Time Series Analysis of Fish Recruitment in the Central Pacific

Abstract

We analyze the monthly fish recruitment series in the central Pacific Ocean (453 observations, 1950–1987) from the `astsa` R package, paired with the Southern Oscillation Index (SOI). Using spectral analysis, SARIMA modeling, and regression with ARMA errors, we characterize recruitment dynamics and investigate the SOI–recruitment association. A SARIMA model with seasonal differencing captures the dominant annual cycle. The contemporaneous SOI effect is not significant after accounting for autocorrelation; cross-correlation analysis reveals a 5–7 month lagged relationship, but the lagged regression also fails to reach significance due to strong internal dynamics, highlighting a limitation of regression with ARMA errors and motivating state-space approaches.

1 Introduction

Understanding the temporal dynamics of fish populations is central to fisheries science and management. Recruitment—the number of new individuals entering a population—is a key demographic process driven by environmental conditions such as ocean temperature and atmospheric circulation patterns, making it inherently variable and challenging to predict (Shumway and Stoffer 2017).

This report analyzes the monthly recruitment series for fish in the central Pacific Ocean, spanning January 1950 through September 1987 (453 observations), available in the `astsa` R package. We pair this with the Southern Oscillation Index (SOI), a measure of El Niño/La Niña activity, to investigate potential environmental drivers of recruitment variability.

Our analysis proceeds in three stages: (1) exploratory data analysis and spectral analysis to characterize temporal structure; (2) SARIMA modeling to capture trend and cyclical dynamics; and (3) regression with ARMA errors to examine the SOI–recruitment association, including a lagged specification motivated by cross-correlation analysis. This classical approach provides a phenomenological baseline that complements mechanistic models used in modern stock assessment, such as TMB (Kristensen et al. 2016) and POMP (King, Nguyen, and Ionides 2016). Our methodology follows the course notes (Ionides 2026) and the textbook treatment in Shumway and Stoffer (2017, Ch. 4).

1.1 Relationship to Previous STATS 531 Projects

Several previous STATS 531 midterm projects have analyzed environmental time series with similar methodology. A W25 project on Lake Malawi water levels (STATS 531 W25 Midterm Project 08 2025) applied SARIMA with seasonal differencing—a pattern we also encounter. A W24 project on Detroit precipitation (STATS 531 W24 Midterm Project 07 2024) used covariate regression;

peer reviewers noted the importance of lagged effects, which we address here. A W16 project on PM2.5 and temperature (STATS 531 W16 Midterm Project 06 2016) demonstrated regression with ARMA errors and likelihood ratio testing; we adopt the same framework. Our contribution focuses on fisheries recruitment and connects classical methods to state-space models.

2 Exploratory Data Analysis

The recruitment series contains 453 monthly observations from 1950 to 1987. Figure 1 shows the raw time series alongside the SOI.

