

# GIS Methods for Identifying Wintering Marine Bird Hotspots in the U.S. Salish Sea

---



LaTourrette, P. (2008)

*University of Washington  
Caleb Anderson and Chris Domschke  
August 2014*



## Recommended Course of Action

This report describes geospatial analysis methods for identifying wintering marine bird hotspots and possible impacts in the Salish Sea. ESRI ArcGIS tools are used to analyze 20 years of observation data to determine spatial and temporal changes in population hot spots and possible relationships to habitat improvement sites, urban and commercial development, fishing, pollution, and shoreline improvements. The relationship of wintering bird populations to possible impacts is unclear due to the complexity of the migratory bird ecological system, but continued effort and additional implementation of suggested GIS techniques could lead to a better understanding of why wintering bird populations and hot spot locations are changing. Despite the complexity of the system this study develops methods, as proof of concept, for analyzing the proximity of marine bird populations to potential impact sources. It is suggested that relationships to impacts be analyzed in both an individual species and guild method, and by a species richness method, though the former is considerably less complicated and thus a less resource intensive method. The relationships of marine birds to impacts should be measured based on the distance from marine bird hot spots to potential impact locations. A regression analysis can be used to compare the proximity of marine bird hot spots to impact locations. To accomplish this, the input dataset must meet the assumptions of the regression method used. Most of the stock ArcGIS methods assume a linear relationship between dependent and independent variables, and if this is not the case non-parametric test exclusive of ArcGIS should be considered. Species richness analysis poses its own problems, but can also be used in regression analyses, and may provide a better picture of marine health and ecosystem utility. It is suggested that the proof of concept methods in this report be used when attempting to determine areas of high concern for management actions. However, future development of techniques will be required to accurately depict the system state and needs. These developments are discussed in the discussion section as well as in the business case, and although the developments may seem time consuming they will add valuable information to the ecosystem health knowledge of the Salish Sea.

## Table of Contents

Recommended Course of Action .....	ii
Table of Figures.....	2
Table of Tables.....	3
Acknowledgements .....	3
1 Introduction .....	4
2 Design and Methods.....	9
2.1 Data Collection .....	9
2.2 Data Pre-Processing.....	11
2.3 Single Species and Guild Analysis.....	12
2.3.1 Creating a Geodatabase.....	12
2.3.2 Calculating Effort.....	13
2.3.3 Aggregate Species Observations .....	13
2.3.4 Multi Year Species Observations, Density and Hot Spots .....	13
2.4 Species Richness Analysis.....	16
2.5 Hot Spot Analysis .....	16
3 Results .....	19
4 Discussion .....	22
5 Business Case for Implementation and Development.....	26
6 Literature Cited.....	30
7 Appendices .....	32
7.1 Appendix A: Model Builder Tool Diagrams .....	32
7.2 Appendix B: Density and Hot Spot maps for all Grebe species in 3 year blocks.....	35
7.3 Appendix C: Thresholds Matrix showing some of the interconnected processes and states that can lead to or detract from a stable population of marine birds in the Salish Sea.....	48

## Table of Figures

Figure 1.1 Overview of the Salish Sea Ecoregion, and the more specific project study area as limited to the Washington State, USA portion of the ecoregion. ....	5
Figure 1.2 Social Ecological Systems Table. Shows focal scale of the study and the corresponding potential biophysical, economic, and social values for attaining the project goals. ....	6
Figure 1.3 Percent changes in mean abundance among feeding guilds for 35 common overwintering marine birds in the Salish Sea between 1978/79 and 2003-2005 (Essington et al. 2012 ) .....	7
Figure 2.1 2014 PSAMP monitoring flight path.....	10
Figure 2.2 Sample area of 300 meter buffer used for clipping observations points. Only data within the buffer were used in analyses.....	11
Figure 2.3 <i>Create Geodatabase</i> Tool graphic user interface.....	12
Figure 2.4 <i>Calculate Effort</i> Tool, graphical user interface. ....	14
Figure 2.5 <i>Species Analysis</i> Tool, graphical user interface.....	15
Figure 2.6 <i>3 yr Group</i> Tool, graphical user interface.....	16
Figure 2.7 Frequency distribution graphic from ArcMap showing large numbers of zeros in the data for species richness values (zero-inflated data). Approximately 20% of the values are zeros. Note the missing data around 11 is an artifact of the display, not a real hole. ....	17
Figure 2.8 ESRI Hot Spot Analysis Tool, graphical user interface.....	18
Figure 3.1 The Geodatabase schema , including all contained feature classes.....	20
Figure 3.2 Mean and standard deviation of the three year summed species richness values for all shoreline data. Year displayed is the last year of the three year block.....	22
Figure 5.1: Workflow diagram describing personnel, timeline and task that have been accomplished and need to be accomplish to meet the goals of the project statement.....	29
Figure 7.1 <i>Create Geodatabase</i> Tool, Model Builder model view. This figures shows the Model Builder design of the create geodatabase tool, use to create a stock geodatabase for future analyses and tools.....	32
Figure 7.2 <i>Calculate Effort</i> Tool, Model Builder model view. This figure shows the Model Builder design of the calculate effort tool, used to calculate the amount of survey effort per cell in the analysis. ....	33
Figure 7.3 <i>Species Analysis</i> Tool, Model Builder model view. This figure shows the Model Builder design of the species analysis tool, used to combine count values per unit effort across the analysis grid.....	33
Figure 7.4 <i>3 yr Group</i> Tool, Model Builder model view. This figure shows the Model Builder diagram for the tool 3 yr group which is used in analyzing multiple year groupings for counts by species or by guild. The output is a polygon feature class of counts per unit effort for the species across the 3 year range. ....	34

## Table of Tables

Table 3.1 This table displays the varying distances of total flight length, shoreline buffered flight length, and percentage of average shoreline effort (Shoreline Percentage of Effort). The total distance remains fairly stable, but does show variations up to almost 20%. ....21

## Acknowledgements

Our thanks to Trina Bayard, of Audubon Washington, and Joe Evenson, of the Washington Department of Fish and Wildlife, for their contributions of data and knowledge, as well as their unending support.

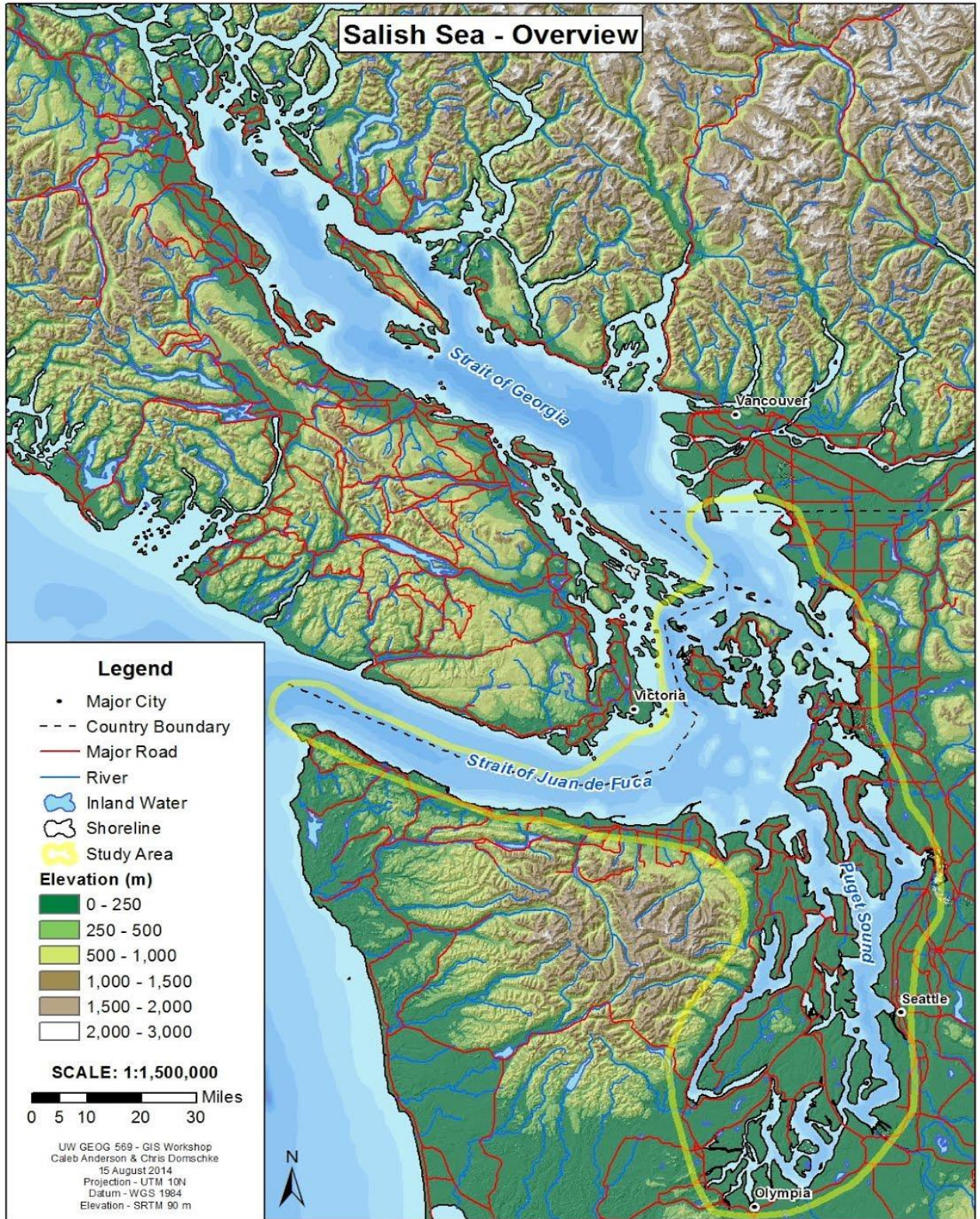
## 1 Introduction

As upper trophic predators, marine birds can provide valuable insight into the health of the marine ecosystem. In addition to their tight linkage with prey species, the availability of good time-series data sets and the general abundance of many species make marine birds relatively useful indicator species (Pearson and Hamel 2013). The Washington State Department of Fish and Wildlife (WDFW) has been using aerial transects to monitor marine bird populations in the Washington portion of the Salish Sea since 1994. This, continuous time-series, data allows for the creation of an index of abundance for marine bird species that can be compared across years, and allows for monitoring of trends in abundance.

The Salish Sea is a combination of inland marine waters extending from the north end of the Strait of Georgia and Desolation Sound in British Columbia, Canada, to the south end of the Puget Sound in Washington State, USA, and as far west as the mouth of the Strait of Juan de Fuca (Washington State Department of Natural Resources 2009). According to the Washington department of ecology, "Since 1960...the number of people living in the 12 counties bordering Puget Sound has more than doubled – from about 1.8 million to more than 4.4 million residents in 2008. In fact, 67 percent of Washington State’s entire population lives in the Puget Sound region." (Department of Ecology 2014). The increase in population has led to subsequent increases in pollution, fishing, marine traffic, developed land, and other impacts that affect migratory birds. For the purposes of this study, and due to the nature of the primary data source, the study area is limited to observations from the Washington State, USA, portion of the Salish Sea (Fig. 1.1).

Over the last 100 years there have been significant changes to the Salish Sea ecosystem and resultant impacts on migratory birds. Many species are in part, or fully dependent, on the marine waters of the Salish Sea ecosystem. "172 bird and 37 mammal species depend on the Salish Sea marine ecosystem for habitat or food; of these, 42% (72) of the bird and 78% (29) of the mammal species are highly dependent on the marine ecosystem. This means that 58% of the birds and 22% of the mammals using the Salish Sea marine ecosystem are co-dependent on marine and terrestrial ecosystems for significant life history needs, which highlights the importance of ensuring that marine conservation efforts in the Salish Sea do not stop at the high-tide mark." (Gaydos and Pearson 2011). Even though much of the focus for management has been on marine areas, the impacts to birds come from various terrestrial and marine sources which are all part of a very complex system. There is an increasing push to study indicators of health for both marine and terrestrial birds in the region as indicators of ecosystem health (Pearson and Hamel 2013).

Figure 1.1 Overview of the Salish Sea Ecoregion, and the more specific project study area as limited to the Washington State, USA portion of the ecoregion.



Some of the major components that affect this system in ways that are not limited to the Salish Sea marine area include (Anderson et. al. 2009)<sup>1</sup>:

- Air and water contamination
- Fisheries declines
- Changes in food availability or habitat along migratory routes for migratory species
- Climate change
- Anthropogenic landscape modifications, such as shoreline modifications

**Figure 1.2 Social Ecological Systems Table. Shows focal scale of the study and the corresponding potential biophysical, economic, and social values for attaining the project goals.**

Social-Ecological Systems Table				
Goal: Stable or improving populations of over-wintering marine bird species in the Salish Sea.				
Geographically		Biophysical	Economic	Social
Above Focal Scale (Extra-system effects)	Pacific Flyway and Salish Sea Ecosystem	Viable stop-over for migrating marine bird species (Audubon 2014).	Value of ecotourism to the areas. Healthy populations may signify healthy marine ecosystems, i.e. fisheries etc. (Parsons et. al. 2008)	Exchange of ideas and community with ecotourists and other local communities to protect marine avian species.
Focal Scale(System)	Salish Sea dependent Marine Bird Communities -- Specifically in WA State	Annually stable or improving over-wintering marine bird numbers and/or richness across the focal area (Project Partners' Goal).	Stable fisheries, maintained property values, high level of intrinsic value in biodiversity (SOS 2013).	Benefits of biodiversity to local communities: birders, fisheries, real estate (Pimentel et. al. 1997).
Finer Scale (Agents)	Localized Over-wintering Concentrations of Marine Birds	Annually varying, but stable populations of over-wintering marine birds. (WDFW transect count numbers)	Costs v. benefits of restoration activities and protective activities. Significance of healthy marine communities/ ecosystems.	Localized value of biodiversity. Dependence of community on healthy marine ecosystem. Benefits of restoration activities to community values and happiness.

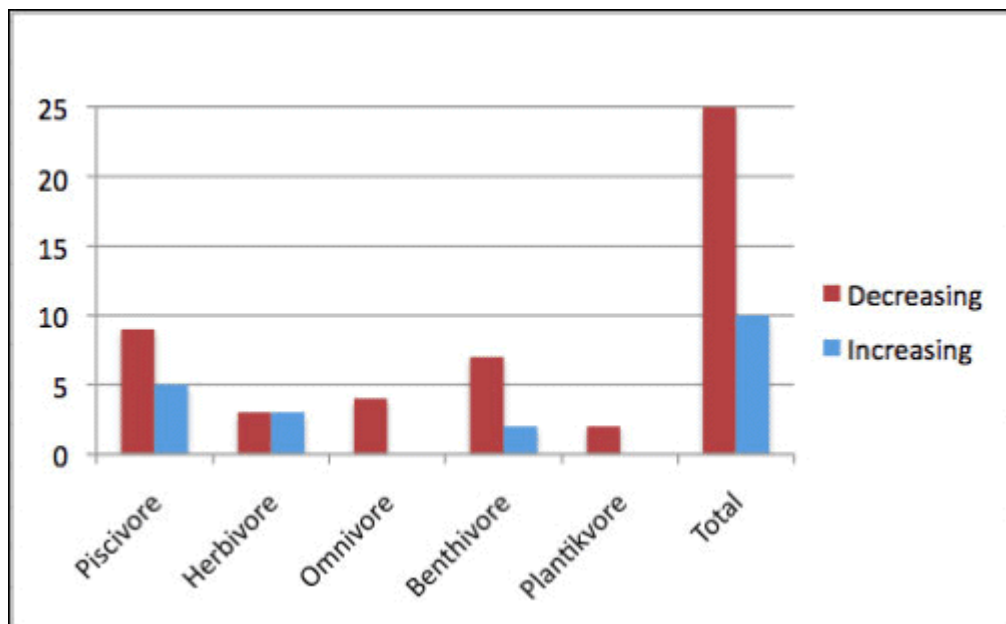
<sup>1</sup> Some of the implications of the processes listed here, and interconnectivity of the processes can be visualized in the threshold matrix located in appendix C.



The focal area for the study is the portion of the Salish Sea that is in Washington State, USA. However, effects on the health of marine bird populations extend beyond that scale, and are impacted by influences both from above the focal scale (Pacific Flyway migratory route) and from the finer scales (Sub-basin or smaller regions within the Salish Sea). In fact it may be easier to determine finer scale impacts than impacts within the focal region (Fig. 1.2). Never-the-less, the index of abundance trends that will be looked at in this study are focused on the Washington portion of the Salish Sea, which is primarily confined to Puget Sound and the Strait of Juan de Fuca.

Having determined that marine bird abundance is an important indicator of marine health, it is imperative to understand current trends. According to Bower (2014), a few wintering marine bird species have experienced increases in population, but the majority of change in bird populations in the Salish Sea has been in steady decline (Fig. 1.3). As this decline continues, it becomes more important to better understand the ecological system in which marine birds are included and what impacts are causing the decline. Current declines have been linked to decrease in forage fish populations, declines in eel grass bed area and quality as forage fish habitat, and derelict fishing gear, as well as changes in wintering locations (Bayard 2014, Therriault et. al. 2009, Plummer et. al. 2013, Good et. al. 2009, Wilson et. al. 2013).

**Figure 1.3 Percent changes in mean abundance among feeding guilds for 35 common overwintering marine birds in the Salish Sea between 1978/79 and 2003-2005 (Essington et al. 2012 )**



Geospatial analyses can be key tools in analyzing the index of abundance measures from the, long time-series, aerial transect data collected by WDFW. ESRI ArcGIS (10.2) tools,

such as hot spot analysis using the Getis-Ord  $G_i^*$  Statistic, define spatial relationships of high and low count values to determine areas that can then be used in analyzing the proximity of birds to potential impacts. Nothing more than cursory, exploratory regression analysis was completed in this project in an attempt to isolate potential impacts. However, techniques and methods have been laid out to analyze marine bird population hotspots by species, by guild, and by species richness.

