7.1

USING A KNOWLEDGE BASED FORECASTING SYSTEM TO ESTABLISH THE LIMITS OF PREDICTABILITY

Harvey Stern*

Bureau of Meteorology, Melbourne, Vic., 3001, Australia

1. INTRODUCTION

Weather forecasts provided to the public are steadily improving (Figure 1(a)) and six years ago, Stern (1998, 1999) presented the results of an experiment to establish the limits of that predictability. The experiment involved verifying a set of forecasts for Melbourne (Australia) out to 14 days. These forecasts were based upon a subjective interpretation of the ensemble mean output of the NCEP[1] Numerical Weather Prediction (NWP) model.

The verification data suggested that, at that time, routinely providing or utilising day-to-day forecasts beyond day 4 would be inappropriate. However, the data also suggested that it might have been possible to provide some useful information about the likely weather up to about a week in advance for some elements and in some situations. Shortly thereafter, in April 1998, the Victorian Regional Forecasting Centre (RFC) commenced a formal (official) trial of day-to-day forecasts for Melbourne out to day 7.

There have been considerable advances in NWP modelling since then, and also in associated techniques for statistically interpreting the NWP model output utilising objective methods. Stern (2004a) has recently demonstrated that the skill displayed by the trial maximum temperature forecasts is superior to that of climatology (even) at day 7 (Figure 1(b)).

In the light of the skill displayed by the official trial forecasts, the Bureau of Meteorology (BoM) recently commenced routinely issuing a forecast out to day 7 to the public each evening. Predictions for days 5, 6, and 7 are couched in general terms.

2. BACKGROUND

The BoM routinely issues its three-month Seasonal Climate Outlook (SCO) to the public on about the middle day of each month prior. To illustrate, the September to November 2004 outlook (Figure 2) was issued on 17 August, 2004.

Corresponding author address: Harvey Stern, Victorian Regional Office, Bureau of Meteorology, Box 1636M, Melbourne, 3001, Australia; e-mail: H.Stern@bom.gov.au

[1]National Center for Environmental Prediction

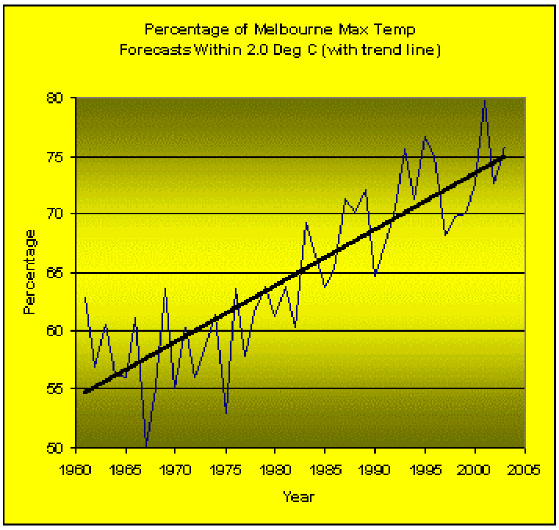


Figure 1(a) Long-term trend in accuracy of Melbourne maximum temperature forecasts 1961-2003, as measured by the percentage of forecasts within 2 deg C.

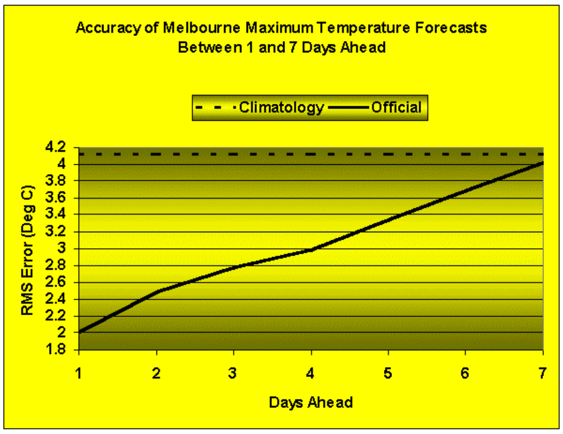


Figure 1(b) Accuracy of day 1 to day 7 Melbourne maximum temperature forecasts 1998-2003, as measured by the Root Mean Square (RMS) error (after Stern, 2004a).