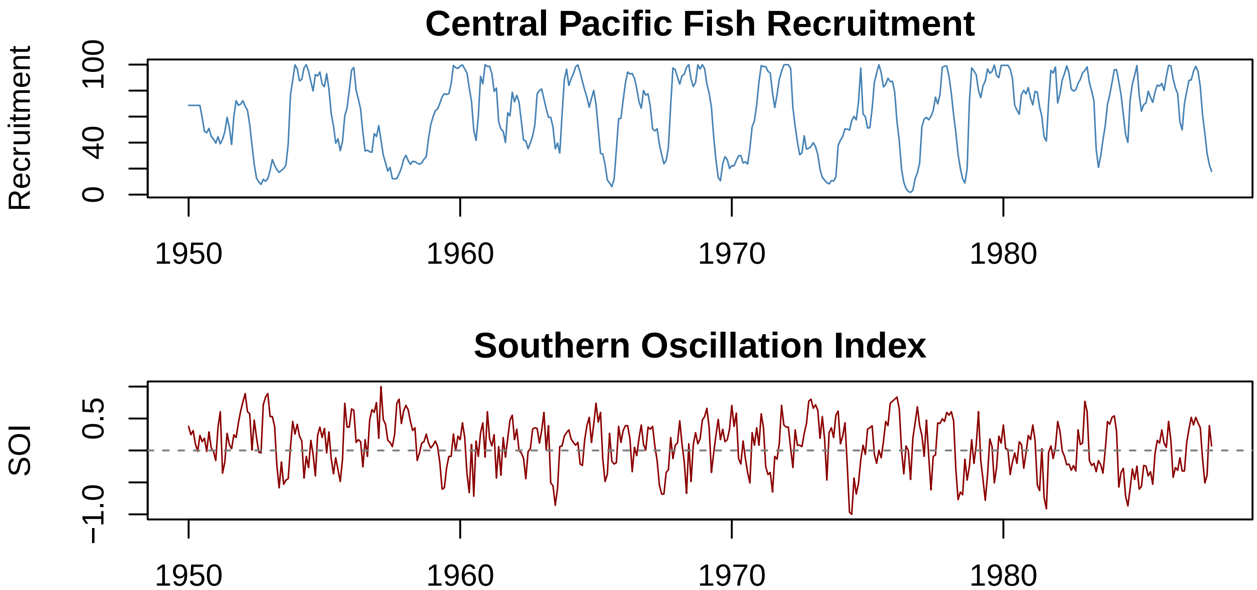

Figure 1: Monthly fish recruitment in the central Pacific (top) and Southern Oscillation Index (bottom), 1950–1987.

Recruitment ranges from 2 to 100 with a mean of 62.3 and standard deviation of 28. The series exhibits strong cyclical behavior and right-skewed variability, so we apply a log transformation to stabilize the variance. The log transformation produces a more symmetric distribution (Figure 7). A seasonal subseries plot (Figure 6) reveals modest seasonal variation, and an STL decomposition (Figure 8) confirms a slowly varying trend with multi-year oscillations, a stable seasonal component, and substantial irregular variation.

3 Frequency Domain Analysis

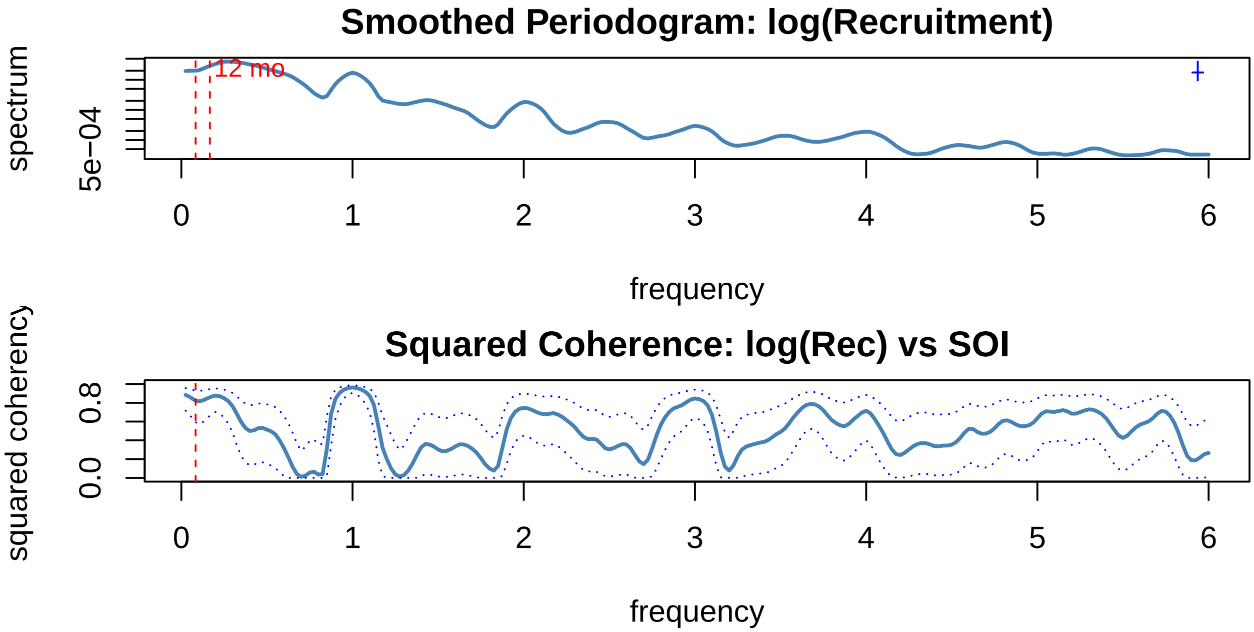

Figure 2: Smoothed periodogram of log recruitment (top) and squared coherence between log recruitment and SOI (bottom). Dashed red lines mark annual and semi-annual frequencies.

The periodogram (Figure 2, top) confirms a dominant peak near 1/12 cycles per month (annual cycle), along with substantial low-frequency power reflecting multi-year oscillations. The squared coherence (bottom) shows elevated coherence at low frequencies, indicating that recruitment variability shares structure with ENSO at interannual time scales. This motivates the regression analysis in Section 5.

