

# Coho Salmon Analysis in British Columbia

**Group members: Shiyi Chen, Marina Tian, Hanwen Zhang**

## Introduction

The Coho salmon (*Oncorhynchus kisutch*), an ecologically and economically essential species, exhibits a distinct spatial distribution within the freshwater and marine ecosystems of British Columbia. This region provides a diverse array of habitats that are critical at different life stages of the Coho salmon, from spawning to maturity. Understanding the spatial patterns of Coho salmon occurrences can provide valuable insights into their behavior, habitat preferences, and population dynamics, which are crucial for conservation efforts (Quinn, 2005).

The primary aim of this study is to analyze the spatial distribution of Coho salmon occurrences to detect patterns that may suggest clustering, randomness, or dispersion at various spatial scales. Such spatial analysis is foundational for assessing habitat quality, identifying critical ecological zones, and designing conservation strategies that support the persistence of Coho salmon populations (Sandercock, 1991; Beamish and Mahnken, 2001). Furthermore, it assists in understanding the potential impacts of environmental change on these patterns, which is essential given the challenges of habitat degradation and climate change (Schindler et al., 2008). The analysis leverages the pair correlation function (PCF), a sophisticated tool that offers insight into the spatial structure of point patterns by measuring the variation in point intensity over distance, thus identifying clustering or dispersion tendencies.

By employing spatial statistical methods, including point pattern analysis and spatial distribution modeling, we intend to answer the following questions:

1. What is the spatial distribution of Coho salmon within British Columbia?
2. How do these observed patterns align with known ecological preferences and behaviors of Coho salmon, such as the selection of spawning sites or foraging areas?
3. Can the spatial patterns identified through the pcf inform the management and conservation strategies for Coho salmon in British Columbia, considering the species' ecological significance and conservation status?

By integrating these approaches, this study aims to contribute valuable knowledge to the conservation biology of Coho salmon and provide actionable insights for their sustainable management.

## Methods

- Data Description

The data for this study consists of point records of Coho salmon occurrences in British Columbia, sourced from comprehensive field surveys. Each record includes geographic coordinates (decimalLatitude, decimalLongitude), which mark the specific location of Coho salmon observations. Additional attributes in the dataset encompass temporal markers (year, month, day), depth of capture or observation, and relevant event details.

- Analytical Workflow

### 1. Preprocessing and Data Cleaning

The raw data were initially loaded into R. Observations with missing longitude or latitude were removed to ensure the integrity of the spatial analysis. The dataset was filtered to include only observations within the latitudinal and longitudinal bounds of British Columbia and labeled with the state province 'BC' to refine the focus of the study area.

### 2. Spatial Object Creation

Using the ``sf`` package, the cleaned data were transformed into Simple Features (sf). The ``st_as_sf`` function was employed to create the sf object, specifying latitude and longitude as the coordinate columns.

### 3. Spatial Transformation and Point Pattern Creation

Data were projected from geographic to planar coordinates suitable for spatial analysis using ``st_transform``. The ``spatstat`` package was utilized to convert the sf object into a planar point pattern (ppp) object, essential for point pattern analysis.

### 4. Exploratory Spatial Analysis

An exploratory analysis using plots of the point data over the map of British Columbia. The ``ggplot2`` package, was used for creating these initial visualizations.

### 5. Spatial Point Pattern Analysis

- The ppp object was analyzed to calculate the pair correlation function using ``PCF`` from ``spatstat``, which identifies clustering or dispersion at various

distances. Ripley's K-function was computed with `Kest`, applying border correction to mitigate edge effects.

## 6. Simulation Envelopes

Envelopes for the K-function were generated through Monte Carlo simulations (99 simulations were used for increased confidence) using the `envelope` function, which benchmarks the empirical K-function against a CSR pattern.

## 7. Density Estimation

Kernel density estimates were produced using both standard and adaptive methods via the `density` function in `spatstat`, visualizing hotspots of occurrence.

## 8. Quadrat Counts and Morisita's Index

A quadrat count analysis was performed to supplement the continuous density estimates with discrete spatial counts. Morisita's index was calculated for various quadrat sizes to assess spatial dispersion.

