

한국 멸치어업의 어획량 분석과 예측 ARIMA 모델 및 스펙트럼 해석

박해훈 · 윤갑동*
동해수산연구소 · 부산수산대학교*

Analysis and Prediction of Anchovy Fisheries in Korea ARIMA Model and Spectrum Analysis

Hae-Hoon PARK and Gab-Dong YOON*

East Sea Fisheries Research Institute · *National Fisheries University of Pusan, Korea

Forecasts of the monthly catches of anchovy in Korea were carried out by the seasonal Autoregressive Integrated Moving Average (ARIMA) model and spectral analysis. The seasonal ARIMA model is as follows:

$$(1 - 0.431B)(1 - B^{12})Z_t = (1 - 0.882B^{12})e_t$$

where: Z_t =value at month t ; B^p is a backward shift operator, that is, $B^p Z_t = Z_{t-p}$; and e_t =error term at month t , which is to forecast 24 months ahead the anchovy catches in Korea. The prediction error by the Box-Cox transformation on monthly anchovy catches in Korea was less than that by the logarithmic transformation. The equation of the Box-Cox transformation was $Y' = (Y^{0.58} - 1)/0.58$. Forecasts of the monthly anchovy catches for 1991~1992, which were compared with the actual catches, had an absolute percentage error (APE) range of 1.0~63.2%. Total observed annual catches in 1991 and 1992 were 170,293 M/T and 168,234 M/T respectively, while the predicted catches were 148,201 M/T and 148,834 M/T (APE 13.0% and 11.5%, respectively). The spectrum analysis of the monthly catches of anchovy showed some dominant fluctuations in the periods of 2.2, 6.1, 10.2, 12.0 and 14.7 months. The spectrum analysis was also useful for selecting the ARIMA model.

Key words : forecast, time series analysis, ARIMA, spectrum analysis, Box-Cox transformation

서 론

최근 치어 등의 남획과 어로기술의 발달을 남용함으로써 어업자원이 감소하는 경향이 있어 이를 보호하기 위해 어업별로 어선 어구의 규모나 수, 어장 및 어기의 제한, 어획물 또는 그물코 크기의 제한, 어획량의 제한 같은 규제를 실시하고 있다. 또한 어장 이용방식을 조정하든가 어가 유지를 위해 어획량을 조절하는 것도 무차별한 어획으로부터 수산자원을 보호하는 한 방법이다. 이와 같이 수산 자원을 적절하게 이용하고 적절한 수요 공급을 위해서는 자원량이나 어획량을 파악하고 예측할 필요가 있다.

어획량을 예측하는 데는 해양학적, 생물학적, 사회 환경 및 경제적인 여러 요인들에 의해 복잡한 영향을 받는다. 생태계 모델이나 다중 모델을 이용하여 정량적으로 예측하는 것은 여러 가지 불확실한 점들 때문에 아직 크게 실용화되지는 못하고 있다 (Yoo and Zhang, 1993; Stergiou, 1990). 수온이나 産卵親魚量 등을 변수로 하여 어획량과의 관계를 중회귀분석이나 주성분분석으로 구하든가 (川上, 1981), 또 수온, 수온 경도, 동물성부유생물량을 변수로 하여 어획량과의 회귀식을 중회귀분석으로 구한 예도 있으며 (국립수산진흥원, 1992; Park and Choi, 1995), 백터 자기회귀를 이용한 방법 등이 있다. 이렇게 하려면 실제로 오

랜 기간 동안의 여러 가지 요소가 포함된 자료가 축적되어 있을수록 좋겠지만 현재 실정은 그렇지 못하다.

이에 비해 시계열분석은 시계열자료인 어획량 자료 이외의 다른 자료가 없더라도 이들 사이의 상관관계나 스펙트럼 분석을 통하여 변동 상태를 파악하거나 예측할 수 있다. ARIMA (Autoregressive Integrated Moving Average) 기법은 과거 경제분야에서 많이 적용되어 왔지만, 최근 수산분야에도 응용되고 있다 (Stergiou, 1990; Yoo and Zhang, 1993; Ohnishi et al., 1995).