4 SARIMA Model Selection

4.1 ACF/PACF Analysis and Differencing

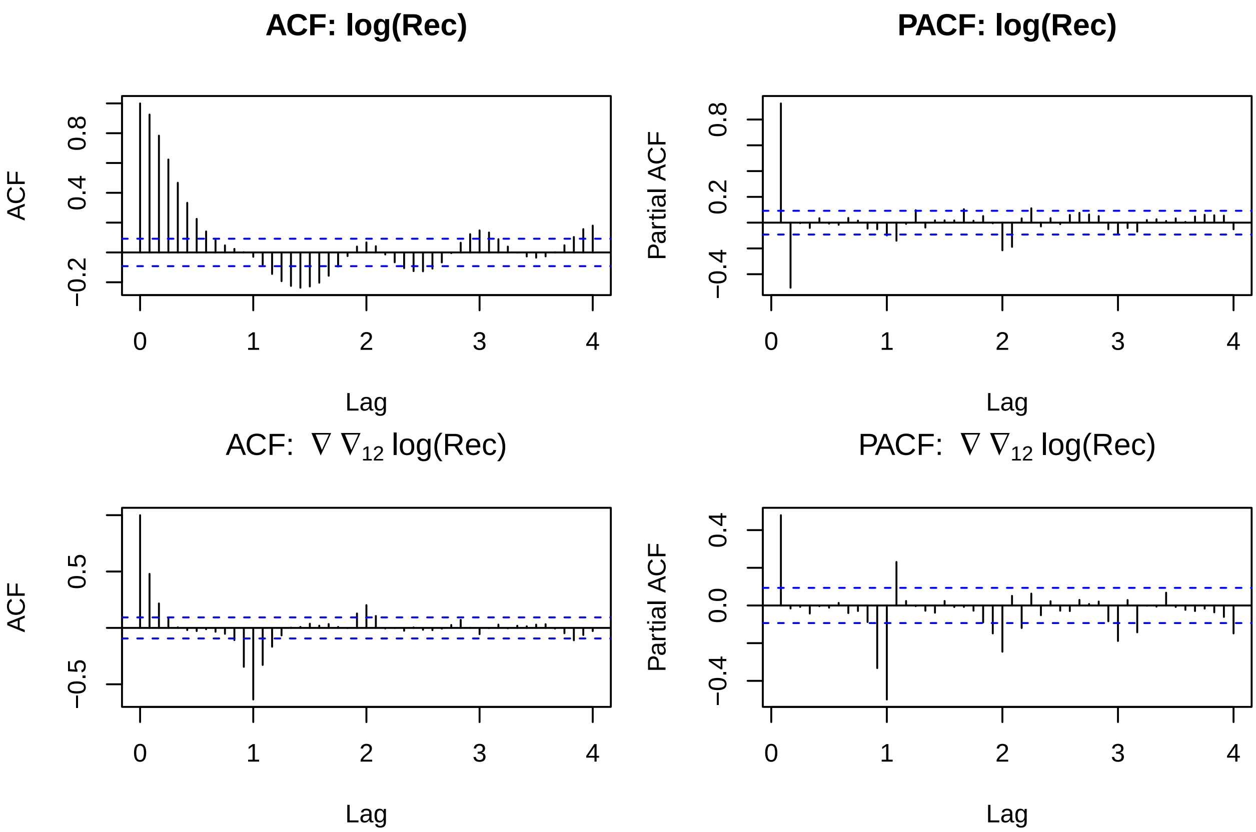

Figure 3: Sample ACF and PACF of log recruitment (top) and of the differenced series $\nabla\nabla_{12}\log(\text{Rec})$ (bottom).

The ACF of log recruitment (Figure 3, top left) decays slowly with a sinusoidal pattern, indicating non-stationarity. After applying $d = 1$ and seasonal differencing $D = 1$ (period 12), the ACF and PACF (bottom) show rapid decay with a significant negative spike at seasonal lag 1 in the ACF (suggesting seasonal MA(1)) and significant spikes at lags 1–2 in the PACF (suggesting low-order AR terms). An augmented Dickey–Fuller test confirms stationarity ($p < 0.01$).

4.2 Model Comparison

We compare eight SARIMA$(p, d, q) \times (P, D, Q)_{12}$ models with $d = 1$, $D = 1$ (Table 2). The best candidate by AIC is SARIMA(1,1,0)(1,1,1) (AIC = -129.94).

For comparison, `auto.arima()` selects SARIMA(2,0,1)(2,0,0)_{12} with AIC = -145.76. This stationary model ($d = 0$, $D = 0$) uses higher-order AR terms to capture persistence. We proceed with our best differenced model for parsimony, while acknowledging the stationary alternative achieves a lower AIC.

4.3 Diagnostics

We select **SARIMA(1,1,0)(1,1,1)** (lowest AIC among differenced candidates). Parameter estimates are in Table 1.

Table 1: Parameter estimates for the selected SARIMA model.

