COMBINING HUMAN AND COMPUTER GENERATED FORECASTS USING A KNOWLEDGE BASED SYSTEM

Harvey Stern *

Bureau of Meteorology, Australia

ABSTRACT

Stern's 2005 paper "Defining cognitive decision making processes in forecasting: a knowledge based system to generate weather graphics", presented an analysis of Day 1 to Day 7 rainfall and temperature forecasts during a 100-day real-time trial conducted from February to May 2005. The analysis suggested that adopting a strategy of combining human and computer-generated predictions has the potential to lift the percentage variance explained by human predictions (the current official forecasts) of rainfall and temperature by 7.90%. Forecast verification data from a new 100-day real-time trial, conducted from August to November 2005, on a new set of independent data, was analysed. For this new trial, the knowledge based system (now) modified in order to mechanically combine human and computer-generated predictions and, therefore, to (now) take into account forecasters' valuable domain and contextual knowledge, was utilised. The trial confirmed the conclusion presented in the previous paper, with the percentage variance explained by human predictions (the official forecasts) being lifted by 7.72%. This increase in accuracy arises because:

- In most circumstances, the combining strategy leaves the system's forecasts almost identical to the human (official) forecasts (the percentage variance of the official forecasts explained by the combined forecasts was 77.17%); whilst,
- In those few circumstances when the combining strategy substantially changes the human (official) forecasts, the system's forecasts usually represent an improvement.

There is an increasing interest in the question of what might be the appropriate future role for the human in the forecast process. The results presented here suggest that the future role of human forecasts may be as an input to a system that mechanically combines human predictions with computer generated forecasts.

1. PREFACE

'Consider mechanically integrating judgmental and statistical forecasts instead of making judgmental adjustments to statistical forecasts

…Judgmental adjustment (by humans) of (automatically generated statistical forecasts) is actually the least effective way to combine statistical and judgmental forecasts … (because) judgmental adjustment can introduce bias (Mathews and Diamantopoulos, 1986) (see also, Stern (1996), who documents forecaster over-compensation for previous temperature errors)

…The most effective way to use (human) judgment is as an input to the statistical process

… Cleman (1989) reviewed over 200 empirical studies on combining and found that mechanical combining helps eliminate biases and enables full disclosure of the forecasting process. The resulting record keeping, feedback, and enhanced learning can improve forecast quality' (Sanders and Ritzman, 2001).

2. INTRODUCTION

Sanders and Ritzman (2001) highlight the difficulty associated with utilising (human) judgment as an input to the statistical process 'when the (human) forecaster gets information at the last minute'.

In generating the predictions presented here, the strategy is therefore:

* *Corresponding author address:* Harvey Stern, Bureau of Meteorology, Box 1636, Melbourne, Vic., 3001, Australia; e-mail: h.stern@bom.gov.au

- To take judgmental (human) forecasts (derived with the benefit of knowledge of all available computer generated forecast guidance); and,
- To input these forecasts into a system that incorporates a statistical process to mechanically combine the judgmental (human) forecasts and the computer generated forecast guidance;

Thereby immediately yielding a new set of forecasts.

In this context, the purpose of the present work is:

1. To evaluate the new set of forecasts; and,
2. To document the increase in accuracy achieved by that new set of forecasts over that of the judgmental (human) forecasts.

3. BACKGROUND

Some 30 years ago, Snellman (1977) lamented that whereas the initial impact of guidance material was to increase the accuracy of predictions on account of a healthy human/machine 'mix', operational meteorologists were losing interest and that the gains would eventually be eroded by what he termed the 'meteorological cancer'.

Snellman suggested that producing automated guidance and feeding it to the forecaster who 'modifies it or passes it on', encourages forecasters 'to follow guidance blindly' and concluded by predicting an erosion of recent gains.

Hindsight informs us from forecast verification statistics that the erosion of gains did not take place. In fact, the accuracy of forecasts continued to increase - see, for example, Stern (2005a, 2005c).

Nevertheless, evidence is emerging that the increasing skill displayed by the guidance material is rendering it increasingly difficult for human forecasters to improve upon that guidance (Mass and Baars, 2005; Ryan, 2005).

