.
E- mail: dgonguez@hydromet.gov.bz

## 2. DATA

*a. Sources and Types*

The 1200 UTC radiosonde ascents for the station operated by the National Meteorological Service of Belize NMS (Belize) (WMO station identification code number 78583) for the period 1st June 2002 to 31st December 2003 formed the main database from which stability indices were collected for analysis. These soundings were obtained from an archive located on the University of Wyoming Atmospheric Science Department's website. (URL: http://weather.uwyo.edu/upperair/sounding.htm.). Soundings within this period corresponded to a total of 127 thunderstorm events. Only 1200 UTC soundings were considered since they represented a dormant, undisturbed boundary layer.

*b. Thunderstorm Data*

If thunderstorms were occurring at the time of release or just prior to release then these soundings were not added to the database. These soundings would not have been necessary since they would inform of an event that was already taking place.

Only thunderstorms observed within +12 hours of radiosonde observing time i.e. 1200 UTC were considered. Observations of thunderstorm occurrences were obtained from records of forecast discussions as produced by local forecasters at NMS (Belize). Such discussions are generated shortly after the forecast is issued. In situations where it was unclear from these discussions that thunderstorms did occur then the local observations of cumulonimbus (Cb) tops or distant Cbs were then taken into consideration. Only thunderstorms occurring with the physical borders of the country were accounted for since the forecast discussions dealt with only the forecast area of responsibility of NMS (Belize).

*c. Indices, buoyancy and moisture variables*

The initial suite of variables used in this study included both stability indices, buoyancy, shear and moisture variables. These stability indices include the Showalter Index or SI (Showalter, 1953), Lifted Index or LI (Galway, 1956), severe weather threat or SWEAT (Miller et al, 1971) and K Index or KI. Measures of Buoyancy in the forms of the Convective Available Potential Energy or CAPE (Moncrief and Miller, 1976) and Convective Inhibition or CIN were also given early consideration. The Bulk Richardson Number or BRC (Weismann and Klemp, 1982) and precipiatable water (PWAT) taken throughout the entire duration of the sounding were also part of the initial database. Normalized CAPE or NCAPE (Blanchard, 1998), defined as the total CAPE divided by the depth of the buoyant layer, initially showed some potential as an indicator for thunderstorm development.

These indices and buoyancy measures were also computed using virtual temperatures to account for any moisture present. The nomenclatures were then altered to reflect the virtual temperature considerations. The LI was changed to LFTV, CAPE to CAPV, CIN to CINV, normalized CAPE to NCAPV and the bulk Richardson number to BRNV. Although moisture considerations are inherently built into the Showalter, K index and severe weather threat their designations were also altered for uniformity and consistency to SHOW, KINX, and SWET respectively.

## 3. METHOD

*a. Multivariate Discriminant Analysis*

Discriminant function analysis was used to determine which indices, moisture and buoyancy variables best predicted whether a thunderstorm would occur. One of the purposes of discriminant analysis (DA), among others, is to discard variables which are of little importance to group distinctions. Stepwise DA was easily computed with the statistical software package SPSS. Stepwise procedures select the most correlated independent variable first, remove the variance in the dependent, then select the second independent which most correlates with the remaining variance in the dependent. This selection goes on until an additional independent variance does not increase the squared correlation denoted by $R^2$ in most statistical literature. It is not within the scope of this paper to further elaborate on this statistical process.

After applying multivariate discriminant analysis the initial group of eight variables was reduced to six. The final suite included: SHOW, LFTV, SWET, KINX, CAPV and PWAT. These formed the categories for which threat levels were determined as described in the following section.

