

Bayesian Echo Classification for Australian Single-Polarization Weather Radar with Application to Assimilation of Radial Velocity Observations

S. J. RENNIE, M. CURTIS, J. PETER, A. W. SEED, AND P. J. STEINLE

Centre for Australian Weather and Climate Research, Bureau of Meteorology, Melbourne, Victoria, Australia

G. WEN

Meteorological Observation Center, China Meteorological Administration, Beijing, China

(Manuscript received 2 November 2014, in final form 2 March 2015)

ABSTRACT

The Australian Bureau of Meteorology's operational weather radar network comprises a heterogeneous radar collection covering diverse geography and climate. A naïve Bayes classifier has been developed to identify a range of common echo types observed with these radars. The success of the classifier has been evaluated against its training dataset and by routine monitoring. The training data indicate that more than 90% of precipitation may be identified correctly. The echo types most difficult to distinguish from rainfall are smoke, chaff, and anomalous propagation ground and sea clutter. Their impact depends on their climatological frequency. Small quantities of frequently misclassified persistent echo (like permanent ground clutter or insects) can also cause quality control issues. The Bayes classifier is demonstrated to perform better than a simple threshold method, particularly for reducing misclassification of clutter as precipitation. However, the result depends on finding a balance between excluding precipitation and including erroneous echo. Unlike many single-polarization classifiers that are only intended to extract precipitation echo, the Bayes classifier also discriminates types of nonprecipitation echo. Therefore, the classifier provides the means to utilize clear air echo for applications like data assimilation, and the class information will permit separate data handling of different echo types.

1. Introduction

The use of radar observations for data assimilation (DA) in NWP is growing with the development of high-resolution NWP. Quality control is vital for data assimilation because the impact of a few bad observations can be substantial (Rabier et al. 1996), damaging a forecast. For DA and other quantitative applications of radar data, quality control (QC) that provides flexibility depending on the application is desirable. Many echo identification algorithms have been developed in recent years, particularly those utilizing dual-polarization parameters (e.g., Bachmann and Zrnić 2008; Dixon et al. 2005; Koistinen et al. 2009; Schuur et al. 2003). These are able to discern various echo types, including different hydrometeor types. Echo identification without dual

polarization is difficult but necessary for assimilation of observations from single-polarization radar.

The Australian Bureau of Meteorology (BoM) has recently upgraded selected parts of its single-polarization network to Doppler capability. This, along with the BoM's development of high-resolution (1.5 km) NWP, means that the assimilation of radar observations is desirable to improve the model initialization and reduce spin-up time (Dance 2004; Salonen et al. 2011; Sun 2005; Zhao et al. 2008). The BoM's high-resolution limited area models (LAMs) use the Australian Community Climate and Earth-System Simulator (ACCESS) NWP system (Puri et al. 2010). The BoM is developing the assimilation of radial velocity observations for the LAMs over Australia's capital cities (ACCESS-City systems), and it requires a means to select observations that provide good wind estimates. Unfortunately, the BoM is some years away from dual polarization in the operational weather radar network, so QC methods for single-polarization radars must be used to extract suitable observations for assimilation.

Corresponding author address: S. J. Rennie, Bureau of Meteorology, GPO Box 1289, Melbourne VIC 3001, Australia.
E-mail: s.rennie@bom.gov.au

TABLE 1. List of the classes used for manual classification: class number, class name, and class abbreviations as used for figure labels, etc.

Class No.	Class	Class abbreviation
1	Convective precipitation	con
2	Shallow convective precipitation	shc
3	Stratiform precipitation	str
4	Insects	ins
5	Smoke (bushfires)	smk
6	Chaff	chf
7	Macroaerofauna (birds/bats)	brd
8	Permanent ground clutter	pe
9	AP ground clutter	gc
10	AP sea clutter	ap
11	Sidelobe sea clutter	sl
12	Second-trip echo	2tp

Most single-polarization methods focus on removing unwanted (i.e., nonprecipitation) echo from radar data. Discrimination between precipitation and nonprecipitation has been shown using a neural network (Lakshmanan et al. 2007). Biological bloom patterns were incorporated into the neural network to improve removal of biological echo (Lakshmanan et al. 2010). Statistical pattern classification to remove anomalous propagation from single-polarization radar was demonstrated by Moszkowicz et al. (1994) to be effective.

Recently, Peter et al. (2013) developed a naïve Bayes classifier (NBC) to discriminate anomalous propagation (AP) sea clutter and precipitation, using reflectivity and feature fields based on reflectivity. These feature fields included echo top height, vertical gradients, spin (Steiner and Smith 2002), and texture (Hubbert et al. 2009; Kessinger et al. 2005). Here the structure of the NBC has been extended to classify a wide range of echo types and to use Doppler information, including radial velocity and spectrum width. The prior probabilities for some classes are decided by geographical information, such as distance from the coast and probability of detection maps.

This paper explains how the classifier was developed and trained using a manually classified dataset. The NBC is assessed against the training dataset for a quantitative measure of its efficacy. It has also been implemented so that it runs routinely on radar data from the BoM network, and so it can be qualitatively assessed by regular inspection of the classification results. The classifier is tested against an existing method for precipitation identification that it will replace. The application of the classification information to Doppler radar data assimilation is discussed. Finally, possible advances to the NBC are examined.

2. The classifier

An NBC selects the most likely class of a range of classes based on the value of various observed feature

fields and the prior probability that the class will occur. For each class, there is a pdf to describe the likelihood of the range of values for each feature field. Good discrimination relies on these pdfs overlapping as little as possible. The feature fields are also assumed to be independent (the naïve aspect), as useful results have been shown to be obtained with dependent feature fields (Friedman et al. 1997; Peter et al. 2013). The alternatives are to use a few fields that are known to be independent, or a much more complicated implementation that was not anticipated to yield benefits matching the effort involved. The probability of an occurrence of class c based on a range of n feature field values x_1, \dots, x_n is determined by

$$P(c|x_1, \dots, x_n) = \frac{P(x_1, \dots, x_n|c)P(c)}{P(x_1, \dots, x_n)}, \quad (1)$$

where $P(x_1, \dots, x_n|c)$ is the conditional probability of observing a feature field value x_i given it belongs to class c ; $P(c)$ is the prior probability of a given class; and $P(x_1, \dots, x_n)$ is the probability of obtaining a particular value. The denominator term is constant and can be ignored. The classifier is described in more detail in Peter et al. (2013), with the important difference that here $P(c)$ is not assumed equal for all classes.

The NBC has been implemented in the BoM's new in-house radar data handling software (Ancilla). This software contains the framework for all aspects of the classifier: it creates the feature fields, recognizes a range of pdfs to describe each feature field, accepts the prior probabilities, and runs the classifier. It also contains tools to train the classifier, by aggregating feature field values from the training dataset to create histograms. Additionally, a tool to visualize and to manually class radar volumes is provided. The classes selected to be used by the classifier are listed in Table 1, along with their abbreviation and number, which are used throughout this paper. This is considered to be a comprehensive list of the major echo types that are typically seen on Australian weather radars.

a. Radar data

The training dataset contains around 200 radar volumes from a range of radars, mostly from 2012 although events from 2009 through 2013 were used to provide examples. The selected radars were primarily Doppler radars, many of which are within the Sydney test bed area for the BoM's ACCESS-City development. Most are S-band radars and the rest are C band (Fig. 1). All radars make plan position indicator scans over 14 elevations between 0.5° and 32° , with 1° azimuthal resolution. The range resolution for Doppler radars is 250 or

