Application of Artificial Neural Network Forecasts to Predict Fog at Canberra International Airport

DUSTIN FABBIAN AND RICHARD DE DEAR

Department of Physical Geography, Macquarie University, North Ryde, New South Wales, Australia

STEPHEN LELLYETT

New South Wales Regional Office, Australian Bureau of Meteorology, Darlinghurst, New South Wales, Australia

(Manuscript received 6 December 2005, in final form 2 June 2006)

ABSTRACT

The occurrence of fog can significantly impact air transport operations, and plays an important role in aviation safety. The economic value of aviation forecasts for Sydney Airport alone in 1993 was estimated at $6.8 million (Australian dollars) for Quantas Airways. The prediction of fog remains difficult despite improvements in numerical weather prediction guidance and models of the fog phenomenon. This paper assesses the ability of artificial neural networks (ANNs) to provide accurate forecasts of such events at Canberra International Airport (YSCB). Unlike conventional statistical techniques, ANNs are well suited to problems involving complex nonlinear interactions and therefore have potential in application to fog prediction. A 44-yr database of standard meteorological observations obtained from the Australian Bureau of Meteorology was used to develop, train, test, and validate ANNs designed to predict fog occurrence. Fog forecasting aids were developed for 3-, 6-, 12-, and 18-h lead times from 0600 local standard time. The forecasting skill of various ANN architectures was assessed through analysis of relative operating characteristic curves. Results indicate that ANNs are able to offer good discrimination ability at all four lead times. The results were robust to error perturbation for various input parameters. It is recommended that such models be included when preparing forecasts for YSCB, and that the technique should be extended in its application to cover other similarly fog-prone aviation locations.

1. Introduction

Weather forecasts are often viewed as the most important services provided by the meteorological profession (Leipper 1995). These forecasts however possess no intrinsic economic value unless they influence the behavior of individuals and organizations whose activities are sensitive to weather conditions (Murphy 1994). The Australian Bureau of Meteorology (BoM) provides Terminal Aerodrome Forecasts (TAFs) for all major airports in Australia with each capital city TAF being valid for 24 h. These forecasts are used by the commercial airlines and Air Services Australia (formerly the Civil Aviation Safety Authority) for flight planning, in-flight decision making, and optimization of airport operations. BoM staff routinely review their forecasting performance to seek ongoing improvement, and this study is part of a coordinated effort between Macquarie University and the BoM in relation to the latter's National Fog Project.

2. Fog and aviation

The main use of TAFs by airlines and aircraft operators is for flight planning, both pre- and intraflight. The core component of this planning is the pilot's "alternative fuel" decision (pretakeoff). Such decisions are based on TAFs issued for the intended destination. When certain criteria are exceeded, legislation requires that additional fuel be uplifted should extra flying time or diversion to an alternate landing site be necessary. When these criteria are not exceeded, depending upon the fuel policy of the airline, the pilot must decide whether to carry the additional fuel. The profit margins of most airlines are relatively low; they have been previously estimated at approximately 2% (Bonné 2005).

Corresponding author address: Richard de Dear, Dept. of Physical Geography, Division of Environmental and Life Sciences, Macquarie University, North Ryde, NSW 1209, Australia. E-mail: rdedear@laurel.ocs.mq.edu.au

DOI: 10.1175/WAF980.1

Unexpected costs can therefore shift a flight from making money for the airline, to incurring a substantial loss. Leigh (1995) calculated the estimated loss for a diverted Qantas long-haul flight with an intended destination of Sydney that was diverted to Brisbane at U.S. $31,000.

3. Choice of study site

Canberra is the political capital of Australia and is located in the Australian Capital Territory, approximately 300 km southwest of Sydney. The automatic weather station at Canberra International Airport (YSCB) is located 6 km east of the city center. The choice of YSCB for this study was based on four main criteria. First, Canberra is Australia's foggiest capital city, receiving an average of 46 fog days per year (Malley et al. 2003). Second, the majority of the fogs occurring in Canberra are of the simple radiation type, which obviates the need for the artificial neural network (ANN) to discern different fog genesis processes. Third, as a result of Canberra's political significance, population, and infrastructure, there has been a constant BoM presence for an extended period of time. The observational record for this location is of high quality. Foggier locations (e.g., Richmond in New South Wales) do not have an observational record of a similar length or quality. Last, YSCB is also of strategic importance as it serves as an important alternate destination for Australia's busiest airport, Sydney's Kingsford Smith International Airport.

Malley et al. (2003) provide a detailed climatology of fog events in Canberra. May, June, and July are the foggiest months on average while December is the least foggy (see Fig. 1).