## 9. Inhomogeneous Point Pattern Analysis

To account for potential variation in point intensity, inhomogeneous versions of the PCF and K-function were calculated. These functions provided insight into whether observed clustering or dispersion was uniform.

# Results

- Spatial Distribution of Coho Salmon

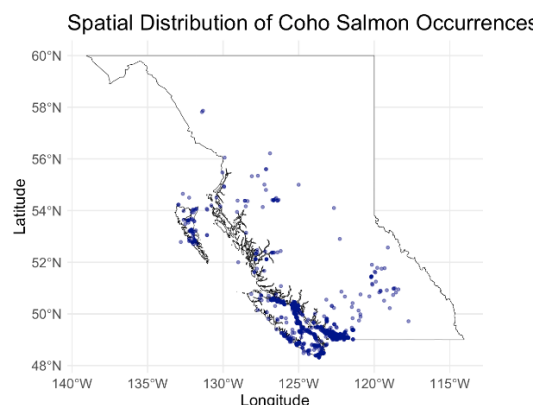


Figure 1 Spatial Distribution of Coho Salmon Occurrences

The spatial point pattern analysis revealed a non-random distribution of Coho salmon in British Columbia. There is a clear concentration of salmon occurrences along the coastline and river systems. This pattern is typical for salmon, as they are anadromous fish that migrate to their natal river systems to spawn. The occurrences are not uniformly spread out but seem to follow certain pathways, which could represent rivers and streams that are suitable habitats for Coho salmon. (Figure 1)

- Pair Correlation Function (PCF)

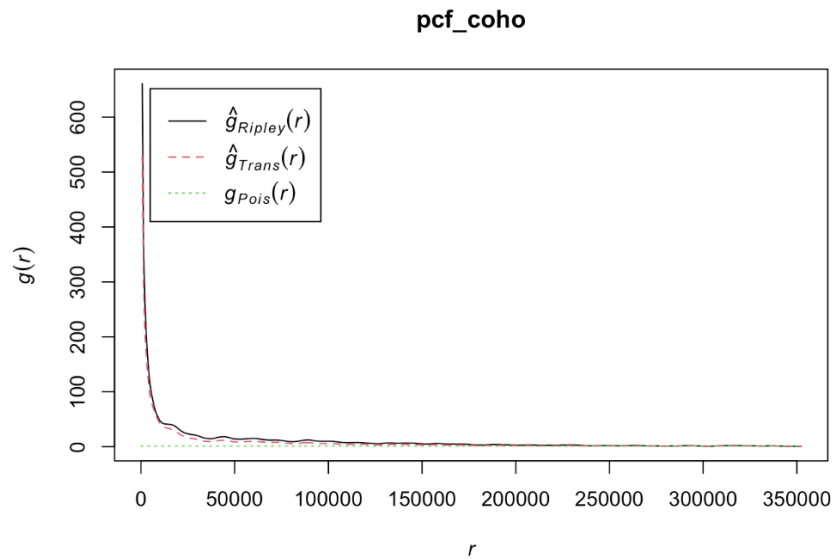


Figure 2

In the plot (figure 2), the pair correlation function drops sharply at very small distances and quickly approaches the line representing complete spatial randomness. This suggests that: 1. At very short distances, there is a strong inhibition effect – points are

less likely to be found very close to each other, which could be due to territorial behavior, resource competition, or physical space limitations. 2. At larger distances (beyond where  $g(r)$  approaches 1), the distribution of points does not significantly deviate from what would be expected by chance (assuming a Poisson process). In other words, there is no strong evidence of clustering or dispersion at these scales.

