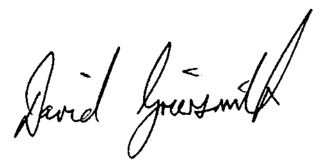
Discussion

17. It could be possible to create a real time data base for a Regional performance measurement system using AIFS, with a simple GUI and graphical display of data for management information purposes, decision-making and alert generation. The data could be placed on the WWW.

18. As a way forward, a prototype measurement system (i.e. perhaps phase 1 of 2) could be established for a defined part of AIFS which involves a well-defined client grouping. This might enable an assessment of the benefits or otherwise of the system.



(David Griersmith)
SRSA
17 July 1997

wpwin6.1: regionperfmon1.wpd

VERIFICATION OF FORECASTS FROM THE 1960S TO THE 1990S

HARVEY STERN

Victorian Regional Office
Bureau of Meteorology
Box 1636M
Melbourne 3001
Australia

ABSTRACT

The author's recent work on the accuracy of weather forecasts is summarised. Predictions examined include those of temperature, rainfall, aviation significant weather and extended outlooks. The verification of seasonal predictions and worded components of forecasts are also discussed.

Verification statistics may be used to maximise the performance of automated guidance; to enable individual forecaster characteristics to be evaluated; to show how human emotions and other factors adversely influence the quality of predictions; and, to assist management involved in policy formulation.

It is concluded that technological advances reflected in the development of systems such as the Australian Integrated Forecast System (AIFS) create some wonderful opportunities to use verification statistics in a continuous improvement sense by providing direct and immediate feedback into the forecast process and may even change that process.

"In the more stringent form of the scientific method, forecasting tests hypotheses; forecast errors cause hypotheses to be revised, and these are then tested by further forecasts. At first glance, meteorology, a forecasting discipline, is uniquely suited to be scientific ... In the day-to-day operations of meteorological services the constant stream of predictions (some of which are wrong) should quickly demolish or modify doubtful hypotheses. In fact, feedback between prediction and hypothesis is rare" (Ramage, 1993).

1. INTRODUCTION

One of the earliest studies of the impact of meteorological advances on forecast performance was carried out by Scott (1873). He examined the impact of the burgeoning telegraph network upon the issue of gale warnings for British waters. A positive impact was found. This example underlines the main purpose of forecast verification, which is to establish what impact developments in the science have had on the accuracy of the product in order to establish whether or not work is proceeding in the correct direction. However, nearly a century later, a person of the status of the Director of the South African Weather Bureau (Schumann, 1950 a&b) wrote that "nobody is able to state categorically to what extent forecasting practice has benefited by the improved methods and bigger budgets".

The purpose of this paper is to provide analyses of the accuracy of forecasts of various weather elements. The focus of the analyses presented is largely Melbourne. This is in order to take advantage of the author's experience covering more than two decades at the Bureau of Meteorology's (BoM) Victorian Regional Office (VRO) in Melbourne. This experience provides the author with a practical and significant background in interpreting some of the trends in accuracy in terms of changes in procedures at the VRO. The author has previously documented trends in the accuracy of forecasts both from an Australia-wide (Stern, 1979a, 1980a, 1986 a&b) and overseas (Stern, 1980b) perspective.

Ramage (1993) proposes that the "scientific impact of weather forecasting on understanding weather could be enormous ... (but) we have allowed meteorology to split into two sub-disciplines separated by the activity that could have provided the scientific cement - forecasting". He suggests that poor forecasts are a consequence of poor understanding and presents an iterative procedure whereby understanding (that is, accepted theory) can be modified and then lead to forecast improvement. This is diametrically opposed to what he terms "the bureaucratic viewpoint ... (that) understanding already exists". With that mindset, deficiencies in the current pathway are seen as temporary - more powerful computers being regarded as means (in the future) to overcome problems in the numerical processing; more data being regarded as a means (in the future) to overcome inaccurate specification of the initial state. An iterative procedure is proposed whereby verification of forecast performance may be used in a systematic manner as a vehicle to evaluate and modify theory which then feeds directly back into the forecast process.

2. AN ANALYSIS OF MELBOURNE'S TEMPERATURE ERRORS

To begin, a statistical analysis of a thirty five year record of Melbourne maximum temperature forecast errors is presented. An error in a maximum temperature forecast is defined as occurring if there is more than a 5°C difference between the observed maximum temperature and the previous afternoon's forecast. The frequency of these errors is shown to display a strong seasonal variation (Fig 2.1). The frequency peaks during the summer month of December and is at a minimum during the winter months of June and July. Fig. 2.1 shows that the seasonal variation in the frequency of errors is similar to the seasonal variation in the variability of maximum temperature. The seasonal variation in the error frequency is, therefore, considered to be largely a consequence of greater difficulty being associated with predicting maximum temperature during the seasons with greater variability.

Since the mid-1970s, the annual frequency of errors has ***halved*** - from about 30 per annum to about 15 per annum (Fig. 2.2). The reduction in error frequency is explained by the expansion of remote sensing sources of data, in particular, satellite imagery, and the increasing sophistication of Numerical Weather Prediction (NWP) models, and the implementation of techniques to statistically interpret their output in terms of local weather elements such as the Analogue Statistics Model (ASM) (Stern, 1980 b&c; Dahni *et al*, 1984; Dahni, 1988) and Model Output Statistics (MOS) (Woodcock, 1984). The ASM was first implemented operationally in the VRO as a part of the Automated Regional Operations System (AROS) (Barclay and Butt, 1988). It later came to be referred to as the Generalised ASM (GASM) (Dahni and Stern, 1995; Dahni, 1995 & 1996 a,b&c, Stern, 1996a), when it was implemented as a part of the Australian Integrated Forecast System (AIFS) (Love, 1994, Gigliotti, 1995).

With regard to satellite imagery, there was once-daily visual imagery during the late 1960s, twice daily imagery (including infrared imagery) by the middle 1970s and three-hourly imagery from the Japanese Geostationary Meteorological Satellite (GMS) from the late 1970s followed by satellite derived winds and temperature soundings. The sharp, albeit temporary, improvement in 1979/1980 may be

attributed to the influx of buoy data associated with the First GARP (Global Atmospheric Research Project) Global Experiment (FGGE) (Hicks *et al.*, 1979; Pendlebury, 1979; West, 1979; Zillman, 1981).

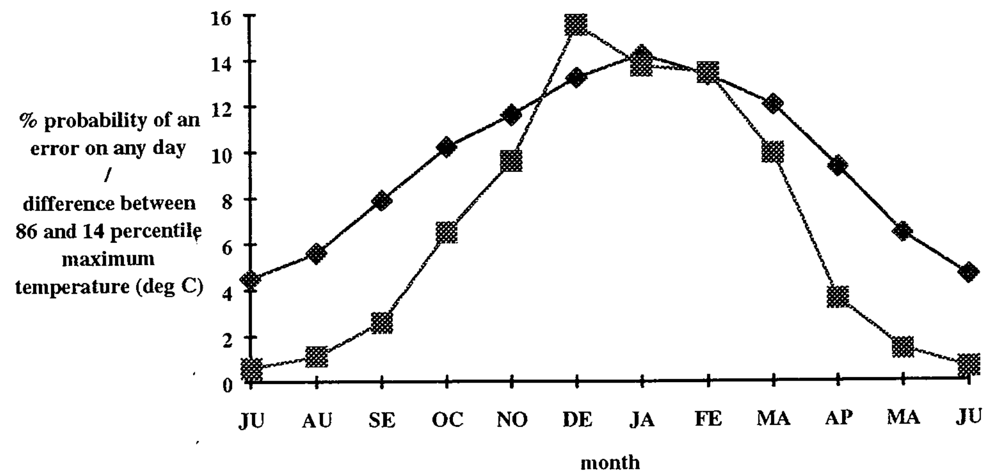


Fig 2.1 Seasonal variation of (1) the likelihood of an error in the maximum temperature forecast occurring on any day (squares) and (2) the variability of the maximum temperature illustrated by the difference between the 86 percentile and the 14 percentile maximum temperatures (diamonds) (from Stern, 1996a&b).

However, between the early 1960s and the early 1970s a slight increase in error frequency is evident. This is attributed to the fact that during that period the majority of a group of very experienced forecasters were replaced (as a consequence of retirement or transfer) by a less experienced group. By the early 1970s, there was a team with mostly less than the 4 years experience level designated by Gregg (1969) as that needed before gains from "additional" experience are no longer made.

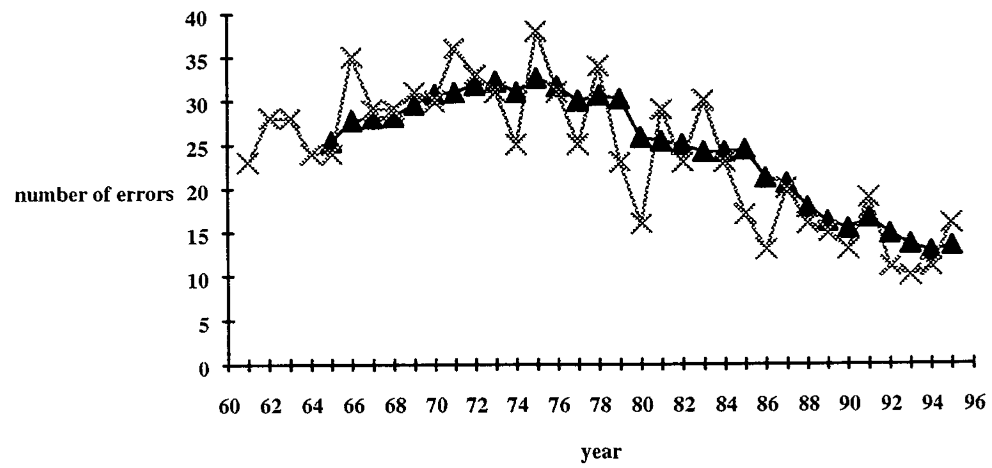


Fig 2.2 Inter-annual variation in the number of Melbourne maximum temperature errors in 12 month periods ending in June of the indicated year (crosses) with the trend illustrated by 5-year running means (triangles) also ending in June of the indicated year (after Stern, 1996a&b).

In the late 1980s, the VRO Regional Forecasting Centre (RFC) commenced a program of issuing maximum temperature forecasts for Melbourne out to four days. The accuracy of these temperature forecast outlooks increased rapidly and Fig 2.3 shows that 2-day outlooks of maximum temperature now possess a superior level of accuracy to that of the 1-day temperature forecasts in the 1960s, while the accuracy of 3-day and 4-day outlooks are approaching that level of accuracy. This trend reflects improvements in understanding of broadscale evolution of synoptic patterns, attributable largely to advances in global NWP modelling.

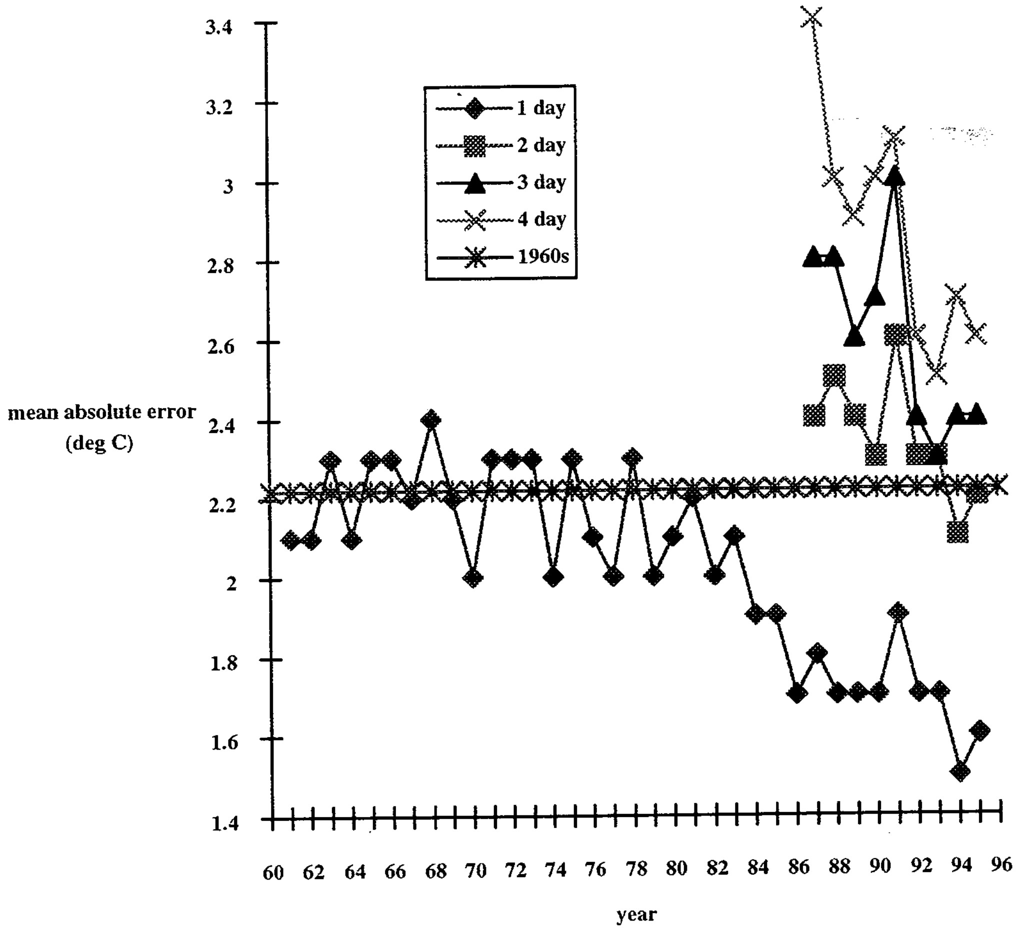


Fig 2.3 The trend in the mean absolute error (°C) of maximum temperature outlook forecasts in 12-month periods ended in June of indicated year (the straight line represents the mean absolute error in 1-day forecasts during the 1960s, while the other lines, from lower to upper, respectively represent the mean absolute errors in 1, 2, 3 and 4 day outlook forecasts) (from Stern, 1996a).

3. VERIFICATION OF PRECIPITATION FORECASTS

The issue of precipitation forecasts, both predicting amount and occurrence is now discussed. Fig 3.1 depicts trends in the accuracy of forecasts of the occurrence, or otherwise, of precipitation in Melbourne, compared to the "no skill" persistence forecast. The depiction suggests a slight, albeit erratic, fall in forecast performance over the period. Fig 3.1 also indicates that morning forecasts (minimal lead time) are better than the previous evening's predictions, and that both morning and evening forecasts are only a few percentage points ahead of persistence. Fig 3.2 depicts trends in the accuracy of Quantitative Precipitation Forecasts (QPF) in Melbourne, compared to the "no skill" persistence forecast. The depiction suggests a fall in forecast performance into the 1970's followed by a rise. The performance at the end of the period, however, is only marginally ahead of that at the beginning. The fall in performance into the 1970s may be attributed to the same reason as was attributed to the fall in performance at maximum temperature prediction during the

early years, that is, inexperience of the forecast team. Fig 3.2 also shows that morning forecasts (in spite of minimal lead time) were, overall, little better than the previous evening's predictions throughout the period. Both morning and evening forecasts during some periods were no better, and sometimes worse, than persistence.

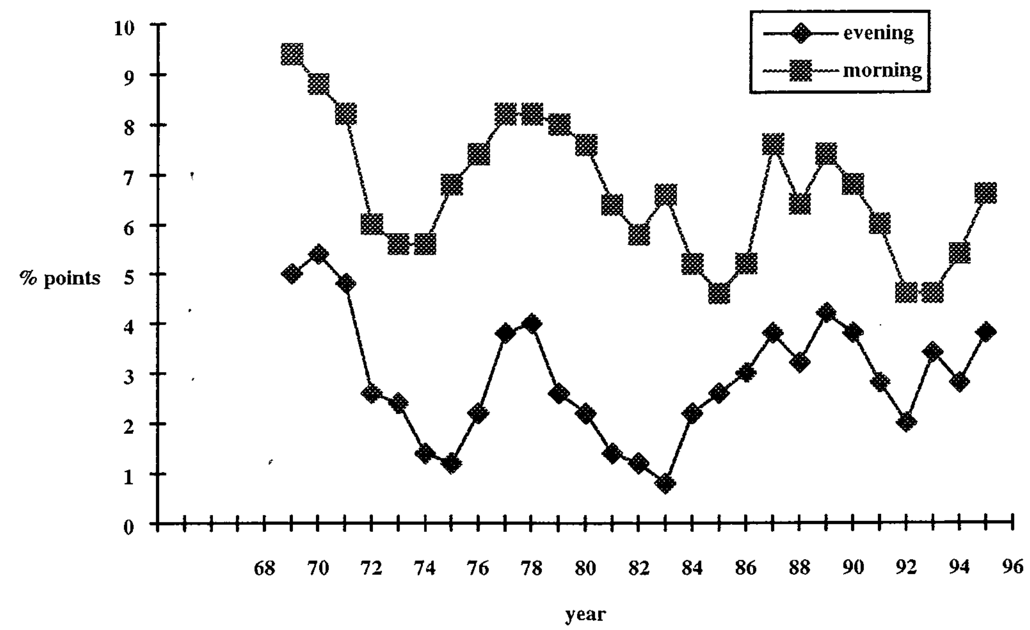


Fig 3.1 Trend in skill of the previous evening's (diamonds) and morning's (squares) forecasts of the occurrence, or non-occurrence, of rainfall. This trend in skill is represented by 5 year running means, ended in the indicated year, of the number of percentage points over that of persistence, of the percentage of rainfall forecasts correct in terms of rain / no rain.

