Article

시계열모형을 이용한 굴 생산량 예측 가능성에 관한 연구

남종오 · 노승국*

한국해양수산개발원 수산업관측센터
(121-915) 서울시 마포구 매봉산로 45

A Study on Forecast of Oyster Production using Time Series Models

Jong-Oh Nam and Seung-Guk Noh *

Fisheries Outlook Center, Korea Maritime Institute
Seoul 121-915, Korea

Abstract : This paper focused on forecasting a short-term production of oysters, which have been farmed in Korea, with distinct periodicity of production by year, and different production level by month. To forecast a short-term oyster production, this paper uses monthly data (260 observations) from January 1990 to August 2011, and also adopts several econometrics methods, such as Multiple Regression Analysis Model (MRAM), Seasonal Autoregressive Integrated Moving Average (SARIMA) Model, and Vector Error Correction Model (VECM). As a result, first, the amount of short-term oyster production forecasted by the multiple regression analysis model was 1,337 ton with prediction error of 246 ton. Secondly, the amount of oyster production of the SARIMA I and II models was forecasted as 12,423 ton and 12,442 ton with prediction error of 11,404 ton and 11,423 ton, respectively. Thirdly, the amount of oyster production based on the VECM was estimated as 10,425 ton with prediction errors of 9,406 ton. In conclusion, based on Theil inequality coefficient criterion, short-term prediction of oyster by the VECM exhibited a better fit than ones by the SARIMA I and II models and Multiple Regression Analysis Model.

Key words : oyster, forecast, multiple regression analysis model, seasonal autoregressive integrated moving average model, vector error correction model

1. 서 론

우리나라 천해양식 수산물의 대표품목 중에 하나인 굴은 다른 어떤 양식 품목보다도 양식의 역사가 깊다. 또한 생산량 규모도 세계적 수준과 비교해서 결코 뒤처지지 않는다. 2010년 기준 우리나라 천해양식 굴 생산량은 29만여 톤으로 중국 다음으로 많은 세계 2위 수준이다. 하지만 이러한 생산 규모 및 양식의 오랜 역사에도 불구하고, 그 양을 예측할 수 있는 모형은 아직 개발되지 못한 실정이다. 이에 본 연구는 다양한 시계열분석 기법을 활용하여

우리나라의 천해양식 굴 생산 함수를 추정하고 이를 바탕으로 굴 생산량 예측을 시도해 보고자 한다.

사실 시계열 자료를 이용하여 특정 종속변수의 향후 변화를 예측하는 이론들은 거시·금융 분야에서 활발히 논의되어왔다. 국내 농수산 분야에서는 이와 관련된 연구가 다소 미진한 편인데, 이용 가능한 자료의 한계와 생물의 특성에 따른 예측의 어려움이 원인이었다. 국내 농업 분야의 생산량과 관련한 예측모형의 경우, 1990년대 각종 생육모형을 이용한 예측이 주를 이루었다. 이후 강과 김(2000)이 다변량통계기법과 지역별 회귀모형을 이용하여 쌀 예상 생산량을 추정하였다. 이와 신(2000)은 조건부 자기회귀모형을 설정하여 송이버섯의 생산량을 예측하였

*Corresponding author. E-mail : bluesyno@kmi.re.kr

다. 장 (2000)은 기상 요인을 고려한 주요 식량작물에 대한 생산량 예측모형을 제시하였으며, 고 (2003)도 감자의 생산형태별로 생산량을 예측할 수 있는 모형을 개발하였다. 또한 최근에는 김 등 (2010)이 중기선행관측을 위한 농축산물의 작형별 수급모형 및 예측평가시스템을 개발하기도 하였다.

한편, 우리나라 수산물의 생산량 예측과 관련된 논문으로는 박과 이 (1997)가 우리나라 연근해산 수산물에 대한 생산량 예측모형을 제시하였다. 그리고 김 (2000)은 우리나라 전체 수산물 및 부류별(어류, 패류)·어종별(갈치, 명태, 멸치, 고등어, 조기, 오징어) 생산합수로부터 그 예측량을 제시하였다. 다음으로 조 (2005)도 전체 수산물 및 어업형태별·부류별 수산물 생산량 예측모형을 제시하였다. 그 외 박과 윤 (1996)이 시계열 분석을 이용하여 우리나라 명태어업의 어획량을 예측하기도 하였다. 하지만 양식수산물 중 김, 미역, 굴 등과 같이 연산별로 생산되는 품목에 대한 생산합수 추정 및 생산량 예측모형은 전무한 실정이며, 이러한 품목의 생산량 예측과 관련된 국내 논문도 거의 없는 상황이다.

그러나 외국의 경우에는 수산 분야에서도 생산량의 예측에 대한 다양한 연구들이 이루어져왔다. 대표적으로 Czerwinski et al. (2007)이 Box-Jenkins 모형과 인공신경망모형을 이용하여 단위 어획노력당 어획량에 대한 단기 예측모형을 도출하였고, Koutroumanidis et al. (2006)은 시계열 데이터와 ARIMA 모형을 이용하여 수산물 생산량을 예측하는 모형을 구축하였다. 이 외에도 Mendelssohn (1981), A Noble (1991), Stergiou et al. (1997), Potter et al. (2004), Chesoh et al. (2008), Wertheimer et al. (2009) 등의 연구가 있는데, Box-Jenkins 모형을 이용한 수산물 생산량 예측에 관한 연구가 주를 이루었다.