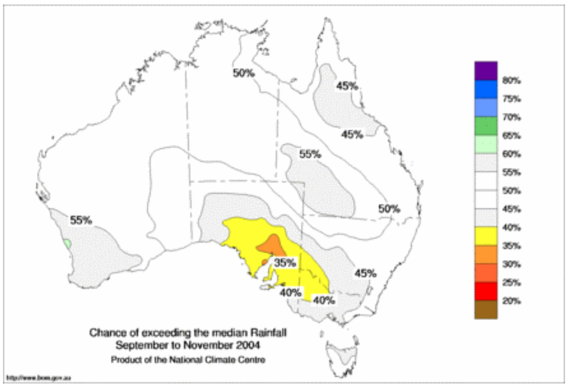


Figure 2 Rainfall outlook for September to November 2004.

The work of Lorenz (1963, 1969a&b, 1993) suggests that there is a 15-day limit to day-to-day predictability of the atmosphere. Furthermore, ongoing increases in the accuracy of NWP model output continue to be evident. It was, therefore, considered appropriate to now repeat Stern's (1998, 1999) experiment.

This was done to enable assessment of whether or not there may now be scientific justification to prepare day-to-day forecasts for the day 8 to day 15 period, with a view to providing a "link" between the day 1 to day 7 forecast and the three-month SCO.

Incidentally, Stern (1999) also investigated whether, even though the day 5 to day 14 forecasts for the *individual* days displayed little skill at that time, they might provide an indication of *overall* weather conditions during the day 5 to day 14 period.

Analysis of the data revealed that there was some basis for an affirmative response to this question. However, the relationship between the forecast weather and observed weather was not strong (level of significance between 0.2% and 13%).

Interestingly, with increases in the accuracy of NWP model output, the South African Weather Service now provides such a product (Figure 3).

3. A LONG RANGE GLOBAL FORECASTING SYSTEM

RFC forecaster Stuart Coombs recently alerted the author to anecdotal evidence that the output of the NOAA[2] GFSlr[3] NWP Model displayed considerable skill, and that, on occasions, it had predicted significant events even towards the end of the forecast period (day 16).

[2]National Oceanic and Atmospheric Administration
[3]Global Forecasting System long range

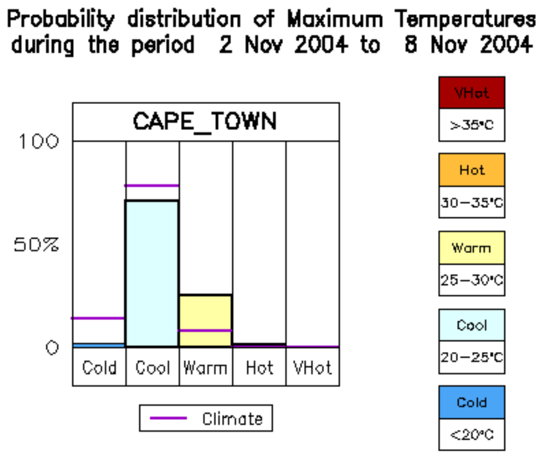


Figure 3 Probability distribution of Week 2 Cape Town maximum temperature, showing an enhanced likelihood (compared with climatology) of warm weather, and a diminished likelihood (compared with climatology) of cool to cold weather (source: http://www.weathersa.co.za/fcastProducts/ExtendedRange/Images/CAPE_TOWN_TX.gif (as at 04UTC, 27 October, 2004).

This GFSlr output includes forecast data every 12 hours from forecast hour 192 (day 8) to 384 (day 16) on a 2.5 degree latitude/longitude grid covering the globe. The data is updated 4 times per day. An illustration of the output of the system is presented at Figure 4, and, for a more complete view, one may refer to:
http://www.arl.noaa.gov/ready/metdata.html.

The current paper presents a preliminary study of the skill displayed by forecasts derived from 100 twice-daily "runs" of the GFSlr model (base analyses between 12UTC on 5 August, 2004, and 00UTC on 24 September, 2004). The paper also makes reference to forecasts derived on base analyses subsequent to the period of preliminary study.

4. INTERPRETING MODEL OUTPUT IN TERMS OF LOCAL WEATHER

Over recent years, Stern (2002, 2003, 2004b&c) has been involved in the development of a knowledge based weather forecasting system. Illustrations of components of its output are presented at Figures 5 and 6; and, for a more complete view of the system, one may refer to:
http://www.weather-climate.com/knowledge.html.

The knowledge based forecasting system was utilised to objectively interpret the output of the GFSlr model statistically in terms of local weather at Melbourne (maximum temperature, minimum temperature, probability of precipitation, and amount of precipitation) in order to rigorously establish current limits of predictability.

