12B.2 DEVELOPMENTS IN AUSTRALIAN DOPPLER RADAR DATA CLASSIFICATION AND ASSIMILATION

Susan J. Rennie*, Justin Peter, Mark Curtis, Alan Seed, and Peter Steinle
Centre for Australian Weather and Climate Research, Australian Bureau of Meteorology,
Docklands, Vic., Australia

1 INTRODUCTION

The Australian Bureau of Meteorology has upgraded the Doppler capability in a number of its operational radars in recent years. With this new observation source, the assimilation of radial velocity is developing in conjunction with kilometer-scale numerical weather prediction. Radar provides one of the few observation types that can yield good coverage at these resolutions. However, quantitative applications like data assimilation require good quality control.

Australian radars detect a range of echo types, including precipitation, aerofauna, chaff, smoke, ground clutter and sea clutter. Some of these may be used for wind estimation, if the bias is sufficiently small or quantified. Precipitation use is well established for wind estimation, though it may be biased by the fall velocity of the hydrometeors. Insect echo may be used for wind estimation in data assimilation if it is demonstrated that the independent flight is negligible compared to the observation error (Rennie et al. 2010; Rennie et al. 2011). Smoke, though rare, can produce substantial echo from bush fires. The most extreme fires have a reflectivity and echo height comparable to precipitation. Chaff when ejected from an aircraft does not reflect the wind velocity, so should be avoided. Echo from the ground or sea (either directly or anomalous propagation (anaprop or AP)) or from birds or bats cannot be used for wind estimation. These echo types are observed at various radars depending on their geography and climate.

The objective for developing this classification algorithm is to use the available information to identify echo types, to the stage where observations of radial wind can be reliably selected for data assimilation. The simplest methods involve simple thresholds to remove unwanted echo (e.g. low reflectivity, echo top height). More advanced methods combine dual-polarization parameters with fuzzy logic or Bayesian classification. The first is inadequate and the second is not possible with the Bureau of Meteorology's current radar network. The approach used here is to develop a Bayesian-based method using all available information.

The classification scheme is a Naïve Bayes Classifier (Peter et al. 2013) intended for both Doppler and non-Doppler radars. Using knowledge of the geography, observation of the occurrence of echo types, and probability of detection (POD) maps, the classification scheme applies a logical refinement of class prior probabilities to the Naïve Bayes Classifier. The minimum objective is to classify observations as okay to assimilate, possible to assimilate, and to be excluded. The results must also improve on an existing method of thresholding to separate precipitation from non-precipitation. This paper focuses on the classification method applied to the Doppler radars only.

2 AUSTRALIA'S DOPPLER RADARS

Australia has 15 Doppler radars in its collection of over 60 radars. These are located in capital cities and areas where severe weather is common (Figure 1). Geographically, they are in tropical and temperate regions, and most are coastal. They are a mixture of C-band and S-band, and 1° or 2° beam width. Each produces 14 PPI scans between 0.5° and 32°, every 6 or 10 minutes, with 250 m or 500 m range resolution and 1° angle resolution. Each ray alternates PRF, for on-site dealiasing that yields a Nyquist velocity between ~13 m s^{-1} and 52 m s^{-1}, depending on PRF choice. The signal undergoes on-site processing including noise-correction, SQI thresholding and zero-velocity filtering, before the corrected reflectivity and Doppler velocity are sent off-site. Some Doppler radars also return uncorrected reflectivity and spectrum width, which provides further information that may be used for classification.

Permanent ground echo is a problem because there are 'holes' in the scan where permanent echo has been consistently removed by the zero-velocity filter, but surrounding these areas a returned signal is relatively frequent. Side-lobe sea clutter also occurs frequently in some regions. Therefore a conventional POD map is of some use, provided that often-present clear-air echo is weaker than the ground echo. The zero-isodop filtering also removes some precipitation echo which can adversely affect classification.