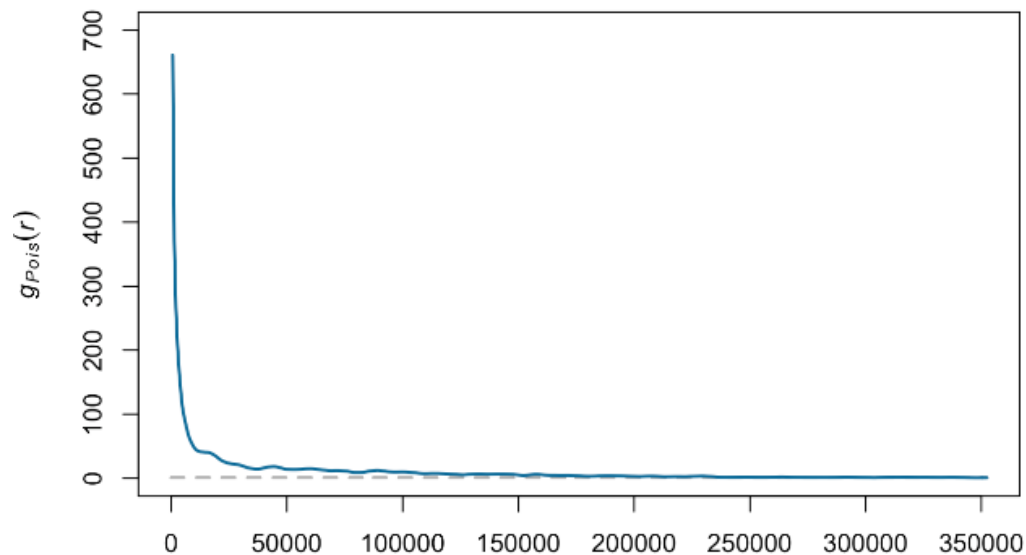


Figure 3

The estimator of the pair correlation function also assumes homogeneity. Here again, we can relax this assumption via the `pcfinhom()` function. (Figure 3)

