Modeling Sockeye Salmon Populations in the Kruzgamepa River, Alaska, in Relation to Dynamic Fishery Management Strategies

December 2025

Abstract

The Kruzgamepa River (also known as the Pilgrim River) of the Seward Peninsula in Northwest Alaska is home to a predominantly Native Alaskan subsistence fishery. Their fishery is critical for local food security and the preservation of indigenous fishing practices. The health of the fishery also has significant impacts on the Native Alaskan communities which rely on it and local ecosystem. Our report aims to model sockeye salmon counts through singular spectrum analysis (SSA) to implement well-rounded fishery management decisions in the region, as well as give some indication of the health of local ecosystems. We conclude that SSA is effective, providing reasonable weekly average estimates and forecasts based on the 2003–2014 data. In particular, the SSA-based time series model properly estimates the overall trend likely affected by changes in catch limits, and two distinct seasonal components possibly caused by two sockeye salmon runs.

1 Background and Research Objective

Sockeye salmon have been a cornerstone of human and animal diets for millennia in North America. Native Alaskan communities subsist on these nutrient-rich fish, but overfishing and changes in seasonal weather patterns in recent decades have threatened local sockeye salmon populations. Sockeye salmon are also vital to coastal ecosystems like that of the Seward Peninsula, where more than 130 species of animals feed on salmon [3], including large mammals like bears and orcas.

Our study aims to analyze time series data of sockeye salmon counts in the Kruzgamepa River (also known as the Pilgrim River) on the Seward Peninsula, a waterway in the Nome census area, Alaska, for modeling and forecasting future salmon runs. The Kruzgamepa River begins at Salmon Lake and empties into the Imuruk Basin where it eventually drains into the Bering Sea. The indigenous Inuit people (specifically the Inupiat tribe) have occupied the land surrounding the Seward Peninsula for at least 4000 years, and rely on subsistence pursuits of hunting and fishing to live. For example, surveys from Kotzebue, a city slightly north of the Seward Peninsula, show salmon make up roughly 13% of subsistence harvests by residents [4]. The time series model we develop in our study will be used to gain insights into the health of the local ecosystem over time, and forecasting will be used to guide expectations and management decisions for future seasons.

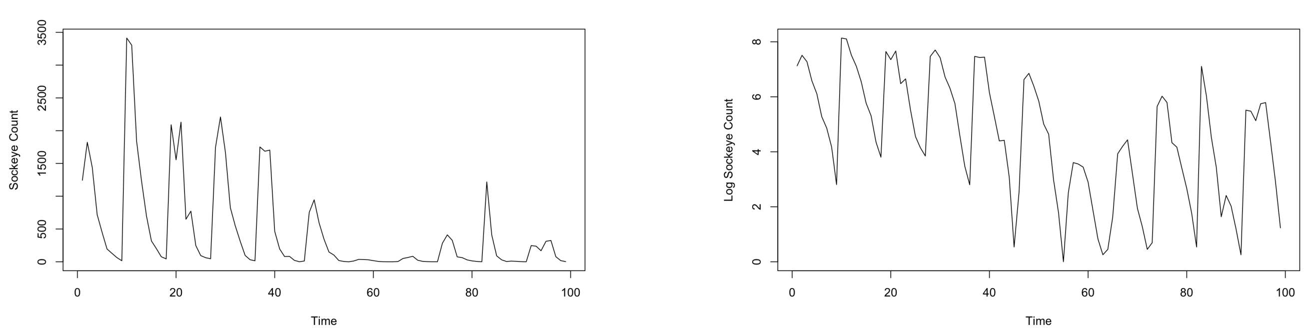
2 Data

2.1 Data Collection

The dataset was collected by the United States Geological Survey (USGS) and the Alaska Department of Fish and Game (ADFG), and is available on the USGS website [1]. The dataset recorded the daily river temperature, river height, and count of sockeye salmon passing through a fishing weir in the Kruzgamepa River from late June to mid-September in 2003–2014 (12 years). Note that the fishing weir was operated during the summer by the Norton Sound Economic Development Corporation.

