A Cluster-Based Method for Hydrometeor Classification Using Polarimetric Variables. Part II: Classification

GUANG WEN
LACS, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

ALAIN PROTAT AND PETER T. MAY
Centre for Australian Weather and Climate Research, Bureau of Meteorology, Melbourne, Victoria, Australia

WILLIAM MORAN
University of Melbourne, Melbourne, Victoria, Australia

MICHAEL DIXON
National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 10 April 2014, in final form 11 August 2015)

ABSTRACT

Two new algorithms for hydrometeor classification using polarimetric radar observations are developed based on prototypes derived by applying clustering techniques (Part I of this two-part paper). Each prototype is defined as a probability distribution of the polarimetric variables and ambient temperature corresponding to a hydrometeor type. The first algorithm is a maximum prototype likelihood classifier that uses all prototypes attributed to the different hydrometeor types in Part I. The hydrometeor type is assigned as the prototype with the highest likelihood when comparing the polarimetric variables and temperature with each prototype. The second algorithm is a Bayesian classifier that uses the probability density functions (PDFs) as derived from the prototype set associated with the identical hydrometeor type. The posteriori probability in the Bayesian method is calculated from a combination of the PDFs and the prior probability, the maximum of which corresponds to the most likely hydrometeor type. The respective merits of the two techniques are discussed. The two classifiers are applied to CP-2 S-band radar observations of two hailstorms that occurred between 16 and 20 November 2008, including the so-called Gap storm, which produced a devastating microburst and large hail at the ground. Results from the classifiers are compared with those derived using the well-established National Center for Atmospheric Research fuzzy logic classifier. In general, good agreement is found, yielding overall confidence in the robustness of the new classifiers. However, large differences are found for the melting ice and ice crystal categories, which will need to be studied further.

1. Introduction

Polarimetric weather radar is effective for the discrimination of different types of hydrometeors because the polarimetric radar measurements are sensitive to habits, shapes, sizes, and other microphysical properties of the particles. The accurate determination of hydrometeor types is crucial for many weather applications, including quantitative rainfall-rate estimation (Giangrande and Ryzhkov 2008; Cifelli et al. 2011), cross validation between spaceborne radars and ground-based polarimetric radars (Chandrasekar et al. 2008), cloud microphysics research (Vivekanandan et al. 1999), and detection of hail in severe weather (Zrnić et al. 2001; Ryzhkov et al. 2013).

Most of the hydrometeor classification algorithms in the literature are implemented on the basis of a semi-empirical rule-based fuzzy logic classification method, using predefined and type-dependent classification boundaries of polarimetric variables (Straka et al. 2000). In this approach, the boundaries are adjusted by means...
of membership functions designed for each hydrometeor type, sometimes overlapping for different types. The classification algorithms have been improved in the last few decades. Zrnić et al. (2001) comprehensively analyzed the usefulness of polarimetric variables in the fuzzy logic classification by comparing the classification performance achieved using a limited number of variables with that achieved using the full set of variables. They also suggested including environmental temperature as a constraint for the discrimination of hydrometeors. Liu and Chandrasekar (2000) developed a procedure for the adjustment of the parameters in membership functions and associated weights by means of neural network feedback. Schuur et al. (2003) developed three types of classifiers—namely, a winter classifier, a summer classifier, and a clutter classifier—by using different sets of polarimetric variables for each purpose. Lim et al. (2005) balanced the metrics of false negative and false positive classification errors by combining product and additive inferences, and started using the freezing level and weighting factors as classification criteria. Park et al. (2009) modified and refined previous classification algorithms by introducing a confidence factor, matrix weights, a melting-layer designation system, and vertical continuity checks. Marzano et al. (2007) and Dolan and Rutledge (2009) developed membership functions based on model simulations using T-matrix scattering calculations. In addition to the above-mentioned fuzzy logic methods, some researchers applied statistical decision theory to exploit the probability density functions (PDFs) of polarimetric variables associated with the different hydrometeor types (Marzano et al. 2008; Li 2011). This class of study still heavily relies on T-matrix backscattering simulations of hydrometeor microphysics.

While physically consistent results have clearly been obtained using the above-mentioned classification methodologies, we believe that there is room for improvement in the robust classification of hydrometeor types using polarimetric variables because of the following:

1) Algorithms in the literature generally consider the polarimetric variables and environmental temperature singly or in pairs, but they may not be completely independent.

2) Algorithms in the literature apply predetermined PDFs, for example, modified beta functions and trapezoidal functions. The preassumed shape of these PDFs may introduce errors.

3) The existing hydrometeor classification algorithms do not naturally take into account the convolution of the true PDFs with specific radar noise characteristics. Different levels of radar noise and other calibration errors could lead to errors in the assumed PDFs, and subsequently affect the classification.

In this two-part paper, we describe a new method to overcome some of these issues. In a previous paper (Wen et al. 2015, hereafter Part I), the cluster analysis is used to objectively derive the PDFs of the polarimetric variables for different hydrometeor types identified using our current understanding of the polarimetric signatures of these hydrometeor types. First, several individual clusters are derived from a series of volumetric scans. These are then validated and labeled in order to build prototypes, which are the distributions of these clusters. Each prototype (corresponding to a hydrometeor type) is represented by a Gaussian distribution associated with a weight, and the total PDF for each hydrometeor type is subsequently regarded as a Gaussian mixture composed of a set of these prototypes.

