Introduction: Long-Term Population Dynamics of Kenai River Late-Run Sockeye

Executive Summary: Kenai River Late-Run Sockeye Population Dynamics (1979–2025)

This report provides a formal time-series analysis of the late-run sockeye salmon (*Oncorhynchus nerka*) in the Kenai River using a comprehensive 46-year dataset. By employing both **Ordinary Least Squares (OLS) regression** and the non-parametric **Mann-Kendall Trend Test**, the study identifies significant long-term shifts in population biomass, migration timing, and run distribution.

Key Findings

- **Sustained Population Growth:** The analysis confirms with high statistical certainty ($p < 0.0001$) that the late-run escapement is in a state of growth. The **Sen's Slope** estimate indicates an average annual increase of approximately **26,263 fish per year**.

- **Significant Phenological Delay:** The peak of the sockeye run (P50) is arriving nearly **8 days later** than it did in the late 1970s. This represents a shift of approximately **1.73 days per decade**, a trend that is highly statistically significant ($p = 0.00072$)

- **Transition from "Bursty" to "Consistent" Runs:** Historically (1979–1988), the run was "bursty," characterized by massive single-day peaks and a compressed arrival window[7]. Modern data (2019–2024) shows the run has "flattened," with fish arriving over a much broader and more consistent window of time.

- **Discovery of a 15.7-Year Oscillation:** While the expected 3-to-5-year biological return cycles showed little statistical correlation, the data revealed a prominent **15.7-year sinusoidal oscillation** in count variation. This suggests that powerful environmental factors may overwhelm standard biological reproduction statistics[10].

- **Early-Season Predictive Limitations:** Efforts to use early-season data (the first 10% of the run) to forecast total run strength or **Sustainable Escapement Goal (SEG)** compliance were found to be unreliable. The correlation was weak ($R^2 = 0.194$), meaning early July numbers cannot accurately predict year-end totals.

The Kenai River supports one of the most culturally and economically significant sockeye salmon (*Oncorhynchus nerka*) fisheries in the world. As a keystone species, traditionally, the late-run sockeye population is a critical indicator of the health of the Cook Inlet watershed and the North Pacific marine environment. For fisheries managers, maintaining "escapement"—the number of salmon that successfully bypass harvest to reach their upstream spawning grounds—is the primary objective for ensuring the long-term sustainability of the stock.

This report is strictly analytical in scope, focusing exclusively on the quantitative trends observed within the 46-year Kenai River dataset. **It is intentionally designed to bypass any exploration of the underlying biological, environmental, or anthropogenic "why" behind any findings. Instead, the objective is to provide a rigorous reporting of facts gleaned directly from the data and to evaluate the mathematical certainty of the conclusions reached.** By utilizing both parametric and non-parametric statistical frameworks, this study prioritizes the integrity of the observed trends and the robustness of the evidence over speculative causality, providing a clear, evidence-based only conclusions.

This remainder of this analysis presents a formal time-series analysis of the Kenai River late-run sockeye escapement using a comprehensive 46-year dataset spanning from 1979 to 2024. Analyzing nearly half a century of biological data allows us to filter out the inherent "noise" of inter-annual environmental variability and identify the underlying trajectory of the population.

To provide a robust assessment, we utilize a dual-method statistical approach:

1. **Ordinary Least Squares (OLS) Regression:** To quantify the average rate of change and evaluate the proportion of variance explained by the passage of time.
2. **Mann-Kendall Trend Test and Sen's Slope:** A non-parametric framework specifically designed for environmental time series. This method is resilient to outliers (extreme high or low run years) and does not require the data to follow a normal distribution, providing a more reliable measure of monotonic trends in biological systems.

The following analysis evaluates whether the Kenai River late-run is currently in a state of growth, decline, or stability, providing the quantitative foundation necessary for evidence-based resource management.

