

UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

MACHINE LEARNING FOR CLASSIFYING BIOLOGICAL RADAR
ECHOES WITH S-BAND POLARIMETRIC RADAR

A DISSERTATION
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
DOCTOR OF PHILOSOPHY

By
Precious Kizeze Jatau
Norman, Oklahoma
2023

MACHINE LEARNING FOR CLASSIFYING BIOLOGICAL RADAR
ECHOES WITH S-BAND POLARIMETRIC RADAR

A DISSERTATION APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Tian-You Yu, Chair

Dr. Valery Melnikov, Co-Chair

Dr. Robert Palmer

Dr. Joseph Havlicek

Dr. Jeffrey Kelly

Dr. Justin Metcalf

Dr. Amy McGovern

*Dedicated to the late Dr. Raymond Jatau. You paved the path before me. All I
have done is follow.*

Acknowledgments

My highest gratitude goes to God, who has actualized what once seemed like fantasy. Like any other journey, there have been challenges, but always punctuated by His even greater providence. He is the reason I am the person I am today.

Many people have helped me along the way. First and foremost, I would like to express my sincere gratitude to my advisors: Tian-You Yu and Valery Melnikov. Their tutelage has been foundational to my formation as a researcher. I appreciate their availability to meet with me every week, and willingness to share their wealth of experience in radar engineering, meteorology, signal processing and aeroecology. I am also thankful for all the constructive feedback, that helped me think about problems in new innovative ways. I am proud of the work we have done together and I hope you can keep inspiring future generations of students.

It has taken a village to raise this researcher. I am especially grateful for the help provided by every member of my committee during this project: I learnt about Machine Learning from Amy McGovern; weather radar theory from Robert Palmer; and signal and image processing from Joseph Havlicek. I am also appreciative of Justin Metcalf, under whose mentorship I grew as a teacher and developed a better understanding of probability and statistics. And last by not least, Jeffrey Kelly who has been instrumental in my understanding of bird and insect biology. If I have seen further, it has only been because I stood on the shoulders of giants.

I am thankful for Advanced Radar Research Center. Being a part of the largest academic radar program in the nation has been an honor. I am especially grateful for my friends and colleagues: Yoon Kim, Cesar Salazar Aquino, Hyeri Kim, Nawaf Almuqati, Jorge Luis Alva Alarcon, Syed Jehangir, Min-Duan Tzeng, Ayano Ueki and Hsiu-Wei Hsu. I also want to recognize the friendly security officers that keep us safe, the janitors who keep our stations clean, and the secretaries and managers who tirelessly sustain the lab. Within the broader OU community, special thanks to Rodney Keele for the constant support throughout my graduate studies and the friendly baristas at the Bizzell Library's Bookmark Cafe who made many tedious days a little less stressful.

Of the giants in my life, none stands taller than the late Dr. Raymond Jatau, my father in Heaven. We might have only shared 19 years together, but in that time, you set a blue print for the rest of my life with your hard work and humanity. I hope this Ph.D makes you proud. I am also grateful for the sacrifices of my mother, Comfort Jatau and the support of my siblings: Halley, Favour, Nansih, Hendrix, Diana and Christina.

Funding for this study was provided in part by the NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA110OAR4320072.

Table of Contents

Acknowledgments	v
Table of Contents	x
List of Tables	xi
List of Figures	xvii
Abstract	xviii
1 Introduction	1
1.1 Motivations	1
1.2 Review of Biological Echo Detection Using Radar	4
1.3 Problem Statement	7
1.4 Contributions	9
1.5 Scope of the Dissertation	10
2 Fundamentals of Weather Radars and Machine Learning	12
2.1 Radar Fundamentals	13
2.1.1 Important Radar Parameters	13
2.1.2 Point Target Radar Equation	15
2.1.3 Weather Radar Equation	17
2.1.4 NEXRAD	20
2.1.5 Level II Products	22

2.2	Clear-air Echoes	27
2.2.1	Characteristics of Clear-air Echoes	28
2.2.2	Daily Cycle of Clear-air Echoes	30
2.3	Machine Learning (ML) Methods Used in this Dissertation	33
2.4	Velocity Azimuth Display (VAD)	36
3	A Bird-Insect Ridge Classifier (BIRC) for S-band Polarimetric Weather Radar	38
3.1	Introduction	38
3.2	Data Collection	39
3.2.1	Selection of Bird and Insect Scans	39
3.2.2	Selection of Radar Variables for ML algorithms	40
3.3	Feature Processing to Prepare Inputs	41
3.3.1	Texture	42
3.3.2	Blob Coloring and Minor Region Removal to Extract Migration Echoes	43
3.3.3	Reference with Respect to Target Azimuth	45
3.3.4	Averaging Bird and Insect Cases	47
3.3.5	Input Features for Classifiers	50
3.4	Machine Learning Methods	51
3.5	Metrics	51
3.6	Model Training and Validation	53
3.7	Performance	54
3.8	Case Studies and Discussions	57
3.8.1	Bird/Insect Migration Test Cases	59
3.8.2	Bird Migration with Aliased Velocities	60
3.8.3	KHTX Bird Roosts	62
3.8.4	KTLX Bird Roosts	64
3.9	Recommendations for Operational Implementation	65

3.10	Conclusions	67
4	Application of the Bird-Insect Ridge Classifier for VAD Wind Estimation	71
4.1	Introduction	71
4.2	Proposed Method	74
4.2.1	Detecting Biological and Weather Echoes	75
4.2.2	Classifying Biological Echoes into Birds and Insects	77
4.2.3	VAD Algorithm	77
4.3	Data Preparation	79
4.3.1	NEXRAD Data	79
4.3.2	RAP as Ground Truth	80
4.4	Wind Bias for Evaluating VADs	82
4.5	Identifying the Source of VAD Biases	83
4.5.1	Insects-to-Birds Ratio of Range Rings	84
4.5.2	Analysis of Wind Biases	88
4.6	Improving Wind Estimates from Contaminated VADs	91
4.6.1	Birds-Insects Bias Differences for Identifying Improvement Conditions	91
4.6.2	Analysis of Bird-Insect Bias Differences	95
4.6.3	VAD Improvement	97
4.6.4	Analysis of improvements	98
4.7	Recommendations and Conclusions	103
5	Conclusions and Future Work	105
5.1	Summary of Contributions	105
5.1.1	Classification of Bird and Insect Echoes	105
5.1.2	Application of BIRC for VAD Wind Estimation	106
5.2	Future Work	108

5.2.1	Improvement of Bird/Insect Classification	108
5.2.2	Validation of VAD Products	109
	References	110
	Acronyms	127
	Appendix A A Supplement to: Application of the Bird-Insect Ridge Classifier for VAD Wind Estimation	129

List of Tables

2.1	System specifications of WSR-88D [61, 62]	22
3.1	Confusion matrix for bird detection	52
3.2	95% confidence interval for the ridge classifiers' and decision trees' metrics on migration data. The possible inputs are Dual-Pol(DP), their textures(ΔDP), sector($sect$) and range interval(RI)	55
3.3	Model comparison results. Each cell is labelled R (ridge) or D (decision tree) depending on which model performs better	57
3.4	Summary of case studies	59
4.1	Performance metrics of BIRC	77
4.2	Summary of data for bird migration seasons	79
4.3	Distance between KOHX, KHTX and closest RAP and sounding stations.	81

List of Figures

1.1	Sunken US Airways Flight 1549 in the aftermath of a bird strike [23].	3
2.1	NEXRAD coverage of Contiguous US [59].	21
2.2	Calculation of φ_{DP} for three distinct pulses. [68]	27
2.3	Average Reflectivity below 2 km versus time for a nocturnal case. Sunset is ~ 2 Z while sunrise is ~ 11 Z. Data was collected using the Cimarron radar on May 31, 1999. [70]	31
2.4	Part of decision tree, containing three attributes	34
2.5	VAD applied to obtain speed and direction of migration. Echoes are found to move toward 70° with a speed of 10 ms^{-1}	36
3.1	Flowchart of feature processing	41
3.2	Blob coloring with minor region removal to extract large scale mi- gration echoes. (a) Reflectivity (in dBZ) of bird migration echoes. (b) mask of bird migration echoes. (c) gates containing birds ex- tracted using mask. (d) Reflectivity (in dBZ) of insect echoes (e) mask of insect echoes (f) gates containing insects extracted using mask	44
3.3	VAD analysis to reorient velocity to be relative to the target as- pect for 70 km range gates. The blue line shows filtered V_r while the green line is the sine fit. (a) Initial VAD finds targets to be migrating toward 13.73° (b) Reoriented fit shifts measurements so migration is toward 0°	47

3.4	Averaged dual polarization variables as a function of ϕ_{aspect} for the training set. (a) Z_{DR} vs ϕ_{aspect} (b) φ_{DP} vs ϕ_{aspect} (c) ρ_{HV} vs ϕ_{aspect} . Birds are in blue and insects red. Rows represent ranges of 15, 30 and 45 km from the radar. Training set distributions of averaged (d) Z_{DR} (e) φ_{DP} , and (f) ρ_{HV} for birds and insects.	48
3.5	Ridge classification results for bird migration observed with KTLX radar at 04:13 UTC on 2 May 2015. BIR represent birds and INS represent insects.	60
3.6	Ridge classification results for insect migration observed with KTLX radar at 17:08 UTC on 11 July 2019.	60
3.7	Bird migration with velocity aliasing observed with KTLX at 04:41 UTC on 03 March 2018. (a) Ridge classifier (RC) using dual polarization variables. (b) $0.5^\circ V_r$ shows heavy aliasing north and south of radar. (c) RC using dual polarization variables and their textures. (d) RC using dual polarization variables and range interval. (e) Hydrometeor Classification Algorithm (HCA). (f) RC using dual polarization variables, their textures and range interval.	61
3.8	Bird roosts observed with KHTX at 11:15 UTC on 11 August 2015. (a) RC using dual polarization variables. (b) $0.5^\circ Z$ scan showing bird roosts (west and north-west), insects (around the radar) and weather echoes (north-east and south). (c) RC using dual polarization variables and their textures. (d) RC using dual polarization variables and range interval. (e) HCA. (f) RC using dual polarization variables, their textures and range interval.	62
3.9	Same as Figure 3.8 but for roosts observed with KTLX at 11:47 UTC on 8 August, 2017. Four bird roosts were identified north-west, north-east and south-east of the radar	64

3.10	Weights of base ridge classifier. Positive (negative) weights for a variable imply that the probability of being a bird is directly (inversely) proportional to the variable.	67
4.1	VAD wind biases averaged by height and time of day for (a) nocturnal bird migration observed by KOHX (Nashville, TN) during May 2018. (b) diurnal bird migration observed by KHGX (Houston, TX) during May 2022. The black dashed lines indicate sunset around 1 UTC and sunrise around 11 UTC. Local time is UTC - 5.	72
4.2	Proposed VAD modifications based on BIRC's predictions. WX stands for weather while BIO represents biological echoes	75
4.3	Correcting misclassifications of biological echoes from the Hydrometeor Classification Algorithm (HCA). (a) HCA predictions for nocturnal birds from the 0.5° elevation. (b) ρ_{HV} for nocturnal birds. (c) HCA predictions after correction. (d) HCA predictions for afternoon insects from the 2.5° elevation. (e) ρ_{HV} for afternoon insects (c) HCA predictions after correction.	76
4.4	Comparison of local VAD on biological echoes with level 3 VWP for KOHX, May 2018. The median speed and direction differences are -0.88 ms^{-1} and 0.72° respectively	78
4.5	Daily averages of relative bird, insect and weather proportions for the (a) May 2018 KOHX nocturnal migration season (b) May 2022 KHGX diurnal migration season. Local time is UTC - 5	80
4.6	Comparison of RAP and sounding data for KOHX, May 2018. The median speed and direction differences of 0.10 ms^{-1} and 4.5° respectively.	82

4.7	(a) Radial velocity for a nocturnal bird migration case. (b) The u (grey) and v (pink) wind components estimated from VAD have large deviations from RAP's u (blue) and v (red) components. (c) Large wind biases due to deviations measure the bird's airspeed. (d) Radial velocity for a daytime insect swarm. (e) Insects' VAD trace the RAP wind with small deviations. (f) Small wind biases measure the insect's airspeed.	84
4.8	(a) BIRC predictions for a nocturnal bird migration case. (b) The number of bird gates exceeds those for insects across all heights. (c) Insects-to-birds ratio is low ($\sim 22\%$) because birds dominate insects. (d) BIRC predictions for an afternoon insect swarm. (e) The number of insect gates exceeds those for birds across all heights. (f) Insect-to-bird ratio is high ($\sim 90\%$) indicating insect dominance.	86
4.9	Insects-birds ratios averaged by height and time of day for the (a) KOHX nocturnal migration season (b) KHGX diurnal migration season. The black dashed lines indicate sunset and sunrise around 1 and 11 UTC. Local time is UTC - 5	87
4.10	Daily averages of total bird and insect gates within a VAD ring. From left to right, columns correspond to nocturnal and diurnal migration seasons. From top to bottom, rows correspond to bird and insect gates count.	88
4.11	(a) - (b) Wind biases averaged by height and the insects-birds ratio. (c) - (d) Scatterplots for wind bias against insects-birds ratio. From left to right, columns correspond to nocturnal and diurnal migration seasons. Biases can be seen to reduce as the insect-bird ratio increases (i.e fewer birds, more insects)	89

4.12	Comparisons of biases from bird dominated (blue) and insect dominated (pink) rings for the (a) KOHX nocturnal migration season. (b) KHGX diurnal migration season.	90
4.13	(a) BIRC predictions for an insect swarm case. (b) Radial velocities from 620 <i>m</i> height. (c) Biases from predicted birds and insects. (d) Difference between bird and insect biases. Red points are from insects and blue are from birds.	92
4.14	(a) BIRC predictions for a mixed case. (b) Radial velocities from 620 <i>m</i> height. (c) Biases from predicted birds and insects. (d) Difference between bird and insect biases.	93
4.15	(a) BIRC predictions for a bird dominated case. (b) Radial velocities from 780 <i>m</i> height. (c) Biases from predicted birds and insects. (d) Difference between bird and insect biases.	94
4.16	Averages of bird-insect bias differences by height and time of day, for the (a) KOHX nocturnal migration season (b) KHGX diurnal migration season. Local time is UTC - 5	95
4.17	Bird-insect bias difference averaged by height and insect-to-bird ratio for the (a) KOHX nocturnal migration season (b) KHGX diurnal migration season.	96
4.18	Height by insects-birds ratio averages of improvement and the fraction of VADs lost. From left to right, columns correspond to nocturnal and diurnal migration seasons. From top to bottom, rows correspond to improvement and the fraction of VADs lost.	99
4.19	Height by time of day averages of improvement and the fraction of VADs lost. From left to right, columns correspond to nocturnal and diurnal migration seasons. From top to bottom, rows correspond to improvement and the fraction of VADs lost.	100

4.20	Analysis of insect filtering for the KOHX nocturnal migration season using different confidence thresholds τ . From top to bottom, rows correspond to improvements and fraction of VADs lost. . . .	102
4.21	Same as Figure 4.20, but for the KHGX diurnal season	103
A.1	Relative bird, insect and weather proportions for KOHX, May 1 - 12, 2018	129
A.2	Same as Figure A.1, but for May 13 - 31, 2018.	130
A.3	Biological VAD-wind biases for each day during the KOHX (Nashville, TN) May 2018 nocturnal migration season. The black vertical line denotes 0 UTC.	131
A.4	Same as Figure A.3 but for insects-birds ratios	131
A.5	Same as Figure A.3 but for total bird gates	132
A.6	Same as Figure A.3 but for total insect gates	132
A.7	Same as Figure A.3 but for bird-insect bias differences	133
A.8	Same as Figure A.3 but for improvements	133
A.9	Relative bird, insect and weather proportions for KHGX, May 1 - 12, 2022	134
A.10	Same as Figure A.9, but for May 13 - 31, 2022.	135
A.11	Biological VAD-wind biases for each day during the KHGX (Houston, TX) May 2018 nocturnal migration season. The black vertical line denotes 0 UTC.	136
A.12	Same as Figure A.11 but for insects-birds ratios	136
A.13	Same as Figure A.11 but for total bird gates	137
A.14	Same as Figure A.11 but for total insect gates	137
A.15	Same as Figure A.11 but for bird-insect bias differences	138
A.16	Same as Figure A.11 but for improvements	138

Abstract

The S-band WSR-88D weather radar is sensitive enough to observe biological scatterers like birds and insects. However, their non-spherical shapes and frequent collocation in the radar resolution volume create challenges in identifying their echoes. We propose a method of extracting bird (or insect) features by coherently averaging dual polarization measurements from multiple radar scans, containing bird (insect) migration. Additional features are also computed to capture aspect and range dependence, and the variation of these echoes over local regions.

Next, ridge classifier and decision tree machine learning algorithms are trained, first only with the averaged dual pol inputs and then different combinations of the remaining features are added. The performance of all models for both methods, are analyzed using metrics computed from the test data. Further studies on different patterns of birds/insects, including roosting birds, bird migration and insect migration cases, are used to investigate the generality of our models. Overall, the ridge classifier using only dual polarization variables was found to perform consistently well across all these tests. Hereafter, this model is called the Bird-Insect Ridge Classifier (BIRC).

Enhancements of the Velocity Azimuth Display (VAD) Wind Profile (VWP) are explored by integrating BIRC to generate three new VAD products namely: the insects-birds ratio, VAD focused on birds and insects focused VADs. Two main findings are drawn from experiments on these products. First, the insects-

birds ratio is found to have an inverse relationship to wind biases, confirming bird contamination as a cause of the latter. Second, VAD wind biases can be reduced by focusing VAD on insects instead of all biological echoes.

It is recommended that BIRC can be used on the WSR-88D, for classifying biological echoes from the HCA as birds or insects. Furthermore, the insects-birds ratio, bird VAD and insect VAD products can be incorporated into the VWP. To the best of our knowledge, this is the first machine learning classifier that has been demonstrated to simultaneously classify diverse patterns of bird and insect echoes, and improve clear-air VAD wind estimation by incorporating taxonomic information.

Chapter 1

Introduction

1.1 Motivations

Aerial biota play a critical role in our ecological system. Thus, it is important to study the interactivity between humans, climate and these animals. Such broad scale studies would require information collected with fine spatial and temporal resolutions, over large areas and long time periods. [1–5]. Fortunately, the existing US Next Generation Radar (NEXRAD) network meets those requirements and has been shown to observe biological scatterers [6–11]. NEXRAD has wide coverage with 160+ Weather Surveillance Radar - 1988 Doppler (WSR-88D) systems deployed across the continental US, Alaska, Hawaii and Puerto Rico. WSR-88Ds were upgraded to have dual-polarization capability in 2013, enabling the classification of various hydrometeor types and non-weather echoes. Each WSR-88D senses the environment as raw time series signals, referred to as level I data. Further processing is performed on this raw data to obtain and record level II data, comprised of three legacy single polarization variables: reflectivity factor Z , Doppler velocity V_r and spectrum width σ_v , and three dual polarization variables namely: differential reflectivity Z_{DR} , cross-correlation coefficient ρ_{HV} and differential phase φ_{DP} .

Though primarily designed for weather, the WSR-88D can also detect biological scatterers such as birds, insects and bats [12], offering a rich repository of historical information on aerial biota. NEXRAD has generally archived over 25 years worth of data [5, 13]. It can be adapted as a biological surveillance network and leveraged to address important issues in meteorology, aviation, epidemiology and aeroecology.

One of these problems is the contamination of clear-air wind estimates from the Velocity Azimuth Display (VAD) Wind Profile (VWP). VAD approximates wind speed and direction by fitting a sinusoidal model to radial velocities at a given height. As a result, the quality of estimates depends on the nature of echoes. For example, passive flyers like insects are primarily driven by the wind and can be good wind tracers. Birds, on the other hand, have been found to contaminate radar derived winds [14] with their larger airspeeds [15]. Therefore, the differentiation of both taxa would enable improvements to radar wind estimation and potentially provide new information about bird migration. This would have far reaching impacts on meteorological observations, given the wide use of VWP among the WSR-88D products [16].

Another potential application of biological surveillance is the mitigation of bird strikes. Collisions between birds and aircraft often result in injury/death for the birds, damage for the aircraft or both. Bird strikes pose a serious aviation hazard for low-level flights [17]. The primary concern is human safety, though significant financial losses are additionally incurred, due to damage of equipment and delay of operations [18]. The annual cost has been estimated at \$1.2 billion world-wide [19]. Large scale seasonal migration poses a particular threat for aviation [20], because of the difficulty in monitoring the occurrence of migrating birds at the local scale of airport responses [18]. A popular example of bird strikes with a

migratory specie is the “The Miracle on the Hudson”. On January 15, 2009, US Airways Flight 1549 encountered a flock of Canada Geese shortly after take off. Some of these birds were ingested into both engines leading to a shutdown and loss of thrust. The pilot, Captain Chesley Sullenberger, ultimately saved the lives of all 155 people on board by the unorthodox, but heroic decision to land the plane on the Hudson river. A picture of the submerged aircraft is shown in Figure 1.1. Subsequent studies found that the geese came from a migrant population, and the strike had occurred outside the population’s breeding range [21]. Tracking the movement of migrating birds on a broad scale can provide critical information for averting bird strikes [17, 22] in the future.



Figure 1.1: Sunken US Airways Flight 1549 in the aftermath of a bird strike [23].

Aerial biota have also been found to be disease carriers. For example, a zoonotic origin has been considered as a plausible explanation for the recent COVID-19 pandemic [24]. Bats were the supposed culprit in this case. Furthermore, the long distance transport of migrants has frequently been associated

with outbreaks of agricultural pests and insect-borne plant or animal diseases [25, 26]. Large-scale radar monitoring can improve our understanding of the spread of avian diseases by allowing a detailed mapping of migratory flyways [17, 27].

There is a huge potential to solve important problems by monitoring and tracking different biological echoes individually on NEXRAD. However, despite many advances in characterizing hydrometeor types [28–30], the classification of biological taxa is still an active research field [22, 31–33]. For example, the Hydrometeor Classification Algorithm (HCA) has 8 classes for weather compared to one generic biological class [29]. A similar biological classification model is needed.

This work aims to develop a biological classifier for NEXRAD. The classifier would need to operate on singular gates to maximize its scope of usage. For instance, migrating birds might cover several kilometers of airspace while emerging roosts might be confined to one gate. A gate level classifier would detect bird presence in both scenarios. Furthermore, applications of the classifier for improving VAD winds are investigated, by accounting for the source of measured velocities. It is hoped that findings from this study will be useful for researchers and industry partners interested in radar aeroecology. The layout of the rest of this chapter is as follows. First, a review of previous literature on detecting biological echoes is provided, followed by the problem statement and a summary of contributions. The final section presents the scope of this dissertation.

1.2 Review of Biological Echo Detection Using Radar

Various methods have been explored to detect biological echoes with radars [32–38]. However, much less work has been done on taxonomic classification. This section presents an overview of existing approaches for biological echo recognition,

as well as discussions on their viability for operational use on NEXRAD.

Automated techniques have been developed for identifying bird contamination in the lowest elevation scan based on reflectivity and Doppler velocity [39, 40], and extracting height profiles of bird’s density, speed and direction at C-band [17]. Neither are applicable to non-bird echoes nor singular range gates. Taking a different approach, polarimetric spectral analysis was explored for distinguishing birds from insects in multimodal spectra [14]. However, the required level-I time series data is not typically available from the operational WSR-88D. Radar Cross Section (RCS) were also used for discriminating individual targets observed at close proximity to non-polarimetric W and X-band research radars [15]. NEXRAD, by contrast, offers coarser spatial resolution at S-band, which is insufficient for resolving point targets.

A different approach uses fuzzy logic to detect biological echoes. A popular example is the operational HCA, which separates biological echoes from ground clutter, and weather [29], without differentiating taxa. Focusing more on avian fauna, another fuzzy logic algorithm was developed for separating birds, precipitation and ground clutter echoes [32]. Insects are not identified in that study. Around the same time, a binary bird-insect classifier was developed to supplement the HCA [22], although a challenge is posed by its dependence on reflectivity causing occasional confusion of dense insect aggregations with birds [41]. This dissertation builds on findings from the binary classifier, and seeks to address its limitations.

Machine Learning (ML) is becoming widely adopted in weather radar, in large part due to work on better understanding its black boxes [42]. It has found applications in digital beamforming [43], velocity dealiasing [44, 45] and monitoring biological echoes [5, 31, 41, 46–48]. Bird roost detection in particular, is a pop-

ular use case of ML because of their distinct toroidal shapes. More broadly, the detection of any radar echo can be framed as an image recognition task, where scans correspond to images and range gates correspond to pixels. There are three main approaches to image recognition. They include image classification, object localization and semantic segmentation.

In image classification, the presence or absence of an object is predicted in a given image. Following this approach, convolutional neural networks were used in [46] to detect whether an individual radar image contained at least one Purple Martin or Tree Swallow roost, with correct determinations made about 90% of the time. However, the model was designed for one orientation of birds and would not be applicable to non-roosting birds or other types of biological echoes. Additionally, the model was not designed to localize roosts within radar images.

The second approach is object localization, where an object’s presence and location are simultaneously predicted in a given image. Based on this approach, a single frame detector was developed in [47] that locates and tracks roosts within radar images. Building on this work, an Artificial Intelligence (AI) system with an interface for human screening was also developed to collect research-grade data on roosting birds [5]. Similar to the previous case, the usage of these models is limited to roosting birds. Furthermore, non-bird echoes like precipitation and wind farms, which often assume the prototypical roost shape, were a leading source of false detections.

The third approach is semantic segmentation, where each pixel is classified. Following this approach, random forests were used in [31] to develop a variety of segmentation models. The model performances were evaluated for combinations of the six level-II products with 8 output classes (7 biological and 1 meteorological). However, the models were trained with the primary goal of exploring the variance

of known types of biological scatterers and rain within the radar parameter space. They were not developed for production level use [31].

1.3 Problem Statement

The US network of S-band polarimetric weather radars can detect signals from birds and insects. In identifying the source of these echoes however, the current most precise designation, is an all-encompassing biological class provided by the HCA. As such, a gate level classifier is needed to further distinguish biological echoes by taxa. This classifier should be operationally deployable on NEXRAD, in order to take advantage of the latter’s space-time coverage and resolution. Bird echoes can be monitored on a broad scale, potentially providing critical information for averting bird strikes and ornithological studies. Conversely, insect monitoring can enable improved wind estimation, in the way of their velocities and support entomological efforts.

Certain design considerations should be made in developing the biological classifier. The first involves the selecting the appropriate recognition task i.e classification, localization or segmentation. Classification would make predictions on the radar scan level, while localization operates on local regions of range gates. While these approaches have been used to detect biological echoes [5, 46, 47], image/region level prediction limits the scope of these models to specific broad patterns. Like the HCA, a biological classifier for NEXRAD should ideally perform semantic segmentation, where predictions are made at the gate level, making it applicable to a wider variety of biological activity, from broad scale migration events to local roost detection. The biggest anticipated challenge is obtaining ground truth data. One range gate can contain multiple birds, insects and even

other types of echoes at the same time. Furthermore, biota can cover a large area (hundreds of kilometers) and so verifying the exact composition of each gate becomes unfeasible.

The second consideration concerns the selection of output taxa. In clear-air conditions, birds and insects are the most pervasive boundary layer radar echoes observed by NEXRAD [15, 17, 37, 49]. Thus, a minimum biological classifier should be able to separate these two classes. A major challenge is posed by their strongly non-spherical shapes [12]. Moreover, polarimetric measurements heavily depend on their size, shape and orientation [12, 50]. Thus, even in the case of single-species ensembles, polarimetric quantities can vary significantly depending on the azimuthal orientation [8]. This leads to frequently similar features for bird and insect echoes. An example of this similarity is seen in Z_{DR} , which provides information on the shape of scatterers in a resolution volume. The Z_{DR} of Purple Martin colonies have been found to range from -4 to 6 dB [51] while insects have been observed with Z_{DR} values between 2 and 9 dB [52]. This complicates the differentiation of both classes.

Furthermore, a biological classifier should account for the current features of NEXRAD. Owing to their design for weather monitoring, the WSR-88D has a beamwidth of 0.88° (0.96°) at 3 GHz (2.7 GHz) with a range resolution of 250 m. As such, individual range gates are significantly larger than any single bird or insect. This complicates the interpretation of some radar variables for biological classification. For instance, birds are generally larger than insects. Moreover, Z depends on the size of targets within the radar resolution volume. However, Z is also dependent on the number of targets, which means that a large Z value could be from a large bird or a dense aggregation of insects. Another intuitive way birds might be distinguished from insects is by their higher self-generated flight

velocity. However, biological targets also leverage the underlying wind field to aid their flight. As a result, passively flying insects on a windy day could migrate with larger velocities than actively flying birds in a mild wind field. Radial projection and aliasing further complicates the derivation of independent flight speeds from the Doppler velocity.

In light of these considerations, it is hypothesized that an ML based biological classifier can be implemented for NEXRAD. The proposed model would perform segmentation of range gates into birds and insects. Potential input features would include Level II variables and their constructed derivatives. A cost-benefit assessment will be conducted to finalize the best inputs for the model. As for the challenge of obtaining labelled ground truth data, novel data collection techniques will be investigated. Finally, the proposed model will be evaluated for monitoring broad scale migration of biological echoes and improving the quality of VAD wind estimation.

1.4 Contributions

The main contribution of this work is the design, development, and verification of an ML-based bird-insect classifier for NEXRAD, that makes predictions at the range gate level, and is applicable to a wide range of biological activity, including bird roosts, mass bird/insect migration and mixed cases. Furthermore, the classifier was integrated into VAD to generate three new products that can be used to approximate wind contamination levels, obtain more accurate wind estimates from insects and track bird migration velocities. By best knowledge, no other classifier has been demonstrated to improve both HCA's classification of biological echoes and VAD wind estimation.

Other contributions were made during the course of this research. First, a novel data preparation method was proposed for developing gate level supervised machine learning classifiers for birds and insects. This involved extracting bird (insect) migration blobs, remapping radar measurements to their azimuthal aspects and coherently averaging multiple scans to under-emphasize unwanted echoes. Second, HCA misclassifications of biological echoes as weather were discovered during classifier evaluations. A correction procedure was proposed based on ρ_{HV} and its efficacy demonstrated for bird migration cases. Finally, bird (insect) gate counts within height intervals, were used to approximate the population of migrating flocks (swarms). These counts could find applications in aero-ecological studies.

1.5 Scope of the Dissertation

This dissertation is divided into 5 chapters, which details the training/evaluation of the bird-insect classifier, and its application for VAD wind estimation. Chapter 2 discusses fundamental concepts relevant for understanding polarimetric weather radars, the history and nature of clear air echoes, VAD, and machine learning methods used in this work.

Chapter 3 presents the development cycle of the bird-insect ML classifier for HCA’s biological class. First, a technique for extracting bird/insect echoes from migration events is proposed. Using features elaborated in [7, 52], low precipitation scans containing mass bird migration were collected. Echoes would be predominantly birds, though few insects might be found in the remaining gates. Thus, a novel method is proposed to further emphasize bird signatures by mapping radar measurements to their azimuthal aspects, and coherently averaging

multiple remapped scans. The same procedure is repeated to collect insect data from scans dominated by insect swarms. The second phase involved training and evaluating the classifier. Combinations of different inputs features were explored on two machine learning algorithms to generate multiple candidate models. A final model is then selected based on three criteria, including quantitative analysis on the labelled test data, qualitative case studies involving a diverse range of bird/insect activity and model complexity, defined by the number of required inputs. The goal was parsimony in selecting the best performing model.

Chapter 4 explores potential enhancements of VAD wind estimation, using the classifier developed in Chapter 3. Three new products are developed here. They include VAD focused on birds and insects, and the relative proportion of both echoes. Using these products, experiments are conducted over two bird migration periods, in order to study the effect of both taxa on VAD wind bias levels. Lastly, improvement strategies are explored for bird contaminated VADs.

The final Chapter 5 attempts to lay the groundwork for future research with discussions on the limitations of the proposed approaches and suggestions for future improvements.

Chapter 2

Fundamentals of Weather Radars and Machine Learning

In 1900, Nikola Tesla wrote “When we raise the voice and hear an echo in reply, we know that the sound of the voice must have reached a distant wall, or boundary, and must have been reflected from the same. Exactly as sound, so an electrical wave is reflected. we may determine the relative position or course of a moving object such as a vessel at sea, the distance travelled by the same or its speed ” [53]. Radars operate on this fundamental principle of transmitting radio waves and inferring target properties from changes in the received echo. They can operate day and night, irrespective of weather conditions. They might also transmit continuous waves or pulses, with single or dual-polarization.

This work uses the NEXRAD weather surveillance network to observe birds and insects, chosen because of several important features. First, the constituent 160+ pulsed Doppler radars provide good spatial and temporal coverage of the US, collecting data every 5-10 minutes [54]. Moreover, NEXRAD has dual polarization capability, which additionally provides useful information on target’s shape, concentration and diversity of physical properties in a volume, on top of the location, power and velocity normally measured by basic single polarization radars.

A bird-insect classifier for NEXRAD will be subsequently developed and ap-

plied for wind estimation. This chapter details the necessary materials and methods. It is organized as follows: an overview of the WSR-88D fundamentals is provided in Section 2.1, followed by discussions on the nature and origin of clear-air echoes in Section 2.2; Section 2.3 describes the ML algorithms explored for classifying bird and insect echoes followed by the final Section 2.4, describing the VAD technique used for estimating winds as well as bird/insect migration velocities.

2.1 Radar Fundamentals

2.1.1 Important Radar Parameters

A pulse Doppler radar transmits a short burst of radiation, which is intercepted and reflected by a target. Subsequently, the radar received echo is used to deduce the scatterer's distance and velocity. For instance, suppose a round trip time T_R for a target, its range R is the spatial equivalent of this time, calculated as

$$R = \frac{cT_R}{2} \quad (2.1)$$

where c is the speed of light. Furthermore, the impingement of the target by the transmitted pulse generates electromagnetic vibrations according to the Doppler effect. Relative to the incident wave, stationary targets would vibrate at the same frequency, approaching targets at higher frequencies and receding targets with lower frequencies. The Doppler frequency f_d measures this frequency shift,

$$f_d = - \frac{2V_r}{\lambda} \quad (2.2)$$

where λ is the radar wavelength and V_r is the radial velocity.

Several parameters are considered during radar design, in order to maximize the quality of range and velocity detections. One of these parameters is the radar pulse width τ , which determines the minimum separation distance ΔR in range, for which two scatterers are resolvable. This distance is called the range resolution and is improved (degraded) by shorter (longer) pulse widths. Their mathematical relationship is given by

$$\Delta R = \frac{c\tau}{2} \quad (2.3)$$

In practice, radar echoes are contaminated by noise from the system and the environment. As such, multiple pulses are transmitted and their received wave forms, averaged to improve the Signal to Noise Ratio (SNR) of measurements. This transmission routine is described by the Pulse Repetition Time (PRT), T_p or the Pulse Repetition Frequency (PRF), $f_p = \frac{1}{T_p}$. The PRT affects the maximum unambiguous range and velocity measurable by the radar. Range is deduced from the time delay caused by targets, relative to when the last pulse was transmitted. Thus, echoes from a previous pulse m arriving after a new pulse $m + 1$ has been sent out, would be measured with an aliased delay of $T_R - mT_p$ instead of the true delay of T_R . A longer listening period T_P raises the maximum unambiguous range R_{un} , expressed as

$$R_{un} = \frac{cT_p}{2} \quad (2.4)$$

Conversely, the resolution of fast moving targets requires pulses to be collected in quick succession. Therefore, the maximum unambiguous velocity V_{un} that the radar can measure is inversely proportional to T_p ,

$$V_{un} = \pm \frac{\lambda}{4T_p} \quad (2.5)$$

The result of Equations (2.4) & (2.5) is the well known radar dilemma where large R_{un} and V_{un} are desired, however any choice of T_p simultaneously maximizes one and minimizes the other. Ultimately, there are no solutions only trade-offs.