이에 본 연구에서는 지금까지 상대적으로 소홀히 다루어져 왔던 연산 품목 중의 하나인 굴의 생산량에 대해 다양한 계량기법(다중회귀분석, SARIMA, VECM)을 이용하여 단기 예측을 시도하였다. 그리고 그 결과를 모형별로 비교·분석하여 예측력을 검증해봄으로써 시계열모형을 이용한 굴 생산량 예측 가능성을 제시해 보았다.

끝으로 본 연구의 구성은 우선 본 장에서 연구의 목적 및 필요성, 그리고 본 연구와 관련된 선행 연구들과 연구범위를 간략히 언급한다. 다음으로 2장에서는 굴 생산량 예측을 위한 굴 산업의 생산 현황을 분석하고, 이어서 3장에서는 굴 생산량 예측을 위한 모형설정에 대한 이론과 분석 자료를 제공하고, 이를 바탕으로 그 예측량을 모형별로 제시해 보고자 한다. 끝으로, 4장은 결론 부분으로써 본 논문을 간략히 요약하고, 논문의 실증분석에서 나타난 한계점들을 언급하면서 글을 맺고자 한다.

2. 굴 생산 현황 분석

우리나라 굴 양식은 최초 1897년 원산만에서 살포식 방법에 의해 시작된 것으로 본다. 이 후, 1907년 전남에서 종묘 살포식과 수하방식이 도입되었고, 1923년에는 경남 가덕도에서 규립식 살포법이 도입되어 굴양식이 본격화되었다. 특히 1965년 대일 김 수출의 급감으로 김 어장이 굴 어장으로 전환되는 상황을 맞이했고, 굴 수하식 양식장이 대규모로 조성되면서 우리나라 굴 산업은 생산과 수출 면에서 크게 성장해 왔다.

Fig. 1은 우리나라 천해양식 굴(각굴 기준)의 연도별 생산금액 및 생산량 추이를 나타낸 것이다. 굴 생산량은 1990년 이후 조금씩 감소하였으나, 2000년대 들어서는 다시 증가하는 추세를 보이고 있다. 그러나 2008년부터 굴 생산량은 다시 감소하고 있는 것으로 나타났다. 이러한 추세가 보여주듯이 굴 생산량은 연도별로 매우 불규칙하게 변동하고 있음을 알 수 있다. 이에 대한 가장 큰 원인은 굴 양식의 생산성이 기상과 해황 등과 같은 자연 조건 및 어장의 환경에 크게 좌우되기 때문이다. 그러므로 굴 생산량의 예측력 강화를 위해서는 굴 생산성에 상당한 영향을 미치는 수온이나 강수량 등의 자연변수 이용이 반드시 필요하다.

한편 굴 생산금액 추이를 살펴 볼 때, 굴 생산금액은 국내 굴 생산량 변동에 상당히 민감한 반응을 보임을 알 수 있다. 다시 말해, 굴의 생산량이 증가하면 초과 공급이 발생하여 굴의 단가가 하락하고, 역으로 굴 생산량이 감소하면 초과 수요가 발생하여 가격이 상승하는 현상을 보여 왔다. 한 예로, 2007년 이후 굴의 생산량이 크게 감소함에 따라 초과 수요가 발생하여, 생산단가가 크게 상승해 생산금액 또한 큰 폭으로 상승한 것을 알 수 있다(Fig. 1). 그러므로 굴 생산량 예측 시 월별 위판가격(월별 생산금액/월별 생산량)의 변수 등도 고려할 필요가 있다.

다음으로 Fig. 2는 우리나라 천해양식 굴의 월별 생산

Source: Korean Statistical Information Service (<http://kosis.kr>)

Fig. 1. Trends in volume and value of annual oyster production

Source: Korean Statistical Information Service (<http://kosis.kr>)

Fig. 2. Trends in volume and value of monthly oyster production

량 및 생산금액 추이를 나타내고 있는데, 굴의 주기성을 뚜렷하게 확인할 수 있다. 이는 굴의 생물적 특성과 소비 습성, 양식방법 등에서 기인한다고 볼 수 있다.

다시 말해, 굴은 여름철에 산란을 하기 때문에 채묘와 양식 시설의 설치가 여름에서 초가을까지 이루어진다. 따라서 이 시기에는 생산이 거의 이루어지지 않는다. 굴이 본격적으로 생산되는 시기는 초가을이 지난 10월경부터 12월이며, 익년 1월에 접어들어서는 생산이 최고조에 달했다가 2월 이후 차츰 감소하여 6~7월경에는 거의 출하가 되지 않는다.

이에 굴 생산량은 연별로 뚜렷한 주기성을 가지는 음(-)의 2차 곡선의 형태를 보인다. 그리고 월별·분기별 생산량도 Fig. 3에서 보듯이 12월을 정점으로 서로 상이한 생산 수준을 보이므로 그 양을 예측하기 위해서는 연별 주기성과 월별·분기별 생산수준의 정확한 분석이 필요하다. 아울러 이렇게 반복되는 주기적인 생산량(Y)을 예측하기 위해서는 충분한 시계열 자료의 확보가 요구된다.