멸치어업은 한국 주요 어업의 하나이다. 본 연구에서는 과거 22년간 어획된 멸치의 월별 어획량 자료를 이용하여 첫 20년간의 자료로써 ARIMA 모형을 통해 한국 멸치어업에서의 어획량을 분석하여 적절한 모형을 구하고, 이것을 이후 2년(24개월) 동안의 어획량을 월별로 예측하여 실제 어획량과 비교하며, 스펙트럼 분석으로 순환변동 성분을 구하고자 한다.

자료 및 방법

1. 자료

본 연구에 사용된 자료는 1971~1992년의 월별 멸치 어획량 자료이다 (농림수산통계연보, 1993). 자료의 비정상성 (nonstationary)을 줄이기 위해 차분 (difference)이나 로그 (대수) 변환 (log transformation) 및 멱변환 (power transformation)을 하는데, 여기서 어획량 자료를 각각 로그 (log) 변환과 Box-Cox 변환 시켜 비교 분석하였다.

2. ARIMA 모형

p차 자기회귀(AR: autoregressive)모형은 시점 t에서의 시계열값 Z_t 가 p개의 과거 값들과 오차항 e_t 의 선형결합으로 나타내진다.

$$Z_t = \varphi_1 Z_{t-1} + \varphi_2 Z_{t-2} + \dots + \varphi_p Z_{t-p} + e_t \quad (1)$$

단, e_t 는 평균 0, 분산이 σ^2 인 백색잡음이다. 후진연산자 (backward shift) B를 사용하여 표현하면,

$$(1 - \sum_{k=1}^p \varphi_k B^k) Z_t = e_t \quad (2)$$

$$\varphi(B) Z_t = e_t$$

$$\text{단, } \varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$$

또한, q차의 이동평균(MA: moving average) 모형 Z_t 는 일반 선형모형에서 e_t 의 항이 유한 q개로 증속될 때의 모형을 말한다. 즉,

$$Z_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (3)$$

$$Z_t = \theta(B) e_t \quad (4)$$

$$\text{단, } \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

위의 두 과정 즉 AR(p) 모형과 MA(q) 모형을 하나의 시계열 모형으로 포함시켜 나타낼 때 이를 ARMA(p, q) 모형이라 한다. 즉,

$$Z_t = \varphi_1 Z_{t-1} + \varphi_2 Z_{t-2} + \dots + \varphi_p Z_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (5)$$

이를 간단히 표현하면,

$$\varphi(B) Z_t = \theta(B) e_t \quad (6)$$

또, 시계열이 비정상적(nonstationary)인 경우 각 시점에서 차분을 취하면 정상적(stationary)인 시계열로 바꿀 수 있는데, d차 차분을 다음과 같이 정의하자.

$$W_t = \nabla^d Z_t = (1 - B)^d Z_t \quad (7)$$

차분을 취한 시계열에 ARMA 모형을 적용할 때 이것을 ARIMA (Autoregressive Integrated Moving Average) 모형이라 하고, 시계열 자료에 계절사차 s인 계절성이 있을 때 이를 포함시켜 나타내면, 일반적으로 ARIMA(p, d, q)(P, D, Q)s로 표시하고 다음과 같은 식으로 표현된다.

$$\varphi_p(B) \Phi_p(B^{12}) W_t = \theta_q(B) \Theta_q(B^{12}) e_t \quad (8)$$

$$W_i = \nabla^d \nabla_{12}^D Z_i$$

여기서, W_i 는 계절적, 순차적으로 차분한 것을 나타낸다.

