The accuracy of weather forecasts for Melbourne, Australia

Harvey Stern*

Bureau of Meteorology, Melbourne, Australia

This is an abridged version of a paper that was published in METEOROLOGICAL APPLICATIONS Meteorol.Appl.15: 65–71 (2008)

The complete version of the paper is published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/met.67

*** Correspondence to: Harvey Stern, Bureau of Meteorology, Melbourne, Australia. Email: h.stern@bom.gov.au**

ABSTRACT: An analysis of the accuracy, and trends in the accuracy, of medium-range weather forecasts for Melbourne, Australia, is presented. The analysis shows that skill is evident in forecasts of temperature, rainfall, and qualitative descriptions of expected weather up to 7 days in advance. The analysis also demonstrates the existence of a long-term trend in the accuracy of the forecasts. For example, Day-3 forecasts of minimum temperature in recent years (average error ~ 1.6 °C) are as skilful as Day-1 forecasts of minimum temperature in the 1960s and 1970s, whilst Day-4 forecasts of maximum temperature in recent years (average error ~ 2.0 °C) are more skilful than Day-1 forecasts of maximum temperature in the 1960s and 1970s. It is suggested that this trend may be largely attributed to: a combination of (1) enhancements in the description of the atmosphere's initial state provided by remote sensing and other observational technologies; (2) advances in broad-scale numerical weather prediction (NWP); and (3) improvements in the forecast process that are supported by good organizational management, including developing and implementing new prediction techniques, and careful succession planning.

1. Introduction

With the ongoing availability and increasing capacity of high-performance computing, improved techniques for data assimilation and new sources and better use of satellite information, improvements in the skill of Numerical Weather Prediction (NWP) systems have been well documented (refer, for example, to Wilks, 2006). Although one might expect that these improvements would naturally translate into improved public weather forecasts of surface temperature, precipitation, and qualitative descriptions of expected weather, quantitative assessment of the improvement in forecasts of these weather elements are not generally available. The primary aim of the current study is to provide such an assessment to serve:

1 as a quantitative history of improvements in weather forecasting, and,

2 as a benchmark and current state of the art of actual weather forecasting for Melbourne, Australia.

2. Background

Some years ago, Stern (1999a) presented the results of a 1997 experiment to establish the then limits of predictability. The experiment involved verifying a set of subjectively derived quantitative forecasts for Melbourne up to 14 days. These forecasts were based upon an interpretation of the National Center for Environmental Prediction (NCEP) ensemble mean predictions. The verification data suggested that, at that time, routinely providing skilful day-to-day forecasts beyond Day-4 would be difficult, but that it might be possible to provide some useful information on the likely weather up to about one week in advance for some elements and in some situations. The data also suggested that in some circumstances even the 3- to 4-day forecasts would lack skill.

In April 1998, the Victorian Regional Forecasting Centre (RFC) of the Australian Bureau of Meteorology commenced a formal trial of forecasts for Melbourne up to 7 days. Dawkins and Stern (2003) presented analyses of results of these forecasts, which show an increase in forecast skill over the first 4 years of the trial. Since then, in addition to advances in NWP, there have also been improvements in techniques for statistically interpreting the NWP model output for weather variables utilizing objective methods. Using data from the formal trial, as well as a set of experimental forecasts for beyond 7 days, Stern (2005) found that, for the first time, there was preliminary evidence of some skill up to Lorenz's 15-day limit (Lorenz, 1963, 1969a, b, 1993), particularly for temperature.

3. Results and discussion

Until the 1980s, temperature forecasts in Australia were prepared for just the next 24 h. At about that time, worded forecasts and predictions of maximum temperature up to 4 days were first issued to the public. From the late 1990s, this service was extended to minimum temperature. Experimental worded forecasts up to 7 days, with corresponding predictions of minimum temperature, maximum temperature, and rainfall amount, were also commenced. Around 2000, these predictions were made available to special clients and, since early in 2006, they have been issued officially to the public. Table I presents a summary of the current level of accuracy of Melbourne Day-1 to Day-7 forecasts (based on the most recent 12 months of data – ended 31 May 2007).

