A SYSTEM FOR AUTOMATING IDENTIFICATION OF BIOLOGICAL ECHOES IN NEXRAD LEVEL II RADAR DATA

by

Reginald Marshall Mead

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Computer Science

MONTANA STATE UNIVERSITY
Bozeman, Montana

November, 2009

APPROVAL

of a thesis submitted by

Reginald Marshall Mead

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Division of Graduate Education.

Dr. John Paxton

Approved for the Department of Computer Science

Dr. John Paxton

Approved for the Division of Graduate Education

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.

Reginald Marshall Mead

November, 2009

DEDICATION

I dedicate this thesis to my family and friends, especially those friends who constantly encouraged me to put off my thesis until "next semester".

ACKNOWLEDGEMENTS

I would like to thank Dr. Rick Sojda and Dr. John Paxton for inviting me to work on this project and for providing plenty of guidance for me along the way. I would also like to thank Dr. Robb Diehl and the University of Southern Mississippi for their support of data and expertise. I would like to thank Manuel Suarez for his many contributions to the project. I would like to thank Karin Sinclair, Bob Thresher and the National Renewable Energies Lab for their support.

Funding Acknowledgment

This work was kindly supported by grants from the National Renewable Energies Lab and the USGS. However, any opinions, findings, conclusions, or recommendations expressed herein are those of the author and do not necessarily reflect the views of NREL or the USGS.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Since its inception in the mid twentieth century, radar ornithology has provided scientists with new tools for studying the behavior of birds, especially with regards to migration. A number of studies have shown that birds can be detected using a wide variety of radar devices. Generally, these studies have focused on small portable radars that typically have a finer resolution than large weather surveillance radars. Recently, however, a number of researchers have presented qualitative evidence suggesting that birds, or at least migration events, can be identified using large broad scale radars such as the WSR-88D used in the NEXRAD weather surveillance system. This is potentially a boon for ornithologists because NEXRAD data covers a large portion of the country, is constantly being produced, is freely available, and is archived back into the early 1990s.

A major obstacle is that identifying birds in NEXRAD data currently requires having a trained technician manually inspect a graphically rendered radar sweep. The immense amount of available data makes manual classification of radar echoes infeasible over any practical span of space or time.

In this thesis, a system is presented for automating this process using machine learning techniques. This approach begins with classified training data that has been interpreted by experts or collected from direct observations. The data is preprocessed to ensure quality and to emphasize relevant features. A classifier is then trained using this data and cross validation is used to measure performance. The experiments in this thesis compare neural network, naïve Bayes, and k-nearest neighbor classifiers. Empirical evidence is provided showing that this system can achieve classification accuracies in the 80^{th} to 90^{th} percentile.

CHAPTER 1
INTRODUCTION

Motivation

Free and easily accessible WSR-88D weather radar data from the Next Generation Radar (NEXRAD) system has opened new doors for many in the scientific community [1, 2]. In particular, radar ornithologists have the opportunity of studying bird migration behavior from a new perspective [3]. Radar, in general, allows for such things as monitoring bird movements at night and the NEXRAD system could provide the ability to track migration over a much broader geographical scale. The NEXRAD system also provides useful [4] historical data, with most sites archiving data back into the early 1990s.

Currently, one of the biggest obstacles to utilizing this data for radar ornithology is the lack of an automated approach for identifying radar scans that include birds. Instead, researchers must painstakingly look through graphically rendered radar scans and manually identify those containing echoes with the characteristics of birds. Clearly, this can be a lengthy and expensive approach. It is also impractical when dealing with data on the scale of the NEXRAD system, which is constantly acquiring more data. A single site alone, completing one volume scan every ten minutes will produce over 52,000 volume scans in one year. This results in over 200GB of data a year for just one site.

This thesis describes an algorithmic approach to classifying biological and nonbiological echoes in NEXRAD radar data using machine learning techniques. It also provides empirical results of applying the algorithm in various settings.

Scope

This thesis is confined to separating biological echoes from nonbiological echoes. Although a distinction could be drawn between birds and insects or even between various types of birds such as waterfowl and passerines, the focus of this thesis is simply differentiating between echoes of interest (something biological) and everything else. Likewise, no attempt is made to classify nonbiological echoes into subcategories such as rain, snow, dust, etc.

Outline of Thesis

This thesis is organized as follows. Chapter 2 provides a review of the background literature and a description of the structure of NEXRAD level II data. Chapter 3 describes the process developed. Empirical results are provided in Chapter 4. This thesis concludes with a discussion concerning the implications of the results as well as directions for future research in Chapter 5.

CHAPTER 2
BACKGROUND

This chapter will provide background information necessary for understanding the problem being investigated in this thesis, beginning with a review of the literature from the radar ornithology field, with emphasis on work related to the WSR-88D radar installations in the NEXRAD surveillance network. This is followed by a detailed discussion of the operation of the WSR-88D as well as the format of the Level II data provided by the National Climatic Data Center (NCDC) NEXRAD repository.

Literature Review

Prior to the availability of radar, studying bird migration was a painstaking and error prone task, often requiring field observers to record bird counts acquired through visual observation. One particular problem with this approach is that some birds, such as passerines (songbirds) and waterfowl, often do a large portion of their migration at night [5, 6, 7].

Moon watching [8], the process of estimating nocturnal migration by recording bird counts as they fly across the disk of the moon, was a partial solution to this problem, but it suffered from its own drawbacks. Moon watching is particularly limited by the small amount of observable space, the requirement of a full moon, and clear skies. [9]. In general, visual observation of bird migrations is limited by visibility issues relating to distance, light and weather conditions.

Radar provided a tool whereby many of the limitations of direct visual observation could be avoided. Radar ornithology, the study of birds by radar, is almost as old as radar itself. In fact, it was the early radar technicians in the post World War

II era that first began noticing anomalous radar echoes that could not be properly attributed to a source. It didn't take long for these early pioneers to determine that the source of these "angels" was actually birds [6, 10].

Many studies have shown the effectiveness of studying birds using a wide variety of radars including portable marine and military radars, small surveillance radars, and larger weather radars [11, 12, 1]. Radars are separated into classes based on operating frequency / wavelength. These classes are designated by letters that were originally used by the military to maintain secrecy. For the purpose of radar ornithology, only the X, C, L and S band radars are of interest. X band radars use the smallest wavelengths, with operating frequencies of roughly 8 to 12 GHz. This roughly corresponds to wavelengths in the 2.4 to 3.8 cm range. C band radars operate in the 4 to 8 GHz range with wavelengths of roughly 3.8 to 7.5 cm, S band radars operate in the 2 to 4 GHz range with wavelengths of 7.5 to 15cm and L band radars operate in the 1 to 2 GHz range and have wavelengths of roughly 15 to 30 cm [13, 9]. C and S band radars are optimal for studying passerines, while L band and larger wavelengths will start to wrap around small birds and suppress them in the data [9]. Radars with smaller wavelengths are better for studying smaller specimens and some radars, such as X band tracking radars, can even be used to measure wing-beat patterns [14].

Radar ornithologists were quick to see the potential when the U.S. Weather Bureau (later to become the National Weather Service) started deploying a network of WSR-57 weather surveillance radars [15]. These S band 10 cm radars provided broader coverage than any previous radar network that was open to the scientific community. The real boon came to the academic community in the 1990s when the US government started replacing the WSR-57 with the WSR-88D. In addition to the base reflectivity product provided by the WSR-57, the WSR-88D provided Doppler information in the form of two new data products: mean radial velocity and spectrum width [16]. This

new network became known as the Next Generation Radar network or NEXRAD. As the internet became ubiquitous in the academic community, a new digital distribution network provided researchers with a fast and free method of attaining data from the NEXRAD network [17].

Since its inception, a number of papers have been published showing that the NEXRAD WSR-88D is an effective tool for studying bird migration and behavior [18, 19, 20, 21, 22, 23, 24]. The field of Chiropterology, the study of bats, has also benefited from this research as new studies are coming out indicating that the NEXRAD network may also be useful in studying the flight patterns of bats [25].

It should be noted that ornithologists are not the only group interested in understanding the way that individual or groups of birds appear in and affect NEXRAD data. There is also interest in the meteorological community, where scientists would like to know how birds affect estimates of wind and other parameters. It has been shown that bird contamination can have a measurable effect on various meteorological products produced by the weather radar surveillance network [26, 27, 28].

Although there have been a rising number of papers relating to the identification of birds in radar data, applications of machine learning techniques in this area are somewhat more sparse. Most deal with clutter removal (where clutter includes birds) through the use of neural networks [29, 30, 31]. In general, these applications have targeted smaller radars with different operational parameters than the WSR-88D, or they make use of the signals coming directly from the radar unit that are not available to researchers using the processed Level II data. An exception to this is a paper by Krajewski et al. [32] in which they use neural networks to identify anomalous propagation in WSR-88D Level II data. Although similar in goal, identifying anomalous propagation is, however, a much different problem from identifying birds.

WSR-88D Data

Like all radar, the WSR-88D sends out repetitions of electromagnetic pulses and measures the returned signal to sense foreign elements. Different properties of these foreign elements can be ascertained by looking at different aspects of the returned signal, namely signal strength and Doppler shift. These properties give us the basic data moments of reflectivity, radial velocity and spectrum width. These basic moments combined with radar operational status make up Level II data [16]. Increasing levels indicate increasing amounts of processing. Level I data includes the analog signals produced by the base radar equipment and are not recorded. Level III and higher data include derived products that are calculated from the Level II data [16]. Level II is the lowest level archived and available to the general public [17].

A single pulse repetition senses objects within the radar's beam, which is either roughly 1° or 0.5° in width for the WSR-88D. For weather surveillance radars to be useful, they need to have a full 360° view of the surrounding air space. This is accomplished by rotating the radar about a vertical axis while continually sending out pulse repetitions. At its lowest level, the radar beam is nearly level with the horizon (0.5° above the horizon). The elevation angle of the radar can be increased to probe higher levels of the atmosphere.

As previously mentioned, NEXRAD radars produce three basic data moments: reflectivity, radial velocity and spectrum width. Reflectivity is a measure of the returned signal strength, whereas radial velocity and spectrum width are based on the Doppler shift of the returned pulses, measuring object movement.