Parameter	Estimate	SE	z-value
ar1	0.3612	0.0464	7.78
sar1	-0.1142	0.0526	-2.17
sma1	-0.9784	0.0735	-13.31

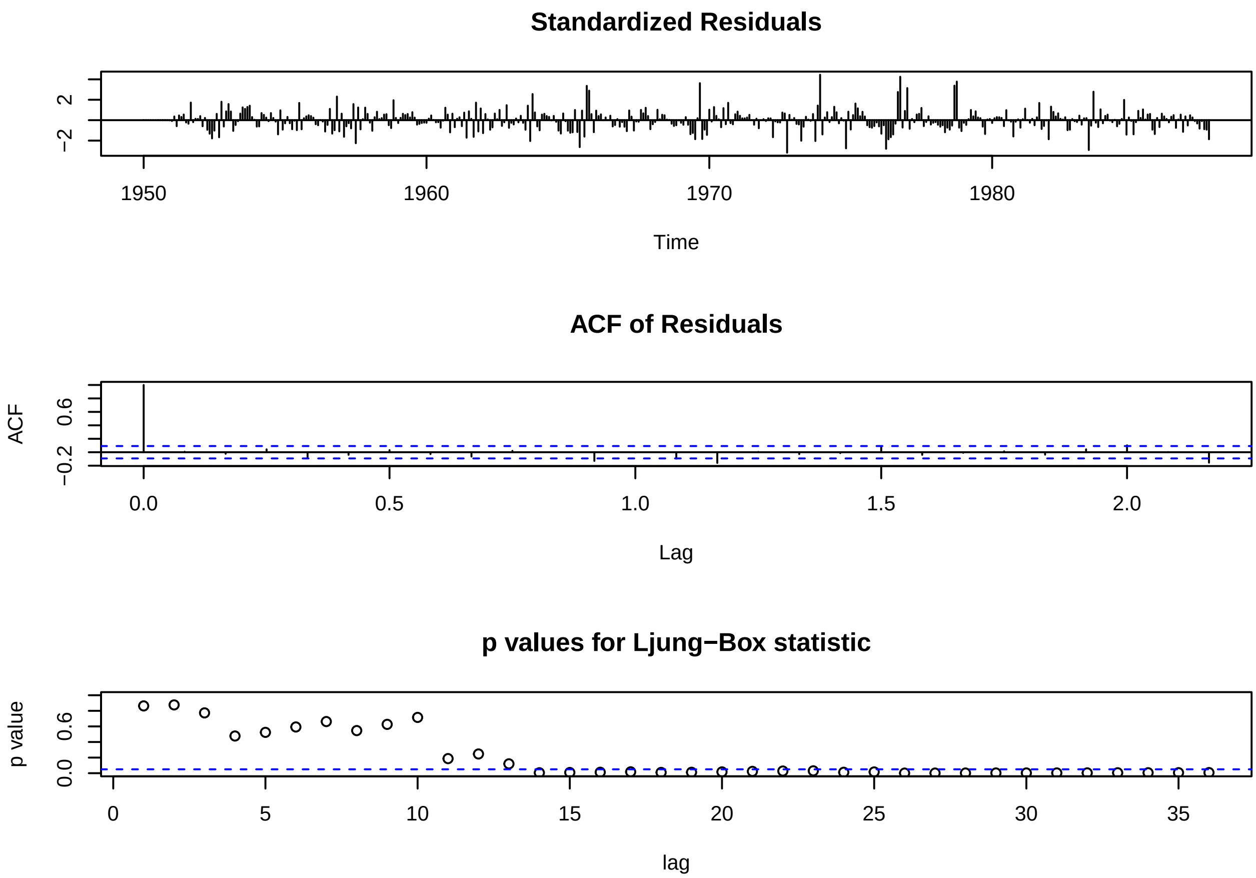

Figure 4: Diagnostic plots: standardized residuals (top), ACF of residuals (middle), and Ljung-Box p-values (bottom).

The residual ACF (Figure 4) shows no significant autocorrelation at low lags. Ljung–Box tests give $p = 0.094$ at lag 12 and $p = 0.004$ at lag 24, with p-values decreasing at higher lags, suggesting some residual structure consistent with `auto.arima()`'s lower AIC. The Q-Q plot (Figure 9) shows approximate normality with mild right-tail deviation.

5 Environmental Association: Recruitment and SOI

We model the SOI–recruitment association using regression with ARMA errors (Shumway and Stoffer 2017, Ch. 4), following the framework in STATS 531 W16 Midterm Project 06 (2016):

$$\log(\text{Rec}_t) = \beta_0 + \beta_1 \, \text{SOI}_t + \eta_t, \quad \eta_t \sim \text{ARMA}(p, q)$$

The best-fitting model uses ARMA(2,0) errors (AIC = -111.72). The SOI coefficient is $\hat{\beta}_1 = 0.031$ (SE = 0.025, $z = 1.24$). A likelihood ratio test yields $\Lambda = 1.54$ ($p = 0.215$). The contemporaneous effect is **not statistically significant** after accounting for autocorrelation.

