A SYSTEM FOR AUTOMATED FORECASTING GUIDANCE

H. Stern
Regional Office, Bureau of Meteorology, Melbourne*

(Manuscript received June 1979; revised September 1980)

ABSTRACT

An algorithm capable of providing automated guidance for the prediction of all elements included in a public weather forecast is described. A number of analogues to the initial circulation, or to a prediction of the next day's circulation, are retrieved and the guidance is based on a statistical analysis of associated observed weather.

During the spring of 1979 a pilot model employing the analogue statistics approach, which was designed specifically to provide guidance for Melbourne and only during the month of October, was subjected to trial. The performance of the system during this period is shown, using both statistical and subjective verification techniques, to be not significantly different from that of the official forecasts. The output of the model was not made available to operational meteorologists. Evidence is presented that suggests the operational availability of the model's output during the spring of 1979 would have improved the Bureau product.

INTRODUCTION

During the past decade there has been considerable work towards the development of systems to provide the operational meteorologist with automated forecasting guidance (Klein 1970; Weigel 1973; Bengtsson 1976; Klein 1976). Some of this guidance has been in worded form (Glahn 1970; Lonnqvist 1973; Glahn 1976; Annette 1977). A variety of approaches have been applied and these include the use of analogue retrieval techniques (Thormeyer 1971; Jarrel et al. 1975; Annette 1978). Wagoner (1973) notes that 'a large portion of the forecaster's subjective prognosis is nothing more than an analog procedure. He simply searches his mind for situations similar to the one presently confronting him. These are then converted into a modified forecast by mentally determining the average outcome of all the situations'. It is the purpose of this paper to describe the development of a system of automated forecasting guidance that employs such a procedure. The characteristics of the models that were designed in association with the project are presented in detail by Stern (1979).

* This work was conducted under the supervision of the Department of Meteorology, University of Melbourne.

141
DEVELOPMENT OF THE MODEL

A preliminary system to provide worded weather forecasts for the city of Melbourne was developed during 1974 and the early part of 1975. This model recovers a single analogue and the forecast is a description of the weather observed on the day following that of the most analogous synoptic situation to the current synoptic situation. The model was subjected to a test under simulated operational conditions during October 1975. The forecasts produced by the system were found to be significantly poorer than those obtained using manual methods.

It is the author's opinion that the failure of the single analogue model to provide useful guidance arose largely from its treatment of the forecasting problem as one with a unique solution and not as one with a spectrum of possible solutions and, by implication, a 'most probable' solution. To illustrate - only a limited number of data are used to describe the state of the atmosphere for the purpose of analogue retrieval; hence, associated with any particular set of data there is a spectrum of possible weather events dependent upon the possible values of variables not specified by this data set. Improved results may be obtained if a set of analogues are retrieved and statistical techniques, such as regression analysis, are applied to this set in order to yield a forecast. Indeed, regression analysis is used by the algorithm developed to establish the relationship between weather elements required to be forecast and variables not used to describe the state of the atmosphere for the purpose of analogue retrieval. By doing this, the non-linearity in the evolution of weather events is allowed for. It was also considered that the accuracy of the guidance provided by the model could be greatly increased if it had the facility to withdraw analogues to prognoaes of the next day's synoptic situation. This facility would enable influences from outside the area over which analogues to the current synoptic situation are retrieved to be considered.

Regression techniques have been used as a means of predicting a variety of weather elements (Bermowitz 1975; Bocchieri et al. 1974; Carter and Glahn 1976; Hammons et al. 1976). Paegle (1974) stratified data according to synoptic type and developed equations for the prediction of the probability of precipitation for each of these synoptic types. Each time a forecast is required the appropriate equation is applied. It is the author's view, however, that the analogue statistics approach would be superior to that of Paegle. This is because each time a forecast is required a new equation is developed on data from those situations most relevant to the forecast situation. The analogue statistics approach has been used hitherto in the designing of systems to forecast the movement of tropical cyclones (e.g. Jarrel et al. 1975). It is believed, however, that this is the first time it has been used in the designing of a system for the prediction of all elements in a public weather forecast.

