FINAL REPORT

AVIAN INFLUENZA RISK ASSESSMENT FOR THE UNITED STATES: MODELING PATHWAYS OF DISEASE SPREAD BY WILD BIRDS

July 2009 Contract #07-7100-0228-CA9

Submitted to USDA-WS-APHIS-National Wildlife Research Center

Principal Investigators/Editors: Paul F. Doherty, Jr. and Kenneth R. Wilson

The Risk Assessment Committee (in alphabetical order):
Amy J. Davis[1]
Paul F. Doherty, Jr.[1]
Matthew L. Farnsworth[2]
Alan B. Franklin[3]
Michele M. Gilbert[1]
Jennifer A. Hoeting[4]
Robert G. McLean[3]
Andrew A. Merton[4]
Toni J. Piaggio[3]
Ryan S. Miller[2]
Gordon C. Reese[1]
Colleen T. Webb[5]
Kenneth R. Wilson[1]

[1]Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, USA
[2]USDA-APHIS-VS-Centers for Epidemiology and Animal Health, Fort Collins, USA
[3]USDA-APHIS-WS-National Wildlife Research Center, Fort Collins, USA
[4]Department of Statistics, Colorado State University, Fort Collins, USA
[5]Department of Biology, Colorado State University, Fort Collins, USA

Preface

The global spread of highly pathogenic H5N1 avian influenza (AI) is thought to be partially mediated by waterfowl. Although this strain of AI has not arrived in North America, the ramifications of such an introduction are enormous, including negative effects on the US poultry industry. To help plan for the possibility of an AI introduction our work plan focused on identifying potential routes of introduction and subsequent spread of AI by waterfowl in the United States. We set out to develop and use spatially explicit, mathematical and statistical risk assessment models with empirical data from 1) environmental and cloacal samples collected from wild birds in 2006 as prescribed in the national strategy "An early detection system for Asian H5N1 Highly Pathogenic Avian Influenza", 2) banding and band recovery data from a variety of waterfowl species, 3) distribution of poultry operations in the U.S., and 4) genetic analyses of detected avian influenza subtypes, as available within the allotted timeframe. This final report is in the form of three sections. The entire Risk Assessment Committee was involved in frequent discussions that guided the research from beginning to end.

Section 1 is the majority of the report and is a detailed description of the primary methods and results. Categorizing waterfowl migration patterns is difficult, so we developed and explored several metrics, namely waterfowl banding and recovery incidence, minimum spanning tree, a cluster metric and a satellite metric. We applied these metrics to different functional groups of waterfowl (e.g., dabblers) as well as individual species to help identify areas of North America that potentially should receive additional surveillance activities. If a particular species, or functional group, was thought to be spreading a disease, areas of the country and associated poultry operations could be identified through linkages from our modeling efforts, and ultimately notified. J. Hoeting, A. Merton, C. Webb, and A. Davis were the lead team members in preparing this section.

Section 2 of the report is a draft manuscript entitled "Linking wildlife movements to disease spread: implications for highly pathogenic avian influenza surveillance." This section distills the results of Section 1 to address one of the main themes our committee focused upon. P. Doherty and A. Merton were the lead team members in preparing this section.

Section 3 is in the form of 2 abstracts from presentations and ongoing work that has partially been supported by this project. The first abstract is focused directly on the question of whether there is evidence that waterfowl migration could spread low pathogenic AI as well as the genetic relatedness of low pathogenic AI across the country. We are currently applying logistic regression models in a model selection framework to predict the pattern of AI prevalence in waterfowl across a large portion of the U.S. in space and time. Within this framework, we estimated the probability of an AI positive bird as a function of the waterfowl migration metrics, age and sex of the bird, week of AI sample collection, and ambient temperature. The second abstract pertains to investigating whether phylogenic information in the low pathogenic AI samples could shed light on the likelihood of paths of introduction, as well as relatedness of AI across the country.

Table of Contents

Preface **1**

Section 1 - Modeling movements of Avian Influenza in North American Waterfowl **3**

Chapter 1: Measuring Waterfowl Movement - Network Analysis Overview **3**

1.1 Motivation and Available Data 3

1.2 Network Measures 6

Chapter 2: Waterfowl Banding and Recovery Data **6**

2.1 Overview 6

2.2 All Harvest vs Direct Harvest 7

2.3 Lattice Data 7

2.4 Species and Functional Groups 7

Chapter 3 Flow **8**

3.1 Motivation 8

3.2 Flow Figures 8

Chapter 4 Minimum Spanning Trees **18**

4.1 Motivation 18

4.2 Minimum Spanning Tree Figures 18

Chapter 5 Clustering and Satellite Analysis **28**

5.1 Motivation 28

5.2 Cluster and Satellite Figures 28

Chapter 6: Understanding the Metrics **47**

6.1 Overview 47

6.2 Example: Node 1498 47

6.3 Example: Node 868 48

Chapter 7 Direct Harvest to Identify Flow, MST, Clusters, and Satellites **49**

7.1 Overview 49

Literature Cited 86

Section 2 - Linking Wildlife Movements To Disease Spread: Implications For Highly Pathogenic Avian Influenza Surveillance **87**

Abstract **87**

Introduction **87**

Methodology **88**

Results **91**

Discussion **93**

Literature Cited **94**

Section 3 – Abstracts of Ongoing Research **104**

Abstract 1: Phylogenetic analysis of large-scale surveillance of avian influenza A virus from wild bird fecal samples across the United States **104**

Abstract 2: Avian influenza in waterfowl: Assessing the potential spread and impact on poultry in the United States based on wild waterfowl movement. **105**

Acknowledgments **106**

Section 1 - Modeling movements of Avian Influenza in North American Waterfowl

Chapter 1: Measuring Waterfowl Movement - Network Analysis Overview

1.1 Motivation and Available Data

Migration of waterfowl is hypothesized to contribute to the introduction and spread of avian influenza (AI) in North America. For example, AI could be introduced by waterfowl migrating from Asia into Alaska or from Europe into Greenland and Eastern Canada. Characterization of waterfowl migration patterns can help us assess the location-specific risk of avian influenza outbreaks in the United States, especially in relation to poultry operations.

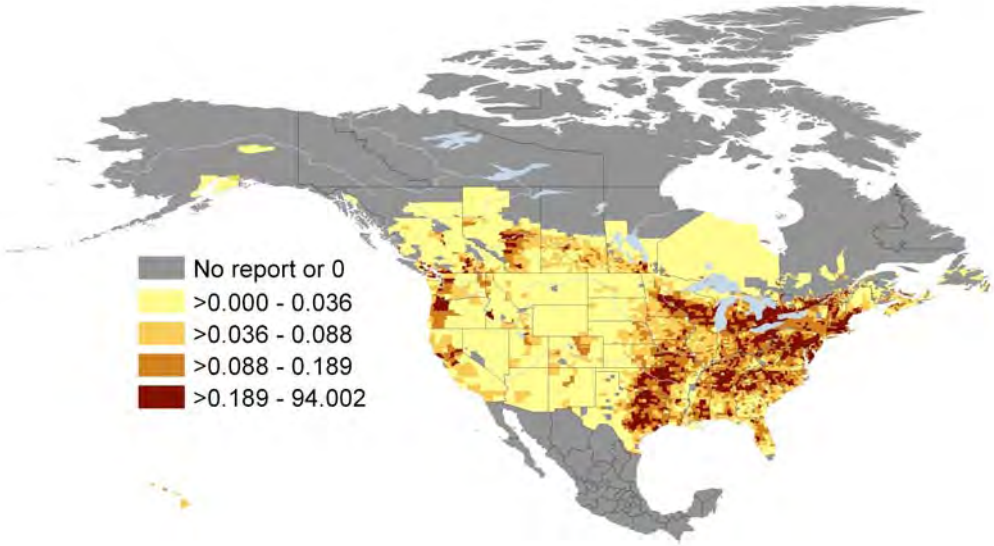
Waterfowl migration patterns can be characterized using a network perspective. Nodes in the network represent locations, and edges represent the number of birds that travel between two locations. Connectivity of the network can potentially impact how frequently or severely a location is exposed to AI. Connectivity can be measured at multiple levels including local, flyway, and continental scales. In order to better understand waterfowl migration, we developed five network metrics: Banding Flow, Recovery Flow, Minimum Spanning Tree, Cluster, and Satellite measures.

1.1.1 Banding and Recovery Data

We generated each of the five network metrics using waterfowl banding and recovery information from the United States Geological Service Patuxent Wildlife Research Center Bird Banding Laboratory (BBL). Waterfowl banding/recovery data include spatial and temporal information on waterfowl movement by species. These data are collected with coordinates of the banding and recovery locations and summarized to the 10-min block. For our study, we created a 200-km grid across North America and aggregated individual banding/recovery data to the nearest grid point. These grid points formed the nodes of our waterfowl migration network. Recorded movements of birds between nodes create edges (e.g., banded at one location and recovered at another). Edges can be weighted according to the number of birds moving along the edge (i.e., between two nodes). See Chapter 2 for more details.

1.1.2 Poultry location data

Poultry location data were obtained from the United States Department of Agriculture (USDA). Unlike the banding and recovery data, specific location data are not available and the level of resolution is to the county level. We summarize these data in four ways: total number of poultry farms per square mile (Fig. 1.1 Top), the total number of broiler farms (Fig. 1.1 Bottom), and the total number of small and large broiler farms (Fig. 1.2 Top and Fig. 1.2 Bottom, respectively). The poultry data maps can be compared with the waterfowl connectivity maps discussed below to help identify and prioritize areas of the country for increased surveillance and management action.



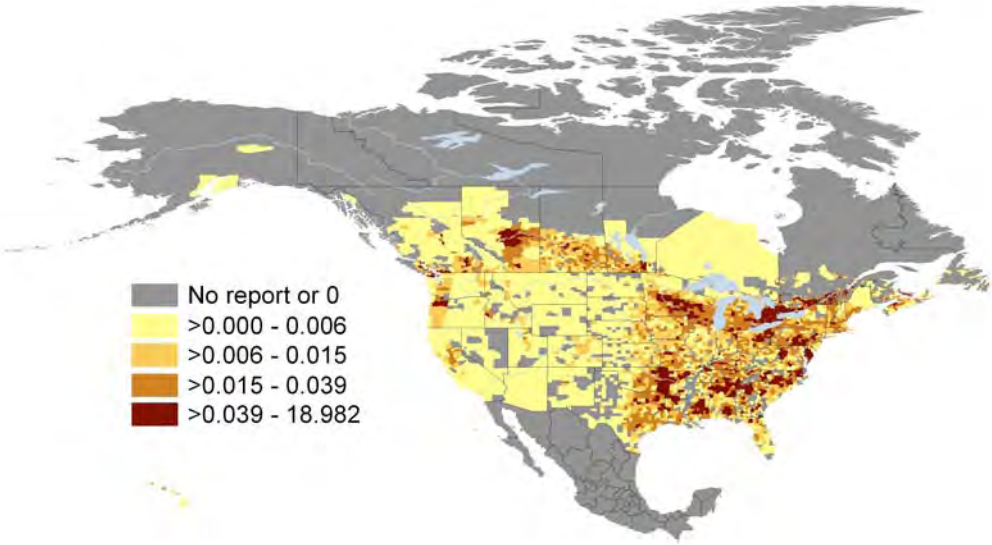
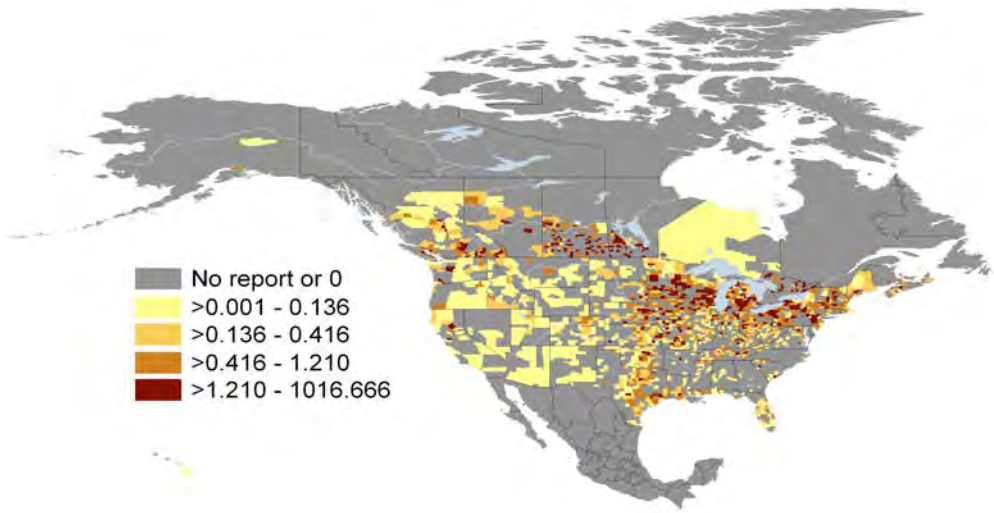


Figure 1.1: Average number (farms per square mile) of poultry (Top) and broiler farms (Bottom) in North America.



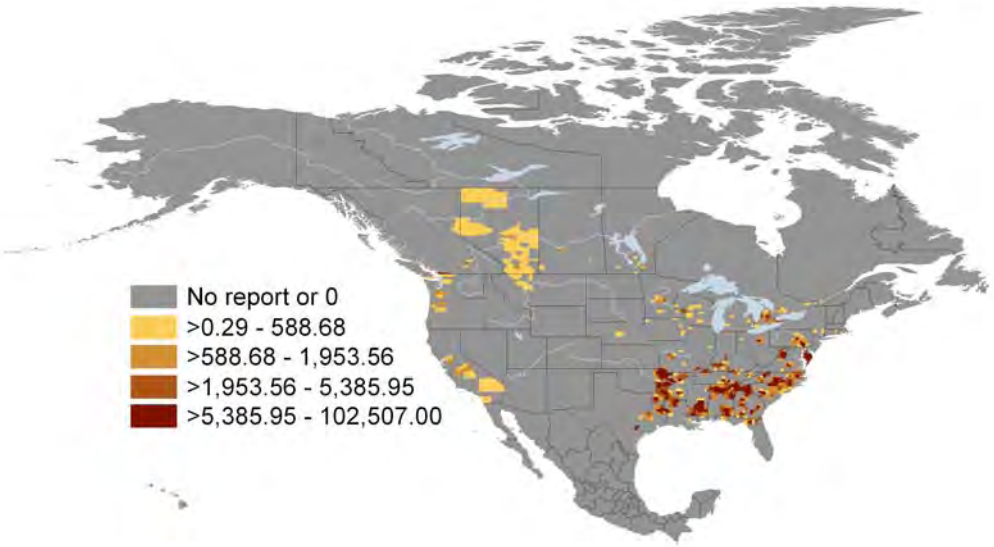


Figure 1.2: Average number of broiler chickens per square mile on small (<5000 birds, top) and large (≥5000 birds, bottom) farms in North America.

1.2 Network Measures

1.2.1 Banding and recovery flow

Banding flow and recovery flow are indices of the number of birds at a node. Banding flow is the number of birds banded at a node, and recovery flow is the number of birds recovered at a node. These flow measures are an attempt to capture local level connectivity and mixing. Areas with higher concentrations of waterfowl are hypothesized to have more mixing of birds and a higher possibility of AI transmission. Higher rates of contact among birds may occur directly or indirectly via feces and shedding of AI virus into shared water sources.

1.2.2 Minimum spanning tree

The Minimum Spanning Tree (MST) is a global or whole network measure of connectivity (Urban and Keitt 2001). The MST is the smallest subgraph or portion of the entire network for which all of the nodes in the network remain connected. There are multiple MSTs, or multiple ways to subset the network and still have all nodes remain connected.

1.2.3 Clusters and satellites

Clusters group nodes together based on similar connectedness patterns, which in some cases resemble flyways. Clusters represent an intermediate network scale in between the local and global or full network scales. Within a cluster, most birds entering or leaving a node will come from other nodes in the same cluster. However, many nodes will also have a number of birds entering from or leaving to nodes in other clusters. The strength of these intercluster connections relative to the intracluster connections describe the satellite metric, which is a measure of mixing between clusters, i.e., mixing at an intermediate scale in the network and on the landscape. Nodes that are important satellites could be important for movement of AI between clusters.

Chapter 2: Waterfowl Banding and Recovery Data

2.1 Overview

Waterfowl banding and recovery data were used to develop the metrics. These data were obtained from the Bird Banding Laboratory (BBL) of the United States Geological Service Patuxent Wildlife Research Center. This is a large, publicly available set of data, and can be accessed from (http://www.pwrc.usgs.gov/bbl/). Each record in the database contains information for an individual bird. Relevant data fields include species, sex, age (hatch year or after hatch year), date of banding, time of banding, and coordinates of banding location (to the 10-min block). If the bird was recovered, usually by hunter harvest, information on the date of recovery, time of recovery, and coordinates of the recovery location are also available. Other data fields are also available, but were not used in this analysis.

According to the USGS Patuxent Wildlife Research Center, hunters recover approximately 13.5% of the total game birds banded. The data used in these analyses include only hunter-recovered birds. Other types of recoveries, such as road kill and re-captured birds were excluded in order to reduce variability in the data.

In addition to the BBL data fields, we generated three data fields for our analyses: functional group, time to harvest, node number. Birds were assigned to the functional groups based on their

species. See Section 2.4 for details. Time to harvest was calculated as the number of weeks between banding and recovery. Node number is the number of the 200-km grid point to which an individual bird was assigned. A more complete description of the lattice is given in Section 2.3.

2.2 All Harvest vs Direct Harvest

Two different sets of data were used in this analysis. The All Harvest data set includes all birds that were recovered from 2002 to 2006, regardless of when they were banded. The Direct Harvest data set includes all birds recovered from 2002 to 2006 that were harvested within 52 weeks of being banded. Approximately 40% of the recorded birds are direct harvests. The Direct Harvest data set is less noisy than the All Harvest data set. Fewer outliers (i.e., birds with unusual trajectories) occur in the Direct Harvest data set as these birds have had less time to travel to more locations.

2.3 Lattice Data

Each banding or recovery entry is recorded at a specific location. In order to create our network, we generated a 200-km lattice over North America using a Great Circle Distance (GCD) projection. Tradeoffs occur in any choice of projection. Because we were most interested in bird movement, we chose GCD in order to preserve distances between nodes in a data set that span the continental scale. For convenience in our projection, we used 40° North latitude and 100° West longitude as the starting point. These are round numbers and their intersection falls roughly in the middle of the continental U.S. We generated the lattice using a GCD projection by identifying locations that are approximately 200 km away in each of the 4 cardinal directions (N, E, S, W) and continued until a large rectangular window spanning the latitudes of approximately 18° to 80° North and longitudes 50° West to 170° East was filled. Our lattice extended to 170° East to include the Aleutian Islands of Alaska. Specific locations were aggregated to the nearest grid point in order to associate individual records with nodes. Network metrics were calculated at the 200-km scale. We additionally investigated 100-km and 10-minute scales, but the 200-km scale gave the most reasonable level of detail while minimizing noise in the data.

2.4 Species and Functional Groups

Waterfowl species were assigned functional groups based on expert knowledge. Species assignments are shown in Table 2.1. For our analysis, we examined the functional groups: Dabblers, Dark Geese, Light Geese, and all Waterfowl. We defined all waterfowl as all Dabblers, Dark Geese, and Light Geese. Additionally, we examined five Dabbler species individually: Mallard, Blue-winged Teal, American Black Duck, Northern Pintail, and Wood Ducks. Each network metric was calculated for each of the species and functional groups given above (9 total) for both All Harvest and Direct Harvest data sets.

Table 2.1: Description of Functional Groups

Functional Group	Species Included	Scientific Names
Dabblers	American Black Duck	*Anas rubripes*
	Green-winged Teal	*Anas crecca*
	American Widgeon	*Anas americana*
	Blue-winged Teal	*Anas discors*
	Cinnamon Teal	*Anas cyanoptera*
	Gadwall	*Anas crecca*
	Mallard	*Anas platyrhynchos*
	Mallard x Black Duck Hybrid	
	Mottled Duck	*Anas fulvigula*
	Northern Pintail	*Anas acuta*
	Northern Shoveler	*Anas clypeata*
	Unidentified Teal	
	Wood Duck	*Aix sponsa*
Dark Geese	All Canada Goose subspecies	*Branta canadensis*
	Brant	*Branta bernicla*
	Greater White-fronted Goose	*Anser albifrons*
Light Geese	Snow Goose	*Chen caerulescens*
	Ross's Goose	*Chen rossii*

Chapter 3 Flow

3.1 Motivation

Flow metrics potentially capture local contact rates at the node level.

3.1.1 Defining the metric

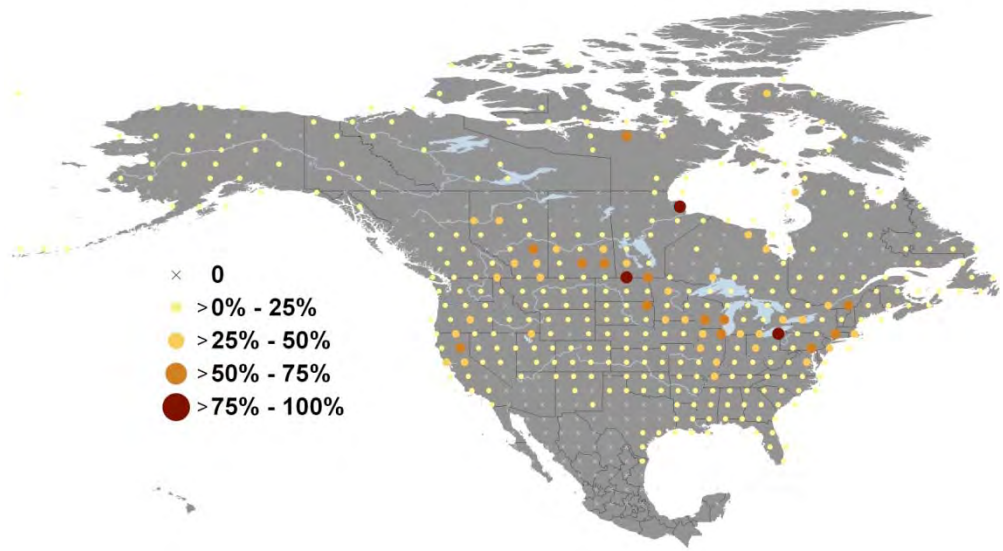
The banding flow metric is the total number of birds banded at a node (either All Harvest or Direct Harvest). The recovery flow metric is the total number of birds recovered at a node. In the following network flow figures (Figures 3.1-3.9), larger circles indicate large numbers of birds.

3.1.2 Alternative metrics

We also generated several alternative metrics based on incidence measures described by Hanski (1994). We calculated net flow for each node as the number of birds banded minus the number of birds recovered at each node, standardized by the maximum banded minus recovered in the data set. We also calculated several versions of Hanski's incidence function, a standard measure of local connectivity in the ecology literature. These alternative metrics were highly correlated with each other and with banding and recovery flow. Because banding and recovery flow are the simplest metrics, we report them here.

3.2 Flow Figures

The resulting figures identify areas that are potentially important in the spread of AI as waterfowl migrate from winter to summer activity areas. As can be seen, there is not a consistent pattern of flow for all species.



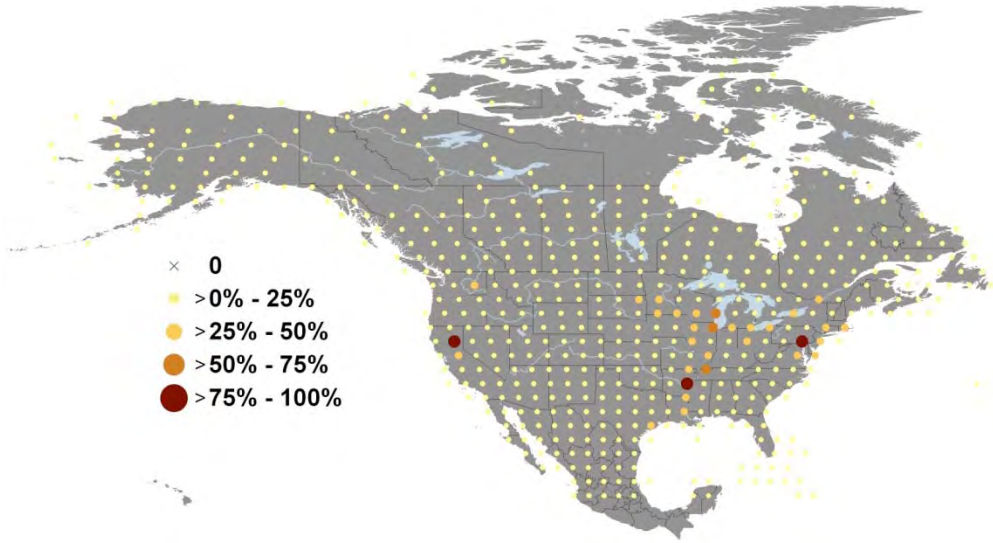
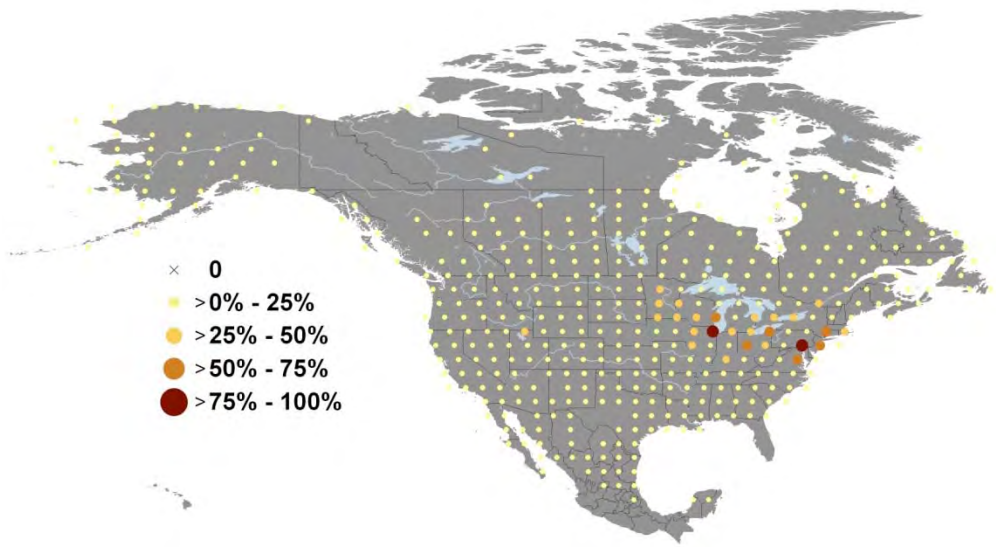


Figure 3.1: Banding (top) and recovery (bottom) network flow for all waterfowl harvested 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).



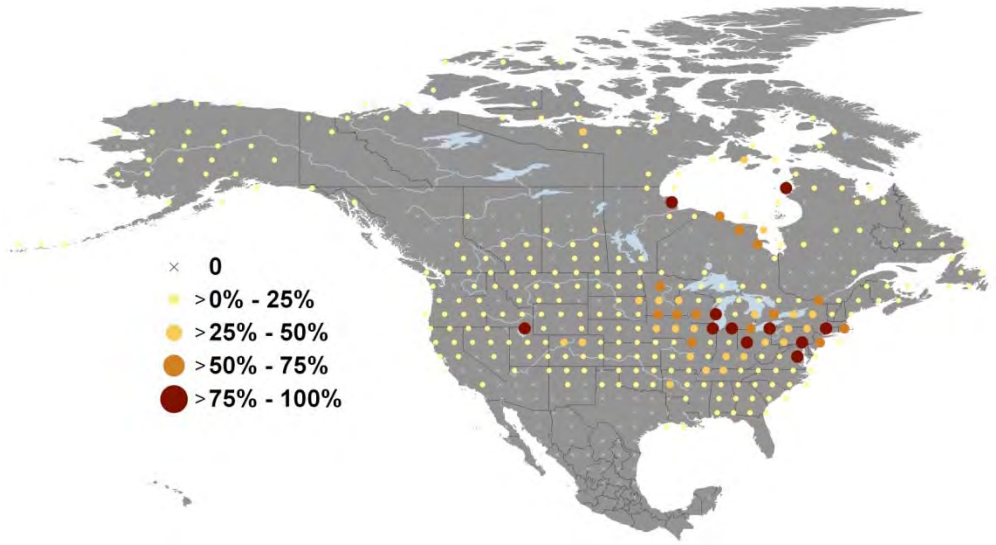
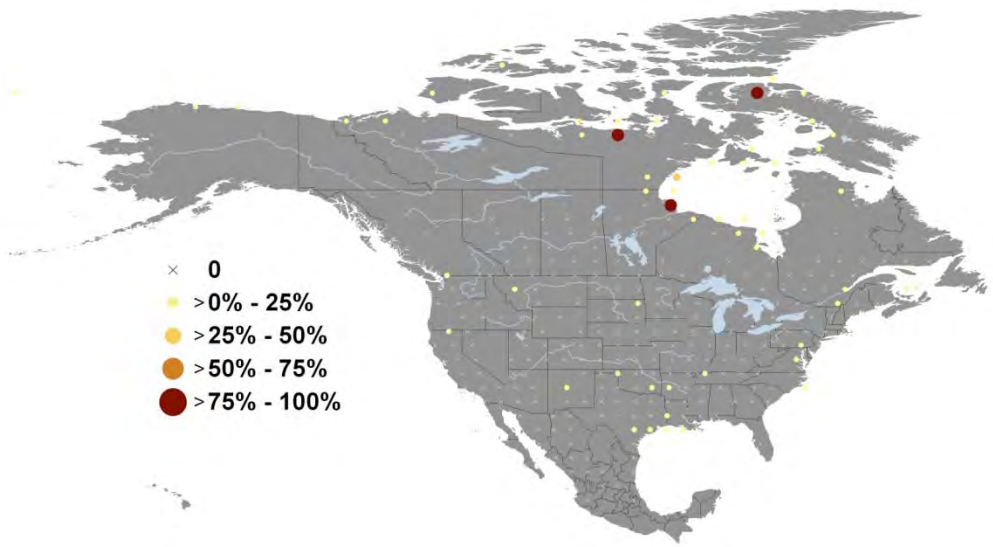


Figure 3.2: Banding (top) and recovery (bottom) network flow for dark geese harvested 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).



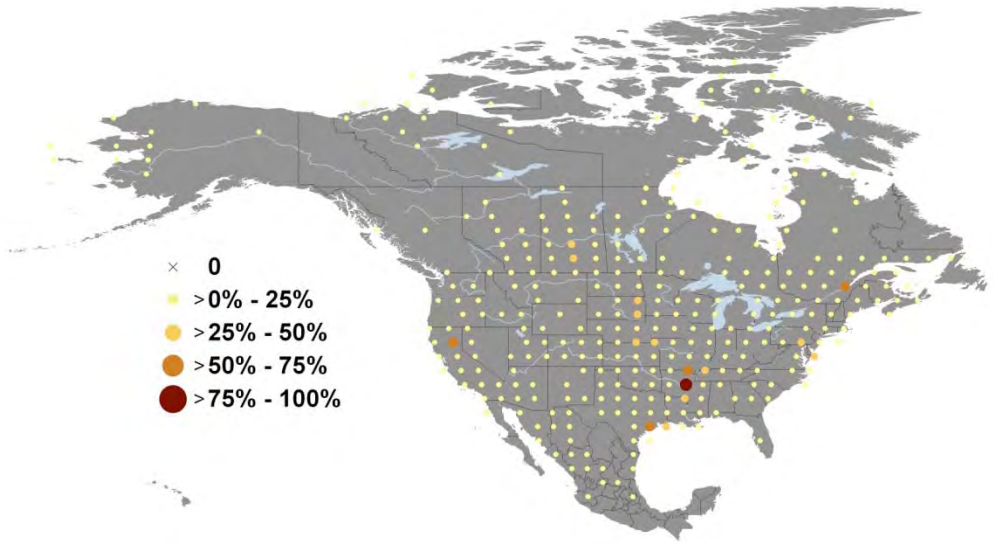
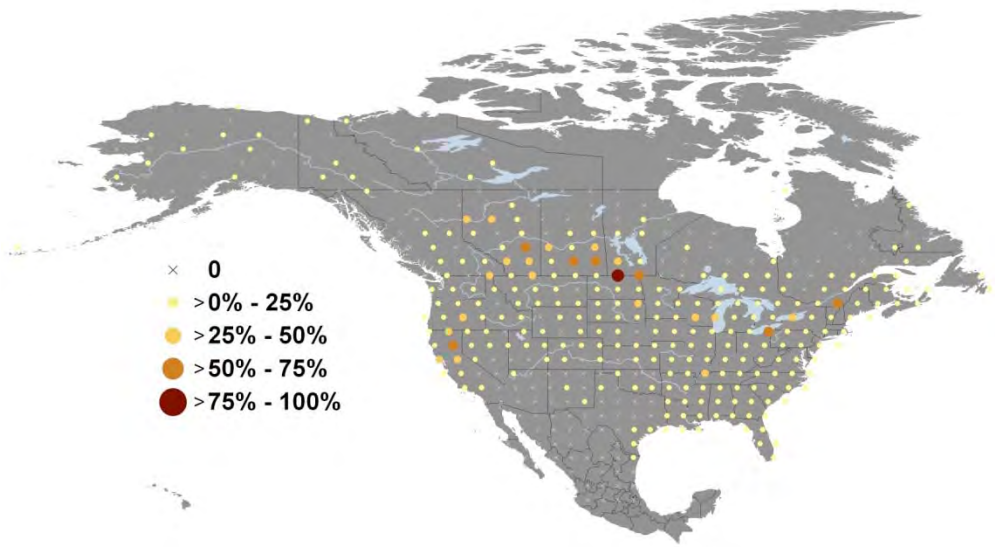


Figure 3.3: Banding (top) and recovery (bottom) network flow for light geese harvested 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).



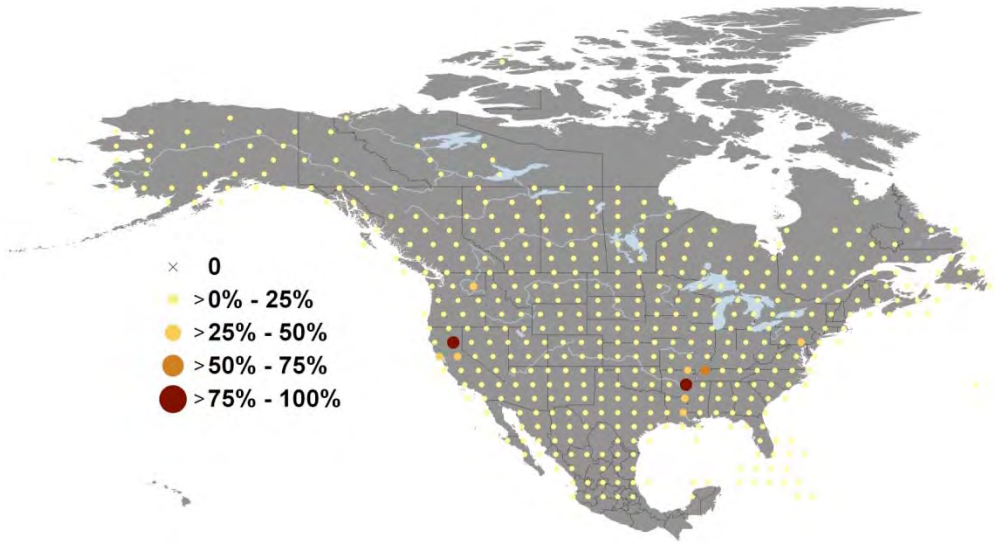
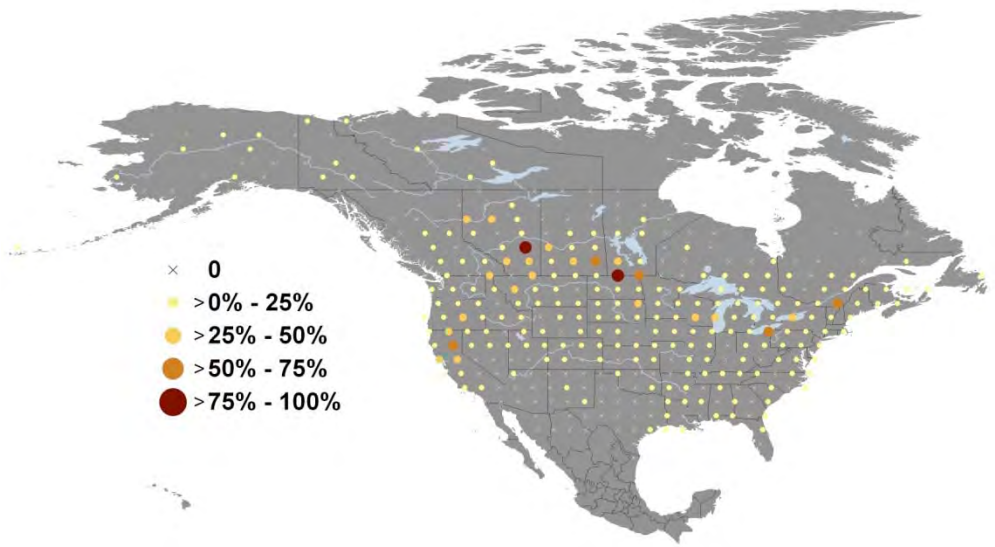


Figure 3.4: Banding (top) and recovery (bottom) network flow for dabblers harvested 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).