Figure 1 shows reflectivity data of meteorological phenomenon for an example NEXRAD Level II scan. This scan was rendered using the NOAA Weather and Climate Toolkit [33]. Raw reflectivity is denoted by units of Z, where Z is a signal

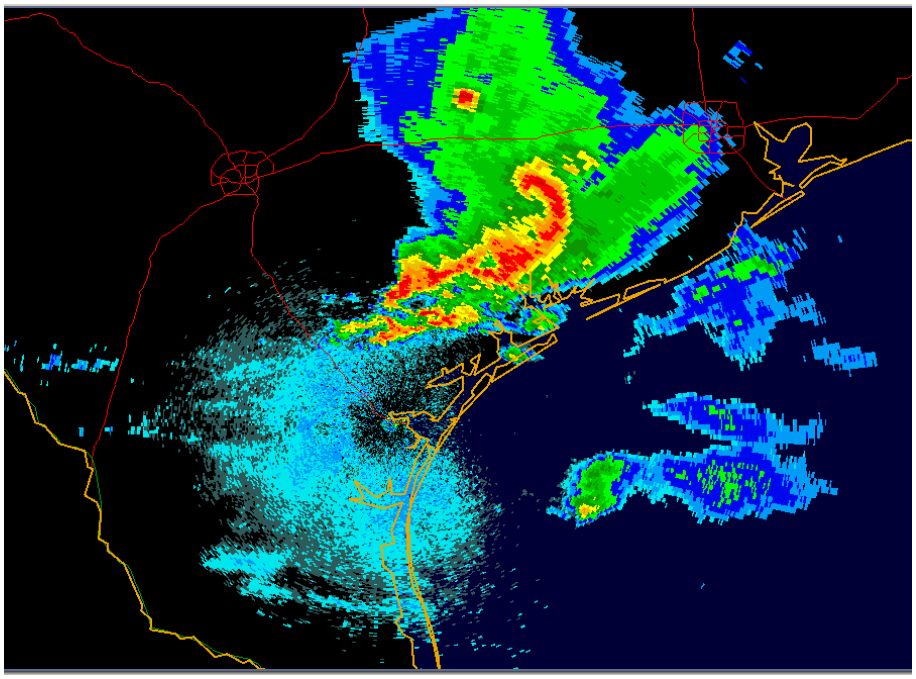


Figure 1: Example NEXRAD Scan Showing Reflectivity. Blues and greens depict lower signals while yellows and reds are indications of stronger signals.

strength scale used in meteorology. Due to its nature, reflectivity is usually converted to a logarithmic scale and provided in decibels of Z (dBZ). In general, larger values of reflectivity are measures of more pulsed data being re3turned. This usually indicates larger objects (or larger numbers of small objects, e.g. rain drops). The range of values for reflectivity depends largely on the current scanning mode being used, but the system has an overall range of -28 to +75 dBZ.

Mean Radial Velocity measures the average velocity in ms^{-1} at which sensed entities in the radar beam are moving towards or away from the radar. Figure 2 shows radial velocity values for the same scan shown in Figure 1. Because it is a component measurement, this metric only measures true ground speeds for objects moving directly parallel with the radar's beam. Likewise, objects moving perpendicular to the radar's beam will have zero radial velocity. One might think that the nature of

the radial velocity measurement would decrease its usefulness, and at the local level this might be true, but when radial velocity values are viewed for an entire scan, its usefulness is apparent. As can be seen in Figure 2, the boundary between green and red values roughly defines a line perpendicular to the direction of movement. In this figure, shades of green represent negative values. Perhaps counter intuitive, negative values indicate a component vector moving towards the radar. Positive values, shown in red, indicate that the component values are moving away from the observer.

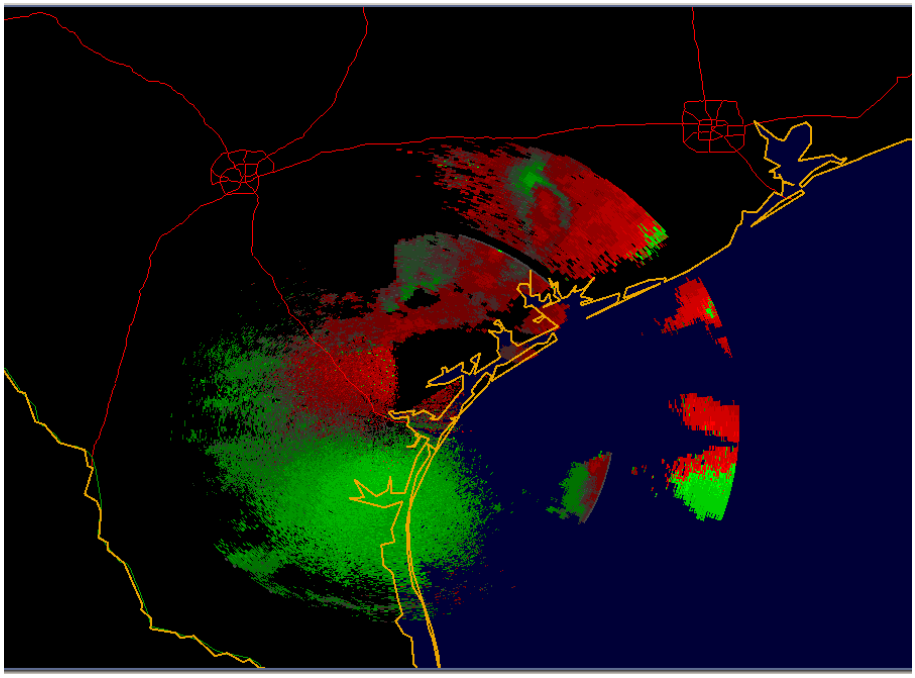


Figure 2: Example NEXRAD Scan Showing Radial Velocity. Shades of green and red indicate movement towards or away from the radar respectively.

The last of the three basic radar products is spectrum width. As mentioned above, the radial velocity measurement produced by the radar is an average of the radial velocity components of all objects within the radars beam at a given range. Spectrum width, illustrated in Figure 3, gives us a metric indicating the amount of

variation in radial velocity among those individual objects. High spectrum width values indicate high variability.

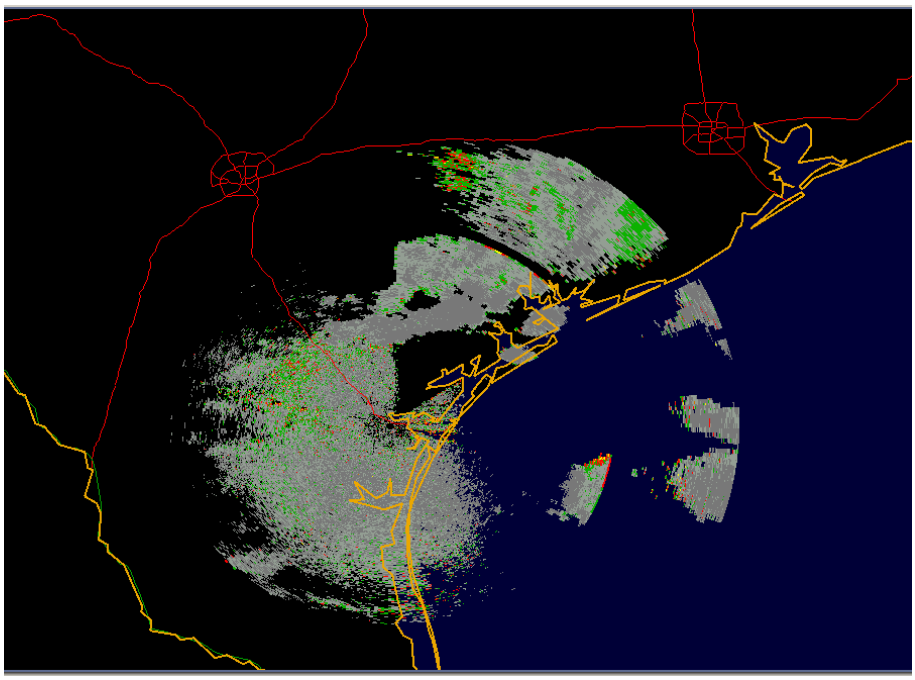


Figure 3: Spectrum Width

Typically, data is acquired for a subset of possible elevation angles depending on current atmospheric conditions (i.e. clear air or precipitation). This subset is known as the *Volume Coverage Pattern*(VCP). A WSR-88D using a given VCP will begin at the lowest elevation angle, perform a scan, raise to the next highest elevation angle specified, perform another scan, etc. At the end of the sequence the radar returns to the lowest elevation angle and repeats the pattern. In addition to specifying which elevation angles to use, the VCP also specifies the pulse length, the rotation speed, and the data moments to be collected for each scan.

The fact that the VCP may indicate that only a subset of the possible data moments may be collected for a particular scan should be emphasized. Although both reflective and Doppler information can be collected simultaneously, it is often

broken out into two or more separate scans. The reasoning is that the pulse repetition frequency (PRF) that results in the best reflectivity data in terms of accuracy and maximum range is not necessarily best for Doppler data. One cannot assume that a given scan contains both reflective and Doppler data or that the data contained is the best that was collected for that elevation angle. Details for the various possible volume coverage patterns is provided in Appendix A.

NEXRAD data is conceptually organized into four hierarchical components that describe the three dimensional space around the radar site [34]. Figure 4 illustrates how these components fit together. The largest logical unit is the volume, representing the entire three dimensional space around a radar site, as sampled by the current volume coverage pattern. The shape of a volume can be roughly conceptualized as a half sphere. This half sphere is then divided into a set of two dimensional sweeps or scans. Each sweep contains data at a specific elevation angle. Each sweep can be further divided into a set of rays much like the spokes on a tire. Each ray contains data at a certain azimuth. Finally, the ray is divided into a set of bins or pulse volumes [2, 35]. Pulse volumes can be identified by their range, or distance, from the radar site.

Each pulse volume contains a subset of the three basic data moments defined above: reflectivity, radial velocity and spectrum width. Each pulse volume also has the two positional attributes of range and azimuth. These two attributes plus the base three attributes make up the primary five attributes against which the classifiers learn [2].

The resolution of NEXRAD data depends on the build of the radar system at the time it was acquired. Until mid-2008, most radar sites were based on Build 9, which provided one degree of beam width and pulse volume lengths of 1000 m for reflectivity data and 250 m for Doppler data [34]. Beginning in 2008, NEXRAD stations started

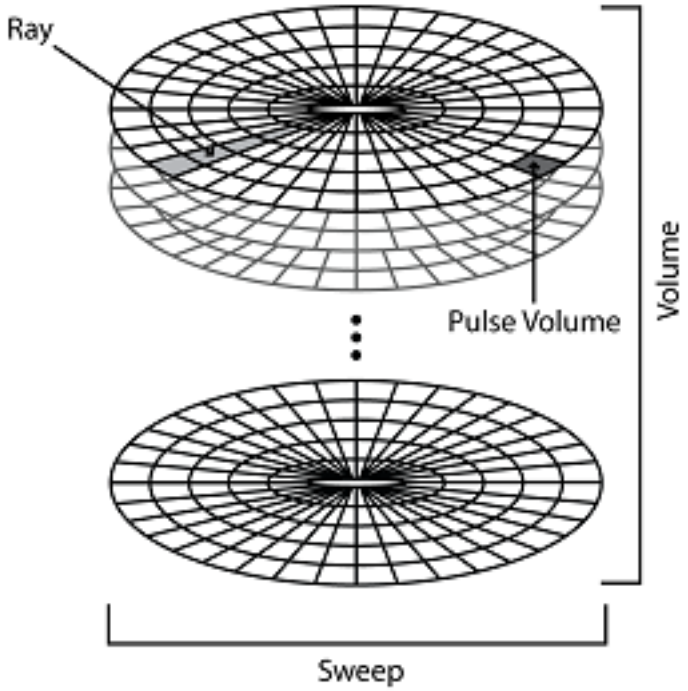


Figure 4: Data Organization For NEXRAD Volume Scans

moving to Build 10. Build 10 provides improved resolution by decreasing the beam width to 0.5 degrees and by making all pulse volumes 250 m in length [36]. Starting in 2010, NWS plans to start upgrading sites to Build 11 [37], providing support for dual polarization. This is expected to provide at least two more data moments that can distinguish between rain, sleet, hail and presumably different types of birds.