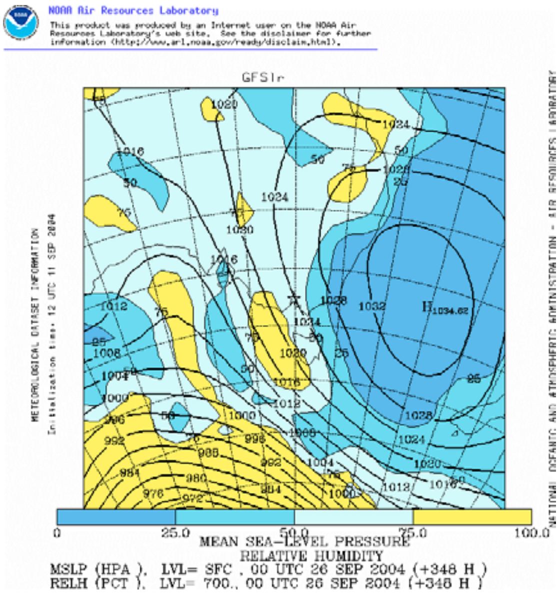
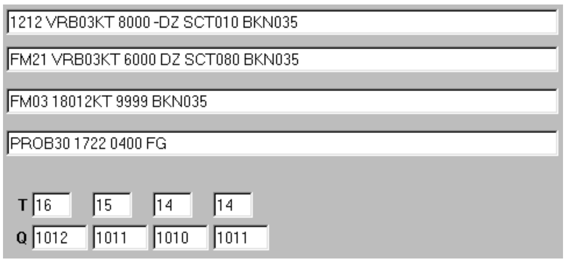


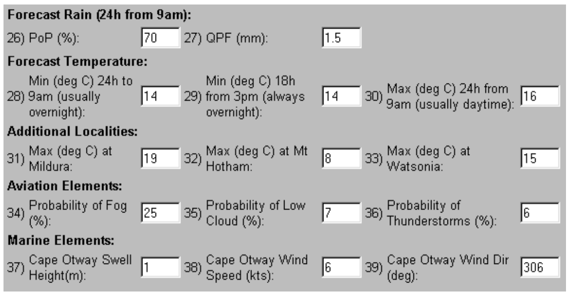
Figure 4 Global Forecasting System long range (day 14.5) forecast of MSL Pressure and 700 hPa Relative Humidity for 00 UTC 26 September, 2004.



1212 VRB03KT 8000 -DZ SCT010 BKN035
FM21 VRB03KT 6000 DZ SCT080 BKN035
FM03 18012KT 9999 BKN035
PROB30 1722 0400 FG

T	16	15	14	14
Q	1012	1011	1010	1011

Figure 5 Illustration of the Terminal Aerodrome Forecast (TAF) component of the output of the knowledge based system.



Forecast Rain (24h from 9am):

26) PoP (%):	70	27) QPF (mm):	1.5		

Forecast Temperature:

28) Min (deg C) 24h to 9am (usually overnight):	14	29) Min (deg C) 18h from 3pm (always overnight):	14	30) Max (deg C) 24h from 9am (usually daytime):	16

Additional Localities:

31) Max (deg C) at Mildura:	19	32) Max (deg C) at Mt Hotham:	8	33) Max (deg C) at Watsonia:	15

Aviation Elements:

34) Probability of Fog (%):	25	35) Probability of Low Cloud (%):	7	36) Probability of Thunderstorms (%):	6

Marine Elements:

37) Cape Otway Swell Height(m):	1	38) Cape Otway Wind Speed (kts):	6	39) Cape Otway Wind Dir (deg):	306

Figure 6 Illustration of some quantitative components of the output of the knowledge based system.

5. CONSISTENCY OF OUTPUT

Examination of the output of "runs" of the GFSlr model reveals a modest, but useful, level of consistency of output from one "run" to the next. Indeed, there were a number of occasions when consistent advance notice was given by the GFSlr model to the prediction of a number of unusual events. For example, Figure 7 shows that the warm day on 20 September, 2004, was anticipated well in advance by the GFSlr model, as interpreted by the knowledge based system.

However, even this relatively low level of jerkiness in the forecasts from one "run" to the next would be unsatisfactory, were a decision made to issue them to the public.