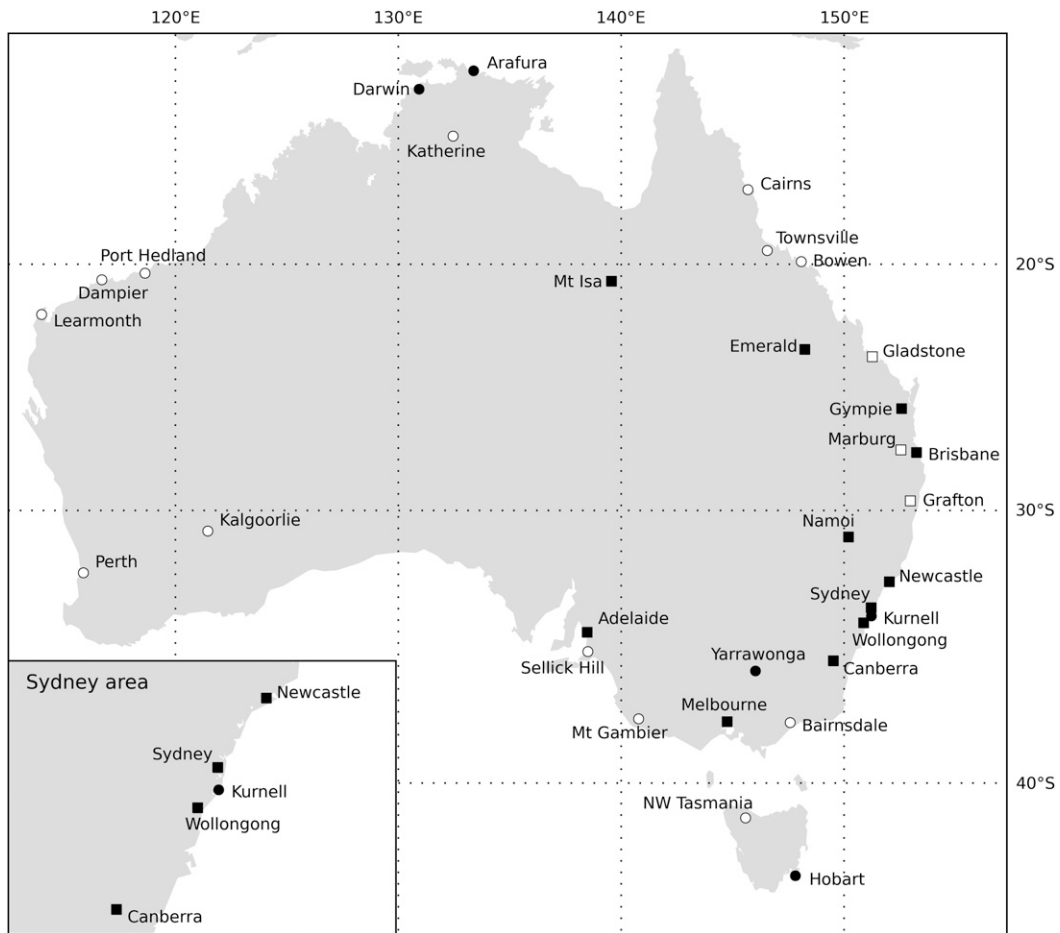


FIG. 1. Map of dedicated weather watch radars in the BoM network. C-band radars are marked with filled or open circles (● or ○), S-band radars are marked with open or filled squares (□ or ■). Radars from which Doppler information is returned have solid black markers (● or ■); radars without Doppler information are open white (□ or ○). Note that Newcastle and Canberra were upgraded to Doppler around the end of the period of the training data.

500m and is up to 1000m for non-Doppler radars. Beamwidth may be 1° or up to 2°. The Nyquist velocities vary but are typically 26, 39, or 52 m s⁻¹ and may change periodically. The maximum range is between 150 and 300 km. Specifics of the radars used for the training dataset are included in Table 2. Note that on-site radar processing includes a Doppler zero-velocity filter to remove ground echo and applies a signal quality index (lag-one correlation coefficient) threshold.

Volumes were selected to cover the range of classes and to include multiple examples of each class from multiple radars. Only one radar (Wollongong) provided spectrum width during the bulk of the period from which the training dataset was drawn, so effort was made to manually classify all classes using Wollongong data. A second-trip echo was not recorded at Wollongong, and AP ground clutter was assumed to have the same spectrum width as permanent ground clutter. More recently

other radars also started to supply spectrum width, so this parameter is now used to classify echo from several radars. A summary of radar volumes used to create the training dataset is shown in Table 2, which contains the number of volumes from each radar that contributed to each class's training data. For most classes and feature fields, the amount of classed data seemed to capture the climatology and further additions to the training dataset only slightly altered the resultant histogram of feature field values. Echo top height is particularly difficult to capture, as it is somewhat quantized by beam elevation, and the rare classes are also difficult to describe.

The radar volumes are stored in Hierarchical Data Format, version 5 (HDF5) following the Operational Programme for the Exchange of Weather Radar Information (OPERA) Data Information Model (Michelson et al. 2011). Manual classification was done by creating a "CLASS" field within the volume, visualizing the

volume on screen with the Ancilla viewing/editing graphical user interface (GUI), and using other fields (e.g., reflectivity, radial velocity) to identify and “paint” the echoes in the CLASS field with the appropriate class value as per Table 1. Values from each feature field could then be extracted according to the value of the CLASS field. Echo types were identified by an expert user and only pixels with known echo type were classified. Note that the classifier does not need to discriminate accurately between convective and stratiform precipitation types, so the manual classification did not need to perfectly separate these precipitation types.

b. Feature fields, histograms, and pdfs

The feature fields used by the classifier include those moments recorded by the radar, and the fields derived from them. The non-Doppler feature fields were explored by Peter et al. (2013), from which the present version was developed, though with texture kernels extended to two dimensions.

Reflectivity (DBZH), radial velocity, and spectrum width are the potential raw fields. Radial velocity itself is not used because it is not informative, especially since Doppler filtering is already applied for ground clutter removal. Spectrum width (WAVG) is averaged using a Gaussian kernel across adjacent azimuths because the pulse repetition frequency (PRF) alternates with azimuth and spectrum width was found to depend on this. Various feature fields were derived from reflectivity and radial velocity.

Texture T of a field X (Hubbert et al. 2009; Kessinger et al. 2005) is calculated by

$$T = \left[\sum_j^N \sum_i^M (X_{i,j} - X_{i-1,j})^2 \right] / (N \times M) \quad (2)$$

and measures the squared difference of X (e.g., reflectivity) between adjacent pixels within a kernel $N \times M$ (where $N = M$ for this work). Reflectivity texture (ZTEX) was calculated with a kernel of 11×11 . Radial velocity texture (VTEX) was calculated with a kernel of 15×15 . The kernel sizes were selected to optimize the difference between values for different classes (Rennie et al. 2014).

Spin (Steiner and Smith 2002) is defined as a measure of the change in sign of the reflectivity difference between adjacent bins within a kernel. Specifically, the value of spin is the number of valid fluctuations as a percentage of the number of possible fluctuations within the kernel. The valid measurable spin fluctuation fulfills the following conditions for successive bins X_{i-1} , X_i , and X_{i+1} :

$$\text{sign}\{X_i - X_{i-1}\} = -\text{sign}\{X_{i+1} - X_i\} \quad (3)$$

$$\frac{|X_i - X_{i-1}| + |X_{i+1} - X_i|}{2} > \text{spin threshold}. \quad (4)$$

Reflectivity spin (SPIN) used a threshold of 3 dBZ and a kernel of 19×19 . The texture and spin are similar, so the different kernel size for texture and spin of reflectivity helped to make these more independent. The ZTEX–SPIN correlation coefficient was 0.55, the highest of any pair of feature fields. Results using only one of these were slightly worse for clutter detection (not shown).

The vertical gradient of reflectivity (VGR) is a measure of the difference in reflectivity between bins of the same along-ground range and azimuth at adjacent elevations divided by the difference in altitude of the beam centers. Reflectivity is smoothed with a 3×3 Gaussian filter before this is calculated.

Echo top height (ETH) is the beam center altitude at which the vertical profile of reflectivity drops below some threshold. Beam height is calculated using the standard $4/3$ (effective) Earth radius approximation (e.g., Doviak and Zrnić 1993, p. 21). Two thresholds were used: 4 dBZ was used for ETH and -5 dBZ was used for ETH2. ETH2 was only used if ETH did not exist for the same location. Using two thresholds was found to give slightly better results than either alone. A low threshold is necessary to maximize the coverage of this feature field; a high threshold would make ETH unavailable for the weaker echo types.