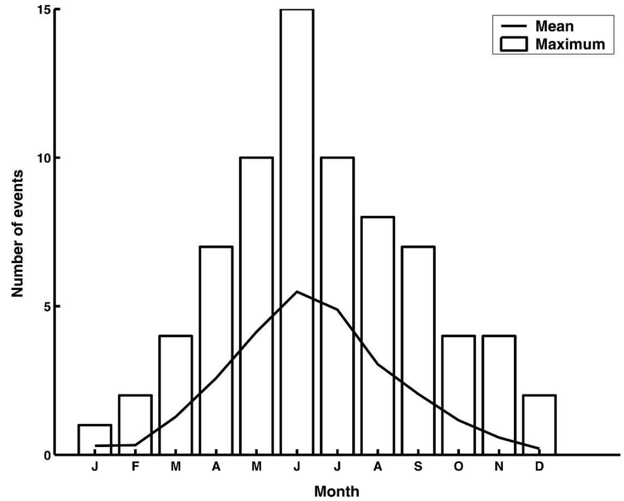

Fig. 1. Climatology of fog events per month (YSCB, 1960–2004)

Although Canberra is Australia's foggiest capital city, there still exists a disproportionate number of fog days (to no-fog days) during the fog season. This inequality leads to unequal class sizes in the training and validation datasets used. Many authors have outlined the problems such differences can cause when attempting to develop ANN-based forecasting tools (e.g., Parikh et al. 1999). The major problem that can arise when developing an ANN based on such data stems from the manner by which the backpropagation algorithm works. Flawed networks can result from the employment of the mean-squared error function with a batch training routine. In cases where the class sizes are grossly unequal, the network may converge on a solution by which the largest class is always forecast. In essence, this is a forecast with no skill, but with a very low overall error.

In this study, several different training approaches were tried as solutions to this problem. The first approach, as suggested by Parikh et al. (1999), involved adjusting the ratio of fog events to no-fog events in the developmental datasets. This did raise concerns regarding the final employment of the trained ANNs within an operational context where the fog to no-fog ratio employed within network development would not accurately reflect the eventual observational data employed.

Ultimately, however, each alternative approach failed to provide for a more skillful forecasting system. In the end, it was found that employing the raw training data provided for the best performing ANNs.

4. ANNs and the multilayer perceptron

Although ANNs have existed for over 50 yr, their first use in atmospheric science was only as recent as 1986 (Gardner and Dorling 1998; Rummelhart and McCelland 1986). Neural networks are artificial intelligence tools that have recently been applied to diverse forecasting problems such as tornado detection (Marzban and Stumpf 1996) and quantitative precipitation forecasting (Hall et al. 1999). These systems were first designed to mimic human learning processes (Lippmann 1987). ANNs are an example of "spoken data" methods; their structure and, thus, relationship between inputs and outputs are determined by the historical dataset to which they are exposed (Mehrotra et al. 1997). If suitably "trained" on a set of examples, they can theoretically "learn" or extract a link between any input and corresponding output data (Pasini et al. 2001). ANNs are known as "universal function approximators" because of their capacity to approximate virtually any continuous nonlinear function with arbitrary accuracy (Tasadduq et al. 2002). ANNs can thus be used to solve many complex problems in which a priori knowledge is incomplete or unavailable (Tresp et al. 1995).

Gardner and Dorling (1998) suggest the application of a numerical approach if adequate data, computing resources, and understanding of the problem are available. The complexity of many meteorological processes is such that our understanding, and thus the forecasting skill of many numerical techniques, are inherently limited. A further limitation of current deterministic approaches is the scale on which they operate. The Mesoscale Limited Area Prediction System (MESOLAPS; Puri et al. 1998) is a dynamic atmospheric numerical weather prediction (NWP) model run by the BoM as part of their forecasting regime. This system, referred to as MESOLAPS05 herein, has a spatial resolution of 0.05° over the Canberra locale, which is generally too coarse for accurate modeling of subgrid-scale phenomena such as fog.

Visibility forecasting is considered a complex prediction problem in meteorology because of the difficulty in quantifying the relative importance of synoptic, mesoscale, and local factors (Pasini et al. 2001). The occurrence of fog is so strongly dependent on local conditions that broadscale application of physical dynamics alone can never be sufficient to forecast it with a high degree of accuracy and consistency. While such physical reasoning can provide a good indication on the *likelihood* of fog, the decision to place fog on a TAF is most often dependant on the forecaster's local knowledge and past experience (Regano 1997). This local knowledge is not readily quantified and thus is almost impossible to incorporate into NWP approaches such as MESOLAPS05 at the grid scales currently available. It could be expected that an increase in accuracy and consistency of prediction from such models will be seen as our understanding of microphysics is improved and grid scales approach several hundreds of meters; however, this is likely to be several years off at least.

The most commonly applied ANN architecture in the atmospheric sciences is the multilayer perceptron (MLP; Gardner and Dorling 1998). An MLP consists of a set of neurons or nodes organized into an input layer, a number of "hidden layers," and an output layer. The input and output layers are connected by hidden layers. The inclusion of these additional layers marks a significant milestone in ANN development, as their addition allows for the MLP to approximate nonlinear functions. Each node's inputs are passed through a transfer function that acts to fit the neuron's outputs to a specific value range and functional shape.

MLPs were first proposed after it became clear that simple perceptrons (ANNs whose input-layer nodes are directly linked through weights to the output layer) performed poorly in comparison with other methods (Hsieh and Tang 1998). This is because simple perceptrons are only suitable for modeling linear relationships, which is why the inclusion of hidden layers marks a milestone in ANN development. There exist many different forms of MLPs in the literature, but most can be said to vary according to five main criteria: (i) direction of information flow (feedback or feedforward), (ii) training method, (iii) learning algorithm, (iv) number of hidden layers, and (v) error function.

5. Current fog forecasting techniques

BoM aviation forecasters, who, under federal legislation, are the government-appointed suppliers of aviation forecasts for YSCB, use an ensemble of aids when deciding whether to forecast fog for any particular TAF issue. These aids range in complexity from simple historical analog tools (e.g., Regano 1997) to satellite-derived observations and limited area dynamic NWP models (e.g., Puri et al. 1998). Ultimately the final decision is still heavily based on the forecaster's experience, preferences, and expert interpretation of the forecasting aids. The forecasting aids developed in this study are designed to complement existing approaches in an effort to increase overall fog forecasting skill.

