International Environmental Modelling and Software Society (iEMSs)
2010 International Congress on Environmental Modelling and Software
Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada
David A. Swayne, Wanhong Yang, A. A. Voinov, A. Rizzoli, T. Filatova (Eds.)
http://www.iemss.org/iemss2010/index.php?n=Main.Proceedings

New software methods in radar ornithology using WSR-88D weather data and potential application to monitoring effects of climate change on bird migration

Reginald Mead[a], John Paxton [a], Richard S. Sojda [b]

[a]*Department of Computer Science, Montana State University, Bozeman, MT 59717 USA (mead/paxton@cs.montana.edu)*

[b]*Northern Rocky Mountain Science Center, U.S.D.I.-Geological Survey, Bozeman , MT 49715 USA (sojda@usgs.gov)*

Abstract: Radar ornithology has provided tools for studying the movement of birds, especially related to migration. 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 such data cover a large portion of the United States, are constantly being produced, are freely available, and have been archived since the early 1990s. A major obstacle to this research, however, has been that identifying birds in NEXRAD data has required a trained technician to manually inspect a graphically rendered radar sweep. A single site completes one volume scan every five to ten minutes, producing over 52,000 volume scans in one year. This is an immense amount of data, and manual classification is infeasible. We have developed a system that identifies biological echoes using machine learning techniques. This approach begins with training data using scans that have been classified by experts, or uses bird data collected in the field. The data are 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. We compared neural networks, naive Bayes, and k-nearest neighbor classifiers. Empirical evidence is provided showing that this system can achieve classification accuracies in the 80th to 90th percentile. We propose to apply these methods to studying bird migration phenology and how it is affected by climate variability and change over multiple temporal scales.

***Keywords**:* NEXRAD; Machine Learning; Climate Change; snow goose

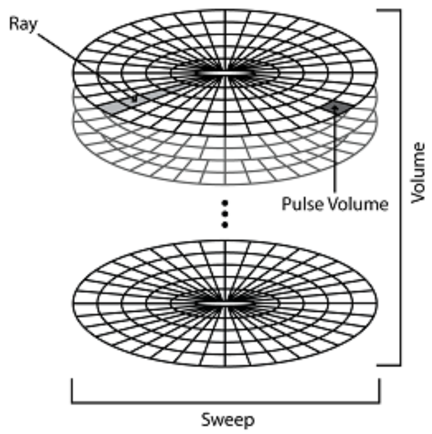
1. INTRODUCTION

Radar ornithology has provided tools for studying the movement of birds, especially related to migration (as demonstrated by O'Neal et al. [2010]). Using radar, scientists can observe birds at night and other times when visibility is poor. Scientists can also use radar to follow the movements of birds over vast distances, provided they have the right kind of equipment.

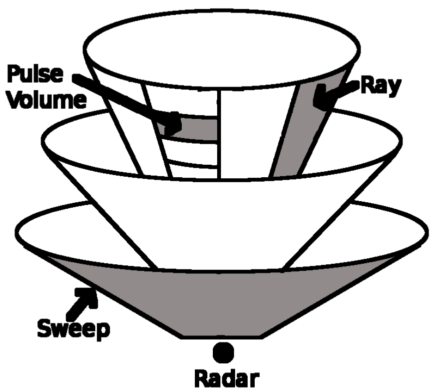
Despite the potential for broad scale studies of migration using radar, studies of this nature are impeded by the cost of acquiring radar equipment and employing field workers to operate the equipment, as well as by the logistical difficulty of simultaneously collecting data for a large geographic area.

In recent years, progress has been made in using broad scale weather surveillance radars in a radar ornithology context. The NEXRAD network of WSR-88D radar stations is particularly appealing to scientists because NEXRAD data offer broad coverage of the continental United States, are freely available, constantly updating, and archived back to the 1990s. Using NEXRAD data mitigates many of the costs typically associated with radar ornithology, but offers new challenges that scientists must address in order to use the data effectively. The first challenge is that NEXRAD data are generally much coarser than data from radars typically used for radar ornithology. Identification of individual birds is not possible; instead the WSR-88D provides radar information for an entire region of space known as a pulse volume. It is not unusual for pulse volumes to be roughly a cubic kilometer. Despite this potential loss of granularity, Diehl and Larkin [2002] have provided qualitative evidence indicating that migration events can be visually identified in WSR-88D sweeps. Further development of their techniques in combination with technological-based observations (portable X-band radar and infrared imagery) has further demonstrated the use of NEXRAD for quantifying bird movements. As scientifically useful as their techniques are, they had not been automated for detecting birds in the NEXRAD data.