*b. Threat levels*

Using the entire dataset of 127 thunderstorm events from June 1st 2002 to December 31st 2003 maximums, minimums, means and median values for the six selected variables or indices were calculated. Table 1 below shows the calculated values for the specified variables

Table 1. Values of maximum, minimum, mean and medians for six selected variables.

| Var | SHOW | LFTV | SWET | KINX | CAPV | PWAT |
|---|---|---|---|---|---|---|
| Max | 5.7 | 1.81 | 311.4 | 42.5 | 4894 | 66.69 |
| Min | -5.1 | -10.9 | 71.42 | 11.0 | 47.62 | 33.34 |
| Mean | 0.3948 | -5.33 | 209.38 | 30.09 | 2396.4 | 54.46 |
| Median | 0.44 | -5.77 | 212.85 | 31.2 | 2530 | 52.59 |

Ranges were then found using the difference between maximum and minimum values. These ranges were further subdivided into five categories of threat levels for each index, moisture or buoyancy variable. Each individual category was then given a rating based on the threat level with 0 being the lowest and 4 the largest. Tables 2 to 7 depict the different risk levels and ratings for the individual variables.

Table 2. Risk level and rating for SHOW

| Range | Risk | Rating |
|---|---|---|
| Greater than 5.7 | Extremely low or unlikely | 0 |
| 2.2 to 5.6 | Low or slight | 1 |
| -1.5 to 2.1 | Moderate | 2 |
| -5.1 to –1.4 | High or strong | 3 |
| Less than –5.1 | Extremely high | 4 |

Table 3. Risk level and rating for LFTV

| Range | Risk | Rating |
|---|---|---|
| Greater than 1.81 | Extremely low or unlikely | 0 |
| -2.42 to 1.8 | Low or slight | 1 |
| -6.66 to –2.43 | Moderate | 2 |
| -10.9 to –6.67 | High or strong | 3 |
| Less than –10.9 | Extremely high | 4 |

Table 4. Risk level and rating for SWET

| Range | Risk | Rating |
|---|---|---|
| Less than 71.42 | Extremely low or unlikely | 0 |
| 71.5 to 151.4 | Low or slight | 1 |
| 151.5 to 231.4 | Moderate | 2 |
| 231.5 to 311.4 | High or strong | 3 |
| Greater than 311.4 | Extremely high | 4 |

Table 5. Risk level and rating for KINX

| Range | Risk | Rating |
|---|---|---|
| Less than 11.0 | Extremely low or unlikely | 0 |
| 11.1 to 21.5 | Low or slight | 1 |
| 21.6 to 32.0 | Moderate | 2 |
| 32.1 to 42.5 | High or strong | 3 |
| Greater than 42.5 | Extremely high | 4 |

Table 6. Risk level and rating for CAPV

| Range | Risk | Rating |
|---|---|---|
| Less than 47.6 | Extremely low or unlikely | 0 |
| 47.7 to 1663.1 | Low or slight | 1 |
| 1663.2 to 3278.6 | Moderate | 2 |
| 3278.7 to 4894.1 | High or strong | 3 |
| Greater than 4894.1 | Extremely high | 4 |

Table 7. Risk level and rating for PWAT

| Range | Risk | Rating |
|---|---|---|
| Less than 33.34 | Extremely low or unlikely | 0 |
| 33.35 to 44.46 | Low or slight | 1 |
| 44.47 to 55.58 | Moderate | 2 |
| 55.59 to 66.69 | High or strong | 3 |
| Greater than 66.69 | Extremely high | 4 |

Sturtevant (1995) uses six categories of threat levels for Lifted and SWEAT indices and seven for the Showalter index. These ranged from thunderstorms unlikely to "Yikes!!" and "Head to the storm shelter" on the upper extreme.

For each sounding a value of the index falls with a particular range yielding a characteristic rating. The six rating values were then summed to yield an accumulated risk level or potential. This values varied from zero to twenty-four as shown in Table 8.

Table 8. Accumulated risk level or potential

| Range | Potential |
|---|---|
| 0 | Extremely low |
| 1 to 8 | Low or slight |
| 9 to 16 | Moderate |
| 17 to 24 | High or strong |
| 24 | Extremely high |

The following is an example of how the accumulated risk level is derived. Consider the following: SHOW=1.26
LFTV=-4.48
SWET=195.2
KINX=29.3
CAPV=1939
PWAT=42.95.

Then from Tables 2 to 7 the following ratings were obtained: SHOW=2
LFTV=0
SWET=2
KINX=2
CAPV=2
PWAT=1.