Fig. 3. Trends in volume of oyster production by month and season

3. 예측모형 및 실증분석

예측모형

다중회귀분석모형

굴 생산량을 예측하기 위한 다중회귀분석에는 굴 생산량에 영향을 주는 설명변수로써 시차 및 더미 변수를 포함한 굴 위판가격, 생산량, 수온과 강수량을 고려하였다¹⁾. 특히 본 연구에 적용된 식 (1)의 다중회귀분석모형은 굴 생산량에 영향을 줄 수 있는 다양한 변수들에 대한 사전 점검을 거친 후²⁾, 모형의 시계열 안정성과 다중공선성 문제 등을 해결하기 위해 더미변수를 제외한 다른 변수들은 증감률로 전환하였다.

$$\ln Yratio = \alpha_0 + \alpha_1 \ln UPratio(-1) + \alpha_2 \ln Yratio(-1) + \alpha_3 \ln Yratio(-12) + \alpha_4 D2 + \alpha_5 D3 \quad (1)$$

$\ln Yratio$: 로그 변환한 월별 굴 생산량 증감률

$\ln UPratio(-1)$: 로그 변환한 전기의 굴 위판가격 증감률

¹⁾천해양식 굴은 사료를 먹고 자라는 양식어류와 달리 자연에서 생성된 영양성분이나 해황에 따라 생산량이 크게 변동하므로 가뭄, 이상 수온, 장마, 태풍, 해적생물 등의 자연·환경적 요인의 변화가 굴 작황 및 생산량에 영향을 줄 수 있다(한국농촌경제연구원 1988). 또한 최근 가뭄 및 이상 수온에 따른 굴 양식 피해 사례는 다양한 매체를 통해 보도된 바 있다(남해안신문 2012, 수산업관측센터 2011, 연합뉴스 2011, 전남일보 2008, 한국수산업 2008). 이에 본 논문에서는 시계열 자료의 확보가 가능한 수온과 강수량을 자연·환경적 요인 변수로써 사용하였다.

²⁾다중회귀분석모형을 선정하기에 앞서 본 연구는 굴의 대체재로 보이는 홍합 생산가격, 바지락 생산가격과 굴 도매가격, 굴 소매가격, 강수량, 수온 등의 변수들에 대해 시계열의 안정성을 가지는지 그리고 변수 간 다중공선성의 문제는 없는지를 분석하고, 이들 변수들을 포함한 모형의 유의성을 검정해 보았다. 또한 다중회귀분석모형 내의 시차변수는 종속변수와 설명변수 간의 시차상관분석을 통해 선정되었다. 분석 결과, 월별 홍합 생산가격, 바지락 생산가격, 굴 소매가격, 굴 도매가격은 굴 월별 생산량에 대해 유의미한 변수가 되지 못하였을 뿐만 아니라 굴 위판가격과 굴 도매가격, 그리고 굴 도매가격과 굴 소매가격 간에는 다중공선성의 문제가 발생하였다. 그리고 시차상관분석을 통해 선정된 시차변수는 t 기의 로그 변환한 월별 굴 생산량 증감률에 상대적으로 시차 상관성이 높게 나타난 $t-1$ 기와 $t-12$ 기의 로그 변환한 굴 생산량 증감률과 $t-1$ 기의 로그 변환한 굴 위판가격 증감률이 선정되었다. 이상으로 본 논문에서 선정된 다중회귀분석모형은 변수의 유의성, 변수 간 다중공선성, 시계열의 안정성 등의 문제를 해결하는 일련의 과정을 거쳐, 로그 변환한 굴의 월별 생산량 증감률과 월별 위판가격 증감률이 시차변수로 선정되었다. 아울러 수온과 강수량은 t 기의 이상 수온과 가뭄이 굴 생산량에 미치는 영향을 파악하기 위해 시차를 고려하지 않은 더미변수를 선정하였다.

lnYratio(-1): 로그 변환한 전월의 굴 생산량 증감률
 lnYratio(-12): 로그 변환한 12개월 전의 굴 생산량 증감률
 D2: 수온 더미변수, $D2 > |Mean + SD|: 1$,
 $D2 < |Mean + SD|: 0$
 D3: 강수량 더미변수, $D3 > |Mean + SD|: 1$,
 $D3 < |Mean + SD|: 0$

여기서 lnYratio는 종속변수이며, lnUPratio(-1), lnYratio(-1), lnYratio(-12)는 독립변수로 설정하였다. 또한 D2와 D3는 월별 이상 수온과 이상 강수량을 나타내는 더미변수로써 D2와 D3의 “1”은 이들 월별 값이 “수온과 강수량의 평균에 표준편차를 더하거나 빼 합의 절대값($D2 > |Mean + SD|: 1$, $D3 > |Mean + SD|: 1$)” 보다 클 경우를 의미한다³⁾. 특히, D2와 D3의 더미변수를 t 기로 선택한 이유는 최근 이상 기후에 따른 환경의 변화를 가능한 빠르게 반영해 주기 위한 것으로써 그 자료는 국립해양조사원과 기상청이 발표하는 수온 및 강수량 자료를 이용하였다⁴⁾.

SARIMA 모형

시계열 자료의 동태적 움직임을 안정적 시계열로 모형화하는데 주로 사용되는 ARIMA 모형은 현 시점의 자료가 미래에 대한 예측정보를 전혀 포함하지 않고 있다는 백색잡음과정(white-noise process)을 전제로 하고 있다(김과 장 2008)⁵⁾.

ARIMA 모형 식별은 1단계로 계열상관도(correlogram)와 자기상관함수(ACF) 및 편자기상관함수(PACF) 등을 이용하여 시계열의 안정성을 확보한 다음, 2단계로 상기 계열상관도와 ACF, PACF 등을 이용하여 ARIMA 모형의 적합한 차수를 결정한다. 그리고 3단계로, 선정된 모형으로부터 추정된 모수를 이용하여 도출된 예측 결과를 모형선택의 기준이 되는 AIC(Akaike Information Criterion), SBC(Schwarz Bayesian Criterion), HQ(Hannan-Quinn Criterion) 등을 이용하여 모형 적합도를 검정한 후 적합도가 나쁘면 다시 식별 단계로 돌아가고, 역으로 적합도가

좋으면 예측모형으로 선택하게 된다.