Regano's (1997) historical analog fog forecasting aid produces real-time forecasts of fog probabilities for any Australian synoptic weather station based on its past climatic history. According to BoM staff, the aid generally works well when there are sufficient analogs from which to calculate fog probabilities. Forecasts based on analog sample sizes of less than 10 events (due to a short available record) are generally unusable. This method is therefore ill suited to sites where the digitization of records is incomplete, for example, YSCB.

High-resolution satellite image enhancement algorithms developed by the BoM are able to detect low cloud and fog from the Advanced Very High Resolution Radiometer sensor on board the polar-orbiting National Oceanic and Atmospheric Administration satellites (Matthews et al. 2002). While these algorithms generally work well under clear-sky conditions, they are also unable to detect low cloud or fog that is present under higher cloud (Matthews et al. 2002).

MESOLAPS05 provides gridded output for several domains over the Australian continent including one covering Canberra. It also provides specific hourly forecasts of wind and humidity fields for selected synoptic stations in Australia that include Canberra. Forecasters often regard MESOLAPS05 as providing the best general indication of how the synoptic situation is likely to evolve over the proceeding few hours. While 0.05° spacing is regarded as high resolution for any NWP model, it is nevertheless extremely difficult to accurately and consistently replicate the subgrid-scale processes involved in the formation, persistence, and eventual dissipation of mist/fog (Huang et al. 2002).

6. Methods

a. Data

The data used in this study came from two distinct forms of meteorological report archived within BoM databases. The first and longest duration time series was for Synoptic Surface Observation Reports (SYNOPs). The SYNOP dataset comprised 3-hourly observations from 1960 through the end of 2003. The second report type was hourly Meteorological Aerodrome Reports (METARs). These contained observations from 1985 through the end of 2003. The primary use of METAR observations was to minimize the number of missing observations in the SYNOP time series.

Based on an analysis of monthly fog frequency for YSCB (see Fig. 1), the fog season was defined as the period from 1 April through 31 August. The development of ANN models was constrained to this window of time only.

b. Input variables

From the 17 variables available in the SYNOP database, 8 were chosen to be input directly into the final choice of MLPs. The choice of inputs was determined by three main criteria: (i) statistical inferences—here the population characteristics of each potential input parameter associated with a fog and no-fog event were compared; if significant differences were found, one could suggest that this parameter would provide some discriminatory ability to the final ANN; (ii) domain knowledge, and (iii) data availability. However, the availability of potential MLP inputs was the determining factor. Because of problems handling missing observations outlined below in section 7, suitable variables that suffered from a large degree of incompleteness were not selected. All input data employed in this study were observed at the exact locality for which the forecasts were being made—YSCB.

Each of the eight parameters finally chosen was available at seven 3-hourly time steps, from 0900 through 0300 local standard time (LST) the following day. Classifications of fog events were made upon the basis of 0600 LST observations. Each lead time MLP was developed independently from the rest. First, the skill of MLPs employing each set of eight observations at specific time steps alone was determined. The addition of previous time steps, provided in unison with the latest available observations (e.g., providing MLPs with both 0300 and 2400 LST observations, determined by the lead from 0600 LST) was subsequently investigated. In all cases this was not found to significantly add skill. The next stage involved the inclusion of rate-of-change parameters, each for a different temporal resolution, in combination with the chosen time step inputs. This addition was found to be beneficial in terms of the final MLP forecast skill. Finally, indices derived from the eight observational variables were also tested in order to determine whether their inclusion produced a significant increase in MLP skill. Those examined included dewpoint depression, zonal and meridional components of wind (U.S. Environmental Protection Agency 2000), relative humidity, mixing height estimations (Carroll 2005), and stability (Turner 1964). However, all were found to be insignificant in terms of MLP forecasting skill.

The eight parameters sourced directly from the SYNOP database and input into all of the final MLPs included the following:

- Dry-bulb and dewpoint temperature (°C): The occurrence of fog can be identified by the point at which dry-bulb and dewpoint temperatures converge. Trends in the relative differences between these two parameters toward or away from zero are an indicator of the likelihood of fog formation and fog presence. These predictors were employed by themselves as well as being combined to calculate relative humidity (an input which was eventually excluded from the final choice of parameters).
- Wind speed ($m\ s^{-1}$): Wind speed is a strong determinant of fog. High wind speeds can act to dissipate mist before it forms into a thicker layer of fog (Oke 1987). Low wind speeds allow for turbulent mixing, which spreads cooling vertically, deepening the fog layer (Sturman and Tapper 1996).
- Mean sea level pressure (MSLP; hPa): MSLP was included because of its association with boundary layer inversions, radiative cooling at the surface, clear skies, and the presence of anticyclones. An increase in MSLP, or persistent high values of this parameter, would provide an indication of a synoptic situation conducive to fog formation.
- Wind direction (degrees true): Dependent on the lead time at which it is sampled, wind direction is either an indicator of synoptic situation (at a long lead) or local conditions (short lead). When em-

ployed in unison with wind speed, forecasters can derive an indication of fog likelihood based on the characteristics of both parameters. For example, at the longer lead time, afternoon southeasterlies are usually indicative of a synoptic situation not conducive of fog. However, at a shorter lead time (during nocturnal periods), mild southeasterlies are symptomatic of katabatic drainage flows. While these flows are often necessary for fog formation, they occur almost every night during the year, and as such their presence offers little predictive capability.