A forecasting model based on analogue statistics was then devised. The model was designed to provide guidance only during the month of October. This month was selected because during October a fairly even balance of summer and winter synoptic types is observed and the weather in Melbourne is particularly changeable. If a successful forecasting system could be devised for October, it should be successful in other months. The main features of the analogue statistics system are summarised in Fig. 1. The operator of the model determines whether analogues to the current or
forecast situation are retrieved, the number of such analogues, and the weight to be given to each of the parameters used to specify the situation. Up to four regression equations may be established - for the prediction of minimum temperature, maximum temperature, probability of fog occurrence, and 'weather type' (0 = Fine; 1 = Marginal; 2 = Rainy). Terms in the equations are selected by the model's operator and may include parameters for tomorrow that must be forecast.

The algorithm is for the Melbourne 'Tomorrow' forecast, which is issued at 1630 EST. These forecasts are valid from midnight to midnight and include a prediction of overnight minimum temperature and daytime maximum temperature. They are distributed to interstate forecasting centres.

Incidentally, Woodcock (1980) is working on the development of an automated system to predict Sydney's maximum temperature. It establishes regression relationships on data stratified by means of an analogue retrieval technique. Woodcock has shown that equations developed on stratified data perform better than equations developed on non-stratified data.

**The procedure for analogue retrieval**

The means by which the model retrieves analogous synoptic situations to the current or forecast synoptic situation is now presented. The data base of the system covers each of the 527 synoptic situations leading up to each October day from 1957 to 1973 inclusive. Each synoptic pattern is described in terms of the surface flow, the 500 mb level flow, and the change in the surface flow during the preceding 24 hours. The three fields are defined by data from the twelve southern Australian stations - Perth, Albany, Esperance, Forrest, Kalgoorlie, Adelaide, Mount Gambier, Woomera, Cobar, Nowra, Laverton, and Hobart. Their locations are given in Fig. 2. It is considered that data from these stations describe most synoptic features relevant to Melbourne's weather. The data types used to describe the basic atmospheric circulation systems and their rate of movement are:

(i) Departure of the 1200 EST MSL pressure at each station from the mean of the twelve 1200 EST MSL pressures (or, alternatively, from the mean of a subset of the twelve 1200 EST MSL pressures);

(ii) Departure of the 0900 EST 500 mb level height at each station from the mean of the twelve 0900 EST 500 mb level heights (or, alternatively, from the mean of a subset of the twelve 0900 EST 500 mb level heights);

(iii) 0900 EST 900 mb level wind direction and speed at each station;

(iv) 0900 EST 500 mb level wind direction and speed at each station;

(v) Change in MSL pressure, during the 24-hour period ended 1200 EST (on the day of the situation), at each station; and

(vi) Vector change in 900 mb level wind, during the 24-hour period ended 0900 EST (on the day of the situation), at each station.

In the event of any data item not being recorded by a station, the value of this item was determined from synoptic scale analyses.
Fig. 1 A schematic diagram indicating some of the procedures that comprise the analogue statistics method.

Fig. 2 Location of stations whose data are used for the selection of analogues or referred to elsewhere in the text.
Data types (i) and (iii) define the surface flow, data types (ii) and (iv) define the mid-tropospheric flow, while data types (v) and (vi) define the change in the surface flow during the preceding 24 hours.

The operator of the model possesses the facility during the analogue retrieval procedure, on the basis of experience with using the algorithm, to vary the weight given to:

- the relative value of the stations;
- the data types;
- the twelve MSL pressures upon which the mean MSL pressure is calculated, during that calculation; and
- the twelve 500 mb level heights upon which the mean 500 mb height is calculated, during that calculation.

**The similarity parameter**

The basis for analogue selection is a 'similarity' parameter that was developed by the author. The parameter provides a measure of the difference between two synoptic situations defined in terms of the data types referred to earlier. The rationale behind its design was to develop a 'score' that represents the sum of the root mean square differences between the data types used to specify the situations.