CHAPTER 3
METHODOLOGY

This chapter introduces a system that employs a two stage process for automating the identification of biological radar echoes. Pulse volumes are the units of interest in this system. Training and classification is performed at the pulse volume level, without the use of any sweep level information. First, expert classified training data is used to construct a database that will be used in all subsequent experiments. Second, classification or validation tasks are performed by retrieving data from the database and processing it with the classification and validation modules. The first stage is typically a onetime step while the second stage is performed multiple times for various data subsets and experiment designs. New training data is often loaded into the database as it becomes available using the first stage of the process.

The first stage of this process is illustrated in Figure 5. It begins with raw Level II NEXRAD data from the NCDC data service [17]. This raw data has a metadata file associated with it that contains classification information obtained from radar ornithology experts or by direct observation by field agents. Both of these files are read by the data parser. The data parser understands the compression algorithm and underlying binary format of the raw data. It is capable of reading both Build 9 and the new Build 10 NEXRAD data formats [34, 36]. All the relevant information is extracted from the raw file and loaded into a unified data structure. This abstracted data is then tagged with the sweep or pulse volume level classification information contained in the metadata file.

After the data has been parsed, it then enters the first stage of preprocessing. This step performs all experiment invariant preprocessing, such as data cleanup and feature

extraction. Preprocessing is covered in more detail in the following section. Once the initial preprocessing has been performed, the data is sent to the Data Uploader for storage in a MySQL database. The Data Uploader understands the schema for the NEXRAD database and performs a mapping between long term persistent storage and the data structures used for short term experiment tasks.

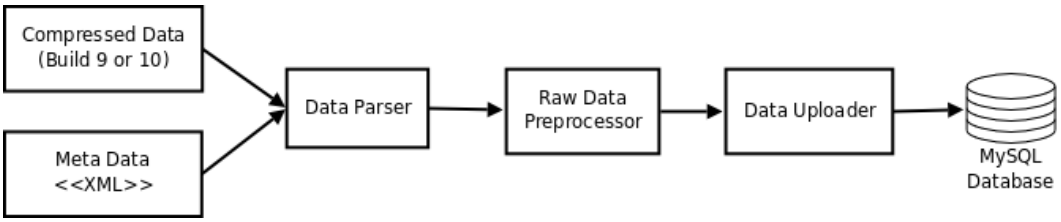
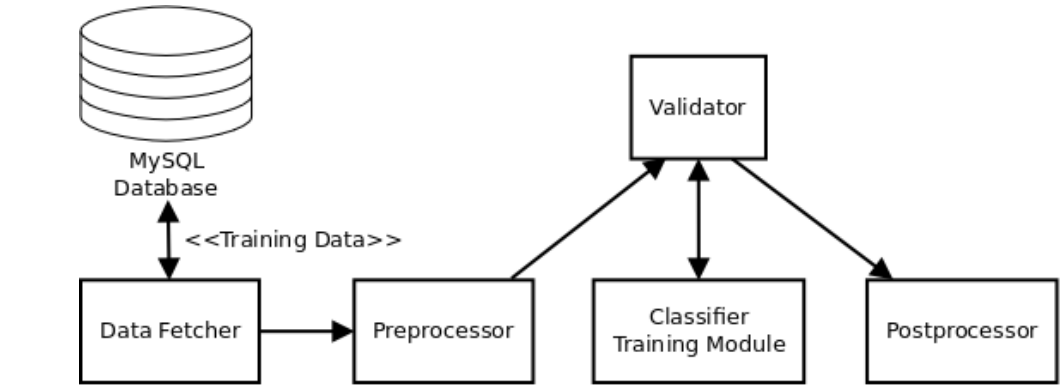


Figure 5: Stage 1: Initial Data Processing

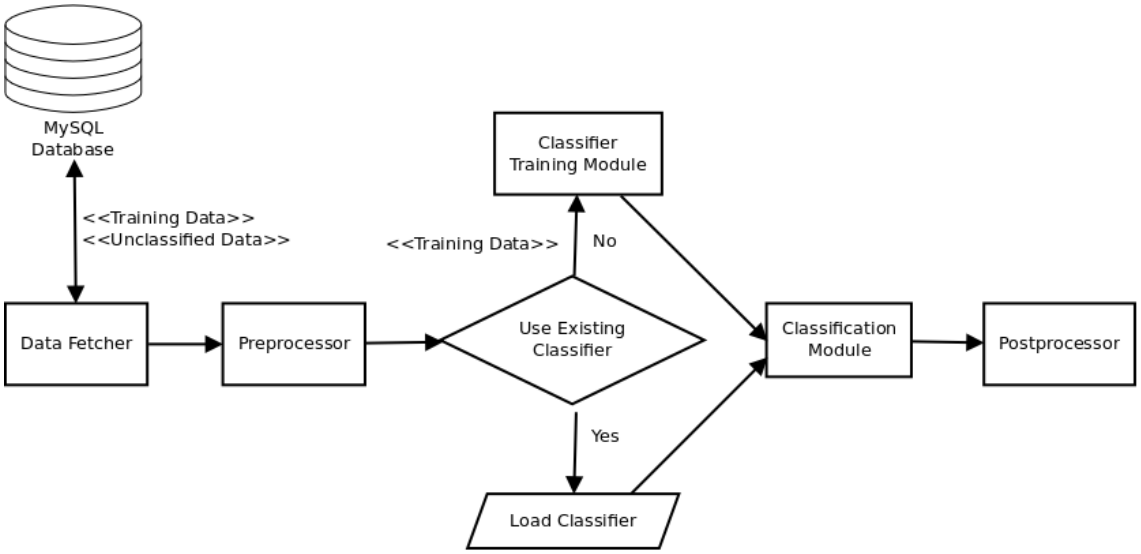
Classification and validation tasks are performed in the second stage of the process. As shown in Figure 6, this stage begins by retrieving the specific data necessary for the current task from the database using the Data Fetcher. The Data Fetcher is similar to the Data Uploader, but moves data in the opposite direction, querying the database and constructing appropriate data structures from the query results. Next, the data enters the second stage of preprocessing. This stage performs all experiment specific preprocessing, including feature selection and transformation (e.g. discretization).

After the second stage of preprocessing, the data is ready to be consumed by the classifier. This takes one of two forms depending on the type of task being performed. If a validation task is being performed, as in Figure 6(a), the only type of data being used is training data and the Validator must separate this into separate training and testing datasets. The other type of task, shown in Figure 6(b), is a classification task, in which the system is simply being used to classify data of an unknown type. In this case, the data is separated into classified and unclassified datasets when it comes

out of the database. The classifier is either loaded from disk or is trained using the classified data, after which the unknown data can be classified.



(a) Validation



(b) Classification

Figure 6: Stage 2

After classification and validation, the data is optionally sent to the Postprocessor. The Postprocessor simply takes the results of the tasks and converts them into a more useful form. This usually includes creating a text or html report and possibly graphically rendering the classified sweeps.

The steps of preprocessing, classification and validation are explained in more detail below.

Data Preprocessing

Before classifier training can take place, a series of steps must be performed on the data to improve the quality of the classifier produced. Most of these steps involve removing data that is unnecessary, untrustworthy, or corrupt.

As mentioned above, data from the NEXRAD system comes in the form of a volume scan, so the first step is to remove from the volume scan the pulse volumes corresponding to all elevation angles that are of no interest. It is unlikely that any significant quantity of birds will be found in all but the bottommost sweep when investigating most local bird movements. Therefore, the 0.5° scan is saved and the rest are discarded. The second step is to remove pulse volumes that are known to cause problems. This includes pulse volumes within 20 km of the radar as well as pulse volumes further than 145 km from the radar. The pulse volumes closest to the radar are removed because they tend to contain the most contamination from ground clutter [38]. Pulse volumes more distant than 145 km are removed due to range-velocity ambiguity, also known as the Doppler dilemma [39]. As a result of this property of Doppler radars, radial velocity readings beyond a certain distance are not guaranteed to be accurate. Therefore, ambiguous data is simply discarded.

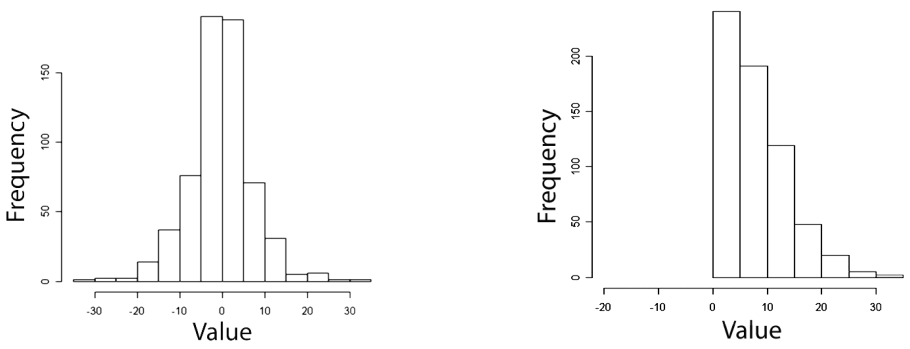
The third step is to realign adjacent scans within the same elevation angle. As mentioned in the previous chapter, WSR-88D radars often perform multiple scans at a particular elevation angle to improve the quality of information produced. Because lower elevation angles tend to have more interference from ground clutter, it is important to get accurate radial velocity readings so that clutter removal algorithms do

a better job of cleaning up the signal. Therefore, the radar will typically perform one scan at a certain pulse repetition frequency to accurately find the position of sensed entities and then performs one or more scans at a different pulse repetition frequency to get accurate velocity values. Realignment becomes an issue because the radar is a mechanical device and will not sweep exactly 360°. This causes the azimuths of the rays to be slightly different. This is not a problem when only dealing with reflectivity and Doppler data separately, but it is an issue when conceptually thinking of one pulse volume as containing all three data moments. In this case, the misaligned scans must be sampled to produce accurate composite estimates for the conceptual pulse volume.