The efforts of this study are primarily focused on method development and recommendations for future GIS implementation when working specifically with aerial transect data. Using the described geospatial analysis and data manipulation techniques will lead to data output that can be used to determine spatial relationships with potential impacts.

## 2 Design and Methods

WDFW aerial transects from the Puget Sound Ambient Monitoring Program (PSAMP), were the primary data source used in all of the study analyses (Essington et. al. 2011.). These data consisted of a file of recorded flight locations, and a file of all recorded observations for each year of surveys. The data were compiled and manipulated for spatial analysis using Microsoft Access, and spatial analysis was performed using ESRI ArcGIS 10.2.

WDFW and Audubon Washington were interested in looking at various impacts to marine health and the spatial distribution of marine birds to these impacts. Though this analysis did not look at the spatial distribution of birds in relation to impacts, the primary methods of data processing leading up to this analysis were designed and automated for future analyses. The main products toward this goal are tools developed to analyze hot spots of survey counts by species or guild, and to analyze species richness values across the study area.

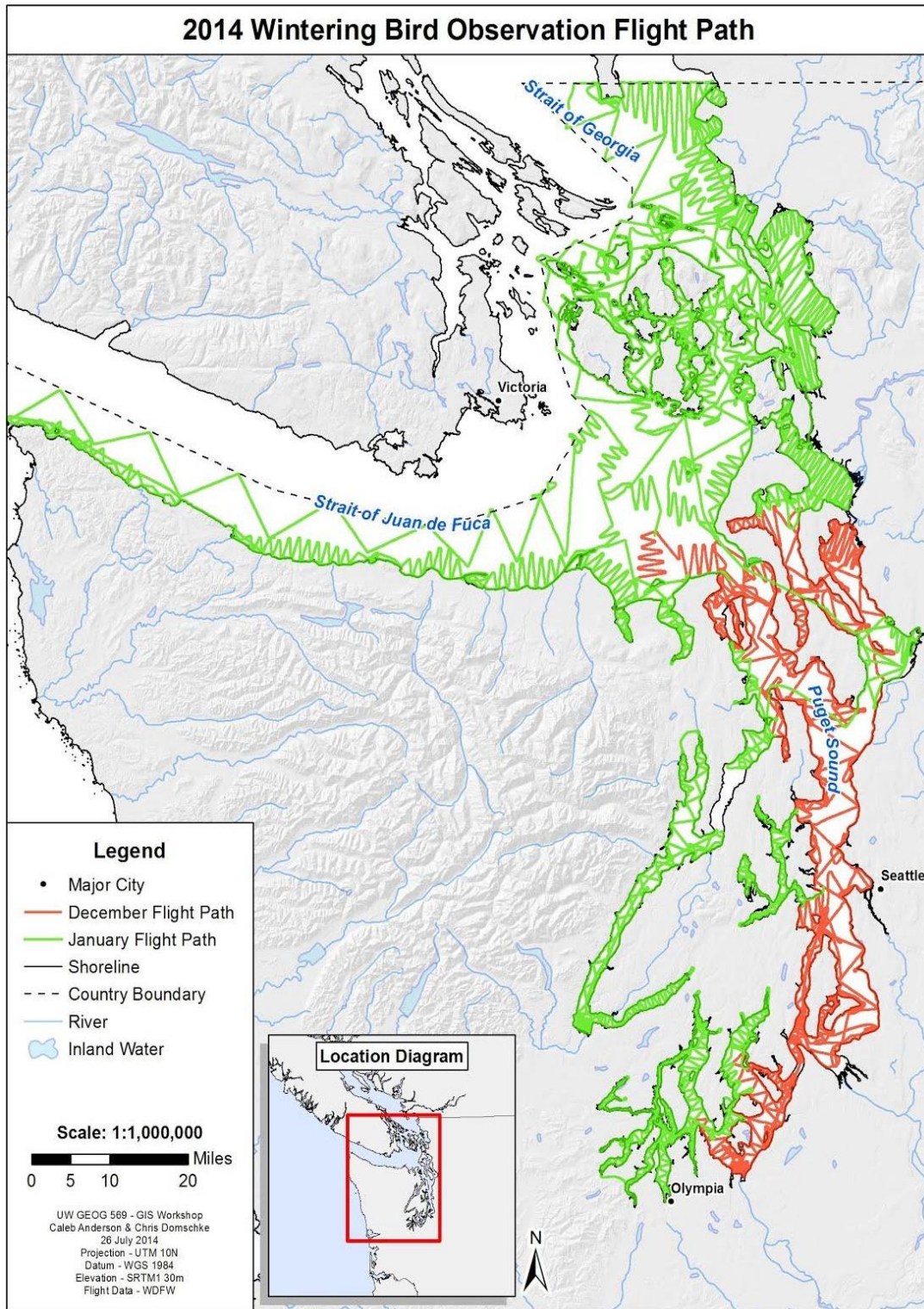
### 2.1 Data Collection

Wintering marine bird observation data (PSAMP) provided by WDFW were the primary data used in the following analyses. This data was collected using a single engine plane containing two observers and a recorder which flew designated routes. The routes consisted of shoreline transects and open water transects which were flown in a “zigzag” pattern (Fig 2.1). The seating configuration of the plane was the pilot and data recorder in the front two seats and an observer in each of two rear seats. Each observer surveyed a 50 meter transect, for a combined transect width of 100 meters when both observers were active (Joe Evenson, pers. comm.).

The data were collected into two files, an observation file and a LOG file. The LOG files record the GPS locations of the plane, along with the observer identifier and whether or not the observer was surveying at the corresponding location, as well as other relevant data that can be used in sightability calculations. The observation file contains four letter species codes and count values for all birds observed. Each observation has a corresponding location in the LOG file, but a LOG location is not limited to a single observation. Non-avian species and some inanimate objects were recorded in the observations data using their corresponding codes – these data were removed prior to analysis.

The WDFW PSAMP data consisted of surveys from the years 1994 to 2014. The data were provided as Microsoft Excel files and were restructured in order to establish a uniform format. The files were formatted to the specifications provided by WDFW and compiled in a Microsoft Access Database.

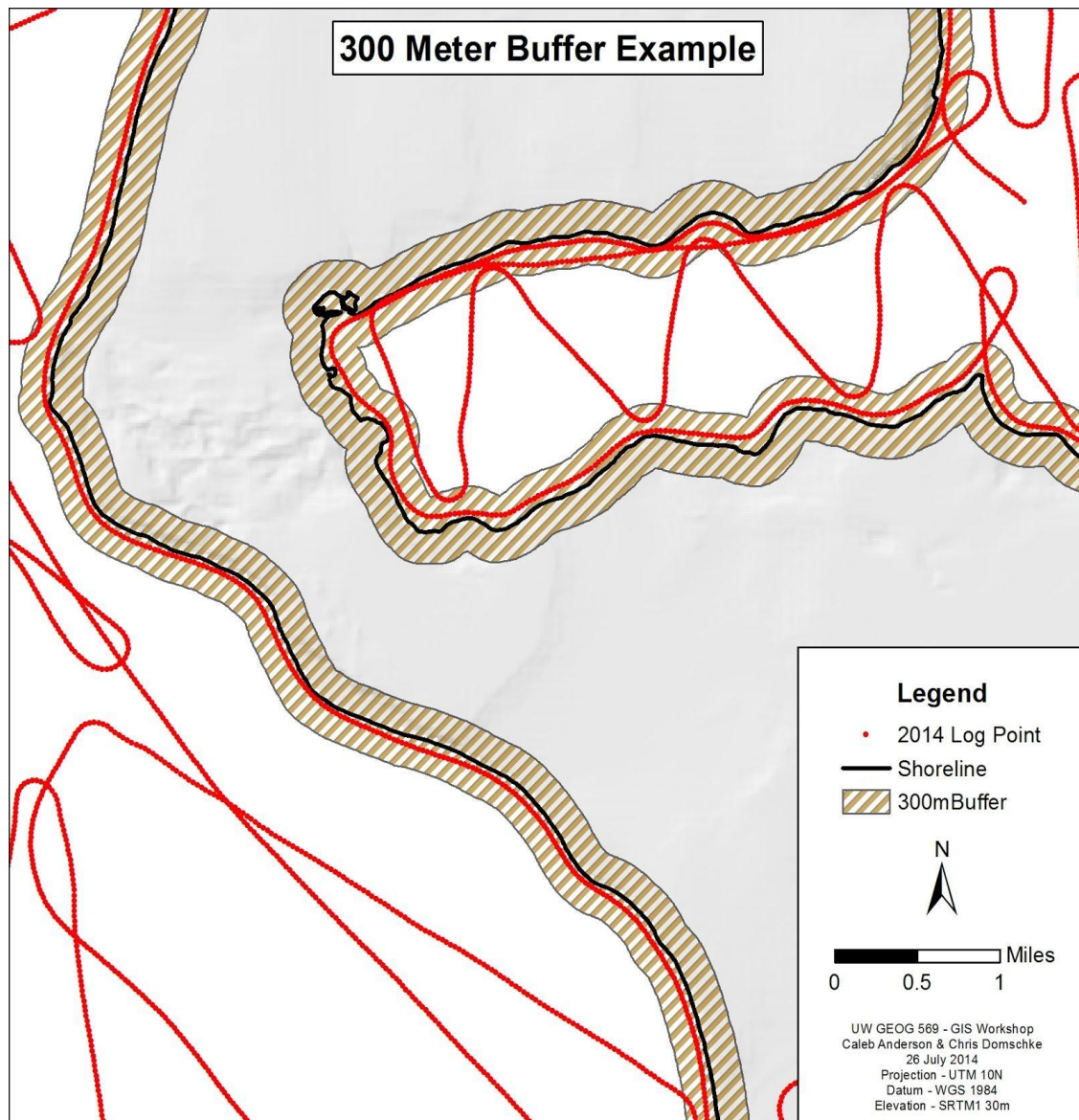
Figure 2.1 2014 PSAMP monitoring flight path.



## 2.2 Data Pre-Processing

The observation data required manipulations prior to being useable in GIS analysis of hot spots. Due to the need to interpolate values across the entire surface of the study area to accurately analyze off-shore data, only shoreline data was analyzed. It was infeasible to manually separate the shoreline transect locations from the off-shore transect locations, owing to the lack of attribute data that would designate this difference. Therefore, it was decided to create a 300 meter shoreline buffer and to clip the data points to this buffer. In this way most of the open-water data points were removed (though some remained), while maintaining the majority of the shoreline data points (Fig. 2.2). Every year of data was clipped to this buffer.

Figure 2.2 Sample area of 300 meter buffer used for clipping observations points. Only data within the buffer were used in analyses.



In addition to buffering all observations to within 300 meters of the shoreline, all observations needed to be aggregated to a grid. These aggregations were not always done in the same manner, but the grid was identical in all processes. A polygon grid of one kilometer by one kilometer cells was created. This grid was also clipped to the 300 meter buffer. Data used in analyses were then aggregated to this grid of cells.

## 2.3 Single Species and Guild Analysis

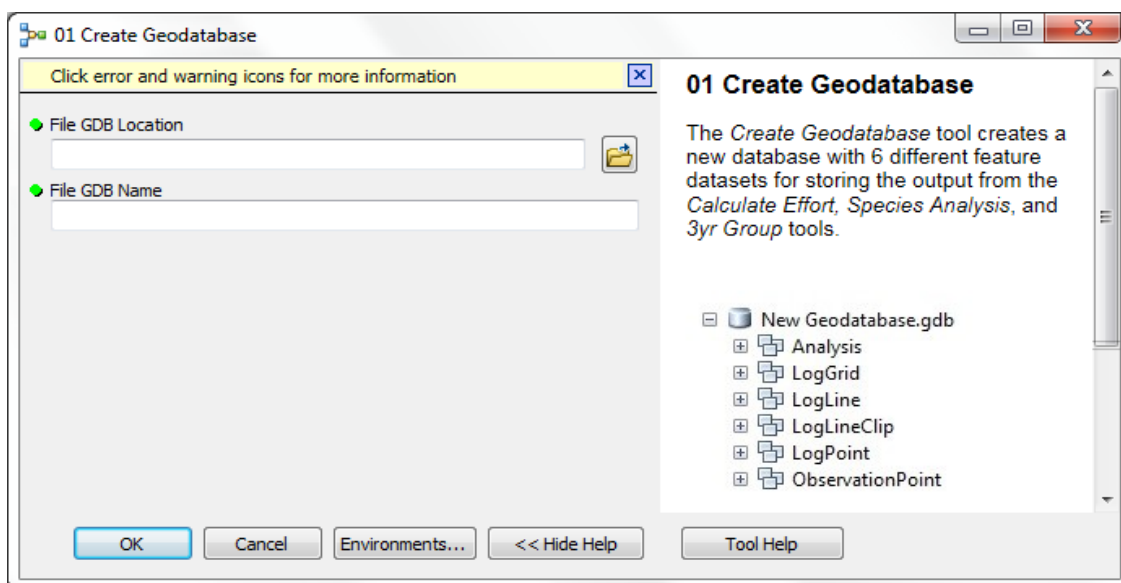
Several tools (ArcToolbox Model Builder models and python scripts) were developed in order to analyze density and hot spots for individual species, species groups, and species richness both for individual years and multiple year blocks. These tools include:

- Create Geodatabase (ArcToolbox Model Builder model)
- Calculate Effort (ArcToolbox Model Builder model)
- Species Analysis (ArcToolbox Model Builder model)
- Species Richness (Python Scripts)

### 2.3.1 Creating a Geodatabase

The *Create Geodatabase* tool creates a file geodatabase with six feature datasets for storing the outputs of the models (Fig. 2.3). This tool creates a geodatabase on the host machine with all of the necessary files to complete by species and by guild hot spot analysis. By using this tool to create a geodatabase, the other tools and scripts designed for this study are able to locate the files needed to run, and have a container for storing outputs. This tool requires the user to input a location in which to create the geodatabase, and to input a unique name for the geodatabase.

Figure 2.3 *Create Geodatabase* Tool graphic user interface.



### 2.3.2 Calculating Effort

The location of some transects vary from year to year and in order to account for that difference and to account for variations in sightability, a count per unit effort value was calculated. The method used for determining this value is built into the ArcToolbox models. The user will need to know the location of the LOG and observations tables, the location of the buffer polygon and grid polygon, the location of the output database (created in the previous tool), and the year of data on which they would like to run the tool.

The tool, *Calculate Effort*, runs several processes in order to create the following from the Log and Observation Tables (Fig. 2.4).

1. Observation Point Feature Class – displays the observation x,y data and exports it to a new feature class.
2. Log Point Feature Class – displays the LOG x,y data and exports it to a new feature class.
3. Transect Line Feature Class – creates a line feature class from the Log Point Feature Class.
4. Shoreline Transect Line Feature Class – clips the Transect Line Feature Class to the shoreline buffer of 300 meters.
5. Grid Polygon Feature Class with Transect Length Per Cell – calculates the distance of transect from the Shoreline Transect Line Feature Class that intersects with each grid cell in the analysis grid.

### 2.3.3 Aggregate Species Observations

A second tool used for determining count per unit effort, *Species Analysis*, associates the count field from the observation point feature class with the output grid from the *Calculate Effort* tool (Fig. 2.5). A field is added to the new grid with a value for species count per kilometer of effort (CountPerKmEff). This field is used in both hot spot and density analysis.

For this tool the user will need to enter the year for analysis, the study name and an optional SQL query to designate particular species. They will also need to know the location of the observation point feature class, the buffer polygon, the LOG grid and the output geodatabase. Most of these files are contained in the previously created geodatabase.

### 2.3.4 Multi Year Species Observations, Density and Hot Spots

The *3yr Group* tool is used to compare bird observations as 3 year groups instead of one year at a time (Fig. 2.6). This tool merges the output grid from the *Species Analysis* tool for the designated years. The count per kilometer effort field is averaged over the three years and the resulting value is used in hot spot and density analysis of multiple year blocks. This model also runs the built-in ESRI *Kernel Density* and Hot Spot *Hot Spot Analysis (Getis-*

Ord Gi\*) tools. The resulting outputs were used to create a map set of Grebe hot spot and density analysis (Appendix B).

For this tool the user will need to enter the study name, an output name for the multiyear block (Year Group Name), and the years to include in the block, which should be consecutive. The user will also need to know the location of the grid polygon feature class and the output geodatabase.

Figure 2.4 Calculate Effort Tool, graphical user interface.