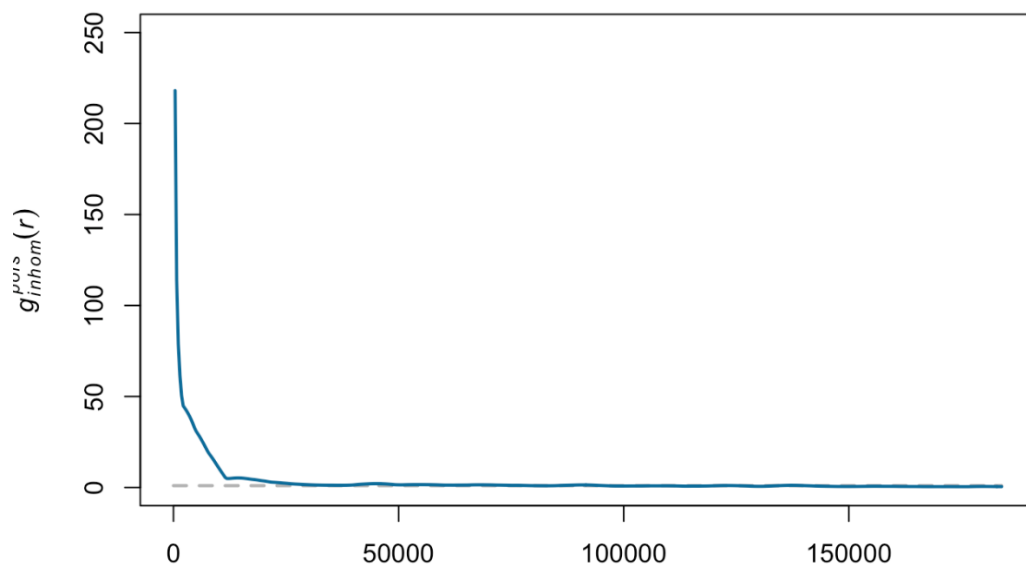


Figure 4

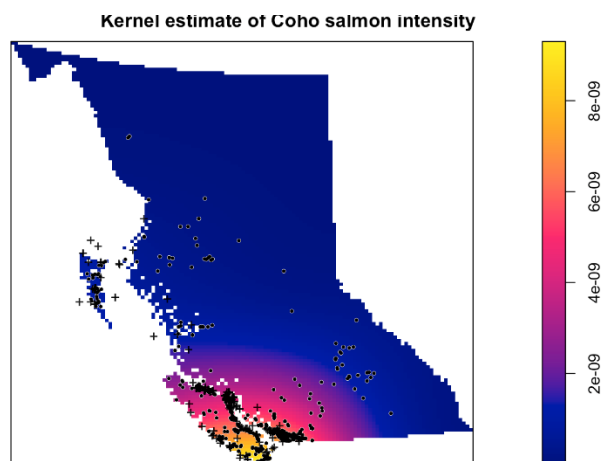
From the plot (Figure 4), Sharp Peak at the Beginning: The plot shows a very high value of  $g_{inhom}(r)$  at very small distances, which rapidly decreases to values below 1. This peak suggests an extremely high level of local clustering at small scales. It could indicate that points (e.g., Coho salmon observations) are much more likely to be found near each other than expected under an inhomogeneous Poisson process at these distances. Below the Line of Randomness: As the value of  $g_{inhom}(r)$  drops below 1, it indicates inhibition or regularity, meaning that points are less likely to be found at these distances compared to a random pattern. In ecological terms, this could suggest some form of territorial behavior, environmental constraints, or other mechanisms that lead to regular spacing among individuals beyond the initial clustering distance. Returning Towards Randomness: The function seems to approach the dashed line (which typically represents  $g(r)=1$ ) as  $r$  increases. If  $g_{inhom}(r)$  stabilizes at 1 for larger distances, it would suggest that the points are randomly distributed at these larger scales when accounting for inhomogeneity.

- Quadrat Counts and Kernel Density Estimation

```
## Chi-squared test of CSR using quadrat counts
##
## data:
## X2 = 15596, df = 66, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
## Quadrats: 67 tiles (irregular windows)
```

*Figure 5*

The quadrat count analysis revealed significant non-randomness in the spatial distribution of Coho salmon. From Figure 5 The chi-squared test of Complete Spatial Randomness (CSR) using quadrat counts yielded a p-value < 2.2e-16, indicating a highly significant deviation from CSR ( $X^2 = 15596$ ,  $df = 66$ ). This suggests that the occurrences of Coho salmon are not uniformly distributed across the study area but are instead clustered in specific regions, a pattern that is supported by the visualization of the data.



*Figure 6*

The kernel density estimation of Coho salmon intensity provided a visualization of occurrence hotspots (Figure 6). The color scale reflects the estimated intensity of Coho salmon occurrences, with cooler colors indicating lower densities and warmer colors indicating higher densities. The highest densities are depicted in yellow, indicating the most likely areas to find Coho salmon, indicating regions of potential ecological importance or preferred habitat. The density is highest at the southern edge of the study area, suggesting this region has more favorable conditions for Coho salmon or has been sampled more intensively.

