A Spatial Analysis of Waterbird Distribution

DATA 589

April 28, 2024

Hexuan Zhang, Xunyan Zhao, Shengyao Li

Introduction:

The data underpinning this study comprise a comprehensive dataset named "BC Coastal Waterbird Survey," which encapsulates the geographical distribution of waterbird observations across the province of British Columbia (BC). This region, known for its vast and rugged coastline, presents a diverse array of marine habitats conducive to waterbirds, ranging from sheltered inlets to expansive offshore environments. The data are a collection of observational records that provide insights into waterbird population dynamics and spatial distribution patterns within BC.

In avian ecology, spatial distribution is a critical lens through which we can gain insights into the behaviors and habitat preferences of waterbirds. Our study harnesses the "BC Waterbird Discovery Points Data" to analyze waterbird distribution patterns in British Columbia's diverse coastal regions. The data transformation into a point pattern process (ppp) is a key step in visualizing the spread of waterbirds across BC, providing a foundation for examining spatial arrangements and distribution patterns.

A significant focus of this study is to determine whether waterbirds exhibit a uniform distribution across the study area. The implications of these findings are substantial, informing conservation strategies and enhancing our understanding of waterbird ecology. Additionally, the research seeks to identify hotspots for waterbird observations within BC, offering practical benefits to ornithologists and birdwatchers. By locating areas with a higher frequency of waterbirds, the research could uncover critical zones for feeding, breeding, or migration.

This study aims to offer a nuanced understanding of waterbird distribution in BC, with potential implications for conservation and natural resource management. Through spatial statistical analyses, we hope to contribute meaningful insights into the conservation and ecological study of waterbirds.

This study investigates spatial autocorrelation to understand how waterbird sightings in British Columbia are spatially related. By assessing whether the presence of waterbirds in one location can predict their presence nearby, we aim to uncover patterns that influence waterbird clustering. A central technique in this research is Kriging, a geostatistical method used to estimate waterbird counts in unsampled areas, thus offering a comprehensive view of their distribution. The predictive surface generated by Kriging will inform future research directions and conservation planning.

Overall, the study aims to enhance our understanding of waterbird ecology through spatial analysis, aiding in the strategic conservation of these vital species within BC's marine habitats.

Method:

Part 1. Visualization of Waterbird Dataset

To visualize the data in a given BC map, our first method utilizes the "BC Waterbird Discovery Points Dataset," which contains observations of waterbirds in the maritime regions of British Columbia (BC), Canada. The dataset includes essential geographic coordinates for each observation—decimalLongitude and decimalLatitude—which are crucial for mapping and spatial analysis to understand the distribution patterns of waterbirds. The data does not contain any personal identifiers or sensitive information, ensuring compliance with ethical standards for wildlife research.

The analysis begins with a check for data integrity to ensure all necessary variables, especially geographic coordinates, are complete and error-free. This step is crucial because missing or incorrect coordinate data could significantly impact the results of the spatial analysis.

Once data integrity was confirmed, the observations were converted into a spatial object using R's sp package, setting the coordinate reference system to WGS84, the standard for GPS technology. This transformation facilitates accurate geospatial analysis and ensures compatibility with other spatial data layers.

After data preparation, the study employed Geographic Information System (GIS) tools for visualizing and analyzing the spatial distribution of waterbird observations. The primary software used for this visualization was ggplot2 in conjunction with the sf package, which are both well-documented and widely used in the statistical and ecological research communities for their robust capabilities in handling and visualizing spatial data (Pebesma, 2018; Wickham, 2016).

The key steps in the visualization process included plotting the BC provincial boundary as a spatial framework using data from the "BC_Parks.Rda" data file.

Part 2. Uniformity

Defining the Observation Window:

Using the command window_data <- as.owin(DATA$Window), the geographical boundaries for the waterbird observations are established by converting the spatial boundary data into an object recognized by spatstat (owin). This window defines the area within which the waterbird observations are analyzed, ensuring that all spatial analyses remain constrained within the actual observation region.

Creating the Point Process Model:

Coordinates are extracted from the observations_sf object, a simple feature collection, with the st_coordinates function. These coordinates are then used to construct a point process model (observations_ppp) using ppp(x = coords[,1], y = coords[,2], window = window_data). This transformation allows for subsequent spatial statistical analyses, leveraging the robust features of the spatstat package to handle and analyze point data accurately.