The similarity parameter $R$ is defined as follows:

Suppose it is required to compare synoptic situation $j$ with synoptic situation $k$. Refer to data type $t$, associated with station $u$, from situation $v$, as $N_{tvu}'$. Refer to the weight given to data type $t$ as $W_t'$, refer to the weight given to station $u$ as $W_u'$ and refer to the standard (scalar or vector) deviation about the (scalar or vector) mean of data type $t$ as $S_t'$. The value of the similarity parameter comparing the two synoptic situations is then given by

$$R_{jk} = 100 \sum_{t=1}^{6} W_t \sum_{u=1}^{12} \left( \left| N_{tu} - N_{tu} \right| \right)^{2,1/2}$$

where the differences are scalar if $t$ is a scalar quantity, the differences are vector if $t$ is a vector quantity,

$$\sum_{t=1}^{6} W_t = 12, \text{ and } \sum_{u=1}^{12} W_u = 12.$$  

The factor 100 is a convenient scaling factor.
The role that the $S_t$ play in normalising the data types relative to each other may be seen from Eqn 1. The $S_t$ are determined using another computer model as follows: refer to the mean (scalar or vector) of items of data type $t$ associated with the 527 synoptic situations (17 months) as $M_t$ where

$$M_t = (527)^{-1} \sum_{v=1}^{527} \sum_{u=1}^{12} W_{uv} N_{tuv} \quad \ldots 2$$

the summations being scalar if $t$ is a scalar quantity and vector if $t$ is a vector quantity.

The $S_t$ are then given by

$$S_t = (527)^{-1} (12)^{-1} \left\{ \sum_{v=1}^{527} \sum_{u=1}^{12} W_{uv} \left( N_{tuv} - M_{t} \right)^2 \right\}^{1/2} \quad \ldots 3$$

where the differences are scalar if $t$ is a scalar quantity and the differences are vector if $t$ is a vector quantity.

The model retrieves the most analogous synoptic situation to the current or forecast situation, according to the criteria of selection and the weightings used, by comparing the situation with each of the 527 situations in turn. The similarity parameter $R$ is calculated for each comparison and the 'best' analogue is regarded as that one associated with the smallest value of $R$. Similarly, the 'second best' analogue is regarded as that one associated with the second smallest value of $R$ and the 'third best' analogue is regarded as the one associated with the third smallest value of $R$. The process is repeated until the required number of analogues, that is, the number specified by the operator of the model, has been retrieved.

Operating the model

The model may be directed to derive regression equations for the purpose of obtaining predictions of quantitative aspects of the next day's weather. Terms to be included in the regression equations are selected by the operator from a wide variety of variables including parameters for the next day that must be forecast. The variables were chosen for their ability to play a significant role in determining various aspects of the next day's weather or because they appeared, on the basis of the author's experience, to be highly correlated with these aspects. Incidentally, selection of variables in this way provides only a subjective base to the system. A selection based on variance reducing potential would raise the objectivity, enhance the acceptance by operational meteorologists, and allow systematic future development. The author considers this to be an area where further work on the model should be directed.
In summary, the operator decides upon the number of analogues to be retrieved, the weight to be given to each of the data types and stations in the analogue retrieval procedure, and the terms to be included in the regression equations. Finally, he may, if he is reasonably confident about a particular aspect of the next day's weather, 'force' the system to predict that aspect along the lines he wishes. The machine is then used only to forecast aspects of the weather about which significant uncertainty exists. While this arbitrary approach to the use of the model is a justifiable use of the information available, in any objective test of the system it is not valid to apply this facility. It will be seen that when the model was tested, a completely objective approach was used.

**A TEST OF THE MODEL**

During the spring of 1978, and again during the spring of 1979, the analogue statistics model was subjected to trials in order to assess its potential skill. The trials were effectively conducted in a 'real-time' manner but the model was not 'run' until some time after the time of issue of the Bureau prediction. Examples of the system's guidance are:

**For 14 October 1978**

*Weather:* A little rain during the morning. Showers later.
*Wind:* Moderate northwesterly winds turning light southwesterly.
*Min:* 10°C  *Max:* 17°C  *Rainfall:* 3 mm
*Precis:* Little Rain
*Confidence:* Average

