Fog Forecasting for Melbourne Airport Using a Bayesian Decision Network

TAL BONEH
Faculty of Information Technology, Monash University, Clayton, and Bureau of Meteorology, Melbourne, Victoria, Australia

GARY T. WEYMOUTH, PETER NEWHAM, RODNEY POTTS, AND JOHN BALLY
Bureau of Meteorology, Melbourne, Victoria, Australia

ANN E. NICHOLSON AND KEVIN B. KORB
Faculty of Information Technology, Monash University, Clayton, Victoria, Australia

(Manuscript received 5 January 2015, in final form 29 June 2015)

ABSTRACT

Fog events occur at Melbourne Airport, Melbourne, Victoria, Australia, approximately 12 times each year. Unforecast events are costly to the aviation industry, cause disruption, and are a safety risk. Thus, there is a need to improve operational fog forecasting. However, fog events are difficult to forecast because of the complexity of the physical processes and the impact of local geography and weather elements. Bayesian networks (BNs) are a probabilistic reasoning tool widely used for prediction, diagnosis, and risk assessment in a range of application domains. Several BNs for probabilistic weather prediction have been previously reported, but to date none have included an explicit forecast decision component and none have been used for operational weather forecasting. A Bayesian decision network [Bayesian Objective Fog Forecast Information Network (BOFFIN)] has been developed for fog forecasting at Melbourne Airport based on 34 years’ worth of data (1972–2005). Parameters were calibrated to ensure that the network had equivalent or better performance to prior operational forecast methods, which led to its adoption as an operational decision support tool. The current study was undertaken to evaluate the operational use of the network by forecasters over an 8-yr period (2006–13). This evaluation shows significantly improved forecasting accuracy by the forecasters using the network, as compared with previous years. BOFFIN-Melbourne has been accepted by forecasters because of its skill, visualization, and explanation facilities, and because it offers forecasters control over inputs where a predictor is considered unreliable.

1. Introduction

Fog is a dense mass of small water droplets suspended in the air near the ground. Fog can vary in vertical depth from a few to hundreds of meters. Its density reduces visibility from a few kilometers to near zero. The definition of fog according to the World Meteorological Organization (WMO) is horizontal visibility of less than 1000 m on the ground (WMO 1966). For the practical purposes of an airport, fog is defined as very small water droplets that reduce visibility to a level so low as to impede or prevent aircraft operations (ICAO 2007).

Fog forecasts are critical for airlines and have significant economic consequences. A single unforecast fog event at a major Australian airport can lead to a serious safety hazard, cause passenger inconvenience, and cost millions of dollars (Leigh 1995). About 20% of fog events at Melbourne Airport, Melbourne, Victoria, Australia, were not forecast in the 1999–2005 period leading up to the current study (Newham et al. 2009). If fog is forecast, all aircraft must carry enough fuel to reach an alternative airport or enable the aircraft to maintain a holding pattern (Australian Government Civil Aviation Safety Authority 2010). The cost of an unforecast fog is generally much higher than that of a false alarm, but the exact figures have not been disclosed, and...
would vary with aircraft, route, fuel policies, and other factors (Keith and Leyton 2007; Keith 2003). A simple solution might be to always prepare for fog conditions. However, these preparations are expensive and therefore should be minimized.

One of the weather reports used by the aviation industry for planning airport operations is the terminal area forecast (TAF), which applies to the area within 5 nautical miles (n mi; 1 n mi = 1.852 km) of an airport. By international regulations, a PROB FOG (i.e., fog) forecast statement must be included in the TAF if the risk of its occurrence is considered to be at or above 30%. The TAFs for all major airports in Australia are provided by the Australian Bureau of Meteorology (BoM). TAFs are prepared every 6 h and are valid for 30 h (e.g., the 0612 TAF is issued at 0600 UTC), and can be amended at any time. An additional local product that supplements the TAF is Code Gray (CG), which reports a low (less than 30%) yet significant (more than 5%) chance of fog (Miao et al. 2012). This product can be issued at around 1500 local time (LT) and can then be updated until late evening and is available to all airlines.

Making decisions about weather forecasts involves uncertainties, which result from incomplete knowledge, missing or poor quality data, uncertainty in observation, or uncertainty in supportive guidance tools. Numerical weather prediction (NWP) underpins much weather prediction; however, the spatial resolution of the NWP models limits the scale of the atmospheric processes that these models can represent. Moreover, the formation of fog is a complex local phenomenon that depends on a delicate balance of local factors, making the problem of fog forecasting difficult. In some countries such as the United Kingdom, a high spatial density of visibility observations can make high-resolution NWP more capable for short-term fog forecasts (e.g., Clark et al. 2008). In Australia, with fog being a relatively rare event, such observations are mostly uneconomic and hence too sparse to provide significant benefit to NWP.

To fill the gap between the need to predict fog reliably and the support provided by NWP for this task, a range of individual weather forecasting guidance tools that complement NWP have been developed. Tools used in Australia are based on statistics, including neural networks that provide only yes/no guidance for fog forecasts rather than a fog probability (Fabbian et al. 2007), fuzzy logic (Miao et al. 2012), and logistic regression (Stern and Parkyn 2001). These tools do not fully resolve the difficulty in forecasting fog.

The task of forecasters is to integrate the different pieces of information offered by observations, NWP, and other guidance tools and to make a final decision on what to forecast. Thus, the prediction of fog relies heavily on the forecaster’s judgment, skill, and ability to deal with the uncertainty involved, potentially leading to inconsistencies between different forecasters and between different TAF issue times. Moreover, the forecaster’s subjective considerations cannot easily be tracked for future validation and therefore are difficult to systematically improve. On the other hand, increasing financial, staffing, legal, and regulatory pressures increase the importance of accurate and structured forecast processes to support consistent, defensible output and to ensure the incorporation of forecasters’ local knowledge in fog forecasts. It follows that supporting the forecasters in the process of decision-making using state-of-the-art information technology is crucial. One such technology is Bayesian decision networks (BDNs).