- Ripley's K-Function

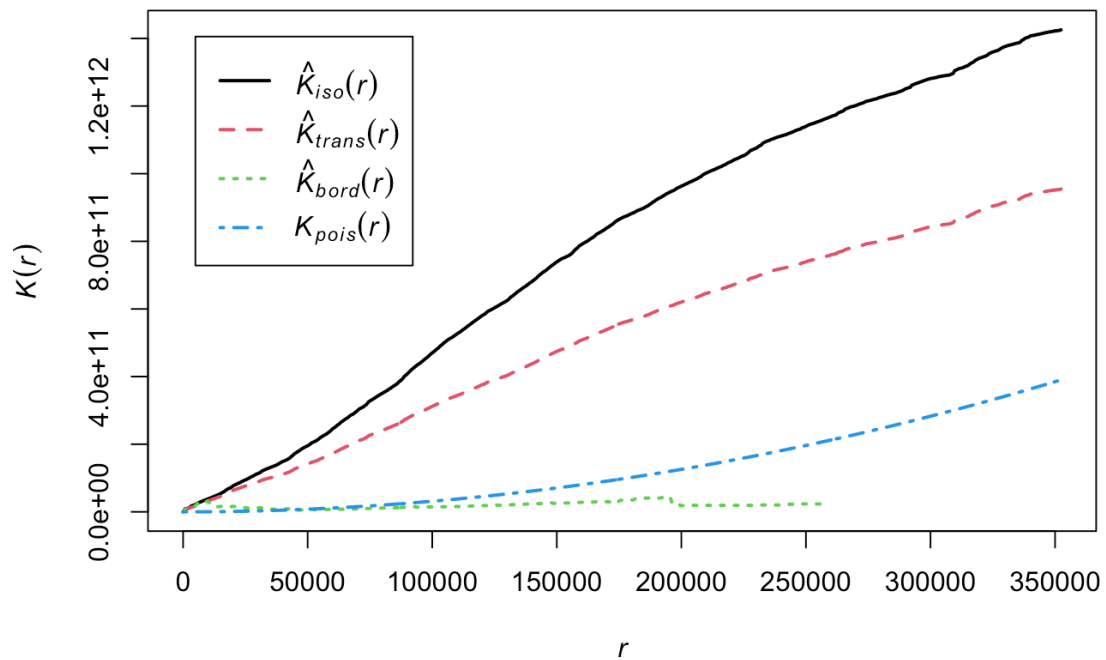


Figure 7

From Figure 7, The solid black line is well above the blue dash-dotted line, especially at smaller values of  $r$ . This suggests that Coho salmon exhibit a much more clustered distribution than would be expected if they were randomly distributed within the study area. This could be due to environmental factors, social behavior, or habitat preferences that cause the salmon to cluster in certain areas. Edge Effects: The difference between the black, red, and green lines indicates the impact of edge corrections on the estimation of spatial clustering. The fact that these lines are distinct but follow a similar trend suggests that edge effects are indeed present in the data but that the general pattern of clustering is robust to these effects. Scale of Clustering: The degree of clustering (indicated by the steepness of the black line) seems to increase with distance at smaller scales, leveling off as the distance increases. This might mean that Coho salmon tend to form groups or clusters at certain scales, perhaps corresponding to suitable habitat features or social behaviors like schooling.

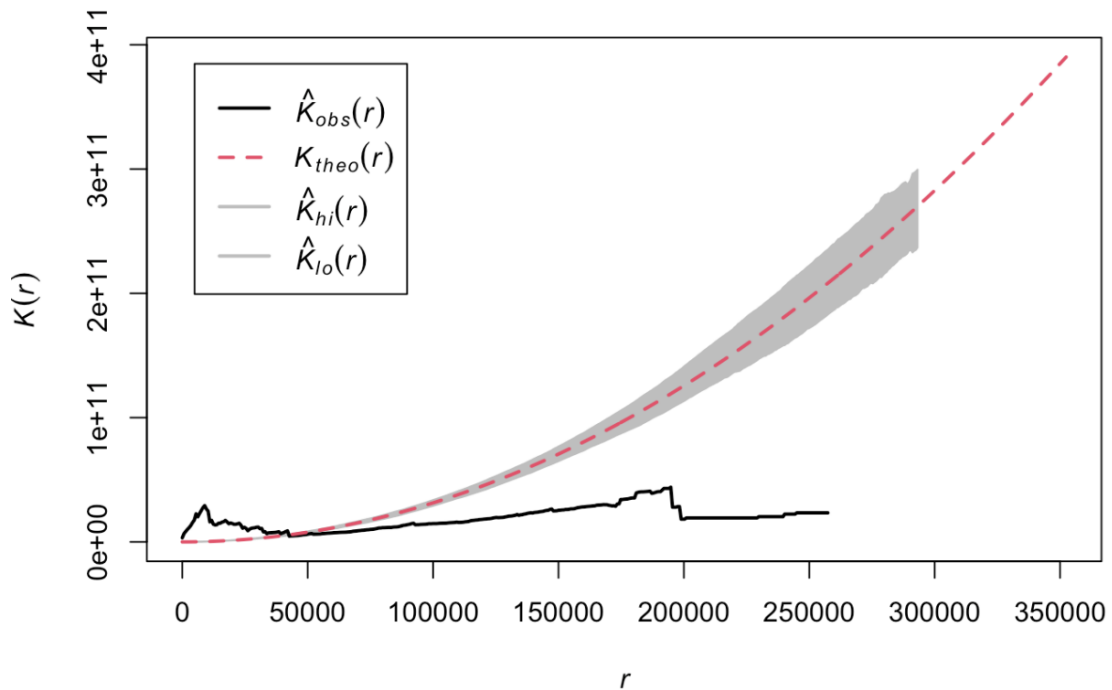
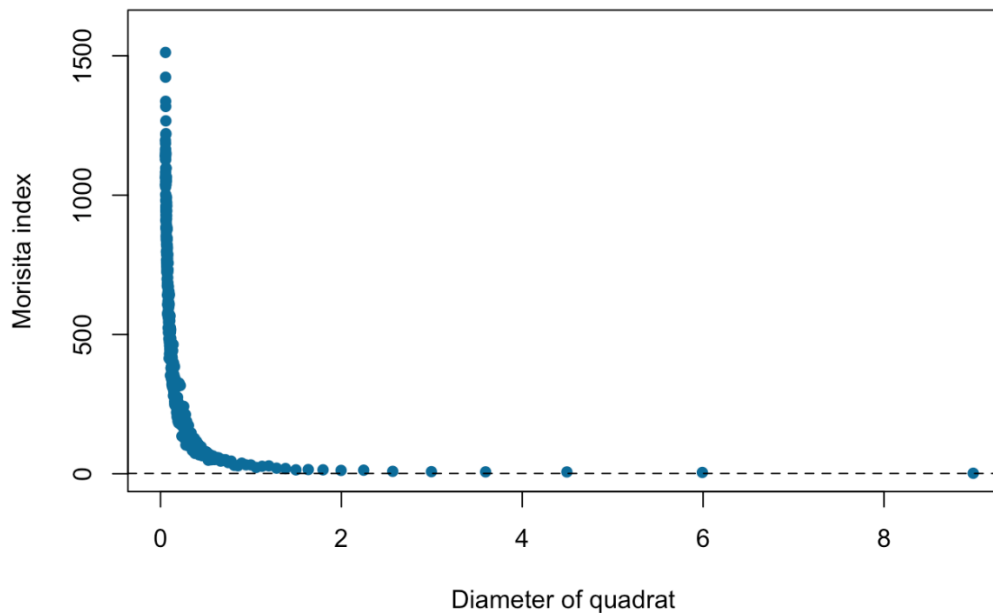