2.1.2 Point Target Radar Equation

The radar equation relates the received power to various radar and target parameters. Consider a transmitter that radiates power P_t isotropically. This power will spread on the surface of an expanding sphere, so that the power density at a given range R is $\frac{P_t}{4\pi R^2}$. However, the WSR-88D uses a parabolic dish antenna to concentrate radiation into a narrow beam. Along the axis of this beam, the ratio of the resulting concentrated power to isotropic power per unit area, provides a measure of the antenna's gain G . A target located at range R with an RCS of σ will intercept a power of $\frac{P_t G}{4\pi R^2} \sigma$, and assuming no absorption, reflect this power in all directions. Back at the radar, the received power P_r is given by

$$P_r = \frac{P_t G}{4\pi R^2} \sigma \frac{1}{4\pi R^2} A_e = \frac{P_t G A_e}{(4\pi)^2 R^4} \sigma \quad (2.6)$$

where A_e is the effective aperture of the antenna [55]. The relationship between antenna gain and effective area, and the radar wave length were shown from theoretical considerations [55, 56] as

$$A_e = \frac{G \lambda^2}{4\pi} \quad (2.7)$$

Substituting Equation (2.7) into Equation (2.6), and accounting for power

losses, the received power becomes

$$P_r = \frac{P_t G^2 \lambda^2}{(4\pi)^3 R^4 L} \sigma \quad (2.8)$$

as shown in [53, 55, 57, 58], where L is the loss factor, including the effect of wave guide losses, antenna inefficiencies, beam attenuation, receiver bandwidth limitations and other factors. For dish antennas, gain can be approximated using

$$G = \frac{\pi^2}{\theta_B^2} \quad (2.9)$$

where θ_B is the half power beam width (in radians).

The radar Equation (2.8) depends on scatterers being in the far field of the antenna. Radar beams are initially transmitted with diameters equal to the antenna, but progressively change shape due to diffraction effects. By a distance of $\frac{2D_a^2}{\lambda}$, where D_a is the antenna's diameter, the beam assumes the idealized conical shape and Equation (2.8) holds. This is around 1445 m for the WSR-88D. Conversely, when $R < \frac{2D_a^2}{\lambda}$, the beam width is larger than what is assumed by the radar equation and without additional considerations, overestimates the received power. It can however be used approximately, by assuming a beam width equal to D_a , up till a distance of D_a/θ .

The overall evolution of a radar pulse can be summarized at the transmission and reception stages. Weather radars transmit a pulse $x(t)$, which has the form,

$$x(t) = a(t) \text{rect}\left(\frac{t}{\tau}\right) \exp[-j(2\pi f_t t + \psi_t)] \quad (2.10)$$

where $a(t)$ is the amplitude of the wave, f_t is the carrier frequency, and ψ_t is the transmitter phase. For a target with range R and radial velocity V_r , the

received IQ signal can be represented as

$$V(f, t) = \frac{|A|}{2} \text{rect} \left(\frac{(t - \frac{2R}{c})}{\tau} \right) \exp \left[-j \left(\frac{4\pi}{\lambda} R + \frac{4\pi}{\lambda} V_r t - \psi_t - \psi_s \right) \right] \quad (2.11)$$

where ψ_s is the phase shift after scattering.

2.1.3 Weather Radar Equation

The simple radar equation for a point target was developed in the previous section. However, in real world applications of weather radar, the beam illuminates a volume filled with many scatterers. In this case, the received signal is the sum of reflections from these scatterers, accounting for phase. Its power can be expressed as

$$P_r = \frac{P_t G^2 \lambda^2}{(4\pi)^3 R^4 L} \sum_i \sigma_i \quad (2.12)$$

where $\sum_i \sigma_i$ is total back-scattering cross section from the volume. Assuming a uniform distribution of scatterers, this total RCS can be rewritten as

$$\sum_i \sigma_i = \Delta V \sum_{\Delta v} \sigma_j \quad (2.13)$$

where Δv is represents a unit of volume and ΔV is the total volume.

The smallest volume within which all echoes are integrated is called the resolution volume (or cell). Its angular and range dimensions are defined by the beam width and pulse width respectively. NEXRAD produces a symmetrical beam, with a width θ_B related to the antenna diameter and radar wavelength, as shown below

$$\theta_B = \frac{1.27\lambda}{D_a} \quad (2.14)$$

[53]. The pulse width determines range resolution. Thus, the resolution cell has dimensions of $R\theta_B$ by $R\theta_B$ by ΔR , and a total volume ΔV approximated as

$$\Delta V = \pi \left(\frac{R\theta_B}{2} \right)^2 \frac{\Delta R}{2} \quad (2.15)$$

Substituting (2.15) into (2.13), and the resulting total radar cross section into (2.12), the received power becomes

$$P_r = \frac{P_t G^2 \lambda^2 \theta_B^2 \Delta R}{512 \pi^2 R^2 L} \sum_{\Delta V} \sigma_i \quad (2.16)$$

[55]. A power overestimation problem was subsequently observed in experiments using known $\sum_i \sigma_i$. This was addressed by a more rigorous derivation, using a paraboloidal antenna with circular cross section, assuming a Gaussian representation for the power per unit area within the main lobe, accounting for differences between two-way and one-way antenna power density patterns, and assuming the interception of the entire beam by targets [55, 58]. The received power becomes

$$P_r = \frac{P_t G^2 \lambda^2 \theta_B^2 \Delta R}{1024 (\ln 2) \pi^2 R^2 L} \sum_i \sigma_i \quad (2.17)$$

Since meteorological radars are designed to observe water drops, the Rayleigh approximation for targets much smaller than the radar wavelength, is assumed to hold. Here, the target diameter $D < \frac{\lambda}{16}$ (~ 6.25 mm for NEXRAD at $\lambda = 10$ cm) and RCS increases with the target's size. Mathematically, this RCS can be expressed as

$$\sigma_{Ray} = \frac{\pi^5}{\lambda^4} |K_m|^2 D^6, \quad (2.18)$$

where K_m is

$$K_m = \frac{(m^2 - 1)}{(m^2 + 2)}, \quad (2.19)$$

and m is the complex refractive index of water. For water at microwave wavelengths, $|K_m|^2$ is about 0.93.

Reflectivity factor Z is the total back scattering cross section per unit volume. Using the Rayleigh approximation in Equation (2.18), Z can be expressed as

$$Z = \frac{\sum_i D_i^6}{\Delta V} = \frac{\lambda^4}{\pi^5 |K|^2} \frac{\sum_i \sigma_{Ray}}{\Delta V} \quad (2.20)$$

When the target size exceeds the Rayleigh limit, the calculated Z becomes an effective value. Given their interchangeable usage in common parlance, it is worth noting that reflectivity factor is different from reflectivity. Reflectivity is defined as

$$\eta = \frac{\sum_i \sigma_i}{\Delta V} \quad (2.21)$$

Substituting (2.20) into (2.17), yields

$$P_r = \frac{P_t G^2 \theta^2 \Delta R \pi^3 |K|^2}{1024 (\ln 2) R^2 \lambda^2 L} Z \quad (2.22)$$

If a resolution volume contains a single target near the beam axis, its RCS can be obtained by equating Equation (2.22) and (2.8), as shown below

$$\sigma = \frac{\pi^6 |K|^2 R^2 \theta^2 \Delta R}{16 (\ln 2) \lambda^4} Z = 80.6 \frac{R^2 \theta^2 \Delta R}{\lambda^4} Z \quad (2.23)$$

Alternatively, Equation (2.23) gives the total RCS of scatterers in a volume. Substituting Equation (2.9) into Equation (2.22), the received power becomes

$$P_r = \frac{P_t \Delta R \pi^7 |K|^2}{1024 (\ln 2) R^2 \lambda^2 \theta^2 L} Z \quad (2.24)$$

The echo voltage received from a volume of scatterers can be represented as

$$V(T_p) = \frac{1}{\sqrt{2}} \sum_{i=1} A_i W_i \exp \left[-j \left(\frac{4\pi}{\lambda} R + \frac{4\pi}{\lambda} v_r T_p - \psi_t - \psi_s \right) \right] \quad (2.25)$$

where $\frac{|A_i|}{\sqrt{2}}$ is the prefilter echo amplitude and W_i is a range dependent weighting factor [53].

2.1.4 NEXRAD

NEXRAD is a network consisting of 160+ pulsed Doppler WSR-88Ds, operated by the National Weather Service. Figure 2.1 shows the distribution of stations across the US. Each radar is capable of detecting a point target with 4 cm² RCS (equivalent to a single bird) at 100 km. Other system specifications are presented in Table 2.1.

The WSR-88D senses the environment by collecting 360° azimuthal sweeps at different elevations, using a rotating parabolic dish antenna. Data collection is performed at two resolutions, depending on meteorological conditions and local policy [60]. The first is standard resolution, which involves azimuthal sampling every 1 degree, amounting to 360 radials per sweep. Super resolution, on the other hand, involves sampling every 0.5 degree for a total of 720 radials per elevation. This is normally used for the lowest 2 or 3 elevations. In both cases, range

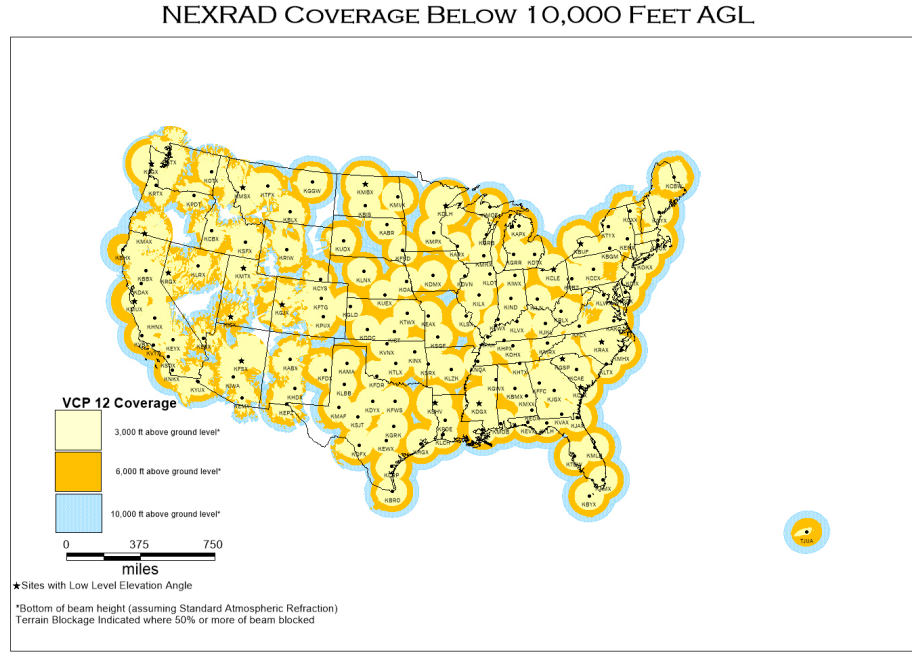


Figure 2.1: NEXRAD coverage of Contiguous US [59].

resolution is 250 m, reflectivity data are collected up to 460 km, while Doppler and dual polarization data are collected up to 300 km [60].

WSR-88D operates in different PRT modes depending on the elevation angle. A split cut mode is used for low elevations, where two successive sweeps are completed. The first sweep uses a long PRT to retrieve the unambiguous power from which reflectivity and dual-pol variables are estimated. This is also known as the contiguous surveillance mode. The second contiguous Doppler sweep uses a short PRT to generate more pulses necessary for the estimation of velocity, spectrum width and reflectivity. At intermediate elevation angles, a batch mode is used, where the radar completes one sweep of interlaced long and short PRTs. Long PRTs are used to recover unambiguous power while short PRTs are used for estimating range folded velocities. Higher elevations only use the Doppler mode since range folding is not an issue.

Furthermore, NEXRAD has two operational modes depending on the estimated precipitation levels namely: clear air and precipitation modes. Each mode contains different Volume Coverage Patterns (VCP), which are specific routines consisting of azimuthal sweeps at distinct elevation angles. Precipitation mode has VCPs 11, 12, 21, 121, 211, 212 and 221 while clear air mode has VCPs 31 and 32. Further information on VCPs can be found in [60].

Table 2.1: System specifications of WSR-88D [61, 62]

Transmitter	
Operating frequency	2.7 – 3 GHz
Wavelength	10.0 – 11.1 cm
Transmit power (peak)	750 kW
Polarization	Dual (simultaneous H and V transmit/receive)
Pulse width	1.57, 4.7 μ s (235 – 705 m)
Antenna	
Diameter	8.5 m (parabolic dish)
3 dB beamwidth (at 2850 MHz)	0.925°
Gain (2850 MHz)	45.5 dB
First sidelobe	-29 dB
Maximum Rotation rate	30°/s
Receiver	
Sensitivity, long pulse	-7.5 dBZ _e at 50 km
Sensitivity, short pulse	-23.0 dBZ _e at 25 km
Gate Spacing	250 m
A/D convertor bits	16 bits
Point target detection	4 cm ² at 100 km

2.1.5 Level II Products

The WSR-88D has three main data levels. Level I data is the raw time series waveform received by the Receiver/Signal Processor (RSP) [63]. In native form, level I data is not used operationally, but is useful for the development and verification of algorithms. Data collection is therefore only on special occasions determined by the Radar Operations Center (ROC). Level I data is processed to obtain and

record level II data, initially comprised of three legacy single polarization variables. However, following the upgrades to dual polarization completed in 2013 [7], three dual-pol variables are also produced using measurements from Horizontal (H) and Vertical (V) polarized waves. Level III data are the output products from the Radar Product Generator (RPG) [64]. They are used for the detection of atmospheric features like precipitation, storms, turbulence and wind, operational forecasting and data research analysis [64, 65]. An extensive list of the over 40 level III products can be found in [66].

This work focuses on level II products for detecting bird and insects. They include three base moments: reflectivity factor Z , Doppler velocity V_r and spectrum width σ_v , and three dual polarization variables namely: differential reflectivity Z_{DR} , cross-correlation coefficient ρ_{HV} and differential phase φ_{DP} . The rest of this section details how these level II products are estimated.

For a given resolution volume, the radar collects M samples of raw voltages i.e $V(0), V(1), \dots, V(M-1)$. The time evolution of these signals is measured by the Auto Correlation Function (ACF), defined as

$$\hat{R}(lT_p) = \begin{cases} \frac{1}{M-|l|} \sum_{m=0}^{M-|l|-1} V^*(m)V(m+l), & \text{for } |l| \leq M-1. \\ 0, & \text{otherwise.} \end{cases} \quad (2.26)$$

where l is the time lag. Larger ACF values at a given time lag suggests stronger correlation of a signal with its time shifted version. Wider ACFs, as a function of l , imply a slower decorrelation rate. In practice, the narrowest ACF is the delta function of white noise, due to its independence from sample to sample.

The base moments are obtained from the horizontal channel, using ACFs at

lags 0 and 1, calculated in Equations (2.27) and (2.28) respectively. $\hat{R}(0)$ estimates average power.

$$\hat{R}(0) = \frac{1}{M} \sum_{m=0}^{M-1} |V(m)|^2 \quad (2.27)$$

$$\hat{R}(1) = \frac{1}{M-1} \sum_{m=0}^{M-1} V^*(m) V(m+1) \quad (2.28)$$

Dual-pol variables, on the other hand, are estimated from voltages in both horizontal H and vertical V channels. Hence, a cross correlation at lag 0 is additionally computed as

$$\hat{R}_{HV}(0) = \frac{1}{M} \sum_{m=0}^{M-1} V_H^*(m) V_V(m) \quad (2.29)$$

Reflectivity Factor

Reflectivity factor Z is directly proportional to the size or abundance of scatterers in a resolution volume (see Equation (2.20)). It is estimated from the mean returned power in the horizontal channel S_H , as shown in Equations (2.30) and (2.31) below.

$$Z = \text{Log}_{10}(S_H) \quad (2.30)$$

$$S_H = \hat{R}(0) - N \quad (2.31)$$

In practice, further corrections are needed for factors like range, radar specifications and system losses.

Doppler Velocity

Doppler velocity V_r is determined from the power weighted mean frequency shift of targets in a resolution volume. It measures the velocity component along radials and is thus also called radial velocity. A target moving parallel to the beam, would have V_r equivalent to its speed. Conversely, targets moving perpendicular to the radar beam, would have V_r close to zero. Furthermore, negative values imply approaching targets while positive values indicate recession.

Radial velocity can be estimated using pulse pair processing, which is a simple and computationally efficient approach based on the lag 1 ACF, that does not require expensive Fourier calculations. Mathematically,

$$V_r = -\frac{\lambda}{4\pi T_p} \arg[\hat{R}(1)] \quad (2.32)$$

Spectrum Width

Spectrum width σ_v measures the consistency of phase changes between successive pulses. It provides information on the variation of radial velocities within a resolution volume. For instance, higher σ_v can be expected from turbulent meteorological conditions. Spectrum width can be represented by

$$\sigma_v = \frac{V_{un}\sqrt{2}}{\pi} \left| \ln \left(\frac{R(0) - N}{R(1)} \right) \right|^{1/2} \quad (2.33)$$

Differential Reflectivity

Differential Reflectivity Z_{DR} is the difference between power returned from the vertical and horizontal channels. The shape of targets can be inferred from Z_{DR} . For instance, spherical scatterers that return equal power in both channels will

have $Z_{\text{DR}} \sim 0$ dB, targets with horizontal major axis alignment will have $Z_{\text{DR}} > 0$ dB, while negative Z_{DR} indicate targets with vertically aligned major axis [67]. Mathematically,

$$Z_{\text{DR}} = 10 \text{Log}_{10} \left(\frac{S_H}{S_V} \right) \quad (2.34)$$

where S_V is the average power in the vertical channel.

Differential Phase

Electromagnetic waves accrue phase shifts as they propagate through atmospheric media. The differential phase, φ_{DP} is the difference of these shifts from the vertical and horizontal channels. It is proportional to the concentration and size of particles in a volume [67]. For example, a range gate that is densely filled with precipitation will have a higher φ_{DP} than a sparsely filled one. Differential phase can be calculated as the angle subtended by the vector multiplication of pulses from the H and V channels, as shown in Equation 2.35 and Figure 2.2 below.

$$\varphi_{\text{DP}} = \angle \hat{R}_{\text{HV}}(0) \quad (2.35)$$

Cross-correlation Coefficient

Cross-correlation coefficient ρ_{HV} measures the consistency of returned power and phase between both channels. It captures the diversity of physical target characteristics like type, shape and orientation [67]. A resolution volume with homogeneous targets would have $\rho_{\text{HV}} \sim 1$. For example, rain has been observed with $\rho_{\text{HV}} > 0.92$ while biological scatterers have $\rho_{\text{HV}} < 0.83$ [29]. This is consistent with rain having more uniform shape and distribution, in comparison to

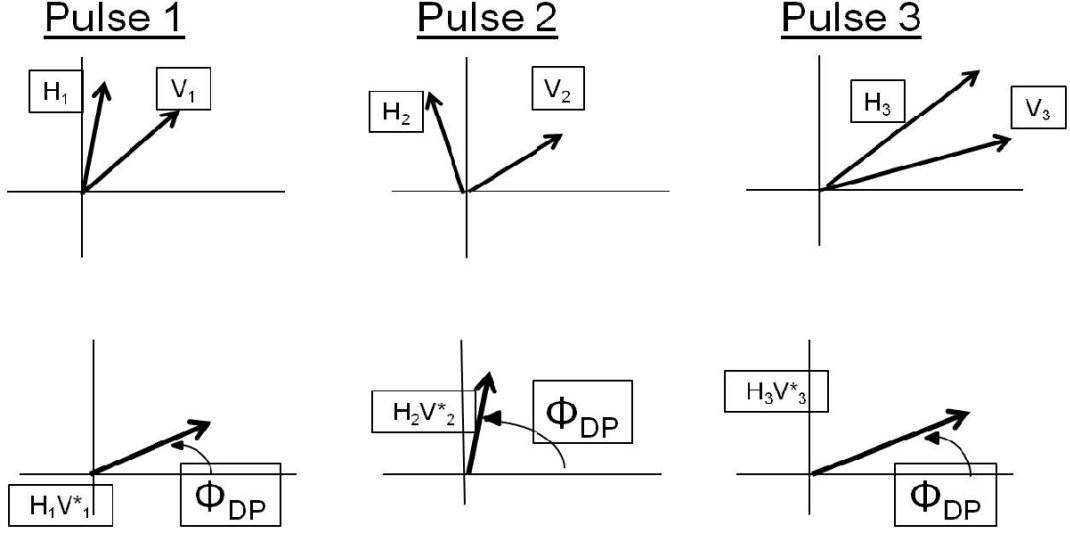


Figure 2.2: Calculation of φ_{DP} for three distinct pulses. [68]

the diversity of biological scatterer sizes and orientations. Volumes comprised of chaotic scatterers will have $\rho_{HV} \sim 0$. Mathematically,

$$\rho_{HV} = \frac{|\hat{R}_{HV}(0)|}{[S_H S_V]^{1/2}} \quad (2.36)$$

2.2 Clear-air Echoes

Clear-air echo is radar return from a region of the atmosphere with no clouds or precipitation [69]. Existing research has identified four main causes of such return. They include birds, insects, particulate matter and spatial fluctuations of refractivity. Birds are especially common during migration seasons, where their significant speeds of $10\text{-}20 \text{ ms}^{-1}$ [70], have been found to contaminate radar wind estimates. Insects are more ubiquitous and generally wind borne, except in cases of congregated alignment where the group generates its own velocity [71]. As such, they can be used to trace wind motion. Birds and insects are the predominant

source of boundary layer clear-air echoes [15, 17, 37, 49], especially at S-band.

Particulates like smoke and dust occasionally contribute to clear air returns, however, a dense aggregation is needed to create a detectable cloud because of their minuscule sizes [70]. Other returns arise from refractivity fluctuations. An example is turbulent Bragg scatter, where the mixture by turbulent flow across refractive index gradients, creates a field of refractivity perturbations that reflects radar signals. Their presence has been confirmed by agreements between expected and measured reflectivity, using radars of different wavelengths [72]. Bragg scatter is weak at S-band. Other specular reflections off refraction gradients have been observed using vertical pointing radars [73, 74], however S band wavelengths are too short to cause such reflections.

The rest of this section further describes general characteristics of clear-air echoes. An underlying trend observed for predominating birds and insects is respective nocturnal and daytime activity. As such, discussions on the daily cycle of clear-air echoes are also presented herein.

2.2.1 Characteristics of Clear-air Echoes

Clear-air echoes are typically called 'angels' in literature [70]. They have been attributed to birds by ornithologists [6, 75] and meteorologists [52, 76, 77], however, other studies have proposed insects [49]. Clear-air echoes can occur as isolated targets or contiguous layers of Z [70]. Furthermore, they are often granulated, with different graininess across night and day [78]. Nighttime returns are typically coarser, indicative of larger particulates [70].

Symmetrical patterns have been observed in clear-air PPIs (Plan Position Indicator) of dual-polarization variables [28]. This can be explained by the changing

RCS of non-spherical targets with radar viewing angle. During mass movement of these scatterers towards a particular azimuthal direction, the rotating antenna samples different aspects as they approach, fly over/around, and recede the radar location. Thus, the aforementioned symmetry patterns have been attributed to mass aligned movement of birds [79] and insects [6].

Expanding rings of Z is another feature of clear-air echoes. They usually appear around morning time during certain periods of the year. Gravity waves were initially postulated as causing these rings [80], however, recent research has shown that this is due to birds leaving their roosting sites, evidenced by persistent emanation from the same location [6, 55, 75]. Other causes of Z rings include dusk bat departure from roosting sites and convective cells [53] often observed as stagnant 1 - 3 km diameter rings [70].

Seasonal fluctuations have been observed in clear-air Z . In the Great Plains, late spring often has the strongest values during nighttime, with daily fluctuations as high as 20 dBZ [70]. This high nocturnal Z is coincident with expected bird migration. Furthermore, thin lines of Z are commonly observed in daytime clear-air returns. They have been attributed to insects gathering at meteorological boundaries [49]. Large and sharp refractive index gradients also occur at these boundaries, however, specular reflection can be ruled out because of the short wavelength at S-band. Other studies analyzed the vertical flight of scatterers using profiling airborne radar data [81]. They found insect plumes were often collocated with updrafts, and that micro-insects tend to fall or fly down against updrafts with an average speed of $0.5 \pm 0.2 \text{ ms}^{-1}$.

2.2.2 Daily Cycle of Clear-air Echoes

Changes in the nature of clear-air echoes have been observed with the time of the day [22, 52, 70, 78]. For instance, daytime returns recorded higher Z_{DR} and lower φ_{DP} than nighttime [52]. These were attributed to birds and insects respectively. Another work discovered a unique daily cycle in clear-air returns [70]. Daytime Z was found to be weak and confined to the lowest 1 km. Nighttime Z , on the other hand, was much stronger, typically ≥ 10 dBZ, but reaching up to 25 dBZ in exceptional cases. Moreover, sunrise and sunset were notable inflection points. Steadily decreasing nighttime Z hits a minimum at sunrise after which there is a rapid increase to modest daytime values. A similar dip was observed at sunsets, followed by rapid increase in Z and height of returns. An example of this cycle is shown in Figure 2.3, which shows average Z below 2 km measured by the Cimarron radar on the night of May 31, 1999. This cycle implies a clear change in nature (probably taxa) of scatterers between day and night. More recent research [22] has attributed this phenomenon to nocturnal bird migration and daytime insect domination.

Nocturnal Clear-air Echoes

Migratory birds have been found to travel long distances at night. Thus, nocturnal echoes during bird migration seasons are often bird dominated. The National Oceanic and Atmospheric Administration’s (NOAA) Environmental Technology Lab (ETL) considers this a severe problem, and routinely flags low level nighttime radar wind profiler measurements as bird contaminated [70, 82–84]. Moreover, similar issues have been reported for NEXRAD’s VAD wind profiles [77, 85] and long wavelength wind profilers [84]. The latter recorded differences as large as 15

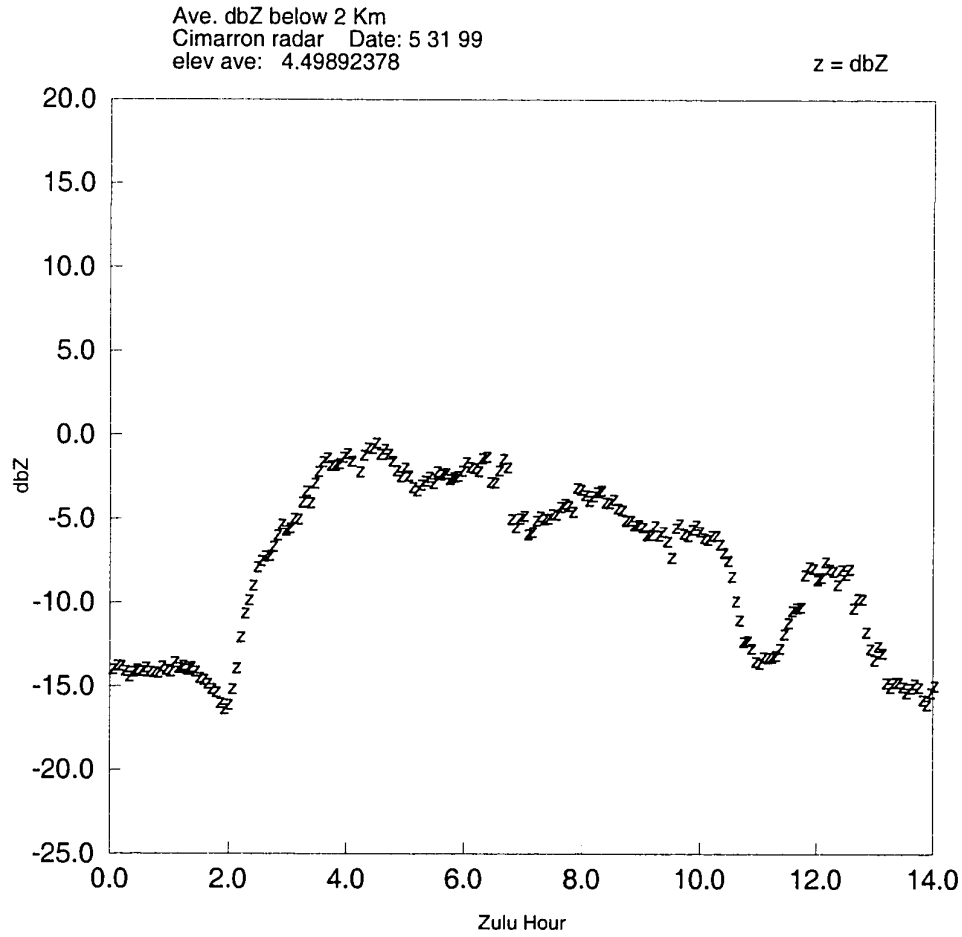


Figure 2.3: Average Reflectivity below 2 km versus time for a nocturnal case. Sunset is ~ 2 Z while sunrise is ~ 11 Z. Data was collected using the Cimarron radar on May 31, 1999. [70]

ms^{-1} .

Though often bird dominated, the presence of insects should not be discounted in nocturnal echoes. A count of nighttime echoes with a 1.5 m resolution FM-CW radar, found a density of 1 echo per 12 meter cube over a depth of 500 m [86]. This implies about 46 billion members over the state of Oklahoma, which are unlikely to be solely from birds [70]. Furthermore, birds have been observed with Z ranging from 5 to 15 dBZ [6]. Specifically, one bird in a radar probe volume

can account for 10 dBZ of echo [77]. This would imply an improbable 500 million birds at the instant of a radar scan, based on a probe volume of 100-meter cube and 1 bird per volume over the state of Oklahoma, and through a depth of 3 km [70]. Once again, other scatterers must be present in nocturnal echoes to explain this number.

Daytime Clear-air Echoes

Early studies in the 1940's concluded that most clear-air echoes are caused by insects [87]. This was based on the difficulty of creating refractive index inhomogeneities that are strong enough to be perceived by radar, and visual confirmation of insect presence coincident with radar observations. Subsequent work have further constrained predominant insect activity to daytime.

Bernard-like cells with unusually high reflectivity, have been seen during daytime, concurrent with observations of an abnormal number of airborne ants [35]. Another work used 3.22 and 0.86 cm radars to study clear-air Z from the Convective Boundary Layer (CBL) during the day in Colorado. They found scatterers to be wind borne, based on differences $\sim 0.2 \text{ ms}^{-1}$ between VAD and anemometer wind measurements. Refractive index gradients were ruled out because of high observed Z of -15 to 5 dBZ (for the 3.22 cm radar), and an absence of Z maxima near inversion heights. Thus, these daytime clear-air echoes were attributed to insects, seeds and particulates. Other studies using different wavelength radars, have come to a similar conclusion [49].

Despite the aforementioned trend, it should be noted that birds can be present during the day. For example, diurnal migrants favour daytime for their travels. The following quote from [70] emphasizes the diversity of avian activity: “Meteorologists should probably accept that it is possible for birds to be migrating at

any time of the day or night, on any day of the year, with any relationship to the weather and in any direction”.

2.3 Machine Learning (ML) Methods Used in this Dissertation

Traditionally, fuzzy logic has been used for classifying weather radar echoes. However we opted for a supervised ML approach because models trained this way can easily be updated as new data is collected. These ML methods learn a model that minimizes prediction errors on the training data. As such, they generally improve as more data is provided. This is not necessarily the case with fuzzy logic where the classification rules and membership functions are manually defined. Consequently, an established ML pipeline would allow future research efforts to be focused on much needed data collection and verification. Two supervised ML approaches are explored in this dissertation. They include the ridge classifier and decision tree.

The ridge classifier learns a linear function of input variables that achieves the best separation between classes in the output. We used the SGDClassifier [88] in scikit-learn. Let us consider a model with only dual polarization variables as inputs. For a single range gate, the function is given as

$$f(w, b) = \begin{vmatrix} w_{Z_{DR}} & w_{\Phi_{DP}} & w_{\rho_{HV}} \end{vmatrix} \begin{vmatrix} Z_{DR} \\ \Phi_{DP} \\ \rho_{HV} \end{vmatrix} + b \quad (2.37)$$

where $w_{Z_{DR}}$, $w_{\Phi_{DP}}$, $w_{\rho_{HV}}$ are the weights for Z_{DR} , Φ_{DP} , and ρ_{HV} respectively and

b is a bias term. It can also be expressed as

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (2.38)$$

where $\mathbf{w}$ is the weight vector and $\mathbf{x}$ is the input vector. The goal is to find parameters that minimizes the log loss error given by

$$L(y_i, f(x_i)) = \log(1 + \exp(-y_i f(x_i))) \quad (2.39)$$

where y_i is the label for each training example. A scaled L2-norm of the weights is added to the above loss in order to penalize weight explosions. The final loss function is given by

$$C(\mathbf{w}, b, \alpha) = \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i)) + \alpha \|\mathbf{w}\|_2^2 \quad (2.40)$$

where n is the number of training examples and α controls the effect of the weight penalty. The learning process uses stochastic gradient descent [89] on w and b , and a search on α to find values that minimize $C(w, b, \alpha)$.

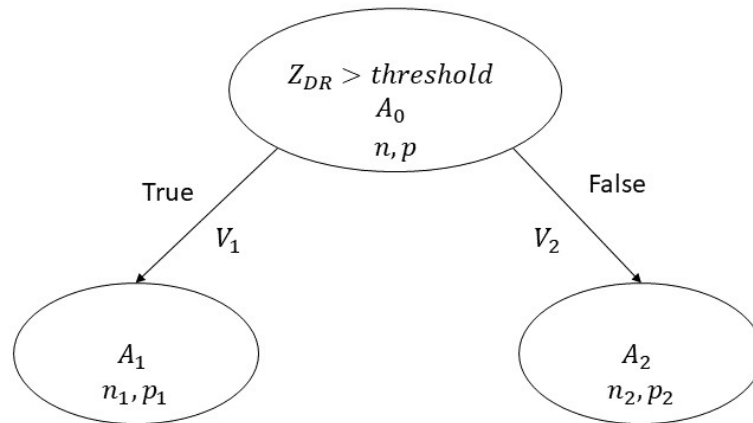


Figure 2.4: Part of decision tree, containing three attributes

Our second technique, decision trees, learn rules to recursively partition data so that samples with the same labels are grouped together. We used the Decision-TreeClassifier [88] from scikit-learn. A part of a decision tree is shown in Figure 2.4. Let us define some terms before discussing how decision trees learn. An attribute A is a question asked about the data e.g is $Z_{DR} > \text{threshold}$? Answers to this question, like True or False, are called values V_k and are used to partition the data set. There are also two classes c_j containing p positive examples (birds) and n negative examples (insects). The entropy of an attribute measures its homogeneity. It is defined as

$$E(p(c_j), \dots, p(c_m)) = \sum_{j=1}^m -p(c_j) \log_2 p(c_j) \quad (2.41)$$

where $p(c_j)$ is proportion of the j th class. High entropy indicates a uniform distribution over classes while low entropy indicates the dominance of some classes. Information gain measures the reduction in entropy for a given split. It is defined as

$$G(A) = E(p(c_j) \dots p(c_m)) - \sum_{k=1}^l \frac{n_k + p_k}{n + p} E(p(c_j) \dots p(c_m) | V_k) \quad (2.42)$$

where n_k and p_k are the number of positive and negative examples respectively in the k th split. In other words, $G(A)$ is the difference between the entropy before a split and the mean entropy after the split. Decision trees learn by finding attributes that maximizes $G(A)$. They can also be thought of as a graph, comprising nodes and edges. Nodes in a tree can be any/both of two types: parent nodes which are attributes, and children nodes, the offspring of parent nodes. Terminal nodes are exclusively children nodes. We applied minimal cost complexity pruning [90] which regularizes the tree by penalizing the number of terminal nodes.

2.4 Velocity Azimuth Display (VAD)

VAD is a wind estimation technique developed in the sixties [91, 92], which involves analyzing radial velocities as a function of azimuths at a given height. These velocities form a sinusoid, from which wind speed and direction can be determined. Figure 2.5 below shows an example of a VAD sine wave with a peak around the 70° azimuth. During mass bird (insect) migration, VAD can be used to estimate their common migration velocity. This section describes the model used for such retrievals in this dissertation.