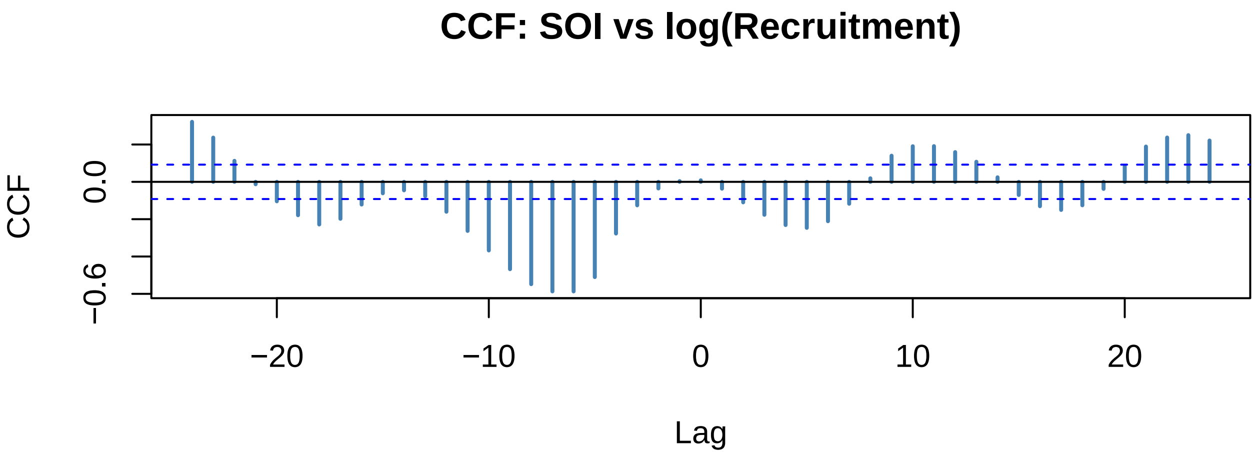

Figure 5: Cross-correlation function between SOI and log recruitment. Negative lags indicate SOI leading recruitment.

The CCF (Figure 5) reveals the strongest association at negative lags of approximately 5–7 months, indicating SOI leads recruitment. This motivates a lagged regression.

5.1 Lagged SOI Regression

Based on the CCF, we fit a regression with SOI_{t-6} as the covariate:

With ARMA(2,0) errors, the lagged SOI coefficient is $\hat{\beta}_1 = 0.016$ (SE = 0.026, $z = 0.61$). The likelihood ratio test gives $\Lambda = 0.37$ ($p = 0.54$). Like the contemporaneous model, the lagged effect is not significant. While the CCF shows a clear bivariate association, the strong AR dynamics in the recruitment series absorb most predictable variation, leaving little for SOI to explain. This is a known limitation of regression with ARMA errors for highly autocorrelated series (Shumway and Stoffer 2017, Ch. 4), and motivates state-space models where SOI can enter as a mechanistic driver.

6 Conclusions

The central Pacific fish recruitment series exhibits strong cyclical behavior with an approximately annual period, detectable through both time- and frequency-domain methods. A SARIMA(1,1,0)(1,1,1) model adequately captures temporal dynamics among differenced candidates, though `auto.arima()` suggests a stationary alternative with lower AIC. The SOI–recruitment relationship illustrates a limitation of regression with ARMA errors: while the CCF reveals a clear 5–7 month lagged association, neither the contemporaneous nor the lagged regression detects a significant effect, because the strong AR dynamics absorb most predictable variation.

These classical methods provide a phenomenological description but do not incorporate biological mechanisms. State-space models—TMB (Kristensen et al. 2016) via Laplace approximation and POMP (King, Nguyen, and Ionides 2016) via sequential Monte Carlo—can incorporate SOI as a mechanistic driver of recruitment, and comparing these approaches is a natural extension for the final project.

Acknowledgments

Analysis was conducted using R (R Core Team 2020) with the `astsa`, `forecast`, and `tseries` packages. Claude (Anthropic) was used for debugging. Course notes (Ionides 2026) and the textbook (Shumway and Stoffer 2017) informed the methodology throughout.

Bibliography

Ionides, Edward. 2026. "Notes for STATS 531, Modeling and Analysis of Time Series Data." https://ionides.github.io/531w26/.

King, Aaron A, Dao Nguyen, and Edward L Ionides. 2016. "Statistical Inference for Partially Observed Markov Processes via the R Package pomp." *Journal of Statistical Software* 69 (12): 1–43.