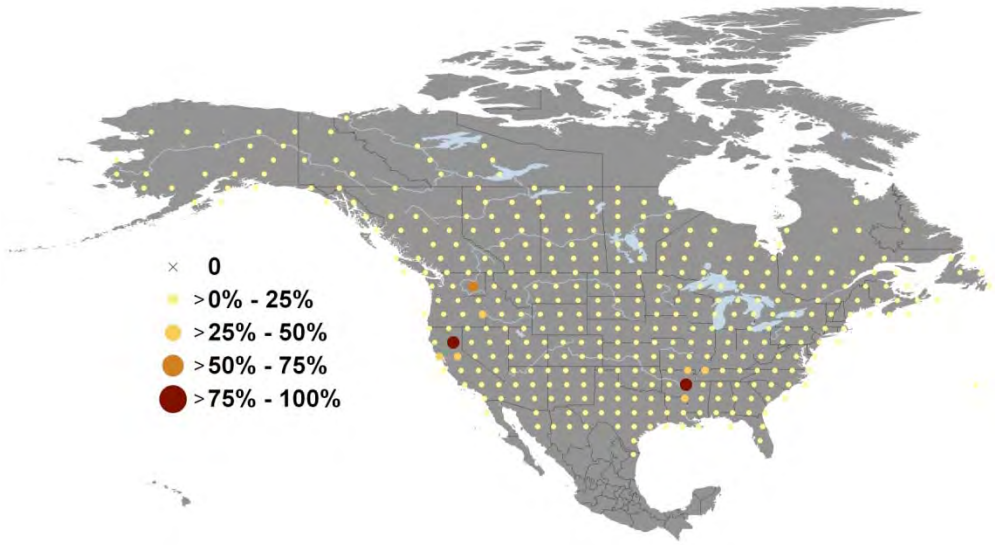


Figure 3.5: Banding (top) and recovery (bottom) network flow for Mallards harvested 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

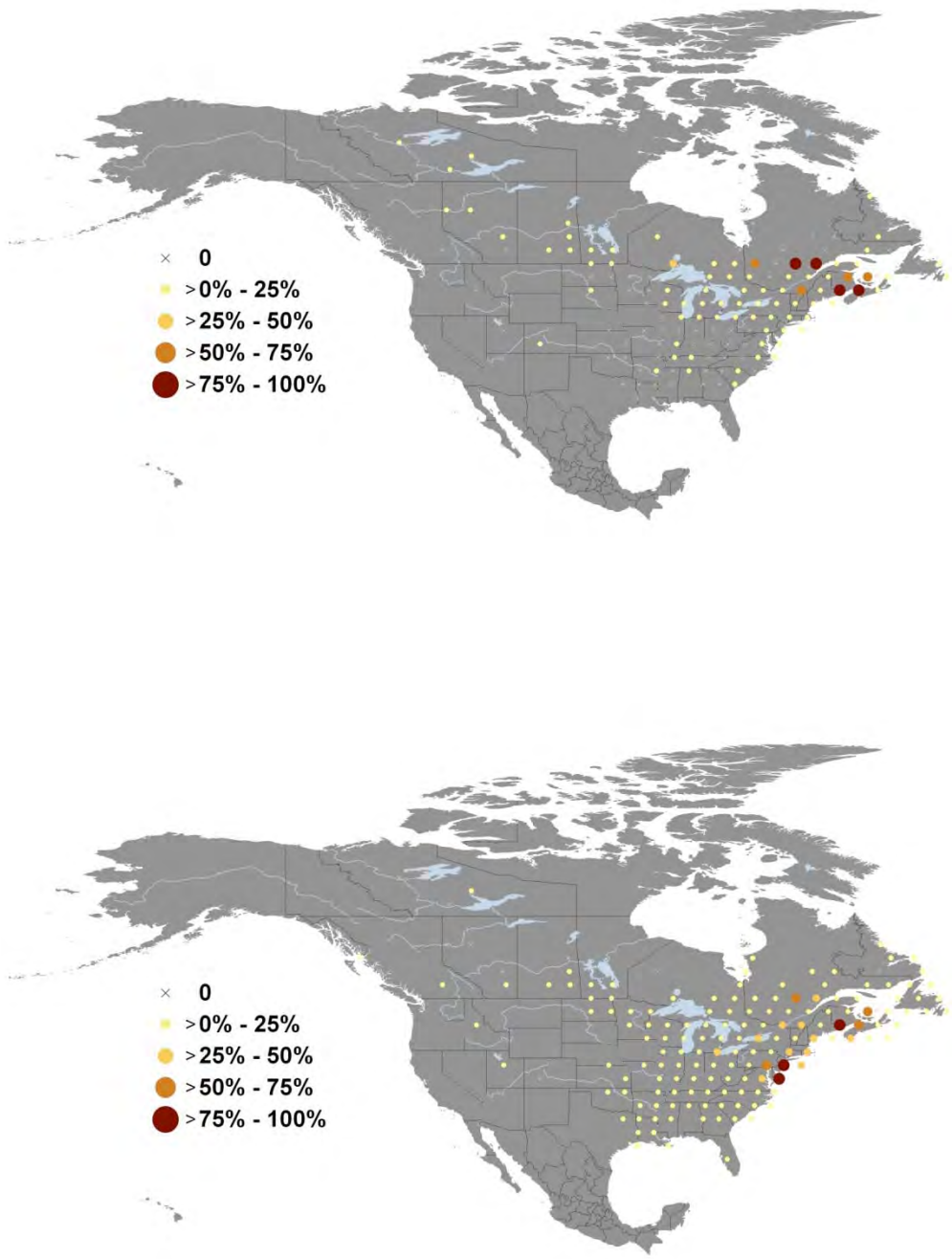
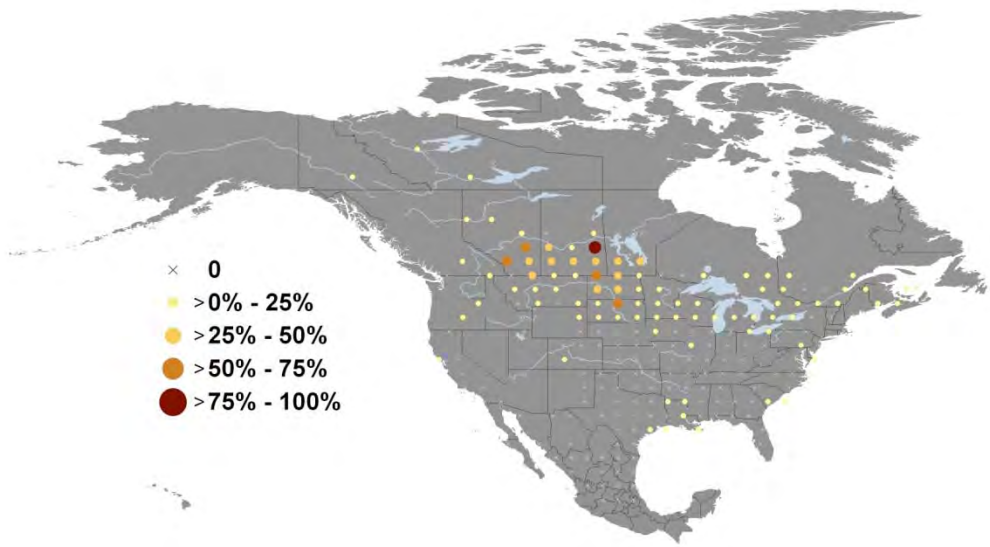


Figure 3.6: Banding (top) and recovery (bottom) network flow for American Black Ducks harvested 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).



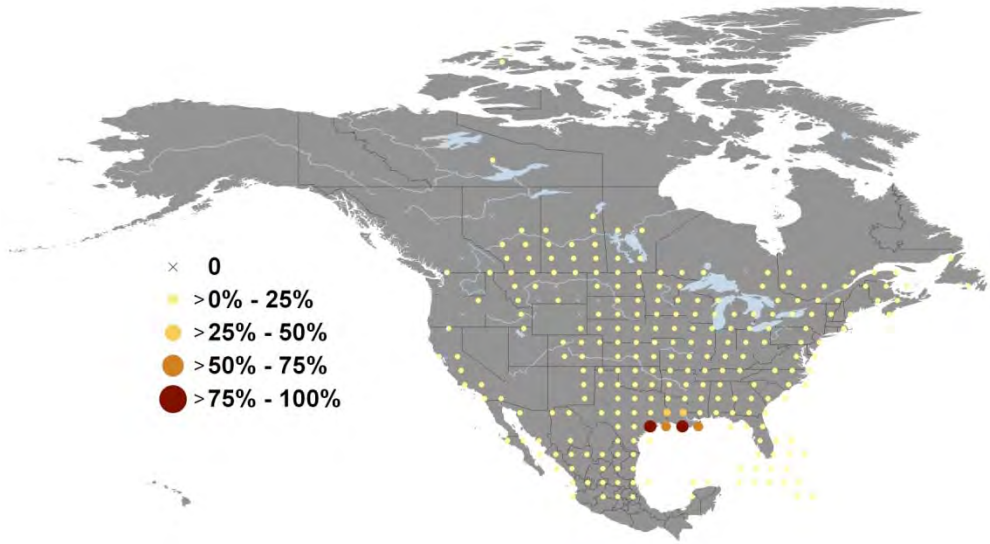
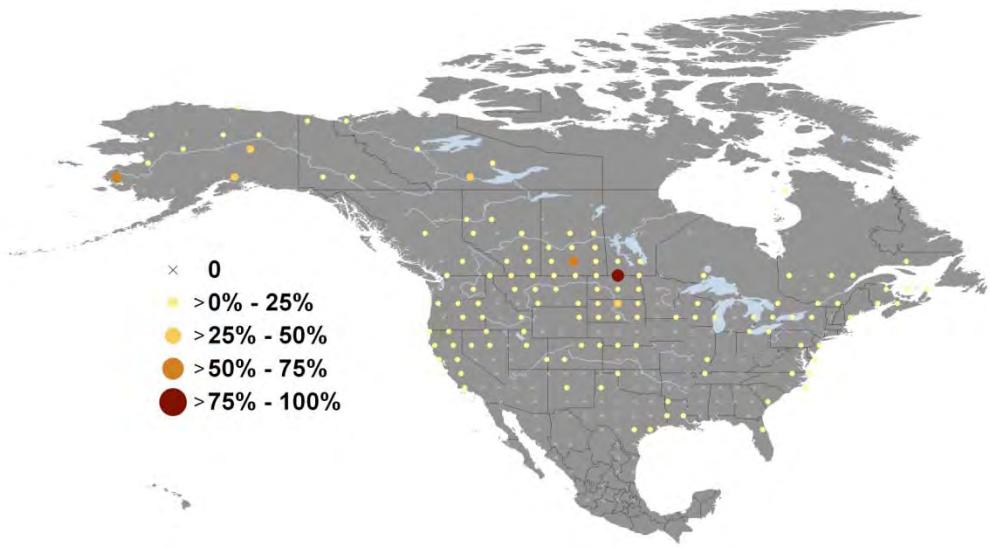


Figure 3.7: Banding (top) and recovery (bottom) network flow for Blue-winged Teal harvested 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).



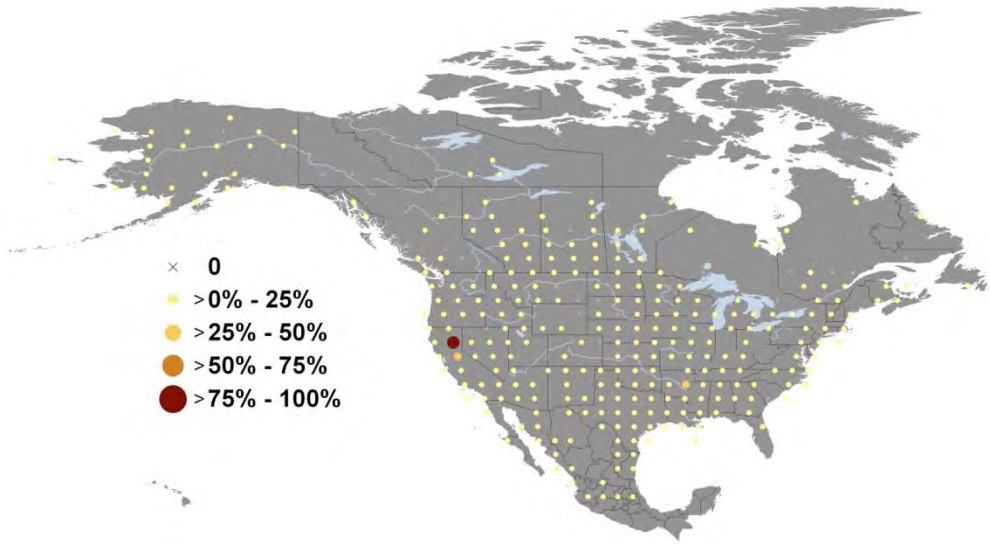
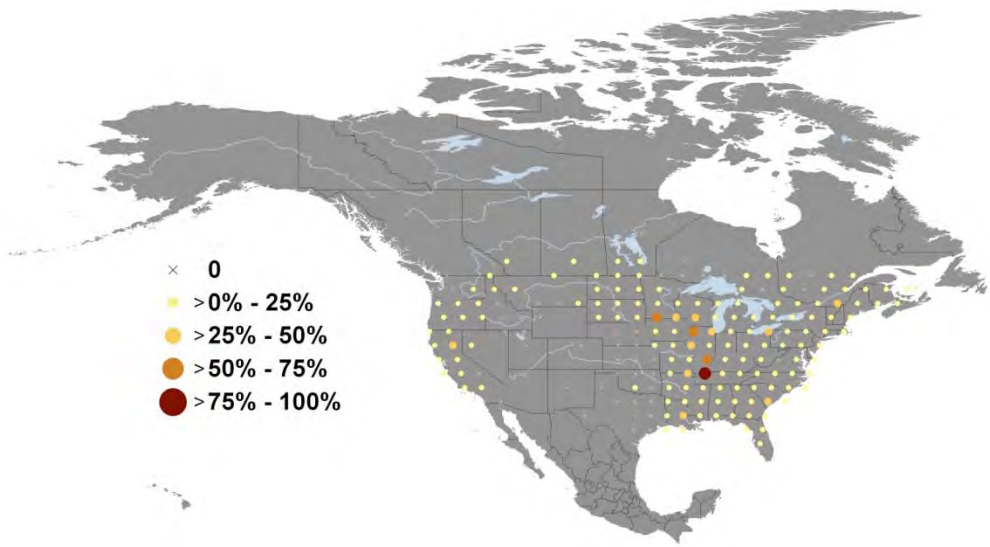


Figure 3.8: Banding (top) and recovery (bottom) network flow for Northern Pintails harvested 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).



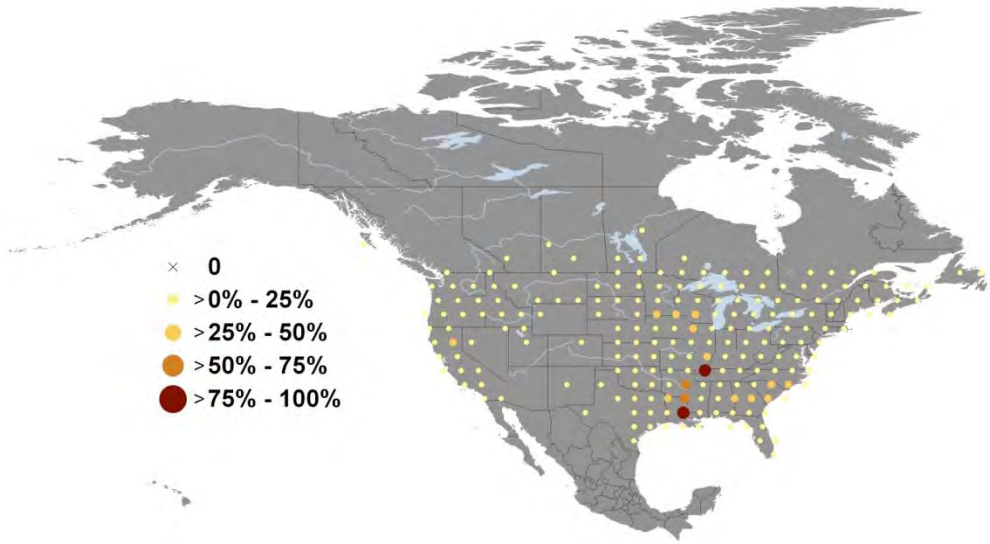


Figure 3.9: Banding (top) and recovery (bottom) network flow for Wood Ducks harvested 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

Chapter 4 Minimum Spanning Trees

4.1 Motivation

The Minimum Spanning Tree (MST) (Ahuja et al. 1993) identifies a minimum set of nodes within the network necessary to maintain full network connectivity, i.e., to maintain at least one connection to every node. Multiple MSTs can occur within a network.

4.1.1 Defining the metric

Finding all MSTs for a large network, such as the bird migration network, is computationally prohibitive. Therefore, we used an algorithm to find multiple plausible MSTs and report a summary MST as described below.

For our MST selection algorithm, we randomly select a node, eliminated it from the network, and searched to see whether all other nodes in the network remained connected. If all other nodes in the network remain connected, then the sampled node remains removed, and another node is randomly selected. If all other nodes in the network do not remain connected, the sampled node is put back in the system and another node is randomly selected. The process continued until no additional nodes can be removed from the system.

The resulting MST is dependent on the random process that created it. Therefore, the process was repeated 100 times. Each time a node is included in an MST, it is assigned a value of one. Otherwise, it is assigned a zero. Nodes with fewer than five birds aggregated to them were also assigned zeros in order to reduce noise in the MST. To generate the MST statistic the zeros and ones were summed for each node across the 100 replicates, and then divided by 100 to give a final MST value for each node that is between zero and one.

4.2 Minimum Spanning Tree Figures

In the MST plots shown below, larger circles indicate a larger MST statistic. In other words, the larger the circle then the more often that node was selected for the MST and the smaller the circle the less often the node was selected for the MST. As with the flow metric, the larger circles for species indicate areas that are potentially important in the spread of AI and the patterns vary by species.

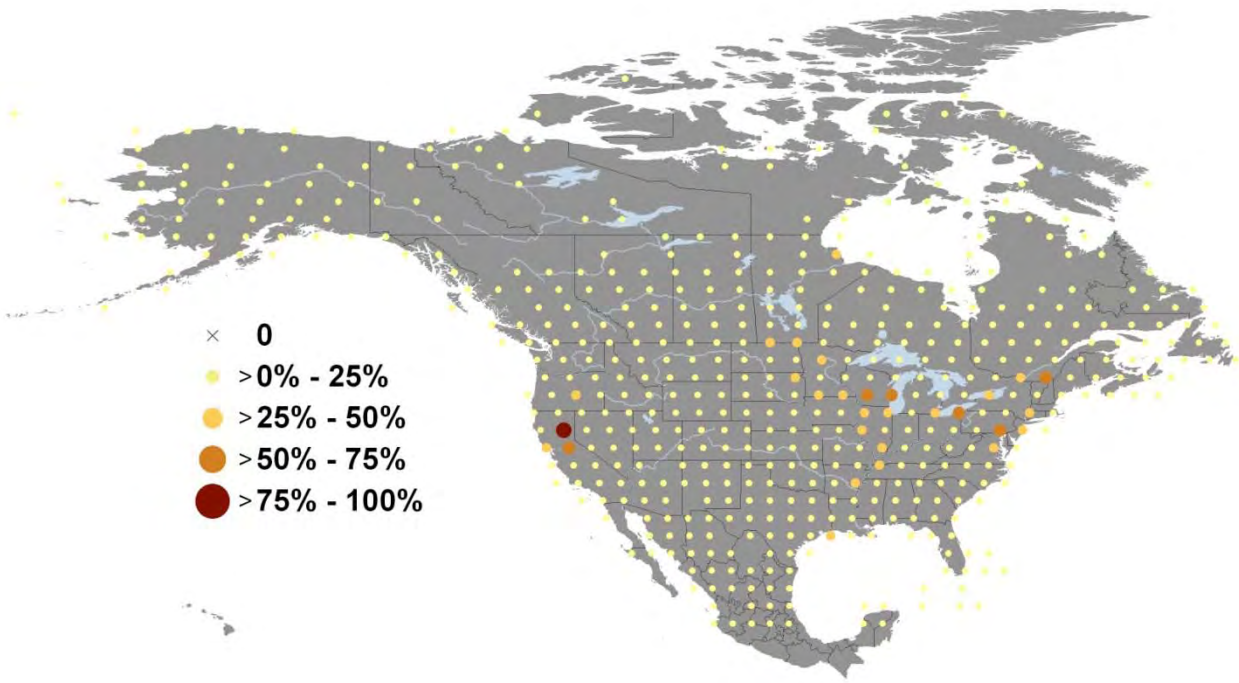


Figure 4.1: The minimum spanning tree (MST) for all harvested waterfowl from 2002 thru 2006. The larger the circle, the more often that node was selected for the MST (larger circles indicate a larger MST statistic and are expressed as quartiles).

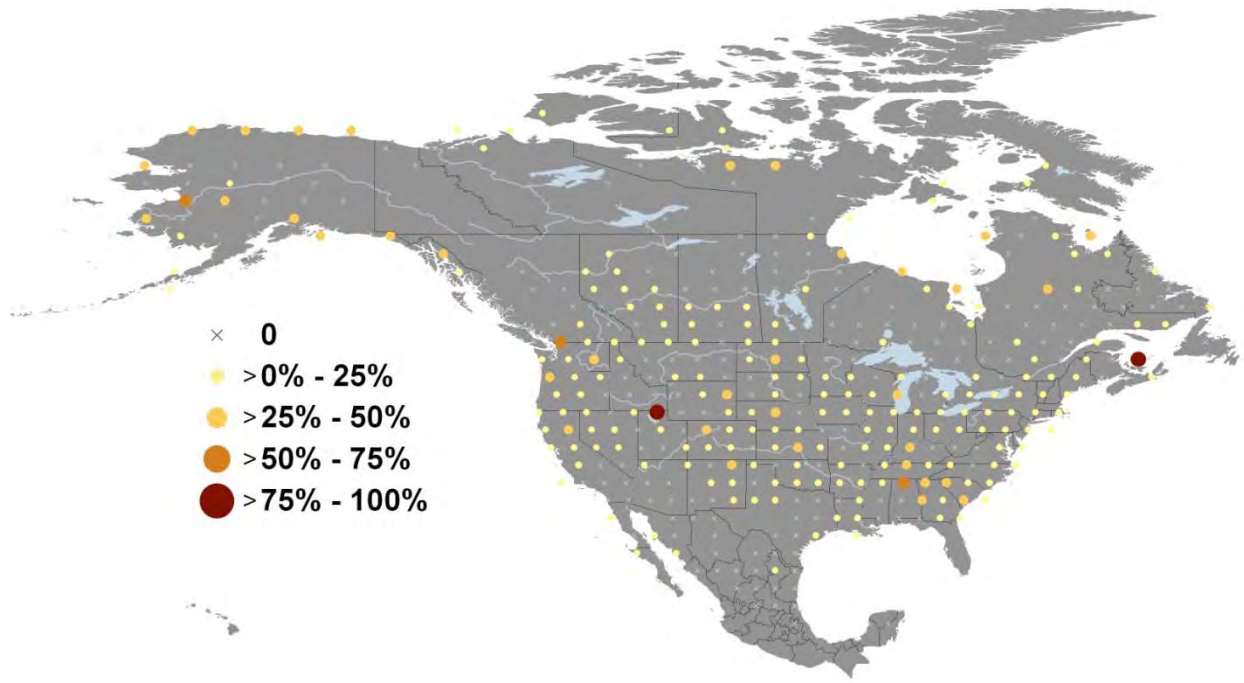


Figure 4.2: The minimum spanning tree (MST) for dark geese harvested from 2002 thru 2006. The larger the circle, the more often that node was selected for the MST (larger circles indicate a larger MST statistic and are expressed as quartiles).

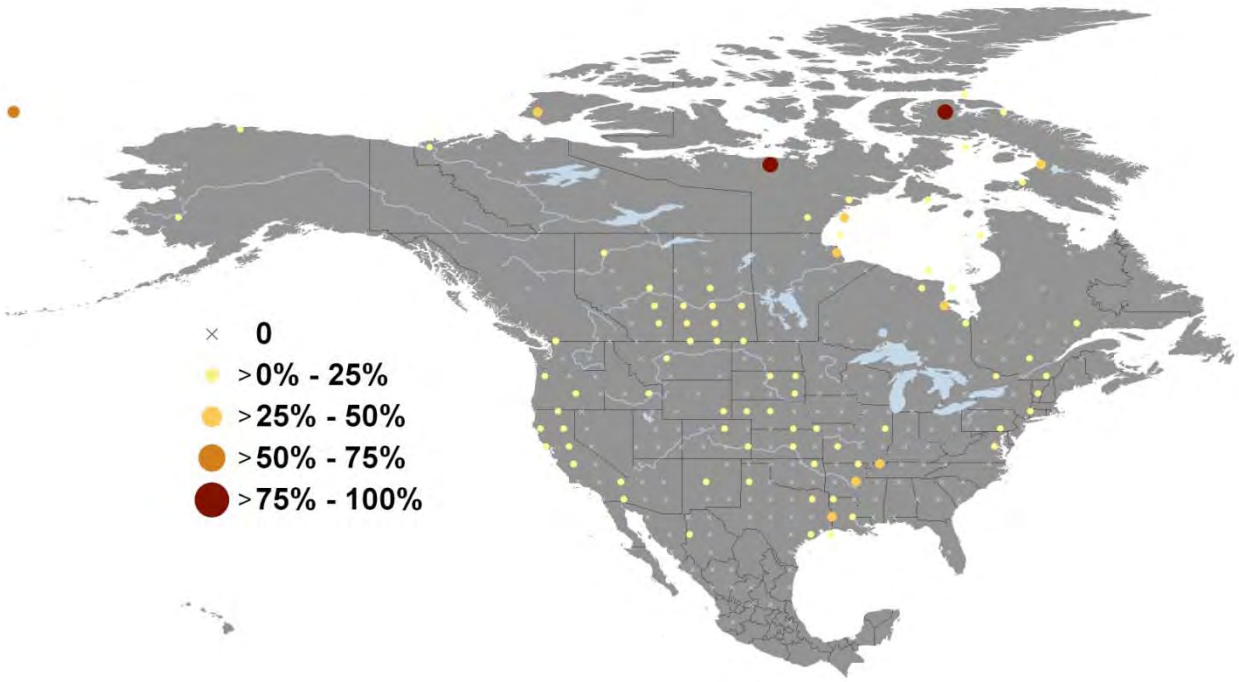


Figure 4.3: The minimum spanning tree (MST) for light geese harvested from 2002 thru 2006. The larger the circle, the more often that node was selected for the MST (larger circles indicate a larger MST statistic and are expressed as quartiles).

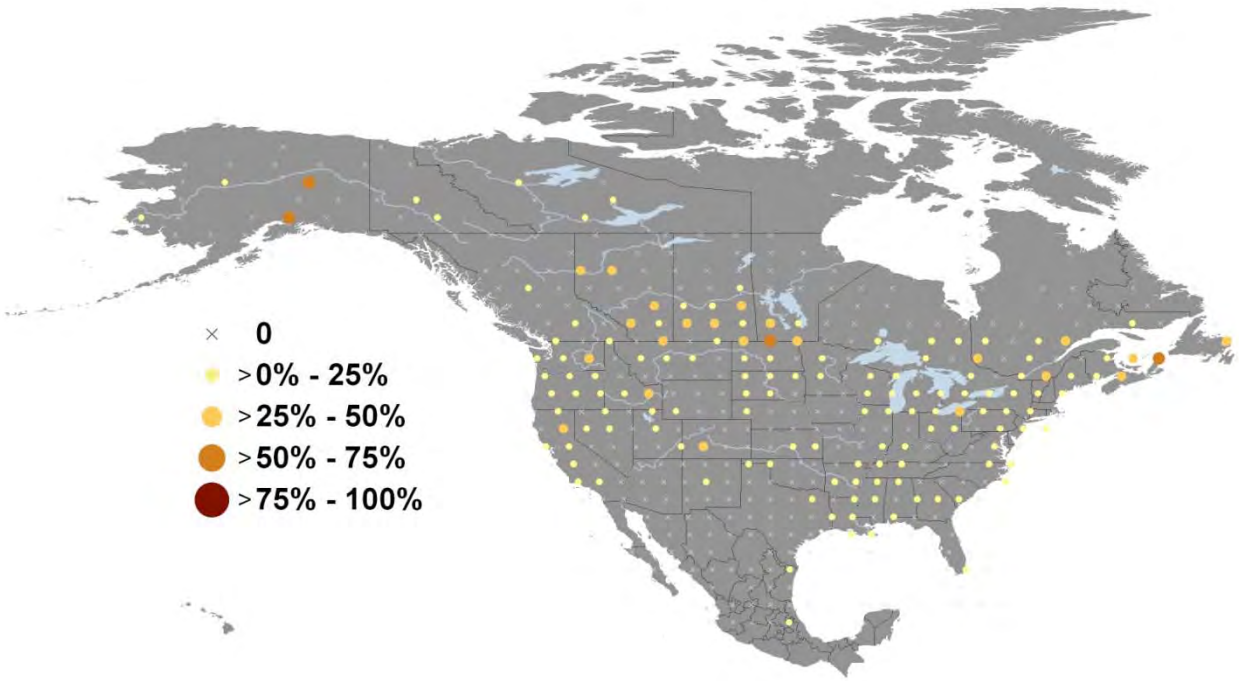


Figure 4.4: The minimum spanning tree (MST) for dabblers harvested from 2002 thru 2006. The larger the circle, the more often that node was selected for the MST (larger circles indicate a larger MST statistic and are expressed as quartiles).

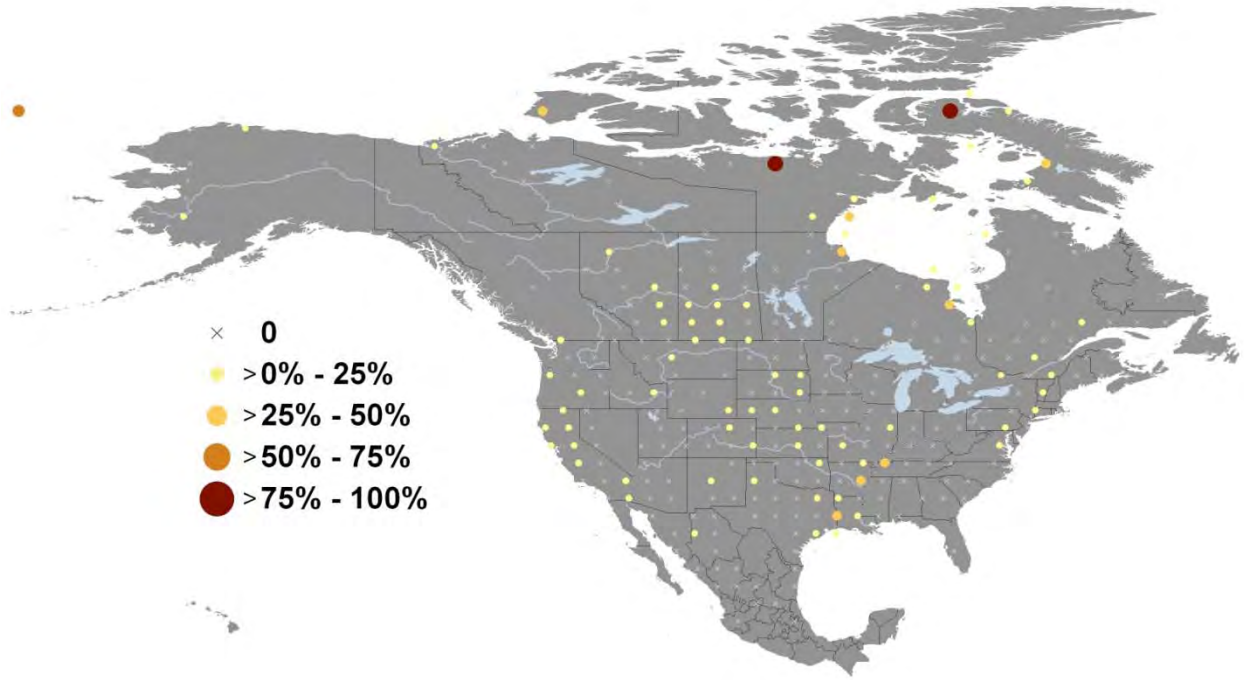


Figure 4.5: The minimum spanning tree (MST) for Mallards harvested from 2002 thru 2006. The larger the circle, the more often that node was selected for the MST (larger circles indicate a larger MST statistic and are expressed as quartiles).

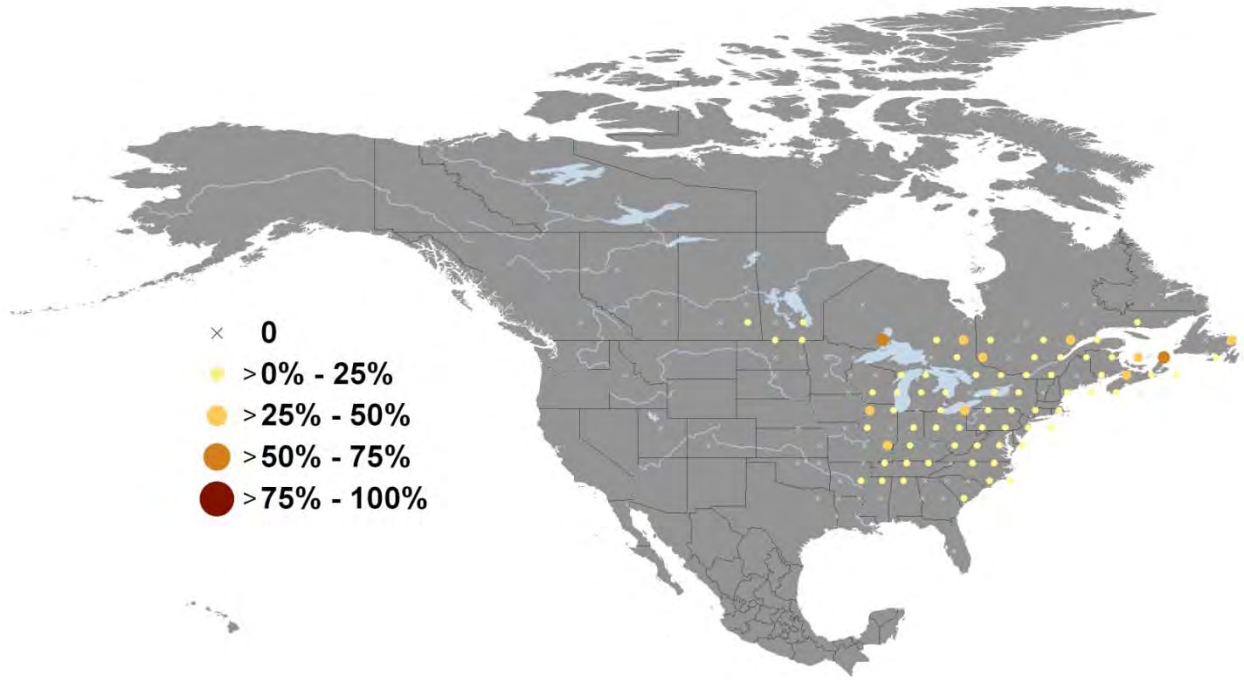


Figure 4.6: The minimum spanning tree (MST) for American Black Ducks harvested from 2002 thru 2006. The larger the circle, the more often that node was selected for the MST (larger circles indicate a larger MST statistic and are expressed as quartiles).