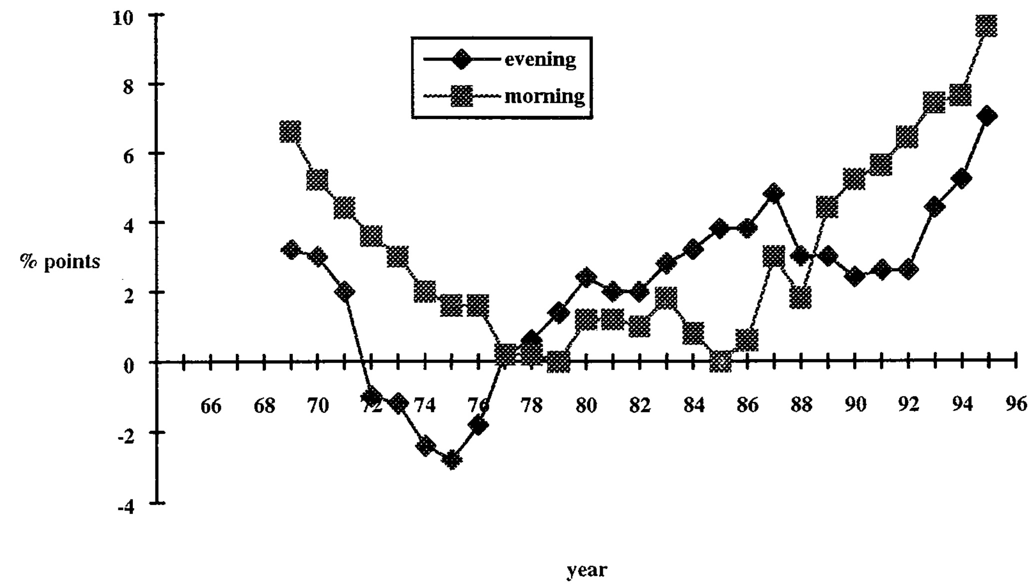


Fig 3.2 As for Fig 3.1, but for the percentage of QPFs in the correct rainfall range.

4. VERIFICATION OF FORECASTS OF AVIATION WEATHER

Records of the accuracy of forecasts at the Melbourne Airport terminal (Tullamarine) have been maintained since it opened during the early 1970s (Shanahan, 1972&1995). There are three categories of aviation-significant weather for which records of forecasting performance have been kept. These categories are thunderstorms, fog and low cloud.

As these events are all rare events, measures such as "percentage correct" on an "observed/not observed" basis will be dominated by the frequency of occurrence of the event for the particular season. For this reason, the measures of forecast skill that have been chosen are two of those not dominated in that way and that are readily able to be extracted from contingency tables. They are the "Probability of Detection" (POD) and the "False Alarm Ratio" (FAR). These measures, which are used for describing the skill at forecasting rare events, were first described by Peirce (1884) after Finley (1884) "wrongly" claimed a high success rate at tornado forecasting by simply reporting the total percentage of correct (observed and forecast, not observed and not forecast) predictions. The measures are defined by Doswell *et al.* (1990), these definitions being reproduced in Table 4.1.

There is substantial risk to life and property if aviation-significant weather is not forecast and it occurs, hence, operational meteorologists seek to achieve a high POD. However, considerable expense is incurred by airlines each time there is such a forecast of aviation-significant weather, hence, operational meteorologists also seek to achieve a low FAR. Although these measures do provide an indication of forecast skill, they suffer from the disadvantage, as pointed out by Doswell *et al.* (1990), of not giving credit for null events (not observed/not forecast).

The Bureau of Meteorology divides the aviation forecast verification data into three lead times - short (1-7 hr), medium (7-13 hr) and long (13-19 hr) (Shanahan, 1972&1995). The trend in the accuracy of 1-7 hr lead time forecasts for the Melbourne Airport terminal is indicated in Fig 4.1, which presents for all aviation-significant elements combined, 5-year running means of POD and FAR. This combining of all aviation-significant elements is done in order to obtain overall trends, notwithstanding that some trends in specific elements might be masked as a result. It might be suggested that the errors be dis-aggregated into their various components. However, aviation significant weather elements occur only rarely. As a result, the data are too few for reliable conclusions to be drawn were they to be dis-aggregated. Furthermore, the primary purpose of the present analysis is to examine trends in forecasting performance from the point of view of the client. Given that the client's primary interest is whether or not the Airport is to be closed on account of aviation-significant weather, than in the nature of that weather, it would be inappropriate to dis-aggregate the errors. Fig 4.1 depicts a slight, albeit erratic, decline in POD, that fall accelerating towards the end of the period (deteriorating forecast performance). Fig 4.1 also depicts a slight, albeit erratic, rise in FAR over the period (deteriorating forecast performance), and a temporary reversal in the early 1980s of the performance trend highlighted above (a rise in POD, accompanied by a steady FAR).

Table 4.1 Definitions of the forecast performance measures "Probability of Detection" (POD) and "False Alarm Ratio" (FAR) (from Doswell *et al.*, 1990).

Forecast/ Observed	No	Yes
No	W	Y
Yes	Z	X

with **POD** = x/(x+z),and **FAR**=y/(x+y).

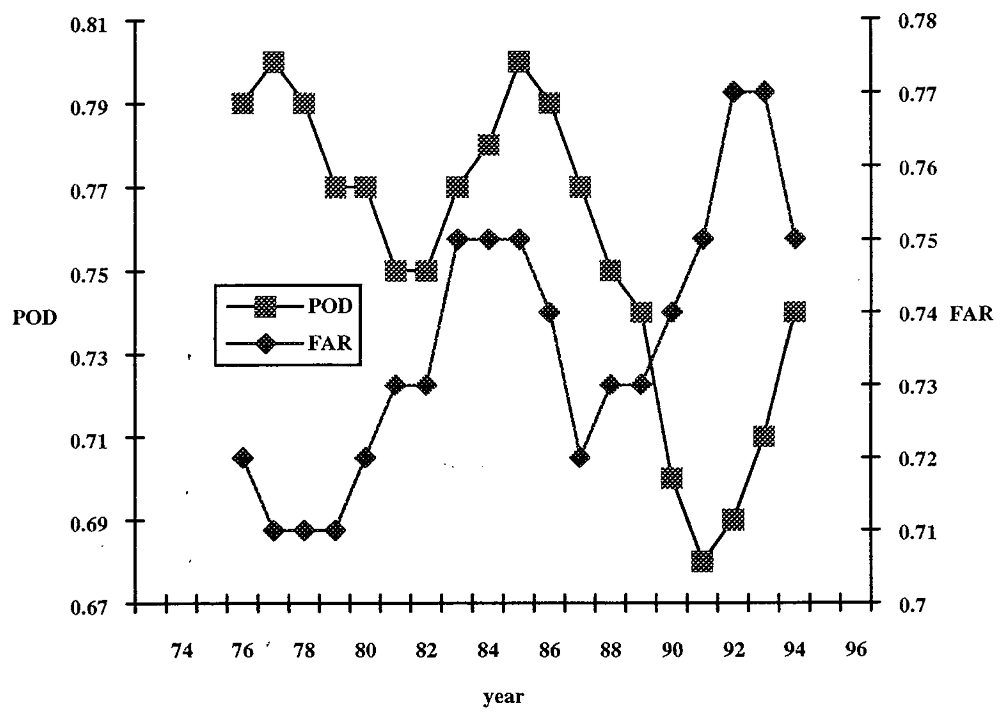


Fig 4.1 Trend in the accuracy of 1-7 hr Melbourne Airport forecasts, all elements combined, depicted by 5 year running means (ended in the indicated year) of probability of detection (POD) (squares) and false alarm ratio (FAR) (diamonds) (from Stern, 1996a).

The trend in the accuracy of 7-13 hr lead time forecasts for the Melbourne Airport terminal is indicated in Fig 4.2, which presents for all aviation-significant elements combined, 5-year running means of POD and FAR. Fig 4.2 depicts similar trends to those depicted for 1-7 hr lead time forecasts, namely, a slight, albeit erratic, decline in POD (deteriorating forecast performance), a slight, albeit erratic, rise in FAR (deteriorating forecast performance), and a temporary reversal in the early 1980s of the performance trend highlighted above (a rise in POD, accompanied this time by a fall in FAR).

The trend in the accuracy of 13-19 hr lead time forecasts for the Melbourne Airport terminal is indicated in Fig 4.3, which presents for all aviation-significant elements combined, 5-year running means of POD and FAR. Fig 4.3 depicts a different trend in the POD to those depicted for shorter lead time forecasts, namely, a slight, albeit erratic, rise (improving forecast performance). However, a slight, albeit erratic, rise in FAR is also detected over the period (deteriorating forecast performance), so counteracting the trend shown by POD, and once again, in the early 1980s, a temporary rise in POD, accompanied by a temporary fall in FAR.

In summary, a number of key conclusions may be drawn. Firstly, there is an improvement in forecast performance for all lead times during the early 1980s - this might be attributable to the development of semi-objective forecasting aids for fog (Goodhead, 1978) and low cloud (Keith, 1978) just prior to the early 1980s, and underlines the importance of even semi-objective guidance. Secondly, there is no overall trend in the forecast performance in long lead time (13-19 hr) predictions. Thirdly, there is a trend towards deteriorating forecast performance in shorter lead time (1-7 hr and 7-13 hr) prediction. One encouraging feature evident in Figs 4.1, 4.2 and 4.3 is an improvement in forecast performance according to all measures at the very end of the verification period. This improvement may be attributed to a decision to transfer aviation forecasting staff to the RFC in the VRO, where full advantage might be taken of the knowledge of all the VRO RFC forecasting staff, which was not possible at the remote airport location.

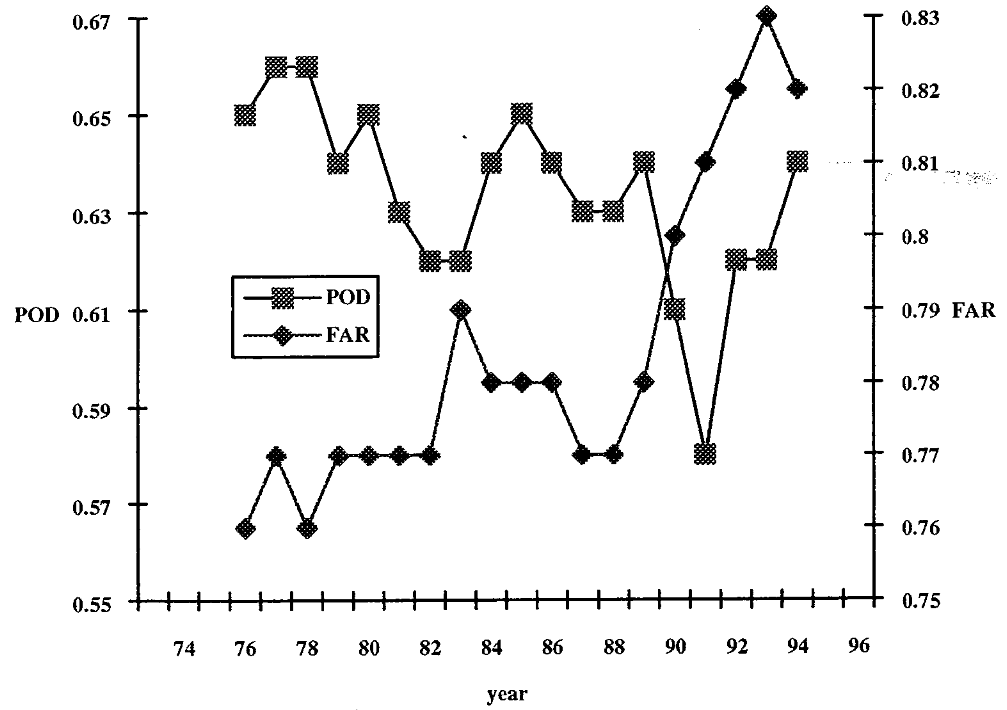


Fig 4.2 As for Fig 4.1, but for 7-13 hr Melbourne Airport forecasts (from Stern, 1996a).

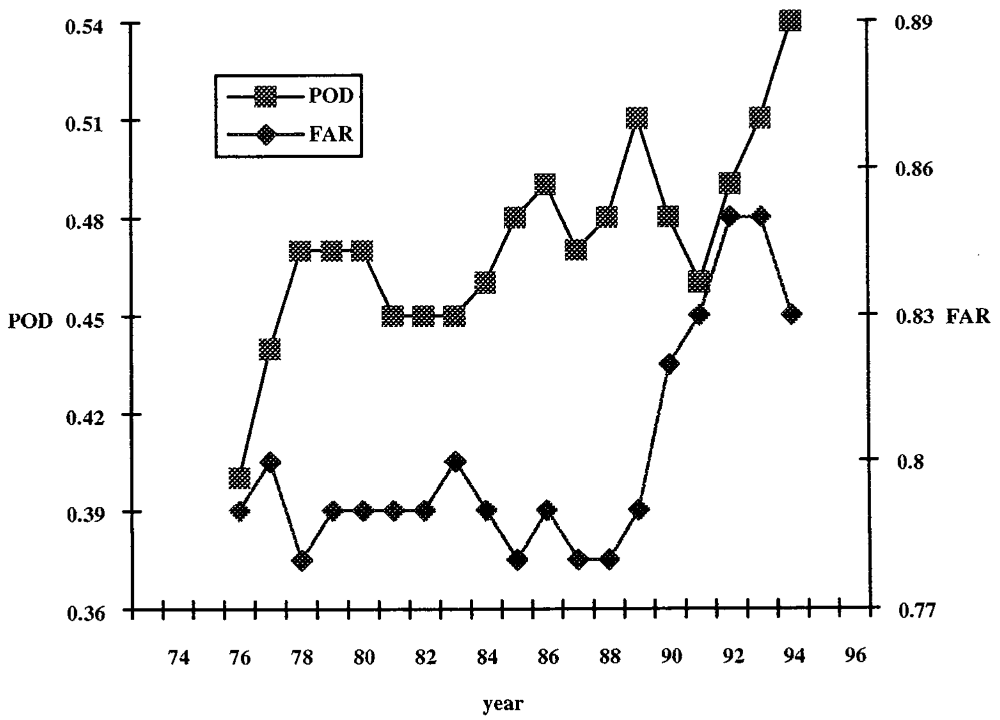


Fig 4.3 As for Fig 4.1, but for 13-19 hr Melbourne Airport forecasts (from Stern, 1996a).

5. EXTENDED OUTLOOKS

The VRO currently provides 1 - 4 day weather forecasts to the general public. These forecasts are issued each afternoon by the VRO RFC in Melbourne. The forecasts are based upon an interpretation, in terms of local weather, of the output

of various global and Australian region NWP models, using a combination of GASM, MOS and other guidance

The National Center for Environmental Prediction (NCEP) currently produces a 15-day global ensemble average prognosis. That the NCEP prognosis extends out to 15 days is consistent with the work of Lorenz (1963, 1969a&b, 1993), which suggests a 15-day limit to day-to-day predictability of the atmosphere.

The primary purpose of this Section is to present preliminary results of an experiment (Stern, 1998) which involves verifying a set of quantitative forecasts for Melbourne out to 14 days. These forecasts are prepared based on the NCEP output when available, for days 5 to 14, and on the official forecasts for days 1 to 4.