2.2 Data Cleaning

The length of the observation periods differed slightly from year to year with 2012 having an abnormally low amount of observations. For that reason, the year 2012 was removed and the lengths of the observation periods were shortened to the next shortest period, resulting in a period length of 63 days per year after standardization. Then, linear interpolation was performed on the missing values (less than 1.2% of data) so that analysis could be performed on an uninterrupted time series. Weekly averages were taken after the linear interpolation, resulting in 9 weekly averages per year over 11 years for a total of 99 observations. Finally, a log transformation was performed on these 99 observations to make the distribution of data more symmetric with a narrower range of values.



(a) Original time series.

(b) Log-transformed time series.

Figure 1: The original and log-transformed weekly average sockeye salmon count time series (2003-2014 without 2012) from the Kruzgamepa River.

3 Methods

3.1 Singular Spectrum Analysis

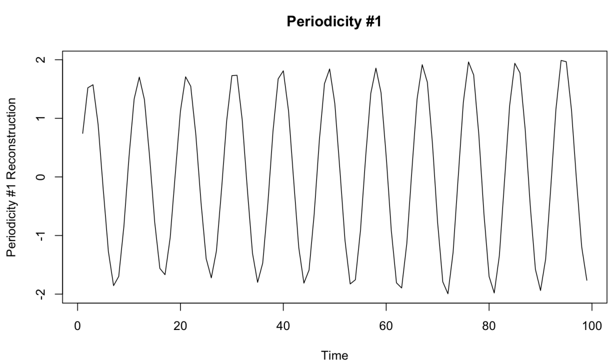
The singular spectrum analysis (SSA) aims to "decompose the observed time series into the sum of interpretable components with no a prior information about the time series structure" [2]. Implementing the SSA means identifying the time series components as trend, seasonality, or noise. The time series decomposition can be achieved by examining the relationship between eigenvalues and eigenvectors from a particular trajectory matrix resulting from the time series through embedding. For example, a single high-magnitude eigenvalue may indicate trend in the time series, while a pair or eigenvalues of a similar magnitude, corresponding to a pair of similar looking eigenvectors, typically indicates a possible unique seasonality (Fig. 4b). Furthermore, each of these eigenvector pairs typically forms a regular polygon, where its number of vertices represents the period (Fig. 5a). Lastly, the non-noisy parts of w-correlation matrix plot indicates that the first seven eigenvalues are significant (Fig. 5b). Thus, we conclude that the first, fourth and fifth eigenvalues correspond to a trend, the second and the third eigenvalues correspond to the first major period, and the sixth and seventh eigenvalues correspond to the second major period.

3.2 Autoregressive Integrated Moving Average Model

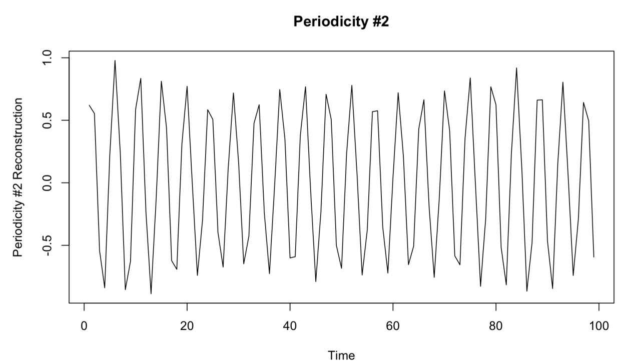
In addition, various autoregressive integrated moving average (ARIMA) models, each one involving an autoregressive component, differencing, and a moving average component, were created and tested [5]. However, these ARIMA models were deemed ineffective and were thus omitted from the report.

4 Results

The SSA performed on the log-transformed weekly average sockeye salmon count time series turned out to be effective for modeling and reasonable for forecasting. The two seasonal components identified earlier are reconstructed below, with the first having a period of roughly 9 weeks (Fig. 2a), and the other having a period of roughly 4.5 weeks (Fig. 2b).