Another obstacle to using NEXRAD data is the sheer magnitude of data available. The network of 159 radars is constantly producing and archiving data, with each site acquiring new volume scans roughly every five to ten minutes. A single site could produce over 52,000 scans in a single year. The system produces terabytes of new information every year. Manually searching through such a vast amount of data is impractical for studies encompassing large geographic areas or any considerable span of time.



(a) Conceptual Layout



(b) Realistic Layout

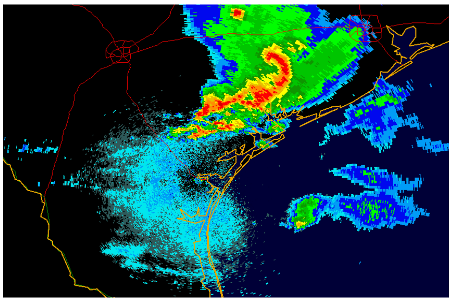
Figure 1: NEXRAD Hierarchical Data Structure

A potential solution to the problem of sifting through NEXRAD data has recently been proposed by Mead et al. [2008] in the form of a system that employs machine learning classifiers to automatically identify biological echoes in the WSR-88D Level II output. Initial results indicated that the system was capable of identifying mid to large scale passerine migration events.

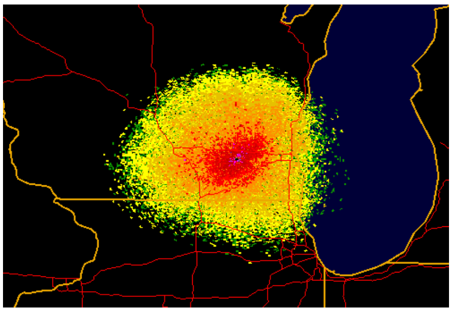
In this paper, we provide further empirical results for the system initially proposed by Mead et al. [2008] by using an expanded training dataset as well as by using lesser snow goose (*Chen caerulescens*) training data. The latter was collected using direct 3-dimensional observation by researchers in the field on over 1.5 million geese. We begin with a brief overview of NEXRAD data in Section 2. In Section 3 we describe our methodology and briefly describe the machine learning framework we are using. Our results are provided in Section 4. We then discuss potential applications of our methodology in Section 5, and conclude with a general discussion of this work in Section 6. Here, we explain how this system could be used to study changes in bird migration phenology related to climate change using the long term archive of weather radar data in the United States.

2. BACKGROUND

NEXRAD radars scan a three dimensional space around the radar that can be conceptualized as a half sphere. Level II data (the least processed data freely available via the NCDC distribution network) organize this space in a hierarchy as illustrated by Figure 1. At the outermost level is the volume. The volume is divided into a number of sweeps. As the name implies, a sweep contains data for one full sweep of the radar. It can be visualized as a disk, with subsequent sweeps being stacked on top of lower sweeps. Each sweep is taken at a particular elevation angle, so in reality, sweeps resemble flower petals more than stacked disks. Each sweep is divided into a number of rays like the spokes on a bicycle wheel and each ray is divided into a number of pulse volumes. The pulse volume is the smallest spatial unit and it is the three dimensional space within a pulse volume that the radar characterizes. Further details regarding Level II data are provided by Crum et al. [1993].