These experimental forecasts are verified against "climatology". A variety of verification measures are used, with a view to establishing a possible limit to predictive capability. These measures are

*Root Mean Square (RMS) error of the minimum temperature forecasts, *af*;
*RMS Error of the maximum temperature forecasts, *bf*;
*RMS Error of the quantitative precipitation forecasts, *cf*;
*Percentage rain/no rain forecasts correct, *df*; and,
*Brier skill score about probability of precipitation forecasts, *ef*.

For each of the forecast days 1 - 14 inclusive, these measures are calculated. They are then compared with corresponding measures of the performance of climatology (ac, bc, cc, dc, ec) and combined into a series of skill scores for each of the day 1 to 14 forecasts. Firstly, the "temperature skill score" is defined as

$$100x((ac/af)+(bc/bf)-2)/2$$

Secondly, the "rainfall skill score" is defined as

$$100((cc/cf)+sqrt(df/dc)+sqrt(ec/ef)-3)/3$$

Thirdly, the "joint skill score" is defined as the mean of the two others.

Fig 5.1 shows that skill at predicting both temperature and rainfall decline, albeit unsteadily, as one moves from day 1 to day 14, the latter more rapidly than the former. The unsteady character of the declines is probably a consequence of the experiment being in its early days – the experiment only began in May 1997, the first forecast verified being that based on 20 May data, the most recent forecast verified being that based on 19 July data. The decline might be expected to become smoother as the numbers of data increase. At the present stage of the experiment, all that may be said is that the "jury is still out" about whether or not extending the official forecast period beyond day 4 would be justified.

However, it already is quite clear from Fig 5.1 that there is little justification for outlooks beyond day 7. This renders inappropriate, for example, the tendency for traders in agricultural commodities to utilise such predictions in their work. As an illustration of this tendency, refer to the Wall Street Journal of 22 June 1992, which reported that "the possible development of a high pressure ridge", depicted in the 10^{th} day of the US National Weather Service's model, sparked "renewed fears of a drought in the central Midwest (and) drove grain futures prices higher at the Chicago Board of Trade". The model's prediction proved to be incorrect and the Wall Street Journal of 3 July 1992 (eleven days later) reported that "heavy rain

... helped alleviate short-term drought fears and drove grain futures prices lower at the Chicago Board of Trade".

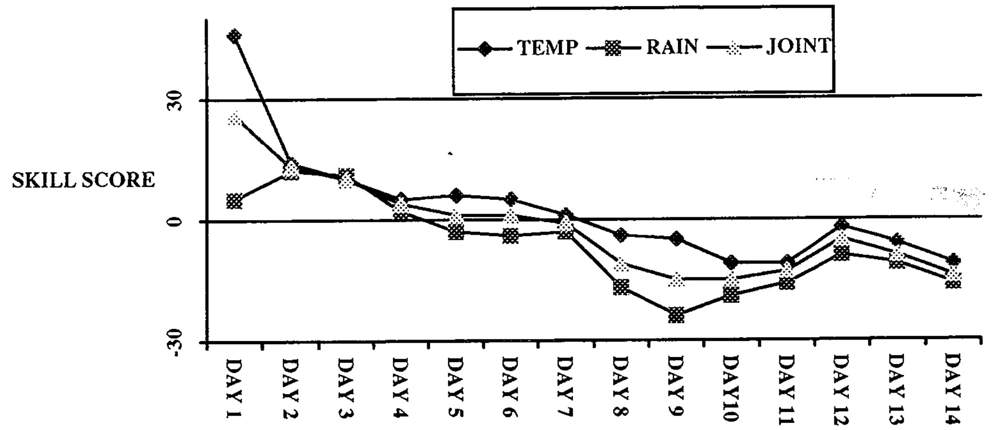


Fig 5.1 Temperature, rainfall and joint skill scores, associated with forecasts for days 1 to 14, compared with climatology. A positive value of the skill score suggests a forecast displaying skill in excess of climatology, whereas a negative skill score suggests the reverse.

6. THE ECONOMIC IMPACT OF SEASONAL PREDICTIONS

The accuracy of the Victorian component of the BoM's seasonal climate outlook (SCO) service is now reported upon. In Victoria, major users include Bureau customers such as fire control authorities, authorities involved in managing reservoirs from the point of view of the provision of irrigation waters and flood mitigation, and the farming community. The outlook service consists of two components. These are a general press release which, in qualitative terms and usually with the aid of a map, indicates areas likely to experience particular rainfall outcomes (below average/average/above average) and a publication for specialist clients which, in addition to the above, provides a probability distribution of particular rainfall outcomes.

Stern and Williams (1988, 1989) specifically investigated the feasibility of providing Victorian fire control authorities with advice as to the likely severity of forthcoming fire seasons (usually December to March). To achieve that end, the existence of a relationship between the El Niño Southern Oscillation (ENSO) phenomenon and Australian rainfall was noted. A focus was then placed on Melbourne rainfall, this focus identifying that the relationship was "particularly marked" during spring and somewhat weaker in summer (Fig 6.1), and that there was a consequent link between ENSO and fire risk (Fig 6.2). This link arises because an extended dry period is a necessary, although not sufficient, precursor to a severe fire weather season in Victoria (Foley, 1947). On this basis, Stern and William's (1988, 1989) work led to the BoM VRO adopting the practice of providing a briefing to fire control authorities about the likely severity of forthcoming fire seasons.

Since 1989, the BoM has provided a *publicly issued* national seasonal outlook (BoM, 1989) in qualitative terms (map and text). Initially, it was restricted to particular times of the year (June to December) and to limited areas of the country (mainly inland eastern and northern Australia) associated with relatively higher rainfall / SOI correlation coefficients. The outlook indicates the expected precipitation regime on a three-category basis (below average (deciles 1 - 3), average (deciles 4 - 7), or above average (deciles 8 - 10). It now applies to all

seasons and most areas (with the exception of those designated as "seasonally dry") and includes a quantitative component (% probability of each precipitation category). It can be demonstrated (Stern, 1995a&b, 1996a) that randomly generated outlooks are correct, on the one in three category basis used here, on 34% of occasions, the low percentage (well below 50%) simply being a consequence of having to select one of three possible outcomes (instead of two). It can also be demonstrated that the assumption of climatology (average) is correct on 40% of occasions. Since the SCO service was introduced in 1989, an outlook based on the assumption of persistence (that the preceding three months' rainfall anomaly would be repeated in the subsequent three-month period) would have been correct on 44% of occasions.

Fig 6.3 summarises the monthly variation in the accuracy (on the three-category basis) of that part of the qualitative component of the outlook valid for part or all of Victoria. Each outlook that relates to a part of the State is verified, and its performance is attributed to all months of the particular outlook period. Fig 6.3 shows that in the second half of the calendar year the national seasonal outlook has out-performed each of the "no-skill" measures. This corresponds with BoM's (1989) expectation as to that part of the year with highest skill and is only slightly different to that anticipated by Stern and Williams (1988, 1989). The performance depicted in the January to June period is a reflection of the weakness of the rainfall/ENSO relationship during the first half of the calendar year. This is consistent with Latif *et al.*'s (1994) finding that "...a seasonal dependence in predicability (of different regional anomalies known to be associated with ENSO) is well established..." So, the improved accuracy during the latter half justifies the issue to fire authorities of a statement of expected severity of the forthcoming fire season.

Fig 6.4, which depicts the seasonal cycle of one-season theoretical Australian continental-average canonical correlation skill for precipitation forecasts, and whose shape is remarkably similar to that of Fig 6.3, suggests that the basis for this extended service may have been more due to a demand for such a service rather than an improvement in the techniques used. The message from the above illustrations is that service providers need to be cautious about implementing new forecast products without an accompanying forecast of forecast skill (Tennekes *et al.*, 1987).

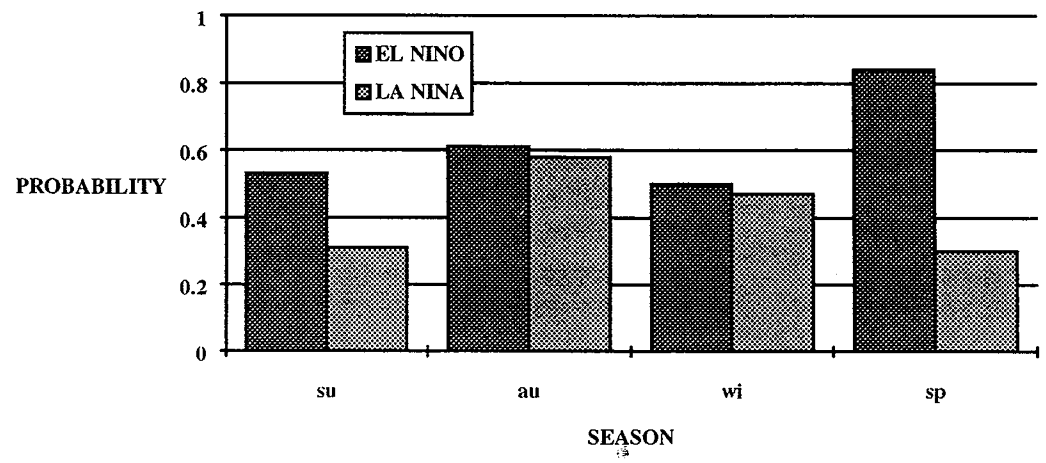


Fig 6.1 Relationship between Melbourne seasonal rainfall and the ENSO phenomenon, depicting the probability that Melbourne seasonal rainfall (su=summer, au=autumn, wi=winter, sp=spring) is below the median, depending upon whether or not an ENSO or LA NINA event is operating (from Stern and Williams, 1989, 1996a).

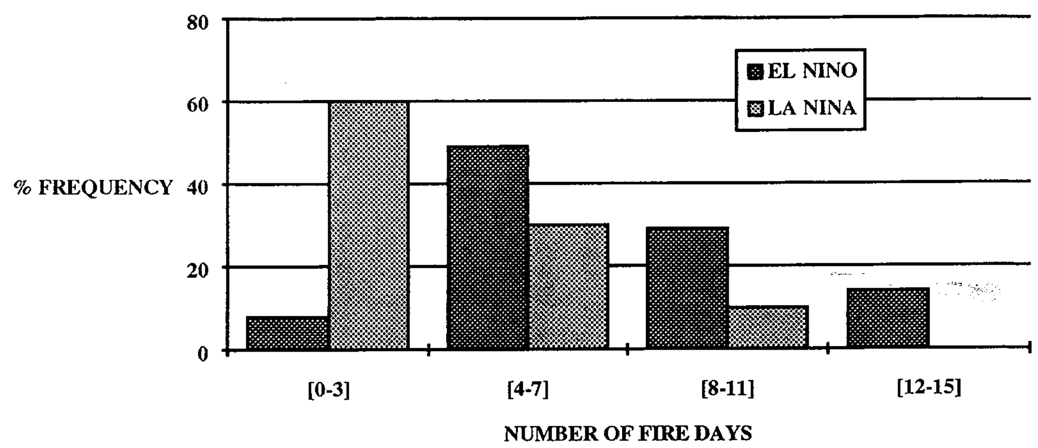


Fig 6.2 Frequency distribution, across four ranges, of the number of fire risk days at Melbourne in a season, given an ENSO or LA NINA event (from Stern and Williams, 1989, 1996a).

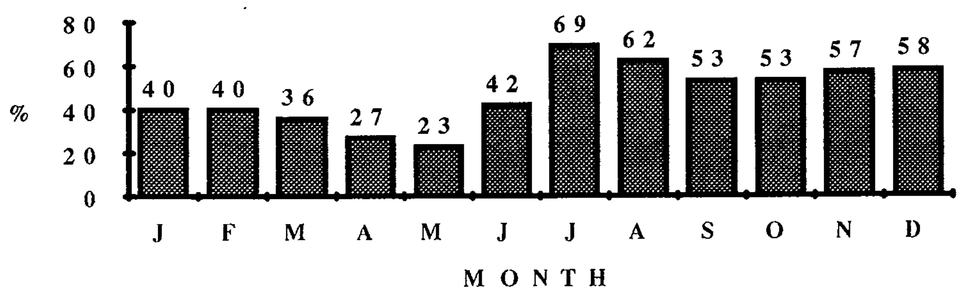


Fig 6.3 Monthly variation in the percentage of correct SCOs (from Stern, 1995 a&b).

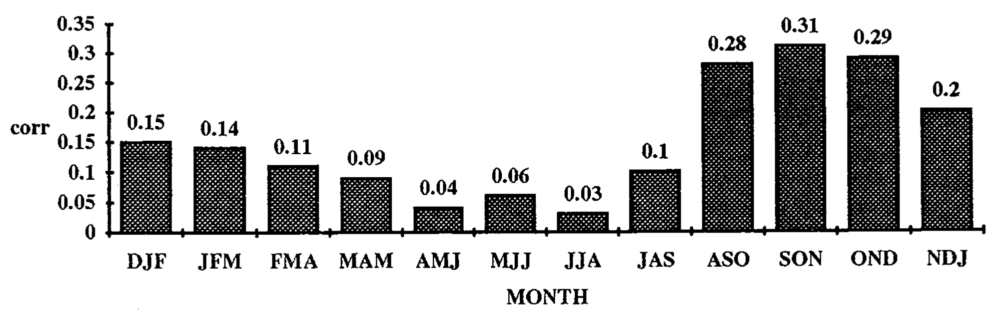


Fig 6.4 The seasonal cycle of one-season theoretical Australian continental-average canonical correlation skill for precipitation forecasts (from Barnston and Smith, 1996).

The subject of forecast verification may be examined directly from the point of view of the economic value of the forecasts to users. This has been addressed by other authors from, for example, the point of view of the value of frost forecasts (Katz *et al.*, 1982), practical utility (Murphy, 1990), management decision making (Spetzler and Staëler von Holstein, 1975; Woodcock, 1976; Wilks, 1995), the value of climate forecasts to agriculture (Stone and Hammer, 1995), Climate Change (Industry Commission, 1991), and gains to the whole economy (Thompson, 1962). Ehrendorfer and Murphy (1988) show how weather forecasting systems may be evaluated in terms of a range of criteria. An Australian example is that of Stern (1996a), who compared the difference in the risk/return

relationship associated with official BoM predictions, and theoretical seasonal rainfall predictions based on the assumption of climatology (Fig 6.5).

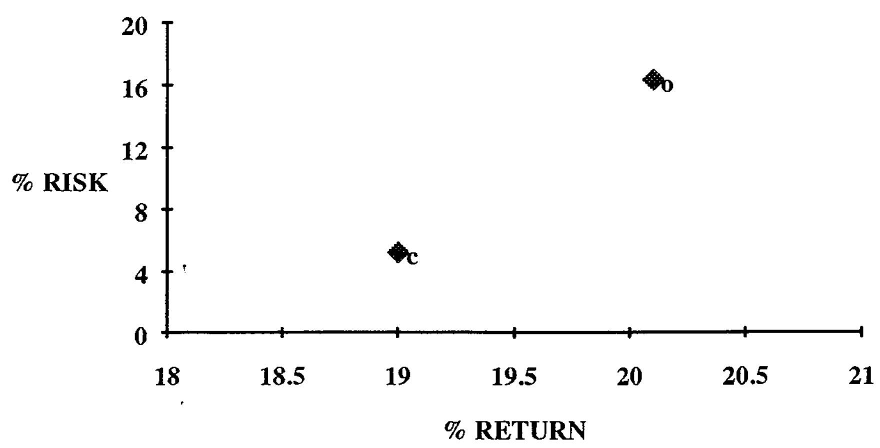


Fig 6.5 Comparing the difference in the risk/return relationship associated with official BoM predictions (o) and theoretical seasonal rainfall predictions based on the assumption of climatology (c) (from Stern, 1996a).

7. VERIFYING WORDS USED IN WEATHER FORECASTS

It is also necessary to deal with the verification of worded components of forecasts. Subjective systems of verification of the words in forecasts have been in use for considerable time. During the 1980s, the BoM employed a so-called "reasonable person" scheme which was based on an independent assessor giving a rating to the forecast according to his or her perception of the usefulness of the prediction. While having the advantage of focussing upon client perceptions of the product's value (the assessor was usually a member of the general public), the scheme nevertheless suffers from the disadvantage of being dependent upon the assessor's subjective opinion - change the assessor, the evaluation could change. As a result, there is no unambiguous basis for comparing forecasts from different sources or establishing whether or not there is a trend towards improvement. A semi-objective system was designed by Colls *et al.* (1981) but the lack of objectivity in determining what was observed renders this system as failing on similar grounds to the "reasonable person" scheme.