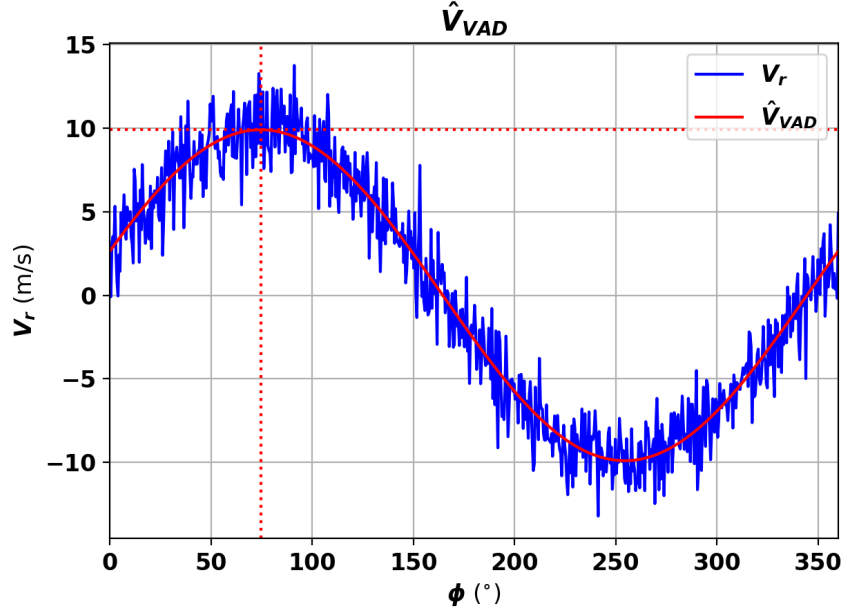


Figure 2.5: VAD applied to obtain speed and direction of migration. Echoes are found to move toward 70° with a speed of 10 ms^{-1}

Assuming uniform migration of organism taxa c at height h and time t , the common migration velocity is recovered by harmonic analysis of radial velocities in the region. First, a sinusoidal model $\hat{V}_{VAD}$ is fitted to these radial velocities,

$$\hat{V}_{VAD}(h, \phi, t, c) = |V_{mig}(h, t, c)| \cos(2\pi f \phi + \delta(h, t, c)) \quad (2.43)$$

where $|V_{mig}(h, t, c)|$ is their migration speed, ϕ is the radar's azimuth, f is frequency (assumed to be $\frac{1}{360}$ cycles per degree for a uniform wind field) and δ is the phase offset.

The direction of migration $\phi_{mig}(h, t, c)$ is defined as the common destination of organisms, calculated as the azimuth with the maximum radial velocity (Equation (2.44)). This maximum velocity is the migration speed (Equation (2.45))

$$\phi_{mig}(h, t, c) = \arg \max_{\phi} \hat{V}_{VAD}(h, \phi, t, c) \quad (2.44)$$

$$|V_{mig}(h, t, c)| = \hat{V}_{VAD}(h, \phi_{mig}, t, c) \quad (2.45)$$

Altogether, the migration velocity $\vec{V}_{mig}(h, t, c)$ is defined as,

$$\vec{V}_{mig}(h, t, c) = |V_{mig}(h, t, c)| \angle \phi_{mig}(h, t, c) \quad (2.46)$$

In the example of Figure 2.5, the red line is the $\hat{V}_{VAD}$ sine fit, from which migration speed and direction are found to be 10 ms^{-1} and 70° respectively. Thus, $\vec{V}_{mig} = 10 \text{ ms}^{-1} \angle 70^\circ$.

Chapter 3

A Bird-Insect Ridge Classifier (BIRC) for S-band Polarimetric Weather Radar

3.1 Introduction

The main hindrance to realizing biological applications of NEXRAD lies in its inability to distinguish biological echoes by taxa. The operational HCA categorizes birds and insects, for example, into a broad biological class. In this chapter, a machine learning model is developed to further classify these biological echoes as birds or insects. A report on the development process is presented herein. Roughly, this involves four parts namely: data preparation, model training and evaluation, case studies and final recommendations.

For data preparation, dual polarization radar scans containing separate large scale bird and insect migration were collected in Section 3.2, after which their textures were computed in Section 3.3.1. Next, blob coloring was used to extract these migration echoes in Section 3.3.2 and VAD, applied to obtain a new azimuth relative to the aspect of birds (insects) in Section 3.3.3. Based on the reoriented azimuths, multiple bird (insect) scans are averaged to reduce contamination by insects (birds), in order to derive training inputs for the classifiers (Section 3.3.4). These inputs are described in Section 3.3.5.

The next part involves model training and evaluation. Two supervised ML algorithms are investigated: ridge classifier and decision tree. The necessary evaluation metrics are described in Section 3.5, followed by the training process for both ML methods using different input combinations (Section 3.6). Section 3.7 concludes this stage with performance assessments of generated models using metrics computed on the test data. In the third part, further case studies (Section 3.8) on migration and bird roosting cases are used to analyze performance on new scenarios. Based on the aforementioned assessments, a final model is selected and recommendations for its operational implementation are presented in Section 3.9.

3.2 Data Collection

3.2.1 Selection of Bird and Insect Scans

Finding the ground truth for biological radar data, at each individual range gate can be challenging. First, a radar resolution volume is much larger than individual biological targets. As a result, one volume could contain multiple birds, insects and even other types of echoes at the same time. Second, biological echoes can span hundreds of kilometers and so it is impossible to verify the exact composition of scatterers within each radar range gate. Our approach was to collect multiple scans of mass bird (insect) migration in clear-air, where we expect to obtain the highest possible number of range gates containing birds (insects). A substantial part of nocturnal echoes in spring and fall are contributed by migrating birds [52]. Such migration is characterized by a large area of echoes centered around the radar site with velocities moving in the same direction (highly aligned). Further analysis of Purple martins in [7] has shown that they have modes around 0 dB for Z_{DR} and 110° for φ_{DP} . On the other hand, insects are routinely observed in

clear-air during warm seasons in Oklahoma, reaching peak intensity in the late afternoon [52]. Analyses have also shown that insects have modes around 6 dB for Z_{DR} , with most echoes saturating at the 8 dB limit of the WSR-88D and φ_{DP} modes around 70° [7]. It should be noted that both set of echoes show azimuthal dependence for most cases. Furthermore, bird migration echoes typically occupy a larger area compared to insect migration echoes at the 0.5° degree elevation scan.

We used these properties to select 90 clear-air scans from the 0.5° elevation of KTLX (Oklahoma City, OK). Specifically, 45 PPIs contained nocturnal bird migration from the spring period of 3-5 UTC during April - May, 2015 - 2017. The other 45 PPIs contained mass daytime insect swarms collected over 16-23:59 UTC, June - August, 2017 - 2019. These scans were determined as clear-air because few or no precipitation echoes were observed. As such, the majority of echoes were due to biological migration activity. This is important for the subsequent extraction of migration echoes by blob coloring. Finally, all gates with range less than 10 km were excluded to reduce contamination by ground clutter.

3.2.2 Selection of Radar Variables for ML algorithms

In general, the size of birds is larger than insects. Additionally, since Z depends on the size of targets within the radar resolution volume, it could be used to differentiate bird and insect echoes. However, Z also depends on the number of targets. As such, a particular Z value could be due to one large bird or many small insects. Because of this ambiguity in interpretation, Z was excluded from the inputs to the model. Similarly, birds typically move faster than insects. However, biological targets leverage the underlying wind field to aid their flight. As a result, passively flying insects on a windy day could migrate with larger velocities than

actively flying birds in a mild wind field. Furthermore, radial projection and aliasing complicates the interpretation of V_r . Thus, V_r is excluded as an input to the model though it is used to recover measurements from the target’s aspect. Signal-to-noise ratio from biological scatterers are frequently low for the reliable measurement of spectrum width (σ_v). Due to this high noise contamination, σ_v is also excluded. None of the single polarization variables are used in training our models.

Dual polarization variables have been used for the identification of biological echoes [7, 22, 31–33, 52]. In this work, they are used in training our model to distinguish between birds and insects. Another factor that motivated their exclusive selection is the volume coverage pattern (VCP) of the WSR-88D. Biological echoes are predominant at the lowest antenna elevation scan of 0.5° in clear-air. At this elevation, the WSR-88D completes two sweeps, about 30 seconds apart. The first (surveillance) sweep measures the dual polarization variables and Z . The second (Doppler) sweep measures the legacy single polarization variables. Using variables from both sweeps could introduce errors. Assuming a target flies along a radial at a speed of 10 ms^{-1} , it would have migrated 300 m (more than the length of one range gate) between both sweeps. The use of only dual polarization measurements ensures temporal coherence.

3.3 Feature Processing to Prepare Inputs



Figure 3.1: Flowchart of feature processing

In this section, feature processing is performed on the collected bird and insect scans to prepare inputs for training and evaluating the classifiers. The flowchart of this process is shown in Figure 3.1. First, the textures of all dual-pol variables are computed for each scan. Next, blob coloring and minor region removal are used to extract only range gates containing migration echoes in radar scans, followed by VAD analysis to find the direction of migration. The next step is critical. Ideally, we would desire a bird migration scan, for example, to be purely comprised of bird echoes. However, they usually also contain few insect echoes. We proposed a way of improving the purity of our training data, by averaging multiple scans, along range and the target’s aspect. This emphasizes the bird signatures of interest and diminishes unwanted returns, thus making the ground truth more accurate. The same procedure is used for insect migration scans containing few unwanted bird echoes.

3.3.1 Texture

Many image operations are performed on a local section defined by a window. Such windows are usually described with respect to a reference pixel, where the result of any computation is output. In our case, the reference pixel is the middle one. Textures computed based on such an operation, characterizes the spatial variation (e.g standard deviation, variance, root-mean-square deviation) of radar variables within a neighborhood [7, 30]. Here, texture is calculated as the mean absolute deviation within an 8-connected window. The latter comprises of pixels top, left, right and down of the reference pixel, as well as the four pixels in the diagonals. Mathematically, at a given radar gate with range r and azimuth angle ϕ , the texture of a variable x is calculated as

$$\Delta x_{r,\phi} = \frac{1}{N-1} \sum_{i=-1}^1 \sum_{j=-1}^1 |x_{r,\phi} - x_{r+i,\phi+j}|, \quad (3.1)$$

where i is the range gate offset, j is the azimuth offset, and N is the window size. Calculations were performed only at gates that contained values in all neighbors.

3.3.2 Blob Coloring and Minor Region Removal to Extract Migration Echoes

Blob coloring is an image processing method used to identify connected groups of pixels with the same value [93, 94]. It is applied to create a mask indicating migration echoes. Let us define some relevant terms before describing the algorithm. All definitions are with respect to a binary image where a pixel either contains a target (pixel value 1) or background (pixel value 0). A region (or blob) is a group of contiguous pixels with the same value. Two types of windows were applied in this study. The first is a 4-connected neighborhood which refers to the reference pixel and pixels above, left, right and below this reference. The other window is the 8-connected neighborhood already described in the previous section. Another operation performed is dilation, which involves iterating a window over an image and setting the pixels of interest at each step as the OR of all pixels within the window. The result of this operation is that regions of target pixels are expanded.

Data for bird and insect migration were collected for clear-air days, which are characterized by a large area of biological echoes centered around the radar and usually against a clear background. An example for bird migration Z is shown in Figure 3.2(a) within a maximum range of approximately 150 km. Note that Z is only chosen for demonstration, it is not used at any other point of this study.

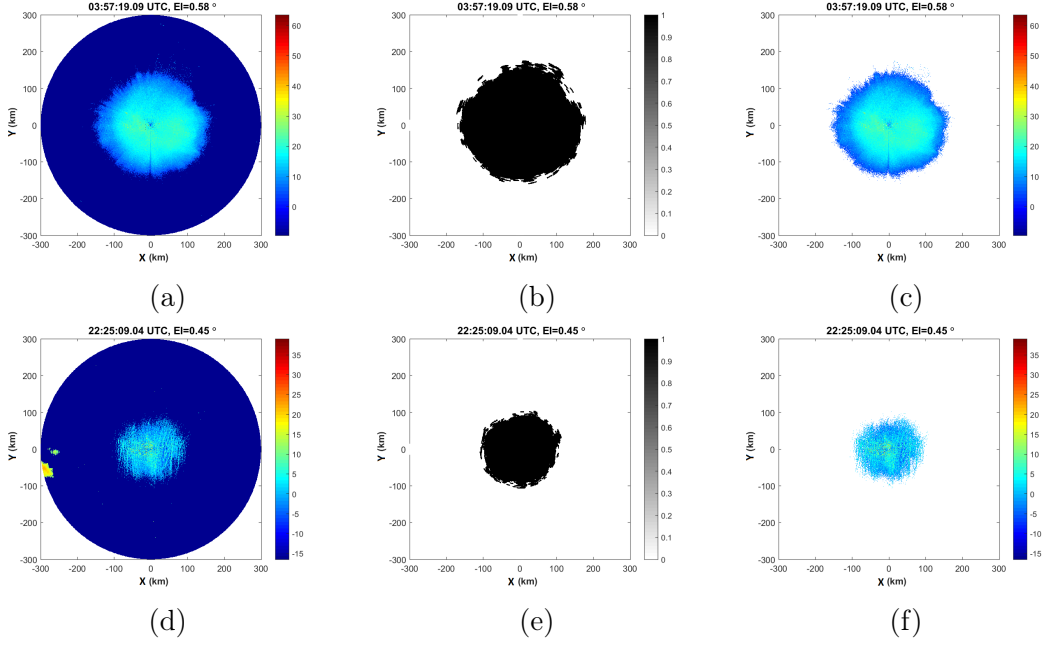


Figure 3.2: Blob coloring with minor region removal to extract large scale migration echoes. (a) Reflectivity (in dBZ) of bird migration echoes. (b) mask of bird migration echoes. (c) gates containing birds extracted using mask. (d) Reflectivity (in dBZ) of insect echoes (e) mask of insect echoes (f) gates containing insects extracted using mask

The radar scan can be considered as an image I where rows correspond to ranges, columns correspond to azimuth angles and range gates correspond to pixels. The blob coloring with minor region removal algorithm is implemented as follows. First, the radar image I is binarized by setting all gates containing echoes to 1 while the remaining gates are set to 0. Usually migration echoes farthest from the radar are disconnected from other echoes close to the radar. The second step involves dilating the binarized image twice to connect the fringe echoes to the main blob. The dilated image J is given as

$$J = (I \oplus B) \oplus B \quad (3.2)$$

where $\oplus$ is the dilation operator, and B is the 8-connected window. In the

third step, a region labelling algorithm [93] is applied to identify the different target regions in J . The size of all identified regions are then computed as the number of pixels they contain. Next, minor region removal [93] is applied, where the largest target region is kept and all other target regions are set to background. The result is a single major region of target echoes. Often times, this major region contains few isolated holes of background pixels. These holes are plugged, by complementing the image, repeating the blob coloring with minor region removal algorithm and re-complementing the image [93]. The resulting mask M is a binary image containing one solid target blob (shown in Figure 3.2(b)) corresponding to the region containing migration echoes. The final image K (Figure 3.2(c)) is extracted by the element wise multiplication of the map M and the original image I . This is expressed as

$$K = I \odot M \quad (3.3)$$

where $\odot$ represents the multiplication operation. This image would contain the migrating birds. All echoes within 10 km are excluded to reduce ground clutter contamination. The remaining echoes are collected as birds samples. The same procedure is repeated for insect cases. Figure 3.2(d) shows Z for insects with a minor precipitation region west of the radar. The generated map excludes this minor region (Figure 3.2(e)). The final extracted echoes would contain insects (Figure 3.2(f)) and those beyond 10 km are collected as insect samples.

3.3.3 Reference with Respect to Target Azimuth

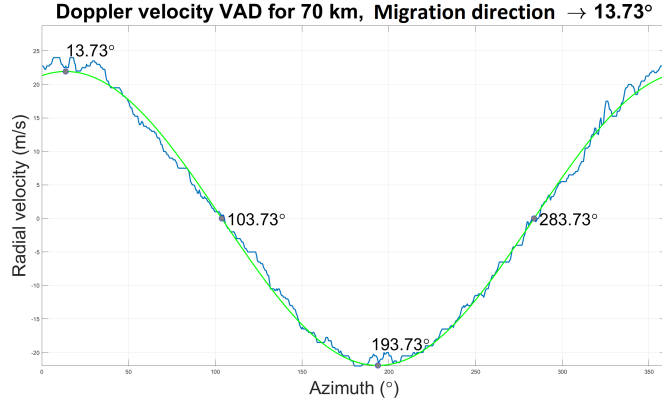
We defined aspect as the orientation of the target's body with respect to the radar beam. Because of the non-spherical shape of biological targets, their radar returns

would depend on the aspect observed. As such, methods for identifying biological echoes will have to account for this dependence. Cases of wide spread alignment can leverage traditional VAD [6, 53, 91] or azimuthal patterns in the correlation coefficient [8] to recover aspect information. We used VAD to rotate the dual-pol variables and textures, so they become a function of their aspect azimuth (ϕ_{aspect}). First, a sinusoid is fit to V_r at every range using Equation (2.43), and the common azimuthal destination of echoes, $\phi_{migration}$, calculated using Equation (2.44). This direction captures measurements from the tail aspect. Scattering from other azimuthal aspects can be deduced by the lag from $\phi_{migration}$ as

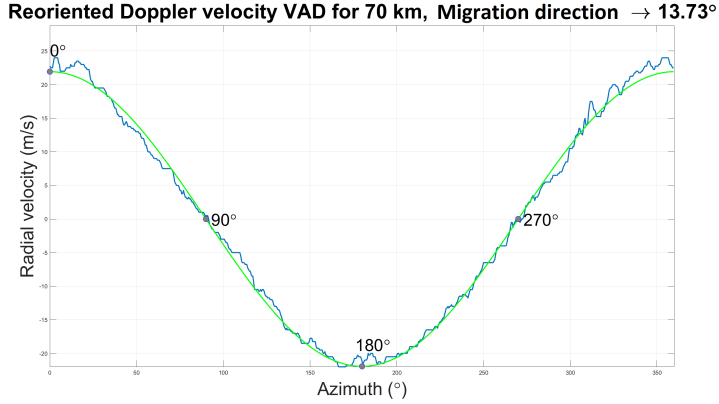
$$\phi_{aspect} = \phi - \phi_{migration} \quad (3.4)$$

such that a ϕ_{aspect} of 0° represents the tail region of biota, 90° represents the left wing region and 180° represents the head region.

An example for this procedure is shown in Figure 3.3. Figure 3.3(a) shows the VAD at range 70 km for one of the scans in the training set. The blue line is the filtered velocity obtained by applying a 10th order one dimensional median filter on V_r . The green line is the fitted $\hat{V}_r$. Migration was found to be toward 13.73° . The radial velocity w.r.t to the target $V_r(\phi_{aspect})$, shown in Figure 3.3(b), is obtained by shifting $V_r(\phi)$ to the left by 13.73° . Migration would be toward ϕ_{aspect} of 0° . This process is applied at every range ring to find the migration direction and rotate all dual polarization measurements and their textures. All measurements are now relative to the aspects of the targets.



(a)



(b)

Figure 3.3: VAD analysis to reorient velocity to be relative to the target aspect for 70 km range gates. The blue line shows filtered V_r while the green line is the sine fit. (a) Initial VAD finds targets to be migrating toward 13.73° (b) Reoriented fit shifts measurements so migration is toward 0°

3.3.4 Averaging Bird and Insect Cases

To reduce the contamination of bird migration cases by insect echoes and vice versa, multiple scans are averaged. Following blob coloring and rotation of the collected scans and their textures, they are grouped into 3 batches. Let us call them batches A, B and C. Each batch contains 15 randomly selected scans per class (a total of 30 scans per batch). We will focus on A though all steps discussed equally apply to B. Each scan will have different azimuths, so we created a new

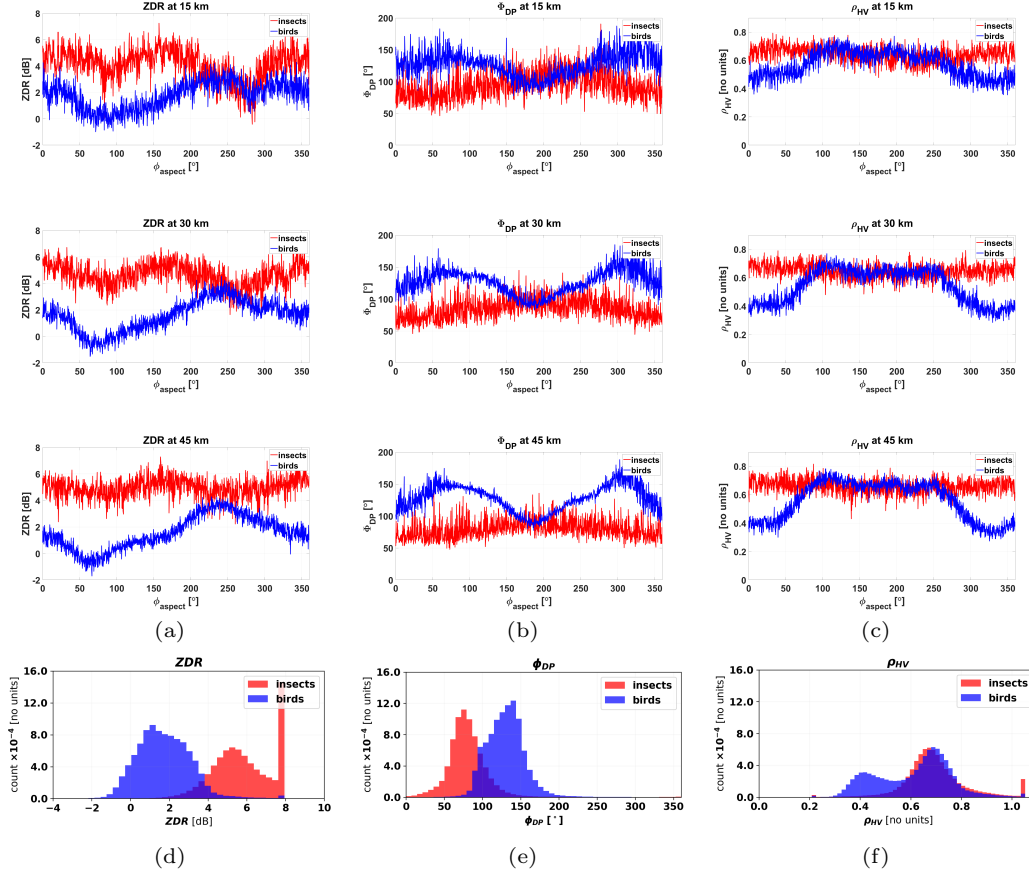


Figure 3.4: Averaged dual polarization variables as a function of ϕ_{aspect} for the training set. (a) Z_{DR} vs ϕ_{aspect} (b) ϕ_{DP} vs ϕ_{aspect} (c) ρ_{HV} vs ϕ_{aspect} . Birds are in blue and insects red. Rows represent ranges of 15, 30 and 45 km from the radar. Training set distributions of averaged (d) Z_{DR} (e) ϕ_{DP} , and (f) ρ_{HV} for birds and insects.

range and aspect azimuth grid both starting at 0 and with resolutions of 250 m and 0.5° respectively. All scans were interpolated to this common grid. The new 15 scans for birds (insects) are then averaged. In the last step, all range gates in the resulting averaged scans from A and B are combined to form the training set, containing 1,711,624 samples: 57% bird and 43% insect cases. Batch C is used as the test set. It is not averaged so that it represents the kind of measurements we expect from the WSR-88Ds. The test set contained 9,402,821 range gates with 60% bird and 40% insect cases.

Some visualizations of the averaged training cases are shown in Figure 3.4. The blue curve is for birds and the red for insects. Each plot in Figure 3.4(a)-3.4(c) is for a dual polarization variable against the target aspect at specific ranges. From top to bottom, rows correspond to measurements 15, 30, and 45 km from the radar. From left to right, columns correspond to Z_{DR} , φ_{DP} and ρ_{HV} respectively. The averaging procedure shows that dual-pol variables have a strong dependence on ϕ_{aspect} and exposes clear delineations between birds and insects. The distribution of the whole training set is also shown in Figure 3.4(d)-3.4(f) for Z_{DR} , φ_{DP} and ρ_{HV} respectively. These results are consistent with previous literature. Analysis in [51] found that echoes attributed to birds (Purple Martins) had Z_{DR} between -4 and 6 dB. In our case, the averaged Z_{DR} (shown in Figure 3.4(a) and Figure 3.4(d)) for birds is generally low, between -2 and 4 dB. The highest value is around 230° (between the head and right wing) and the lowest around 75° (between the tail and left wing). Insects were found to have high Z_{DR} (up to 10 dB) in [52]. Our averaged insect Z_{DR} is also generally higher with most gates between 3 and 8 dB. Interestingly, the values dip below the bird Z_{DR} values between ϕ_{aspect} of 230° and 300° . φ_{DP} (Figure 3.4(b)) for birds are generally higher than insects, with peaks around 50° and 300° . This is consistent with the observed symmetry of φ_{DP} about the direction of migration [52]. Insects have lower φ_{DP} values (Figure 3.4(e)). ρ_{HV} for bird migration have been observed to have low values corresponding to tail-on viewing angles and high values for head-on angles [8, 51]. This can be seen in the sinusoid-like pattern in Figure 3.4(c), with high values (around 0.7) between 60° and 250° and low values (around 0.4) otherwise. The bimodal ρ_{HV} distribution for birds (Figure 3.4(f)) is also due to this sinusoidal pattern. Insects generally have slightly higher ρ_{HV} than birds with a mean value around 0.7.

After the averaging procedure, both the training and test data sets are nor-

malized. The mean and standard deviation for each variable was computed from the 60 scans in batches A and B. They are used to normalize each variable by mean centering and scaling by their standard deviation. This equalizes the scale of all variables and improves training convergence [95]. The same procedure was applied to normalize their textures.

3.3.5 Input Features for Classifiers

The normalized dual polarization variables and their textures, are used as input features for the classifiers. Additionally, inspection of the data revealed that variables varied gradually with range and ϕ_{az} . Thus two new discrete features were created to capture this variation. The first is range interval, which refers to 10 km bins. The second is sector, which refers to 20° sectors computed from ϕ_{az} . So for example, a range gate at 25 km and with ϕ_{az} of 85° would belong to the 3rd range interval and the 5th sector. Both features are one hot encoded to convert them to indicator vectors, as shown below.

$$range\ interval_i = \begin{cases} 1, & \text{if } i = \lfloor range/10 \rfloor. \\ 0, & \text{otherwise.} \end{cases} \quad (3.5)$$

$$sector_j = \begin{cases} 1, & \text{if } j = \lfloor \phi_{az}/20 \rfloor \\ 0, & \text{otherwise.} \end{cases} \quad (3.6)$$

All echoes collected in this study was from 10 to 230 km, so *range interval* would contain 22 elements. The first element is ignored since we did not consider ranges below 10 km. *sector* would contain 18 elements.

3.4 Machine Learning Methods

More complex neural networks have been successfully applied to detect [46] and track roosting birds [47], however there are some drawbacks. First, they were not designed to make classification on a single radar range gate, and rather use a rendered image of a full radar scan as input. Second, they are trained to specifically detect birds engaging in roosting behaviour. As such, these networks cannot be generalized to other patterns of bird activity or types of biological echoes. A more robust algorithm for differentiating biological echoes is needed.

In this study, we investigate two supervised ML algorithms for distinguishing birds and insects, that are (i) capable of using inputs from a single range gate and as such, adaptive to different orientations of biota (ii) able to provide a probability that a range gate contains birds (or insects). (iii) easy to retrain as more data is collected. These algorithms are the ridge classifier and decision tree described in Section 2.3. Furthermore, they form the basis of more sophisticated neural networks and random forests. Therefore, with understanding of their performance, complexity can be added as required. This way we keep the models as useful and as simple as possible.

3.5 Metrics

We used four metrics to assess the quality of our classifiers. They are accuracy (ACC), True Positive Rate (TPR), True Negative Rate (TNR) and Area Under Curve (AUC). Table 3.1 below shows the confusion matrix for our classification problem. Birds are used as the positive class, so true positives (TP) are birds that are correctly classified as birds, false positives (FP) are birds classified as insects, false negatives (FN) are insects classified as birds and true negatives (TN) are

insects correctly classified as insects. Each instance corresponds to a range gate.

Table 3.1: Confusion matrix for bird detection

Classifier output	True labels	
	Birds	Insects
Birds	TP	FP
Insects	FN	TN

Accuracy is perhaps the most popular metric for evaluating classifiers. It is the proportion of the total correct predictions. However, ACC can be biased by the distribution of classes. Assuming a data set contains 90% positive and 10% negative examples, a naive classifier that simply predicts all positives will have an ACC of 90%, though it would not be useful for making any future predictions. Other metrics are needed to better evaluate classifier performance. TPR is the proportion of correct predictions for the positive class and TNR, the proportion of correct predictions for the negative class. Our naive classifier would have a TPR of 100%, but a TNR of 0% (bad). Ideally all three metrics should be 100%. They are calculated as shown in the Equations (3.7) - (3.9) below. Binary classifiers usually predict a probability (or score) for the positive class and then a threshold is applied to obtain the final class. The Receiver Operating Characteristics (ROC) curve plots TPR against the False Positive Rate (FPR) for varying probability thresholds [96]. FPR is $1 - TNR$. At a threshold of 0, TPR is 100% (good) and FPR is 100% (bad). At a threshold of 1, both are 0%. The goal of the ROC curve is to find an intermediate threshold that maximizes TPR and minimizes TNR. The Area Under Curve (AUC) metric summarizes the area under the ROC curve [96]. Good classifiers should have an AUC close to 100%.

$$ACC = \frac{TP + TN}{TP + FN + FP + FN}, \quad (3.7)$$

$$TPR = \frac{TP}{TP + FN}, \quad (3.8)$$

$$TNR = \frac{TN}{FP + TN}, \quad (3.9)$$

3.6 Model Training and Validation

Classifiers are sensitive to the class distribution of the training set. Thus, we applied class weights [88] to samples to balance their effect on the loss function. They are computed as

$$w(c_j) = \frac{\sum_j^m N(c_j)}{m \times N(c_j)} \quad (3.10)$$

where $N(c_j)$ is the number of examples in class c_j and m is the total number of classes. For each machine learning method, 8 models are trained using different combinations of inputs. First, a base model is trained on only dual polarization variables and then different combinations of the remaining features are added to investigate their effect on performance. It should also be noted that not all inputs features can always be obtained from the radar scan. For example, sector is calculated using a sinusoid fit to the velocity of migration echoes. These echoes are mostly composed of a single species moving in a particular direction. In cases with diverse species moving in different directions, the sinusoid fit will not be possible and sector cannot be recovered. Velocity aliasing could also prevent the recovery of sector. Similarly, the calculation of texture for a given gate depends on the presence of targets in all neighboring gates. In an alternative scenario, this calculation will not be possible. In such cases however, the base model can always

be used.

K fold cross validation [97] was used to tune the model hyperparameters. In this method, the data set is divided into K folds, model training is performed on K-2 folds, validation on 1 fold and testing on 1 fold. Since we already held out a test set (batch C), training was performed on K-1 folds and validation on 1 fold. The whole process is repeated K times where each fold is used as training and validation once. A total of 5 folds were used. After cross validation, the hyperparameters that have the best performance are chosen for each model. Final training is performed using the selected hyperparameters and the full training set. An ROC curve is then generated and a critical threshold found, such that it maximizes TPR and TNR. This threshold would be used to convert predictions into classes. The training process is stochastic so each run produces slightly different results. To have a robust assessment of performance, 30 independent training runs are repeated for each model. All the trained models are then evaluated on the test data. Confidence intervals for each metric is calculated using the bootstrapping percentile method where each metric is computed from an iteratively chosen random sample of the test data [46, 98]. We computed each metric for 100 iterations based on 1000 randomly chosen samples. The 100 metrics for 30 repeated runs form a total of 3000 estimates. The confidence interval is found as the 2.5% and the 97.5% points of the distribution [46, 98].

3.7 Performance

The 95% confidence interval for the model metrics are shown in Table 3.2. All the ridge classifiers are predictive with $ACC > 0.81$, $TPR > 0.82$, $TNR > 0.77$ and $AUC > 0.86$. Sector is expected to greatly improve results however its addition

Table 3.2: 95% confidence interval for the ridge classifiers' and decision trees' metrics on migration data. The possible inputs are Dual-Pol(DP), their textures(ΔDP), sector($sect$) and range interval(RI)

	ACC	TPR	TNR	AUC
Ridge classifiers				
DP	0.814 - 0.858	0.832 - 0.892	0.778 - 0.844	0.868 - 0.911
$DP + \Delta DP$	0.849 - 0.891	0.874 - 0.924	0.808 - 0.874	0.909 - 0.943
$DP + sect$	0.812 - 0.858	0.822 - 0.886	0.784 - 0.85	0.869 - 0.911
$DP + RI$	0.815 - 0.86	0.836 - 0.894	0.774 - 0.844	0.869 - 0.912
$DP + \Delta DP + sect$	0.849 - 0.891	0.87 - 0.924	0.81 - 0.874	0.91 - 0.943
$DP + sect + RI$	0.813 - 0.858	0.826 - 0.884	0.78 - 0.848	0.869 - 0.912
$DP + \Delta DP + RI$	0.849 - 0.891	0.872 - 0.926	0.81 - 0.872	0.91 - 0.944
$DP + \Delta DP + sect + RI$	0.85 - 0.891	0.87 - 0.922	0.812 - 0.876	0.91 - 0.944
Decision trees				
DP	0.8 - 0.856	0.786 - 0.892	0.762 - 0.852	0.872 - 0.92
$DP + \Delta DP$	0.778 - 0.852	0.786 - 0.902	0.718 - 0.846	0.866 - 0.925
$DP + sect$	0.751 - 0.815	0.658 - 0.8	0.79 - 0.88	0.831 - 0.884
$DP + RI$	0.75 - 0.82	0.668 - 0.794	0.79 - 0.882	0.792 - 0.868
$DP + \Delta DP + sect$	0.713 - 0.804	0.628 - 0.79	0.768 - 0.866	0.793 - 0.873
$DP + sect + RI$	0.684 - 0.762	0.518 - 0.686	0.814 - 0.894	0.711 - 0.796
$DP + \Delta DP + RI$	0.714 - 0.813	0.7 - 0.828	0.676 - 0.844	0.751 - 0.871
$DP + \Delta DP + sect + RI$	0.669 - 0.796	0.572 - 0.732	0.71 - 0.87	0.702 - 0.832

to ridge classifiers cause marginal changes to performance. It slightly improves TNR, slightly reduces TPR and does not seem to have a noticeable effect on AUC. To understand the role of sector, let us consider classification using Z_{DR} from a range gate 30 km from the radar (Figure 3.4). If this gate has $Z_{DR} > 4$ dB, it should be classified as insects. If $Z_{DR} < 0.5$ dB the echoes should be classified as birds. However, Z_{DR} values within 0.5 – 4 dB can belong to any class. In this scenario sector can further differentiate between the classes. If the undetermined gate falls in the 80 – 100° aspect (or sector 5), then it should be classified as insects. But if it falls in the 220 – 240° aspect (sector 12), Z_{DR} cannot accurately detect the correct class because the values of both classes overlap. Observing the 15 km range gates in Figure 3.4, bird and insect echoes can easily be separated by at

least one variable across all aspects except around $200 - 250^\circ$ where all variables overlap. Similar to the previous example, sector would not improve classification in this region. This could be the reason for the marginal results due to sector. However, it should be noted that the clear delineation between both classes at most aspects, so that classification can be effectively performed without sector, is due to the previous averaging along aspect angles. Addition of range interval marginally changes performance, improving ACC, TPR and AUC, and reducing TNR. The addition of texture generally improves the model metrics.

The decision tree models are also predictive for TPR, TNR and AUC but perform significantly worse on TNR with some models having values around 0.5. This seems to coincide with models using range interval and/or sector. A possible cause could be that its binary decision making tends to prefer classifying whole range intervals (or sectors) as one class in contrast to ridge classifiers that only learn a probability adjustment. Using smaller range intervals and sectors might mitigate this problem. Similar to the ridge classifiers, incorporation of texture generally improves the model metrics. However this might not generalize to non-migration cases. Recall that labels were provided based on the dominant migrating taxa. Textures have the effect of averaging measurements derived over a 3×3 neighborhood, so would emphasize the dominant class leading to better metrics for migration cases. However, for non-migration cases with a heterogeneous mix of scatterers, texture could lead to mis-classifications.