Figure 8

From Figure 8, Initial Values: At the smallest distances, the observed K-function follows closely to the theoretical line, suggesting that within very small distances, the distribution of salmon might not differ significantly from randomness. Deviation from Randomness: As distance increases, the observed K-function diverges from the theoretical line, rising above the upper confidence bound. This indicates the clustering of Coho salmon at those scales, which means salmon are more aggregated than would be expected if they were randomly distributed. Confidence Intervals: The fact that the observed line moves outside the confidence envelope at larger distances reinforces the suggestion of non-random clustering. The broader the gray envelope, the greater the uncertainty about the pattern at that scale, likely due to the variability in bootstrapped simulations. Ecological Insight: For Coho salmon in British Columbia, this could imply that environmental factors or biological behaviors lead to aggregation at certain spatial scales. This could be driven by habitat preferences, spawning behaviors, or other ecological interactions.

- Morisita's Index





*Figure 9*

From Figure 9, Extreme Clumping at Small Scales: The very high Morisita's Index values at small quadrat sizes suggest that Coho salmon are extremely clumped or aggregated at these scales. This could be reflective of spawning behavior, where Coho salmon congregate in specific areas that provide suitable conditions for laying eggs. It might also indicate areas of high habitat quality where salmon prefer to stay, such as regions with ample food supply, appropriate substrate for spawning, or optimal water temperatures.

Steep Decline in Index with Quadrat Size: As the size of the quadrats increases, the index rapidly decreases, indicating that the clumping is not as intense over larger areas. This could mean that while Coho salmon are very likely to be found close together in small, localized areas (e.g., specific parts of a stream or river), they are more evenly distributed when you consider larger stretches of habitat. This pattern is consistent with the behavior of salmon which may have specific, localized spawning sites but are more dispersed across their overall range.

Stabilization of the Index: For larger quadrat sizes, where the Morisita's Index values approach and just exceed 1, the spatial distribution of salmon begins to resemble a random pattern. This transition from clumped to a seemingly random distribution at larger scales may suggest that across the broader regions of British Columbia, Coho salmon do not show strong preferential clustering; instead, their distribution is dictated by the availability of suitable habitats scattered throughout the region.