To verify predictions of precipitation, worded forecasts have been assigned to one of five categories:

1 Fine or Fog, then fine (no precipitation, either specifically referred to or implied in the forecast);

2 Mainly fine or Change (no precipitation specifically referred to in the forecast, but the wording implies that it is expected);

3. Drizzle or Shower or two (light precipitation expected at some time during the forecast period);

4. Showers or Few showers (moderate, intermittent precipitation expected during the forecast period);

5. Rain or Thunder (moderate to heavy precipitation expected during the forecast period).

Figure 1(a) shows verification of Day-1 to Day-7 forecasts of measurable precipitation (0.2 mm or greater) over a 24-h midnight-to-midnight period expressed as a probability. For category 1, 'Fine or Fog, then fine', small probabilities indicate skilful forecasts. For category

5, ‘Rain or Thunder’, large probabilities indicate skilful forecasts. It can be seen that a forecast of ‘Fine or Fog, then fine’ (category 1) is associated with only about a 5% chance of precipitation at Day-1, and that even for Day-7, there is only about a 25% chance of precipitation occurring following a forecast of category 1 weather. For category 5, even 7 days in advance, the results indicate an 80% chance of precipitation when the forecast is indicating ‘Rain or Thunder’.

Figure 1(b) shows verification of Day-1 to Day-7 forecasts of precipitation expressed as amount of precipitation. It can be seen that a forecast of ‘Fine or Fog, then fine’ for Day-7 is associated with an average fall of only about 1 mm of rain, whilst a forecast of ‘Rain or Thunder’ is associated with about 3.5 mm of rain.

Figure 2(a) and (b) shows, respectively, 12-month running (calculated over the preceding 365 days) average errors of the minimum and maximum temperature forecasts, for which data back to the 1960s are available. The graphs show a clear long-term trend in the accuracy of these forecasts. For example, Day-3 forecasts of minimum temperature in recent years (average error ~ 1.6 °C) are as skilful as Day-1 forecasts of minimum temperature in the 1960s and 1970s, whilst Day-4 forecasts of maximum temperature in recent years (average error ~ 2.0 °C) are more skilful than Day-1 forecasts of maximum temperature in the 1960s and 1970s.

Figure 2(c) and (d) show, respectively, scatter plots of forecast versus observed maximum temperatures during the first 10 years and last 10 years of available data. That the plot of the data for the last 10 years (Figure 2(d)) is very much less scattered than the plot of the data for the first 10 years (Figure 2(c)) underlines the increasing accuracy of the forecasts.

4. Concluding remarks

This paper documents the trends in accuracy and the current skill level of forecasts of weather elements at Melbourne, Australia. The city is famous for its highly variable weather and thus provides a challenge for day-to-day weather forecasting. Day-3 forecasts of minimum temperature are currently as skilful as Day-1 forecasts of minimum temperature in the 1960s and 1970s, whilst Day-4 forecasts of maximum temperature are currently more skilful than Day-1 forecasts of maximum temperature in the 1960s and 1970s. By Day-7, there is of course, reduced skill, however. Figure 1(a) and (b) demonstrate that worded forecasts of precipitation, even at Day-7, possess positive skill.

Stern (1996) suggested that improvements in weather forecasts are likely related to improved capability in predicting the broad-scale flow, and to maintaining forecaster experience in the forecast office. The former can be largely attributed to a combination of an enhancement in the description of the atmosphere’s initial state provided by remote sensing and other observational technologies, and to advances in broad-scale NWP. The latter may be related to improvements in the forecast process that are supported by good organizational management, including developing and implementing new prediction techniques, and careful succession planning. To achieve further improvement in the prediction of weather, an ongoing commitment to research into NWP, specification of the atmosphere, and to maintaining forecaster experience in the office – the importance of forecaster experience is underlined by the results of a study by Gregg (1969) – seems desirable.