One approach to address this "jerkiness" may be to regard the individual forecasts as members of an ensemble. For example, the output of the GFSlr model's four most recent runs may then be averaged.

To illustrate, the output of the GFSlr model's 850 hPa temperature and 700 hPa relative humidity are among the data that are input into the knowledge based system in order to obtain a weather forecast.

Firstly, regarding the 850 hPa temperature, Figure 8(a) depicts the mean and standard deviation (uncertainty) of the Melbourne 850 hPa temperature forecast by the 20 October, 2004, "runs" - 00UTC, 06UTC, 12UTC and 18UTC. It also shows a 2nd Order Line of Best-Fit polynomial based on the standard deviation (sd) data, which suggests that there is an overall increase in uncertainty associated with the forecasts of 850 hPa temperature as one moves from day 1 (Oct-21) to day 16 (Nov-5).

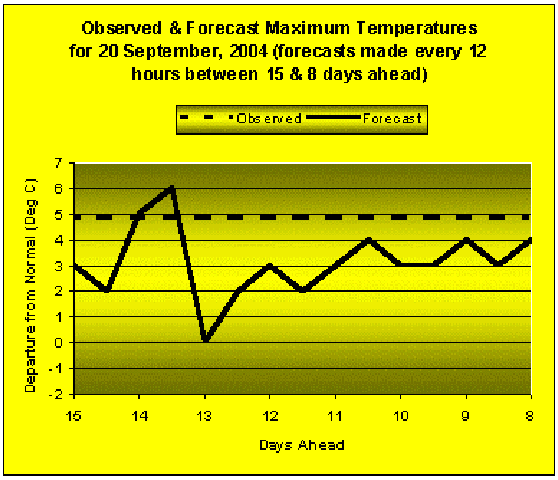


Figure 7 Observed and forecast maximum temperatures for 20 September, 2004.

Secondly, regarding the 700 hPa relative humidity, Figure 8(b) depicts the mean and standard deviation (uncertainty) of the Melbourne 700 hPa relative humidity forecast by the 20 October, 2004, "runs" - 00UTC, 06UTC, 12UTC and 18UTC. It also shows a 2nd Order Line of Best-Fit polynomial based on the standard deviation (sd) data, which suggests that there is an overall increase in uncertainty associated with the forecasts of relative humidity as one moves from day 1 (Oct-21) to day 16 (Nov-5).

To illustrate the skill displayed by the maximum temperature forecasts close to the event, Figure 9(a) depicts the day-to-day fluctuations in the departure from normal of the day 8 and day 8.5 forecast and observed maximum temperatures, whilst Figure 9(b) compares departures from normal of observed and forecast maximum temperatures 8 and 8.5 days in advance.

6. INDEPENDENCE OF FORECAST DATA SETS

There are 22 forecast data sets, each comprising predictions of:

- Minimum temperature (*Min*);
- Maximum temperature (*Max*);
- Quantity of Precipitation Forecast (*QPF*); and,
- Probability of Precipitation (*PoP*).

Four of these data sets correspond to the official BoM forecasts for 1, 2, 3, and 4 days ahead.

Three of these data sets correspond to official trial forecasts for 5, 6, and 7 days ahead.

Fifteen of these data sets correspond to the forecasts based on the interpretation of the output of the GFSlr NWP model for 8, 8.5, 9, 9.5 … 15 days ahead.

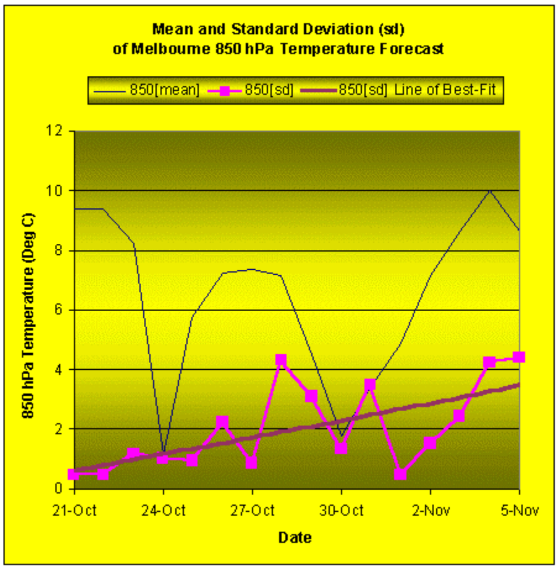


Figure 8(a) Mean and standard deviation (uncertainty) of the Melbourne 850 hPa temperature forecast by the 20 October, 2004, "runs" - 00UTC, 06UTC, 12UTC and 18UTC, and a 2nd Order Line of Best-Fit polynomial based on the standard deviation (sd) data.