Box-Jenkins에 의한 시계열 모형화 방법은 모형을 식별 (Model identification)하고, 모수를 추정 (Parameter estimation)한 후 적합성을 진단 (Diagnosis checking)하는 것이다. 시계열의 자기상관함수 (ACF: autocorrelation function)와 부분자기상관함수 (PACF: partial autocorrelation function)를 분석하여 MA, AR의 차수를 정하고, 시계열에 비정상성이 있으면 계절적, 비계절적 차분을 하여 정상 시계열로 만든 후 분석, 예측한다.

3. 스펙트럼 분석

시계열 속에 포함되어 있는 순환변동은 자기상관함수 (ACF)나 부분자기상관함수 (PACF)로는 파악하기 힘든 반면, 푸리에 변환을 이용하면 쉽게 파악할 수 있다. 시계열 함수 $f(t)$ 의 푸리에 변환은 다음과 같이 정의된다.

$$F(W) = \int_{-\infty}^{\infty} f(t) \exp(-j\omega t) dt \quad (9)$$

위 식을 월별 통계량과 같은 이산자료를 취급하기 위하여, 위 식은 자료의 총 갯수가 N 일 때 다음과 같은 이산형태로 바꾸어진다 (황 · 이, 1989).

$$F(k) = \sum_{n=0}^{N-1} f(n) \exp(-j2\pi kn/N) \quad (10)$$

월별 자료인 경우 샘플링 간격은 1개월이 된다. 평균이 있는 경우에 자료로부터 평균을 빼고 분석하는 것은 주파수 성분이 0인 성분을 빼고 분석하는 것과 같다. 그리고, 자료의 주파수 성분을 구할 때 완전한 주기의 자료를 취급하느냐 아니냐에 따라 주파수 성분에 차이가 나므로, 여기선 각 연도별로 1월에서 12월까지의 자료를 취급한다 (Otnes and Enochson, 1978).

결과 및 고찰

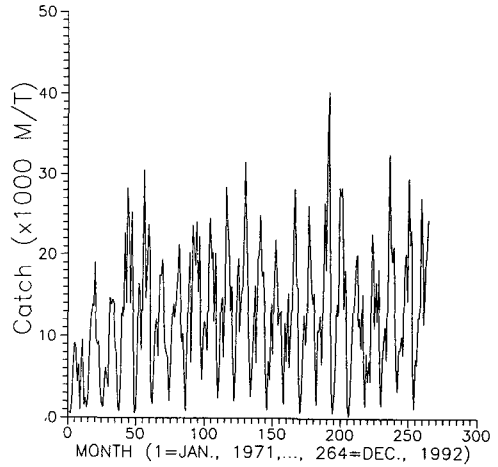


Fig. 1. Monthly anchovy catches in Korea, 1971 to 1992.

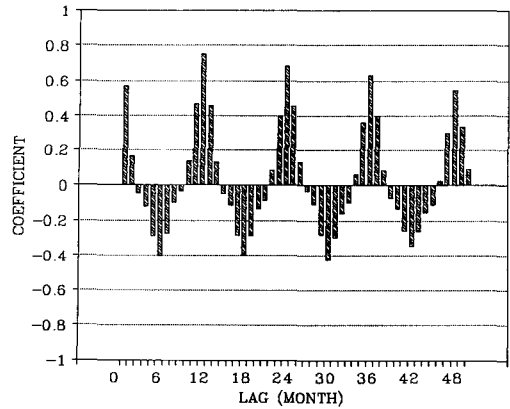


Fig. 2. Estimated autocorrelations (ACF) to lag 50 for the log-transformed anchovy catches.

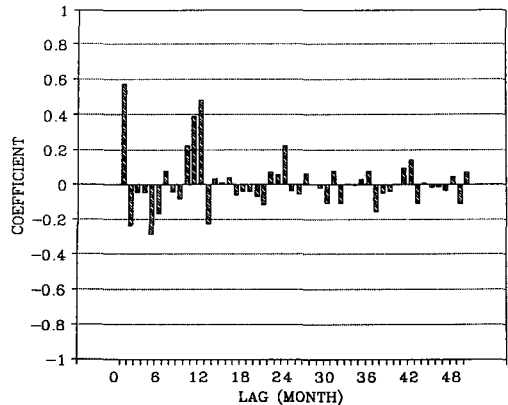


Fig. 3. Estimated partial autocorrelations (PACF) to lag 50 for the log-transformed anchovy catches.