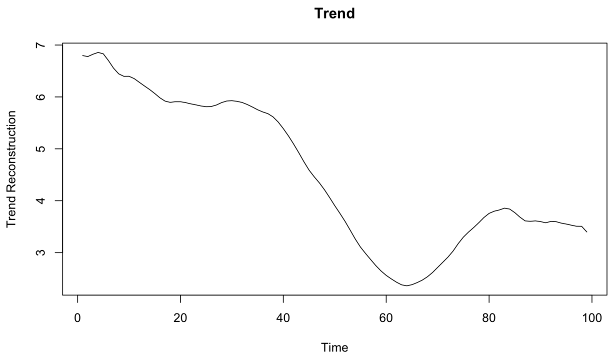
(a) First seasonal component.



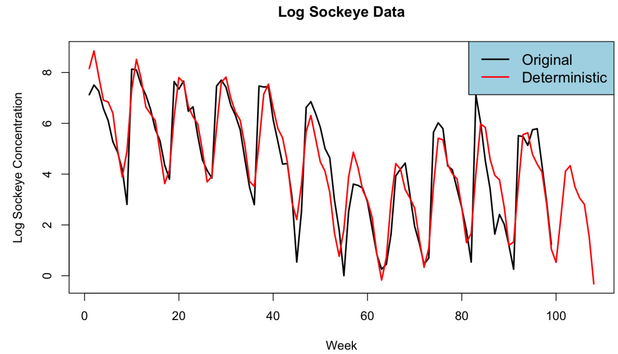
(b) Second seasonal component.

Figure 2: Reconstructed seasonal components using SSA.

In addition, the trend component (Fig. 3a) shows an initial decrease in the weekly average sockeye salmon counts followed by a later increase. Moreover, both the autocorrelation function (ACF) (Fig. 7a) and partial autocorrelation function (PACF) (Fig. 7b) of the residuals show no significant pattern in lags. Similarly, the portmanteau test with p-values all exceeding the 5% significance level for the first seven lags implies no significant autocorrelation in the residuals (Fig. 6a). A Q-Q plot of the residuals (Fig. 6b) indicates the residuals are roughly normal, as does the Shapiro-Wilk normality test, with a p-value of 0.07462. Lastly, the reconstruction of the time series by combining the trend and seasonality components appears to closely follow the original data (Fig. 3b), and 1- to 12-month-ahead forecasts at the end of the same plot appear to have a downward trend.



(a) Reconstructed trend component using SSA.



(b) Log-transformed weekly average sockeye salmon count time series (in black) and reconstructed time series with 1- to 9-week-ahead forecasts (in red).

Figure 3: Reconstructed trend component (left) and reconstructed time series with forecasts (right).

5 Discussion

The SSA-based time series modeling is an effective strategy for modeling the weekly average sockeye salmon count data at the Kruzgamepa River. The model identified two seasonal components based on the 2003–2014 data; a period of 9 weeks corresponding to a yearly periodicity as there are only 9 weekly average observations per year, and a period of 4.5 weeks corresponding to a period a little greater than one month. These two seasonal components likely coincide with the conventional understanding that sockeye salmon runs have a primary and secondary run each summer. Our analysis shows that the secondary run is significantly lower in volume and harder to detect in the time series data, but is still present.

The model also reveals a trend in the data, where there is a clear decline in yearly run volume for the first seven years. The decline is likely explained by overfishing, as ecological conditions had little variation over the time period with consistent river temperatures. To prevent overfishing, catch limits have been introduced and waived since 1985 based on escapement and returns recorded by the ADFG, as the decline in run volume had real-world impacts for food security in the region. Specifically, fishery managers dynamically introduce catch limits to manage population health and local subsistence needs. More recently in the data, one can see the recovery in run volume by the increasing trend from around 2010 onwards, indicating that there may have been some reduction in catch limits by the ADFG.

The trend indicates the local sockeye salmon population, and therefore local ecosystems, are fairly healthy at the end of the data period (i.e., 2014). Despite that, the counts towards the end of the data are still significantly below historical weekly averages and the counts in the beginning of the observation period (i.e., 2003–2005). Thus, constant monitoring of the river is of value, as an upward trend in global ocean and river temperatures could continue to impact the overall decline in the sockeye salmon population. If the population counts kept dropping, as implied by the model-based forecasts, it could prove devastating to the Inupiat people that rely on the fish for sustenance. At the same time, these forecasts must be interpreted with caution due to the limited number of weekly average counts available in the time series data used for modeling. Further research would ideally model sockeye salmon counts in conjunction with local weather and climate data, such as river temperature, as ecosystem conditions are known to influence sockeye salmon spawn rates.

