INCREASING FORECAST ACCURACY BY MECHANICALLY COMBINING HUMAN AND AUTOMATED PREDICTIONS USING A KNOWLEDGE BASED SYSTEM

Harvey Stern *

Bureau of Meteorology, Australia

ABSTRACT

There is an increasing interest in the question of what might be the appropriate future role for the human in the forecast process. It is asserted that computer-generated forecasts are unable (by themselves) to fully replicate the decision-making processes of human forecasters. Similarly, it is also asserted that human forecasters are unable (by themselves) to optimally integrate into the forecasting process, guidance from computer-generated predictions. However, there is the accepted mathematical concept that two or more inaccurate but independent predictions of the same future events may be combined to yield predictions that are, on the average, more accurate than either of them taken individually. Automated and human forecasts might be expected to "bring to the table" different knowledge sets, and this suggests the development of a weather forecasting system that mechanically combines human and computer-generated predictions.

This paper reports on the evaluation of a knowledge based system, modified in order to mechanically combine human and computer-generated predictions. The system's output is firstly evaluated over a "real-time" trial of 100 days duration. The trial reveals that forecasts generated by mechanically combining the predictions explain 7.7% additional variance of weather (rainfall amount, sensible weather, minimum temperature, and maximum temperature) over that explained by the human (official) forecasts. In the light of the results of the 100-day trial, a number of minor modifications are made to the system and the trial is then continued. After 365 Day-1 to Day-7 forecasts, that is, 2555 individual predictions, the average lift in percentage variance of weather explained is 7.9% over that explained by the current official forecasts.

With computer-generated forecasts unable to fully incorporate human forecasters' valuable domain and contextual knowledge, there should be a need for the human forecaster well into the future. That future role may be as an input to a system that mechanically combines human predictions with computer-generated forecasts.

1. INTRODUCTION

"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, 1990) (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

---

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

About the author: Dr Harvey Stern holds a Ph. D. from the University of Melbourne, for a dissertation entitled *Statistically based weather forecast guidance*. Dr Stern also holds a Graduate Diploma in Applied Finance and Investment from the Securities Institute of Australia. His 1992 paper, *The likelihood of climate change: a methodology to assess the risk and the appropriate defence*, which explored the application of options pricing theory to climate change, was the first paper to be published on what later became known as *weather derivatives*.

process. The resulting record keeping, feedback, and enhanced learning can improve forecast quality" (Sanders and Ritzman, 2001).

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, 2005b). 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 (Baars and Mass, 2005; Ryan, 2005).

Stern (1980a) outlines a possible approach to the determination of an optimal human-machine mix. He refers to Sanders (1973), who investigated the skill displayed by daily temperature and precipitation forecasts made in the Department of Meteorology at the Massachusetts Institute of Technology. Sanders found that "few if any individuals who made a substantial number of forecasts outperformed consensus on the average." Stern also refers to Thompson (1977), who, noting Sanders' work, suggested that "an objective and quantitative method" be used to reach a consensus (bearing in mind) … the incontrovertible fact that two or more inaccurate but independent predictions of the same future events may be combined in a very specific way to yield predictions that are, on the average, more accurate than either of them taken individually." Danard (1977) comments on Thompson's method, whilst Danard et al. (1968) discusses the subject of optimally combining independent estimates from a numerical analysis perspective.

Both Thompson (1977) and Danard (1977) discuss how one may optimally combine two forecasts if the assumption that they are independent is made. If this assumption is made, it is implied that ρ, the correlation coefficient between errors produced by the two forecasting methods, is equal to zero. Methods of forecasting a particular weather element are usually based on similar physical principles and therefore the sets of errors produced by the methods tend to be quite highly correlated. Indeed, Danard (1977) acknowledges "that ρ=0 is likely to be more valid for two measurements than for two predictions." Therefore, the validity of the approaches of Thompson (1977) and Danard (1977) is limited by their assumption that ρ=0.

With this in mind, Stern (1980a) suggests that the approach maybe applied under the assumption that ρ is not equal to zero by applying multiple linear regression to forecast verification data, in order to minimise forecasts errors. Stern (1980a&b) and Stern and Dahni (1981 & 1982) (refer also to Dahni et al. (1984)) subsequently demonstrated that forecasts would be improved were one to simply average predictions from different sources. In the context of the foregoing, this assumes that the predictions from the different sources are equally skilful. This is not an unreasonable assumption - to justify unequal weights there needs to be 'strong evidence to support unequal weighting' (Armstrong, 2001b). Indeed, a common method for combining individual forecasts is to calculate an equal weighted average of individual forecasts' (Stewart, 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. Nevertheless, 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.

In recent years, there has been considerable effort directed towards how to optimally combine forecasts from different sources (for example, refer to Aksu and Gunter (1992), Vislocky and Fritsch (1995), Brown and Murphy (1996), Ebert (2001), Etherton (2004), Ryan (2005), Woodcock and Engel (2005), and Stern (2006a)).

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'. The purpose of the present paper is to describe the development of a system that mechanically combines judgmental (human) forecasts (derived with the benefit of knowledge of all available computer generated forecast guidance) and computer generated forecasts guidance and to evaluate the accuracy of the new set of forecasts and to compare it with the accuracy achieved by the judgmental (human) forecasts.

2. A KNOWLEDGE-BASED SYSTEM

The present author has recently been involved in the development of a knowledge based weather forecasting system (Stern, 2002a, 2003, 2004a, 2004b, 2005a, 2005b, 2005c, 2005d, 2006a). The system, in its various guises, has variously been utilised to automatically generate worded weather forecasts for the general public, terminal aerodrome forecasts (TAFs) for aviation interests, and marine forecasts for the boating fraternity[1].

The knowledge based system generates these products by using a range of forecasting aids to interpret Numerical Weather Prediction (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, 1998, 1999, 2000, 2001), thunderstorm probability (Stern, 2004b), wind direction and speed, and swell (Dawkins, 2002).

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

The system was recently utilised to establish the limits of predictability, in a paper published in a previous edition of the Australian Meteorological Magazine (Stern, 2005b).

3. COMBINING FORECASTS

Stern (2005c) showed that human and automated forecasts are poorly correlated (the overall percentage variance of human forecasts explained[2] by the automated forecasts being only 45.9%). This poor correlation 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).