Bayesian networks (BNs; Pearl 1988) have a solid mathematical basis in probability theory and are now widely used as an intuitively appealing, practical representation of knowledge for reasoning under uncertainty (section 3b). When applied to weather forecasting, BNs probabilistically combine information about predictors to produce guidance, where predictors may be other guidance, physical weather variables, and subjective forecaster input. Several BNs for probabilistic weather prediction have been reported, all of which, however, lack an explicit forecast decision component. An early example, the Hailfinder system (Abramson et al. 1996) was developed and constructed using domain expert knowledge to predict severe weather phenomena, but was never tested. The Sydney, New South Wales, Australia, harbor sea-breeze BN (Kennett et al. 2001) was tested but never used operationally. Cano et al. (2004) looked at data-mining methods to automatically construct BNs for meteorological applications, but again, operational deployment was not reported.

To utilize BNs for decision-making, BNs can be extended to create BDNs using utility theory, by adding an explicit representation of possible decisions and the utilities of possible outcomes (Howard and Matheson 1981). BDNs combine probabilities with costs and benefits, and hence they support forecaster decision-making, such as whether to issue a fog forecast in a TAF.

We have developed a BDN for fog forecasting at Melbourne Airport. It combines probabilities with relative costs of false alarm and unforecast fog events. Its input variables are known fog predictors and the Stern–Parkyn guidance tool (Stern and Parkyn 2001). The development of the network was based on 34 years’ worth of data (from 1972 to 2005) and is described in detail in Boneh (2010). The model was adopted in 2006 as an operational decision support tool for forecasters called the Bayesian Objective Fog Forecast Information Network (BOFFIN; hereafter referred to as BOFFIN-Melbourne).
In this paper we present the results of its operational use by forecasters over an 8-yr period (2006–13), which clearly demonstrate their forecasting performance improved from that prior to the BDN. To our knowledge, this is the first report of long-term operational use of a BN (decision net or otherwise) for weather forecasting.

2. Background: Fog at Melbourne Airport

Melbourne Airport is the major airport in Victoria and the second busiest airport in Australia (30 million passengers and 218 000 aircraft movements in 2013; Melbourne Airport 2014). During the period 1972–2005, there was an average of 12.4 fogs per year.

Fog can occur in a light wind situation or with stronger winds during or following rainfall, given other favorable atmospheric conditions. The most common light wind situation for fog formation is the Melbourne eddy, which results from a generally easterly synoptic airflow modified by topography (Fig. 1). This transports moist air from Port Phillip Bay (to the south) and circulates it over the airport in light, locally westerly winds (Spillane 1978; Abbs 1986; Newham 2004). In a cloud lowering to the ground situation, drizzle from low cloud evaporates as it falls, moistening and cooling the air below to saturation, leading the cloud to develop downward. Wind speed (for which the pressure gradient is a proxy) is often relatively high in this case, making low-cloud formation more likely than fog. Occasionally, fog may form at Melbourne Airport when a low-level southerly airflow is blocked by the topography to the north, resulting in stagnant air that mixes and/or cools to produce fog. Less commonly, sea fog may advect northward to the airport (Newham et al. 2007). Keith (1991) investigated the meteorology of fog and low cloud at Melbourne Airport, and provided some nomograms for their prediction. He also noted the importance of katabatic flows in retarding fog formation and persistence at Melbourne Airport, suggesting observations that could be used to improve fog forecasting.
A number of guidance tools exist to support forecasters in Melbourne with the task of fog prediction, varying in both complexity and underlying method. The Regano case–based reasoning tool (Regano 1997) uses synoptic observations of a number of weather elements (e.g., dewpoint, temperature, and wind) to predict fog likelihood. The tool searches a database for analogous cases with values that approximately match the current observations for these predictors. The number of matching cases with fog is divided by the total number of matching cases, giving an estimate of the probability of fog under the given conditions. A verification of the Regano tool at Melbourne Airport revealed that many fog events were associated with either a low number of analogs or no match at all (Newham 2004). These and other problems have prevented the use of this tool at Melbourne Airport.

A nowcasting technique by Weymouth (2006), extending that of Eyre et al. (1984), is used for detecting fog and low cloud from polar-orbiting and geostationary satellite imagery at night. This approach uses the difference in emissivity of water droplets between (nominal) 3.8- and 11-μm wavelengths to detect water cloud. Additions include an approximate distinction between very low cloud or fog and higher water–based cloud, from infrared temperature contrast (Weymouth 2006). The technique provides limited lead time; however, it can provide additional information for forecasters such as the extent of fog and how far it is from the airport. It has been utilized to identify fog predictors in northern Australia (Zeschke 2010).

The Stern–Parkyn synoptic typing technique (Stern and Parkyn 2001) uses logistic regression at 1500 LT to estimate the probability of fog at Melbourne Airport. It uses pressure gradients, which are categorized into 50 bins according to flow direction, strength, and cyclonicity, together with other predictors (e.g., temperature and dewpoint at the airport). The evaluation of this technique showed that the frequency of fog increased with increasing fog probability, indicating some predictive ability. However, the majority of events still occurred with a forecast probability of less than 15%. If a forecaster were to define this as the cutoff criterion to forecast or not forecast fog, using the Stern–Parkyn technique about 58% of fog events would be unforecast. It is less accurate during summer, for synoptic types for which fog is rare, and for fog events associated with rainfall (Newham 2004).