The next step is to remove pulse volumes with reflectivity values that have been flagged as bad or range-folded. A bad value, simply means that the signal strength of the received echo is less than the signal-to-noise ratio. Generally, this simply indicates that a particular pulse volume is empty. Range-folded values occur when echoes for a radar pulse return to the radar after the next pulse has been sent out.

The fifth step is to mitigate the effects of bad or range-folded radial velocity and spectrum width values. Because these attributes (collectively known as the Doppler attributes) are typically acquired in a separate scan from the reflectivity data, it is possible to have corrupted Doppler data for a location that contains valid reflectivity data. Tests have shown that reflectivity is the most useful attribute for classification, so the decision was made to simply set the radial velocity and spectrum width attributes to zero when they are corrupted if valid reflectivity data is present. By doing this, the algorithm can still exploit the reflectivity data even if the lesser attributes are missing or corrupted.

The sixth step enhances the base set of attributes by adding nine second order moments that are derived from the primary attributes. These derived attributes are



(a) High Kurtosis (b) High Skewness

Figure 7: Skewness and Kurtosis Examples

statistics that describe the neighborhood of the pulse volume. For each of the three base attributes, variance, skewness, and kurtosis are calculated over a 25x25 pulse volume neighborhood [40]. Variance is the familiar measure from classical statistics and simply measures the variation of values within the sampled neighborhood. Kurtosis measures the peakedness of a given distribution. Skewness, as the name implies, is a measure of skew in a distribution. To illustrate these statistics, Figure 7(a) provides an example of a distribution with a high kurtosis value and Figure 7(b) provides an example of a distribution with a high amount of skew.

Haykin et al. [29] calculated skewness and kurtosis statistics for the raw in-phase (I) and quadrature (Q) channel samples from an L band radar. In modern digital radar systems the Doppler data is provided in digital form with real (I) and imaginary (Q) components [13]. The idea was to use these statistics to measure the deviation from Gaussianity. In that study, the statistics were calculated for a single pulse volume in the same way that mean radial velocity is calculated. Unfortunately, the Level II WSR-88D data being used by this system does not contain the original I and Q channel samples. Instead, these statistics are calculated on a neighborhood of pulse volumes with the idea that they might help detect boundaries and other phenomena.

Initial experiments showed these derived features to be beneficial for classification, so they were permanently added to the process.

Window size was chosen by maximizing information gain for the derived attributes while varying the size of the neighborhood from 3x3 to 25x25. A 25x25 window produced the best results. It is possible that a larger window could further improve classification accuracy, but due to computational considerations, a 25x25 window was the largest size investigated.

The final step is to remove any attributes deemed insignificant or detrimental, as determined by sensitivity testing. This can include derived or base attributes. This step may or may not be applied depending on the experiment and is conditionally applied depending on the classifier used. This is necessary because some classifiers do not weight attributes by significance. As a result, insignificant features can actually confuse some classifiers, such as K-Nearest Neighbor and naïve Bayes, by diluting the effects of more significant features.

Classification

Classification is performed at two different levels. At the lowest level, machine learning techniques are used to classify non-empty pulse volumes as either biological or nonbiological. These classifications are then aggregated to provide a classification at the sweep level. Classification at the pulse volume level is performed by selecting the class with the highest probability (provided by the classifier). Sweep level classification is slightly different. Instead of using a majority rule to classify the sweep, a threshold is chosen. Only sweeps that contain a certain percentage of biological echoes are classified as biological. A 70% threshold was used for the experiments described in Chapter 4. The purpose of using a threshold is to bias the system towards making

errors of omission rather than errors of comission. If the purpose of this system is to reduce the number of radar sweeps that must be manually examined, it makes sense to bias the system against flagging nonbiological sweeps as biological.

Three classifiers were investigated in this thesis: K-Nearest Neighbor, Naïve Bayes, and Neural Networks. Each machine learning technique has pros and cons that are discussed in the following sections.

K-Nearest Neighbor

The K-Nearest Neighbor classifier operates on the assumption that data instances belonging to the same class will have similar attribute values. Therefore, it treats attribute values as coordinates in N-dimensional space, where N is the number of attributes [41]. In the context of the preprocessed data described previously, N can be as high as fourteen when including the three base attributes, two spatial attributes and all nine derived attributes. Given a set of correctly classified training instances, classification of an unknown instance can be achieved by simply finding the K nearest training instances, where K is some integer greater than or equal to 1, and using their respective classifications to assign a class to the new instance. Figure 8 shows an example using two dimensions. There are a wide variety of variations on the basic K-Nearest Neighbor classifier using different distance metrics and different schemes for determining a class given the classes of the nearest neighbors, but most use this same basic process.

The K-Nearest Neighbor classifier is a lazy learner which means that the bulk of the required computation is done during classification. This differs from the other two classifiers investigated, which do most of their computational processing during training. Training for the K-Nearest Neighbor classifier is a bit of a misnomer because it is generally limited to just storing the training instances.

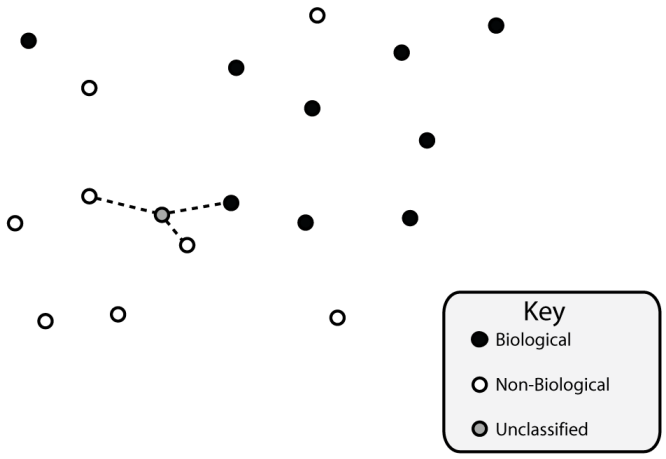


Figure 8: K-Nearest Neighbor

Although this classifier is intuitive and fairly easy to implement, it suffers from a number of shortcomings [42]. To start with, the K-Nearest Neighbor classifier is sensitive to noisy training data. Unfortunately, much of the training data used in this thesis is noisy due to the nature of radar data. Another issue is that the classification accuracy of this classifier can degrade substantially with the inclusion of irrelevant or low value attributes. The reason is that a greater number of attributes increases the dimensionality of the feature space which leads to a lengthening of the distance between related data instances. The last, and perhaps most detrimental, shortcoming is that the computational demands are sensitive to the number of training points as well as the number of attributes. For each new data instance that needs to be classified, distances to a large subset or possibly the whole set of training instances must be calculated. These distance calculations also become more costly as the number of dimensions rises.

The experiments in this thesis used the IBk classifier provided by the Weka machine learning library [43]. This K-Nearest Neighbor implementation was used with the classic Euclidean distance function.

Naïve Bayes

The Naïve Bayes classifier [44] is a simple yet surprisingly effective classifier based on Bayesian inference. At its core, the classifier leverage Bayes' theorem:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \tag{1}$$

This theorem is usually used in the context of calculating the probability of an event A given some evidence B. In this context, such a probability can be found by multiplying the probability of event A by the probability of seeing evidence B, given event A, and dividing by the probability of evidence B. In the context of classifying pulse volumes, this theorem is used to calculate the probability of a pulse volume belonging to a given class given its set of attributes. This can be done using Bayes' theorem by calculating the probability of seeing the set of attributes for the particular class, the probability of seeing the given class, and the probability of seeing the set of attributes. All three of these probabilities can theoretically be computed from the training data [44].

Unfortunately, calculating the joint probability for a set of attributes can be computationally expensive and requires a massive amount of training data. The training data used in this thesis can contain up to 14 multivalued attributes and two classes. Intuitively, it might be unlikely to have even one training instance with exactly the same combination of attributes, and many such training instances would be needed for each class in order to accurately estimate the probability of seeing that particular combination. This combinatorial explosion makes using true Bayesian inference infeasible for most practical applications.

In order to avoid this computational overhead, the Naïve Bayes classifier makes the assumption that the attributes are conditionally independent of each other. This

assumption allows the complex probability to be reduced to a product of probabilities for the individual attributes as shown in Equation 2.

$$P(D|A,B,C) = \frac{P(A|D) * P(B|D) * P(C|D) * P(D)}{P(A,B,C)} \tag{2}$$

For classification, only maximization of class probabilities is required. It is only necessary to discover which potential class has the highest probability given the attributes. Determining the actual probabilities as shown in Equation 2 is not required. Instead, only the numerator is required and the denominator can be removed. This is valid because the denominator is invariant when only changing the class D. Given this reduction the maximization problem given in Equation 3 can be rewritten as Equation 4 [44].

$$Y = \text{MAX}(P(D|A,B,C)) \tag{3}$$

$$Y = \text{MAX}(P(A|D) * P(B|D) * P(C|D) * P(D) \tag{4}$$

The conditional independence assumption is where the "naïve" portion of the name comes from. Although this is a rather profound assumption in most practical applications, studies [45] have shown that in many cases it has little effect on the outcome. The simplicity and effectiveness of this classifier [41] makes it an obvious benchmark for classification experiments. The disadvantage of using this classifier is that it uses a purely probabilistic model and by definition the conditional independence assumption means that it cannot exploit any relationships that may exist between attributes.

Naïve Bayes does not weight attributes by importance so it can suffer from degradation of performance when attributes are included that aren't helpful in predicting a

pulse volume's class. For experiments that use this classifier, it is important to make use of the preprocessing step that removes unnecessary attributes.

In order to determine which attributes should be included when using the Naïve Bayes classifier, a version of greedy search [46] is performed. First, information gain [44] values are calculated for each of the attributes. These information gain values determine the order in which attributes are added. Starting with the attribute with the highest information gain, a series of experiments are performed and the average accuracy is recorded. The next attribute is then added and the previous step is repeated. This continues until the average accuracy starts to decrease. At this point the search is stopped and the attributes that have already been added to the evaluation set become the subset of attributes used in later experiments.

Table 1: Attribute Subset Selection

Attribute	**Accuracy**
Reflectivity	81.25%
Reflectivity Skewness	87.50%
Spectrum Width	87.50%
Radial Velocity Skewness	87.50%
Range	87.50%
Radial Velocity	87.50%
Reflectivity Kurtosis	93.75%
Reflectivity Variance	62.50%

Table 1 shows the results of the attribute selection procedure just described. The average accuracy remains the same or rises while adding the first seven attributes. When the eighth attribute is added the accuracy takes a large dip. This triggers the cut off criteria and only the first seven attributes are used. It may seem strange that the accuracy remains the same while adding the middle four attributes. This is not entirely unexpected because the statistic we are using for our greedy search, information gain, is just a heuristic and does not directly correlate with accuracy.