Kristensen, Kasper, Anders Nielsen, Casper W Berg, Hans Skaug, and Bradley M Bell. 2016. "TMB: Automatic Differentiation and Laplace Approximation." *Journal of Statistical Software* 70 (5): 1–21.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Shumway, Robert H, and David S Stoffer. 2017. *Time Series Analysis and Its Applications: With R Examples*. 4th ed. Springer.

STATS 531 W16 Midterm Project 06. 2016. "Association Between PM2.5 and Temperature Difference." https://ionides.github.io/531w16/midterm_project/project6/midterm_project.html.

STATS 531 W24 Midterm Project 07. 2024. "Precipitation in Detroit." https://ionides.github.io/531w24/midterm_project/.

STATS 531 W25 Midterm Project 08. 2025. "Water Level in Lake Malawi." https://ionides.github.io/531w25/midterm_project/.

7 Supplementary material

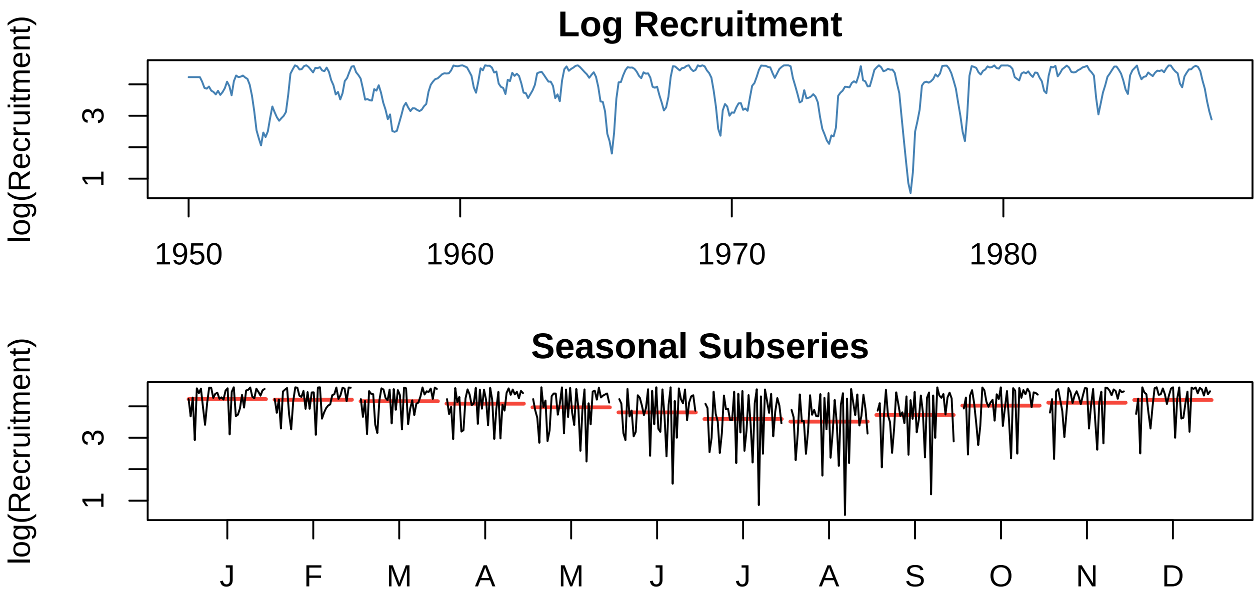

Figure 6: Log-transformed recruitment series (top) and seasonal subseries plot (bottom).

Table 2: SARIMA model comparison by AIC and BIC.

Model	AIC	BIC	logLik
SARIMA(1,1,0)(1,1,1)	-129.94	-113.59	68.97
SARIMA(1,1,1)(1,1,1)	-128.12	-107.68	69.06
SARIMA(2,1,0)(1,1,1)	-128.09	-107.66	69.05
SARIMA(2,1,1)(1,1,1)	-126.93	-102.41	69.47
SARIMA(3,1,1)(1,1,1)	-125.30	-96.69	69.65
SARIMA(1,1,1)(0,1,1)	-125.03	-108.68	66.51
SARIMA(2,1,0)(0,1,1)	-125.02	-108.67	66.51
SARIMA(2,1,1)(0,1,1)	-123.73	-103.30	66.87