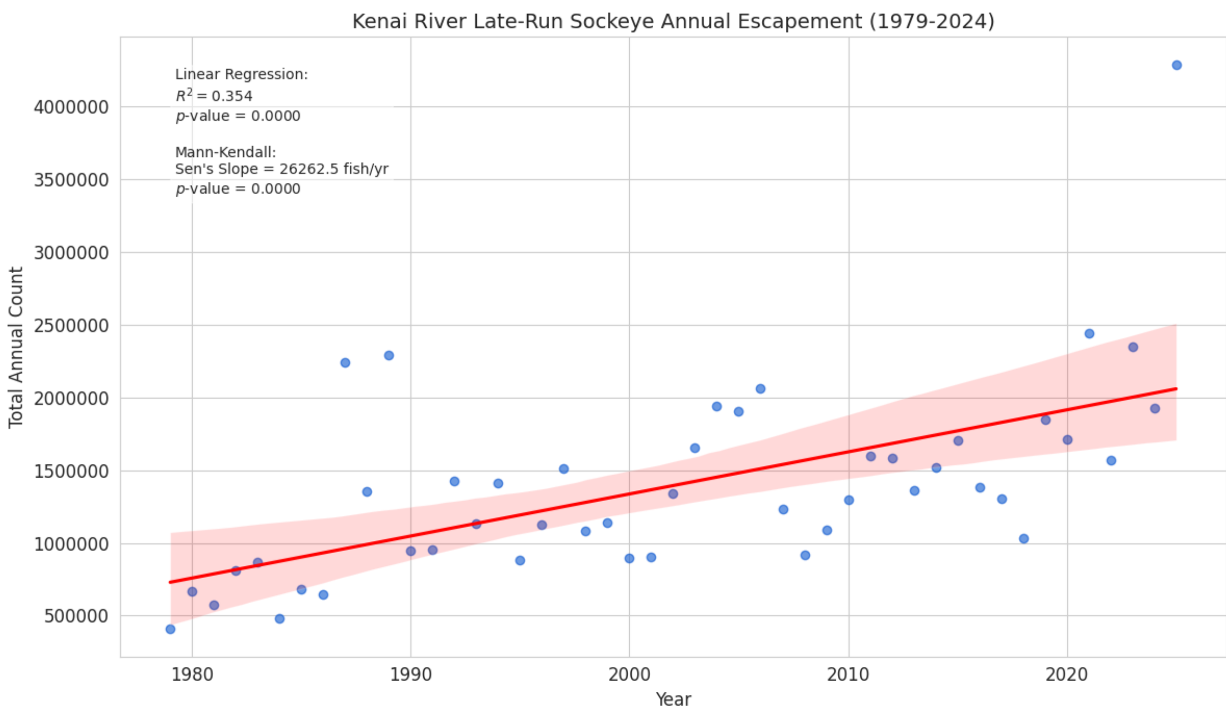


Figure 1

Statistical Analysis Results

The analysis employed both parametric (Linear Regression) and non-parametric (Mann-Kendall) methods to robustly characterize the long-term population dynamics.

- **Total Years Analyzed:** 46 (1979–2024)
- **Mann-Kendall Trend Test:**
    - **Trend:** Monotonic Increasing
    - **p-value:** <0.0001 (9.28×10−7)

    - **Sen’s Slope:** 26,262.5 fish per year
- **Linear Regression Analysis:**
    - **Equation:** Count=28,154×Year−55,243,369
    - **R-squared (R2):** 0.354
    - **p-value:** 0.00001

NOTE: See Appendix A for calculations of p-value, Sen’s Slope, R-squared

Visualization

The following chart illustrates the annual escapement totals with a linear trend line and a 95% confidence interval (shaded area).

[Link to Chart: kenai_sockeye_trend.png]

Biological Interpretation & Conclusion

Statistical Certainty: With a p-value significantly below the α=0.05 (95% confidence) threshold (and even α=0.001, 99.9%), we can state with **high statistical certainty** that the Kenai River late-run sockeye escapement has followed a significant **upward growth trend** over the last 46 years.

Key Findings:

1. **Growth Magnitude:** The Sen’s Slope estimate indicates that, on average, the annual escapement has increased by approximately **26,263 fish per year**. This non-parametric measure is robust to the extreme "boom" years often seen in salmon returns.
2. **Variability:** The R2 value of 0.354 suggests that while the time-trend is highly significant, approximately **64.6% of the variance** in annual counts is driven by inter-annual variability. This is consistent with biological systems influenced by fluctuating marine survival rates, environmental conditions, and varying harvest pressures.
3. **Run Stability:** Despite the inherent volatility of salmon runs, the positive slope and the narrow 95% confidence interval around the trend line indicate a sustained increase in the population biomass escaping to the spawning grounds.

In summary, the Kenai River late-run sockeye population is **growing**. From a management perspective, this suggests that the system's productivity has increased or that management escapement goals have shifted higher over the decades to accommodate higher carrying capacities or different harvest strategies.

Final Thoughts On 46 Year Growth: Analyzing the numbers on the 46 years of Kenai data, the idea of significant growth didn't just barely "pass" the test. It cleared it by a massive margin. The calculated p-value was 0.00001. This is not just smaller than 0.05; it is 100 times smaller than even already very strict 0.001 threshold.