Therefore, even though the accuracy does not rise while adding these attributes, there is no guarantee that the accuracy will not rise later in the procedure. This is evident in this example by the increase in accuracy resulting from inclusion of reflectivity kurtosis. It is true that a similar or possibly better accuracy could possibly be achieved by leaving out some of the attributes that show no improvement, but that would require a more computationally expensive search than the greedy search that was used for this procedure. This may be a possible cause for further research in the future.

The classifier implementation used in this investigation was the Naïve Bayes classifier provided by the Weka 3 machine learning library [43].

Neural Network

Neural networks [47, 48] are a popular machine learning technique used in real world applications [49]. Although neural networks may require more overhead in the implementation and training stages, they tend to be fast and effective for many practical applications. They also lend themselves to parallelization and hardware implementations. This framework, based loosely on the neural networks that make up the human brain, can quickly be trained to map input patterns to output patterns. By repeatedly iterating over a set of training examples, the neural network can learn to emulate even complex non-linear functions [50].

Neural networks are composed of a number interconnected processing units called neurons. Neurons take a set of inputs and produce a single output. The neurons maintain a set of weights for each of the inputs and the behavior of the neuron can be changed by manipulating these weights. The neuron functions by summing or multiplying the weighted inputs and then passing them through a thresholding

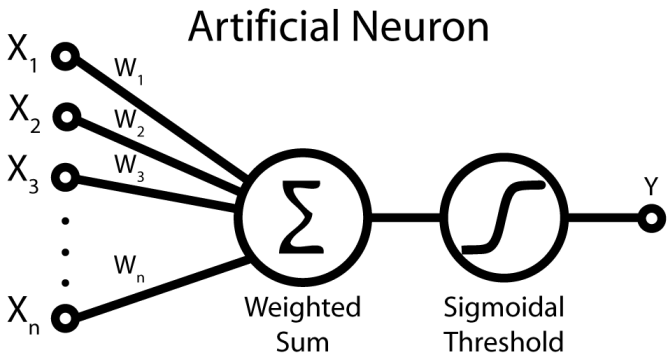


Figure 9: Artificial Neuron

function, such as the sigmoid given in Equation 5. Figure 9 shows the components of an artificial neuron.

$$S(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

The way that neurons are connected can have a significant effect on the operation of the network. As such, there are many different possible architectures for a neural network and some architectures are more effective for a given problem domain than others. A survey of the different types of architectures is provided by Rojas [47].

A simple feed forward network architecture is used in this system. Figure 10 illustrates a typical layout for a feed forward neural network. The neurons in this type of network are organized into layers. Typically, there is an input layer, an output layer and one or more hidden layers. Starting from the input layer, each neuron is connected to every neuron in the adjacent layer.

The power of a neural network is realized when an appropriate learning algorithm is employed to learn the weights that correctly produce the desired output given a specific input. Backpropagation [51, 52] is a common learning algorithm for this task.

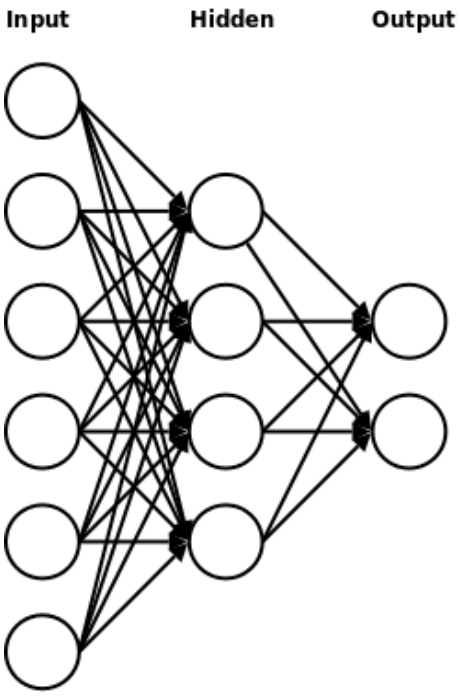


Figure 10: Feed Forward Neural Network

This algorithm uses a gradient descent strategy that calculates derivatives in order to propagate errors back through the network and update weights accordingly. Back-propagation gives these networks the ability to automatically learn complex functions simply through repeated exposure to training examples.

For the problem being investigated, the pulse volume attributes serve as the inputs to the neural network and the classification (1 for biological, 0 for nonbiological) is the output. It should be noted that “1” and “0” are simply nominal class values. The internal mechanics of the neural network use the more standard -1 to 1 scale for outputs.

The disadvantages to using a neural network include (1) increased implementation overhead, (2) difficulty in deducing any decision making knowledge from inspection

of the weights, and (3) the network can take a prohibitively long time to converge to a solution on complex problems [44].

Despite these setbacks, neural networks have found their way into a wide range of real world applications because they can learn complex relationships between the attributes used as inputs and once they have been trained, classification can be performed quickly [49].

The backpropagation network implemented by the multilayer perceptron classifier in Weka [43] was used for this investigation.

Validation

As with any classification problem, validation is of the utmost importance. Most validation schemes involve separating the data into a training set and a testing set. The classifier is constructed using the training data and then its accuracy is validated using the testing set. This type of scheme can help avoid the situation where the classifier simply memorizes the training data. Classifiers that fall into this trap will appear to perform well when tested against the training data but usually perform poorly on new data. This simple scheme does not work well, however, when the amount of data is limited because using too much data for validation does not leave enough training data to produce a good classifier; and, using too little data does not do a good job of validating the classifier's accuracy.

Cross validation [53] offers a good alternative to the simple validation scheme described above. With cross validation, the available data is evenly divided into a number of subsets known as folds. Validation then occurs in an iterative fashion. On each iteration, one fold is held out for validation and the remaining folds are used for

training. At the end, each fold has been held out exactly once, and the results are averaged together to obtain an overall accuracy.

All of the experiments in this thesis use tenfold cross validation, but it is employed in two different ways. The original dataset was classified at the sweep level, so the 40 training sweeps are divided into ten folds, each containing four sweeps. More recent experiments, have been working with goose data that is actually classified at the pulse volume level by workers in the field. For these experiments, the training dataset consists of 1352 pulse volumes, so each fold contains 135 pulse volumes.

A related factor that can have an effect on the quality of the classifier is the balance of classes within the data. An unbalanced dataset can lead to results that are not representative of the classifier's actual effectiveness. For instance, if 80% of the instances in a dataset belong to class 1, a classifier could simply classify every instance as class 1 and it would appear to have an accuracy of 80%. Obviously, this classifier could produce widely varying results when used on real-world data.

To avoid balance issues, random over sampling (ROS) and random under sampling (RUS)[54] are employed to balance the dataset. This scheme randomly removes instances of the majority class and randomly duplicates instances of the minority class. This procedure works well for correcting mildly unbalanced datasets, but highly unbalanced datasets can still lead to problems due to the lack of diversity in the resulting minority class. The unbalanced training data used in Chapter 4 contains 30% to 40% nonbiological training examples and 60% to 70% biological echoes.

In order to truly have a balanced validation procedure, it is also important to make sure that the resulting folds have roughly the same distribution of classes as the composite dataset (the dataset resulting after ROS and RUS have been performed). Stratification is a process that randomly swaps training examples between folds until

all folds have the same class distribution. Stratified tenfold cross validation is used whenever possible.

Software Source Code and Documentation

The process described above was implemented in Java and leverages the Weka machine learning library [43]. All source code and documentation for this project is available as part of the nexrad-mltoolkit project on Google Code [55].

CHAPTER 4

RESULTS

In the following sections, empirical results are provided for a number of experiments using this system in several different contexts. The results for the original set of training data, which will henceforth be referred to as the *Passerine Dataset*, are provided in the first section. This dataset contains a set of classified sweeps for Lincoln, IL and surrounding areas. The training data is classified at the sweep level by experts with experience identifying biological echoes in Doppler radar sweeps. Then results for geese data, which will be referred to as the *Geese Dataset* are provided. The majority of this data was collected around Aberdeen, SD and Hastings NE. This training data is classified at the pulse volume level and has been provided by researchers in the field. Finally, the results of applying this process on a large scale are briefly discussed in the third section. Radar data for an entire fall and spring migration season was collected for three sites including Milwaukee, WI, Green Bay, WI, and Great Falls, MT. This data is referred to as the *Migration Dataset* in the following sections. For these experiments, the classifiers provided by the Weka machine learning framework [43] were used.

Passerine Data

The Passerine Dataset originally contained 40 training sweeps provided by Dr. Robb Diehl, University of Southern Mississippi. This data was classified at the sweep level. Each sweep was a prototypical example of either biological or nonbiological echoes. Theoretically, none of the sweeps contained a mixture of both echo types. Accuracy and time estimates for the three classifiers investigated are provided in

Table 2. Typically, timing results are not included in academic publications, because they can be affected by so many outside influences. These timing results are provided simply as anecdotal evidence of the difference in computation between the K Nearest Neighbor classifier and the other two.

Table 2: Original Diehl Dataset

Classifier	**Correctly Classified Sweeps**	**Experiment Duration**
K Nearest Neighbor	39 / 40	191 min
Naïve Bayes	39 / 40	5 min
Neural Network	40 / 40	17 min

This dataset was eventually supplemented with more sweeps from Dr. Diehl as well as some sweeps from Mr. Manuel Suarez of the USGS. After the dataset was balanced using ROS / RUS, the new dataset contained 100 classified sweeps. With the additional training data, it became apparent that using K Nearest Neighbor was infeasible due to its high computational cost. Therefore, experiments after this point focused on comparing Naïve Bayes with the Neural Network. The results for the new dataset are shown in Table 3.

Table 3: Expanded Diehl Dataset

Classifier	**Correctly Classified Sweeps**
Naïve Bayes	83 / 100
Neural Network	96 / 100

Geese Data

The focus of the Geese Dataset is to facilitate a set of experiments for which ground truth is known. Unlike the sweeps in the Passerine Dataset which are classified at the sweep level by radar ornithologists, the Geese Dataset is classified at the pulse volume level. Biological classifications are drawn from direct observations of geese in the

airspace recorded by researchers in the field. This data includes latitude, longitude, altitude (AGL), number of geese, direction of flight, and weather. Nonbiological training instances are collected for the same geographic area by utilizing historical weather data to find precipitation events occuring at times when birds are not present. Most of these observations come from the Aberdeen, SD area. This dataset is the first to almost exclusively use the new higher resolution data provided by NEXRAD Build 10. Performing ROS / RUS produced a balanced dataset containing 1352 pulse volumes. The results of ten test runs have been averaged to provide the confusion matrices in Table 4 and Table 5.

