

시계열분석을 이용한 한국 명태어업의 어획량 예측 : AIC

박 해 훈 · 윤 갑 동*

동해수산연구소, 부경대학교*

(1996년 5월 25일 접수)

Prediction of Walleye Pollock, *Theragra chalcogramma*, Landings in Korea by Time Series Analysis : AIC

Hae - Hoon Park and Gab - Dong Yoon*

East Sea Fisheries Research Institute, *Pukyong National University

(Received May 25, 1996)

Abstract

Forecasts of monthly landings of walleye pollock, *Theragra chalcogramma*, in Korea were carried out by the seasonal Autoregressive Integrated Moving Average (ARIMA) model. The Box - Cox transformation on the walleye pollock catch data handles nonstationary variance. The equation of Box - Cox transformation was $Y' = (Y^{0.31} - 1)/0.31$. The model identification was determined by minimum AIC (Akaike Information Criteria). And the seasonal ARIMA model is presented

$$(1 - 0.583B)(1 - B^1)(1 - B^{12})Z_t = (1 - 0.912B)(1 - 0.732B^{12})e_t$$

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 walleye pollock landings in Korea. Monthly forecasts of the walleye pollock landings for 1993~1994, which were compared with the actual landings, had an absolute percentage error (APE) range of 20.2~226.1%. Total observed annual landings in 1993 and 1994 were 16,610M/T and 10,748M/T respectively, while the model predicted 10,748M/T and 8,203M/T (APE 37.0% and 23.7%, respectively).

서 론

생태계 파괴가 일어나면 이를 회복하는데 드는 시간은 파괴에 소요된 시간보다 더 걸린다는 것이 알려져 있다. 따라서, 근래의 무분별한 치어 등의 남획과 어로기술 발달의 남용 및 자원량을 고려하지 않고 과다어획하므로써 어업자원이 감소하고 있어 이를 보호하기 위한 방안이 요구되고 있다.

과다어획은 어가하락을 일으키기도 하는 반면에, 적정어획은 적절한 소득을 올릴 수 있을 뿐만 아니라 어업을 지속시킬 수 있기 때문이다. 이를 위해 어선 및 어구의 규모나 수, 어장 및 어기의 제한, 어획물의 종류 또는 그물코 크기의 제한, 어획량의 제한 같은 규제를 실시하고 있다. 또, 어장의 이용 방식을 조정하든가, 어가 유지를 위해 어획량을 조절하는 것도 자원 보호의 한 방편이 된다. 이렇게

하기 위해서는 먼저 자원량이나 어획량을 파악하고 예측할 필요가 있는데, 여기서는 어획량의 예측을 다루고자 한다.

어획량의 예측은 해양학적, 생물학적, 사회 환경 및 경제적인 여러 요인들에 의해 복잡한 영향을 받는다. 수온이나 産卵親魚量 등을 변수로 하여 어획량과의 관계를 중회귀분석이나 주성분분석으로 구한 예나(川上, 1981), 또 수온, 수온경도, 동물성 부유생물량을 변수로 하여 중회귀분석으로 구한 예가 있다(국립수산진흥원, 1992). 이러한 방법에는 여러가지 축적된 자료가 많을수록 좋지만, 아직까지 얼마 축적되지 않은 실정이다.

이에 비해 시계열분석은 시계열자료 이외에 다른 자료가 없어도 이들 사이의 상관관계나 스펙트럼 분석을 통하여 변동 상태를 파악하고 예측할 수 있다. ARIMA(Autoregressive Integrated Moving Average) 기법은 최근 수산분야에도 응용되고 있는데(Stergiou, 1990 ; 유와 장, 1993 ; Ohnishi *et al.*, 1995; 박과 윤, 1996), 이것은 먼저 모형을 식별하고 모수를 추정한 후 적합성을 진단하는 것이다. 많이 쓰이고 있는 Box-Jenkins에 의한 시계열 모형화 방법은 AR 및 MA 모형에는 쉽게 적용되나, 이의 큰 약점은 ARMA모형을 식별하지 못한다는 것이다. 따라서, ARMA 모형의 식별과 적합성 검토에 많은 반복적인 작업이 따를 수 있으며, 주관적인 요소가 들어가는 단점이 있으므로, 여기서는 객관적이고 능률적인 식별방법으로 최소 AIC(Akaike Information Criteria)를 이용하고자 한다.