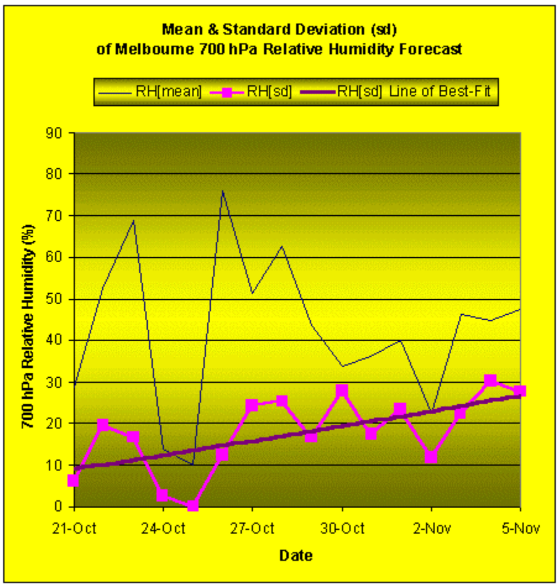


Figure 8(b) Mean and standard deviation (uncertainty) of the Melbourne 700 hPa relative humidity (RH) forecast by the 20 October, 2004, "runs".

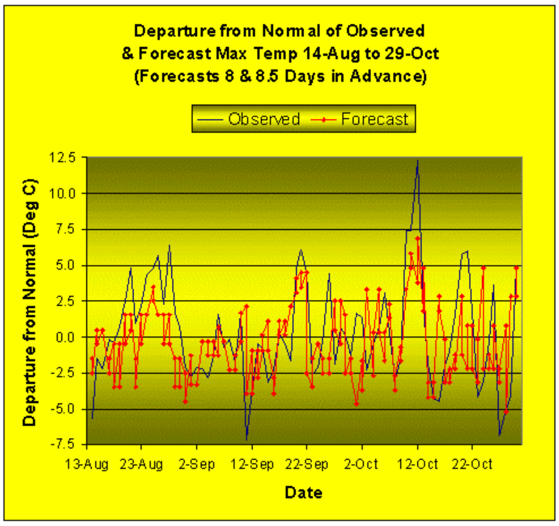


Figure 9(a) Day-to-day fluctuations in the departure from normal of observed and forecast maximum temperatures 8 and 8.5 days in advance. The cold days of 14 Aug and 11 Sep were well anticipated, as also were the warm days around 25 Aug, 20 Sep and 12 Oct. This depiction of day-to-day fluctuations extends beyond the period of preliminary study.

However, the elements of each of these 22 forecast data sets are not truly independent. This lack of independence arises from the fact that weather patterns often persist for several days.

Now, Figure 10 shows that the best-fit curve for all of the Regression Coefficients 'b' in the *Max*, *Min*, *PoP*, and *QPF* Equations:

(Observed departure from normal[4])=

a + b(Observed departure from normal a number of days before)

where a and b are constants (1)

suggests that 'b' is positive until day 4.

One may, therefore, deduce that only every fourth data element is truly independent. The numbers of degrees of freedom utilised to establish confidence limits in the analyses that follow are all, therefore, reduced to 1/4 of what they would have been, had all the data elements been truly independent.

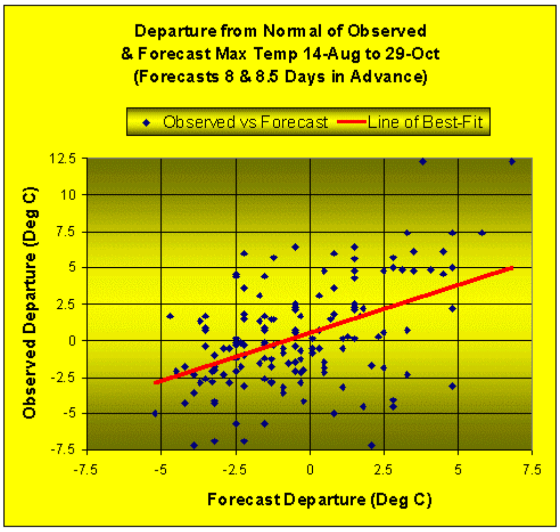


Figure 9(b) Comparison between the departures from normal of observed and forecast maximum temperatures 8 and 8.5 days in advance.