Stern (1980b) developed a standard system of terminology linked directly with the coded observations, as recorded in the official observation book (BoM, 1977) and is presented, in full, in Stern's (1980b) Appendix A. This was done for the ASM in order to "exclude the possibility of an ambiguous interpretation" which would have made it "difficult to (evaluate) the degree of success" (Dobryshman, 1972) of the ASM in comparison to the officially issued forecasts. The format for the system of terminology is a description based on weather and cloud, followed by a description based on wind, followed by minimum and maximum temperatures and rainfall amount, followed by a precis and clarification of preceding components. Stern's (1980b) Appendix C includes an objective scheme to "translate" official forecasts into that terminology and a consequential objective means to evaluate both types of product. That system of terminology has formed the basis of the textual component of subsequent upgrades of the ASM - see Dahni (1988) and Dahni and Stern (1995) for details.

8. FEEDBACK

As indicated in the introduction, the emphasis of this paper is upon feeding back verification data into the forecast process. This is now discussed in some detail, with a concentration on temperature forecasts because of the availability of a long period of relevant data.

A system of synoptic classification for southeastern Australia was employed to stratify the errors synoptically. The system is first referred to by Treloar and Stern (1993) and is described in detail by Stern (1996a). In summary, the basis for the synoptic classification is simply the direction and strength of surface flow and whether that flow is cyclonic or anticyclonic.

The frequency of errors is shown to be synoptically dependent (Fig. 8.1 and Fig 8.2). The synoptic class associated with the greatest ***number*** of errors (6.9% of the total) is the frequently occurring (4.4% of the total) "moderate NNE anticyclonic" synoptic class. This is possibly because of difficulty with assessing the likelihood of a sea breeze as the associated anticyclone moves to the east (particularly early in summer, when the lag in the warming of the ocean, compared to the land, provides a greater potential for sea breezes), difficulty with assessing the potential of upslide cloud (also early in summer, when the lag in the warming of the ocean (compared to the land) provides a greater potential for strong east-west thermal gradients), and this synoptic class occurring more than twice as frequently during the (more changeable) summer months than during winter. Other synoptic classes associated with a particularly high number of errors include "strong NNE anticyclonic", "moderate ENE anticyclonic", "moderate WNW cyclonic" and "strong WNW cyclonic".

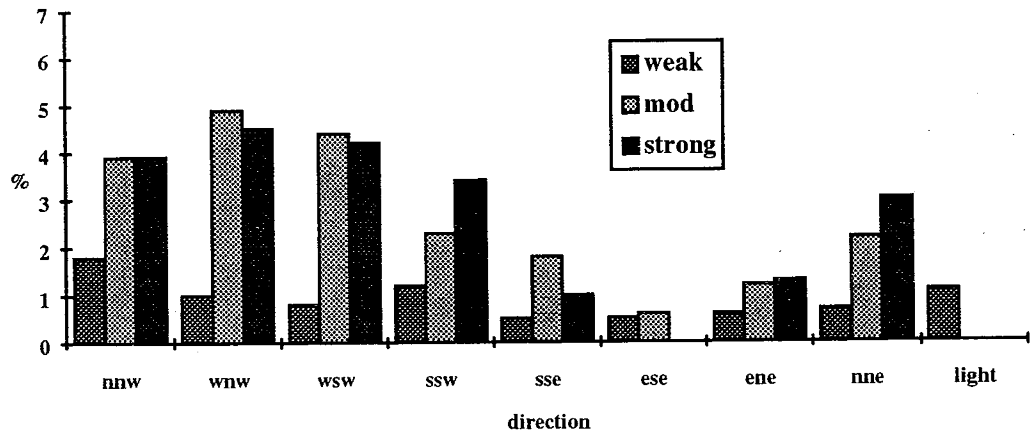


Fig 8.1 The percentage of errors associated with cyclonic synoptic classes. Each group of three columns represents, from left to right, weak, moderate and strong synoptic classes (from Stern, 1996a&b).

The monthly peak in error frequency arises in December, even though the monthly peak in maximum temperature variability occurs in January. This is, in part, a consequence of the "moderate NNE anticyclonic" synoptic class being associated with fewer errors in the latter month (possibly due to land-ocean temperature contrasts reducing, as the summer advances, so leading to lesser difficulty with assessing the potential for sea breezes and upslide cloud). It is also, in part, a consequence of errors associated with the "strong WSW cyclonic" and "strong SSW cyclonic" classes. These classes, while not being among the main

overall contributors to forecast errors, emerge as a major contributor specifically during the early summer, being also associated with fewer errors in the latter month (possibly due to skill at predicting frontal passages increasing as the summer advances). However, other classes, for example, the "strong ENE anticyclonic" class, show an apparent decrease in skill.

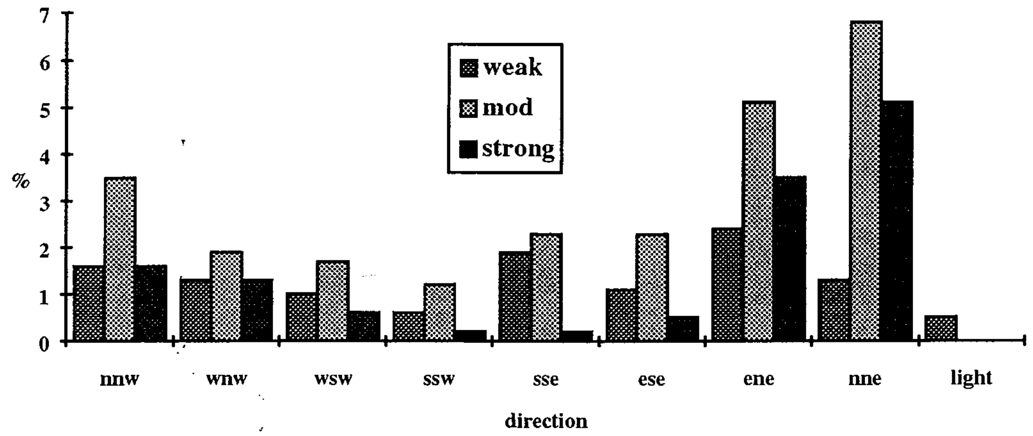


Fig 8.2 As for Fig 8.1, but for anticyclonic synoptic classes (from Stern, 1996a&b).

A better indication of difficulty associated with each synoptic class may be an analysis of errors carried out to the exclusion of the 5 middle months of the year (May to September, inclusive, or 42% of the data) when few errors (66, or only 8% of the total) occurred. By this approach, the tendency for common "winter types" to be wrongly regarded as "not difficult" would be eliminated - the word "wrongly" is used because the only reason they appear to be associated with a low percentage of errors is due to their occurrence during the winter months, when the temperature variability is relatively low. A parameter suitable for reflecting the true indication of forecast difficulty would therefore be, for each of the synoptic classes:

(number of October to April errors) / (number of October to April members of that type)

The value of this parameter is depicted, for each class, in Figs 8.3 and 8.4. Figs 8.3 and 8.4 show that there is greater forecast difficulty with cyclonic synoptic classes than with anticyclonic synoptic classes. This is possibly the result of problems with forecasting the synoptic evolution of accompanying troughs and fronts. Specifically, the "weak NNW cyclonic" class is seen to be the most difficult synoptic class - possibly a result of the forecasters facing the dual problems of predicting the onset of a fairly weak approaching front or trough, and predicting the occurrence, or otherwise, of a sea breeze. Furthermore, it is found that the anticyclonic synoptic classes with which there are the greatest difficulties are those with WNW flow - possibly the difficult issue here lies in the fact that a slight shift in that flow (say, to the WSW, that is, from offshore to onshore) can result in a significant temperature difference.

The "strong ESE cyclonic" synoptic class is found to be the least difficult synoptic class. This is possibly the result of this situation being usually associated with a well developed cloud band and little risk of the flow turning offshore, on account of it often being followed by the development of an east coast low (Stern, 1979a, 1989). Finally, the tendency for some anticyclonic synoptic classes with onshore flow to be associated with a relatively high level of difficulty compared with that for cyclonic synoptic classes is depicted. This is possibly because, in anticyclonic situations, there is the absence of a warmer airmass ahead of the front (or trough) to enhance the land/sea thermal contrast and so to assist the onshore component of the flow.

There appears to be a strong tendency towards "bias" (under-forecast or over-forecast) in the case of many of the synoptic classes. This is depicted in Figs 8.5, 8.6 and 8.7 which, respectively, indicate by flow direction the number of over-estimates and under-estimates for light or weak, moderate, and strong classes. In summary, most of the errors with types having flow from the southerly direction, as well as the weak or moderate ENE (susceptible to sea breezes and upslide processes), are over-estimates. By contrast, most of the errors from other flow directions are under-estimates. Of particular note is that of the 64 errors over the 35 year period that occurred in association with one of the six classes with flow from the SSW, none of these errors were under-estimates. This may be a consequence of forecasters never seriously over-estimating the strength of a cool change, but occasionally under-estimating its strength, or even failing to predict its arrival at all.

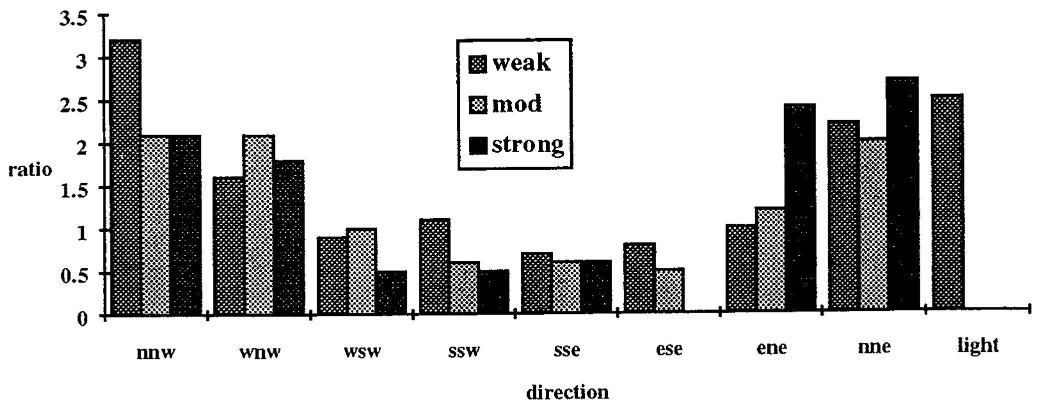


Fig 8.3 Forecast difficulty associated with cyclonic synoptic classes indicated by the ratio (% of summer errors)/(% of summer classes). Each group of three columns represents, from left to right, weak, moderate and strong synoptic classes (from Stern, 1996a&b).

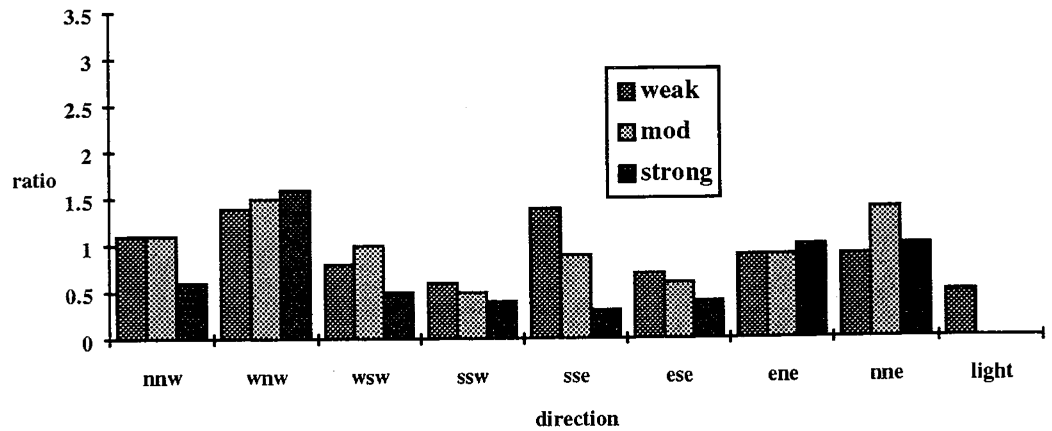


Fig 8.4 As for Fig 8.3, but for anticyclonic synoptic classes (from Stern, 1996a&b).

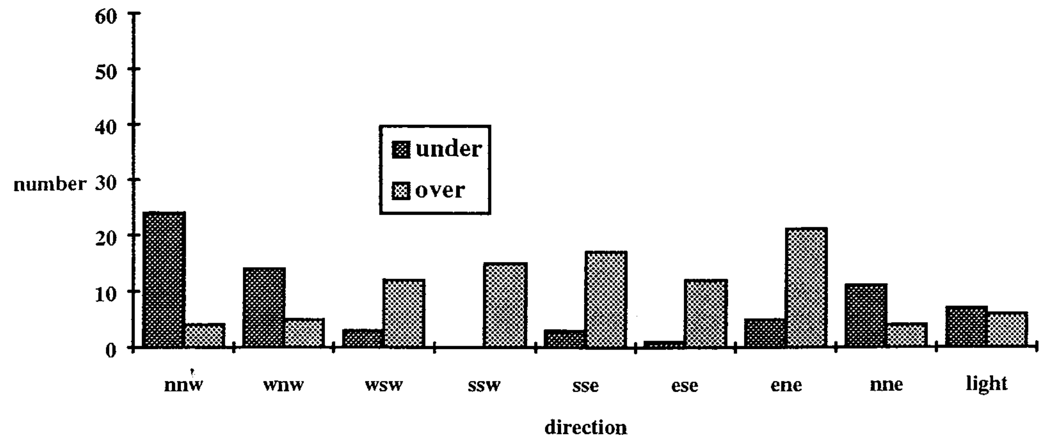


Fig 8.5 The number of over-estimates and under-estimates associated with weak or light synoptic classes (from Stern, 1996a&b).

An alternative approach to the analysis of the relationship between forecast difficulty and synoptic class is to employ the ratio depicted in Figs 8.3 and 8.4, but calculated for the previous day's synoptic class. In this approach, the synoptic class of the day on which the forecast is issued is used instead of the synoptic class of the day for which the forecast applies. As a result, the capacity of forecasters to predict the next day's synoptic class is taken into account. Figs 8.8 and 8.9 depict the values of that ratio for the previous day's cyclonic and anticyclonic classes and show that there is greater forecast difficulty with anticyclonic synoptic classes than with cyclonic synoptic classes, the reverse of the case with the "same" days' classes. This is possibly the result of cyclonic classes having mostly well established troughs and fronts, which have passed across the area by the morrow. Also shown is that the more difficult cyclonic synoptic classes are those light and with the flow from the NE quadrant - possibly on account of these having the least defined patterns, whereas the easier cyclonic patterns are those strong and from the southerly sector - probably a result of these patterns, once established, would almost certainly persist from that direction for at least another day.

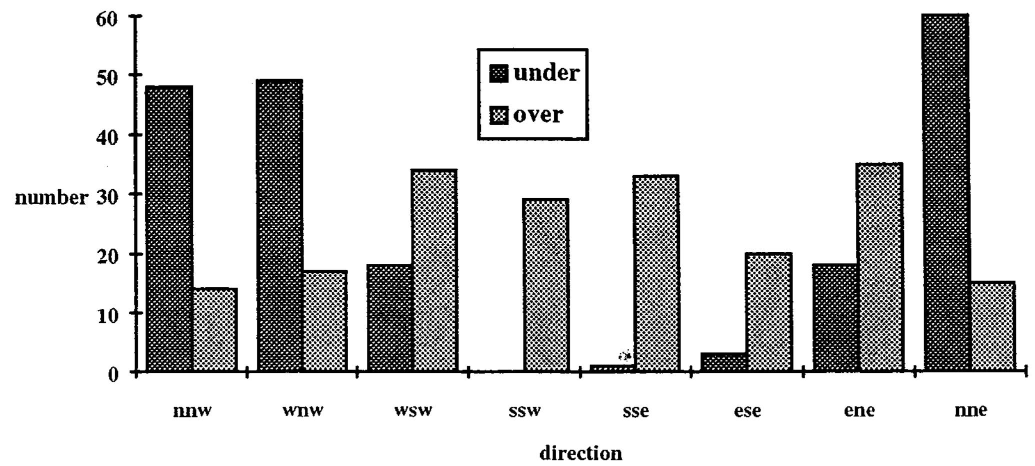


Fig 8.6 As for Fig 8.5 but for moderate synoptic classes (from Stern, 1996a&b).