* *Corresponding author address:* Susan Rennie, Centre for Australian Weather and Climate Research, Australian Bureau of Meteorology, GPO Box 1289, Melbourne VIC 3001 Australia; email: s.rennie@bom.gov.au

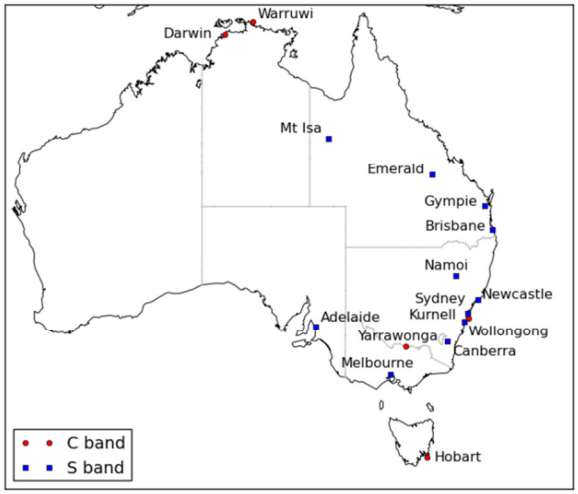

Figure 1 Location of the Australian Doppler radars. Red circles are C-band radars, blue squares are S-band radars.

3 THE CLASSIFICATION SCHEME

The classification scheme utilizes a range of feature fields, including texture (Hubbert et al. 2009; Kessinger et al. 2004) and spin (Steiner; Smith 2002) of reflectivity and velocity, spectrum width, echo top height, reflectivity, vertical gradient of reflectivity. The feature fields used are detailed in Table 1. Each feature field required a pdf describing the range of values for each class. To create this a manually classified data set was created using over 190 radar volumes across a selection of radars (from the whole network). Where need was apparent, individual pdfs were created for subsets of radars with similar operating characteristics. The values for each class were aggregated and normalized histograms produced. A range of pdfs were empirically fit to the histogram and the optimal fit was chosen (Figure 2). The pdfs (including composite pdfs) are triangular, trapezoid, normal, inverse normal, log normal, skew normal, truncated normal, gamma, inverse gamma, Laplace, composite of Laplace and normal (symmetric about y axis), composite of two Laplaces, composite of Laplace and skew normal, and composite of two log normals (log binormal).

The classes manually classified (as per Table 2) include precipitation (stratiform, convective and shallow convective), chaff, insects, birds and bats (dusk/dawn dispersal), permanent ground clutter, anaprop ground clutter, side-lobe sea clutter, anaprop sea clutter, smoke and second-trip echo. The second trip echo was ultimately ignored for being rare and not clearly distinguishable. Smoke may also be ignored, e.g. in winter in the temperate regions of Australia. Note that stratiform and convective precipitation aren't expected to be discriminated in this classification, but were given separate classes to help characterize the range of the feature field values they can encapsulate, and to better balance the prior probabilities to frequency of occurrence. It was also considered that values in the tails of distributions may not be reliable for classification, so the pdfs are truncated in some cases and the feature field not used if the value is too extreme. For example, reflectivity below −10 dBZ or

Table 1. Feature fields used. Smoothing is done with a 3x3 Gaussian kernel.

Field	Function of	Description
DBZH	reflectivity	reflectivity
EHGT	smoothed reflectivity	echo top height to 4 dB
EHGT2	smoothed reflectivity	echo top height to −5 dB, where EHGT does not exist
WAVG	spectrum width	spectrum width from weighted average using adjacent beams
VTDL	smoothed reflectivity	vertical gradient of reflectivity
ZTEX	reflectivity	variation of reflectivity in 2D kernel of 11×11
VTEX	dealiased radial velocity	variation of velocity in 2D kernel of 15×15
SPIN	reflectivity	change in sign of reflectivity gradient in 2D kernel of 19×19

Table 2. Classes, their abbreviations, the number of points used for pdfs, the number of volumes, days and radars from which examples were drawn.

Classes	Abbrev.	Num. pixels	Num. volumes	Num. days	Num. radars
Convective precipitation	con	2853238	31	26	11
Stratiform precipitation	str	7638167	26	22	9
Shallow convection	shc	503052	9	10	7
Chaff	chf	193622	34	7	4
Insects	ins	8399359	61	29	12
Birds/bats	brd	49785	22	10	5
Permanent ground clutter	pe	145173	53	30	11
AP ground clutter	gc	373813	21	12	6
Side-lobe sea clutter	sl	735582	64	30	7
AP sea clutter	ap	366761	20	9	5
2nd trip echo	2tp	19417	5	2	1
Smoke	smk	645165	32	9	6

above 50 dBZ is not used. Echo top height above 12 km is ignored. Classification is made with whichever features are available.