Acknowledgements

The author thanks colleagues Noel Davidson and Mark Williams for encouraging this work; internal Bureau of Meteorology reviewers Bob Seaman, Tony Bannister, and Evan Morgan; and two anonymous Meteorology Applications reviewers for their helpful advice; and Terry Adair and Robert Dahni, for their development of the forecast verification datasets.

Bibliography

Dawkins SS, Stern H. 2003. Trends and volatility in the accuracy of temperature forecasts. In the International Conference on Southern Hemisphere Meteorology and Oceanography, Wellington, New Zealand, 24–28 Mar., 2003.

de la Lande J, Hagger RJ, Stern H. 1982. Melbourne forecasts–good or bad? Meteorology Australia 2(1): 2.

Gregg GT. 1969. On Comparative Rating of Forecasters. ESSA Technical Memorandum WBTM SR-48.

Lashley S, Lammers A, Fisher L, Simpson R, Taylor Weisser J, Logsdon D. 2008. Observing verification trends and applying a methodology to probabilistic precipitation forecasts at a National Weather Service forecast office. 19thConferenceonProbabilityandStatistics. American Meteorology Society: New Orleans, 20–24 Jan., 2008.

Lorenz EN. 1963. Deterministic, non-periodic flow. Journal of the Atmospheric Sciences20: 130–141.

Lorenz EN. 1969a. Atmospheric predictability as revealed by naturally occurring analogues. Journal of the Atmospheric Sciences26: 636–646.

Lorenz EN. 1969b. The predictability of a flow which possesses many scales of motion. Tellus21: 289–307. Lorenz EN. 1993. The Essence of Chaos. University of Washington Press: Seattle.

Sanders F. 1979. Trends in skill of daily forecasts of temperature and precipitation. Bulletin of the American Meteorological Society 60: 763–769.

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

Stern H. 1980. An increase in the skill of Australian temperature forecasts. AustralianMeteorologicalMagazine28: 223–228.

Stern H. 1986. A trend in the skill of Australian temperature forecasts. MeteorologyAustralia4(2): 1–4.

Stern H. 1996. Statistically based weather forecast guidance. Ph. D. Thesis, School of Earth Sciences, University of Melbourne.

Stern H. 1999a. An experiment to establish the limits of our predictive capability for Melbourne. AustralianMeteorologicalMagazine48: 159–167.

Stern H. 1999b. Statistically Based Weather Forecast Guidance, Meteorological Study 43, Bureau of Meteorology: Australia.

Stern H. 2005. Establishing the limits of predictability at Melbourne, Australia, using a knowledge based forecasting system and NOAA’s long-range NWP model. AustralianMeteorologicalMagazine54: 203–211.

Wilks DS. 2006. Statistical Methods in the Atmospheric Sciences, 2nd edn. Elsevier: New York.

Table 1: The current level of accuracy of Melbourne's Day-1 to Day-7 forecasts derived from 12 months of data ended 31 May 2007.

Where the verification statistic, '% variance explained' is quoted, the statistic refers to the 'percentage variance of the observations explained' by the predictions derived from a regression relationship between observed and forecast values.

The forecasts of √(rain amount), rather than (rain amount), are verified. This is because the distribution of the observed variable 'rain amount' is highly skewed, the variable √(rain amount) being preferred for forecast verification purposes on account of its more normal distribution.

Element	**Verification Parameter**	**Day-1**	**Day-2**	**Day-3**	**Day-4**	**Day-5**	**Day-6**	**Day-7**
√ (Rain Amount)	% Variance Explained	39.8	36.2	30.6	25.0	18.4	10.9	6.6
Min Temp	% Variance Explained	74.8	59.0	53.7	47.4	28.3	23.0	13.9
Max Temp	% Variance Explained	79.7	71.7	62.4	55.7	40.6	31.1	20.5
Thunder	Critical Success Index (%)	27.9	26.1	25.6	16.3	11.6	9.5	5.1
Fog	Critical Success Index (%)	35.3	28.9	15.9	11.4	4.9	2.4	0.0

Figure 1a The probability of precipitation occurring following the use of various phrases in Melbourne's forecasts (average over all cases 42.1%).