명태 어업은 한국 주요 어업의 하나이다. 본 연구에서는 과거 24년(1971-1994)간 어획된 명태의 월별 어획량 자료를 이용하여 첫 22년간의 자료로써 AIC를 이용하여 적절한 ARIMA 모형을 구하고, 이것을 이후 2년(24개월) 동안에 대해 월별 어획량을 예측하여 실제 어획량과 비교하고자 한다.

자료 및 방법

1. 자료

본 연구에 사용된 자료는 1971-1994 년간 명

태어업의 월별 어획량 자료이다(농림수산통계연보, 1995). 자료의 비정상성을 줄이기 위해 어획량 자료에 Box-Cox변환을 시켜 분석하였다. 왜냐하면, Box-Cox 변환은 일반적으로 쓰고 있는 로그(대수)변환을 포함하고 있으므로 모형선택과 예측 오차에 있어서 보다 작은 오차를 줄 수 있기 때문이다(박과 윤, 1996).

2. ARIMA 모형과 AIC

시점 t 에서의 시계열값 Z_t 가 p 개의 과거값들의 선형결합으로 나타내지는 p 차 자기회귀(AR : autoregressive)모형과 오차항 e_t 가 유한 q 개로 증속될 때의 q 차의 이동평균(MA : moving average)모형을 하나의 시계열 모형으로 포함시켜 나타낼 때 이를 ARMA(p, q) 모형이라 한다(유와 장, 1993 ; 박과 윤, 1996). 즉,

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (1)$$

이를 간단히 표현하면 다음과 같다.

$$\phi(B) Z_t = \theta(B) e_t \quad (2)$$

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

$$W_t = \nabla^d Z_t = (1 - B)^d Z_t \quad (3)$$

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

$$\phi_p(B)\Phi_P(B^{12})W_t = \theta_q(B)\Theta_Q(B^{12})e_t \quad (4)$$

여기서, W_t 는 계절적, 순차적 차분한 것을 나타낸다.

시계열 모형화 방법은 Fig. 1에서와 같이 모형을 식별(Model identification)하고, 모수를 추정(Parameter estimation)한 후 적합성 진단(Dia-

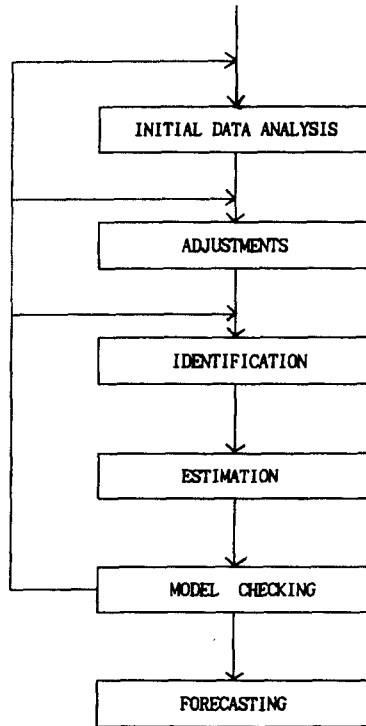


Fig. 1. A general paradigm for univariate time series modelling.

gnosis checking)을 반복하는 것이다. 시계열에 비정상성이 있으면 계절적, 비계절적 차분을 하여 정상 시계열로 만든 후 분석, 예측한다. 모형식별에 있어 Box-Jenkins에 의한 방법은 ARMA모형의 식별에는 어려움이 따르고 주관적인 요소가 들어가는 단점이 있으므로, 여기서는 객관적인 식별방법으로 최소 AIC(minimum Akaike Information Criteria)를 이용하고자 한다. AIC는 다음과 같이 정의된다.

$$AIC(p, q) = \ln[\sigma^2(p, q)] + 2(p+q)/n \quad (5)$$

단, $\sigma^2(p, q)$ 은 모형에 적합시켰을 때 백색잡음의 분산추정값이다.

결과 및 고찰

Fig. 2는 1971~1994년까지의 한국 명태어업의 월별 어획량이다.