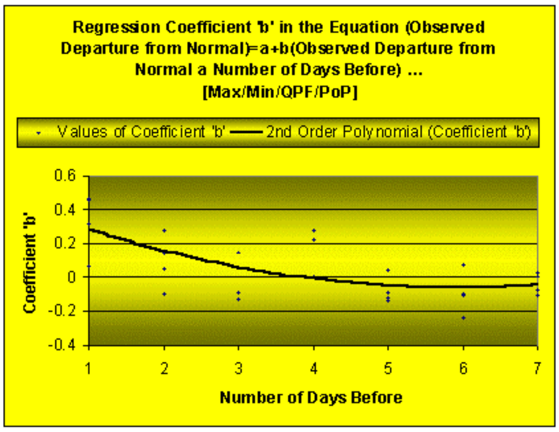


Figure 10 The best-fit curve for all of the Regression Coefficients 'b' in the *Max*, *Min*, *QPF* and *PoP* Equations 1.

7. ANALYSIS OF RESULTS

7.1 *Minimum temperature*

Figure 11 depicts three best-fit 2nd order polynomial curves, about the Regression Coefficients

[4]In this context, the normal *QPF* is regarded as the √(mean daily precipitation for a particular month), whilst the observed *Precipitation Amount* is either √(observed precipitation), if precipitation is observed, or 0 (if precipitation is not observed). Furthermore, in this context, the normal *PoP* is regarded as the monthly proportion of days with precipitation, whilst the observed *PoP* is either 100%, if precipitation is observed, or 0% (if precipitation is not observed).

'b' in the Equations derived on data from the 22 *Min* forecast data sets:

(Observed *Min* departure from normal)=

a + b(Forecast *Min* departure from normal a number of days in advance)

where a and b are constants (2)

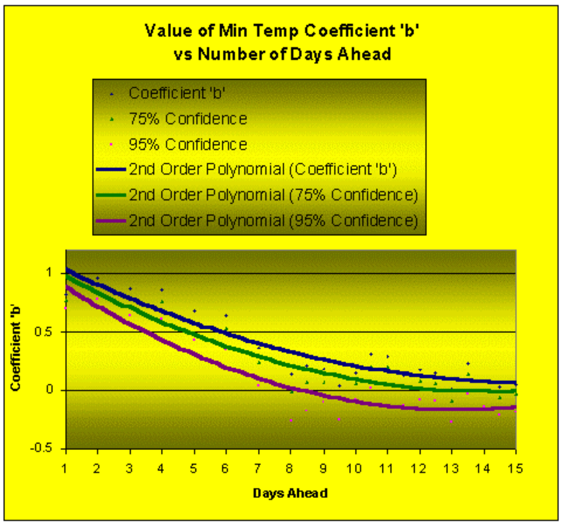


Figure 11 Confidence limits for the regression coefficients 'b' in the *Min* equations. In calculating confidence limits, the number of degrees of freedom is reduced by a factor of 1/4. Positive values of 'b' suggest skill.

The three curves are:

- Top curve: Regression coefficients 'b';
- Middle curve: 75% lower confidence limit for regression coefficient 'b'; and,
- Bottom curve: 95% lower confidence limit for regression coefficient 'b'.

The curves show that:

- It is more likely than not that there is skill at forecasting minimum temperature out to 15 days ahead;
- It is three times more likely than not that there is skill at forecasting minimum temperature out to 12 days ahead; and,
- One can be 95% confident that there is skill at forecasting minimum temperature out to 8 days ahead.

The 'b's represent the proportion of the forecast departure from normal to utilise, should one wish to achieve optimal forecast skill. Hence, by way of example, for forecasts for 8 days ahead, although the significance of the skill is high (at the 95% level), the magnitude of that skill is not (the proportion of the forecast departure from normal to utilise is only 0.35).