	Predicted Class	
Actual Class	Nonbiological	Biological
Nonbiological	619.9 (91.7%)	56.1 (8.3%)
Biological	284.9 (42.14%)	391.1 (57.86%)
Accuracy = 74.9%		

Table 4: Naïve Bayes Classifier Results for Goose Dataset

	Predicted Class	
Actual Class	Nonbiological	Biological
Nonbiological	565.1 (83.59%)	110.9 (16.41%)
Biological	183.2 (27.1%)	492.8 (72.9%)
Accuracy = 78.3%		

Table 5: Neural Network Classifier Results for Goose Dataset

The confusion matrix [56] allows us to see the number of correct classifications as well as the number of false positives and false negatives. These results show slightly lower accuracy than that achieved with the passerine data. This could be the result of additional noise resulting from geolocation problems. One further point of interest is that the neural network not only has a higher overall classification accuracy, but it also has a much better classification accuracy for biological echoes in particular.

Two Season Migration Data

The Migration Dataset is intended to test this algorithm's ability to process large amounts of data. Hourly sweeps were acquired for Milwaukee, WI, Green Bay, WI, and Great Falls, MT spanning a time period of 180 days split between fall 2007 and spring 2008. This dataset is not intended to test the classifier's accuracy, as it is composed entirely of unclassified data. Never-the-less, a subset of the results can still be examined to get an idea of how the classifier is performing.

Unlike earlier classification tasks, it does not make sense in this case to classify every sweep as biological or nonbiological. Some sweeps are bound to be empty or nearly empty and giving such a sweep either classification is misleading. Therefore, a threshold of at least 10,000 non-empty pulse volumes must be reached for a sweep to receive a classification, otherwise it is simply classified as "No Data". This threshold was chosen after considering pulse volume count histograms for both the Passerine and Migration datasets. For reference, the average number of non-empty pulse volumes in a biological sweep from the Passerine Dataset is almost 37,000.

This dataset includes 12,437 sweeps. A neural network was trained using the passerine data and the total time to classify all sweeps was 78 minutes. The breakdown for classifications is provided below.

This breakdown is plausible for what might be expected during a seasonal migration. Almost half of the sweeps show little to no activity with the other half containing roughly two-thirds biological with one-third nonbiological.

Only the Milwaukee migration data is considered for further analysis below because it covers an overlapping area with Green Bay and this particular area has been studied by a number of experts. These experts can provide valuable information, such as time frames when large migration should be detected. The breakdown for

(a) Composite

KGRB, KMKX & KTFX

Classification	**Sweeps**
No Data (Empty)	5886
Nonbiological (Precipitation)	2150
Biological	4401

(b) KMKX

KMKX (Milwaukee)

Classification	**Sweeps**
No Data (Empty)	1739
Nonbiological (Precipitation)	485
Biological	1989

Table 6: Migration Data Class Distribution

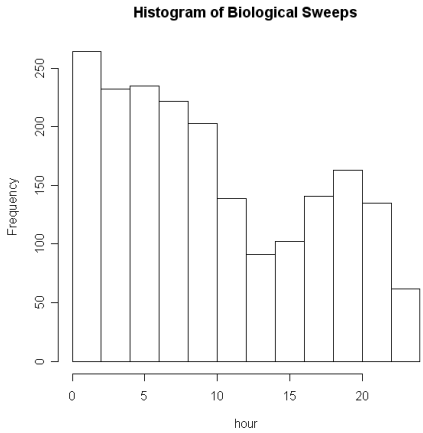
Milwaukee is shown below. The breakdown is similar to the composite figures above, and seems plausible.

The histograms for hour of sweep acquisition can also provide insight into the performance of this technique. Figure 11 provides the histograms for biological, non-biological, and no data classifications.

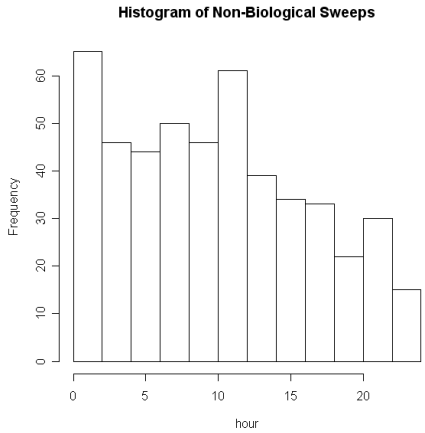
The first obvious observation in all three graphs is a departure from uniformity or normality. This is important because it is not what one would expect to see if the difference between these classifications and a random set of classifications was not statistically significant.

Looking specifically at the biological histogram, another interesting characteristic is observed. A large portion of the biological sweeps fall in the 4:00 to 5:00 GMT range. This is consistent with expert opinions regarding frequent migration event times [38].

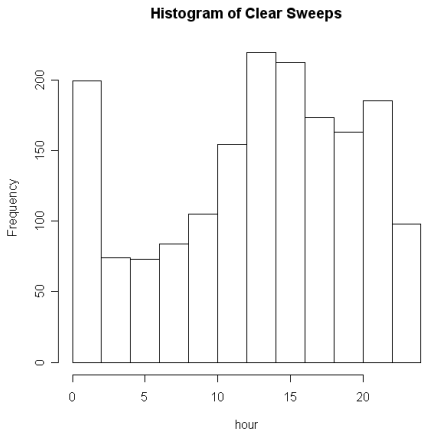
Another interesting statistic to examine is the frequency of class change when considering sweeps in chronological order. One would expect sweeps before or after



(a) Biological



(b) Nonbiological

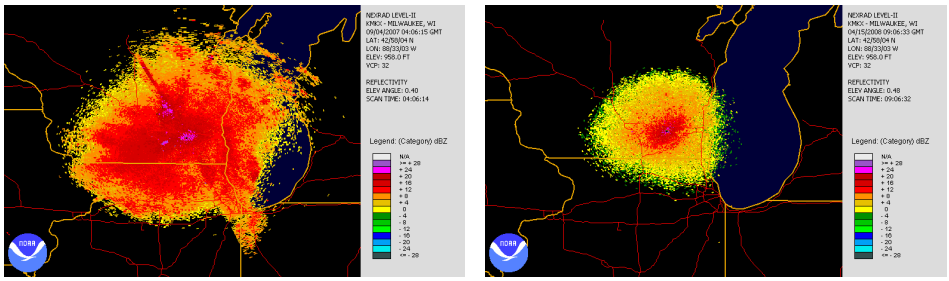


(c) Clear-Air

Figure 11: Sweep Hour Histograms

a sweep with a particular classification to have a higher probability of sharing that classification.

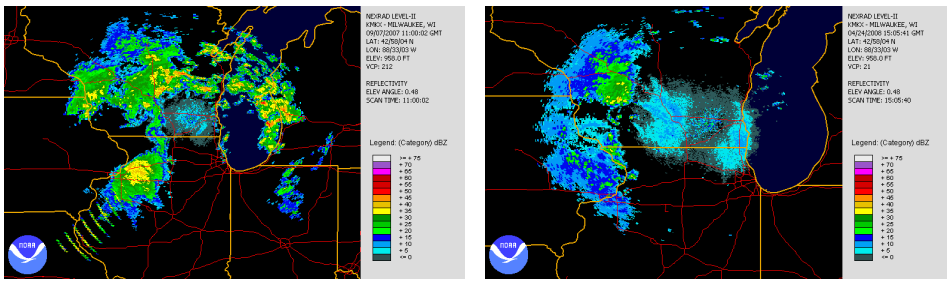
Counting the number of times the classification changes for the Milwaukee data produces a frequency count of 599 for 4213 sweeps. This means that there are long stretches of uniform classification, which is what one would expect to see for long migration or precipitation events. For reference, the average frequency count for the same amount of randomly generated classes is about 2800. This is further evidence that the algorithm is, indeed, identifying pulse volumes associated with birds.



(a) Sample 1 (b) Sample 2

Figure 12: Example Biological Sweeps

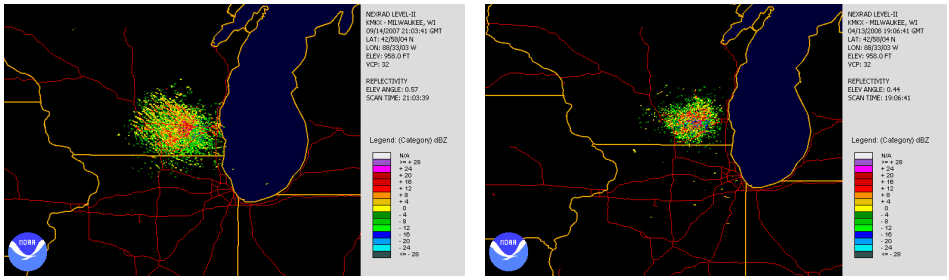
As a final investigation tool, several sweeps can be rendered and visually inspected to validate the classification. Figures 12 through 14 show example sweeps selected from each classification. These examples were randomly selected.



(a) Sample 1 (b) Sample 2

Figure 13: Example Nonbiological Sweeps

The sweeps classified as biological in Figure 12 show the tell-tale signs of bird activity [38]. The elliptical shape and distribution apparent in this sweep is typical for bird events. This is in contrast to the nonbiological sweeps seen in Figure 13. These sweeps have irregular shapes which are typical of precipitation events. Finally the empty sweeps shown in Figure 14 are indeed mostly empty, containing only a small number of echoes close to the radar site. These echoes are probably from clutter that is present at lower altitudes near the radar.



(a) Sample 1 (b) Sample 2

Figure 14: Example Clear Air Sweeps

The examples in Figures 12 through 14 are admittedly anecdotal, but they show that this technique appears to be working in the desired manner. This particular experiment demonstrates that this system is able to process large amounts of real data in an acceptable amount of time.

CHAPTER 5
CONCLUSION

Prior to this research, there were no published accounts of applying artificial intelligence methodology to identification of birds in NEXRAD weather radar data. Now, the process described makes it possible for researchers to utilize the wealth of new and archived weather radar data in a new, automated way. The goal of this research was to produce a system that could substantially improve upon the time required for radar ornithologists to manually classify the echoes contained in radar sweeps. This goal has been achieved.

After being trained, both the Naïve Bayes and Neural Network classifiers are capable of classifying all the pulse volumes contained in a sweep in a matter of seconds. As seen in the previous section, the data from an entire migration season for one station can be processed in a matter of hours.

Furthermore, detecting biological echoes in NEXRAD radar data with a classification accuracy in the mid 80^{th} to mid 90^{th} percentile range is feasible using machine learning techniques. The comparison of classification techniques revealed that both the Naïve Bayes classifier and the Neural Network are suitable candidates for real world application. Although Naïve Bayes had a slight edge in computational performance, most of those savings were in training time, which should only need to be done a few times in a practical application. In terms of accuracy, the Neural Network produced better results. All things considered, the Neural Network would seem to be the better choice for an automated data classification system.