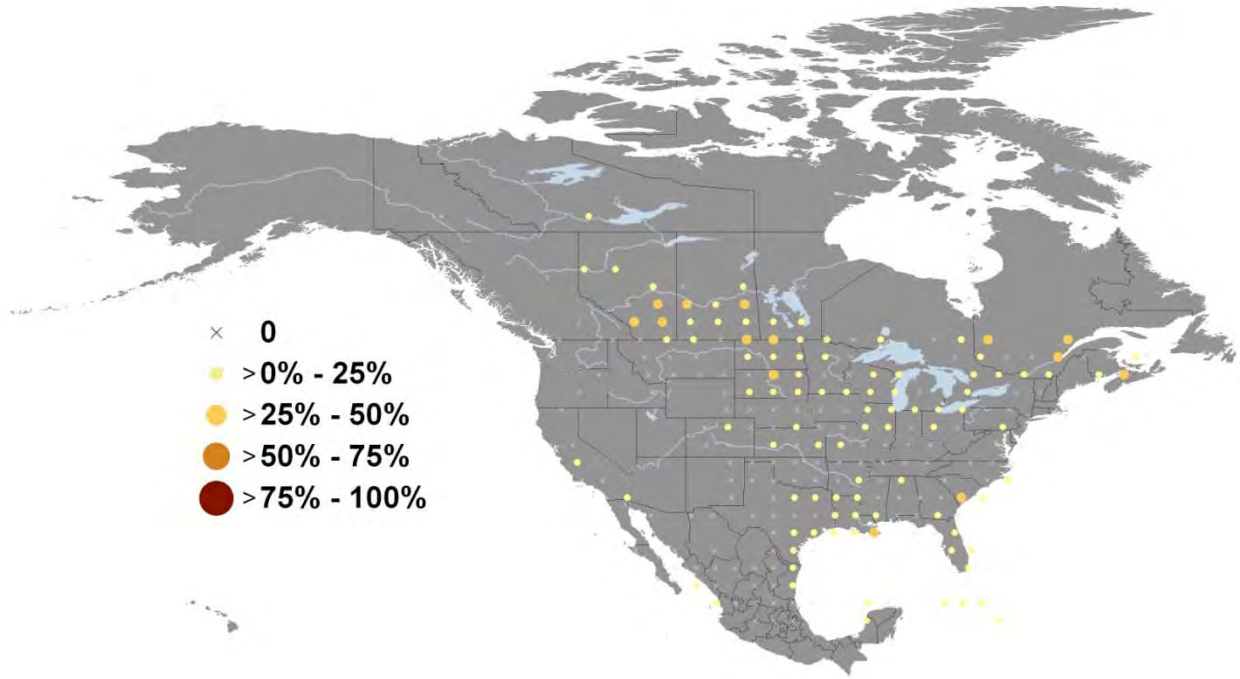


Figure 4.7: The minimum spanning tree (MST) for Blue-winged Teal harvested from 2002 thru 2006. The larger the circle, the more often that node was selected for the MST (larger circles indicate a larger MST statistic and are expressed as quartiles).

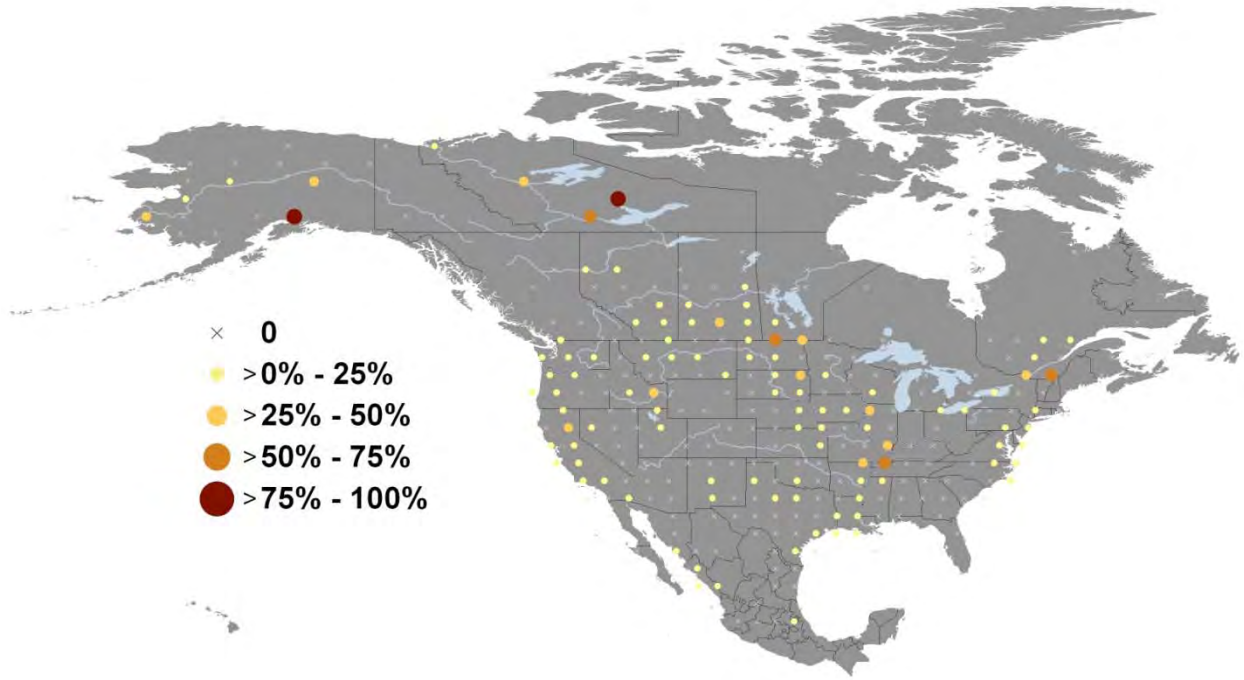


Figure 4.8: The minimum spanning tree (MST) for Northern Pintails harvested from 2002 thru 2006. The larger the circle, the more often that node was selected for the MST (larger circles indicate a larger MST statistic and are expressed as quartiles).

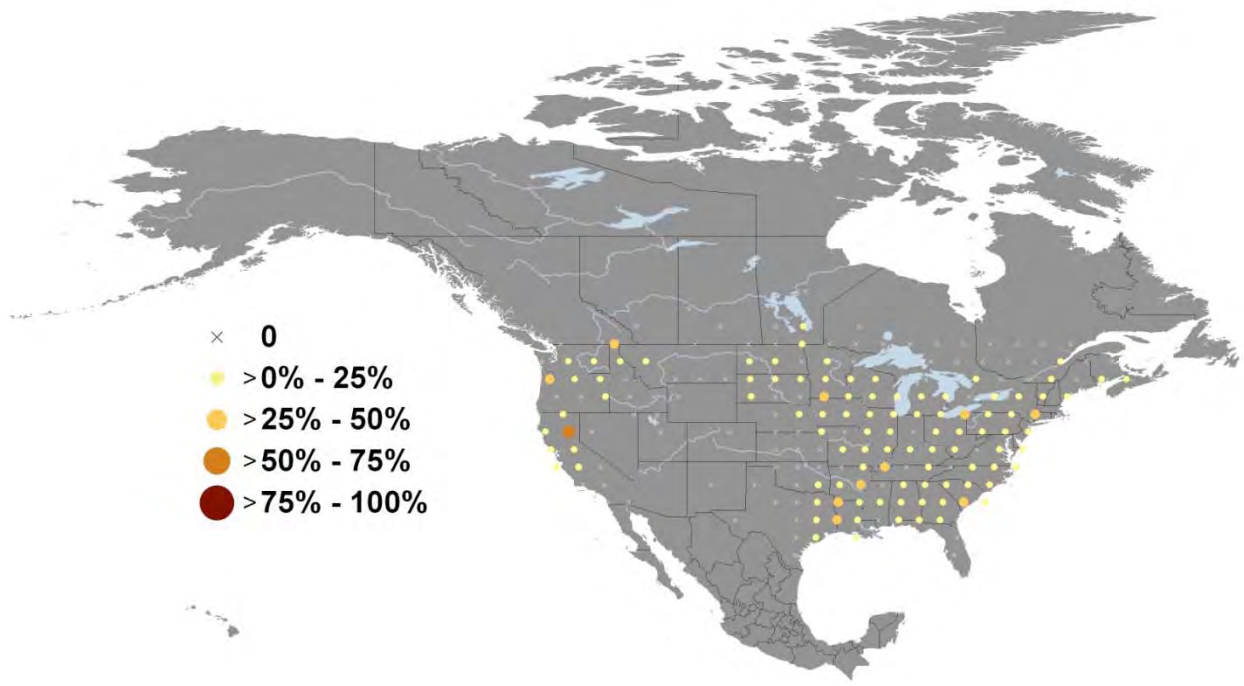


Figure 4.9: The minimum spanning tree (MST) for Wood Ducks harvested from 2002 thru 2006. The larger the circle, the more often that node was selected for the MST (larger circles indicate a larger MST statistic and are expressed as quartiles).

Chapter 5 Clustering and Satellite Analysis

5.1 Motivation

Clusters are subsections of the network that have similar movement patterns. In other words, a cluster is a set of nodes among which there is more movement within the set than with between nodes not contained within the cluster. Satellites describe nodes that can have membership in multiple clusters and are likely to be important in movement of birds between clusters.

5.1.1 Cluster metric

We used an annealing algorithm to assign nodes to clusters. The pairwise connectivity between two nodes is the number of birds that were banded at the first node in the pair and harvested at the second. Hence, more birds traveling between two nodes results in strong pairwise connectivity. The algorithm initially identifies the strongest pairwise connection and assigns those two nodes to cluster number one. The algorithm then continues on to the next strongest pairwise connection. If one of the nodes in this connection is already assigned to cluster number one, then the second node in the pair will also be assigned to cluster one. If neither node has been assigned a cluster then a second cluster is created. This process is repeated until all nodes have been assigned to a cluster. The algorithm finishes by merging clusters with a mean inter-connectivity that exceeds the mean intra-connectivity. (The mean is defined to be the total number of connections divided by the total number of potential connections available between the two clusters.) We also generated clusters using an alternative algorithm. Clusters were generally similar between the two algorithms, so only the annealing algorithm results are reported here.

5.1.2 Satellite metric

The satellite analysis is done in conjunction with the cluster analysis above. Once nodes have been assigned to their primary clusters as described above, the number of birds that travel between each node and nodes in non-primary clusters are counted. The number of birds moving along these intercluster connections are summed for each node and divided by the total number of birds moving to or from the same node, resulting in a value for the satellite metric. Satellite nodes were first described in social networks where they are referred to as bridging ties (Granovetter 1973, Fernandez and Gould 1994).

5.2 Cluster and Satellite Figures

In the cluster plots, the color of each node corresponds to the cluster it belongs to. Clusters as outlined often align with flyways that have been identified for North American waterfowl. In the satellite plots, the color of each node corresponds to the primary cluster it belongs to, and the size of the circle corresponds to the value of the satellite metric (larger circles indicate nodes with more intercluster connections).

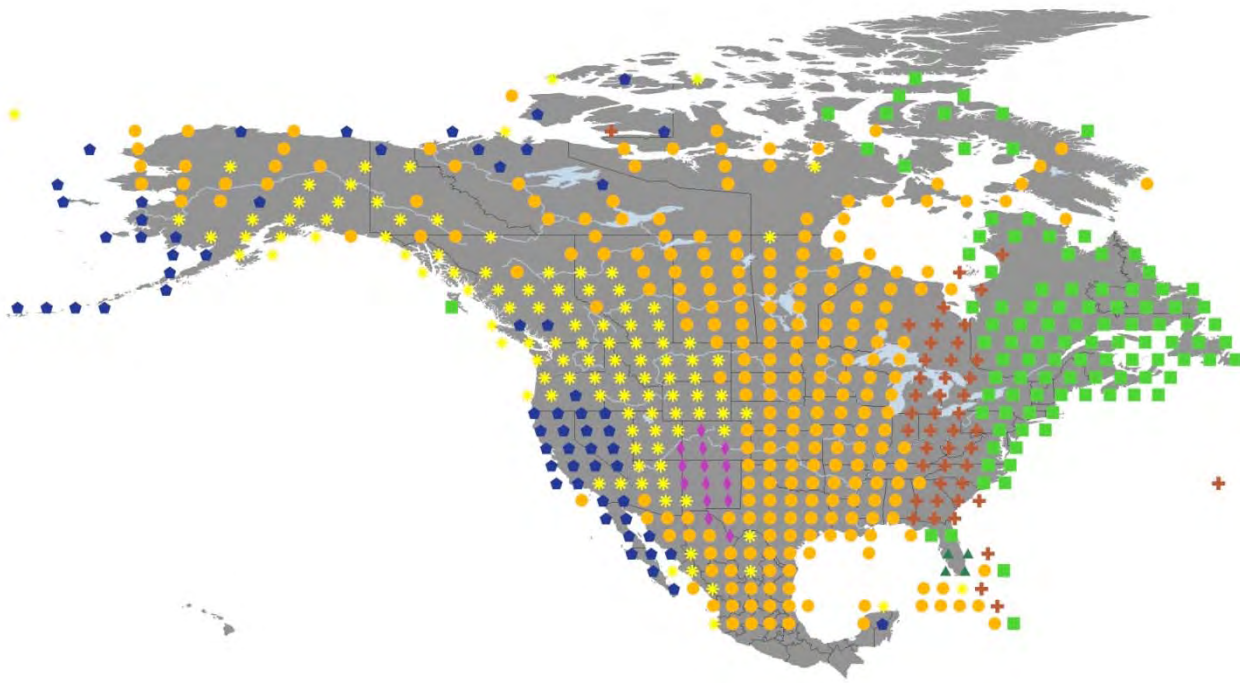


Figure 5.1: Cluster plot for all waterfowl harvested from 2002-2006. Seven clusters, each represented by a different symbol/color combination.

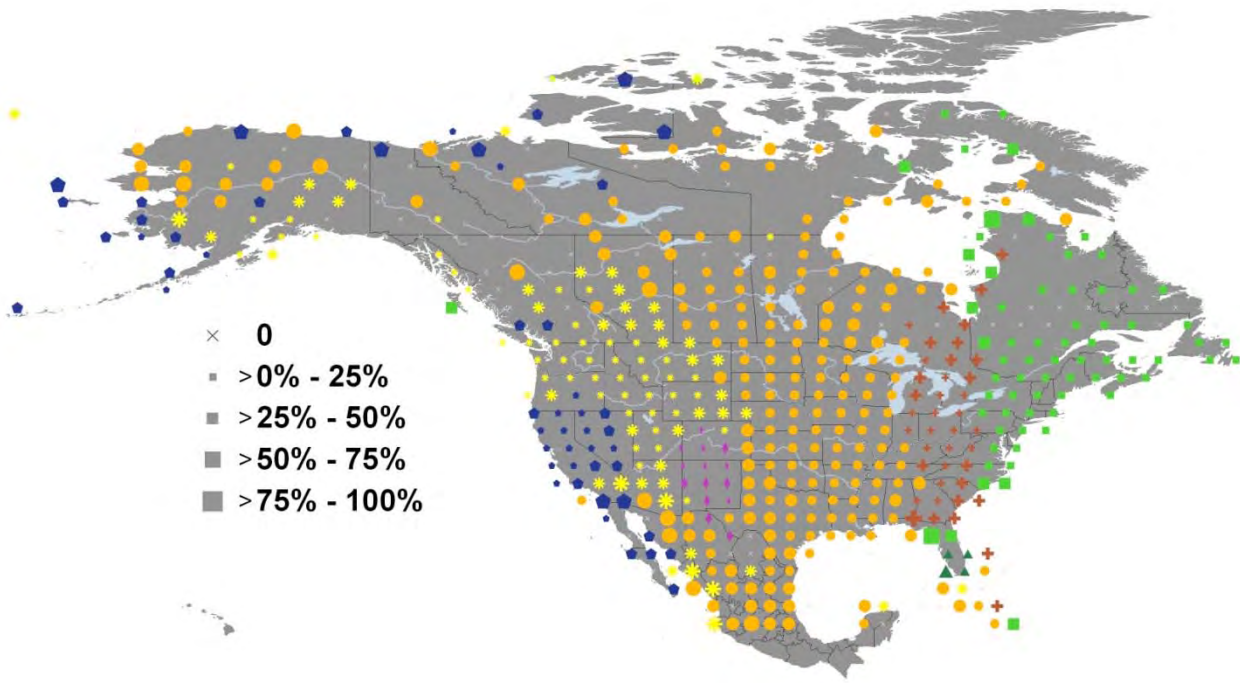


Figure 5.2: Satellite plot for all waterfowl harvested from 2002-2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

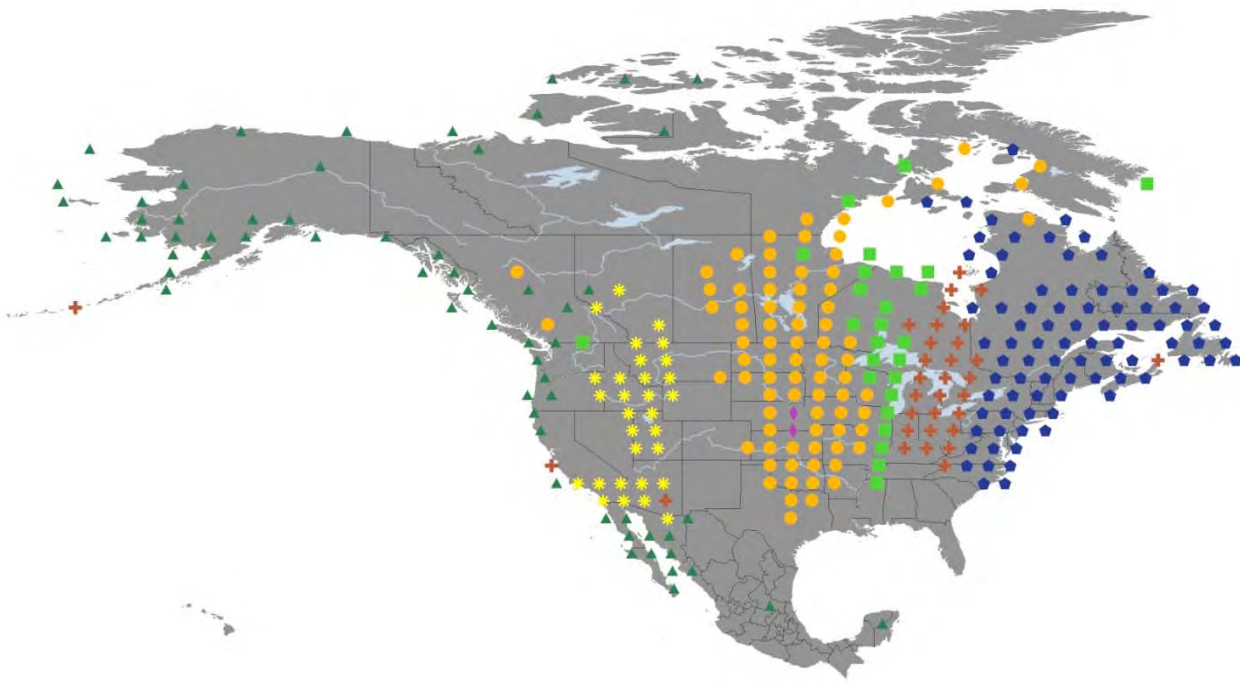


Figure 5.3: Cluster plot for dark geese harvested from 2002-2006. Seven clusters, each represented by a different symbol/color combination.

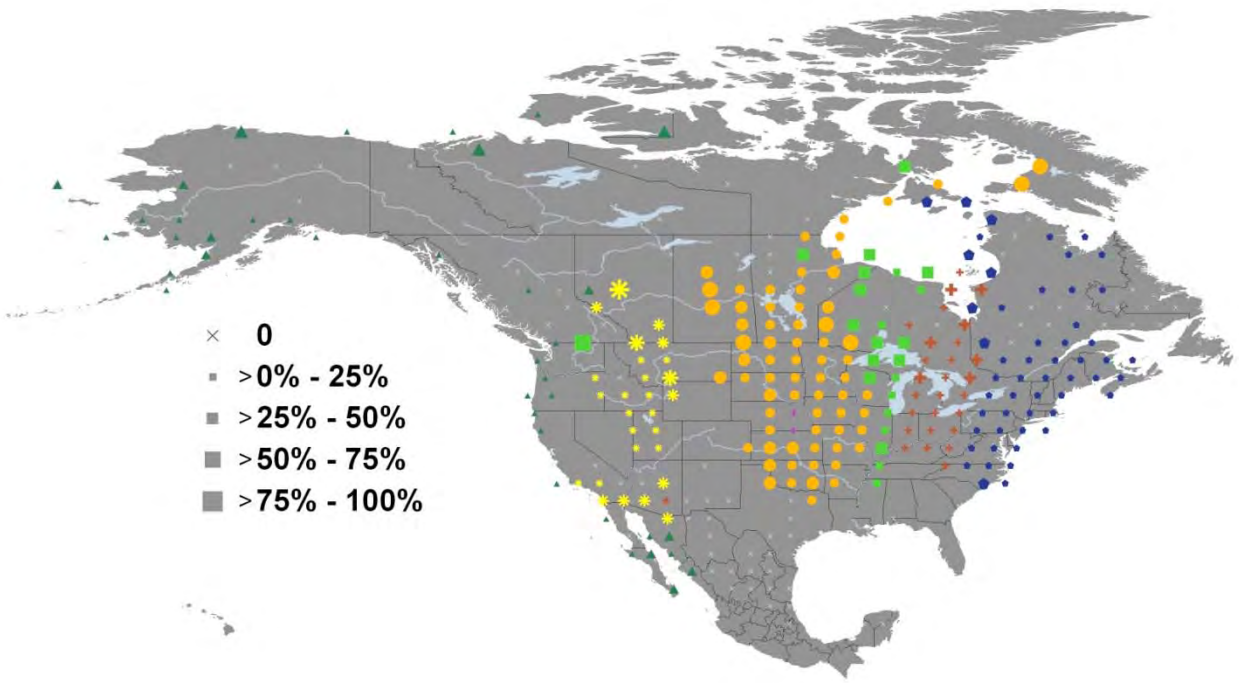


Figure 5.4: Satellite plot for dark geese harvested from 2002-2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

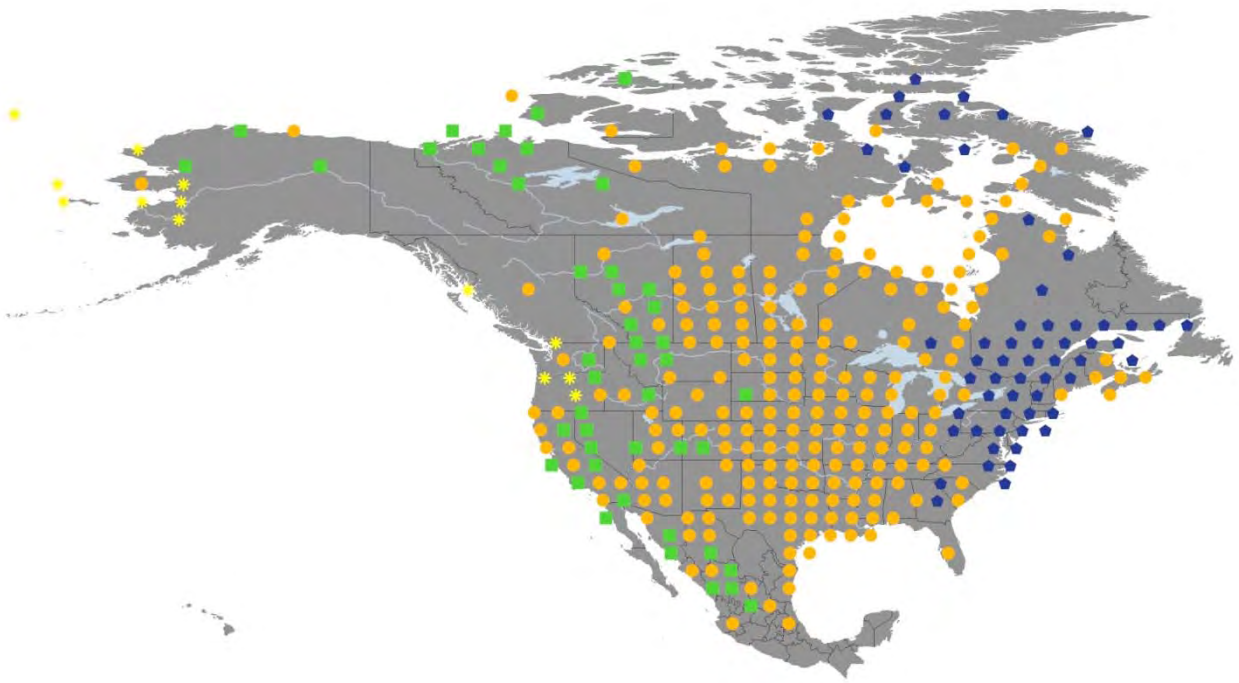


Figure 5.5: Cluster plot for light geese harvested from 2002-2006. Four clusters, each represented by a different symbol/color combination.

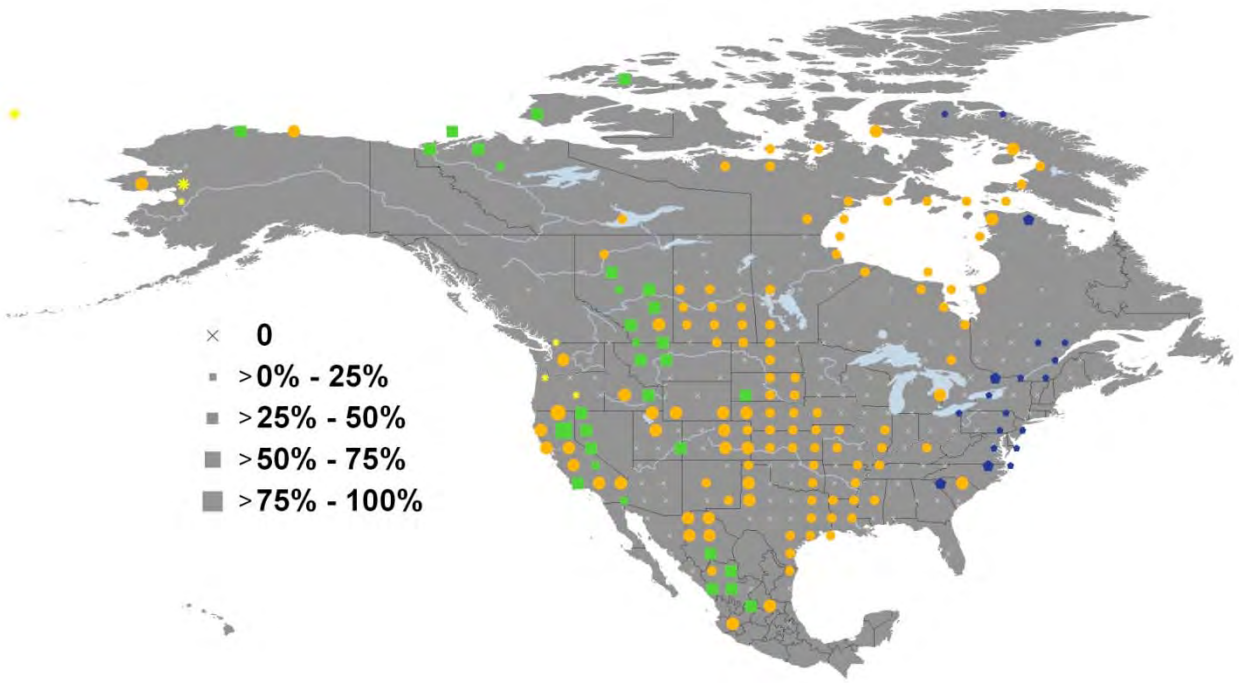


Figure 5.6: Satellite plot for light geese harvested from 2002-2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

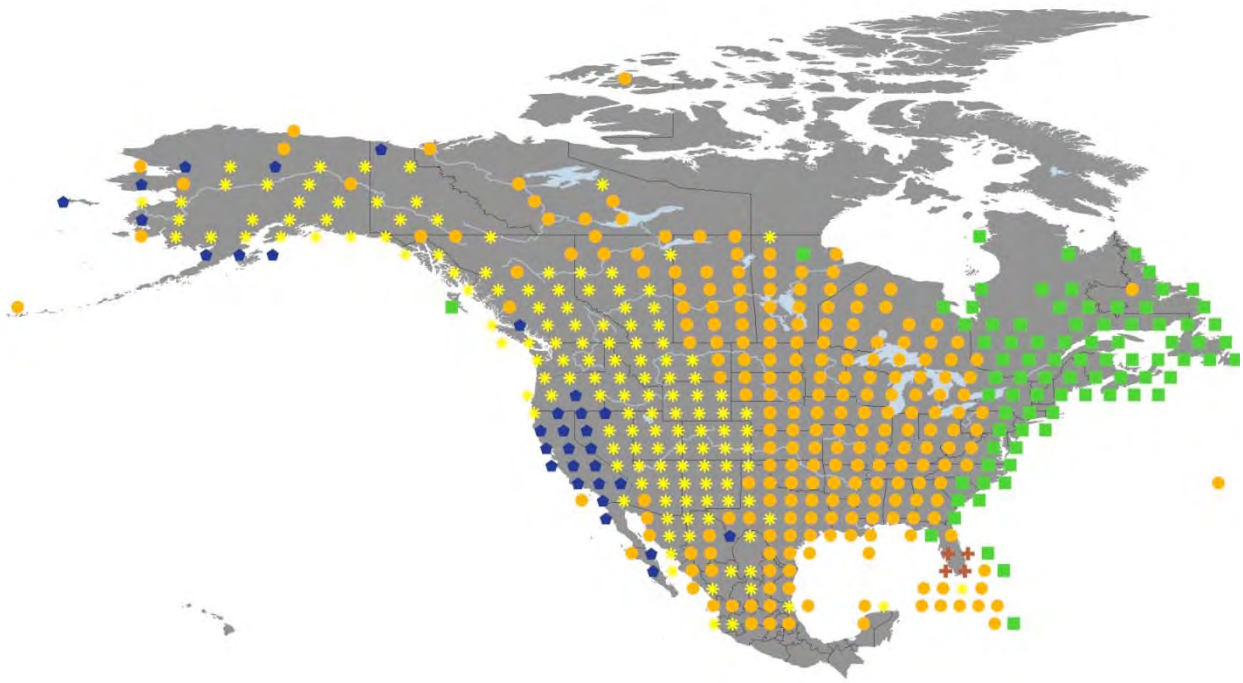


Figure 5.7: Cluster plot for dabblers harvested from 2002-2006. Five clusters, each represented by a different symbol/color combination.

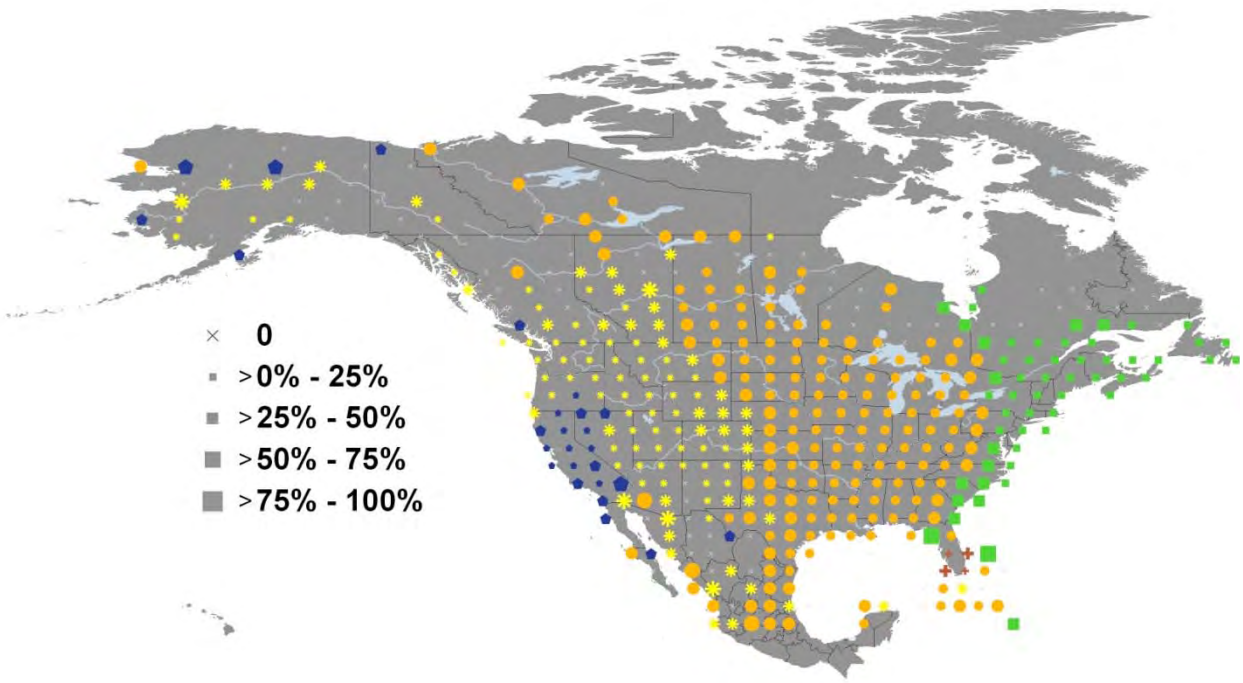


Figure 5.8: Satellite plot for dabblers harvested from 2002-2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

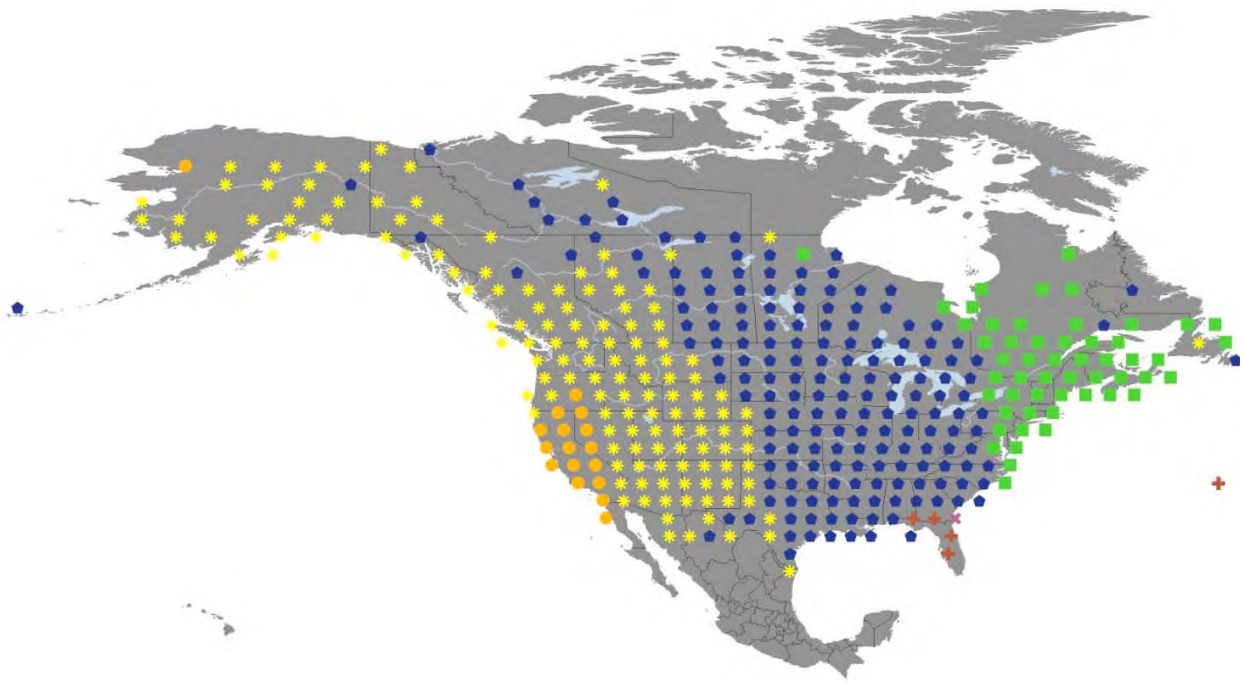


Figure 5.9: Cluster plot for Mallards harvested from 2002-2006. Five clusters, each represented by a different symbol/color combination.