본 연구에서 선택된 굴 생산량 예측모형도 이러한 백색잡음과정을 만족시키기 위해 통상적인 ARIMA(p, d, q)에 시차 P 인 계절 AR과 시차 Q 인 계절 MA, D 차 계절 차분한 모형인 ARIMA(0,0,0)(1,1,1)₁₂를 취한다⁶⁾.

$$(1 - \Phi_1 B^{12})(1 - B^{12})Y_t = c + (1 - \theta_1 B^{12})e_t \quad (2)$$

여기서 Y_t 는 월별 굴 생산량을, B 는 후행연산자, 그리고 $B^k Y_t$ 는 Y_{t-k} 를 의미한다. 또한 $E(Y_t)$ 는 0이며, e_t 는 백색잡음으로 평균이 0이고, 분산이 σ_e^2 인 다음과 같은 $e_t \sim N(0, \sigma_e^2)$ 정규분포를 가진다.

VECM 모형

현재 t 기의 불균형이 차기 $t+1$ 기에 그 일부가 조정되어짐이 가능한 VECM모형은 분석에 이용된 변수가 단위근이 존재하여 이들 시계열이 불안정한 것으로 판명되더라도 변수 간 선형결합함수가 공적분성을 가지면 예측을 위한 시계열 분석이 가능하게 된다.

본 연구는 월별 굴 생산량 변수와 상호 인과성을 가질 만한 변수들을 Granger Causality 검정을 통해 우선 선정한 후⁷⁾, 이들 변수의 단위근 존재 유무 및 변수 간 공적분성 존재 유무 등을 검정한 후 월별 굴 생산량의 예측에 적합한 VECM모형을 설정한다. VECM모형의 내생변수로는 월별 굴 생산량과 굴 위판가격을 선정하였으며, 그 외 외생변수로서 통영지역의 수온과 강수량의 전기 더미변수를 사용하였다.

이에 식 (3)은 불안정한 시계열로 두 변수 간 공적분성이 있는 VECM 모형이다.

$$\Delta Y_t = a_0 + \beta_1(Y_{t-1} - \alpha UP_{t-1}) + \sum_{i=1}^k a_{1i} \Delta Y_{t-i} + \sum_{i=1}^k a_{2i} \Delta UP_{t-i} + a_3 D2(-1) + a_4 D3(-1) + \varepsilon_{yt}$$

³⁾수온의 더미변수는 평균에 표준편차를 더하지만 하고 빠지는 않았는데, 이는 평균에 표준편차를 빼 범위를 넘어선 수온과 평균에 양(+)과 음(-)의 표준오차를 더한 범위 내의 수온이 굴 생산량의 변화에 통계적으로 유의미한 변화를 보이지 않았기 때문이다.

⁴⁾여기서 다중회귀분석모형으로부터 굴의 월별 생산량을 예측하기 위해서는 t 기에 확보 가능한 설명변수들의 과거자료를 주로 이용하나, 더미변수인 수온과 강수량만은 가능한 최근 변화를 반영해 주기 위해 t 기의 자료를 이용하였다. 그로 인해 굴 생산량을 예측하는 2011년 9월은 국립해양조사원과 기상청이 각각 제시한 수온과 강수량의 예상치를 이용하였다.

⁵⁾백색잡음과정의 전제를 만족시키려면 연속시계열의 오차항 시계열 $\{e_t\}$ 가 연속 시계열에 있는 각각의 값들에 대해 평균이 0이고 ($E(e_t) = E(e_{t-1}) = \dots = 0$), 분산이 모든 i 와 j 에 대해 일정하며 ($E(e_t^2) = \dots = \sigma^2$), 오차항에 공분산이 모든 j 에 대해 ‘0’이어야 한다 ($Cov(e_t, e_{t-1}) = Cov(e_{t-j}, e_{t-j-i}) = 0$).

⁶⁾계절차분한 ARIMA(0,0,0)(1,1,1)₁₂ 모형 외 추가적으로 계열상관도의 자기상관함수와 편자기상관함수 양측 모두에 나타난 유의미한 스파이크를 제거하기 위해 식 (2)에 AR(12)와 MA(12)를 추가한 모형도 제시한다.

(1 - $\Phi_1 B$)(1 - $\Phi_1 B^{12})(1 - B^{12})Y_t = c + (1 - \theta_1 B)(1 - \theta_1 B^{12})e_t$

⁷⁾인과성을 분석해 본 변수로 차분된 월별 굴 생산량, 차분된 월별 굴 위판가격, 차분된 굴 도매가격, 수온, 강수량에 대해서도 시도해 보았으나, 분석 결과, 차분된 굴 도매가격, 수온, 강수량은 상호 인과성을 가지지 않는 것으로 나타났다.

$$\Delta UP_t = b_0 + \beta_2(Y_{t-1} - \alpha UP_{t-1}) + \sum_{i=1}^k b_{1i} \Delta Y_{t-i} + \sum_{i=1}^k b_{2i} \Delta UP_{t-i} + b_3 D2(-1) + b_4 D3(-1) + \varepsilon_{upt} \quad (3)$$

Y_t : 월별 굴 생산량

UP : 월별 굴 위판가격

$D2(-1)$: 전기의 수온 더미변수

$D3(-1)$: 전기의 강수량 더미변수

실증분석

자료분석

본 논문의 분석에 사용된 우리나라 굴 생산량 및 위판가격의 관측치 수는 260개로 1990년 1월부터 2011년 8월까지의 통계청 어업생산동향조사 월별자료를 이용하였다. 그리고 다중회귀분석의 설명변수로 사용된 수온과 강수량은 굴을 가장 많이 생산하는 통영지역을 대상으로 한 자료로써 수온은 국립해양조사원 자료를, 그리고 강수량은 기상청 자료를 이용하였다.