4. A KNOWLEDGE BASED SYSTEM

Over recent years, the present author has been involved in the development of a knowledge based weather forecasting system (Stern, 2002, 2003, 2004a, 2004b, 2005a, 2005b, 2005c, 2005d). Various components of the system may be used to automatically generate worded weather forecasts for the general public, terminal aerodrome forecasts (TAFs) for aviation interests, and marine forecasts for the boating fraternity.

The knowledge based system generates these products by using a range of forecasting aids to interpret NWP model output in terms of such weather parameters as precipitation amount and probability, maximum and minimum temperature, fog and low cloud probability (Stern and Parkyn, 2001), thunderstorm probability (Stern, 2004b), wind direction and speed, and swell (Dawkins, 2002).

For example, Stern's 2005b forecasts in weather graphic format (Figure 1) are generated from an algorithm that has a logical process to yield HTML code by combining predictions of temperature, precipitation, wind, morning and afternoon weather, and special phenomena (thunderstorm, fog), with features of the forecast synoptic type (strength, direction, and cyclonicity of the surface flow).

5. THE TRIAL OF FEBRUARY TO MAY 2005

Stern (2005b) conducted a 100-day trial (Feb 14, 2005 to May 24, 2005) of the performance of the knowledge based system at predicting Melbourne’s weather, with twice-daily forecasts being generated out to seven days in advance.

During the trial, the overall percentage variance of observed weather explained by the forecasts so generated (the system's forecasts) was 43.24% compared with 42.31% for the official forecasts. That the knowledge based system achieved some success in its attempt to replicate the cognitive decision making processes in forecasting is confirmed by the closeness of the overall percentage variances explained by the two sets of forecasts.

Specifically for precipitation, the percentage variance explained by the quantitative precipitation forecasts and probability of precipitation forecasts so generated was 26.78% compared with 25.07% explained by the official forecasts[1].

On a rain/no rain basis[2], the percentage of correct forecasts so generated was 78.82% compared with 77.64% of the official forecasts.

However, the overall percentage variance of official forecasts explained by the system's forecasts was only 45.91%, indicating that the system's forecasts were not highly correlated with the official forecasts. This was made up of 63.59% of the variance of officially forecast temperature, and 28.23% of the variance of officially forecast precipitation.

This indicates, that, on a day-to-day basis, there are significant aspects of the processes employed in deriving the official forecasts that are not taken into account by the system's forecasts (in all likelihood what Sanders and Ritzman (2001) refer to as 'domain knowledge'), and vice versa.

Sanders and Ritzman (2001) define 'domain knowledge' as 'knowledge practitioners gain through experience as part of their jobs' and make particular reference to that component of domain knowledge named 'contextual knowledge, which is the type of knowledge one develops by working in a particular environment.' 'The quality of domain knowledge is affected by the forecaster's ability to derive the appropriate meaning from the contextual (or environmental) information' (Webby et al., 2001).

Combining forecasts by mathematically aggregating a number of individual forecasts increases the reliability of forecasts (Kelley, 1925; Stroop, 1932) and averages out unsystematic errors (but not systematic biases) in cue utilization.

A common method for combining individual forecasts is to calculate an equal weighted average of individual forecasts' (Stewart, 2001). However, under some conditions unequal weights make sense 'if you have strong evidence to support unequal weighting' (Armstrong, 2001b).

---

[1] The official *Amount of Precipitation* forecasts are expressed in terms of rainfall ranges and, for verification purposes, the *Amount of Precipitation* forecast is taken to be the mid-point of the range forecast:

Range 0 = No precipitation; Range 1 = 0.2 mm to 2.4 mm (1.3 mm); Range 2 = 2.5mm to 4.9mm (3.7 mm); Range 3 = 5.0mm to 9.9mm (7.5mm); Range 4 = 10.0mm to 19.9mm (14.9mm); Range 5 = 20.0mm to 39.9mm (29.9mm); Range 6 = 40.0mm to 79.9mm (59.9mm); and, Range 7 = 80.0mm or more (119.9mm).

The official *Probability of Precipitation* forecasts were taken to be Stern's (1999) interpretation of the words utilised in the official forecasts. The validity of this interpretation was verified on the basis of the data collected during the trial of February to May 2005 and modified, for subsequent application, on the basis of this verification (refer to Table 1).

[2] For verification purposes, it is said that there has been rain on a particular day when at least one of the 0300, 0600, 0900, 1200, 1500, 1800, 2100, or 2400 Melbourne CBD present or past weather observations include a report of precipitation, with a recording of at least 0.2 mm during the preceding three hours.