A range of standard and composite pdfs were available to fit to the histograms. Functional pdfs allow for better representation of undersampled classes and for avoiding artifacts from histogram binning. These were chosen empirically and not from an expectation that the climatology of the feature field would conform to a particular pdf. The following pdfs were included:

- Trapezoidal distribution, which linearly increases to a plateau
- Normal (Gaussian) distribution
- Inverse normal distribution
- Lognormal distribution
- Skew-normal distribution
- Truncated normal distribution
- Exponential distribution
- Gamma distribution
- Inverse gamma distribution
- Laplace distribution
- Laplace-normal distribution, a composite pdf with the Laplace distribution centered on 0 combined with normal distributions located at plus–minus their mean; this can be set to exist in only the positive domain
- Laplace–Laplace distribution, a composite of two collocated Laplace distributions

- Laplace–skew-normal distribution, a composite of a Laplace distribution and a skew-normal distribution, both centered anywhere
- Log-binormal distribution, a composite of two lognormal distributions

For each feature field, values for each class were aggregated and histograms created. The normalized histograms were created, selecting a number of bins following [Izenmann \(1991\)](#); that is, $W = 2(\text{IQR})N^{-1/3}$, where W is the bin width, IQR is the interquartile range, and N is the number of data points. Thus, the number of bins $n_{\text{bins}} = (\text{maxval} - \text{minval})/W$ rounded up to the nearest integer. Some processing was required to produce reasonably smooth histograms. Reflectivity is provided in radar-dependent rounded values at intervals that periodically decrease with increasing value. This would produce a very uneven histogram highly dependent on bin choice. Therefore, reflectivity and functions of reflectivity were dithered to reduce the effect of having rounded values when creating the histogram. Dithering was accomplished by adding or subtracting a random quantity to each value, which spreads each rounded value to within half the interval to its neighbor values. The result is a smooth histogram with narrow bins, better suited for fitting pdfs. Spikes occur in the histograms under two circumstances: where $\text{VGR} = 0$ because the reflectivity was often identical between elevations and where the interval between rounded reflectivity values changed. Since these artifacts are not meaningful to the distribution, their removal was accomplished by fitting a pdf, deleting outliers, and then interpolating the histogram across the space using the pdf values.

The pdfs were fit to the histograms using Python software. The SciPy statistics package provided some pdfs and the remainder were manually coded. The SciPy optimization package was used to perform a least squares fit (using FITPACK) to each histogram to find the optimal parameters for the each pdf. Initial guesses were calculated using the histogram data to ensure the correct local minimum was near the start point for the fit optimization. The best-fit pdfs for each histogram based on root-mean-square residuals and the Kolmogorov–Smirnov statistic were plotted overlaying the histograms. These were visually verified, and the best representation based on both statistics was noted. Sometimes the simpler pdf was chosen if two pdfs gave identical fits. In a few instances a different pdf was chosen to be more realistic. For example, if the “best” pdf was unrealistic because it had no tail (e.g., trapezoid), then another pdf was selected. For a few difficult cases where the optimization did not appear to

automatically reach the right local minimum and no good fits were found, the pdf parameters were manually derived to achieve an accurate representation of the histogram (ETH/ETH2: ap; WAVG: str and chf; see class abbreviations in [Table 1](#)). The parameters for the best fits were then inserted into the Ancilla classification scheme.

For classification, feature field values at the tails of pdfs were converted to NaN (not a number) so that these would not be used. This is partly because there is doubt that the tails are well fit and partly because extremes may not be indicative of class; for example, extreme VTEX values may result from velocity dealiasing errors and should therefore be ignored.

Full details of the pdfs used can be found in [Rennie et al. \(2014\)](#), and they are shown in [Fig. 2](#). Generally there is not great separation between the classes, though there are some cases where classes are quite different. For example, insect echo and permanent echo typically have low echo top height. Permanent ground clutter can have high ZTEX and WAVG.

c. Prior probabilities

The final requirement for the NBC is the prior probability of a class $P(c)$. In theory this might be the climatological occurrence of a class, but in practice that would mean that rare classes would almost never be identified, even if they composed the majority of echoes in a scan. For data assimilation it may be more important to identify and remove these rare classes. The prior probabilities are therefore selected to behave as weights rather than true probabilities. For best classification results, the least number of possible classes should be permitted to the classifier for any pixel (by setting some prior probabilities to 0). The climatology and features of the different classes that could affect the prior probability are discussed below.

Precipitation (three classes: con, shc, str) is very common and its prior probabilities reflect this. Insect echo (ins) is also very widespread, although it is not observed by the less sensitive radars and at locations where migrating insects are less numerous, including colder climates and over the ocean. Macroaerofauna (birds and bats: brd) are usually localized and sporadic in appearance, typically as dusk or dawn dispersals. Australia lacks the large-scale bird migrations seen in some parts of the world; the radars do not see scans dominated by bird echoes with the resultant widespread velocity signal (e.g., [Dokter et al. 2011](#)). Smoke (smk) and chaff (chf)—the other “clear air” echo types—are sporadic but when present can dominate the radar scan for hours (and occasionally longer). Our experience is that chaff is typically released over the ocean, or

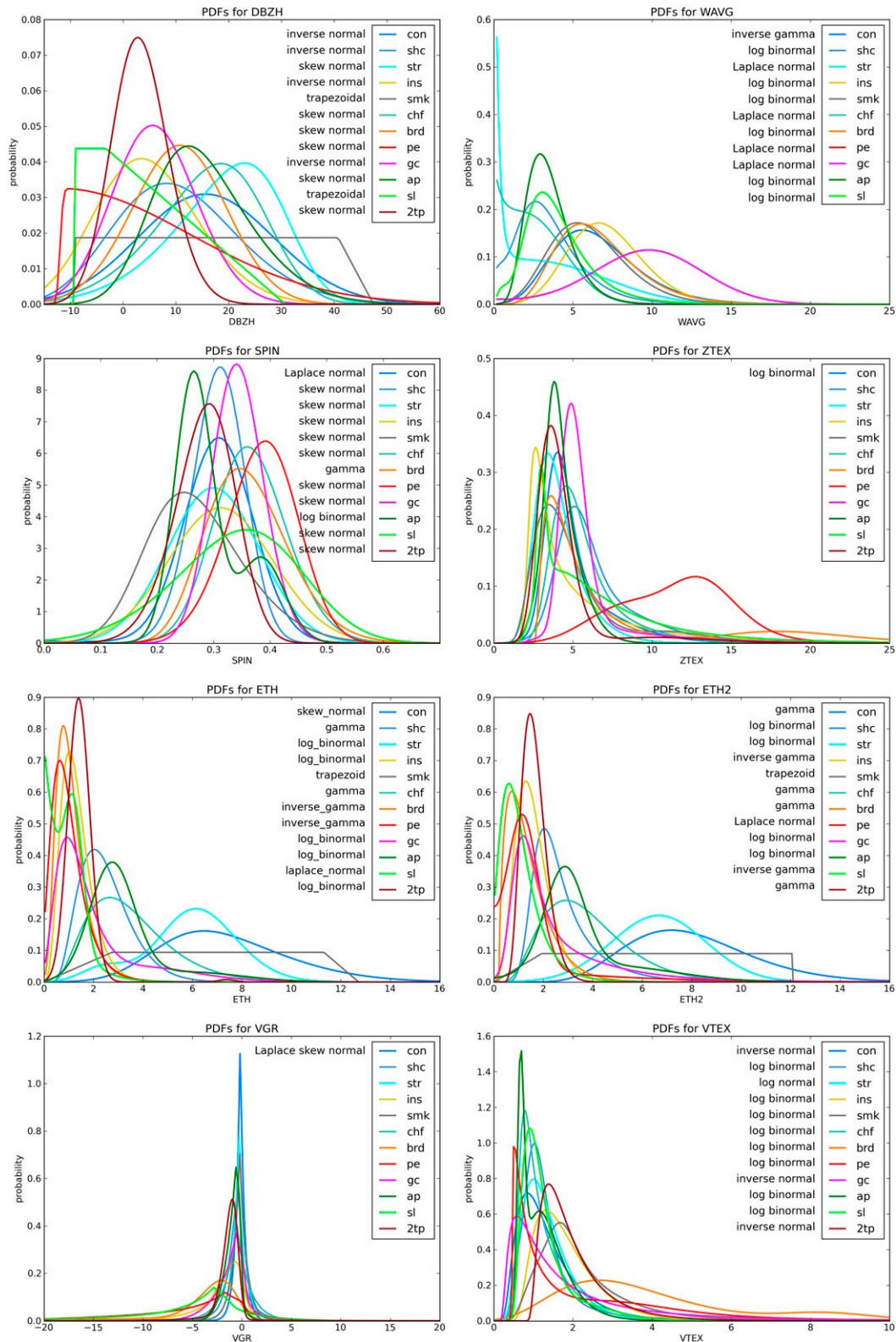


FIG. 2. Pdfs for the eight feature fields used to classify echo types. Pdfs are noted, for each class where different.

