J2.7 AN EXPERIMENT TO ESTABLISH THE LIMITS OF OUR PREDICTIVE CAPABILITY

Harvey Stern*
Bureau of Meteorology, Melbourne, Australia

1. **INTRODUCTION**

1.1 Background

The Victorian Regional Office (VRO) of the Australian Bureau of Meteorology (BoM) currently provides 1 to 4 day weather forecasts to the general public. These forecasts are issued each afternoon by the VRO Regional Forecasting Centre (RFC) in Melbourne. The forecasts are based upon an interpretation, in terms of local weather, of the output of various global and Australian region Numerical Weather Prediction (NWP) models.

This interpretation is carried out using a combination of the Generalised Analogue Statistics Model (GASM) (Stern, 1980a&b; Stern and Dahni, 1981; Dahni *et al*, 1984; Dahni, 1988; Dahni and Stern, 1995; Stern, 1996; Dahni, 1996), Model Output Statistics (MOS) (Woodcock, 1984) and other guidance.

There is potential for extending the application of such forecast guidance schemes because interpreting local weather in terms of the synoptic flow is readily automated by statistical methodologies. Indeed, Brooks (1995) wrote 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".

That this prediction by Brooks may soon become a reality is supported by Figure 1. Figure 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. This reduction in performance is possibly a consequence of less attention being able to be directed towards the lower priority long-term predictions.

However, Figure 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.

The National Center for Environmental Prediction (NCEP) currently produces a 15-day global ensemble average prognosis (Toth and Kalnay, 1993; Traction and Kalnay, 1993; Climate Diagnostic Center, 1997). 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.

1.2 Purpose

This paper has a two-fold purpose:

(1) To present preliminary results of an experiment, which involves verifying a set of quantitative forecasts for Melbourne out to 14 days. These forecasts are based on a subjective interpretation of the NCEP output, when available, for days 5 to 14, and of the official forecasts for days 1 to 4; and,

(2) To use the results to assess whether or not extending the period of the official forecasts beyond 4 days might be justified. After all, "verification allows forecasters to know, quantitatively and objectively, how well they are doing, and in what ways they can improve their product" (Doswell, 1995).

2. **VERIFICATION METHODOLOGY**

2.1 Verification measures

The experimental forecasts are verified against "climatology" and a randomly generated set of forecasts. 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*;

---

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

RMS Error of the maximum temperature forecasts, *bf*;

RMS Error of the quantitative precipitation forecasts (QPF) in Rainfall Ranges, as defined by Stern (1980a), Ranges 0, 1, 2, 3, etc. being respectively less than 0.2 mm, 0.2-2.5 mm, 2.6-5 mm, 5.1-10 mm, etc., *cf*;

Percentage rain/no rain (R/NR) forecasts correct, *df*; and,

Brier score (Brier, 1950) about probability of precipitation (PoP) forecasts, as modified in accordance with how it is now "used almost universally" (Wilks, 1995), *ef*.

For each of the forecast days 1 to 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.

2.2 Verification skill scores

Firstly, the ***"minimum temperature skill score"*** (Figure 3) is defined as

$$100x((ac/af)-1)$$

Secondly, the ***"maximum temperature skill score"*** (Figure 4) is defined as

$$100x((bc/bf)-1)$$

Thirdly, the ***"QPF skill score"*** (Figure 5) is defined as

$$100x((cc/cf)-1)$$

Fourthly, ***"R/NR skill score"*** (Figure 6) is defined as

$$100x(\sqrt{(df/dc)}-1)$$

Fifthly, the ***"Brier skill score"*** (Figure 7) is defined as

$$100x(\sqrt{(ec/ef)}-1)$$

The square root (√) of (df/dc) and (ec/ef) are required, in order that they be directly comparable to (ac/af), (bc/bf) and (cc/cf) (which intrinsically do include a square root function).

These are then combined into a ***"temperature skill score"*** (Figure 8), defined as

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

and, a ***"rainfall skill score"*** (Figure 9), defined as

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

and, a ***"joint skill score"*** (Figure 10), defined as the mean of the temperature and rainfall skill scores.

2.3 Interpreting words

In order to objectively verify the official forecasts, it was necessary to interpret worded components of the 1 to 4 day forecasts in terms of their meaning in terms of QPF, R/NR and PoP. Stern (1980a) 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 (1980a) Appendix A. 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 (1980a) Appendix C includes an objective scheme to "translate" official forecasts into that terminology and a consequential objective means to evaluate the official forecast. Table 1's interpretation of worded components of the 1 to 4 day forecasts is based on the aforementioned scheme.