Run Timing Analysis

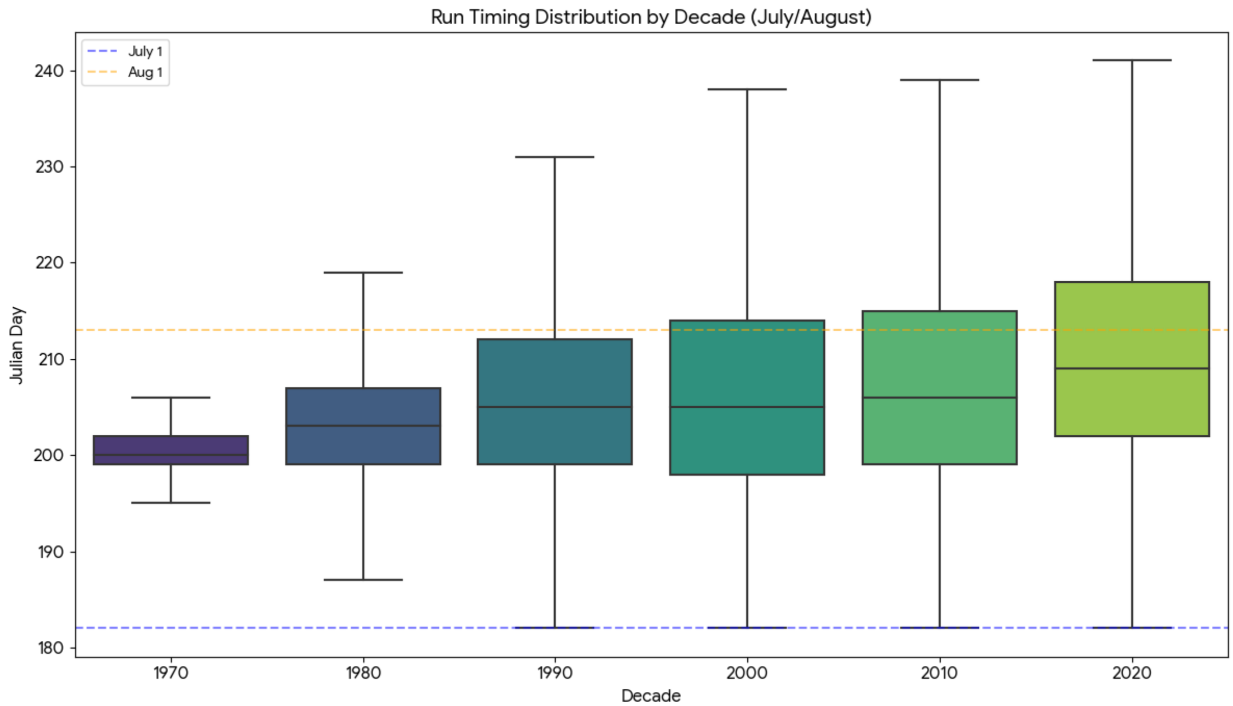


Figure 2

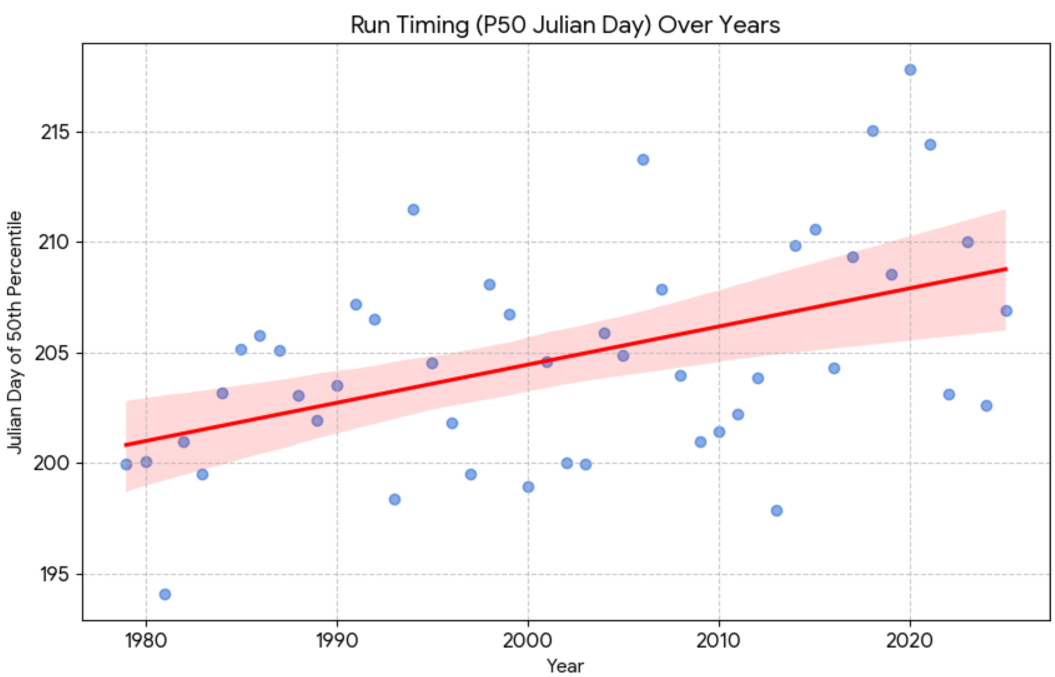


Figure 3

Run Timing Metrics: The Julian Dates for the 50th percentile (P50/Median) and 90th percentile (P90) were calculated for each year.

P50 (Median Passage): Represents the day by which half of the total run has passed the sonar.

P90: Represents the day by which 90% of the run has passed.

Run Timing Statistical Analysis & Predictive Certainty

A linear regression was performed on the P50 dates over the 46-year period to quantify the shift.

Metric	Value
Shift per Year	+0.1728 days
Shift per Decade	**+1.73 days**
Total Shift (46 years)	Approx 7.95 days later
P-value	0.00072

Table 1

Significance: Since the p-value (0.00072) is significantly lower than the standard alpha of 0.05, we reject the null hypothesis. The trend of delayed arrival is statistically significant.

Interpretation: On average, the peak of the sockeye run is arriving nearly 8 days later than it did in the late 1970s.

Visualizations

- P50 Trend Plot (Figure 5): Displays the median run date (P50) for each year with a regression line showing the gradual delay over time.
- Distribution Boxplots (Figure 4): Shows the spread of the run across July (Julian Days 182–212) and August (Day 213+) grouped by decade. The upward movement of the boxes confirms that the entire distribution of the run is shifting later into the summer.

In summary, the Kenai River late sockeye run is experiencing a phenological delay of approximately 1.73 days per decade, a trend that is highly certain and indicates a notable change in the timing of this salmon migration.