The Melbourne Airport 0606 TAF Fog Forecasting Excel Sheet (0606TAF) is a framework tool that displays the results of observations and other guidance tools and helps guide the forecaster’s thinking. It has four components: observations; objective guidance, which includes guidance tools and calculated information; subjective guidance, which includes forecasters’ estimation of different aspects of the atmosphere; and a subjective decision component, which includes the forecaster’s decision regarding fog in light of the displayed information. In 2009, 0606TAF was replaced by the Fog–Forecast Decision Support System (Fog–FDSS), a web-based application that serves as the interface to the BDN. Fog–FDSS is currently accepted and used by all forecasters. Prior to 2006, the forecasters had to subjectively integrate this information and make the final decision on what to predict.

3. Constructing a Bayesian decision network for Melbourne Airport

To integrate existing tools and information objectively and consistently, and thus complement existing approaches, the BoM developed the BOFFIN decision support tool for fog forecasting at the major airports in Australia. BDN technology was used, as a state-of-the-art mathematical technology that offers a powerful computational model that can be used for reasoning with uncertain data and offers many advantages for weather forecasting: it provides an explicit representation of uncertainty, a graphical representation of domain relationships, information from multiple scales and formats, and a normative integration of objective and subjective information. The latter is important for weather forecasting where domain knowledge and understanding is often incomplete and expert opinion may play a significant role.

a. Modeling the fog forecasting process for Melbourne Airport

As a first step in building our BDN, we modeled the fog forecasting process and identified the main predictors for fog occurrence, namely moisture (Moisture), gradient (Gradient), lapse rate (LapseRate), month (Month), length of night (LengthOfNight), rain (Rain), and Stern–Parkyn. These variables other than Month are inherently continuous. However, when forecasters are forming their conceptual models for reasoning, they are often interested in significant and discrete ranges of data, such as those that are identified with very favorable (vfav), favorable (fav), and unfavorable (unfav) conditions for fog to form. Accordingly, variables were divided into ranges or states that are statistically related to different fog probabilities using data from different types of meteorological information archived in the BoM databases. A dataset of 12054 cases, one per day, was collected from 1972 to 2005, where each case includes 1) the values for the input variables, which are
derived from weather stations (sometimes missing), and 2) whether fog occurred between 1500 and 1500 LT the next day, derived from direct airport observations. In this 34-yr dataset, fog occurred in 410 cases (3.4% of days).

Table 1 depicts the fog forecasting variables (predictors) and their states after discretization. Further details of the discretization are in Boneh (2010).

1) Moisture availability upstream is a good indicator of whether or not sufficient low-level moisture will be available for the formation of fog because it is often advected toward the airport overnight in conditions conducive to fog, for example, from the Laverton area (Fig. 1) via the Melbourne eddy (Abbs 1986). Moisture includes pairs of seasonally varying absolute humidity $T_d$ (°C) and temperature $T$ (°C), with $T - T_d$ being related to relative humidity. These pairs for Laverton were discretized as in the example for April–May in Fig. 2. Two series of pairs were defined: for days that ended with or without fog. The two series were placed on a grid with axes $x = T_d$ (°C) and $y = T$ (°C). The probability of fog was then calculated in a grid box around each point. Curves were drawn on the grid, based on the forecaster’s opinion, to separate areas with different fog probabilities, while making sure that the selected areas include sufficient data for robustness. With regard to moisture, curves included
around 80% of fog cases in the vfav, 15% in the fav, and 5% in the unfav areas. For the warmer months, November–March, the 2100 LT observations are more representative and for the other months the 1800 LT observations are more useful.

2) Pressure gradient, which is measured using stations roughly 100 km north, south, east, and west of Melbourne Airport, is a good indicator of the synoptic situation and correlates well with the Melbourne eddy. Thus, it serves as a proxy for wind direction, fetch, and speed and, in turn, indicates how favorable conditions are for fog formation. The pressure gradient is correlated differently for fog in rain and no-rain conditions. In rain conditions, stronger pressure gradients can allow fog, as evaporative cooling of rain on the ground leads to enhanced near-ground stability and hence more calm (and moist) surface conditions conducive to fog. North–south and east–west pressure gradient pairs were plotted on a grid for fog and no-fog (nofog) cases, and the grid was discretized, using an approach similar to that for moisture.

3) The presence or development of a low-level inversion typically increases the likelihood of fog formation. A low-level inversion helps to trap moist air at low levels and prevents a katabatic flow from the ranges to the north. The katabatic flows tend to have lower relative humidity and frequently prevent or clear fog. We have found that the presence of an inversion, and hence the likelihood of fog, is indicated by a proxy for lapse rate—the decrease in temperature with height, that is, the temperature at an elevated site near the source of katabatic flows (Kilmore Gap, elevation 527.8 m) minus the temperature at Melbourne Airport (elevation 113.4 m; Fig. 1). We used the lapse rate proxy at 2100 LT in the absence of a more reliable indication of inversion strength and height available at the time.

4) The value of predictors conducive to fog can vary at different times of year, as can the frequencies of fog. Longer nights, which increase cooling, and dew deposition make fog more likely overall to occur in the cooler months as do greater soil moisture and other seasonal factors. Hence, we used Month (Table 1; 4a) and LengthOfNight (Table 1; 4b) as predictors of fog.

5) Traditionally, for predicting fog forecasters consider two situations: rain and no rain. Rainfall influences the other predictors, increasing near-surface atmospheric

---

![Figure 2](https://example.com/figure2.png)

**Fig. 2.** Observed moisture conditions discretized as unfavorable, favorable, and very favorable for fog for April and May, at 1800 LT, before fog and no-fog days. Observations plotted toward the lower right tend to have higher relative humidity, while observations to the right have higher dewpoints (high absolute humidity). Some dewpoint observations early in the dataset were only recorded to the nearest degree.
moisture, cooling, and stability. Thus, fog can occur postrain in conditions otherwise unlikely to produce it (e.g., unfavorable conditions of the pressure gradient, moisture, and lapse rate). In practice, rainfall above or below 4.5 mm gave useful information on fog being more or less likely.