(a) Non-biological Sweep



(b) Biological Sweep

Figure 2: Visually Rendered NEXRAD Sweeps

NEXRAD Level II data provide three data moments (or features): reflectivity, mean radial velocity and spectrum width. Reflectivity is a measure of signal strength, mean radial velocity is the component of sensed velocities moving towards or away from the radar and spectrum width is a measure of the variability of the radial velocity moment. Figure 2 provides an example of how an actual reflectivity sweep is visually rendered. The sweep shown in Figure 2(a) shows a typical non-biological sweep. In this case, we see the irregular shape typical of precipitation. Figure 2(b), on the other hand, shows a sweep that experts would probably classify as biological. The sweep shows a symmetrical, spherical shape with a texture that is consistent with birds landing or taking off during a migration event. More information on weather radar data moments is provided by Doviak and Zrnic [2006].

3. METHODOLOGY

Our methodology begins with training data that are visually classified by experts or are extracted from observations by researchers in the field. Observers used global positioning systems, geographic information systems, and digital rangefinders to gather 3-dimensional observations (latitude, longitude, altitude [AGL]) on flocks of lesser snow geese associated with major migration staging areas, such as those birds at Sand Lake National Wildlife Refuge in northeastern South Dakota and Freezout Lake Waterfowl Management Area in northwestern Montana. Data collected included, number of geese, direction of flight, overall shape of flock, and weather conditions. Because such large numbers of geese were being observed in rapid succession, observers had to take written field notes and transcribe them later to a spreadsheet. The data from the spreadsheet were then error checked and transferred to a mysql database where they was stored. Before this data can be used by our system, they must be preprocessed in a number of ways. Preprocessing selects the 0.5°

sweeps from the full volume scan, merges those sweeps, removes clutter prone pulse volumes within 20 km of the radar site, removes radial velocity ambiguous pulse volumes beyond 145 km, removes pulse volumes with bad or range-folded values, and calculates a set of second order moments that statistically characterize the neighborhood of a pulse volume. After preprocessing the data to clean it up and highlight important information, the training data consist of a series of pulse volumes that have been labeled as either biological or non-biological. Each pulse volume is characterized by the three data moments described earlier and also by several statistical measures that provide information about the neighborhood of a pulse volume. These measures are variance, skewness and kurtosis as described by Joanes and Gill [1998]. These statistics are calculated for each of the three base data moments.

Following the preprocessing step, a machine learning classifier, either k-nearest neighbor, naive Bayes or a neural network, is trained and then the desired task is performed. Mitchell [1997] provides an overview of machine learning techniques, including those used in our system. We used a k value of three for the k-nearest neighbor classifier. The neural network had one hidden layer consisting of three nodes and was trained over the course of 300 epochs, using a learning rate of .3. To avoid over-fitting, the neural network reserved five percent of the training set for validation.

Typical tasks include system validation using ten-fold cross validation and the classification of unlabeled data. We use ten-fold cross validation as described in Kohavi [2005]. Ten-fold cross validation allows us to estimate the accuracy of our system in a robust manner. This is accomplished by dividing the training data into ten “folds”. Accuracy results are calculated for each fold by using the remaining nine folds to train the classifier. The accuracy results for all ten folds are then averaged to estimate the overall accuracy of our system. The folds for the snow goose data were stratified to retain class distribution, but the folds for the passerine data were not. An in depth description of our methodology is provided by Mead et al. [2008] and Mead [2009]. Random over sampling and random under sampling was used to randomly copy or delete training instances in order achieve balanced data sets containing an equal number of biological and non-biological training examples.

4. RESULTS

In this section, we provide some empirical results for the system we have developed. To measure the effectiveness of our system, we tested it in two different contexts. First, we tested the system using passerine data that have been classified at the sweep level by experts in the field of radar ornithology. After this, we attempted to provide better ground truth by testing our system on snow goose data collected from researchers in the field.

4.1 Passerine Data

Mead et al. [2008] describe how this system was originally trained and tested using expert classified passerine data. Results for experiments using the original dataset are provided in Table 1. The percentages indicate the number of correctly classified sweeps (biological and nonbiological) divided by the total number of sweeps classified. The original dataset used training examples that experts consider obvious and

Table 1: Original Diehl Dataset

Classifier	Correctly Classified Sweeps
K Nearest Neighbor	95% ± 0.7% [94.3, 95.7]
Naïve Bayes	97.5% ± 0.4% [97.1, 97.9]
Neural Network	95% ± 0.7% [94.3, 95.7]

Table 2: Expanded Diehl Dataset

Classifier	Correctly Classified Sweeps
K Nearest Neighbor	91% ± 0.5% [90.5, 91.5]
Naïve Bayes	77% ± 1.1% [75.9, 78.1]
Neural Network	82% ± 1.2% [80.8, 83.2]

were therefore arguably easier to learn. All values are provided with 95% confidence intervals to help assess statistical significance.