Under the framework of the cluster-based method described in Part I, two classification algorithms—namely, a prototype-level maximum prototype likelihood classifier (MPLC) and a distribution-level Bayesian classifier—are developed for hydrometeor-type classification using polarimetric variables. As a natural extension to the prototype approach taken in Part I, we propose a MPLC in which the similarity between data and each prototype is evaluated in terms of a likelihood function in a high-dimensional feature space constructed by the polarimetric variables and temperature. Since our method is incremental—that is, we get additional prototypes as new data becomes available—the number of prototypes grows. When the set of prototypes is deemed large enough, it can be used to approximate the PDFs of polarimetric variables corresponding to each hydrometeor type. Such work has been described in detail in Part I. A Gaussian mixture was adopted to model the PDFs of polarimetric variables and temperature for a given hydrometeor type with minimal bias. The Gaussian mixture should have better performance than a parameterization of the PDFs with a single shape, such as a multivariate Gaussian function and other shapes used in statistical classifiers. These PDFs are then used in a Bayesian classifier.

As described in Part I, the cluster-based method has some advantages:

1) The identification of different hydrometeor types comes directly from the clustering technique rather than ad hoc assumptions. The PDFs are obtained from the real radar data and computed by considering all the variables together.

2) It is an incremental method, in the sense that more prototypes can be added as more radar data are processed. The accuracy of the classification can be
improved by generating more prototypes as more data becomes available.

3) Differences in radar noise characteristics are naturally included in the derivation of the PDFs for different radar types or locations.

In section 2, we describe the characteristics of the CP-2 radar and the measurements used in this study. In addition, a summary of the clustering-derived prototypes from Part I is provided. The MPLC and Bayesian classifiers are introduced in section 3. In section 4, the classifiers are used to process the polarimetric observations of two severe hailstorms over the Brisbane area in Queensland, Australia. The classification results are then benchmarked against the National Center for Atmospheric Research (NCAR) fuzzy logic classifier (Vivekanandan et al. 1999). Finally, we provide a summary in section 5.

2. Radar measurements and cluster-derived prototypes

This study uses radar observations collected by the CP-2 S-band dual-polarization radar, which is located near Brisbane. This radar can scan in both PPI and RHI modes. The scanning strategy used for this study is a series of 9- or 10-tilt volumetric PPI scans (typically scanning at 0.5°, 1.0°, 1.5°, 2.0°, 3.0°, 4.0°, 6.0°, 9.0°, 12.0°, and 17.0°), with several vertically scanned RHIs every half hour. The temporal resolution for PPI scanning is 6 min, the azimuthal resolution is 1°, the radial resolution is 150 m, and the maximum range is 142 km. For RHI scanning, the range resolution is the same, the elevation resolution is 0.25°, and the maximum elevation is 38°. The CP-2 radar characteristics are detailed in Table 1.

The four radar variables utilized for this study are as follows: 1) reflectivity factor at horizontal polarization (Z_h), 2) differential reflectivity (Z_{dr}), 3) cross-correlation coefficient between horizontally and vertically polarized radar signals at zero lag (ρ_{hv}), and 4) specific differential phase (K_{dp}). Term K_{dp} is estimated from the measured differential phase shift (φ_{dp}) over 14 successive range gates (2 km) using a linear regression–based method (Bringi and Chandrasekar 2001). In addition, the environmental temperature (T_{env}), usually obtained from a nearby operational radiosonde or potentially from numerical weather prediction, is also used. The vertical profile of temperature is an important parameter because it allows for a simple and powerful discrimination between ice-phase and liquid-phase hydrometeors. Note the temperature used in this study is a single profile collected about 20 km away from the radar, up to 5 h before or after the radar observations.

Therefore, this information should be used with caution, as discussed in Part I.

A flowchart of the entire hydrometer classification system is illustrated in Fig. 1. The polarimetric data are postprocessed by a quality control procedure proposed by V. N. Bringi and M. Thurai (2008, personal communication). This eliminates the radar echoes from nonmeteorological scatterers, including anomalous propagation, ground clutter, birds, insects, etc. This postprocessing step also utilizes variables not used by the classifiers, such as textures of Z_h, Z_{dr}, and ϕ_{dp}. The classification system consists of two major components: the prototype generation unit and the hydrometeor classification unit. As presented in detail in Part I and outlined here, the prototype generation unit produces a set of prototypes for each of the hydrometeor types based on cluster analysis. A two-step clustering procedure with K-means and expectation–maximization (EM) clustering is iteratively operated with prior information obtained by searching peaks and valleys in data histograms to generate well-behaved clusters. These clusters are associated with a class type and promoted as prototypes using the following procedures: 1) identify the locations of the clusters in a storm and 2) compare their statistics to well-established classification boundaries and thresholds in the feature space. In addition, the clusters are not modified by the comparison with classification boundaries. A renormalization process is introduced to organize prototypes derived from multiple radar volumetric scans. A Kullback–Leibler divergence approach is implemented to reduce the number of prototypes while retaining the PDFs derived from all of the original prototypes. The PDFs with the reduced number of prototypes should produce consistent classification results as the original PDFs for any given dataset.

Using this technique, we obtained 449 prototypes from 25 volumetric PPI scans between 0618 and
0642 UTC 16 November 2008, 0200 and 0254 UTC 19 November 2008, and 0654 and 0748 UTC 24 March 2013. These include several instances of hail and heavy rain. It was important to include cases where hail was reported to make sure that hail and a rain/hail mixture, which are critical for severe weather warning, were correctly handled.