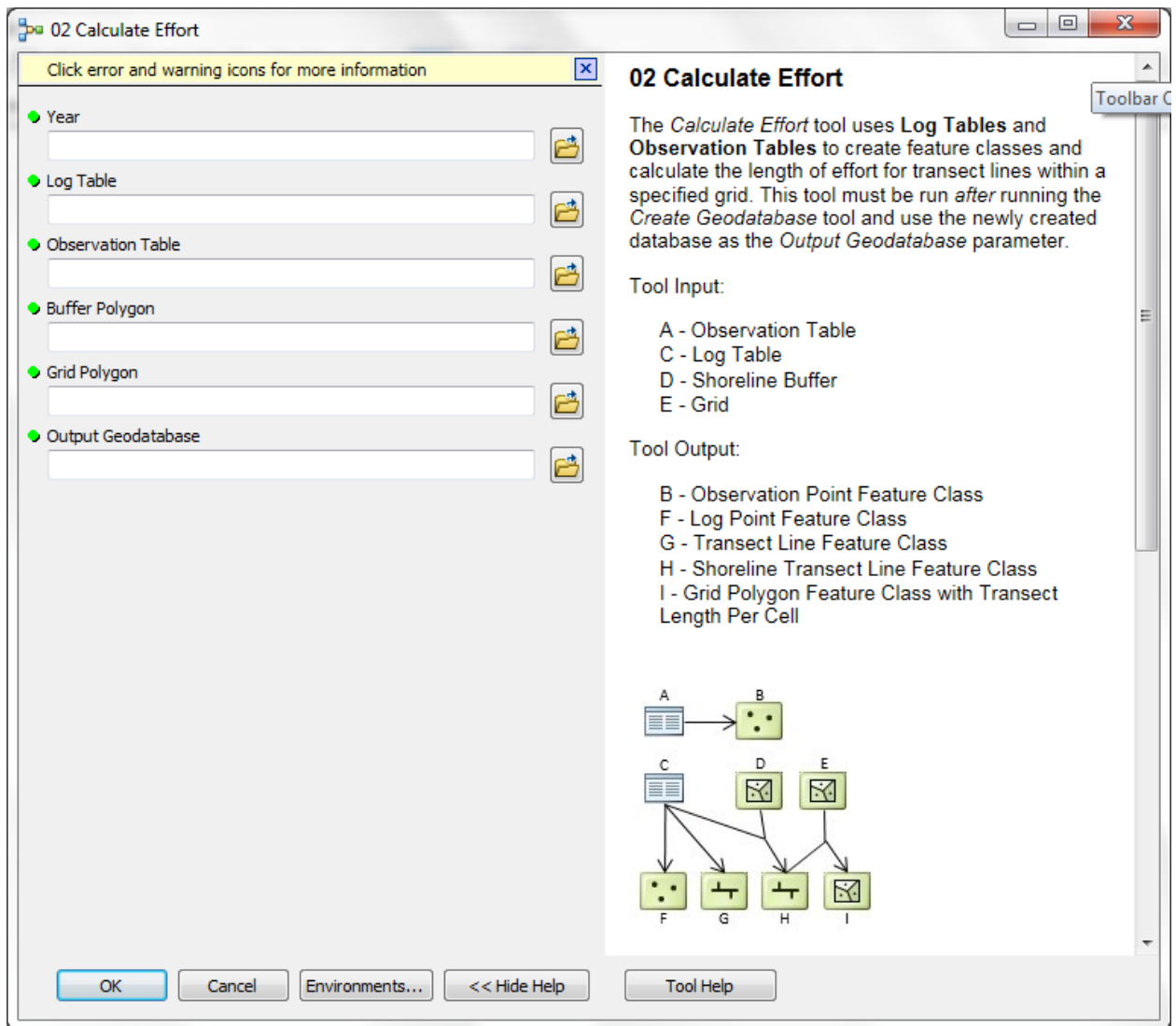




Figure 2.5 *Species Analysis Tool*, graphical user interface.

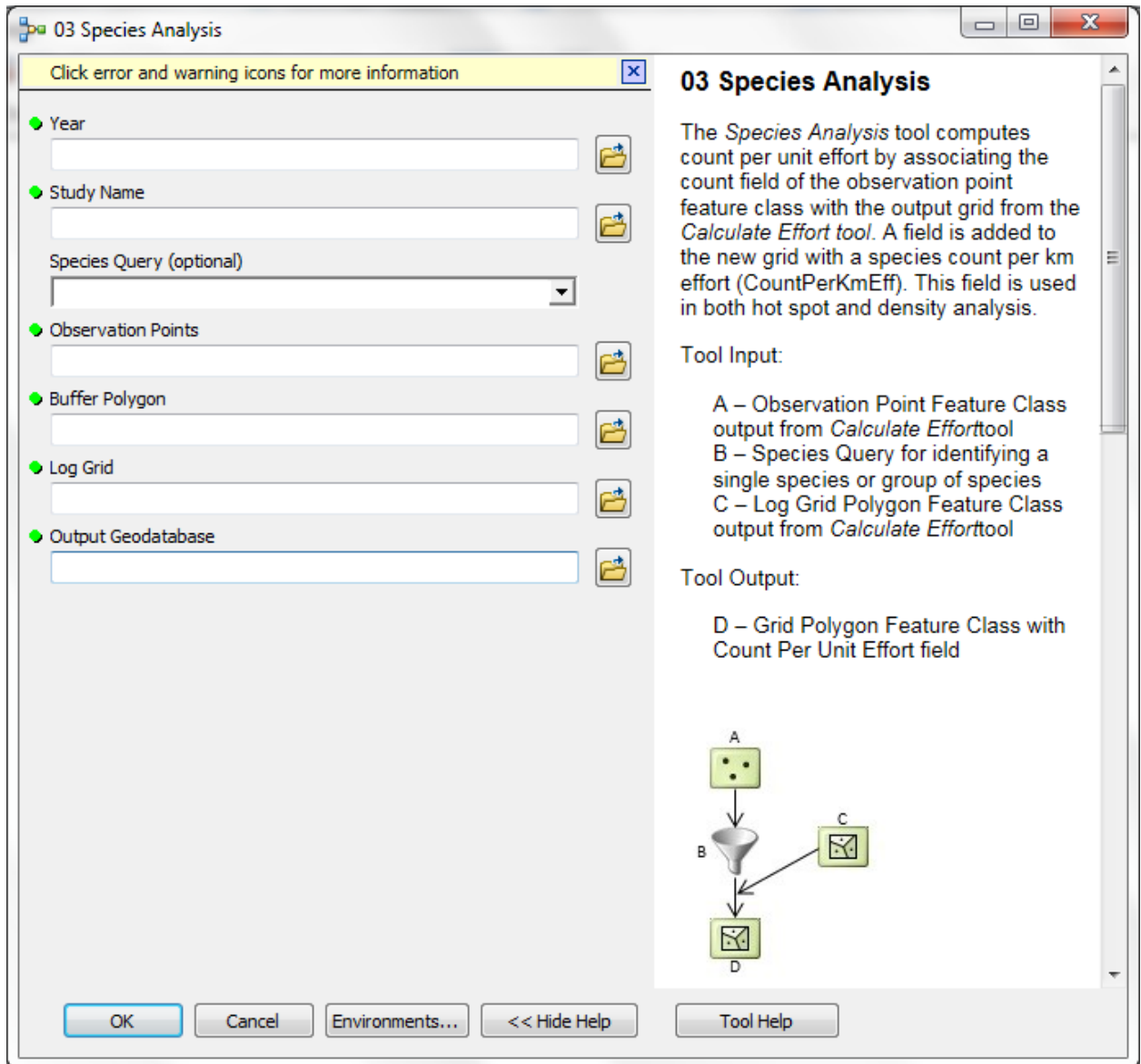
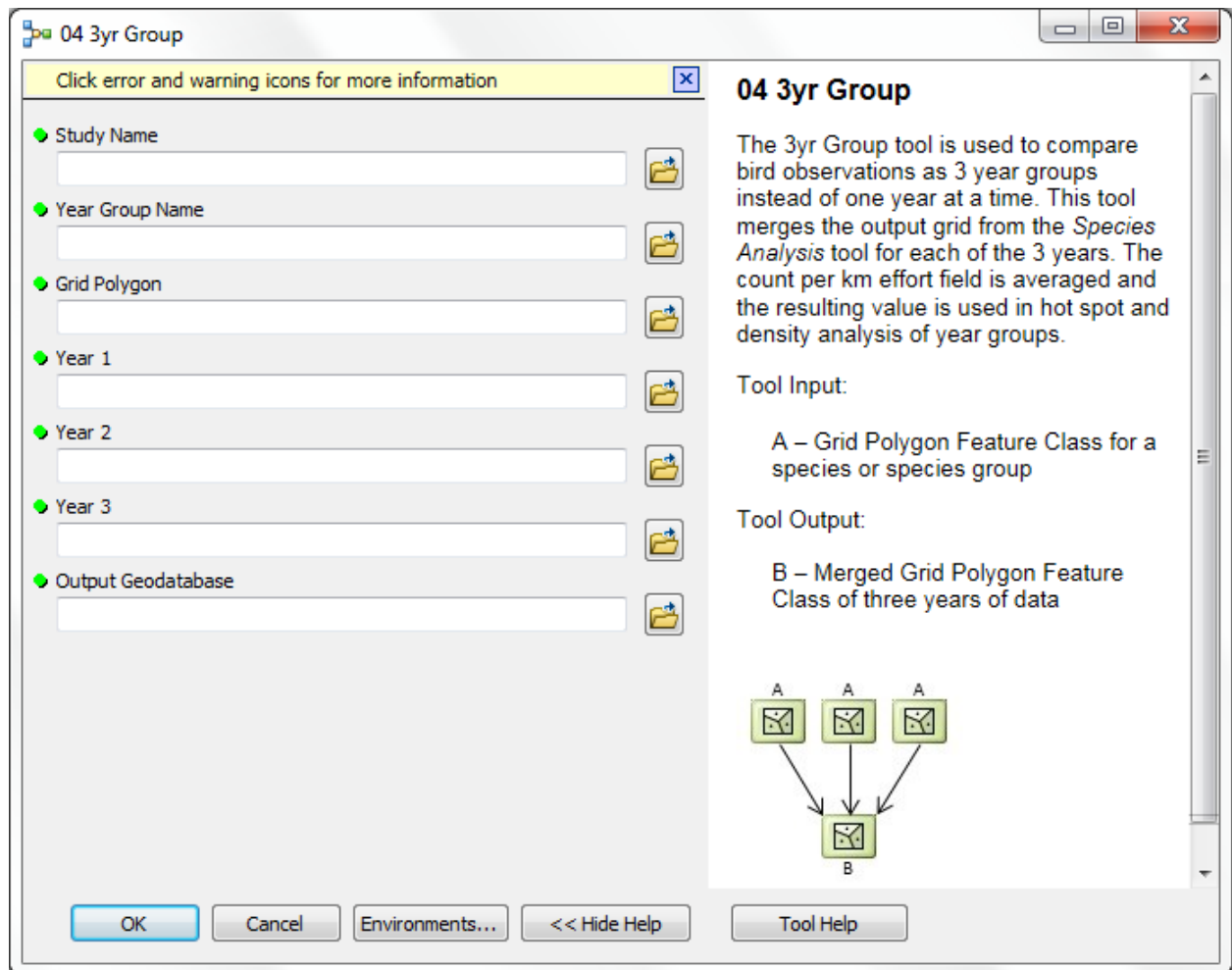


Figure 2.6 3 yr Group Tool, graphical user interface.



## 2.4 Species Richness Analysis

The final tool created for marine wintering bird shoreline analysis is a Python script that analyzes species richness, by three year groupings. This tool requires the user to add all “SpeciesRichness” tables from the corresponding Access Database to a map document, and the user must enter the workspace and the map document locations in the tool. The tool will then export Species Richness feature classes for all tables in the map document.

## 2.5 Hot Spot Analysis

One of the primary goals of the project was to create a hotspot analyses by species, by species groups, and by species richness. When conducting hot spot analysis using the ESRI *Hot Spot Analysis (Getis-Ord Gi\*)* tool, there are key options for determining the spatial conceptualization, distance method, and distance band (Fig. 2.8). These parameters all influence the neighborhood of analysis for each grid cell in the model. Defining these

parameters should be entirely based on biologically relevant values determined for the specific input species or guild. The parameters used in this design were inverse distance conceptualization of spatial relationships, Euclidean distance method, and a one kilometer distance band. Again, these parameters should be set based on best available knowledge of the habits and home ranges of the subjects.

When looking at “hot spots” for species richness, the built-in tools were not able to handle the complexity of the type of analysis that was required. Instead all zero-value cells were removed from analysis as the data is, by nature, zero-inflated (Fig. 2.7) – this included any cells that had zero survey effort in any of the years included in the three year block (Agarwal et. al. 2001, Matt Johnson, pers. comm.). Since zeros were removed, cold spots were not analyzed. Hot spots were then defined as cells with values higher than the upper limit of a 95% confidence interval.

**Figure 2.7** Frequency distribution graphic from ArcMap showing large numbers of zeros in the data for species richness values (zero-inflated data). Approximately 20% of the values are zeros. Note the missing data around 11 is an artifact of the display, not a real hole.

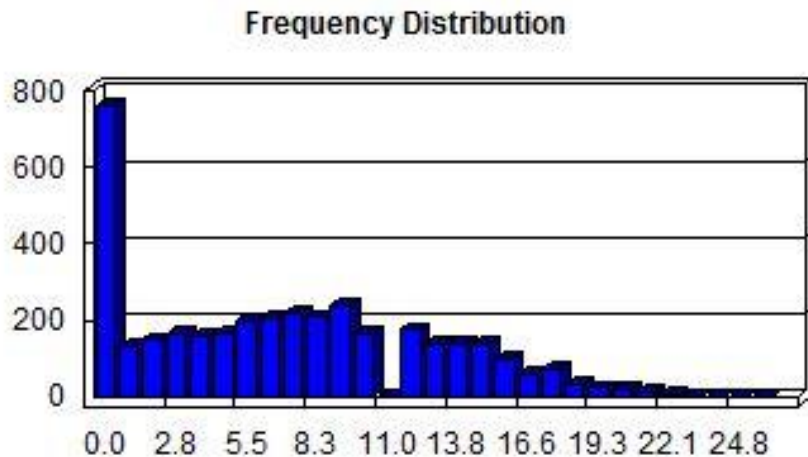
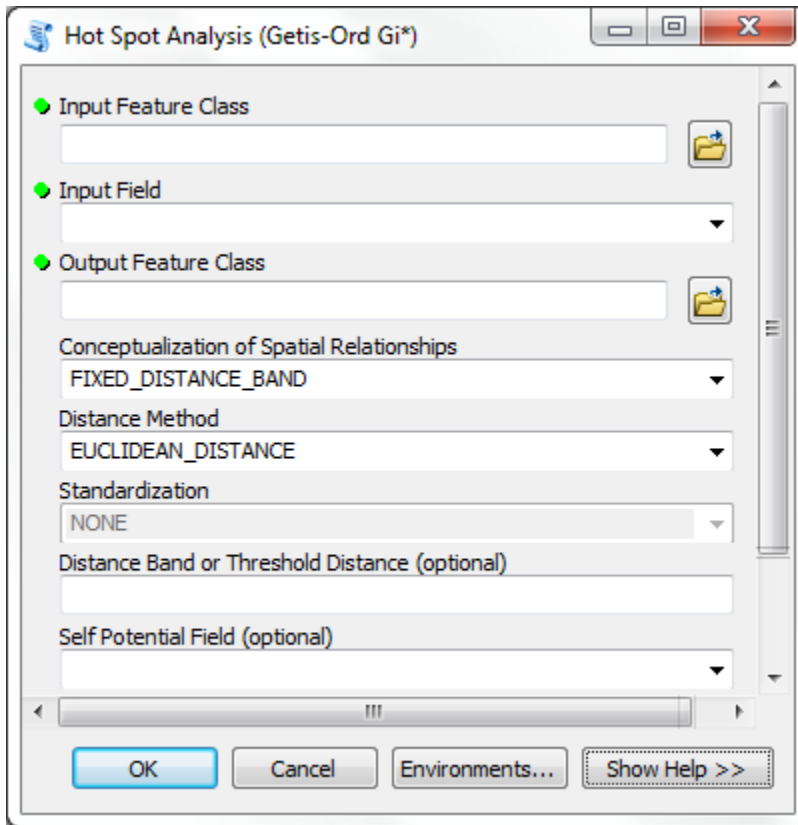


Figure 2.8 ESRI Hot Spot Analysis Tool, graphical user interface.



### 3 Results

There are four main results from this study: a Microsoft Access database containing all years of PSAMP data, an ESRI File Geodatabase and toolbox containing the results of the survey effort scripts for use in future analysis, a methodology for analyzing hot spots for the PSAMP aerial transect data, and an analysis of the current trends in species richness.

Prior to any additional developments a Microsoft Access database had to be developed to contain the data. The Access database allows for data manipulations using SQL, outside of the ESRI ArcGIS environment. This is an especially important characteristic for analyzing species richness data. The Access database allows the user to determine unique instances of data combinations, simplifying the species richness analysis. The Access database contains all of the original WDFW aerial transect LOG and observation data files, as well as species richness tables for all years, and for all three year groupings for species richness. Also included in the database are queries that can be modified to create new combinations of species richness tables, as well as those used to remove extraneous data prior to processing.

Figure 3.1 displays the geodatabase schema created by running the tools for extracting data from the LOG and observation Microsoft Access database tables into a file geodatabase and downloading data from various sources for the map background<sup>2</sup>. The geodatabase was completed and populated to serve as an initial dataset for future analyses. All of the effort feature classes and map base layers are contained in the geodatabase. The geodatabase also serves as a container for future analysis feature datasets and classes. Each run of the Species Richness script will output all feature classes to this database by default.