Performing Quadrat Analysis:

The quadratcount function divides the study area into a 10x10 grid and counts the number of observations within each grid cell: Qua <- quadratcount(observations_ppp, nx = 10, ny = 10). This method quantifies the spatial distribution of observations and is particularly useful for visualizing and initially assessing spatial uniformity.

The uniformity of the distribution is formally tested using quadrat.test(Qua), which applies a statistical test to evaluate whether observations are equally distributed across the grid cells. The results from this test (quad_test) provide statistical evidence regarding the spatial uniformity of waterbird observations.

Visualization:

Visualization techniques are applied to effectively communicate the results of the quadrat analysis and the spatial distribution of observations. The command plot(Qua, main="Quadrat count of bird observations") generates a visual representation of the grid counts, highlighting areas with varying densities of observations. Adding points(observations_ppp) to the plot overlays the actual points on the grid, offering a detailed view of where observations are concentrated.

Part 3. Density Estimation

The utilization of kernel density estimation is directly aimed at addressing the research question: "In which areas are waterbirds most likely to be found in British Columbia?" By analyzing the density of observations, this method provides significant insights into potential hotspots where waterbird presence is highest, suggesting regions that may be critical for waterbird feeding, breeding, or congregating. This analytical approach not only helps in wildlife management and conservation efforts but also guides researchers and wildlife enthusiasts in targeting their observation efforts more effectively.

This methodological approach, leveraging both the spatstat package and GIS visualization tools, ensures that the workflow is comprehensive and capable of providing detailed insights into the spatial distribution of waterbirds, facilitating the identification of key areas for further study and conservation.

Following the initial setup and data preparation, the analysis progresses to a more detailed examination of waterbird distribution using kernel density estimation, a method well-suited for identifying density patterns of point data across a geographic landscape. The primary aim here is to utilize the spatial statistical capabilities of the spatstat package in R to determine where waterbirds are most densely clustered within British Columbia.

Density Calculation:

The function density(observations_ppp) calculates the kernel density estimation for the point pattern object observations_ppp, which represents waterbird observations. This function applies a smoothing kernel to each point, spreading its influence over a region defined by a bandwidth parameter, thus estimating the density of points (waterbird observations) throughout the study area. This density is stored in lambda_u_hat_value, which quantifies the likelihood of finding waterbirds at any given location within the geographic window.

Visualization of Density:

The kernel density estimation is visualized using plot(lambda_u_hat_value, main = "Density Estimation via Kernel Method"). This plot provides a visual representation of waterbird density across BC, highlighting areas of higher and lower waterbird concentrations. The main goal of this visualization is to identify areas where waterbirds are most likely to be observed, aiding in ecological understanding and conservation planning.

Overlay Observational Points:

To enhance the visualization and provide context to the density estimates, observational points are overlaid on the density plot:

plot(observations_ppp, pch = 16, cex = 0.6, col = "yellow", add = TRUE) initially places yellow points representing the actual locations of waterbird observations, giving a direct visual indication of where observations were recorded.

plot(observations_ppp, pch = 16, cex = 0.5, col = "blue", add = TRUE) adds another layer of points in blue, potentially to differentiate between different types of observations or to highlight specific features of the data.

Part 4. Variogram (Spatial Autocorrelation)

We used a variogram to determine the Spatial Autocorrelation of the birds in BC.

Data Preparation:

The initial step involves loading the necessary R packages, including gstat and sp, for spatial analysis. The dataset is loaded into R and stored in a data frame named count_by_coordinates. Missing values in the dataset are handled using the na.omit function to ensure data integrity.

Spatial Data Handling:

Spatial coordinates (decimalLatitude and decimalLongitude) and bird counts are extracted from the dataset.A spatial points data frame (spdf) is created to combine the spatial coordinates and bird count data.The SpatialPoints and SpatialPointsDataFrame functions from the sp package are utilized to create spatial objects.

Variogram Analysis:

The empirical variogram, which describes the spatial autocorrelation of the data at various distances, is computed using the variogram function from the gstat package. The variogram is visualized to understand the spatial dependence structure of the bird count data.

Variogram Modeling:

A theoretical variogram model is fitted to the empirical variogram using the fit.variogram function. In this workflow, a spherical model is employed as an example. The model parameters, such as sill, range, and nugget, are specified in the variogram model.

Visualization:

The fitted variogram model is plotted along with the empirical variogram to assess the goodness of fit and visualize the spatial dependence of the bird count data.