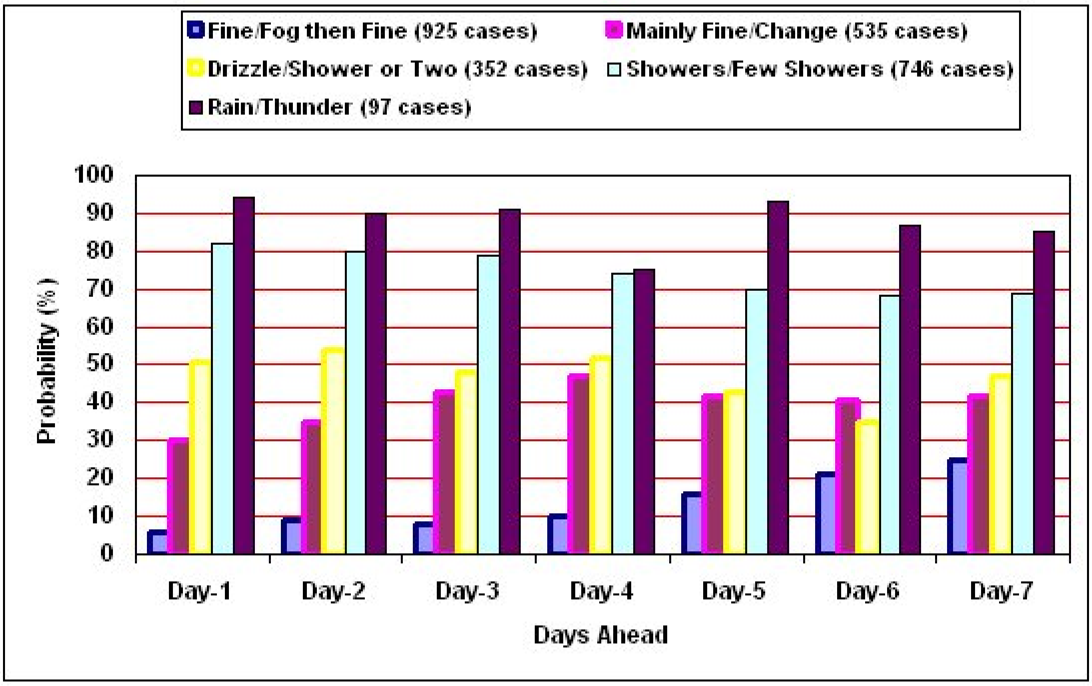

Figure 1b The amount of precipitation (mm) occurring following the use of various phrases in Melbourne's forecasts (average over all cases 1.6 mm).