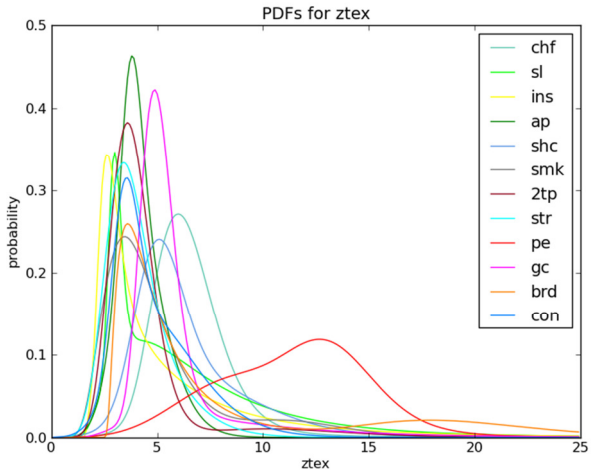

Figure 2 PDFs for ZTEX, from empirical fits of various functions.

The prior probabilities were next determined. As we abandoned the assumption of equal prior probabilities for all classes, they are treated as weights and selected empirically. The prior probabilities can be spatially variable within the scan. Land-sea masks were created, and the distance from the coast calculated (positive over land and negative over water). The prior probability of sea clutter over land is zero, and conversely ground clutter over sea, with overlap at the coastline. Aerofauna echo has zero probability offshore. Chaff is only seen at a few radars, and is typically released over the ocean, or over little-inhabited areas; its probability is zero at radars where it has not been observed. Recurring ground and sea echo is identified with POD and the prior probability is based on this. For ground clutter, if POD>0.2, then its probability is 0.1, else 0. Side-lobe sea clutter was given a minimum probability of 0.05 everywhere offshore, plus its POD. Precipitation can occur anywhere, as can smoke (blown offshore). Typical prior probabilities are listed, where the typical POD may be 0.4 in affected regions:

con:	0.4	str:	0.5	shc:	0.25		
ins:	0.4	smk:	0.1	chf:	0.05	brd:	0.2
gc:	0.1	pe:	POD	ap:	0.1	sl:	POD

4 RESULTS

The classification scheme is tested by application to the manually-classified data set. Results are shown in the contingency tables, which hereafter have the manual classes in rows and generated classes in columns, so the diagonal represents an accurate result. Table 3 shows often better than 50% success at detecting the correct class (when combining precipitation classes). Note that 2nd trip echo has been discarded from the generated classes. Collapsing the classes into three levels of desirability (Table 4), which is how the classification would ultimately be used, shows nearly 90% of precipitation is correctly identified. Important to note is the false alarm rate (FAR) of non-precipitation being identified as precipitation (italicized in Table 4). A low FAR is preferable for data assimilation. Since the representation of each class in the training data set is not in proportion to its occurrence in reality, this does not indicate the frequency with which precipitation would be affected by contamination. However, it does indicate the weaknesses. For example, any anaprop sea clutter may cause some contamination.

At present, resources don't permit creation of an independent manually-classified data set to test this against. The success of the scheme is assessed by monitoring the classification results. Some random examples are shown in Figure 3, Figure 4, and Figure 5, which demonstrate varying success at identifying a range of classes. Generally these are considered to be successful.

Table 3. Results of classing the manually trained dataset. Bold indicates a correct identification, and italics indicate a major incorrect identification.

		Automatic classification											
Manual classification	%	con	shc	str	ins	smk	chf	brd	pe	gc	ap	sl	2tp
	con	**41.9**	**6.8**	**34.7**	1.1	9.0	0.1	0.3	1.0	4.3	0.6	0.2	0.0
	shc	**4.1**	**48.3**	**10.9**	6.1	4.2	0.3	9.1	3.8	4.8	6.8	1.6	0.0
	str	**29.4**	**8.1**	**54.7**	1.0	2.7	0.2	0.6	0.8	1.0	1.0	0.4	0.0
	ins	0.8	5.7	1.5	**68.0**	2.0	0.2	7.2	*10.2*	3.5	0.1	0.9	0.0
	smk	*38.1*	*12.0*	9.4	*11.3*	**14.6**	0.0	3.2	5.2	6.3	0.0	0.0	0.0
	chf	4.5	*22.1*	*29.4*	9.8	1.2	**20.0**	3.3	0.3	0.1	7.2	2.1	0.0
	brd	0.0	3.7	0.4	*21.7*	0.2	0.0	**49.1**	*15.6*	5.8	0.0	3.5	0.0
	pe	1.1	7.9	0.9	9.5	0.8	0.0	6.8	**69.5**	2.8	0.0	0.7	0.0
	gc	2.4	*28.6*	4.8	*19.5*	3.4	0.0	7.9	6.4	**26.8**	0.0	0.1	0.0
	ap	5.6	*20.5*	*14.5*	0.0	1.4	6.1	0.0	0.0	0.0	**47.0**	4.9	0.0
	sl	0.5	7.1	1.7	0.2	4.8	2.2	0.0	0.2	0.1	1.9	**81.4**	0.0
	2tp	1.0	*17.1*	0.7	*56.2*	0.6	0.2	3.7	0.6	1.2	2.9	*15.8*	0.0