Sanders and Ritzman (2001), in their discussion of how best to combine forecasts, suggest that '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). It may be asserted that an automated system is unable (by itself) to fully replicate the decision-making processes of human forecasters and, similarly, human forecasters (by themselves) are unable to optimally integrate into the forecasting process automated forecasting guidance. The only way to preserve forecasters' valuable domain and contextual knowledge as an integral component of the forecasting process, while simultaneously incorporating automated forecasting guidance into that process, may therefore be to utilise

[1] The system's components have been extensively documented in a series of recent papers presented to American Meteorology Society (AMS) Conferences over the past few years (those without access to hard copy or CDs of AMS Conference Proceedings may download these papers from the AMS Website)

[2] The verification statistic, 'percentage variance explained' is most easily understood in the context of regression. For example, "… the smaller the variability of the residual values around the regression line relative to the overall variability, the better is our prediction … if there is no relationship between the X and Y variables, then the ratio of the residual variability of the Y variable to the original variance is equal to 1.0. If X and Y are perfectly related then there is no residual variance and the ratio of variance would be 0.0. In most cases, the ratio would fall somewhere between these extremes, that is, between 0.0 and 1.0. 1.0 minus this ratio is referred to as R-square … if we have an R-square of 0.4 then … we have explained 40% of the original variability, and are left with 60% residual variability. Ideally, we would like to explain most if not all of the original variability. The R-square value is an indicator of how well the model fits the data (e.g., an R-square close to 1.0 indicates that we have accounted for almost all of the variability with the variables specified in the model)" (StatSoft, Inc., 2006). Where the verification statistic, 'percentage variance explained' is quoted in the present paper in the context of evaluating 'overall' performance of the forecasting system across a number of different weather parameters, the statistic refers to the arithmetic average 'percentage variance explained' by predictions of these parameters.

a system that mechanically combines the automated and human predictions.

4. MODIFYING THE SYSTEM

Stern's (2005c) paper "Defining cognitive decision making processes in forecasting: a knowledge based system to generate weather graphics", delivered to the American Meteorology Society's *21st Conference on Weather Analysis and Forecasting and 17th Conference on Numerical Weather Prediction*, 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 human (official) forecasts explained 42.3% of the variance of the observed weather, whilst (by itself) the automated (knowledge based) system explained only a slightly greater 43.2% of the variance (that is, only an additional 0.9%) of the observed weather.

Post-analysis suggested that were one to adopt a strategy of combining human and computer-generated predictions one has the potential to lift the percentage variance explained by human predictions (the current official forecasts) of weather to a much greater 50.2%, that is, by 7.9% (Figure 1). This suggestion from the post-analysis was encouraging but needed to be validated by a fresh real-time trial conducted on a new set of independent data[3].

With a view to validating the assertion that mechanically combining automated and human predictions might lead to an increase in accuracy, the knowledge-based system was modified to mechanically combine human and computer-generated predictions so that it now took into account forecasters' valuable domain and contextual knowledge. Sanders and Ritzman (2001) refer to their 1992 study (Sanders and Ritzman, 1992), 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 process of integrating human and computer-generated forecasts is illustrated in Figure 2 as it applies to forecasts of Probability of Precipitation from which are derived forecasts of sensible weather. The inputs are averaged, which assumes that the predictions from the different sources are equally skilful. This is not an unreasonable assumption - to justify unequal weights there needs to be 'strong evidence to support unequal weighting' (Armstrong, 2001b). It cannot be emphasised too strongly that the modified system is not a system of forecast guidance, nor can it be used as a system of forecast guidance by the human forecasters, because the human (official) prediction is now, itself, an input into the modified system.

Note that the system's sensible weather predictions arises from an algorithm that interprets its generated probability of precipitation (POP) and synoptic type (Treloar and Stern, 1993; Stern and Parkyn, 1999) (refer to PANEL 1 below), and, conversely, the implied human (official) Probability of Precipitation arises from an algorithm that interprets the human (official) sensible weather predictions. This approach is similar to what Scott and Proton (2004) refer to as the creation of "anchor grids" from which to generate additional grids. What this entails is for forecasters within GFE/IFPS to create an anchor grid such as probability of precipitation (POP) and derive other forecast grids such as weather from the PoP grid. The forecaster thereby leverages a grid of values of one weather element to systematically extract grids of other weather elements via a set of algorithms.

PANEL 1

```
… else if (type=="2" && popfinal<70)
{iconmorningzero="Drizzle.";
iconafternoonzero="Cloudy.";}
else if (type=="2" && popfinal<80)
{iconmorningzero="Shower.";
iconafternoonzero="Cloudy.";}
else if (type=="2")
{iconmorningzero="Shower.";
iconafternoonzero="Shower.";}
//TYPE 3 WEAK CYCLONIC NNW
else if (type=="3" && popfinal<5)
{iconmorningzero="Sunny.";
iconafternoonzero="Sunny.";}
else if (type=="3" && popfinal<20)
{iconmorningzero="Sunny.";
iconafternoonzero="Partly Cloudy.";}…
```

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 précis in terms of Probability of Precipitation and Amount of Precipitation. This algorithm was

[3] In order to establish the validity of the assertions made in this paper, it was important to test the system "real-time" on a new set of data. As Sharov (2006) states: "…models may work well with the data to which they were fit, but show no fit to other data sets … To solve this problem, the concept of validation was developed. Model Validation is testing the model on another independent data set."

derived from Stern's (1999) proposed interpretation and a verification of the official précis 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 forecast of sensible weather arise largely from the system's generated Probability of Precipitation, and, conversely, the human (official) Probability of Precipitation, arises from an algorithm that interprets the (official) worded précis, any verification of the Probability of Precipitation may also be regarded as representing a verification of forecast sensible weather.

The system generates forecasts for 56 localities in Central Victoria, and therefore potentially caters for some 4 million people (or 20% of the Australian population). Figure 3 presents an illustration of its output by depicting the 3 June 2006 forecast for Mt St Leonard, which is in hilly country about 80 km east of Melbourne (Map 1).

5. THE REAL-TIME TRIAL

Forecast verification data from a new 100-day real-time trial of the system so modified, conducted from 20 August to 27 November 2005, on a new set of independent data, was then analysed. The analysis included verification of forecasts of a range of weather elements (rainfall, temperature, wind, thunder and fog) and confirmed the assertion that mechanically combining automated and human predictions might lead to an increase in accuracy, with the percentage variance explained by human predictions (the official forecasts) being lifted by 7.7% over that explained by the current official forecasts.

The lift in accuracy arose because in most circumstances, the combining strategy left the system's forecasts almost identical to the human (official) forecasts (the percentage variance of the official forecasts explained by the combined forecasts was 77.2% - a considerable increase over the 45.9% achieved previously), whilst in those few circumstances when the combining strategy substantially changed the human (official) forecasts, the system's forecasts usually represented an improvement.