6) The Stern–Parkyn technique (Stern and Parkyn 2001) analyzes the synoptic situation for the next morning (as described earlier) and its probability output provides a good indication for fogs that occur in an easterly gradient flow.

BOFFIN produces two outputs: the probability of fog occurring and the categorical forecast [say no fog (saynofog), CG, and PROB FOG (fog needs to be put on TAF; fogOnTAF)] to issue.

b. Bayesian decision networks: Background

A Bayesian network (Pearl 1988) is a directed, acyclic graph whose nodes represent the random variables (discrete and/or continuous) in the problem. Directed arcs connect pairs of nodes, representing the direct dependencies (which are often causal connections) between variables. The nodes pointing to a node $X$ are called its parents, and are denoted $\text{Parents}(X)$. The relationship between variables is quantified by conditional probability distributions (CPDs) associated with each node, denoted $[X \mid \text{Parents}(X)]$, where the state of the child nodes depends on the combination of the values of the parent nodes. In nodes without parents, the CPDs hold the marginal (prior) probability for the node. The full joint probability distribution (JPD) is given by the product of the CPDs; more formally, for a BN with nodes $X_1, \ldots, X_n$, then

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \text{Parents}(X_i)).$$

Thus, a BN is a mathematical model providing a factorized and compact representation of the full JPD. In this paper, we only build BNs with discrete variables, discretizing continuous variables where necessary and hence the CPDs become conditional probability tables (CPTs).

Given a network consisting of nodes, arcs, and CPTs, users can add observations by setting the values of any combination of nodes in the network. This evidence $e$ propagates through the network, producing a new posterior probability distribution $P(X \mid e)$ for each variable, a process termed probabilistic inference. A number of efficient algorithms, based on Bayes’s theorem, perform probabilistic updating, providing a combination of predictive, diagnostic, and explanatory reasoning. Of note, the direction of arcs identifies the nature of dependencies between nodes; evidence however propagates in both directions along the arcs.

BNs can be augmented with decision nodes, representing the possible decisions that can be made, and utility nodes, representing the value (i.e., costs and benefits) of combinations of decisions and outcomes nodes. The resultant BDN is a graphical representation and implementation of decision theory (Savage 1954), combining probability theory and utility theory. For a fuller treatment of Bayesian decision networks and their application, see Korb and Nicholson (2010).

c. The BOFFIN-Melbourne BDN

BOFFIN-Melbourne was constructed according to our previously published methodology for BDN construction (Boneh 2010; Korb and Nicholson 2010, p. 463), using the Netica BN software (Norsys Software Corporation 1995). Figure 3 (top) shows the BOFFIN-Melbourne structure with rectangular boxes depicting the input predictor nodes (Moisture, Gradient, LapseRate, Month, LengthOfNight, Rain, and Stern–Parkyn), the output prediction node representing whether or not a fog will occur overnight (Fog), and the decision node representing the possible forecasts (Forecast). The utility node is depicted by a hexagon. The possible states for each node are those determined by the process described in section 3a above.

An inherent challenge in building a BDN is the need to represent a problem domain with the usual trade-off between the complexity of the model (and associated computational requirements) and its faithfulness (and hence usefulness). In particular, the sizes of the CPTs are exponential in the number of parents as the child node models the dependencies between its parents, with consequent effects on the amount of data required for training. The process of building such a model is called knowledge engineering and involves graphical modeling choices that must capture the most significant domain relations, assumptions, and their quantification while controlling complexity. For example, one may manage CPT sizes by limiting the number of node states, modeling only significant dependencies between nodes, or simply reversing the directions of the arcs. In practice, knowledge engineering often involves both expert elicitation and automated learning from data. In weather modeling, all variables are usually related, and capturing all these dependencies can create an over-complicated structure. This problem is magnified when modeling a rare event such as fog, since data and knowledge are scarce, requiring the final model to be simplified.

The BOFFIN-Melbourne structure reflects the required trade-off between the desired expressiveness of the model and the need to achieve manageable CPTs.
FIG. 3. Final BOFFIN-Melbourne model with the a priori distribution (before entering any observations for a specific case) and the CPT for the Fog node. (top) BDN for the Fog node at Melbourne airport. Node descriptions are provided in Table 1. The boxes represent the nodes with the variable name (across top), their possible states (left side), and the distribution over the states as both numbers (center) and visualized as horizontal bars (right side). The output decision is depicted as a box labeled Forecast with the possible decisions and their calculated expected utility; the horizontal bar indicates the optimal decision. The hexagon node labeled Utility Function defines a utility function, $U(Fog, Forecast)$, for the possible combinations of its parents, that is, (fog, saynofog), (fog, CG), (fog, fogOnTAF), (nofog, saynofog), (nofog, CG), and (nofog, fogOnTAF). Arcs indicate direct influence between nodes. (bottom) The CPT of the Fog node, which has two parents: Rain and LengthOfNight.
To this end, only the most important relationships between nodes were explicitly modeled by arcs. A simple causal structure, with all predictor nodes as parents of the Fog node, was infeasible because of the large size of the Fog CPT and the limited training data. To overcome this problem, the number of parents of Fog was reduced to only LengthOfNight (modeled from Month by reducing 12 states to only 5) and Rain, capturing that fog formation depends significantly upon rain and length of night.