Should at least one of the 0300, 0600, 0900, 1200, 1500, 1800, 2100, or 2400 Melbourne CBD present or past weather observations include a report of precipitation, but with a recording of only a 'trace' during the preceding three hours, the day is not regarded as 'rain day'. However, in this circumstance, for the purposes of verifying the forecast *Amount of Precipitation*, the amount fallen is regarded as being 0.1mm, and for the purposes of verifying the forecast *Probability of Precipitation*, the *Probability of Precipitation* is regarded as 50%.

Should at least one of the 0300, 0600, 0900, 1200, 1500, 1800, 2100, or 2400 Melbourne CBD present or past weather observations include a report of distant precipitation, but with a recording of 0.0mm during the preceding three hours, the day is not regarded as 'rain day'. In this circumstance, for the purposes of verifying the forecast *Amount of Precipitation*, the amount fallen is regarded as being 0.0mm, and for the purposes of verifying the forecast *Probability of Precipitation*, the *Probability of Precipitation* is regarded as 25%.

Krishnamurti et al. (1999) found that weather forecasts based on a combined forecast using weights based on regression were more accurate than combined forecasts with equal weights.

6. COMBINING FORECASTS

Regarding the two sets of forecasts as partially independent and utilising linear regression to optimally combine the estimates of minimum temperature, maximum temperature, precipitation amount, and precipitation probability, Stern (2005b) demonstrated a lift in the overall percentage variance of observed weather explained.

This result suggested that adopting such a strategy of optimally combining the official and system predictions has the potential to deliver a set of forecasts that are substantially more accurate than those currently issued officially.

Indeed, the overall percentage variance of observed weather explained (an excellent measure of the usefulness of the forecasts) was lifted (by the consensus forecasts) to 50.21% from 43.24% (system) and 42.31% (official), a lift of 7.90% from that achieved by the official forecasts.

The accuracy increases because 'Combining is most effective when the forecasts combined are not correlated and bring different kinds of information to the forecasting process' (Sanders and Ritzman, 2001) and that although 'both (human) intuitive and (computer) analytic processes can be unreliable … different kinds of errors will produce that unreliability' (Stewart, 2001).

What these data suggested was that adopting a strategy of combining predictions has the potential to deliver a set of forecasts that explain as much as 7.90% more variance than that explained by forecasts currently issued officially.

In fact, forecast verification data from a new real-time trial presented in the sections that follow, demonstrate that a substantial increase in accuracy is, indeed, achievable, were one to adopt such a strategy.

7. MODIFYING THE SYSTEM

The knowledge based system has been modified so that it now automatically integrates judgmental (human) forecasts and the computer generated guidance, thereby incorporating the forecasters' valuable contextual knowledge into the process. Sanders and Ritzman (2001) refer to their 1992 study, in which they demonstrated that judgmental forecasts based on contextual knowledge were significantly more accurate than those based on technical knowledge (and) … were even superior to (a) … statistical model.' The knowledge based system, so modified, underwent a 'real-time' trial, the results of which are evaluated in the present paper.

This process of integrating human and computer generated forecasts is illustrated for *Probability of Precipitation* estimates in Figure 2.

Stern (1999) published a proposed interpretation of words used in forecasts in terms of *Probability of Precipitation* and *Amount of Precipitation.*

The system includes an algorithm that interprets the (official) worded precis in terms of *Probability of Precipitation* and *Amount of Precipitation*. This algorithm was derived from Stern's (1999) proposed interpretation and a verification of the official precis that was conducted during the trial of February to May 2005.

By way of illustration, an extract of the probability (%) algorithm, and an extract of the amount (mm) algorithm, are respectively given in Tables 1 and 2.

Because the system's weather icons (Figure 1) arise largely from the system's generated *Probability of Precipitation*, and, conversely, the human (official) *Probability of Precipitation*, arises from an an algorithm that interprets the (official) worded precis, any verification of the *Probability of Precipitation* may also be regarded as representing a verification of forecast *sensible weather*.

8. THE TRIAL OF AUGUST TO NOVEMBER 2005

The new 100-day trial, conducted with a fresh set of data (Aug 20, 2005 to November 27, 2005), of the performance of the modified system involves daily forecasts being generated out to seven days in advance.