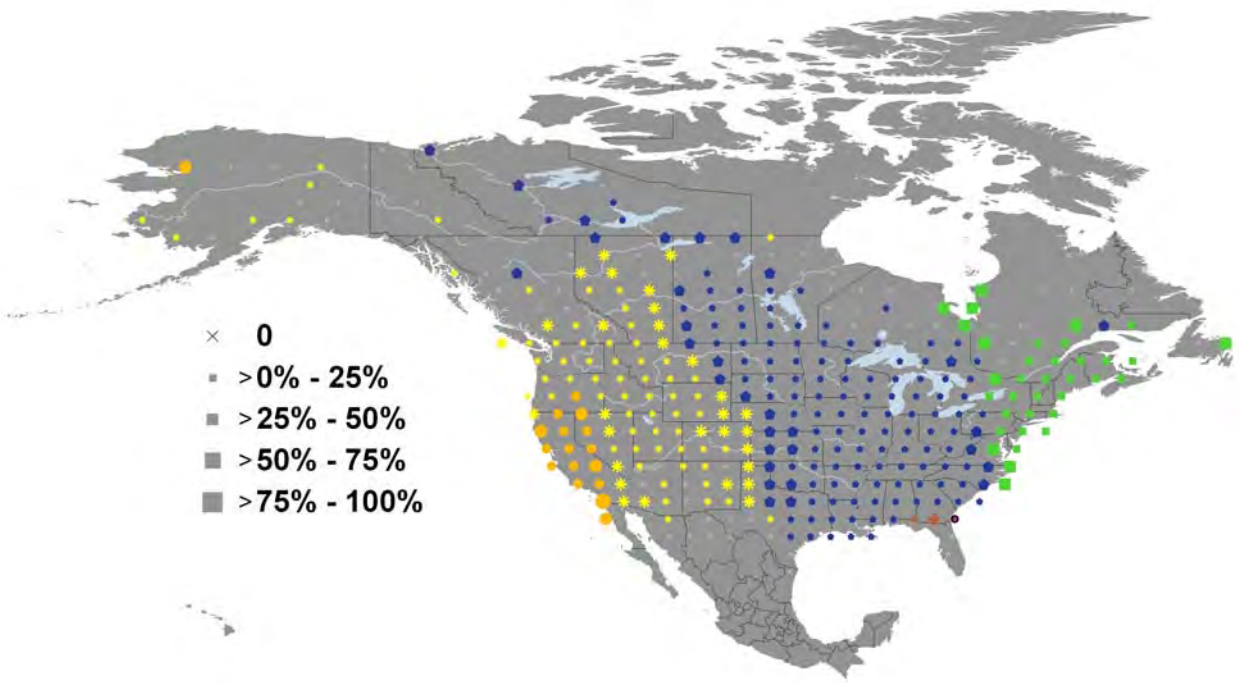


Figure 5.10: Satellite plot for Mallards harvested from 2002-2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

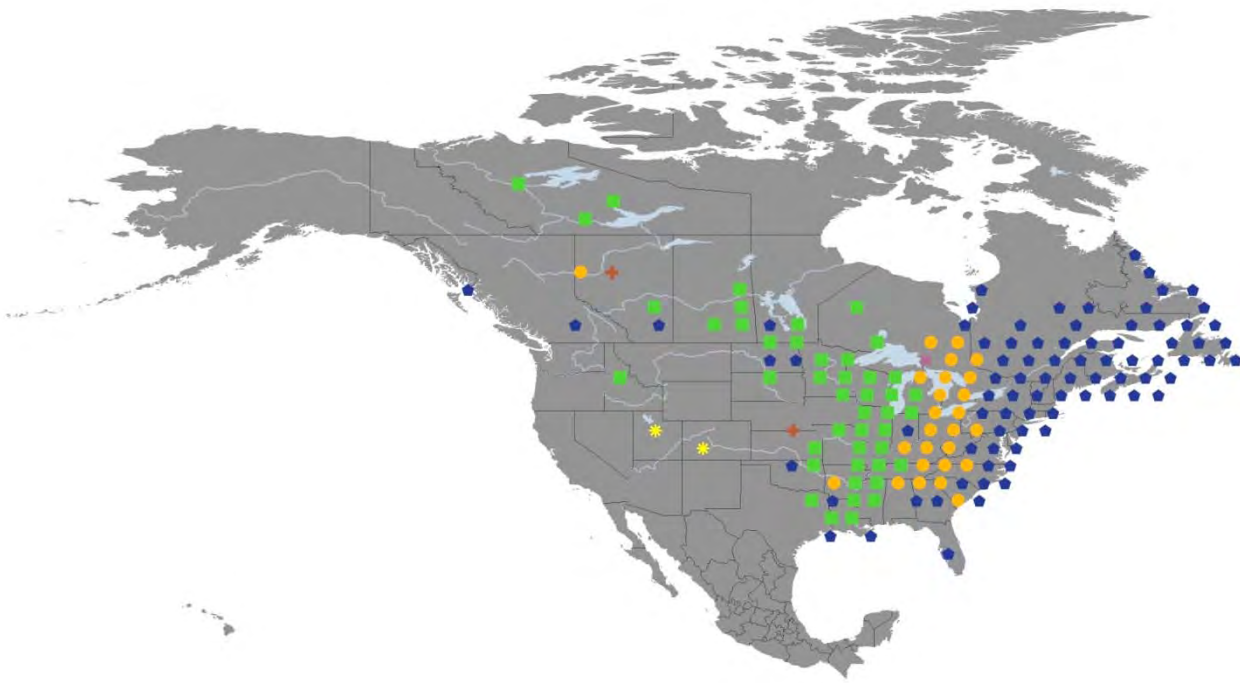


Figure 5.11: Cluster plot for American Black Ducks harvested from 2002-2006. Five clusters, each represented by a different symbol/color combination.

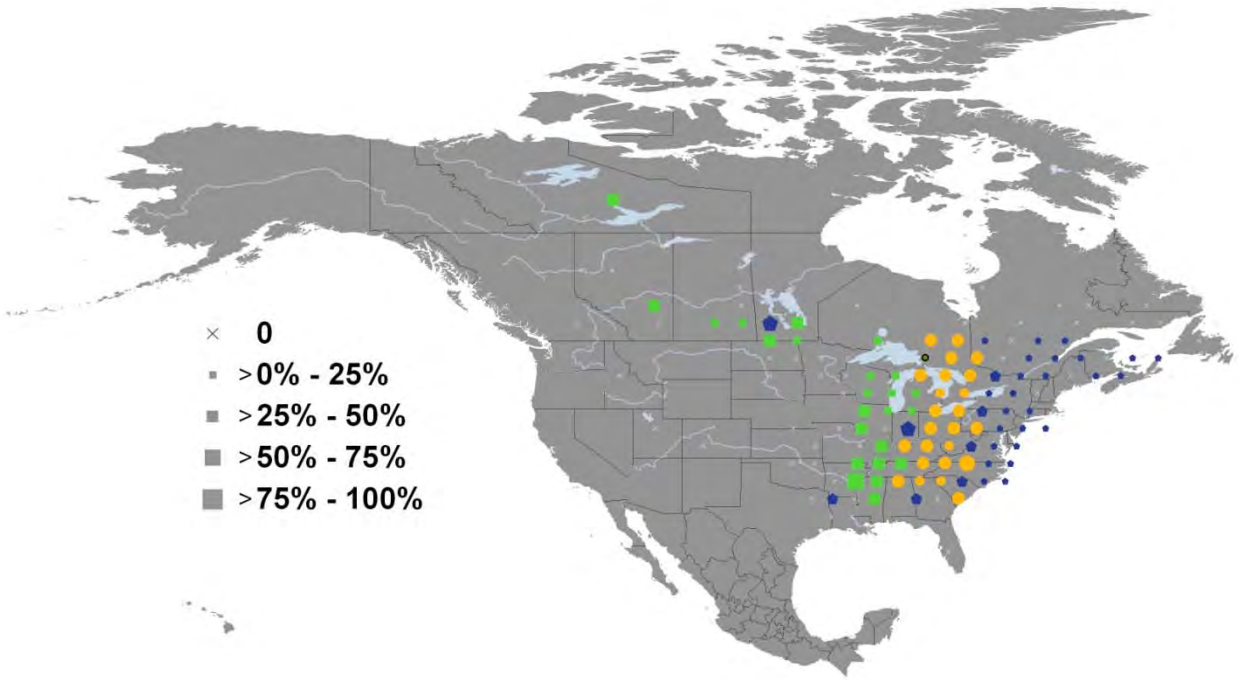


Figure 5.12: Satellite plot for American Black Ducks harvested from 2002-2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

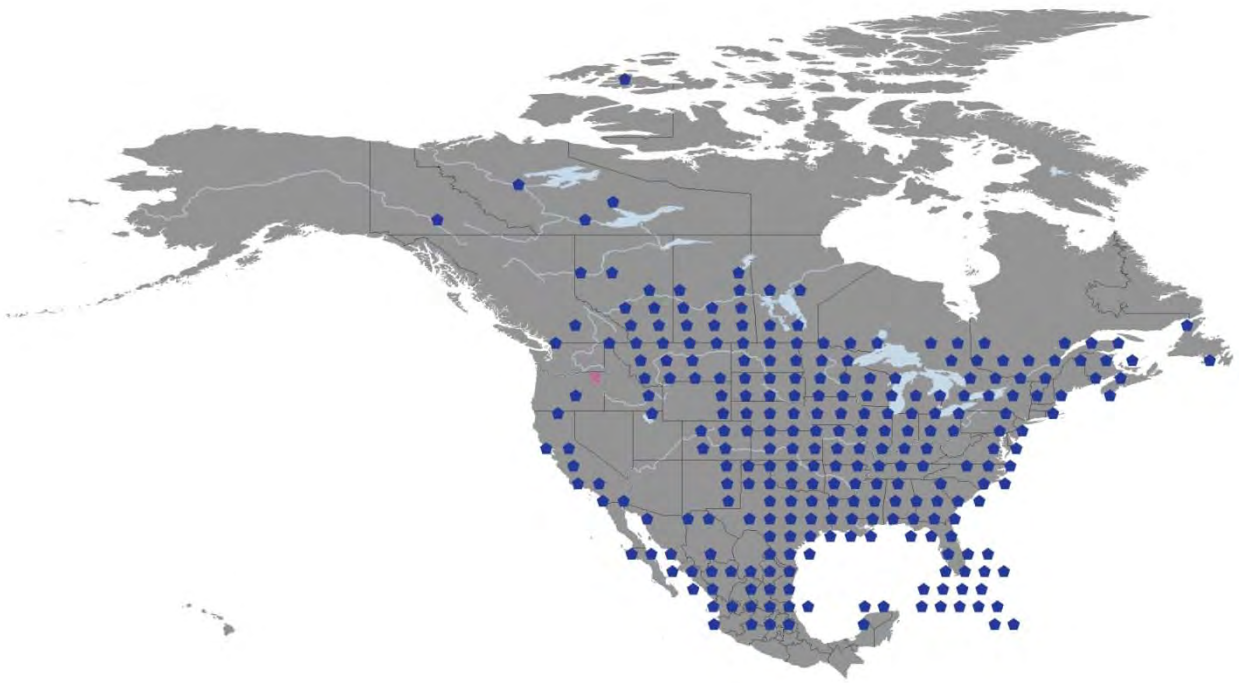


Figure 5.13: Cluster plot for green-winged Teal harvested from 2002-2006. One cluster was generated.

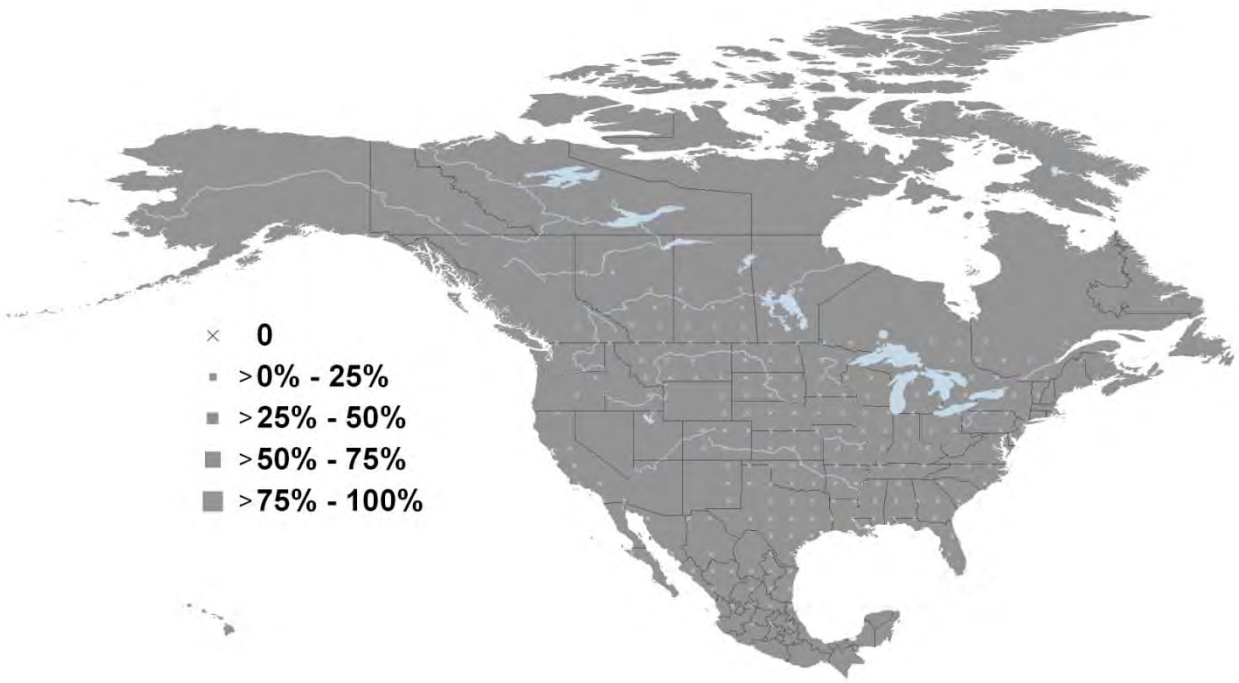


Figure 5.14: Satellite plot for Green-winged Teal harvested from 2002-2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

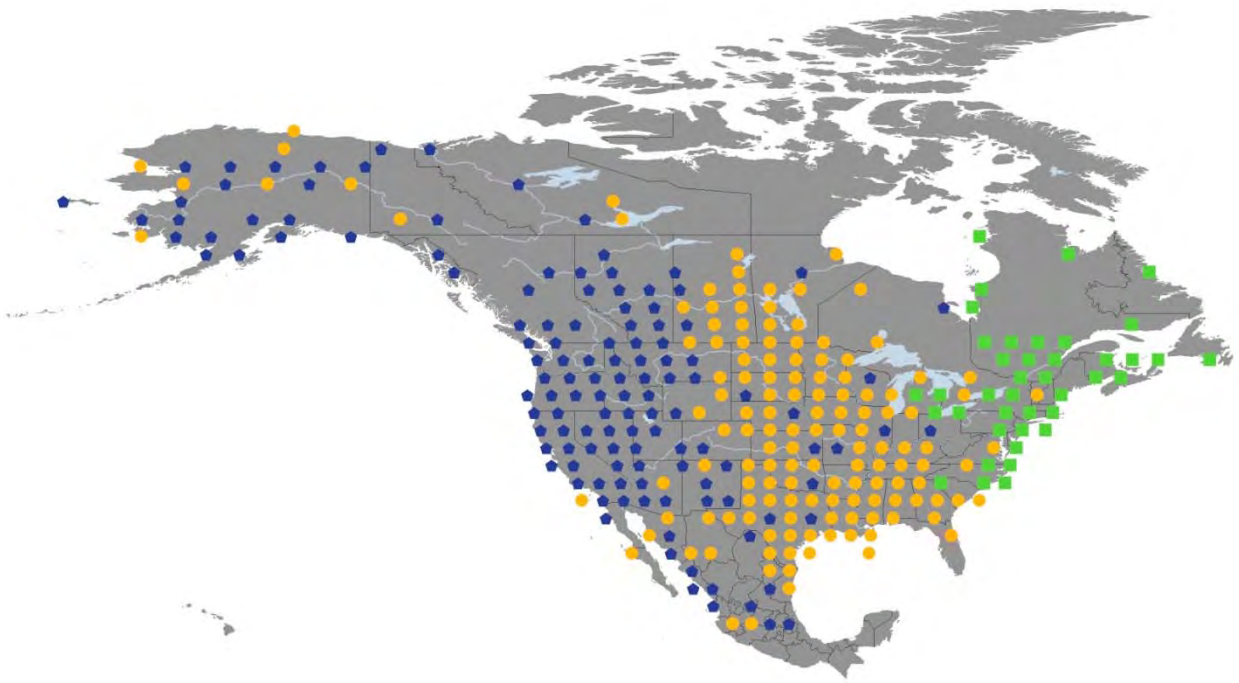


Figure 5.15: Cluster plot for Northern Pintails harvested from 2002-2006. Three clusters, each represented by a different symbol/color combination.

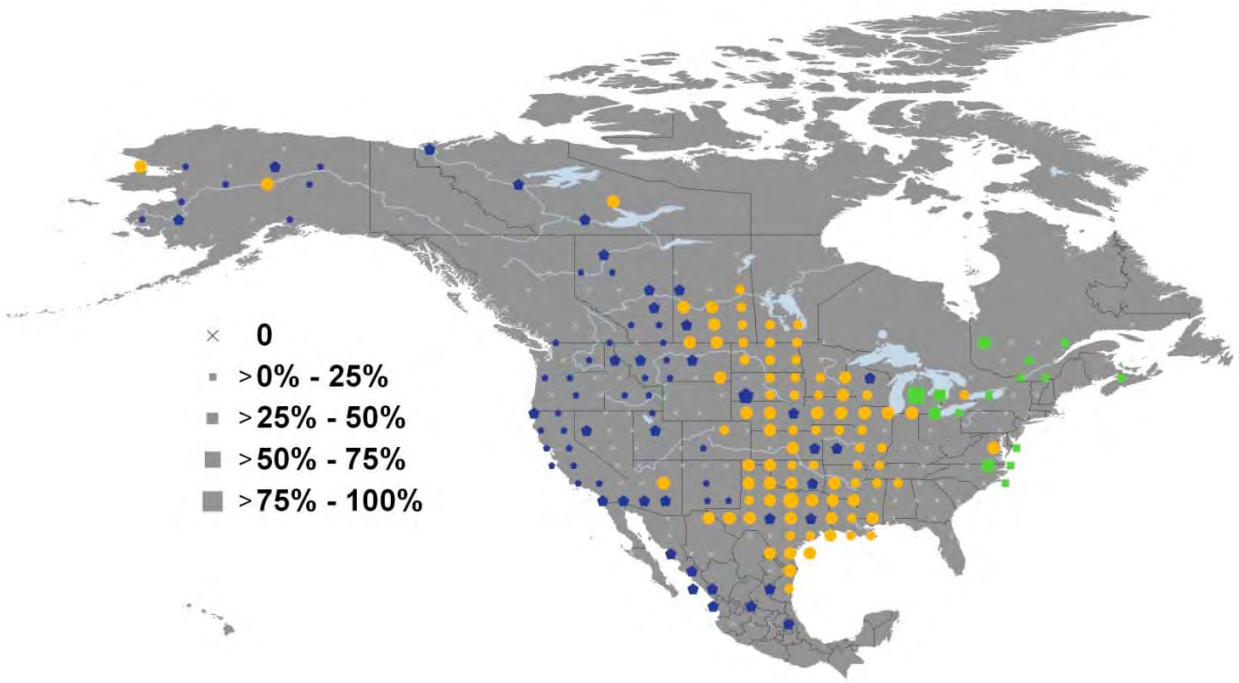


Figure 5.16: Satellite plot for Northern Pintails harvested from 2002-2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

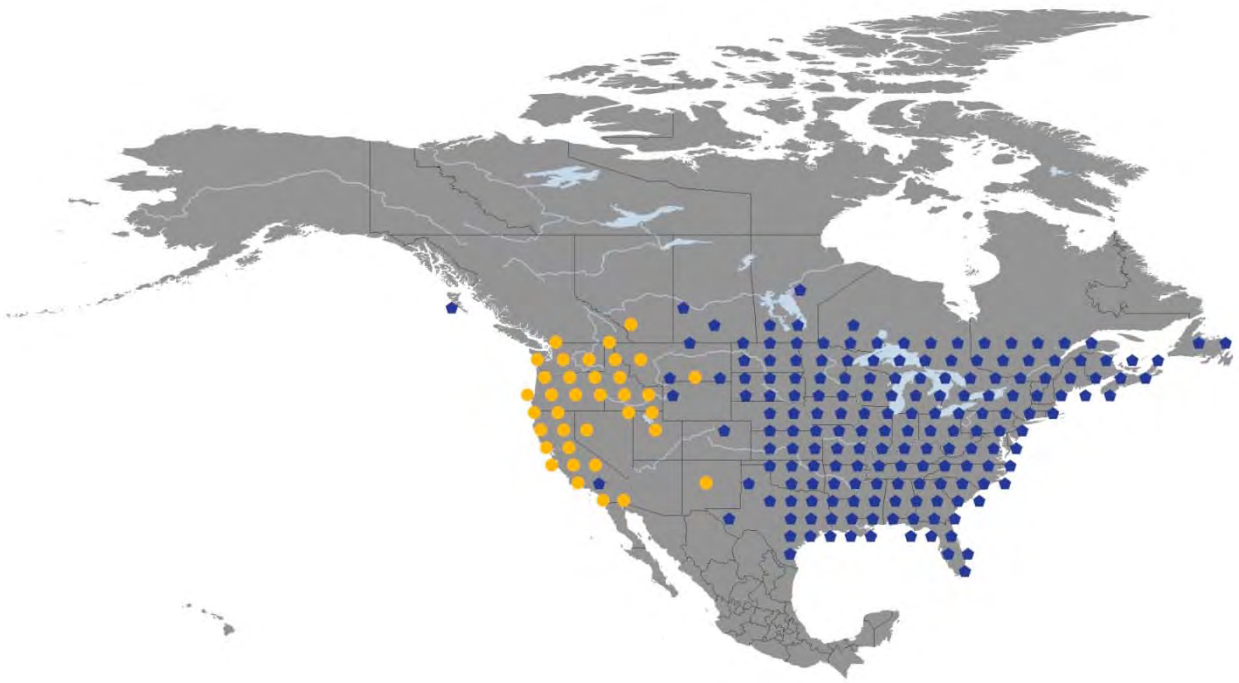


Figure 5.17: Cluster plot for Wood Ducks harvested from 2002-2006. Two clusters, each represented by a different symbol/color combination.

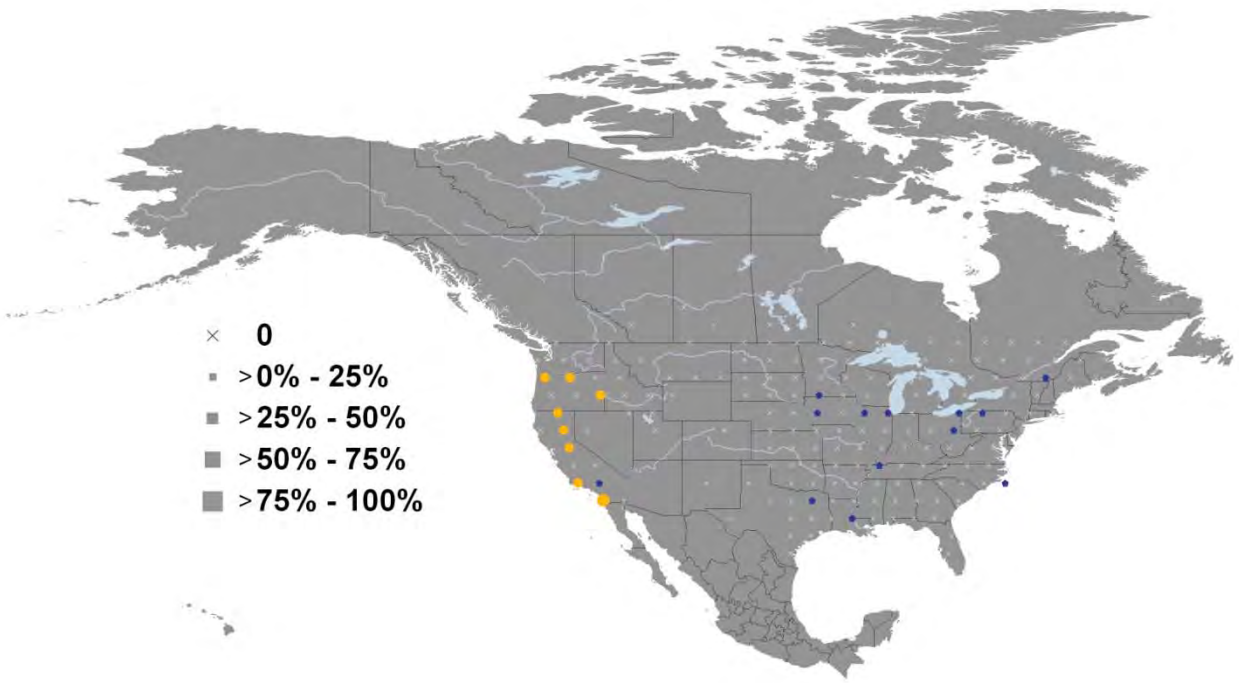


Figure 5.18: Satellite plot for Wood Ducks harvested from 2002-2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

Chapter 6: Understanding the Metrics

6.1 Overview

Each network metric captures connectivity at different levels within the network. The Minimum Spanning Tree describes nodes important in the overall connectivity of the network at the continental scale. The two flow metrics describe nodes with high concentrations of birds and potentially high levels of contact locally. Clusters describe a naturally occurring, intermediate scale in the network, and satellites describe points of connection among clusters. These metrics can be examined alone to investigate the potential for mixing, connectivity and spread of AI at a particular spatial scale. The metrics can also be examined as a whole, in order to determine which locations are important in the introduction and spread of AI across a range of scales.

6.2 Example: Node 1498

Figure 6.1 shows the movement of mallards from where they were banded in Alaska (node 1498) to where they were eventually recovered. Table 6.1 gives the values of each metric for Mallards at node 1498 (all the harvest data was used for this example). Looking across all of the metrics for one node allows us to gauge the node's importance in the bird migration network. In this instance, the Minimum Spanning Tree value is 1.00, indicating that this node was necessary to define a minimum spanning tree. Hence node 1498 is very important with respect to maintaining connectivity of the network. More than 230 mallards originally banded at node 1498 were recovered between 2002 and 2006 whereas less than 30 banded mallards were recovered at node 1498 over the same five-year period. Since the total number of banded mallards recovered throughout North America between 2002 and 2006 exceeds 106,000, the relative flows, both banding and recovery, are small. Node 1498 resides in the flyway that spans western North America, which includes many nodes in Alaska, the Pacific Northwest, and the interior Western United States. The satellite value for this node is 0.197 indicating that there is roughly a 20% probability of a bird from this node either coming from or going to another cluster.

Table 6.1: Metric values for Mallards for node numbers 1498 and 868.

Node	Location	MST	Banding flow	Recovery flow	Cluster	Total Clusters	Satellite
1498	Alaska	0.54	232	28	4	5	0.197
868	New Jersey	0.01	1137	899	3	5	0.061

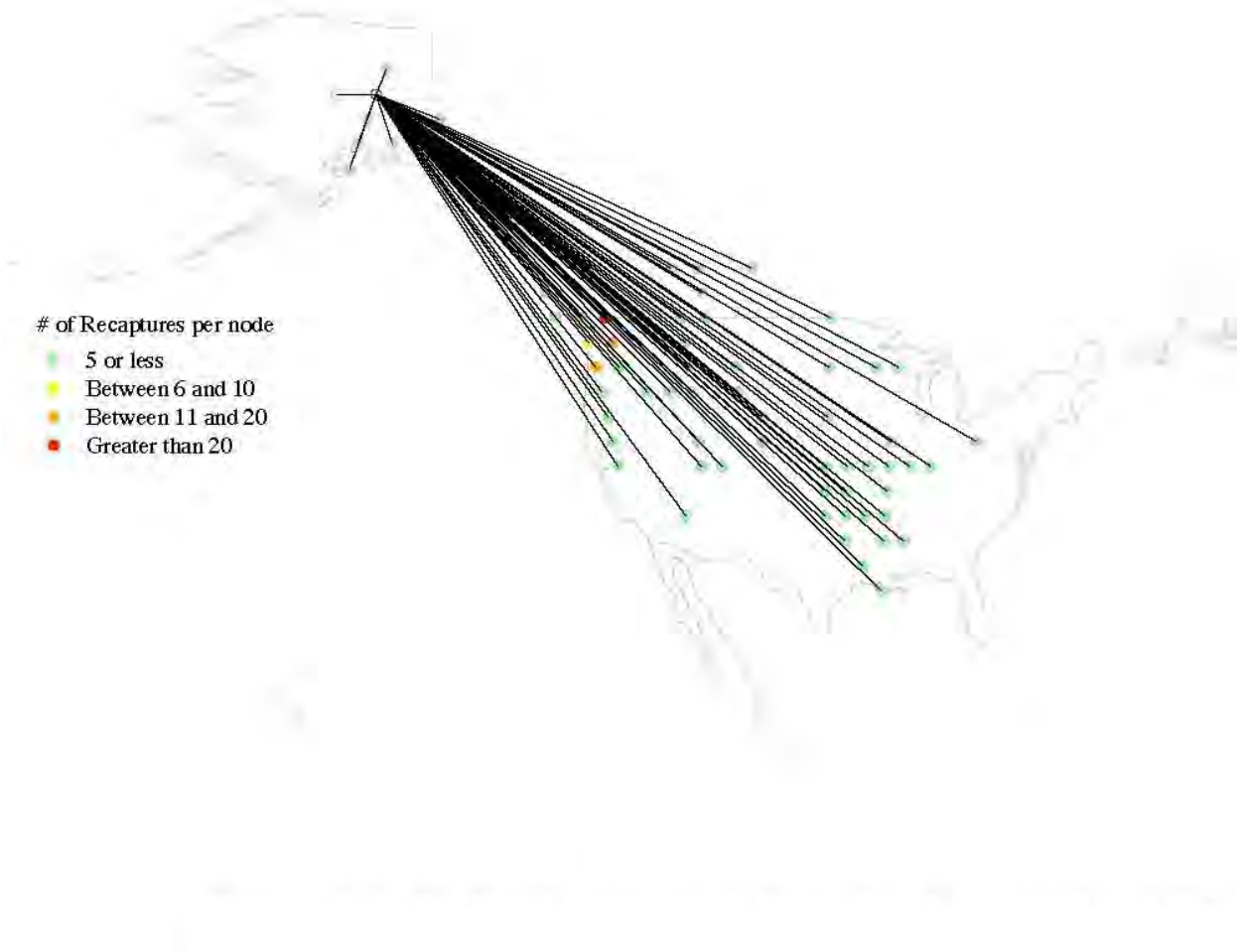


Figure 6.1: Mallard movement from banding station 1498.

6.3 Example: Node 868

Figure 6.2 shows the movement of mallards banded at node 868. Table 6.1 summarizes the values of each metric for all-harvest Mallards at node 868. The Minimum Spanning Tree value for this node is 0.00, indicating that this node was not included in any of 100 simulated MSTs. Therefore, node 868 can be considered of lesser importance in the overall connectivity of the network. The banding flow, approximately 1140, indicates that approximately 1% (1140/106,000) of all recovered mallards are harvested at node 868. Node 868 is within a cluster, which includes New England, New York to Maryland, and the eastern provinces of Canada. The satellite value for this node is 0.061, indicating that there is a 6% probability of mixing with other of a bird from this node either coming from or going to another cluster.

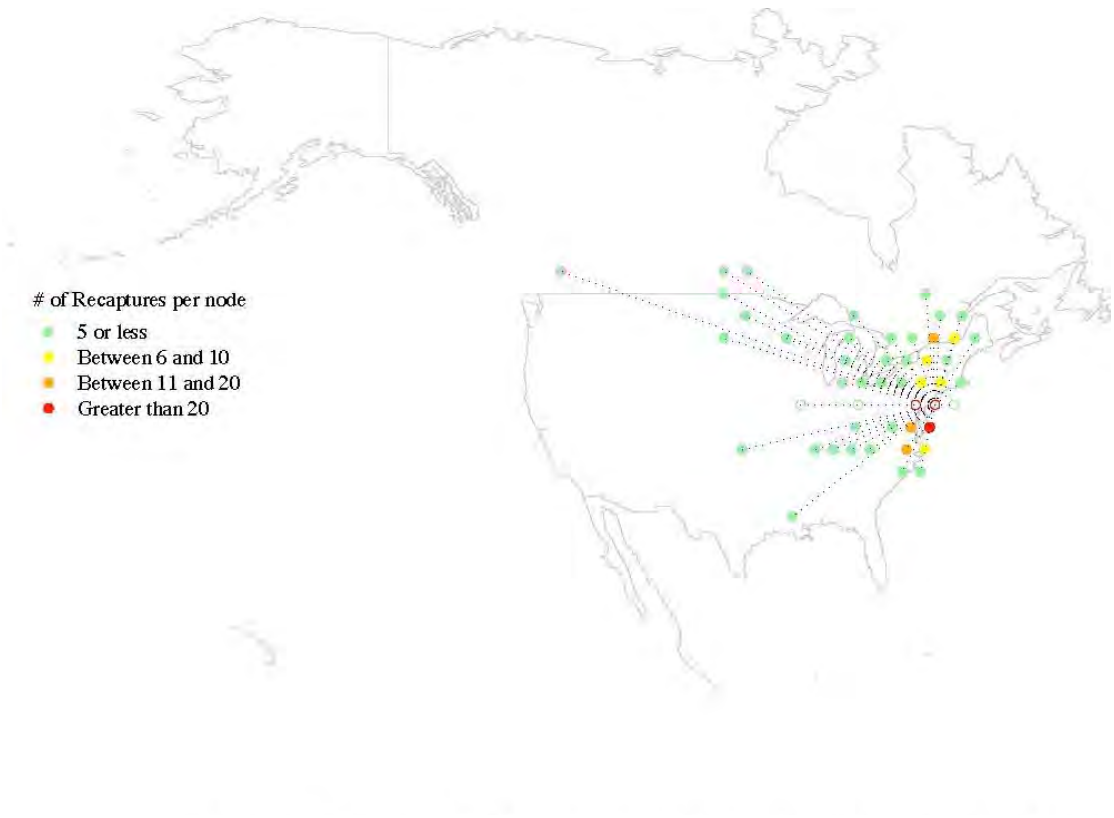


Figure 6.2: Mallard movement from banding station 868.

Chapter 7 Direct Harvest to Identify Flow, MST, Clusters, and Satellites

7.1 Overview

The following figures were generated using the direct harvest data, i.e., birds that were recovered within 52 weeks of banding in the years 2002-2006. The data thus represents a more immediate, within year, response. In comparing these figures with those in the previous chapters, which relied on all harvested birds during 2002-2006 (e.g., figures 7.1 compared to 3.1, etc.), the patterns for flow and cluster are fairly similar and more variable for MST and satellite metrics.