굴 생산량의 기초통계량을 살펴볼 때, 평균, 최댓값, 최솟값은 각각 약 18,531톤, 62,720톤, 7톤이었다. 또한 수온의 기초통계량은 평균이 16.09°C이었고, 최댓값과 최솟값은 각각 26.93°C, 4.1°C이었다. 그리고 강수량은 평균이 122.60 mm이었으며, 최댓값과 최솟값은 각각 794.80 mm, 0 mm이었다.

또한 이들 변수의 계절별 변화를 살펴보면 Fig. 4와 같이 생산량 및 강수량은 월별 변동폭이 크게 나타났으나, 수온은 변수의 특성상 상대적으로 낮은 월별 변동폭을 보였다. 한편, 이들 세 변수의 계절적 변동을 살펴보면 변수 모두 계절성이 뚜렷한 것으로 나타났으며, 생산량의 계절동향은 수온과 강수량의 계절 동향과 역 방향으로 움직이고 있다는 것도 알 수 있었다.

모형분석

굴 생산량 예측을 위한 앞서 언급한 예측모형에 기초하

Table 1. Basic statistics

	Unit : ton, °C, mm, %, EA		
	Volume	Water Temperature	Precipitation
Mean	18,530.97	16.09	122.60
Median	17,578	16.45	74.85
Maximum	62,720	26.93	794.80
Minimum	7	4.1	0
Std. Dev	14,993.04	6.09	125.91
Skewness	0.58	-0.06	2.04
Kurtosis	2.60	1.66	8.78
Probability	0.000263	0.000056	0.000000
Sum	4,818,053	4,185.86	31,877.20
Observations	260	260	260

Source: Korean Statistical Information Service (<http://kosis.kr>), Korea Meteorological Administration (<http://web.kma.go.kr>), Korea Hydrographic and Oceanographic Administration (<http://www.khoa.go.kr>)

여 분석된 결과를 모형별로 살펴보면 다음과 같다.

① 다중회귀분석모형

우선, 다중회귀분석모형에 사용된 모든 변수에 단위근 검정을 시도해 본 결과, 5% 유의수준에서 모든 변수에 단위근이 존재하지 않는 것으로 나타났다. 왜냐하면 모든 변수의 t -통계량이 5% 유의수준의 critical values인 -2.8738 보다 낮은 것으로 나타났기 때문이다. 따라서 이들 변수는 모두 5% 유의수준에서 안정적인 시계열을 가지고 있다고 판단할 수 있다.

다음으로, 굴 생산량 예측을 위해 적용된 다중회귀분석 모형의 변수 간 다중공선성 유무를 점검해 본 결과, 변수와 상수(C)의 VIF(Variance Inflation Factors)가 모두 10 이하로 다중공선성이 의심되지 않는 것으로 나타났다.

시차 및 더미 변수를 활용한 다중회귀분석모형의 추정 결과, 조정된 결정계수 값이 0.514로 독립변수로부터 설명 가능한 비중이 약 51%임을 알 수 있었다. F-통계량은

Fig. 4. Trends in oyster production volume, water temperature, precipitation by month and season

53.083으로 ‘설명변수들의 계수가 모두 0이 아니다’라는 귀무가설을 1% 유의수준에서 채택하였다.

또한 다른 조건이 일정하다는 가정 하에 전기의 굴 위판가격 증감률이 1% 증가할 때, 굴 생산량 증감률은 1.029% 증가한 것으로 나타났다. 이는 로그 변환한 전기의 굴 위판가격 증감률의 변화에 로그 변환한 굴 생산량 증감률이 다른 변수들에 비해 상대적으로 민감함을 의미하는 것으로 로그 변환한 전기의 굴 위판가격 증감률이 상승하면, t 기의 로그 변환한 굴 생산량 증감률도 다소 증

감함을 의미한다.

다음으로 다른 조건이 일정할 때, 전월 및 12개월 전의 굴 생산량 증감률이 1% 증가할 때, 굴 생산량 증감률은 각각 0.012%, 0.610% 증가하는 것으로 나타났다. 즉, 로그 변환한 전월 및 12개월의 굴 생산량 증감률은 로그 변환한 t 기의 굴 생산량 증감률에 양(+)의 영향을 주기는 하나 그 증감 폭은 작았음을 알 수 있었다. 그러나 로그 변환한 전월의 굴 생산량 증감률 변수의 통계적 유의성은 낮게 나타났다.

끝으로 다른 조건이 일정할 때, 통영지역의 20년간 평균 강수량에 음(-)과 양(+)의 표준편차를 더한 t 기의 이상 강수량(D3)은 로그 변환한 굴 생산량 증감률에 음(-)의 영향을 주는 것으로 나타났다. 이는 이상 강수량의 계절이 장마철인 5~8월에 집중되므로 굴 생산량은 9월을 제외한 6~8월에 여전히 감소하는 경향을 보이기 때문이다.

② SARIMA 모형

우리나라 굴 생산량의 월별 추이 및 계열상관도를 분석해 본 결과, 월별 굴 생산량(Y)의 시계열 자료에 단위근이 존재하며, 분산의 불안정성이 의심되는 것으로 나타났다.

이에 분산의 불안정성과 단위근 문제를 해결하기 위하여 월별 굴 생산량의 연별 주기성을 고려하여 1차 적분 후 12기간 계절 차분을 시도하였다. 그 결과, 12개월 계절 차분한 D12Y($Y_t - Y_{t-12}$)는 단위근이 존재하지 않는 것으로 나타났으며, D12Y의 추세도 시계열의 안정성을 확보한 것으로 나타났다.