TABLE 3. The typical prior probabilities for each class. Variations are radar dependent as noted.

Class	Prior probability	
Convective precipitation	0.4	
Shallow convective precipitation	0.25	
Stratiform precipitation	0.5	
Insects*	0.4 over land (to 8 km offshore)	0 over sea
Chaff***	0.05 over land	0.15 over sea
Smoke (bushfires)*	0.1	
Birds/bats*	0.3 over land	0 over sea
Permanent ground clutter	POD over land	0 over sea
AP ground clutter	0.1 over land (to 1 km offshore)	0 over sea
AP sea clutter	0.1 over sea	0 over land
Sidelobe sea clutter	1.2 × POD over sea	0 over land
Second-trip echo	0	

* Prior probability of clear air echo types was set to 0 for radars with low reflectivity resolution and low sensitivity.

** Prior probability of chaff set to 0 if radar is not near a known chaff release site. Chaff is usually released over the ocean.

occasionally inland in areas of low population, and near air force bases, so it has not been observed at all radars.

Ground and sea echo are discriminated by the distance from the coast, which has been defined as positive over land and negative over ocean. The prior probabilities are altered by a distance-from-coast threshold (which is also applicable to aerofauna echo). This ensures that echoes can only be classified as either ground or sea clutter in any location, or both along the coastline.

There are two types of ground and sea clutter: permanent and AP. Permanent ground (pe) and beam edge [sidelobe (sl)] sea clutter have prior probabilities as a function of the probability of detection (POD) maps created for each radar. Although clutter filtering is applied on-site, there remains some ground clutter echo, which in the POD map creates haloes around the holes where ground clutter is consistently removed. Typical POD values for these haloes are 10%–50%. The sidelobe sea clutter typically appears in a wedge shape near the radar, where the sidelobes or beam edges are not blocked by topography, and the POD value can reach 80% or higher. AP ground (gc) and sea (ap) clutter are both sporadic, and some radars are climatologically more prone than others, especially those with a wider beam.

A few different schemata were created based on different radar types and locations, including whether chaff had been seen at that radar. The prior probabilities are described in Table 3. Some are constant values and some spatially vary as a function of POD and/or distance from the coast (land/sea discrimination). Second-trip echo

was ultimately excluded because of its rarity, so its prior probability is 0.

3. Assessment of the classifier

The primary means of assessing the NBC are quantitatively against the training dataset and qualitatively by monitoring its output over time. The ideal method to quantitatively test the NBC would be against an independent classified dataset, for example, another manually classified dataset. However, the resources to create another manually classed dataset were not available, as this is a laborious and time-consuming task. Nevertheless, the size and diversity of the training dataset should aid the representativeness of the results, such that a similar outcome might be expected for any radar volume from the network. An assessment against an independent classification using dual-polarization is made at the end of this section.

The assessment against manual classification is made by simply comparing the manual and automatic (NBC generated) classes. The results are tabulated in confusion matrices, with the manual class in rows and the automatic class in columns. The full result of testing against the training dataset is shown in Table 4. Values are converted to percentages, so rows add to 100%.

Note that the percentage of classification is an indication of how well the classifier performed on a class, not how often a class will “contaminate” other classes when the classification is incorrect. For example, if 30% of sea clutter were misclassified as rainfall, this does not mean that rainfall will be contaminated with sea clutter 30% of the time. If sea clutter is only present occasionally, then occasionally some (30%) of the sea clutter may contaminate the precipitation observations. The misclassifications are dominated by chaff, because chaff contributes many pixels to the training dataset but is classified poorly because it is rare and so has a low prior probability.

Some interesting conclusions can be drawn from the results in Table 4. The classifier does not effectively distinguish between precipitation types. Smoke is not well classified, because it has a low prior probability and a large proportion of the training data came from the Melbourne Black Saturday 2009 bushfires, where the height and reflectivity of the smoke plumes were comparable to convective storms. Smoke constrained to the convective boundary layer is more often classified as shallow precipitation or insects. Chaff is also difficult to classify, as it evolves from a high-spatial-variability line to a low-variability cloud as it disperses, so it is challenging to characterize throughout its lifetime. Chaff is mostly misclassified as stratiform precipitation toward

TABLE 4. Results of the classification applied to the training dataset. The manual (man) and automatic (aut) classes are in rows and the classifier output classes are in columns. Values are percentages and the rows add up to 100%. The total number of classed pixels is shown in the Counts column. The diagonal of correct classification is shown in bold.

man\aut	con	shc	str	ins	smk	chf	brd	pe	gc	ap	sl	2tp	Counts
con	55.1	7.7	29.9	1.3	3.7	0.2	0.1	0.6	1.0	0.3	0.2	0.0	2 812 443
shc	3.2	64.7	9.8	9.6	0.7	2.5	0.9	2.3	1.2	2.8	2.1	0.0	497 081
str	13.7	7.8	69.9	1.2	2.4	2.6	0.2	0.7	0.4	0.8	0.4	0.0	7 538 183
ins	0.4	5.4	2.0	70.9	1.7	0.1	5.5	9.9	2.9	0.1	1.0	0.0	8 166 986
smk	14.9	15.4	21.5	13.0	26.3	0.0	2.1	4.7	2.0	0.0	0.0	0.0	631 508
chf	6.1	18.6	27.0	6.6	0.4	29.9	1.6	2.2	0.6	5.2	1.8	0.0	1 187 655
brd	0.0	4.2	0.9	26.5	0.8	0.0	46.6	13.2	5.0	0.0	2.7	0.0	48 210
pe	2.4	6.9	0.3	9.5	0.7	0.0	9.0	64.4	5.9	0.0	0.8	0.0	132 677
gc	8.4	27.8	2.9	20.5	0.9	0.8	7.4	5.4	25.7	0.0	0.0	0.0	367 101
ap	3.1	17.4	20.2	0.0	0.9	11.8	0.0	0.0	0.0	41.5	4.9	0.0	365 141
sl	1.0	6.4	1.1	0.1	2.9	1.3	0.0	0.2	0.0	1.7	85.2	0.0	704 312
2tp	0.6	15.5	1.2	56.9	0.7	0.2	3.5	0.6	1.3	2.4	17.1	0.0	19 176

the end of its lifetime. Birds are most often mistaken for insects or permanent ground clutter because the echo is typically shallow and near the radar. The POD greatly assists recognizing permanent ground clutter and sidelobe sea clutter, though misclassification as insects and precipitation, respectively, are most common. The AP echoes are poorly classified, probably because their low prior probability and high echo top height (given that the path of the beam is not known, so it is assumed to be much higher) mean that AP echo is often mistake for precipitation. It is apparent that the types of echo most difficult for an observer to distinguish from precipitation are also the most poorly classified.

The classification is not expected to be used in such detail as given in Table 4; for example, it is not intended to discriminate types of precipitation, so these classes may be combined. Echo types have been grouped into three “superclasses” in Table 5: precipitation, clear air that may yield useful radial velocity (smoke and insects), and other echo (chaff, birds, and ground/sea clutter). The disuse of chaff for wind estimation is discussed in section 4. The aggregation of classes shows that over 90% of precipitation is identified correctly (Table 5). Clear air echo is reasonably well identified, and 35% of all clutter is identified as precipitation. Much of this clutter is from chaff, which contributed a large proportion of the training data (counts in Table 4). When interpreting the tables of superclasses, it must be remembered that the values are biased by the size of each class’s contribution to the superclass in the training dataset (which does not represent climatology) but inter-comparison between such tables remains useful.