Evaluation of the forecasts prepared during the 100 days of the trial shows that the overall percentage variance of official forecasts explained by the system's forecasts is now lifted to 77.17% (from 45.91% previously), demonstrating that, in most circumstances, the combining strategy leaves the system's forecasts almost identical to the official forecasts..

This is made up of 84.37% of the variance of officially forecast temperature (63.59% previously), and 69.97% of the variance of officially forecast precipitation (28.23% previously).

Furthermore, the overall percentage variance of observed weather explained (a sound measure of the usefulness of the forecasts) is now lifted by the system to 40.15% from 32.43% (official) – a rise of 7.72%, which is close to the 7.90% lift suggested previously by Stern's (2005b) consensus forecasts, demonstrating that, in those few circumstances when the combining strategy substantially changes the official forecasts, the system's forecasts usually represent an improvement on the official forecasts.

This substantial increase in accuracy arises because:

- In most circumstances, the combining strategy leaves the system's forecasts almost identical to the human (official) forecasts; whilst,
- In those few circumstances when the combining strategy substantially changes the human (official) forecasts, the system's forecasts usually represent an improvement on the human (official) forecasts.

Figure 3 shows that the overall percentage variance of the observed weather explained is lifted by between 5% and 10% at most lead times.

Specifically for precipitation, the percentage variance explained is lifted by the system to 32.98% (made up of 40.33% for *Probability of Precipitation* and 25.63% for *Amount of Precipitation*) from 23.73% (official) - made up of 29.30% for *Probability of Precipitation* and 18.17% for *Amount of Precipitation*.

On a rain/no rain basis, the percentage of correct forecasts generated by the system is lifted by the system to 76.14% from 70.43% (official).

The root mean square error (rmse) of the √(*Amount of Precipitation* forecast) is reduced by the system to 0.973 mm from 1.108 mm (official). The rmse of the √(*Amount of Precipitation* forecast) is a preferred verification parameter to (*Amount of Precipitation* forecast) in order to reduce the skewness in the distribution of the latter.

Figure 4 shows that the overall percentage variance of the observed precipitation explained is lifted by between 6% and 12% at most lead times.

Specifically for temperature, the percentage variance explained is lifted by the system to 47.32% (made up of 45.83% for minimum temperature, and 48.81% for maximum temperature) from 41.13% (official) - made up of 41.58% for minimum temperature, and 40.68% for maximum temperature.

The rmse of the temperature forecasts generated by the system was 2.604 deg C (made up of 2.634 deg C for minimum temperature, and 2.573 deg C for maximum temperature) compared with 2.775 deg C (made up of 2.704 deg C for minimum temperature, and 2.845 deg C for maximum temperature) for the official forecasts.

Figure 5 shows that the overall percentage variance of the observed temperature explained is lifted by between 5% and 9% at most lead times. Only at Day-1, is the overall percentage variance of the observed temperature explained not lifted.

These results indicate that, on a day-to-day basis, what Sanders and Ritzman (2001) refer to as 'domain knowledge', is now taken into account by the system.

9. OTHER WEATHER ELEMENTS

The system also develops predictions of other weather elements (without directly utilising the combining process), and predictions for other localities. These include:

- Forecasts of 9am and 3pm wind speed and direction at Melbourne Airport. The system's forecasts of wind speed explain 47.73% of the variance of the observed wind speed (compared with 48.96%

explained by the official forecasts) and predict (within half an octant) the wind direction on 72.34% of occasions (compared with 73.40% of the official forecasts).
There is considerable potential for an increase in accuracy of the wind speed forecasts. Averaging the system and official wind speed forecasts would lift the percentage variance explained to 56.46%. There is also considerable potential for an increase in accuracy of the wind direction forecasts. Averaging the system and official wind direction forecasts would lift the percentage predictions of wind direction (within half an octant) to 74.47%.