- Total cloud amount (oktas): A prerequisite for fog formation is a negative net radiation balance, as inferred by the occurrence of clear nocturnal skies.
- Surface visibility (km): Visibility measurements provide short-term *nowcasting* guidance as the ambient air temperature approaches its dewpoint and mist begins to form.
- Rainfall (mm in past 3 h): The timing of recent rainfall is important to fog formation. Rainfall that has occurred the previous afternoon can provide the requisite moisture for fog formation. On a shorter time scale, rain can inhibit the formation of fog by causing local turbulence within the boundary layer.

In addition to this final choice of eight input parameters, 0600 LST observations of dry-bulb temperature and MSLP pressure, as well as a persistence parameter, were also integrated into the final choice of inputs for the MLPs. This addition resulted in a statistically significant increase in skill for the majority of the lead-time MLPs tested. The inclusion of 0600 LST observations was aimed at partially negating the loss of skill through the extension of the lead time. As forecast accuracy invariably increases with a decrease in lead time, the employment of forecast values for a lead of 0 h offers scope for increased forecasting skill. This type of forecast can be sourced from NWP models such as MESOLAPS05. However, NWP approaches often suffer from inconsistencies in their forecasting ability of second-order parameters, that is, those associated with the inherent limitations of their grid spacing (e.g., rainfall, cloud cover). The 0600 LST MSLPs and dry-bulb temperatures are examples of the more accurately forecast variables at Canberra (BoM Staff 2005, personal communication). To determine the maximum skill obtainable from the MLPs, it was decided to use past observations (i.e., "perfect" forecasts). As forecast error has no doubt decreased chronologically, the use of archived forecast data was not found to be as useful as initially envisaged in maximizing MLP skill. However, given that forecasts, by their nature, are inherently imperfect, the sensitivity of the MLPs fog forecast skill to forecast errors in these 0600 LST observed input parameters was tested by randomly assigning errors of a controlled magnitude. This process allowed for the behavior of the MLPs when error is added to these parameters to be determined.

To maximize the efficiency of MLP training, all input data were normalized to fit the range $[-1, 1]$, which in this case was to fit the tan-sigmoid transfer function. Scaling all inputs to the same range prevented the MLP from placing unwarranted emphasis on those inputs with a larger mean magnitude, for example, MSLP (Mehrotra et al. 1997).

The target variable in this study was a binary indicator of either fog or no fog. Fog events were not defined according to World Meteorological Organization specifications (visibility less than 1 km). Instead, the final target class was a combination of fog events (hereby defined as occurrences of surface visibility of less than 2 km) and mist events (occurrences of visibility of less than 5 km). These definitions were used after consultation with BoM officials, in accordance with the aviation needs of Canberra airport.

c. Neural network

In accordance with the development of a feedforward, backpropagated MLP during batch-supervised learning, the data used in this study were partitioned into training, testing, and verification subsets.

Intraseasonal variability was removed from the training, testing, and validation sets in order to increase forecasting skill. This was achieved by randomizing the inputs and their associated target (output) values, a common practice during MLP training (e.g., Reusch and Alley 2002). Randomization of the input–output sequence was achieved through the pseudorandomization technique outlined in Demuth and Beale (2001).

Mehrotra et al. (1997) explain that, by randomizing input data, the correlation present between inputs and outputs presented consecutively is removed, which in turn can potentially increase MLP skill. In addition, randomizing allows for the MLP to be trained on a time series spanning a "longer" (in a chronological sense) period of time. It was hoped that this approach would remove any possibility of bias in results that could have been obtained had the networks been trained on older observations, and validated on more recent data. This approach was only considered reasonable as there was no significant change in any of the parameters detected over the time series employed, for example, such as that due to the anthropogenically forced enhanced greenhouse effect. Randomization ensures that the testing and validation sets conform to the same limits as the training set.

Within any neural network platform there exists a plethora of different MLP architectures, setups, training algorithms, and weight initialization functions. Gardner and Dorling (1998) provide a comprehensive literature review of the applications of ANNs within the atmospheric sciences. The majority of the studies cited employed a feedforward MLP. The architecture chosen in this study, for reasons expounded upon in section 7b, was a feedforward MLP with two hidden layers. The ANNs employed the Levenberg–Marquardt training (trainlm function in MATLAB) algorithm (Hagan and Menhaj 1994). Trainlm provided a learning rate that was rapid enough to maintain efficiency during the various training stages, while being more thorough than quicker algorithms (Demuth and Beale 2001).

The development of MLPs comprises both training and validation stages. The size (defined here by the number of nodes in each of the two hidden layers) of each MLP was incrementally increased, as larger networks are generally able to simulate more complex functions. Larger networks, however, are also more prone to overparameterization. As Reusch and Alley (2002) explain, the optimal size of the network is dependant on both the complexity of the forecasting problem and the architecture chosen. It is thus a significant factor in how well the MLP works. Too many nodes can lead to "overfitting" of the training set while too few will result in poor MLP performance.