7.2 *Maximum temperature*

Figure 12 depicts three best-fit 2nd order polynomial curves, about the Regression Coefficients 'b' in the Equations derived on data from the 22 *Max* forecast data sets:

(Observed *Max* departure from normal)=

a + b(Forecast *Max* departure from normal a number of days in advance)

where a and b are constants (3)

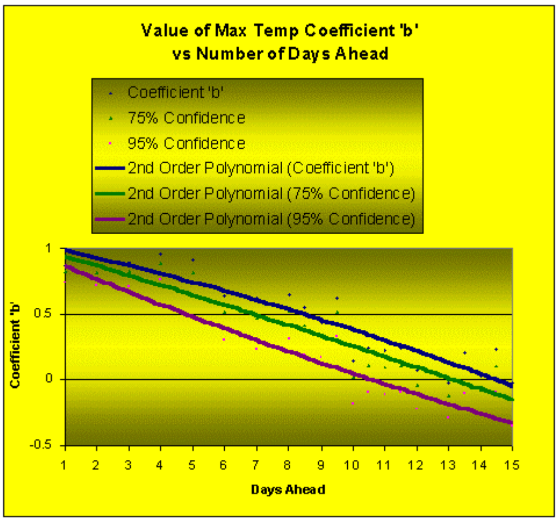


Figure 12 Confidence limits for the regression coefficients 'b' in the *Max* equations. In calculating confidence limits, the number of degrees of freedom is reduced by a factor of 1/4. Positive values of 'b' suggest skill.

The three curves are:

- Top curve: Regression coefficients 'b';
- Middle curve: 75% lower confidence limit for regression coefficient 'b'; and,
- Bottom curve: 95% lower confidence limit for regression coefficient 'b'.

The curves show that:

- It is more likely than not that there is skill at forecasting maximum temperature out to 14 days ahead;
- It is three times more likely than not that there is skill at forecasting maximum temperature out to 13 days ahead; and,

- One can be 95% confident that there is skill at forecasting maximum temperature out to 10 days ahead.

As with the case for minimum temperature, the 'b's represent the proportion of the forecast departure from normal to utilise, should one wish to achieve optimal forecast skill.

Figure 12 shows that, for maximum temperature forecasts for 8 days ahead, the significance of the skill (better than the 95% level) and the optimal proportion of forecast departure from normal to utilise (0.55) are both higher than corresponding values for minimum temperature.

Furthermore, Figure 12 shows that the optimal proportion of forecast departure from normal to utilise for day 7 is 0.63. This value is greater than the corresponding value derived by Stern (2004a) for day 7 (0.511) using 1998-2003 data from the official trial. One may interpret this to be suggesting that there has been an improvement in the accuracy of the official trial forecasts since the 1998-2003 period.

7.3 *Quantitative Precipitation Forecast*

Figure 13 depicts three best-fit 2nd order polynomial curves, about the Regression Coefficients 'b' in the Equations derived on data from the 22 *QPF* forecast data sets:

```
(Observed Precipitation Amount departure from normal)=

a + b(Forecast QPF departure from normal a number of days in advance)

where a and b are constants             (4)
```

The three curves are:

- Top curve: Regression coefficients 'b';
- Middle curve: 75% lower confidence limit for regression coefficient 'b'; and,
- Bottom curve: 95% lower confidence limit for regression coefficient 'b'.

The curves show that:

- It is more likely than not that there is skill at forecasting precipitation amount out to 14 days ahead;
- It is three times more likely than not that there is skill at forecasting precipitation amount out to 10 days ahead; and,
- One can be 95% confident that there is skill at forecasting precipitation amount out to 6 days ahead.

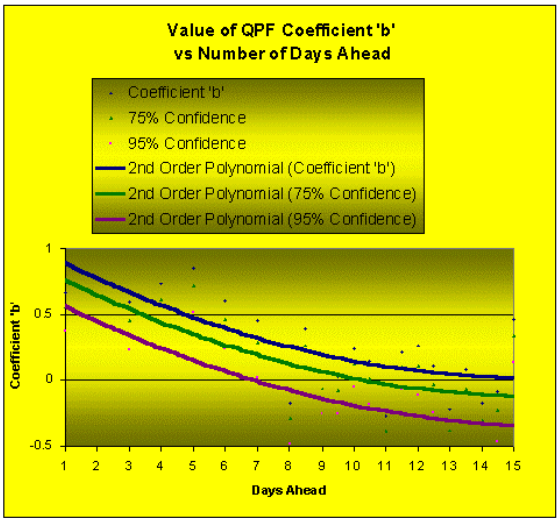


Figure 13 Confidence limits for the regression coefficients 'b' in the *QPF* equations. In calculating confidence limits, the number of degrees of freedom is reduced by a factor of 1/4. Positive values of 'b' suggest skill.