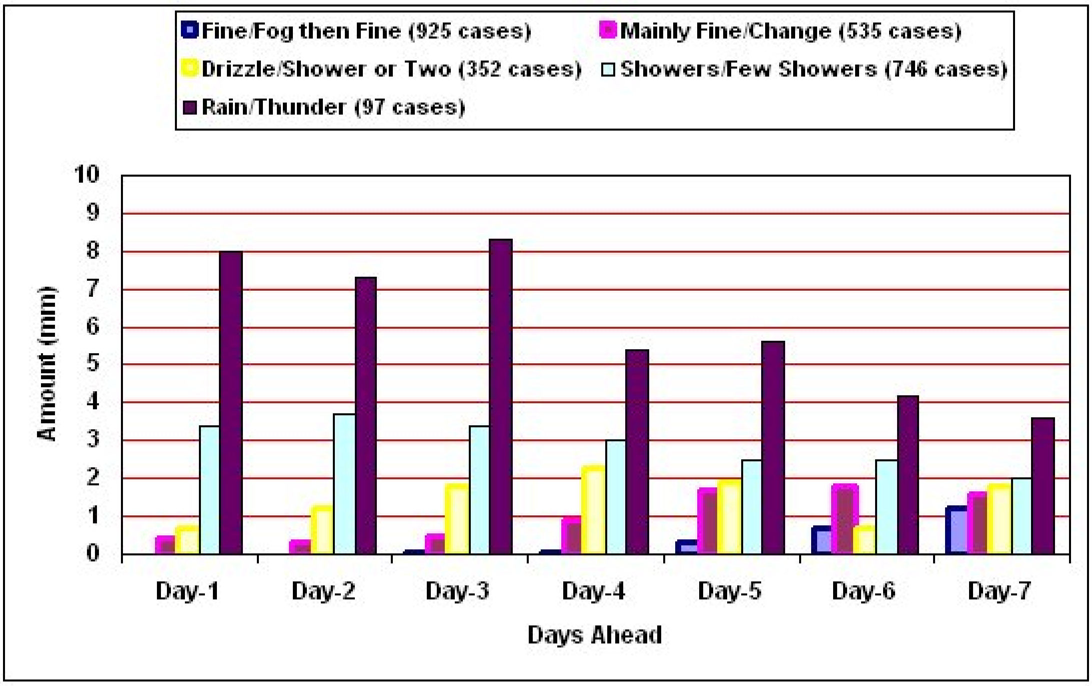

Figure 2a Trend in the accuracy of Melbourne's minimum temperature forecasts (°C) during preceding 365 days for Day-1, Day-2, ... , Day-7.