Another important step during the training process is the need to train multiple iterations of the same configuration commencing from different initial conditions (Reusch and Alley 2002). This ensures that a particular MLP setup is not evaluated based on a single training run where the backpropagation algorithm became trapped within a local minimum in the data. Using random initialization of the MLP's biases and weights before training ensured that each training run commenced at a different position within the error surface. Five different initializations were found to satisfy both time constraints and the requirement to assess each configuration thoroughly.

Inherent to the method by which the backpropagation algorithm trains an MLP is the risk of overtraining where the MLP calculates a function to fit both the signal *and* the noise in a particular training set. In this sense the MLP has *memorized* the training set as opposed to discerning the underlying patterns (Mehrotra et al. 1997). Once an MLP has been overtrained, it loses the ability to generalize new data beyond the training set. Demuth and Beale (2001) suggest two methods to improve generalization when training MLPs: regularization and early stopping. Regularization involves an alternate training algorithm with a modified performance function employing a Bayesian framework to estimate the optimal parameters during training. In this study, early stopping was found to be the more effective and was, thus, employed.

Edwards and Murray (2000) explain that the method of early stopping assumes that MLPs learn first the signal and then the noise within the data. Early stopping involves training the MLP on a training set, while monitoring the model's error in an independent verification set (Reusch and Alley 2002). Training was stopped once the error started to decrease in the training set, but not in the validation set (Mehrotra et al. 1997). While the verification set was not used by the MLP during training to adjust weights, the error for this set was still monitored, and thus one cannot assume that the entire training process was completely independent from this verification data (Hewitson and Crane 1994). To reliably assess the skill of each MLP, a separate testing set that had not been employed at all during the training phase was used.

The performance of any MLP-based forecasting aid is heavily reliant upon the relative characteristics of the training and validation sets. Through altering the characteristics of these datasets so that there is a stronger (weaker) correlation between them, one can artificially increase (decrease) the calculated skill for the same forecasting aid. Cross validation was employed to provide a more robust measure of the MLP architecture's true skill. This process involved partitioning the data into 2/3 randomly selected training cases and using the remaining 1/3 for network validation. This process was repeated m times. The cross-validated value of MLP skill was then the average of the m values computed. In this study $m = 5$ was found to produce robust results within reasonable time constraints.

A major shortfall of the artificial neural network approach (indeed all statistical approaches) is the lack of a clear method for dealing with missing input values. The data record used in this study, as with most observational records of any substantial length, suffered from incomplete or missing observations. Of the 15 794 observation days available, 6855 (over 43%) had at least one missing value among the eight input parameters. Vamplew et al. (1996) provide a good summary of the various techniques available, including editing the backpropagation algorithm to be able to cope with missing values (e.g., Tresp et al. 1995) and omitting those days with missing values entirely (e.g., Reusch and Alley 2002). In this study the exclusion and estimation of missing values were employed to overcome this problem. The results of both approaches are discussed in section 7a.

d. Lead times

MLPs were trained to produce forecasts for 3-, 6-, 12-, and 18-h lead times from 0600 LST. These lead times correspond to the BoM TAF issues: the 6-h lead corresponds to the 1400/0800 TAF, the 12-h lead to the 0800/0200 TAF, and the 18-h lead to the 0200/2000 TAF (TAF notation for XXXX/YYYY: XXXX indicates the time in Greenwich mean time that the TAF is issued and YYYY the time for which it is valid until).

e. Skill assessment

As the output from the MLP is in the form of a single value scaled to fit between 0 and 1, a threshold value is required to decide when to declare a fog forecast. The choice of this threshold has a direct effect on the MLP skill: increase the threshold, and the probability of detection (hit rate) and false alarm rate both increase. Hence, simply applying the probability of detection and false alarm rate to an aid is restrictive, as these skill scores are dependant on the threshold criterion at which they were calculated. Signal detection theory, through relative (receiver) operating characteristic (ROC) curves, removes any dependence on this arbitrary threshold decision. This analysis, and subsequent linkages drawn with the cost/loss ratios of the end user, can be employed to determine their optimum threshold choice.

Construction of the ROC curves in this study was achieved by calculating separate contingency tables for a series of threshold values spanning the range of the output node's transfer function. Corresponding pairs of probability of detection H and false alarm rate F values are then calculated for each threshold value in the series. The ROC curve thus is one that transects each of these points. The use of ROC curves for the assessment of skill is based on a comparison of the ROC curve with both an ideal forecasting aid ($H = 1$ and $F = 0$), and a random forecasting aid ($H = F$). The diagonal line in a typical ROC plot represents this random forecasting (i.e., no skill—akin to flipping a coin) and the aim is to produce an aid whose ROC curve is as far away from this diagonal as possible. Calculation of the *area* underneath the ROC curve (in this study using the trapezoid rule, simply termed A) is used to quantify skill and compare ROC curves.

7. Results and discussion

a. Missing values

Various strategies for handling missing values were explored. The first was to exclude those selected inputs that suffered from the largest degree of incompleteness. The two most problematic inputs with regard to completeness were visibility and total cloud at 2400 LST. The only MLP that employed these inputs was the one for forecasts at a 3-h lead time. Removal of these two variables resulted in a reduction in the ROC curve mean area, significant at $\alpha = 0.10$, when a two-sample t test for the difference of the means was applied. This outcome was a result of preliminary analyses, obtained using the small samples sizes available due to logistical constraints.

Furthermore, if suitable methods of estimation could be found, then a larger percent of the available record could subsequently be employed to develop MLPs, potentially increasing the predictive capability.