To obtain the accumulated risk level these six ratings were then summed to yield a value of 9. From Table 8 this value of 9 corresponds to moderate potential for thunderstorm development.

## 4. PERFORMANCE EVALUATION

### *a. Assessment Measures*

In order to evaluate the forecast skill of the indices seven statistical measures were utilized. These included False Alarm Rates (FAR), False Alarm Ratio (FARatio), Hit Rate (HR), Proportion of Perfect Forecasts (PPF), Critical Success Index (CSI) and Bias (BS).

With a categorical variable, the forecast is for an occurrence or nonoccurrence of a particular event. Categorical forecasts, like rain/no rain, thunderstorm/no thunderstorm or hail/no hail can then be easily simplified to a yes/no statement. So too can observations be placed in one of two categories or bins (event observed/not observed). Using the designation H for all "hits" this would define a yes forecast or the event was predicted and it did occur. Also M used for a "missed" forecast indicates all incorrect no forecast that the event would not occur i.e. not forecast but observed. Then F would indicate a "false alarm" or all incorrect yes forecasts. And finally, Z indicates all correct no forecasts. This is more clearly shown in Table 9.

Table 9. Forecast verification table

| Event | observed | not observed |
|---|---|---|
| **forecast** | H | F |
| **not forecast** | M | Z |

By definition the false alarm rate FAR=F/(F+Z) is the proportion of forecasts of the event when it did not occur. For a forecast it would be ideal if the FAR could be minimal or negligible.

The limits of FAR then are 0 to 1 with zero being a perfect score. The false alarm ratio defined as FARatio= F/(F+H) is sensitive only to false predictions and not to missed events. By under-forecasting the number of events this score can always be increased but at the cost of more missed events. The limits are also the same as that of FAR.

The hit rate (HR) or probability of detection is the proportion of perfect yes forecasts. It measures the success of the forecast in correctly predicting the occurrence of events, HR=H/(H+M). This verification measure is sensitive only to missed events and not to false alarms. By issuing a large number of forecasts on the assumption that a greater number will be correct, would effect an increase in HR. However, this is usually achieved at the cost of more false alarms. The limits of HR lie between 0 and 1 with 1 being a perfect score.

The proportion of perfect forecast is defined as

PPF=(H+Z)/n where n is H+F+M+Z -the total number of forecasts. The upper limit to this measure is 1.

CSI, the critical success index also known as the threat score and Gilbert score (GS) is equal to the total number of correct event forecasts (H) divided by the total number of event forecasts + number of misses i.e. CSI=H/((H+F+M). CSI is not affected by the number of non-event forecasts that were not observed (correct rejections). Higher values of CSI indicate greater success in forecasting the event.

For categorical forecasts, bias is estimated by the ratio of the total number of events forecast to the total number of events observed. The bias BS=(H+F)/(H+M) measures the relative frequency of predicted and observed events, without regard to accuracy. A perfect score indicates that the predicted event is the same as the observed. BS=1 implies no bias, So too BS<1 suggests under-forecasts and BS>1 over-forecasts. Ideally it would be best to achieve a BS=1 by minimizing both F and M.

*b. Performance Assessment Results*

The scientific goal of verification is to focus on learning about the different aspects of the quality of the forecast. A governing principle of verification is that no single measure exists that provides complete and authoritative information about the quality of a forecast product. All scoring systems are deficient in some way or another. This simply means that it is necessary to use a variety of measures to obtain reasonably complete and convincing evaluation measures. This also means that it is important to be aware of the limitations of the various scores so that they are not interpreted and used incorrectly.

In this study accumulated risk levels for the period January to November 2004 were calculated using the seven indices mentioned in section 3.b. Parallel assessments were done for the KINX for the same time period. This data set comprised 222 observations or upper air soundings. *It was found that values of 13 and above for the accumulated risk level illustrated significant levels of confidence that thunderstorms would occur.* Table 9 is a comparison between all correct “no” forecasts (Z), hits (H), false alarms (F) and missed (M) forecasts for the aggregate indices and the single (KINX).