Limitations and Future Directions

Although the results provided are encouraging, the system described in this thesis suffers from several limitations:

- Restrictive training data format
- Coarseness of output
- Individual pulse volume geolocation issues

Training Data Format

The initial classification scheme for training data was restrictive in that classifications were made at the sweep level. This constraint limited the amount of potential training data, because radar sweeps rarely contain only one type of echo. Even the sweeps that have been designated as only biological or only nonbiological probably suffer from a certain amount of noise caused by other echo types.

Another potential issue stems from the use of these sweeps for validation. Sweeps that are prototypical examples of biological or nonbiological events are understandably the easiest to classify and this might have resulted in overly optimistic results.

In the future, a tool will be needed that allows experts to classify portions of a sweep. This could be accomplished with a simple graphical interface that allows a user to draw polygons around certain regions of a sweep. A more advanced tool might leverage image processing techniques to automatically segment a sweep. In this scenario, an expert would be prompted to provide classifications for each of the identified regions.

Coarseness of Output

This issue is closely tied to the issue described in the previous section. Because the training sweeps used for validation were classified at the sweep level, the system was initially designed to aggregate pulse volume classifications to produce a sweep level classification. However, this is fairly restrictive because only allowing entire sweeps to be classified as biological or nonbiological is unlikely to be useful in most real applications.

Improvements in this area would provide finer details for classified sweeps such as number and size of biological regions in addition to the percentage of echoes classified as biological. This information could then be combined with a graphical rendering to produce more useful results.

Pulse Volume Geolocation

The ground truth provided by the observations from the Geese Dataset adds further credibility to this process, but many obstacles to utilizing these observations were discovered. Resolving these issues will be part of the focus of future research. The first issue is calculating the beam height at the location of the observation. Any observed birds must be flying within the beam for the observation to be useful. In order to do this, several things are necessary: an accurate method of calculating the height of the beam, an accurate way to estimate the height above ground level of the birds and an accurate way to calculate the height of the observer. These calculations are further complicated by differences in the reported height above sea level given by different models of the earth [57].

Another issue is that most sweeps occur at roughly five minute intervals. Observers in the field have no way of knowing when the radar is actually detecting what they

are seeing. This becomes a problem when birds are flying at high speeds because they can cross pulse volume boundaries. Likewise, when birds are ascending or descending they can fly into or out of the beam. The issue of crossing pulse volume boundaries is exacerbated by the new Build 10 systems that offer finer resolution by decreasing the size of pulse volumes.

Finally, mapping a pulse volume to a geographical location can require some complex math due to the curvature of the earth and the angle of the beam. Intuitively, this problem is worse as an observer gets farther from the radar station. Current work on improvements to the system will help resolve the problems just discussed.

In addition to addressing the limitations described in the previous section, there are a number of aspects of this research that could be expanded upon:

- Independent validation of results
- Increased number of potential classes
- Incorporation of sweep level features
- Integration with GIS systems
- Application of System

Independent Validation of Results

More research is required to corroborate the results found in this thesis. Ideally, this corroboration would come from an independent study. One possible method of doing so would be to use this process to find areas of high bird use and then confirm these results using bird data collected from field researchers (similar to the Geese Dataset). Some form of independent validation is going to be required for this tool to gain acceptance in the broader scientific community.

Expanded Classification Types

Increasing the number of classes is one way that the system could be substantially improved. The current use of only biological and nonbiological classes is obviously restrictive. Possible class types could include rain, snow, insects, passerines, waterfowl, etc. The ability to differentiate between types of birds would be particularly useful for researchers with a particular focus, such as studying goose migration.

Sweep Level Features

Classification based on sweep level features is an obvious extension of this research. The process described in this thesis is purely based on pulse volume level classification. Although the process can produce classifications for entire sweeps, no sweep level features are considered when producing that classification. Instead, it is simply an aggregation of the classifications made for the pulse volumes contained in that sweep.

Possible sweep level information that could be useful for classification might include texture and distribution features for groups of echoes defined by segmentation or clustering. Another possible extension would be to feed entire neighborhoods of pulse volumes into the classifier. This would be an intermediate level of information, somewhere between sweep and pulse volume level depending on the size of the neighborhood.

GIS Integration

Another obvious question, that should be considered in the future, is how the process developed in this thesis can be utilized by researchers. Scientific data visualization will be necessary in order to present the information produced by this system in a way that can be understood by scientists and other interested parties. A logical next step would be to provide a mechanism, whereby the results of this process can

be utilized by GIS systems. This integration could allow maps to be made of hotspots or show how migration trends change over time.

Applications

This system could be used to advance understanding of avian behavior in a number of practical applications. Studying migration phenology and migration corridors are two examples of applications that are currently getting a lot of attention.

Climate change is currently a topic of great interest. One possible way of studying climate change is to look for changes in migration phenology occuring in the past several decades. Archived WSR-88D data could provide exactly the data needed for such a study. Using this system, researchers could identify shifts in migration timing resulting from changes in climate.

Identification of migration corridors is another practical application that could benefit from this system. In addition to interest from the ornithology field, a better understanding of migration corridors would benefit wind energy projects with an interest in reducing bird-turbine collisions. This could be done by either placing facilities in areas with less migration or by only operating the facilities during times that do not coincide with migration events.

REFERENCES CITED

[1] S.A. Gauthreaux Jr, C.G. Belser. Radar ornithology and biological conservation. *The Auk*, 120(2):266–277, 2003.

[2] R.H. Diehl, R.P. Larkin. Introduction to the WSR-88D (NEXRAD) for ornithological research. *Bird conservation implementation and integration in the Americas. Proceedings of the Third International Partners in Flight Conference*, pages 20–24, 2002.

[3] R.M. Mead, J. Paxton, R.S. Sojda. Identifying biological echoes in radar scans using machine learning. *Proceedings of the 4th Biennial Meeting of iEMSs*, 2008.

[4] J.M. Ruth, W.C. Barrow, R.S. Sojda, D.K. Dawson, R.H. Diehl, A. Manville, M.T. Green, D.J. Krueper, S. Johnston. Advancing migratory bird conservation and management by using radar: an interagency collaboration. *Open-File Report 2005*, 1173:1–12, 2005.

[5] G.H. Lowery Jr. A quantitative study of the nocturnal migration of birds. *University of Kansas Museum of Natural History*, 3:361–472, 1951.

[6] S.E. Eastwood. *Radar ornithology*. Methuen, 1967.

[7] P. Berthold. *Control of bird migration*. Springer, 1996.

[8] G.H. Lowery, R.J. Newman. Direct studies of nocturnal bird migration. *Recent Studies in Avian Biology, A. Wolfson, ed. Urbana: University of Illinois Press*, pages 238–263, 1955.

[9] B. Bruderer. The Study of Bird Migration by Radar Part 1: The Technical Basis*. *Naturwissenschaften*, 84(1):1–8, 1997.

[10] D. Lack, G.C. Varley. Detection of birds by radar. *Nature*, 156:446, 1945.

[11] S.E. Hobbs. A radar signal processor for biological applications. *Meas. Sci. Technol.*, 2:415–418, 1991.

[12] B. Bruderer. The study of bird migration by radar part 2: Major achievements. *Naturwissenschaften*, 84(2):45–54, 1997.

[13] M.I. Skolnik. *Radar handbook*. McGraw-Hill Professional, 1990.

[14] B. Bruderer, P. Steidinger. Methods oí Quantitative and Qualitative Analysis of Bird Migration with a Tracking Radar.

[15] S.A. Gauthreaux Jr. Weather radar quantification of bird migration. *BioScience*, 20(1):17–20, 1970.

[16] T.D. Crum, R.L. Alberty, D.W. Burgess. Recording, archiving, and using WSR-88D data. *Bulletin of the American Meteorological Society*, 74(4):645–653, 1993.

[17] National Climatic Data Center. NCDC NEXRAD Data Inventory. http://www.ncdc.noaa.gov/nexradinv/.

[18] S.A. Gauthreaux Jr, C.G. Belser. Displays of bird movements on the WSR-88D: Patterns and quantification. *Weather and Forecasting*, 13(2):453–464, 1998.

[19] J.E. Black, N.R. Donaldson. Comments on Display of Bird Movements on the WSR-88D: Patterns and Quantification. *Weather and Forecasting*, 14(6):1039–1040, 1999.

[20] R.P. Larkin, W.R. Evans, R.H. Diehl. Nocturnal Flight Calls of Dickcissels and Doppler Radar Echoes over South Texas in Spring (Llamadas durante vuelos nocturnos sobre Texas de Spiza americana durante la primavera y el eco en un radar Doppler). *Journal of Field Ornithology*, 73(1):2–8, 2002.

[21] S.A. Gauthreaux Jr, C.G. Belser. Bird movements on Doppler weather surveillance radar. *Birding*, 35(6):616–628, 2003.

[22] R.H. Diehl, R.P. Larkin, J.E. Black. Radar observations of bird migration over the Great Lakes. *The Auk*, pages 278–290, 2003.

[23] R.P. Larkin. Radar techniques for wildlife. *Techniques for wildlife investigations and management, 6th edition: Bethesda, Md., The Wildlife Society*, pages 448–464, 2005.

[24] S.A. Gauthreaux, C.G. Belser, C.M. Welch. Atmospheric trajectories and spring bird migration across the Gulf of Mexico. *Journal of Ornithology*, 147(2):317–325, 2006.

[25] J.W. Horn, T.H. Kunz. Analyzing NEXRAD doppler radar images to assess nightly dispersal patterns and population trends in Brazilian free-tailed bats (Tadarida brasiliensis). *Integrative and Comparative Biology*, 48(1):24, 2008.

[26] J.M. Wilczak, R.G. Strauch, F.M. Ralph, B.L. Weber, D.A. Merritt, J.R. Jordan, D.E. Wolfe, L.K. Lewis, D.B. Wuertz, J.E. Gaynor, i in. Contamination of wind profiler data by migrating birds: Characteristics of corrupted data and potential solutions. *Journal of Atmospheric and Oceanic Technology*, 12(3):449–467, 1995.

[27] S.A. Gauthreaux Jr, D.S. Mizrahi, C.G. Belser. Bird migration and bias of WSR-88D wind estimates. *Weather and Forecasting*, 13(2):465–481, 1998.

[28] K.W. Schulze. An investigation into the contamination of WSR-88D VAD wind profile output by migrating birds. 2004.

[29] S. Haykin, C. Deng. Classification of radar clutter using neural networks. *IEEE Transactions on Neural Networks*, 2(6):589–600, 1991.