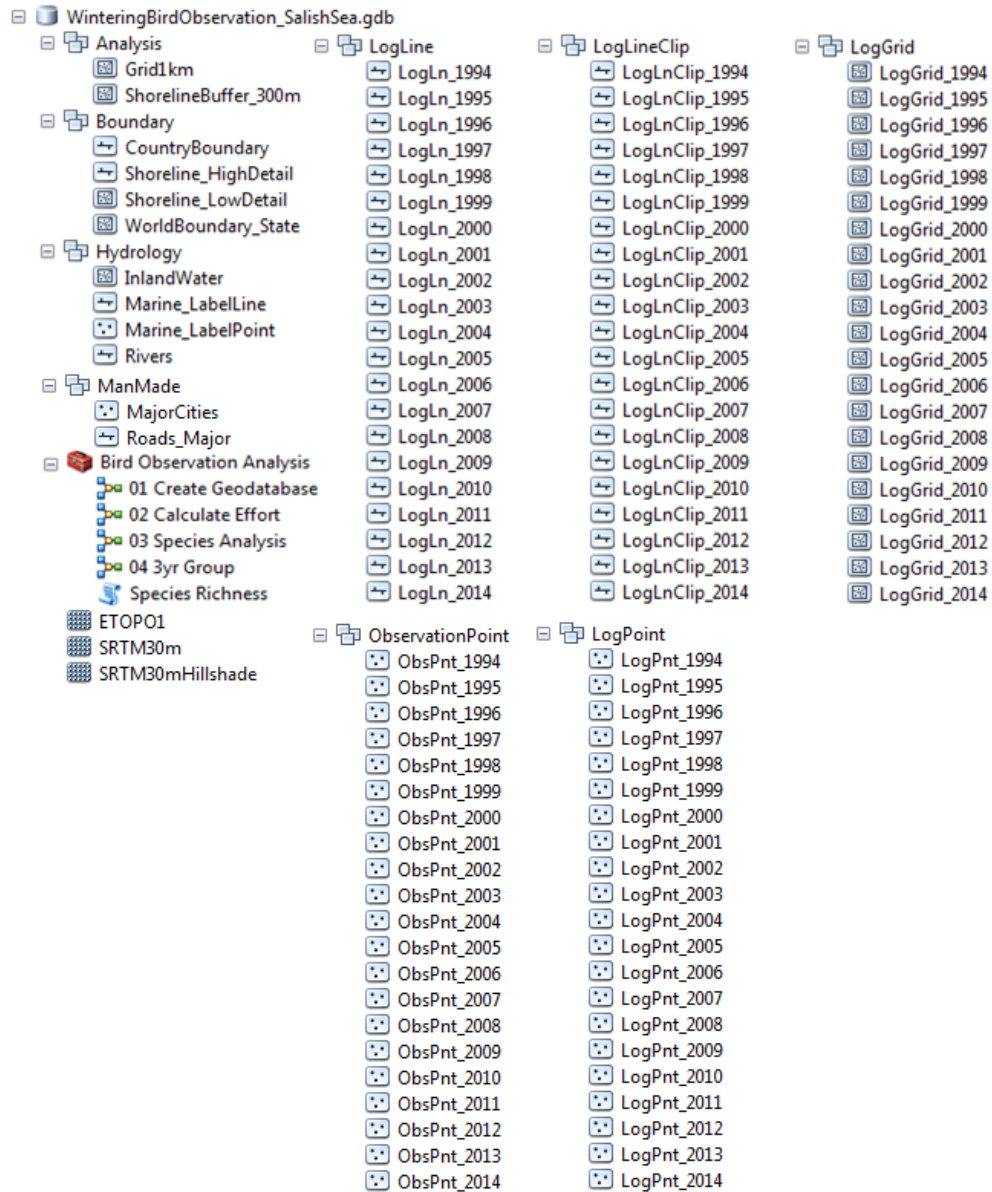
The toolbox within the geodatabase contains all of the tools necessary to analyze hot spots by species or by guild. These tools run through the processes to get from input datasets to hot spot analysis feature classes, and can also create density analyses. The Species Richness tool will create a feature class of species richness values for the input grid. The user can then examine the distribution of the data by displaying symbology that represents the appropriate section of the distribution. Step by step instructions on tool and script usage are available, but have not been included in this report<sup>2</sup>. Map outputs for these tools can be viewed in Appendix B.

A tabular comparison of yearly survey effort can be found in Table 3.1. It was produced with the *Calculate Effort* tool used to analyze each year of flight log data and displays the variation in the distance flown across years.

---

<sup>2</sup> Contact Chris Domschke at [cdomsch@uw.edu](mailto:cdomsch@uw.edu) or Caleb Anderson at [caleband@uw.edu](mailto:caleband@uw.edu).

Figure 3.1 The Geodatabase schema , including all contained feature classes.

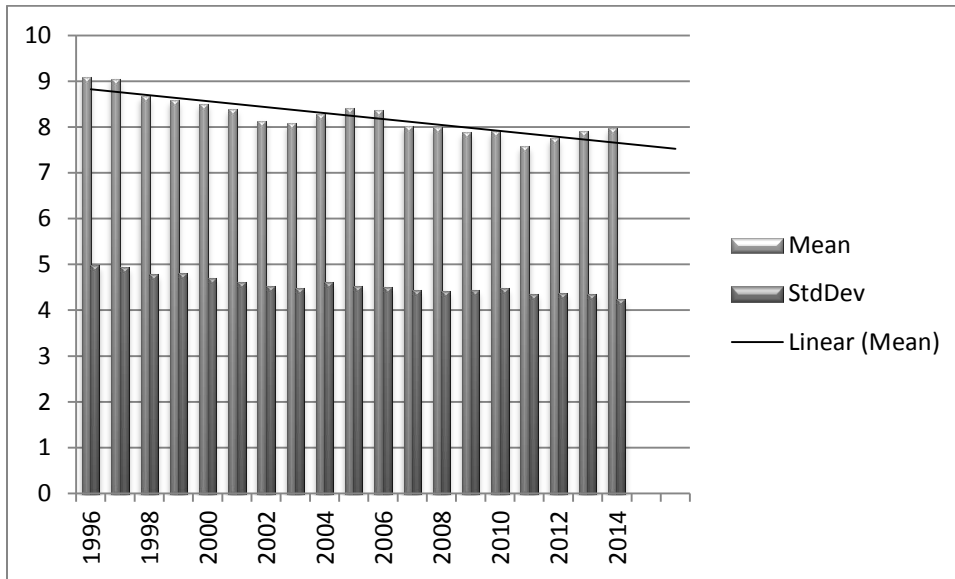


**Table 3.1** This table displays the varying distances of total flight length, shoreline buffered flight length, and percentage of average shoreline effort (Shoreline Percentage of Effort). The total distance remains fairly stable, but does show variations up to almost 20%.

<b>Year</b>	<b>Total "On Effort" Flight Length</b>	<b>Shoreline "On Effort" Flight Length</b>	<b>Shoreline Percentage of Effort</b>
1994	6,410	3,292	102.03%
1995	6,827	3,395	105.22%
1996	6,545	3,355	103.98%
1997	6,591	3,540	109.71%
1998	6,666	3,845	119.16%
1999	6,410	3,053	94.62%
2000	6,288	3,200	99.18%
2001	6,254	3,078	95.39%
2002	6,495	3,257	100.94%
2003	7,039	3,320	102.89%
2004	6,532	3,152	97.69%
2005	7,070	3,233	100.20%
2006	7,515	3,150	97.63%
2007	3,970	2,822	87.46%
2008	6,571	3,160	97.94%
2009	7,208	3,148	97.56%
2010	6,784	3,118	96.63%
2011	6,746	3,115	96.54%
2012	6,797	3,192	98.93%
2013	6,764	3,167	98.15%
2014	6,763	3,167	98.15%

Species richness was compared across rolling three year blocks beginning with years 1994-1996 and ending with the year 2014. To arrive at this data species richness was first summed across all three years in each cell of the analysis grid. The average of these summed values is displayed in Figure 3.2 along with the standard deviation, and a trend line for visual effect.

**Figure 3.2 Mean and standard deviation of the three year summed species richness values for all shoreline data. Year displayed is the last year of the three year block.**



## 4 Discussion

Three types of analytical methods were examined for looking at marine bird concentrations. From the simplest methodology to the most complex they were, species density analysis, species hot spot analysis and species richness distribution analysis. Each analysis has its positives and negatives.

Prior to delving into the pros and cons of the methodologies it is important to note that a number of tools were created to help in the data processing for the analyses. These tools have the ability to accept different grid scales and multiple years of data. The reasons for creating these models and tools was to be able to pass them along to other GIS users, streamline the process in order to reduce the chance of human error, and to save time. Because there are currently 21 years of wintering bird observation data for the Salish Sea, running each process without model or script automation would not have been possible given the timeline for the project. However, despite their utility the tools are not entirely without flaw.



The model and script creation process was time consuming as well, but mostly because of bug related issues having to do with the use of ESRI tools in their model builder. Upon researching several problems with the use of tools in the model, it was discovered that it was a bug that currently had no other solution than a work-around. In some cases, a one step process became a four step process in order to achieve the same result. Currently, the *3yr Group* tool has a persistent bug where between 5 and 55 random cells will not be included in the final analysis steps because of null values. These null values are generated during the calculation that averages the count per effort over the three years  $(([\text{CountPerKmEff}] + [\text{CountPerKmEff}_1] + [\text{CountPerKmEff}_2]) / [\text{Year\_Count}])$ . The fact that the cells are random each time the tool runs is an indication that the problem is not the input but the process. This issue will require further trouble shooting and possible tool restructuring. Although some cells were left out of the analysis in this manner, they did not contain values and were fewer than one percent of total cells, so their loss should not affect the outcome of the following analysis methods.

Density analysis is a fairly simple design. The kernel density tool is used in ArcMap to create maps that represent high density numbers of individuals across the survey. These maps provide no statistical significance, but they do accurately display the spatial distribution of individuals of a species for the survey years being analyzed. The visual representations of these maps allow the user to observe large population movements or changes over a time series (Appendix B). The map series included in the appendix clearly shows the decrease in total abundance across the survey area and the relative decreases in abundance at many locations. To reiterate, this analysis is a good way to visualize changes in abundance and location of populations over time, but does not provide any statistical significance in regards to hot spot locations.

Progressing in complexity, hot spot analysis for species or guilds does produce a map of statistically significant hot spots. These are high value cells that are surrounded by other cells of high value. The Getis-Ord  $G_i^*$  statistic is used in the analysis of hot spots (Mitchell 2009). This statistical method weights the values of the analysis cell based on the neighboring cell values. The major benefit of using this method of analysis is that it produces statistically significant results. However, there are complications with this methodology since a neighborhood must be defined for the purposes of this analysis. In cases working with biological systems, such as this one, the neighborhood should be biologically relevant. That is to say, the neighborhood should be such that it represents the natural history of the species or guild being analyzed. For instance, if the species in question has a large home range, or moves large distances in a given timeframe then the neighborhood would be fairly large, whereas if the species in question stays fairly localized or a specific temporal distribution is being analyzed the neighborhood would be correspondingly small. The second complication, in hot spot analysis, is directly related to

the neighborhood problem. It is the selection of a distance concept for the weights given to values in the neighborhood. There are three primary options that could apply to marine bird systems. First, the inverse distance relationship whereby cells more distant from the analysis cell have decreasing weight in the hot spot calculation. Second, the inverse distance squared conceptualization in which more distance cell weights will decrease in an exponential fashion rather than a linear fashion thus decreasing their importance to the analysis cell. Lastly, the fixed distance band conceptualization in which all cells within the neighborhood are given the same weight and contribute equally to the analysis cell value. As with the neighborhood, the distance conceptualization should be biologically relevant to the system at question.

The data being used in the analyses in this study should be considered instantaneous measures of abundance, or indices of abundance based on transect data. One of the assumptions of transect data is that there are no double counts. The data used in this study may violate this assumption in some cases due to the nature of the shoreline buffer clipped transect data, but when the data is interpolated across the entire surface this assumption should hold. In the case of current analyses there should not be enough double counted data to significantly violate the assumption, but future adjustments should be considered. The benefit of having annually collected transect data is that it enables establishment of an index of abundance for analyzing trends in the population data. In this case, since all of the input data was clipped to the same buffer, i.e. handled in the same fashion, the index of abundance can be analyzed without large concerns. Fortunately, these concerns are primarily confined to the analysis of species hot spots. The quantities of individuals in the density analysis should make small numbers of double counts insignificant. That is not to say that the next method for discussion, species richness distribution analysis, does not have complications; however since only one instance of a species is required to be included in species richness analysis double counts should be inconsequential.

In species richness distribution analysis, a confidence interval for the distribution of species richness values is the primary means of analysis. Species richness is the number of species present in a given area; in the case of this analysis that area is a grid cell one square kilometer in size. Species richness can be computed for a single year of data or a multiple year block of data. Processing for a single year of data is relatively simple, requiring aggregation of species observations to a cell and then listing of the species present within that cell. When looking at multiple year blocks of data, all years of observations within the block must be aggregated to the cell level prior to species richness values being calculated. If multiple year block values were calculated as averages of species richness across the years then it is possible that species could be lost in the analysis. For instance if three species are found in a cell each of two years that does not mean that they are the same three species each year – the true species richness in that two year block could be between

three and six unique species. The same logic can be extended to as many years as are in the multiple year block. In the current analysis, blocks of three years were examined.

The same logic as noted above also makes hot spot analysis on species richness values inappropriate. In the previous case, species may not have been the same between years in the multiple year blocks. In the case of hot spots, species may not be the same between cells, even if the values of neighboring cells are equivalent. For instance in a neighborhood of nine cells (the analysis cell and its eight neighbors) the species richness for every cell may be three, however without knowing which species are present in the neighborhood all that can be said is the species richness for that neighborhood is between three and (3 species\* 9 cells) 27. For a hot spot analysis to be undertaken on species richness values, a fixed distance neighborhood would be required and the value of each neighborhood would have to be calculated separately from the built-in ArcGIS tool. It is possible to examine species richness hot spots in this manner, but it is beyond the scope of this study.

A further complication when looking at species richness data is the problem of zero-inflated data. For an explanation of how zero values were handled in this data, see the *Methods* section of this paper. Suffice it to say that for the purposes of this analysis all cells with zero effort in any year, or zero species, richness are removed from analysis. The removal of the zeros in this fashion make analysis of cold spots impractical. However, the zeros had to be removed in this fashion to avoid negatively biasing the mean of the species richness observation data, and to allow for analysis of the distribution of species richness values (Matt Johnson, pers. comm.). In future analyses the data should be adjusted for zero-inflation, possibly using the methods of Agarwal et. al. (2001).

Despite these complications, examining the location of cells with species richness in excess of two standard deviations from the mean, i.e. the values which can be said to be above the mean with more than 95% certainty is a valuable exercise for determining areas of high concentrations of species. Understanding the locations of these areas of high biodiversity can be important in land management actions (Luoto et. al. 2004). Areas with high biodiversity can also indicate highly diverse and/or productive sections of the ecosystem valuable for protection and ecosystem services.

For future development, it has been suggested by Luoto et. al. that biodiversity needs to be measured on a finer scale than many national bird species datasets that look at 10 kilometer by 10 kilometer cell size areas (2004). However, Audubon in conjunction with the Farallon Institute for Advanced Ecosystem Research and the Canadian Wildlife Services are analyzing potential marine important bird areas using this scale. Their current large scale marine bird population studies in California are using 10 kilometer square grid cells and a 30 kilometer neighborhood distance band hot spot analysis in an attempt to identify

Marine Important Bird Areas(Sydeyman et. al. 2013). However, due to the differences in project scale and scope, a direct comparison of techniques is impractical. Among other issues, there are many islands and bays in the Puget Sound. Some of these islands have different kinds of habitats and bird populations on different sides and a hot spot neighborhood that is too large will include the entire island in the calculation. Opposing shores of a bay can also have different habitats and bird populations. A hot spot neighborhood that is too large will include areas that do not reflect the true nature of the cells that do have a statistical significance over 95%. Despite arguments to the contrary, analyzing species abundance at this scale may be more appropriate for management actions, particularly on a state or national scale. That said, a finer scale analysis could be useful for a more localized scope, but it is necessary to keep in mind that as the scale decreases, the quantity and quality of survey data or model data must increase or there can, and most likely will, be a mismatch in scale. It is imperative that the scale of analysis be at a scale useful to the management goals of those undertaking the analysis. This will require some tradeoffs between relative costs and benefits of survey effort and analysis power.

## **5 Business Case for Implementation and Development**

The initial project statement for this study was to analyze the proximity of marine bird concentrations (hot spots) to potential impacts – a simple set of goals that requires a complex suite of processes. In this study the dataset was predefined as the WDFW aerial transect data from the PSAMP project. These are annually collected observation data from aerial transects. After determining the context and format of the data it was decided that a Microsoft Access<sup>3</sup> database would be the most efficient tool available for data management. While compiling the data in the database, a conceptual foundation was laid for creation of a hot spot type analysis that could be used in future impact analyses. The decision was made to pursue hot spot analyses for individual species and guilds, and for species richness. The built-in tool “Hot Spot Analysis” was the chosen method for individual species hot spot analysis. During this phase hot spot analysis was considered for species richness analysis, though it was later determined that probability distributions would be a more appropriate approach.

While conceptualizing methods, a geodatabase schema was designed as a container for the background datasets and the analysis datasets to be used in the analytical process. This geodatabase was built, and data acquired to populate it included base layer feature classes. Empty feature classes were built into the schema to accommodate analysis layers and outputs. Data from the database was preprocessed for compatibility and then imported to the geodatabase.

---

<sup>3</sup> The authors did not have access to more complex DBMS systems, though other systems such as Oracle could make interactions with ArcGIS more manageable.

After creation of the geodatabase and database, analytical methods were explored, as proof of concept, to determine their applicability in answering the project statement. As mentioned, appropriate methods were determined for two primary types of analysis. Tools were created to automate the processes of data manipulation and analysis. Compiled tools output hot spot analyses for individual species and/or guilds, as well as density analyses for the same groupings. Both products allow the user to better understand the distribution of species counts across space and time. A script was compiled to automate the species richness analysis. It too is a proof of concept that requires considerable user input to create viable products other than those already created and contained in the geodatabase.

All of the products created for this study – the Access database, the ESRI geodatabase, and the proof of concept tools and analyses – will help the user to further the goals defined by the project statement. The tools and scripts can be modified by users with mid-level knowledge of ArcGIS and Python to develop future products along the current study design path.

Future development depends on the primary method of describing the system that is decided upon by the user. The choice needs to be made to pursue individual species and guild representations of marine health, or species richness representations. The former is less time consuming, while the latter may serve as a better indicator to environmental health and utility. In either case, to answer the project goals, regression analyses based on proximity of marine birds to impacts will be the next step in the larger progression. The results of this analysis will help define the locations to be considered healthy or unhealthy and which locations potentially require management actions.

Whichever path is chosen, data will need to be collected on impacts, and will need to be analyzed for compatibility with distance analyses. The primary requirement for the new datasets will be the ability to determine distance from species locations to the impact location. Additional requirements include scale of the data; small scale data can usually be combined to cover larger areas, but large scale data cannot generally be divided to facilitate analysis. Also, the measurement regime for the impacts needs to be on a useful gradient for analysis. If individual species or guild analysis is the chosen method, then analysis of impacts can begin immediately after impacts data collection. The suggested method for this analysis is geographically weighted regression analysis between bird hot spots and impact areas<sup>4</sup>. There are caveats to considering this type of regression analysis: first, it is not useful if there is no regionality in the relationship, second, this is still a form of linear regression and therefore assumes a normal distribution of data, and in the case of this analysis it would assume that site selections is linearly related to distance from impact location. If these caveats do not hold for the data other non-parametric analyses should be considered.

Should the decision be made to pursue species richness analysis, the next step would be to interpolate the PSAMP data across the entire surface of the study area yielding species

---

<sup>4</sup> For more information on this type of analysis consult *The Esri Guide to GIS Analysis Volume 2: Spatial Measurements and Statistics* by Andy Mitchell, p.218.