**For 23 October 1978**

*Weather:* Fine and sunny.
*Wind:* Fresh north to northeasterly winds.
*Min:* 16°C  *Max:* 29°C  *Rainfall:* Nil
*Precis:* Sunny, Windy.
*Confidence:* High

It may be noted that the model includes in its output a quantitative precipitation forecast as well as providing an indication of the level of confidence that can be had in the accuracy of the guidance. The model does not print out the entire forecast - parts of the prediction are obtained by manual reference to particular files. These files are not computer files - they are tables of numbered words, phrases, and sentences that correspond to various aspects of objectively derived descriptions of the weather associated with appropriate analogues. The 1978 trial is described in detail by Stern (1979). It showed that the model is 'capable of providing guidance of some skill'. The remainder of the paper is devoted to discussing the 1979 trial.

Although the model's data base is restricted to October situations, the trial was conducted from 1 September to 30 November inclusive - for 91 days. The results of a trial, including calendar months other than the one upon whose data the model was developed, might be expected to underestimate its skill. Nevertheless, it was considered that the benefit to be gained from a longer trial, in terms of the significance of the results, outweighed this disadvantage.
The weights to be given to the parameters used in the analogue retrieval step, and the terms to be included in the regression equations during the statistical analysis step, were determined by a set of conditions related to the data. This set of conditions was decided upon at the beginning of the trial and was not changed at all during its course in order to obtain a consistent set of results. The conditions may be summarised as follows:

(a) Analogues were obtained using only MSL pressure data over southeastern Australia, initial and forecast, the forecast data being obtained from the operational Australian Region Primitive Equation (ARPE) model prognosis (McGregor et al. 1978) valid for 1000 EST the next day. The number of analogues retrieved is 25. Note that the prognoses are verified against data centred at about 0900 EST. Therefore, in order to avoid a timing bias in the forecasts, MSL pressure data obtained from the prognoses were taken from points 2° longitude to the west of the stations (assuming an average eastward movement of pressure systems of about 16° longitude per day, i.e. an eastward movement of 2° longitude between 0900 EST and 1200 EST). The prognoses were not adjusted to take into account development because it was considered that this would have only a marginal effect on the prognosis over a period of three hours.

(b) No terms were specified to be included in a prediction equation for tomorrow's 'weather type'. In such circumstances, the system 'averages' the weather types observed on the days of the analogues in order to obtain a prediction.

(c) Terms in the minimum temperature prediction equation were today's maximum, today's dew-point, a forecast of tomorrow's 1000-700 mb thickness at Laverton (which was related to the ARPE model's predictions of 850 mb level temperature), and tomorrow's weather type. A term related to low-level flow was not included in the equation because the analogues are selected on the basis of surface circulation and hence there is little variation of this parameter between the analogues.

(d) Terms in the maximum temperature prediction equation were today's maximum, a forecast of tomorrow's 1000-700 mb thickness at Laverton, and tomorrow's weather type.

(e) Terms in the probability of fog occurrence prediction equation (only derived when tomorrow's weather type is predicted to be '0') were today's maximum and today's dew-point depression.

(f) If the recent trajectory of the airmass over southeastern Australia had been from the north, dew-point readings at certain upstream stations exceeded particular values, and the SLVH model (Falconer and Fitt 1977), as driven by the ARPE model, predicted low or negative saturation deficits near Melbourne, the analogue statistics system was 'forced' to predict no less than a certain minimum amount of precipitation (e.g. in these circumstances a minimum of 5 mm is forecast if the ARPE model predicts negative saturation deficits within 120 nautical miles of Melbourne, west of Melbourne's longitude). Note that although upper air data
were not used in the retrieval of analogues to the predicted
circulation, vertical motion processes were partly taken into
account by means of the use of the SLYH model forecasts of
saturation deficit.

The 91 forecasts produced by the model were compared with the official
Melbourne observations. The skill displayed during the trial period by the
model's forecasts, forecasts issued officially by the Bureau, and a set of
persistence forecasts, will now be discussed.

The performance of the system during its trial

The purpose of this sub-section is to tabulate the relative accuracy of the
forecasts produced by the system using ARPE model data objectively, and the
subjective Bureau predictions. The accuracy of quantitative and
qualitative aspects of the forecasts during the spring of 1979 is
summarised in Table 1. Verification data specifically for the month of
October are also presented.