Fig. 1은 1971~1992년까지의 한국 멸치어업의 월별 어획량 자료이다.

Fig. 1에서 보는 바와 같은 자료의 비정상성 분산을 줄이기 위해 로그(대수) 변환을 하여 구한 ACF와 PACF를 Fig. 2 및 Fig. 3에 각각 나타내었다.

Fig. 2와 Fig. 3에서 보는 바와 같이 계절성과 비정상성이 나타나므로, 비계절적 차감과 계절적 차감을 하여 구한 ACF와 PACF는 Fig. 4와 Fig. 5와 같다.

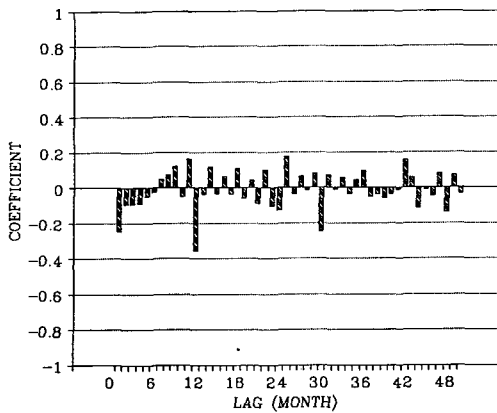


Fig. 4. Estimated ACF after nonseasonal first differencing and seasonal first differencing for the log-transformed anchovy catches.

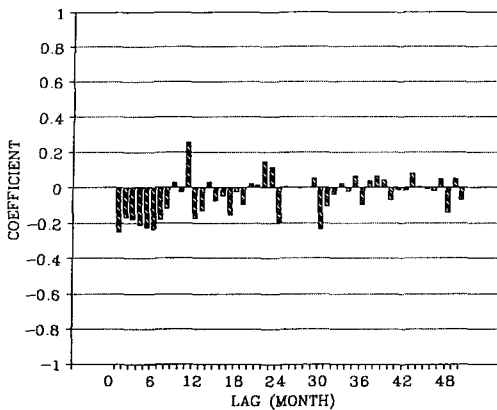


Fig. 5. Estimated PACF after nonseasonal first differencing and seasonal first differencing for the log-transformed anchovy catches.

위의 그림들로부터 다음과 같은 모형을 정하였고,

$$\text{ARIMA}(1, 1, 1)(2, 1, 1)_{12}$$

이 모형에 대해 잔차의 ACF 및 PACF는 오차한계를 넘는 부분이 있었으나, 정규성 검정이나 수정된(Modified) Box-Pierce 통계량과 RUN 검정은 모두 이 모형의 적합함을 보여 주었으므로 평균 0인 백색잡음이라 볼 수 있다(김, 1993). 과다적합화 분석을 해본 결과 다른 모형에 비해 이 모형이 더 적절함을 보여주므로, 추정된 모수를 ARIMA 모형에 대입시키면 다음과 같다.

$$(1-0.381B)(1-0.027B^{12}+0.141B^{24})(1-B^1)(1-B^{12})Z_t = (1-0.968B)(1-0.727B^{12})e_t \quad (11)$$

한편, 자료를 분석하는 데 있어서 일반적으로 로그(대수) 변환을 많이 하지만, 이것은 Box-Cox 변환의 특수한 경우이므로, 한국 멸치어업의 어획량 자료에 대해 Box-Cox 변환식 (12)을 적용하여 가장 적절한 변환을 구한 결과는 식 (13)과 같았다(Sokal and Rohlf, 1981).

$$Y' = (Y^a - 1)/a, \quad a \neq 0 \quad (12)$$

$$= \ln Y, \quad a = 0$$

$$Y' = (Y^{0.58} - 1)/0.58 \quad (13)$$

Box-Cox 변환을 하여 구한 멸치 어획량 자료의 ACF에 계절성이 나타나므로(여기 표현하지는 않음) 계절차분하여 구한 ACF와 PACF를 Fig. 6과 Fig. 7에 나타내었다.