The labeled prototypes are illustrated in Fig. 2 with a color corresponding to each hydrometeor type. The hydrometeors found by our cluster-based approach match the hydrometeor categories considered in the NCAR fuzzy logic method, except for the graupel and rain mixture. Our results in Part I indeed indicate that the clustering technique could not identify a rain/graupel mixture class specifically, despite the fact that cases with large areas of rain/graupel mixture identified by the NCAR classifier were included in our training dataset (more details will be given in section 4). This suggests that there is no sufficient information content in the four polarimetric variables used in our method to allow for accurate discrimination between rain and a rain/graupel mixture. The large overlap of membership functions in fuzzy logic techniques for these two hydrometeor types (Vivekanandan et al. 1999) seems to confirm this point as well.

Following the naming convention of the NCAR classifier given in Table 2, these 449 prototypes were distributed as follows: 70 prototypes for light rain, 55 for moderate rain, 37 for heavy rain, 110 for dry snow, 23 for wet snow, 106 for ice crystals, 20 for graupel/small hail, 16 for rain/hail mixture, and 12 for hail. Overall, Fig. 2 shows that the discriminating power of the $Z_h-Z_{dr}$ pair is particularly high, with little overlap overall between the prototypes corresponding to the different hydrometeor types, especially for hail (H) and the rain/hail mixture (R/H), the different types of ice hydrometeors, and the different rain intensities. Also, when there is large overlap in the $Z_h-Z_{dr}$ two-dimensional space, such as between ice aggregates (IA) and melting ice (MI) or between moderate rain (Rm) and graupel/small hail (G/Hs), another input variable provides the way to discriminate between these hydrometeor types (temperature for these specific cases).

3. Algorithms

The cluster-based method is a statistical way to characterize the relationship between polarimetric variables and hydrometeor types as well as the classification boundaries based on radar observations. The correlations between input polarimetric variables are naturally considered and incorporated into the classification. It is an incremental method in the sense that prototypes can be added as additional data are being processed. An operational implementation of this technique could be to run the clustering technique once a month in order to progressively refine the PDFs of polarimetric variables for each hydrometeor type as the radar collects a larger variety of cases. Under this framework, the prototype-level MPLC and the distribution-level Bayesian classifier are developed for hydrometeor classification using CP-2 radar observations. A flowchart of the classification procedure is illustrated in Fig. 3. Before the classification, we assume that a set of prototypes representing the hydrometeor types under consideration are available, such as the prototypes produced in Part I. The MPLC and Bayesian classifiers utilize this set of the cluster-derived prototypes. We describe these two classifiers in detail below.

a. Maximum prototype likelihood classifier

The MPLC makes direct use of all the prototypes individually. A block diagram of the MPLC is shown in Fig. 4. The similarity between a data point and each prototype is evaluated in terms of a likelihood function in high-dimensional feature space. The prototype with the largest likelihood is the most similar to the data point. Finally, a class label is assigned to this data
point according to the hydrometeor type of this prototype.

The MPLC involves a set of the total \(N_p\) prototypes \(\mathcal{P} = \{\mathcal{P}_{j,C}\}_{j=1}^{N_p}\), in which the \(j\)th prototype is modeled as a Gaussian distribution, \(\mathcal{P}_{j,C} = \mathcal{N}(\mu_j, \Sigma_j)\), where the \(\mu_j\) and \(\Sigma_j\) are the mean and covariance matrix of \(\mathcal{P}_{j,C}\), respectively. The corresponding hydrometeor type is denoted by \(C \in \mathcal{C}\), where \(\mathcal{C} = \{C_i\}_{i=1}^{N_C}\) (\(N_C\) is 10 in our case). In addition, a weight \(\alpha_j\) is assigned to the \(j\)th prototype \(\mathcal{P}_{j,C}\) in the EM clustering, as described in Part I. The weights satisfy the condition \(\sum_{j=1}^{N_p} \alpha_j = 1\). The conditional probability, \(\Pr(X | \mathcal{P} = \mathcal{P}_{j,C})\), of a data point \(X\) given the \(j\)th prototype \(\mathcal{P}_{j,C}\) can be calculated by

\[
\Pr(X | \mathcal{P} = \mathcal{P}_{j,C}) = \mathcal{N}(X; \mu_j, \Sigma_j)
\]

\[
\Pr(X | \mathcal{P} = \mathcal{P}_{j,C}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_j|}} \exp\left[ -\frac{1}{2} (X - \mu_j)^T \Sigma_j^{-1} (X - \mu_j) \right].
\]  

(1)

In (1), \(d\) is the number of input data fields (five in our case) and the superscript “\(T\)” represents the vector transpose. The similarity between \(X\) and the prototype \(\mathcal{P}_{j,C}\) can be measured in terms of a likelihood function, \(\alpha_j \Pr(X | \mathcal{P} = \mathcal{P}_{j,C})\). As shown in Fig. 4, a maximum likelihood criterion is used to obtain an estimate of the prototype, \(\mathcal{P}_{\text{max}}\), which is the most similar to \(X\) on the basis of the existing information contained in the prototype set. This can be given as

\[
\mathcal{P}_{\text{max}} = \arg \max_j \alpha_j \Pr(X | \mathcal{P} = \mathcal{P}_{j,C}).
\]  

(2)

where “\(\arg \max\)” stands for the argument of the maximum—that is, the set of prototypes of the given argument for which the given function attains its maximum value.