The Fog CPT is \( P(\text{Fog} | \text{Rain}, \text{LengthOfNight}) \) (Fig. 3, bottom). The fog predictors Moisture, Gradient, LapseRate, and Stern–Parkyn were treated as independent indicators of fog and modeled as children of Fog, with some potential small cost to accuracy. In addition, only conceptually important and CPT-manageable dependencies between predictors were represented by arcs (e.g., from Rain to Gradient). Together, these simplifications reduced the number of parameters from about 6000 to 98.

Following the completion of the network structure, the CPTs were learned from the data using the counting-learning algorithm (Spiegelhalter and Lauritzen 1990), as implemented in Netica.

Figure 3 depicts the final model with the a priori distribution before entering any new observations for a specific case. The distribution for the Fog node reflects the low occurrence of fog at Melbourne airport [i.e., \( P(\text{Fog} = \text{fog}) = 3.3\% \)]. Note also that the distribution shown for the only node without parents (Month) is exactly its CPT and the marginal probability of a given day being in each month.

Figure 4a depicts the fog prediction model with a limited number of observations Month = May, Moisture = fav, and Stern–Parkyn = 30–100 (each clearly set to 100%). The distribution over the other nodes (Fog, Rain, Gradient, and LapseRate) changed according to these observations. The probability of fog increased from 3.3% to 31%, indicating fog is now much more likely. The probabilities of Gradient = vfav, Rain = <4.5 in., and LapseRate = <2.5 in. also increased when these observations were incorporated into the network. This change in distribution suggests what can be expected from subsequent evidence and is particularly useful to the forecasters when some data are missing.

To extend the BN to a decision network, the so-called utilities of the possible outcomes and decisions need to be defined, BOFFIN-Melbourne includes one decision node (Forecast) with three decisions (saynofog, CG, and fogOnTAF), while the outcome of interest is whether fog occurred or not, represented by the two-state Fog node. Each decision and outcome combination is given a relative value, here by the forecasters, who know them only qualitatively. For example, they know that the cost of missed forecasts is much greater than that of false alarms. Thus, a saynofog decision while fog occurred is the worst combination and therefore received the lowest utility, whereas a fogOnTAF decision with fog having occurred is the best combination, and received the highest utility. In addition, utilities were calibrated to achieve relative observed fog events of approximately 5% in the CG decision and approximately 25% in the fogOnTAF decision.

The computed so-called expected utilities (combining probabilities and utilities) associated with each possible forecast decision, given no evidence, are shown in Fig. 3 within the Forecast node box. Note that the optimal decision is CG, reflecting the conservative approach to fog forecasting; even though the probability of fog (3.3%) is quite low when nothing is known, the optimal decision is still to issue a CG fog warning. Given any new evidence, the expected utility of each possible decision is recalculated, revising the optimal decision.

Figure 4 depicts examples of BOFFIN-Melbourne in use for two specific scenarios, one conducive to fog, one not, both with evidence for three input nodes. In the fog (Fig. 4a) case, described above, the observations increase the probability of fog to 31% and as a consequence the decision with the highest expected utility, and thus the optimal decision, is revised to become fogOnTAF. The nofog case (Fig. 4b) includes the observations Month = December, LengthOfNight = from November to January, Moisture = unfav, Stern–Parkyn = 1–2, and Gradient = fav. These observations decrease the probability of fog to a very low 0.067% and as a consequence the optimal decision is revised to saynofog.

These examples illustrate that BDNs easily handle partial information and missing observations. Following the addition of newly available observations, it calculates the posterior probability for all unobserved nodes and the expected utility for the decisions according to the observations; these changes are visible to the forecaster. The network also enables the forecaster to simulate reverse inference. For example, entering the condition fog will lead to calculation of the probability of the observations that would lead to its prediction (such as Moisture, etc.). These BDN properties, plus the

---

1 These percentages represent the ratios between the number of fog events that were correctly forecast CG (but not fogOnTAF) and the total number of CG forecasts (but not fogOnTAF), and between the number of fog events that were correctly forecast fogOnTAF and the total number of fogOnTAF forecasts, respectively.
FIG. 4. Examples of evidence propagation. Observations indicate a situation conducive to (a) fog and (b) nofog. Shown in (a) is the model with the evidence Month = May, Moisture = fav, and Stern–Parkyn = 30–100; the new calculated posterior probabilities for the rest of the network; and the expected utilities for each of the decisions. The fogOnTAF is the decision with the highest expected utility (6.815 48), indicated by the horizontal bar. Shown in (b) is the model with the evidence Month = December, Moisture = unfav, Stern–Parkyn = 1–2, and Gradient = fav; the new calculated posterior probabilities for the rest of the network; and the expected utilities for each of the decisions. The saynofog is the decision with the highest expected utility (6.815 48).
careful discretization of inputs into informative ranges that are statistically related to different event probabilities, help the forecasters (i) to understand the network better (and therefore accept it), (ii) to assess how representative and influential a particular observation is, and (iii) to switch quickly between different types of scenarios. From a technical perspective, this technology is robust and can effectively deal with problems such as missing data and model overfitting, making it particularly suited to forecasts of rare events. In addition, it is possible to enter uncertain input as a likelihood vector, where each number in the vector represents the probability of an observation given a state of interest, as when measuring a state with a faulty sensor. This technology is able to provide continuous outputs even with discrete node states. Taken together, BDNs are ideal for using probabilistic information from ensemble runs of an NWP model at some potential cost to complexity and maintenance overheads.