Part 5. Spatial Interpolation by Kriging

Data Preprocessing:

The workflow begins with preprocessing the data to ensure the quality and accuracy of subsequent analyses. This step includes removing any observations with missing values using the na.omit function in R, which is crucial to avoid computational errors in spatial statistics.

Defining the Interpolation Grid:

An interpolation grid is defined to cover the spatial extent of the study area, with a grid size of 0.01 degrees. This grid size may need to be adjusted based on the dataset's spatial distribution to balance detail and computational efficiency.

Spatial Data Framing:

The SpatialPointsDataFrame from the sp package is utilized to integrate the geographic coordinates with bird counts, transforming the data into a format suitable for spatial analyses.

Kriging Interpolation:

Kriging interpolation is performed using the krige function from the gstat package. The model is based on a spherical variogram model fitted to the empirical variogram of the bird counts. This step involves predicting bird counts at unsampled locations to create a continuous surface of bird distributions.

Results:

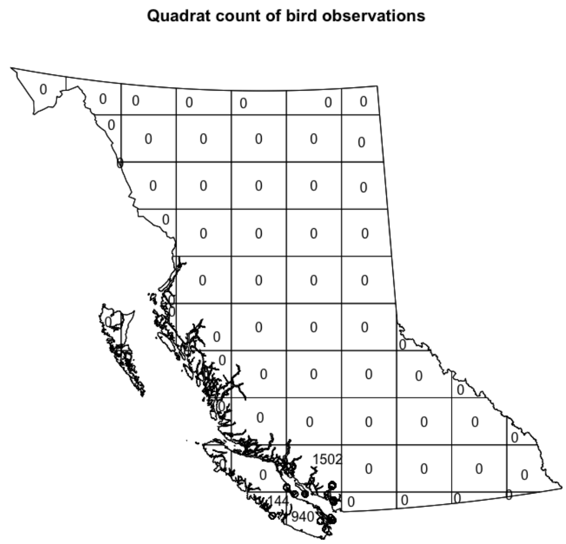
Part 1. Visualization of Waterbird Dataset

The coordinates are marked along the axes, with longitude (X-axis) ranging from 135°W to 115°W and latitude (Y-axis) extending from 48°N to 60°N.

The blue points, which represent bird observation locations, are concentrated in the southern part of the province, particularly in coastal areas and islands.

Detailed outlines of coastlines and islands are present, highlighting the geographic features.

Part 2. Uniformity:



Descriptions of the diagram: The map of British Columbia is overlaid with a grid of quadrats, with the counts of bird observations noted in each quadrat.

The numbers within the quadrats seem to vary significantly, with some quadrats having zero counts and others having counts in the hundreds.

The southern coastal regions, especially along the border and where islands are present, have higher counts of observations. This is indicated by quadrats with higher numbers such as "1502" and "940".

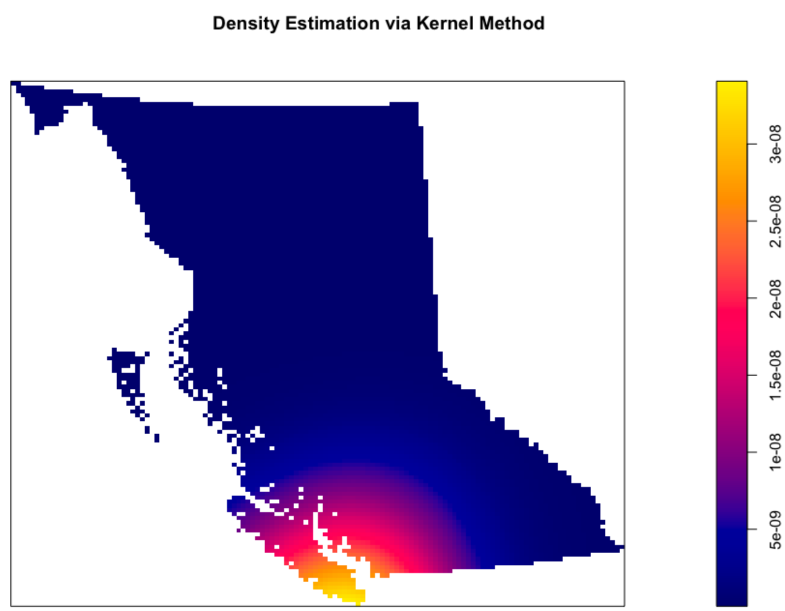
The interior and northern parts of the province have far fewer observations, with many quadrats showing zero counts, which indicates no observations or a very low frequency of bird occurrences in these quadrats.