References

[1] Carey, M.P., & Zimmerman, C.E. (2017). *Count of Sockeye Salmon (Oncorhynchus Nerka), River Temperature, and River Height in the Pilgrim River, Nome, Alaska, 2003-2014*, USGS, 5 May 2017. Available at https://doi.org/10.5066/F71834PF.

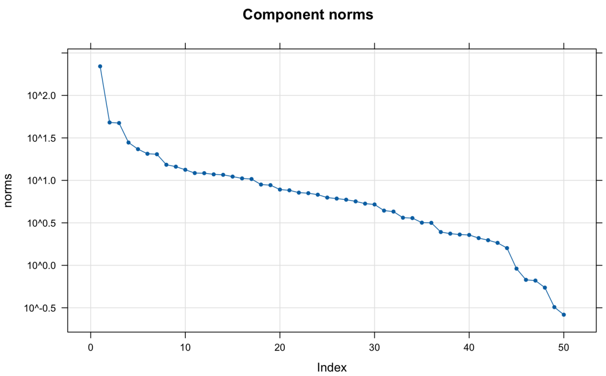
[2] Golyandina, N. & Korobeynikov, A. (2014). Basic singular spectrum analysis and forecasting with R. *Computational Statistics & Data Analysis*, 71, 934–954.

[3] Repasch, M. (2020). *Why Are Salmon Important To The Ecosystem*. Ocean Blue Project. Accessed November 13, 2025.
Available at oceanblueproject.org/the-plight-of-the-wild-salmon/.

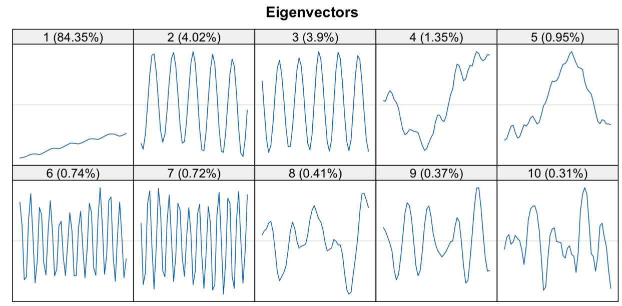
[4] SASAP: State of Alaska Salmon and People (2025). *At a Glance: Salmon as part of Kotzebue's rich portfolio of resources*. Accessed December 15, 2025.
Available at alaskasalmonandpeople.org/region/kotzebue/.

[5] Shumway, R.H, & Stoffer, D.S. (2016). *Time Series Analysis and Its Applications*, Springer, New York.

6 Appendix

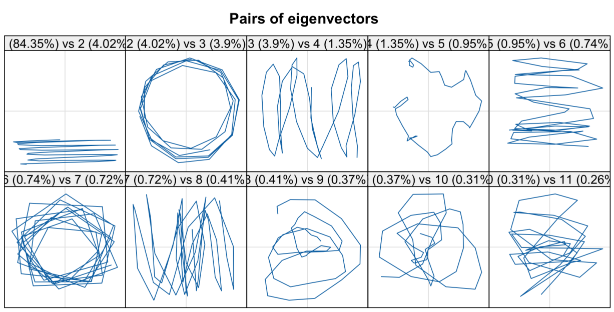


(a) SSA component norms representing the square roots of the eigenvalues from the trajectory matrix: The component norms before the tail off (the first seven of them) are considered significant for modeling the data.

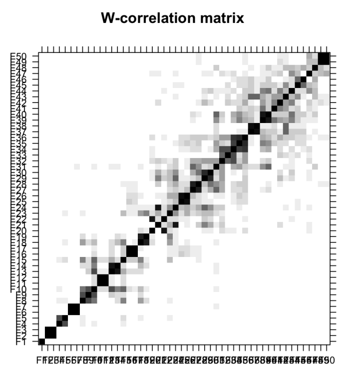


(b) SSA model eigenvectors from the trajectory matrix: Similar reconstructed shapes show eigenvector pairs suitable for modeling seasonal components.

Figure 4: SSA-based time series decomposition.