To determine the class label for \(X\), the hydrometeor type \(C_{\text{max}}\) corresponding to the most similar prototype \(\mathcal{P}_{\text{max}}\) is retrieved from the prototype set \(\mathcal{P}\) and subsequently assigned to \(X\).

b. **Bayesian classifier**

The PDFs of polarimetric variables and temperature derived in Part I can be used to construct a Bayesian classifier. A block diagram of the architecture for the Bayesian classifier is shown in Fig. 5. Under the Bayesian interpretation, the compatibility of a data point with the given hydrometeor type can be considered as a conditional probability of the data point \(X\) given the hydrometeor type \(C \in \mathcal{C}\); this is referred to as a likelihood function. This probability may be calculated in terms of the PDFs as

\[
\Pr(X | C) = \sum_{\mathcal{P}_{j,C} = C} \alpha_j \Pr(X | \mathcal{P}_{j,C}) = \sum_{\mathcal{P}_{j,C} = C} \alpha_j \mathcal{N}(X; \mu_j, \Sigma_j).
\]  

(3)

**Fig. 2.** Prototype means of (a) \(Z_h-Z_\text{dr}\), (b) \(Z_h-K_{\text{dp}}\), (c) \(Z_h-P_{\text{hv}}\), and (d) \(Z_h-T_{\text{env}}\) for different hydrometeor types. Prototypes were collected from 25 radar volumes between 0618 and 0642 UTC 16 Nov 2008, 1400 and 1454 UTC 19 Nov 2008, and 0654 and 0748 UTC 24 Mar 2013.
where $C(P_j)$ is the hydrometeor type of the prototype $P_j \in P$. In (3), $\alpha_j'$ is the normalized weight,

$$\alpha_j' = \frac{\alpha_j}{\sum_{C(P_j) = C} \alpha_k}.$$  (4)

A prior probability is also required by the Bayesian classification, as illustrated in Fig. 5. The prior probability $Pr(C)$ includes any extraneous and other information available before the current dataset under scrutiny has been obtained. In the current configuration, $Pr(C)$ is assigned uniformly for every hydrometeor type. During the Bayesian classification process, the hydrometeor types with the second- and third-highest posteriori probabilities are examined if the classification is inconsistent with ground observations or basic physical principles. The prior probability for the hydrometeor types with the posteriori probabilities larger than the correct one will be reduced by half. It is the objective of a future study to incorporate local weather knowledge and the spatial and temporal correlations between range gates into the modeling of the prior probability.

Once $Pr(X \mid C)$ and $Pr(C)$ have been estimated, the conditional probability $Pr(C \mid X)$ of the hydrometeor type given a data point can be calculated using Bayes’s rules. The conditional probability, $Pr(C \mid X)$, also referred to as the posteriori probability in Bayesian statistics, is an optimal way to combine prior information with the new dataset. The posteriori probability can be given as

$$Pr(C \mid X) = \frac{Pr(C)Pr(X \mid C)}{\sum_{C=C_1}^{C_{N_C}} Pr(C)Pr(X \mid C)}.$$  (5)

The maximum of the posteriori probability is calculated in order to obtain the maximum a posteriori (MAP) classification. This can be expressed as

$$C = \arg \max_C Pr(C \mid X).$$  (6)

c. Comparison between the MPLC and Bayesian classifiers

The cluster-derived prototypes can be partitioned into multiple prototype sets, each corresponding to a given hydrometeor type. This is mathematically formulated as

$$\mathcal{P} = \{P_{C_i}\}_{i=1}^{N_C},$$  (7)

where $P_{C_i}$ is a prototype set containing all prototypes corresponding to a particular hydrometeor type $C_i \in C$; that is, $P_{C_i} = \{P_j\}_{j=1}^{N_{C_i}}$, where $N_{C_i}$ is the number of prototypes in the prototype set $P_{C_i}$ and $\sum_{i=1}^{N_C} N_{C_i} = N_P$.

We consider the conditional probability $Pr(C = C_i \mid P = P_j)$ of a particular hydrometeor type $C_i$ given any prototype $P_j$ as

$$Pr(C = C_i \mid P = P_j) = I_{C_j}(P_j) = \begin{cases} 1 & P_j \in P_{C_i} \\ 0 & P_j \notin P_{C_i} \end{cases},$$  (8)

where $I_{C_j}$ is an indicator function and $P_{C_i}$ is the set of the prototypes labeled as the hydrometeor type $C_i$. This reflects the fact that each prototype is mapped to only one hydrometeor type in the current procedure of promoting a cluster to a prototype in the prototype generation unit, as described in Part I. Therefore, the classification techniques identify the dominant hydrometeor type for each radar resolution volume.

Under the Bayesian classification, we need to compute the posteriori probability $Pr(C = C_i \mid X)$.
According to the properties of the marginal and conditional probability functions, \( \Pr(C = C_i \mid X) \) is given by

\[
\Pr(C = C_i \mid X) = \frac{1}{C_{\text{max}}} \sum_{j=1}^{N_p} \Pr(C = C_i, P = P_j \mid X) = \frac{N_p}{C_{\text{max}}} \sum_{j=1}^{N_p} \Pr(C = C_i \mid P = P_j, X) \Pr(P = P_j \mid X).
\]

(9)

Since \( \Pr(C = C_i \mid P = P_j, X) = \Pr(C = C_i \mid P = P_j) \), we obtain

\[
\Pr(C = C_i \mid X) = \frac{1}{C_{\text{max}}} \sum_{j=1}^{N_p} \Pr(C = C_i \mid P = P_j, X) \Pr(P = P_j \mid X).
\]

(10)

Substitution of (8) in (10) then gives

\[
\Pr(C = C_i \mid X) = \frac{1}{C_{\text{max}}} \sum_{P_j \in \mathcal{P}_{C_i}} \Pr(P = P_j \mid X).
\]