앞과 같이 모형을 정하고, 추정된 모수를 대입하면 식 (14)와 같다.

$$(1-0.431B)(1-B^{12})Z_t = (1-0.882B^{12})e_t \quad (14)$$

식 (14)의 모형에 대한 잔차의 정규성 검정은 잔차들과 정규값 간의 상관계수는 0.994였으며, 이것은 5% 유의수준에서 정규성이 채택되며, 잔차의 ACF와 PACF는 오차한계를 넘는 부분이 있었으나(김, 1993) 거의 백색잡음임을 나타내었고, 수정된(Modified) Box-Pierce 검정은 Lag 12에서 χ^2 값은 13.8(DF=10), Lag 24에서 χ^2 값은 29.3(DF=22)였으므로 유의수준 5%에서 이 모형은 기각되지 않으며, RUN 검정은 런

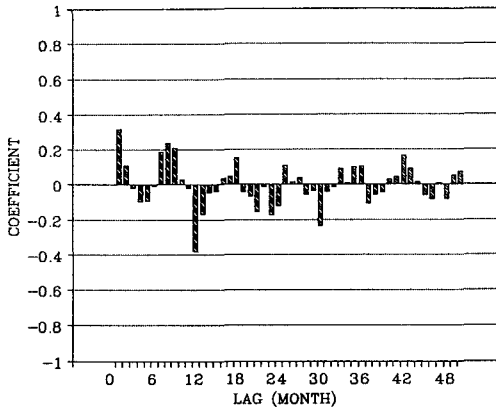


Fig. 6. Estimated ACF after seasonal first differencing for the Box-Cox transformed anchovy catches.

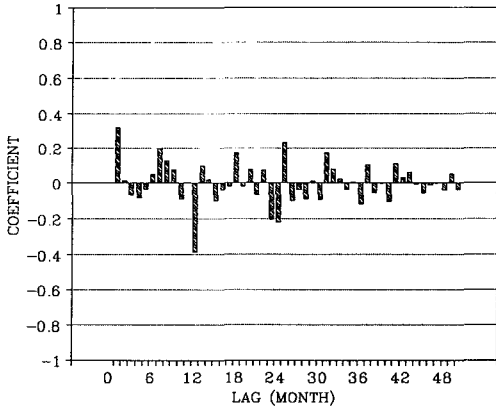


Fig. 7. Estimated PACF after seasonal first differencing for the Box-Cox transformed anchovy catches.

의 갯수가 128이었고 기대치는 114.57이었으며, 이것은 유의수준 5%에서 독립성이 있다. 따라서, 잔차는 평균 0인 백색잡음이라 볼 수 있다.

다음은 스펙트럼 성분을 살펴보자. 원자료인 Fig. 1을 보면 전반적인 추세는 보이지 않으므로, 어획량 자료의 평균값을 제거한 후 Fourier 변환한 것의 Real 성분이 Fig. 8에 나타나 있다.

Real 성분인 진폭성분이 중앙점을 기준으로 대칭되어 있으므로, 자료 해석은 총갯수가 N일 때 N/2 점까지만으로 해석이 가능하다.