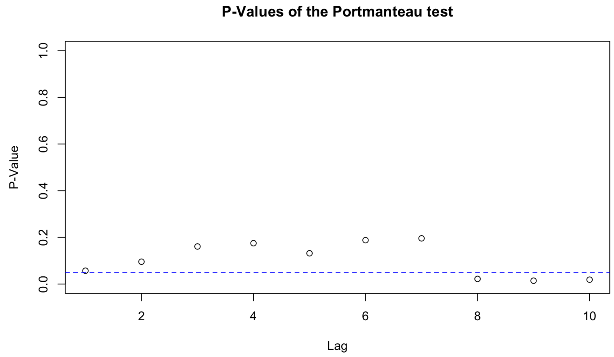


(a) SSA model eigenvector pairs: The number of vertices in each polygon represents the period for each seasonal component.

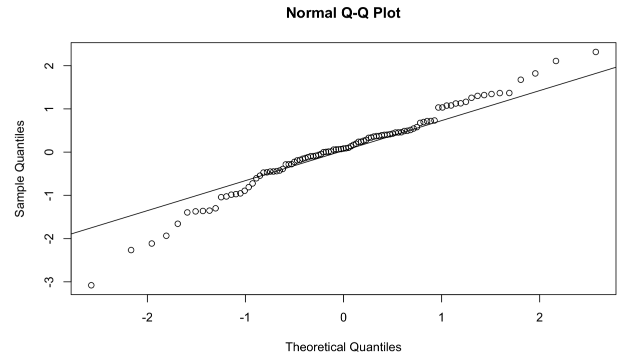


(b) SSA model w-correlation matrix: The eigenvectors corresponding to the eigenvalues with low w-correlation coefficients become noisy/fuzzy in this plot (i.e., after the seventh in this case), and these eigenvectors are considered insignificant for modeling the data.

Figure 5: SSA eigenvector groupings and noise floor determination.

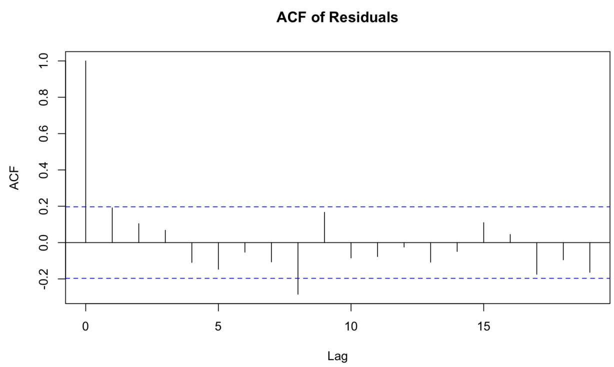


(a) The portmanteau test on the autocorrelation of the SSA-based model residuals: The p-values above the blue dashed line representing the 5% significance level are not considered statistically significant.

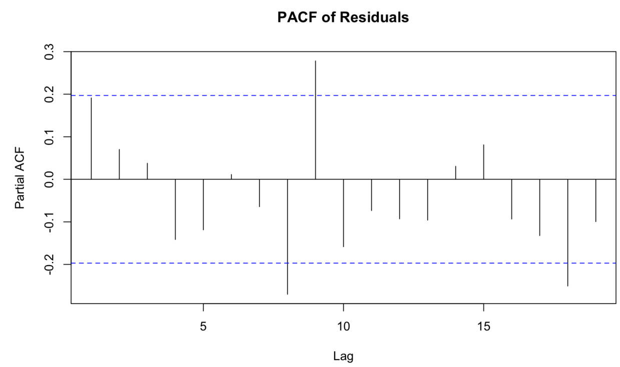


(b) Normality of the SSA-based model residuals: Adherence of the points to the straight line indicates normality.

Figure 6: Autocorrelation and normality of the SSA-based model residuals.



(a) The ACF plot of the SSA-based model residuals.



(b) The PACF plot of the SSA-based model residuals.

Figure 7: The ACF and PACF of the SSA-based model residuals: Because the spikes at each lag are within the blue dashed lines or barely exceed them, it is safe to conclude that all the major dependence structures from the original time series are captured by the SSA-based model, and thus the residuals are roughly consistent with the white noise process.