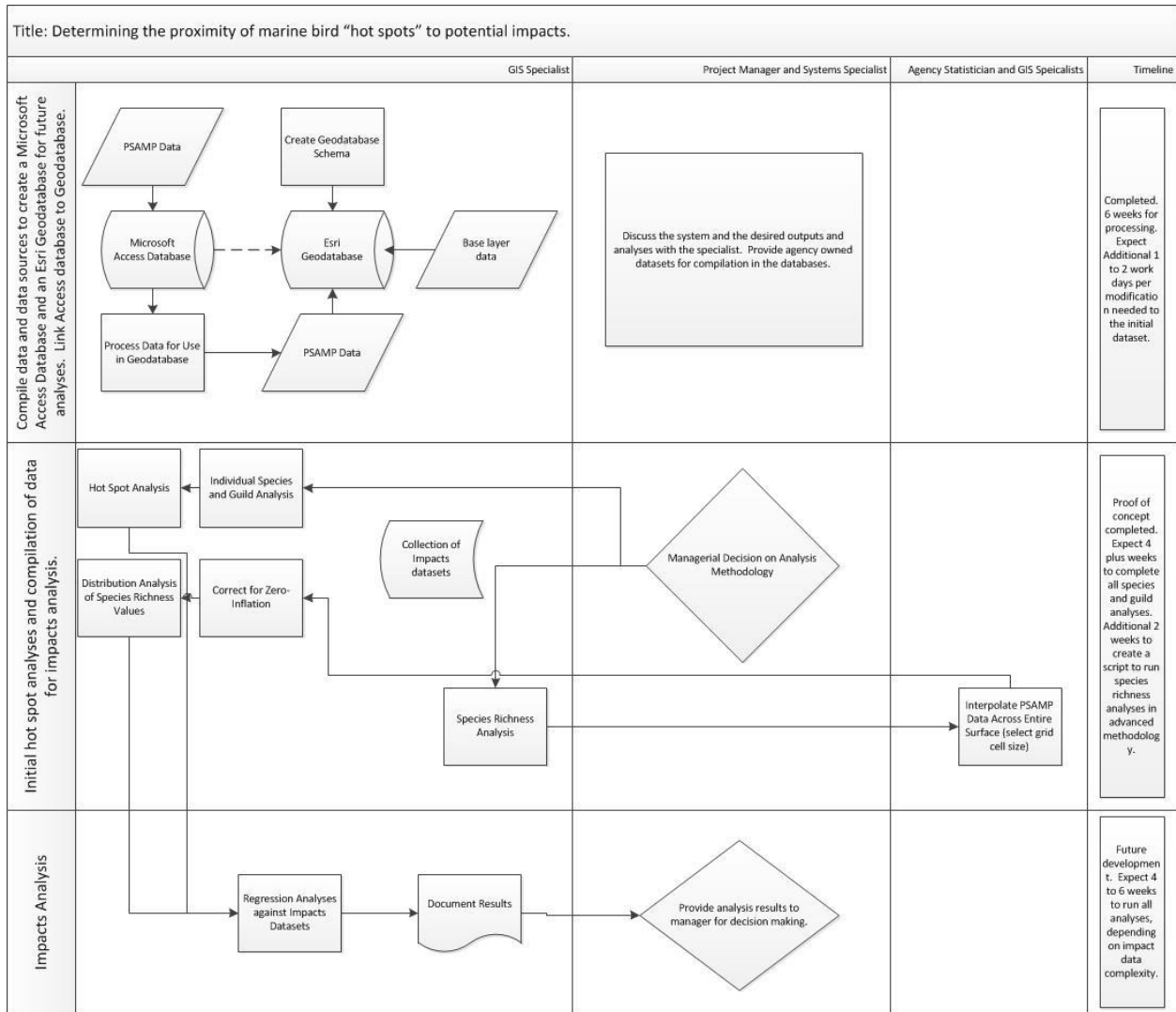
richness values for each grid cell. The grid cell size should be defined to approximate management area objectives<sup>5</sup>. For instance, if management is on a 5 square kilometer scale, then this should be the size of the grid cells. The interpolated values would then need to be corrected for zero-inflation (Agarwal et. al. 2002). At this point the distribution of the values can be examined for the desired confidence values, and with the desired probability distribution – if non-parametric. High value cells from this surface can then be used in regression analyses against distance from potential impacts.

After this point, the areas suggested as high impacts by the regression analyses should be examined on the ground and referred to agency personnel for management actions if appropriate. Figure 5.1 shows the probable workflow and estimated timeframe to accomplish all processes listed above. Time needed should be considered as forty hour work weeks.

---

<sup>5</sup> For more alternative methodology not presented due to its complexity and proof of concept nature contact Chris Domschke at [cdomsch@uw.edu](mailto:cdomsch@uw.edu) or Caleb Anderson at [caleband@uw.edu](mailto:caleband@uw.edu).

**Figure 5.1: Workflow diagram describing personnel, timeline and task that have been accomplished and need to be accomplished to meet the goals of the project statement.**



## 6 Literature Cited

- 2013 *State of the Sound: A Biennial Report on the Recovery of Puget Sound*. 2013. Tacoma, WA: Puget Sound Partnership.
- Agarwal, Deepak K., Alan E. Gelfand, and Steven Citron-Pousty. 2002. "Zero-Inflated Models with Application to Spatial Count Data." *Environmental and Ecological Statistics* 9 (4): 341–55. doi:10.1023/A:1020910605990.
- Anderson, Eric M., John L. Bower, Nysewander, David R., Evenson, Joseph R., and Lovvorn, James R. 2009. "Changes in Avifaunal Abundance in a Heavily Used Wintering and Migration Site in Puget Sound, Washington, During 1966-2007." *Marine Ornithology* 37: 19–27.
- Bayard, Trina. 2014. "Guest: How to Reverse Decline of Marine Birds." *Seattle Times / Opinion*.  
[http://seattletimes.com/html/opinion/2024321376\\_trinabayardopedmarinelifexxxml.html](http://seattletimes.com/html/opinion/2024321376_trinabayardopedmarinelifexxxml.html).
- Bower, John L. 2009. "Changes in Marine Bird Abundance in the Salish Sea: 1975 to 2007." *Marine Ornithology* 37 (March): 9–17.
- "Coastal Stewardship - Marine Birds and Forage Fish." 2014. *Audubon Washington*. Accessed August 26. <http://wa.audubon.org/coastal-stewardship-marine-birds-and-forage-fish>.
- Essington, Tim, Terrie Klinger, Tish Conway-Cranos, Joe Buchanan, Andy James, Jessi Kershner, Ilon Logan, and Jim West. 2011. "Marine Birds." *Encyclopedia of Puget Sound*. <http://www.eopugetsound.org/node/21241>.
- Gaydos, J.K., and Pearson, S.F. 2011. "Birds and Mammals That Depend on the Salish Sea: A Compilation." *Northwestern Naturalist* 92: 79–94.
- Good, Thomas P., June, Jeffery A., Etnier, Michael A., and Broadhurst, Ginny. 2009. "Ghosts of the Salish Sea: Threats to Marine Birds in Puget Sound and the Northwest Straits from Derelict Fishing Gear." *Marine Ornithology* 37: 67–76.
- Luoto, Miska, Virkkala, Raimo, Heikkinen, Risto K., and Rainio, Kalle. 2004. "Predicting Bird Species Richness Using Remot Sensing in Boreal Agricultural-Forest Mosaics." *Ecological Applications* 14 (6): 1946–62.
- Mitchell, Andy. 2009. *The Esri Guide to GIS Analysis Volume 2: Spatial Measurements & Statistics*. Redlands, CA: Esri Press.
- Parsons, Matt, Ian Mitchell, Adam Butler, Norman Ratcliffe, Morten Frederiksen, Simon Foster, and James B. Reid. 2008. "Seabirds as Indicators of the Marine Environment." *ICES Journal of Marine Science: Journal Du Conseil* 65 (8): 1520–26. doi:10.1093/icesjms/fsn155.
- Pearson, S.F., and Hamel, N.J. 2013. *Marine and Terrestrial Bird Indicators for Puget Sound*. Olympia, WA: Washington Department of Fish and Wildlife, and Puget Sound Partnership.
- Pimentel, David, Christa Wilson, Christine McCullum, Rachel Huang, Paulette Dwen, Jessica Flack, Quynh Tran, Tamara Saltman, and Barbara Cliff. 1997. "Economic and Environmental Benefits of Biodiversity." *BioScience* 47 (11): 747–57. doi:10.2307/1313097.



- Plummer, Mark L., Harvey, Chris J., Anderson, Leif E., Guerry, Anne D., and Ruckelshaus, Mary H. 2013. "The Role of Eelgrass in Marine Community Interactions and Ecosystem Services: Results from Ecosystem-Scale Food Web Models." *Ecosystems* 16: 237–51. doi:10.1007/s10021-012-9609-0.
- "State Board on Geographic Names Approves 'Salish Sea.'" 2009. *Washington State Department of Natural Resources*.  
[http://www.dnr.wa.gov/ResearchScience/News/Pages/nr09\\_177.aspx](http://www.dnr.wa.gov/ResearchScience/News/Pages/nr09_177.aspx).
- Sydemann, William J., Losekoot, Marcel, Santora, Jarrod A., Thompson, Sarah Ann, Distler, Trisha, Weinstein, Anna, Smith, Melanie A., Walker, Nathan, and Morgan, Ken H. 2013. *Hotspots of Seabird Abundance in the California Current: Implications for Important Bird Areas*. Farallon Institute for Advanced Ecosystem Research, Audubon California, Audubon Alaska, and Canadian Wildlife Services.
- Therriault, Thomas W., Hay, Douglas E., and Schweigert, Jacob F. 2009. "Biological Overview and Trends in Pelagic Forage Fish Abundance in the Salish Sea (Strait of Georgia, British Columbia)." *Marine Ornithology* 37: 3–8.
- "Threats to Puget Sound." 2014. *Department of Ecology*. Accessed August 26.  
[http://www.ecy.wa.gov/puget\\_sound/threats.html](http://www.ecy.wa.gov/puget_sound/threats.html).
- Wilson, Scott, Eric M. Anderson, Amy S. G. Wilson, Douglas F. Bertram, and Peter Arcese. 2013. "Citizen Science Reveals an Extensive Shift in the Winter Distribution of Migratory Western Grebes." *PLoS ONE* 8 (6): e65408.  
doi:10.1371/journal.pone.0065408.

## 7 Appendices

### 7.1 Appendix A: Model Builder Tool Diagrams

Figure 7.1 *Create Geodatabase Tool*, Model Builder model view. This figure shows the Model Builder design of the create geodatabase tool, use to create a stock geodatabase for future analyses and tools.

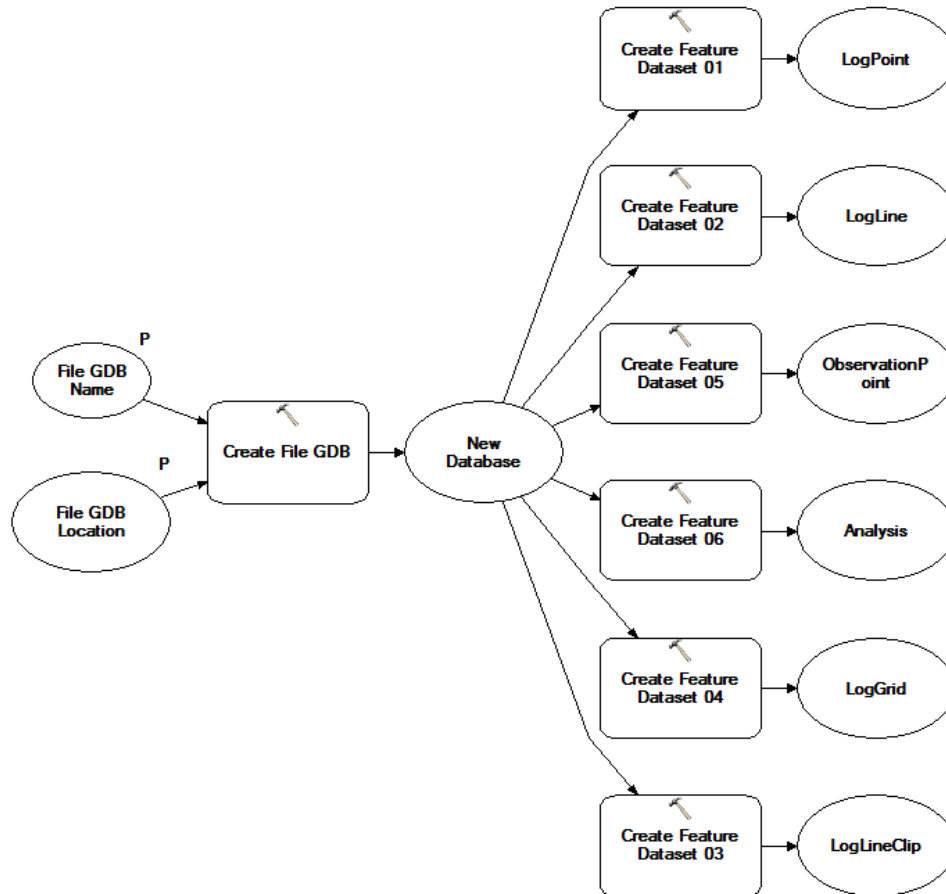


Figure 7.2 *Calculate Effort Tool*, Model Builder model view. This figure shows the Model Builder design of the calculate effort tool, used to calculate the amount of survey effort per cell in the analysis.

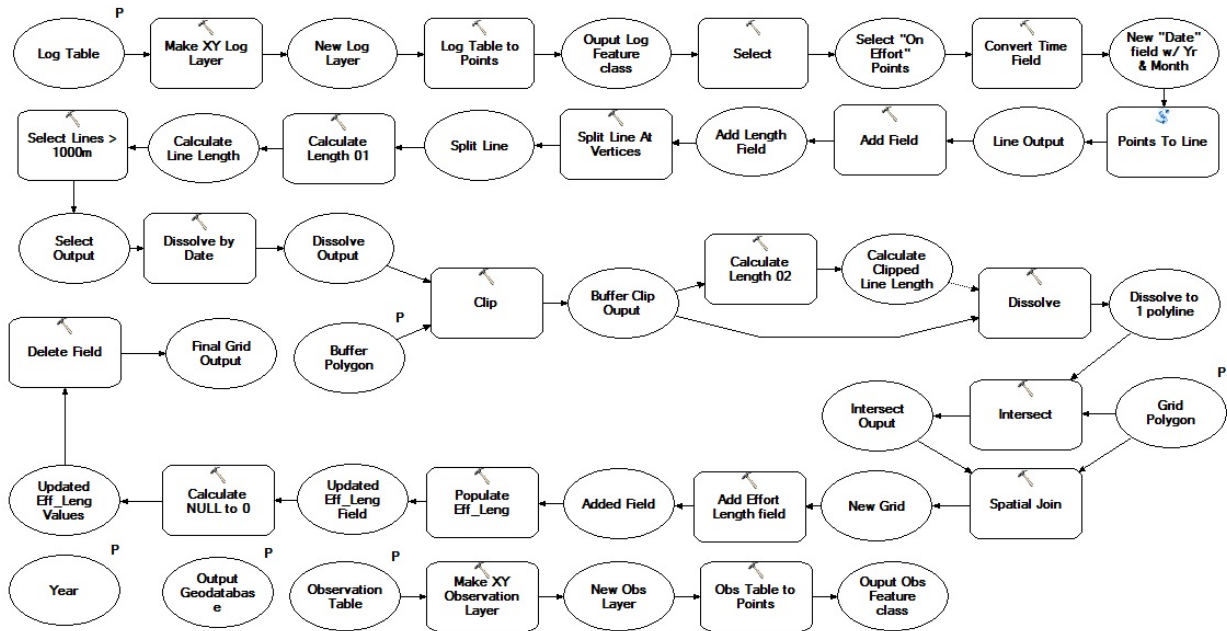


Figure 7.3 *Species Analysis Tool*, Model Builder model view. This figure shows the Model Builder design of the species analysis tool, used to combine count values per unit effort across the analysis grid.