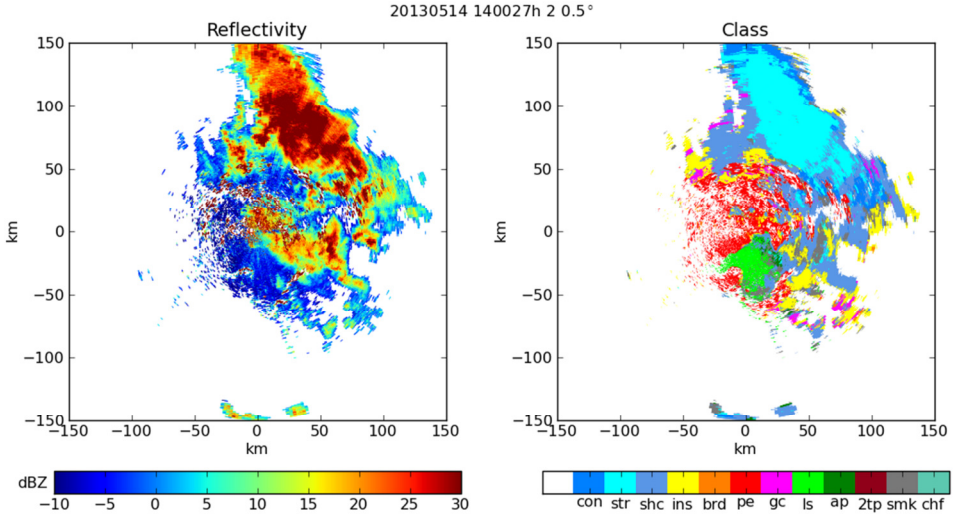

Figure 3 Melbourne, 14 May 2013 at 1400 UTC. Classification of precipitation (blues) and clutter (red and green).

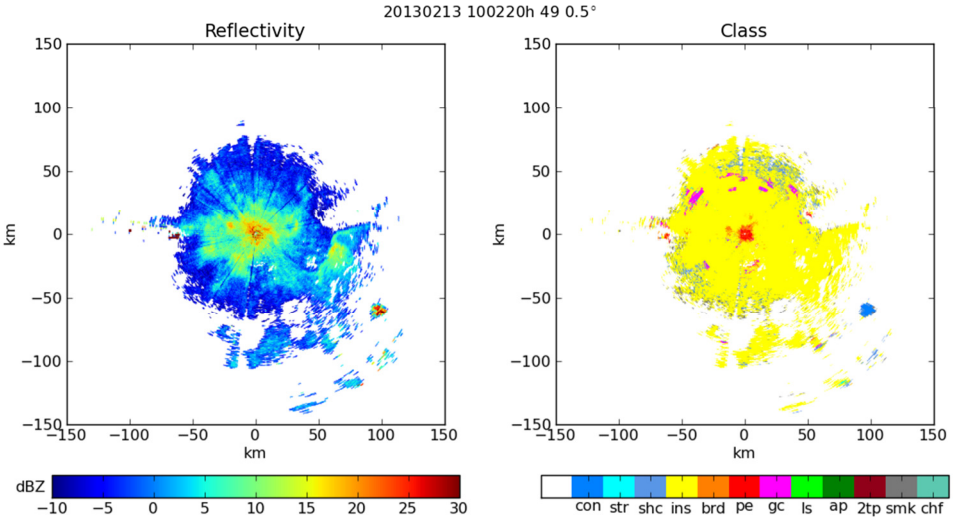

Figure 4 Yarrawonga, 13 February 2013 at 1000 UTC. Classification of nocturnal insects and isolated convective shower.