Table 9. Comparison of contingency values for collective indices and single index for period January-November 2004

| Contingency values | Composite indices | KINX |
|---|---|---|
| Z | 158 | 149 |
| H | 26 | 17 |
| F | 16 | 24 |
| M | 22 | 32 |

From the table above it is depicted that for this particular dataset the composite indices/moisture and buoyancy variables clearly outperformed the single valued index in forecasting thunderstorm events. Hits were higher while false alarms and missed events were lower that those of the single index. Although the ideal would be to correctly forecast all events, in real life the failure to forecast a storm that occurred will have far more dramatic consequences than forecasting a storm that did not occur. While it would be desirable to maximize H, minimizing values of F would be far more important than those of M.

Using values from Table 9 performance evaluators as described in section 4.a. were calculated for both the aggregate indices and the single-valued index. All the statistical performance assessments taken collectively indicate an overall better capability of the collective indices to forecast thunderstorms occurring within 24 hours of observation time. This is clearly demonstrated by the assessment scores in table 10.

Table 10: Statistical measures used for comparison of performances of collective and single valued index.

| Statistical Measures | Collective | Single (KINX) |
|---|---|---|
| FARatio | 0.381 | 0.585 |
| FAR | 0.093 | 0.139 |
| HR | 0.542 | 0.349 |
| PPF | 0.784 | 0.703 |
| CSI | 0.406 | 0.233 |
| BS | 0.875 | 0.837 |

False alarms and ratios were significantly lower for the aggregate indices as compared with the KINX. With hit rates and proportion of perfect being higher it could be inferred that the combination indices performs better than the single values index. This is further reinforced by the higher critical success index.

Although the results are encouraging false alarms and missed forecasts are still rather large as shown in table 9. The following section deals with examples of false alarms and missed events.

*c. Select examples*

*Case 1. 12$^{th}$ April 2004*

The Easter Monday morning's sounding as shown in Figure 1.indicated a dry profile with a well-defined inversion near the surface. PWAT values of 38.62 mm and CAPV at 856.6 J/kg were rather low. During the course of the day surface temperatures rose to near 104 F over parts of the interior. The synoptic situation was a cold front extending across the central Gulf of Mexico to the northern tip of the Yucatan peninsula.

(**REMAINDER OF THIS PAGE INTENTIONALLY LEFT BLANK**)

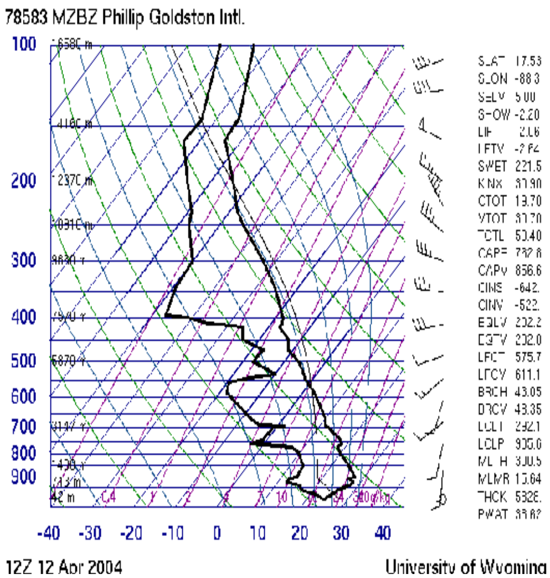


Figure 1. Atmospheric profile for Belize at 1200 UTC 12 April 2004. (Courtesy University of Wyoming, Atmospheric Science Department)

Considering the threat or risk levels using tables 2 to 7 the six values taken singly indicated a low to moderate threat of thundershowers. However, later in the evening severe thunderstorms with hail were reported. These were associated with pre-frontal lift along with warm boundary layer temperatures.

With an accumulated risk level of 10 and the single valued index (KINX) both indicating a moderate threat of thunderstorm activity this case indicates a failure of the scheme to adequately predict the possibility of thunderstorms. This missed event reinforces the need for additional guidance material to analyze the likelihood of thunderstorms. In very dynamic and rapidly changing synoptic/mesoscale conditions the indices from the morning's sounding are of limited use.