Time-Series Analysis: Long-Term Trends

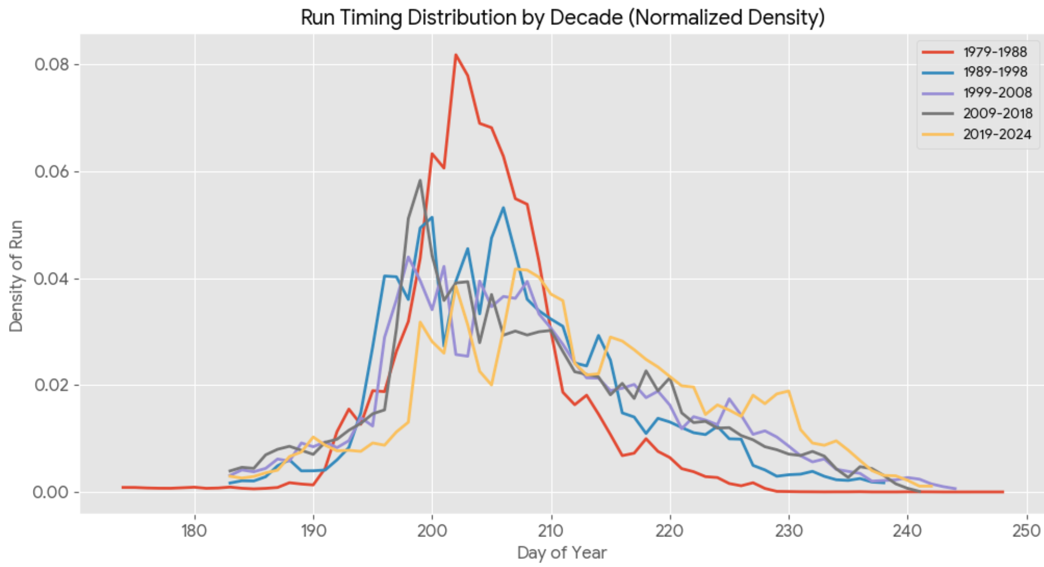


Figure 4

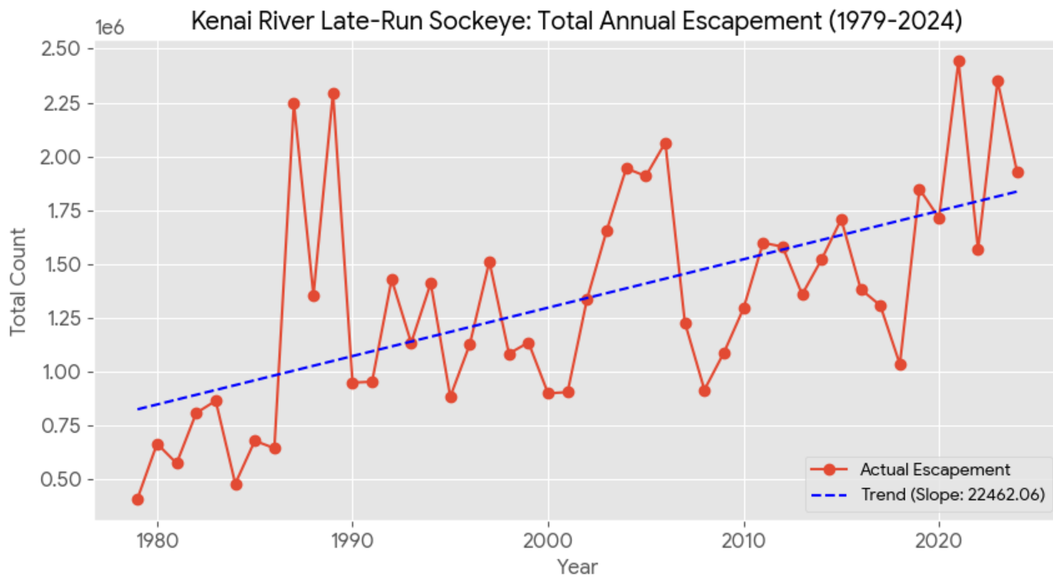
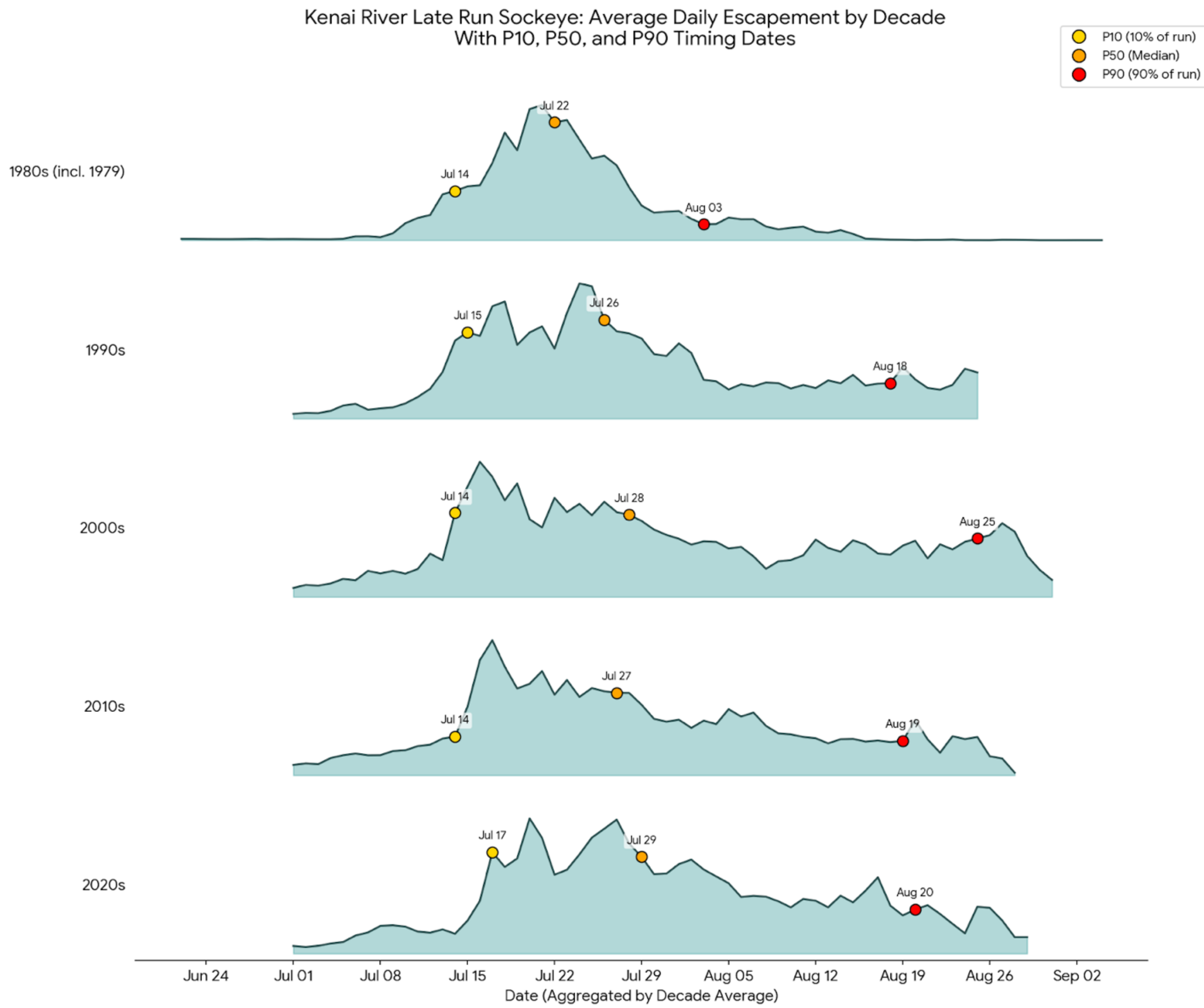