In the light of the results of the 100-day trial, a number of minor modifications were made to the system. For example, from 27 December 2005, Day-1 wind forecasts were generated by a combining procedure – previously, they had been solely computer-generated. The real-time trial was then continued. After 365 Day-1 to Day-7 Melbourne forecasts from 20 August 2005 to 19 August 2006, inclusive, that is, 2555 individual predictions, the lift in percentage variance (of rainfall amount, sensible weather, minimum temperature, and maximum temperature) explained was to 41.3% from 33.4%, that is, an increase of 7.9% over that explained by the current official forecasts, the addition of more independent data further affirming the result suggested by the earlier work. The percentage variance of the official forecasts explained by the combined forecasts during the year was 76.8%.

There was an overall lift shown in the accuracy of forecasts (with the performances of all elements taken together) for each lead-time (Figure 4). There was also an overall lift shown in the accuracy of forecasts (with the performances at all lead times taken together) of each weather element (Table 3). The significance of that lift in accuracy is illustrated, for example, in Figure 5 (for Day-1 to Day-7 maximum temperature forecasts), which places the 0.33 deg C decrease in RMS error achieved by the system in the context of the performance of human (official) forecasts of that element.

A feature of the verification data presented is its comprehensive nature – covering not just forecasts of rainfall and temperature, but also sensible weather, fog, thunder and wind, and the 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.

With regard to the accuracy of forecasts of rare weather elements, the combining process was shown to lift the Critical Success Index (Wilks, 1995) for predictions of fog, from 15.5% to 17.8%, and for predictions of thunder, from 17.9% to 21.6%.

With regard to the accuracy of forecasts of wind, the percentage variance of wind speed explained by the 9am and 3pm Day-1 forecasts from 27 December 2005, when combined wind forecasts were first generated, was lifted from 47.5% to 54.3%, and the percentage correct forecasts of wind direction was lifted from 68.3% to 71.2%.

Although a lift did not occur in every single instance when the verification data was analysed with all lead times taken separately and all weather elements taken separately, a lift occurred in most instances. One of the very few exceptions, the case of Day-1 forecasts of fog, where the Critical Success Index for the combined forecasts (27.3%) was substantially below that of the official forecasts (35.3%), is worthy of comment. The inability (of the combining process) to improve on the Day-1 official forecasts of fog may very well be a consequence of the effort that the forecasting personnel of the Victorian Regional Office (and others) have invested over the years into short term fog and low cloud forecasting at Melbourne Airport (Goodhead, 1978; Keith, 1978; Stern and Parkyn, 1998, 1999, 2000, 2001; Newham, 2004). This effort may have resulted in such a high level of pre-existing human forecast skill at short-term predicting of fog, that mechanically combining human fog forecasts with automated fog forecasts actually caused a decline in accuracy.

6. PLACING A VALUE ON THE FORECASTS

What is particularly interesting about the verification data is that the combined forecasts are more consistent than the official forecasts. In a 1992 paper presented to the 5th International Meeting on Statistical Climatology (Stern, 1992; refer also to Stern, 2005e; Stern, 2006b), the author introduced a methodology for calculating the cost of protecting against the onset of global warming. The paper, "The likelihood of climate change: A methodology to assess the risk and the appropriate defence", was presented to the meeting held in Toronto, Canada, under the auspices of the American Meteorology Society (AMS). In this first application of what later was to become known as 'weather derivatives' (Stern, 2001a,b,c,&d; Stern, 2002b&c; Dawkins & Stern, 2003&2004; Stern and Dawkins, 2004), the methodology used options pricing theory from the financial markets to evaluate hedging and speculative instruments that may be applied to climate fluctuations. In a related paper, Stern and Dawkins (2003) applied forecast verification data in a weather risk management context to price a financial instrument to guarantee the accuracy of a short-term weather forecast. What now follows is an application of options pricing theory where one uses the theory in a weather risk management context.

The theory shows that the more consistent forecasts are from one day to the next, between Day-7 (when they are first issued) and Day-1 (the final issue), the cheaper are the prices of option contracts that one may wish to purchase to protect against the eventuality that the forecasts might be incorrect. The implication from this is that, the more consistent forecasts are from one day to the next, the more valuable are the forecasts (Appendix A.1).

The American Marketing Association (2006) notes that "a "competitive advantage" exists when there is a match between the distinctive competences of a firm and the factors critical for success within the industry that permits the firm to outperform its competitors. Advantages can be gained by having the lowest delivered costs and/or differentiation in terms of providing superior or unique performance on attributes that are important to customers."

From the foregoing, it may be said that the value of a series of weather forecasts with a low volatility, that is, a series of forecasts that display a high level of consistency from one day to the next, is greater than the value of a series of forecasts with a high volatility. This is because the cost of protecting against the possibility of such weather forecasts being incorrect by adopting a strategy of purchasing weather derivatives is lower. This means that sellers of weather derivatives, who utilise low volatility forecasts to price their call and put options, are provided with a competitive advantage over sellers of weather derivatives who utilise high volatility forecasts. This arises because sellers of weather derivatives who utilise low volatility forecasts being able to charge lower, and, therefore, more competitive, prices to purchasers of weather derivatives who wish to use those weather derivatives to protect against the possibility of the weather forecasts being incorrect.

Hence, the data presented in Table 3, in showing the combined forecasts are more *consistent* than the official forecasts, are also showing that the combined forecasts are more *valuable* than the official forecasts with:

- o The consistency (RMS inter-diurnal change from Day-7 to Day6 to … to Day-1 forecast) associated with combined forecasts of rainfall amount being to 0.44mm$^{0.5}$ (this RMS inter-diurnal change being well below the 0.65mm$^{0.5}$ associated with the official forecasts).
- o The consistency associated with combined forecasts of sensible weather being 11.6% (this RMS inter-diurnal change being well below the 18.9% associated with the official forecasts);
- o The consistency associated with combined forecasts of minimum temperature being 1.17ºC (this RMS inter-diurnal change being well below the 1.36ºC associated with the official forecasts); and,
- o The consistency associated with combined forecasts of maximum temperature being 1.36ºC (this RMS inter-diurnal change being well below the 1.86ºC associated with the official forecasts).

Furthermore, that the combined forecasts are more *accurate* than individual currently available predictions taken separately, also provides the small to medium sized companies involved in weather broadcasting with a potential competitive advantage (O'Donnell *et al*., 2002) over their peers should they choose to adopt a strategy of mechanically combining these existing predictions.

And, there is a multiplicity of existing predictions to choose from. For example, one may choose, as in the case of the present work, to combine computer generated forecasts from a statistical interpretation of the output of the NOAA Global Forecasting System (GFS) (NOAA, 2006), with human forecasts for Melbourne from the Australian Bureau of Meteorology (Bureau of Meteorology, 2006). Alternatively, in order to provide a purely Australian context, one may replace the NOAA output as the computer generated component with the Australian Bureau of Meteorology's *Optimal Consensus Forecasts* (Australian Weather News, 2006).