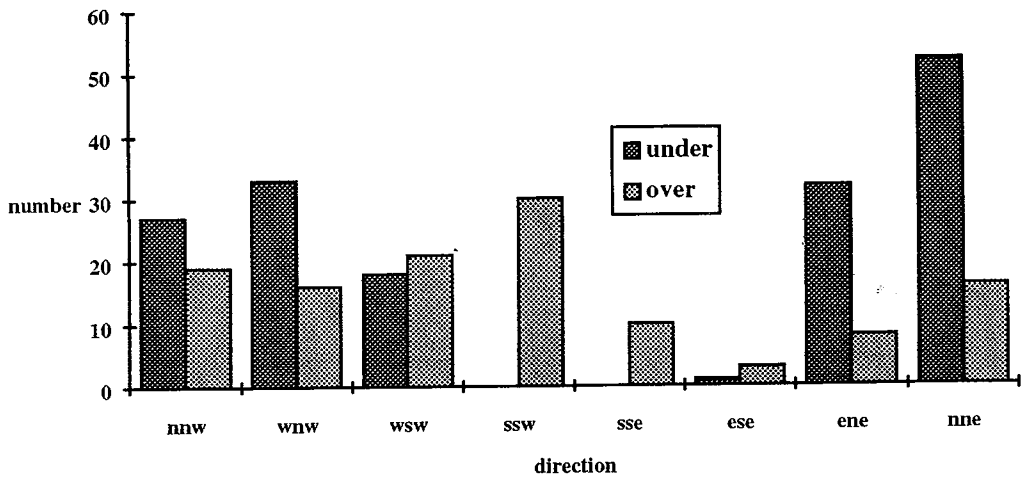


Fig 8.7 As for Fig 8.5 but for strong synoptic classes (from Stern, 1996a&b).

There also appears to be a tendency towards "bias" (under-forecast or over-forecast) in the case of some of the previous day's synoptic classes, although the effect is not as great. This is depicted in Figs 8.10, 8.11 and 8.12 which, respectively, indicate by flow direction the number of over-estimates and under-estimates for light or weak, moderate, and strong classes. In summary, most of the errors with types having flow from the northerly half and a moderate wind strength are over-estimates. Specifically, there is a tendency to over-estimate the next day's maximum temperature when the class on the day that the forecast is made is from the ENE and strong.

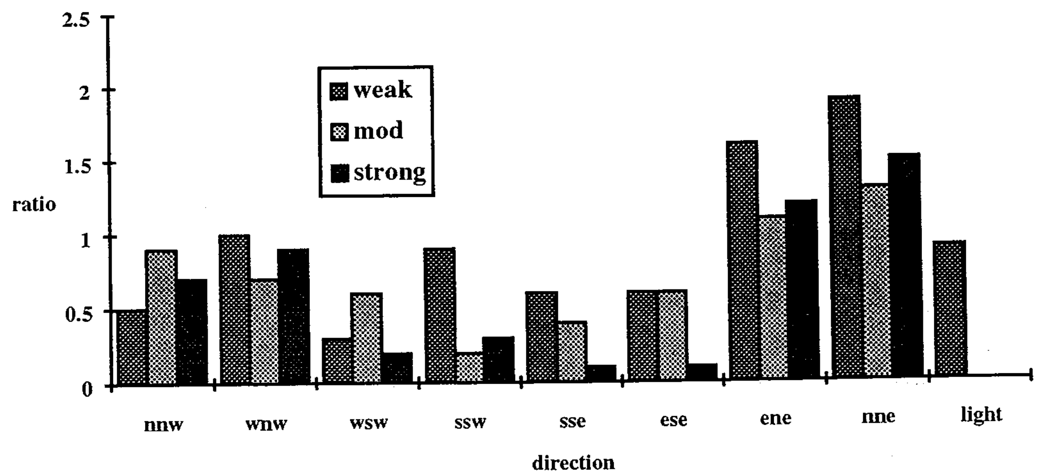


Fig 8.8 Forecast difficulty associated with the previous day's cyclonic synoptic classes indicated by ratio (% of summer errors)/(% of summer classes) (from Stern, 1996a&b). Each group of three columns represents, from left to right, weak, moderate and strong synoptic classes.

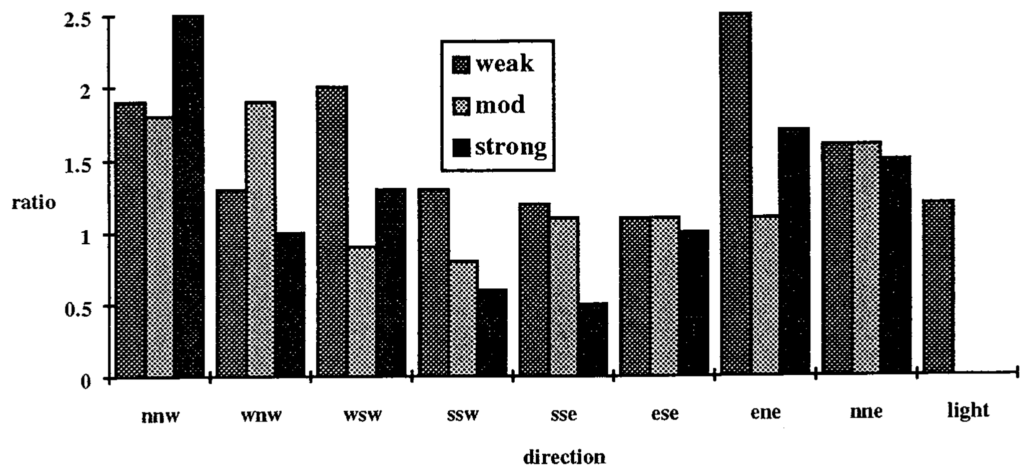


Fig 8.9 As for Fig 8.8 but for anticyclonic synoptic classes (from Stern, 1996a&b).

The historical trend is also synoptically dependent, as will be seen in the data presented shortly. The mean annual number of errors associated with moderate or strong wind synoptic classes falls from 25 per annum in the early 1970s to fewer than 10 per annum towards the end of the thirty five year period (Fig 8.13) - as errors accompanying moderate or strong wind synoptic classes form the overwhelming majority, the reasons given for the overall trend (early in this subsection) apply to these subsets of synoptic classes.

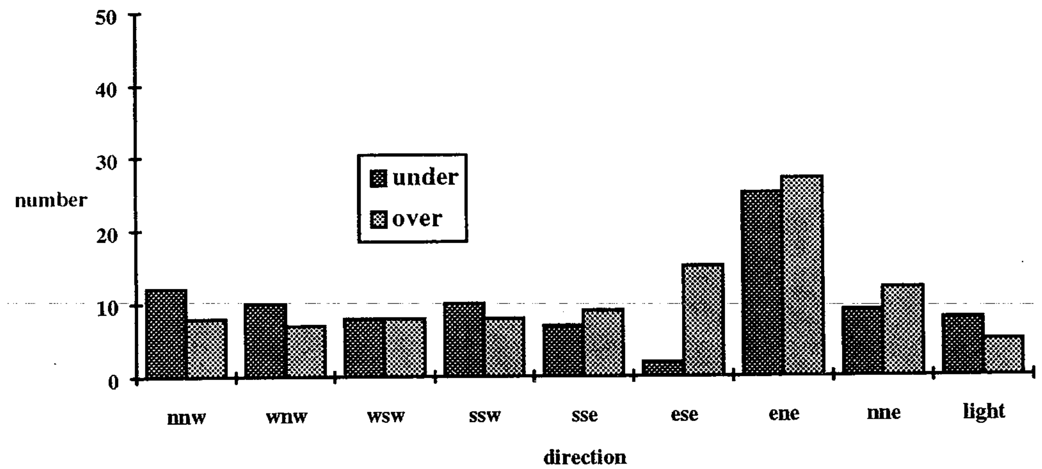


Fig 8.10 The number of over-estimates and under-estimates associated with the previous day's weak or light synoptic classes (from Stern, 1996a&b).

However, improvement has only recently become evident with light or weak wind synoptic classes. The trend with light or weak wind synoptic classes actually showed an extended period of deterioration - 4 per annum during the 1960s and early 1970s, rising to 6 per annum during the late 1970s and early 1980s, and back to 4 per annum since then (Fig 8.14). This may be due to there having been little increase in the understanding of mesoscale processes during a period when synoptic scale NWP has made considerable advances (late 1970s and early 1980s), these advances leading to a reduced focus on mesoscale aspects of weather forecasting. Improvement later in the period may be a result of the increased focus on mesoscale aspects of weather forecasting, such as the operational implementation of finer-scale NWP models and the analysis system MCIDAS (both of which could have had an impact on sea breeze prediction and the forecasting of other mesoscale phenomena), such focus occurring in the wake of devastating bushfires in 1983.

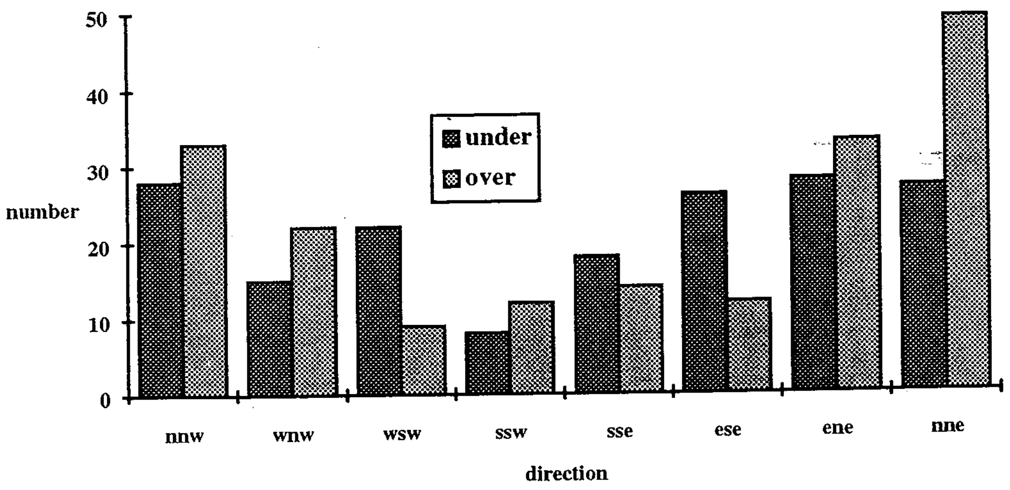


Fig 8.11 As for Fig 8.10 but for moderate synoptic classes (from Stern, 1996a&b).

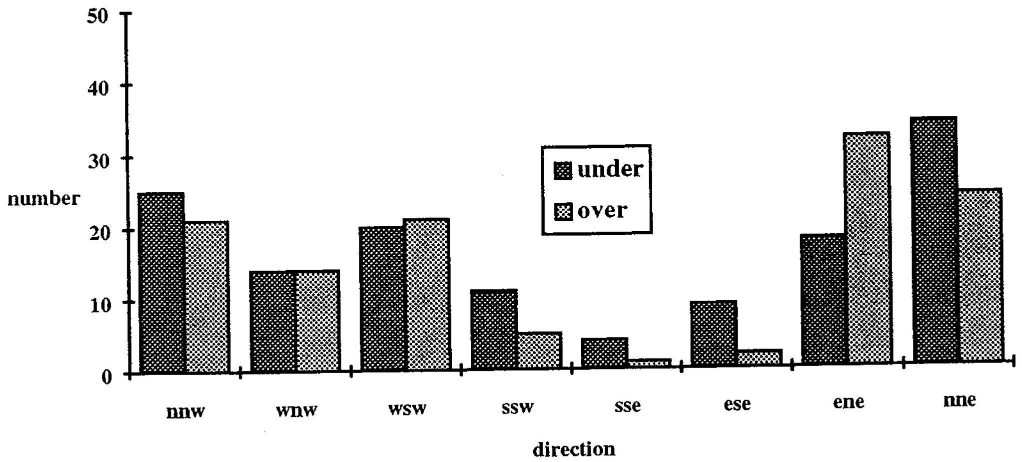


Fig 8.12 As for Fig 8.10 but for strong synoptic classes (from Stern, 1996a&b).

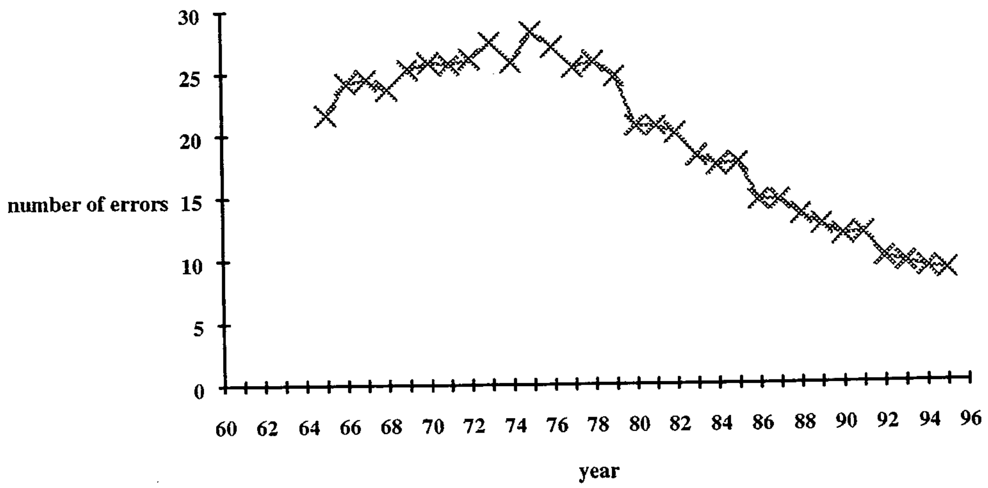


Fig 8.13 The trend in the annual number of Melbourne maximum temperature errors associated with moderate or strong wind synoptic classes, illustrated by 5-year running means ended in June of indicated year (from Stern, 1996a).

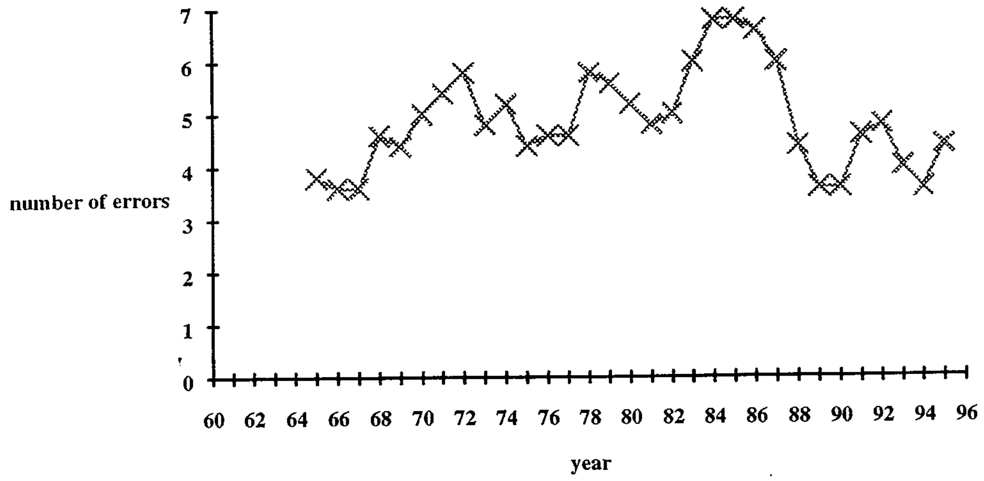


Fig 8.14. As for Fig 8.13, but for weak or light wind synoptic classes (from Stern, 1996a).

Section 3's perspective regarding rainfall forecasts contrasts with encouraging conclusions obtained regarding trends in the accuracy of maximum temperature forecasts, particularly those associated with synoptic classes accompanied by broad scale forcing; and also the accuracy of long lead time aviation forecasts, that is, those where broad scale forcing is important. What is emerging is a conclusion that where the physics of the weather producing processes has to be considered in terms of dynamic theory (that is, precipitation processes and meso-scale processes accompanying maximum temperature forecasting in weak or light wind synoptic classes, and short-term aviation forecasts), that theoretical understanding has, so far, not been realised in terms of a demonstrable improvement in the accuracy of weather forecasts. There is this lack of realisation, notwithstanding the computing resources being devoted to models attempting to reproduce mathematically that physical dynamic theory in a weather forecasting context. It is suggested that the "missing link" to connect the improved theoretical understanding to an improved forecasting performance might be found statistically.