Table 1  A summary of the model's performance during the spring of 1979

<table>
<thead>
<tr>
<th>Verification parameter</th>
<th>Combined data for September, October, and November (91 cases)</th>
<th>Data specifically for October (31 cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model</td>
<td>Bureau</td>
</tr>
<tr>
<td>r.m.s. error of maximum temperature forecasts (°C)</td>
<td>2.77</td>
<td>2.45</td>
</tr>
<tr>
<td></td>
<td>[+0.31]</td>
<td>[+0.14]</td>
</tr>
<tr>
<td>r.m.s. error of minimum temperature forecasts (°C)</td>
<td>1.85</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td>[-0.43]</td>
<td>[-0.47]</td>
</tr>
<tr>
<td>r.m.s. error of quantitative precipitation forecasts (Ranges)</td>
<td>1.10</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td>Skill score</td>
<td>6.70</td>
<td>6.96</td>
</tr>
</tbody>
</table>

Note: (1) Values of verification parameters that are significantly
different (at the 5% level) from the model value are given in
italics.

(2) Two values are given for the r.m.s. error of the Bureau's
quantitative precipitation forecasts. The first one given is for
the 18-hour estimate; the second one given is for the 24-hour
estimate.

(3) Mean biases (forecast minus observed) of the temperature
forecasts are given in square brackets.
A number of points about various aspects of the verification statistics will now be made.

(a) Temperature

The table shows that, although both sets of the model's forecasts of minimum and maximum temperature were superior to estimations based on the assumption of persistence, they were not significantly different to those issued officially.

(b) Precipitation amount

Each prediction produced by the system is accompanied by a quantitative precipitation forecast (QPF) for the midnight to midnight period. The Bureau does not prepare a QPF for this period. Nevertheless one is prepared for the city, at 0600 EST for the ensuing 18 hours and another, for the metropolitan area, is prepared at 2000 EST the previous evening for the 24-hour period commencing 0900 EST. The Bureau QPFs are not issued to the general public and are given in precipitation ranges - Range 0 represents no precipitation, Range 1 represents precipitation up to 2.5 mm, and Range x, where x is an integer greater than one, represents precipitation greater than \(2.5(2^{x-2})\) mm and up to \(5.0(2^{x-2})\) mm.

The table shows that both sets of the model's forecasts of precipitation quantity were superior to corresponding persistence estimates. Although it is not strictly valid to compare verification statistics of predictions for slightly different time periods, the table does suggest that the quality of the model's forecasts was not significantly different from that of those issued officially.

(c) Overall skill

No comprehensive and objective scheme for the verification of public weather forecasts has been described in the literature to date. The author therefore derived such a scheme and this is discussed in detail by Stern (1979). It evaluates the overall accuracy of a public weather forecast in terms of weather, wind, and temperature, in order to give it a skill score. The maximum possible score is ten. The main features of the scheme are now summarised:

- Up to four points may be awarded for the 'weather' part of the forecast - one for correctly indicating the occurrence or non-occurrence of significant phenomena, one for correctly timing the occurrence of those phenomena, one for accurately indicating the main feature of the day's weather (for example, cloud cover in the event of fine conditions, intensity of rain in the event of wet conditions), and one if the weather part of the forecast is perfect in every respect.

- Up to two points may be awarded for the 'wind' part of the forecast - one for direction and one for speed.

- Up to two points may be awarded for each of the maximum and minimum temperature forecasts - two if the forecast is in error
by 2°C or less, one if the forecast is in error by 3°C, and one if the forecast is in error by 4°C and the observed temperature is 25°C or higher.

A serious error in any part of the forecast, for example, a temperature prediction error of 6°C or more, results in the total being reduced to four, irrespective of the skill that may have been displayed in other parts of the forecast.

The table shows that, although both sets of the model's forecasts were superior, overall, to corresponding sets based on the assumption of persistence, they were not significantly different from those issued officially.