The results from the evaluation against the training dataset are similar to a qualitative assessment of applying the classifier to other radar volumes. Since the NBC was implemented in Ancilla to run in real time, all

radar files are output with classification information. This output has been monitored for months by automated plotting of the reflectivity and class. Two examples are shown, with the raw reflectivity, the classification, the reflectivity of precipitation classes, and the velocity of precipitation and clear air classes. The first example is a difficult case with AP sea clutter (Fig. 3). Insects and sidelobe sea clutter are well identified, but only parts of the AP clutter are identified as clutter (of any variety). The second example (Fig. 4) has showers crossing the Sydney radar. The showers are correctly classified as precipitation, though very small showers where the spatial variability is high are classified as clutter.

The purpose of creating the classification algorithm is to extract useful radar observations for any required application. To be considered successful, it must at least improve on an existing method used by the BoM. Previously, a thresholding algorithm had been used to extract rainfall for quantitative precipitation estimation and nowcasting. This method keeps only echo ≥ 5 dBZ. Echoes are excluded where a comparison between adjacent elevations indicates ground clutter has been removed (a sharp increase in reflectivity with elevation). Echo top heights less than 2 km (for a 5-dBZ threshold) are also excluded as nonprecipitation. This method was

TABLE 5. Collapsed classification results derived from Table 4. Precipitation represents all three precipitation classes (%). Clear air includes insects and smoke. Clutter contains the remaining classes. The diagonal of correct classification is shown in bold.

	Precipitation	Clear air	Clutter	Counts
Precipitation	91.1%	4.2%	4.6%	10 847 707
Clear air	11.0%	70.3%	18.7%	8 798 494
Clutter	34.9%	8.0%	57.2%	2 824 272

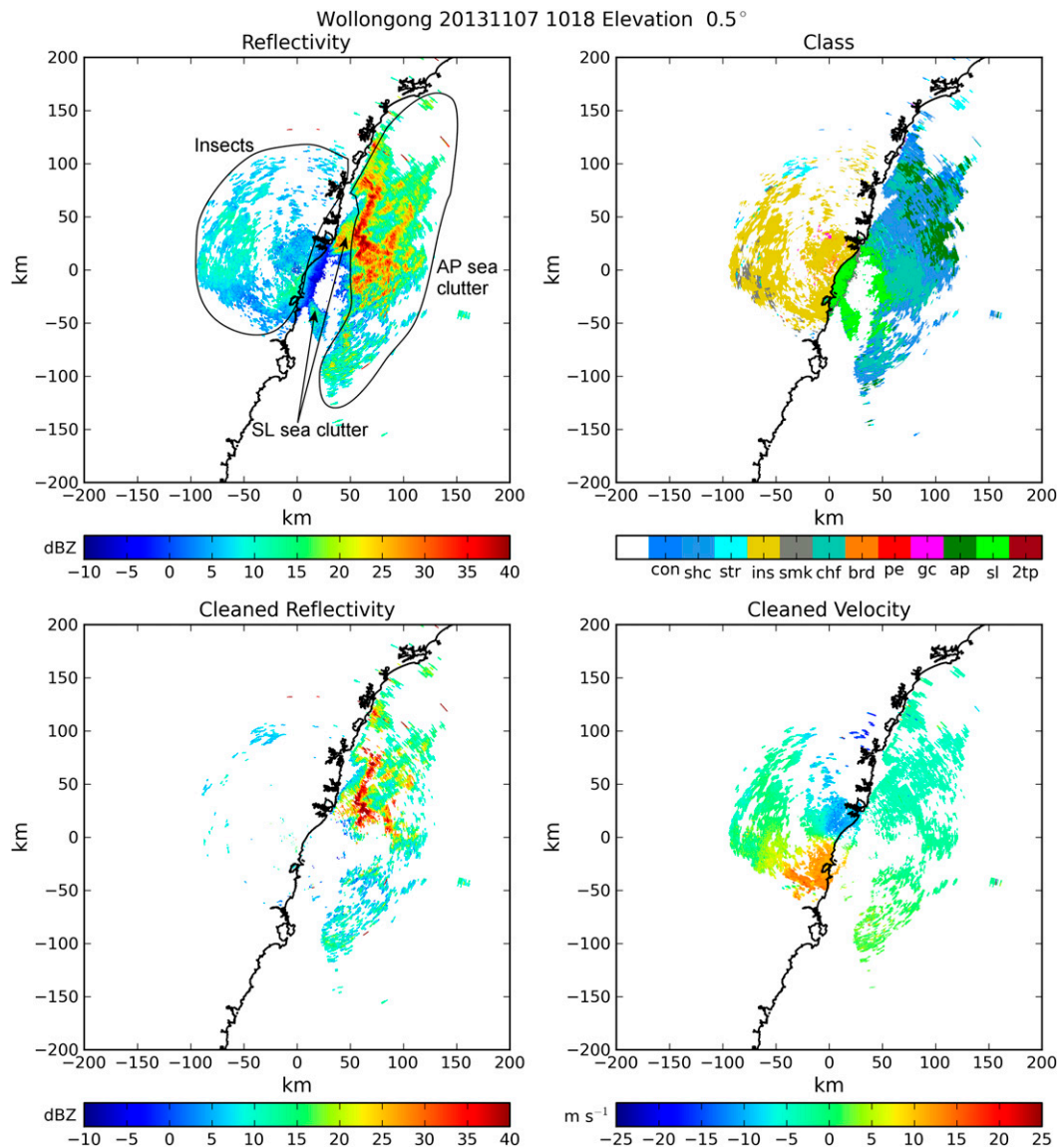


FIG. 3. Example of the classification of a case of AP sea clutter from Wollongong at 1018 UTC 7 Nov 2013. Data are from the 0.5° elevation scan. The reflectivity panel shows the raw radar data, with echo types coarsely labeled. The class panel shows the NBC output. The cleaned reflectivity panel shows the observations that would be selected by applications that use precipitation echo. The cleaned velocity panel shows observations from classes that may provide wind estimation (precipitation, insects, and smoke).

applied to the training dataset and the results were compared (Table 6) with NBC results for precipitation and nonprecipitation classes and echoes ≥ 5 dBZ. All clear air echoes were included as clutter. The results indicate that the NBC is slightly better at detecting precipitation and that it substantially reduces clutter misclassification from 47% to 21%. Overall, excluding echo < 5 dBZ results in a slightly higher proportion of all classes being classified as precipitation; that is, higher reflectivity echo is more likely to be classified as precipitation. On the other hand, 10%–60% of all clutter

echo (depending on type) in the training dataset is < 5 dBZ (e.g., DBZH pdfs in Fig. 2); so, if the 5-dBZ threshold is used, the amount of clutter contamination will be reduced.