That the trend in aviation-weather forecast performance is best (or "least worst") for the longest lead time seems to partially match the improvement (since the early 1970s) in the case of temperature forecasts for moderate or strong wind synoptic classes (where broad scale forces dominate) and may be explained in similar terms. The trend towards deteriorating forecast performance in shorter lead time predictions seems to partially match the deterioration in the case of temperature forecasts for light or weak wind synoptic classes (where meso-scale forces dominate) and may also be explained in similar terms. It could also be speculated that the introduction of the "Trend Type Forecast" (TTF), just prior to the acceleration in the deterioration of the 1-7 hr lead time predictions, which allowed deficiencies in the first 3 hours of the Terminal Airport Forecast (TAF) to be catered for without issuing an amendment, diminished the perceived requirement for accurate shorter term TAFs (Shanahan, 1995).

9. MAXIMISING THE PERFORMANCE OF AUTOMATED GUIDANCE

Stern (1996a) illustrates how optimal operating conditions of automated guidance might be determined on an interactive basis, using several examples. The examples relate to four grids of Mean Sea Level Pressure values for analogue retrieval, operating the ASM in both the initial data (ID) mode (set of analogues to initial synoptic pattern retrieved), in accordance with the Classical Statistical Method (CSM) and the Perfect Prog (PP) mode (set of analogues to forecast synoptic pattern retrieved). The four grids are, firstly, a grid covering an area extending from Forrest, near the Head of the Bight across southeastern Australia, secondly, a grid covering southern Australia, thirdly, a grid covering Victoria, Tasmania and the adjacent oceans, and, fourthly, a grid covering both Southern Australia and the oceans further to the south. Performances of the ASM are evaluated at predicting a range of weather elements for Melbourne. These weather elements are maximum and minimum dry bulb temperatures, 9am and 3pm dewpoint temperatures, POPs and QPFs, probability of thunderstorms and fogs, as well as forecasts of wind strength and cloud amount. Table 9.1 presents the results of an evaluation conducted in this manner. Further details of the analyses are presented by Stern (1996a).

Table 9.1 Comparative evaluation of the ASM performances, as measured by RMS Error (°C) for maximum and minimum dry bulb temperatures (MAX, MIN), 9am and 3pm dewpoint temperatures (DP9, DP3), Brier Score for POPs/RMS Error for QPFs (PRECIP), Brier Score for probability of thunder and fog (THUN, FOG), and maximum wind gust (knots) and cloud amount (WIND, CLOUD), for, (a) Employing the four grids operated in both initial data (ID) and perfect prog (PP) modes (1991 - 1993), (b) Combining results from ID (CSM) and PP modes and (c) Corresponding official VRO forecasts[1] (from Stern, 1996a).

(a)	MAX	MIN	DP9	DP3	PRECIP	THUN	FOG	WIND	CLOUD
ID1	3.02	1.78	2.30	2.92	.212/.874	.0196	.0152	5.98	2.40
ID2	2.83	1.78	2.32	2.84	.207/.876	.0196	.0156	5.86	2.36
ID3	3.21	1.73	2.31	2.98	.219/.883	.0198	.0156	5.81	2.41
ID4	2.68	1.79	2.25	2.82	.207/.881	.0200	.0160	5.61	2.34
PP1	2.21	1.75	2.31	2.79	.179/.793	.0194	.0154	4.18	2.26
PP2	2.53	1.83	2.33	2.91	.183/.800	.0196	.0152	4.54	2.27
PP3	2.18	1.75	2.28	2.83	.177/.791	.0194	.0156	4.13	2.23
PP4	2.53	1.84	2.33	2.82	.180/.798	.0202	.0158	4.47	2.22

(b)	MAX	MIN	DP9	DP3	PRECIP	THUN	FOG	WIND	CLOUD
ID1-4	2.94	1.77	2.30	2.89	.211/.879	.0197	.0156	5.82	2.38
PP1-4	2.37	1.79	2.31	2.84	.180/.796	.0197	.0156	4.33	2.25

(c)	MAX	MIN	DP9	DP3	PRECIP	THUN	FOG	WIND	CLOUD
VRO	2.22	1.68	n/a	n/a	n/a	n/a	n/a	n/a	n/a

10. EVALUATING INDIVIDUAL FORECASTER CHARACTERISTICS

Knowledge, reputation, experience and qualifications of the human forecaster are assumed to impact upon the accuracy of the predictions, and the nature of such impacts has been the subject of much research. Sanders (1973) compared the skill

[1]VRO forecast data for maximum and minimum temperatures on 10 Oct 1991, 11 Oct 1991 and 26 Oct 1993 not available; no forecast data for other elements available.

displayed by students and professors at weather forecasting and detected no significant difference, finding that "few if any individuals who made a substantial number of forecasts outperformed consensus on the average". This seems to be confirmed by the results of similar studies conducted by Gedzelman (1978) and Bosart (1983). Gregg (1969) evaluated comparative skill displayed by a group of experienced weather forecasters. In this context, an experienced weather forecaster was defined as one who had worked in the profession for more than four years. Gregg (1969) found little difference in the skill displayed by members of the group, attributing that lack of difference to the forecasters having access to the same tools, to the same data and to them using the same techniques. Ramage (1993) evaluated a group of experienced forecasters at the Royal Observatory in Hong Kong, evaluating each forecaster separately and also "could find no difference (in skill)". Ramage writes that "statistics ... (have) threatened ... (the) belief that requiring forecasters to have college degrees would lead to better forecasts" while Roebber and Bosart (1996), who studied the relative contribution of education and experience to forecast skill, found that "forecast skill is largely determined by experience". The results of a less rigorous investigation is also interesting. Ryan (1987), in an evaluation of the relative performance of BoM forecasters at the VRO RFC did not find that their skill was stratified according to qualifications (in fact, a Technical Officer, without a University Degree, ranked second).

Dickinson (1995) and Smith (1995) have been involved in a project aimed at evaluating weather forecasting performances at the VRO RFC. One of the project's aims was to investigate skill at predicting the probability of precipitation (POP) at Melbourne during three different time periods, interest having been generated in this topic by the work of Fraedrich and Leslie (1987), which, in turn, followed Miller and Leslie's (1984 & 1985) work. The periods for which the POP estimates apply, and corresponding issue times, are presented in Table 10.1.

Table 10.1 Periods for which the POP estimates apply, and corresponding issue times (local time) (from Stern, 1996a).

Period	Issue Times
2000 (previous evening)	0600-1800
0600 (same day)	0600-1200
0600 (same day)	1200-1800

The verification of forecasts expressed in terms of probability presents some difficulties (Brier, 1950; Brier and Allen 1951; Epstein, 1969a; Murphy, 1971, 1973 & 1977) as also does their interpretation (Murphy and Winkler, 1979 & 1992; Murphy *et al.*, 1980; Murphy and Brown, 1983). Notwithstanding these difficulties, when outcomes over a 42 month period, encompassing a sample size of 2734 POP forecasts between October 1991 to March 1994, were analysed, a number of interesting results emerged. The group undertaking the trial, involving experienced shift supervisor meteorologists, displayed a very slight bias towards over-forecasting the POP. Illustrating this point is the result of the verification exercise showing that the mean of the 2734 POPs was 31.4%, compared with precipitation being observed on 30.2% of the corresponding periods - a difference of only 1.2%. However, employing a two-tail test on the difference between the means showed that the difference was not significant at the 5% level, and Fig 10.1

clearly shows that, overall, the forecasters displayed considerable skill at predicting POPs.

One of the group was a meteorologist, whose reputed bias towards predicting dry weather was so well entrenched in the minds of colleagues that he was known as the "optimist". However, the "optimist" was found to exhibit in his POPs no less pessimism than the rest of the group. In fact, the data suggests that the "optimist" might be slightly more "pessimistic" than the remainder of the group, taken as a whole. Illustrating this point is the result of the verification exercise showing that the mean of his 520 POPs being 32.0% compared with precipitation being observed on 30.0% of the corresponding periods - a difference of 2.0% (corresponding data for his colleagues taken as a separate group being 31.2%, 30.3% and 0.9%). However, a two-tail test on the difference between the means showed that the difference was not significant at the 5% level.

The bias of the group varied across the range of forecast POPs (Fig 10.1). When POPs of less than 50% were predicted, the mean of the 1813 POPs was 10.8%, compared with precipitation being observed on 14.6% of occasions, the difference of 3.8% representing a bias towards under-forecasting the POP. When POPs of 50% or more were predicted, the mean of the 921 POPs was 71.9%, compared with precipitation being observed on 61.0% of the corresponding periods, the difference of 10.9% representing a bias towards over-forecasting the POP. The bias of the "optimist" across the range of forecast POPs displayed characteristics similar to those of his colleagues (Fig 10.2). When POPs of less than 50% were predicted by the "optimist", the mean of his 341 POPs was 8.9% compared with precipitation being observed on 13.2% of occasions, a bias towards under-forecasting of 4.3% (corresponding data for his colleagues taken as a separate group being 11.2%, 14.9% and 3.7%). When POPs of 50% or more were predicted by "the optimist", the mean of his 179 POPs was 75.9%, compared with precipitation being observed on 62.0% of the corresponding periods, a bias towards over-forecasting of 13.9% (corresponding data for his colleagues taken as a separate group being 70.9%, 60.8% and 10.1%).

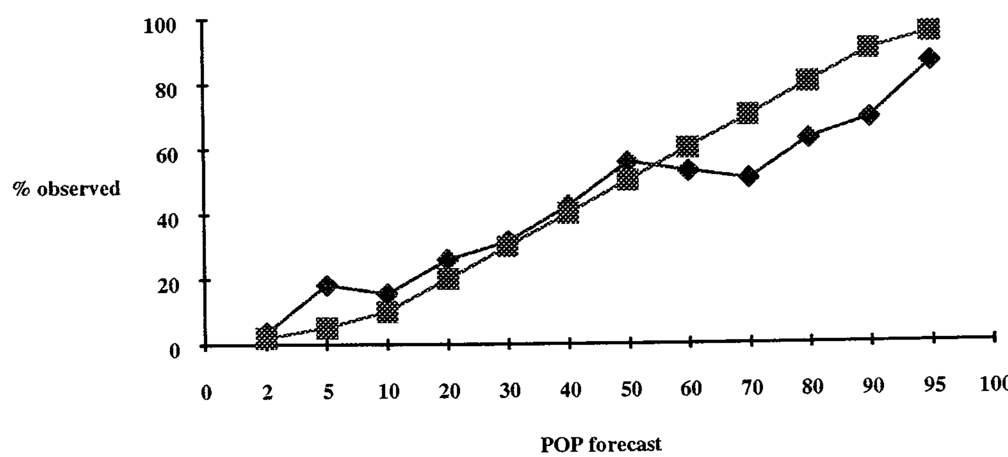


Fig 10.1 Percentage of occasions that precipitation was observed when various POPs forecast (diamonds) compared with a perfect set of forecasts (squares). The scales are somewhat stretched at the extremities to better depict the performance of the very high and very low POP forecasts (from Stern, 1996a).

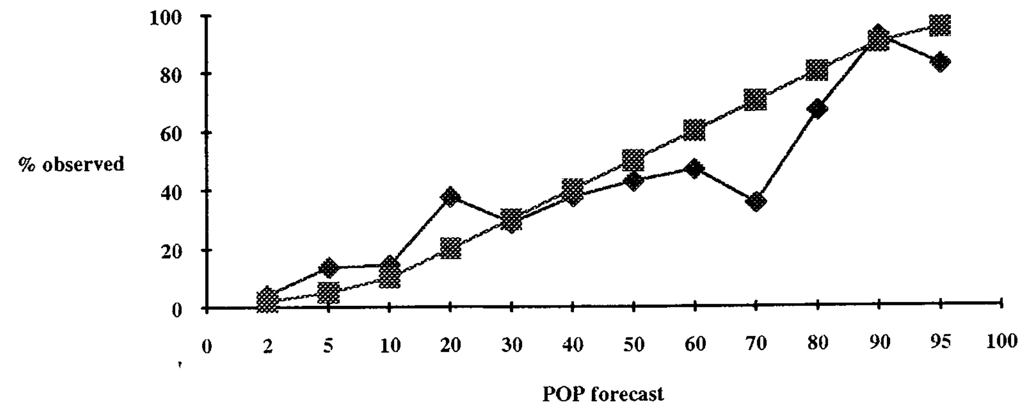


Fig 10.2 Percentage of occasions that precipitation was observed when various POPs forecast by "the optimist" (diamonds) compared with a perfect set of forecasts (squares). The scales are somewhat stretched at the extremities to better depict the performance of the very high and very low POP forecasts (from Stern, 1996a).

11. EMOTIONS AND OTHER INFLUENCES ON PREDICTIONS

The purpose of this Section is to discuss how human emotions and other influences upon the prediction process can act to adversely effect the accuracy of a prediction. Allen (1981) employs decision analysis theory "as a means to describe, understand and improve decision strategies used in weather forecasting". He observes that the human weather forecaster's decisions "are influenced by a variety of factors including tradition, ease of access to information and information processing resources, time constraints, perceived penalties and perceived payoffs". Allen notes that "some of these factors lead to the use of sub-optimal decision strategies ... (with the consequence that) ... present levels of forecast accuracy are lower than they might be. Section 11 includes a study of the impact of an error on the subsequent decision strategies of forecasters.

An error in maximum temperature is defined as "more than a 5°C difference between the observed maximum temperature and the previous afternoon's forecast". A sub-set of Melbourne maximum temperature errors, those associated with an error the previous day, is analysed. The analysis is illustrated in Fig. 11.1, which presents frequency distributions of errors when there was an error on the previous day. Over a 35 year period from 1960 to 1995, there were 114 instances of errors on consecutive days. Of these pairs, there were 70 cases (61%) of the sign of the error reversing, that is, an over-estimate followed by an under-estimate or an under-estimate followed by an over-estimate. This proportion is shown, by applying a two-tail test, to be different from the "expected" 50% at the 5% level of significance. In order to explain this apparent anomaly, it is instructive to study the example of the error of 16 February, 1983, and to relate the anomaly to the behaviour of stock market investors.

The anomaly may be attributed to forecasters "over-compensating" for the preceding error, that is, placing a high degree of focus on the previous error and having a concern that it may be repeated even if there is no real basis for having that concern. For example, on the day of one of Victoria's most destructive bushfires (16 February, 1983) the maximum temperature prediction was an under-estimation (observed 43°C, predicted 34°C). This under-estimation was preceded by an over-estimation for the preceding day, 15 February, 1983 (observed 31°C, predicted 38°C). Concern about the error of 15 February 1983, in combination with uncertainty as to whether cloud observed on 15 February would persist into the next day, may have resulted in inadequate weight being given to the potential for very high temperatures suggested by the Australian Region Primitive Equation (ARPE) NWP model (McGregor *et al.*, 1978) output for the 16th. The ARPE output for 9am on 16 February 1983, is of the synoptic class "moderate WNW anticyclonic" and suggests an 850 hPa temperature near Melbourne of about 23° C. Retrospective interpretation of the ARPE model output by the ASM indicates a likely maximum temperature at Melbourne, in this situation, of 41.8°C.

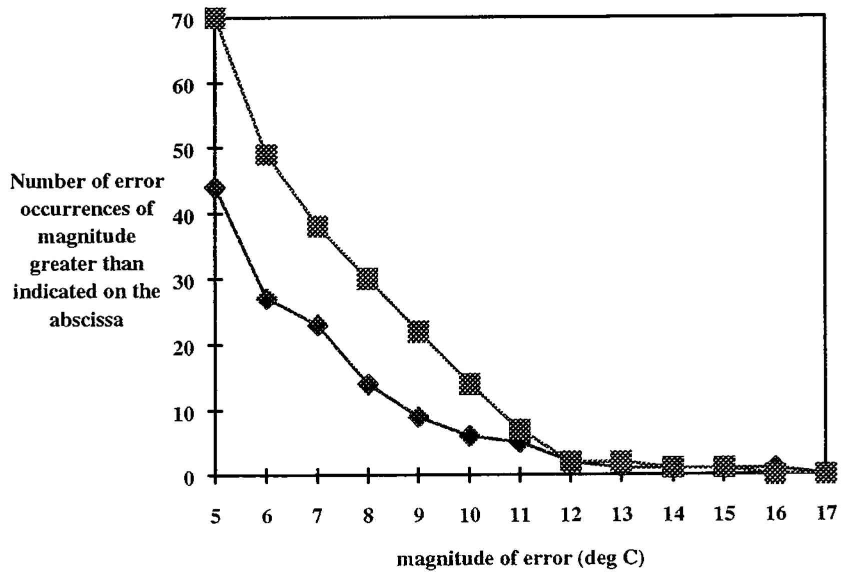


Fig 11.1 Frequency distributions of errors when there was also an error on the previous day, divided into (1) cases when the sign of the error on the second day was the same as that of the first (diamonds) and (2) cases when the sign was opposite (squares) (from Stern (1996a&b).