위 자료에 대해 가장 적절한 Box-Cox 변환식

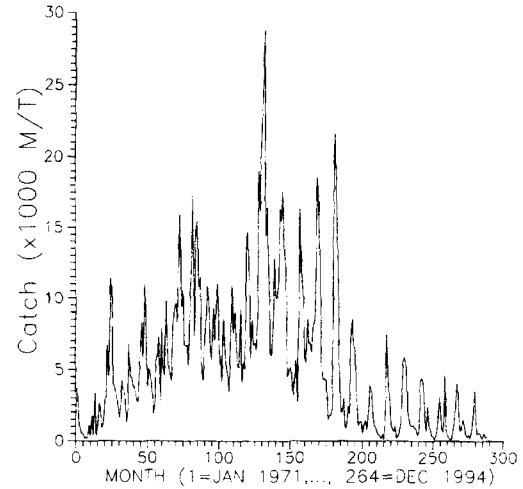


Fig. 2. Monthly walleye pollock landings in Korea, from 1971 to 1994.

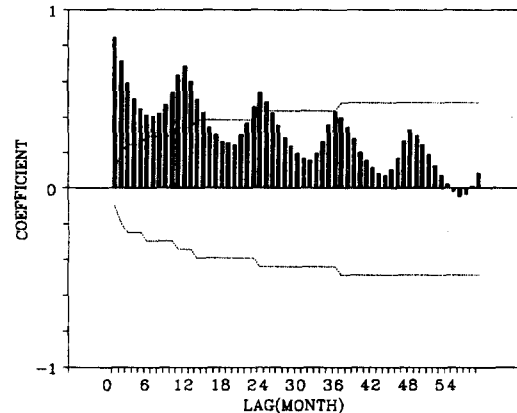


Fig. 3. Estimated ACF to lag 60 for the Box-Cox transformed walleye pollock landings.

을 구하면 식 (6)과 같다(Sokal and Rohlf, 1981).

$$Y' = (Y^{0.31} - 1)/0.31 \quad (6)$$

단, Y는 원자료값이다.

Box-Cox 변환시킨 자료의 ACF(autocorrelation function)와 PACF(partial autocorrelation function)가 Fig. 3과 Fig. 4에 나타나 있는데, 계절성과 비정상성이 나타남을 알 수 있다.

비계절적 차감과 계절적 차감을 하여 구한 ACF와 PACF는 Fig. 5 및 Fig. 6과 같다.

위의 그림들로부터 $P=0, Q=1$ 라 두고, p와 q를

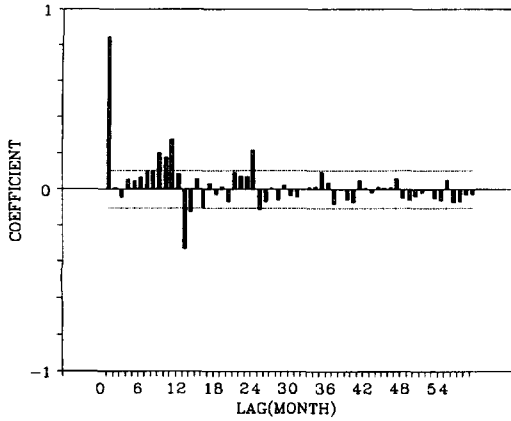


Fig. 4. Estimated PACF to lag 60 for the Box - Cox transformed walleye pollock landings.

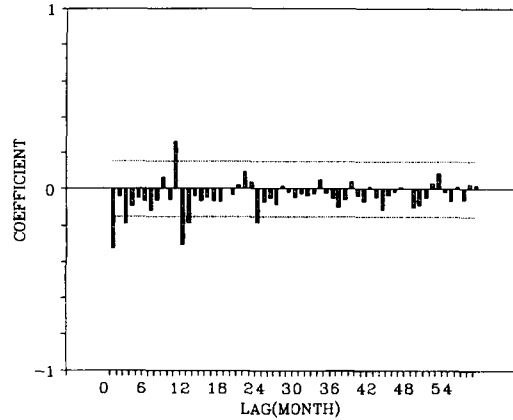


Fig. 6. Estimated PACF after nonseasonal first differencing and seasonal first differencing for the Box - Cox transformed pollock landings.