7. CONCLUDING REMARKS

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). Computer-generated forecasts are unable (by themselves) to fully replicate the decision-making processes of human forecasters. Similarly, human forecasters (by themselves) are unable to optimally integrate into the forecasting process, guidance from computer-generated predictions. However, the work presented here demonstrates that mechanically combining human and computer-generated predictions (in order to "bring to the table" the two different knowledge sets) results in a set of forecasts that are more accurate than those currently issued officially.

With automated systems unable (by themselves) to fully incorporate forecasters' valuable domain and contextual knowledge, there should be a need for the human forecaster well into the future. This role may be to provide input to a system that mechanically combines human predictions with computer generated forecasts.

8. PROPOSED FUTURE WORK

Stern's (2005b) *Australian Meteorological Magazine* paper, "Establishing the limits of predictability at Melbourne, Australia, using a knowledge based forecasting system and NOAA's long-range NWP model", suggested that, for the first time, we have emerging evidence that there may now be some forecast skill out to Lorenz's suggested 15-day limit (to day-to-day predictability of the atmosphere), particularly for temperature. With this background, for future work, it is proposed to extend the system to Day-10, utilising forecasts based upon "climatology" for generating Day-8 to Day-10 predictions, in lieu of the human (official) forecasts (because these human forecasts are not currently prepared for those lead times).

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, to Mark Williams, for discussions on preserving forecast consistency, to Tom Haydon, for assistance with data extraction, and to Blair Trewin, Graham Mills, and Terry Hart, and to my other Bureau of Meteorology colleagues, for their helpful discussions and feedback.

9. REFERENCES

Aksu, C, and S. I Gunter, 1992: An empirical analysis of the accuracy of SA, OLS, ERLS and NRLS combination forecasts. *International Journal of Forecasting*, 8, 27–43.

American Marketing Association, 2006: Dictionary of marketing terms. The website http://www.marketingpower.com was accessed on 15 July 2006.

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).

Australian Weather News, 2006: The Australian Bureau of Meteorology's automated *Optimal Consensus Forecasts* are available via the website http://www.australianweathernews.com/forecast_OCF.htm, which was accessed on 15 July 2006.

Baars, J. and Mass, C. F., 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*. Amer. Meteor. Soc., Washington, DC, 1-5 Aug., 2005.

Black, F. 1976: The Pricing of Commodity Contracts, *Journal of Financial Economics*, 3, 167-79.

Black, F. and Scholes, M. 1973. The pricing of options and corporate liabilities, *J. Political Economy*, 81, 637–54.

Brown, B. G, and A. H Murphy, 1996: Improving forecasting performance by combining forecasts: The example of road-surface temperature forecasts. *Meteor. Appl*., 3, 257–265.

Bureau of Meteorology, 2006: The Australian Bureau of Meteorology's human (official) forecasts for Melbourne are available via the website http://www.bom.gov.au/products/IDV10450.shtml, which was accessed on 15 July 2006.

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

Danard, M., 1977: Comments on "How to improve accuracy by combining independent forecasts", *Mon. Weath. Rev*., 105, 1198-1199.

Danard, M., Holl, M.M., and Clarke, J.R., 1968: Fields by correlation assembly – a numerical analysis technique, *Mon. Weath. Rev*., 96, 141-149.

Dahni, R.R., de la Lande, J. and Stern, H., 1984: Testing of an operational statistical forecast guidance system, *Aust. Meteor. Magazine*, 32, 105-106.

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

Dawkins, S. S., and Stern, H., 2003: Cashing in on the weather: how agriculture might protect against adverse seasons using weather derivatives. *Australia New Zealand Climate Forum: Climate Serving Agriculture, Palmerston North, New Zealand, 19-21 Mar., 2003.*

Dawkins, S. S., and Stern, H., 2004: Managing weather risk during major sporting events: the use of weather derivatives. Presented at *2nd International Workshop on Climate, Tourism and Recreation,* Orthodox Academy of Crete, Kolimbari, Crete, Greece, 8-11 Jun., 2004. Published in Berichte des Meteorologischen Institutes der Universitat Freiburg Nr. 12. A. Matzarakis, C. R. de Freitas and D. Scott (Eds.) Advances in Tourism Climatology. Freiburg, November 2004. pp 166-173.

Ebert, E. E, 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation, *Mon. Weath. Rev*., 129, 2461–2480.

Etherton, B. J., 2004: Model consensus and ensemble weighting for spot forecasts. *17th Conference on Probability and Statistics in the Atmospheric Sciences*, Am. Meteor. Soc. , Seattle, Washington, USA 11-15 Jan., 2004.

Goodhead, H. M., 1978: An objective method for predicting fog and mist at Melbourne Airport. *Airmet Conference*, 9-10 Feb., 1978, Canberra Meteor. Soc., Roy. Meteor. Soc. (Aust. Branch), Canberra, Australia.

Keith, R., 1978: Formation of low cloud at Melbourne Airport in a lee trough. *Airmet Conference*, 9-10 Feb., 1978, Canberra Meteor. Soc., Roy. Meteor. Soc. (Aust. Branch), Canberra, Australia, 39-44.

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).

Mathews, D. P. and Diamantopoulos, A., 1990: Judgmental revision of sales forecasts: effectiveness of forecast selection, *Journal of Forecasting*, 9, 407-415 (refer to Armstrong, 2001a, 411).

Merton, Robert C., 1973: Theory of rational option pricing, *Bell Journal of Economics and Management Science*, 4 (1), 141-183.

Newham, P., 2004: Fog forecasting for Melbourne Airport. *3rd Conference on Fog, Fog Collection, and Dew*, Cape Town, South Africa, 11-15 October, 2001.

Numa Financial Systems, 2006: The *Numa Option Calculator* is available via the website http://www.numa.com, which was accessed on 15 July 2006.

O'Donnell, A., Gilmore, A., Carson, D. and Cummins, D., 2002: Competitive advantage in small to medium-sized enterprises, *Journal of Strategic Marketing*, 10, 205-223.

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

Risk Glossary, 2006. The website http://www.riskglossary.com was accessed on 2 July 2006.

Treloar, A. B. A. and Stern, H., 1993: A climatology and synoptic classification of Victorian severe thunderstorms, *4th International Conference on Southern Hemisphere Meteorology and Oceanography*, March 29 to April 2, 1993, Hobart, Australia, American Meteorological Society.

Sanders, F., 1973: Skill in forecasting daily temperature and precipitation: Some experimental results. *Bull. Am. Meteor. Soc*., 54, 1171-1179.

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, *Bull. Am. Meteor. Soc*., 58, 1036-1044.

Scott, A. C. and Proton, V. J., 2004: The necessity of a coherent and consistent forecast methodology in the IFPS/GFE Era, *20th Conference on Interactive Information and Processing Systems*, Seattle, Washington, USA 11-15 Jan., 2004.