By this we could say that the birds are not distributed uniformly throughout BC.

```
Warning message:
Some expected counts are small; chi^2 approximation may be inaccurate
> plot(Qua, main="Quadrat count of bird observations")
> points(observations_ppp)
> quad_test

	Chi-squared test of CSR using quadrat counts

data:
X2 = 68990, df = 63, p-value < 2.2e-16
alternative hypothesis: two.sided

Quadrats: 64 tiles (irregular windows)
>
```

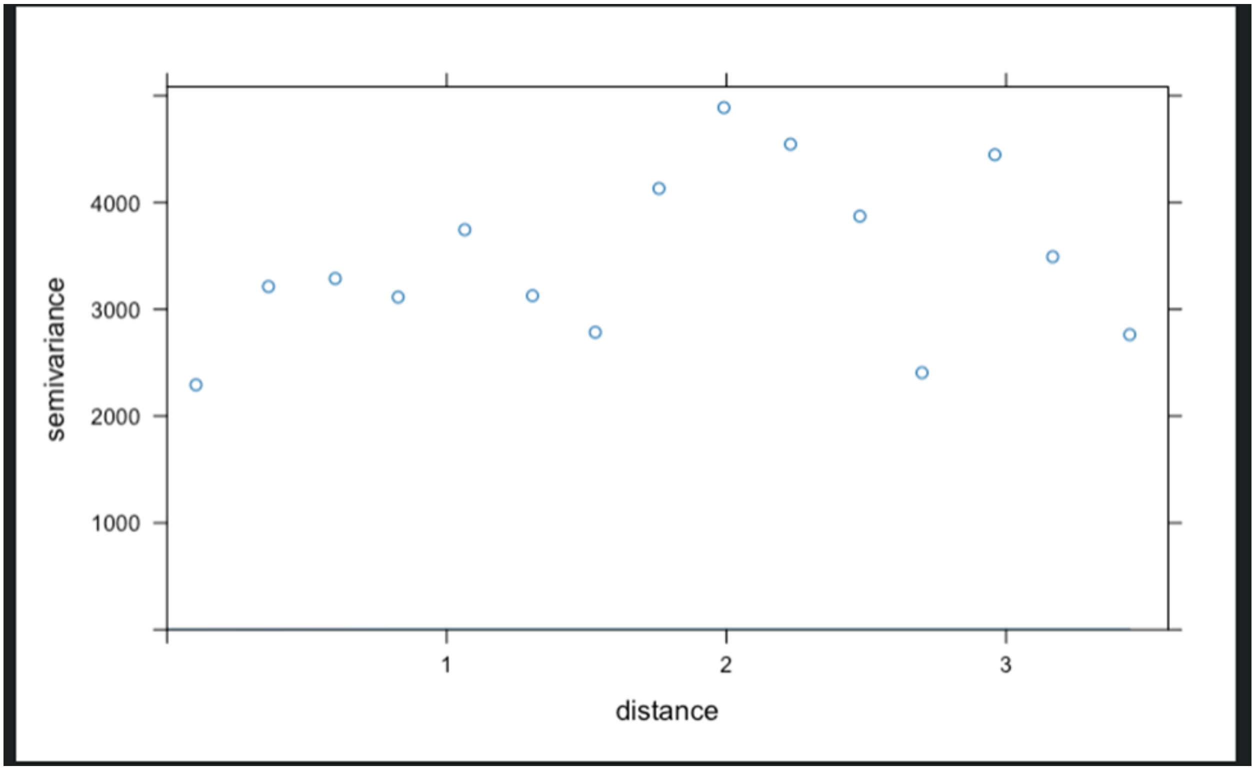
The low p-value suggests a meaningful departure from uniformity.

Part 3. Density estimation



The image presents a kernel density estimation of bird sightings, which is a technique to analyze spatial data. Colors from blue to yellow indicate increasing densities of bird observations, with hotspots evident in certain areas, suggesting regions of high avian activity. This method is valuable in ecological studies for identifying patterns in bird distribution. The yellow to red areas in the figure represent the highest density areas, mainly concentrated in the bottom center area of the figure. This area of high density may indicate a breeding site, a migratory stopover, or a location where food is abundant for a particular bird population.