따라서 본 연구의 SARIMA 모형은 ARIMA(0,0,0)(1,1,1)₁₂로 선정하였다. 하지만 계열상관도에 한 개의 유의미한 스파이크가 존재함에 따라 이것을 제거하기 위하여 AR(12)와 MA(12)를 ARIMA(0,0,0)(1,1,1)₁₂ 모형에 추가해 보기도 하였다.

분석 결과, 월별 굴 생산량 예측을 위한 SARIMA I과 II

Table 2. Results of unit root test

Augment Dickey Fuller	Variables	t-Statistic	Prob*
Test Statistic	LnYratio	-3.329	0.0147
	LnUPratio(-1)	-7.654	0.0000
	LnYratio(-1)	-3.338	0.0143
	LnYratio(-12)	-3.362	0.0133
	D2	-3.547	0.0076
	D3	-12.053	0.0000
Test critical values	5% level	-2.874	

Note : *means Mackinnon (1996) one-side p-value./Null hypothesis (Ho) is that unit root exists.

Table 3. Test of multicollinearity

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
LNUPratio(-1)	0.059	1.479	1.476
LNratio(-1)	0.003	1.571	1.538
LNratio(-12)	0.002	1.170	1.145
D2	0.002	1.567	1.231
D3	0.002	1.275	1.110
C ⁸⁾	0.000	1.473	-

Table 4. Results of multiple regression analysis model

Variable	Coefficient	Std. Error	t-Statistic	Prob.	R ² (adj. R ²)	F-statistic
LnUPratio(-1)	1.029***	0.244	4.219	0.000	0.524 (0.514)	53.083***
LnYratio(-1)	0.012	0.055	0.221	0.826		
LnYratio(-12)	0.610***	0.047	12.935	0.000		
D2	0.043	0.043	0.998	0.319		
D3	-0.124**	0.050	-2.487	0.014		
C	0.022	0.019	1.166	0.245		

Note : ***, **, * mean that coefficients of variables are statistically significant under 1%, 5%, and 10% critical values respectively

⁸⁾VIF에는 centered와 uncentered VIF의 형태가 있는데, 만일 원식(original equation)이 절편인 상수항(C)을 가지지 않으면 uncentered VIF만 Eviews에서 보여준다.

Fig. 5. Stationarity and heteroskedasticity analysis of Y time series data

Table 5. Results of unit root (Augmented Dickey-Fuller) test

Variable	Level	Prob.
D12Y	-15.0903	0.000000

Note : Null hypothesis was rejected under 1% critical value. Null hypothesis is that D12Y has a unit root

모형 모두 조정된 결정계수 값이 각각 0.056, 0.246으로 낮게 나타나 예측을 위한 모형으로써 적합성이 떨어짐을 알 수 있었다. 이는 월별 굴 생산량이 가지고 있는 연별 주기성과 월별·분기별 생산수준의 변동에 기인한 것으로 ARIMA 모형의 적합한 차수 선택이 곤란했기 때문이다.

Fig. 6. Stationarity analysis of D12Y time series data

Table 6. Results of SARIMA models

Model	Variable	Coefficient	Std. Error	t-Statistic	Prob.	R ² (adj. R ²)	F-statistic
SARIMA I	C	196.90	537.64	0.3662	0.7145	0.064 (0.056)	8.418
	AR(1)	-0.90***	0.023	-38.739	0.0000		
	MA(1)	0.99***	0.004	271.583	0.0000		
SARIMA II	C	215.760	246.592	0.875	0.3825	0.259 (0.246)	20.220
	AR(1)	0.094	0.102	0.918	0.3596		
	AR(12)	0.033	0.090	0.372	0.7100		
	MA(1)	0.078	0.080	0.976	0.3300		
	MA(12)	-0.623***	0.076	-8.152	0.0000		

Note : *** means that coefficients of independent variables are statistically significant under 1% critical value

Table 7. Results of Granger causality test

Null hypothesis	Observations	F-statistic	Prob.
UP does not G-C Y	258	3.075	0.0479
Y does not G-C UP		3.319	0.0378

Note : lag 2 was included to test Granger Causality and null hypothesis was rejected under 5% critical value

Table 8. Results of unit root (Augmented Dickey-Fuller) test

	Level	Prob
Y	-0.0472	0.6661
UP	-2.3060	0.0207

Note : Y and UP variables have unit roots under 1% critical value

Table 9. Results of Johansen cointegration test

Data Trend	None	None	Linear	Linear	Quadratic
Test Type	No Intercept	Intercept	Intercept	Intercept	Intercept
	No Trend	No Trend	No Trend	Trend	Trend
Selected (0.05 level*) Number of Cointegrating Relations by Model					
Trace	1	2	2	2	2
Max-Eig	1	2	2	2	2
*Critical values based on MacKinnon-Haug-Michelis (1999)					
Information Criteria by Rank and Model					
Data Trend	None	None	Linear	Linear	Quadratic
Rank or	No Intercept	Intercept	Intercept	Intercept	Intercept
No. of CEs	No Trend	No Trend	No Trend	Trend	Trend
Log Likelihood by Rank (rows) and Model (columns)					
0	-4552.696	-4552.696	-4552.684	-4552.684	-4552.646
1	-4510.461	-4493.932	-4493.920	-4490.123	-4490.122
2	-4508.644	-4473.030	-4473.030	-4467.270	-4467.270
Akaike Information Criteria by Rank (rows) and Model (columns)					
0	35.8329	35.8329	35.8485	35.8485	35.8638
1	35.5330	35.4112	35.4189	35.3970	35.4048
2	35.5501	35.2865	35.2865	35.2570*	35.2570
Schwarz Criteria by Rank (rows) and Model (columns)					
0	36.0551	36.0551	36.0984	36.0984	36.1416
1	35.8107	35.7028	35.7245	35.7164	35.7381
2	35.8834	35.6475	35.6475	35.6458*	35.6458

또한 모형 I에 AR(12)와 MA(12)를 추가한 모형 II는 모형 I 보다 모형의 설명력은 상대적으로 나은 편이나 모형에 이용된 변수의 통계적 유의성은 떨어지는 것으로 나타났다.