(11)

As can be seen in (11), the posteriori probability of a hydrometeor type for any given data point in the cluster-based Bayesian classification is equal to the sum of the posteriori probabilities of all prototypes corresponding to this hydrometeor type given the data point. The decision rule for the Bayesian classifier can then be expressed as

\[
C = \arg \max_{C_i} \sum_{P_j \in \mathcal{P}_{C_i}} \Pr(P = P_j \mid X).
\]

(12)

If the local neighborhood is emphasized using appropriate weights of the prototypes closest to the data point being analyzed, the Bayesian classifier is reduced to the MPLC formulation, namely,

\[
C = \arg \max_{C(P)} \Pr(P = P_j \mid X).
\]

(13)

It can be concluded that the MPLC and Bayesian classifiers differ from each other in two aspects:

1) The local neighborhood is emphasized in the MPLC since the prototype with the highest likelihood is picked, whereas the contributions of all prototypes corresponding to an identical hydrometeor type are taken into account in the Bayesian classifier through the use of the full PDFs of polarimetric variables.

2) In the Bayesian method, a prior probability needs to be defined. If the prior probabilities are assigned uniformly, the Bayesian classifier becomes similar to the MPLC.

The MPLC requires much less previous radar data to train the classifier in comparison with the Bayesian algorithm. The advantage of the Bayesian approach is that it may be used to introduce temporal and spatial correlations of polarimetric signatures in terms of a prior probability and to incorporate information from other sources (such as local weather knowledge). If the training dataset is relatively small, we would recommend adopting the MPLC, but the Bayesian classifier is preferable if the number of prototypes is large enough to produce smooth PDFs for all hydrometeor types. In what follows, we show
and quantitatively compare the results from both techniques.

d. Operational implementation considerations

Since radar datasets are often large, we also need to consider and compare the computational complexity of the two classifiers. To reduce the number of computations for each of the classifiers, we take the natural logarithm for the computation of the posteriori probabilities in the processing of each classifier. The MPLC becomes equivalent to maximizing a similarity measure that is given by

\[
\text{Similarity} = \log(\alpha_j) - \frac{1}{2} \log(|\Sigma_j|)
\]

\[
-\frac{1}{2} (X - \mu_j)^T \Sigma_j^{-1} (X - \mu_j).
\] (14)

By considering the computations of the similarity measure and the maximum likelihood, the time for computation of the MPLC for a dataset with \( M \) points may be expressed as

\[
t_{\text{MPLC}} = MN_P (t_{\text{maha}} + t_{\text{ML}} + t_{\text{other}}),
\] (15)

where \( t_{\text{maha}} \) is the time for computations of the squared Mahalanobis distance, \( (X - \mu_j)^T \Sigma_j^{-1} (X - \mu_j) \); \( t_{\text{ML}} \) is the time for computation of the maximum likelihood; and \( t_{\text{other}} \) is the time consumed by other operations. The computational complexity of the MPLC can therefore be defined as \( O(MN_P) \), where the expression \( O(*) \) represents the growth rate in order of a mathematical function.

Similarly, after taking a natural logarithm for the posteriori probability in the Bayesian classifier, we obtain

\[
\log \Pr(C \mid X) = \log \Pr(C) - \log \Pr(X)
\]

\[
+ \log \left\{ \sum_{C(P) = C} \exp \left[ \log(\alpha_j') - \frac{d}{2} \log(2\pi) \right] \right\}
\]

\[
- \frac{1}{2} \log(|\Sigma_j|) - \frac{1}{2} (X - \mu_j)^T \Sigma_j^{-1} (X - \mu_j).
\] (16)

The time for computation of the log-posteriori probability and MAP can then be expressed as

\[
t_{\text{Bayes}} = MN_P (t_{\text{maha}} + t_{\text{exp}} + t_{\text{other}}) + MN_C (t_{\text{prior}} + t_{\text{MAP}}),
\] (17)

where \( t_{\text{exp}} \) is the time for computing the exponential and \( t_{\text{prior}} \) is the time for obtaining the logarithm of the prior probability. Because the computation of the Mahalanobis distance dominates, the computational complexity of the Bayesian classifier may also be expressed as \( O(MN_P) \). Therefore, the MPLC and the Bayesian classifier have a comparable computational complexity and our experiments confirm that the time required for the two classifiers to process the same dataset is reasonably close. In our experiments, both algorithms (using 449 prototypes) require about 30 s for a single PPI sweep and 2-3 min for a full volumetric scan (9-10 sweeps).

4. Results

The validation of hydrometeor classification techniques is difficult. In the literature, researchers have developed both qualitative and quantitative methods for such validation. Liu and Chandrasekar (2000) and Lim et al. (2005) compared their classification results to airborne in situ microphysical observations from the T-28 aircraft using optical array probes, such as the 2D cloud particle measurement probe (2D-C), the high-volume particle sampler, and the hail spectrometer. Although the spatial coverage and resolution of such observations are very different from that of radar, this is probably the most objective way to validate such classification methods. However, aircraft observations within dual-polarization radar coverage are rare and generally expensive. May and Keenan (2005) systematically evaluated C-band classification results with a combination of 50- and 920-MHz wind profiler estimates of rain and hail for the radar pixels over the profiler. Marzano et al. (2008) simulated the backscattering characteristics of multiple types of precipitation and subsequently produced a labeled training set to test the accuracy of the algorithms of hydrometeor classification. Using such simulated observations, quantitative scores were derived to assess the performance of different classifiers. Schuur et al. (2012) compared Automated Surface Observing System ground observations to the results of precipitation classification based on polarimetric data and the Rapid Update Cycle model. Both Dolan and Rutledge (2009) and Snyder et al. (2010) compared the classification results of X-band radar data to S-band radar data and analyzed the radar intercomparison qualitatively. Zrnić et al. (2001) and Park et al. (2009) developed procedures for verification and self-consistency based on intuition, precipitation microphysics, and conceptual models. In the study by Al-Sakka et al. (2013), a validation dataset was built subjectively by a human expert and then used to obtain quantitative scores to measure the accuracy of the classification.