Part 4. Variogram (Spatial Autocorrelation)




```
print(fit)
```

	model <fctr>	psill <dbl>	range <dbl>
1	Nug	0.5285531	0.0000000
2	Sph	1.0193704	0.9887494

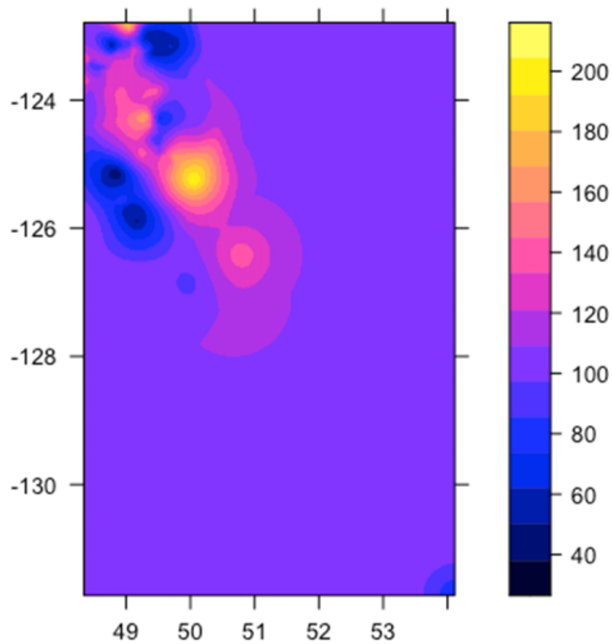
2 rows

The graph shows semi-variation versus distance, an important measure of spatial autocorrelation. The semi-variation value in the plot increases with distance, which indicates that the similarity between sample points decreases with the distance between them. This is a common phenomenon in spatial analysis and usually points to positive spatial autocorrelation of the sample data.

Besides, in my variogram analysis, the "Nugget" model parameter "psill" is approximately 0.5286, which indicates that a substantial portion of the variability cannot be explained by spatial structure and may be due to random noise or measurement error. The "range" parameter is 0, suggesting the presence of a Nugget effect in my data. As for the "Spherical" model, the "psill" is about 1.0194, representing the total variance that can be explained by the spatial structure, and the "range" is approximately 0.9887, indicating that there is significant spatial correlation between observations within this distance.

From these analyses, I can understand that there is a significant spatial correlation in the number of waterbirds within a certain distance (approximately 0.9887 units). Beyond this range, the spatial correlation weakens and is no longer significant. The presence of the Nugget effect suggests that, in addition to the spatial model, other factors such as environmental variables or stochastic elements are influencing the distribution of waterbirds.

Part 5. Spatial Interpolation by Kriging



This image is a waterbird count distribution map created by kriging interpolation from variogram models. Colors indicate predicted waterbird counts: red and yellow for higher counts and blue for lower. It also pinpoints high-density locations with latitude and longitude coordinates. In this image, the darker colors in the central part of the image indicate a higher density in that area.

Discussion

From this information we can draw some conclusions:

The spatial distribution of waterbirds is not random but shows some degree of spatial structure, which may be related to environmental characteristics, habitat conditions or food availability.

Areas of high density may point to waterbird hotspots, which may attract large numbers of waterbirds by providing good habitat conditions or food resources.

Through these models and maps, we can better understand the distribution patterns of waterbirds and provide support for conservation and management decisions, such as establishing protected areas and conducting monitoring projects.

Interpolated maps can help researchers and conservationists identify areas to focus next fieldwork on, especially those where bird populations appear to be high or where changes are significant.

Overall, these analytical tools and results provide a basis for further research that may point to the need for more complex ecological models or conservation measures.

And to improve our project, we could make some Time Series Analysis. In detail, considering the impact of temporal factors on waterbird distribution is crucial. Long-term time series data can help understand population dynamics and the effects of environmental changes.

Reference

Noonan, M. (2024). Spatial autocorrelation. DATA 589: Spatial Statistics. University of British Columbia.

Noonan, M. (2024). Spatial autocorrelation 2: Kriging. DATA 589: Spatial University of British Columbia.

Noonan, M. (2024). Assignment 02 - Exploratory Data Analysis. DATA 589: Spatial University of British Columbia.

Noonan, M. (2024). Assignment 03 - Fitting and Validating Poisson Point Process Models. DATA 589: Spatial University of British Columbia.