4. Methods

a. Data: Operational use

Cross-validation trials were performed for 1999–2005, in which the BDN outperformed the operational forecasts of the same period (Boneh 2010), leading to its operational deployment at Melbourne Airport since 2006. The BDN runs at two forecast times: 1500 and 2100 LT (possibly also at 1800 LT in winter). At each forecast time, the available observations are input into the BDN. Month and LengthOfNight are known variables, but for those variables that are not yet available, 0000 UTC base-time NWP-predicted values are used as inputs. More specifically, we note the following:

1) At the 1500 LT run, the latest observations are used for the Gradient, whereas LapseRate, Rain, Moisture, and Stern–Parkyn input are NWP predictions.
2) At the 2100 LT run, the 1500 LT observations are used for Gradient, the 2100 LT (or 1800 LT) observations are used for Moisture, the 2100 LT observations are used for LapseRate, and Rain and Stern–Parkyn input are NWP predictions.

In addition, forecasters can override both the observations and the NWP predictions at any of the run times if they think that they are not representative of the present meteorological situation. A summary of inputs is available in Table 1.

b. Evaluation methodology

BOFFIN-Melbourne operational decisions were archived during its operational use from 2006 to 2013. To evaluate the performance of the BDN, three sets of decisions were collected for the two forecast times 1500 and 2100 LT (Table 2). The first was a set that provides an upper limit of automated performance for BOFFIN-Melbourne with omniscient knowledge (BDN-Omniscient) using the perfect prognosis method, showing how well BOFFIN-Melbourne could possibly perform with all observational variables utilized even when their values were unavailable in real time. For example, it included the observation for actual rainfall in the forthcoming forecast day, a figure that was unknown at any of the operational forecast times. Even omniscient BOFFIN-Melbourne cannot forecast perfectly, because the model and observations do not capture every relevant aspect of the atmosphere.

The second set was the decisions BOFFIN-Melbourne generated alone in real time, unaided by the forecasters (BDN-Autonomous). These decisions were based on the available observations at the time and NWP data for predictions (e.g., rainfall amount until 0900 LT the next morning). The performance of this set reveals the real-time skill of BOFFIN-Melbourne without the forecasters’ aid, demonstrating its ability to generate automated decisions. The third set (BDN-Forecasters) comprised the operational decisions. These were based on BOFFIN-Melbourne, but the forecasters could

---

2 The Stern–Parkyn input is a combination of available 1500 LT observations and predictions of the 0900 LT (next morning) MSLP.
replace observations or predictions with their own estimations represented in Table 2 as Obs–frcst and NWP–frcst, respectively. This set demonstrates the operational skill of the forecasters when using BOFFIN-Melbourne. The difference between the BDN-Autonomous set and the BDN-Forecasters set addresses the question of what extra value, if any, forecasters add when they adjust the inputs to BOFFIN-Melbourne. Table 2 summarizes the sources of data used in each set at each time.

c. Evaluation scores

Traditionally, fog forecasting skill may be evaluated using a contingency table that shows the number of forecasted fog and nofog cases and the actual occurrences of fog and nofog events (Fig. 5). This creates four combinations of forecasts: hit, where an event was forecast and occurred; miss, where an event was not forecast but occurred; false alarm, where an event was forecast but did not occur; and correct negative, where an event was not forecast and did not occur. Different skill scores can be calculated using the contingency table, the most commonly computed for fog forecasting are the probability of detection (POD), which is the ratio between the hits and the number of fog events, and the false alarm ratio (FAR), which is the ratio between the number of false alarms and the number of fog forecasts. POD and FAR range from 0 to 1, with 1 being a perfect POD, and 0 being a perfect FAR (Fig. 5). When the output of the forecast is the probability of fog, then different probability thresholds are used as a cutoff between the decisions fog or nofog, and the POD and FAR are calculated for each of the selected thresholds. Mathematically, the receiver operating characteristic (ROC) represents the trade-off between the POD and the false alarm rate (denoted as $F$ in Fig. 5). Thus, it visually represents the trade-offs between hits and false alarms for the different probability thresholds. The most common summary score of these trade-offs is the area under the ROC curve (AUC; Mason and Nicholas 1999). We use these metrics (POD, FAR, and AUC) to evaluate BOFFIN-Melbourne against previous operational fog forecasting performance.

BOFFIN-Melbourne was evaluated with regard to three decisions, that is, saynofog and the two warning levels: CG and fogOnTAF. Contingency tables were calculated for each of the warning levels, forecast times, and collected sets (Table 3). The POD and FAR results of BOFFIN-Melbourne were calculated from each of the contingency tables and compared with results of the years before the use of the BDN (1999–2005 set), as shown in Table 4.

For the calculation of the significance of the differences between the PODs and FARs from the different contingency tables, we followed the recommendation from a review of standard methods by Fagerland et al. (2015) to use the Newcombe hybrid score interval...
This method is based on the Wilson score confidence interval for a single proportion (Wilson 1927). A confidence level of 0.95 was defined as significant, unless stated otherwise. AUCs were calculated using continuous probability outputs from the BDN’s Fog node. The AUCs of the different sets were compared using the DeLong algorithm (DeLong et al. 1988) as implemented in the R package (R Core Team 2014) pROC (Robin et al. 2011).

5. Results

The operational forecasting results for 7 yr (1999–2005) with 92 fog days out of 2557 days were calculated and used as the baseline for the assessment of the skill of BOFFIN-Melbourne during the period of its operational use (2006–13) with 105 fog days out of 2922 days. A comparison between the POD, FAR, and AUC scores for these periods is shown in Table 4.

The BDN-Forecasters scores (the operational scores, 2006–13) for CG at 1500 and 2100 LT were better than those from 1999 to 2005, with a significantly higher POD and no significant change in FAR. BDN-Forecasters scores for TAF at 1500 and 2100 LT showed a trend for improvement over the 1999–2005 period, with a somewhat lower FAR (significant at 1500 LT at the 0.90 but not at the 0.95 confidence level), and nonsignificant improvement in the BDN-Forecasters POD. The improvements in the AUC over 1999–2005 at 1500 and 2100 LT were statistically significant, thus indicating improved performance for both forecast times.