Figure 11 depicts the frequency with which precipitation occurred in association with various PoPs. It provides retrospective support to Table 1's interpretation of the worded components in terms of PoP as well as to the forecasters' skill at composing these components. However, Figure 11 suggests that some of the words which have been assigned PoPs of 60% and 70% might better have been assigned PoPs of 50% and 60%, respectively.

Figure 12 depicts the mean precipitation which occurred in association with various QPFs. It provides retrospective support to Table 1's interpretation of the worded components in terms of QPF as well as to the forecasters' skill at composing these components.

3. **DISCUSSION**

3.1 Forecast performance

Figures 3 to 10 show that skill at predicting both temperature and rainfall decline, albeit unsteadily, as one moves from day 1 to day 14. 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 29 August data. The decline might be expected to become smoother as the numbers of data increase.

The data depicted in Figures 3 to 10 show that, on most measures:

Forecast performance declines to a level which is inferior to climatology by about day 5;

Forecast performance declines further from day 5 to about day 10, when it is substantially inferior to climatology; but that,

Forecast performance then improves slightly from about day 10 to day 14.

3.2 Explanation

The decline in overall forecast performance to an ***inferior*** level to climatology is attributed to the forecasts based on climatology being not likely to include errors as large as the forecasts based on the full range of possibilities, as would a set of randomly generated forecasts. For comparison, skill scores for a set of randomly generated forecasts are quoted in the captions beneath Figures 3 to 10 inclusive.

The decline in forecast performance from day 5 to about day 10 is attributed to there being some residual skill present in forecasts for some elements and in some synoptic situations during the early part of that component of the forecast period. That there is this skill is supported by the data presented in Figure 13. Figure 13 shows that, for example, when there is confidence about whether or not precipitation is likely, either on the part of the official forecaster (up to day 4), or due to a strong signal from the NCEP ensemble average (from day 5), this confidence is justified. Figure 13 also shows that, when there is little confidence about whether or not precipitation is likely, this reduced confidence is justified to the extent that estimates from day 3 onwards are inferior to climatology.

The improvement in forecast performance from about day 10 to day 14 is attributed to the NCEP ensemble average prognosis tending towards climatology at the end of the forecast period.

3.3 Some implications

In summary, it seems unlikely that ***routinely*** extending the official day-to-day forecast period beyond day 4 can be justified at present. Nevertheless, it might be possible to provide useful information about the likely weather for up to about a week in advance for some elements and in some situations.

Traders in agricultural commodities regularly utilise longer-term day-to-day predictions in their work. For example, the 22 June 1992 Wall Street Journal reported that “the possible development of a high pressure ridge”, depicted in the 10th day of the US National Weather Service's (NWS) model, sparked “renewed fears of a drought in the central Midwest (and) drove grain futures prices higher at the Chicago Board of Trade”. If the data presented for Melbourne is regarded as representative of other locations, this practice cannot be justified. Indeed, the NWS model prediction, referred to earlier, proved to be incorrect and the Wall Street Journal of 3 July 1992 reported that “heavy rain … helped alleviate short-term drought fears and drove grain futures prices lower at the Chicago Board of Trade”.

4. **CONCLUSION**

4.1 Findings

A rigorous forecast verification methodology has been described, which may be used to guide decision-makers about the validity (or otherwise) of providing (or utilising) longer-term day-to-day weather predictions.

The verification data derived using the methodology suggest that, at present:

(1) ***Routinely*** providing or utilising day-to-day forecasts beyond day 4 would be inappropriate; but,

(2) It might be possible to provide some useful information about the likely weather up to about a week in advance ***for some elements and in some situations***. By contrast, in some circumstances it may not be possible to provide useful information even for days 3 and 4.

4.2 Future work

The conclusions presented here are based on a short period of (mostly) winter data for only one place. It is therefore planned to continue the experiment and, later on, to extend it to other localities. This is in order to establish whether or not the conclusions may be applied generally, that is, to other seasons and to other localities.

Table 1 Interpretation of words in terms of PoP (%) and QPF (mm). R/NR is defined by PoP being above or below 50%. it is set at a 50/50 chance if PoP is 50%. The second figure in a column applies if precipitation is referenced only in one of the pre-noon / post-noon periods.