Over-compensation for recent events is present in other fields of prediction. An illustration is presented using a 120 year data set of "turning points" in the Australian stock market. It is shown that over-compensating for previous events causes reversals in trends of prices. Taking monthly average prices on the Australian stock market since 1875 (Australian Stock Exchange (ASX) Ltd, 1991), there have been 64 reversals in excess of 10% (natural logarithm) over the 120 year period. On average, each reversal occurs at prices 19% higher than the preceding reversal. If one assumes that the distribution of reversal levels (compared to the previous reversal level) is normal, then the frequency distribution of reversal occurrences is expected to be as indicated by the line connecting the squares in Fig 11.2. However, the observed distribution is as indicated by the line connecting the diamonds in Fig. 11.2, which peaks at 47% when the % change (since the previous reversal) is zero. That 47% proves to be significantly greater, at the 5% level of significance, using a one-tail test, than the 33% expected with a normal distribution. It is suggested that the skewness, evident in the distribution of the turning points, arises because investors over-focus on the price of the preceding reversal. This over-focus on the preceding reversal leads to the price of the next reversal being closer, than otherwise would be the case, to the price of the preceding reversal. This occurs, even though that price (of the next reversal) may bear no relation to the "real" value of the stock market at the time of its occurrence.

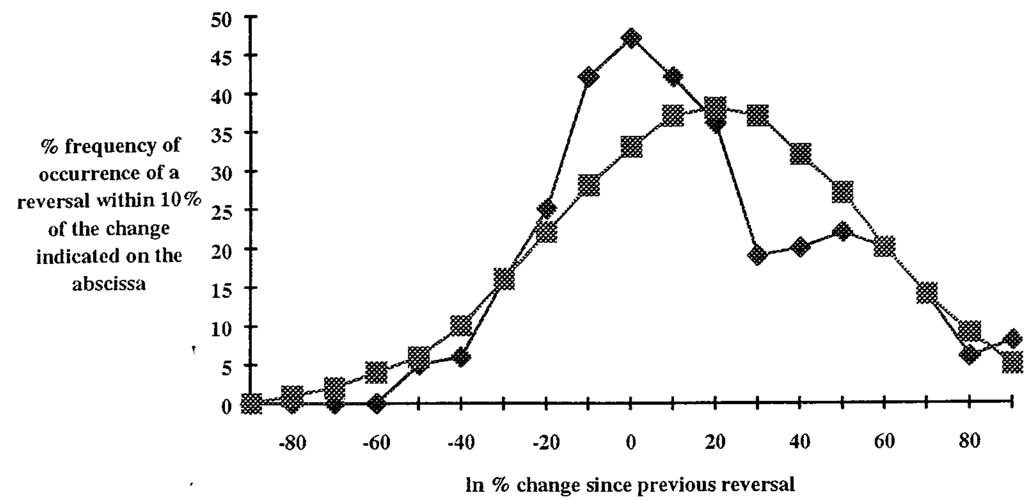


Fig 11.2 Frequency distribution of reversal occurrence level compared to that of the previous reversal level. Line connected by diamonds is the observed distribution. Line connected by squares assumes a normal distribution (from Stern, 1996a&b).

There are other potentially adverse influences the accuracy of a forecast. Sometimes biases are introduced. For example, the impact of requiring to divert aircraft leads to a conservative approach to forecasting at terminals. This conservative approach is illustrated by Fig 11.3, which shows the mean probability of detection (POD) and false alarm ratios (FAR) of aviation significant weather at Melbourne Airport. With the exception of the shortest time period, the FAR is greater than the POD.

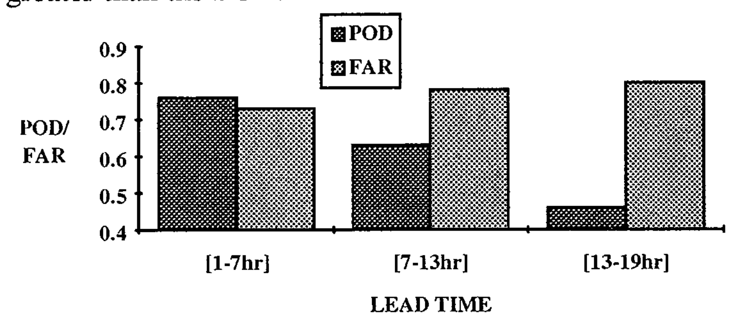


Fig 11.3 The probability of detection (POD) and false alarm ratio (FAR) for aviation significant weather at Melbourne Airport for various forecast lead times (1972-1990) (from Stern, 1996a).

12. CONCLUSIONS

Firstly, a statistical analysis of a thirty five year record of Melbourne maximum temperature forecast errors has been presented. It has been shown there has been an overall trend towards improvement in maximum temperature forecasting, this trend being less evident with weaker gradient synoptic classes. That the trend is less evident with such classes may be a consequence of advances in NWP models impacting upon forecasts of the broadscale flow (these NWP model forecasts interpreted by statistically based forecast guidance), but that when one is dealing with a situation where mesoscale effects are important, there has not been such advances in knowledge or techniques.

That this trend towards improvement extends to outlooks, with 2-day outlooks of maximum temperature now possessing a superior level of accuracy to that of the 1-day temperature forecasts 25 years ago and the accuracy of 3-day and 4-day outlooks approaching that level of accuracy, may be a consequence of advances in broadscale NWP techniques, reflected in the improved performance of global models.

Secondly, it is shown that the picture for other weather elements is less encouraging. The scant improvement being evident in predicting rainfall or short-term forecasts of aviation-significant weather, may be a consequence of there having been few advances in the understanding of the physical processes involved. It underlines the need for better understanding of the theory associated with the mesoscale.

With regard to the human role in forecasting a range of studies by a number of authors demonstrate that forecast skill is more related to experience than to theoretical knowledge, as measured by qualifications. It is relevant to these studies, that the accuracy of Melbourne maximum temperature forecasts associated with synoptic situations where broadscale forcing was important temporarily declined at the same time as the departure of a group of experienced forecasters. It is also relevant that there was a subsequent increase in accuracy as the replacement group gained experience. Given that throughout the entire period under consideration predictions of broadscale synoptic flow improved, it is suggested that the component of experience that is important in forecasting is that of interpreting the broadscale flow in terms of local weather elements (in this particular case, temperature). Because this "experience" can be readily incorporated into a statistically based forecast guidance scheme, there is potential for accuracy increase by extending the application of such forecast guidance schemes. Indeed, interpreting local weather in terms of the broadscale is readily automated by statistical methodologies. Perhaps, this is what Brooks (1995) is driving at, when he writes that "technology, which initially allowed humans to make routine weather forecasts, will soon close that avenue of human endeavour (and thereby permit) concentration on severe events. It is likely that those of us working in meteorology in the second half of the 20th century have seen a kind of human involvement in forecasting that will be open only briefly (Brooks, 1995). That this prediction by Brooks may soon become a reality is supported by Fig 12.1. Fig 12.1 suggests that, overall, whereas human forecasters are capable of significantly improving upon computer generated guidance for short-term predictions, that capability is much reduced for long-term predictions. However, Fig 12.2 indicates that where forecasters focus on a particular location, as they might be expected to do for Melbourne, the State capital, that capability is somewhat preserved. Furthermore, in another study of Melbourne maximum temperature forecast data (Stern, 1994, 1996a) found that forecasters in the VRO RFC possess skill at discriminating between good and flawed NWP output.

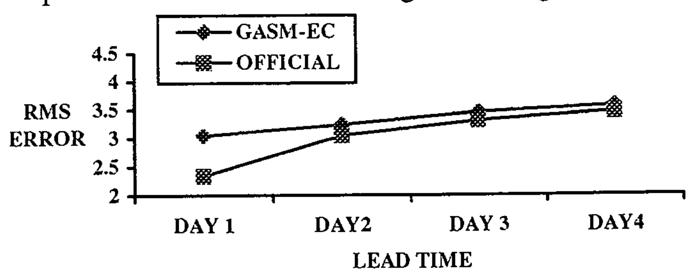


Fig 12.1 Chart illustrating the extent that January 1997 to June 1997 VRO RFC official maximum temperature forecasts, for Melbourne and sixteen other Victorian centres, issued one, two, three and four days ahead, improved upon the guidance generated by GASM using the ECMWF global model output (Setek, 1997).

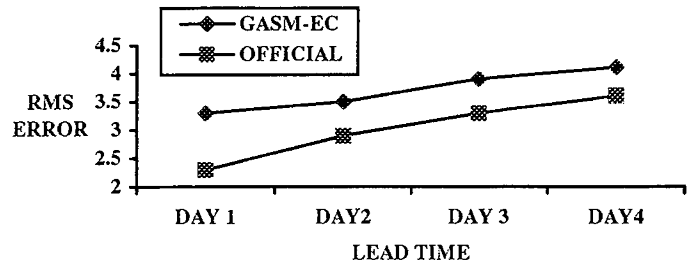


Fig 12.2 Chart illustrating the extent that January 1997 to June 1997 VRO RFC official maximum temperature forecasts, for Melbourne alone, issued one, two, three and four days ahead, improved upon the guidance generated by GASM using the ECMWF global model output (Setek, 1997).

It is also interesting that Noone and Stern (1995) have also noted that the GASP NWP performed best at predicting rainfall associated with larger scale systems. These factors all underline the need for better understanding of the theory associated with the mesoscale.

The results of a study of predictions prepared by experienced forecasters showed that subjective perceptions of forecasters' skill profiles are sometimes later proven to be incorrect. The implication of this finding is that individual forecasters have the potential to improve on their performance if their "true" skill profile was made available to them. It was also noted that human emotions have a major adverse impact on forecasting. The implication of that finding is that if related systematic errors can be identified and eliminated, there is also potential for an improved forecasting performance.

Historical trends in forecast accuracy are a function of a range of factors and, in the case of some weather elements, periods of declining accuracy have occurred. It is suggested that those declines might be a consequence of the lack of a systematic approach, such as that proposed by Ramage (1993), meaning that temporary gains are sometimes later lost through no clear understanding of what caused the gains in the first place.

Using techniques for synoptic clustering it has been shown how understanding of atmospheric processes and forecasting errors may be enhanced. Using the ASM in its PP and CSM modes, data from the analogues may be statistically analysed to yield quantitative and qualitative aspects of a weather forecast. The ASM's operating conditions could also be determined on the basis of what synoptic class grouping the current synoptic situation is a member of. Because of the general and moveable nature of the synoptic classes, being simply defined by the direction, strength and curvature of the surface flow, the conditions can be progressively modified by a feedback process linked to verification of the guidance. In summary, it is proposed that a system of weather forecasting guidance might be devised that fully integrates verification outcomes.

It is proposed that the area of forecasting likely to be particularly sensitive to the accuracy improvement capable of being delivered by statistical methods is that of the short term prediction of weather elements which the aviation industry finds significant. The reasons why that area of weather forecasting is so nominated are that present accuracy levels of such predictions are inferior to those of the 1970s and that the only period of temporary improvement during the 1980s can be related to the development of statistically based semi-objective forecasting aids.

That the present level of accuracy is inferior to that of the 1970s should not necessarily be taken to mean that one adopts the step of returning to the procedures and practices of two decades ago. At that time, forecasts for the aviation industry were prepared at Melbourne Airport whereas most are currently prepared in the RFC in the city. Although such a return might be based upon providing a sharper focus on the forecasting problems at the Airport, there is another approach that might yield more succcessful and more long-lasting returns.

This other approach is to extend the ASM to cover the prediction of aviation sensitive weather elements. A Classical Statistical Method (CSM) based component could be incorporated into the system (as with the "pilot" model or using Markov Chains). This is because short term forecasting of these elements is likely to benefit most from a CSM approach, for inherent in the short term forecasting of aviation significant weather is a reduced requirement for accurate prediction of the broadscale evolution. This other approach is proposed because the attempts at the development of statistical forecasting methods for Melbourne Airport in the past that led to the only significant upturn in forecast performance during the verification period were, in any case, constrained by the greater part of the historical half hourly meteorological reports not having been computer archived. Research suggests that forecaster skill is largely related to forecaster experience. This "experience" can be incorporated into a statistically based forecast guidance scheme. As a result, taking the simple step of computer archiving those data should allow the development of statistically based guidance possessing the sophistication, and having the impact on forecast accuracy of, for example, that which the temperature forecast guidance provided by the ASM did.

There are other areas of future work suggested by the major findings. In regard to implementing the "positive feedback" approach to forecasting technique development in order to preserve gains, these are firstly, to develop skill profiles of individual forecasters with a view to "feeding back" that information into the forecast process and secondly, to conduct studies to identify overall systematic bias that might be attributed to the emotions of human forecasters. Furthermore, it is proposed that there be a test conducted to establish whether the current moves towards standardising the characteristics of NWP components will diminish the benefit to a forecasting office of receiving a range of different NWP models' output by reducing their differences from each other and, hence, also reducing the potential gain from applying methods that optimally combine their outputs.

In conclusion, the consequences of taking advantage of the technological opportunities created by systems such as AIFS, would be both an increase in the accuracy of weather forecasts as well as an increase in theoretical understanding. This paper opened with Ramage's (1993) proposal that weather forecasting might be the cement to bring about communication between the theoretical and practical sides of meteorology. I now close with a similar proposal from Doswell (1995):

"Verification allows forecasters to know, quantitatively and objectively, how well they are doing, and in what ways they can improve their product. In these days of "total quality management" and "quality assurance," what company would sell a product whose quality went unchecked? Today's quality-conscious market wouldn't tolerate that for long. Yet we continue to do pitifully inadequate verification of our forecasts, public and ... we do virtually nothing systematic to follow-up that verification seeking to find out how to improve our forecasts. This latter task, to which I refer as "loop-closing," is essential and yet remains virtually ignored in the forecasting business. How often do forecasters get to go back and re-evaluate a situation gone bad with an eye to preventing similar occurrences in the future? ... Virtually never. In my view of this topic, it's absurd to think that a well-run company producing specialized forecasts would not have an aggressive quality management program" (Doswell, 1995).

ACKNOWLEDGEMENT

The author received considerable assistance from a great number of people, both at the Melbourne University School of Earth Sciences and the Bureau of Meteorology, while carrying out the various pieces of work summarised in this paper.

REFERENCES

Allen GA (1981) Forecaster skill development. *Meteor. Note 123, Bur. Met., Australia,* 21pp.

Australian Stock·Exchange (ASX) Ltd (1991) Indices and yields book, 90 pp.

Barnston AG and Smith TM (1996) Statistical prediction of global continental temperature and precipitation to a year in advance using global SST. *Preprints, 13th Conf. on Probability and Statistics in the Atmospheric Sciences, 21-23 February 1996, San Fransisco, Amer. Meteor. Soc.*

Barclay PA and Butt IC (1988) Automation of regional forecasting centres in Australia. *Preprints, 4th Int. Conf. on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology. Anaheim, Calif., 1-5 February 1988, Amer. Meteor. Soc., 21-24.*

BoM (1977) Recording and encoding weather observations.

BoM (1989) Explanatory Notes. Seasonal Climate Outlook.

Bosart LF (1983) An update on trends in skill of daily forecasts of temperature and precipitation at the State university of New York at Albany. *Bull. Amer. Meteor. Soc.,* 64:346-354.

Brier GW (1950) Verification of forecasts expressed in terms of probabilities. *Mon. Wea.. Rev.,* 78:1-3.

Brier GW and Allen RA (1951) Verification of weather forecasts. In *Compendium of Meteorology, T F Malone ed., Amer. Meteor. Soc.,* 841-848.

Brooks H (1995) Human forecasters and technology. From *http://www.nssl.uoknor.edu/~brooksl*

Colls K Mason IB and Daw FA (1981) A forecast verification for public weather forecasts. *Aust. Meteor. Mag.,* 29:9-23.