Sharov, A., 2006: On-line lectures, *Department of Entomology, Virginia Tech, Blacksburg, Virginia, USA*. The *Quantitative Population Ecology* website http://www.ento.vt.edu/~sharov/PopEcol/lec4/autocor.html was accessed on 30 September 2006.

StatSoft, Inc., 2006: Residual Variance and R-square. This account is available via the *StatSoft, Inc.* website http://www.statsoft.com/textbook/stmulreg.html#cresidual, which was accessed on 30 September 2006.

Stern, H., 1980a: 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., 1980b: A system for automated forecasting guidance. *Aust. Meteor. Magazine*, 28, 141-154.

Stern, H. 1992: The likelihood of climate change: a methodology to assess the risk and the appropriate defence, *5th International Meeting on Statistical Climatology*, June 22-26, 1992, Toronto, Canada, Amer. Meteor. Soc.

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. Mag.*, 48, 159-167.

Stern, H., 2001a: The application of weather derivatives to mitigate the financial risk of climate variability and extreme weather events. *Aust. Meteor. Mag*., 50.

Stern, H. 2001b: The application of weather derivatives to mitigate the financial risk of climate variability, extreme precipitation events and extreme temperature events. *Symposium on climate variations, the oceans and societal impacts, 81st Annual Meeting, Amer. Meteor. Soc., Albuquerque, New Mexico, 14-19 Jan., 2001.*

Stern, H., 2001c: What price the weather? Outlook fine for temperature derivatives. *JASSA (Journal of the Securities Institute of Australia)*, 2-6.

Stern, H., 2001d: Weather derivatives. *Australia Risk*, August 2001.

Stern, H., 2002a: 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., 2002b: Using weather derivatives to mitigate financial risk. *Derivatives Week*, July 1, 2002.

Stern, H., 2002c: Wheat Hedge. *Australia Risk*, August 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: Establishing the limits of predictability at Melbourne, Australia, using a knowledge based forecasting system and NOAA's long-range NWP model. *Aust. Meteor. Mag*., 54:203-211.

Stern, H., 2005c: 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*. Amer. Meteor. Soc., Washington, DC, 1-5 Aug., 2005.

Stern, H., 2005d: Generating quantitative precipitation forecasts using a knowledge based system, *17th BMRC Modelling Workshop*, Melbourne, Australia, 3-6 Oct., 2005.

Stern, H., 2005e: Evaluating the cost of protecting against global climate change: options pricing theory and weather derivatives, *16th Conference on Climate Variability and Change*, San Diego, California, USA 9-13 Jan., 2005.

Stern, H., 2006a: Combining human and computer generated weather forecasts using a knowledge based system, *22nd Conference on Interactive Information and Processing Systems*, Atlanta, Georgia, USA 27 Jan. - 3 Feb., 2006.

Stern, H., 2006b: The application of financial market mathematics to translating climate forecasts into decision making, *3rd International Conference on Climate Impacts Assessments (TICCIA)*, Cairns, Australia, 24-27 Jul., 2006.

Stern, H. and Dahni, R, 1981: Further testing of Stern's (1980) system for automated forecasting guidance. *Aust. Meteor. Mag*., 29, 69-70.

Stern, H. and Dahni, R., 1982: An automated system for forecast guidance. Abstracts, *4th Conf. on Science Technology*, Melbourne, August 1982 (ANZAAS-AIST).

Stern, H. and Dawkins, S. S., 2003: Pricing a financial instrument to guarantee the accuracy of a weather forecast. Third Conference on Artificial Intelligence Applications to Environmental Science, Long Beach, California, USA 9-13 Feb., 2003; *Refer also (for a report on the presentation) to* At the Annual Meeting ... Pricing forecast guarantees. *Bull. Am. Meteor. Soc.,* 84:325-326.

Stern, H., and Dawkins, S. S., 2004: Weather derivatives as a vehicle to realise the skill of seasonal forecasts. *15th Conference on Global Change and Climate Variations & 14th Conference on Applied Climatology,* Seattle, Washington, USA 11-15 Jan., 2004.

Stern, H. and Parkyn, K., 1998: Synoptic climatology of fog at Melbourne Airport. Abstracts, *ANZ Climate Forum*, Perth, Australia, 1998, 51.

Stern, H. and Parkyn, K., 1999: Predicting the likelihood of fog at Melbourne Airport, *8th Conference on Aviation, Range and Aerospace Meteorology*, Amer. Meteor. Soc., Dallas, Texas, 10-15 Jan., 1999.

Stern, H. and Parkyn, K., 2000: Low cloud at Melbourne Airport: A synoptic climatology leading to a forecasting technique, *AMOS 2000, incorporating 7th National AMOS conference & 5th Australasian conference on the Physics of Remote Sensing of Atmosphere and Ocean*, The University of Melbourne, Melbourne, 7-9 February 2000.

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; Amer. Meteor. Soc*., 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).

Thompson, P. D., 1977: How to improve accuracy by combining independent forecasts. *Mon. Weath. Rev.*, 105, 228-229.

Treloar, A. B. A. and Stern, H., 1993: A climatology and synoptic classification of Victorian severe thunderstorms. *4th International Conference on Southern Hemisphere Meteorology and Oceanography*, Mar. 29 - Apr. 2, 1993, Hobart, Australia, *Am. Meteor. Soc*.

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*.

Woodcock F., and Engel, C., 2005: Operational consensus forecasts. *Weather and Forecasting*, 20, 101-111.

Vislocky, R. L., and J. M. Fritsch, 1995: Improved model output statistics forecast through model consensus. *Bull. Am. Meteor. Soc*., 76, 1157-1164.

Map1. The system generates forecasts for 56 localities in Central Victoria, including all of the localities shown on the map, except for Fawkner Beacon and St Kilda Harbour Royal Melbourne Yacht Squadron (RMYS). Mt St Leonard, a sample forecast for which is depicted in Figure 3, is shown located in hilly country, some 80 km to the east of Melbourne.

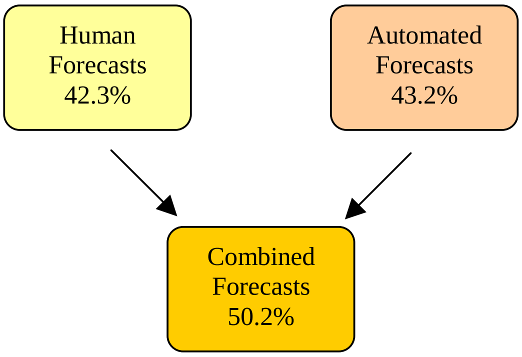

Figure 1. Stern (2005c) showed that the process of combining human (official) and automated forecasts had the potential to yield a set of predictions that is far more accurate than current official forecasts. The human (official) forecasts explained 42.3% of the variance of the observed weather (rainfall amount, sensible weather, minimum temperature, and maximum temperature), whilst (by itself) the automated (knowledge based) system explained only a slightly greater 43.2% of the variance of the observed weather. However, adopting a combining strategy has the potential to lift the overall percentage variance explained to a much greater 50.2%.