An expanded training dataset was developed in order to measure the system's effectiveness when given more difficult training examples. The results for the expanded dataset are provided in Table 2.

4.2 Snow Goose Data

The focus of the snow goose dataset is to facilitate a set of experiments for which ground truth is known. Unlike the sweeps in the passerine training data, which were classified at the sweep level by radar ornithologists, the snow goose training data were classified at the pulse volume level. These classifications are drawn from actual observations of snow geese in the airspace. Overall, we gathered observations on over 1,500 flocks representing over 1.5 million snow geese. Most of these observations on which we report, here, come from the Aberdeen, South Dakota area. This dataset is the first to almost exclusively use the new higher resolution data provided by NEXRAD Build 10. After random over sampling and random under sampling was performed we had a balanced dataset consisting of 828 non-biological pulse volumes and an equal number of biological pulse volumes. The results of ten test runs have been averaged to provide the confusion matrices in Tables 3 and 4. Tolerances next to each value define 95% confidence intervals.

Table 3: Naïve Bayes Classifier Results for Goose Dataset

	Predicted Class	
Actual Class	Non-biological	Biological
Non-biological	45.2% ± .004%	4.8% ± .004%
Biological	21.6% ± .003%	28.4% ± .003%
Accuracy = 73.6% ± .005%		

Table 4: Neural Network Classifier Results for Goose Dataset

	Predicted Class	
Actual Class	Non-biological	Biological
Non-biological	42.5% ± .018%	7.5% ± .018%
Biological	15.9% ± .014%	34.2% ± .014%
Accuracy = 76.7% ± .003%		

Table 5: K-Nearest Neighbor Classifier Results for Goose Dataset

	Predicted Class	
Actual Class	Non-biological	Biological
Non-biological	44.7% ± .001%	5.3% ± .001%
Biological	11.7% ± .004%	38.3% ± .004%
Accuracy = 83.0% ± .005%		

The confusion matrix allows us to see the number of correct classifications as well as the number of false positives and false negatives. These results show a lower classification accuracy for the geese data than for the passerine data. The results of validating the system against the goose data also show that the neural network outperformed naive Bayes in both overall classification accuracy and specifically classification of biological echoes.

Performance differences between the classifiers used are the result of the different inductive biases that the classifiers use to generalize training data and thereby make predictions for new data.

5. POTENTIAL APPLICATIONS

After this system has been trained and an acceptable accuracy rate is achieved, we propose that it can be used to investigate certain aspects of avian behavior and migration. As an example, we have used the system to classify sweeps from three radar stations over the course of two migration seasons. The stations are KMKX Milwaukee, KGRB Green Bay and KTFX Great Falls. The seasons we selected are fall 2007 and spring 2008. We collected the first sweep of every hour for three months for each of the seasons. Specific months were chosen based on recommendations from a radar ornithology expert. The classifier used for this example was a neural network that had been trained using the snow goose data.

By plotting the average number of biological pulse volumes per sweep for a given day, we can produce graphs such as those shown in Figures 3 and 4. These graphs are scatter plots which have been smoothed using locally weighted scatter plot smoothing (LOWESS). These graphs could be used to gather information about starting and ending migration dates.