- Forecasts of the rare weather elements - thunderstorms[3] and fog[4]. The *Critical Success Index* (Wilks, 1995) of the system's forecasts of these elements is 0.000 for fog (the system failed to forecast fog on the occasions when it occurred), and 0.221 for thunderstorms. The *Critical Success Index* was 0.042 and 0.224 for official forecasts of fog and thunderstorms, respectively.
  There is considerable potential for an increase in accuracy of the rare weather element forecasts. From Figure 6a, it may be seen that the verification data suggests:
  1) Reducing the probability criterion under which there is a categorical reference to fog by the system from 15% to 5% (when also accompanied by *Probability of Precipitation* of 25% or less – to exclude potential drizzle situations); and,
  2) Reducing the probability criterion under which there is a categorical reference to thunderstorms by the system from 25% to 5% (when also accompanied by *Probability of Precipitation* of 50% or more);

  would lift the *Critical Success Index* of the system's forecasts of these elements to 0.085 for fog, and 0.266 for thunderstorms.
  That the probability criteria were set too high became apparent during the early stages of the trial, and the system was therefore modified to operate with 5% probability criteria from Day-43.
  An alternative approach would be to examine the relationship between the probability criterion and the precentage profit to be gained from protecting against the occurrence of one of these rare weather elements (Personal Communication: Ross Keith). This is illustrated in Figure 6b, which suggests (for the case of the cost of protection being one fifth the financial loss suffered if the event occurs without protection) an alternative view that:
  1) For fog, the probability criterion should be set to 6%; and,
  2) For thunderstorms, the probability criterion should be set to 12%.
- Forecasts for a number of other Central District localities. Verification of the maximum temperature component of these forecasts reveals that, expressed as an expected departure from Melbourne's maximum temperature, the mean absolute error of the system's forecasts was 0.961 deg C, compared with 1.103 deg C for the official forecasts.

10. CONCLUDING REMARKS

Stern's (2005b) paper "Defining cognitive decision making processes in forecasting: a knowledge based system to generate weather graphics", presented the results of a 100-day trial which suggested that adopting a strategy of combining human and computer-generated predictions has the potential to deliver a set of forecasts that explain about 7.90% more variance than that explained by forecasts currently issued officially.

---

[3] For verification purposes, it is said that there has been a thunderstorm in the metropolitan area during a particular day when at least one of the 0300, 0600, 0900, 1200, 1500, 1800, 2100, or 2400 Melbourne CBD and/or Melbourne Airport observations include a report of cumulonimbus with an anvil and/or lightning and/or funnel cloud and/or thunder (with or without precipitation) – refer to Stern (1980).

[4] For verification purposes, it is said that there has been fog in the metropolitan area during a particular day when at least one of the 0300, 0600, 0900, 1200, 1500, 1800, 2100, or 2400 Melbourne CBD and/or Melbourne Airport observations include a report of fog (including shallow fog) and/or distant fog.

Forecast verification data from a new real-time trial, conducted on the knowledge based system (now) modified in order to mechanically combine human and computer-generated predictions and, therefore, to (now) take into account forecasters' valuable domain and contextual knowledge, was analysed.

The analysis, the results of which are summarised in Table 3, confirmed the conclusion presented in the previous paper, showing that more than 7% extra variance was explained (over that explained by human predictions) therefore demonstrating that a substantial increase in forecast accuracy is, indeed, achievable, were one to adopt such a strategy of combining human and computer-generated predictions.

There is an increasing interest in the question of what might be the appropriate future role for the human in the forecast process (Stewart, 2005).

The results presented here suggest that the future role of human forecasts may be as an input to a system that mechanically combines human predictions with computer generated forecasts.

11. REFERENCES

Armstrong, J. S., 2001a: Principles of forecasting: a handbook for researchers and practitioners. *Kluwer Academic Publishers*.

Armstrong, J. S., 2001b: Combining forecasts (refer to *Armstrong, 2001a, 417-439).*

Cleman, R.T., 1989: Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5,* 559-583 (refer to *Armstrong, 2001a, 411).*

Dawkins, S. S., 2002: A web-based swell and wind forecasting tool. *9th Conference of the Australian Meteorological and Oceanographic Society, Melbourne, Australia, 18-20 Feb., 2002.*

Kelley, T. L., 1925: The applicability of the Spearman-Brown formula for the measurement of reliability. *Journal of Educational Psychology, 16,* 300-303 (refer to *Armstrong, 2001a, 95).*

Krishnamurti, T. N., Kishtawal, C. M., LaRow, T. E., Bachiochi, D. R., Zhan Zhang, Williford, C. E., Gadgil, S., and Surendran, S., 1999: Improved multi-modal weather and seasonal climate forecasts from multimodel "superensemble". Science, 285, 1548-1550 (refer to *Armstrong, 2001a, 423)*.