③ VECM 모형

월별 굴 생산량 예측을 위해 서로 인과성을 가진 변수들의 검정 결과 월별 굴 위판가격과 월별 굴 생산량 간에 상호 인과성이 존재하는 것으로 나타났다.

VECM 모형의 정확한 설정을 위한 단위근 검정 결과는 Table 8에 나타나 있다. 검정 결과, 월별 굴 위판가격과 월별 굴 생산량은 1% 유의수준 하에서 단위근이 존재하는 것으로 나타났다.

모형의 내생변수인 월별 굴 생산량과 위판가격이 level 수준에서 단위근을 가지므로 이들 변수의 공적분 분석을 시도해 보았다.

공적분 분석 결과, 이들 변수 간에는 두 개의 공적분을 가지는 것으로 나타났다. 특히 자료의 속성을 정확히 파악하기 위해 옵션을 모두 설정하여 모형을 분석해 보았을 때, AIC나 SBC가 최소화되는 모형은 선형의 상수항과 추세를 가진 모형이었다. 이에 본 분석에서는 선형의 상수항 및 추세를 가진 모형을 택하여 VECM 분석을 시도하였다.

다음으로 요한센 공적분 검정을 통해 선택된 선형의 상수항 및 추세를 포함한 lag 2의 VECM모형을 분석하면 다음과 같다.

여기서 lag 2를 선택한 이유는 Y와 UP의 시차상관 분석결과, 시차 1에서 높은 상관관계를 보이다가 다시 감소하는 경향을 보였기 때문이다. 분석 결과, 두 변수 간의 공적분성이 존재함을 공적분 방정식의 변수별 t-통계량, 4.057과 -2.777을 통해 알 수 있었다. 그리고 두 모형의 조정 결정계수는 각각 0.210, 0.204로 굴 예측량 추정치 대략 21%의 설명력을 가지는 것으로 나타나 모형의 설명력이 SARIMA 모형과 같이 낮게 나타났다. 또한 모형 적합도의 진단기준이 되는 AIC와 SBC는 모형별로 각각 21.583, 21.695와 13.964, 14.074로 월별 굴 위판가격 모형이 월별 굴 생산량 모형보다 더 적합한 것으로 나타났다.

예측 결과

월별 굴 생산량 예측을 위해 제시한 다중회귀분석모형과 SARIMA 모형 I, II 그리고 VECM 모형의 단기예측 결과는 Table 11과 같다. 분석 결과, 다중회귀분석모형, SARIMA 모형 I, II, VECM 모형 각각의 2011년 9월 굴

Table 10. Results of VECM

Cointegrating Eq ⁹⁾	Coint Eq. (1)	
Y(-1)	1	
UP(-1)	20.6451 (4.0576)	
@TREND(90M01)	-39.9571 (-2.7772)	
C	-23245.42	
Error Correction:	D(Y)	D(UP)
Coint Eq. (1)	-0.492 (-6.597)	-0.008 (-5.101)
D(Y(-1))	0.200 (3.010)	0.007 (4.572)
D(Y(-2))	0.149 (2.248)	0.004 (2.374)
D(UP(-1))	9.840 (3.625)	-0.179 (-2.984)
D(UP(-2))	0.720 (0.254)	-0.107 (-1.708)
C	950.537 (1.041)	32.065 (1.586)
D2(-1)	450.232 (0.202)	-204.627 (-4.164)
D3(-1)	-9,007.574 (-3.721)	115.794 (2.160)
R ²	0.232	0.226
[Adj. R ²]	[0.210]	[0.204]
F-statistic	10.728	10.383
Log likelihood	-2,765.491	-1786.309
AIC	21.583	13.964
SBC	21.695	14.074
Log likelihood	-4,544.931	
AIC	35.517	
SBC	35.779	

Note : () is t-Statistic, AIC: Akaike information criterion, SBC: Schwarz Bayesian criterion

생산량 예측치는 1,337톤, 12,423톤, 12,442톤, 10,425톤으로 추정되었다. 예측 결과, 다중회귀분석모형이 다른 세 모형들에 비해 실제치와 예측치 간에 예측오차가 가장 적은 246톤을 보였으나, 예측 적합도 측면에서는 Theil Inequality Coefficient가 가장 낮은 VECM 모형이 상대적으로 좋은 것으로 나타났다. 그러나 VECM 모형 또한 예측력은 좋지 못했다.

한편, 다중회귀분석모형의 경우 월별 굴 생산량 자료의 시계열 안정성을 확보하기 위해 로그 변환한 월별 굴

⁹⁾공적분 식에서 상수항 C와 추세항 @TREDND(90M01)가 포함된 것은 Table 9의 요한센 공적분 검정에서 자료 추세가 선형이며, 상수항과 추세항을 가진 모형이 적합한 것으로 판명되었기 때문이다.