The NBC is used for a radar network in which currently half the radars provide radial velocity and only a subset of those provide spectrum width. Therefore, not all radars have VTEX or WAVG available as feature fields for the classifier. The contribution of the Doppler parameters VTEX and WAVG was assessed by running the classifier with and without these parameters, for

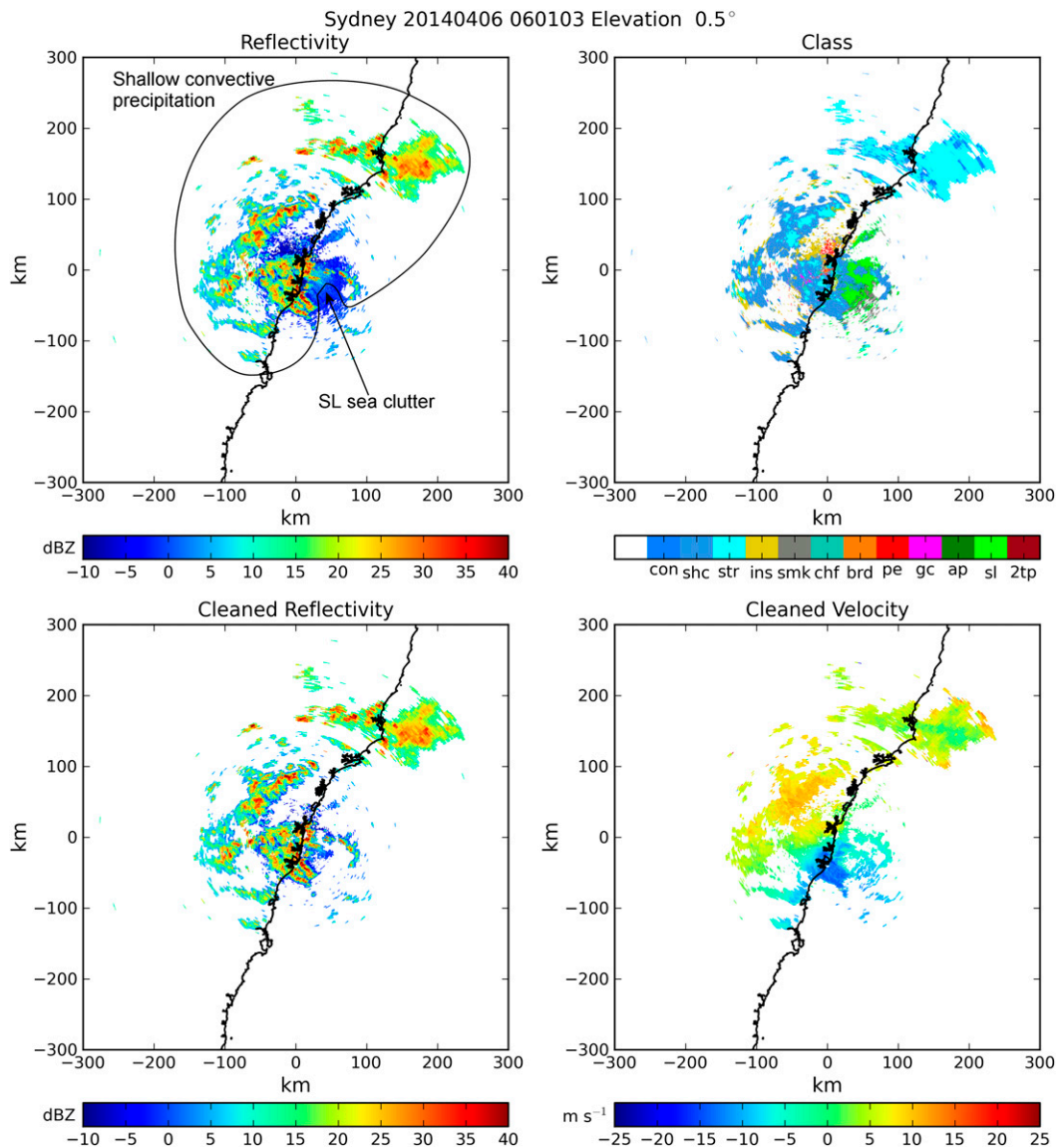


FIG. 4. As in Fig. 3, but for a case of convective showers at Sydney at 0601 UTC 6 Apr 2014. Most of the echo is from precipitation and is correctly identified.

radars that had these parameters available. VTEX was found to have little effect; it improved the identification of clear air by more than 10% but reduced the accuracy of precipitation detection by less than 2% compared with values in Table 5. This is most likely because areas with high VTEX associated with wind shear or dealiasing errors are classified as clutter, not precipitation. It is important not to use VTEX to classify precipitation in tornadoes, for example. The threshold above which VTEX is not used could be lowered if this were a concern; on the other hand, it would not be safe to assimilate the radial velocity from tornadoes, since the BoM’s current NWP cannot resolve that scale. The sample size

of training data with WAVG is small, so results of classification with and without using WAVG are not definitive. It was seen that WAVG improved the classification of some classes, particularly the discrimination between convective and stratiform precipitation, and the identification of AP sea clutter and chaff, in comparison to Table 4. However, WAVG made negligible differences to the accuracy of the superclasses’ classification as per Table 5.

There has been recent effort to develop a dual-polarization (DP) Bayesian algorithm (Wen 2014) using data from the Brisbane Cloud Physics 2 (CP2) C-band research radar (Keenan et al. 2007). This radar,

TABLE 6. Comparison of the Bayesian classification method with an existing threshold-based method, discriminating between precipitation and nonprecipitation for echoes >5 dBZ in the training dataset.

		Bayesian Automatic				Thresholds Automatic	
		Precipitation	Nonprecipitation			Precipitation	Nonprecipitation
Manual	Precipitation	92.5%	7.5%	Manual	Precipitation	90.3%	9.7%
	Nonprecipitation	20.8%	79.2%		Nonprecipitation	46.7%	53.3%

37 km west of the Brisbane radar, is not part of the BoM operational network and does not have equivalent on-site processing applied. Notably, ground clutter filtering is absent and the velocity field is noisier than seen from the BoM operational radars. Several volumes from this radar were classified using DP variables into the following classes: precipitation, biological scatter, ground clutter, sea clutter, and noise. The biological scatter class is equivalent to the insect class. Some light weather echo was classified as noise; this has no equivalent in the present study.

Four cases were examined. The NBC was applied to these four cases using the same schema as for Brisbane; however, WAVG was included and results with and without using VTEX (due to the noisy velocity field) were considered. Ultimately, VTEX was used, relying on its upper threshold to exclude much of the noisy regions. Here the DP classification is treated as “truth.”

The first case (21 November 2008) comprises two volumes with mixed precipitation and strong nocturnal insect echo, one hour apart (1330 and 1430 UTC). Precipitation was classified $>90\%$ correctly. Insect echo was classified predominantly as insects, precipitation, smoke, or birds. Notably, weather over the insect echo resulted in a large ETH, which excluded insects as a possible class, that caused a poor (5%–34%) classification of insects.

The second case (1800 UTC 21 November 2008; Fig. 5) included precipitation, insects, AP sea clutter, and second-trip echo (which DP classified mostly as noise). Correct classifications were precipitation (92%) and insects (34% with 30% as birds and 27% as precipitation). Sea clutter was classified as 24% chaff, 16% insects, and 48% precipitation, and only 6% as sea clutter. This result resembles the example in Fig. 3, where sea clutter is frequently misclassified as chaff

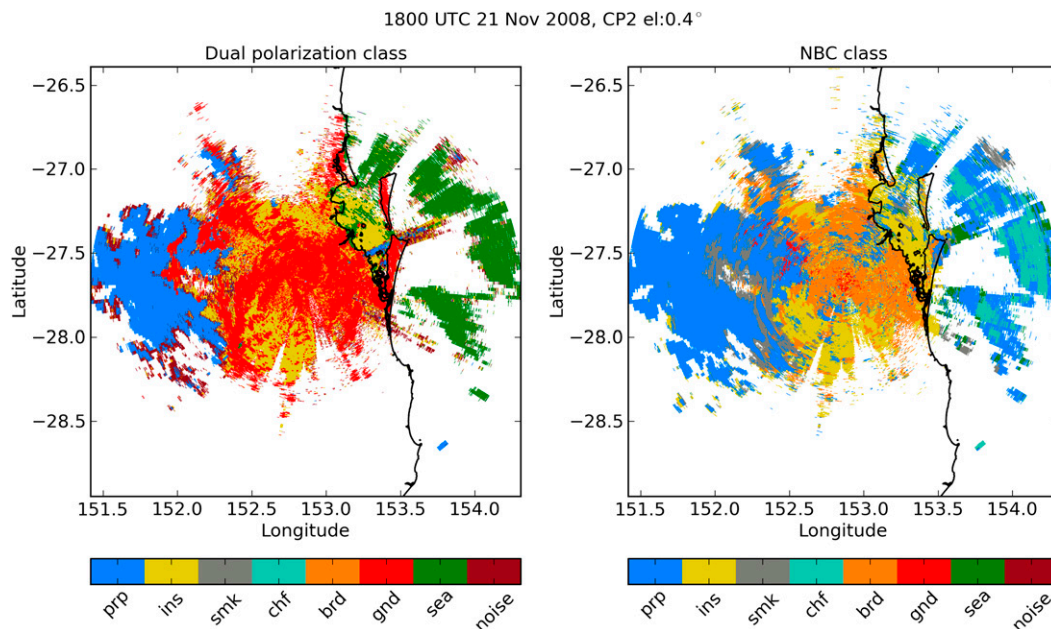


FIG. 5. Comparison of dual-polarization and single-polarization classification for a scan from the CP2 radar, 1800 UTC 21 Nov 2008. The scan contains precipitation (prp), AP sea clutter (sea), insects (ins), and ground clutter (gnd). “Noise” is only classified by the DP method; smoke, chaff, and birds are only classified by the NBC.

nearer to the radar. This appears to be due to the SPIN value, which is bimodal (Fig. 2) and may need further training.