Mass, C. F. and Baars, J., 2005: The performance of National Weather Service forecasts compared to operational, consensus, and weighted model output statistics. *21st Conference on Weather Analysis and Forecasting; 17th Conference on Numerical Weather Prediction. American Meteorological Society, Washington, DC, 1-5 August, 2005.*

Mathews, D. P. and Diamantopoulos, A., 1990: Judgmental revision of sales forecasts: effectiveness of forecast selection. *Journal of Forecasting, 6,* 407-415 (refer to *Armstrong, 2001a, 411).*

Ryan, C., 2005: Implemenation of operational consensus forecasts. *Bureau of Meteorology Analysis and Prediction Operations Bulletin No. 60.*

Sanders, N. R. and Ritzman, L. P., 1992: The need for contextual and technical knowledge in judgmental forecasting (refer to *Armstrong, 2001a, 406).*

Sanders, N. R. and Ritzman, L. P., 2001: Judgmental adjustment of statistical forecasts (refer to *Armstrong, 2001, 405-416*).

Snellman, L. W., 1977: Operational forecasting using automated guidance. *Bulletin of the American Meteorology Society, 58,* 1036-1044.

Stern, H., 1980: The development of an automated system of forecasting guidance using analogue retrieval techniques. M. Sc. Thesis, Department of Meteorology, University of Melbourne, subsequently published in 1985 as *Meteorological Study 35, Bureau of Meteorology, Australia.*

Stern, H., 1996: Statistically based weather forecast guidance. Ph. D. Thesis, School of Earth Sciences, University of Melbourne, subsequently published in 1999 as *Meteorological Study 43, Bureau of Meteorology, Australia.*

Stern, H., 1999: Establishing the limits of predictability at Melbourne, Australia. *Aust. Meteor. Magazine,48,* 159-167.

Stern, H., 2002: A knowledge-based system to generate internet weather forecasts. *18th Conference on Interactive Information and Processing Systems, Orlando, Florida, USA 13-17 Jan., 2002.*

Stern, H., 2003: Progress on a knowledge-based internet forecasting system. *19th Conference on Interactive Information and Processing Systems, Long Beach, California, USA 9-13 Feb., 2003.*

Stern, H., 2004a: Incorporating an ensemble forecasting proxy into a knowledge based system. *20th Conference on Interactive Information and Processing Systems, Seattle, Washington, USA 11-15 Jan., 2004.*
Stern, H., 2004b: Using a knowledge based system to predict thunderstorms. *International Conference on Storms, Storms Science to Disaster Mitigation, Brisbane, Queensland, Australia 5-9 Jul., 2004.*
Stern, H., 2005a: Using a knowledge based forecasting system to establish the limits of predictability. *21st Conference on Interactive Information and Processing Systems, San Diego, California, USA 9-13 Jan., 2005.*
Stern, H., 2005b: Defining cognitive decision making processes in forecasting: a knowledge based system to generate weather graphics. *21st Conference on Weather Analysis and Forecasting; 17th Conference on Numerical Weather Prediction. American Meteorological Society, Washington, DC, 1-5 August, 2005.*
Stern, H., 2005c: Establishing the limits of predictability at Melbourne, Australia using a knowledge based forecasting system and NOAA's long range NWP model. *Submitted to Australian Meteorological Magazine.*
Stern, H., 2005d: Generating quantitative precipitation forecasts using a knowledge based system. *17th BMRC Modelling Workshop, Melbourne, Australia, 3-6 Oct., 2005.*
Stern, H. and Parkyn, K., 2001: A web-based Melbourne Airport fog and low cloud forecasting technique. *2nd Conference on Fog and Fog Collection, St John's, New Foundland, Canada 15-20 Jul.,2001.*
Stewart, N. A., 2005: Forum on the future role of the human in the forecast process. Part 2: Cognitive psychological aspects of expert forecasters (Chairperson: N. A. Stewart). *21st Conference on Weather Analysis and Forecasting; 17th Conference on Numerical Weather Prediction. American Meteorological Society, Washington, DC, 1-5 August, 2005.*
Stewart, T. R., 2001: Improving reliability of judgmental forecasts (refer to *Armstrong, 2001a, 81-106*).
Stroop, J. R., 1932: Is the judgment of the group better than the average member of the group? *Journal of Experimental Psychology, 15,* 550-560 (refer to *Armstrong, 2001a, 95*).
Webby, R., O'Connor, M, and Lawrence, M., 2001: Judgmental time-series forecasting using domain knowledge (refer to *Armstrong, 2001a, 81-106).*
Wilks, D., 1995: Statistical methods in atmospheric sciences. Academic Press.