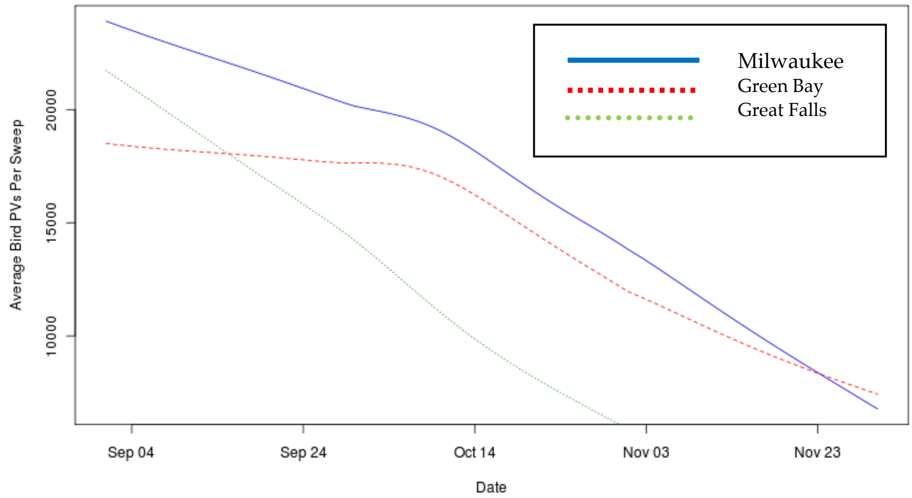


Figure 3: Fall 2007 Migration

Using our system to study migration phenology in relation to any temperature and precipitation trends correlated with climate change is our next research step. Since our system is efficient enough to process multiple seasons of data, an appropriate next step would be to acquire data for several seasons from stations that are near large concentrations of waterfowl (e.g., Great Falls, Montana), process the data, and then extract information on

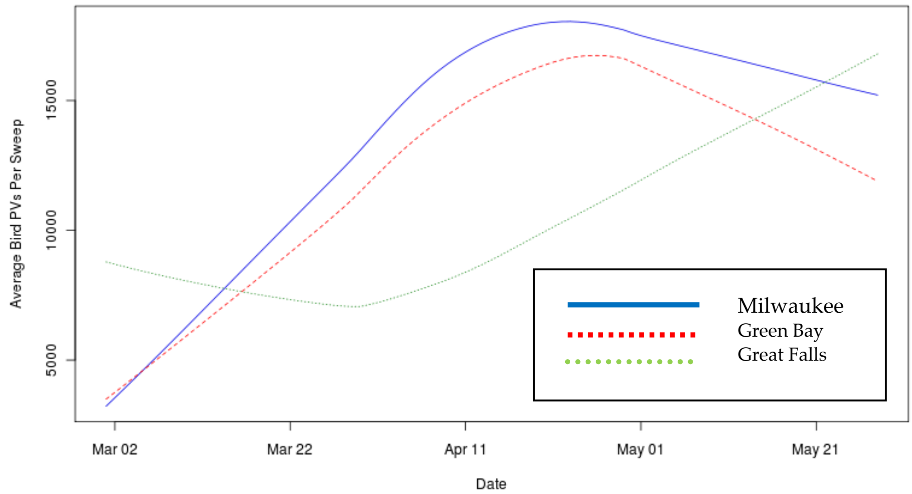


Figure 4: Spring 2008 Migration

any trends that exist over time and space. Such trends might be sensitive to changes in time, e.g., average time of peak migrations, but variability in short term climate patterns (within a year) would have to be analyzed. Similarly, changes in spatial distribution of birds could be examined. For example, the station at Aberdeen, South Dakota could be analyzed in relation to the station to its South, Hastings, Nebraska. The ability to monitor changes in bird migration since the late 1980s or early 1990s, depending on the station, using the freely available NEXRAD data archive is quite exciting. This is especially so when we recognize that weather radar stations are widely distributed across the United States of America.

6. CONCLUSIONS AND FUTURE DIRECTIONS

The results show that this system performs well on passerine identification tasks, achieving accuracy rates in the 80th to 90th percentile for sweep level classification. These rates held up even on the more difficult expanded training dataset. Geese classification accuracy was somewhat lower, having classification accuracies in the high 70th percentile.