An independent two sample t-test is conducted to find the better performing machine learning method. Each metric distribution for equivalent versions of ridge classifiers and decision trees were tested with a significance level of 0.05. The null hypothesis is that both metric distributions have the same mean. All tests were significant ($p < 0.05$). The final results of our t-test is shown in Table

3.3. The cells, labelled R (for ridge classifiers) and D (for decision trees) indicate which model was found to be better for a metric. Our criteria for a method to be better was that it should have higher means on ≥ 3 metrics. The ridge classifiers satisfy this criteria for all input combinations and was chosen as the better method. The best ridge classifier uses dual polarization variables, texture, sector and range interval as inputs with $AUC > 0.91$. It is possible though, that the improvement caused by the added features could be the classifiers over-fitting to migration cases. Thus, additional studies of ridge classifiers on a diverse variety of cases (presented in the next section) are required to understand the effect of these features on performance. Similar discussion on decision trees can be found in the supplementary materials.

Table 3.3: Model comparison results. Each cell is labelled R (ridge) or D (decision tree) depending on which model performs better

	ACC	TPR	TNR	AUC
DP	R	R	R	D
$DP + \Delta DP$	R	R	R	R
$DP + sect$	R	R	D	R
$DP + RI$	R	R	D	R
$DP + \Delta DP + sect$	R	R	R	R
$DP + sect + RI$	R	R	D	R
$DP + \Delta DP + RI$	R	R	R	R
$DP + \Delta DP + sect + RI$	R	R	R	R

3.8 Case Studies and Discussions

In this section, the performance of the ridge classifiers are further tested on four cases. The aims of these analyses include: verification of consistency between the classifier's detections and the available ground truth, investigation of the effect of different input features, and exploration of performance on new scenarios which

differ from the training data. For operational use, we recommend that the ridge classifiers be used in conjunction with the HCA. So the HCA is applied first and if the HCA classifies a range gate as biological, the ridge classifiers are subsequently applied to classify the gate as containing either bird or insect echoes. This configuration is applied to cases in this section, where echoes had not already been identified as biological.

The first case study is for bird and insect migration scans from the test set. This case was chosen to understand the output of the models when all variables are recovered. The second case explores a new bird migration scan where velocities were heavily aliased, so sector could not be recovered. It is also important to investigate the ability of our models to generalize to new situations. One of such scenarios involves the mass exodus of birds from their nighttime roosting locations usually observed as expanding rings of echoes. In the third case, a scan from the KHTX radar (located at Huntsville, AL) containing two groups of roosting birds, insects and precipitation echoes was analyzed. This case is especially important because our training data was collected from a different radar (KTLX) and contained no bird roosts. As a result, the ability to detect these roosts and insect echoes will be strong evidence that it can be extended to new patterns of biological activity and to other radars in NEXRAD. It also demonstrates how precipitation echoes would be handled. The final case also investigates the performance on roosting birds from the KTLX radar. A summary of the aforementioned cases is presented in Table 3.4.

Table 3.4: Summary of case studies

Description	WSR-88D	Location	Date	Time (UTC)
Bird migration	KTLX	OKC, OK	2 May 2015	4:13
Insect swarm	KTLX	OKC, OK	11 July 2019	17:08
Bird migration	KTLX	OKC, OK	3 Mar. 2018	4:41
Bird roosts & insects	KHTX	Huntsville, AL	11 Aug. 2015	11:15
Bird roosts & insects	KTLX	OKC, OK	8 Aug 2017	11:47

3.8.1 Bird/Insect Migration Test Cases

The first case was collected from KTLX at 04:13 UTC on 2 May 2015 composed of migrating birds. It is one of the PPIs in the test set. Since blob coloring with minor region removal had been applied to extract migration echoes, the HCA is not used in this case. Sector could be recovered because of the homogeneity and alignment of scatterers. The classifier outputs a probability of each range gate containing bird echoes. The critical threshold is applied to binarize these probabilities to 0's (insects) and 1's (birds), obtaining the output class. The ridge classifier outputs are shown in Figure 3.5. Birds are colored blue while insects are colored red. Accross the 8 ridge classifiers, birds are detected in 94.4% – 95.4% of classified gates. This is consistent with our hypothesis of migrating birds. Overall, the performance of the base model barely changes as the remaining features (texture, sector and range interval) are added, with at most a 1% difference in proportion of birds detected.

The second case is also a test scan collected from KTLX at 17:08 UTC on 11 July 2019, and composed of migrating insects. Sector was recovered and the HCA was not applied because migration echoes had already been extracted by blob coloring. The ridge classifier outputs are shown in Figure 3.6. Accross the classifiers, insects are detected in 92.3% – 92.7% of classified gates, consistent with our hypothesis of echoes from insects. Like the bird migration case, there is

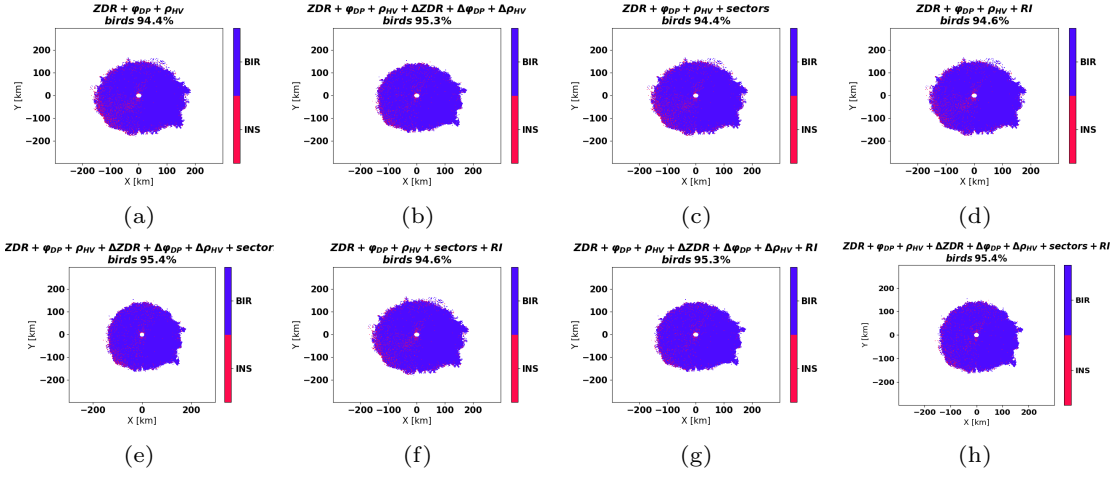


Figure 3.5: Ridge classification results for bird migration observed with KTLX radar at 04:13 UTC on 2 May 2015. BIR represent birds and INS represent insects.

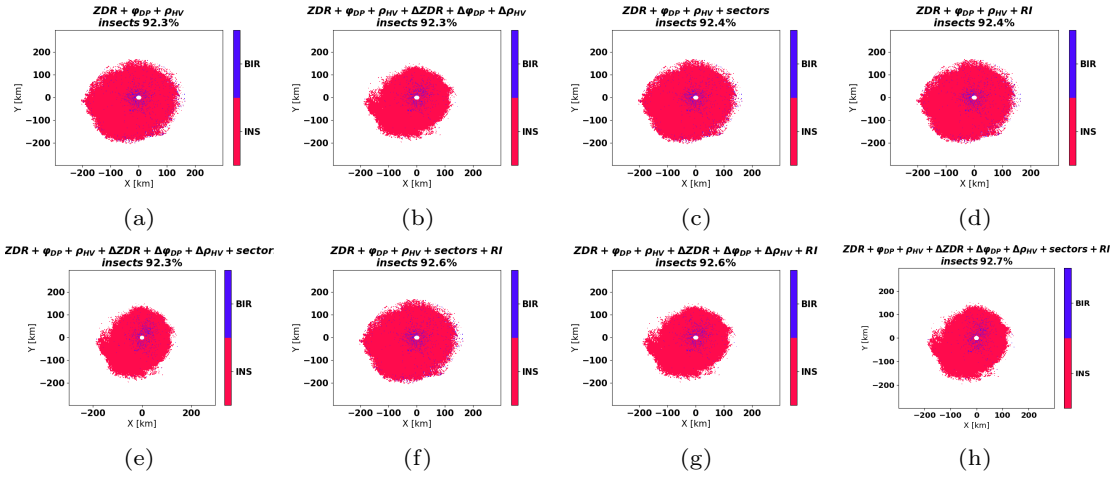


Figure 3.6: Ridge classification results for insect migration observed with KTLX radar at 17:08 UTC on 11 July 2019.

a bare difference of 0.4% in the proportion of insects detected between the base model and those with added features.

3.8.2 Bird Migration with Aliased Velocities

The third case was collected from KTLX at 04:41 UTC on 03 March 2018, composed of nocturnal bird migration during spring. The characteristic symmetry

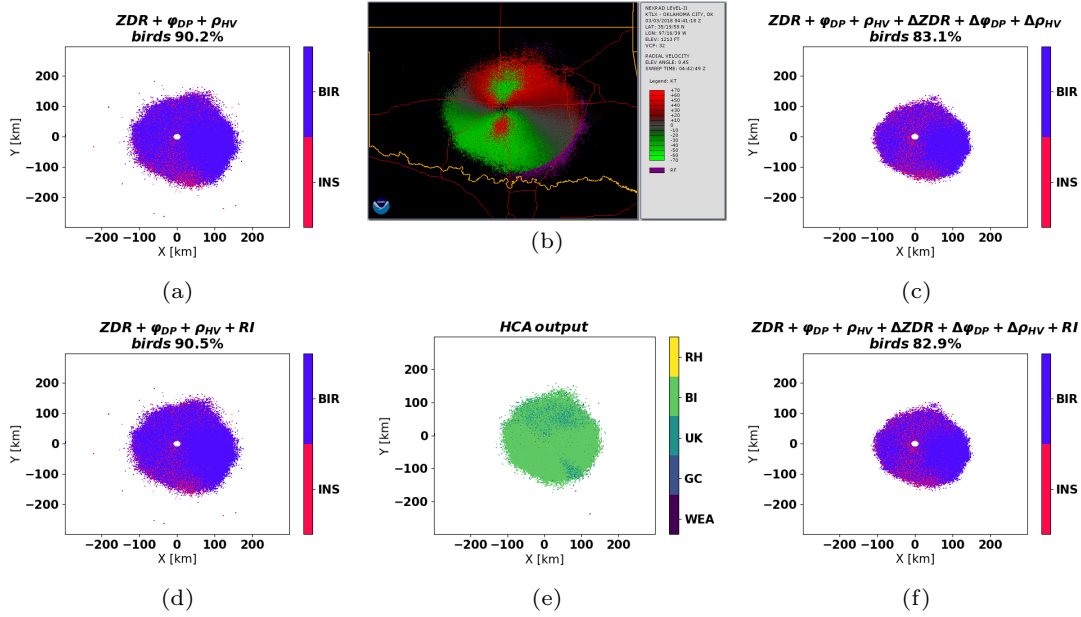


Figure 3.7: Bird migration with velocity aliasing observed with KTLX at 04:41 UTC on 03 March 2018. (a) Ridge classifier (RC) using dual polarization variables. (b) $0.5^\circ V_r$ shows heavy aliasing north and south of radar. (c) RC using dual polarization variables and their textures. (d) RC using dual polarization variables and range interval. (e) Hydrometeor Classification Algorithm (HCA). (f) RC using dual polarization variables, their textures and range interval.

pattern about the migration direction was observed in φ_{DP} (not shown here) with values generally larger than 110° . The radial velocities (shown in Figure 3.7(b)) were also highly aligned, with large magnitudes causing aliasing. Furthermore, the bird/insect fuzzy logic classifier in [22] detected birds as the dominant taxa for this case. The HCA output is shown in Figure 3.7(e). All weather classes are condensed into one resulting in five possible classes which include weather (WEA), anomalous propagation/ground clutter (GC), unknown (UK), biological (BI) and range folded (RH). Most of the echoes are classified as biological with a few gates classified as unknown. We could not recover sector because of heavy velocity aliasing. Our ridge classifiers detects birds as the dominant taxa found in 82.9% – 90.5% of classified gates. Addition of texture to the base classifier

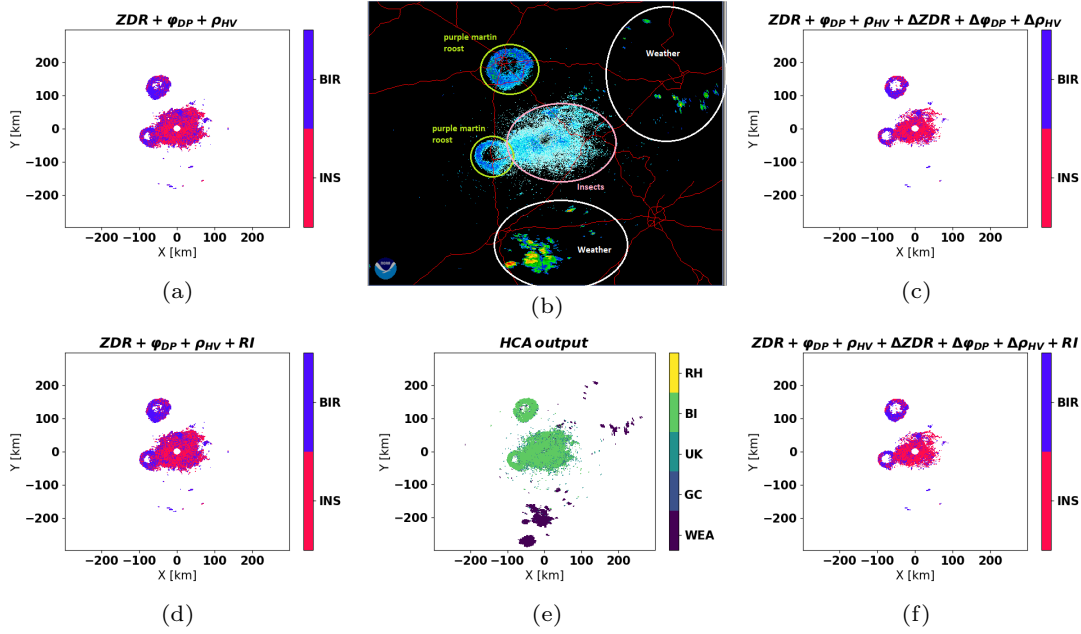


Figure 3.8: Bird roosts observed with KHTX at 11:15 UTC on 11 August 2015. (a) RC using dual polarization variables. (b) 0.5° Z scan showing bird roosts (west and north-west), insects (around the radar) and weather echoes (north-east and south). (c) RC using dual polarization variables and their textures. (d) RC using dual polarization variables and range interval. (e) HCA. (f) RC using dual polarization variables, their textures and range interval.

decreases the proportion of birds by about 7%. The addition of range interval barely changes performance, with an increase of 0.29% in the proportion of birds detected.

3.8.3 KHTX Bird Roosts

The third case was collected from KHTX radar (located at Huntsville, AL) at 11:15 UTC on 11 August 2015. This case demonstrates the performance of the ridge classifiers on bird roosts from a different WSR-88D radar. The Z scan is shown in Figure 3.8(b) with three groups of echoes, labelled based on analysis conducted in [7]. The first group contains two colonies of purple martins engaging in their morning roosts, verified by ground observers from the Purple Martin Conservation

Society [7]. The roosts are located north-west and west of the radar. The second group contains insects around the radar, identified by their comparatively low mean airspeed of 1.8 ms^{-1} and concentration at low heights [7]. Air speed was calculated by vector subtraction of wind speeds (obtained from balloon sounding) from ground speeds obtained from radial velocity [8]. The third group contains weather echoes identified by their near 0 dB Z_{DR} , near 1 ρ_{HV} and φ_{DP} near the calibration offset of 60° [7].

The classifiers are applied as we recommend for operational use. First, the HCA is used to classify range gates into the five condensed classes. The results (shown in Figure 3.8(e)) matches the labels provided in [7] with the bird roosts and insects identified as biological echoes and the precipitation echoes also correctly identified (seen farther south and north-east of the radar in Figure 3.8(e)). Next, the ridge classifiers are applied on only the biological class. Sector could not be recovered here because of the presence of diverse scatterers with different velocities. The base classifier (Figure 3.8(a)) and the classifier with range interval (Figure 3.8(d)) identify the roosts as bird dominated and the insect region as insect dominated. These results are consistent with the bird and insect groups observed in [7]. The models with texture (Figure 3.8(c) and 3.8(f)) similarly identifies the insect region but mis-classify large parts of the roosting birds as insects, most noticeable where the western roost intersect with insect echoes.

Overall, the base classifier's detections matches the labels provided demonstrating the potential of this classifier to be applied to new cases of biological activity and different radars within NEXRAD. The addition of range interval does not have a noticeably effect on performance. The addition of texture seems to degrade performance on fine and hollow features like bird roosts.

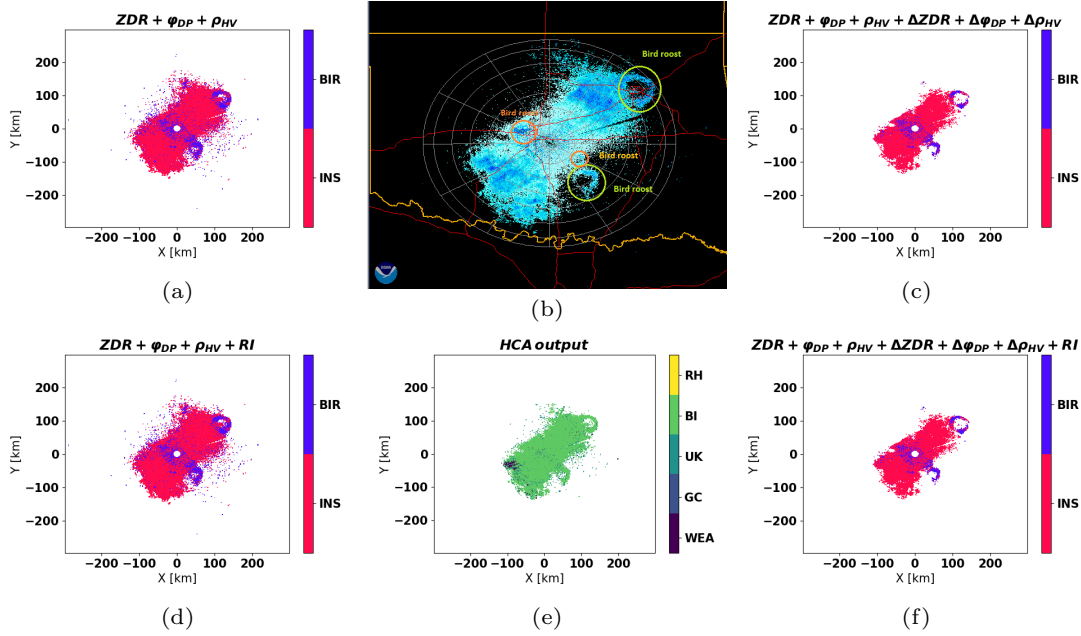


Figure 3.9: Same as Figure 3.8 but for roosts observed with KTLX at 11:47 UTC on 8 August, 2017. Four bird roosts were identified north-west, north-east and south-east of the radar

3.8.4 KTLX Bird Roosts

The final case was collected from KTLX at 11:47 UTC on 8 August, 2017. We identified four bird roosts against an insect background (Figure 3.9(b)) by tracking their expansion across multiple scans. Furthermore, the bird/insect fuzzy logic classifier in [22] detected the western and the two south-eastern roosts as bird dominated. It was not tested on the north-eastern roost because it has a range limit of 100 km. A major issue with this fuzzy logic classifier is that it has a tendency to classify gates containing dense aggregation of insects as birds because it uses Z as one of the inputs. A correction was made using thresholds on Z_{DR} and ρ_{HV} . However, this also led to other complications due to the variation of biological measurements with their viewing angle.

Since our ridge classifiers are based on dual-pol variables, they are robust to

the concentration of scatterers in a range gate. The HCA (Figure 3.9(e)) detects most gates as comprised of biological echoes and the classifiers are applied to these echoes. Sector could not be recovered because of the presence of diverse scatterers. The base classifier (Figure 3.9(a)) and the classifier with range interval (Figure 3.9(d)) detects all the roosts as bird dominated. Addition of texture (Figures 3.9(c) and 3.9(f)) leads to more parts of the roosts being classified as insects. Similar to the previous case, the base classifier’s output matches the provided labels. Their performance on bird roosts are also consistent with the results obtained from the bird/insect fuzzy logic classifier in [22]. Overall, the addition of range interval does not have a noticeably effect on performance while the addition of texture degrades the detection of roosts.

3.9 Recommendations for Operational Implementation

The base classifier which uses only dual-pol variables, showed robust performance across the different case studies involving bird/insect migration, as well as more complicated mixtures of bird roosts with insects. Moreover, all its inputs are readily available from a single range gate, thus further expensive feature processing would not be required. As such, we recommend that the base classifier can be used for differentiating HCA’s biological class as birds or insects on NEXRAD. Hereafter, this classifier is also called the Bird-Insect Ridge Classifier (BIRC). Further applications for improving VAD wind estimation will be provided in Chapter 4. In the meantime, the rest of this section details justifications for selecting the base classifier.

Our performance analysis on the test data revealed that the classifier with the full set of features improved each metric from the base classifier by about

4%. However, because the test data was collected from cases of migration, it is possible that the former is overfitting to one particular scenario and might not extend to new scenarios effectively. Additional case studies in the previous section, provides strong evidence that this might be the case. The base classifier was found to have consistently strong performance across all scenarios. The addition of range interval and sector seems to have a benign effect, probably because they have already been applied to improve the purity of the training data during the averaging step. Texture on the other hand, seems particularly suited to broad scale migration but has worse performance on finer and hollow features like bird roosts. Additionally, the base classifier was demonstrated to be effective on new orientations of biological echoes and a different WSR-88D radar in the KTLX and KHTX bird roost cases.

Based on previous literature and the results illustrated in Figure 3.4, our intuition is that birds generally have a lower Z_{DR} , higher φ_{DP} and lower ρ_{HV} compared to insects. The weights (refer to Equation (2.37)) learned by the base classifier confirms this. They are shown in Figure 3.10. For a given input, a positive weight implies that the probability of being a bird increases as the input increases. Negative weights imply that the probability of being an insect increases as that input increases. The weight magnitude corresponds to its importance on the classification outcome. Z_{DR} has a weight of -2 meaning that the higher its value, the higher the probability of being an insect. Higher φ_{DP} values would increase the probability of being a bird, since it has a weight of +1. ρ_{HV} is more subtle with a weight of -0.1, and so higher values would slightly nudge the classifier toward insects. Also notice how the magnitude of the weights are proportional to the degree of separation between both classes in Figure 3.4. Z_{DR} seems to be the most important with about twice the effect of φ_{DP} and 20 times the effect of ρ_{HV} .

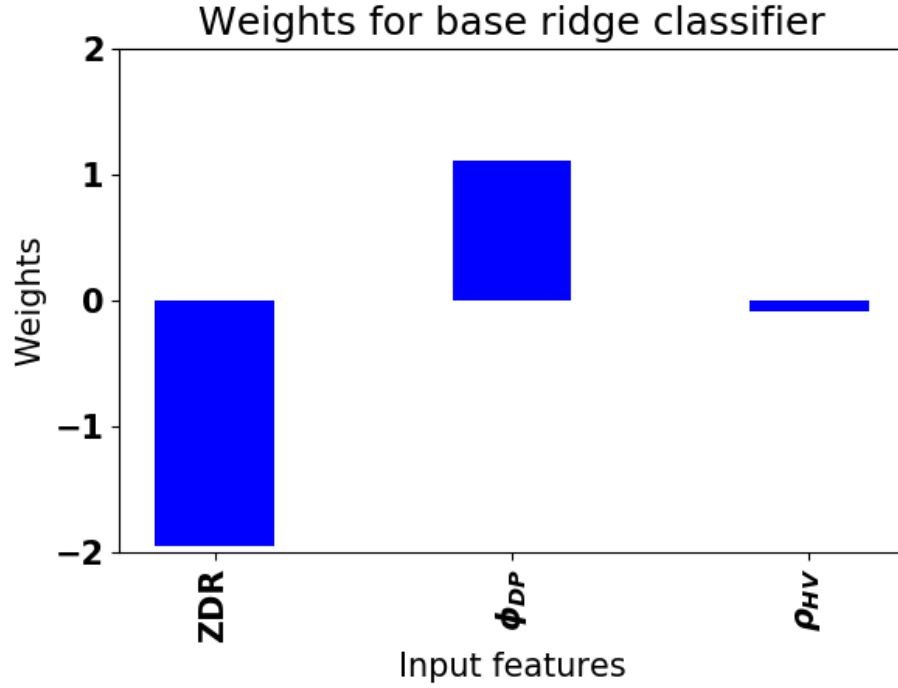


Figure 3.10: Weights of base ridge classifier. Positive (negative) weights for a variable imply that the probability of being a bird is directly (inversely) proportional to the variable.

3.10 Conclusions

The ability of NEXRAD to detect birds and insects offers much promise for meteorological and biological studies. In this work, a classifier is developed for distinguishing bird and insect radar echoes based on dual polarization variables. Unique challenges were presented by the complex scattering off their non-spherical bodies. These were addressed by leveraging cases of large scale highly aligned migration. Measurements relative to the body aspect of biota were extracted as the lag off the mean flight direction, obtained using VAD. Another issue is the reliability of ground truth for training data sets because of the frequent collocation of birds and insects in the radar sampling volume, leading to the contamination of scans

otherwise dominated by a single species. This was addressed by averaging 15 alignment calibrated bird (insect) migration scans to reduce the effect of the less dominant class.

Overall data preparation for the training and test set involves the following steps. First, 45 migration scans were collected for each class. Blob coloring with minor region removal is then applied to extract migration echoes and textures computed. Extracted migration echoes are rotated to obtain measurements relative to the target’s aspect. Next, rotated scans are grouped into 3 batches, each containing 15 scans per class. Within each class in the first two batches, all scans are averaged. Gates from the 4 resulting averaged scans are used as the training samples. The last batch is used as the test set. Furthermore, all gates are grouped into 10 km range intervals and 20° target azimuth sectors. The final inputs considered are dual pol variables, their textures, range intervals and sectors.

Two machine learning methods were explored: ridge classifier and decision tree. Eight models were trained for each method, starting with a base model using only dual polarization variables and then adding other features (range interval, texture and sector). Models were evaluated based on ACC, TPR, TNR and AUC, all computed from the test data set. An independent student t-test comparison showed that the ridge classifiers performed better than decision trees in ≥ 3 metrics. Based on this, the ridge classifiers were selected for classifying bird and insect radar echoes. All the ridge classifiers are predictive with $ACC > 0.81$, $TPR > 0.82$, $TNR > 0.77$ and $AUC > 0.86$. The addition of other features generally show an improvement, with about a 4% increase on these metrics, however later evidence suggests that this is probably due to over-fitting to cases of large scale migration.

Further investigations were conducted to assess the classifiers on more diverse

orientation of biota. The first case contained bird and insect migration scans from the test set. They are chosen because all inputs are available. The ridge classifiers detected birds in 94.4% – 95.4% of range gates for the bird scan and insects in 92.3% – 92.7% of gates for the insect scan, consistent with the expected taxa. Overall, the addition of the remaining features to the base model has a limited impact on performance, with an increase of at most 1% and 0.4% in the proportion of birds and insects detected for their respective cases. The next case contained bird migration with heavily aliased velocities, so sector could not be obtained. The ridge classifiers detected a bird majority, with 82.9% – 90.5% of gates classified as birds. The base classifier and the one with range interval detected the highest proportion of birds. The addition of texture decrease the proportion of birds detected by about 7%

All training data was collected from KTLX. Thus, a classifier’s ability to detect biological patterns from other WSR-88Ds would provide strong evidence that it can be applied accross NEXRAD. Furthermore, the training data contained no bird roosts. The ability to detect these roosts would demonstrate that the classifiers are effective for diverse orientation of biota. The next case explores bird roosts collected by the KHTX radar (located in Huntsville, AL). Previous studies [7] identified groups of bird, insect and weather echoes for this scan. Sector could not be recovered because of the heterogenous mix of scatterers. The base classifier and the one with range interval detects the roosts as bird dominated. Their results also match the groups identified in [7]. The addition of texture seems to degrade performance on the roosts, probably because it is less suited for capturing finer features. The final case also contains bird roosts, collected by KTLX. Four flocks of roosting birds were identified against a background of insect echoes. Additionally, a bird/insect fuzzy logic classifier developed in [22]

identifies three of these roosts (the fourth roost was outside the fuzzy classifier’s range limit) as birds. Again, the base classifier and the one with range interval detect all the roosts as birds. The results also match those in [22].

Overall, the base classifier shows consistently strong performance across all the scenarios considered. The addition of range interval and sector seems to cause no noticeable improvement, while texture seem to favour detecting broader scale migration at the expense of finer features. The base classifier was also demonstrated to be effective when applied to new situations involving roosting birds and a different radar (KHTX), showing its potential to be used on the US national network of S-band polarimetric weather radars. As such, the implementation of this classifier as an accessory to the HCA is recommended. For every range gate, the HCA would first be applied. If the HCA detects biological echoes, the base classifier is then used to further distinguish the gate as containing either bird or insect echoes.

There are many potential improvements to identifying bird and insect radar echoes. The HCA can misclassify biological echoes leading to worse BIRC performance. As such, future verification and improvements of HCA’s predictions would be beneficial. Furthermore, this study has broadly explored birds and insects, however, even within these classes different species can have different scattering characteristics. More work is needed to understand and incorporate this information into a classification model. Ultimately, ground truth validation remains the biggest challenge. One approach being considered is validating the developed classifiers using measurements from other sensors. It is hoped that this research also encourages more data collection and verification efforts.

Chapter 4

Application of the Bird-Insect Ridge Classifier for VAD Wind Estimation

4.1 Introduction

The VWP is one of the most widely used meteorological radar products [16], provided at each of the 160+ weather radar stations comprising NEXRAD. At a given height, the horizontal wind velocity is estimated using VAD under the assumption of a uniform wind field, where a sinusoid is fitted to the Doppler velocities from different azimuth angles [16]. Fundamentally, it assumes that these echoes trace the underlying wind. However, different types of echoes have different wind relative motion.

This study focuses on clear-air conditions, where boundary layer radar echoes are predominantly from birds and arthropods (mostly insects) [15, 17, 37, 49]. Airborne insects are generally good wind tracers [15], producing reliable VAD winds. Returns from flying birds, on the other hand, have been found to contaminate Doppler radar velocities [6, 39, 76, 77, 85], resulting in biased VAD measurements. For instance, the National Centers for Environmental Prediction (NCEP) have suggested nocturnal bird migrants as the most common source of nighttime VAD wind contamination. [39, 99]. This is consistent with the investi-

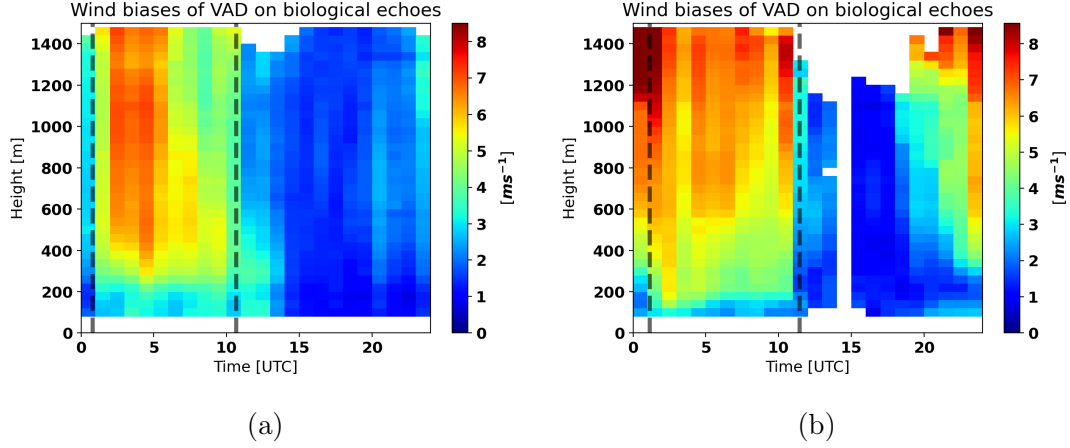


Figure 4.1: VAD wind biases averaged by height and time of day for (a) nocturnal bird migration observed by KOHX (Nashville, TN) during May 2018. (b) diurnal bird migration observed by KHGX (Houston, TX) during May 2022. The black dashed lines indicate sunset around 1 UTC and sunrise around 11 UTC. Local time is UTC - 5.

gations in this work. VAD biases relative to the wind were explored for two bird migration seasons. Further details on the analysis procedure will be presented in Section 4.4. For now, discussions are focused on Figure 4.1, which shows averages of VAD biases across each season, by height and time of day. Local sunset and sunrise are respectively denoted by black dashed lines around 1 and 11 UTC. The results in Figure 4.1(a) are from a nocturnal migration period collected by the KOHX radar (Nashville, TN) during May 2018. It can be observed that large biases ($\geq 5 \text{ ms}^{-1}$) occurred during the nighttime (1 - 11 UTC) and for heights $> 200 \text{ m}$, where migrating birds are expected to be present. Around sunrise, VAD wind biases drastically drop and persist over the course of daytime. This is consistent with a switch to insect domination at the crack of dawn. Figure 4.1(b) shows results from a diurnal migration period collected by KHGX (Houston, TX) during May 2022. A similar trend is observed, except for large biases in the late afternoon (20-24 UTC) and once again, for heights $> 200 \text{ m}$. This is consistent

with higher expected daytime bird activity.

Given the disparate impacts of birds and insects, it would be beneficial to account for taxa in the VAD process. For example, meteorologists can obtain less contaminated wind products from insects [41]. Conversely, VAD on bird flocks can be used to map their migratory pathways and warning systems developed to better mitigate collisions with aircraft, wind farms, off-shore platforms and other man-made structures [17]. More generally, aero-ecological studies of each animal can be enhanced by having more precise information. Despite these benefits, NEXRAD's operational VAD algorithm does not account for biological taxa, because of the difficulty of performing classification in a generalizable manner. Migration patterns are location dependent. Moreover, they characterize expected activity over a longer contiguous period, often different from more diverse bird/insect behaviour across individual scans. It is possible for birds to migrate at any time of the day, month and year, in any direction, with any relationship to the weather, and ultimately, with yearly evolving patterns [70]. And while birds might be expected to have larger biases than insects, birds are capable of wind tracing in situations where it matches their intended velocity and insects, deviating in cases of mass alignment. Besides, interactions between birds and insects are commonplace. In light of the aforementioned considerations, gate level taxonomic classification of biota is needed on NEXRAD.

In order to meet this need, BIRC was developed in Chapter 3 and tested on a variety of cases including bird roosts, mass bird/insect migration and mixed cases, spanning multiple WSR-88Ds located in Oklahoma, Alabama, Tennessee and Mississippi [100]. This chapter investigates the application of BIRC to enhance clear-air VADs by incorporating information on biological taxa. The goal is to identify the source of VAD wind biases and potentially mitigate them. Subsequent

sections further describe the proposed method (Section 4.2), data preparation for the two bird migration seasons (Section 4.3) and the wind bias metric used for evaluation (Section 4.4). Following descriptions of the necessary materials and methods, VAD contamination is studied based on the amount of constituent bird echoes (Section 4.5). Sometimes, these contaminated VADs simultaneously contain uniform insect migration. Improvements of wind estimation by focusing VAD on these insects is explored in the second part of this work (Section 4.6). Final recommendations and conclusions are presented in Section 4.7.

4.2 Proposed Method

The proposed integration of BIRC into the VAD wind estimation process is shown in Figure 4.2. For each VAD ring, HCA detects gates as biological, weather or others. A correction procedure is then applied to rectify misclassifications, after which BIRC is used to bifurcate biological echoes into birds and insects. These classifications are used to generate three products namely: the insects-birds ratio, VAD on birds and VAD on insects.

The insects-birds ratio is explored for identifying bird contamination (or insect tracing) levels in Section 4.5. In the event of distinct simultaneous bird and insect migration, insects should better align with the wind. Thus, improvements derived from the insect VAD product are explored in Section 4.6. The required conditions of coincident and distinct bird/insect migration, are respectively determined using the insects-birds ratio and comparisons between bird and insect VAD biases. The rest of this section further details the configuration of HCA, BIRC and VAD algorithms for this study.