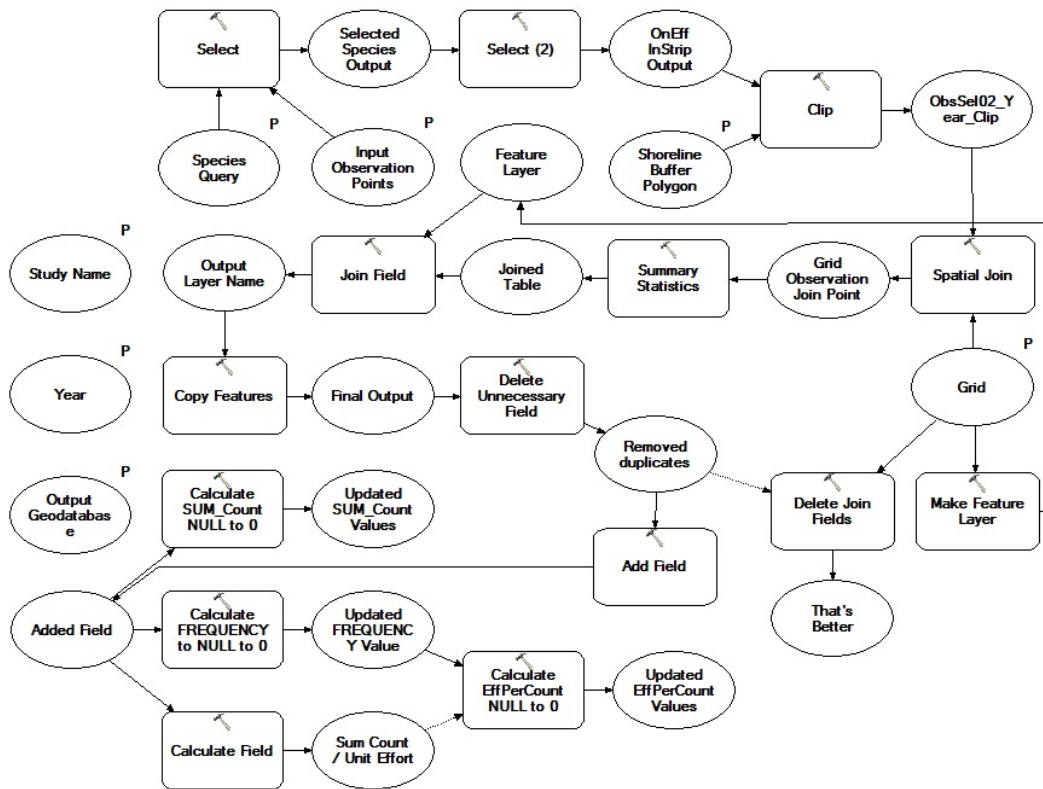
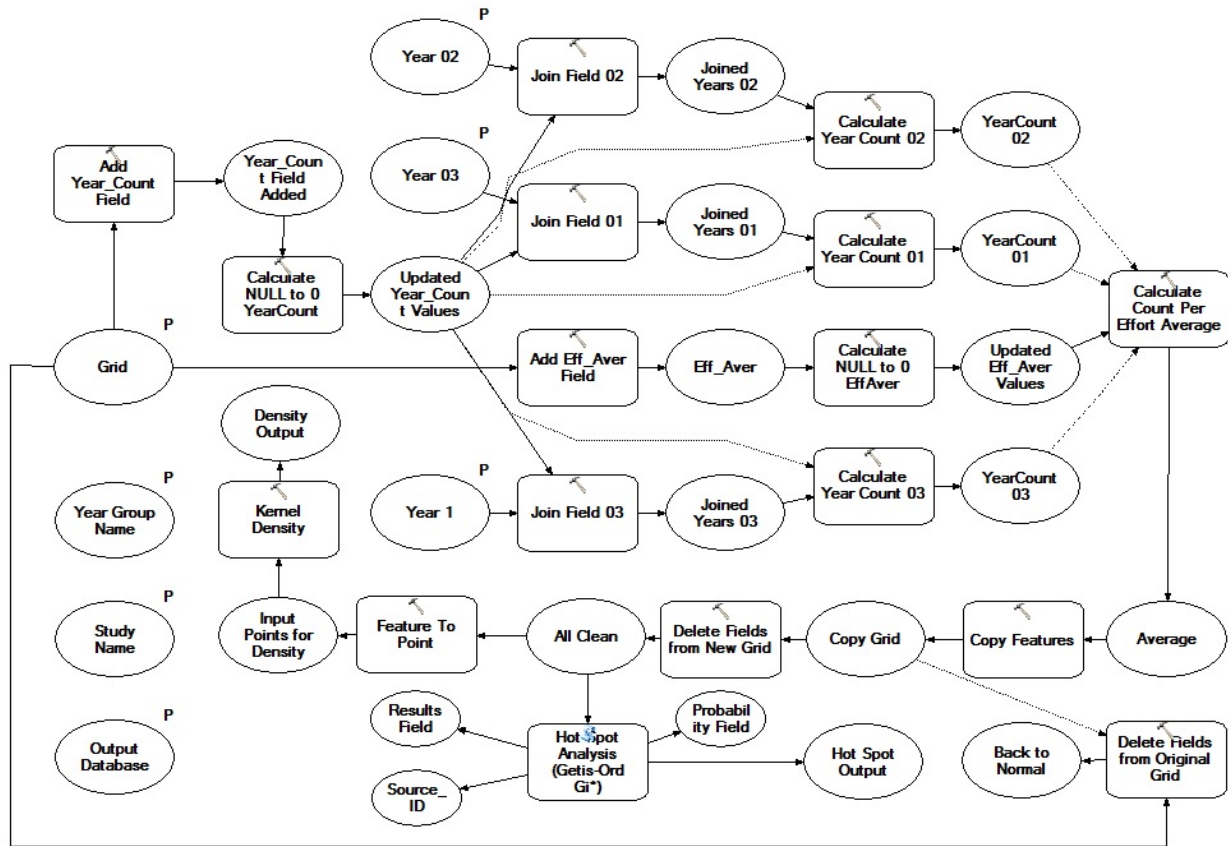
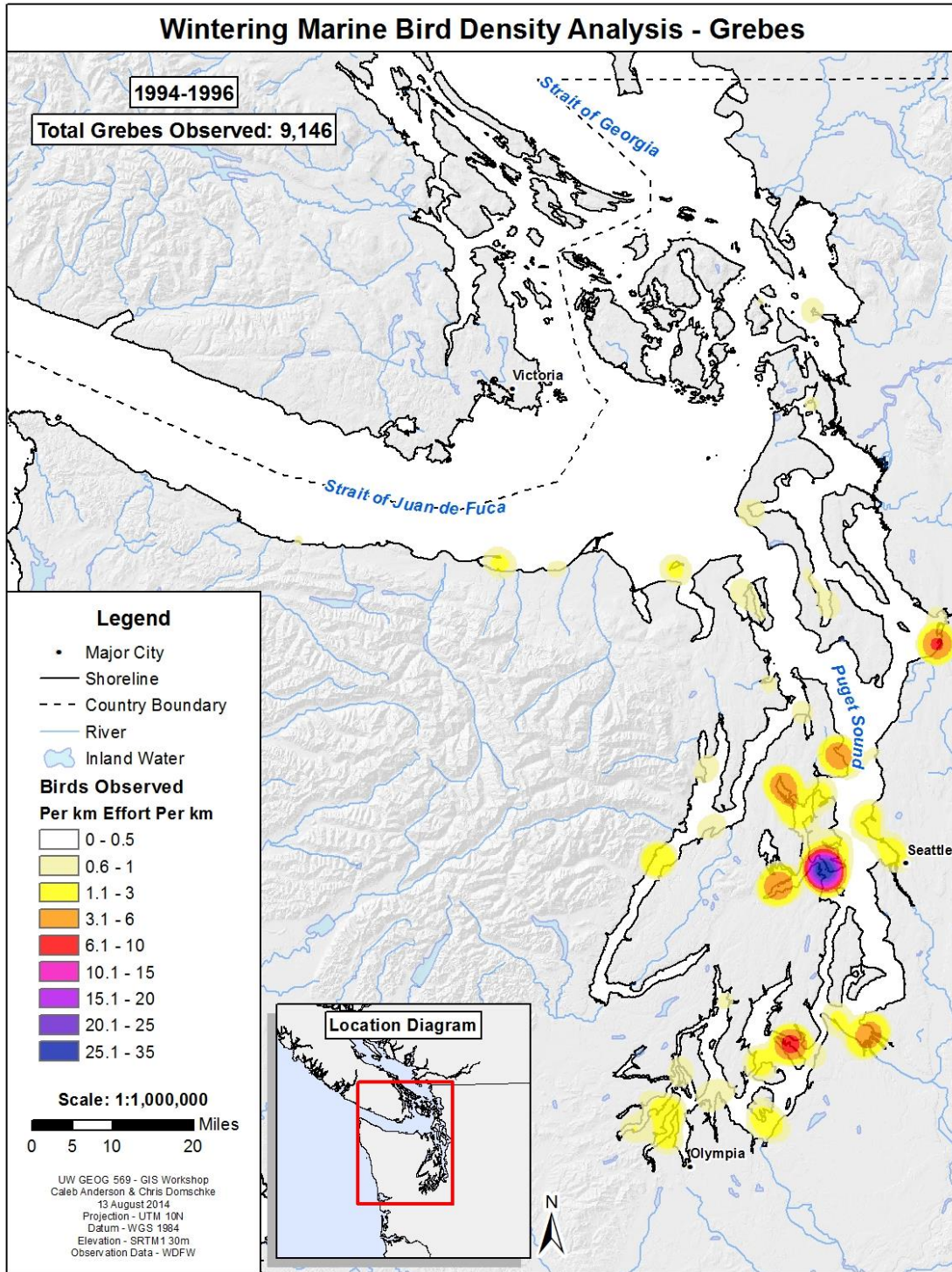


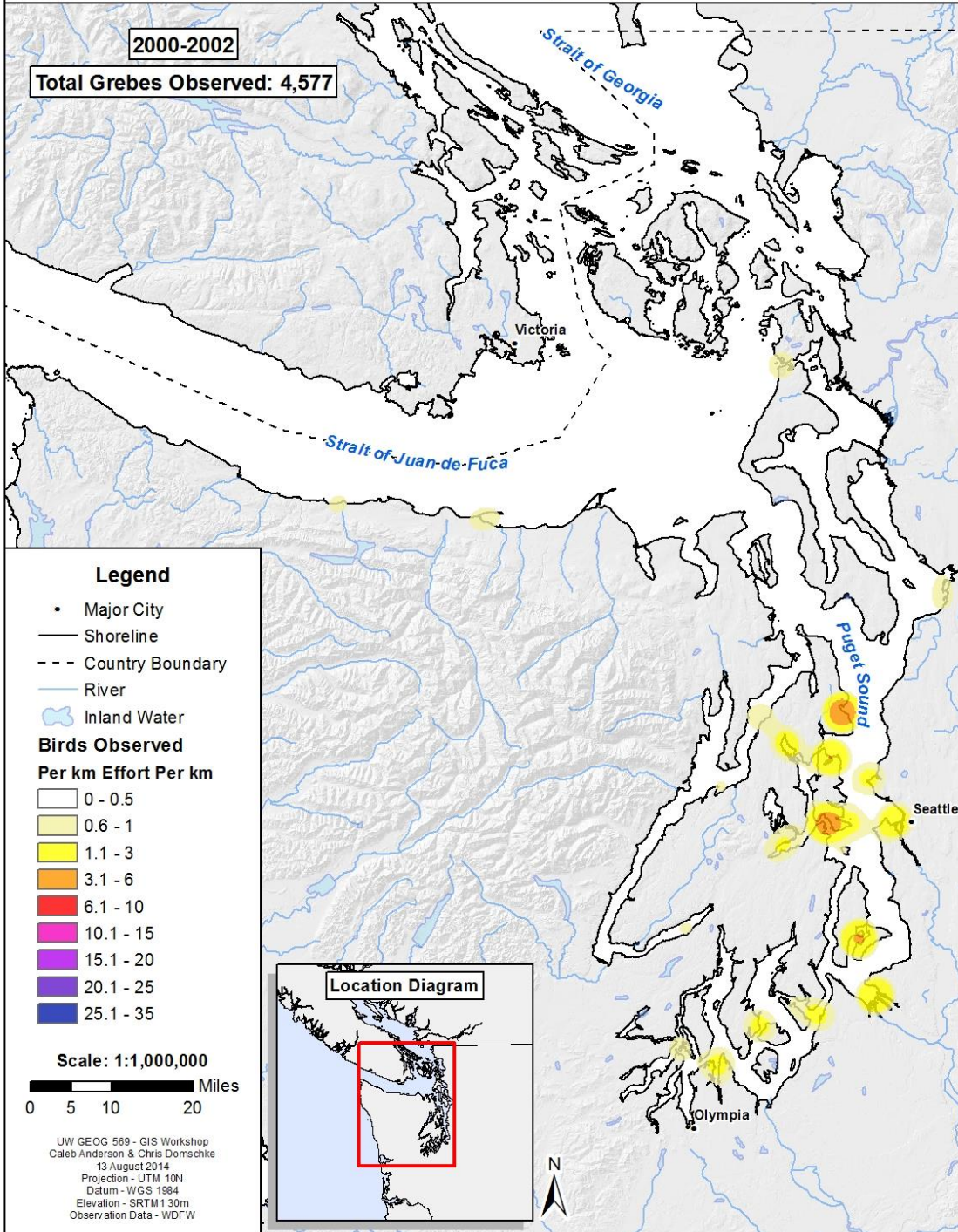
Figure 7.4 3 yr Group Tool, Model Builder model view. This figure shows the Model Builder diagram for the tool 3 yr group which is used in analyzing multiple year groupings for counts by species or by guild. The output is a polygon feature class of counts per unit effort for the species across the 3 year range.



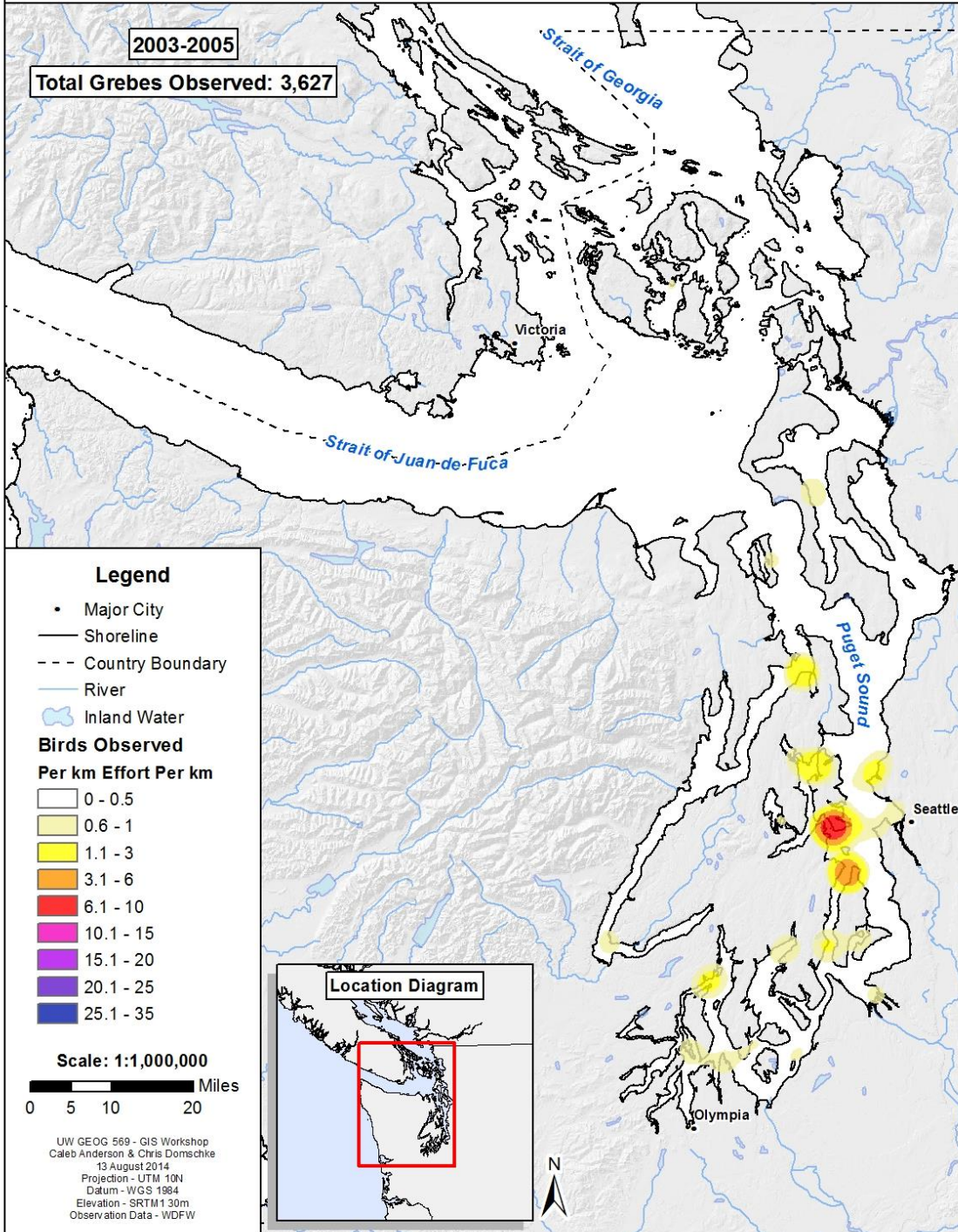
7.2 Appendix B: Density and Hot Spot maps for all Grebe species in 3 year blocks.