Through a structured process of elimination, the best methods of estimating missing values within the dataset were established. Rainfall, total cloud, and surface visibility at 2400 LST were the most problematic variables. Climatological medians and means, extrapolation from prior observations, linear regression, MLPs, and a statistical analog technique similar to that employed by Regano (1997) were all tested as methods of missing value estimation. The use of climatological median values was the most suitable method for rainfall, with over 89% of the recorded observations being equal to their climatological median value, and given that fog occurrence is more sensitive to rain–no-rain data than particular rainfall amounts, this method was deemed suitable. Linear regression employing 2100 LST observations was the best method of estimating total cloud values at 2400 LST. Over 64% of the observations were found within one octa of the estimated value. This level of accuracy was considered sufficient to provide an indication of the clear nocturnal periods required for radiation fog formation. An analog statistical technique provided the best estimates of missing surface visibility values. In this case over 63% of the observations were found to lie within 5% of the estimated value. The level of completeness in the remaining variables was such that methods of estimation were not required.

Sensitivity analyses indicated that the skill of an MLP trained on data containing estimated values was clearly dependent on the quality of the estimated values. Vamplew et al. (1996) reached a similar conclusion. The consequences of their inclusion on MLP forecast skill were assessed through a two-pronged approach.

First, in order to investigate the effect of estimated input data on MLP training, two separate sets of MLPs were created. One set of MLPs was trained on only complete observations, while another was trained on a combination of observed and estimated values. As with every experiment conducted, five-way cross validation was employed here. Thus, five randomly partitioned datasets were composed with both complete, and complete and estimated data. The predictive skill of all MLPs was then calculated on the same validation sets comprising exclusively complete observations. Results indicate that MLPs trained on data containing estimated values did not perform at a markedly lower standard than those trained exclusively on true data. Training MLPs on both complete and estimated values was attempted with the expectation that the more cases in the training set, the better the MLP forecast skill. However, because no benefit was seen in the ROC curves, it was concluded that the best approach was to develop MLPs using only complete observational data; that is, observations with missing values were excluded from both the training and testing datasets.

Second, recognizing that any fog forecast tool being used in an operational context is likely to encounter occasional missing values of one or more input variables, the effect of estimated values in both the training and testing sets was also examined. Two MLPs were trained: one on complete observations and the other on a combination of complete and estimated values. The two MLPs were then tested on a combination of complete and estimated values. Based upon resultant ROC curves, the skill of both MLPs was significantly lower than when tested on complete data alone. However, if the use of estimated input data was deemed necessary by the end user, the performance of the MLP is still satisfactory—thus indicating the technique's robustness to error.

b. Hidden layers

Most previous studies agree that MLPs with two hidden layers are theoretically capable of approximating any arbitrary nonlinear function (Hewitson and Crane 1994; Spining et al. 1994). In an effort to minimize training times and avoid the risk of overparameterization, MLPs were initially trained with a single hidden layer. Then a second hidden layer was added to the MLP architecture, resulting in a statistically significant increase in predictive skill (at $\alpha = 0.10$) when the results of the two networks were compared for various lead times. It was also found that any more than two hidden layers had a negligible effect on MLP skill, but dramatically increased computational times for training.

While an increase in hidden-layer nodes initially resulted in an increase in MLP performance, this increase was only observed up to a point, after which the ROC areas, and therefore the predictive skill, leveled out. At this point no more hidden-layer nodes were added. The final number of hidden-layer nodes varied for the MLP at each lead time, and ranged from 3 to 20 nodes in each hidden layer.

TABLE 1. Cross-validated mean values, A_{mean} (and their standard deviations, A_σ), of area underneath ROC curves.

Lead time from 0600 LST (h)	Forecast (TAF) issue (LST)	Area underneath ROC (A)	
		A_{mean}	A_σ
3	—	0.937	0.007
6	1400/0800	0.858	0.012
12	0800/0200	0.849	0.010
18	0200/2000	0.839	0.014

c. Forecasting skill

Table 1 lists the cross-validated mean values of the areas under the ROC curves for the final MLPs developed to produce fog forecasts at the specified lead times. The larger the area, the better the predictive skill. The individual ROC curves for the best-performing MLPs for each lead time are also supplied (Fig. 2). An area of 1 indicates a perfect forecast while less than 0.5 indicates skill no better than that achieved via random forecasts. Overall results in Table 1 indicate that the MLP approach was useful for fog prediction at YSCB. Also of note is the sharp decrease in skill from the 3- to the 6-h lead time, after which a relatively minor degradation in skill occurs as lead time extends from 6 to 18 h.

Marzban (2004) provides an explanation for the lack of variance between the results for the 3- to the 18-h-lead networks. Employing the arbitrary classification of forecasting models as either "good" or "bad," Marzban (2004) explains that ROC analysis provides for an efficient method by which to separate good models from their poorer equivalents. It does not, however, provide such an effective method for the separation of good models alone.

d. Error perturbations in 0600 LST inputs

As indicated previously, the inclusion of two 0600 LST observational inputs was found to have a beneficial effect on MLP fog forecast skill. Obviously these 0600 LST data would be forecast if the MLP is ever used in an operational setting. It is therefore instructive to test the MLPs' ability to cope with errors in these inputs.