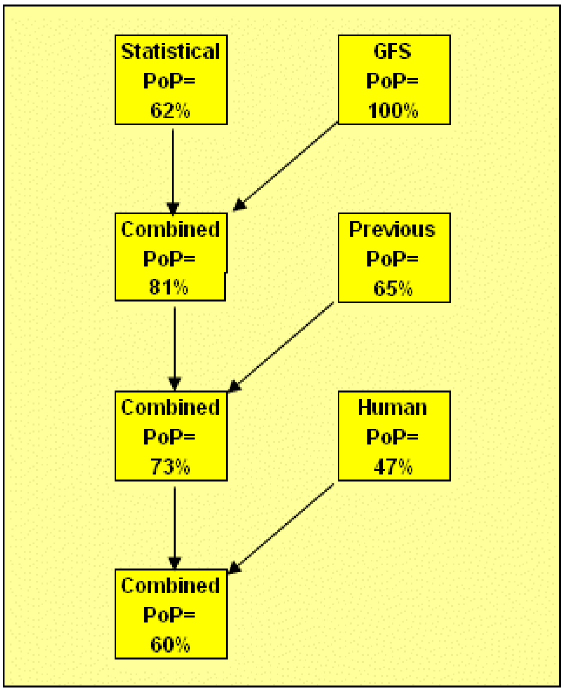

Figure 2 The process of combining human (official) and computer generated (statistical) forecasts of Probability of Precipitation:

- Firstly, the estimate from a statistical model (of 62%) is averaged with the implied estimate from the NOAA Global Forecasting System (GFS) (of 100%) to yield 81% (note that the "implied estimate" from the NOAA GFS is taken to be 100%, where a rainfall amount of at least 0.2 mm is indicated by the NOAA System, 50%, where a rainfall amount of 0.1 mm is indicated, and 0%, where no rainfall is indicated);
- Secondly, this 81% outcome is then averaged with the previous estimate (generated 'yesterday') by the knowledge based system (of 65%) to yield 73% (the benefit of this step lies in it preserving some "memory" of the previous forecast, and hence results in a more consistent series of forecasts between Day-7 and Day-1; note that, for the "previous forecast" for Day-7, the "previous forecast" is taken to be the climatological normal); and,
- Finally, this 73% is then averaged with the implied estimate from the human (official) forecast (of 47%) to yield 60%.

Experimental Mt St Leonard Long Range Weather Forecast

Based Upon Statistical Interpretation of NOAA Data

Combined With Human Forecasts

(http://www.arl.noaa.gov/ready.html)

NOAA Data At: 10am, 3-Jun-2006

Minimum temperatures are for the night/early morning,
maximum temperatures are for the daytime,
whilst precipitation amounts and probabilities
are for the 24 hours from midnight.

Day & Date	Morning	Afternoon	Min Temp (deg C)	Max Temp (deg C)	Precip Amount (mm)	Precip Prob (%)	9am Wind/ 3pm Wind Melb Apt (km/hr)
Sun-4-6-2006	Shower.	Shower.	5	9	2.2	53	SW 8 SSE 15
Mon-5-6-2006	Cloudy.	Sunny.	3	10	0	34	N 8 S 8
Tue-6-6-2006	Fog.	Sunny.	3	10	0	12	N 8 S 8
Wed-7-6-2006	Fog.	Sunny.	2	11	0	14	N 8 SSE 8
Thu-8-6-2006	Sunny.	Partly Cloudy.	3	11	0	19	N 15 N 15
Fri-9-6-2006	Partly Cloudy.	Cloudy.	4	10	0	24	N 25 N 25
Sat-10-6-2006	Cloudy.	Sunny.	4	11	0	29	N 8 N 8

The Weather Icons

Acknowledgement: Bureau of Meteorology & World Meteorological Organisation

Sunny	Dry	Haze	Mist	Fog	Partly Cloudy	Cloudy	Windy	Dust

Possible Shower	Drizzle	Shower	Snow	Rain	Thunder	Cyclone

Figure 3 An illustration of the output of the system - the 3 June 2006 forecast for Mt St Leonard, which is in hilly country some 80 km east of Melbourne (refer to Map 1).

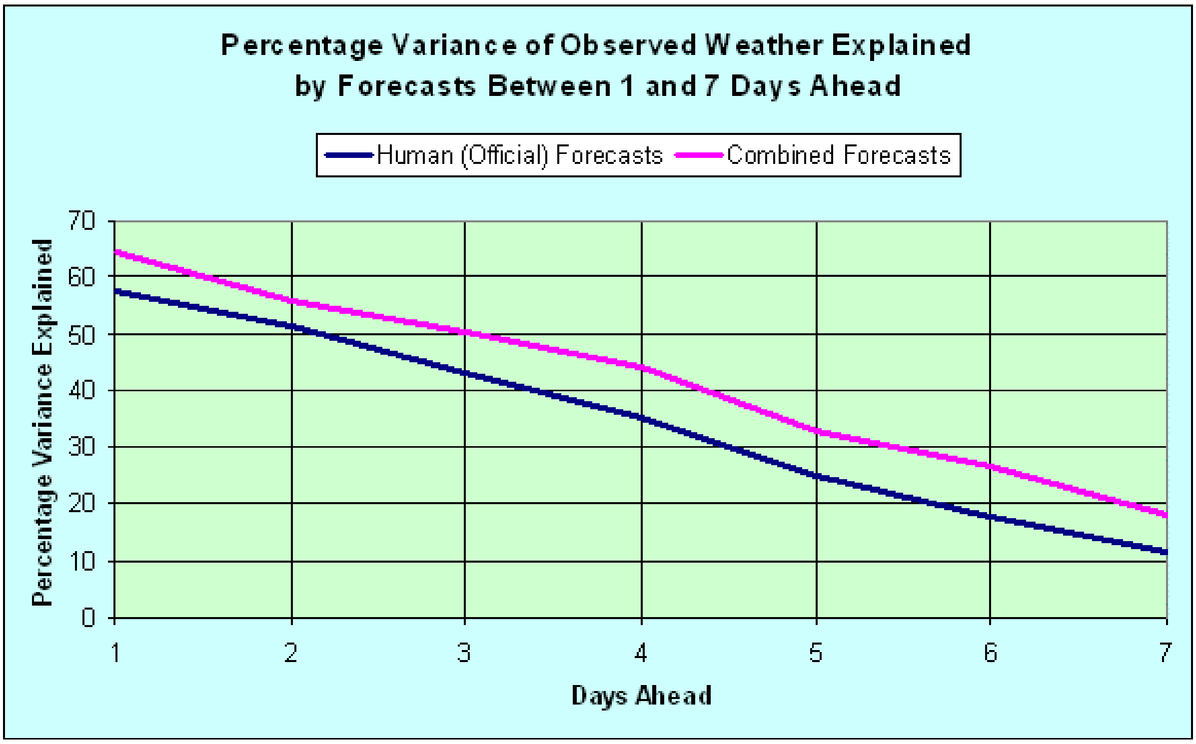

Figure 4 Overall percentage variance of observed weather explained for each lead time (the weather elements rainfall amount, sensible weather, minimum temperature, and maximum temperature, taken together) explained by forecasts between 1 and 7 days ahead. The graph shows that the combined forecasts displayed a lift in the accuracy of forecasts for each lead time.