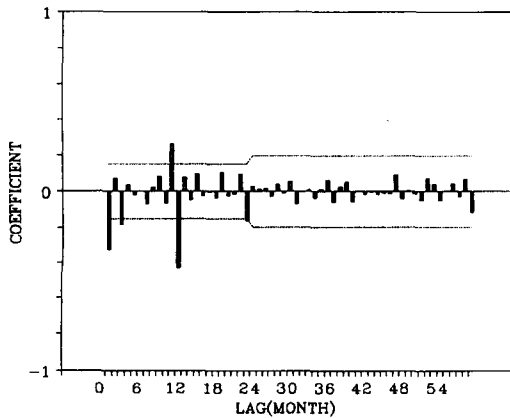


Fig. 5. Estimated ACF after nonseasonal first differencing and seasonal first differencing for the Box - Cox transformed walleye pollock landings.

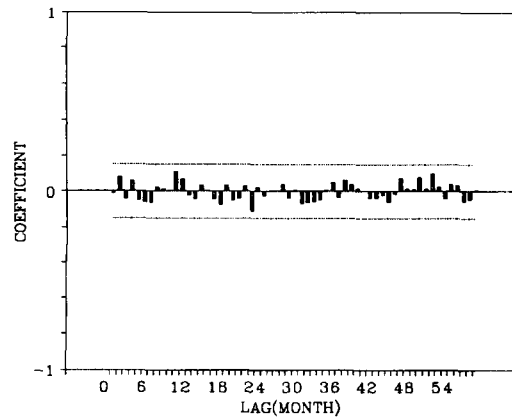


Fig. 7. Estimated ACF of residuals when fitted to the ARIMA model.

각각 0에서 4까지 변화시켜 가면서 구한 AIC값에 자기회귀차수의 과대경향을 검토하여 선택한 ARIMA 모형에 추정된 모수를 대입하면 식 (7)과 같다.

$$(1 - 0.583B)(1 - B^1)(1 - B^{12})Z_t = (1 - 0.912B)(1 - 0.732B^{12})e_t \quad (7)$$

이 모형에 대한 잔차의 ACF 및 PACF는 Fig. 7 및 Fig. 8과 같다.

Fig. 7와 Fig. 8에서 보면 시차 23에서 95%의 오차한계를 넘는 부분이 있으나 거의 백색잡음임을 나타내고 있다. 잔차의 정규성 검정은 잔차들과 정

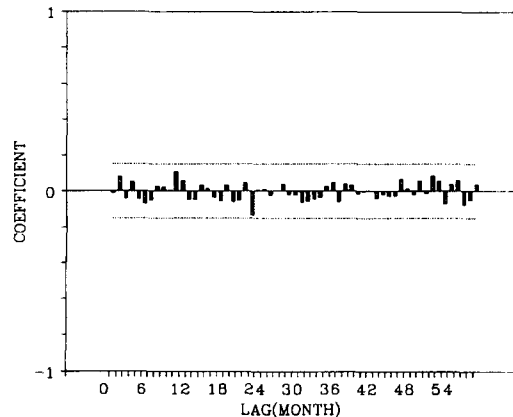


Fig. 8. Estimated PACF of residuals when fitted to the ARIMA model.

규값 간의 상관계수는 0.993이었으며, 이것은 5% 유의수준에서 정규성이 채택되며, 수정된(Modified) Box - Pierce 검정은 Lag 12에서 χ^2 값은 9.8(DF=9), Lag 24에서 χ^2 값은 17.7(DF=21)이었으므로 유의수준 5%에서 이 모형은 기각되지 않으며, RUN 검정은 런의 갯수가 122이었고 기대치는 125.9이었으며, 이것은 유의수준 5%에서 독립성이 있다. 따라서, 잔차는 평균 0인 백색잡음이라 볼 수 있다.

식 (7)로 이후의 1993년부터 1994년까지 2년간(24개월)을 예측한 어획량과 실제 어획량을 나타낸 것이 Table 1과 Fig. 9에 나타나 있다.

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

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

이다. 단, X_t 는 시점 t에서의 실제 어획량이고, F_t 는 예측값이다.