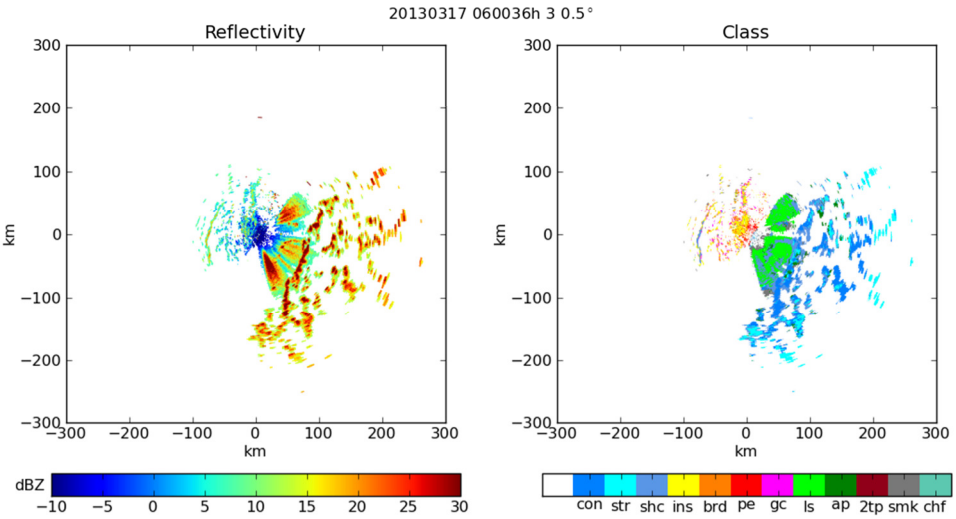

Figure 5 Wollongong, 17 March 2013, 0600 UTC. Classification of convective showers and side-lobe sea clutter, some showers embedded within.

Table 4. Classification results from Table 3 collapsed into three classes. Percentage and total pixels (sum of rows) are shown.

Manual classification	Automatic classification				
	%	precip	clear air	clutter	Totals
	precip	**88.6**	5.7	5.7	10854876
	clear air	*11.6*	**66.8**	*21.5*	8798475
	clutter	*25.9*	9.9	**64.3**	1818974

The minimum requirement for this method is to improve on an existing algorithm used to identify precipitation for QPE/QPF applications at the Bureau of Meteorology. This algorithm uses thresholds of vertical gradients of reflectivity and echo top height to decide whether a pixel is clutter or not. Reflectivity less than 5 dBZ is also discarded (marked as clutter). Applying this method to the training data set yields the results shown in Table 5, with the Bayesian method's results prepared for comparison in Table 6. The most valid comparison is by considering only observations with reflectivity greater than 5 dBZ since that is how the observations are used for QPE.

Table 5. Accuracy of existing method to discriminate precipitation and clutter, applied to training dataset. 'All data' means including anything less than 5 dB as clutter.

	All data			DBZH > =5 dBZ		
%	pr.	cl.	Totals	pr.	cl.	Totals
precip	**79.2**	20.8	10968684	**90.3**	9.7	9623226
clutter	22.5	**77.5**	10902053	41.1	**58.9**	5978132

Table 6. Data from Table 3 collapsed to two classes for comparison with Table 5.

	All data			DBZH > =5 dBZ		
%	pr.	cl.	Totals	pr.	ccl.	Totals
precip	**88.6**	11.4	10854876	**91.1**	8.9	9566022
clutter	14.1	**85.9**	10617449	17.4	**82.6**	5978132

The weaknesses of the old method are manifest where clutter of greater than 5 dB (primarily smoke, chaff, strong insect echo, and anomalous propagation sea clutter) is poorly detected. The new method is better at detecting these echo types, although they are still the most difficult to classify and so cause false alarms. The ability to detect precipitation is similar for the two methods, however.

There is potential for improving the results with other techniques. For example, ensuring continuity of classes where classification is certain. Simple methods to convert localized wrong classifications to the surrounding correct classification could improve the appearance of the results, provided that the 'correct' classification can be selected. Alternatively, use of external information about precipitation, for example forecasts of probability of precipitation (PoP), could also reduce the false alarm rate (Table 7). In this example the following algorithm was applied. For PoP>0.4, use normal prior probabilities, for PoP<0.1 use priors of 0.02, otherwise use half the normal prior probabilities. The detection of precipitation was reduced, but this may be improved by tuning the algorithm.