Words	PoP	QPF
Sunny	10	0
Mainly sunny; Fine	20	0
Becoming sunny; Partly cloudy; Becoming cloudy	30	0
Mainly cloudy; Cloudy	40	0
Chance of precipitation; Cool change (with no precipitation reference); Local showers; Fog or drizzle; Drizzle patches	50	0.2
Little drizzle	60	0.2
Few showers	70-60	1-0.2
Mainly fine; Shower or two	70	1
Becoming fine; Clearing shower or two	60	0.2
Little rain	70-60	1-0.2
Drizzle	80-70	1-0.2
Showers [S]	80-70	5-2
S developing or clearing	70	2
Thunderstorms [T]	80-70	10-5
Snow	90-80	5-2
Sleet	90-80	5-2
Rain [R]	90-80	10-5
R or T, heavy at times	90	20
S, heavy at times	90	10
R at times	90	10

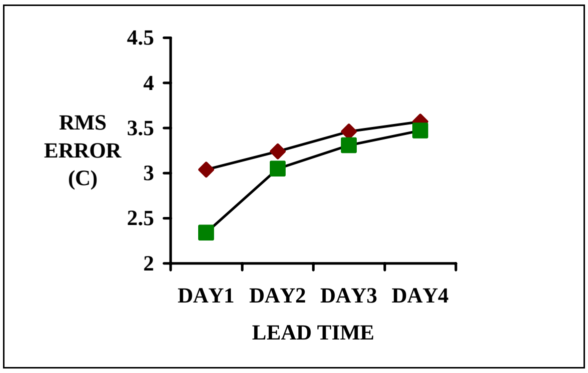

Figure 1 Chart illustrating the extent that official maximum temperature forecasts (squares) for seventeen Victorian centres, issued several days ahead, improved upon the guidance generated by GASM using the European Centre for Medium Range Weather Forecasting (ECMWF) global model output (diamonds). The January 1997 to June 1997 VRO RFC data, upon which this chart is based, were provided by Setek (1997).

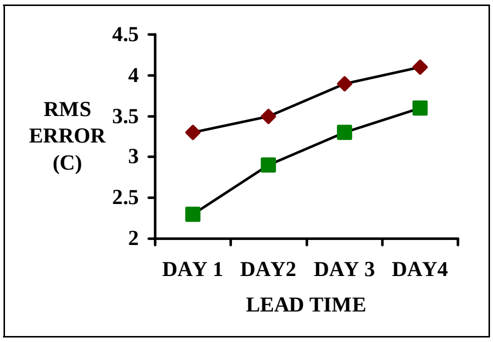

Figure 2 As Figure 1, but for Melbourne alone. The January 1997 to June 1997 VRO RFC data, upon which this chart is based, were provided by Setek (1997).

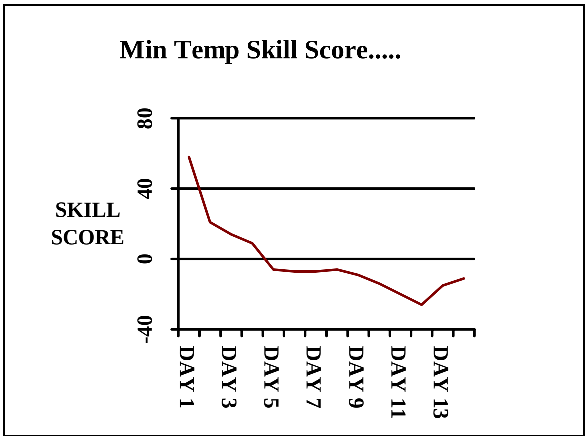

Figure 3 Minimum Temperature Skill Score compared to climatology. Positive values show skill better than climatology (Random = -21).

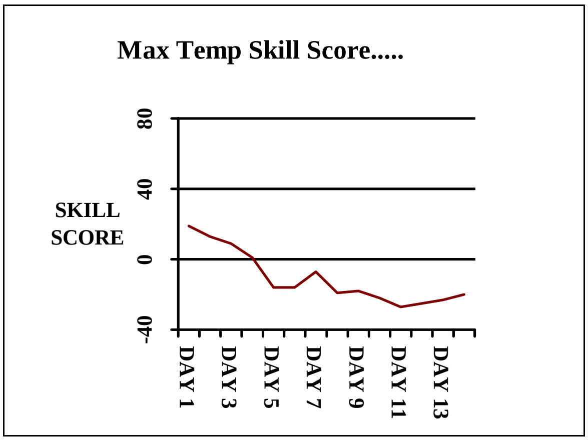

Figure 4 As Figure 3, but for Maximum Temperature Skill Score (Random = -10).