1993년의 어획량의 월별 오차범위는 20.2~226.1%이고, 1994년에는 22.9~208.8%였다. 2년간의 실제 어획량은 1993년에 16,610M/T, 1994년에는 10,748M/T인데, 예측값은 각각 10,459M/T, 8,203M/T이며(오차는 각각 37.0%, 23.7%), 총어획량에 대한 오차는 31.7%였다. 이 값은 그리이스의 mullidae(Stergiou, 1990)나 한국의 갈치(유와장, 1993)의 경우보다 오차가 큰데, 이것은 원자료의 특성과 또한 장기간의 자료를 사용한 때문이라 여겨진다. 한편, Fig. 9를 보면 예측어획량과 실제어획량 사이에 전반적으로 한 시차의 지연이 있다. 이것은 이전의 어업과 최근의 어업시기에 1개월간의 시차가 있음을 의미하는데, 생물학적(플랑크톤 등)요인이나 해황 등의 비생물학적 요인에 의한 어기의 변동일 수도 있으나 앞으로 규명되어야 할 문제이다.

Table 1. Comparison of errors between the actual landings and the forecasts of walleye pollock by the ARIMA model on the Box - Cox transformed data.

		Actual landing	Forecast by ARIMA on Box - Cox transformed data
Range of error by month		20.2 - 226.1%	
Error by year	1993	16,610(M/T)	10,459(M/T)(27.0%)
	1994	10,748(M/T)	8,203(M/T)(23.7)
	total	27,358(M/T)	18,662(M/T)(31.8)

요 약

한국 명태어업에서의 24년(1971 - 1994)간의 월별 어획량 자료에 시계열 분석을 이용하여 어획량을 분석, 예측하였다. 시계열 분석은 다른 생물학적, 해양학적, 사회 경제적인 요소가 없어도 단지 어획량 자료만으로 분석과 예측이 가능하다. 1971~1992년 사이의 월별 명태 어획량의 자료에 대한 적절한 Box - Cox 변환식은 $Y' = (Y^{0.31} - 1)/0.31$ 이었고, AIC(Akaike Information Criteria)를 이용하여 구한 계절 ARIMA 모형은 다음과 같다.

$$(1 - 0.583B)(1 - B^1)(1 - B^{12})Z_t = (1 - 0.912B)(1 - 0.732B^{12})e_t$$

이 모형으로 이후의 2년간(1993년 - 1994년)에 대해 명태의 월별 어획량(24개월)을 예측하였다. 예측 어획량과 실제 어획량과의 월별 오차범위는 20.2~226.1%였다. 실제어획량이 1993년과 1994년에 각각 16,610M/T, 10,748M/T이고, 예측어획

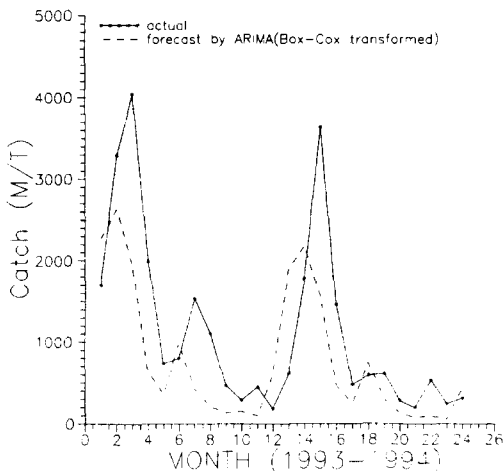


Fig. 9. Comparison between the actual landings and the forecasts by the ARIMA model.

량은 10,459M/T과 8,203M/T이었다(1993년의 오차는 37.0%, 1994년에는 23.7%).

사 의

본 논문을 작성함에 있어 도움 말씀 주신 동해수산연구소 김용문 소장님과 허영희 연구사님께 감사드립니다.

참고문헌

국립수산진흥원(1992) : 원격탐사에 의한 해어황 예보 기술개발(Ⅲ). 과학기술처. 165~226.
박해훈, 윤갑동(1996) : 한국 멸치어업의 어획량 분석과 예측 - ARIMA 모델 및 스펙트럼 해석 -. 한국수산

학회지 29(2), 143~149.
유신재, 장창익(1993) : 시계열 분석에 의한 어획량 예측. - 한국 근해산 갈치를 예로 하여 -. 한국수산학회지 26(4), 363~368.
川上太左英(1981) : 漁業解析入門. 恒星社厚生閣, 294.
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.
Sokal, R.R. and F.J. Rohlf(1981) : Biometry. W.H. Freeman and Company, New York, 859.
Stergiou, K.I.(1990) : Prediction of the Mullidae fishery in the eastern Mediterranean 24 months in advance. Fisheries Research 9, 67~74.