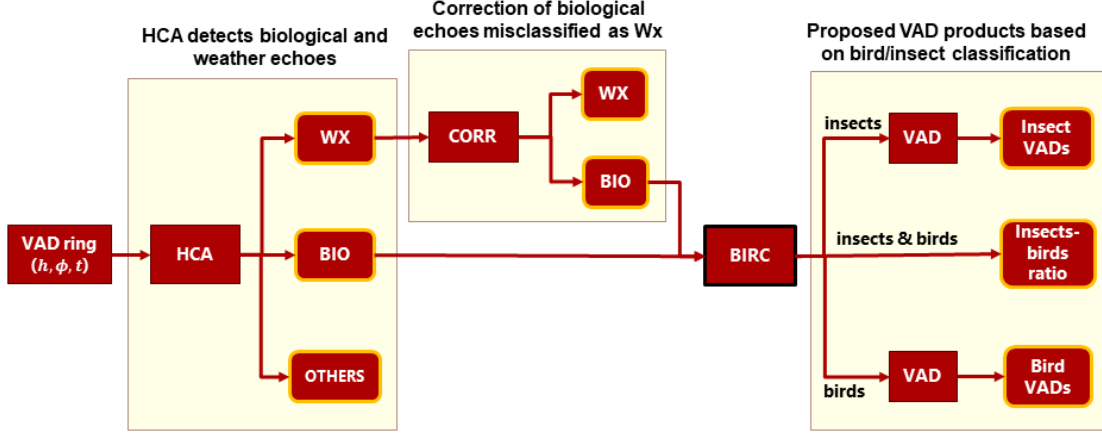


Figure 4.2: Proposed VAD modifications based on BIRC's predictions. WX stands for weather while BIO represents biological echoes

4.2.1 Detecting Biological and Weather Echoes

In this work, HCA is mainly used to identify biological and weather echoes. Weather encompasses the 8 normally provided meteorological classes including dry aggregated snow, wet snow, crystals, graupel, big drops, light and moderate rain, heavy rain and a mixture of rain and hail [29]. In order to ensure clear-air conditions, free from meteorological scatterers, scans with $> 5\%$ of range gates containing weather are excluded from subsequent analysis. Biological echoes, on the other hand are latter classified by taxa.

While the HCA is generally reliable, owing to its rigorous use and testing, one observed issue is the misclassification of some biological echoes as weather. These are addressed by changing "weather" predictions to biological if $\rho_{HV} \leq 0.88$, the threshold selected based on weather echoes been found to have $\rho_{HV} \geq 0.85$ and biological echoes having $\rho_{HV} \leq 0.83$ [29]. Furthermore, correction was limited to heights below 1500 m, where most organisms are found [8].

Two main misclassification patterns were observed in the migration dataset.

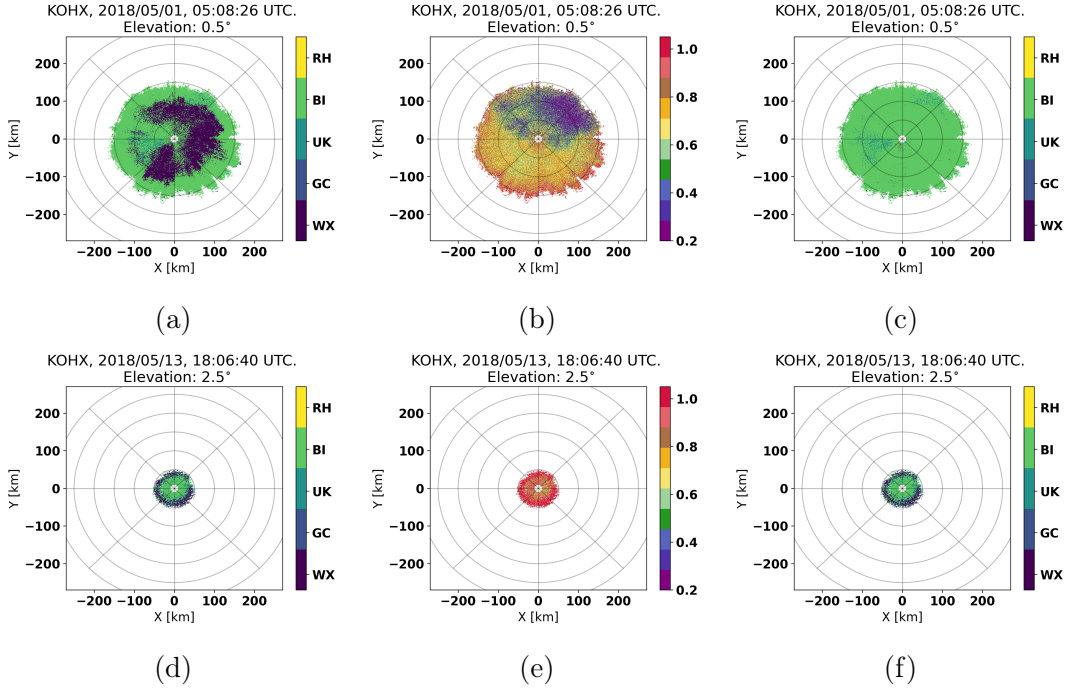


Figure 4.3: Correcting misclassifications of biological echoes from the Hydrometeor Classification Algorithm (HCA). (a) HCA predictions for nocturnal birds from the 0.5° elevation. (b) ρ_{HV} for nocturnal birds. (c) HCA predictions after correction. (d) HCA predictions for afternoon insects from the 2.5° elevation. (e) ρ_{HV} for afternoon insects (c) HCA predictions after correction.

Figure 4.3 provides a demonstration of the correction procedure on these patterns. The first case contains nocturnal bird migration around midnight. This scan was collected by KOHX from the 0.5° elevation at 05:08 UTC on 1 May 2018. HCA's outputs (Figure 4.3(a)) show prevalent "weather" within a 100 km radius of the radar, which is unlikely given their $\rho_{HV} \leq 0.8$ (see Figure 4.3(b)). The correction procedure changes most of these echoes to biological (Figure 4.3(c)). The second case is a more complicated insect scan collected by KOHX at 18:06 UTC on 13 May 2018. At the 2.5° elevation, HCA classifies gates close to the radar as biological, but the farthest ring of echoes are designated "weather" (Figure 4.3(d)). These rings are a peculiar feature of clear-air insect dominated scans in this study and

are likely also insects. However, their $\rho_{HV} \approx 1$ resembles weather (Figure 4.3(e)) and thus correction is ineffective (Figure 4.3(f)). In cases like these, the recovery of fewer insects samples would diminish VAD availability.

4.2.2 Classifying Biological Echoes into Birds and Insects

The BIRC model developed in Chapter 3 is applied to classify biological echoes as birds or insects. Mathematically, the taxa C is predicted based on an optimal threshold of 0.5 on the probability of bird presence p , expressed as

$$C(h, \phi, t) = \begin{cases} 1, & p \geq 0.5 \\ 0, & p < 0.5. \end{cases} \quad (4.1)$$

where h is height, ϕ is azimuth and t is time. $\{C = 0\}$ means insects detected and $\{C = 1\}$ means birds detected. BIRC's performance metrics are recaptured in Table 4.1 below.

Table 4.1: Performance metrics of BIRC

ACC	TPR	TNR	AUC
0.814 - 0.858	0.832 - 0.892	0.778 - 0.844	0.868 - 0.911

4.2.3 VAD Algorithm

Following the identification of biological echoes in clear-air sweeps and their separation by taxa, the next stage is VAD wind estimation. The VAD algorithm developed in Section 2.4 is used for subsequent experiments. Hereafter, the former is also called local VAD to distinguish it from the operational counterpart on NEXRAD.

A few modifications were made to the local VAD for this study. An important condition for harmonic VAD analysis is that the domain is well filled with organisms [8]. This was addressed by combining data from the 4 lowest elevation sweeps ($0.5 - 3.5^\circ$). Furthermore, a minimum sample size of 720 gates and coverage of 75% (based on suggestions in [8]), were ensured prior to VAD. Coverage is defined as the percentage of range gates containing radar echoes. Another important criteria is horizontally homogeneous flight orientation, which is usually disrupted by meteorological echoes [8]. VAD was limited to clear-air scans, defined as those containing $\leq 5\%$ weather. Furthermore, the VAD height grid was defined from 0 to 1500 m, in steps of 40 m.

Local VAD using all biological echoes were verified against NEXRAD's VWP for the KOHX (Nashville, TN) nocturnal migration season of May 2018. The results are shown in Figure 4.4. Both profiles are interpolated to the VAD's height grid, and their speed and direction differences calculated. The median differences are -0.88 ms^{-1} and 0.72° respectively, confirming the veracity of the local VAD algorithm.

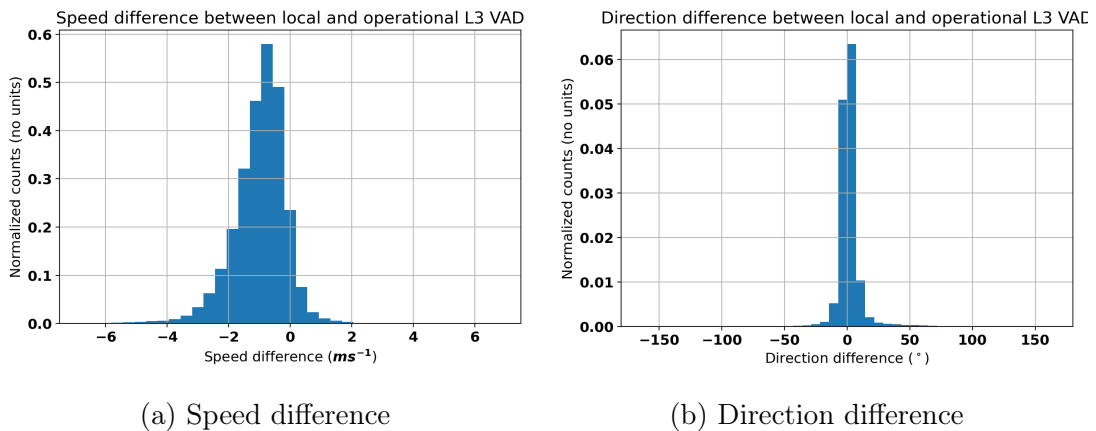


Figure 4.4: Comparison of local VAD on biological echoes with level 3 VWP for KOHX, May 2018. The median speed and direction differences are -0.88 ms^{-1} and 0.72° respectively

4.3 Data Preparation

VADs are evaluated with respect to ground truth wind measurements. Data for the former are obtained from NEXRAD while Rapid Refresh (RAP) [101] model predictions are assumed to approximate the ground truth. This section details the preparation process for both data sets.

4.3.1 NEXRAD Data

Since insects are ubiquitous among clear-air radar echoes, data acquisition is focused on two seasons of bird migration namely: nocturnal and diurnal. Table 4.2 summarizes the data sets for both periods. Nocturnal migrants typically travel during nighttime. Data for this season were collected from the KOHX (Nashville, TN) radar for the whole month of May 2018, including 7380 scans. Level II products are used as inputs to VAD and BIRC, while HCA and VWP predictions are retrieved from Level III. For the diurnal migration season, 6467 scans were collected from the KHGX (Houston, TX) for May 2022. These data sets can be obtained from the National Centers for Environmental Information (NCEI) NEXRAD data archive [102].

Table 4.2: Summary of data for bird migration seasons

WSR-88D	Location	Period	Bird migration	Number of scans
KOHX	Nashville, TN	May 2018	nocturnal	7380
KHGX	Houston, TX	May 2022	diurnal	6467

The distribution of echoes were verified for both migration seasons. In each scan, HCA and BIRC were used to detect the frequency of weather, bird and

insect echoes, and their relative proportions obtained by normalization to sum to 1. Figure 4.5 shows the monthly averages of these proportions by the time of day. Nighttime is generally defined as 1-11 UTC while the remaining period is daytime. Daily interchanging domination of birds during nights and insect during daytimes can be observed during the nocturnal season (Figure 4.5(a)), while bird domination is everpresent during the diurnal season (Figure 4.5(b)). This matches the expected bird migration behaviour. Weather is also low for both seasons being generally $\leq 35\%$. However, subsequent VAD analysis of biological echoes is further constrained to scans with $\leq 5\%$ weather to mitigate meteorological interruptions.

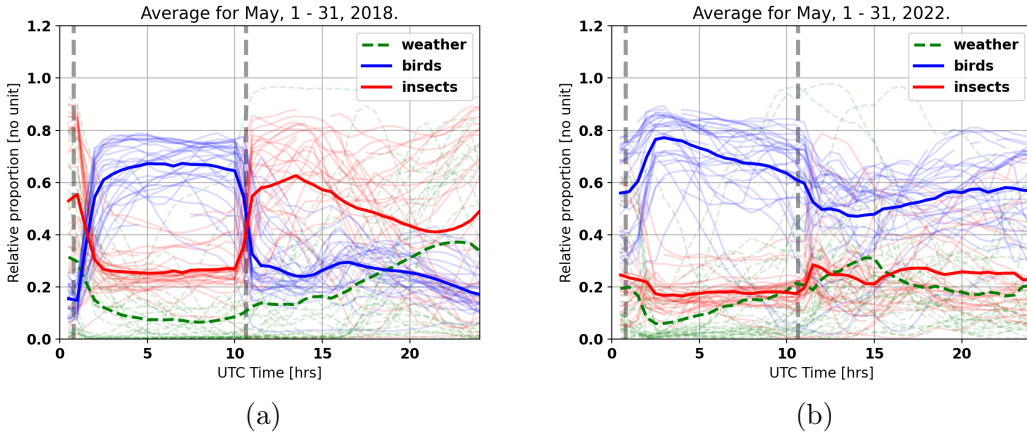


Figure 4.5: Daily averages of relative bird, insect and weather proportions for the (a) May 2018 KOHX nocturnal migration season (b) May 2022 KHGX diurnal migration season. Local time is UTC - 5

4.3.2 RAP as Ground Truth

There are other wind velocity measurements like balloon soundings [103] and RAP [101]. Balloon soundings by the National Weather Service (NWS) measures wind speed and direction as a function of pressure level. Sounding data can be downloaded from the University of Wyoming’s upper air data portal [103].

RAP, on the other hand, is a continental scale assimilation and forecast system operational at NCEP [101, 104]. Alongside VAD, these measurements enable robust wind sensing across the continent, complementing each other with different estimation modalities at different times and locations. In this section, RAP and sounding are evaluated to select the best ground truth reference for VAD based on proximity to radar stations and time synchronization of measurements.

VAD wind estimation stands out because of NEXRAD’s data availability, each of the 160+ stations providing data every 5-10 minutes. While soundings have the advantage of directly measuring the wind, they provide limited coverage. Temporally, balloons are only launched twice everyday, at 00:00 UTC and 12:00 UTC [8]. Spatially, sounding stations are not always close enough to radar sites. For example, the KOHX radar ideally has the Nashville sounding station 0.74 km away whereas KHGX’s closest station is 193.79 km away in Lake Charles, Louisiana (see Table 4.3). Overall, approximately 41% of stations are within a 50-km range and 53%, within a 100-km range [8]. RAP, on the other hand, provides 1-h updates at 50 vertical levels and 13-km horizontal grid spacing [8]. They are generally better synchronized with NEXRAD. For example, the closest RAP points to KOHX and KHGX are 2.68 km and 6.71 km away respectively (see Table 4.3). Because of the better space-time coverage, RAP is used for evaluating VAD wind profiles.

Table 4.3: Distance between KOHX, KHTX and closest RAP and sounding stations.

	KOHX	KHGX
Closest Rapid Refresh point (km)	2.68	6.71
Closest sounding station (km)	0.74	193.79

The accuracy of RAP was further verified using sounding data. Since the Nashville sounding is 0.74 km away, it is assumed to provide the true wind mea-

surements for KOHX. Thus, during May 2018, each sounding was compared to RAP predictions within a two hour period e.g the 12 UTC sounding is compared to predictions within 10-14 UTC. Both profiles are interpolated to a common height grid defined from 0 to 1500 m, sampled every 40 m. Furthermore, RAP predictions are converted from u and v wind components to speed and direction, and subtracted from soundings. The results are shown in Figure 4.6, with median speed and direction differences of 0.10 ms^{-1} and 4.5° respectively, demonstrating the reliability of RAP predictions. Mismatches of 3.13 km in space and ≤ 2 hours in time would contribute to these slight differences.

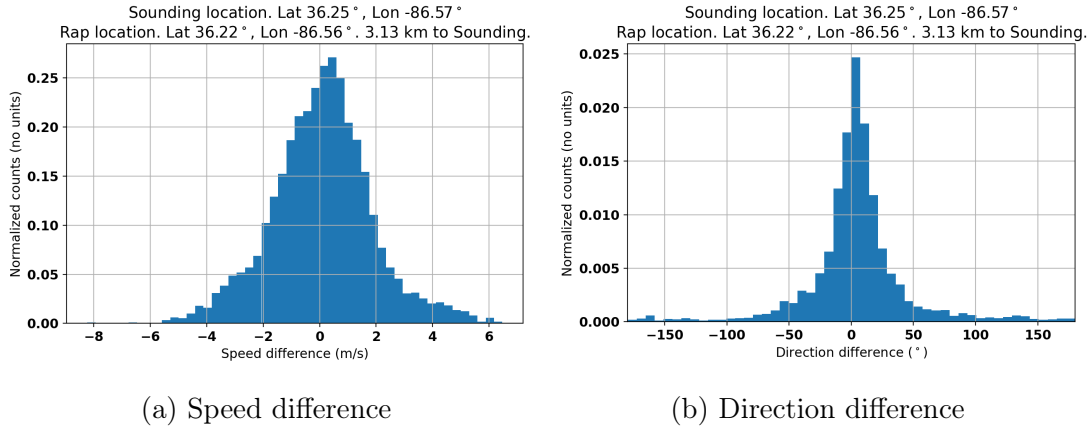


Figure 4.6: Comparison of RAP and sounding data for KOHX, May 2018. The median speed and direction differences of 0.10 ms^{-1} and 4.5° respectively.

4.4 Wind Bias for Evaluating VADs

VAD is an indirect wind estimation technique. It tracks the flight of radar echoes, which in turn, leverage the wind in varying degrees. The discordance between the estimated VAD and the underlying wind is termed wind bias. For aerial organisms, this approximates their *airspeed* [8]. Explained another way, VAD measures

ground speed, the velocity of organisms relative to the stationary earth surface. Ground speed includes contributions from independent flight and the wind. Thus, the flight component can be isolated by removing the wind contribution, resulting in airspeed i.e the velocity of organisms relative to the wind. Wind bias (or airspeed) e is calculated as

$$e(h, t, c) = \|\vec{V}_{mig}(h, t, c) - \vec{V}_{wind}(h, t)\| \quad (4.2)$$

where c is the taxa and $\vec{V}_{wind}$ is the wind velocity, obtained from RAP in this study. A bias of 0 implies perfect wind tracing whereas large biases signify poor wind tracing. The biases for nocturnal and diurnal migration seasons in Figure 4.1 were computed using Equation (4.2). The taxa was all biological echoes.

4.5 Identifying the Source of VAD Biases

This section investigates the cause of VAD bias variations during migration periods (see Figure 4.1). Here VAD is calculated using all biological echoes, and is subsequently also called biological VADs. The problem is elucidated in Figure 4.7, where the VAD wind profiles of two scans from KOHX reveal significant wind deviations in one and agreement in the other. The first scan was collected during nocturnal bird migration around midnight, 05:04 UTC on 8 May 2018. Radial velocities show widespread alignment with a common heading toward the radar’s north-west (Figure 4.7(a)). Furthermore, comparisons between VAD and RAP wind reveal substantial differences, especially between the v components at high altitudes (Figure 4.7(b)). As such, VAD wind biases are large, ranging from ~ 3 to 10 ms^{-1} and increasing with height (Figure 4.7(c)). The second case was collected at 15:47 UTC the next day. Echoes also have a common heading

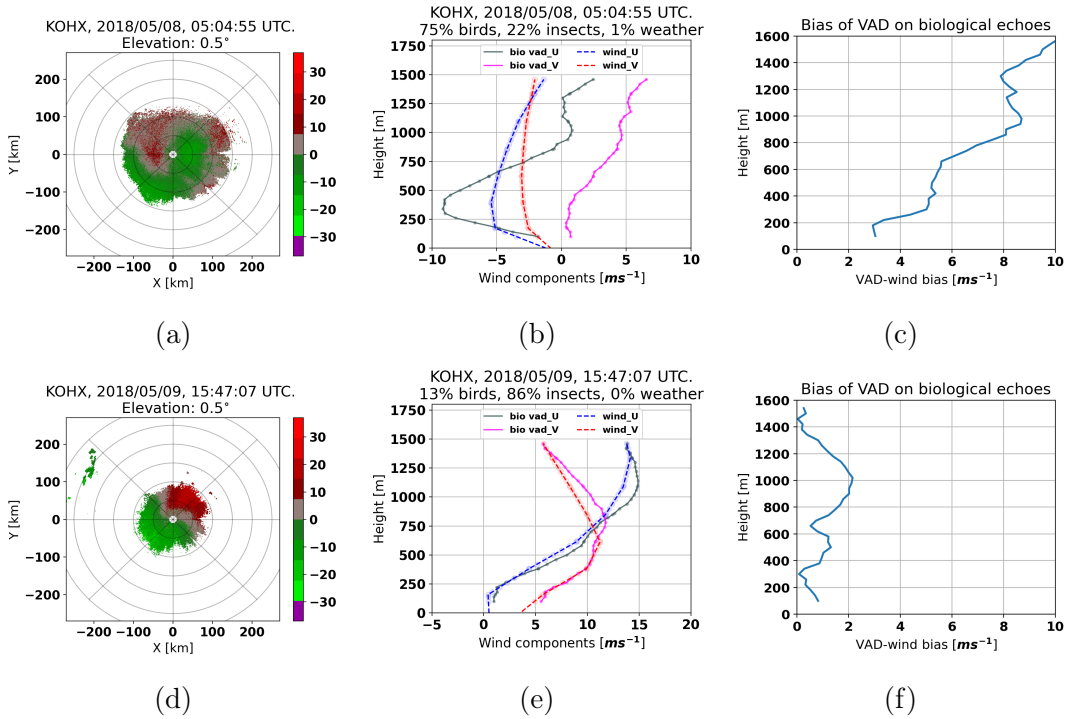


Figure 4.7: (a) Radial velocity for a nocturnal bird migration case. (b) The u (grey) and v (pink) wind components estimated from VAD have large deviations from RAP’s u (blue) and v (red) components. (c) Large wind biases due to deviations measure the bird’s airspeed. (d) Radial velocity for a daytime insect swarm. (e) Insects’ VAD trace the RAP wind with small deviations. (f) Small wind biases measure the insect’s airspeed.

towards the radar’s north-east, but with larger radial velocities (Figure 4.7(d)). RAP reveals these velocities to be the result of stronger underlying winds, which are traced by the echoes’ VADs (Figure 4.7(e)). As such, wind biases are low with values $< 2.1 \text{ ms}^{-1}$ (Figure 4.7(f)).

4.5.1 Insects-to-Birds Ratio of Range Rings

Birds have been reported to contaminate VAD, while insects are generally considered as good wind tracers. Wind bias magnitudes should therefore, be related to their relative distribution: VAD rings with a bird majority and few insects should

have large biases, while insect-dominated rings should be more accurate. To summarize the taxonomic distribution of range rings, an insects-birds ratio $R_i(h, t)$ is defined as

$$R_i(h, t) = \frac{N_i(h, t)}{N_i(h, t) + N_b(h, t)}, \quad (4.3)$$

$$N_b(h, t) = \sum_i^n C(h, \phi_i, t) \quad (4.4)$$

$$N_i(h, t) = \sum_i^n 1 - C(h, \phi_i, t) \quad (4.5)$$

where $N_b(h, t)$ is the number of birds predicted by BIRC and $N_i(h, t)$ is the number of insects. $R_i = 0$ implies that the ring is 100% birds while $R_i = 1$ implies 100% insects.

Insects-birds ratios R_i are investigated for the cases discussed in Figure 4.7. The results are shown in Figure 4.8. For the nocturnal migration case, BIRC classifies a majority, 75% of echoes, as birds (Figure 4.8(a)). The number of bird and insect echoes are counted within each ring using Equations (4.4) and (4.5). Birds are found to outnumber insects across all heights, especially 200 – 1200 m (Figure 4.8(b)), leading to consistently low insects-birds ratios about a median of 22% (Figure 4.8(c)). The second case is estimated to contain 86% insect gates (Figure 4.8(d)). Across all heights, insects outnumber birds, especially within 200 – 800 m (Figure 4.8(e)), resulting in a high median insect-to-bird ratio of 90% (Figure 4.8(f)). Comparing biases to the insects-birds ratio, a pattern begins to emerge. Large biases for the nocturnal migration scan (Figure 4.7(c)) correspond to low insect-bird ratios (Figure 4.8(c)), whereas low biases for the

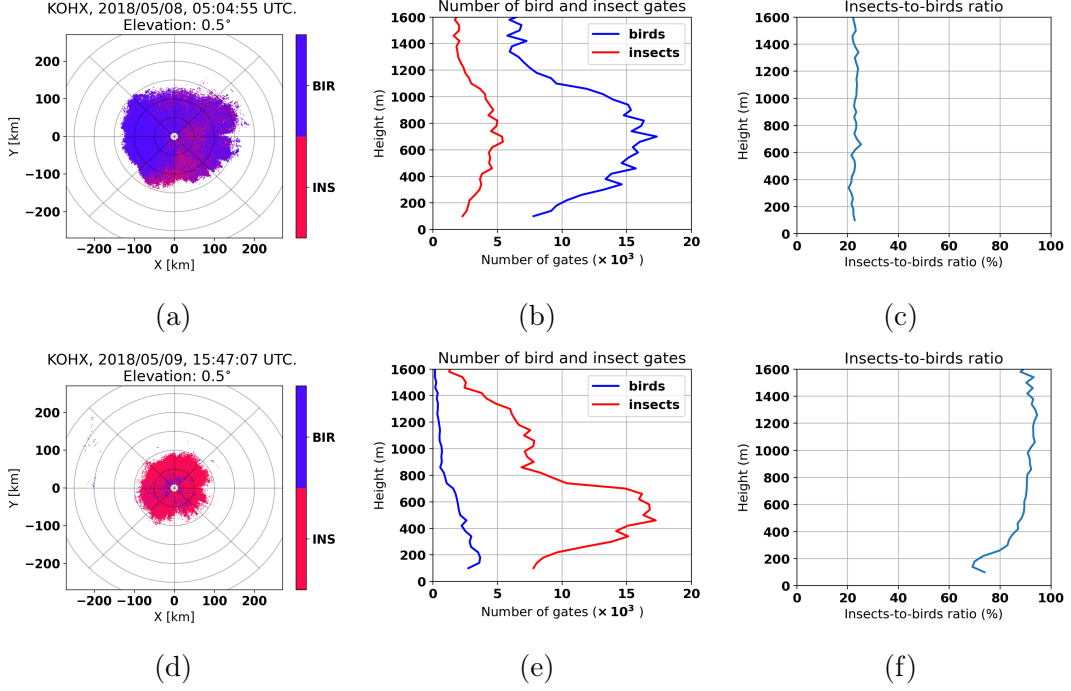


Figure 4.8: (a) BIRC predictions for a nocturnal bird migration case. (b) The number of bird gates exceeds those for insects across all heights. (c) Insects-to-birds ratio is low ($\sim 22\%$) because birds dominate insects. (d) BIRC predictions for an afternoon insect swarm. (e) The number of insect gates exceeds those for birds across all heights. (f) Insect-to-bird ratio is high ($\sim 90\%$) indicating insect dominance.

insects dominated scan (Figure 4.7(f)) correspond to high ratios (Figure 4.8(f)).

Analysis were also conducted for the whole nocturnal and diurnal migration seasons. The height by time of day averages for insects-birds ratios are shown in Figure 4.9. Matching expectations, the nocturnal migration season (Figure 4.9(a)) is revealed to have low ratios ($R_i \sim 28\%$) during nighttime (1-11 UTC) indicative of bird domination, and high ratios ($R_i \sim 77\%$) during daytime due to insects dominating. A notable difference for the diurnal migration season (Figure 4.9(b)) is observed during daytime. Birds seem to emerge late morning ~ 15 UTC and progressively occupy higher altitudes over the rest of the day. By 18 UTC through 1 UTC the next day, ratios are between 30–60% indicating simultaneous bird and

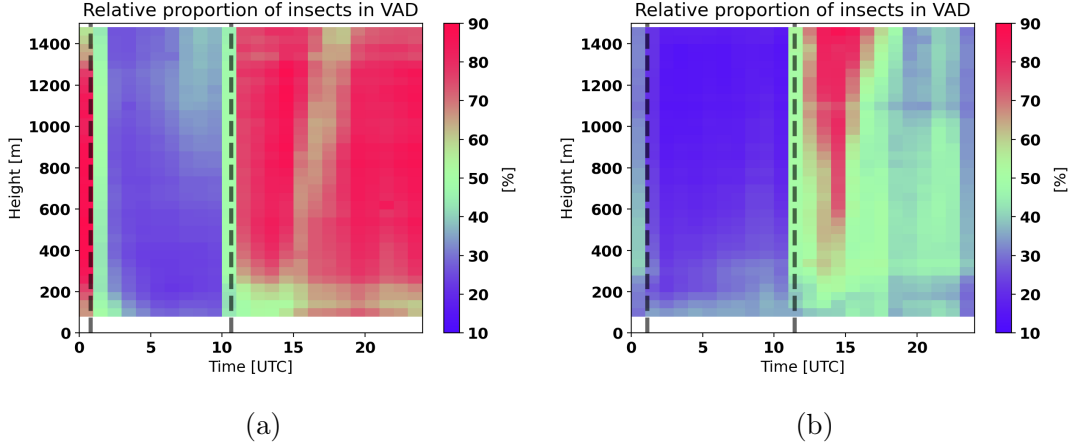


Figure 4.9: Insects-birds ratios averaged by height and time of day for the (a) KOHX nocturnal migration season (b) KHGX diurnal migration season. The black dashed lines indicate sunset and sunrise around 1 and 11 UTC. Local time is UTC - 5

insect activity. This is expected for diurnal bird migrants. Nighttimes still possess the most significant bird activity with $R_i \sim 19.8\%$. Once again, comparisons of biases to insects-birds ratios (Figure 4.1 to Figure 4.9) uncovers a pattern of large biases for low R_i bird dominated rings and low biases in high R_i insects dominated rings.

Additional properties of bird/insect activity are revealed by the daily averages of their gate counts within a VAD ring (Figure 4.10). During the nocturnal season, birds are most frequent within heights of 150 – 1200 m in the nighttime, as shown in Figures 4.10(a), while insects are prevalent at lower altitudes of 150 – 800 m during daytime, shown in Figure 4.10(c). Some differences can be observed for the diurnal season. Birds dominate higher altitudes, frequently occurring within 150 – 1400 m (Figure 4.10(b)). Insects are still found at 150 – 800 m, but are largely absent during daytime bird ascendance (Figure 4.10(d)).

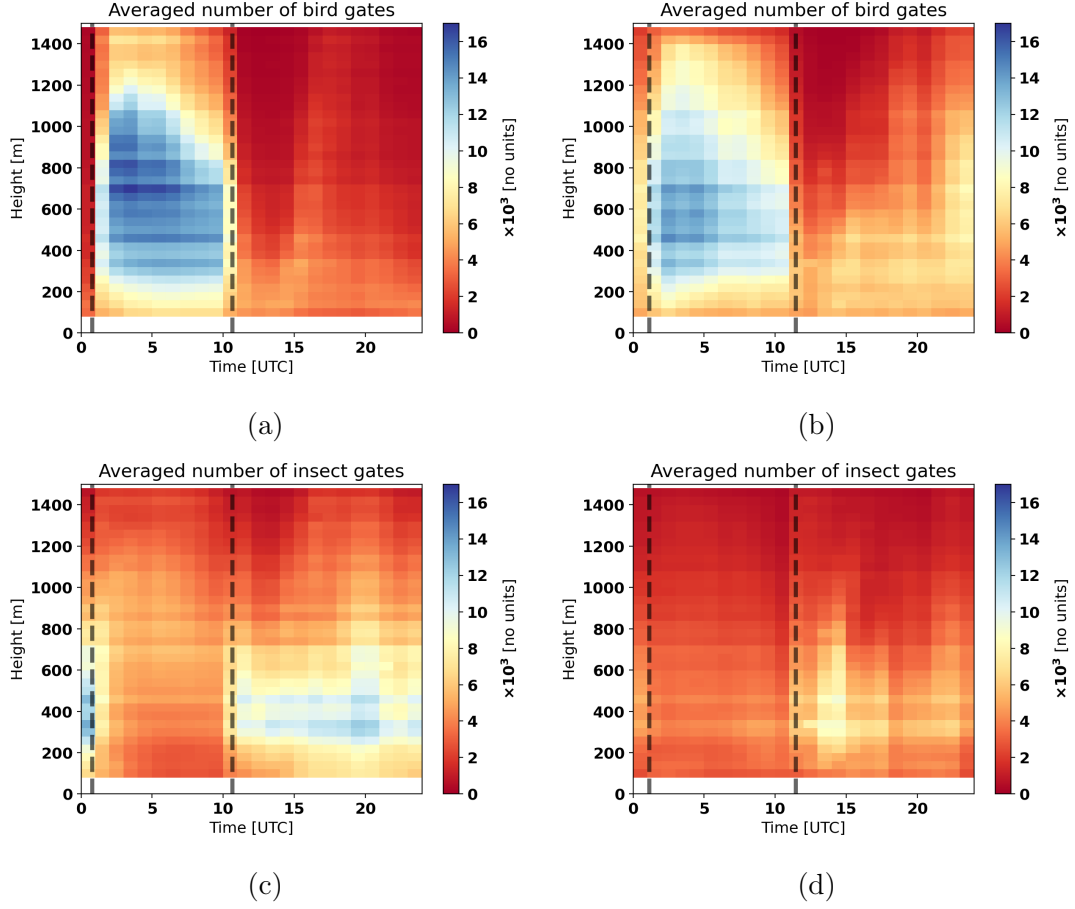


Figure 4.10: Daily averages of total bird and insect gates within a VAD ring. From left to right, columns correspond to nocturnal and diurnal migration seasons. From top to bottom, rows correspond to bird and insect gates count.