Fig. 8에서 보면, 순환변동 성분은 (1)의 2.2개월, (2)의 6.1개월, (3)의 10.2개월, (4)의 12개월, (5)의 14.7개월 성분에서 상대적으로 큰 값을 나타내고 있

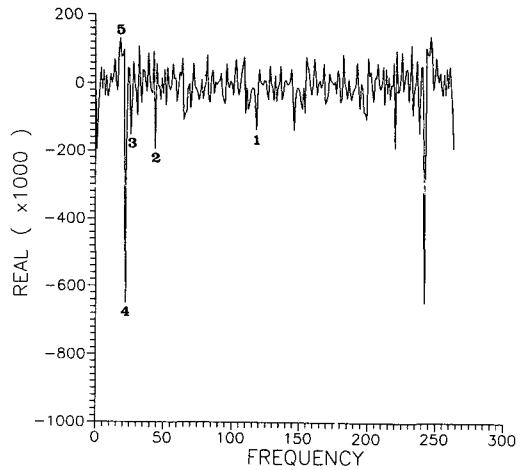


Fig. 8. Fourier transform of anchovy catch after subtracting mean value of it: Real part.

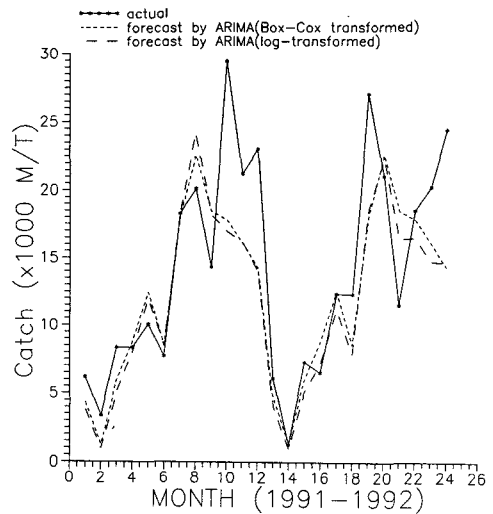


Fig. 9. Comparison between the actual catches and the forecasts by the ARIMA models.

다. 이 순환변동 성분을 고려하면, 위의 두 모형중 후자의 모형이 더욱 적절함을 알 수 있다.

로그 변환 자료에 대한 식 (11)과 Box-Cox 변환시킨 자료에 대한 식 (14)를 이용하여 1991년부터 1992년까지 2년간(24개월)을 예측한 어획량과 실제 어획량을 나타낸 것이 Fig. 9와 Table 1에 있다.

실제어획량과 예측어획량과의 오차(APE: absolute percentage error)를 다음 식으로 표시하면 (Stergiou, 1990),

Table 1. Comparison of errors between the actual catches and the forecasts by the ARIMA models on the log-transformed and the Box-Cox transformed anchovy catches

	Actual landing	Forecast by ARIMA on log-transformed data	Forecast by ARIMA on Box-Cox transformed data
Range of error by month		0.3~71.9%	1.0~63.2%
Error by year	1991	170,293(M/T) (14.3%)	148,201(M/T) (13.0%)
	1992	168,234(M/T) (17.4%)	148,834(M/T) (11.5%)
	total	338,527(M/T) (15.9%)	297,035(M/T) (12.3%)

$$APE = |X_t - F_t| \times 100 / X_t$$

단, X_t 는 시점 t 에서의 실제 어획량이고, F_t 는 예측값이다.

Fig. 9와 Table 1에서 보면, 로그(log) 변환과 Box-Cox 변환 중 후자의 편이 오차가 더 적은 결과를 나타내고 있다. Box-Cox 변환을 시켜 멀치 어획량을 예측했을 때 1991년의 어획량의 오차(APE) 범위는 1.6~63.2%이고, 1992년에는 1.0~60.4%였다. 2년 동안의 실제 어획량은 1991년에 170,293M/T, 1992년에는 168,234M/T인데, 예측값은 각각 148,201M/T, 148,834M/T이며(오차는 1991년과 1992년에 각각 13.0%, 11.5%), 2년간의 총 어획량에 대한 오차는 12.3%였다. 여기서 전체 어획량에 관하여 분석하였지만, 각 해역별로의 자료가 있으면 이것과 해당 해역의 해어황상 태나 사회, 환경 정보 등을 고려함으로써 그 해역에 대해 보다 적절한 분석과 예측이 가능할 것이다(小倉·竹内, 1990).