To assess the robustness of the final models to errors in these 0600 LST input parameters, new validation sets were created in which the 0600 LST observations were randomly perturbed with errors of fixed magnitude. Here, the random element of the perturbation was the

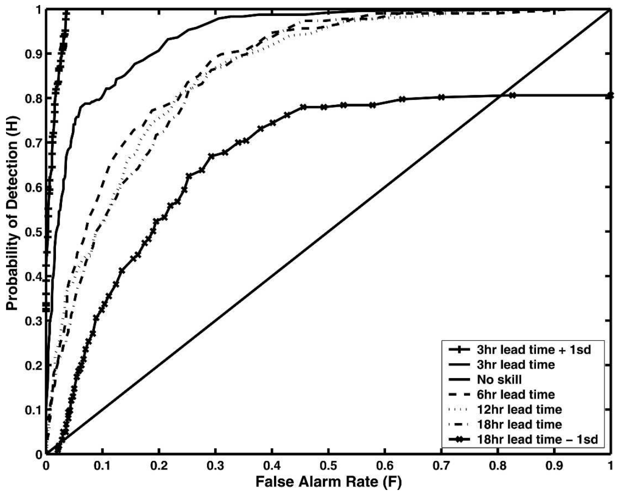

FIG. 2. ROC curves for the optimized Canberra fog forecasting neural nets. Upper (lower) error bounds are provided by adding (subtracting) one standard deviation from the ROC curve of the best (worst) performing network.

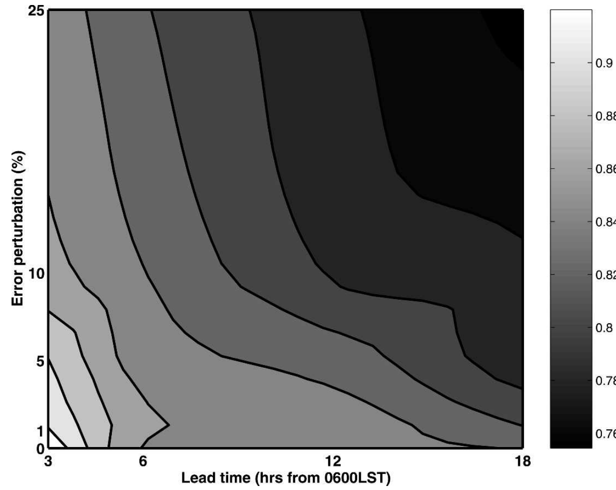

FIG. 3. Impact of errors in the 0600 LST observations upon MLP forecast skill. Contours represent cross-validated mean areas under the ROC curves.

sign in front of the error. The error levels tested were 1%, 5%, 10%, and 25% of the observed values. Areas under the ROC curves show the degradation in skill when the MLPs were exposed to these input errors.

Figure 3 illustrates that, while there is some degradation in skill, all four models appear to be reasonably robust to errors of the magnitude one could expect in the 0600 LST forecast variables when being employed as inputs. Even the 18-h lead-time model at a 25% error perturbation still gives "acceptable" discrimination ability. Also of note is that the most "robust" model to these error perturbations was not necessarily the best performer.

8. Conclusions and further work

a. Conclusions

The development of MLP-based fog forecasting aids for YSCB has been outlined. Results indicate that all four models exhibit good forecasting ability. Error perturbation and ROC curve analyses and domain knowledge of the meteorological processes behind fog led to the selection of effective predictors. The ROC curves indicate that the MLPs are robust to error perturbations in the forecasted inputs. It would appear likely that this approach is well suited to the Canberra area, and the integration of these neural network tools into the BoM's current forecasting procedures for YSCB should provide a tangible benefit.

b. Further work

The scope for further development of this study is significant. Worthwhile issues for further work include the development of MLPs to forecast fog onset and clearance times. More work could fruitfully be applied to the problem of missing values among the neural network's inputs. Extending the error perturbations to include all predictors in the final model would produce an error surface of the overall bounds of the discriminatory ability of the final models at each lead time. Application of fuzzy logic offers one avenue to compare the relative significance of the selected parameters to fog formation. Through the economic evaluation of the developed aids, optimum thresholds on the MLPs' output functions at which to forecast fog can also be determined for specific end users. The integration of current aids into other forecast systems, for example, within an objective forecast consensus scheme, with MLP output as one input also offers significant promise. Further, given sufficient data, the current approach could be applied to other nonlinear meteorological forecast problems, for example, snow forecasting and alpine road ice forecasting. Finally, while preliminary studies were conducted on the skill other neural network architectures could offer, further time devoted to this matter could yield superior results.

The next stage in this research effort is an application of the neural network approach to Australia's busiest international airport: Kingsford Smith in Sydney.

Acknowledgments. The authors thank members of the BoM staff for their valuable input: Sue Lawrence, Chris Matthews, and Ian Robertson of Sydney Airport Meteorological Unit; Clem Davis (now retired) and Jessica Trevena of the Canberra Meteorological Office; David Williams of the New South Wales Regional Office; and Rodney Potts and Gary Weymouth of the Bureau of Meteorology Research Centre, Melbourne.

REFERENCES

Bonné, J., cited 2005: Airlines still struggle with paths to profit. [Available online at http://www.msnbc.msn.com/id/3679292/.]

Demuth, H., and M. Beale, 2001: *Neural Network Toolbox: For Use with MATLAB*. 4th ed. The MathWorks, 844 pp.

Edwards, P. J., and A. F. Murray, 2000: A study of early stopping and model selection applied to the papermaking industry. *Int. J. Neural Syst.,* **10,** 9–18.