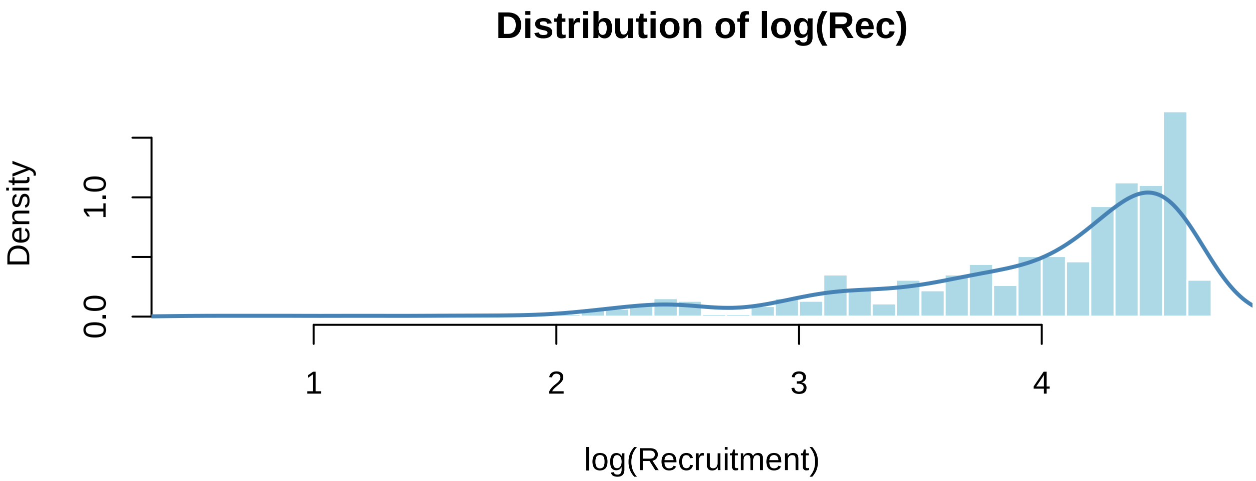

Figure 7: Histogram and density of log-transformed recruitment.

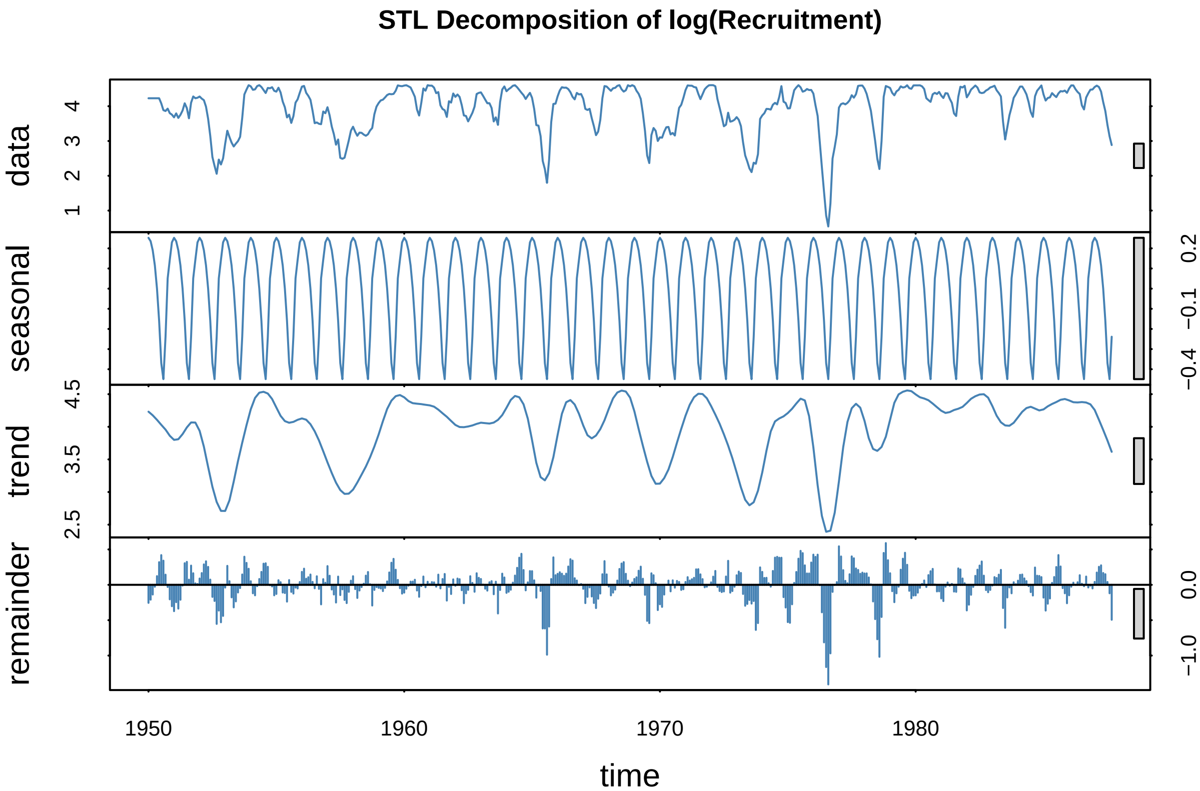

Figure 8: STL decomposition of log recruitment showing trend, seasonal, and remainder components.