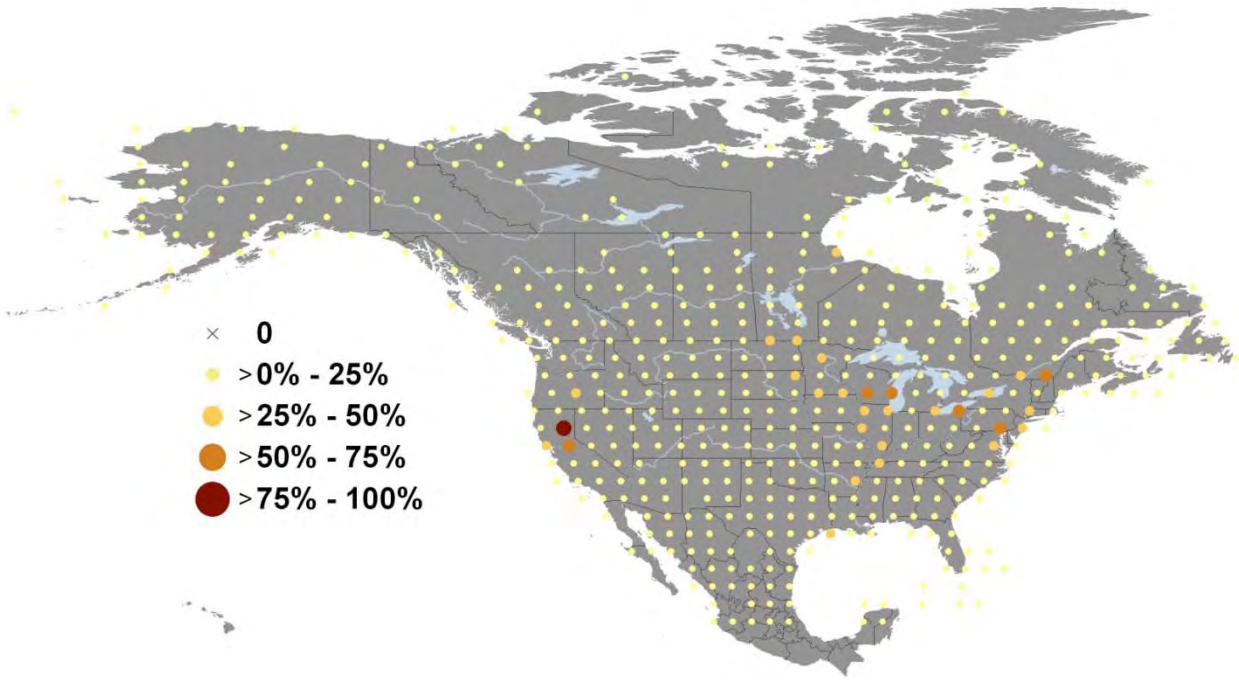


Figure 7.1: Direct harvest network flow for all waterfowl from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

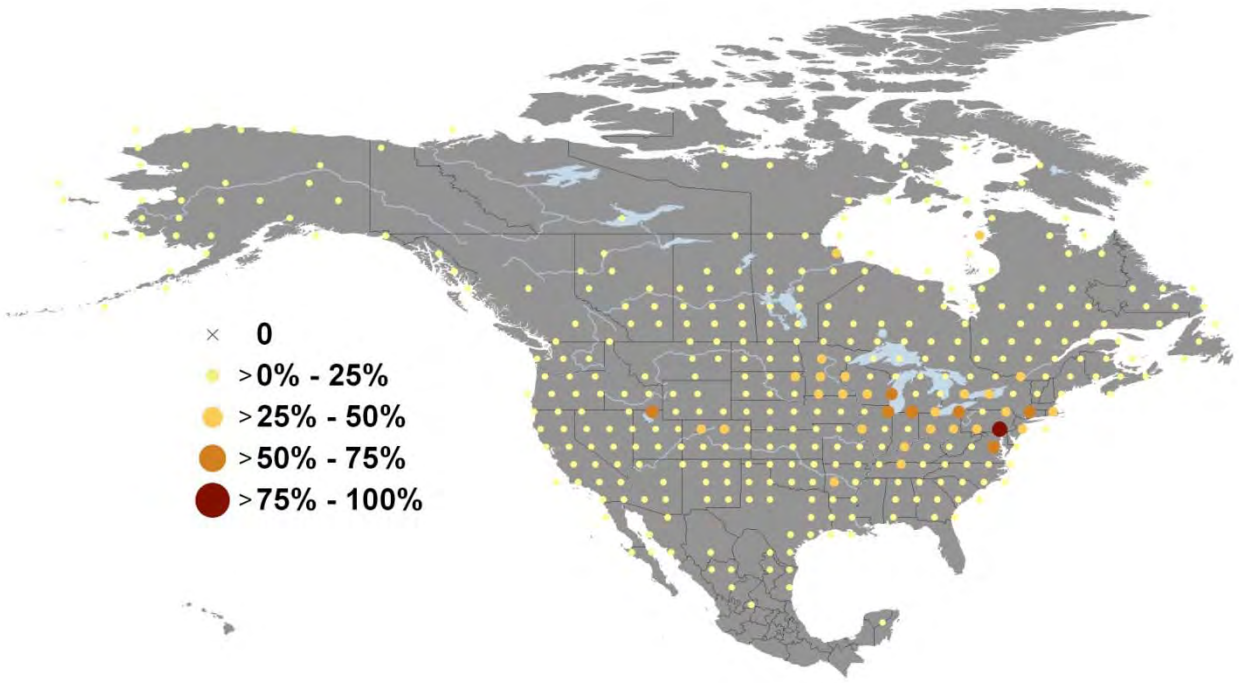


Figure 7.2: Direct harvest network flow for dark geese from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

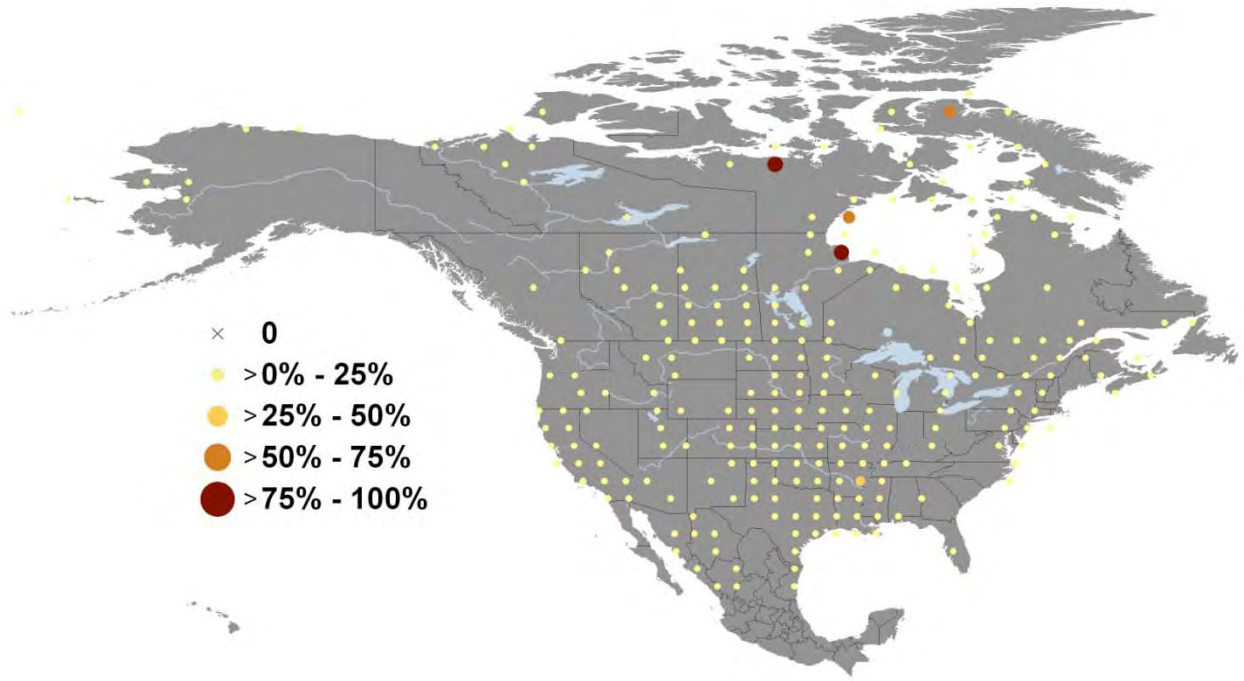


Figure 7.3: Direct harvest network flow for light geese from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

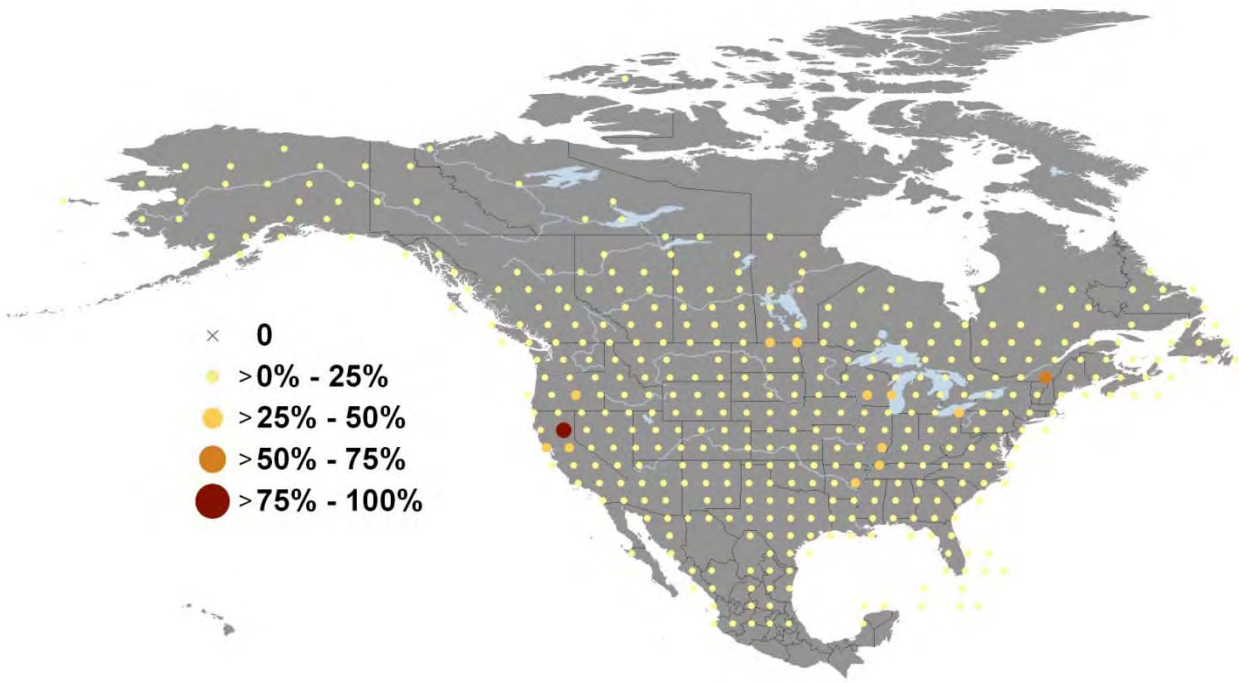


Figure 7.4: Direct harvest network flow for dabblers from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

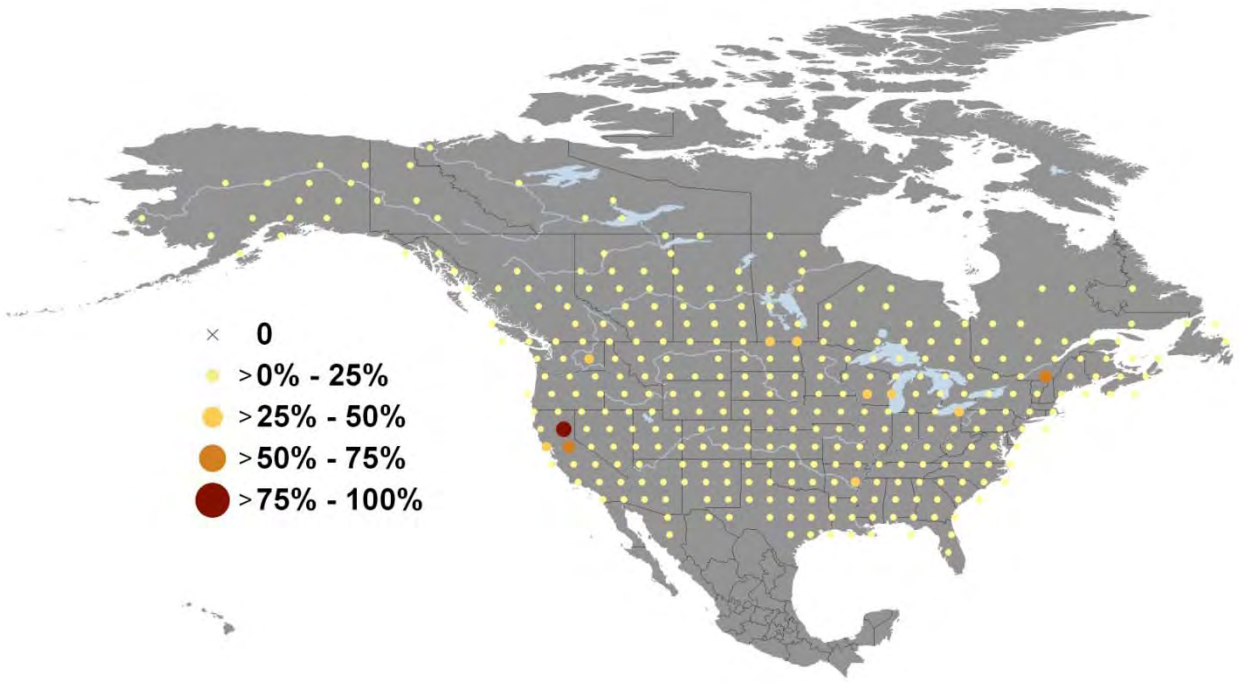


Figure 7.5: Direct harvest network flow for Mallards from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

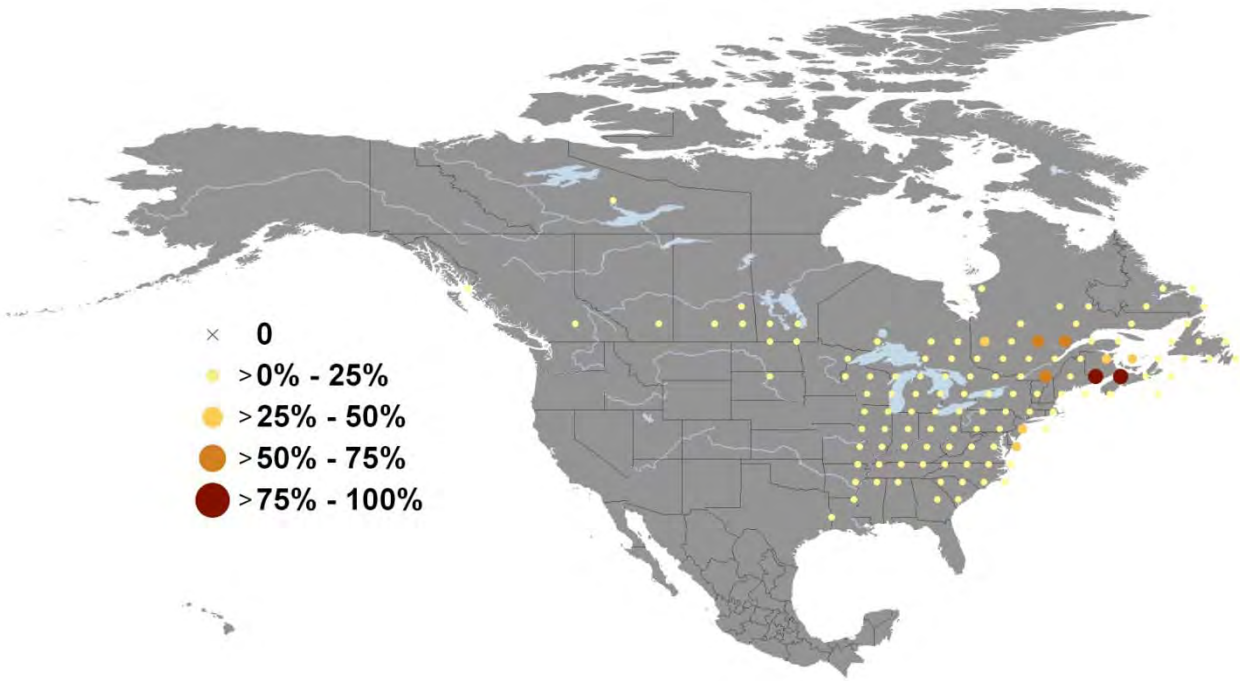


Figure 7.6: Direct harvest network flow for American Black Ducks from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

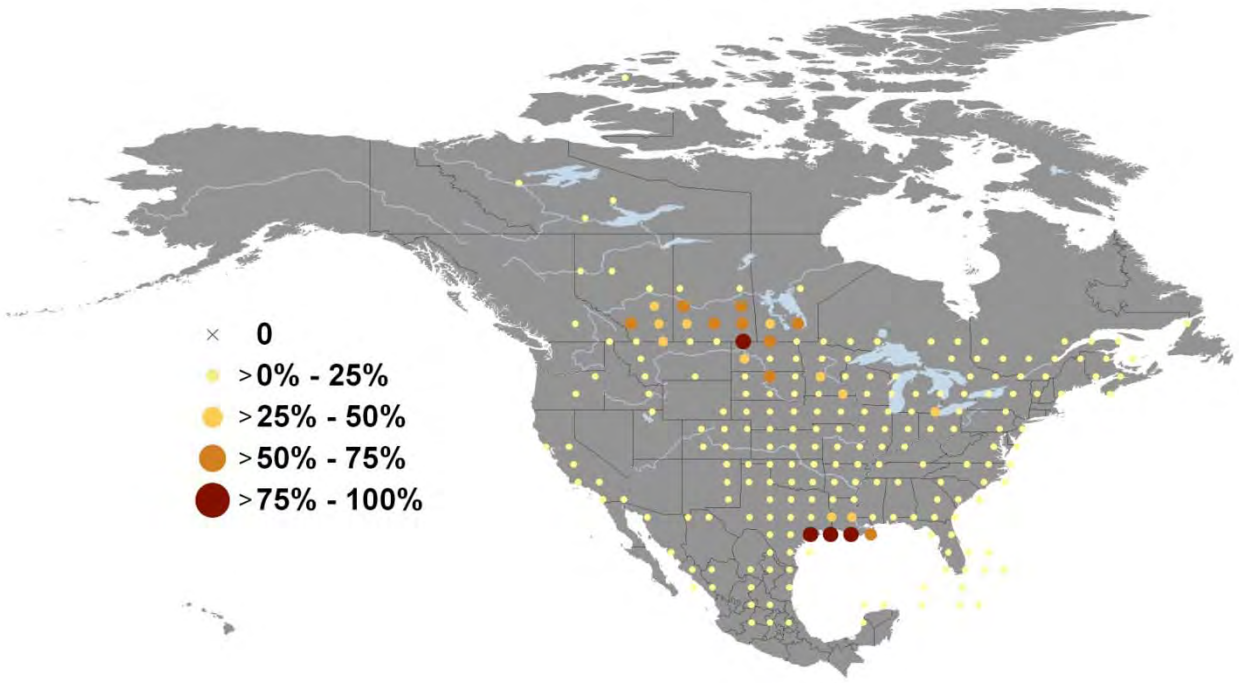


Figure 7.7: Direct harvest network flow for Blue-winged Teal from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

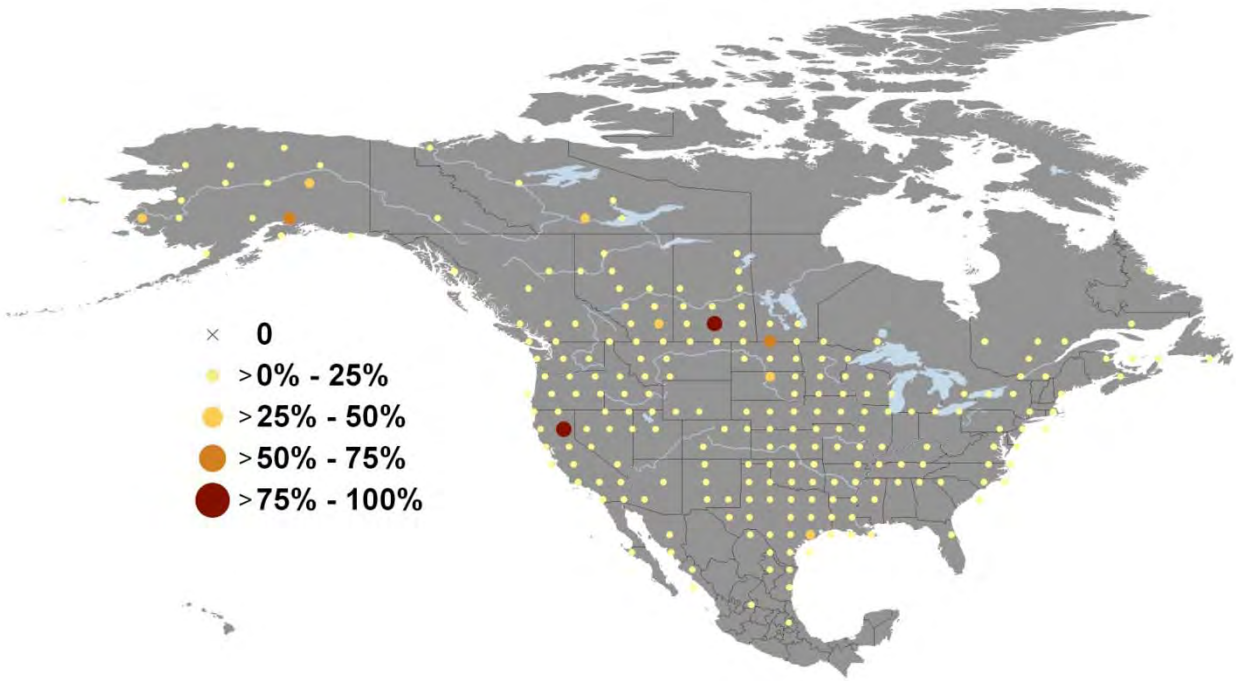


Figure 7.8: Direct harvest network flow for Northern Pintails from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

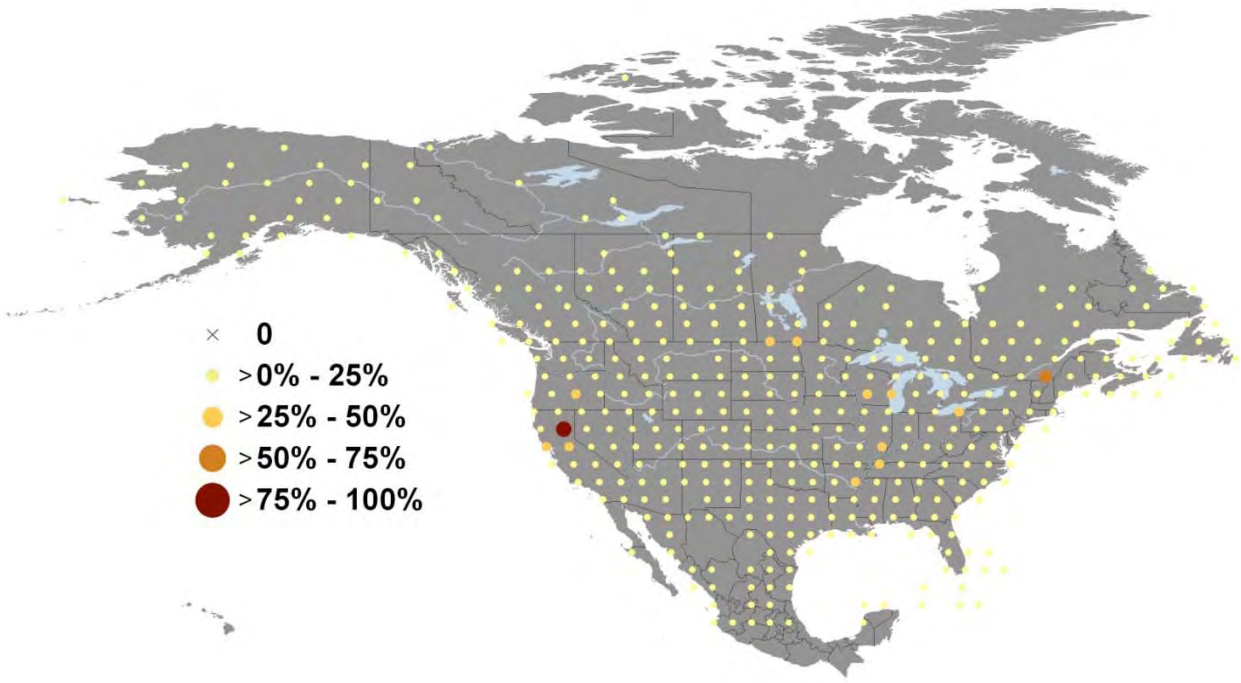


Figure 7.9: Direct harvest network flow for Wood Ducks from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

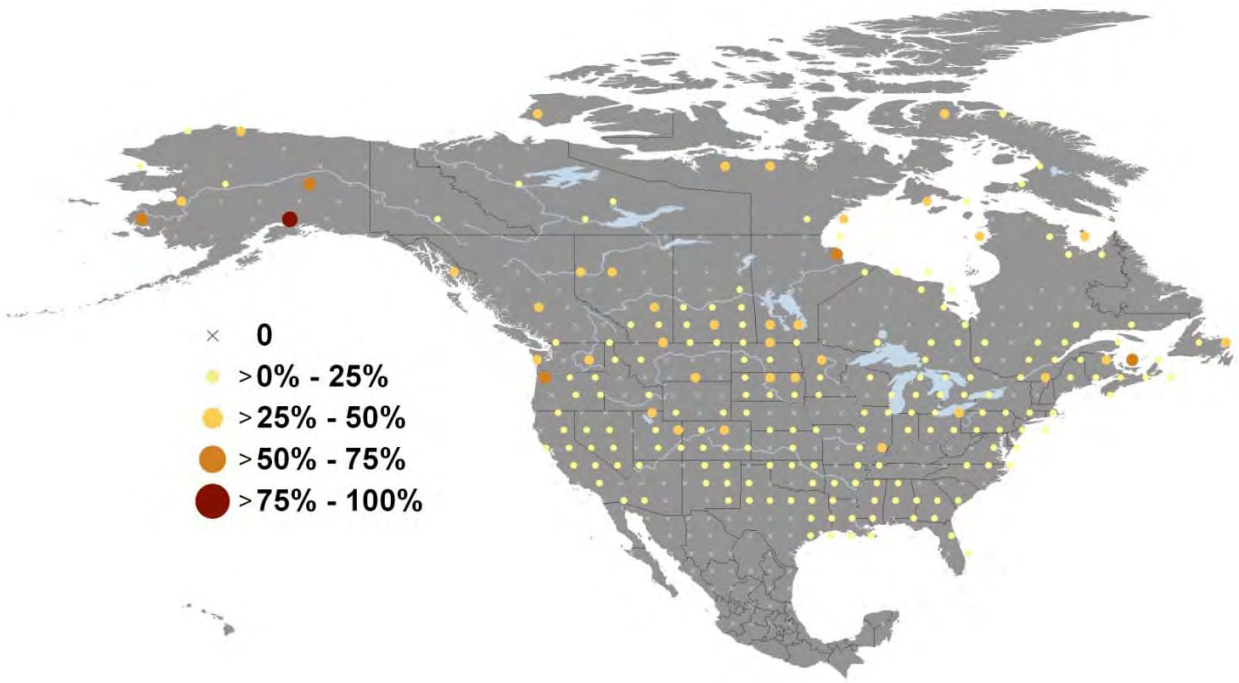


Figure 7.10: Direct harvest minimum spanning tree (minimum 5 birds per location) for all waterfowl from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

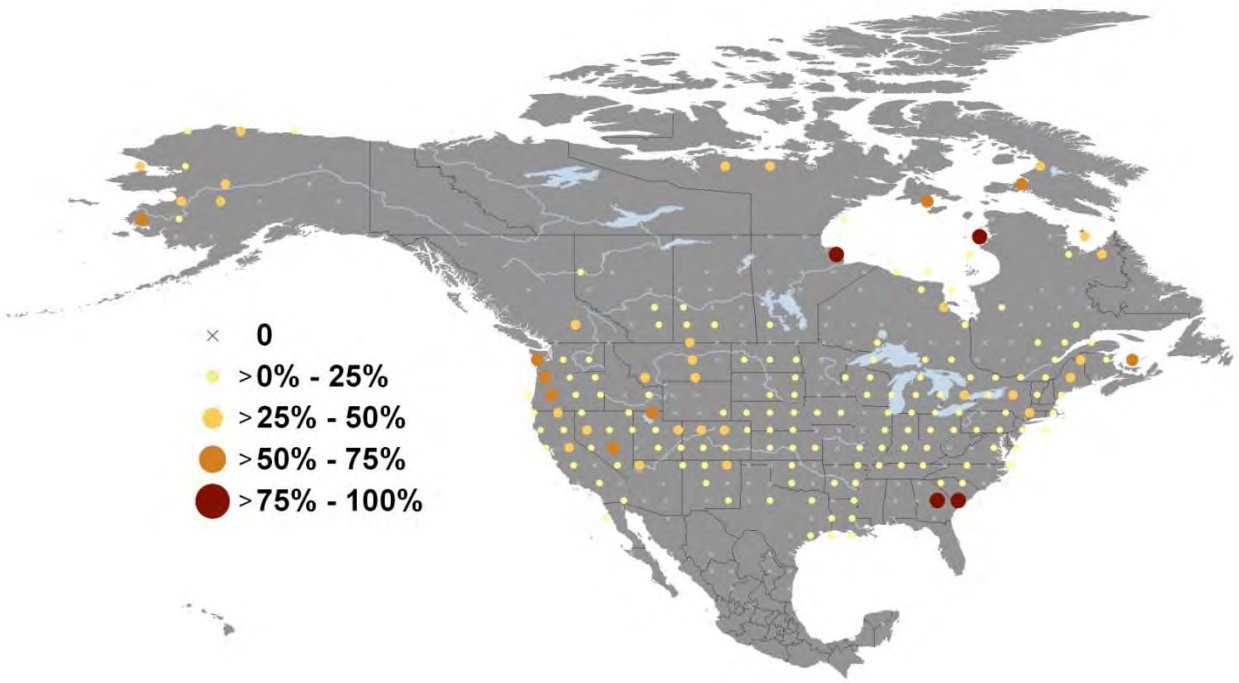


Figure 7.11: Direct harvest minimum spanning tree (minimum 5 birds per location) for dark geese from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

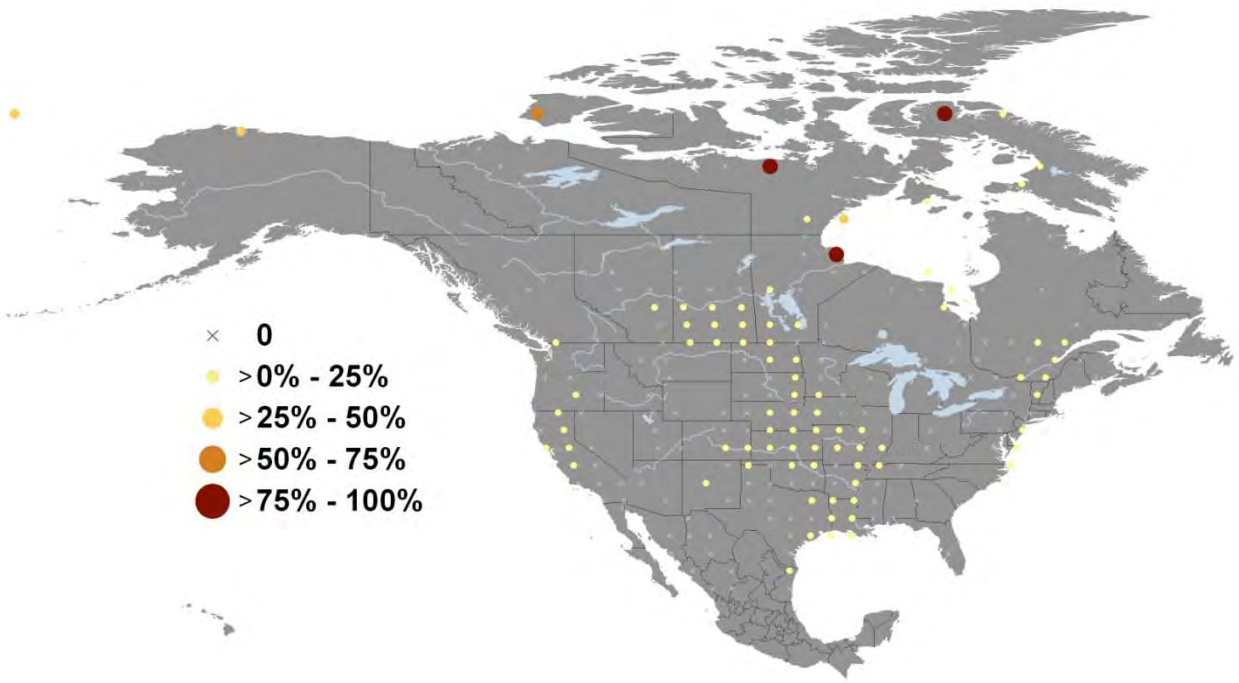


Figure 7.12: Direct harvest minimum spanning tree (minimum 5 birds per location) for light geese from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

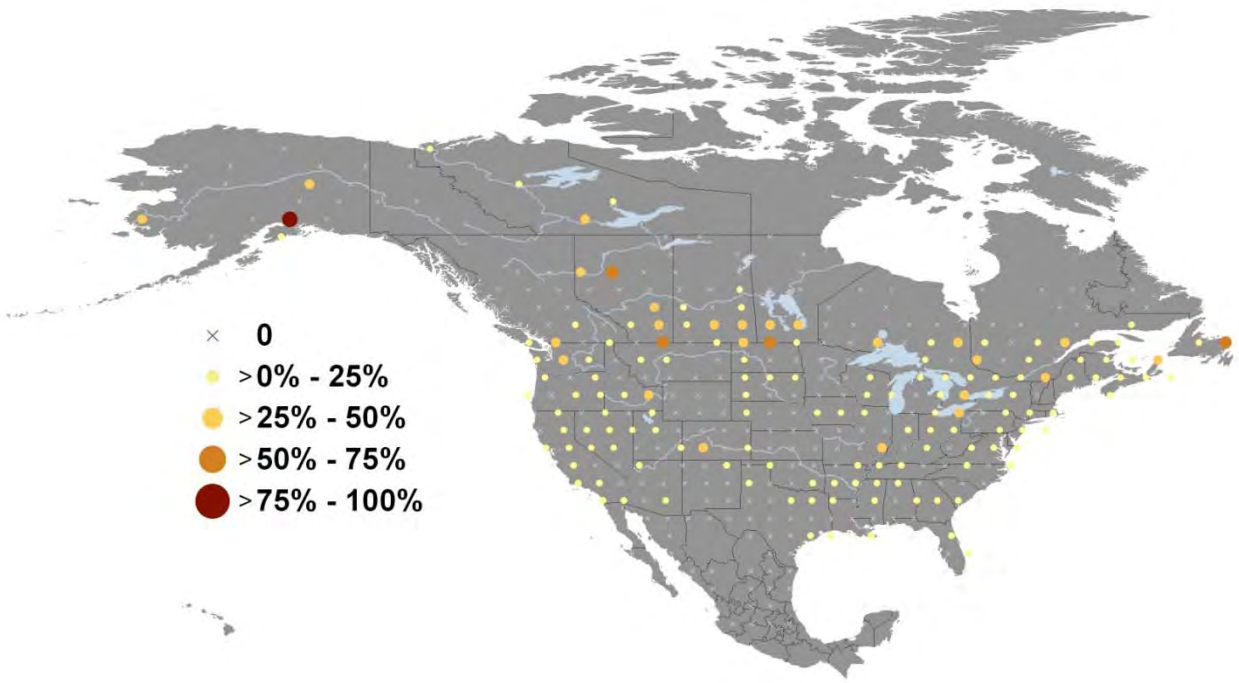


Figure 7.13: Direct harvest minimum spanning tree (minimum 5 birds per location) for dabblers from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

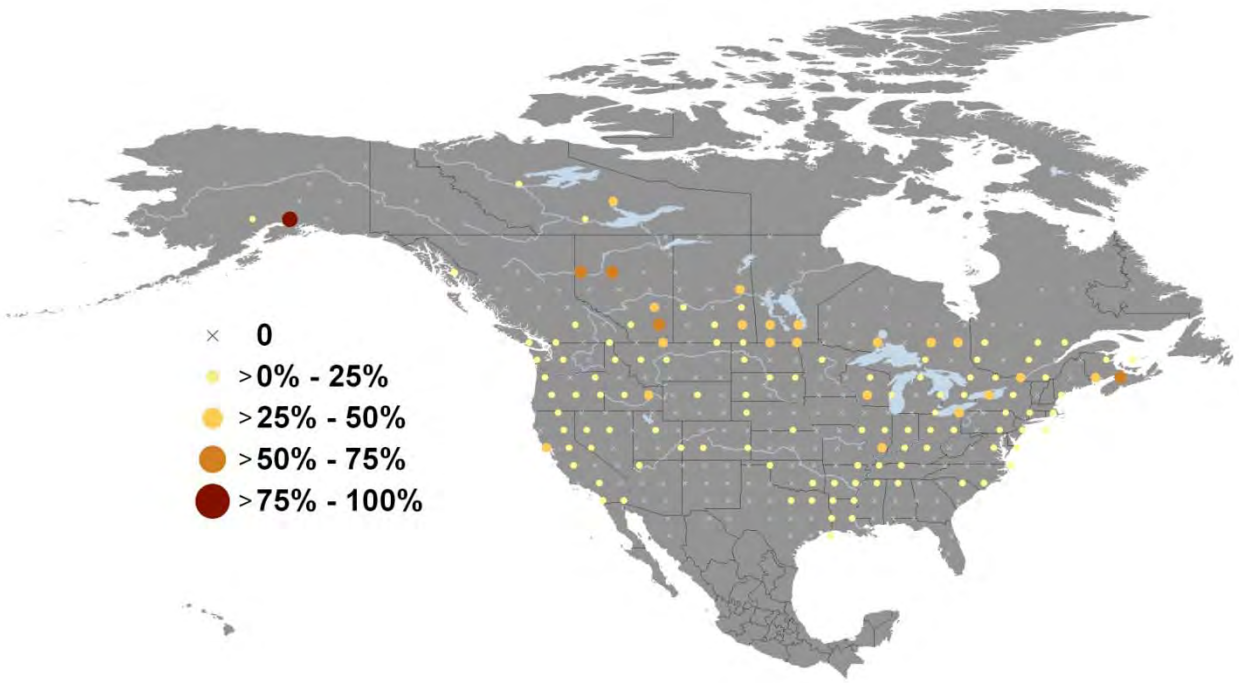


Figure 7.14: Direct harvest minimum spanning tree (minimum 5 birds per location) for Mallards from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

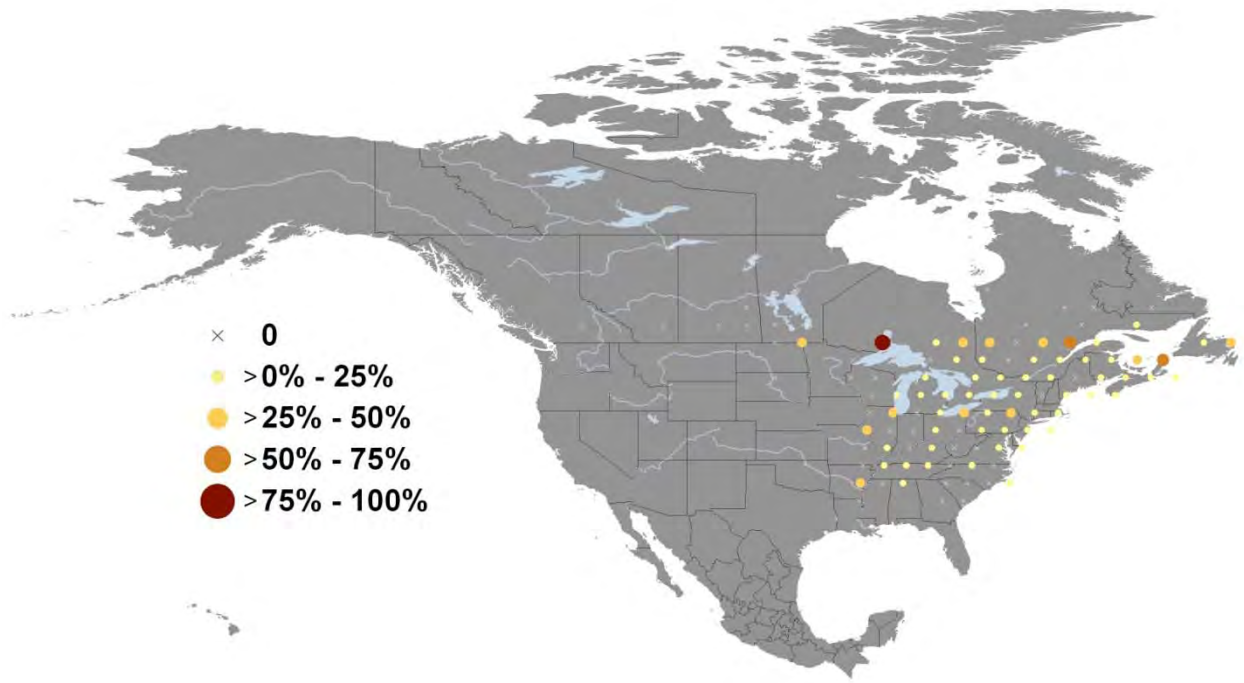


Figure 7.15: Direct harvest minimum spanning tree (minimum 5 birds per location) for American Black Ducks from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