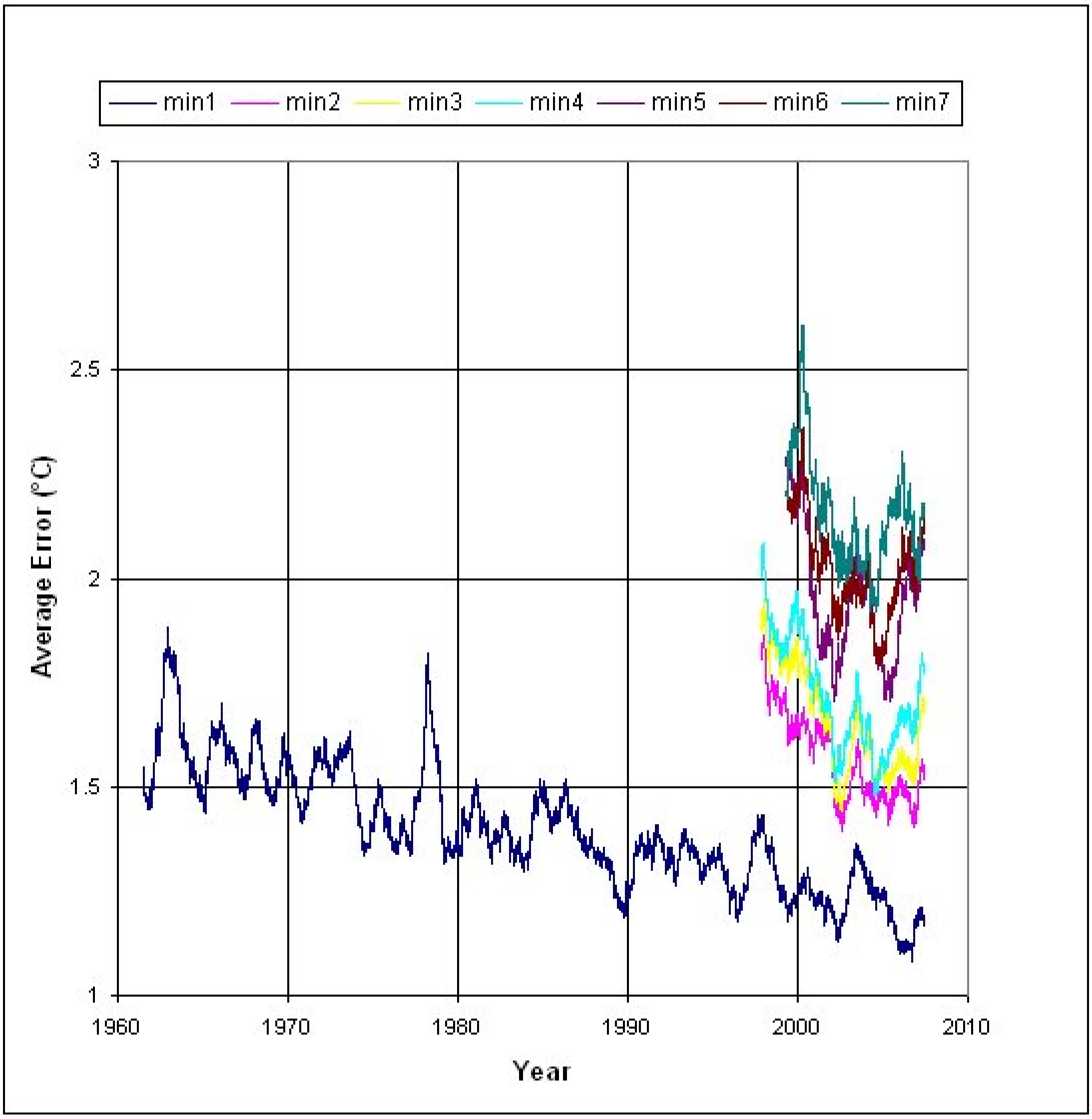

Figure 2b Trend in the accuracy of Melbourne's maximum temperature forecasts (°C) during preceding 365 days for Day-1, Day-2, … , Day-7.