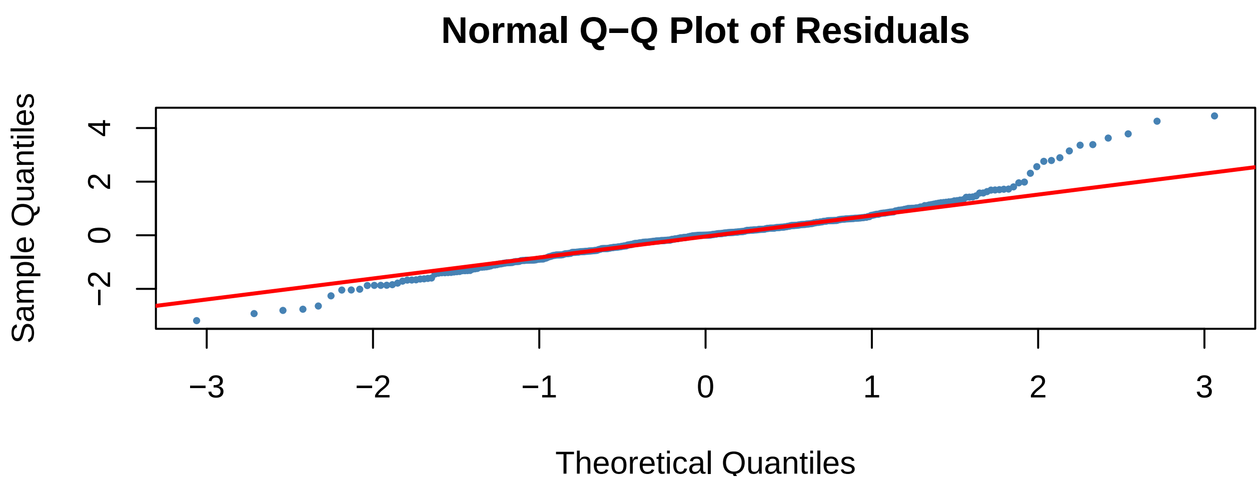

Figure 9: Q-Q plot of standardized residuals from the selected SARIMA model.

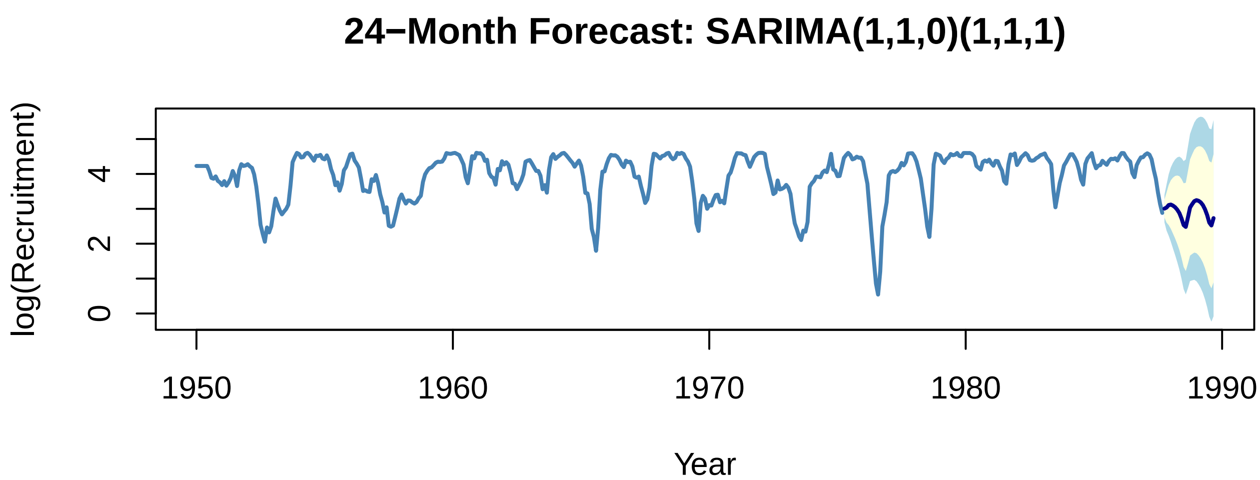

Figure 10: 24-month forecast from the selected SARIMA model with 80% and 95% prediction intervals.

Table 3: Regression with ARMA errors: parameter estimates for contemporaneous SOI model.

Parameter	Estimate	SE	z-value
ar1	1.4193	0.0400	35.48
ar2	-0.5271	0.0402	-13.11
intercept	3.9574	0.0913	43.35
soi_ts	0.0313	0.0253	1.24