A comparison between the BDN-Autonomous and the 1999–2005 scores revealed that the POD for CG at 1500 LT significantly improved with no significant change to FAR. At the TAF warning level, the differences between BDN-Autonomous and 1999–2005 for POD and FAR were not statistically significant. The BDN-Autonomous scores for 2100 LT CG were similar to those at 1500 LT with the exception of a trend toward improvement of the BDN-Autonomous POD over 1999–2005, which was only significant at the 0.90 confidence level. This is reflected in the AUC. The AUC for 1500 LT in BDN-Autonomous was significantly higher than that in 1999–2005, but the improvement in 2100 LT did not reach statistical significance. Overall, AUC results suggest the BDN using data available in real time outperforms pre-BDN official forecasts.

Comparing between BDN-Forecasters and BDN-Autonomous scores for CG at 1500 LT indicated that there was nonsignificant improvement in POD and FAR. A comparison between the TAF at 1500 LT scores revealed nonsignificant improvement in POD due to

<table>
<thead>
<tr>
<th>Method</th>
<th>Time of forecast (LT)</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous</td>
<td>1500</td>
<td>74</td>
<td>578</td>
<td>18</td>
<td>1887</td>
<td>44</td>
<td>129</td>
<td>48</td>
<td>2336</td>
</tr>
<tr>
<td></td>
<td>2100</td>
<td>76</td>
<td>587</td>
<td>16</td>
<td>1878</td>
<td>53</td>
<td>159</td>
<td>39</td>
<td>2306</td>
</tr>
<tr>
<td>BDN-Forecasters</td>
<td>1500</td>
<td>97</td>
<td>906</td>
<td>8</td>
<td>1911</td>
<td>56</td>
<td>103</td>
<td>49</td>
<td>2714</td>
</tr>
<tr>
<td></td>
<td>2100</td>
<td>99</td>
<td>915</td>
<td>6</td>
<td>1902</td>
<td>63</td>
<td>121</td>
<td>42</td>
<td>2696</td>
</tr>
<tr>
<td>BDN-Autonomous</td>
<td>1500</td>
<td>95</td>
<td>938</td>
<td>10</td>
<td>1879</td>
<td>47</td>
<td>143</td>
<td>58</td>
<td>2674</td>
</tr>
<tr>
<td></td>
<td>2100</td>
<td>95</td>
<td>930</td>
<td>10</td>
<td>1887</td>
<td>60</td>
<td>212</td>
<td>45</td>
<td>2605</td>
</tr>
<tr>
<td>BDN-Omniscient</td>
<td>1500</td>
<td>92</td>
<td>783</td>
<td>13</td>
<td>2034</td>
<td>69</td>
<td>218</td>
<td>36</td>
<td>2599</td>
</tr>
<tr>
<td></td>
<td>2100</td>
<td>91</td>
<td>767</td>
<td>14</td>
<td>2050</td>
<td>68</td>
<td>219</td>
<td>37</td>
<td>2598</td>
</tr>
</tbody>
</table>

TABLE 4. POD (%), FAR (%), and AUC results for fog warning levels CG and fogOnTAF at forecast times 1500 and 2100 LT. Contingency tables are in Table 3.
forecaster intervention but BDN-Forecasters FAR significantly outperformed BDN-Autonomous FAR. There was no significant difference in the AUC. For the 2100 LT scores, the only significant difference was in TAF at 2100 LT, with BDN-Forecasters FAR being significantly better than BDN-Autonomous. There was no significant difference in the AUC between the two sets. In summary, all BDN-Forecasters scores were numerically better than BDN-Autonomous scores, with some FAR differences being significant. This result, which might be expected as the BDN models only limited aspects of the atmosphere, indicates the usefulness of forecasters in the process. The spatially and temporally sparse inputs into the BDN are generally proxies for more ideal, but unavailable predictors of fog. For example, Moisture at Laverton at 2100 LT may be unrepresentative of later changes, and it may be that in such cases, forecasters can estimate more appropriate inputs into the BDN or correct for biases in NWP guidance.

We expected BDN-Omniscient might show the best scores because it is based on the actual observations (a perfect prognosis approach) and not on NWP predictions, yet we thought that forecasters might add value as mentioned above. For CG at 1500 LT, there were nonsignificant improvements in BDN-Omniscient POD over the 1999–2005 POD. Differences between BDN-Omniscient FAR and 1999–2005 FAR were also not significant. For TAF at 1500 LT, BDN-Omniscient POD significantly outperformed the 1999–2005 POD and BDN-Autonomous POD, but with no significant change in FAR, and a consequent significant advantage in AUC. Thus, BDN-Omniscient significantly outperformed the BDN-Autonomous and 1999–2005 results at the TAF level, with mixed but nonsignificant results at the CG level.

A comparison between BDN-Omniscient and BDN-Forecasters scores for CG or TAF at 1500 and 2100 LT shows a nonsignificant improvement in POD for BDN-Forecasters with little change in FAR, possibly indicating the usefulness of forecasters in the process for more marginal fogs. For TAF at 1500 and 2100 LT, BDN-Omniscient FAR was significantly outperformed by the BDN-Forecasters FAR, but BDN-Omniscient outperformed BDN-Forecasters in POD (only significant at the 0.90 confidence level), indicating a different POD–FAR trade-off. The overall balances between POD and FAR indicated by AUC were similar for both BDN-Omniscient and BDN-Forecasters.