*Case 2. 10$^{th}$ April 2004*

This case was chosen because it displayed the highest accumulated risk level for the entire year and reflected a successful performance of the collective indices. With SHOW at –2.66 (Rating =3), LFTV=-7.86 (Rating=3), SWET=233.6 (Rating=3), KINX=33.7 (Rating=3), CAPV=3515 (Rating=3) and PWAT=51.02 (Rating=2). This yielded an accumulated rating of 17 which translates to a high risk of thunderstorm. The atmospheric sounding (Figure 2) shows a shallow moist layer near the surface with another from about 15,000 to 25,000 ft.

The weather that afternoon was hot and humid, typical of this time of the year. Later in the evening a large isolated thunderstorm developed over the west central part of the country. The storm moved northeastwards and subsequently weakened after exiting the mainland around 0100 UTC.

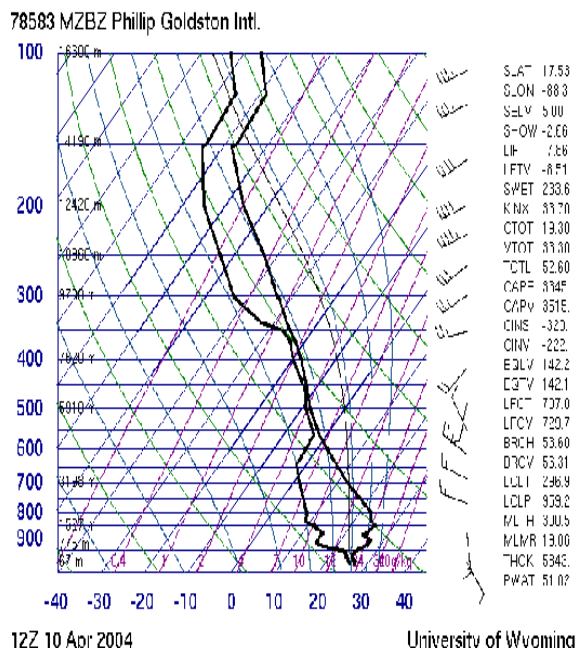


Figure 2. Atmospheric profile for Belize at 1200 UTC on 10 April 2004. (Courtesy University of Wyoming, Atmospheric Science Department)

*Case 3. 27$^{th}$ May 2004*

This case reflects a failure in both the single valued index (KINX) and the aggregate to adequately reflect the thunderstorm potential for that day. The morning's atmospheric profile (See Figure 3) shows a capping inversion at about 7,000 ft (750 mb) and a lack of moisture in the low to mid levels.

As the day progressed skies became cloudy and deep convective cells started to extend vertically. By midday thundershowers affected the southern and central parts of the country with a couple cells in the north. These all but dissipated by 0100 UTC on the 28th May.

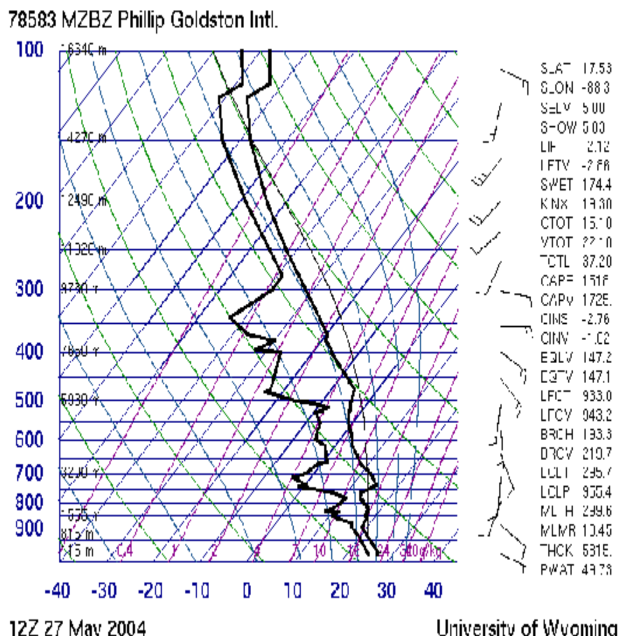


Figure 3. Atmospheric profile for Belize for 27 May 2004. (Courtesy University of Wyoming Atmospheric Science Department)

A KINX value of 19.30 pointed to a low or slight possibility of thunderstorm activity for that day. Meanwhile the accumulated risk level considering the indices, buoyancy and precipitable water values for that sounding was only 10 which at most pointed to a moderate risk of thunderstorms and not the widespread activity experienced that day.