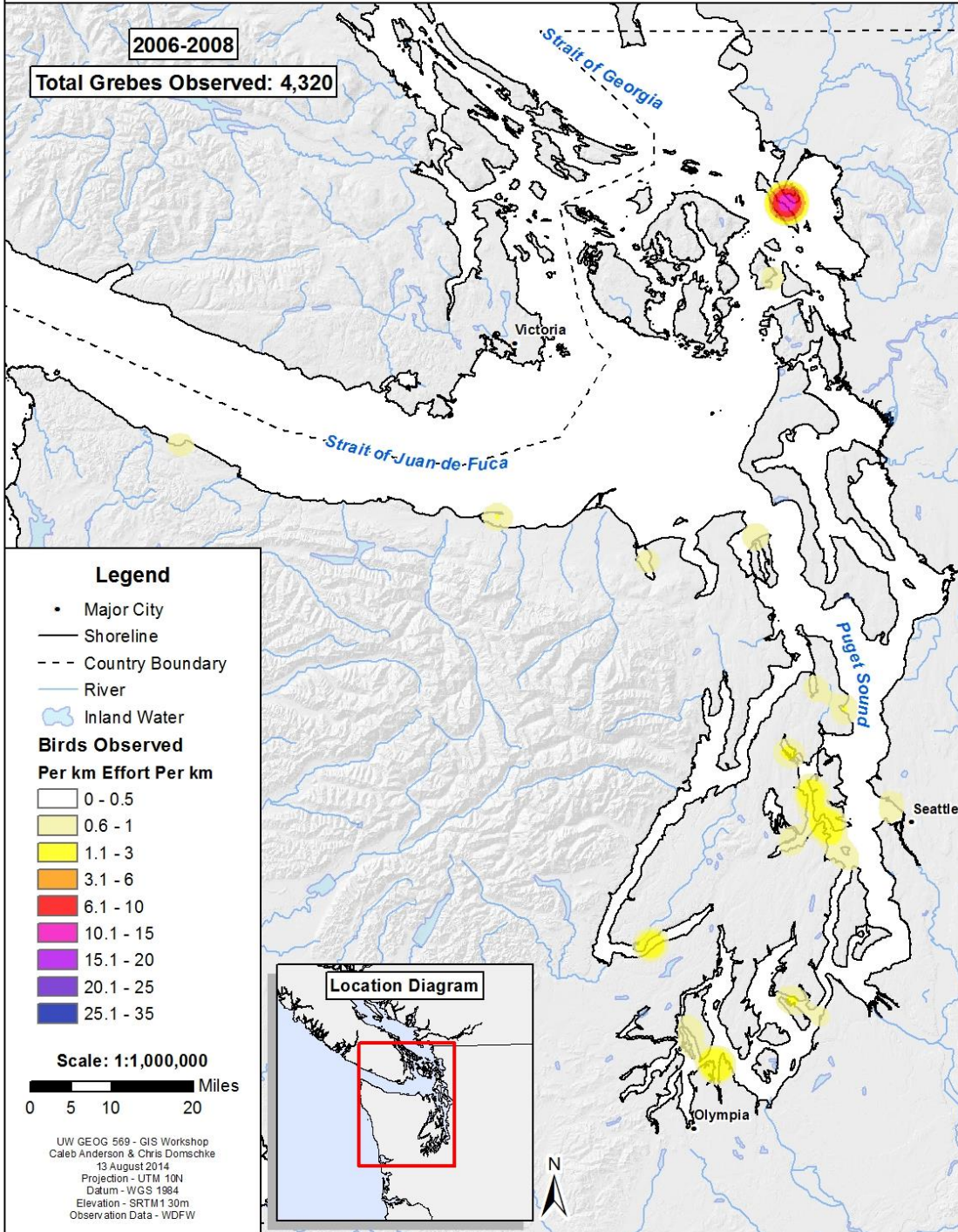
# Wintering Marine Bird Density Analysis - Grebes



# Wintering Marine Bird Density Analysis - Grebes

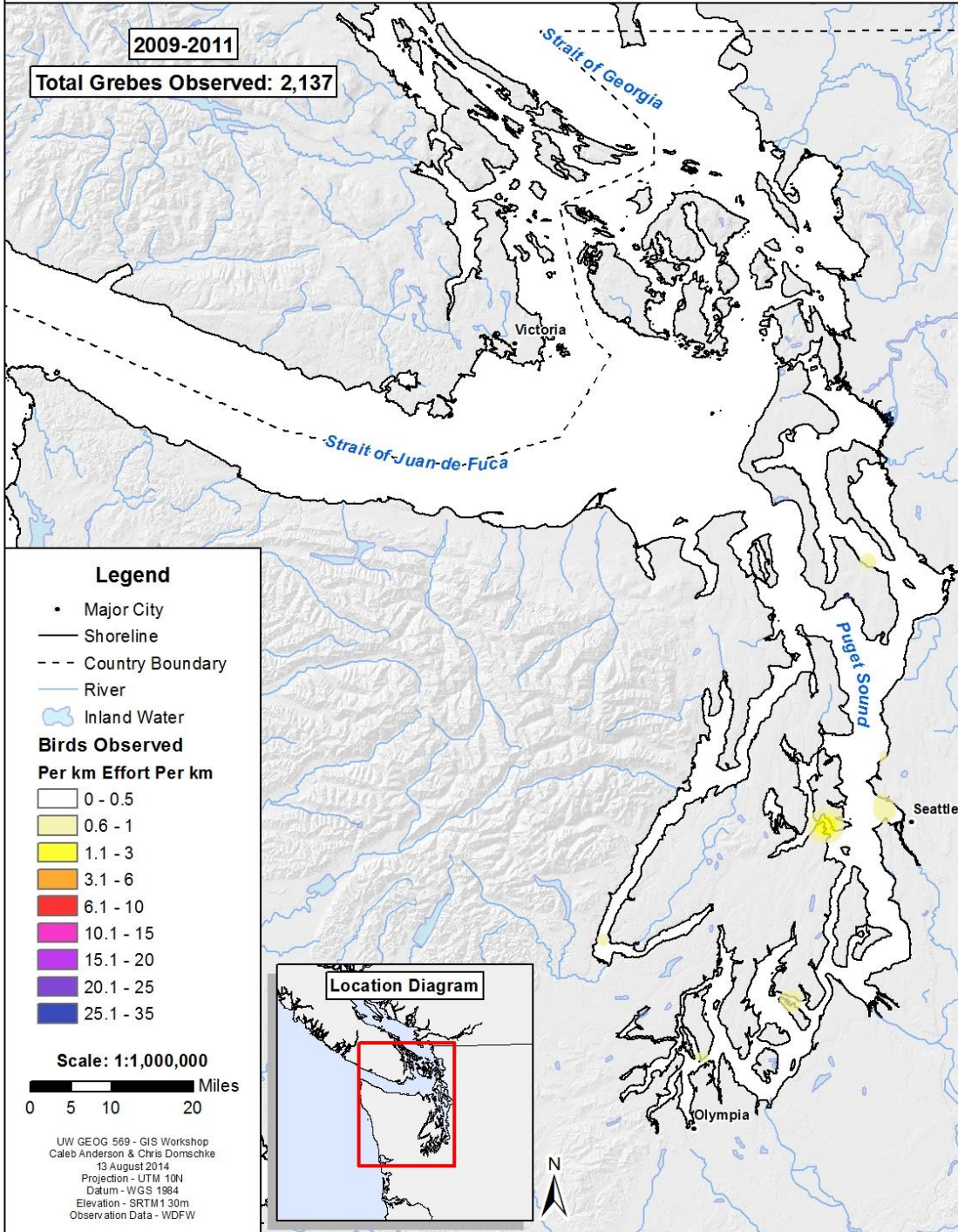


# Wintering Marine Bird Density Analysis - Grebes

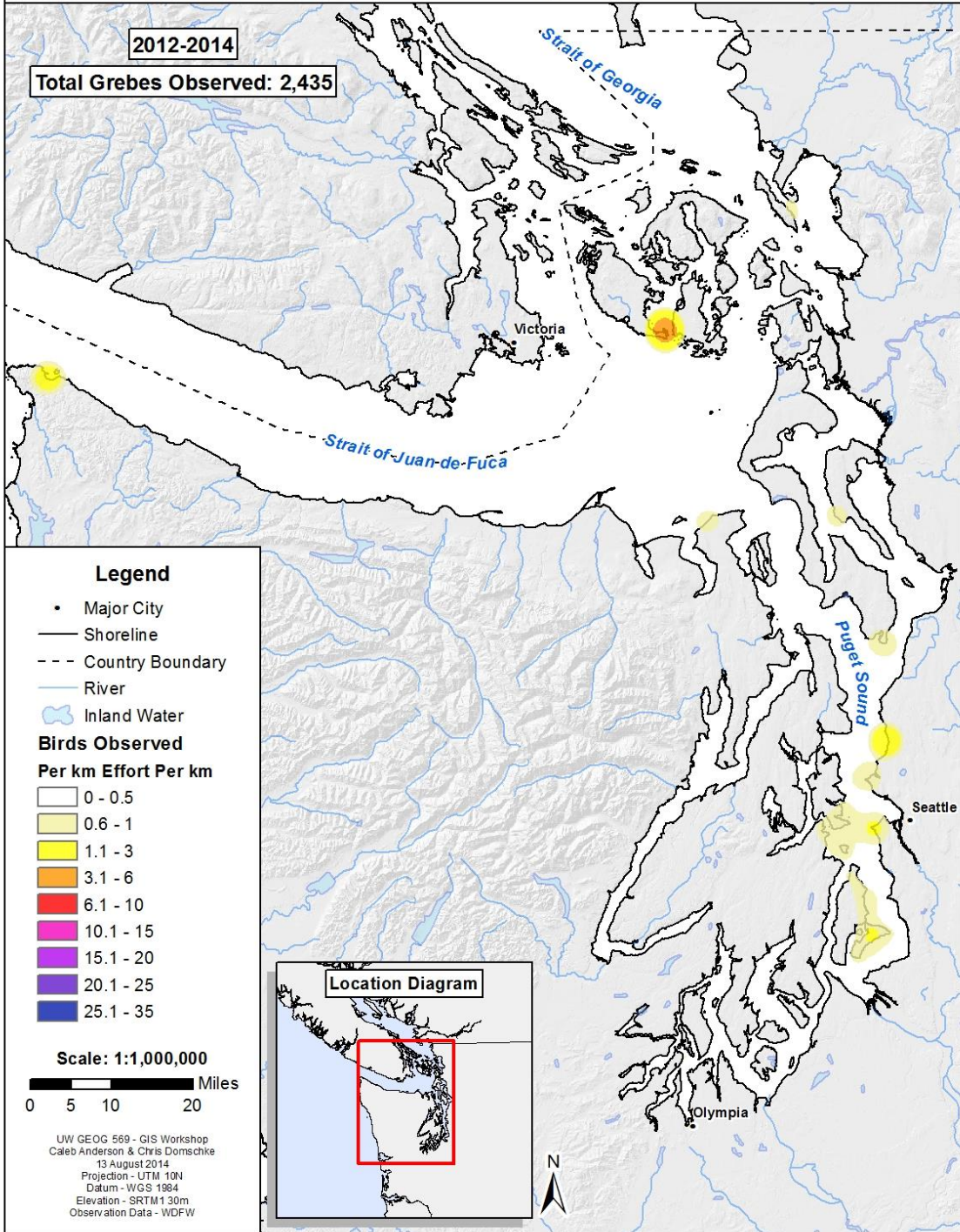




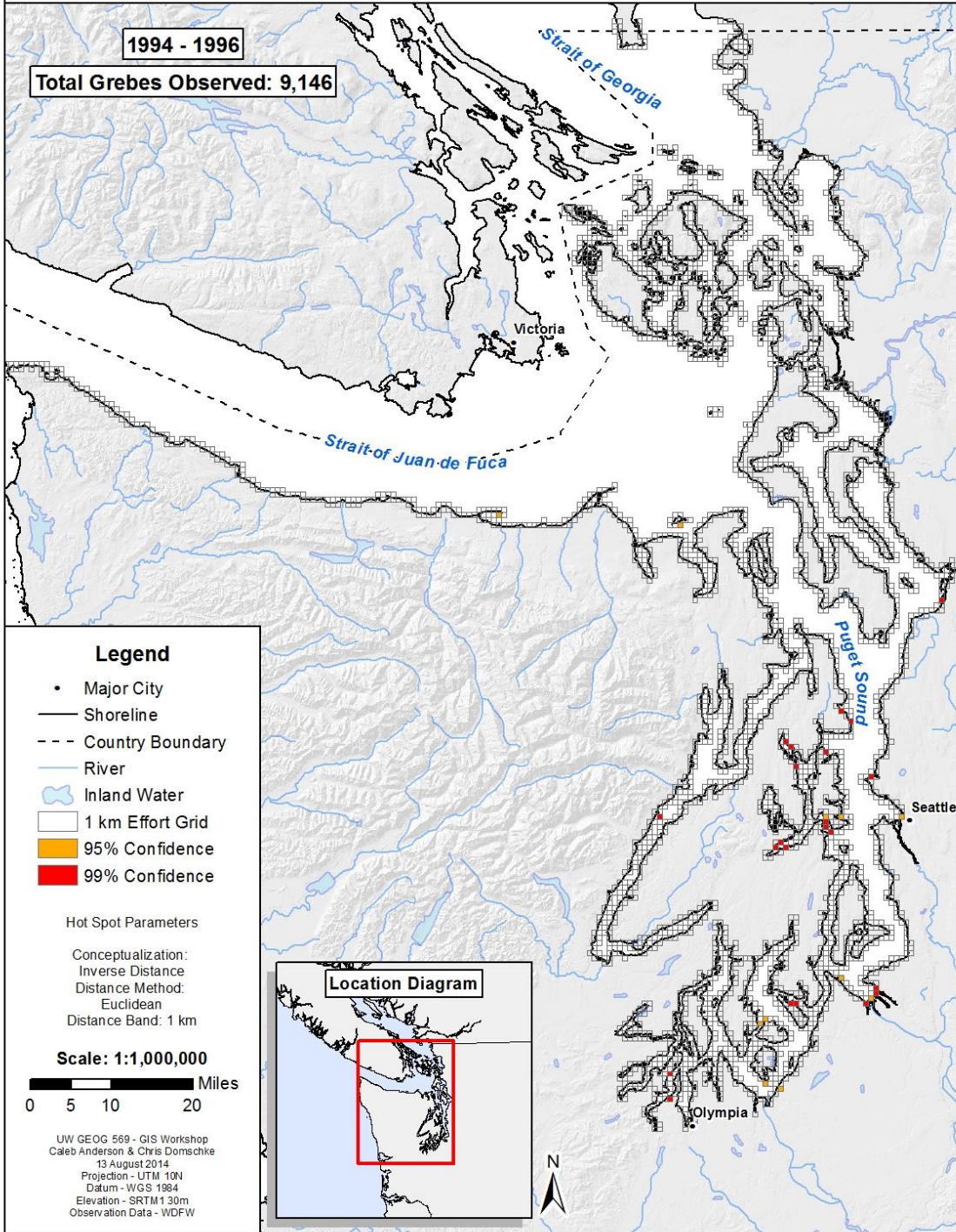
# Wintering Marine Bird Density Analysis - Grebes



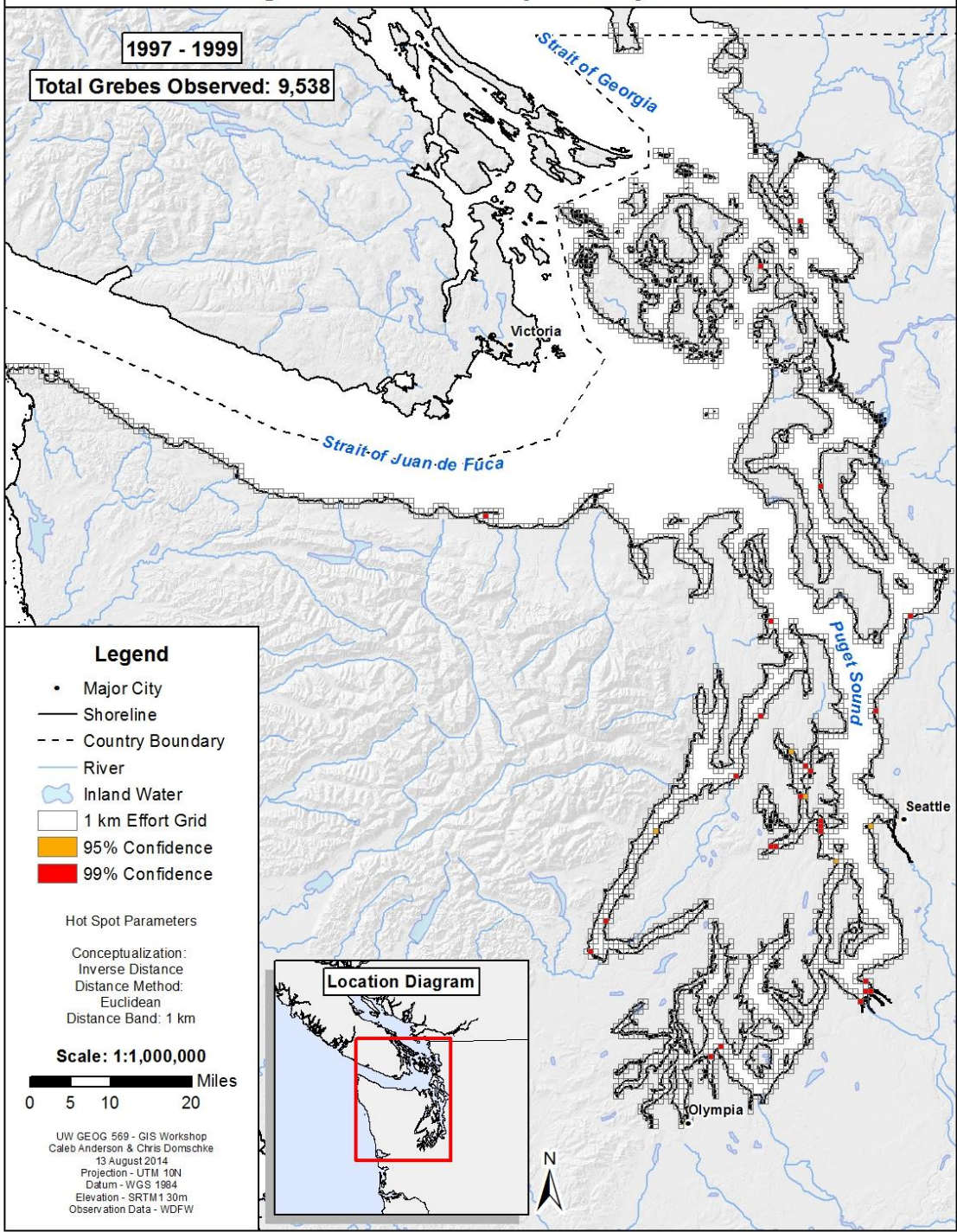
# Wintering Marine Bird Density Analysis - Grebes