4.5.2 Analysis of Wind Biases

VAD wind biases were analyzed based on the insects-birds ratio, in order to verify their hypothesized inverse relationship. The results are shown in Figure 4.11. Figures 4.11(a) - 4.11(b) shows bias averages according to height and the insects-birds ratio, for the nocturnal and diurnal seasons respectively. Both seasons reveal identical trends of decreasing biases for increasing R_i i.e going from left to right at a given height. An accumulation of the diurnal season's biases toward $R_i \approx 0$ can also be observed. This is because of the omnipresence of birds. Scatterplots of

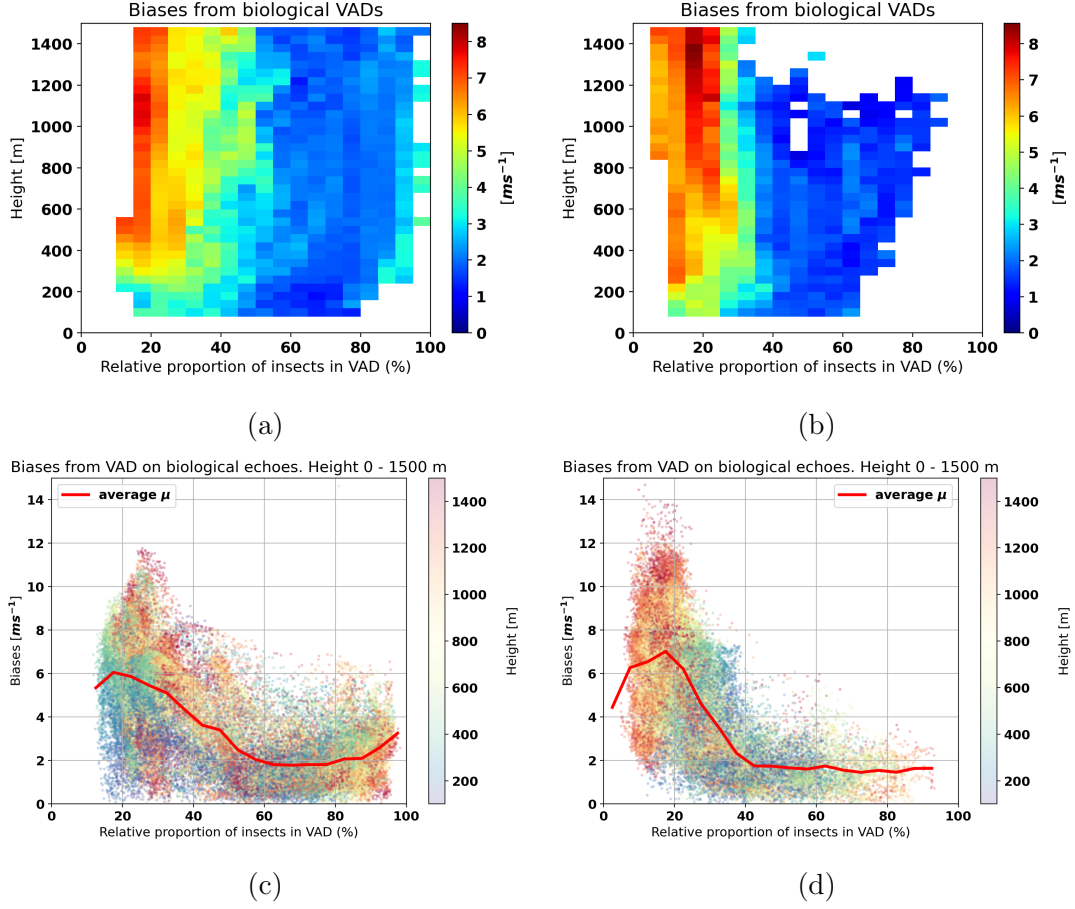


Figure 4.11: (a) - (b) Wind biases averaged by height and the insects-birds ratio. (c) - (d) Scatterplots for wind bias against insects-birds ratio. From left to right, columns correspond to nocturnal and diurnal migration seasons. Biases can be seen to reduce as the insect-bird ratio increases (i.e fewer birds, more insects)

individual VAD biases as a function of R_i for both seasons, are shown in Figures 4.11(c) - 4.11(d). Each point is colored by its height. Following the red line, mean biases generally decrease as R_i increases. This trend can be explained by the capability of birds to generate larger wind independent flight velocities than insects. It also follows that bird biases would have larger variances, as seen in the greater spread of points around $R_i \sim 0\%$ compared to lower variations for $R_i \sim 100\%$.

Several conclusions are drawn from these results. First, birds are found to

be the primary biological source of clear-air VAD biases, consistent with previous findings [39, 99]. Second, BIRC can be applied for directly measuring bird contamination via the insects-birds ratio. A stand out feature of this approach is the use of only dual-pol variables, making it generalizable across NEXRAD. For example, different bias-time patterns between the KOHX nocturnal and KHGX diurnal seasons (Figure 4.1), become identical bias-ratio patterns (Figure 4.11). This means that for a given height, from any WSR-88D, the insects-birds ratio can be computed and the wind contamination level inferred based on this ratio, regardless of time and location.

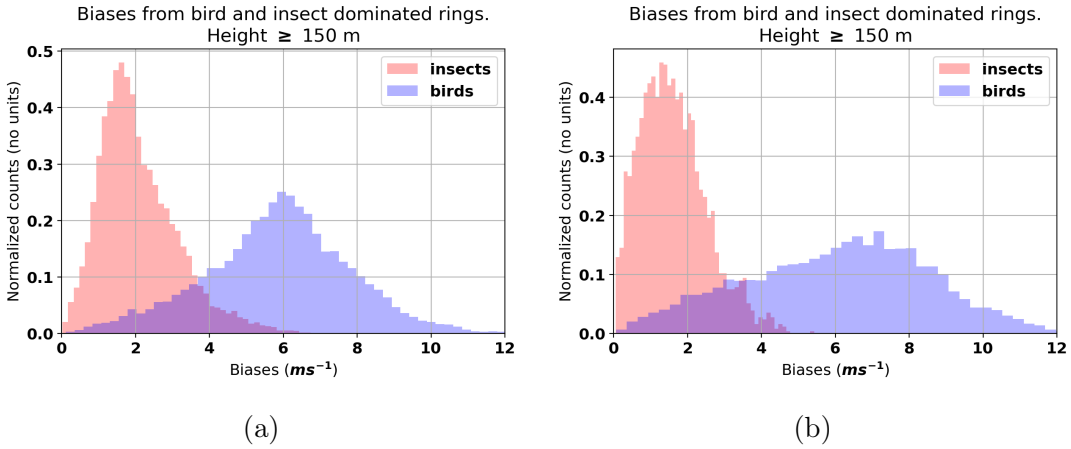


Figure 4.12: Comparisons of biases from bird dominated (blue) and insect dominated (pink) rings for the (a) KOHX nocturnal migration season. (b) KHGX diurnal migration season.

Finally, the results provide further support for the quality of BIRC’s predictions. Previous research has reported birds as having larger biases than insects, with a threshold of 5-7 ms^{-1} separating both groups [7]. To validate these findings, biological VADs are first partitioned into 3 groups including bird dominated, mixed and insect dominated rings, defined by $R_i \leq 40\%$, $40\% < R_i \leq 60\%$ and $R_i > 60\%$ respectively. Biases from bird dominated rings were then compared to

insect dominated rings. The results are shown in Figure 4.12. Medians for both regions were found to be 5.9 ms^{-1} and 1.9 ms^{-1} for the nocturnal season, and 6.2 ms^{-1} and 1.5 ms^{-1} for the diurnal season. Overall, a threshold around $\sim 4 \text{ ms}^{-1}$ was found to separate both groups, similar to the aforementioned findings. The slight bias threshold difference could be due different VAD algorithms.

4.6 Improving Wind Estimates from Contaminated VADs

In the previous section, biases of clear air biological VADs were found to depend on their taxonomic distribution. For example, bird dominated rings typically had large biases $\sim 6 \text{ ms}^{-1}$. However, the remaining gates could contain insect echoes on which, under the right conditions, VAD can be performed to obtain more reliable wind estimates. This process of focusing VAD on wind tracing insects is termed improvement. Conversely, it is possible to obtain better bird VADs for ornithological studies by using only birds. Wind improvements and its necessary preconditions are explored in this section.

4.6.1 Birds-Insects Bias Differences for Identifying Improvement Conditions

Improvement of a bird contaminated ring requires that constituent insects' velocities are different from those of birds. Ideally, there should be two sinusoids, one from each taxa, with birds having wind biases $\sim 4 \text{ ms}^{-1}$ larger than insects. Hence, a bias difference $\Delta e(h, t)$ is defined as

$$\Delta e(h, t) = e_{birds}(h, t) - e_{insects}(h, t) \quad (4.6)$$

where $e_{birds}(h, t)$ is the bias from birds' VAD and $e_{insects}(h, t)$ is the bias from insect's VAD.

Bird-insect bias differences are used to ascertain the viability of improvement. Given a VAD ring, $\Delta e \approx 0$ means bird and insect VADs have equal wind deviations, in which case improvement is not possible. VAD can only be improved when birds have larger biases i.e $\Delta e > 0$. Ideally, $\Delta e \approx 4 \text{ ms}^{-1}$ indicates two sinusoids dominated by birds and insects respectively. Thus, focusing VAD on insects would produce more accurate wind estimates. Insects having larger biases is highly unlikely.

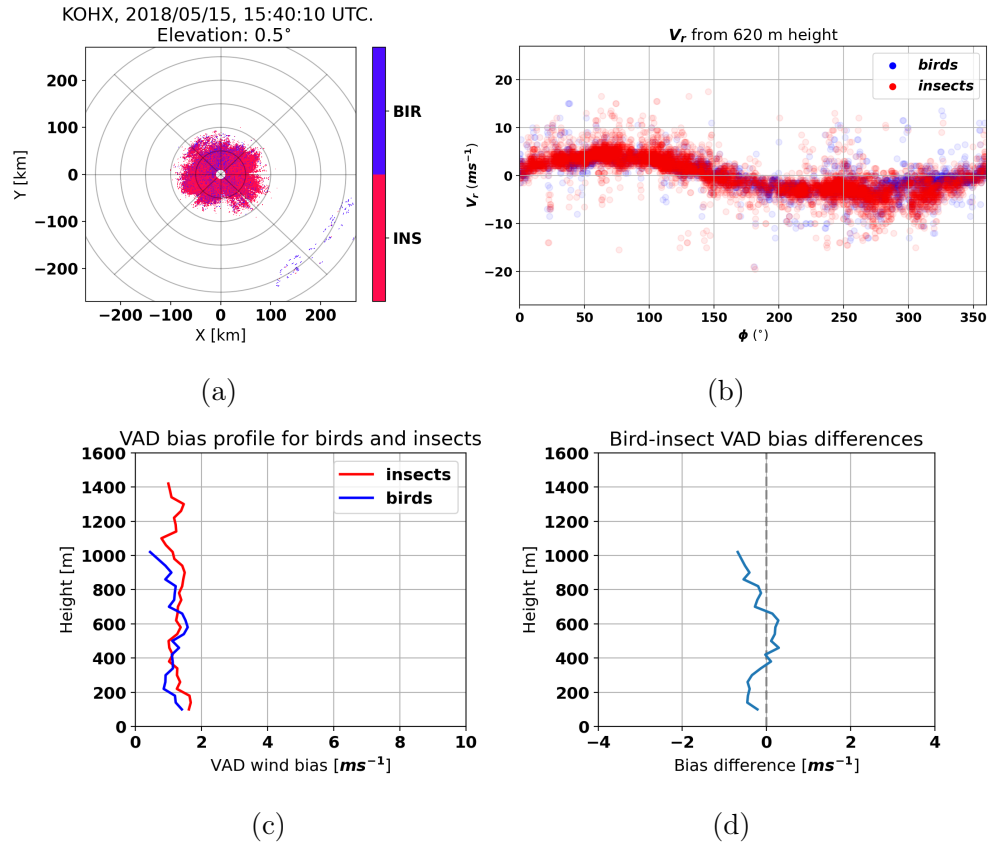


Figure 4.13: (a) BIRC predictions for an insect swarm case. (b) Radial velocities from 620 m height. (c) Biases from predicted birds and insects. (d) Difference between bird and insect biases. Red points are from insects and blue are from birds.

Three examples from KOHX are provided for demonstration. The first scan is insect dominated and collected at 15:40 UTC on 15 May 2018 (Figure 4.13(a)). A sample 620 m height cut of radial velocities reveals one insect dominated sinusoid (shown in Figure 4.13(b)). VAD is performed on these insects and their biases calculated using Equation (4.2). As expected, insects have low biases $< 2 \text{ ms}^{-1}$ (Figure 4.13(c)). The few embedded bird echoes, on the other hand, are largely misclassifications and unrepresentative of distinct other migration. They are estimated with identical biases to insects. As a result, bias differences are $\sim 0 \text{ ms}^{-1}$ (Figure 4.13(d)) and improvement is not feasible.

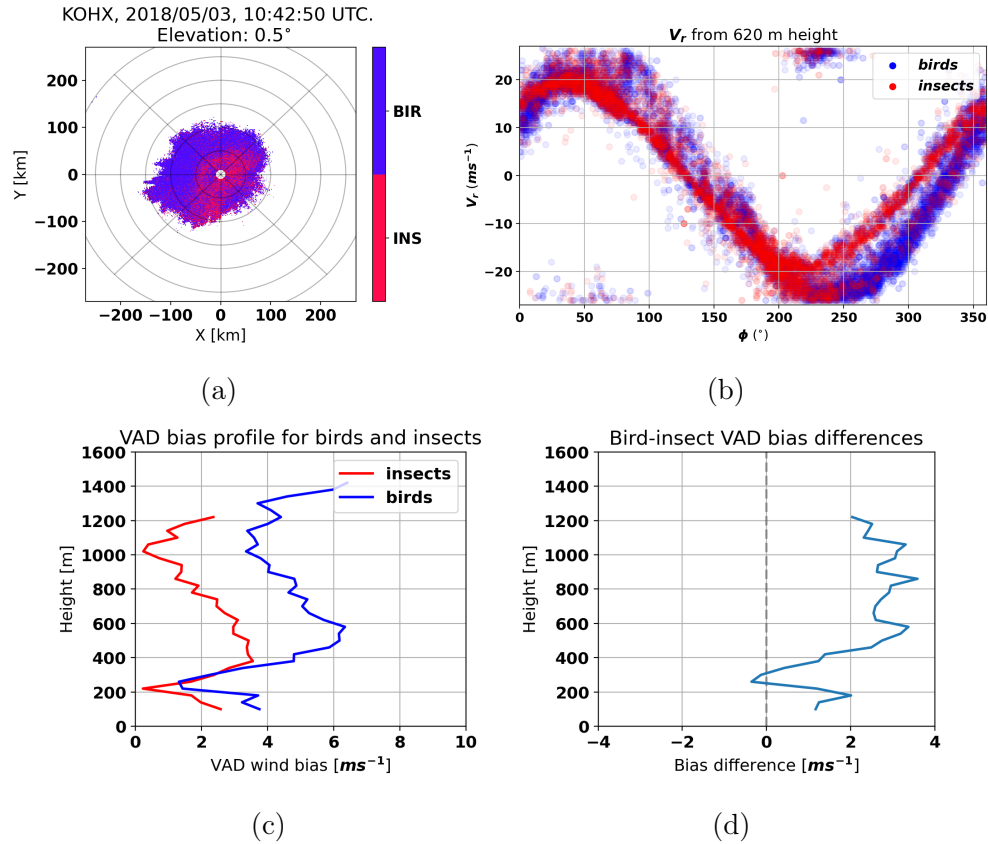


Figure 4.14: (a) BIRC predictions for a mixed case. (b) Radial velocities from 620 m height. (c) Biases from predicted birds and insects. (d) Difference between bird and insect biases.

The second case is a mixed scan collected at 10:42 UTC on 3 May 2018 (Figure 4.14(a)). Velocities cut from the 620 m height reveal simultaneous bird and insect migration, in the form of two sinusoids (shown in Figure 4.14(b)). Biases calculated from their VADs show birds with greater or equal deviations across all heights, especially > 400 m (Figure 4.14(c)). For the latter region, bias differences ranges from $1 - 3.5 \text{ ms}^{-1}$ (Figure 4.14(d)). Insects focused VADs would be more accurate in this case.

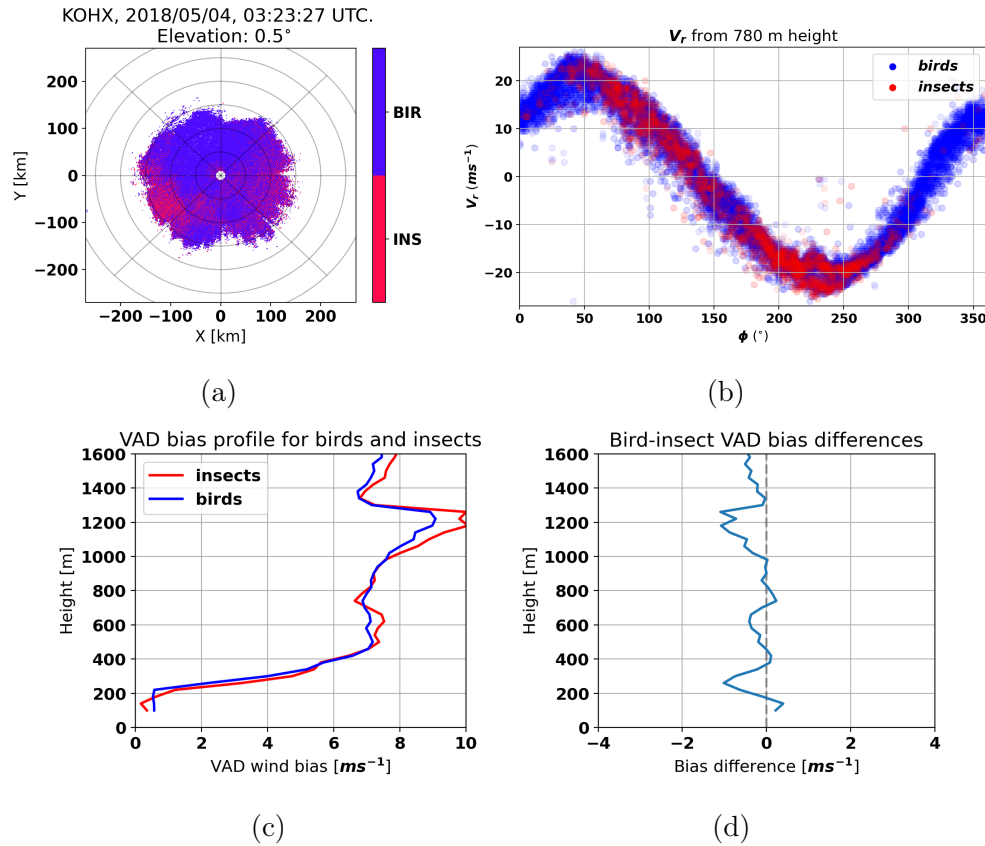


Figure 4.15: (a) BIRC predictions for a bird dominated case. (b) Radial velocities from 780 m height. (c) Biases from predicted birds and insects. (d) Difference between bird and insect biases.

The final case is a bird dominated scan, collected at 03:23 UTC on 4 May 2018 (Figure 4.15(a)). Velocities cut from the 780 m height reveal one bird dominated

sinusoid (Figure 4.15(b)). As expected, these birds have large biases, especially for heights > 400 m where values range from $5 - 9 \text{ ms}^{-1}$ (Figure 4.15(c)). The few insects echoes are largely misclassifications, and as such estimated with identical biases to birds. Bias differences are $\sim 0 \text{ ms}^{-1}$ (Figure 4.15(d)) indicating that improvement is not feasible.

4.6.2 Analysis of Bird-Insect Bias Differences

Bird-insect bias differences, Δe , were analyzed for both migration seasons based on the time of day and birds-insects ratio. For the sake of brevity, $\Delta e > 0$ indicates favourable conditions for improvement while $\Delta e \leq 0$ indicates unfavourable conditions.

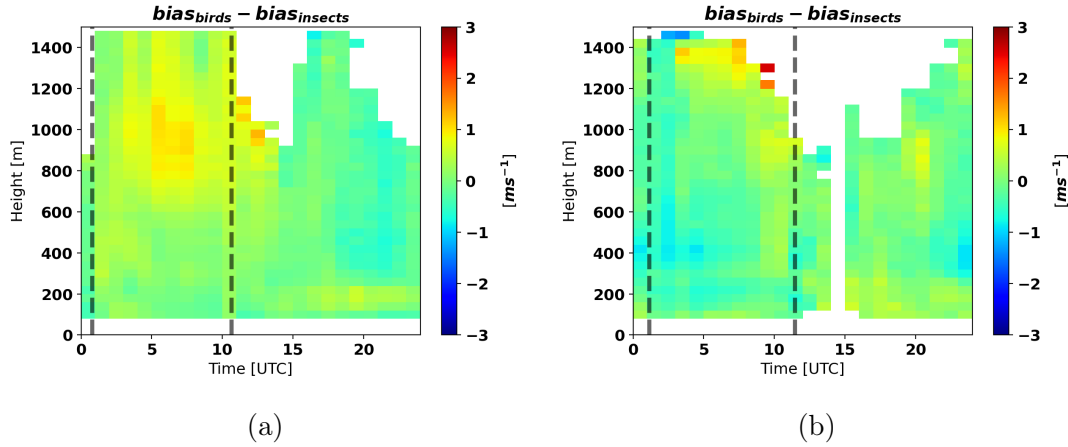


Figure 4.16: Averages of bird-insect bias differences by height and time of day, for the (a) KOHX nocturnal migration season (b) KHGX diurnal migration season. Local time is UTC - 5

For time of day, averages of bias differences are shown in Figure 4.16. It can be observed in Figure 4.16(a) that the nocturnal migration season has the most favourable conditions for wind bias corrections between heights of 600 – 1200 m during nighttime, up till around sunrise (0-12 UTC), coincident with the

apex fringes of bird migration (Figure 4.10(a)). The remaining insect dominated daytime generally has $\Delta e \approx 0$. Averages for the diurnal season are shown in Figure 4.16(b). A noticeable discrepancy is that $\Delta e > 0$ seemingly prevails outside bird contaminated nighttime regions. However, a day-to-day inspection of this region (see Figure A.15), reveals an oscillation between favourable and detrimental conditions. For example, heavily bird dominated nighttime VADs on 15 May 2022 has $\Delta e \approx -2 \text{ ms}^{-1}$, while the same period on 23 May 2022 is moderately bird contaminated and has $\Delta e > 2 \text{ ms}^{-1}$. These daily variations underscore a crucial limitation of understanding complex bird behaviour by time of day.

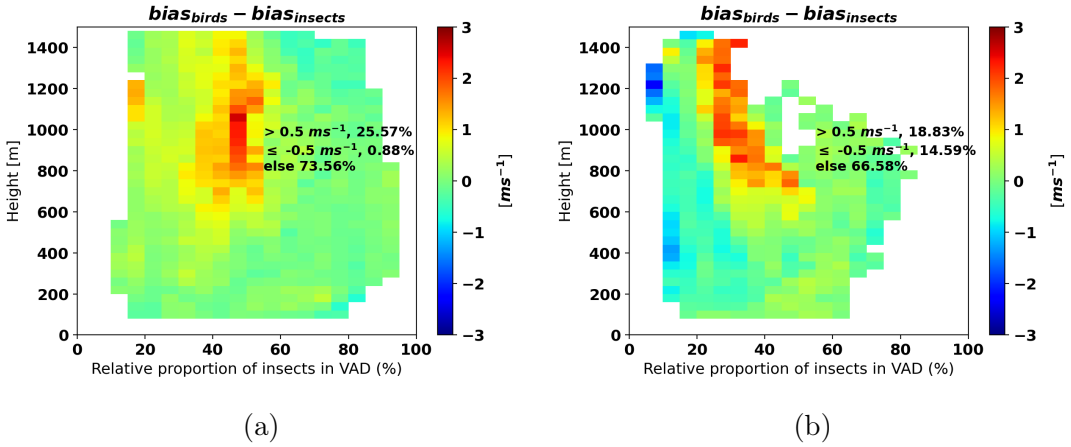


Figure 4.17: Bird-insect bias difference averaged by height and insect-to-bird ratio for the (a) KOHX nocturnal migration season (b) KHGX diurnal migration season.

Wind improvement occurs when a ring contains enough birds for corruption, but enough insects for VAD wind estimation. The birds-insects ratio R_i captures this taxonomic distribution, thus a distinct pattern emerges for Δe analysis based on this metric. The results of this analysis are shown in Figure 4.17. Specifically, nocturnal and diurnal seasons are shown in Figures 4.17(a) and 4.17(b) respectively. For the former season, the most favourable conditions occur in VADs where $30\% < R_i < 60\%$ and $600 < \text{height} < 1400 \text{ m}$. The diurnal season similarly has

the best conditions for $20\% < R_i < 55\%$ within the same height interval. The high Δe region is slightly shifted toward the left due to higher bird contamination, especially at higher altitudes. Moreover, varying bird levels are revealed to cause the aforementioned day-to-day discrepancies during nighttime. Considering the 1200 m height for example, conditions are detrimental ($\Delta e \approx -2 \text{ ms}^{-1}$) for heavily contaminated rings with $R_i \approx 10\%$ and ideal ($\Delta e \approx 2 \text{ ms}^{-1}$) for moderately contaminated rings with $R_i \approx 30\%$.

Overall, the intermediate region of best improvability across both seasons, i.e $20\% < R_i < 60\%$, is consistent with the requirement of simultaneous bird and insect presence. Many insects (few birds) make improvements unnecessary. Too few insects (many birds) make improvements unfeasible.

4.6.3 VAD Improvement

The operational VAD algorithm does not distinguish biological echoes by taxa. Moreover, VAD on insects have been shown to reduce biases in the previous section. Thus, a cost-benefit analysis is conducted to understand the impacts of the proposed approach relative to the existing one i.e focusing VAD on insects instead of all biological echoes. Two metrics are considered. The first is improvement I defined as the reduction in wind biases,

$$I(h, t) = e_{bio}(h, t) - e_{insects}(h, t) \quad (4.7)$$

where $e_{bio}(h, t)$ is the bias from biological VADs. Using the insect subset of echoes would lead to fewer samples and potentially fewer VADs. Hence, the fraction of VADs lost is also defined as,

$$F(h, t) = \frac{C_{bio}(h, t) - C_{insects}(h, t)}{C_{bio}(h, t)} \quad (4.8)$$

where $C_{bio}(h, t)$ is the number of VADs generated using biological echoes and $C_{insects}(h, t)$, the number generated using insects.

4.6.4 Analysis of improvements

Time of day has the limitation of blurring complex daily variations. As such, discussions are primarily focused on analysis by the insects-birds ratio, shown in Figure 4.18. Averages by time of day are still provided in Figure 4.19, as a reference for comparison. Improvements are correlated to bird-insect bias differences, but would have lower magnitudes since differences here are between biological (insects included) and pure insect echoes. Thus, the largest improvements are within 600 – 1400 m heights for both seasons, and $30\% < R_i < 60\%$ for the nocturnal season (Figure 4.18(a)) and $20\% < R_i < 55\%$ for the diurnal season 4.18(b)). The availability of improved VADs depends on the presence of insect echoes. For example, heavily bird dominated rings with $R_i \sim 10\%$ stand out as having significant losses $\sim 80 - 100\%$ during the diurnal season (Figure 4.18(d)). Such bird domination levels do not occur for the nocturnal season (Figure 4.18(c)). Across both seasons, VADs are generally lost for heights > 800 m which are largely absent of insects.

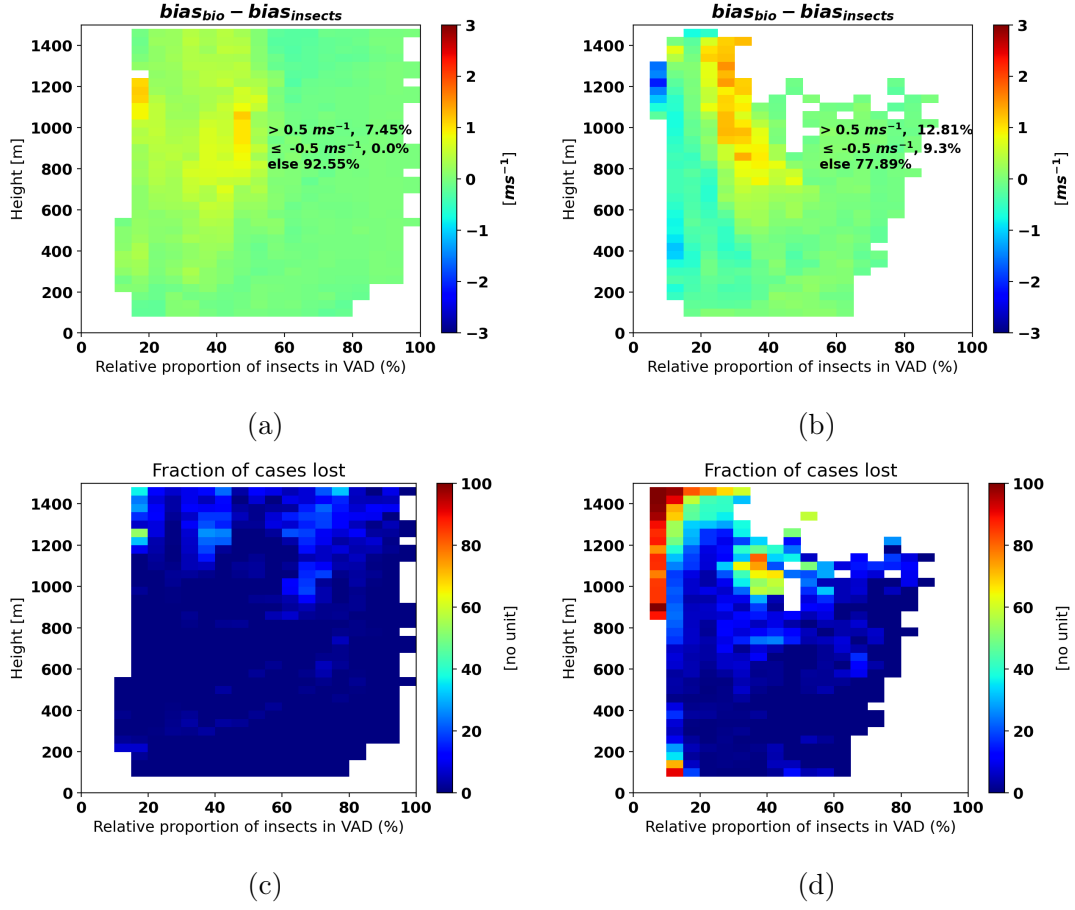


Figure 4.18: Height by insects-birds ratio averages of improvement and the fraction of VADs lost. From left to right, columns correspond to nocturnal and diurnal migration seasons. From top to bottom, rows correspond to improvement and the fraction of VADs lost.

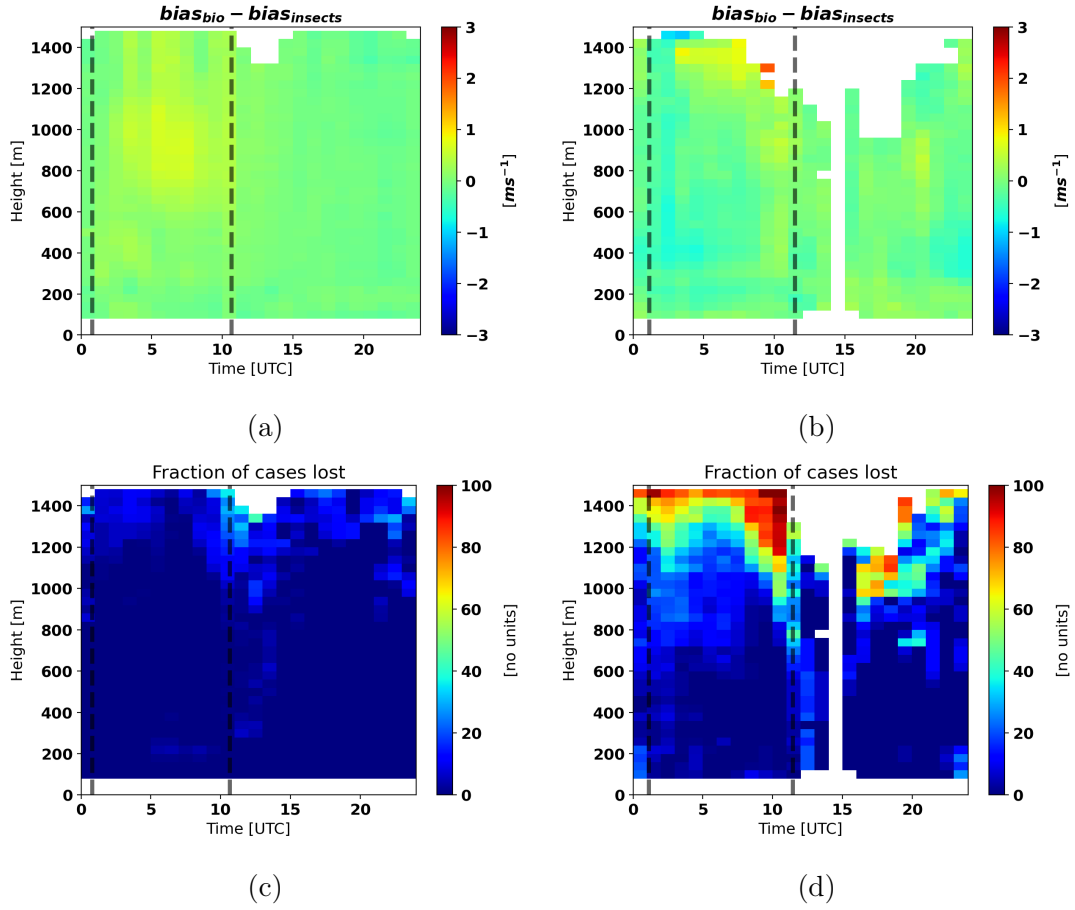


Figure 4.19: Height by time of day averages of improvement and the fraction of VADs lost. From left to right, columns correspond to nocturnal and diurnal migration seasons. From top to bottom, rows correspond to improvement and the fraction of VADs lost.

Filtering Insect Predictions by Confidence

Assuming a case of insects perfectly tracing a uniform wind field, VAD quality would depend on coverage, number of samples and the accuracy of insect detections. BIRC has a TPR of $0.832 - 0.892\%$ meaning that $0.108\% - 0.168\%$ of birds-as-insects misclassifications can be expected. In theory, perfect insect classification should lead to maximum improvements. Recall that BIRC identifies biological gates as insects based on a score p , where $p = 0$ indicates the highest confidence of insect presence and $p = 1$ indicates lowest confidence. This section investigates improvements for varying thresholds on p . BIRC is modified such that

$$C(h, \phi_i, t) = \begin{cases} \text{None}, & p_i > \tau \\ 0, & p_i \leq \tau. \end{cases} \quad (4.9)$$

where τ is the confidence threshold, $\{C = 0\}$ are insects and $\{C = \text{None}\}$ gates are discarded.

Improvements and the fraction of lost VADs, were analyzed for the modified version of BIRC for both seasons. The results for the nocturnal season using $\tau = 0.4, 0.3, 0.2 \& 0.1$ is shown in Figure 4.20. Within the $30\% < R_i < 60\%$ and $600 < \text{height} < 1400$ m region, improvements can be observed to increase as τ decreases from 0.4 to 0.1, with medians ranging from 0.46 ms^{-1} to 0.84 ms^{-1} (Figures 4.21(a) - 4.21(d)). Availability, on the other hand, decreases with τ . This is especially salient for bird dominated ($R_i \approx 0$) and high altitude rings (height > 800 m) where insects are less common. Results for the diurnal season are also shown in Figure 4.21. Because of the ubiquitous presence of birds, the main impact of filtering is that more cases are lost as τ approaches 0.

Overall, these results affirm the hypothesis of better improvements with more confident predictions, demonstrating what is possible with future refinements of bird-insect classification.

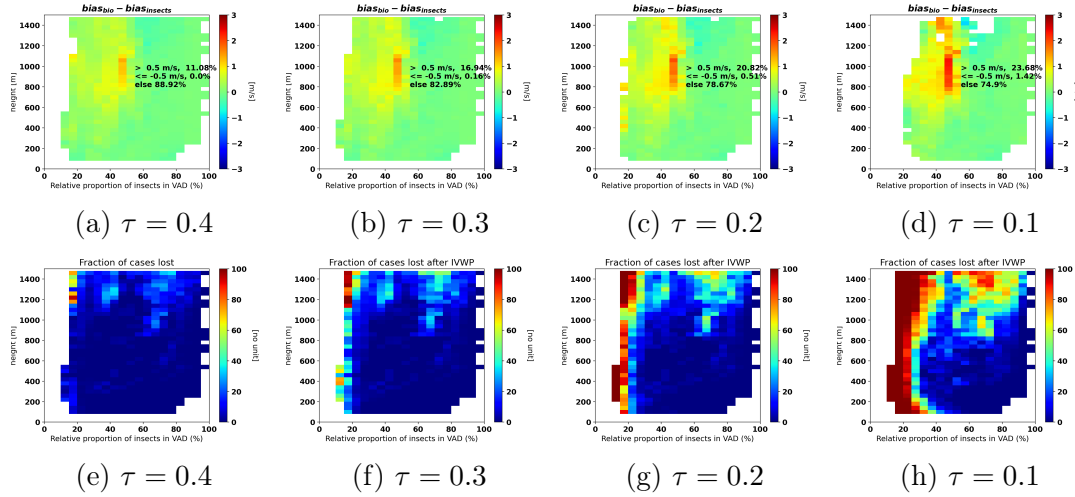


Figure 4.20: Analysis of insect filtering for the KOHX nocturnal migration season using different confidence thresholds τ . From top to bottom, rows correspond to improvements and fraction of VADs lost.