The evaluation of our proposed algorithms is also challenging due to the lack of in situ microphysical measurements for direct comparisons within the CP-2
radar coverage, especially for ice-phase hydrometeors. In this section, our classification results are benchmarked against the well-established and widely used NCAR fuzzy logic classifier. This cannot be considered as a validation, but it is a good way to evaluate how well the cluster-based method compares with a well-established algorithm. It is important to stress here that neither our techniques nor the NCAR technique will be considered as the reference in what follows, in the absence of truly independent in situ measurements. However, simple but robust microphysical rules are used when possible to identify spurious classification in the three different techniques. A simple example of such a rule is that there cannot be a rain/hail mixture classified without having a hail core in the ice phase somewhere prior to the incidence of the rain/hail mixture. Available ground reports are also used to validate the incidence of hail at the ground. However, those reports are not very specific about hail location and timing. In this section, we present classification results from a series of severe thunderstorms that occurred in southeast Queensland between 16 and 20 November 2008.

On 16 November 2008, convective activity first started over northern New South Wales, Australia, in the afternoon and later crossed the border into southeast Queensland. The first storm propagated in a northeastern direction, causing wind and hail damage. This storm then slowed and merged with a second storm. The resulting storm moved northward from Redbank Plains, Queensland, where the CP-2 research radar is located. This culminated in an extremely intense hailstorm at the Brisbane suburb called “The Gap.” The reader is referred to Richter et al. (2014), who published a detailed analysis of this case. On 20 November 2008, a distinct line of storms moved through southeast Queensland, producing mostly wind and hail damage in the Brisbane metropolitan area. Rainfall was intense but over a short period, with fast-moving storms organized into a single line that moved across southeast Queensland from Beaudesert to the Sunshine Coast in less than 3 h (from approximately 0800 to 1100 UTC). The reports from the Bureau of Meteorology indicate hail of at least a cricket ball size (7 cm) and extremely strong wind gusts.

a. 16 November 2008, “Gap storm”

Figures 6a–d shows the CP-2 polarimetric radar observations at 17.8° elevation angle for this event. Three interesting features can be identified. The first is a large pocket of high reflectivity (>50 dBZ) observed near the 0°C isotherm (X: 0 ~ 10 km, Y: 5 ~ 10 km). Term $Z_{dr}$ is approximately 1 dB, and $K_{dp}$ is between 1° and 2° km$^{-1}$. We should therefore expect the classifiers to detect a mixture of rain and hail in this region. In Figs. 6e–g, we present the classification results of the MPLC, Bayesian, and NCAR classifiers. The rain/hail mixture (purple) classification produced by the three techniques in this area agrees very well in terms of location and areal extent, and it also agrees with our initial guess based on the polarimetric signatures. Ground reports from this event indicate that damaging hailstones were observed at several locations along the path of the storm, some as large as golf balls. A large area of the rain/graupel mixture is also produced by the NCAR classifier, surrounding the region where the rain/hail mixture is found. Our classifiers indicate moderate or heavy rain in this same area. As discussed earlier, our clustering method could not identify a rain/graupel mixture class, despite the fact that this case with large areas of rain/graupel in the NCAR classifier was included in our training dataset for the clustering. Therefore, in order to compare our techniques and the NCAR technique, we will consider in what follows that the techniques agree when our techniques classify a radar pixel as heavy rain and the NCAR technique classifies it as rain/graupel or heavy rain. Regarding that specific case, ground reports do not show any occurrence of graupel at the ground, which may indicate that the classification of rain/graupel by the NCAR technique is questionable and warrants further investigations or tuning of the membership functions for that specific class.

The second prominent feature in Fig. 6 is high $Z_h$ (in excess of 60 dBZ) associated with $Z_{dr}$ smaller than 0.5 dB observed above the 0°C isotherm in the north-northwest (at X: -15 to ~ -5 km, Y: 20 ~ 30 km) and north-northeast (at X: 5 ~ 15 km, Y: 20 ~ 30 km) quadrants, which are obvious signatures of hail. In Fig. 6, the classification results indeed show large pockets of hail. Graupel/small hail is also detected by all three classifiers on the edges of this hail region. As discussed in Part I and in the literature, the polarimetric signature of hail with diameters from 5 to 20 mm is very similar to that of graupel with diameters from 0.5 to 5 mm. As a result, we do not attempt to separate between graupel and small hail in our classification, and nor does the NCAR classifier. In this particular case, the spatial distribution indicates that these are likely small hail particles, not graupel, as it surrounds a well-defined hail area. The comparison of the three classifiers in this area shows that the MPLC and Bayesian classifiers identify ice aggregates, while the NCAR classifier shows a large area of graupel/small hail. It is difficult to further quantify which result is closer to the truth; however, again this difference clearly reflects differences for graupel/small hail between the assumed membership functions in the NCAR classifier and the PDF of polarimetric variables derived from our clustering technique.
The third interesting signature observed in Fig. 6 is a cell of negative $K_{dp}$, large $Z_h$, nearly zero $Z_{dr}$, and low $\rho_{hv}$ at $X$: 15 km, $Y$: 20 ~ 25 km. The negative $K_{dp}$ values are often observed within hail regions (Figueras i Ventura et al. 2013). Term $K_{dp}$ can also be negative for vertically oriented ice crystals, which is due to the presence of a strong electric field (Ryzhkov and Zrnić 2007; Hubbert et al. 2014). As compared to $K_{dp}$, the corresponding $Z_{dr}$ generally remains near zero because it is heavily weighted by larger size ice aggregates or crystals that do not align with the electrostatic field. In Figs. 6e–g, all of the three classifiers identified this signature as hail. In addition, Fig. 6 shows an arc of melting ice near the 0°C isotherm where the transition between ice and liquid hydrometeors occurs. The spatial distribution of melting ice is very similar for the three classifiers.