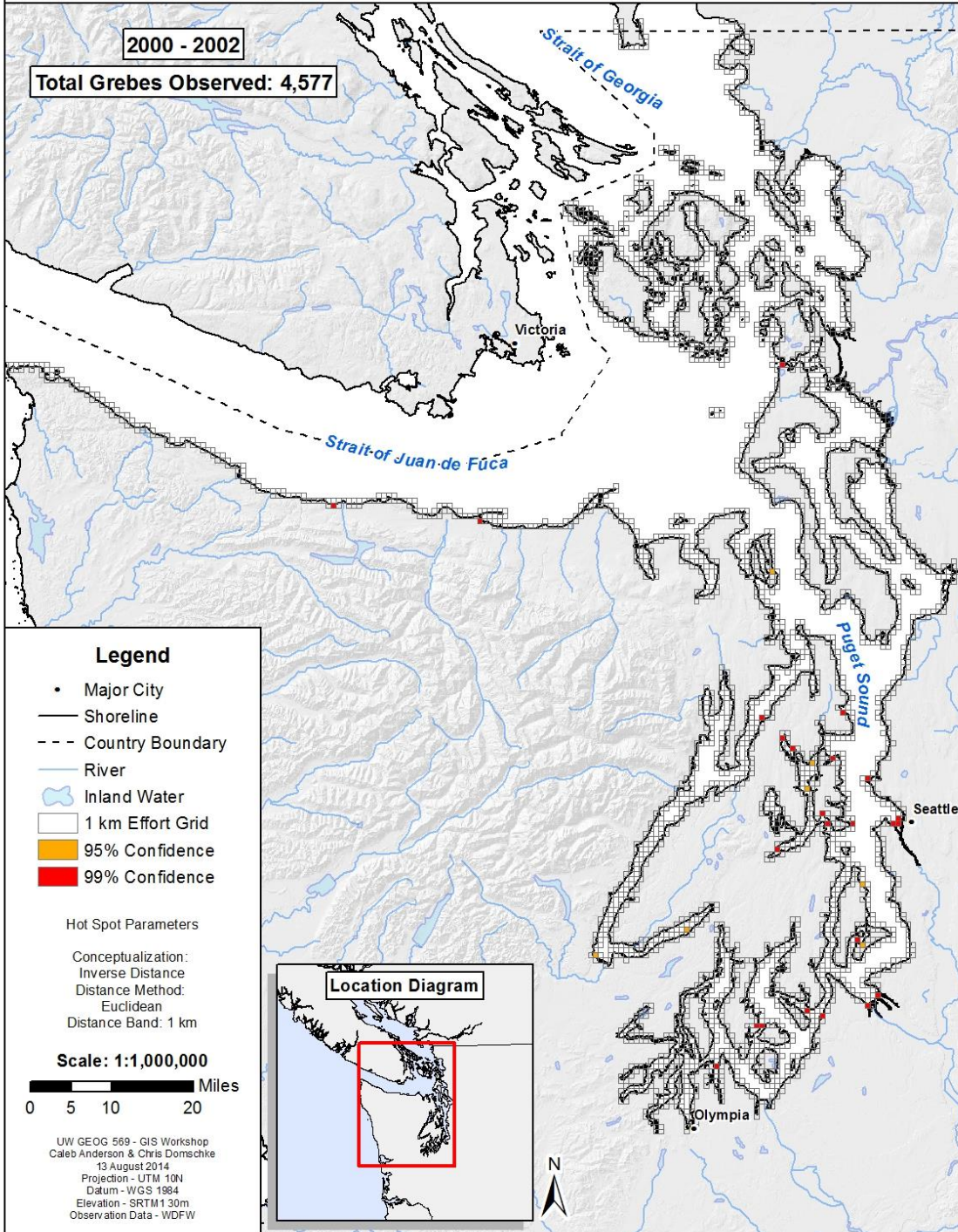
# Wintering Marine Bird Hot Spot Analysis - Grebes



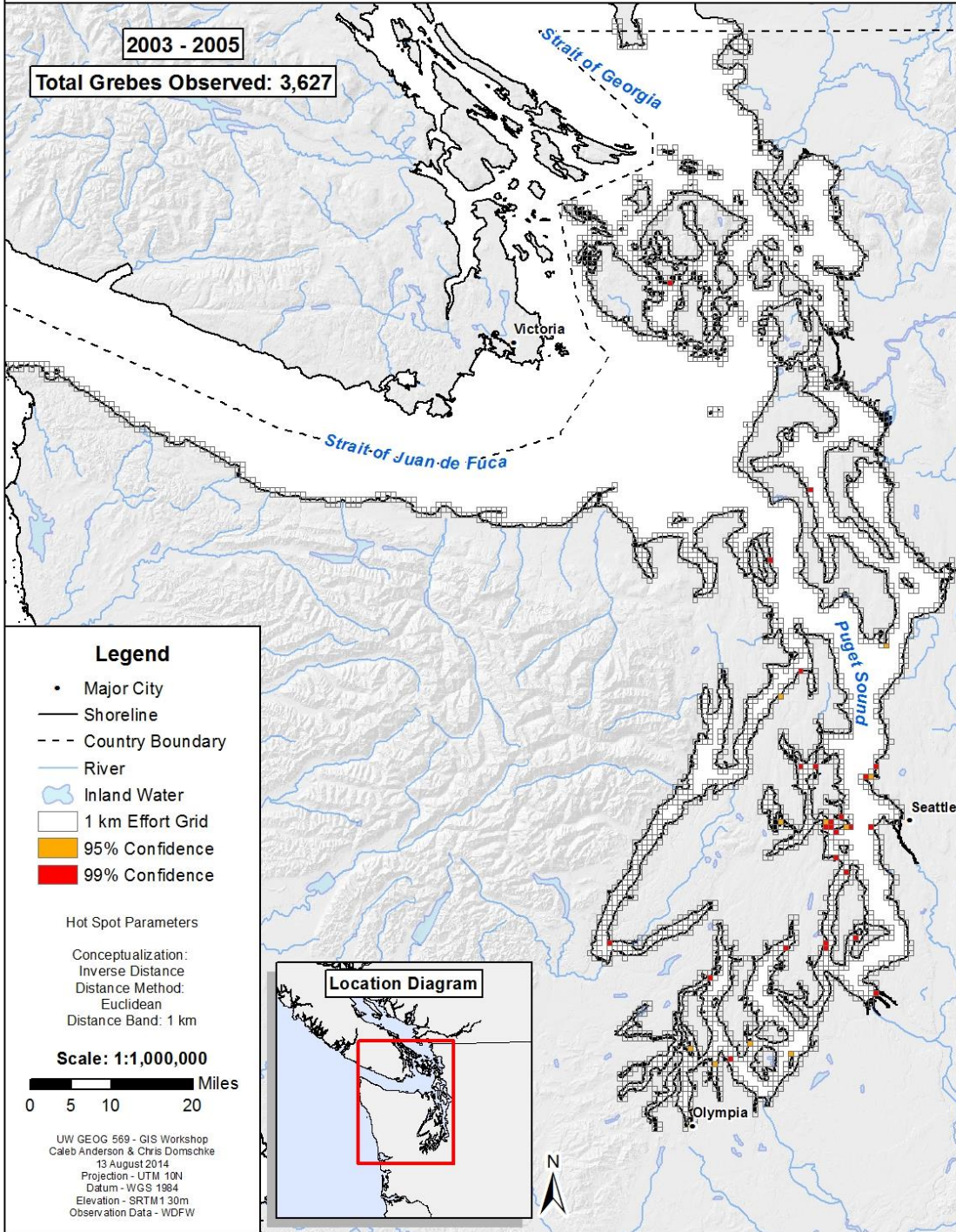
# Wintering Marine Bird Hot Spot Analysis - Grebes



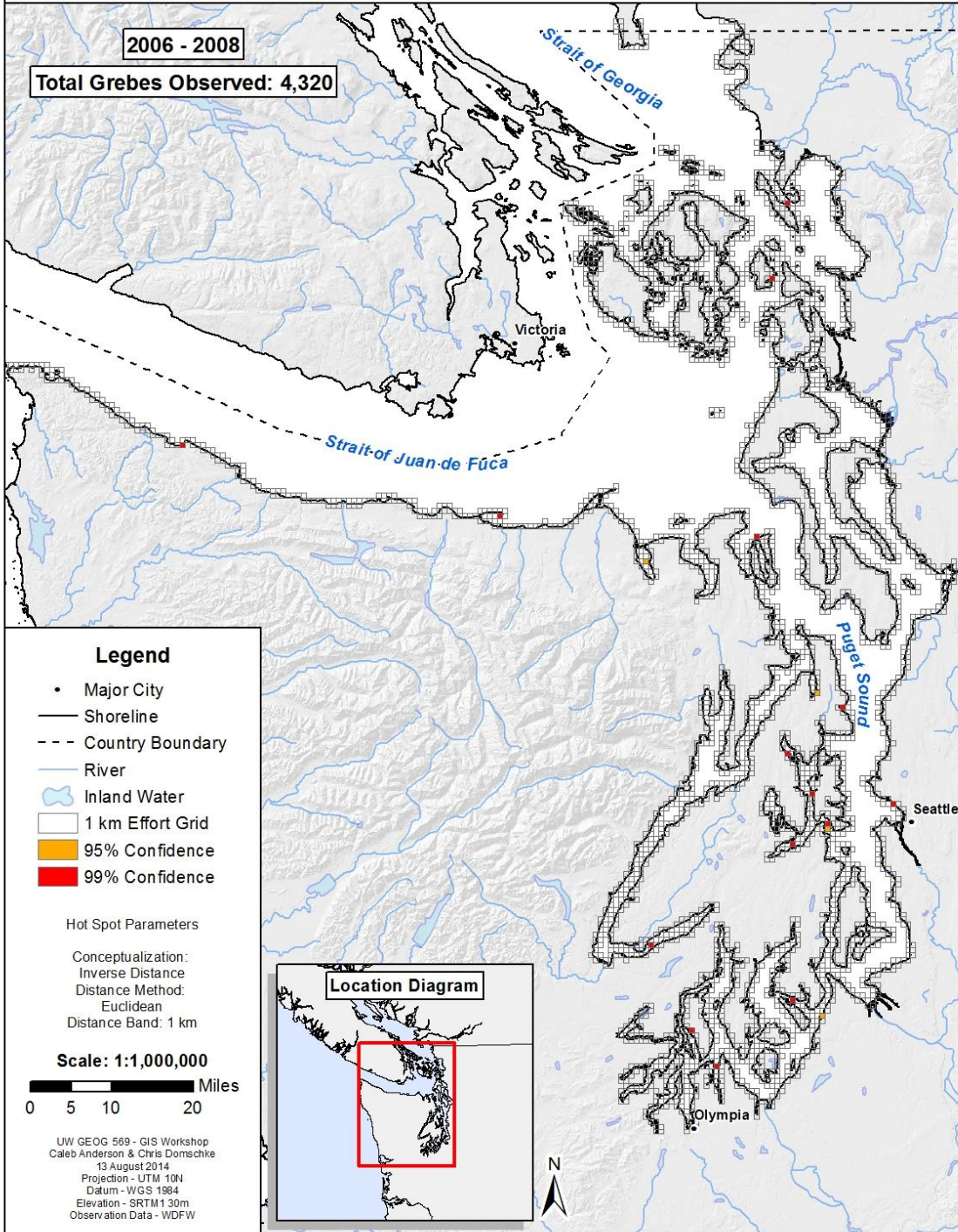
# Wintering Marine Bird Hot Spot Analysis - Grebes



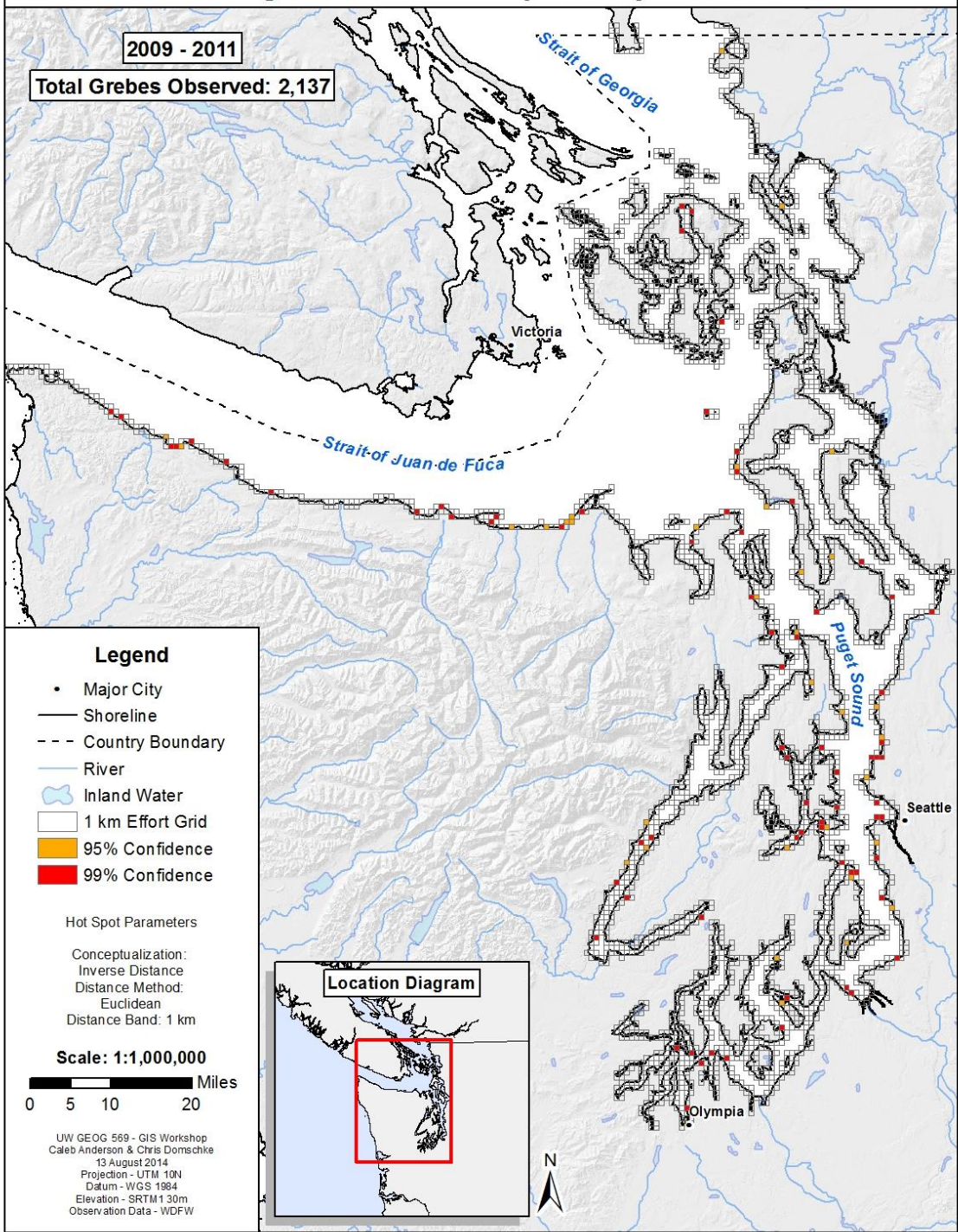
# Wintering Marine Bird Hot Spot Analysis - Grebes



# Wintering Marine Bird Hot Spot Analysis - Grebes

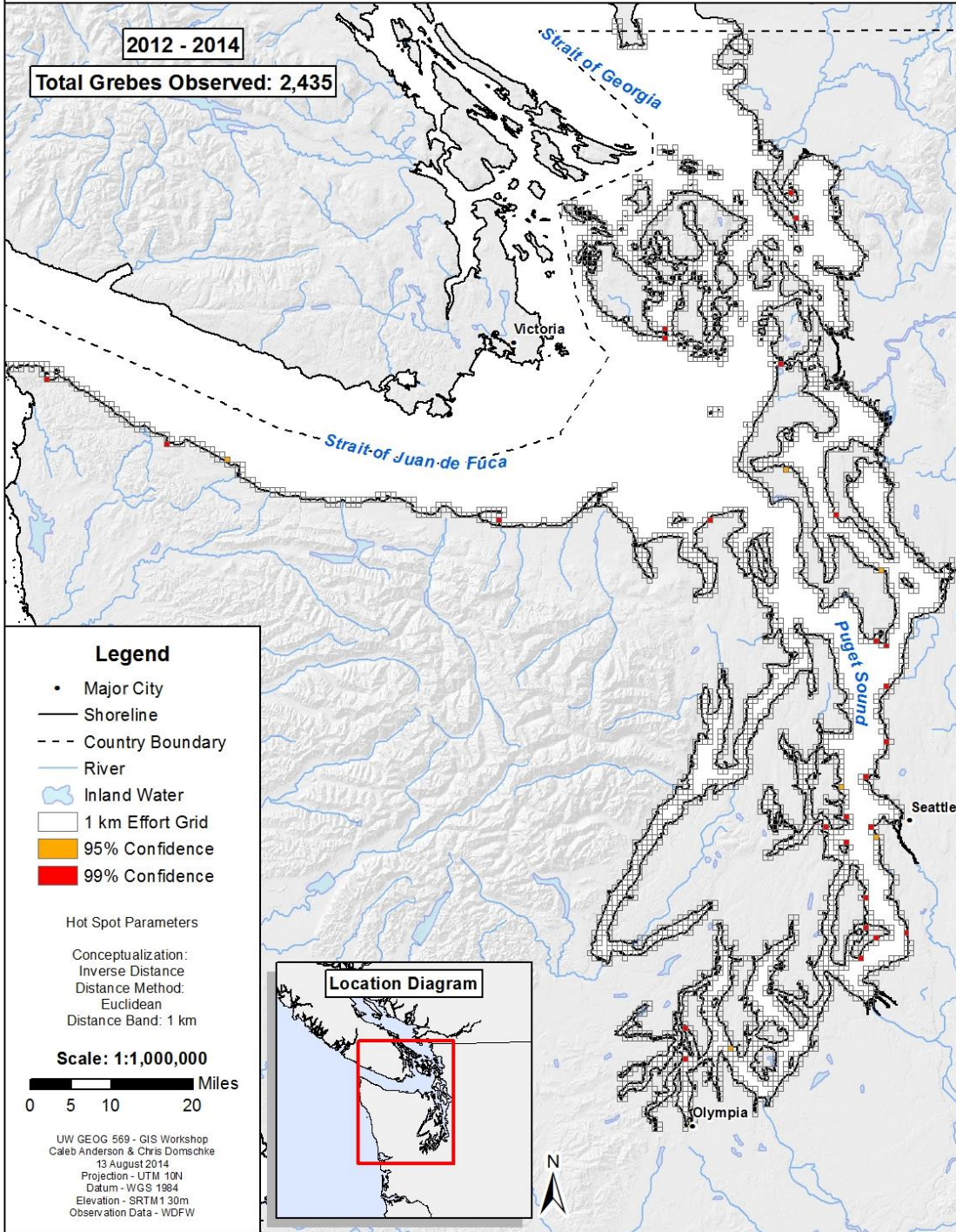


# Wintering Marine Bird Hot Spot Analysis - Grebes





# Wintering Marine Bird Hot Spot Analysis - Grebes



**7.3 Appendix C: Thresholds Matrix showing some of the interconnected processes and states that can lead to or detract from a stable population of marine birds in the Salish Sea.**