*Case 4. 25th June 2004*

On this day the single value index did not perform while the aggregate gave a clear indication that thunderstorms were likely. KINX value for this sounding was a low 18.9 pointing to a low or slight risk of thunderstorms. Meanwhile the collective indices had a value of 13 indicating a moderate risk of thunderstorms.

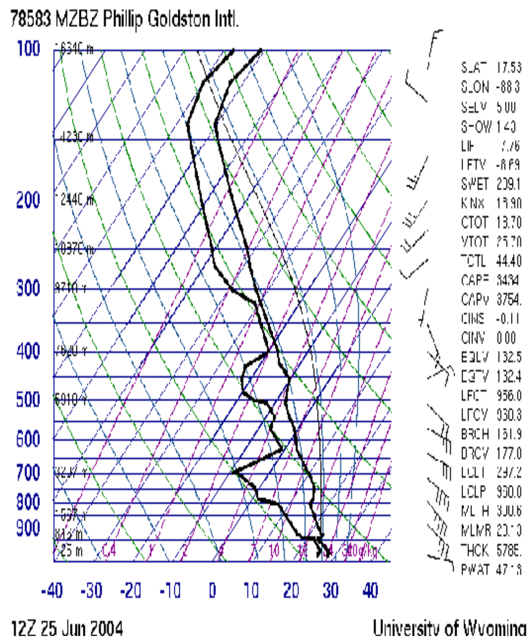


Figure 4. Atmospheric profile for Belize for 25 June 2004 (Courtesy University of Wyoming, Atmospheric Science Department)

Figure 4 above shows much high level moisture, a large "positive" area indicating a high CAPE value and the. convective inhibition was nil. Using the KINX alone on this particular day the forecaster would have missed the development of thunderstorms. However, the high CAPV (3754 J/kg) value and moderate moisture levels (PWAT=47.13 mm) would have hinted at the potential for some thunderstorm activity.

*Case 5. 18th and 19th June 2002*

On Tuesday 18th June a tropical wave which had upper atmospheric support crossed Belize and dumped copious amounts of rainfall resulting in localized flooding near the central portion of the country. In the two days these thunderstorms regenerated themselves over the same area producing an all time one day record of accumulated rainfall of 22.83 inches at one observing station and the capital received a bit more than 13 inches in that same day. The ensuing flood submerged and damaged a bridge on the only artery connecting the capital city with the commercial center of the country thereby severing road transportation between the two cities. High precipitation (HP) thunderstorms (Ahrens 1994) frequently produce heavy precipitation, flash flooding, extreme downbursts. However, these supercells were not considered to be of that variety. Typically HP cells due to

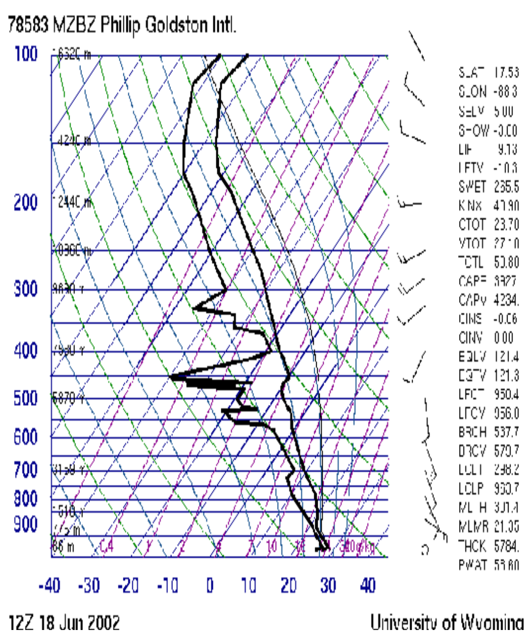


Figure 5. Atmospheric profile for Belize for 18 June 2002. (Courtesy University of Wyoming, Atmospheric Science Department)

precipitation loading would have reduced values of buoyancy. CAPV value in the morning's sounding was 4,234 J/kg. It must be bourn in mind, though, that this sounding was prior to the start of the event. It could be induced that the regeneration of thunderstorms over the same area in a slowly changing synoptic environment was responsible for the deluge of rainfall.