To check the consistency of the vertical continuity of the three hydrometeor classifications further, we show two vertical cross sections at azimuth 25° and 345° in...
Figs. 7 and 8, respectively, which correspond to the two black lines in Fig. 6a. The classification results of the MPLC and Bayesian classifiers are broadly consistent with the NCAR fuzzy logic classifier. The three classifiers detect a large region of hail with a rain/hail mixture underneath in Figs. 7e–g (range: 10–20 and 25 km) and Fig. 8e–g (range: 20–30 km), starting from the top of the echoes and descending to the ground. The NCAR classifier produces few rain/graupel pixels at the highest elevation angle (at range ~16–18 km in Fig. 7) above the hail and rain/hail mixture regions. This suggests again that the membership functions for the rain/graupel class should probably be adjusted for this specific location in the NCAR classifier.

In Table 3, we show confusion matrices in order to quantitatively analyze the differences between the three classifiers. The confusion matrices show the percentages of the number of samples in each cell (normalized by the total number of points along each row). The diagonal elements of the matrices give a measure of how
consistent the classifications are, while the other elements of the matrices indicate the degree of disagreement between classifications. In Table 3a, the diagonal elements of all of the categories achieve the largest values (79% or more) between the MPLC and Bayesian classifiers. This highlights the good consistency between our two techniques (as also observed more qualitatively in Figs. 7 and 8). The main differences are between the different ice particle categories (7% of ice aggregates from the Bayesian classifier are classified as ice crystals by the MPLC classifier), and between melting ice and rain or ice aggregates (melting ice in the Bayesian classifier is classified 11% of the time as rain and 8% of the time as ice aggregates).

The confusion matrices between our two new classifiers and the NCAR classifier are characterized by slightly larger differences. A very good agreement is found for the ice aggregates (86%), rain (81%), rain/hail (79%–86%), and hail (75%) categories (Tables 3b and 3c). The largest differences are found for the melting ice

![Fig. 8. Vertical cross sections of (a) $Z_h$ (dBZ), (b) $Z_{dr}$ (dB), (c) $K_{dp}$ ($\text{km}^{-1}$), (d) $r_{hv}$, (e) MPLC, (f) Bayesian classification, and (g) NCAR fuzzy logic classification at a 345° azimuthal angle measured at 0630 UTC 16 Nov 2008.](image)
category, with low values of the diagonal elements for NCAR–MPLC and NCAR–Bayesian of only 39% and 30%, respectively, indicating that 61%–70% of the radar data classified as melting ice by our techniques are classified differently by the NCAR classifier. In contrast, when the NCAR classifier detects melting ice, our two techniques generally agree. Large differences are also found for the ice crystal category, with the NCAR classifier indicating ice aggregates about 40% of the time when our classifiers indicate ice crystals.

### Table 3. Confusion (matching) matrix between (a) MPLC–Bayesian classifier, (b) NCAR–MPLC, and (c) NCAR–Bayesian classifier between 0554 and 0648 UTC 16 Nov 2008. (The percentages in the confusion matrices represent the proportions of the categories on the row classified as the ones on the column.)

<table>
<thead>
<tr>
<th></th>
<th>Rain</th>
<th>Hail</th>
<th>Rain/hail</th>
<th>Graupel</th>
<th>Ice crystals</th>
<th>Ice aggregates</th>
<th>Melting ice</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bayesian classifier</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rain</td>
<td>95%</td>
<td>0%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
</tr>
<tr>
<td>Hail</td>
<td>2%</td>
<td>97%</td>
<td>0%</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Rain/hail</td>
<td>3%</td>
<td>0%</td>
<td>96%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Graupel</td>
<td>2%</td>
<td>5%</td>
<td>1%</td>
<td>89%</td>
<td>0%</td>
<td>0%</td>
<td>2%</td>
</tr>
<tr>
<td>Ice crystals</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>90%</td>
<td>8%</td>
<td>1%</td>
</tr>
<tr>
<td>Ice aggregates</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
<td>7%</td>
<td>88%</td>
<td>1%</td>
</tr>
<tr>
<td>Melting ice</td>
<td>11%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>2%</td>
<td>8%</td>
<td>79%</td>
</tr>
</tbody>
</table>