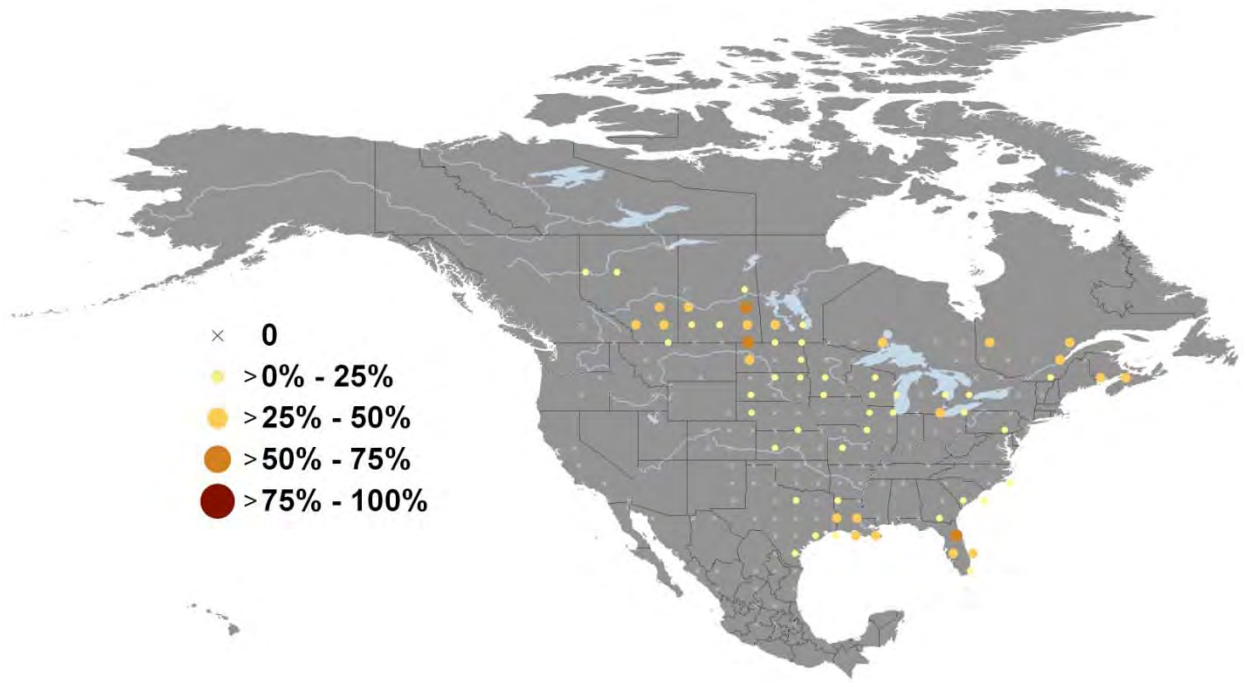


Figure 7.16: Direct harvest minimum spanning tree (minimum 5 birds per location) for Blue-winged Teal from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

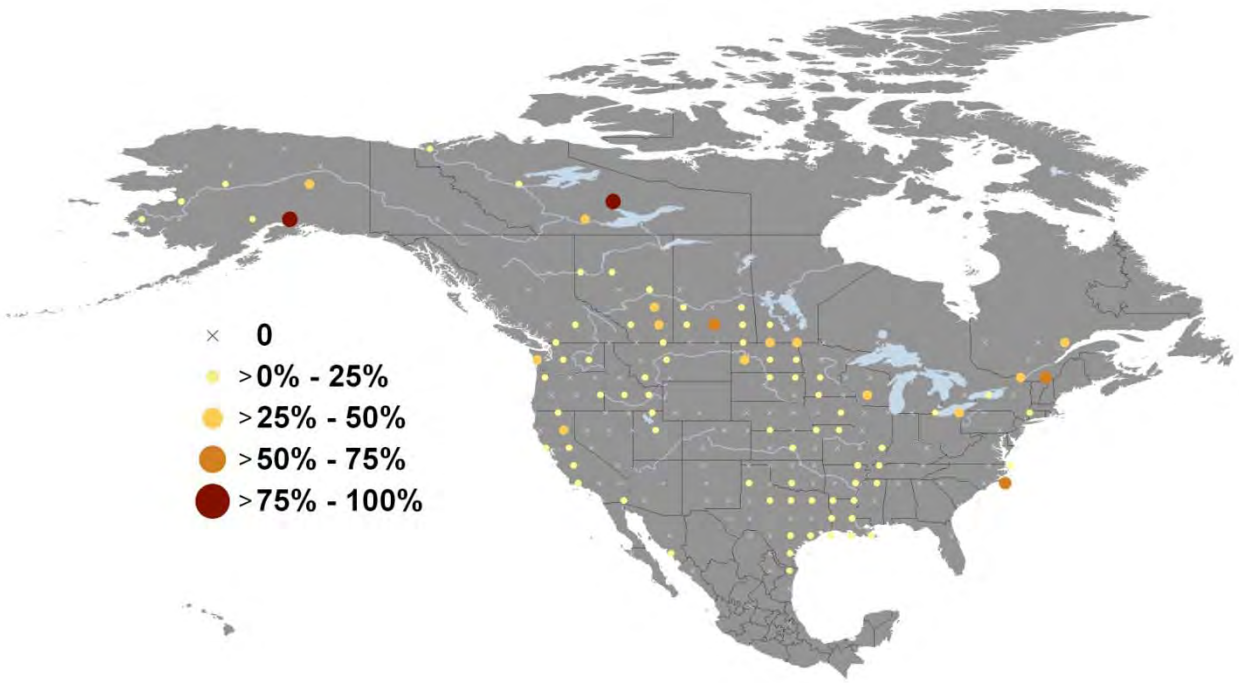


Figure 7.17: Direct harvest minimum spanning tree (minimum 5 birds per location) for Northern Pintails from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

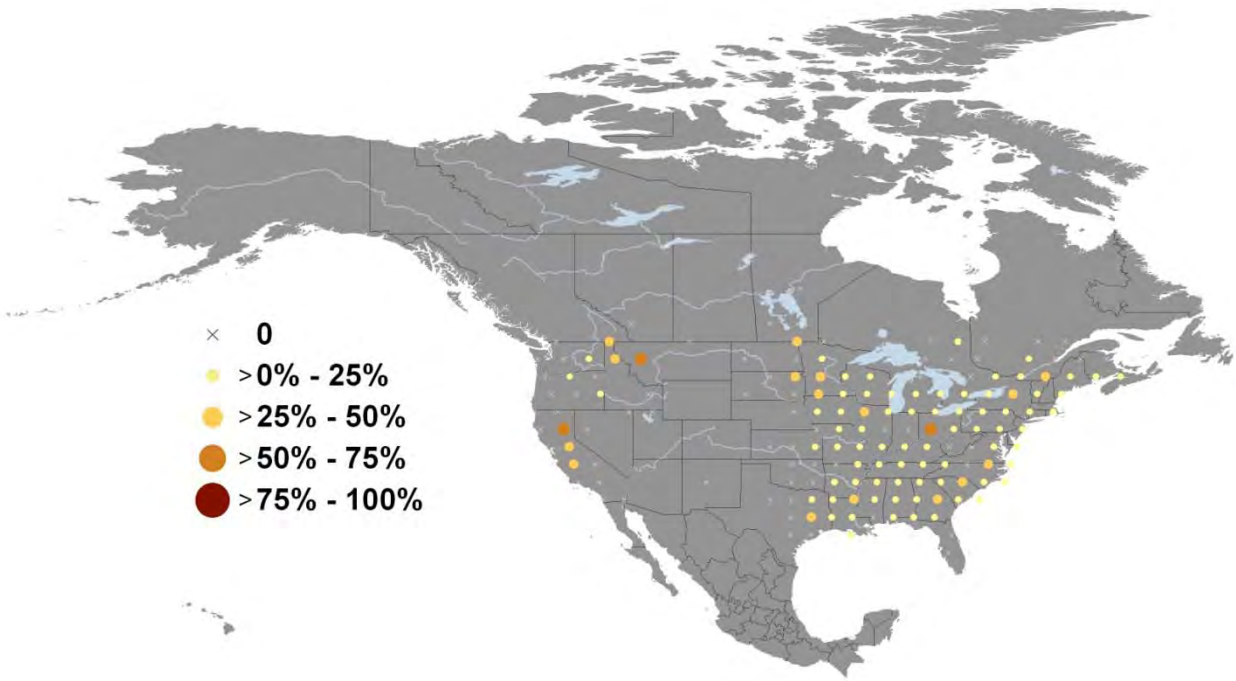


Figure 7.18: Direct harvest minimum spanning tree (minimum 5 birds per location) for Wood Ducks from 2002 thru 2006 (circle sizes represent percent of flow through a node in quartiles).

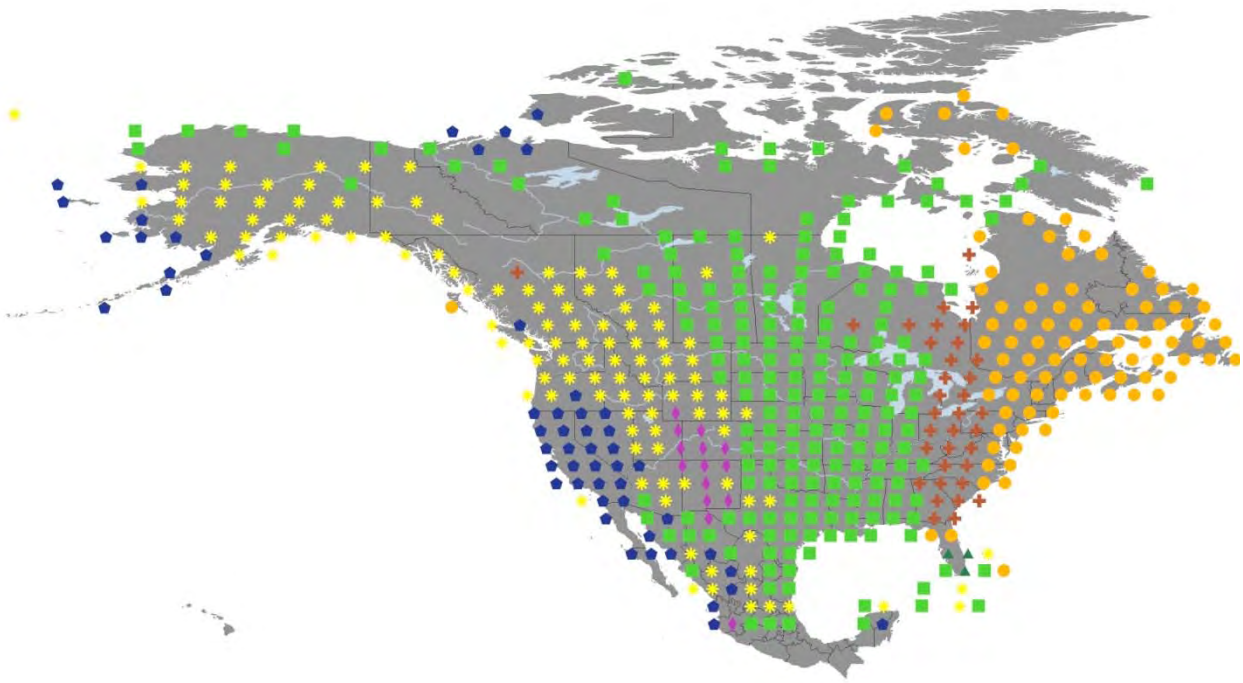


Figure 7.19: Direct harvest cluster plot for all waterfowl from 2002 thru 2006. Seven clusters, each represented by a different symbol/color combination.

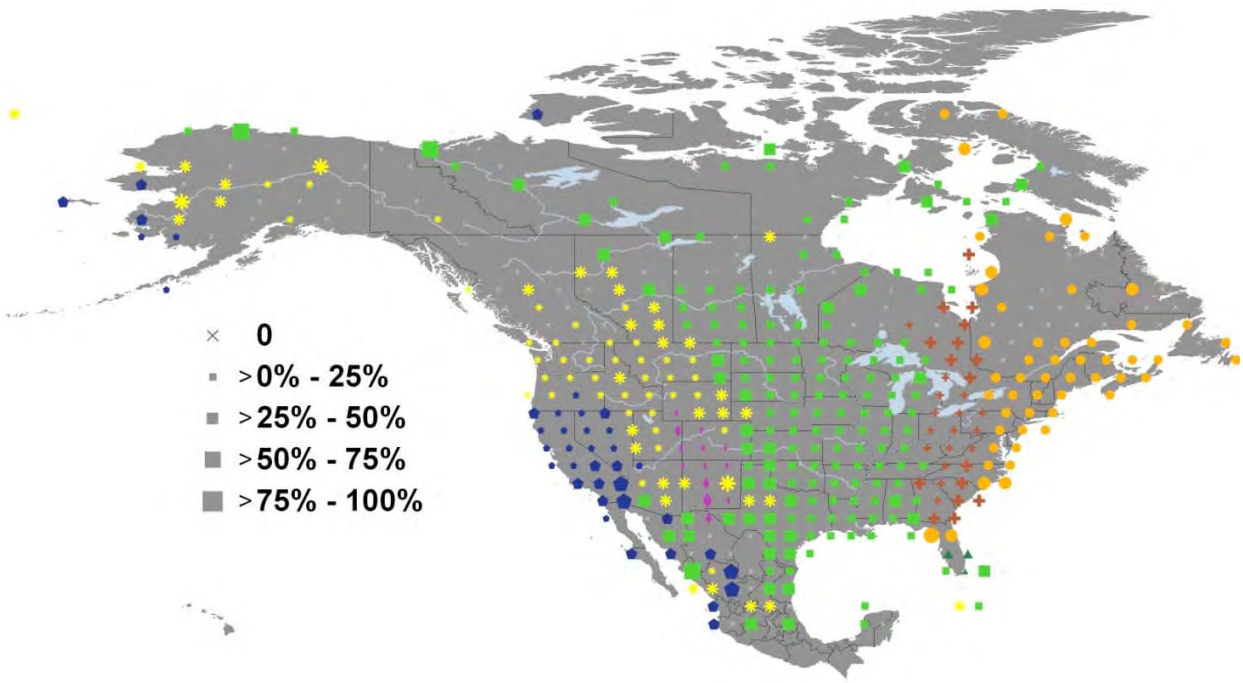


Figure 7.20: Direct harvest satellite plot for all waterfowl from 2002 thru 2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

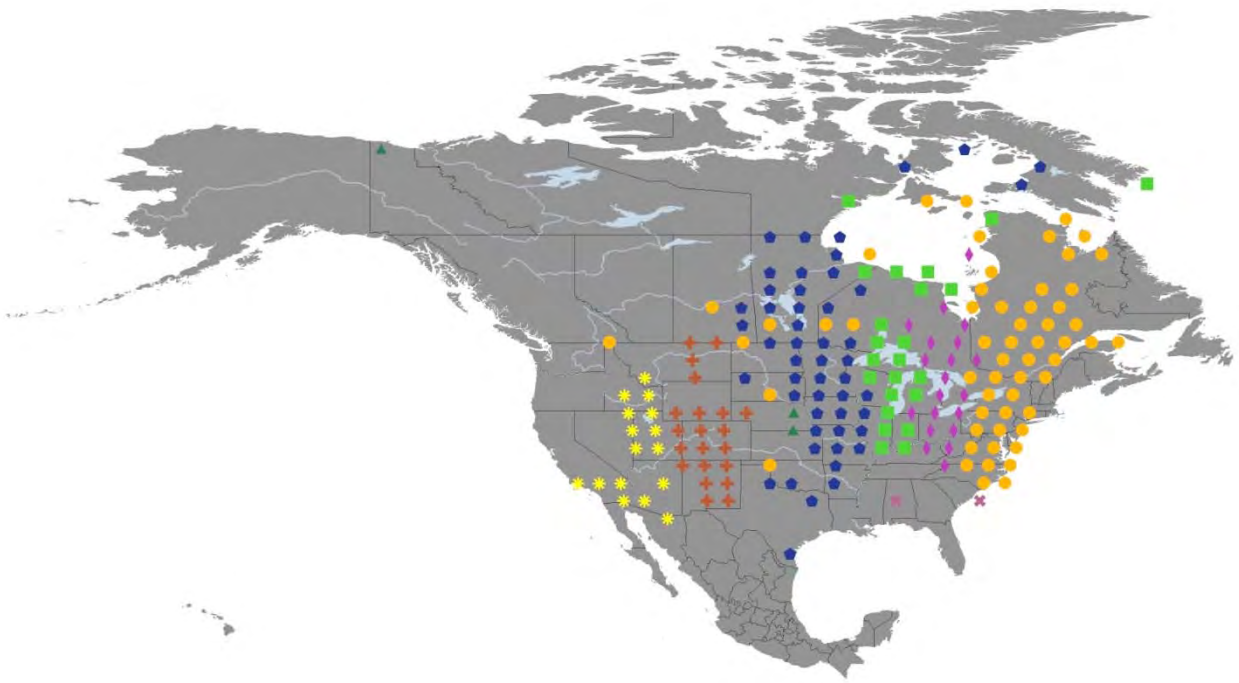


Figure 7.21: Direct harvest cluster plot for dark geese from 2002 thru 2006. Seven clusters, each represented by a different symbol/color combination.

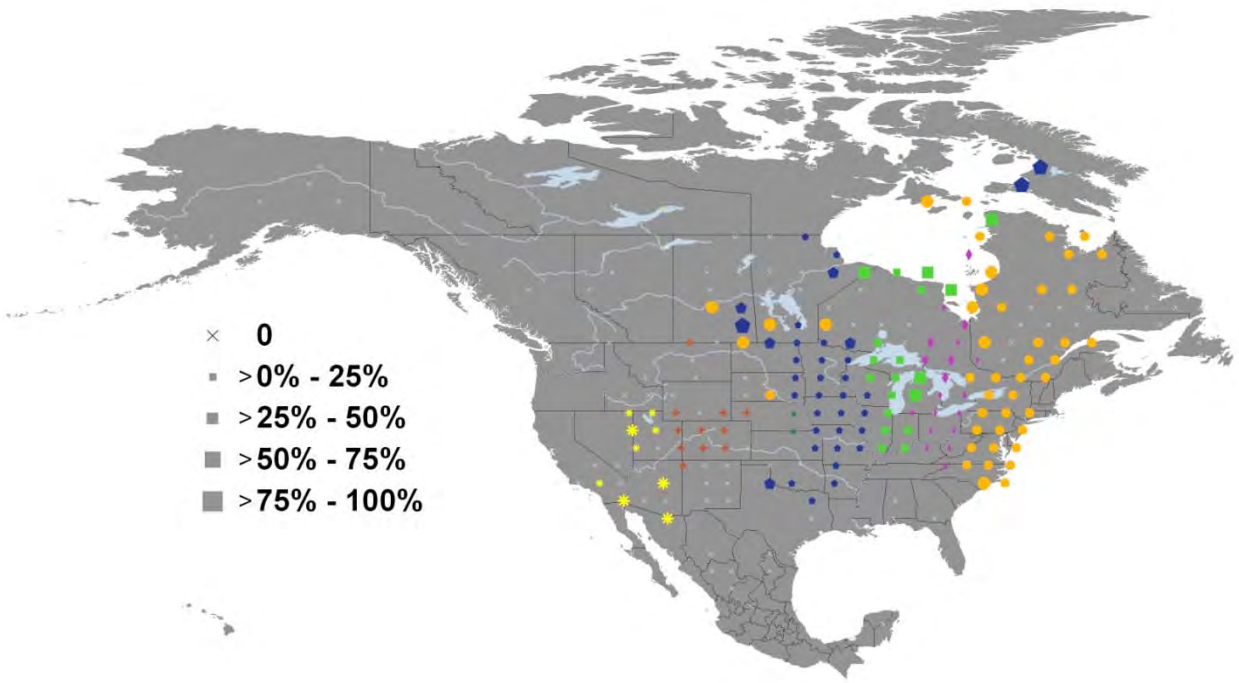


Figure 7.22: Direct harvest satellite plot for dark geese from 2002 thru 2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

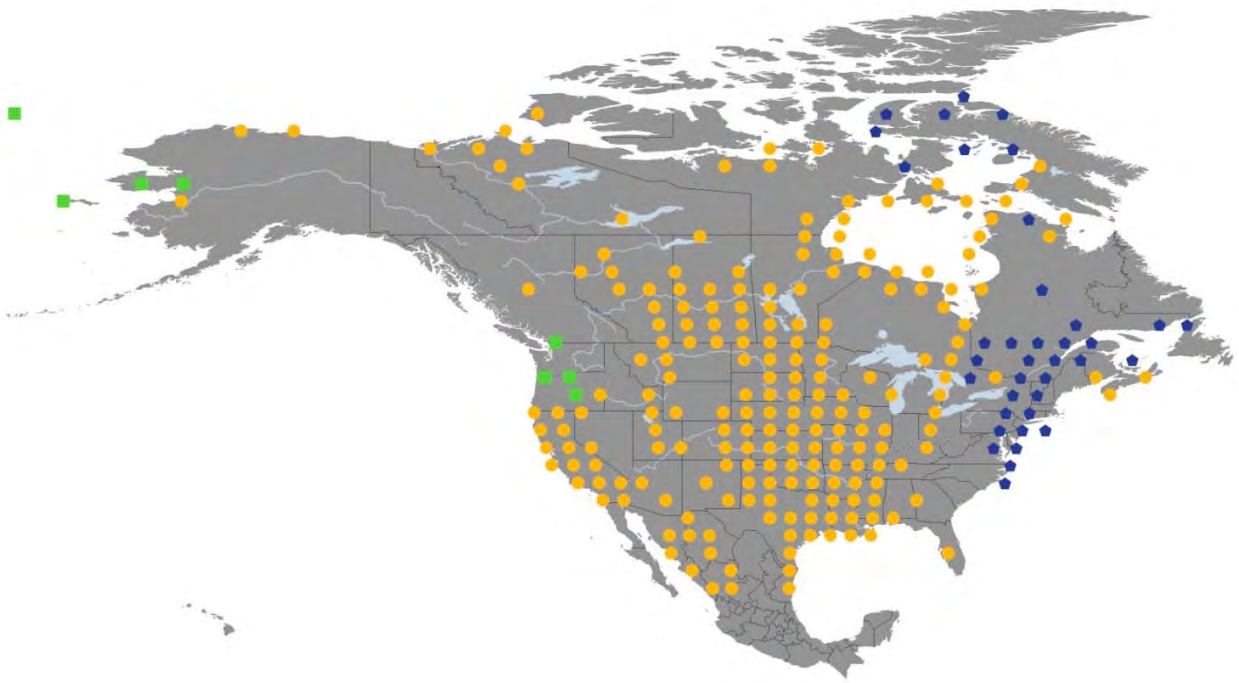


Figure 7.23: Direct harvest cluster plot for light geese from 2002 thru 2006. Three clusters, each represented by a different symbol/color combination.

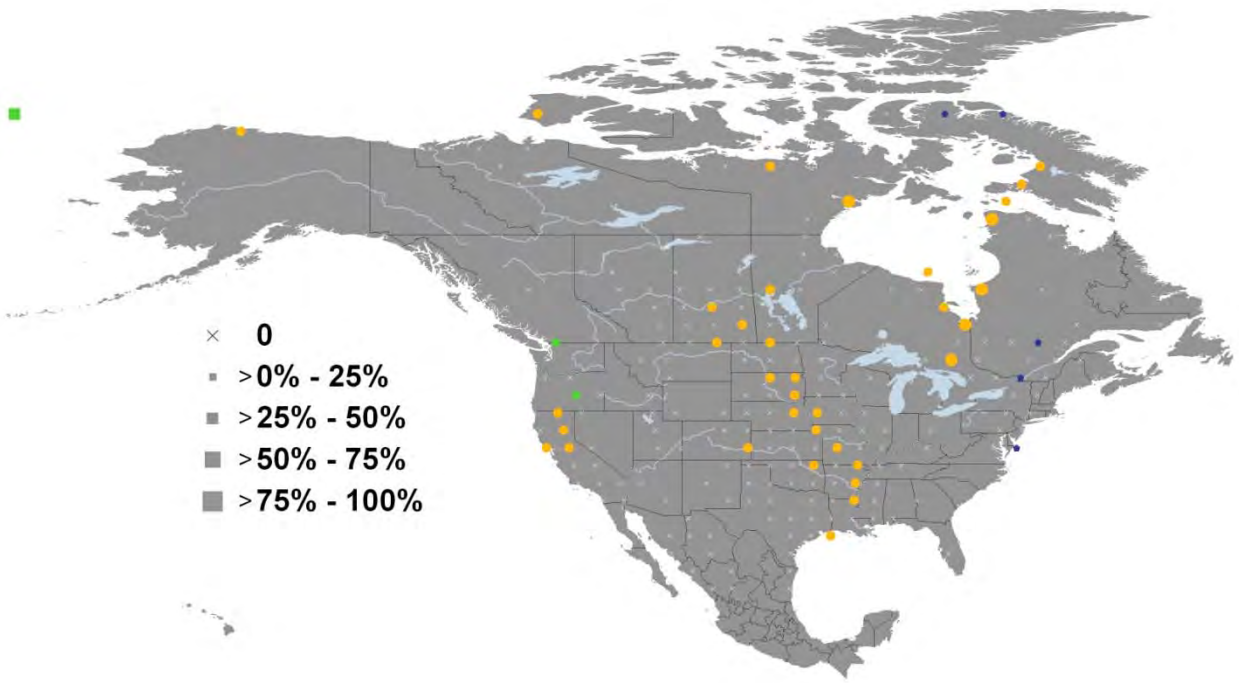


Figure 7.24: Direct harvest satellite plot for light geese from 2002 thru 2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

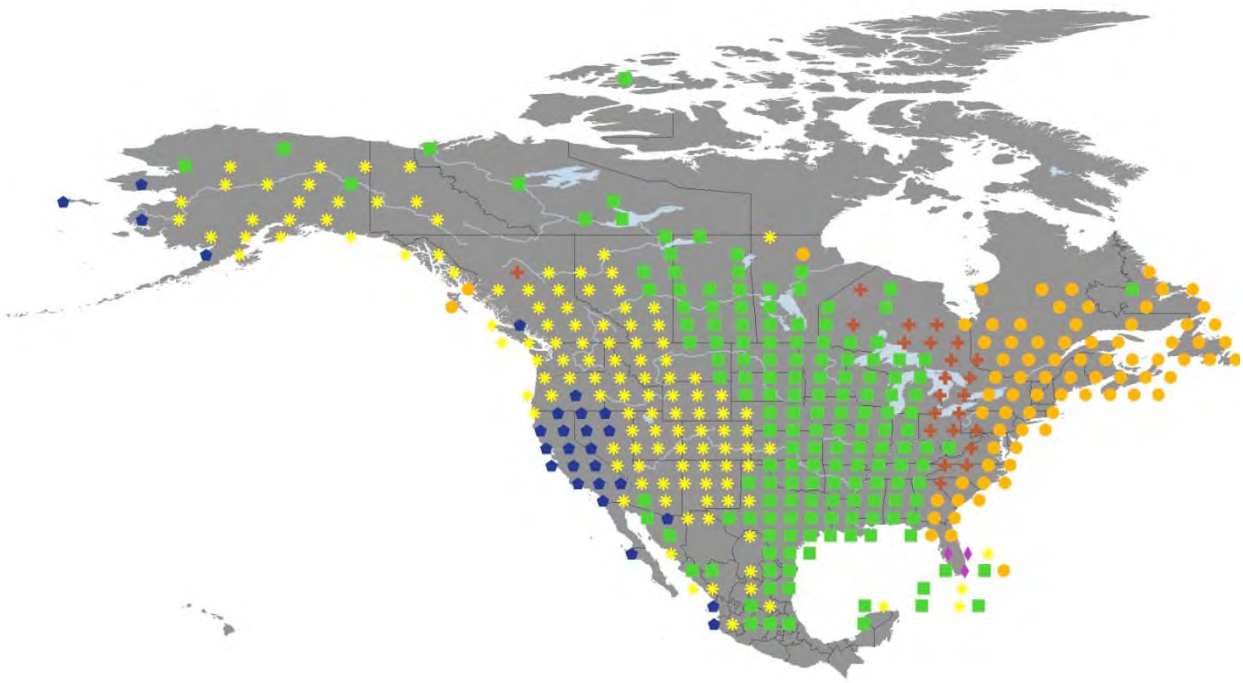


Figure 7.25: Direct harvest cluster plot for dabblers from 2002 thru 2006. Six clusters, each represented by a different symbol/color combination.

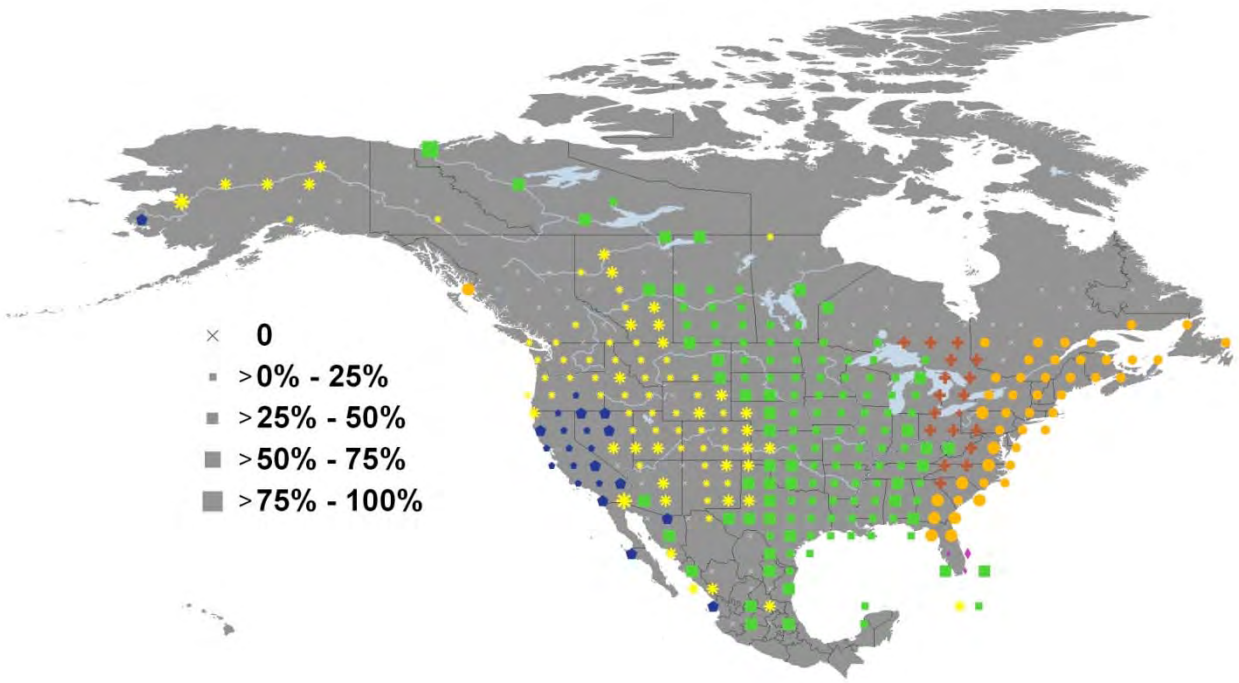


Figure 7.26: Direct harvest satellite plot for dabblers from 2002 thru 2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

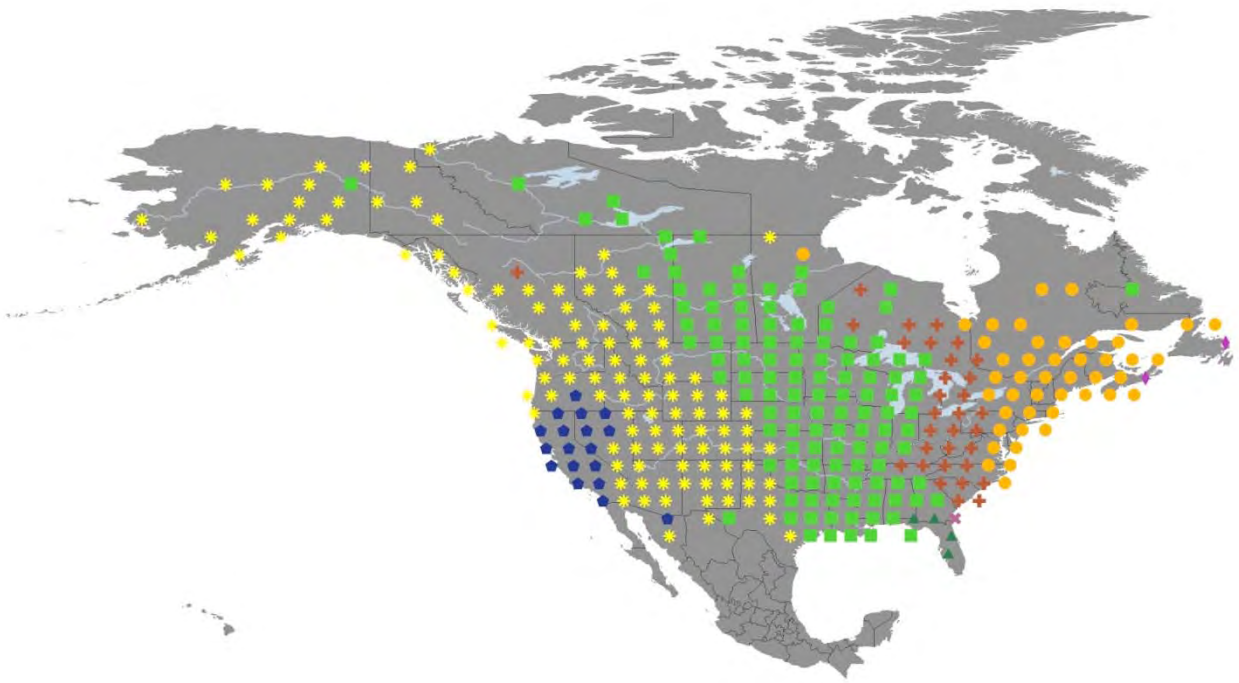


Figure 7.27: Direct harvest cluster plot for Mallards from 2002 thru 2006. Seven clusters, each represented by a different symbol/color combination.