There was no significant improvement in POD nor in FAR at 2100 over 1500 LT for either CG or TAF in any of the sets of results (1999–2005 and the three 2006–13 sets) although there were trends toward improvement in 2100 LT AUC over 1500 LT AUC in all cases. It should also be noted that the AUC scored better in the three sets of results than in the 1999–2005 set, with no significant differences between the three sets themselves.

We conclude that (higher and lower) BDN probabilities are better able to discriminate between more and less conducive conditions for fog than previous fog forecasts, given their improved AUC scores. In less conducive conditions (in which it may be harder to forecast fog occurrence), fog duration is expected to be on average shorter. We show that indeed the BDN associates lower probabilities with shorter fog events as expected (Table 5). The frequency of fogs of shorter average durations may provide a limited indication of forecast difficulty. For example, there was a greater fraction and number of fog events of short durations up to 30, 60, and 90 min yr$^{-1}$, indicating potentially more difficult forecasts, during the 2006–13 period than in 1999–2005.

There were several causes of unforecast fogs and false alarms. A common forecast failure was due to low cloud affecting airport operations when fog was forecast, or vice versa. We found that mesoscale situations and predictors for low cloud were often similar to those for fog, with low cloud being more likely in stronger southerly synoptic flows. A number of unforecast fogs occurred after or during rain, or after a mesoscale air-mass change later than 1800 or 2100 LT that made some or all afternoon or evening predictors less representative of fog likelihood. In addition, cloud is not modeled in the network yet had an effect in some cases. A number of the unforecast fogs were of limited extent, and of short duration, as expected (e.g., from Table 5).

### Table 5. BDN fog probability vs fog duration.

<table>
<thead>
<tr>
<th>BOFFIN fog probability (%)</th>
<th>Mean fog duration (min) during 2006–13</th>
<th>No. of fog events</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–10</td>
<td>118</td>
<td>24</td>
</tr>
<tr>
<td>10–50</td>
<td>141</td>
<td>39</td>
</tr>
<tr>
<td>50–100</td>
<td>300</td>
<td>42</td>
</tr>
</tbody>
</table>

### 6. Conclusions and future work

#### a. Conclusions

The aim of this project was to develop a BDN for fog forecasting at Melbourne Airport that would be accepted by the forecasters and that has equal or better performance compared with historical operational skill. The network has been accepted by forecasters and continuously used since 2006. In terms of our metrics, our aim was to achieve significant improvements in...
AUC and, at the least, either a significant improvement in POD without a significant change in FAR or vice versa. The performance of BOFFIN-Melbourne was compared with previous operational results at two forecast times: 1500 and 2100 LT. Comparison of the AUC showed significant improvement in forecast skill at both forecast times, reflecting improvements in POD and/or FAR, and thus meeting our aims. BOFFIN-Melbourne exhibited good forecasting ability at the intended forecast times.

It is conceivable that some of the improvements in forecast skill utilizing the BDN in 2006–13 relative to 1999–2005 have arisen from greater NWP accuracy in forecast skill utilizing the BDN in 2006–13 relative to intended forecast times. This possibility actually underscores an advantage of the BDN, in being able to directly utilize any future NWP improvements.

b. Future work

Future work will aim to broaden the scope of all aspects of modeling and forecasting fog. Several potential predictors were excluded from the current network. This was done because some, such as temperature inversion, airmass changes, and low-level winds, are accounted for, to some extent, by predictors that had been included. Other predictors, such as low cloud or fog occurring the previous morning, cloud cover, and afternoon visibility, were excluded because they are quite hard to quantify. Moreover, we employed existing discretization rules used by forecasters, the formation of which involves subjective expert judgment, and are time consuming and expensive. Incorporation of sophisticated techniques from statistics and machine learning in the domain will speed the process of objective identification and quantification of potential predictors. This, in turn, may lead to better meteorological understanding of fog formation and better expression of the forecasters’ conceptual models in measurable terms. Some predictors show value at short lead times, and we have found BDN performance improves as we approach nowcasting time frames (Boneh 2010). Taken together, these methodologies will improve the value and usefulness of the model.

BDN technology allows the model to adapt and learn in real time. We decided to freeze the model for operational use, and to update it with lessons learned from the use of the network and archived data only after extensive evaluation and testing. Thus, the online adaptation option will be explored in future research, when implications are better understood.

We found that forecasters preferred to use BOFFIN-Melbourne when it was incorporated into the 0606TAF and Fog–FDSS framework tool, which they already used operationally. Further research on a standard interface that incorporates BDNs into the Graphical Forecasting Editor (GFE: Glahn and Ruth 2003), which has become the main operational framework tool in the BoM, may prove valuable.

A future change in the conceptual modeling of the BDN technology allows the model to adapt and learn in real time. It is conceivable that some of the improvements in forecast skill utilizing the BDN in 2006–13 relative to 1999–2005 have arisen from greater NWP accuracy in forecast skill utilizing the BDN in 2006–13 relative to intended forecast times. This possibility actually underscores an advantage of the BDN, in being able to directly utilize any future NWP improvements.

Acknowledgments. We gratefully acknowledge Belinda Roux and Kathleen Hirst for the topographic map of the Melbourne Airport area, and Mey Manickam for providing data for the evaluation. We thank the reviewers for feedback that has improved the clarity and presentation of the paper. This work was supported by the Meteorological Service Charge from the Australian aviation industry.

REFERENCES


Keith, R., 1991: Results and recommendations arising from an investigation into forecasting problems at Melbourne Airport. Meteorological Note 195, Bureau of Meteorology, Melbourne, Victoria, Australia, 19 pp.


Spillane, K. T., 1978: Atmospheric characteristics on high oxidant days in Melbourne. Clean Air, 12, 50–56.