There are a number of factors that may be negatively affecting classification accuracy on the geese training data. The first is the amount of training data available. The sweep level classifications given to the passerine data allow every pulse volume in the sweep to be used as training data. This results in a training data set with hundreds of thousands of training instances. In contrast to this, the snow goose training data are based on direct field observation for individual pulse volumes and contains less than two thousand training instances. It is possible that the limited amount of training data is affecting the classifier's ability to learn the complex function mapping pulse volume features to the correct class.

One partial solution to our lack of training data is to use semi-supervised learning approach. Under certain conditions, semi-supervised learning can use unlabeled data to improve classifier performance. The availability of large amounts of unlabeled radar data makes semi-supervised learning attractive and we plan to look into incorporating a semi-supervised learning component in the future. Zhu [2005] provides a survey of the semi-supervised learning literature.

Another factor that is potentially affecting classification accuracy is additional noise in the training data resulting from geolocation issues. By geolocation issues, we are referring to the difficulty of associating an observed location with the correct pulse volume provided by the radar. In addition to finding the correct location after taking into account the variable elevation of the radar beam and the curvature of the earth, there is also the issue of timing. The radar equipment takes a certain amount of time to make the 360° sweep and a field observer has no way of knowing how long it will be before the beam is directed at the space that she is observing. This problem is exacerbated by the constantly moving nature of geese in flight. Updates to the NEXRAD system that have increased pulse volume resolution make it more likely that geese in flight could cross the boundary between adjacent pulse volumes in the time between an observation and when the radar is actually passing over that particular area. We have begun to explore ways to observe NEXRAD level II data in near real time. In combination with that, we plan to communicate probable locations of birds to field observers for them to validate visually. Alternatively, we will also attempt to use portable X-band radar to document the location of birds in the airspace, although that will not provide information about species of birds. It is probable that classification accuracy for geese will improve with additional training data that are free of the noise introduced by geolocation errors. In addition, as the precision of NEXRAD data increase fourfold, and recognizing that flocks of snow geese might intersect multiple pulse volumes, the amount of data (numbers of pulse volumes) might also dramatically increase.

Before our work, no one had published information about using algorithms for identification of birds in NEXRAD data, although experts had been able to do so visually.

Such expert identification had been at the sweep level, and were able to do so at the pulse volume level. This indicates, and we have corroborating unpublished data, that the data being returned from a pulse volume for small birds were different than that returned for large birds. This had not been demonstrated before, and we look forward to investigating those differences more fully. We expect that the implementation of polarimetric Doppler radar for weather forecasting will provide further data moments to analyze and further refine our ability to not only detect birds, in general, but to potentially differentiate species.

ACKNOWLEDGMENTS

We would like to thank R. Diehl and his students at the University of Southern Mississippi for their support, data, and expertise. E. Klaas and E. Walsh were outstanding snow goose observers and their dedication to stringent field protocol was outstanding. We would also like to thank M. Suarez of the USDI-Geological Survey for his many contributions, including data organization. W. Schultze, A. Manville, B. Bortner, M. Green (USDI-Fish and Wildlife Service), B. Lonner (Montana Department of Fish, Wildlife, and Parks), S. Stephens (Ducks Unlimited), and D. Otis (USDI-Geological Survey) all contributed invaluable logistic and administrative support. We would like to thank K. Sinclair, and R. Thresher of the National Renewable Energy Lab for encouragement. A. Arnone, J. Sheppard, and C. Schwartz provided peer review that improved the manuscript considerably. This project was funded by the National Renewable Energy Laboratory and the USDI-Geological Survey. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

REFERENCES

Crum, T., R. Alberty, and D. Burgess. Recording, archiving, and using WSR-88D data. *Bulletinof the American Meteorological Society*, 74(4):645–653, 1993.

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

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

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

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

Mead, R. A system for automating identification of biological echoes in nexrad level ii radar data.Master's thesis, Montana State University, 2009.

Mead, R., J. Paxton, and R. Sojda. Identifying biological echoes in radar scans using machinelearning. In *Proceedings of the 4th Biennial Meeting of iEMSs*, 2008.

Mitchell, T. *Machine Learning*. Mac Graw Hill, 1997.

Zhu, X. Semi-Supervised Learning Literature Review. Technical Report, Computer Sciences, University of Wisconsin-Madison, 2005.