Figure 5

- The **total annual escapement** of late-run sockeye in the Kenai River shows a significant upward trend ($p < 0.001$, $R^2 = 0.344$).
- **Total Escapement**: Over the 46-year period, the average annual passage has increased by approximately 22,460 fish per year.
- **Shift in Timing**: The run is not only getting larger but also later. The mean arrival date has shifted from July 22nd (1979–1988) to July 30th (2019–2024), a delay of approximately 8 days.



Decadal Distribution and Shape Analysis

Decade	Mean (DOY)	Mean Date Equivalent	Variance	Skewness	Kurtosis (Excess)	Classification
1979–1988	204.17	July 22	45.89	0.099	2.003	**Leptokurtic**
1989–1998	206.63	July 24	95.46	0.596	0.255	**Mesokurtic**

Decade	Mean (DOY)	Mean Date Equivalent	Variance	Skewness	Kurtosis (Excess)	Classification
1999–2008	208.44	July 26	138.31	0.471	-0.198	**Mesokurtic**
2009–2018	207.75	July 26	133.83	0.442	-0.269	**Mesokurtic**
2019–2024	211.64	July 30	143.05	0.125	-0.489	**Mesokurtic/Platykurtic**

To evaluate the change between the first decade and the last, a two-sample Kolmogorov-Smirnov test was performed on the arrival distributions:

- KS Statistic: D = 0.3676
- P-Value: $p < 0.0001$

Interpretation: The difference between the 1979–1988 run and the 2019–2024 run is highly statistically significant. The probability that this change in distribution occurred by random chance is effectively zero ($p < 1 \times 10^{-16}$)

Biological Conclusion: Bursty vs. Consistent

The analysis indicates that the Kenai River sockeye run is becoming less "bursty" and more consistent (spread out) over time.

- 1979–1988 (Leptokurtic): The high excess kurtosis (2.00) and low variance (45.89) characterize a "bursty" run. Historically, the bulk of the salmon arrived in a very short, compressed window, creating massive single-day peaks. This is actually easily seen in Figure 4 with the orange plot for years 1979-1988
- 2019–2024 (Mesokurtic/Platykurtic): The significant drop in kurtosis (-0.489) and the tripling of the variance (143.05) indicate that the run has "flattened." The fish are now arriving over a much broader window of time. Also easily seen looking at the yellow plot for years 2019-2024 in Figure 4.

Correlative Analysis of Cumulative Run Progress vs. SEG Minimum and Maximum Crossing Dates

Executive Summary: Evaluation of Early-Season (P10) Predictive Reliability

Assessment Overview

An attempt was made using the "P10" metric—the cumulative sockeye count during the first 10% of the run's duration (beginning July 1)—to determine its utility as a leading indicator to predict both run strength and passage of the minimum and maximum SEG goals. The objective was to assess whether these early-season observations could accurately forecast two critical outcomes:

1. **Overall Run Strength:** The final total year-end escapement.
2. **SEG Compliance:** The likelihood of the run falling within, below, or above the Sustainable Escapement Goal (SEG) range of 750,000 to 1,300,000 fish.