A case may be made to support the proposition that the availability of the model's output to operational meteorologists, during the trial period, would have had a positive impact on the overall quality of the official product. Stern (1979), referring to Thompson (1977), advocates that the man-machine mix forecast be obtained by a consensus approach. To illustrate, had official predictions of maximum and minimum temperature been obtained during the period of the trial by simply averaging the machine's estimate and the man's estimate, the r.m.s. error of the official maximum temperature forecasts would have been 2.29°C (instead of 2.45°C) and the r.m.s. error of the official minimum temperature forecasts would have been 1.72°C (instead of 2.03°C). During the month of October specifically, the r.m.s. error of official maximum temperature forecasts would have been 1.98°C (instead of 2.46°C) and the r.m.s. error of official minimum temperature forecasts would have been 1.56°C (instead of 1.81°C).

CONCLUDING REMARKS

This paper has had as its theme the development of a pilot system of automated forecasting guidance. The system developed retrieves a set of analogues to a current or forecast synoptic situation and statistically analyses data associated with the analogues in order to arrive at a prediction. The model provides guidance for the forecasting of all weather elements included in forecasts issued to the general public. In addition, its output contains a quantitative precipitation forecast and a one-word statement of the overall confidence a user might have in the guidance. The system allows for a considerable amount of man-machine interaction in its operation. The optimum method by which this interaction is applied will be approached as experience with using the model is gained.

Lönnqvist (1973) emphasises 'the inadvisability and fallibility of drawing conclusions from very short verification periods'. Nevertheless, the analogue statistics forecasting system was shown during its 'real-time' trial to be capable of providing guidance of skill not significantly different from that of the current official product. Such guidance would be of value to operational meteorologists who, because they are able to consider factors that the model does not take into account, would often be able to improve on it. In fact, the verification data presented suggest that the operational availability of the model's output during the spring of 1979 would have had a positive impact on forecasts issued officially during that season.

Only a small number of the available observational data and dynamic model output are employed by the system to derive a forecast. In the light of this, it is surprising that the system's performance during the trial
period should have approached so closely that of the official product. The results of the study, therefore, give considerable encouragement to those who look forward to the day when forecasters in Australia are provided with automated guidance for the prediction of all weather elements included in a public weather forecast.

The author considers that the results presented justify the carrying out of further work on the model with a view to enabling it to realise its potential in an operational setting. Areas where such work might be undertaken are:

- An investigation determining which situation types are associated with a model performance consistently superior to that of the Bureau, and vice versa, to throw light on methods of optimising model guidance.

- A statistical analysis of the impact on predictive skill of

  (a) varying the conditions under which the model is run (number of analogues retrieved, the weights given to the data types and stations, terms in the regression equations, etc.);

  (b) the quality of the set of analogues as indicated by the maximum value of the similarity parameter R.

- An analysis of the statistical significance of the regression equations and a modification of the algorithm so that selection of variables in the equations would be based on their variance reducing potential.

- An investigation into what aspect of the model is responsible for its skill. To illustrate, is the skill derived from the intrinsic correlation value of the parameters selected, the analogue selection screening procedure, or the system as a whole?

- The elimination of the requirement to obtain parts of the prediction by manual reference to a particular file.

- The extension of the algorithm to other calendar months and to other localities.

ACKNOWLEDGMENTS

I wish to express my appreciation to Dr T.T. Gibson, my supervisor at the University of Melbourne, for his guidance; to Drs W.J. Gibbs and J.W. Zillman, for making available the resources of the Bureau of Meteorology to this project; to Miss M. Bell, who was responsible for the computer programming associated with the study; to Messrs R.J. Hagger and A.F. Magnuson, for their handling of numerous requests for archived data; and to Drs D. Jessen and U. Radok, and Messrs N.E. Davidson, J.D. Jasper, and C.M. Taylor for their valuable suggestions. Finally, I wish to thank Mr R. Maine and the other reviewers of this paper for their helpful advice.
REFERENCES


Wagoner, R.A. 1973. A technique for using historical analogs to forecast the central pressure of tropical cyclones in the western North Pacific Ocean and South China Sea. M.S. Thesis, Department of Meteorology, Texas A & M University, College Station, Texas. 65 pp.