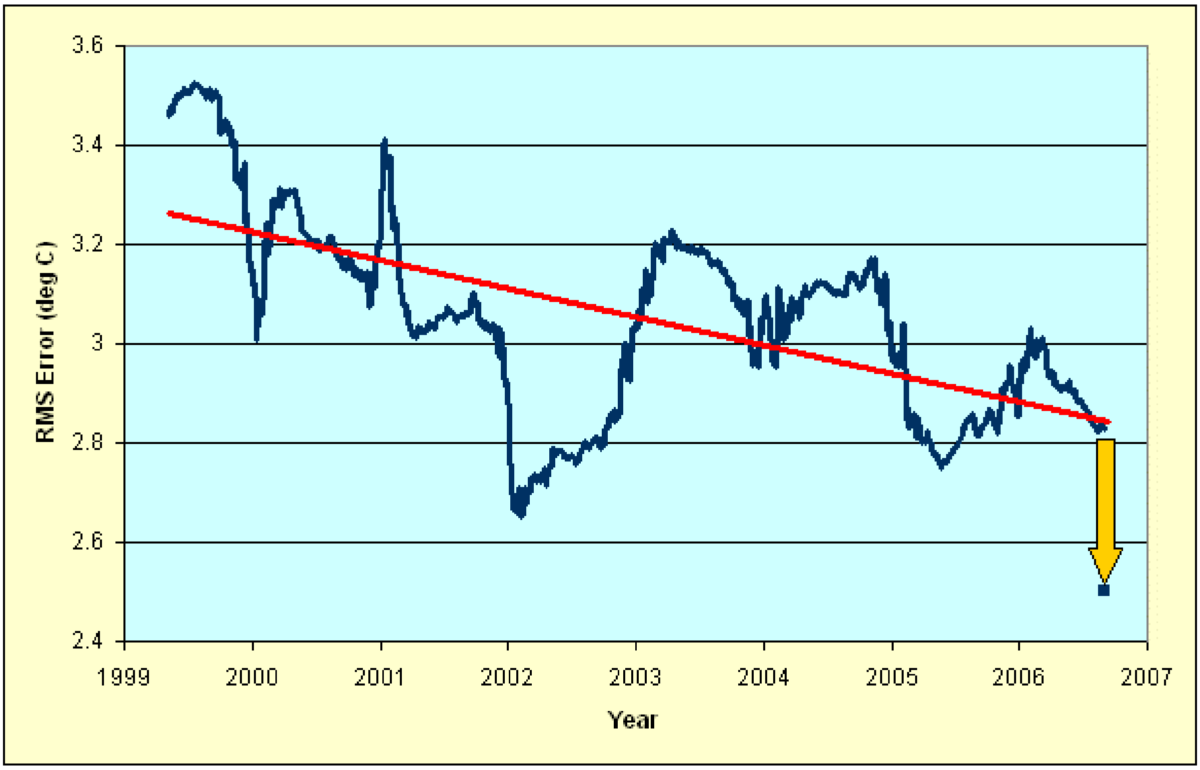

Figure 5 Placing the 0.33 deg C decrease in RMS error of Day-1 to Day-7 maximum temperature forecasts achieved by the system (orange arrow) in the context of the historical performance of human (official) forecasts of that element (*blue line:* RMS Error during the preceding 365 days; *red line:* linear trend).

Table 1 An extract of the probability (%) algorithm. To illustrate the application of the probability algorithm by example, note that a précis of "showers" by the human (official) forecast is, at Day-1, interpreted as a Probability of Precipitation (PoP) of 93%. That interpretation reduces to 82% at Day-4, and reduces further to 75% at Day-7, due to increasing uncertainty as the lead time lengthens.

Day: Précis:	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. To illustrate the application of the amount algorithm by example, note that a précis of "rain" by the human (official) forecast is, at Day-1, interpreted as an Amount of Precipitation of 10 mm. That interpretation reduces to 5 mm at Day-4, and reduces further to 1 mm at Day-7, due to increasing uncertainty as the lead time lengthens.

Day: Précis:	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 Details of the overall lift in forecast accuracy achieved by combining forecasts, demonstrating that there is a lift in the overall performance of the predictions for each individual weather element (lead times Day-1, Day-2, … Day-7 taken together).

Weather Element	Verification Parameter	Official Forecasts	Combined Forecasts	Comment
All Elements	% Variance Explained	33.4	41.3	Combined forecasts are, overall, more accurate
Rain[4] *or* No Rain	% Correct	70.1	76.8	Combining yields more correct forecasts of rain occurrence
√(Rainfall Amount)	% Variance Explained	18.4	23.5	Combined rainfall forecasts are more accurate
...	RMS Error ($mm^{0.5}$)	1.05	0.97	Combined rainfall forecasts are more accurate
...	Forecast Consistency[5] ($mm^{0.5}$)	0.65	0.44	Combined rainfall forecasts are more consistent
Sensible Weather[6]	% Variance Explained	23.7	34.2	Combined forecasts of sensible weather are more accurate
...	Forecast Consistency (%)	18.9	11.6	Combined forecasts of sensible weather are more consistent
Min Temp	% Variance Explained	41.5	47.7	Combined minimum temperature forecasts are more accurate
...	RMS Error (ºC)	2.39	2.27	Combined minimum temperature forecasts are more accurate
...	Forecast Consistency (ºC)	1.36	1.17	Combined minimum temperature forecasts are more consistent
Max Temp[7]	% Variance Explained	50.0	59.7	Combined maximum temperature forecasts are more accurate
...	RMS Error (ºC)	2.82	2.49	Combined maximum temperature forecasts are more accurate
...	Forecast Consistency (ºC)	1.86	1.36	Combined maximum temperature forecasts are more consistent
Thunder[8]	Critical Success Index (%)	17.9	21.6	Combined thunderstorm forecasts are (overall) superior …
…	Probability of Detection (%)	20.6	34.1	With more "hits" …
…	False Alarm Ratio (%)	42.9	62.9	But, at a price of more "false alarms"
Fog[9]	Critical Success Index (%)	15.5	17.8	Combined fog forecasts are (overall) superior …
…	Probability of Detection (%)	19.9	27.3	With more "hits" …
…	False Alarm Ratio (%)	58.9	66.1	But, at a price of more "false alarms"
Wind Speed[10]	% Variance Explained	47.5	54.3	Combined forecasts of wind speed are more accurate
Wind Direction	% Correct Within Half-Octant	68.3	71.2	Combined forecasts of wind direction are more accurate

[4]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).