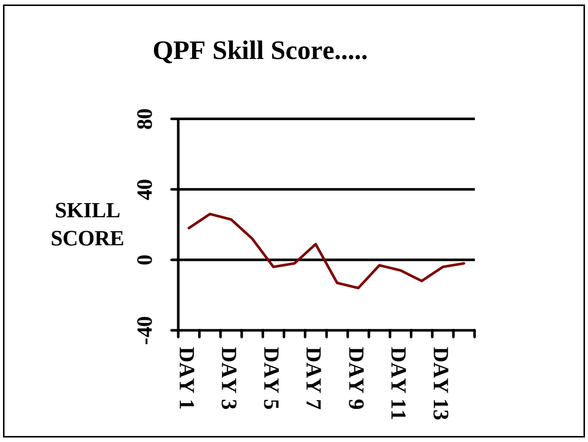

Figure 5 As Figure 3, but for QPF Skill Score (Random = -13).

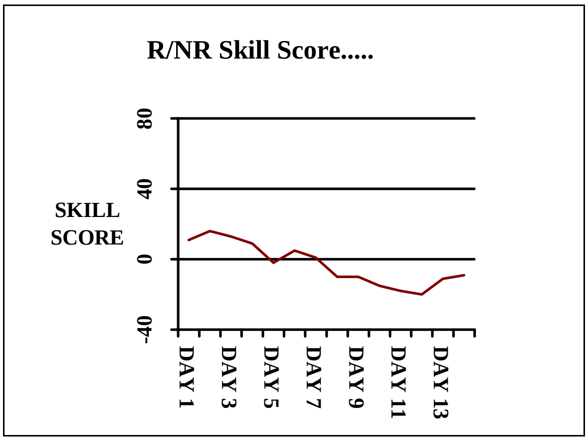

Figure 6 As Figure 3, but for R/NR Skill Score (Random = -12).

Figure 7 As Figure 3, but for Brier Skill Score (Random = -9).

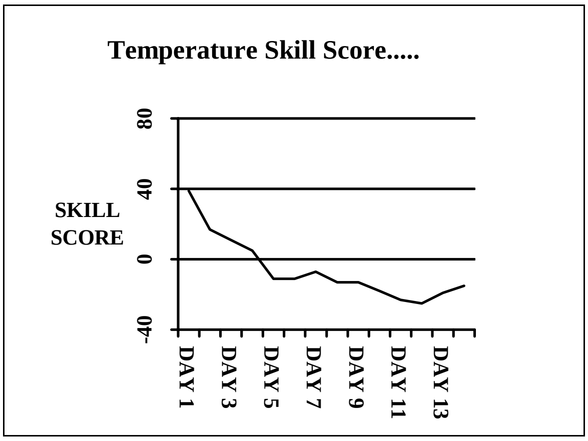

Figure 8 As Figure 3, but for Temperature Skill Score (Random = -15).

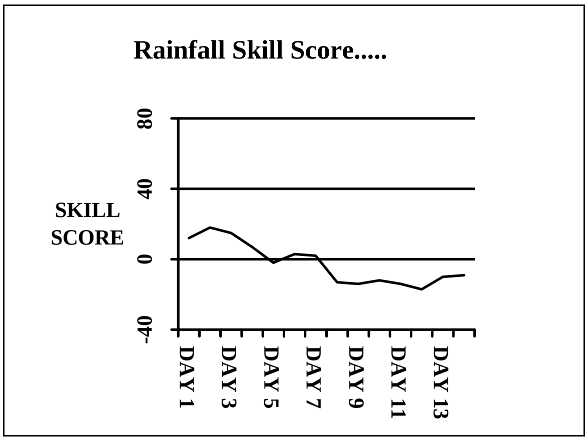

Figure 9 As Figure 3, but for Rainfall Skill Score (Random = -11).

Figure 10 As Figure 3, but for Joint Skill Skill Score (Random = -13).