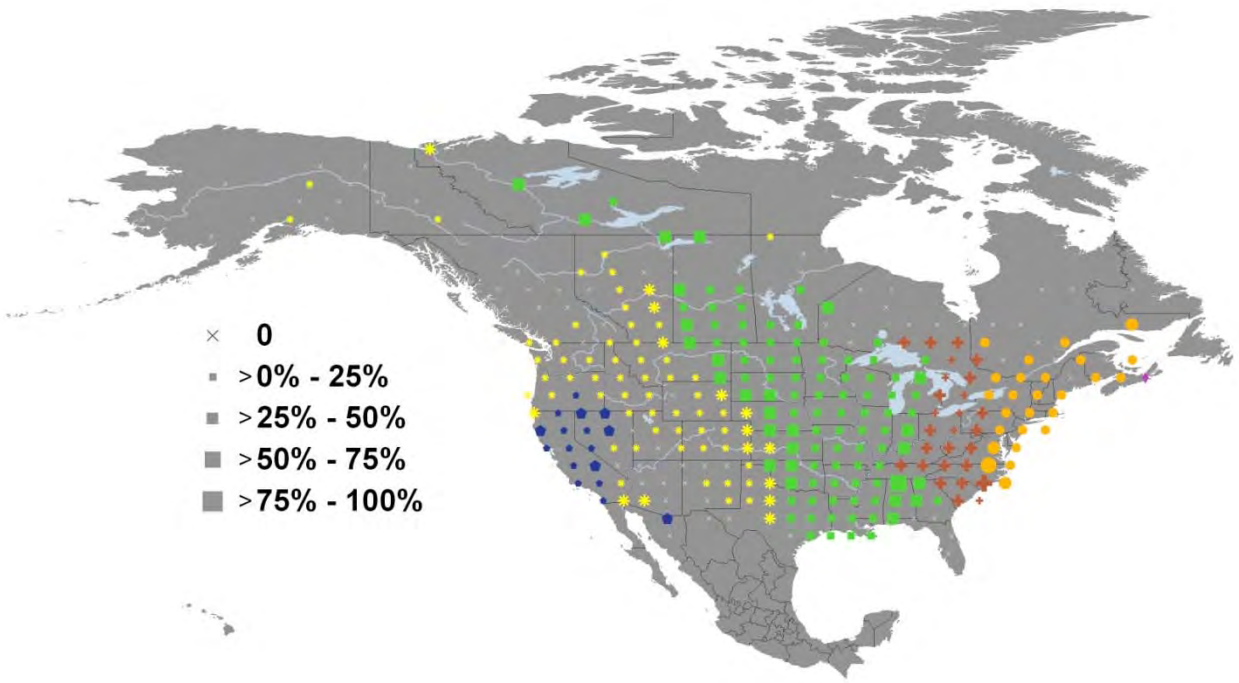


Figure 7.28: Direct harvest satellite plot for Mallards from 2002 thru 2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

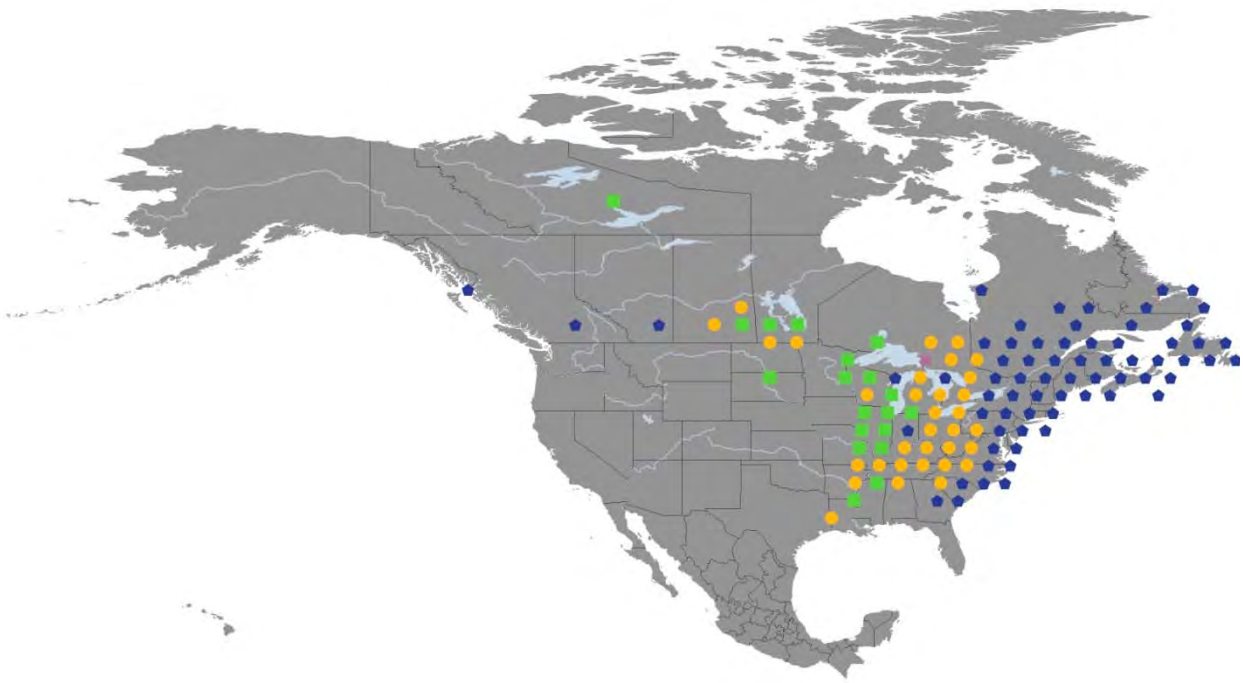


Figure 7.29: Direct harvest cluster plot for American Black Ducks from 2002 thru 2006. Three clusters, each represented by a different symbol/color combination.

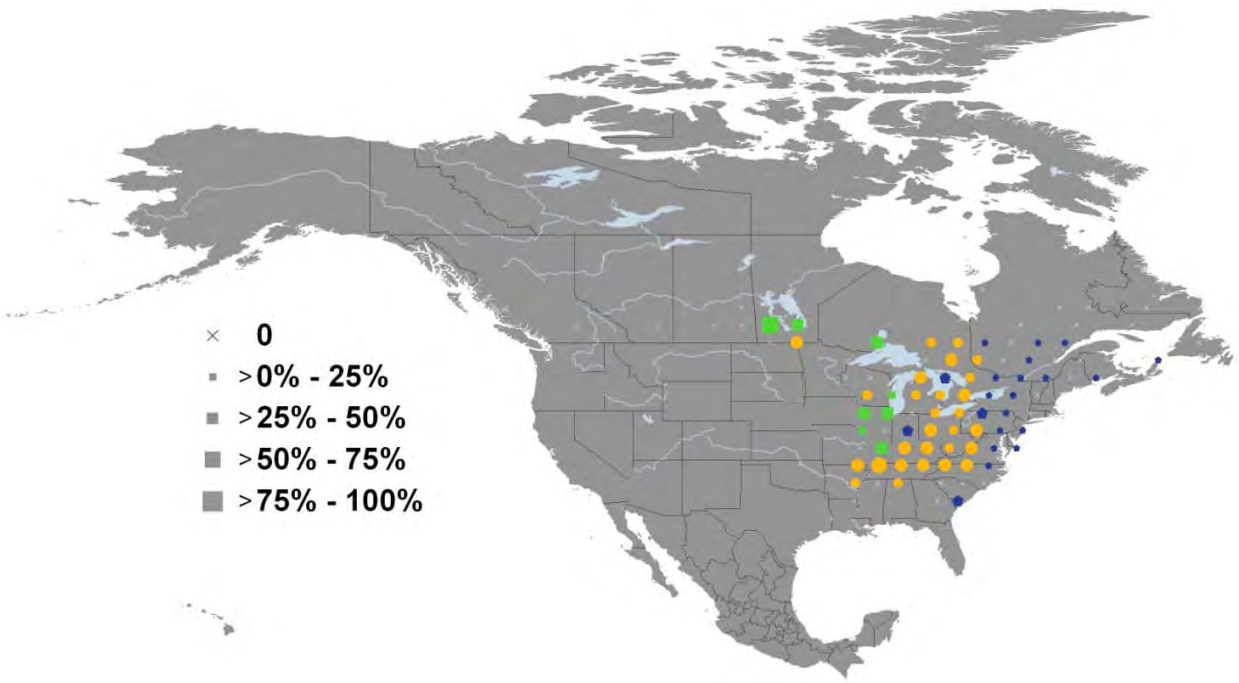


Figure 7.30: Direct harvest satellite plot for American Black Ducks from 2002 thru 2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

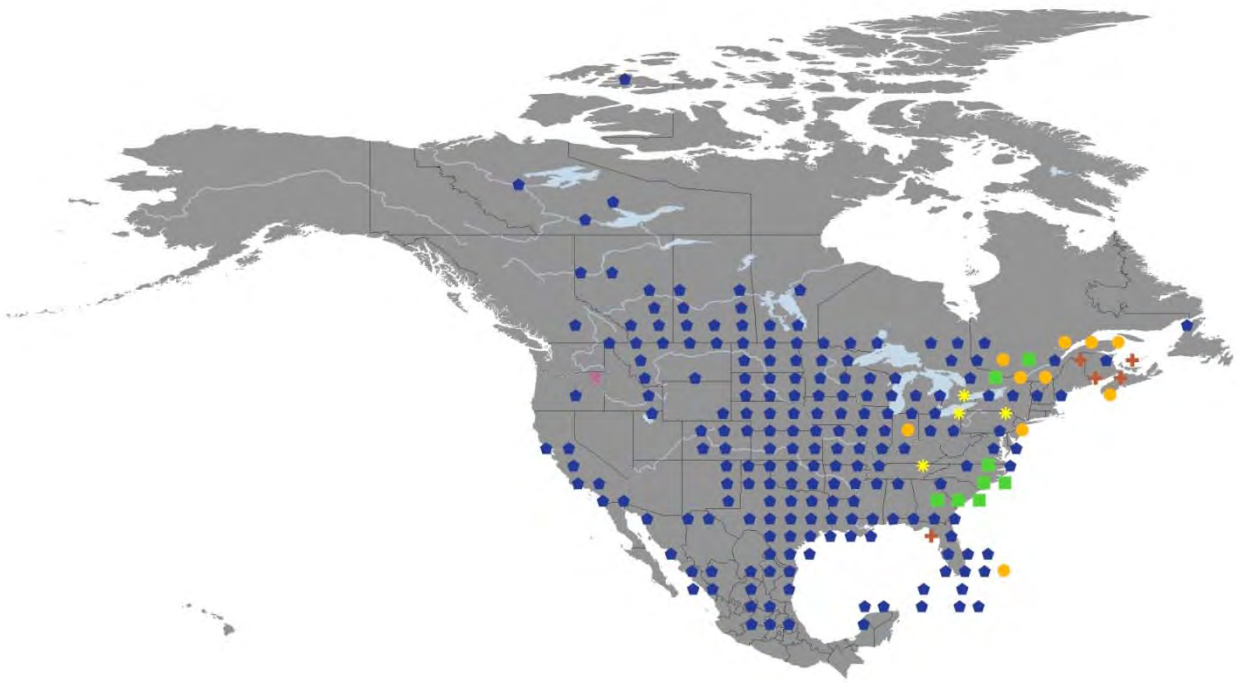


Figure 7.31: Direct harvest cluster plot for Blue-winged Teal from 2002 thru 2006. Five clusters, each represented by a different symbol/color combination.

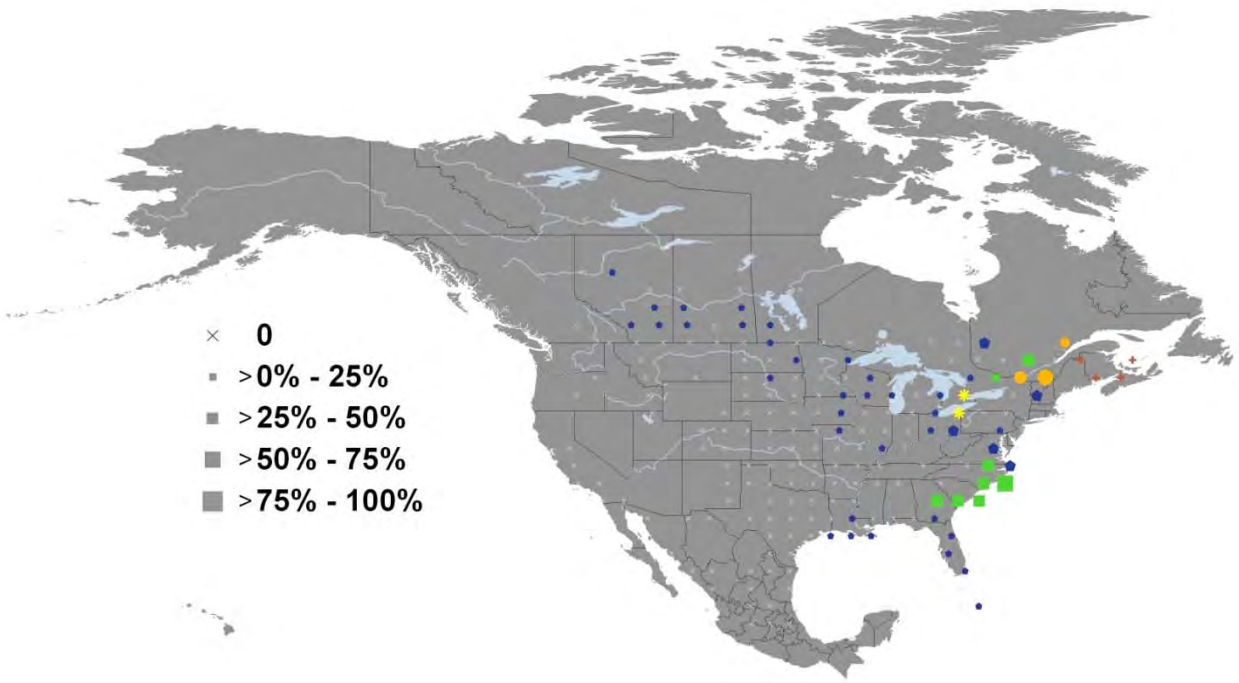


Figure 7.32: Direct harvest satellite plot for Blue-winged Teal from 2002 thru 2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

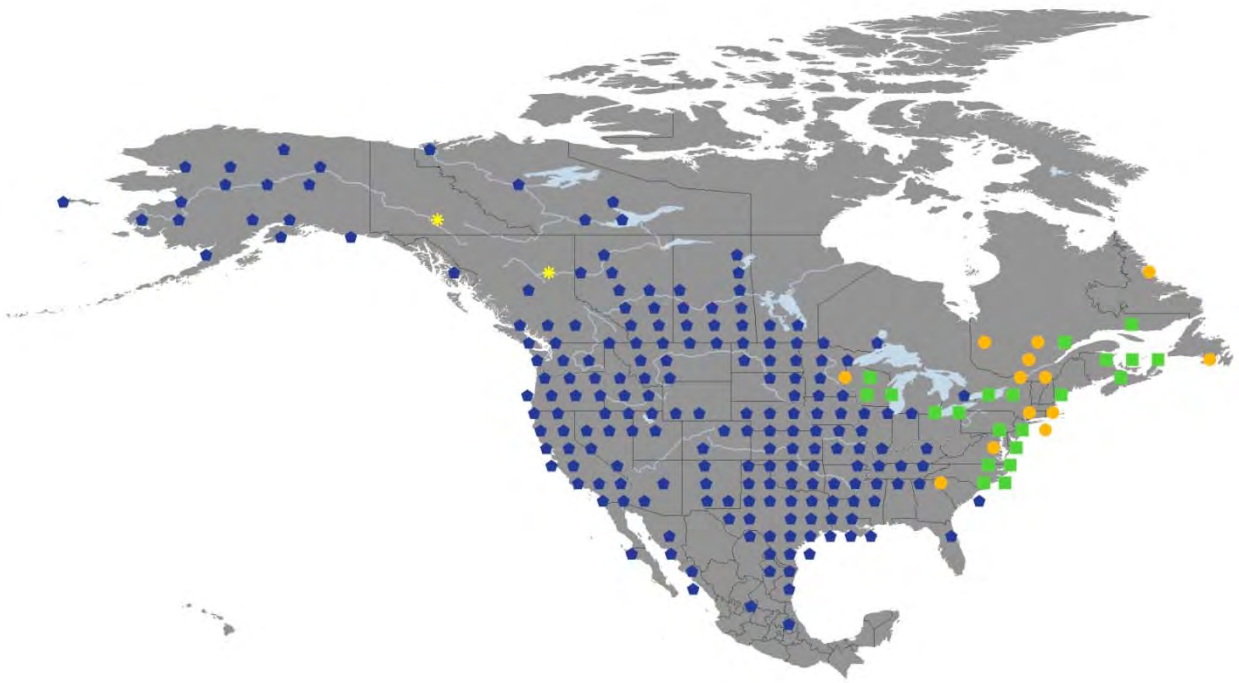


Figure 7.33: Direct harvest cluster plot for Northern Pintails from 2002 thru 2006. Four clusters, each represented by a different symbol/color combination.

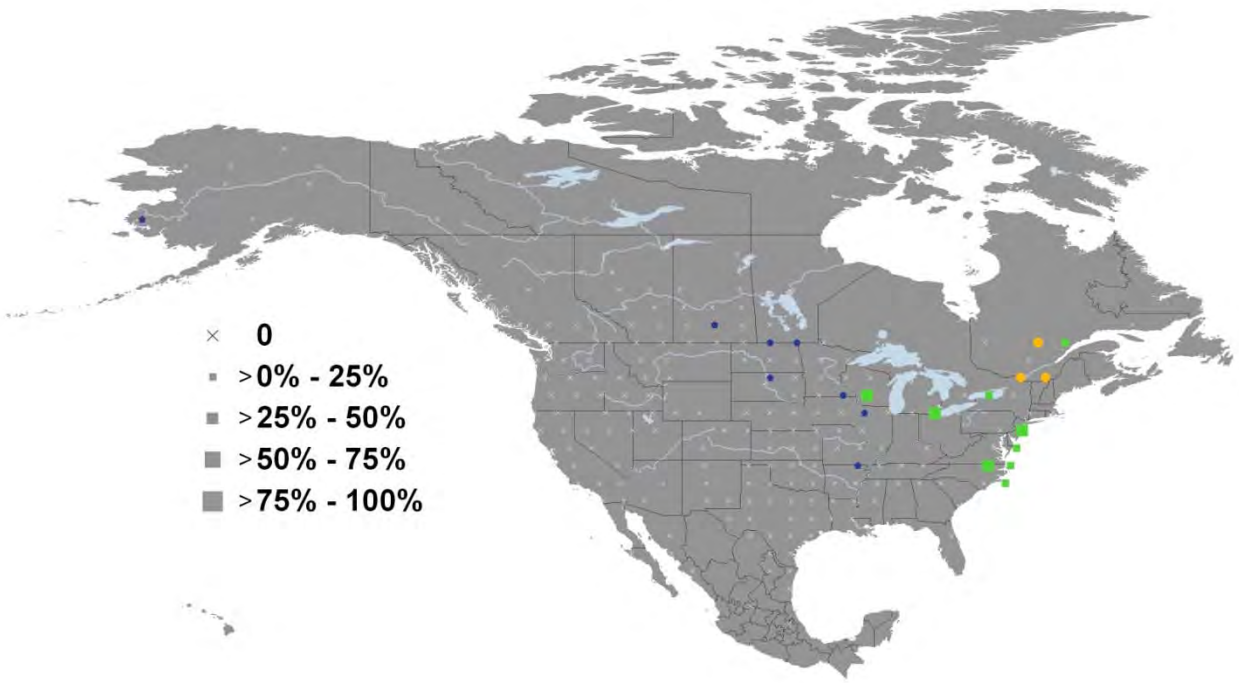


Figure 7.34: Direct harvest satellite plot for Northern Pintails from 2002 thru 2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

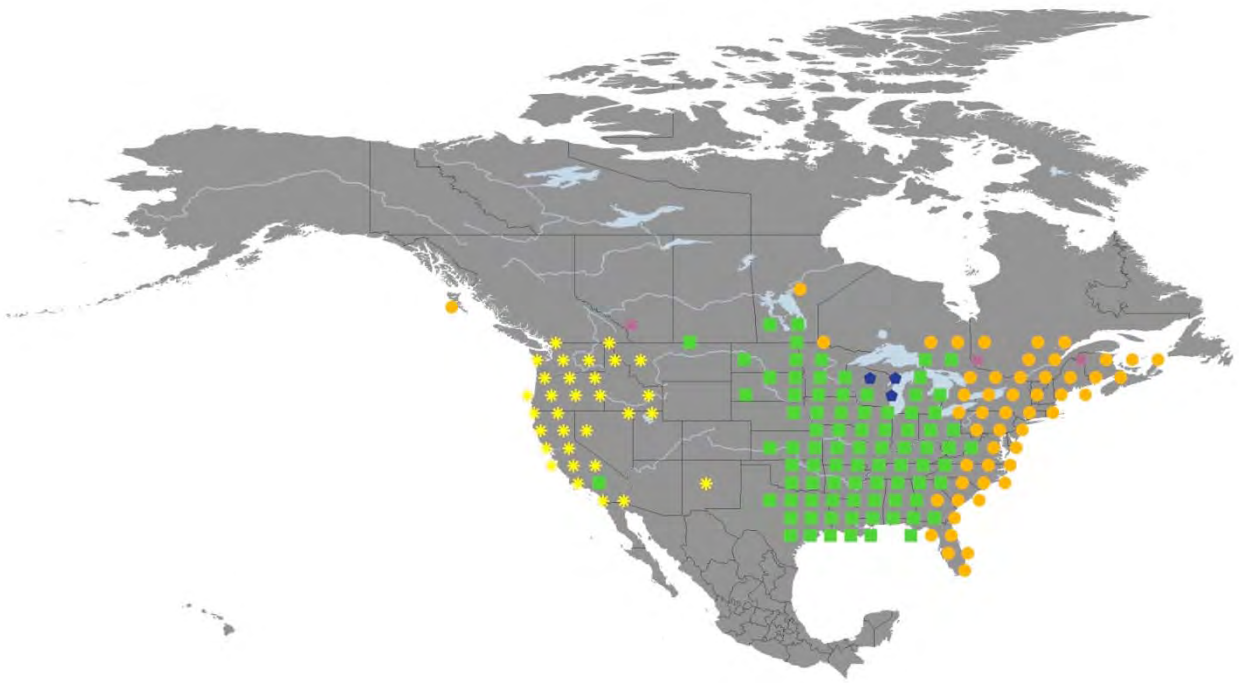


Figure 7.35: Direct harvest cluster plot for Wood Ducks from 2002 thru 2006. Four clusters, each represented by a different symbol/color combination.

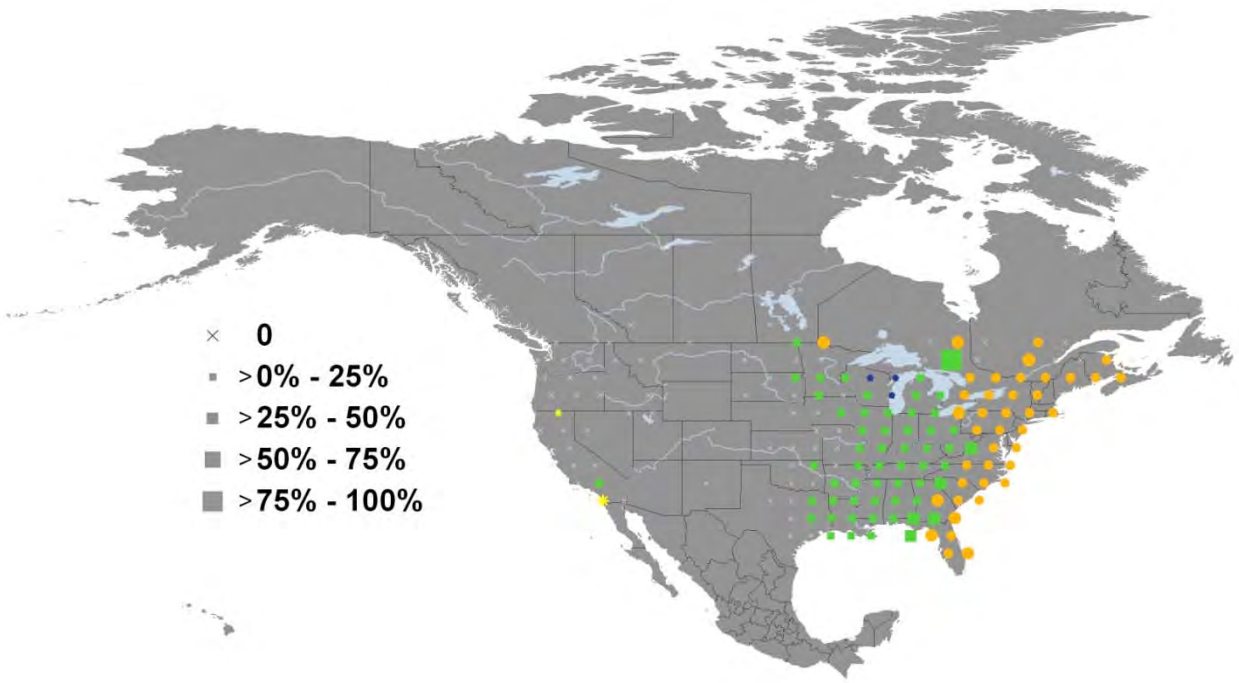


Figure 7.36: Direct harvest satellite plot for Wood Ducks from 2002 thru 2006. Symbol sizes represent the satellite values, i.e., probability of mixing at a node due to birds either coming from or going to another cluster, as quartiles based on all clusters, and the symbol/color combinations represent clusters as in the previous figure.

Literature Cited

Ahuja, R. K., T. L. Magnanti, and J. B. Orlin. 1993. Network flows. Prentice Hall.

Fernandez, R. M., and R. V. Gould. 1994. A dilemma of state power brokerage and influence in the national health policy domain. American Journal of Sociology 99:1455-1491.

Granovetter, M. S. 1973. Strength of weak ties. American Journal of Sociology 78:1360-1380.

Hanski, I. 1994. A practical model of metapopulation dynamics. The Journal of Animal Ecology 63:151-162.

Urban, D., and T. Keitt. 2001. Landscape Connectivity: A graph-theoretic perspective. Ecology 82:1205-1218.

Section 2 - Linking Wildlife Movements To Disease Spread: Implications For Highly Pathogenic Avian Influenza Surveillance

Abstract

Zoonotic infectious disease continues to be of worldwide concern. The role of wildlife movement in the spread of zoonoses is often little understood, but such knowledge is critical for developing effective disease surveillance plans. To illustrate this point, we use an example of efforts to assess the risk of Asian strain H5N1 highly pathogenic avian influenza (HPAI H5N1) being introduced into North America by migratory waterfowl. Wild bird surveillance programs, such as those conducted by the United States Department of Agriculture, have been undertaken to detect introduction of HPAI H5N1. In support of this effort, we performed a network analysis of waterfowl movement across North America using a combination of waterfowl banding and recovery data and location data for poultry production to prioritize areas where surveillance may be most effective at identifying an introduction of HPAI H5N1 into North America relative to poultry production facilities. In addition to the surveillance locations that USDA already monitors, we suggest adding additional effort in places that our analysis identified as areas of high waterfowl mixing and connectivity, such as US-Canadian border areas in the prairie pothole region. Currently some surveillance activities in Canada are being considered for reduction. We suggest that these areas in Canada need increased waterfowl surveillance and that additional collaboration between the North American countries of Canada and the United States is needed. We also use cluster analysis to identify waterfowl flyways through a cluster analysis and identify poultry operations associated with each of these flyways. If a disease outbreak - associated with waterfowl - is detected within these clusters, poultry producers can be notified that an increase in surveillance may be warranted.

Introduction

The role of wildlife in emerging and re-emerging infectious disease continues to garner attention (Daszak et al. 2000, Bengis et al. 2004, Jones et al. 2008). Of considerable concern is the role of wildlife movement on disease spread and transmission to livestock and humans (Kilpatrick et al. 2006). However, understanding wildlife movement at large scales is hampered by the lack of data and rigorous analytical methods for describing such movements (Clobert 2001). When large-scale movement data are available, they can be used to guide surveillance for diseases of zoonotic and agricultural concern.

For example, highly pathogenic Asian strain H5N1 avian influenza (HPAI H5N1) is still considered at risk for introduction into North America. In response, the United States (U.S.) government established an avian influenza surveillance plan (U.S. Interagency Strategic Plan 2006, National Avian Influenza Surveillance Plan 2007) to detect this pathogen in wild migratory waterfowl by collecting cloacal, oropharyngeal, and fecal samples across the U.S. This plan primarily focused on waterfowl (Anatidae) as these animals migrate over long distances and appear to be the primary natural reservoir for type-A influenzas (Webster et al. 1992, Olsen et al. 2006, Koehler et al. 2008, but see Peterson et al. 2008). One component of the surveillance plan directed the US Department of Agriculture (USDA) to collect waterfowl fecal samples. The original, 2006, USDA sampling plan resulted in the collection of 50,000 fecal

samples, 1,000 samples in each of the 50 states. However, because states differ by size and number of waterfowl, changes were suggested for the 2007 sampling season (Farnsworth et al. submitted) to allocate some sampling effort based on band recovery data from different points of entry by migratory waterfowl into the U.S.

Fortunately, the USGS Bird Banding Laboratory maintains one of the largest databases on migratory bird movement in the world with a main focus on waterfowl. Using these data, Farnsworth et al. (submitted) suggested that sampling effort should be unequally allocated across the 50 states, with greater sampling effort in states having higher numbers of waterfowl that originated from the upper Pacific or upper Atlantic flyways and which were subsequently recovered by hunters. The basis for this proposal was the assumption that HPAI H5N1 transmitted by migratory waterfowl could occur through Alaska (i.e., the upper Pacific flyway) or Greenland (i.e., the upper Atlantic flyway).

In this paper, we present proposed refinements to the 2007 fecal sampling plan by incorporating more detailed information on waterfowl movement, connectedness, and mixing contained in the Bird Banding Laboratory data, as well as relate these data to commercial poultry farm locations. We do not assume a particular route of transmission (e.g., Alaska or Greenland) in this analysis. Instead, we use a network approach to identify locations of high waterfowl connectedness and mixing that are potentially important to the spread of avian influenzas and other diseases occurring in waterfowl. Network analyses have recently been used in a number of theoretical and empirical disease studies (Eubank et al. 2004, Meyers et al. 2005, Brooks et al. 2008, Naug 2008). Here, the network is a disease contact network, formed by a set of locations or network nodes. The connections between nodes or edges are determined by the movement of birds between two locations based on banding and recovery information.

The structure of such disease contact networks influences the spread of disease (Eubank et al. 2004, Meyers et al. 2005, Brooks et al. 2008, Naug 2008). For example, disease spreads more quickly if there are “super spreaders,” essentially nodes with relatively many more connections or edges than other nodes in the network (Lloyd-Smith et al. 2005). The structure of disease contact networks can be captured statistically, and structure at multiple scales within the network can be important. Thus, we developed several different connectivity metrics that capture mixing at continental, flyway and local spatial scales.

Our goal was to prioritize areas for increased surveillance of wild birds by identifying areas where a high degree of waterfowl mixing takes place during migration and identifying locations that were highly connected to other locations by waterfowl movement. The network metrics can be examined alone to investigate the potential for mixing, connectivity, and spread of avian influenza at a particular spatial scale. The metrics can also be examined as a whole, in order to determine which locations are important in the introduction and spread of avian influenza across a range of scales. Finally, we have also identified commercial poultry production areas that may be at risk from waterfowl-mediated HPAI H5N1 infection. We present this case study as an example of how rigorous analysis of wildlife movement data can provide guidance for targeted sampling efforts in large-scale surveillance programs.

Methodology

To characterize waterfowl migratory patterns, we used recent (2002-2006) banding and recovery records for waterfowl (Anatidae), in the subfamilies for dabbling ducks (Anatinae), divers (Aythyinae), and geese (Anserinae) contained at the USGS Bird Banding Laboratory. We were primarily interested in dabbling ducks since these species are thought to be very important in the

spread of HPAI H5N1 (i.e., they shed HPAI H5N1 virus but exhibit no morbidity; Brown et al. 2006) and exhibit the highest type-A influenza prevalence worldwide (Fouchier 2006; Fouchier et al. 2007). Further, we selected mallards (*Anas platyrhynchos*) as a primary species of interest because: 1) it is the most abundant waterfowl species in North America, and thus well represented in the Bird Banding Laboratory data and 2) recent work (Keawcharoen et al. 2008) suggests that the mallard may be capable of shedding relatively large quantities of HPAI H5N1 while remaining asymptomatic.

We constructed a 200-km lattice across North America and aggregated banding and recovery (i.e., birds shot by hunters) locations to the center of each lattice grid cell, which we subsequently refer to as a (network) node. The collection of all nodes at which at least one banding and/or recovery occurred formed the contact network. From this network, we derived various metrics to characterize the connectivity and mixing between nodes across the network. In particular, we present measures of the flow of birds through nodes, a minimum spanning tree across the network, clusters (biological flyways) of highly intra-connected nodes and satellite metrics of mixing between flyways.

The original bird banding and recovery location data were reported at a 10-minute resolution (approximately 10x10 km^2). We compared 10-minute, 100-km and 200-km resolutions and found that the 200-km resolution retained sufficient detail to investigate continental-level migratory patterns, while minimizing uninformative noise in the data. We used a great circle distance projection (GCD or gnomonic projection) in constructing the spatial lattice. Tradeoffs occur in any choice of projection. Because we were most interested in bird movement, we chose great circle distance in order to preserve distances between nodes in a data set that is continental scale. For convenience, we used a central point roughly in the middle of the United States, 40° North latitude and 100° West longitude, as the starting point. We generated the lattice using the great circle distance projection by identifying locations that are 200 km away in each of the four cardinal directions and continued until the lattice spanned the latitudes of approximately 18° to 80° North and longitudes 50 ° West to 170° East. Our lattice extended to 170° East to include the Aleutian Islands of Alaska, but not the islands of Hawaii.

Banding and recovery data for the species of interest from 2002 through 2006 were aggregated to the nearest lattice grid point (node) in order to associate individual bird band records with nodes. Network metrics were calculated at the 200-km scale after all aggregation occurred. In general, banding occurs in the late spring and early summer months and recoveries occur in the fall and winter months during waterfowl hunting seasons. For the network measures being considered, the aggregation of the data with respect to time enhances the spatial signal in the connectivity data. We recognize the final allocation of resources for sampling and monitoring must also take into account the time of year but this was not evaluated in our analysis.

Banding and recovery flow- Banding and recovery flow metrics capture local level connectivity and mixing. Areas with higher concentrations of waterfowl are expected to have more mixing of birds and a higher likelihood of avian influenza transmission. Higher rates of contact among birds may occur directly or indirectly via feces and shedding of avian influenza virus into shared water sources.

Banding flow and recovery flow are measures of the number of birds at each node. Banding flow is the number of birds banded at a node, and recovery flow is the number of birds recovered at a node standardized by the maximum count for each metric. In addition to banding

and recovery flow, we investigated several alternative metrics based on incidence measures described by Hanski (1994). These alternative metrics were highly correlated with each other and with banding and recovery flow. Because banding and recovery flow are the simplest metrics, we report them here.

Minimum spanning tree-Minimum spanning trees (Ahuja et al. 1993, Urban and Keitt 2001) can be used to describe network or continental scale mixing and connectivity. In achieving this goal, we were not only concerned with nodes with large numbers of bandings/recoveries, but we were also interested in identifying nodes from which individuals disperse and to which individuals aggregate. By selecting the appropriate subset of nodes, the number of nodes to which surveillance resources need be allocated can be reduced while ensuring that all nodes are represented (by direct or indirect observation).

The minimum spanning tree identifies the minimum subset of network nodes that maintains nearly complete connectivity within the network. Network nodes, selected at random, are pruned one at a time and the resultant network is evaluated to determine whether or not connectivity is maintained. If connectivity is maintained then the node is deleted. Otherwise the node is replaced and a second node tested. This procedure continues until no further nodes can be eliminated and the smallest subset of nodes is identified. We only considered nodes at which at least five individuals were observed (banded or recovered). Furthermore, since minimum spanning tress are not necessarily unique and the analysis requires searching across all possible subsets of nodes, we ran the algorithm numerous times with random starts and found 100 candidate trees for each species/functional group; the resultant figures present the relative proportion of times that each node was included in the minimum spanning tree.

Cluster analysis-Another measure of connectivity is the degree to which sub-collections of nodes (clusters) are more strongly intra-connected relative to the network as a whole. Clusters describe a naturally occurring, intermediate scale in the network. Clusters represent regions (although they may not be *geographically* connected) that individuals are more likely to travel within, i.e., an individual that is banded at a node within cluster i is more likely to be recovered at a node within cluster i; similarly, an individual recovered at some node within cluster i is more likely to have been banded in that same cluster. Such clusters appear to correspond with flyways (Lincoln 1935) or migration corridors (Belrose 1968). To identify clusters we employed an annealing algorithm that weights the "strength" of each pairwise connection by the number of individuals known to have been banded at one of the two nodes and harvested at the other. Pairs of nodes with a strong pairwise connection are characterized by having many individuals being banded and recovered between them (i.e., moving between them).

The algorithm that we used begins by identifying the strongest pairwise connection and assigning the two nodes to cluster number one. The algorithm then proceeds to the next strongest pairwise connection; if one of the nodes has already been assigned to cluster one then the second node of the pair is also assigned to cluster one. Otherwise a second cluster is created. This process is repeated until all of the individual nodes have been assigned to a cluster. The algorithm finishes by merging clusters with a mean inter-connectivity that exceeds the mean intra-connectivity; the mean is defined to be the total number of connections divided by the total number of potential connections available between the two clusters). An important consideration in this analysis is that the approach does not require assumptions about connectivity with respect to proximity, i.e., nodes are not assumed to be part of the same cluster simply because they are

geographically "nearest neighbors," and allow the observed information on waterfowl movement to inform the makeup of the clusters.

Satellite metric-Satellites describe points of connection among clusters and the potential for mixing of birds among flyways. Satellite nodes were first described in social networks where they are referred to as bridging ties (Granovetter 1973, Fernandez and Gould 1994). Once nodes have been assigned to their primary clusters, as described above, the number of birds that travel between each node and other nodes in non-primary clusters are counted. The number of birds moving along these intercluster connections are summed for each node and divided by the total number of birds moving to or from the same node, resulting in a value for the satellite metric.

Poultry production and risk-To illustrate the potential risk of HPAI H5N1 to the poultry industry, we constructed poultry farm location maps at the same 200-km resolution used in the band recovery analysis. These data were extracted from the 2002 USDA National Agricultural Statistics Service poultry survey data (USDA 2002). The raw data included data fields at the county level for the number of farms and the approximate number of birds in inventory on a farm for several poultry production classifications, e.g., broilers, layers, pullets, etc. For the subsequent discussion we restrict attention to broiler chickens, which is the largest segment of the commercial poultry industry in the United States accounting for 86% of all poultry production and 99% of all ready-to-cook chicken products in the United States (Source: Economic Resource Service (USDA)). We analyzed the data for small (<5000 birds per farm) and large (≥5000 birds per farm) broiler farms.

We present two competing views for assessing risk to poultry operations: the total number of poultry at risk and the level of bio-security available at the farm level. The former assumes that the potential number of dollars at risk is large because a positive infection and the subsequent mitigation (generally depopulation of all birds on a farm) would be costly whereas the latter assumes that less secure facilities are at greater risk to infection from outside agents, such as migratory waterfowl. By relating county-scale poultry production to particular clusters (flyways), poultry operations most at risk could be identified and appropriate actions taken in the event of a disease outbreak in wild migratory waterfowl within a given cluster.