The third case (1006 UTC 22 November 2008) had strong insect echo, of which 72% was classified correctly; the remainder was equally classified as precipitation or smoke and birds.

The fourth case (0530 UTC 15 December 2008) contained weak diurnal insect echo, which was classified largely as birds (44%), insects (27%), or precipitation (20%). This volume was predominantly ground clutter, of which 17% was classified correctly (mostly as AP ground clutter).

In all these cases, the DP classification was thorough in detecting ground clutter, but the NBC classifies it poorly—although much was classified as birds, which is another “clutter” type. This is because the NBC was trained on clutter remnants after filtering, not unfiltered clutter. Therefore, a comparison of ground clutter detection is not appropriate.

Apart from ground clutter, the results show the classification behaves similarly to that of the training dataset and to cases such as Figs. 3 and 4. The large proportion of echo classified as birds is a result of the higher VTEX values from noisier velocity.

4. Application to Doppler radar data assimilation

The NBC is being used to identify radial velocity observations suitable for wind estimation, that is, for data assimilation in the high-resolution NWP models. For the BoM’s developmental ACCESS-City LAMs, only echoes classed as precipitation are assumed to be qualified for assimilation. The observations undergo extensive processing after ingestion in the assimilation system [the details of the quality control options are like those in Simonin et al. (2014)], including observation-minus-background checks, removal of isolated pixels, and comparison with neighbors, before spatial averaging to reduce data density. This means that classification is not the only mechanism for removing unreliable velocity observations. Note that for radial velocity, misclassification at long range may not be a large problem because observations far from the radar are not assimilated due to increasing error contributions (Fabry 2010; Simonin et al. 2014). Currently, a range limit of 100 km is applied for assimilated observations.

Radars from the Australian radar network measure substantial clear air echo, which may also be useful for wind estimation (e.g., Fig. 3). By using the class information, the clear air observations can be assessed, for example, by monitoring observation-minus-background statistics for precipitation and clear air separately. Additionally,

clear air echo could be treated independently in the assimilation and could be assigned a different weight for assimilation. As part of the work toward operational assimilation of radial velocities, an examination of observation-minus-background statistics (S. Rennie 2014, unpublished data) from hourly observation processing (without assimilation) over 40 days was made. The results showed that the differences in statistics for clear air and precipitation are not substantial. However, this study also showed that, despite the sporadic clutter types (chf, gc, ap) being the most misclassified (Table 4), it was the continual contribution of small quantities of permanent echo that were substantial enough to dominate the statistics. Future versions of radar quality control include algorithms to more thoroughly remove permanent echoes prior to applying the NBC, rather than relying on the NBC to handle all echo classification. The observation statistics for a QC version that achieves this satisfactorily will be examined before deciding whether to assimilate insect echo.

Chaff is not considered for wind estimation even though it might be supposed to act as a passive tracer. Observed examples of chaff release in Australia (S. Rennie 2013, unpublished data; see example in Rennie 2012) reveal artifacts that suggest that chaff does not give a good wind estimation of the radar sample volume. At release, chaff may show a sharp velocity gradient across the width of the trail, probably due to its velocity at release and movement in the wake of the plane, which should dissipate fairly quickly. However, in cases of extensive chaff release, after a few hours of dissipation a smooth velocity field is not always observed. Coherent chaff trails in close proximity can show velocity variations between the individual trails (S. Rennie 2013, unpublished data). This is hypothesized to be because the chaff only occupies a part of the radar beam, and so the observed velocity depends on the chaff location within the beam. In contrast, precipitation should yield a mean or modal velocity across the radar beam and yield better horizontal continuity. The presence of wind shear could cause a large variation in the observed velocity depending on the height of the chaff. It is only at the last stages of dispersal that chaff appears suitable as a wind tracer. The NBC is intended to identify chaff soon after its release, and later becomes more likely to classify chaff as stratiform precipitation or other clear air echo.

5. Enhancements to the classifier

One of the most effective ways of improving classification by the NBC was to permit spatial variability of the prior probability. This assisted permanent echo

TABLE 7. The prior probability of the precipitation classes based on the probability of precipitation forecast.

Probability of precipitation (PoP)	Prior probability		
	Convective	Shallow convective	Stratiform
PoP > 0.4	0.4	0.25	0.5
0.1 ≤ PoP ≤ 0.4	0.2	0.12	0.2
PoP < 0.1	0.01	0.05	0

detection and limited the number of possible classes at any location, for example, sea clutter only over water. Given the difficulty in identifying echo types at single-polarization radars, the fewer classes to be discriminated by the classifier, the better the result. In Table 4, the accuracy of classification of the permanent echo types (pe and sl) is very high.

The actual probability of any echo within a scan being of a certain echo type is not expected to be constant over time, though the Bayes classifier functions as if this is the case. Therefore, allowing the modification of prior probabilities over time could substantially improve the results. This feature has not been implemented in Ancilla, but its effect was tested by using the training dataset and modifying the prior probability of precipitation.

For each training data volume, the 0–6-h forecast probability of precipitation (PoP) taken from the BoM operational forecast was estimated for that volume. Depending on the value of PoP, the prior probabilities of the precipitation classes were modified to one of three values, representing “highly likely,” “possible,” and “unlikely,” as per Table 7. The result was that the incidence of clutter being classified as precipitation reduced substantially (compared with Table 5), from 34.5% to 14.8%, and the clear air classified as precipitation decreased from 11.1% to 5.0%; the accuracy of precipitation detection also decreased by about 1%. At this stage no further effort to tune the variable prior probabilities has been made, though the authors’ experience is that results are not usually sensitive to small changes in the prior probabilities. Changes to prior probabilities of up to 0.1 usually changed the classification results by only a few percent.

The provision of other information about expected echo types (e.g., reports of fires or chaff, predictions of AP conditions) could also benefit the NBC’s performance. However, a reliable way to automatically determine this information and to pass it to the classifier needs to be developed.

6. Conclusions

The Australian Bureau of Meteorology requires a quality control system for its weather radar network that

can be used to extract observations for data assimilation in high-resolution NWP. The heterogeneous radar network contains single-polarization instruments, some of which provide Doppler observations. A naïve Bayes classifier has been developed to identify various echo types, using reflectivity, echo top height, textures, and gradients of reflectivity (and velocity) and spectrum width if available. The classifier attempts to identify precipitation and various types of clear air echoes and clutter echoes. The ultimate requirement is to discriminate echoes by usefulness, rather than to accurately identify all echo types, since accurate identification of echo types using a single-polarization classifier is difficult.

A quantitative assessment made by applying the classifier to its training dataset suggests that more than 90% of precipitation is correctly identified (as precipitation of some type). Verification by observation of routinely classified radar data in real time supports the quantitative assessment. Nonprecipitation echo is identified with varying degrees of accuracy. Biological echoes are not usually classified as precipitation. Smoke and chaff are difficult to identify, and are often misclassified as precipitation, especially smoke from very large bushfires. AP clutter is also most often misclassified as precipitation, especially when far from the radar. Permanent echo that is identified with the help of a POD map to modify the prior probability is classified fairly accurately in contrast. Overall, the classifier performs better than the baseline threshold-based method that it will replace.