Acknowledgments. To Stuart Coombs, for drawing attention to the skill displayed by the NOAA Global Forecasting System (used in the trial), to Marco Berghege for his suggestion regarding the programming utilised herein, to Neville Nicholls and Frank Woodcock, for their comments on combining forecasts, to Noel Davidson, for his comments on the NWP output, and to my other Bureau of Meteorology colleagues for their helpful discussions and feedback.

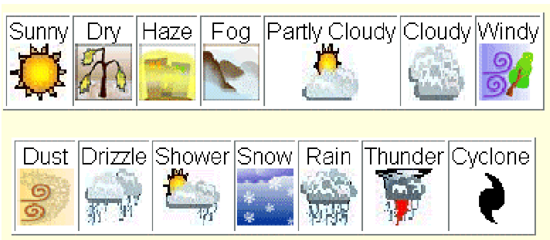

Figure 1. Stern's 2005b weather graphics.

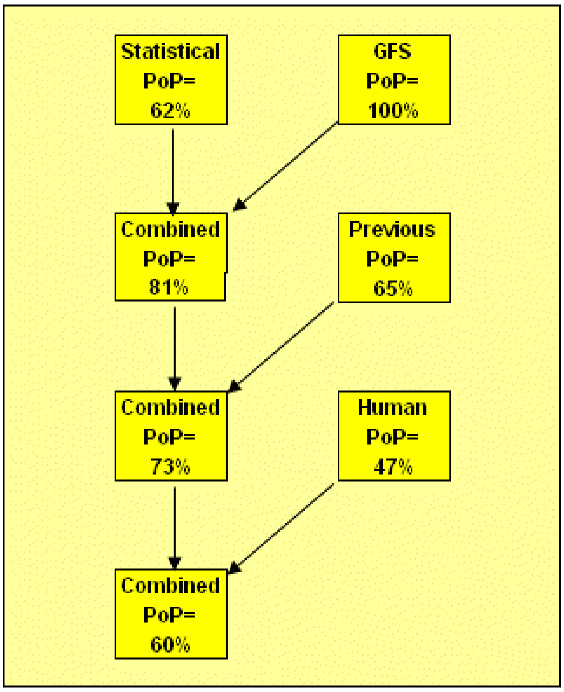

Figure 2. The process of integrating human and computer generated forecasts for *Probability of Precipitation* estimates:

- **Firstly, the estimate from a statistical model (of 62%) is averaged with the implied estimate from the NOAA Global Forecasting System (of 100%) to yield 81%;**
- **Secondly, this 81% outcome is then averaged with the previous estimate (generated 'yesterday') by the knowledge based system (of 65%) to yield 73%; and,**
- **Finally, this 73% is then averaged with the implied estimate from the human (official) forecast (of 47%) to yield 60%.**

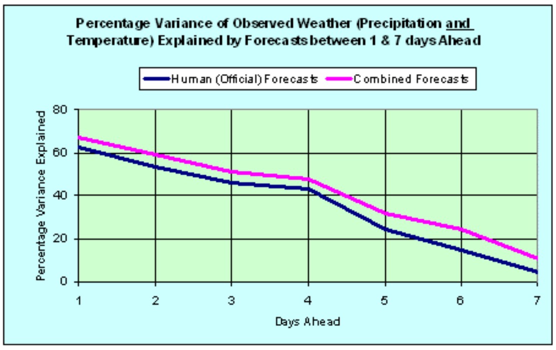

Figure 3. Overall percentage variance of the observed weather explained at different lead times.

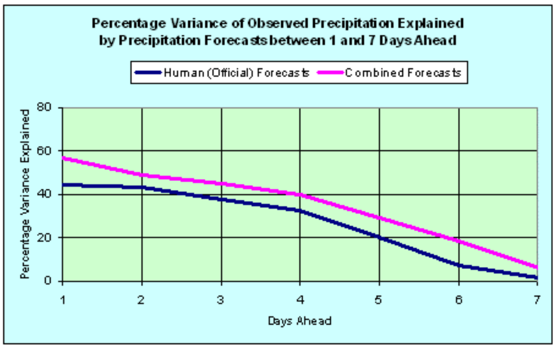

Figure 4. Overall percentage variance of the observed precipitation explained at different lead times.