7.4 *Probability of Precipitation (PoP)*

Figure 14 depicts three best-fit 2nd order polynomial curves, about the Regression Coefficients 'b' in the Equations derived on data from the 22 *PoP* forecast data sets:

```
(Observed PoP departure from normal)=

a + b(Forecast PoP departure from normal a number of days in advance)

where a and b are constants             (5)
```

The three curves are:

- Top curve: Regression coefficients 'b';
- Middle curve: 75% lower confidence limit for regression coefficient 'b'; and,
- Bottom curve: 95% lower confidence limit for regression coefficient 'b'.

The curves show that:

- It is more likely than not that there is skill at forecasting *PoP* out to 12 days ahead;
- It is three times more likely than not that there is skill at forecasting *PoP* out to 10 days ahead; and,
- One can be 95% confident that there is skill at forecasting *PoP* out to 7 days ahead.

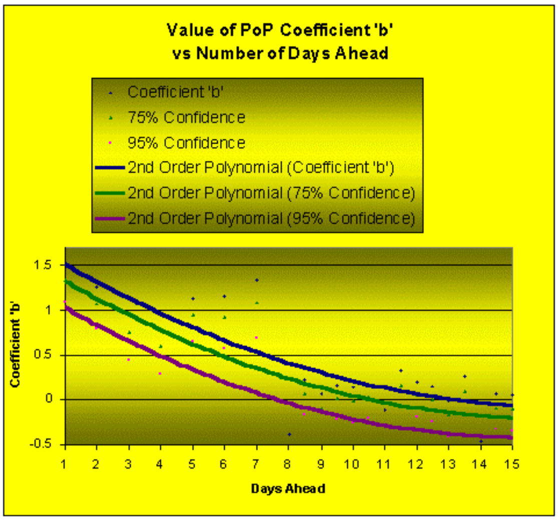


Figure 14 Confidence limits for the regression coefficient 'b' in the *PoP* equations. In calculating confidence limits, the number of degrees of freedom is reduced by a factor of 1/4. Positive values of 'b' suggest skill.

8. CONCLUSION

Analysis of the data suggests that application of the knowledge based system to the interpretation of the Global Forecasting System long range model output yields a set of day-to-day weather predictions that display a modest, but nevertheless potentially useful, level of skill, especially at predicting temperature.

This outcome appears to justify the emergence on the web of extended-period day-to-day forecasts.

Furthermore, even a modest level of forecast skill may be applied to financial market instruments, such as weather derivatives, in order to ameliorate weather-related risk. It may, therefore, be justifiable to prepare such forecasts with a view to using them to ameliorate that risk, and also with a view to providing a "link" between the short-term forecasts and the three-month Seasonal Climate Outlook.

It is planned to analyse further "runs" of the Global Forecasting System long range model, as interpreted by the knowledge based system, in order to add strength to the significance of the conclusions drawn here.

For an updated account of the work in progress readers may go to:
http://www.weather-climate.com/ams2005lr.html

9. REFERENCES

Lorenz, E. N., 1963: Deterministic, non-periodic flow. *J. Atmos. Sci.,* **20**, 130-41.

Lorenz, E. N., 1969a: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.,* **26**, 636-46.

Lorenz, E. N., 1969b: The predictability of a flow which possesses many scales of motion. *Tellus,* **21**, 289-307.

Lorenz, E. N., 1993: The essence of chaos. *University of Washington Press.*

Stern, H., 1998: An experiment to establish the limits of our predictive capability. *14th Conference on Probability and Statistics / 16th Conference on Weather Forecasting and Analysis, Phoenix, Arizona, 11-16 Jan., 1998, Amer. Meteor. Soc.*

Stern, H., 1999: An experiment to establish the limits of our predictive capability for Melbourne. *Aust. Meteor. Mag.,* **48**, 159-167.

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

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

Stern, H., 2004a: Using verification data to improve forecasts. *Australian Meteorology and Oceanography Society 2004 Annual Conference, Brisbane, Queensland, Australia 5-9 Jul., 2004.*

Stern, H., 2004b: Incorporating an ensemble forecasting proxy into a knowledge based system. *20th Conference on Interactive Information and Processing Systems, Seattle, Washington, 11-15 Jan., 2004, Amer. Meteor. Soc.*

Stern, H., 2004c: Using a knowledge based system to predict thunderstorms. *International Conference on Storms, Storms Science to Disaster Mitigation, Brisbane, Queensland, Australia 5-9 Jul., 2004.*