Table 7. Results using probability of precipitation to limit the prior probability of precipitation in the Bayes classifier.

	precip	clutter	Totals
precip	**86.9%**	13.1%	10854876
clutter	5.9%	**94.1%**	10617449

5 APPLICATION TO ASSIMILATION

The Bureau of Meteorology is running trials assimilating radial wind observations from precipitation. The NWP system consists of a 1.5 km resolution domain (variable grid) over the Sydney area, nested in the Australian Community Climate and

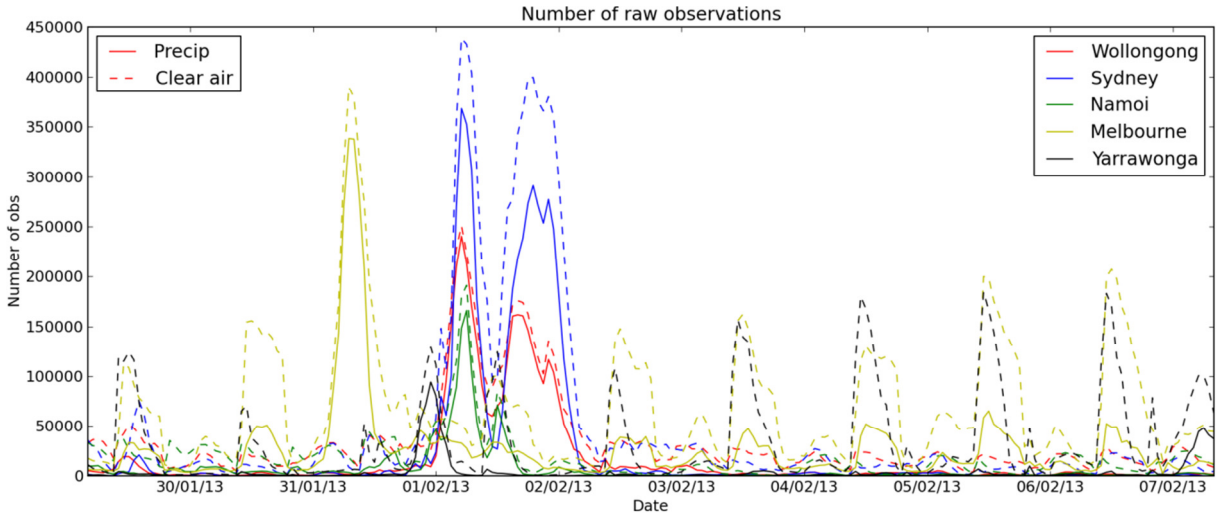

Figure 6 The numbers of precipitation- and clear-air-derived radar observation during a summer period.

Earth System Simulator (ACCESS) regional domain (Puri et al., 2010) at 12 km horizontal resolution. Using the output of the classification scheme, echo only from precipitation is selected for assimilation. Radial wind observations are assimilated hourly using 3D-Var. Before assimilation the raw observations undergo spatial averaging and thinning to reduce the observation density to separation of at least 4 grid spaces. This reduces the number of observations to about 1% of the original number, depending on coverage.

Australian Doppler radars retrieve substantial clear air echo from insects during summer, which may provide a source of observations during dry periods (Figure 6). By classifying these echoes, they can be selectively input into the assimilation system. The initial objective is to produce observation minus background statistics for clear air echo along with precipitation echo. This can be used to investigate the observation error and bias of such observations, and so determine the relative suitability for assimilation. Results to date (Rennie 2013) examining a small number of examples—including when there is strong common orientation behaviour exhibited by insect migrants—have not indicated that the insect flight bias rules out the use of clear air echo for wind estimation. Future trials will be made to investigate the viability of assimilating clear air echo, and the impact on the forecast. By maintaining the class information, insect echo and precipitation echo can be treated independently by the assimilation system.

6 REFERENCES

Hubbert, J. C., M. Dixon, and S. M. Ellis, 2009: Weather radar ground clutter. Part II: Real-time identification and filtering. *J. Atmos. Oceanic. Tech.*, **26,** 1181-1197.

Kessinger, C., S. Ellis, J. Van Andel, J. Yee, and J. Hubbert, 2004: The AP ground clutter mitigatio scheme for the WSR-88D. *20th International Conference on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*.

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. Tech.*, **in press**.

Rennie, S. J., 2013: Common orientation and layering of migrating insects in south-eastern Australia observed with a Doppler weather radar. *Meteorol. Appl.*, **in press**.

Rennie, S. J., A. J. Illingworth, S. L. Dance, and S. P. Ballard, 2010: The accuracy of Doppler radar wind retrievals using insects as targets. *Meteorol. Appl.*, **17,** 419-432.

Rennie, S. J., S. L. Dance, A. J. Illingworth, S. Ballard, and D. Simonin, 2011: 3D-Var assimilation of insect-derived Doppler radar radial winds in convective cases using a high resolution model. *Mon. Wea. Rev.*, **139,** 1148-1163.

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. Tech.*, **19,** 673-686.