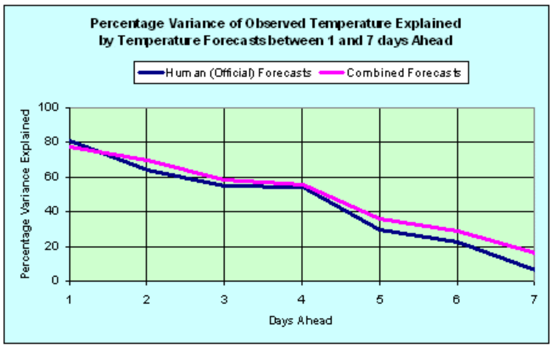

Figure 5. Overall percentage variance of the observed temperature explained at different lead times.

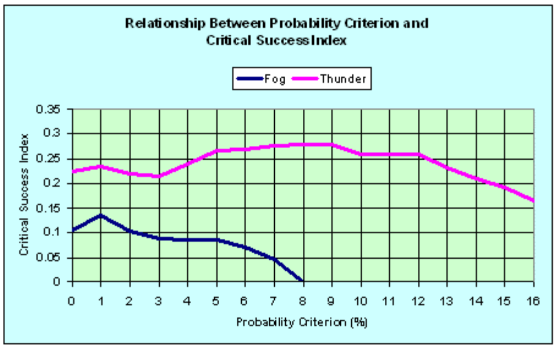

Figure 6a. Probability criterion under which there is a categorical reference to fog and thunder versus *Critical Success Index*.

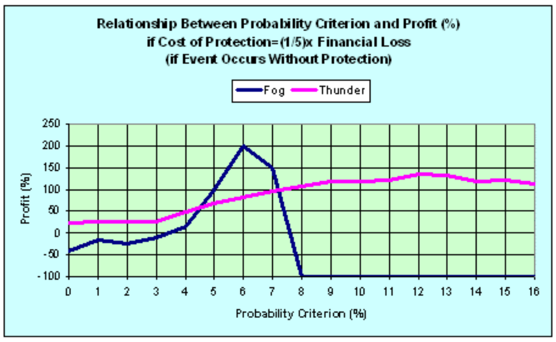

Figure 6b. Probability criterion under which there is a categorical reference to fog and thunder versus *Profit*.

Table 1 An extract of the probability (%) algorithm.

Day: / Precis:	1	2	3	4	5	6	7
Sunny	1%	1%	1%	1%	3%	6%	9%
Partly Cloudy	4%	6%	9%	11%	14%	16%	19%
Cloudy	19%	20%	21%	23%	25%	26%	28%
Becoming Fine	34%	33%	34%	35%	36%	36%	37%
Few Showers	49%	47%	47%	47%	46%	47%	47%
Drizzle Clearing	63%	61%	59%	58%	57%	57%	56%
Showers Clearing	78%	74%	72%	70%	68%	67%	65%
Showers	93%	88%	85%	82%	79%	77%	75%
Rain	99%	99%	97%	94%	90%	87%	84%
Heavy Rain	99%	99%	97%	94%	90%	87%	84%

Table 2 An extract of the amount (mm) algorithm.

Day: **Precis:**	1	2	3	4	5	6	7
Sunny	0	0	0	0	0	0	0
Partly Cloudy	0	0	0	0	0	0	0
Cloudy	0	0	0	0	0	0	0
Becoming Fine	0	0	0	0	0	0	0
Few showers	2	1	1	1	1	1	1
Drizzle Clearing	2	1	1	1	1	1	1
Showers Clearing	2	1	1	1	1	1	1
Showers	5	4	3	2	2	1	1
Rain	10	8	6	5	4	2	1
Heavy Rain	20	16	13	10	7	4	1

Table 3 Summary of Results

	Official	**System**
QPF (% Variance Explained)	18.2	25.6
Precipitation Probability / Sensible Weather (% Variance Explained)	29.3	40.3
Overall Precipitation (% Variance Explained)	**23.7**	**33.0**
Minimum Temperature (% Variance Explained)	41.6	45.8
Maximum Temp (% Variance Explained)	40.7	48.8
Overall Temperature (% Variance Explained)	**41.1**	**47.3**
Overall Temperature & Precipitation (% Variance Explained)	**32.4**	**40.1**