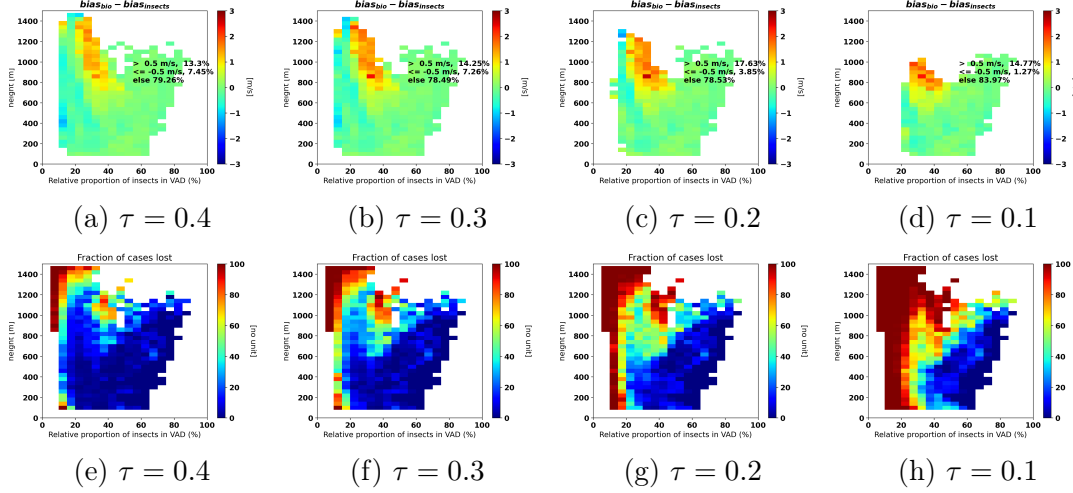


Figure 4.21: Same as Figure 4.20, but for the KHGX diurnal season

4.7 Recommendations and Conclusions

Three new products have been explored for enhancing VADs namely: insects-birds ratio, insect VAD and bird VAD. Insects-birds ratio R_i have been shown to measure bird contamination (or insect tracing) levels by their inverse relationship to wind biases. They can be exapcted as wind tracing scores for clear air VADs: $R_i \sim 0\%$ would indicate poor wind tracing while $R_i \sim 100\%$ indicates reliable wind tracing. Furthermore, thresholds on the insects-birds ratio can be used to demarcate three important regions of bird/insect activity: $R_i < 40\%$ can be bird dominated, $40\% \leq R_i < 60\%$, mixed and $60\% \leq R_i < 100\%$, insect dominated. Usage of the remaining products can be adapted to these regions.

Bird dominated regions have large median biases $\sim 5.5 \text{ ms}^{-1}$ and low improvements $\sim 0 \text{ ms}^{-1}$ indicative of aligned bird migration (shown in Figures 4.11 & 4.18 respectively). As such, bird VADs can be used to retrieve the flock's velocity. VAD on all biological echoes can alternatively estimate bird velocities because of the low impact of the few insects on the model. However, estimates would become

less reliable as R_i increases. Insect dominated regions have low biases $\sim 2.0 \text{ ms}^{-1}$ and low improvements $\sim 0 \text{ ms}^{-1}$ because of good wind tracing. Insect VADs can therefore be used to obtain more accurate wind estimates. Once again, biological VADs can also provide good wind approximation, but become less reliable as R_i decreases. The final mixed region has intermediate biases $\sim 2.85 \text{ ms}^{-1}$ (closer to insects) and maximum improvements, reaching up to 1.5 ms^{-1} . Thus, focusing VAD on insect echoes can reduce wind biases. Several industries would benefit from having these products. Meteorologists can obtain more accurate winds from insect velocities while the aviation industry can leverage bird velocities to understand their migratory activity between airports. Moreover, aero-ecological studies of both taxa would be facilitated by the proposed enhancements.

There are still some areas that can be improved with better echo classification. For example, HCA's misclassification of insects as weather lead to lower insect availability in some cases. To prevent the inheritance of such errors, the taxonomic classifier can be disentangled from the HCA by including new classes for weather and other types of biological echoes. Furthermore, improvements in wind estimation were observed as less confident insect points were filtered prior to the VAD process. Consequently, better bird-insect classification would guarantee more reliable winds.

Chapter 5

Conclusions and Future Work

This dissertation is concluded with a summary of findings and key contributions. Finally, limitations are presented as well as future avenues for improvement.

5.1 Summary of Contributions

5.1.1 Classification of Bird and Insect Echoes

In this study, a Machine Learning model was developed for classifying biological echoes on NEXRAD as either birds or insects. This was done in three stages. First, data was collected from range gates containing birds or insects. Secondly, ridge classifiers and decision tree ML algorithms were trained on different input combinations to generate a variety of models. Finally, a best model is selected by evaluating trade-offs between performance and complexity. Performance was analyzed based on ACC, TPR, TNR and AUC, as well as case studies involving bird flocks, insect swarms and bird roosts. Complexity depends on the number of required inputs. This best model is a ridge classifier using only dual polarization variables. Thus, it is assigned the moniker BIRC.

This study has made a number of contributions to biological radar echo classification. Previous works typically focused on identifying particular patterns of

a particular taxa. For instance, models have been developed for detecting bird roosts, but would not be applicable to singular range gates, non-roosting birds or insects. Other works are applicable to range gates, but not designed for production level use. BIRC can be used on any WSR-88D, to identify both birds and insects, using data from a single range gate, with applicability to diverse scenarios including mass bird migration, insect swarms and mixtures of both echoes. Its efficacy has been demonstrated on WSR-88Ds located in Oklahoma, Alabama, and following subsequent VAD integration, Tennessee and Texas. Another contribution of this study involves data collection. A significant limitation to the development of a biological classifier for NEXRAD, is the difficulty of determining ground truth at the gate level. In order to alleviate this problem, a data preparation method is proposed for extracting bird echoes from migrating bird flocks, and insect echoes from insect swarms.

There are several potential applications of this study’s outcomes. The HCA can be modified to bifurcate biological echoes into birds and insects, and this taxonomic information, incorporated into NEXRAD’s VAD wind profiles. Subsequently, improved wind estimates can be obtained from insects, while migration velocities can be obtained from birds. The latter might be used in bird strike mitigation efforts. Aeroecological studies can also benefit from the broad scale surveillance of biota on NEXRAD.

5.1.2 Application of BIRC for VAD Wind Estimation

In this study, BIRC was applied for VAD wind estimation. At each height, HCA detects biological echoes, which are then classified by BIRC as birds or insects. These predictions are used in the VAD process. Previous studies of VADs from

biological echoes, typically identified broad regions of birds and insects by analyzing radar variables. Here, one VAD ring might be attributed to birds or insects, but not both. This work enables the detection of simultaneously bird/insect activity by making predictions at the gate level. As a result, three new products are obtained from one VAD, including the ratio of bird to insect gates, and two separate VADs for birds and insects.

These products are the main contributions of this work. The insects-birds ratio provides a means of measuring bird contamination (or wind tracing levels). For instance, VADs can be roughly categorized as bird dominated, mixed and insect dominated, with large, medium and low wind biases respectively. Usage of the remaining products can be determined by the aforementioned categories. Bird VADs can be obtained from bird dominated rings, while insect VADs are derived from mixed and insect dominated rings. Another noteworthy contribution is the improvement of wind estimation, especially in the mixed region. Wind bias reductions were observed when VADs are focused on insects instead of all biological echoes. At the time of writing, no similar work on producing bird and insect VADs is known.

The proposed products can be incorporated into NEXRAD's VWP. Insect VADs can serve as a proxy for the wind's velocity, while bird VADs can provide their migration velocities. As stated in the previous section, they can also be used for mitigating bird strikes and aeroecological studies.

5.2 Future Work

5.2.1 Improvement of Bird/Insect Classification

BIRC has shown promising results for classifying birds and insects on NEXRAD. However, there are still areas for future improvements. The main challenge remains lack of actual ground truth verification at each gate. Biological echoes can span several hundreds of kilometers across several heights, as such, manual gate inspection becomes unfeasible. A potential solution could involve the corroboration of radar observations using other sensors. For example, recent work on measuring bird migration explored a multi-sensor array composed of two WSR-88Ds, moon sensors for detecting bird flight across the moon, and Autonomous Recording Units (ARU) for measuring nocturnal flight calls [105].

BIRC can also be modified to standalone. In this work, BIRC operates on HCA’s biological class and can therefore inherit errors from the latter. For instance, HCA was found to misclassify some biological echoes as weather (see Section 4.2.1), effectively eliminating them from BIRC’s consideration. Such error propagation can be mitigated by disentangling future biological classifiers. This would require the addition of new classes for weather and other echoes. Another area of improvement involves more refined taxonomic classification. This work has broadly explored birds and insects, however, different species can have different scattering properties. Thus, classification by species can enable better understanding of biological activity. Other echoes like bats can also be incorporated into the model.

5.2.2 Validation of VAD Products

While evaluations of the proposed VAD products show promising results, ground truth verification remains challenging. RAP predictions were assumed to represent the underlying wind, because of their availability every hour compared to every 12 hours for balloon soundings. However, the latter provides more direct wind measurements. As such, evaluations based on soundings would provide better support for the proposed modifications. The low temporal availability of sounding data would need to be addressed in some way. For instance, special experiments can be conducted where balloons are launched every 4 hours instead.

Certain conditions were ensured for this study. For example, analyses were limited to clear-air conditions where biological echoes are expected to dominate. Conversely, future research can explore VAD wind biases across different meteorological scenarios. Furthermore, this study focused on 1-month periods at 2 WSR-88Ds. Future analysis can be performed for longer periods and extended to new geographical locations.

References

- [1] F. Bairlein, “Migratory birds under threat,” *Science*, vol. 354, no. 6312, pp. 547–548, 2016. DOI: 10.1126/science.aah6647.
- [2] K. V. Rosenberg *et al.*, “Decline of the North American avifauna,” *Science*, vol. 366, no. 6461, pp. 120–124, 2019. DOI: 10.1126/science.aaw1313.
- [3] F. J. Sanderson, P. F. Donald, D. J. Pain, I. J. Burfield, and F. P. van Bommel, “Long-term population declines in afro-palearctic migrant birds,” English, *Biological Conservation*, vol. 131, no. 1, pp. 93–105, 2006. DOI: 10.1016/j.biocon.2006.02.008.
- [4] F. Sánchez-Bayo and K. A. Wyckhuys, “Worldwide decline of the entomofauna: A review of its drivers,” *Biological Conservation*, vol. 232, pp. 8–27, 2019. DOI: <https://doi.org/10.1016/j.biocon.2019.01.020>.
- [5] G. Perez *et al.*, “Using spatio-temporal information in weather radar data to detect and track communal bird roosts,” *bioRxiv*, 2022. DOI: 10.1101/2022.10.28.513761.
- [6] S. A. Gauthreaux and C. G. Belser, “Displays of bird movements on the WSR-88D: Patterns and quantification,” *Weather and Forecasting*, vol. 13,

- p no. 2, pp. 453–464, 1998. DOI: 10.1175 / 1520 - 0434(1998) 013<0453 : DOBMOT>2.0.CO;2.
- [7] P. M. Stepanian, K. G. Horton, V. M. Melnikov, D. S. Zrnić, and S. A. Gauthreaux Jr., “Dual-polarization radar products for biological applications,” *Ecosphere*, vol. 7, no. 11, e01539, 2016. DOI: 10.1002/ecs2.1539.
 - [8] P. M. Stepanian and K. G. Horton, “Extracting migrant flight orientation profiles using polarimetric radar,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 12, pp. 6518–6528, Dec. 2015. DOI: 10.1109/TGRS.2015.2443131.
 - [9] J. F. Kelly, J. R. Shipley, P. B. Chilson, K. W. Howard, W. F. Frick, and T. H. Kunz, “Quantifying animal phenology in the aerosphere at a continental scale using NEXRAD weather radars,” *Ecosphere*, vol. 3, no. 2, art16, 2012. DOI: <https://doi.org/10.1890/ES11-00257.1>.
 - [10] D. N. Bonter, J. Sidney A. Gauthreaux, and T. M. Donovan, “Characteristics of important stopover locations for migrating birds: Remote sensing with radar in the great lakes basin,” *Conservation Biology*, vol. 23, 2009.
 - [11] J. W. Horn and 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*, vol. 48, no. 1, pp. 24–39, Jun. 2008. DOI: 10.1093/icb/icn051.
 - [12] J. Gauthreaux Sidney A., J. W. Livingston, and C. G. Belser, “Detection and discrimination of fauna in the aerosphere using Doppler weather

- surveillance radar,” *Integrative and Comparative Biology*, vol. 48, no. 1, pp. 12–23, May 2008. DOI: 10.1093/icb/icn021.
- [13] S. Ansari *et al.*, “Unlocking the potential of NEXRAD data through NOAA’s big data partnership,” *Bulletin of the American Meteorological Society*, vol. 99, no. 1, pp. 189–204, 2018.
- [14] S. Bachmann and D. Zrnić, “Spectral density of polarimetric variables separating biological scatterers in the VAD display,” *Journal of Atmospheric and Oceanic Technology*, vol. 24, no. 7, pp. 1186–1198, 2007. DOI: 10.1175/JTECH2043.1.
- [15] W. J. Martin and A. Shapiro, “Discrimination of bird and insect radar echoes in clear air using high-resolution radars,” *Journal of Atmospheric and Oceanic Technology*, vol. 24, no. 7, pp. 1215–1230, 2007. DOI: 10.1175/JTECH2038.1.
- [16] J. N. Chrisman and S. D. Smith, “Enhanced Velocity Azimuth Display Wind Profile (EVWP) P 4.7 function for the WSR-88D,” 2009.
- [17] A. M. Dokter, F. Liechti, H. Stark, L. Delobbe, P. Tabary, and I. Holleman, “Bird migration flight altitudes studied by a network of operational weather radars,” *Journal of The Royal Society Interface*, vol. 8, no. 54, pp. 30–43, 2011. DOI: 10.1098/rsif.2010.0116.
- [18] C. Nilsson *et al.*, “Bird strikes at commercial airports explained by citizen science and weather radar data,” *Journal of Applied Ecology*, vol. 58, no. 10, pp. 2029–2039, 2021. DOI: <https://doi.org/10.1111/1365-2664.13971>.

- [19] J. R. Allan, “The costs of bird strikes and bird strike prevention, human conflicts with wildlife: Economic considerations,” in *Proceedings of the Third National Wildlife Research Center Special Symposium*, 2000.
- [20] R. A. Dolbeer, “Height distribution of birds recorded by collisions with civil aircraft,” *The Journal of Wildlife Management*, vol. 70, no. 5, pp. 1345–1350, 2006.
- [21] P. P. Marra *et al.*, “Migratory Canada geese cause crash of US Airways Flight 1549,” *Frontiers in Ecology and the Environment*, vol. 7, no. 6, pp. 297–301, 2009.
- [22] P. Jatau and V. Melnikov, “Classifying bird and insect radar echoes at S-band,” in *35th Conference on Environmental Information Processing Technologies*, Phoenix, United States, Jan. 2019.
- [23] CNN. “Eyewitness views of the Hudson landing.” (2016), [Online]. Available: <https://www.cnn.com/2014/01/15/tech/gallery/hudson-plane-landing-2009-ireport/index.html> (visited on 10/30/2023).
- [24] Z.-W. Ye, S. Yuan, K.-S. Yuen, S.-Y. Fung, C.-P. Chan, and D.-Y. Jin, “Zoonotic origins of human coronaviruses,” *International journal of biological sciences*, vol. 16, no. 10, p. 1686, 2020.
- [25] V. A. Drake and R. A. Farrow, “The influence of atmospheric structure and motions on insect migration,” *Annual Review of Entomology*, vol. 33, no. 1, pp. 183–210, 1988. DOI: 10.1146/annurev.en.33.010188.001151.

- [26] D. E. Pedgley, *Windborne pests and diseases. Meteorology of airborne organisms*. Chichester, UK: Ellis Horwood, 1982.
- [27] A. T. Peterson and R. A. J. Williams, “Risk mapping of highly pathogenic avian influenza distribution and spread,” *Ecology and Society* 13, 2008.
- [28] D. S. Zrnic and A. V. Ryzhkov, “Polarimetry for weather surveillance radars,” *Bulletin of the American Meteorological Society*, vol. 80, no. 3, pp. 389–406, 1999. DOI: 10.1175/1520-0477(1999)080<0389:PFWSR>2.0.CO;2.
- [29] H. S. Park, A. V. Ryzhkov, D. S. Zrnić, and K.-E. Kim, “The hydrometeor classification algorithm for the polarimetric WSR-88D: Description and application to an MCS,” *Weather and Forecasting*, vol. 24, no. 3, pp. 730–748, 2009. DOI: 10.1175/2008WAF2222205.1.
- [30] V. Chandrasekar, R. Keränen, S. Lim, and D. Moisseev, “Recent advances in classification of observations from dual polarization weather radars,” *Atmospheric Research*, vol. 119, pp. 97–111, 2013, ADVANCES IN PRECIPITATION SCIENCE. DOI: <https://doi.org/10.1016/j.atmosres.2011.08.014>.
- [31] S. Gauthreaux and R. Diehl, “Discrimination of biological scatterers in polarimetric weather radar data: Opportunities and challenges,” *Remote Sensing*, vol. 12, no. 3, p. 545, Feb. 2020. DOI: 10.3390/rs12030545.
- [32] B. Radhakrishna, F. Fabry, and A. Kilambi, “Fuzzy logic algorithms to identify birds, precipitation, and ground clutter in S-band radar data using polarimetric and nonpolarimetric variables,” *Journal of Atmospheric and*

- Oceanic Technology*, vol. 36, no. 12, pp. 2401–2414, 2019. DOI: 10.1175/JTECH-D-19-0088.1.
- [33] A. Kilambi, F. Fabry, and V. Meunier, “A simple and effective method for separating meteorological from nonmeteorological targets using dual-polarization data,” *Journal of Atmospheric and Oceanic Technology*, vol. 35, no. 7, pp. 1415–1424, 2018. DOI: 10.1175/JTECH-D-17-0175.1.
- [34] V. A. Drake and D. R. Reynolds, *Radar entomology : Observing insect flight and migration*, English. CABI Wallingford, Oxfordshire ; Boston, MA, 2012, xviii, 489 p., [8] p. of col. plates :
- [35] K. R. Hardy and I. Katz, “Probing the clear atmosphere with high power, high resolution radars,” *Proceedings of the IEEE*, vol. 57, no. 4, pp. 468–480, 1969.
- [36] T. J. Lang, S. A. Rutledge, and J. L. Stith, “Observations of quasi-symmetric echo patterns in clear air with the CSU–CHILL polarimetric radar,” *Journal of Atmospheric and Oceanic Technology*, vol. 21, no. 8, pp. 1182–1189, 2004. DOI: 10.1175/1520-0426(2004)021<1182:OOQEPI>2.0.CO;2.
- [37] R. F. Contreras and S. J. Frasier, “High-resolution observations of insects in the atmospheric boundary layer,” *Journal of Atmospheric and Oceanic Technology*, vol. 25, no. 12, pp. 2176–2187, 2008. DOI: 10.1175/2008JTECHA1059.1.

- [38] V. M. Melnikov, R. R. Lee, and N. J. Langlieb, “Resonance effects within S-band in echoes from birds,” *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 3, pp. 413–416, 2012.
- [39] P. Zhang, S. Liu, and Q. Xu, “Identifying Doppler Velocity Contamination Caused by Migrating Birds. Part I: Feature Extraction and Quantification,” *Journal of Atmospheric and Oceanic Technology*, vol. 22, no. 8, pp. 1105–1113, Aug. 2005. DOI: 10.1175/JTECH1757.1.
- [40] S. Liu, Q. Xu, and P. Zhang, “Identifying Doppler Velocity Contamination Caused by Migrating Birds. Part II: Bayes Identification and Probability Tests,” *Journal of Atmospheric and Oceanic Technology*, vol. 22, no. 8, pp. 1114–1121, Aug. 2005. DOI: 10.1175/JTECH1758.1.
- [41] P. Jatau, V. Melnikov, and T.-Y. Yu, “A machine learning approach for classifying bird and insect radar echoes with S-band polarimetric weather radar,” *Journal of Atmospheric and Oceanic Technology*, vol. 38, no. 10, pp. 1797–1812, 2021. DOI: <https://doi.org/10.1175/JTECH-D-20-0180.1>.
- [42] A. McGovern *et al.*, “Making the black box more transparent: Understanding the physical implications of machine learning,” *Bulletin of the American Meteorological Society*, vol. 100, no. 11, pp. 2175–2199, 2019. DOI: <https://doi.org/10.1175/BAMS-D-18-0195.1>.
- [43] Y.-S. Kim, D. Schwartzman, R. D. Palmer, and T.-Y. Yu, “Fast adaptive beamforming using deep learning for digital phased array radars,” in *2022 IEEE International Symposium on Phased Array Systems and Technology (PAST)*, 2022, pp. 01–07. DOI: 10.1109/PAST49659.2022.9975038.

- [44] M. S. Veillette, J. M. Kurdzo, P. M. Stepanian, J. McDonald, S. Samsi, and J. Y. N. Cho, “A deep learning–based velocity dealiasing algorithm derived from the WSR-88D open radar product generator,” *Artificial Intelligence for the Earth Systems*, vol. 2, no. 3, e220084, 2023. DOI: <https://doi.org/10.1175/AIES-D-22-0084.1>.
- [45] H. Kim and B. Cheong, “Robust velocity dealiasing for weather radar based on convolutional neural networks,” *Remote Sensing*, vol. 15, no. 3, 2023. DOI: [10.3390/rs15030802](https://doi.org/10.3390/rs15030802).
- [46] C. Chilson, K. Avery, A. McGovern, E. Bridge, D. Sheldon, and J. Kelly, “Automated detection of bird roosts using NEXRAD radar data and convolutional neural networks,” *Remote Sensing in Ecology and Conservation*, vol. 5, no. 1, pp. 20–32, 2019. DOI: [10.1002/rse2.92](https://doi.org/10.1002/rse2.92).
- [47] Z. Cheng *et al.*, “Detecting and tracking communal bird roosts in weather radar data,” in *AAAI*, 2020.
- [48] H. Mao *et al.*, “Deep-learning-based flying animals migration prediction with weather radar network,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–13, 2023. DOI: [10.1109/TGRS.2023.3242315](https://doi.org/10.1109/TGRS.2023.3242315).
- [49] J. W. Wilson, T. M. Weckwerth, J. Vivekanandan, R. M. Wakimoto, and R. W. Russell, “Boundary layer clear-air radar echoes: Origin of echoes and accuracy of derived winds,” *Journal of Atmospheric and Oceanic Technology*, vol. 11, no. 5, pp. 1184–1206, 1994. DOI: [https://doi.org/10.1175/1520-0426\(1994\)011<1184:BLCARE>2.0.CO;2](https://doi.org/10.1175/1520-0426(1994)011<1184:BLCARE>2.0.CO;2).

- [50] E. S. Bridge *et al.*, “Technology on the Move: Recent and Forthcoming Innovations for Tracking Migratory Birds,” *BioScience*, vol. 61, no. 9, pp. 689–698, Sep. 2011. DOI: 10.1525/bio.2011.61.9.7.
- [51] M. S. Van Den Broeke, “Polarimetric radar observations of biological scatterers in hurricanes Irene (2011) and Sandy (2012),” *Journal of Atmospheric and Oceanic Technology*, vol. 30, no. 12, pp. 2754–2767, 2013. DOI: 10.1175/JTECH-D-13-00056.1.
- [52] D. S. Zrnic and A. V. Ryzhkov, “Observations of insects and birds with a polarimetric radar,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 2, pp. 661–668, Mar. 1998. DOI: 10.1109/36.662746.
- [53] R. Doviak and D. Zrnić, *Doppler Radar and Weather Observations*. Academic Press, 1993.
- [54] National Weather Service. “About our WSR 88-D Radar.” (), [Online]. Available: https://www.weather.gov/iwx/wsr_88d.
- [55] L. Battan, *Radar Observation of the Atmosphere*. University of Chicago Press, 1973.
- [56] S. Silver, *Microwave Antenna Theory and Design*. New York: McGraw-Hill, 1951.
- [57] M. Skolnik, *Introduction to Radar Systems* (Electrical engineering series). McGraw-Hill, 2001.

- [58] J. R. Probert-Jones, “The radar equation in meteorology,” *Quarterly Journal of the Royal Meteorological Society*, vol. 88, pp. 485–495, 1962.
- [59] Radar Operations Center. “NEXRAD and TDWR Radar Locations.” (2022), [Online]. Available: <https://www.roc.noaa.gov/WSR88D/Maps.aspx> (visited on 09/04/2023).
- [60] U.S. Department of Commerce; National Oceanic and Atmospheric Administration. “Federal Meteorology Handbook No. 11: WSR-88D Meteorological Observations. Part A System Concepts, Responsibilities and Procedures.” (2016), [Online]. Available: <https://www.icams-portal.gov/resources/ofcm/fmh/FMH11/2016FMH11PTA.pdf> (visited on 01/11/2016).
- [61] R. D. Palmer *et al.*, “Observations of the 10 May 2010 tornado outbreak using OU-PRIME: Potential for new science with high-resolution polarimetric radar,” *Bulletin of the American Meteorological Society*, vol. 92, pp. 871–891, 2011.
- [62] NOAA’s National Weather Service, Radar Operations Center. “NEXRAD technical information.” (2022), [Online]. Available: <https://www.roc.noaa.gov/WSR88D/Engineering/NEXRADTechInfo.aspx> (visited on 11/07/2023).
- [63] Radar Operations Center. “NWS WSR-88D Level I Event Data.” (2022), [Online]. Available: https://www.roc.noaa.gov/WSR88D/Level_I/Level1Info.aspx (visited on 11/14/2023).

- [64] Radar Operations Center. “NWS WSR-88D Level I Event Data.” (2022), [Online]. Available: https://www.roc.noaa.gov/WSR88D/Level_III/Level3Info.aspx (visited on 11/14/2023).
- [65] Google Cloud. “NEXRAD L3 real-time and archive data.” (2019), [Online]. Available: <https://console.cloud.google.com/marketplace/product/noaa-public/nexrad-l3>.
- [66] “Radar products available from rpccds (WSR-88D and TDWR products).” (2023), [Online]. Available: https://www.weather.gov/media/tg/rpccds_radar_products.pdf (visited on 11/14/2023).
- [67] M. R. Kumjian, “Principles and applications of dual-polarization weather radar. Part I: Description of the polarimetric radar variables,” *Journal of Operational Meteorology*, vol. 1, pp. 226–242, 2013.
- [68] Warning Decision Training Branch. “Dual-polarization radar principles and systems operations.” (2018), [Online]. Available: <https://training.weather.gov/wdtd/courses/dualpol/documents/DualPolRadarPrinciples.pdf> (visited on 09/04/2023).
- [69] American Meteorological Society. “Glossary of meteorology.” (2012), [Online]. Available: <https://glossary.ametsoc.org/wiki/Welcome> (visited on 11/10/2023).
- [70] W. J. Martin, “Measurements and modeling of the great plains low-level jet,” Available at <https://www.proquest.com/openview/>

e8f71401a2a4575db5fe1b6a7b286a55/1?pq-origsite=gscholar&cbl=18750&diss=y, PhD thesis, University of Oklahoma, Norman, OK, 2003.

- [71] J. Riley, “Collective orientation in night-flying insects,” *Nature*, vol. 253, pp. 113–114, 1975. DOI: <https://doi.org/10.1038/253113a0>.
- [72] R. A. Kropfli, I. Katz, T. G. Konrad, and E. B. Dobson, “Simultaneous Radar Reflectivity Measurements and Refractive Index Spectra in the Clear Atmosphere,” *Radio Science*, vol. 3, no. 10, pp. 991–994, Oct. 1968. DOI: [10.1002/rds1968310991](https://doi.org/10.1002/rds1968310991).
- [73] A. W. Friend, “Continuous determination of air-mass boundaries by radio,” *Bulletin of the American Meteorological Society*, vol. 20, no. 5, pp. 202–205, 1939.
- [74] J. Lane and R. Meadows, “Simultaneous radar and refractometer soundings of the troposphere,” *Nature*, vol. 197, pp. 35–36, 1963. DOI: <https://doi.org/10.1038/197035a0>.
- [75] E. Eastwood, *Radar Ornithology*. London: Methuen and Co., Ltd, 1967.
- [76] K. Jungbluth, J. Belles, and M. Schumacher, “Velocity contamination of WSR-88D and wind profiler data due to migrating birds,” in *27th Conf. on Radar Meteorology*, American Meteorological Society, Vail, CO, 1995.
- [77] T. O’Bannon, “Anomalous WSR-88D wind profiles-migrating birds?” In *27th Conf. on Radar Meteorology*, American Meteorological Society, Vail, CO, 1995.

- [78] K. A. Browning and D. Atlas, “Velocity characteristics of some clear-air dot angels,” *Journal of Atmospheric Sciences*, vol. 23, no. 5, pp. 592–604, 1966. DOI: [https://doi.org/10.1175/1520-0469\(1966\)023<0592:VCOSCA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1966)023<0592:VCOSCA>2.0.CO;2).
- [79] G. Schaefer, “Radar observations of insect flight,” in *Symposia of the Royal entomological Society of London*, 1976.
- [80] F. C. Elder, “Some persistent “ring” angels on high-powered radar,” in *Proceedings of the Sixth Weather Radar Conference*, Cambridge, Massachusetts, Mar. 1957.
- [81] B. Geerts and Q. Miao, “Airborne radar observations of the flight behavior of small insects in the atmospheric convective boundary layer,” 2005.
- [82] D. W. van de Kamp, F. M. Ralph, M. F. Barth, P. A. Miller, J. R. Smart, and L. A. Benjamin, “The new bird contamination quality control check applied to hourly winds from NOAA’s profiler network,” in *Preprints, 28th Conference on Radar Meteorology*, American Meteorological Society, Austin, Texas, 1997.
- [83] P. Miller, M. Barth, J. Smart, and L. Benjamin, “The extent of bird contamination in the hourly winds measured by the NOAA profiler network: Results before and after implementation of the new bird contamination quality control check,” in *Preprints, 1st Symposium in Integrated Observing Systems*, American Meteorological Society, Long Beach, CA, 1997.

- [84] J. M. Wilczak *et al.*, “Contamination of wind profiler data by migrating birds: Characteristics of corrupted data and potential solutions,” *Journal of Atmospheric and Oceanic Technology*, vol. 12, no. 3, pp. 449–467, 1995. DOI: [https://doi.org/10.1175/1520-0426\(1995\)012<0449:COWPDB>2.0.CO;2](https://doi.org/10.1175/1520-0426(1995)012<0449:COWPDB>2.0.CO;2).
- [85] S. A. Gauthreaux, D. S. Mizrahi, and C. G. Belser, “Bird migration and bias of WSR-88D wind estimates,” *Weather and Forecasting*, vol. 13, no. 2, pp. 465–481, 1998. DOI: [https://doi.org/10.1175/1520-0434\(1998\)013<0465:BMABOW>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<0465:BMABOW>2.0.CO;2).
- [86] E. Gossard and R. Strauch, *Radar Observation of Clear Air and Clouds* (Developments in atmospheric science). Elsevier, 1983.
- [87] A. B. Crawford, “Radar reflections in the lower atmosphere,” in *Proceedings of the I.R.E.*, 1949, pp. 404–5.
- [88] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [89] L. Bottou, “Large-scale machine learning with stochastic gradient descent,” in *COMPSTAT*, 2010.
- [90] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Monterey, CA: Wadsworth and Brooks, 1984.
- [91] K. A. Browning and R. Wexler, “The determination of kinematic properties of a wind field using Doppler radar,” *Journal of Applied Meteorology*,

- vol. 7, no. 1, pp. 105–113, 1968. DOI: 10.1175/1520-0450(1968)007<0105:TDOKPO>2.0.CO;2.
- [92] R. M. Lhermitte and D. Atlas, “Precipitation motion by pulse Doppler radar,” in *9th Conf. on Radar Meteorology*, American Meteorological Society, 1961.
 - [93] A. C. Bovik, “Chapter 4 - basic binary image processing,” in *The Essential Guide to Image Processing*, A. Bovik, Ed., Boston: Academic Press, 2009, pp. 69–96. DOI: <https://doi.org/10.1016/B978-0-12-374457-9.00004-4>.
 - [94] P. M. Stepanian, P. B. Chilson, and J. F. Kelly, “An introduction to radar image processing in ecology,” *Methods in Ecology and Evolution*, vol. 5, no. 8, pp. 730–738, 2014. DOI: <https://doi.org/10.1111/2041-210X.12214>.
 - [95] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, “Efficient backprop,” in *Neural Networks: Tricks of the Trade: Second Edition*, G. Montavon, G. B. Orr, and K.-R. Müller, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 9–48. DOI: 10.1007/978-3-642-35289-8_3.
 - [96] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006, ROC Analysis in Pattern Recognition. DOI: <https://doi.org/10.1016/j.patrec.2005.10.010>.
 - [97] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, ser. IJCAI’95, Montreal, Quebec, Canada: Morgan Kaufmann Publishers Inc., 1995, pp. 1137–1143.

- [98] B. Efron and R. Tibshirani, “Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy,” *Statist. Sci.*, vol. 1, no. 1, pp. 54–75, 1986, doi:10.1214/ss/1177013815. DOI: doi:10.1214/ss/1177013815.
- [99] W. G. Collins, “The quality control of Velocity Azimuth Display (VAD) winds at the National Centers for Environmental Prediction,” in *11th Symp. on Meteorological Observations and Instrumentation*, American Meteorological Society, Albuquerque, NM, 2001.
- [100] P. Jatau, V. Melnikov, and T.-Y. Yu, “Detecting birds and insects in the atmosphere using machine learning on NEXRAD radar echoes,” *Environmental Sciences Proceedings*, vol. 8, no. 1, 2021. DOI: 10.3390/ecas2021-10352.
- [101] “RAP, The Rapid Refresh.” (2023), [Online]. Available: [https : / / rapidrefresh.noaa.gov/](https://rapidrefresh.noaa.gov/) (visited on 09/29/2023).
- [102] “NEXRAD Data Archive, Inventory and Access.” (2023), [Online]. Available: <https://www.ncdc.noaa.gov/nexradinv/> (visited on 10/04/2023).
- [103] University of Wyoming. “University of Wyoming atmospheric soundings.” (2023), [Online]. Available: <https://weather.uwyo.edu/upperair/sounding.html> (visited on 09/29/2023).
- [104] S. G. Benjamin *et al.*, “A North American hourly assimilation and model forecast cycle: The rapid refresh,” *Monthly Weather Review*, vol. 144, no. 4, pp. 1669–1694, 2016. DOI: <https://doi.org/10.1175/MWR-D-15-0242.1>.

- [105] A. I. Strand, E. S. Bridge, J. F. Kelly, P. M. Stepanian, D. J. Bodine, and J. R. Soto, “A multi-sensor array for detecting and analyzing nocturnal avian migration,” *PeerJ*, vol. 11, e15622, Aug. 2023. DOI: 10.7717/peerj.15622.

Acronyms

ACC accuracy.

ACF Auto Correlation Function.

AI Artificial Intelligence.

ARU Autonomous Recording Units.

AUC Area Under Curve.

BIRC Bird-Insect Ridge Classifier.

CBL Convective Boundary Layer.

ETL Environmental Technology Lab.

FPR False Positive Rate.

HCA Hydrometeor Classification Algorithm.

ML Machine Learning.

NCEI National Centers for Environmental Information.

NCEP National Centers for Environmental Prediction.

NEXRAD Next Generation Radar.

NOAA National Oceanic and Atmospheric Administration.

NWS National Weather Service.

PPI Plan Position Indicator.

PRF Pulse Repetition Frequency.

PRT Pulse Repetition Time.

RAP Rapid Refresh.

RCS Radar Cross Section.

ROC Radar Operations Center.

RPG Radar Product Generator.

RSP Receiver/Signal Processor.

SNR Signal to Noise Ratio.

TNR True Negative Rate.

TPR True Positive Rate.

VAD Velocity Azimuth Display.

VCP Volume Coverage Patterns.

VWP Velocity Azimuth Display (VAD) Wind Profile.

WSR-88D Weather Surveillance Radar - 1988 Doppler.