Dahni RR (1988) The development of an operational analogue statistics model to produce weather forecast guidance. *Ph. D. Thesis, Department of Meteorology, University of Melbourne, Australia.*

Dahni RR (1995) Techniques development program (September to December 1995). *Attachment to Gigliotti (1995b).*

Dahni RR (1996a) Generalised Analogue Statistics Model Version 1.8 Reference Manual. *Internal Bur. Met., Australia Doc.*

Dahni RR (1996b) Forecast Guidance System User Guide. *Internal Bur. Met., Australia Doc.*

Dahni RR (1996c) Map Archive User Guide. *Internal Bur. Met., Australia Doc.*

Dahni RR and Stern H (1995) The development of a generalised UNIX version of the Victorian Regional Office's operational analogue statistics model. *BMRC Res. Rep. 47, Bur. Met., Australia,* 34 pp.

Dahni RR de la Lande J and Stern H (1984) Testing of an operational statistical forecast guidance system. *Aust. Met. Mag.,* 32:105-106.

Dickinson K (1995) *Personal communication.*

Dobryshman (1972) Review of forecast verification techniques. *WMO Tech. Note No. 120.*

Doswell CA 3rd Davies-Jones R and Keller DL (1990) On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting,* 1:337-345.

Doswell CA 3rd (1995) Meteorology and users. From *http://www.nssl.uoknor.edu/~doswell/*

Ehrendorfer M and Murphy AH (1988) Comparative evaluation of weather forecasting systems: sufficiency, quality and accuracy. *Mon. Wea. Rev.,* 116:1757-1770.

Epstein ES (1969) A scoring system for probability forecasts of ranked categories. *J. Appl. Met.,* 8:985-987.

Finley JP (1884) Tornado prediction. *Amer. Meteor. J.,* 1:85-88.

Foley JC (1947) A study meteorological conditions associated with bush and grass protection strategy in Australia. *Bull. 38, Bur. Met., Australia,* 234 pp.

Fraedrich K and Leslie LM (1987) Evaluation of techniques for the operational, single station, short-term forecasting of rainfall at a midlatitude station (Melbourne). *Mon. Wea. Rev.,* 115:1645-1654.

Gedzelman SD (1978) Forecasting skill of beginners. *Bull. Amer. Meteor. Soc.*, 59:1305-1309.
Gigliotti P (1995) AIFS progress report 7 September 1995. *Internal Bur. Met., Australia Doc.*
Goodhead HM (1978) An objective method for predicting fog and mist at Melbourne Airport. *Airmet Papers, Canberra Meteor. Soc., Roy. Meteor. Soc. (Aust. Branch), Canberra 1978*, 27-38.
Gregg GT (1969) On comparative rating of forecasters. *ESSA Tech. Memorandum WBTM SR-48, October 1969.*
Hicks RA, Keith RJ and Stern H (1979) The impact of FGGE on forecasting in the Victorian region. *Preprints, Australia-New Zealand GARP Symposium, December 17-19, 1979, Melbourne, Australia.*
Industry Commission (1991) Costs and benefits of reducing greenhouse gas emissions. *Rep. No 15, Vol 1, Australian Government Publishing Service*, 251 pp.
Katz RW Murphy AH and Winkler RL (1982) Assessing the value of frost forecasts to orchardists: a dynamic decision-making approach. *J. Appl. Met.*, 21:518-531.
Keith R (1978) Formation of low cloud at Melbourne Airport in a lee trough. *Airmet papers. Canberra Meteor. Soc., Roy. Meteor. Soc. (Aust. Branch), Canberra 1978*, 39-44.
Latif M Barnett TP Cane MA Flügel M Graham ME Storch H von Xu J-S and Zebiak SE (1994) A review of ENSO prediction studies. *Climate Dynamics*, 9:167-179.
Lorenz EN (1963) Deterministic, nonperiodic flow. *J. Atmos. Sci.*, 20:130-141.
Lorenz EN (1969a) Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, 26:636-646.
Lorenz EN (1969b) The predictability of a flow which possesses many scales of motion. *Tellus*, 21:289-307.
Lorenz EN (1993) The essence of chaos. *University of Washington Press*, 227 pp.
Love G (1994) An Australian integrated forecast system. *Int. Conf. on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology. Nashville, Tennessee, 23-28 January 1994, Amer. Meteor. Soc.*
McGregor JL, Leslie LM and Gauntlett DJ (1978) The ANMRC limited-area model: consolidated formulation and operational results. *Mon. Wea. Rev.*, 106:427-438.
Miller AJ and Leslie LM (1984) Short-term single-station forecasting of precipitation. *Mon. Wea. Rev.*, 112:1198-1205.
Miller AJ and Leslie LM (1985) Short-term single-station probability of precipitation forecasting using linear and logistic models. Beitr. Phys. frei. Atmos., 58:517-527.
Murphy AH (1971) A note on the ranked probability score. *J. Appl. Met.*, 10:155-156.
Murphy AH (1973) A new vector partition of the probability score. *J. Appl. Met.*, 12:595-600.
Murphy AH (1977) The value of climatological, categorical, and probabilistic forecasts in the cost-loss ratio situation. *Mon. Wea. Rev.*, 105:803-816.
Murphy JM (1990) Assessment of the practical utility of extended-range ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, 116:89-125.
Murphy AH and Brown BG (1983) Forecast terminology: Composition and interpretation of public weather forecasts. *Bull. Amer. Meteor. Soc.*, 64:13-22.
Murphy AH and Winkler RL (1979) Probabilistic temperature forecasts: the case for an operational program. *Bull. Am. Meteor. Soc.*, 60:12-19.
Murphy AH and Winkler RL (1992) Diagnostic verification of probability forecasts. *Int. J. Forecasting*, 115:1330-1338.
Murphy AH, Lichtenstein S, Fischoff B, and Winkler RL (1980) Misinterpretations of precipitation probability forecasts. *Bull. Amer. Meteor. Soc.*, 61:695-701.
Noone D and Stern H (1995) Verification of rainfall forecasts from the Australian Bureau of Meteorology's Global Assimilation and Prognosis System (GASP). *Aust. Meteor. Mag., 275-286.*
Peirce CS (1884) The numerical measure of the success of predictions. *Science*, 4:453-454.
Pendlebury SF (1979) The effect of drifting buoy data on the forecasting process at the Hobart Regional Forecasting Centre. *Preprints, Australia-New Zealand GARP Symposium, December 17-19, 1979, Melbourne, Australia.*
Ramage CS (1993) Forecasting in meteorology. *Bull. Amer. Meteor. Soc.*, 74:1863-1871
Roebber PJ and Bosart LF (1996) The contributions of education and experience to forecast skill. *Wea. Forecasting, 11:21-40.*
Ryan CJ (1987) *Personal communication.*
Sanders F (1973) Skill in forecasting daily temperature and precipitation. *Bull. Amer. Meteor. Soc.*, 54:1171-1179.

Schumann TEW (1950a) The fundamentals of weather forecasting (Part I). *Weather*, 5:220-224.
Schumann TEW (1950b) The fundamentals of weather forecasting (Part II). *Weather*, 7:248-253.
Scott RH (1873) On some results of weather telegraphy. Quart. J. Meteor. Soc., 1:181-188.
Setek (1997) *Personal communication.*
Shanahan BW (1972) Verification of terminal aerodrome forecasts. *Tech. Note 2, Bur. Met., Australia.* 10pp.
Shanahan BW (1995) *Personal communication.*
Shaw WN (1934) The march of meteorology. Random recollections. *Quart. J. Roy. Meteor. Soc.* 601:101-120
Smith P (1995) *Personal communication.*
Spetzler CS and Staëler von Holstein CAS (1975) Probability encoding in decision analysis. *Manage. Sci.*, 22:340-358.
Stern H (1979a) Trends in the accuracy of Australian temperature forecasts. *Proceedings, Conf. on Value of Meteorological Services, Melbourne, 21-23 February, 1979.*
Stern H (1979b) The east coast low of late August 1974. *Meteor. Note 95, Bur. Met., Australia.*
Stern H (1980a) An increase in the skill of Australian temperature forecasts. *Aust. Meteor. Mag.* 28:223-228.
Stern H (1980b) The development of an automated system of forecasting guidance using analogue retrieval techniques. *M. Sc. Thesis, Department of Meteorology, University of Melbourne* (see also Stern (1985), which refers to a subsequent publication as *Meteor. Study 35, Bur. Met., Australia*).
Stern H (1980c) A system for automated forecasting guidance. *Aust. Meteor. Mag*, 28:141-154.
Stern H (1985) The development of an automated system of forecasting guidance using analogue retrieval techniques. *M. Sc. Thesis, Department of Meteorology, University of Melbourne, Meteor. Study 35, Bur. Met., Australia.*
Stern H (1986a) A trend in the skill of Australian temperature forecasts. *Meteorology Australia*, Vol. 4, No. 2.
Stern H (1986b) A trend in the skill of Australian temperature forecasts. *Extended Abstracts, 2nd Int. Conf. on Southern Hemisphere Meteorology, Amer. Meteor. Soc., Dec 1-5, 1986, Wellington, New Zealand.*
Stern H (1989) Verification of operational numerical prognoses of cyclogenesis off the southeast coast of Australia. *Proceedings, Bur. Met. Res. Centre Workshop "The Role of SSTs in Numerical Modelling of the Atmosphere".*
Stern H (1994) Enveloping predictor parameters in a regression equation – an alternative to Model Output Statistics (MOS). *Abstracts, 10th Conf. Numerical Weather Prediction, Portland, Amer. Meteor. Soc.*
Stern H (1995a) The accuracy of the Victorian component of the Australian Bureau of Meteorology's seasonal outlook service. *Abstracts, 9th Australia New Zealand Climate Forum, Wirrina Cove Resort, South Australia.*
Stern H (1995b) The accuracy of the Victorian component of the Australian Bureau of Meteorology's seasonal outlook service. *Abstracts, Int. Scientific Conf. On Tropical Oceans Global Atmosphere, 2 – 7 April, 1995, Melbourne, Australia.*
Stern H (1996a) Statistically based weather forecast guidance. *Ph.D. Thesis, School of Earth Sciences, University of Melbourne* (approved for publication as a *Meteorological Study, Bur. Met., Australia*).
Stern H (1996b) Statistical analysis of Melbourne maximum temperature forecast errors. *Preprints, 13th Conf. on Probability and Statistics in the Atmospheric Sciences, 21-23 February 1996, San Fransisco, Amer. Meteor. Soc.*
Stern H (1998) An experiment to establish the limits of our predictive capability. *Draft of paper to be submitted to 14th Conf. on Probability and Statistics in the Atmospheric Sciences, 11-15 January 1998, Phoenix, Amer. Meteor. Soc.*
Stern H and Wiliams M (1988) ENSO and summer fire danger in Victoria, Australia. *Abstracts, Conf. On Tropical Meteorology, Brisbane, 1988.*
Stern H and Wiliams M (1989) ENSO and summer fire danger in Victoria, Australia. *Proceedings, 1989 Fire Weather Services Conf., Hobart, Tasmania.*

Stone RC and Hammer GL (1995) The value of climate forecast systems in agricultural production. *Abstracts, 9th Australia New Zealand Climate Forum, Wirrina Cove Resort, South Australia.*

Tennekes H Baede APM and Opsteegh JD (1987) Forecasting forecast skill. *Proceedings, ECMWF workshop on predictability, Reading, 1986.*

Thompson JC (1962) Economic gains from scientific advances and operational improvements in meteorological prediction. *J. Appl. Met.*, 1:13-17.

Treloar ABA and Stern H (1993) A climatology and synoptic classification of Victorian severe weather events. *Preprints, 4th Int. Conf. On Southern Hemisphere Meteorology and Oceanography, Hobart, Australia, March 29 – April 2, 1993.*

Wall St Journal (1992) *As referenced in the text.*

West SJ (1979) Drifting buoy observations as an aid to weather analysis and forecasting off southwest Western Australia. *Preprints, Australia-New Zealand GARP Symposium, December 17-19, 1979, Melbourne, Australia.*

Wilks DS (1995) Forecast value: prescriptive decision studies. In *Economic Value of Weather and Climate Forecasts, RW Katz and AH Murphy eds., Cambridge University Press, Cambridge, UK.*

Woodcock F (1976) The evaluation of yes/no forecasts for scientific and administrative purposes. *Mon. Wea. Rev.*, 104:1209-1214.

Woodcock F (1984) Australian experimental model output statistics forecasts. *Mon. Wea. Rev.*, 112:2112-2121.

Zillman JW (1981) Impact of FGGE buoy system data on southern hemisphere meteorology. *Adv. Space Res.*, 1:197-210. *Cospar, Great Britain.*

Why wait for AIFS?

DAVE WILLIAMS.

Operational Meteorologist. Canberra Meteorological Office.

ABSTRACT

A procedure has been in place in NSW since late 1994 whereby forecasts issued on AROS have been copied across onto a PC. This has allowed the forecasts, principally temperatures, to be archived and subsequently assessed. MOS forecasts have also been considered. Using the methodology of decomposing the mean square error, software has been written to analyse the data. Biases in the MOS guidance have then been isolated and this has allowed the NSW RFC to improve on the guidance to a greater degree than previously, though there is still room for improvement. Canberra's minimum temperature forecasts this winter have been the best they have ever been, at a time when the MOS guidance has been at its worse. There is nothing to stop other regions from employing a similar procedure to archive their forecasts, though the archiving of MOS would be best done using the system of Robert Dahni's.

In the second half of 1994, when MOS became unavailable, the NSW Regional Forecasting Centre's (RFC) minimum and maximum temperature performance declined. This wasn't surprising as it is a well known fact that MOS guidance provides a positive impact on temperature forecasts. Figure 1, taken from "Report 14 - Temperature forecasts during Autumn 1997", shows the strong correlation between MOS and RFC forecasts. Nevertheless it highlighted the urgent need for a new approach in the region, whereby there was more real-time feedback to forecasters. Advances in computing made this readily achievable. When MOS returned in early 1995, and more importantly when it was generated from the output of

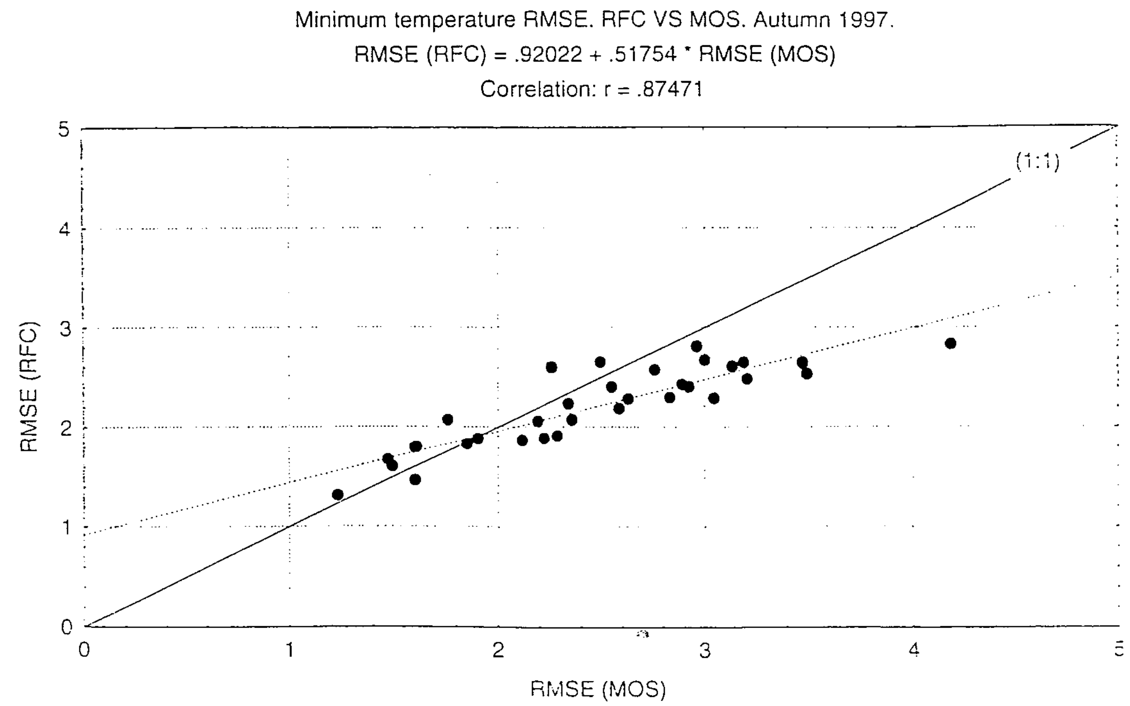


Figure 1: RFC and MOS minimum temperature RMSE's in regional NSW during autumn 1997.