Biological and Environmental Implications: The results may be significant for conservation and management strategies. The initial high clumping could inform biologists and conservationists about critical areas that need protection, especially during the spawning season. This might include protecting specific streams or river segments, managing human activities such as fishing and development, and monitoring water quality and flow conditions.

Consideration of Scale: The fact that the index stabilizes around 1 at larger quadrat sizes is also informative for resource management. It suggests that at the scale of larger river systems or multiple stream networks, Coho salmon might not exhibit strong patterns of aggregation, perhaps due to the vastness and variability of available habitats.

## Discussion

The spatial analysis of Coho salmon occurrences in British Columbia revealed significant patterns that deviate from a model of complete spatial randomness. The pair correlation function (PCF) indicated strong clustering at short distances, which was further substantiated by the results of the quadrat count analysis, as well as the kernel density estimates. The significantly low p-values from the quadrat test of homogeneity suggest that the distribution of Coho salmon is not uniform but instead is highly aggregated in certain hotspots, which were vividly illustrated in the kernel density maps.

The findings align with ecological understandings (*Fortin 2005*) that Coho salmon exhibit specific habitat preferences, especially during spawning when they seek freshwater streams with suitable conditions for egg deposition. These areas are characterized by specific temperature, flow, and substrate composition, often leading to observed clustering. Such ecological behavior may explain the significant intensities of occurrences in regions identified as hotspots on the density maps.

Moreover, the use of inhomogeneous point pattern analysis methods accommodated varying salmon densities across different habitat conditions within the study area. This approach is particularly relevant given the diverse landscape of British Columbia, which includes a complex mosaic of marine and freshwater ecosystems, each with distinct environmental characteristics that can influence salmon behavior and distribution.

The implications of this study are manifold. From a conservation perspective, identifying regions of high salmon density can inform targeted conservation efforts to protect critical habitats, especially in the face of environmental change and anthropogenic impacts. In the context of fisheries management, understanding spatial distribution patterns is vital for developing sustainable harvesting strategies that ensure the long-term viability of Coho salmon populations.

This study contributes to a growing body of literature that utilizes spatial statistics to elucidate patterns of species occurrences. The methods employed here, particularly the application of simulation envelopes in the analysis of Ripley's K-function, demonstrate a robust approach to assessing the significance of observed spatial clustering. Future studies could expand upon this work by integrating environmental covariates, such as water quality and temperature data, to better understand the drivers behind the spatial patterns observed.

In conclusion, the analysis of Coho salmon occurrences in British Columbia has revealed significant clustering, suggesting that intricate ecological processes are at play. These processes are presumably linked to the species' habitat preferences, particularly for reproduction and survival. Such spatial aggregation has notable repercussions for the management and conservation of Coho salmon, underscoring the necessity for spatially informed strategies. The methodologies utilized in this study, including kernel density estimation and Ripley's K-function analysis, have demonstrated their efficacy in discerning these distribution patterns. Understanding these spatial patterns is pivotal not only for the ecological comprehension of Coho salmon but also for the formulation of conservation actions within British Columbia. Given the impact of environmental changes and human activities, there is an imperative to integrate such spatial analyses into adaptive management practices. Looking forward; future research endeavors need to incorporate environmental covariates to enrich our understanding of the factors driving these observed spatial patterns.

## References

Quinn, T.P. (2005). *The Behavior and Ecology of Pacific Salmon and Trout*. University of Washington Press.

Sandercock, F.K. (1991). *Life history of Coho salmon (Oncorhynchus kisutch)*. In C. Groot & L. Margolis (Eds.), *Pacific Salmon Life Histories* (pp. 395-445). UBC Press.

Schindler, D.E., Hilborn, R., Chasco, B., Boatright, C.P., Quinn, T.P., Rogers, L.A., and Webster, M.S. (2008). *Population diversity and the portfolio effect in an exploited species*. *Nature*, 465, 609-612.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley & Sons.

Fortin, M.-J., and Dale, M.R.T. (2005). *Spatial Analysis: A Guide for Ecologists*. Cambridge University Press.