Appendix A

A Supplement to: Application of the Bird-Insect Ridge Classifier for VAD Wind Estimation

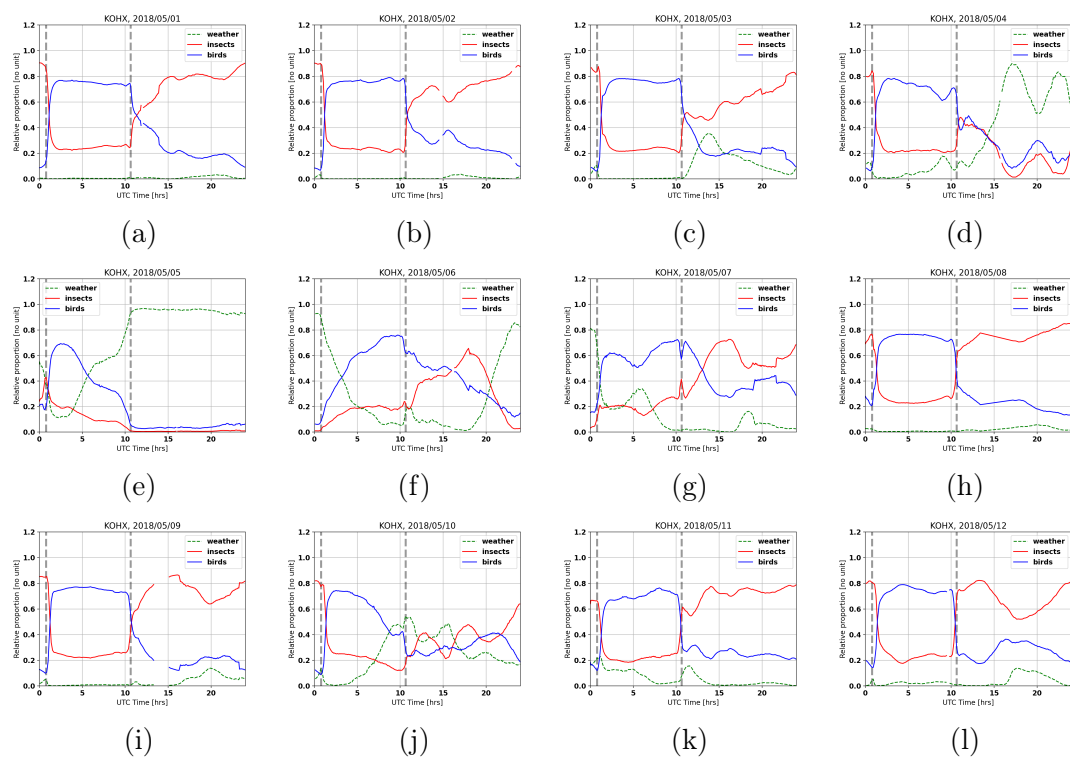


Figure A.1: Relative bird, insect and weather proportions for KOHX, May 1 - 12, 2018

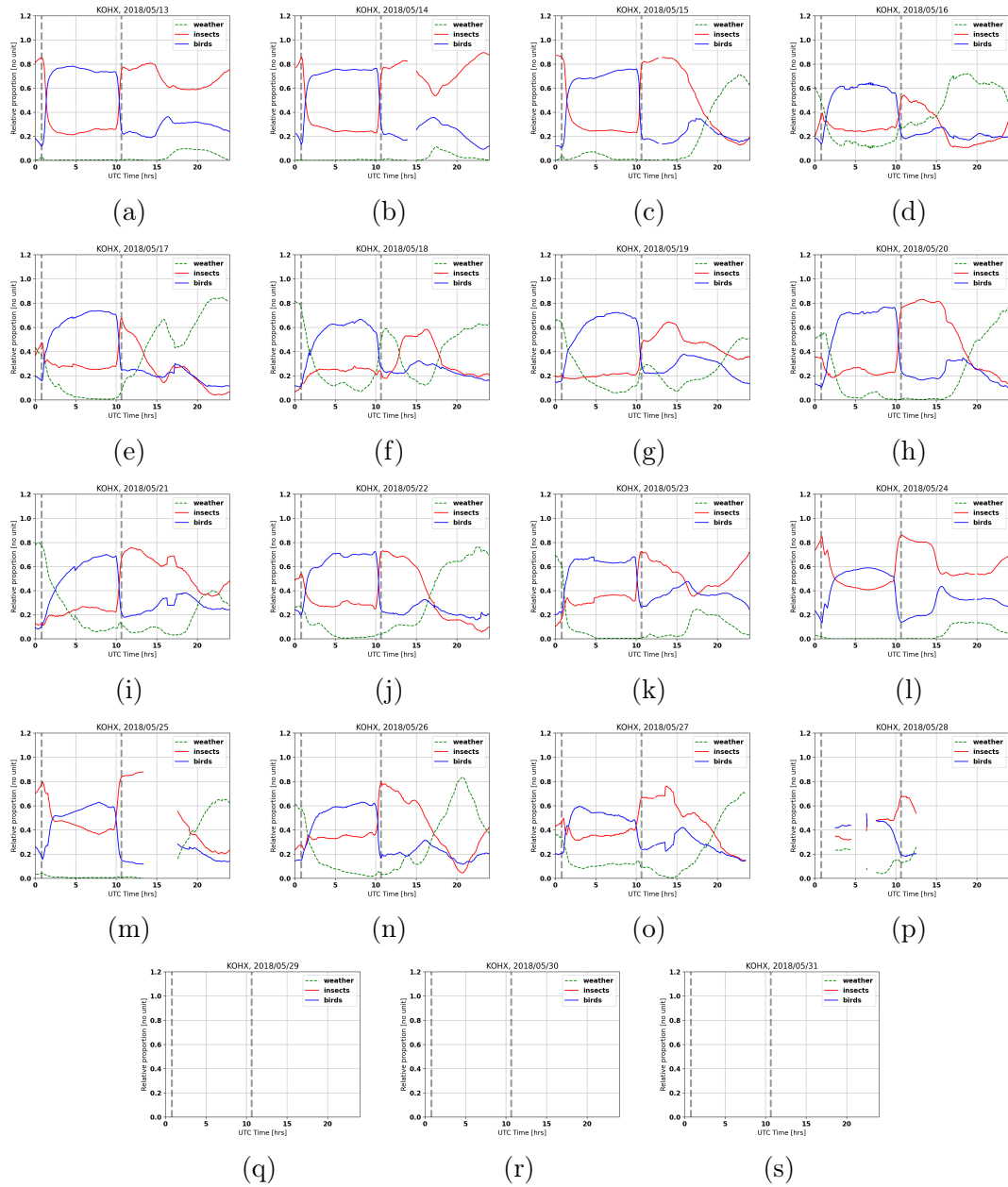


Figure A.2: Same as Figure A.1, but for May 13 - 31, 2018.

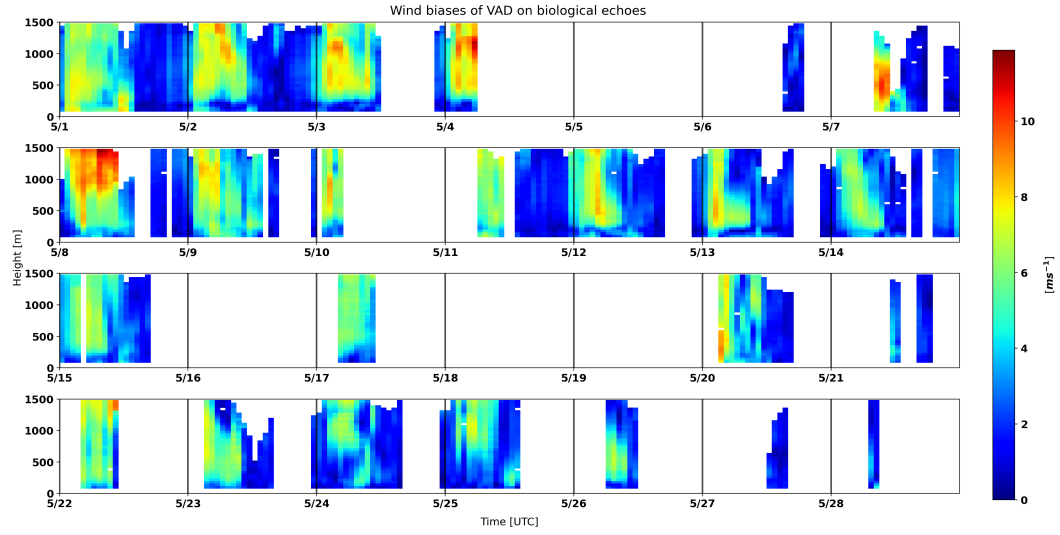


Figure A.3: Biological VAD-wind biases for each day during the KOHX (Nashville, TN) May 2018 nocturnal migration season. The black vertical line denotes 0 UTC.

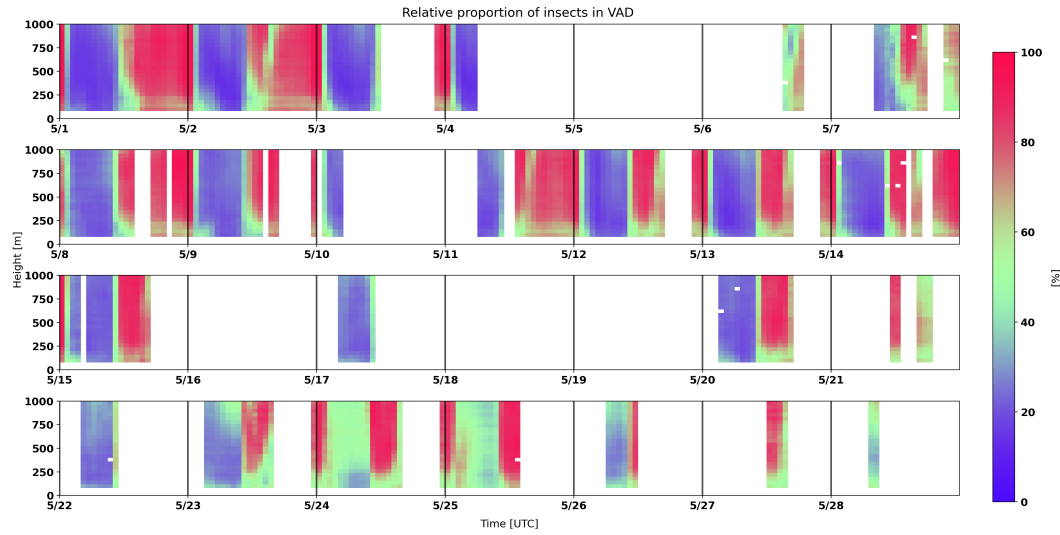


Figure A.4: Same as Figure A.3 but for insects-birds ratios

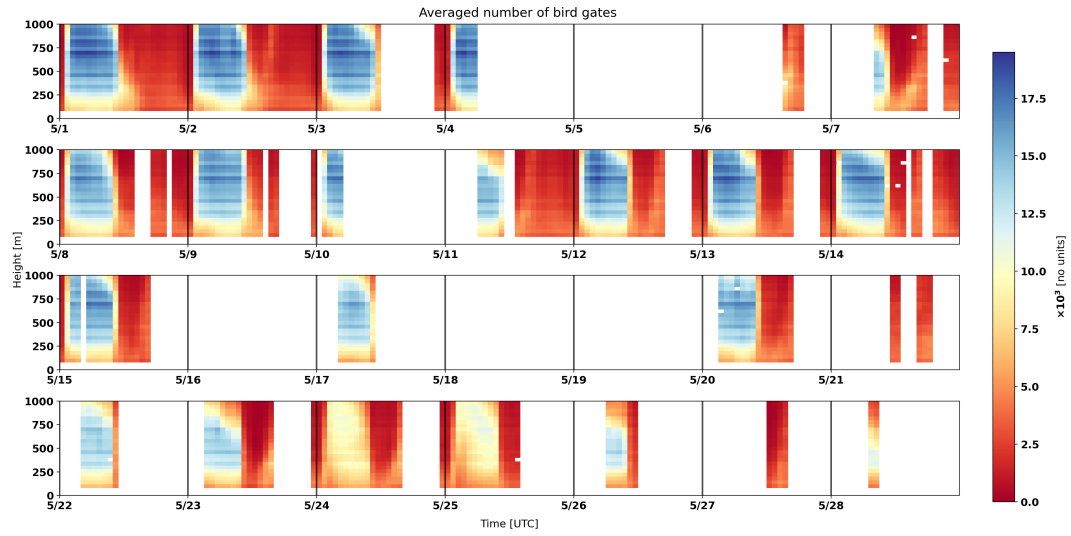


Figure A.5: Same as Figure A.3 but for total bird gates

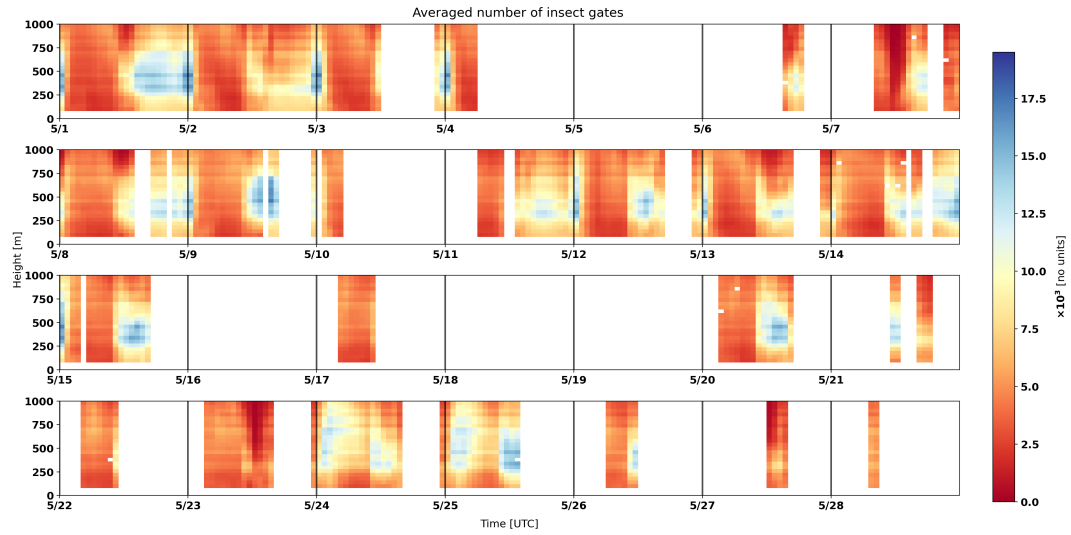


Figure A.6: Same as Figure A.3 but for total insect gates

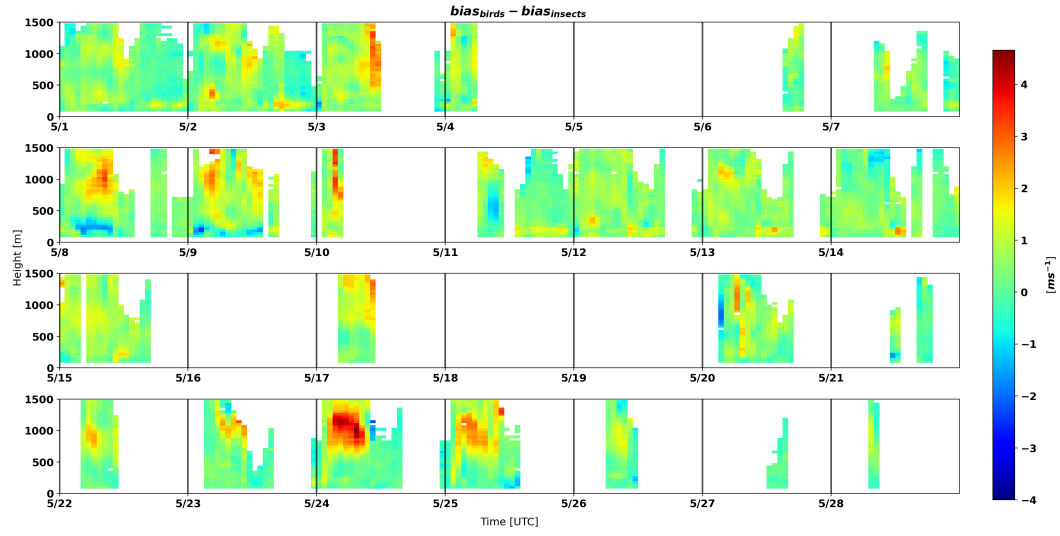


Figure A.7: Same as Figure A.3 but for bird-insect bias differences

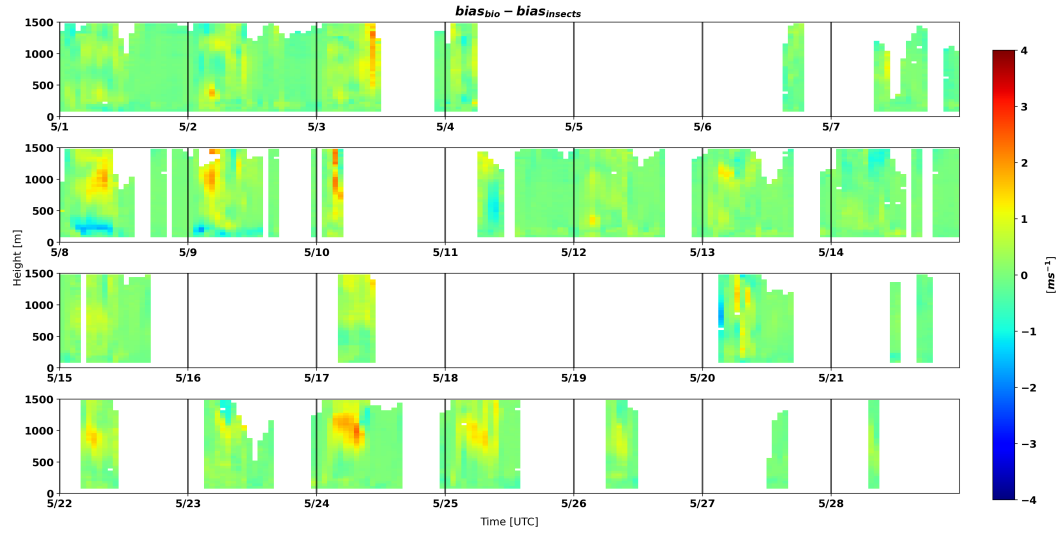


Figure A.8: Same as Figure A.3 but for improvements

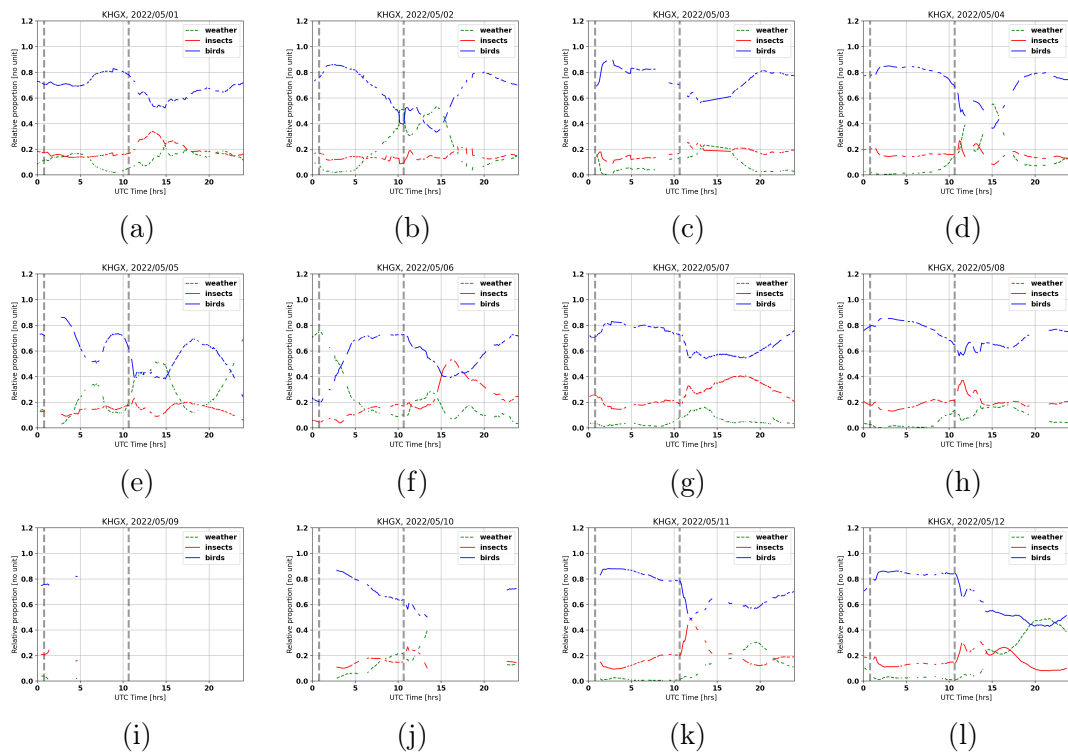


Figure A.9: Relative bird, insect and weather proportions for KHXG, May 1 - 12, 2022

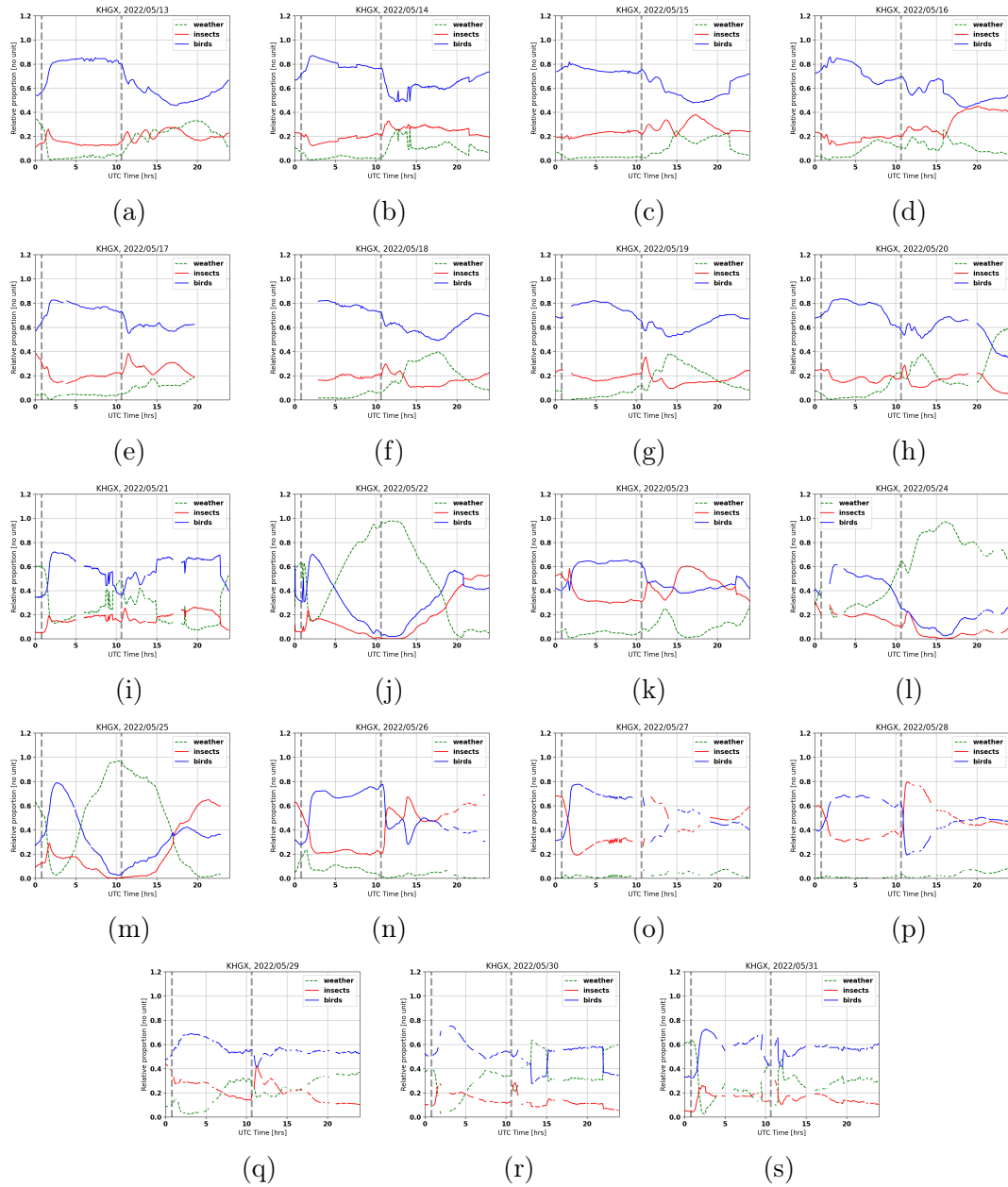


Figure A.10: Same as Figure A.9, but for May 13 - 31, 2022.

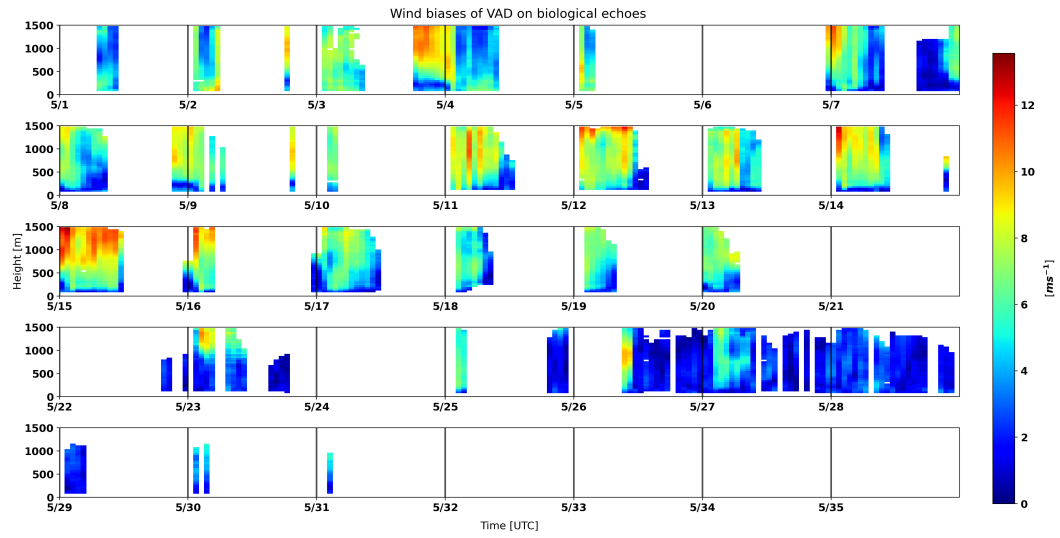


Figure A.11: Biological VAD-wind biases for each day during the KHGX (Houston, TX) May 2018 nocturnal migration season. The black vertical line denotes 0 UTC.

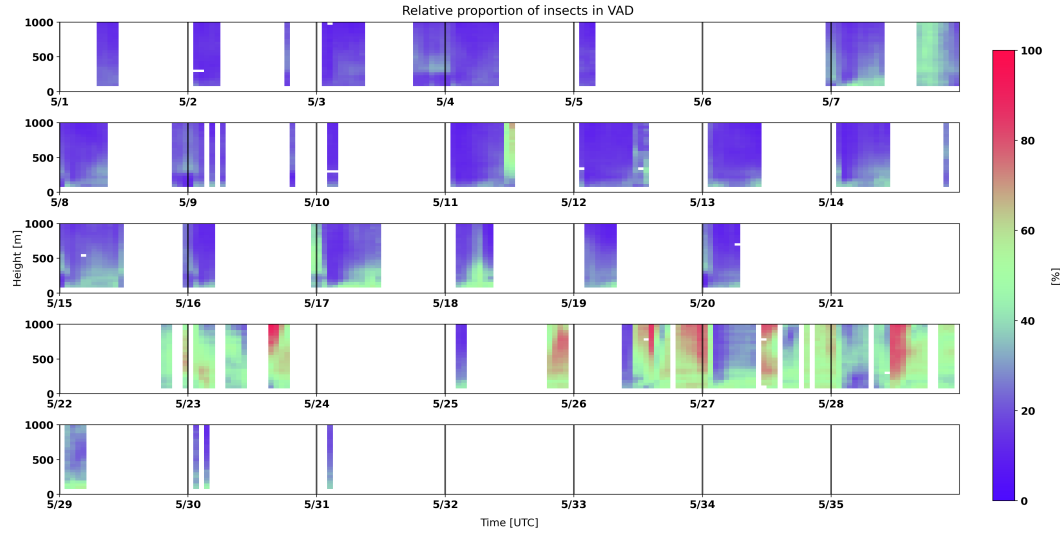


Figure A.12: Same as Figure A.11 but for insects-birds ratios

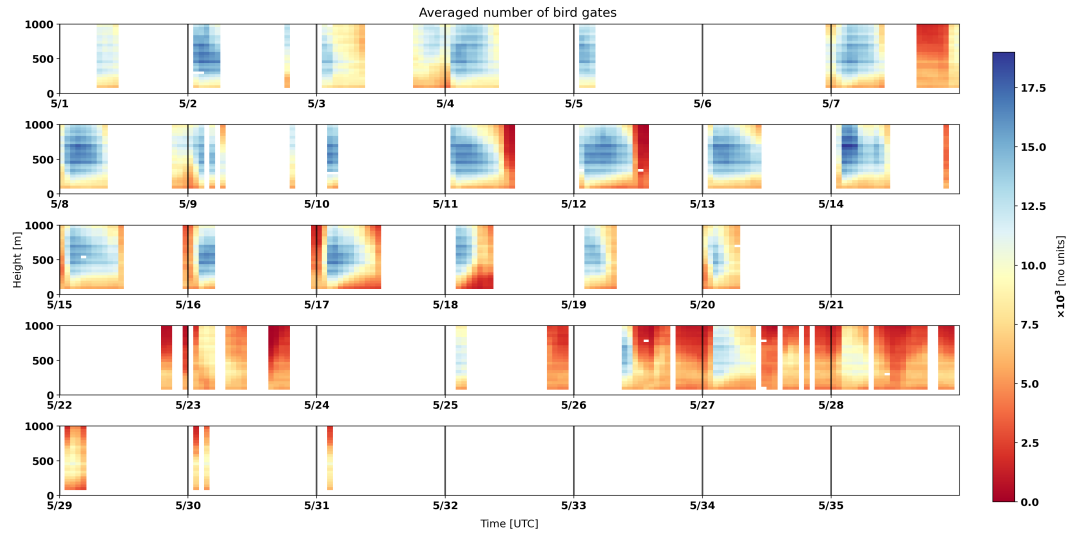


Figure A.13: Same as Figure A.11 but for total bird gates

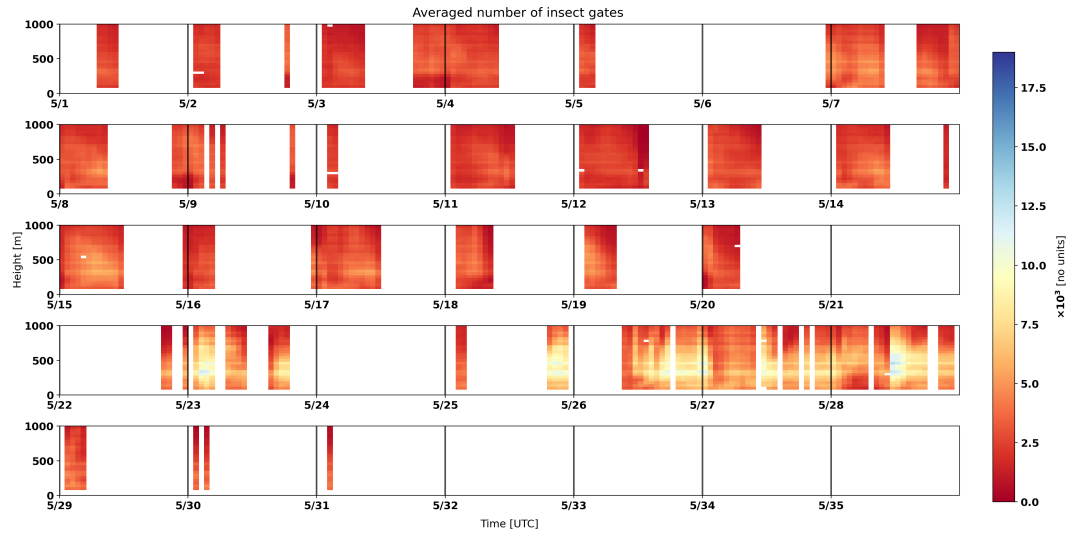


Figure A.14: Same as Figure A.11 but for total insect gates

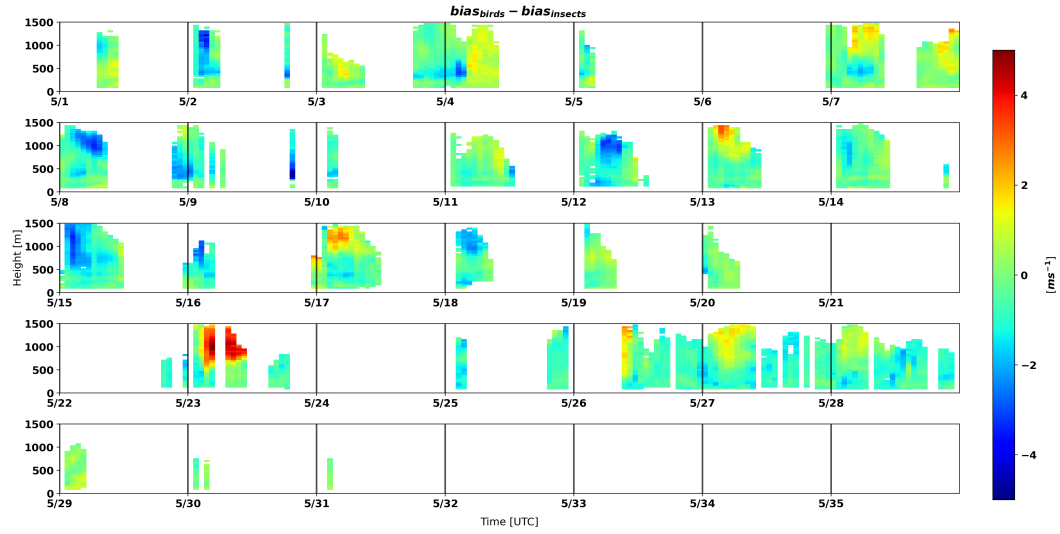


Figure A.15: Same as Figure A.11 but for bird-insect bias differences

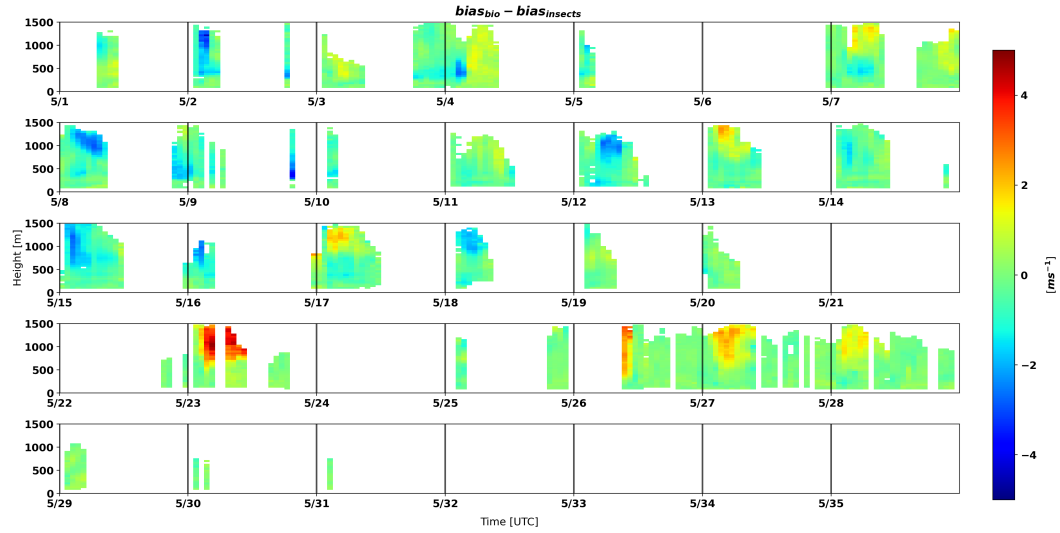


Figure A.16: Same as Figure A.11 but for improvements