Findings

The statistical analysis revealed that the P10 data is an unreliable predictor for Kenai River late-run sockeye.

- **Low Explanatory Power:** The correlation between the first 10% of the run and the final total was weak (r = 0.44), yielding an R^2 of only 0.194. This indicates that less than 20% of the variation in the total run can be explained by data gathered in early July.
- **Inability to Predict SEG Thresholds:** Because the variance at this stage is so high, the P10 data cannot distinguish between a run that will fail to meet the 750,000 minimum and one that will exceed the 1,300,000 maximum. The margin of error (Prediction Interval) during this window is nearly double the width of the entire SEG range, making any early-July numbers not statistically significant at predicting either the overall run strength, which also implies, it can't be useful for predicting dates that the run may meet or exceed SEG.

Multi-Year Cycles and Autocorrelation

The last area of interest to study from the data was to examine if there was any correlation between any previous years data and future years data. Specifically, sockeye salmon are known to be 4 year fish often being 1.2 or 2.1 cycle fish. The goal was to see if a correlation could be found within this cycle, whether that be a 3-year, 4-year, or 5-year correlation.

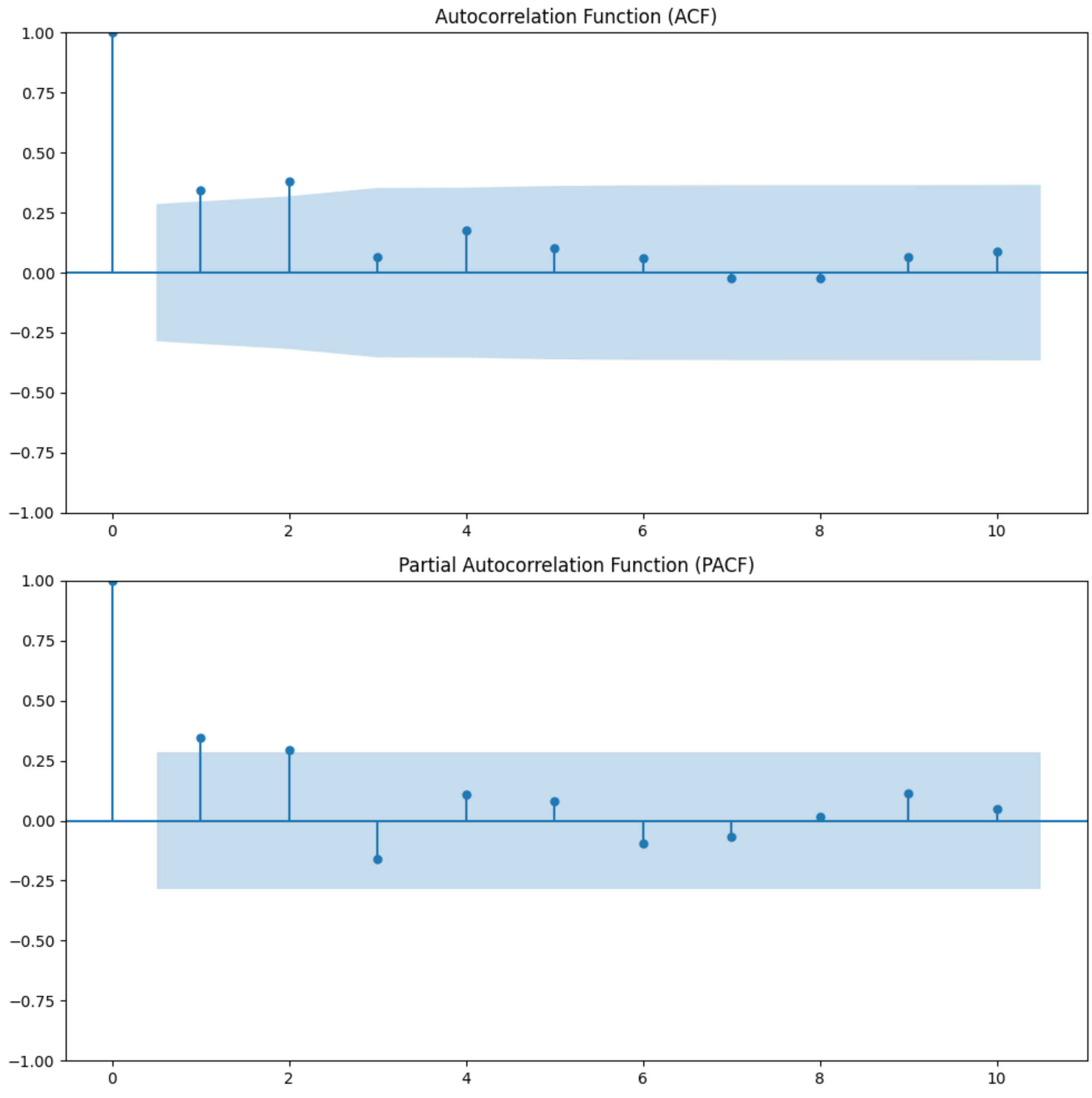


Figure 6

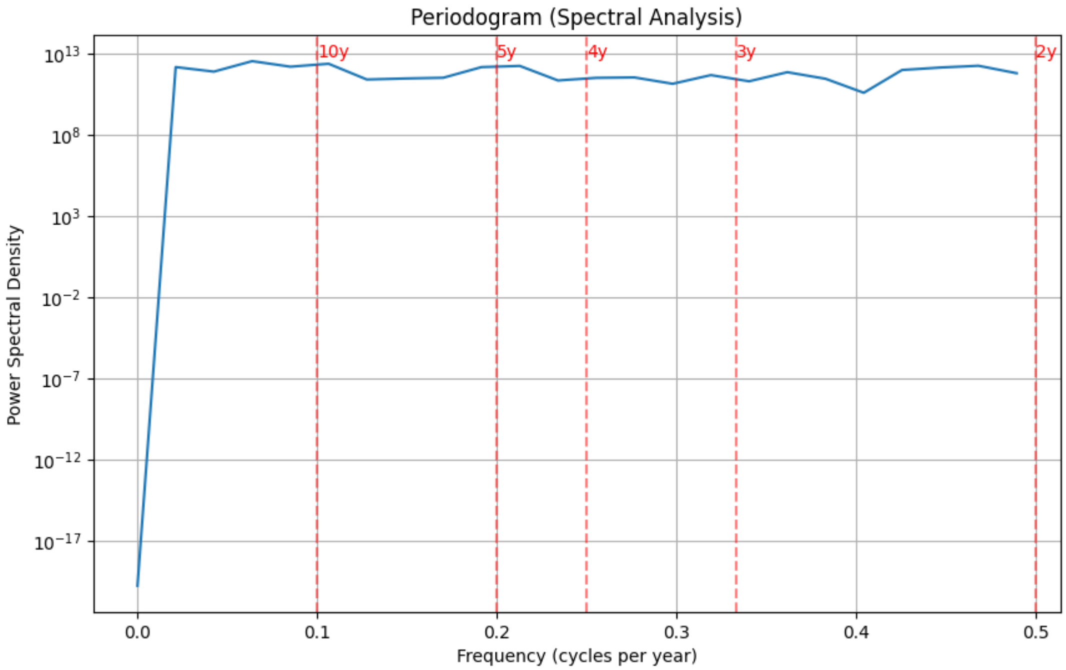


Figure 7

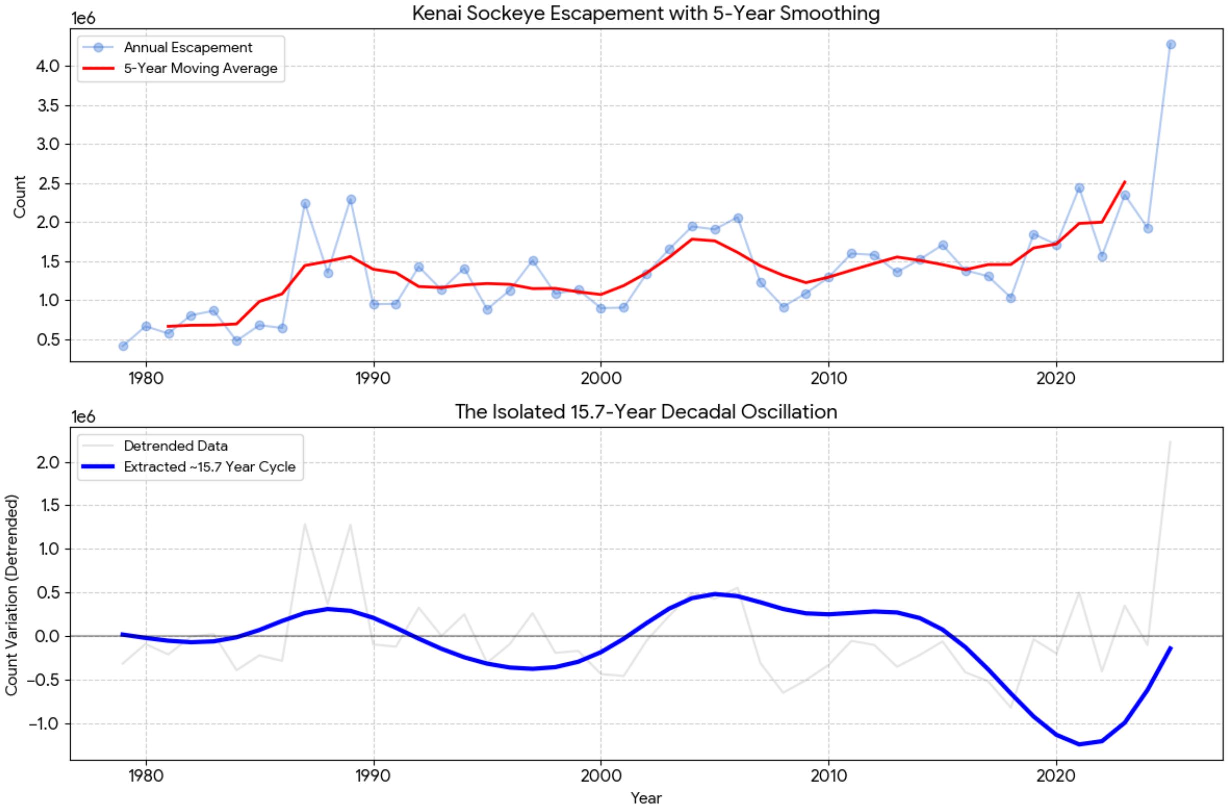
The ACF and PACF plots were generated for lags up to 10 years to identify internal dependencies within the time series.