The classifier output could be improved by tuning the prior probabilities based on external information, such as forecast probability of precipitation. The confirmed presence or absence of any class would reduce the number of classes that the classifier must distinguish. Another option is removing permanent echo prior to classification, since it is relatively easy to locate (as shown by the success of the classifier using the POD map). Methods to remove permanent echo, spikes, and speckle from the radar scans before Bayesian classification have been implemented already in Ancilla (S. Rennie 2014, unpublished data).

Ultimately, the Bayes classifier fulfills the requirement of providing a way to select observations for various applications like data assimilation. The class information is already being used in this way, by selecting radial velocity observations, and is enabling the investigation of whether clear air echo should be used for wind estimation. The classifier could be optimized for other applications, for example, precipitation estimation. Finally, we note that the classifier framework within Ancilla is easily able to be adapted to

dual-polarization radars, in anticipation of future upgrades to the Australian network.

REFERENCES

- Bachmann, S., and D. Zrnić, 2008: Three-dimensional attributes of clear-air scatterers observed with the polarimetric weather radar. *IEEE Geosci. Remote Sensing*, **5**, 231–235, doi:10.1109/LGRS.2008.915745.
- Dance, S. L., 2004: Issues in high resolution limited area data assimilation for quantitative precipitation forecasting. *Physica D*, **196**, 1–27, doi:10.1016/j.physd.2004.05.001.
- Dixon, M., C. Kessinger, and J. Hubbert, 2005: Echo classification and spectral processing for the discrimination of clutter from weather. *32nd Int. Conf. on Radar Meteorology*, Albuquerque, NM, Amer. Meteor. Soc., P4R.6. [Available online at https://ams.confex.com/ams/32Rad11Meso/techprogram/paper_95987.htm.]
- Dokter, A. M., F. Liechti, H. Stark, L. Delobbe, P. Tabary, and I. Holleman, 2011: Bird migration flight altitudes studied by a network of operational weather radars. *J. Roy. Soc. Interface*, **8**, 30–43, doi:10.1098/rsif.2010.0116.
- Doviak, R. J., and D. S. Zrnić, 1993: *Doppler Radar and Weather Observations*. 2nd ed. Academic Press, 562 pp.
- Fabry, F., 2010: Radial velocity measurement simulations: Common errors, approximations, or omissions and their impact on estimation accuracy. *Proc. Sixth European Conf. on Radar in Meteorology and Hydrology*, Sibiu, Romania, ERAD, 7 pp. [Available online at http://www.erad2010.org/pdf/oral/thursday/nwp1/02_ERAD2010_0154.pdf.]
- Friedman, N., D. Geiger, and M. Goldszmidt, 1997: Bayesian network classifiers. *Mach. Learn.*, **29**, 131–163, doi:10.1023/A:1007465528199.
- Hubbert, J. C., M. Dixon, and S. M. Ellis, 2009: Weather radar ground clutter. Part II: Real-time identification and filtering. *J. Atmos. Oceanic Technol.*, **26**, 1181–1197, doi:10.1175/2009JTECHA1160.1.
- Izenmann, A. J., 1991: Recent developments in nonparametric density estimation. *J. Amer. Stat. Assoc.*, **86**, 205–225.
- Keenan, T., J. Wilson, J. Lutz, K. Glasson, and P. T. May, 2007: Rationale and use of the CP2 testbed in Brisbane, Australia. *33rd Conf. on Radar Meteorology*, Cairns, QLD, Australia, Amer. Meteor. Soc., 12B.1. [Available online at https://ams.confex.com/ams/33Radar/techprogram/paper_123253.htm.]
- Kessinger, C., S. Ellis, J. Van Andel, J. Yee, and J. Hubbert, 2005: The AP ground clutter mitigation scheme for the WSR-88D. *20th Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, San Diego, CA, Amer. Meteor. Soc., P1.5. [Available online at <https://ams.confex.com/ams/pdfpapers/86549.pdf>.]
- Koistinen, J., T. Mäkinen, and S. Pulkkinen, 2009: Classification of meteorological and non-meteorological targets with principal component analysis applying conventional polarimetric parameters and their texture. *34th Conf. on Radar Meteorology*, Williamsburg, VA, Amer. Meteor. Soc., 10A.3. [Available online at https://ams.confex.com/ams/34Radar/techprogram/paper_155348.htm.]
- Lakshmanan, V., A. Fritz, T. Smith, K. Hondl, and G. Stumpf, 2007: An automated technique to quality control radar reflectivity data. *J. Appl. Meteor. Climatol.*, **46**, 288–305, doi:10.1175/JAM2460.1.
- , J. Zhang, and K. W. Howard, 2010: A technique to censor biological echoes in radar reflectivity data. *J. Appl. Meteor. Climatol.*, **49**, 453–465, doi:10.1175/2009JAMC2255.1.
- Michelson, D., R. Lewandowski, M. Szczykowski, and H. Beekhuis, 2011: EUMETNET OPERA weather radar information model for implementation with the HDF5 file format. Version 2.1, EUMETNET OPERA Working Doc. WD_2008_03, 36 pp.
- Moszkowicz, S., G. J. Ciach, and W. F. Krajewski, 1994: Statistical detection of anomalous propagation in radar reflectivity patterns. *J. Atmos. Oceanic Technol.*, **11**, 1026–1034, doi:10.1175/1520-0426(1994)011<1026:SDOAPI>2.0.CO;2.
- Peter, J. R., A. Seed, and P. Steinle, 2013: Application of a Bayesian classifier of anomalous propagation to single-polarization radar reflectivity data. *J. Atmos. Oceanic Technol.*, **30**, 1985–2005, doi:10.1175/JTECH-D-12-00082.1.
- Puri, K., and Coauthors, 2010: Preliminary results from Numerical Weather Prediction implementation of ACCESS. *CAWCR Research Letters*, No. 5, CAWCR, Melbourne, VIC, Australia, 15–22.
- Rabier, F., E. Klinker, P. Courtier, and A. Hollingsworth, 1996: Sensitivity of forecast errors to initial conditions. *Quart. J. Roy. Meteor. Soc.*, **122**, 121–150, doi:10.1002/qj.49712252906.
- Rennie, S. J., 2012: Doppler weather radar in Australia. CAWCR Tech. Rep. 55, 52 pp.
- , M. Curtis, J. R. Peter, A. W. Seed, and P. J. Steinle, 2014: Training the Ancilla Naïve Bayes Classification System to provide quality control for weather radar data. CAWCR Tech. Rep. 73, 88 pp.
- Salonen, K., G. Haase, R. Eresmaa, H. Hohti, and H. Järvinen, 2011: Towards the operational use of Doppler radar radial winds in HIRLAM. *Atmos. Res.*, **100**, 190–200, doi:10.1016/j.atmosres.2010.06.004.
- Schuur, T., A. Ryzhkov, and P. Heinselman, 2003: Observations and classification of echoes with the polarimetric WSR-88D radar. NOAA/NSSL Rep., 45 pp.
- Simonin, D., S. P. Ballard, and Z. Li, 2014: Doppler radar radial wind assimilation using an hourly cycling 3D-Var with a 1.5 km resolution version of the Met Office Unified Model for nowcasting. *Quart. J. Roy. Meteor. Soc.*, **140**, 2298–2314, doi:10.1002/qj.2298.
- Steiner, M., and J. A. Smith, 2002: Use of three-dimensional reflectivity structure for automated detection and removal of nonprecipitating echoes in radar data. *J. Atmos. Oceanic Technol.*, **19**, 673–686, doi:10.1175/1520-0426(2002)019<0673:UOTDRS>2.0.CO;2.
- Sun, J., 2005: Convective-scale assimilation of radar data: Progress and challenges. *Quart. J. Roy. Meteor. Soc.*, **131**, 3439–3463, doi:10.1256/qj.05.149.
- Wen, G., 2014: A study on hydrometeor classification with polarimetric variables. Ph.D. thesis, Dept. of Electrical and Electronic Engineering, University of Melbourne, 137 pp.
- Zhao, Q., J. Cook, Q. Xu, and P. R. Harasti, 2008: Improving short-term storm predictions by assimilating both radar radial-wind and reflectivity observations. *Wea. Forecasting*, **23**, 373–391, doi:10.1175/2007WAF2007038.1.