[30] R. Kretzschmar, N.B. Karayiannis, H. Richner. Removal of bird-contaminated wind profiler data based on neural networks. *Pattern Recognition*, 36(11):2699–2712, 2003.

[31] M. Steiner, J.A. Smith. Use of three-dimensional reflectivity structure for automated detection and removal of nonprecipitating echoes in radar data. *Journal of Atmospheric and Oceanic Technology*, 19(5):673–686, 2002.

[32] W.F. Krajewski, B. Vignal. Evaluation of anomalous propagation echo detection in WSR-88D data: A large sample case study. *Journal of Atmospheric and Oceanic Technology*, 18(5):807–814, 2001.

[33] National Oceanic, Atmospheric Administration. NOAA Weather and Climate Toolkit. http://www.ncdc.noaa.gov/oa/wct/.

[34] *Data Documentation for DSI-6500 NEXRAD Level II*, 2005.

[35] G.E. Klazura, D.A. Imy. A Description of the Initial Set of Analysis Products Available from the NEXRAD WSR-88D System. *Bulletin of the American Meteorological Society*, 74(7):1293–1311, 1993.

[36] *Interface Control Document for the RDA/RPG*, 2008.

[37] R.E. Saffle, I. Noblis, F. Church, V.A.M.J. Istok, G. Cate. 10B. 1 NEXRAD PRODUCT IMPROVEMENT–UPDATE 2009. 2009.

[38] R.H. Diehl.

[39] R.J. Doviak, D.S. Zrnic. *Doppler radar and weather observations*. Dover Publications, Mineola, New York, edition second, 2006.

[40] D.N. Joanes, C.A. Gill. Comparing measures of sample skewness and kurtosis. *The Statistician*, pages 183–189, 1998.

[41] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. McLachlan, A. Ng, B. Liu, P.S. Yu, i in. Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1):1–37, 2008.

[42] D.W. Aha, D. Kibler, M.K. Albert. Instance-based learning algorithms. *Machine learning*, 6(1):37–66, 1991.

[43] E. Frank, I.H. Witten. *Data Mining: Practical machine learning tools and techniques with Java implementations*. Morgan Kaufmann, 2005.

[44] T.M. Mitchell. *Machine Learning*. Mac Graw Hill, 1997.

[45] I. Rish. An empirical study of the naive Bayes classifier. *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, pages 41–46, 2001.

[46] S. Russell, P. Norvig. Artificial intelligence: a modern approach. *New Jersey*, 1995.

[47] R. Rojas, J. Feldman. *Neural networks: a systematic introduction*. Springer, 1996.

[48] K.L. Du, M.N.S. Swamy. *Neural networks in a softcomputing framework*. Springer, 2006.

[49] I.A. Basheer, M. Hajmeer. Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods*, 43(1):3–31, 2000.

[50] S. Chen, SA Billings, PM Grant. Non-linear system identification using neural networks. *International Journal of Control*, 51:1191–1214, 1990.

[51] D.E. Rumelhart, G.E. Hinton, R.J. Williams. Learning internal representations by error propagation. *S*, 318:362, 1986.

[52] P.J. Werbos. Beyond regression: New tools for prediction and analysis in the behavioral sciences. 1974.

[53] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence*, Volume 14, pages 1137–1145. Citeseer, 1995.

[54] N. Japkowicz. Learning from imbalanced data sets: A comparison of various strategies, Learning from imbalanced data sets: The AAAI Workshop 10-15, 2000.

[55] R.M. Mead. nexrad-mltoolkit: Machine Learning Tools for Radar Ornithology. http://code.google.com/p/nexrad-mltoolkit/.

[56] J. Han, M. Kamber. *Data mining: concepts and techniques*. Morgan Kaufmann, 2006.

[57] E. Williams. Navigation on the spheroidal earth. *Journal of Navigation. Cambridge*, 2002.

[58] F.M. Handbook. Doppler Radar Meteorological Observations. Part A: System concepts responsibilities, and procedures. Technical Report, Office of the Federal Coordinator for Meteorological Services and Supporting Research (FCC-HI 1A-1991), 1991.

[59] F.M. Handbook. Doppler Radar Meteorological Observations. Part C, WSR-88D products and algorithms. Technical Report, FCM-H11C-1991, Office of the Federal Coordinator for Meteorological Services and Supporting Research, Rockville, Maryland, 1991.

APPENDIX A

VOLUME COVERAGE PATTERNS

Volume Coverage Patterns

A Volume coverage pattern(VCP) is a program telling the radar which elevation angles to scan as well as how to scan them. These patterns are designed to produce the most useful data for current environmental conditions.

The WSR-88D has two general modes of operation: Clear-Air and Precipitation. Intuitively, Clear-Air mode is used in the absence of any precipitation events and can provide more accurate measures of low level signals. Precipitation mode is intended to provide useful information in the presence of precipitation or extreme weather events. Although a radar technician can manually cause the equipment to enter Precipitation mode, it is more common for this mode to be entered automatically when a precipitation event is initially detected.

Table 7 provides general technical information for the volume coverage patterns used by the WSR-88D [58, 59]

Table 7: Volume Coverage Patterns

VCP	**Operational Mode**	**Number of Scans**	**Total Duration**
11	Precipitation	16	5 min
12	Precipitation	17	4.2 min
21	Precipitation	11	6 min
31	Clear Air	7	10 min
32	Clear Air	7	10 min
121	Precipitation	20	5.5 min

VCP 11

VCP 11 is a convection oriented coverage pattern that is designed for both severe and non-severe precipitation events. This pattern provides good overall storm coverage by scanning 14 elevation angles in about five minutes. Split cuts are used for the bottom two elevation angles. Precise elevation angle details are provided in Table 8.

Table 8: Elevation Angles for VCP 11

Scan Number	Elevation Angle	Scan Type
1	0.5	Reflectivity Only
2	0.5	Doppler Only
3	1.45	Reflectivity Only
4	1.45	Doppler Only
5	2.4	Reflectivity and Doppler
6	3.35	Reflectivity and Doppler
7	4.3	Reflectivity and Doppler
8	5.25	Reflectivity and Doppler
9	6.2	Reflectivity and Doppler
10	7.5	Reflectivity and Doppler
11	8.7	Reflectivity and Doppler
12	10.0	Reflectivity and Doppler
13	12.0	Reflectivity and Doppler
14	14.0	Reflectivity and Doppler
15	16.7	Reflectivity and Doppler
16	19.5	Reflectivity and Doppler

VCP 12

Like VCP 11, VCP 12 is intended for use with convection. It is specifically intended for deep convection and provides better vertical resolution and faster scan times than VCP 11. Overlap of lower elevation angles provides better coverage of the lower levels of precipitation events. Table 9 provides specific elevation angles and scan types for VCP 12.

VCP 21

VCP 21 is designed for shallow, generally non-severe precipitation. It is appropriate for slowly changing conditions like stratiform rain and snow. VCP 21 has a slower overall scan time than VCP 11 or 12 and only scans nine elevation angles. Table 10 provides further elevation angle details.

Table 9: Elevation Angles for VCP 12

Scan Number	Elevation Angle	Scan Type
1	0.5	Reflectivity Only
2	0.5	Doppler Only
3	0.9	Reflectivity Only
4	0.9	Doppler Only
5	1.3	Reflectivity Only
6	1.3	Doppler Only
7	1.8	Reflectivity and Doppler
8	2.4	Reflectivity and Doppler
9	3.1	Reflectivity and Doppler
10	4.0	Reflectivity and Doppler
11	5.1	Reflectivity and Doppler
12	6.4	Reflectivity and Doppler
13	8.0	Reflectivity and Doppler
14	10.0	Reflectivity and Doppler
15	12.5	Reflectivity and Doppler
16	15.6	Reflectivity and Doppler
17	19.5	Reflectivity and Doppler

Table 10: Elevation Angles for VCP 21

Scan Number	Elevation Angle	Scan Type
1	0.5	Reflectivity Only
2	0.5	Doppler Only
3	1.45	Reflectivity Only
4	1.45	Doppler Only
5	2.4	Reflectivity and Doppler
6	3.35	Reflectivity and Doppler
7	4.3	Reflectivity and Doppler
8	6.0	Reflectivity and Doppler
9	9.0	Reflectivity and Doppler
10	14.6	Reflectivity and Doppler
11	19.5	Reflectivity and Doppler

VCP 31

VCP 31 is the first Clear-Air mode pattern. It is intended to be more sensitive to low level signals. Unlike precipitation mode patterns, VCP 31 and 32 scan much fewer elevation angles and spend more time doing it. VCP 31 samples five elevation angles in about ten minutes. Elevation angle details are provided in Table 11

Table 11: Elevation Angles for VCP 31

Scan Number	**Elevation Angle**	**Scan Type**
1	0.5	Reflectivity Only
2	0.5	Doppler Only
3	1.5	Reflectivity Only
4	1.5	Doppler Only
5	2.5	Reflectivity Only
6	2.5	Doppler Only
7	3.5	Reflectivity and Doppler
8	4.5	Reflectivity and Doppler

VCP 32

VCP 32 is similar to VCP 31, but it uses a short pulse rather than a long pulse. The short pulse causes VCP 32 to be somewhat less sensitive than VCP 31, but it results in fewer velocity dealiasing failures. Table 12 provides further details.

Table 12: Elevation Angles for VCP 32

Scan Number	**Elevation Angle**	**Scan Type**
1	0.5	Reflectivity Only
2	0.5	Doppler Only
3	1.5	Reflectivity Only
4	1.5	Doppler Only
5	2.5	Reflectivity and Doppler
6	3.5	Reflectivity and Doppler
7	4.5	Reflectivity and Doppler

VCP 121

This VCP was designed as a temporary solution to the problem of excessive range folding and incorrect velocity de-aliasing. These problems are mitigated by performing multiple Doppler scans at the lower elevation angles. Unfortunately, due to the increased scan time, VCP 121 may not be appropriate in situations where conditions are changing rapidly. Individual scan details are provided in Table 13.

Table 13: Elevation Angles for VCP 121

Scan Number	**Elevation Angle**	**Scan Type**
1	0.5	Reflectivity Only
2	0.5	Doppler Only
3	0.5	Doppler Only
4	0.5	Doppler Only
5	1.45	Reflectivity Only
6	1.45	Doppler Only
7	1.45	Doppler Only
8	1.45	Doppler Only
9	2.4	Reflectivity and Doppler
10	2.4	Doppler Only
11	2.4	Doppler Only
12	3.35	Reflectivity and Doppler
13	3.35	Doppler Only
14	3.35	Doppler Only
15	4.3	Reflectivity and Doppler
16	4.3	Doppler Only
17	6.0	Reflectivity and Doppler
18	9.9	Reflectivity and Doppler
19	14.6	Reflectivity and Doppler
20	19.5	Reflectivity and Doppler