The morning's sounding (See Figure 5 above) revealed the level of instability in the atmosphere. All indices taken individually suggested a very unstable environment and the accumulated risk level was 18. This coupled with the approaching synoptic system (i.e. a tropical wave) would have indicated a high potential for thunderstorm outbreaks. Whether these would have been severe HP cells is another avenue needed to be explored.

## 5. SUMMARY and DISCUSSION

*a. Summary*

Statistical measures were used to assess the performance of an aggregate of stability indices, buoyancy parameter and moisture variables in indicating the likelihood of thunderstorms during the12 hour period after the radiosonde observation time. The performance of this collective was compared with that of a single index commonly used as a signature to thunderstorm activity. The collective made up of Showalter, Lifted, SWEAT, K-Index along with CAPE and precipitable water, was found to outperform the single K-Index considered alone. This was exemplified by lower false alarm rates, higher hit rates, success index and proportion of perfect forecasts.

Several case examples showed the utility and value of the collective index as compared with the single valued index. However, these also revealed that the sounding alone could not be used as the only tool to assess the threat of thunderstorm activity. The entire synoptic and mesoscale evolution of the atmosphere must also be taken into consideration with the sounding and its derived diagnostic indices forming a part of the entire forecast package.

*Discussion*

One of the findings of this research reinforces the fact that in a strongly dynamically changing atmosphere, the morning sounding is of limited use- particularly as it relates to rapidly evolving surface processes.

The primary aim of this research was to derive an objective prognostic tool used to develop some competency in forecasting thunderstorms. This would lead into further competencies in the following categories.

- Forecast thunderstorm areas during the forecast period.
- Forecast potential severe thunderstorms and associated weather
- Formulate policy and issue forecasts containing watches/warnings of impending severe thunderstorm events.

The present state of this objective forecast process is at the first bullet mentioned above. It is the author's aspiration to develop a set of forecast tools along the same lines as this research but as applied to the potential for <u>severe</u> thunderstorms. Subsequent to its completion then warning and dissemination protocols will be established for the country of Belize.

(REFERENCES)


Ahrens, C.D.1994 *Meteorology Today: an Introduction to Weather, Climate and the Environment*. Fifth edition. Chapter 16 notes. West Publishing Co.

Blanchard. D.O., 1998: Assessing the Vertical Distribution of Convective Available Potential Energy. *Wea and Forecasting*., **13**, 870-877.

Galway. J.G., 1956: The Lifted Index as a Predictor of Latent Instability. *Bul. Amer. Meteor. Soc.*, **37**, 528-529.

Miller. R.C., A. Binder and R.A. Maddox, 1971: The Use of Computer Products in Severe Weather Forecasting (the SWEAT Index). Preprints. Seventh Conf. On Severe Local Storms, Kansas City, MS, Amer Meteor. Soc.,1-6.

Moncrief. M.W., and M.J.Miller, 1976: The Dynamics and Simulation of Tropical Cumulonimbus and Squall Lines. *Quart. J.Roy. Meteor.Soc*., **102**, 373-394.

Showalter, A. K., 1954: A Stability Index for Thunderstorm Forecasting. *Bul. Amer. Meteor. Soc*.,**34**, 250-252.

Sturtevant, J.S.,1995: The Severe Local Storm Forecasting Primer. 197 pp.

Weismann, M.L., and J.B. Klemp, 1982: The Dependence of Numerically Simulated Convective Storms on Vertical Wind Shear and Buoyancy. *Mon. Wea. Rev*., **110**, 504-520.