요 약

우리나라 멀치어업에서의 1971~1992년 동안의 22년간 월별 어획량 자료를 시계열 분석하여 어획량을 분석, 예측하였다. 시계열 분석은 다른 생물학적, 해양학적, 사회 경제적인 요소가 없어도 단지 어획량 자료만으로 분석과 예측이 가능하다.

첫 20년간인 1971~1990년 사이의 월별 멀치 어획

량 자료를 ARIMA 시계열 모형에 적용시켜 구한 결과는 다음과 같다.

로그(대수) 변환시켰을 때의 ARIMA 모형:

$$(1-0.381B)(1-0.027B^{12}+0.141B^{24})(1-B^1)(1-B^{12})Z_t = (1-0.968B)(1-0.727B^{12})e_t$$

Box-Cox 변환시켰을 때의 ARIMA 모형:

$$(1-0.431B)(1-B^{12})Z_t = (1-0.882B^{12})e_t$$

위의 두 모형중 Box-Cox 변환시킨 것이 로그(대수) 변환시킨 것보다 예측오차가 적었으며, Box-Cox 변환식은 $Y' = (Y^{0.58} - 1) / 0.58$ 이었다. 위의 두 모형 중 후자의 모형을 이용하여 1991~1992년 사이의 월별 어획량을 예측하였다. 예측 어획량과 실제 어획량과의 월별 오차범위는 1.0~63.2% (1991년에 1.6~63.2%이고, 1992년에는 1.0~60.4%)였다. 예측 어획량이 각 연도별로 148,201 M/T과 148,834 M/T인데 비해, 실제 어획량은 170,293 M/T, 168,234 M/T이었다. 2년 동안의 총어획량에 대한 오차는 12.3%였다. 또한 스펙트럼 분석은 순환변동의 주기가 2.2개월, 6.1개월, 10.2개월, 12개월, 14.7개월에서 상대적으로 큰 성분이 있음을 나타내었다. 이 순환변동 성분은 적절한 ARIMA 모형을 결정하는 데도 도움이 된다.

사 의

유익한 조언과 도움 자료를 주신 익명의 심사위원
들께 감사드립니다.

참 고 문 헌

- Ohnishi, S., Y. Matsumiya, M. Ishiguro and K. Sakuramoto. 1995. Construction of time series analysis model effective for forecast of fishing and oceanographic conditions. *Fisheries Science*, 61 (4), 550~554.
- Otnes, R.K. and L. Enochson. 1978. *Applied time series analysis*. Wiley-Interscience, 449 pp
- Park, J.H. and K.H. Choi. 1995. A study on the formation of fishing ground and the prediction of fishing conditions of mackerel, *Scomber Japonicus* Houttuyn. *Bull. Nat. Fish. Res. Dev. Agency*, 49, 25~35 (in Korean).
- Sokal, R.R. and F.J. Rohlf. 1981. *Biometry*. W.H. Freeman and Company, New York, 859 pp
- Stergiou, K.I. 1990. Prediction of the Mullidae fishery in the eastern Mediterranean 24 months in advance. *Fisheries Research*, 9, 67~74.
- Yoo, S. and C.I. Zhang, 1993: Forecasting of hairtail (*Trichiurus lepturus*) landings in Korean Waters by time series analysis. *Bull. Korean Fish. Soc.*, 26(4), 363~368 (in Korean).
- 국립수산진흥원. 1992. 원격탐사에 의한 해어황 예보 기술개발 (III). 과학기술처. 227 pp
- 김원경. 1993. 시계열분석. 경문사, 296 pp
- 小倉通男·竹内正一. 1990. 漁業情報學概論. 成山堂, 190 pp
- 川上太左英. 1981. 漁業解析入門. 恒星社厚生閣, 294 pp
- 황병원·이주신. 1989. 디지털 신호처리. Ohm사, 서울, 335 pp

1995년 11월 20일 접수

1996년 1월 6일 수리