| **NCAR fuzzy logic classifier** |      |      |           |         |              |                |             |
| **MPLC** |      |      |           |         |              |                |             |
| Rain     | 81%  | 0%   | 2%        | 14%     | 0%           | 1%             | 2%          |
| Hail     | 0%   | 75%  | 19%       | 6%      | 0%           | 0%             | 0%          |
| Rain/hail| 20%  | 1%   | 79%       | 1%      | 0%           | 0%             | 0%          |
| Graupel  | 0%   | 5%   | 0%        | 75%     | 0%           | 17%            | 2%          |
| Ice crystals | 0% | 0%   | 0%        | 0%      | 60%          | 39%            | 0%          |
| Ice aggregates | 0% | 0%   | 0%        | 9%      | 3%           | 86%            | 1%          |
| Melting ice | 12% | 0%   | 0%        | 18%     | 8%           | 23%            | 39%         |

| **NCAR fuzzy logic classifier** |      |      |           |         |              |                |             |
| **Bayesian classifier** |      |      |           |         |              |                |             |
| Rain     | 80%  | 0%   | 2%        | 13%     | 0%           | 1%             | 3%          |
| Hail     | 0%   | 76%  | 21%       | 3%      | 0%           | 0%             | 0%          |
| Rain/hail| 12%  | 1%   | 86%       | 1%      | 0%           | 0%             | 0%          |
| Graupel  | 0%   | 11%  | 0%        | 75%     | 0%           | 8%             | 4%          |
| Ice crystals | 0% | 0%   | 0%        | 0%      | 61%          | 38%            | 0%          |
| Ice aggregates | 0% | 0%   | 0%        | 12%     | 1%           | 86%            | 1%          |
| Melting ice | 10% | 0%   | 0%        | 22%     | 7%           | 30%            | 30%         |

### 20 November 2008

*Figure 9* shows a high-resolution RHI from a strong convective storm for which hail was also reported at the ground. This case allows for investigating at high vertical resolution the vertical continuity of the three hydrometeor classification algorithms, which was not possible with the first case. Deep convective cores of 45 dBZ reached 10 km in height, while intense reflectivity > 50 dBZ extended out to 50 km in range and some narrow regions of 55 dBZ extended up to 8 km in height. All classifiers produce regions of hail concentrated between 6 and 8 km and a rain/hail mixture below 6 km in height. Note that all classifiers are able to detect a rain/hail mixture (presumably lifted by the main updraft) at altitudes above the 0°C isotherm altitude and up to 6-km height. Graupel/small hail is classified above and around the hail region by all classifiers as well. In our classification results, ice aggregates predominantly encase the main convective core and melt near the freezing level, which leads to rain below. The NCAR classifier produces a much larger area of graupel/small hail than our classifiers, yielding a graupel/rain mixture below the melting layer. In addition, the NCAR classification results appear to be smoother than our classification results, which reflects the fact that the NCAR algorithm uses additional filtering to obtain the classification. Since both classification results are different but individually consistent vertically, it is not possible to provide further evidence that one classifier is better than the other.
Besides, there are no ground reports available to check for graupel at the ground for that case. Confusion matrices between the MPLC, Bayesian, and NCAR classifiers (not shown for this case) are very similar to the previous case. Another difference between the NCAR classifier and our techniques is the classification of melting ice near the freezing level. In Fig. 9, the identification of melting ice can be further investigated by using the high vertical resolution of these radar data. It is clearly seen that the NCAR classifier produces a spurious layer of rain above melting ice. This shows the NCAR algorithm may rely too heavily on temperature to separate rain and ice. In this case the temperature measurement was collected 2 h after the radar observations. In contrast, a much smaller amount of radar pixels is also wrongly classified as rain above the melting ice layer by our techniques. This could either indicate that the temperature information is not accurate for this case or that the membership functions assumed for melting ice should probably be tuned further in the NCAR classifier.

\[ \text{Fig. 9. RHIs of (a) } Z_h (\text{dBZ}), (b) Z_D (\text{dB}), (c) } K_{dp} (\text{°km}^{-1}), (d) \rho_{hv}, (e) \text{MPLC, (f) Bayesian classification, and (g) NCAR fuzzy logic classification at a 61.3° azimuthal angle measured at 0853 UTC 20 Nov 2008. The dashed line in (a) represents the freezing level.} \]
5. Conclusions

Two new classification algorithms were developed based on clustering-derived prototypes in the companion paper (Part I). The maximum prototype likelihood classifier measured the similarity between a data point and each prototype. The likelihood function was calculated using the distribution of the data point with respect to each prototype. A data point was classified to a hydrometeor type corresponding to the prototype with the maximum likelihood. In the Bayesian approach, we utilized the PDFs for each hydrometeor type. The PDFs and prior probability were combined to produce the posteriori probability. A data point was classified by finding the class with the maximum posteriori probability.

Results obtained with the two classifiers using CP-2 radar observations between 16 and 20 November 2008 were compared with the results obtained with the well-established NCAR fuzzy logic classifier. The 16 November “Gap storm” exhibited hail classifications in both methods, which was consistent with surface reports. The analysis of high-resolution RHI observations of a deep convective hailstorm on 20 November also demonstrated physically consistent hydrometeor classifications by our techniques throughout the radar volume. The results from these two examples were also generally consistent with those from the NCAR fuzzy logic classifier, except for the horizontal extent of graupel/small hail and melting ice, and the discrimination between the ice crystal and ice aggregates categories. From the analysis of the second case, there is some indication that our classifiers produce a more realistic classification of melting ice, as the NCAR classifier was found to produce large amounts of light rain above melting ice, which is not realistic. In situ microphysical observations would be needed to quantitatively confirm these inferences.

Acknowledgments. Michael Whimpey from the Australian Bureau of Meteorology provided assistance with the CP-2 radar data. The authors thank Susan Rennie, Vickal Kumar, and four anonymous reviewers for their constructive comments on this work.

REFERENCES