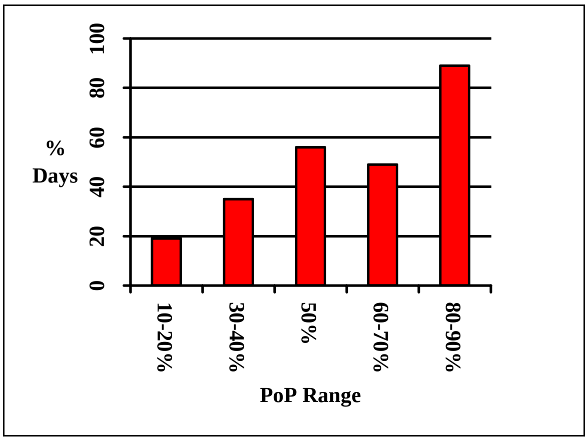

Figure 11 Percentage of days various PoPs were associated with precipitation.

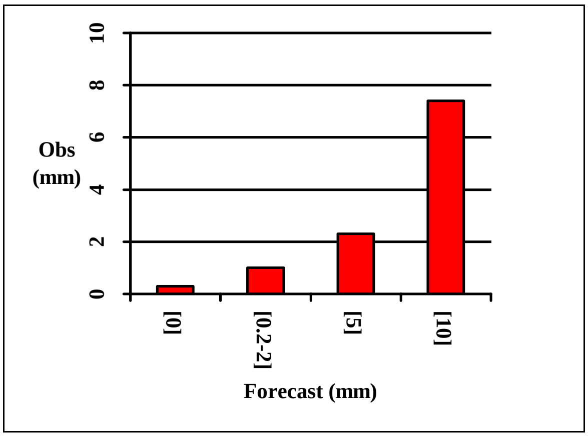

Figure 12 Observed mean precipitation associated with various QPFs.

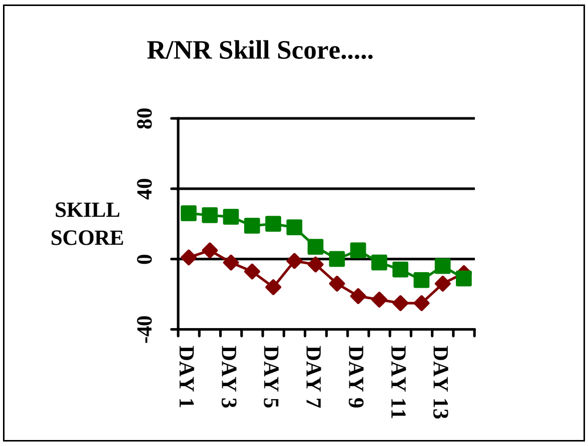

Figure 13 A comparison of the R/NR Skill Scores for high-confidence forecasts (squares) and low-confidence forecasts (diamonds). High-confidence forecasts are those with PoPs of 10-20% or 80-90%. Low-confidence forecasts are those with PoPs of 30-70%.

5. **REFERENCES**

BoM, 1977: Recording and encoding weather observations.

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

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

Climate Diagnostic Center (CDC), 1997: CDC Map Room Weather Products. *From http://www.cdc.noaa.gov/~map/maproom/ENS/ens_desc.html*

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

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

Dahni, R. R., 1996: Generalised Analogue Statistics Model Version 1.8 Reference Manual. *Internal Bur. Met., Australia Doc.*

Dahni, R. R. 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.*

Doswell, C. A. 3rd 1995: Meteorology and users. *From http://www.nssl.uoknor.edu/~doswell/*

Lorenz, E. N., 1963: Deterministic, non-periodic flow. *J. Atmos. Sci*., 20:130-141.

Lorenz, E. N., 1969a: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.,* 26:636-646.

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.*

Setek, M., 1997: *Personal communication*.

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

Stern, H., 1980b: A system for automated forecasting guidance. *Aust. Met. Mag.,* 28:141-154.

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

Stern, H., 1996: Statistically based weather forecast guidance. *Ph. D. Thesis, School of Earth Sciences, University of Melbourne (to be published in 1998/99 as a Meteorological Study, Bur. Met., Australia).*

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.,* 74, 2317-2330

Traction, M. S., and E. Kalnay, 1993: Ensemble forecasting at NMC: Operation implementation. *Weather and Forecasting*, 8, 379-398.

Wall Street Journal, 1992: *As referenced in the text.*

Wilks, D. S., 1995: Statistical methods in the atmospheric sciences. An introduction. *Academic Press, California*, 468 pp.

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

Acknowledgments. I thank my colleagues at the BoM for their helpful reviews of this paper, Ken Dickinson for his comments on the use of PoP values, Geoff Feren for his comments on converting the words into numerical values, and Noel Fitt for his comments on comparing the forecasts with climatology.