Results

We examined approximately 200,000 banding/recovery records from 2002 through 2006. We found that waterfowl often did not stay in a single flyway, but tended to be widely distributed over the continent. For instance, Figure 1 details Mallard hunter recoveries from one particular location in Alaska over a five-year period and demonstrates that although waterfowl flyway (Lincoln 1935, Hawkins et al. 1984) definitions might be a useful political and management designation, individual waterfowl are not faithful to such political designations.

Banding and recovery flow-Banding and recovery flow results varied with respect to species (Figures 2 and 3). We assume that the banding and recovery efforts are representative of where birds are located. For Mallards, banding flow is greatest in the prairie pothole region and Central Valley of California (Figure 2A). Recovery flow is greatest in the Central Valley of California and lower Mississippi River (Figure 3A). Since banding tends to occur earlier in the year and recovery later, this suggests that the prairie pothole region may be an important mixing site early, while the lower Mississippi may be important later in the year. The Central Valley of California

may remain important as a mixing site throughout the year. The prairie pothole region and Central Valley of California also appear to be important local mixing sites for Northern Pintails (*Anas acuta*) (Figures 2B and 3B). No clear patterns, based on banding and recovery, occur for Black Ducks (*Anas rubripes*) (Figures 2C and 3C). Both banding and recovery flow suggest the Great Lakes are an area of high local mixing for dark geese (Figures 2D and 3D).

Minimum spanning tree-The minimum spanning tree results varied with respect to species (Figure 4). For Mallards (Figure 4A), the critical surveillance nodes were distributed along the US-Canadian border, especially in the prairie pothole region as well as in the Central Valley of California. However, for other dabbler species such as Northern Pintail (Figure 4B) and the American Black Duck (Figure 4C), critical surveillance nodes appear in differing regions. For example, Figure 4B (Northern Pintail) illustrates that important nodes at higher latitudes should be included in the surveillance plan. The American Black Duck (Figure 4C) is found primarily in the eastern United States as illustrated by the resultant minimum spanning tree summary. Hence, it is critical that species type be considered in the selection of surveillance sites.

The minimum set of nodes needed to represent all locations at which dabbling ducks were either banded or recovered closely resembles the results presented in Figure 2A, i.e., the nodes tend to be located along the US-Canada border (especially in the prairie pothole region), in Alaska, and the Central Valley of California. This result is driven in part by the large number of Mallards in the database relative to other species.

Unlike the dabblers, the resultant minimum spanning tree for the dark geese (Figure 4D) was more uniform across southern Canada and the contiguous U. S. Furthermore, nodes at the very northern latitudes (dark goose breeding areas) are also deemed to be important for surveillance. A large proportion of the dark geese database was made up of the Canada Goose (*Branta canadensis*); the results are thus largely reflective of this species.

The results from the minimum spanning trees suggest that these nodes of increased waterfowl mixing could be valuable in detecting pathogens being dispersed by waterfowl, depending on the pathogen and waterfowl species, are useful to allocate surveillance efforts across large scales.

Cluster analysis-The results from the cluster analysis varied by species. The standard flyway concept was most supported by Mallards (Figure 5A), for which the most banding and recovery data existed. The mid-continental Mallard flyway can be most strongly detected, as well as the Pacific, Mississippi, and Atlantic flyways. For the Northern Pintails three distinct “flyways” are identified (Figure 5B). Although the American Black Duck resides mostly in the eastern half of the United States there is strong evidence of three distinct clusters or sub-populations as well (Figure 5C).

Satellite metrics-Not surprisingly for most species, relatively high levels of mixing between clusters occur either at the edges of clusters or when there is geographic mixing of flyways, as shown in Figure 5D-F (e.g., Northern Pintail flyways 1 and 2 where red dots mixed within the green dots in the central United States, Figures 5B, E). Northern Pintails appear to have higher levels of mixing between clusters than the other species shown here. Mallards and dark geese appear to have slightly higher levels of mixing between clusters at higher latitudes (Figure 5D).

Poultry production and risk-Figure 6 shows the number of poultry farms associated with the network nodes within the U. S. for large (≥5000 poultry per farm) and small broiler farms (<5000 birds per farm), respectively. The location of some of the largest broiler farms is in the central-southern states (e.g., Arkansas and Georgia) and the Delaware-Maryland-Virginia peninsula.

By comparing the distribution of broiler chickens (Figure 6) with the cluster analysis results (Figure 5A-C) for migratory waterfowl, relative risk to farms of differing size at different locations could be assessed given a particular disease outbreak in a particular species of waterfowl.

Discussion

Understanding animal movements is critical to understanding how zoonotic diseases might spread. By analyzing one of the largest wildlife movement databases in existence, we were able to better understand waterfowl movement, a group of species with high potential to spread highly pathogenic avian influenza viruses. We identified geographic areas that were critical to maintaining the connectedness of the waterfowl migratory population in North America and, thus, areas where surveillance should be targeted. We also identified how these areas (nodes) clustered and how these clusters associated with poultry locations. Although the bird banding and recovery database is extensive, our conclusions are conditional on these data. These data do not come from a probability-based sampling plan, they do not necessarily reflect underlying waterfowl abundances at particular locations, and recoveries are dependent on where hunters have access to particular locations. One future development might be to incorporate these banding data with additional survey data (e.g., May survey flights conducted by the U.S. Fish & Wildlife Service and the Canadian Wildlife Service) to better understand the relationship between banding locations and waterfowl abundance at those locations (J. Nichols, pers comm.). Designing a sampling plan to formally estimate waterfowl movements would also be fruitful. Ultimately, the ability of these network measures to predict avian influenza spread or prevalence should be tested.

With respect to avian influenza surveillance, we suggest that additional sampling effort be allocated to locations in central Canada (Alberta, Saskatchewan, and Manitoba) along the US-Canadian border. These sites ranked as highly connected at multiple scales (local as measured by flow and continental as measured by minimum spanning tree) with many waterfowl likely moving through these areas. In addition, these areas serve as mixing areas for birds originating from several flyways. Canada began conducting an inter-Agency wild bird influenza survey in 2005 and currently has plans for a 2009 survey that will test healthy live wild birds and wild birds found dead (http://wildlife1.usask.ca/en/aiv/index.php, accessed 1 June 2009). We believe that as long as HPAI H5N1 remains a threat, cooperation between the U. S. and Canadian governments is needed, and some focus should be placed on sampling in these locations.

Identifying critical locations, where migratory bird populations have the potential to introduce and exchange HPAI H5N1 and other pathogens across broad geographic areas, have implications for coincident poultry operations. Although, no national level data exist on locations of poultry operations in the United States, which may be at greatest risk for exposure via wild bird pathways (e.g., backyard flocks, small-scale producers, and free ranging producers), we were able to develop a coarse-level portrayal of where poultry surveillance might be enhanced should HPAI H5N1 appear in North American migratory bird populations using our analyses.

Not all species result in similar clusters of nodes, and thus differing groups of poultry operations may be at risk depending on which waterfowl species are carrying a given pathogen.

Because data are not available for specific on farm biosecurity practices we assumed that large broiler farms in the United States, with large numbers of birds at risk, tend to invest more in biosecurity and are presumed to be at lower risk for disease introduction from wild waterfowl mediated avian influenza (Mannelli 2006, MacDonald 2008, USDA 1999, 2000). Hence, it may be less likely that an infection will be transmitted via waterfowl mingling with the local poultry population, but more cost is associated with such risk. However, once avian influenza is introduced to large poultry farms this segment of the industry maybe at greater risk because of more individual bird contacts (Thomas 2003, Tablante 1999). In contrast, small farms are assumed to be less bio-secure because they lack either resources or financial incentives to implement rigorous biosecurity measures (e.g., footbaths, mitigations preventing rodent or wild bird access, truck washes, etc.) (USDA 1999, 2000, 2005; Jacob 2008; Mannelli 2006). In addition, the growth of small scale producers raising antibiotic-free and free range broiler chickens has also added concern about disease emergence in small producers (Jacob 2008). An exception to this assumption is that breeder flocks, which are generally small and highly valued for their genes, tend to maintain very high biosecurity. We were not able to differentiate these flocks in the data, but they are assumed to be a relatively small percentage of broiler flocks.

Utilizing large-scale animal movement databases, in conjunction with network analyses has potential for elucidating disease transmission risk from wild animals to animals of concern. In our case we focused on waterfowl and poultry, but our approach could be taken for many other zoonotic diseases.

Literature Cited

Ahuja, R. K., T. L. Magnanti, and J. B. Orlin. 1993. Network flows. Prentice Hall.

Belrose, F. C. 1968. Waterfowl migration corridors east of the Rocky Mountains in the United States. Volume 61.

Bengis, R. H., F. A. Leighton, J. R. Fischer, M. Artois, T. Morner, and C. M. Tate. 2004. The role of wildlife in emerging and re-emerging zoonoses. Revue Scientifique Et Technique De L Office International Des Epizooties 23:497-511.

Brooks, C. P., J. Antonovics, and T. H. Keitt. 2008. Spatial and temporal heterogeneity explain disease dynamics in a spatially explicit network model. American Naturalist 172:149-159.

Brown, J. D., D. E. Stallknecht, J. R. Beck, D. L. Suarez, and D. E. Swayne. 2006. Susceptibility of North American ducks and gulls to H5N1 highly pathogenic avian influenza viruses. Emerging Infectious Diseases 12:1663.

Clobert, J., editor. 2001. Dispersal. Oxford University Press, Oxford.

Daszak, P., A. A. Cunningham, and A. D. Hyatt. 2000. Wildlife ecology - Emerging infectious diseases of wildlife - Threats to biodiversity and human health. Science 287:443-449.

Eubank, S., H. Guclu, V. S. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. 2004. Modelling disease outbreaks in realistic urban social networks. Nature 429:180-184.

Farnsworth, M. L., W. L. Kendall, J. Doherty, P. F., R. S. Miller, G. C. White, J. D. Nichols, K. P. Burnham, and A. B. Franklin. submitted. Targeted surveillance for highly pathogenic avian influenza in migratory waterfowl across the United States. in review.

Fernandez, R. M., and R. V. Gould. 1994. A dilemma of state power - brokerage and influence in the national health policy domain. American Journal of Sociology 99:1455-1491.

Fouchier. 2006. Global patterns of influenza A virus in wild birds. Science 312:384.

Fouchier, R. A. M., V. J. Munster, J. Keawcharoen, A. D. M. E. Osterhaus, and T. Kuiken. 2007. Virology of avian influenza in relation to wild birds. Journal of Wildlife Diseases 43:S7-S14.

Granovetter, M. 1973. Strength of weak ties. American Journal of Sociology 78:1360-1380.

Hanski, I. 1994. A practical model of metapopulation dynamics. Journal of Animal Ecology 63:151-162.

Hawkins, A. S., R. C. Hanson, H. K. Nelson, and H. M. Reeves, editors. 1984. Flyways - Pioneering Waterfowl Management in North America. US Government, Washington DC.

Jacob, J. P., J. P. Griggs, and J.B. Bender. 2008. Characterization of Small-Scale Antibiotic-Free Broiler Production in Minnesota. Journal of Applied Poultry Research 17: 412-420.

Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak. 2008. Global trends in emerging infectious diseases. Nature 451:990-U994.

Keawcharoen, J., D. van Riel, G. van Amerongen, T. Bestebroer, W. E. Beyer, R. van Lavieren, A. Osterhaus, R. A. M. Fouchier, and T. Kuiken. 2008. Wild ducks as long-distance vectors of highly pathogenic avian influenza virus (H5NI). Emerging Infectious Diseases 14:600-607.

Kilpatrick, A. M., A. A. Chmura, D. W. Gibbons, R. C. Fleischer, P. P. Marra, and P. Daszak. 2006. Predicting the global spread of H5N1 avian influenza. PNAS:0609227103.

Koehler, A. V., J. M. Pearce, P. L. Flint, J. C. Franson, and H. S. Ip. 2008. Genetic evidence of intercontinental movement of avian influenza in a migratory bird: the northern pintail (*Anas acuta*). Molecular Ecology 17:4754-4762.

Lincoln, F. C. 1935. The waterfowl flyways of North America. Volume 342.U.S. Department of Agriculture Circ. 342. 12 pgs.

Lloyd-Smith, J. O., S. J. Schreiber, P. E. Kopp, and W. M. Getz. 2005. Superspreading and the effect of individual variation on disease emergence. Nature 438:355-359.

MacDonald, J.M. 2008. The economic organization of U.S. broiler production. United States Department of Agriculture. Economic Research Service Economic Information Bulletin no. 38.

Mannelli, A., N. Ferrè, and S. Marangon. 2006. Analysis of the 1999-2000 highly pathogenic avian influenza (H7N1) epidemic in the main poultry-production area in northern Italy. Preventive Veterinary Medicine 73: 273-285.

Meyers, L. A., B. Pourbohloul, M. E. J. Newman, D. M. Skowronski, and R. C. Brunham. 2005. Network theory and SARS: predicting outbreak diversity. Journal of Theoretical Biology 232:71-81.

National Avian Influenza Surveillance Plan. 2007. National Avian Influenza Surveillance Plan, U.S. Department of Agriculture, Animal and Plant Health Inspection Service. June 29, 2007.

Naug, D. 2008. Structure of the social network and its influence on transmission dynamics in a honeybee colony. Behavioral Ecology and Sociobiology 62:1719-1725.

Olsen, B., V. J. Munster, A. Wallensten, J. Waldenstrom, A. Osterhaus, and R. A. M. Fouchier. 2006. Global patterns of influenza A virus in wild birds. Science 312:384-388.

Peterson, A. T., S. E. Bush, E. Spackman, D. E. Swayne, and H. S. Ip. 2008. Influenza A virus infections in land birds, People's Republic of China. Emerging Infectious Diseases [serial on the Internet] 10, Available from http://www.cdc.gov/EID/content/14/10/1644.htm.

Tablante, N. L., M. S. Myint, Y. J. Johnson, K. Rhodes, M. Colby, and G. Hohenhaus 2002. A Survey of Biosecurity Practices as Risk Factors Affecting Broiler Performance on the Delmarva Peninsula. Avian Diseases 46: 730-734.

Thomas, M. E., A. Bouma, H. M. Ekker, A. J. M. Fonken, J. A. Stegeman, and M. Nielen. 2005. Risk factors for the introduction of high pathogenicity Avian Influenza virus into poultry farms during the epidemic in the Netherlands in 2003. Preventive Veterinary Medicine 69: 1-11.

Urban, D., and T. Keitt. 2001. Landscape Connectivity: A graph-theoretic perspective. Ecology 82:1205-1218.

U.S. Interagency Strategic Plan. 2006. An Early Detection System for Highly Pathogenic H5N1 Avian Influenza in Wild Migratory Birds. http://www.usda.gov/documents/wildbirdstrategicplanpdf.pdf.

USDA. 1999. Part I: Reference of 1999 Table Egg Layer Management in the U.S. National Animal Health Monitoring System. A. P. H. I. S. United States Department of Agriculture, Veterinary Services. Fort Collins.

_____. 2000. Part II: Reference of 1999 Table Egg Layer Management in the U.S. National Animal Health Monitoring System. A. P. H. I. S. United States Department of Agriculture, Veterinary Services. Fort Collins.

_____. 2002. National Agricultureal Statistics Service Census of Agriculture.

_____. 2005. Part I: Reference of Health and Management of Backyard/Small Production Flocks in the United States, 2004. A. P. H. I. S. United States Department of Agriculture, Veterinary Services. Fort Collins.

_____. 2008. Reference of Management Practices on Small Enterprise Chicken Operations in the United States, 2007. A. P. H. I. S. United States Department of Agriculture, Veterinary Services. Fort Collins.

Webster, R. G., W. J. Bean, O. T. Gorman, T. M. Chambers, and Y. Kawaoka. 1992. Evolution and ecology of influenza-A viruses. Microbiological Reviews 56:152-179.

Figure 1. Dispersion of Mallards from central Alaska (Banding station 1498) from 2002 to 2006.

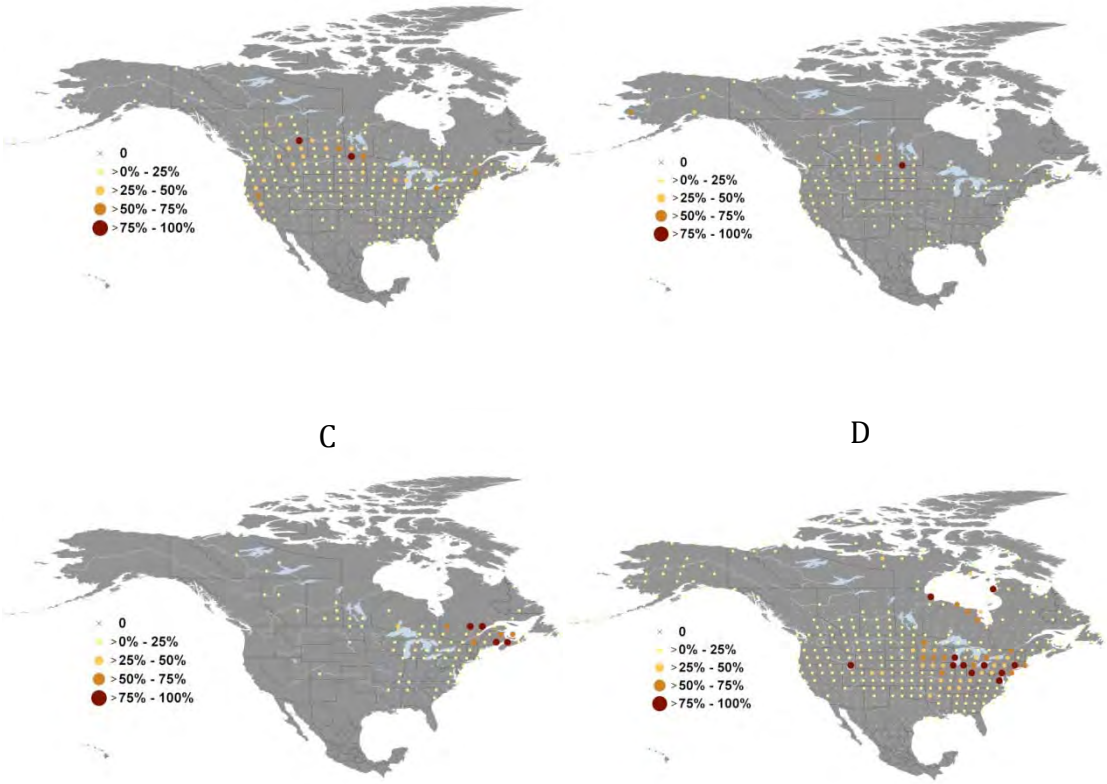


Figure 2. Banding flow. The nodes were identified using banding data from 2002 through 2006. The area of the circle is directionally proportional to the number of birds banded at the node. A) Mallards B) Northern Pintails C) American Black Duck D) Dark Geese.

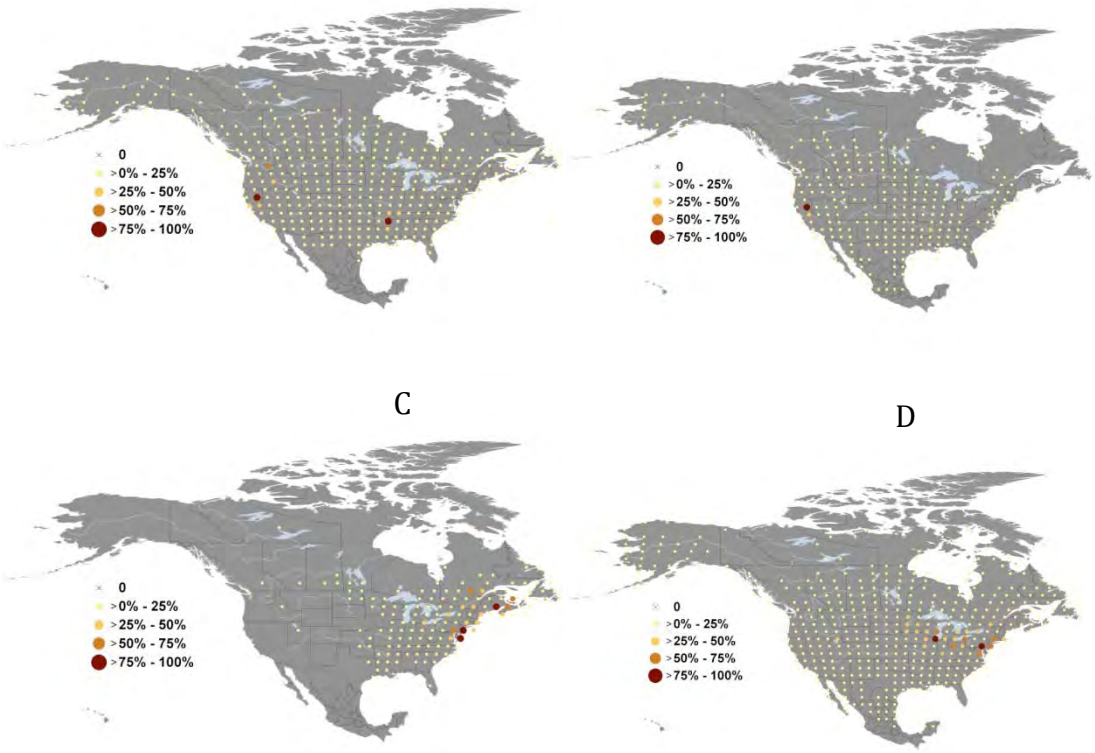


Figure 3. Recovery flow. The nodes were identified using recovery data from 2002 through 2006. The area of the circle is directionally proportional to the number of birds recovered at the node. A) Mallards B) Northern Pintails C) American Black Duck D) Dark Geese.

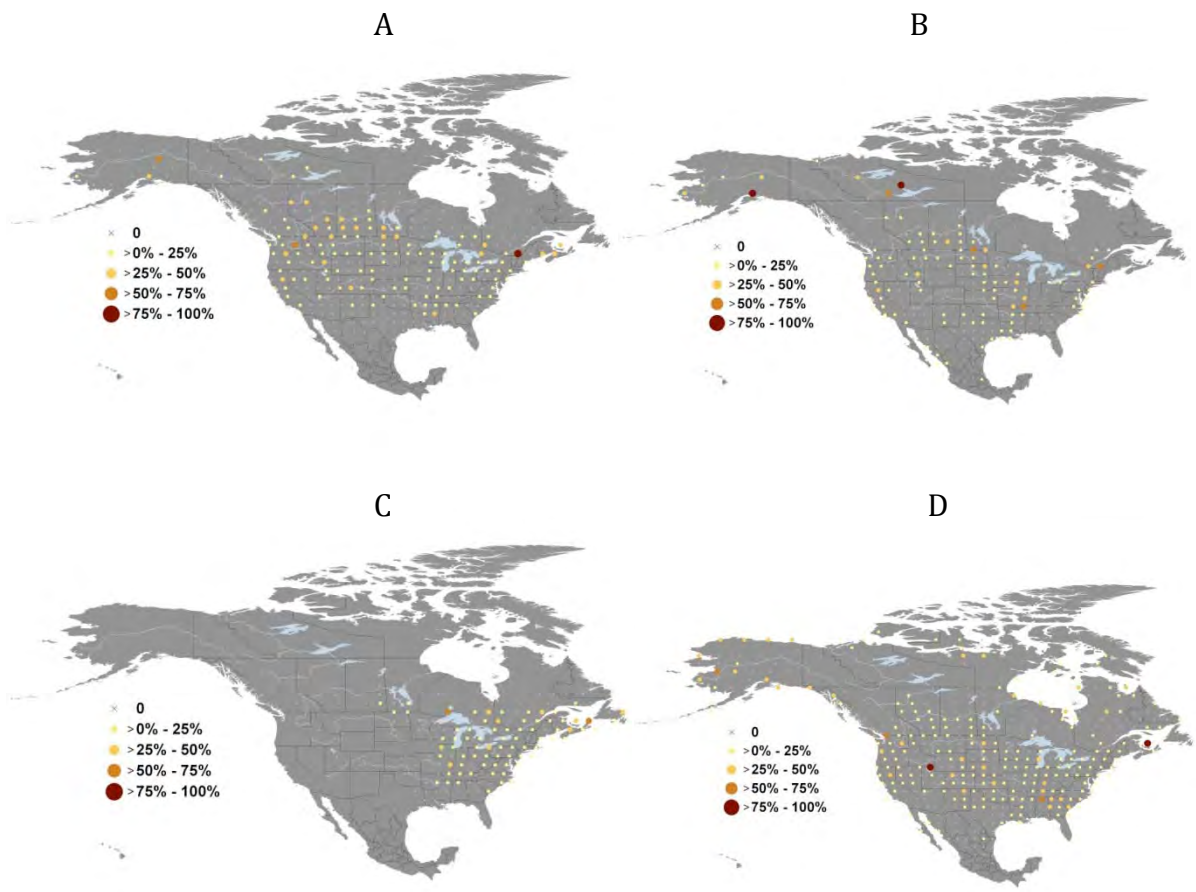


Figure 4. Minimum spanning trees. The nodes were identified using banding/recovery data from 2002 through 2006. The area of the circle is directionally proportional to the number of times the node was included in a minimum spanning tree for the contact network. A) Mallards B) Northern Pintails C) American Black Duck D) Dark Geese.

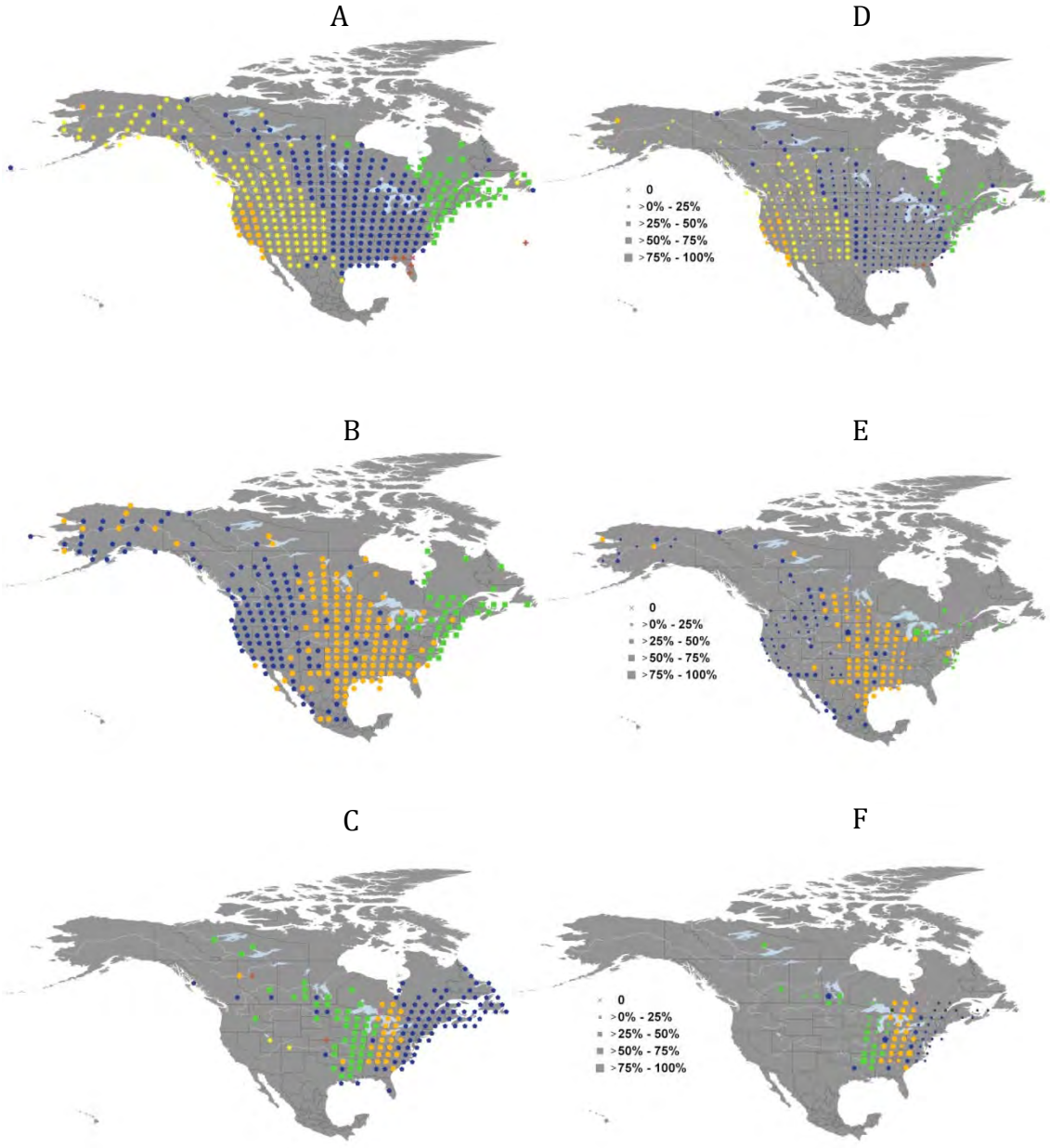
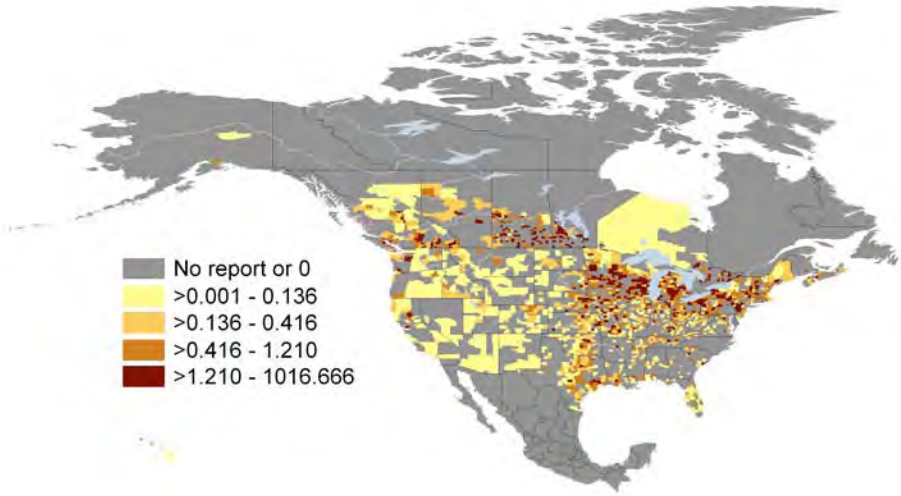


Figure 5. Clusters and satellite nodes. The nodes were identified using banding/recovery data from 2002 through 2006. The color of the circle represents the cluster the node belongs to. In figures D-F, the area of the circle is directionally proportional to the satellite metric representing the level of mixing among clusters. A) Mallard clusters B) Northern Pintail clusters C) American Black Duck clusters D) Mallard satellites E) Northern Pintail satellites D) American Black Duck satellites

A



B

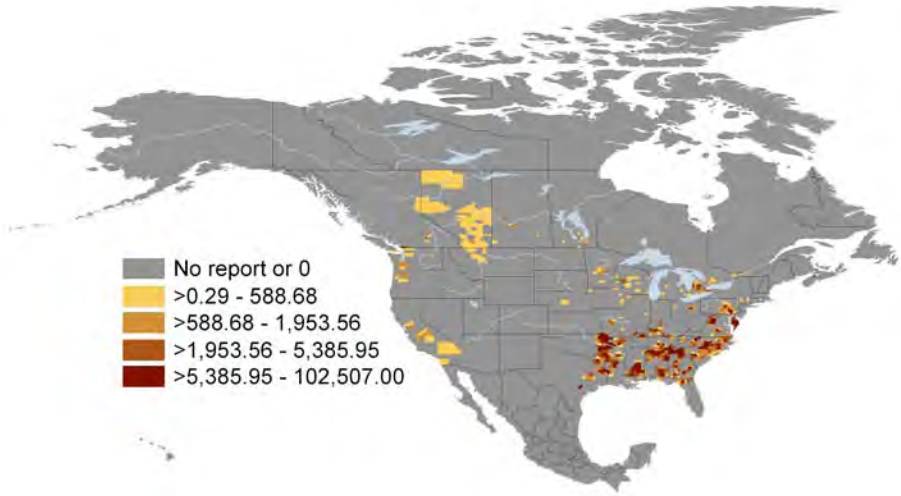


Figure 6. Average number of broiler chickens per square mile on small (<5000 birds, top) and large (≥5000 birds, bottom) farms in North America.

Section 3 – Abstracts of Ongoing Research

Abstract 1: Phylogenetic analysis of large-scale surveillance of avian influenza A virus from wild bird fecal samples across the United States

Researchers: Sergios-Orestis Kolokotronis[1], Theodore Anderson[2], Kaci VanDalen[2], Alan Franklin[2], Antoinette J. Piaggio[2]*

[1]Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY 10024

[2]National Wildlife Research Center, Wildlife Services, APHIS, United States Department of Agriculture, Ft. Collins, CO 80521

In 2006, an interagency avian influenza surveillance effort was initiated and focused on sampling across all 50 states, American Samoa, the Marshall Islands and Guam for the presence of avian influenza viruses. The goal of this surveillance effort was to accomplish early detection of HPAI. In 2006 and 2007 as part of this U.S. interagency strategic plan, USDA-Wildlife Services-National Wildlife Research Center tested approximately 75,000 fecal samples from wild bird fecal samples. The samples collected in this effort were used to generate sequences of the hemagglutinin (H) and neuraminidase (N) segments of the influenza genome. The resulting sequences were used to generate phylogenies of each subtype collected from these wild-bird fecal samples to investigate the following questions: Is there geographical structure in the distribution of H or N subtypes? Is there inflow into the continental US from Europe and/or Asia? Does the evolutionary rate of any H and/or N genes vary in the US? Are there any lineages that show rapid rates of evolution? Is there any evidence for recombination among subtypes?

There were a total of 50,184 fecal samples collected in 2006 from all 50 states) and 25,370 collected in 2007 from select states. There were 419 AI matrix positives from 2006 fecal samples and 340 in 2007 samples. One hundred and ninety-seven isolates were subtyped and all generated some sequence data. We amplified sequences from 12 of the 16 HA subtypes and each of the nine NA subtypes. Some lineages of the H and N show geographical structure and others do not. Preliminary results suggest little AI diversity shared between our sampling areas and Europe or Asia. This result suggests that most AI subtypes circulating in wild birds are endemic to the United States. Final statistical analyses to determine evolutionary rates among subtype lineages and to test for recombination are nearing completion and a manuscript is to be submitted by the end 2009.

Abstract 2: Avian influenza in waterfowl: Assessing the potential spread and impact on poultry in the United States based on wild waterfowl movement.

Researchers: The Risk Assessment Committee

Avian Influenza (AI) is a critical concern in the United States potentially affecting the health of domestic animals, humans, and wildlife. In particular, the financial impact of a highly pathogenic Asian H5N1 AI outbreak in the U.S. domestic poultry industry has been estimated to range from $187 - $623 billion.

In this study, funded by the U.S. Department of Agriculture, we investigated the links between AI, waterfowl bird migration patterns, and domestic poultry farms. Our goals were to 1) determine whether bird migration patterns can be used to predict the risk of AI spreading to and within the United States, and 2) determine the risks of wild waterfowl spreading AI to domestic poultry. Avian influenza was detected via cloacal samples collected between April 2007 and March 2008 from approximately 62,800 individuals (over 150 species) across the United States (including Alaska, Hawaii, and Guantanomo Bay). To assess waterfowl movements, we used North American bird banding and recovery data from the U.S. Geological Survey Bird Banding Laboratory database for birds recovered in 2002—2006. The number of poultry farms and bird inventory was estimated using data from the U.S. Census of Agriculture (2002) and the Canadian Census of Agriculture (2001). For all analyses, we partitioned the data into 200 -km grid cells.

No universally accepted method to assess waterfowl migration patterns exists, so we developed several metrics to measure waterfowl movements and levels of population connectivity in a spatial network of waterfowl migration using the North American bird banding data. Metrics developed included,: 1) bird banding and recovery incidence, which reflects the level of local network connectivity 2) minimum spanning tree, which reflects the minimum number of locations needed to connect the entire network of bird migration (or 95 percent of what?) 3) cluster metric, which identifies locations with similar patterns of bird banding and recovery 4) satellite metric, which that indicates connectivity between clusters.

To investigate the potential importance of waterfowl migration in spreading AI, we used logistic regression in a model selection framework to predict the pattern of AI prevalence in waterfowl across a large portion of the U.S. in space and time. Within this framework, we estimated the probability of an AI positive bird as a function of the waterfowl migration metrics, age and sex of the bird, week of AI sample collection, and ambient temperature.

The results for mallards suggest links between AI patterns and waterfowl movement, ambient temperature, and location of sampling. Based on this evidence for a relationship between waterfowl migration and patterns of AI prevalence among waterfowl, we also investigate the relationships between bird migration patterns, AI prevalence, and domestic poultry populations in order to highlight key areas to concentrate AI surveillance operations to enhance first detection of highly pathogenic Asian H5N1 in the United States.

Acknowledgments

We thank the USDA for funding and support and the USGS Colorado Cooperative Fish and Wildlife Research Unit and Colorado State University for additional support. CTW additionally acknowledges the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health. We thank Viviane Henaux, Bill Kendall, Jim Nichols, Jane Parmley, Susan L. Perkins, Michael Samuel, and Catherine Soos and others for stimulating discussions.