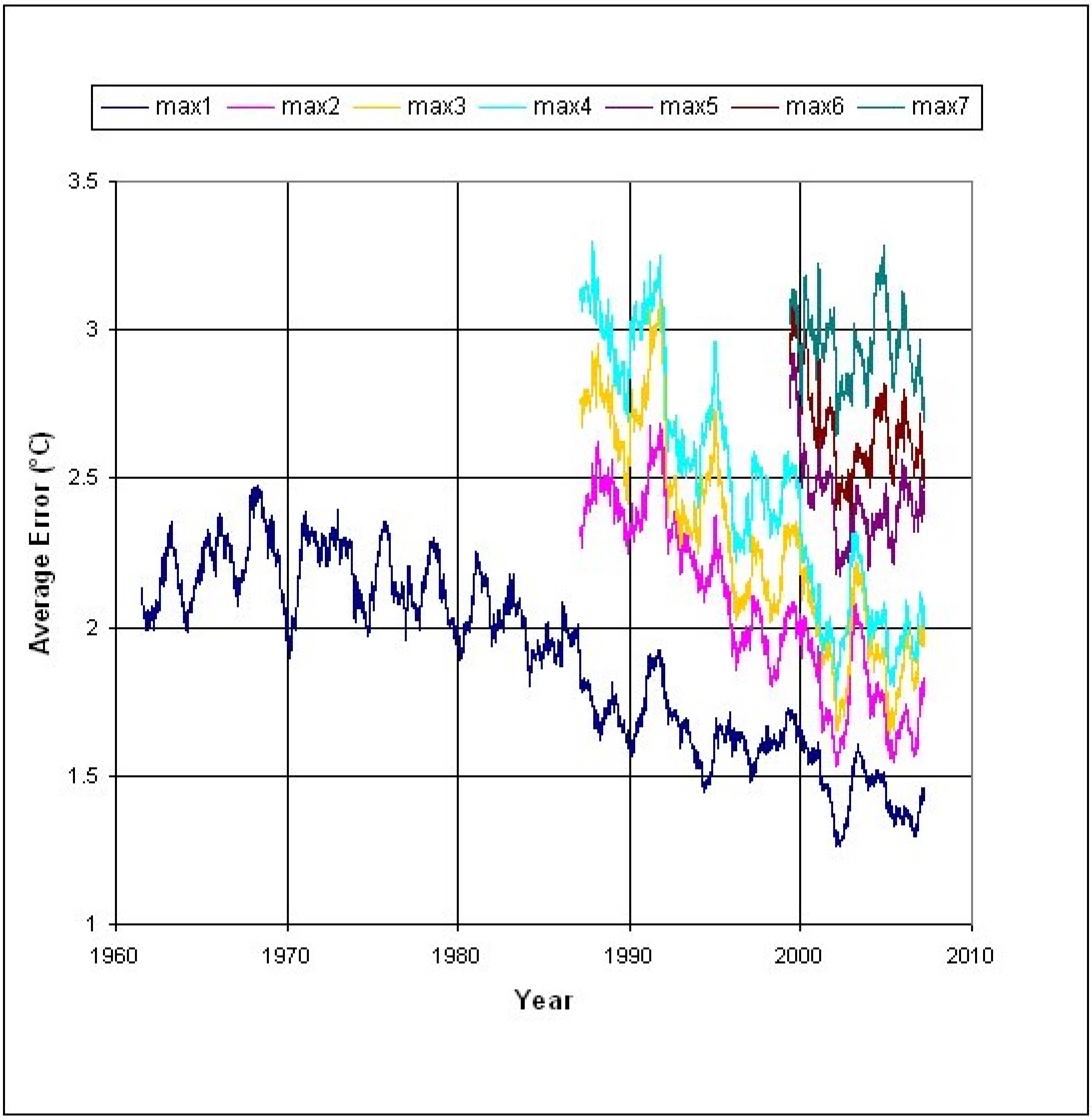

Figure 2c Scatter plot of Day-1 forecast versus observed maximum temperature (°C) during the first ten years of available data (1 July 1960 to 30 June 1970).

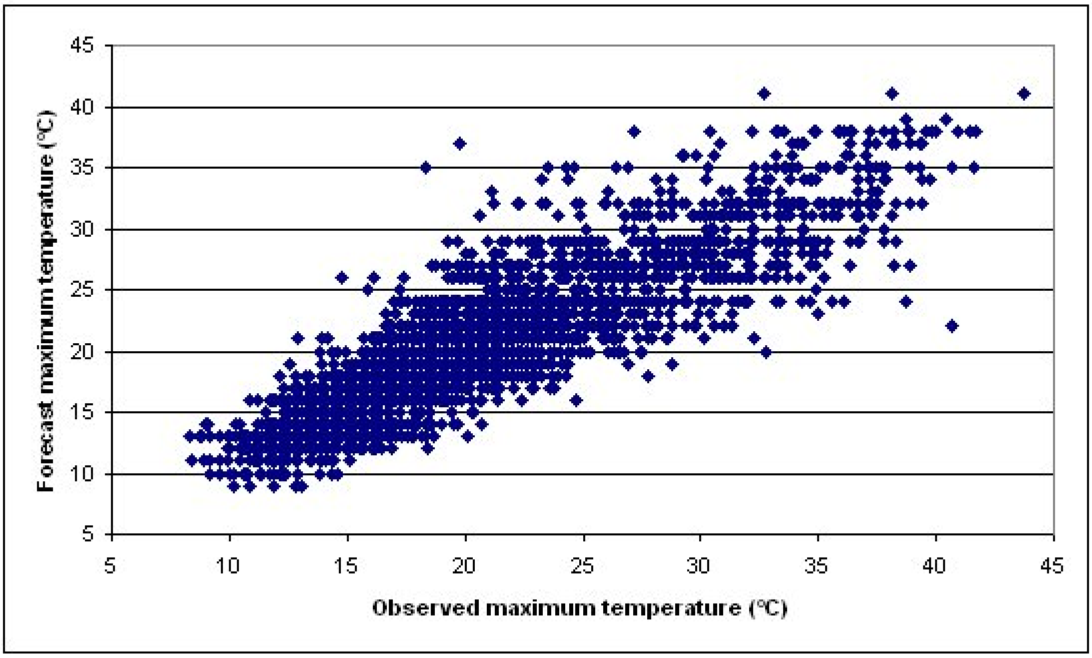

Figure 2d Scatter plot of Day-1 forecast versus observed maximum temperature (°C) during the most recent ten years of available data (1 June 1997 to 31 May 2007).