Gardner, M. W., and S. R. Dorling, 1998: Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmos. Environ.,* **32,** 2627–2636.

Hagan, M. T., and M. B. Menhaj, 1994: Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Networks,* **5,** 989–993.

Hall, T., H. E. Brooks, and C. A. Doswell III, 1999: Precipitation forecasting using a neural network. *Wea. Forecasting,* **14,** 338–345.

Hewitson, B. C., and R. G. Crane, 1994: *Neural Nets: Applications in Geography*. GeoJournal Library, Vol. 29, Kluwer Academic, 194 pp.

Hsieh, W. W., and B. Tang, 1998: Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bull. Amer. Meteor. Soc.,* **79,** 1855–1871.

Huang, X., G. Mills, G. Weymouth, R. Potts, and T. Keenan, 2002: Using high-resolution LAPS model to predict fog. *Extended Abstracts, 14th Modelling Workshop on Modelling and Predicting Extreme Weather Events,* BMRC Research Rep. 90, Melbourne, Australia, Bureau of Meteorology, 25–28.

Leigh, R. J., 1995: Economic benefits of Terminal Aerodrome Forecasts (TAFs) for Sydney Airport, Australia. *Meteor. Appl.,* **2,** 239–247.

Leipper, D. F., 1995: Fog forecasting objectively in the Californian coastal area using LIBS. *Wea. Forecasting,* **10,** 741–762.

Lippmann, R. P., 1987: An introduction to computing with neural nets. *IEEE Acoustics, Speech Signal Process. Mag.,* **4,** 4–22.

Malley, S., Y. Miao, C. Davis, and A. Forrest, 2003: A study of fog statistics and forecasting aids at Canberra Airport. Meteorological Note 219, Australian Bureau of Meteorology, 15 pp.

Marzban, C., 2004: The ROC curve and the area under it as a performance measure. *Wea. Forecasting,* **19,** 1106–1114.

——, and G. J. Stumpf, 1996: A neural network for tornado prediction based on Doppler radar–derived attributes. *J. Appl. Meteor.,* **35,** 617–626.

Matthews, C., S. Lawrence, R. Jardine, and C. Webb, 2002: Fog forecasting at Sydney Airport. Australian Bureau of Meteorology, 6 pp.

Mehrotra, K., C. K. Mohan, and S. Ranka, 1997: *Elements of Artificial Neural Networks*. The MIT Press, 334 pp.

Murphy, A. H., 1994: Assessing the economic value of weather forecasts: An overview of methods, results and issues. *Meteor. Appl.,* **1,** 69–73.

Oke, T. R., 1987: *Boundary Layer Climates*. 2d ed. Muthen and Co., 435 pp.

Parikh, C. R., M. J. Pont, and N. B. Jones, 1999: Improving the performance of multi-layer perceptrons where limited training data are available for some classes. *Proc. Ninth Int. Conf. on Artificial Neural Networks,* Vol. 1, Edinburgh, United Kingdom, Institution of Electrical Engineers, 227–232.

Pasini, A., V. Pelino, and S. Potestà, 2001: A neural network model for visibility nowcasting from surface observations: Results and sensitivity to physical input variables. *J. Geophys. Res.,* **106,** 14 951–14 959.

Puri, K., G. S. Dietachmayer, G. A. Mills, N. E. Davidson, R. A. Bowen, and L. W. Logan, 1998: The new BMRC Limited Area Prediction System, LAPS. *Aust. Meteor. Mag.,* **47,** 203–223.

Regano, L., 1997: A fully automated forecasting aid for fog in Australia. Meteorological Note 212, Australian Bureau of Meteorology, 8 pp.

Reusch, D. B., and R. B. Alley, 2002: Automatic weather stations and artificial neural networks: Improving the instrumental record in West Antarctica. *Mon. Wea. Rev.,* **130,** 3037–3053.

Rummelhart, D. E., and J. L. McCelland, 1986: *Parallel Distributed Processing*. Vol. 1, The MIT Press, 517 pp.

Spining, M. T., J. A. Darsey, B. G. Sumpter, and D. W. Noid, 1994: Opening up the black box of artificial neural networks. *J. Chem. Educ.,* **71,** 406–411.

Sturman, A., and N. Tapper, 1996: *The Weather and Climate of Australia and New Zealand*. Oxford University Press, 476 pp.

Tasadduq, I., S. Rehman, and K. Bubshait, 2002: Application of neural networks for the prediction of hourly mean surface temperature in Saudi Arabia. *Renewable Energy,* **25,** 545–554.

Tresp, V., R. Neuneier, and S. Ahmad, 1995: Efficient methods for dealing with missing data in supervised learning. *Advances in Neural Information Processing Systems 7,* G. Tesauro, D. S. Touretzky, and T. K. Leen, Eds., The MIT Press, 689–696.

Turner, D. B., 1964: A diffusion model for an urban area. *J. Appl. Meteor.,* **3,** 83–91.

U.S. Environmental Protection Agency, 2000: Meteorological monitoring guidance for regulatory modelling applications. EPA-454/R-99-005, U.S. EPA, 171 pp.

Vamplew, P. W., D. Clark, A. Adams, and J. P. Muench, 1996: Techniques for dealing with missing values in feedforward networks. *Proc. Seventh Australian Conf. on Neural Networks,* Canberra, ACT, Australia, University of Tasmania, 250–254.