[5]RMS inter-diurnal change from Day-7 to Day6 to … to Day-1 forecast.

[6]Implied by probability of precipitation estimates.

[7]Forecasts were also prepared for a number of other Central District localities. Verification of the Day-1 maximum temperature component of these forecasts, namely, for those ten places for which official forecasts are issued (note that, for six of these ten places, only Day-1 temperature forecasts are issued officially) reveals that, expressed as an expected departure from Melbourne's maximum temperature, the mean absolute error of the system's forecasts was 0.971ºC, compared with 1.099ºC for the official forecasts.

[8]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 (1980a).

[9]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.

[10]Wind forecast verification data are only provided here for the 236 sets of twice-daily (9am and 3pm) Day-1 forecasts from 27 December 2005, when combined wind forecasts were first generated.

APPENDIX

A.1 Placing a value on the forecasts.

Options pricing theory shows that the more consistent forecasts are from one day to the next, between Day-7 (when they are first issued) and Day-1 (the final issue), the cheaper are the prices of option contracts that one may wish to purchase to protect against the eventuality that the forecasts might be incorrect. The implication from this is that, the more consistent forecasts are from one day to the next, the more valuable are the forecasts.

A challenge in pricing options on commodities is non-randomness in the evolution of many commodity prices. For example, the spot price of an agricultural product will generally rise prior to a harvest and fall following the harvest. Natural gas tends to be more expensive during winter months than summer months. Because of such non-randomness, many spot commodity prices cannot be modelled with a geometric Brownian motion, and the Black-Scholes (1973) or Merton (1973) models for options on stocks do not apply. In 1976, Fischer Black published a paper (Black, 1976) addressing this problem and the Risk Glossary (2006) summarises the result of his work thus:

His (Black's) solution was to model forward prices as opposed to spot prices. Forward prices do not exhibit the same non-randomness of spot prices. Consider a forward price for delivery shortly after a harvest of an agricultural product. Prior to the harvest, the spot price may be high, reflecting depleted supplies of the product, but the forward price will not be high. Because it is for delivery after the harvest, it will be low in anticipation of a drop in prices following the harvest. While it is not reasonable to model the spot price with a Brownian motion, it may be reasonable to model the forward price with one. The assumption that the spot price follows a log-normal process is replaced by the assumption that the forward price follows such a process. From there the derivation is identical to the Black-Scholes formula for evaluating stock options and so the final formula is the same except that the spot price is replaced by the forward - the forward price represents the expected future value discounted at the risk free rate. Black's (1976) option pricing formula reflects this solution, modelling a forward price as an underlier in place of a spot price.

Pricing options on forecast weather elements, which may be employed in a weather risk management context, also requires one to address non-randomness in the evolution of many forecasts of these weather elements. For example, the predicted maximum temperature, for say, 4 days hence, will generally rise as a ridge of high-pressure approaches (anticipating warmer winds from lower latitudes once the ridge passes) even though the current temperature is relatively low. Because of such non-randomness, forecasts of weather elements cannot be modelled with a geometric Brownian motion, and Black's (1976) option pricing formula also now can be applied to forecasts of a weather element. From the foregoing, one may note that, in evaluating option contracts used in the context of applying "weather derivatives" to day to day forecasts in a risk management context, it may be demonstrated that the cost of the "weather derivative" option on a forecast increases as the volatility (σ) of the underlying forecast increases in precisely the same manner that the cost of an option on a forward contract increases as the volatility (σ) of the underlying forward increases. The Black (1976) formula for a call option on an underlying struck at K, expiring T years in the future is $c = e^{-rT}(FN(d_1) - KN(d_2))$ and the put price is $p = e^{-rT}(KN(-d_2) - FN(-d_1))$ where

r is the risk-free interest rate
F is the current forward price of the underlying for the option maturity

$$d_1 = \frac{ln(\frac{F}{K}) + \frac{\sigma^2 T}{2}}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

σ is the volatility of the forward price.
and N(.) is the standard cumulative Normal distribution function.

From the formula, the issue of what values to use is not a trivial one. To illustrate, let us suppose that one wishes to value a European call option (using a dividend yield of 0%) on the Day-1 forecast maximum temperature being above 35°C when:

- The forecast at Day-7 is for a temperature of 32°C,
- The *RMS Inter-Diurnal Change* is 2°C,
- The interest rate is 5%, and

- The pay-off is \$1.00 per each degree Celsius above 35°C.

The methodology is illustrated in PANEL A.1 below:

PANEL A.1

Step 1. To neutralise the impact of the choice of units used, add a large number, say, 1000, to both F and K, which results in F=1035 and K=1032.

Step 2. To neutralise the impact of the units used for the volatility, divide the *RMS inter-diurnal change* (2°C) by the new value for K (1032), which approximates to the 1-day volatility that one would obtain under the assumption that the new forecast follows a log-normal process. This is because the new forecast series is a set of large numbers, and, as a consequence, the

RMS inter-diurnal change ~

$\surd((1/6)x\sum(((\ln(Abs(Day\text{-}6/Day\text{-}7))^2 + \ln(Abs(Day\text{-}5/Day\text{-}6))^2 + \ldots + \ln(Abs(Day\text{-}1/Day\text{-}2))^2\ )$

Step 3. To obtain the annualised volatility, multiply the 1-day volatility obtained at Step 2 by √(365) that, in the current case, yields 3.70%.

Step 4. Go to one of the many option calculators on the WEB (for example, Numa Financial Systems, 2006) to obtain a theoretical European call option value based on a maturity date of 6 days hence (Day-7 to Day-1) to yield \$1.06 as the value of the call option.

The proposition that, when undertaking a defensive strategy of purchasing weather derivatives, the cost of protecting against the possibility of weather forecasts being in error reduces as the forecast consistency increases. To illustrate:

- For an *RMS Inter-Diurnal Change* of 1°C, the value of the call option reduces to \$0.25, but,
- For an *RMS Inter-Diurnal Change* of 3°C, the value of the call option increases to \$1.97, and,
- For an *RMS Inter-Diurnal Change* of 4°C, the value of the call option increases further to \$2.93.