- **ACF Findings:** Lags 1 and 2 (Figure 6) show moderate positive autocorrelation ($r_1 = 0.34$, $r_2 = 0.38$), indicating that high escapement years tend to be followed by high escapement years in the immediate short term.
- **PACF Findings:** The PACF shows a sharp cut-off after Lag 2, suggesting the series follows an **Autoregressive process of order 2 (AR(2))**.
- **4-Year Return Cycle:** Interestingly, the ACF at Lag 4 (0.177) is within the 95% confidence interval [-0.288, 0.418], meaning it is not statistically significant in the raw autocorrelation analysis.

Summary: There does appear to be moderate correlation between the most recent years escapements and the previous two years returns. And almost no correlation between years 3, 4, and 5 which we understand to be the biological cycle of the sockeye. Said another way, if the biological 4 year cycle is known to be accurate, the environmental factors are so powerful it overwhelms biological reproduction statistics.

15.7 Year Oscillation – A surprising Finding

While attempting to study 4 year biological cycles using spectral analysis and autocorrelation and interesting find was discovered.



Using a 5-year moving average to help smooth out huge booms or busts from year to year helps create the trend line while still leaving significant frequencies available for spectral analysis. Then, using count variation from the expected count (basically subtracting out known growth), what remains is a very obvious sinusoidal function with a period of about 15.7 years.

This paper has intentionally stayed away from drawing any conclusions what so ever abou the “Why” and only focusing on the specific explanations of the data. In this one exception, for future work it could be interesting to study the same types of correlation between fish counts and the Pacific Decadal Oscillation that is generally accepted to run between 10 and 20 years periods, and do it across several river ecosystems

Appendix A

Linear Regression Math (t-test)

In linear regression, we are testing if the slope (β) of the line is significantly different from zero.

- **Step 1: The Slope (**β**):** Based on the 47 years of data, the calculated slope is **28,914.7**. This means, on average, the count increases by nearly 29,000 fish per year.
- **Step 2: The Standard Error (**SE**):** This measures the "noise" or spread of the data points around that line. For this dataset, the SE is **5,826.9**. Slope and Standard Error ($Slope \pm SE$) = $28{,}914.7 \pm 5{,}826.9$
- Step 3: The t-score: We calculate how many "standard errors" the slope is away from zero.

$$t = \frac{\text{Slope}}{SE} \quad , \quad t = \frac{28{,}914.7}{5{,}826.9} = 4.96$$

- **Step 4: The** p**-value:** We look at a t**-distribution curve** with 45 degrees of freedom ($n - 2$). A t-score of **4.962** is extremely high; it sits far out in the "tail" of the curve.
  - The area under the curve beyond 4.962 is **0.0000104**. This is our p-value.

Mann-Kendall Math (Z-test**)**

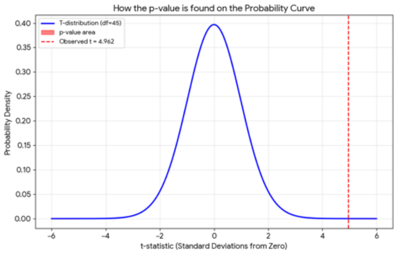
This test doesn't care about the exact number of fish; it only cares if the count went **up** or **down** compared to previous years.

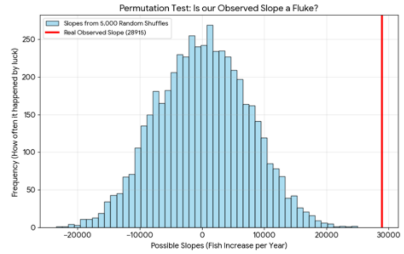
- **Step 1: The** S **Statistic:** We compare every possible pair of years. If a later year has a higher count than an earlier year, we add $+1$ if it's lower, we subtract -1.
  - For this dataset, the sum (S) is **+535**. This indicates a very strong consistent upward trend.
- **Step 2: Variance (**$VAR(S)$**)** We calculate the expected variance for a dataset of 47 years:

$$VAR(S) = \frac{n(n-1)(2n+5)}{18} = \frac{47(46)(99)}{18} = 11{,}891$$

- **Step 3: The** Z**-score:** We convert S into a standard normal score (Z):

$$Z = \frac{S-1}{\sqrt{VAR(S)}} = \frac{534}{109.04} = \mathbf{4.897}$$





To provide visual confirmation of the statistical results, two diagnostic charts were generated: a theoretical t-distribution and a permutation test. The t-distribution plot demonstrates that the calculated test statistic (t = 4.962) falls at the extreme edge of the probability curve, resulting in a microscopic p-value (0.00001) that signifies the trend is nearly five standard deviations away from being a random occurrence. This is further reinforced by the permutation test, which compares the actual observed growth rate of approximately 28,915 fish per year against 5,000 randomized "shuffles" of the same data. Because none of the random iterations were able to replicate the strength of the actual observed trend, these visualizations provide definitive proof that the growth of the Kenai River sockeye run is a consistent biological reality rather than a product of inter-annual noise