Table 11. Results of short-term forecasting of oyster production

	MRAM	SARIMA Model		VECM
		(0,0,0)(1,1,1) ₁₂	(0,0,0)(1,1,1) ₁₂ (1,0,1) ₁₂	
Forecasting volume (ton) (2011.09)	1,337	12,423	12,442	10,425
Actual volume (ton) (2011.09)		1,019		
Forecasting error (ton)	246	11,404	11,423	9,406
Theil IC	0.580	0.858	0.930	0.357

Note : MRAM (Multiple Regression Analysis Model), Theil IC (Theil Inequality Coefficient)

생산량 증감률을 활용하여 예측값을 추정하였다. 이에 월별 굴 생산량 예측값을 추정할 때, 역 환산하는 과정에서 예측값이 다소 과대 추정되어질 수 있는 문제도 있을 수 있다.

한편 SARIMA 모형의 경우, 월별 굴 생산량의 연별 주기성과 월별·분기별 상이한 생산수준에 기인하여 시계열 안정성이 1차 차분 또는 로그 변환을 취함에도 불구하고 시계열의 안정성이 확보되지 않는 것으로 나타났다. 이에 따라 월별 굴 생산량을 12기간 계절 차분한 후 적합한 AR, MA 차수를 선정하다보니 예측력이 좋지 못한 것으로 판단된다.

4. 결론 및 요약

본 연구는 우리나라 천해양식 굴의 월별 생산량을 다중 회귀분석모형, SARIMA 모형, VECM 모형을 이용하여 단기 예측값을 추정해 보았다.

분석 결과, 다중회귀분석모형이 다른 두 모형에 비해 상대적으로 실제값과 예측값 사이의 오차가 적은 것으로 나타났다. 그러나 예측 적합도 측면에서는 VECM 모형이 다중회귀분석모형보다 더 적합한 것으로 나타났다. 하지만 상기 네 모형 모두 예측오차가 큰 것으로 나타났다.¹⁰⁾ 구체적으로 다중회귀분석모형의 2011년 9월 굴 생산 예측값은 1,337톤으로 실제값인 1,019톤보다 246톤이 많은 것으로 분석되었다. 그러나 그 외 SARIMA I, II 모형과 VECM모형은 다중회귀분석모형에 비해 실제값과 예측값 사이의 오차가 크게 나타나 향후 실제값과 추정 예측값 사이의 월별 오차 감정을 통해 모형을 수정·보완해야 될 필요가 있음을 알게 되었다.

다음으로 본 연구의 의의는 굴과 같이 생산의 주기가

연별로 반복되어지고 생산량이 월마다 뚜렷하게 차이가 나는 연산별 품목에 대해 월별 생산수준을 고려한 시차변수와 수온 및 강수량 등을 반영한 시계열모형으로 그 양을 단기 예측해 보았다는 데 의의가 있다. 특히 수산부문에 생산량 예측은 주로 어업별이나 부류별이 아니면 해면어업의 대표어종(고등어, 오징어, 갈치, 조기)에 한정되어 예측 분석이 실시되었으나, 본 연구에서는 연산별로 생산되는 양식품목인 굴의 생산량 예측을 시도해 보았다는 점에서 그 의의가 있다. 특히 향후 연산별 어종인 김, 미역, 다시마 등의 생산량 예측 연구에도 본 연구가 참고자료로써 도움이 될 것으로 기대된다.

또한 본 연구의 한계로써 우선, 연별 주기성과 월별·분기별 생산수준이 상이한 굴 생산량의 경우, 시계열 자료 자체의 안정성을 확보한 후 예측모형을 구축하다보니 모형의 적합도가 낮게 나타나는 문제가 있었다. 그 결과 단변량 변수를 가지고 예측하는 SARIMA I, II 모형의 경우 예측력이 떨어지는 한계가 있었다. 둘째, 굴 생산량을 예측함에 있어 예측에 영향을 줄 수 있는 생물적, 경제적 월별 변수 확보의 어려움을 들 수 있다. 다시 말해, 자연환경적 요인인 수온 및 강수량 외에 굴 생산량에 영향을 줄 수 있는 생물학적 요인들의 축적된 자료 확보가 어렵다 보니 설명력 높은 모형을 구축하는 데 한계가 있었다. 끝으로 굴 생산량 예측모형의 구축에 있어 굴 생산량에 영향을 미치는 굴 시설량을 의미하는 월별 변수로 전환할 수 있는 대리변수(예, 잔여시설량)의 선정이 자료수 부족으로 분석에 적용되지 못한 점 등을 들 수 있다.

끝으로 향후 연구 과제로써 인과성과 단위근 존재를 전제로 한 구조분석인 VECM의 제약으로 인해 의미 있는 변수들이 누락될 수 있는 한계를 극복하기 위해 추가적인 변수(생태 및 환경변수, 시설량 대리 변수 등) 확보를 통

¹⁰⁾ 선정된 굴 생산량 예측 모형에 있어 예측 오차가 높게 나타나는 이유를 찾아보자면, 우선, SARIMA모형은 연별 주기성과 월별·분기별 생산수준의 상이성을 가진 굴 월별 생산량 자료의 시계열 안정성 확보가 어려워 부득이 D12Y차분까지 모형을 확대 분석함에 따라 예측 오차가 크게 발생한 것으로 판단된다. 둘째로, VECM 모형은 변수 간의 상호 인과성 및 개별 변수의 단위근 존재 등을 고려하다 보니 이용 가능한 변수 선정의 폭이 줄어들면서 예측오차의 폭을 좁히지 못한 것으로 여겨진다. 끝으로 다중회귀 분석모형, SARIMA, VECM 모형 모두에 적용되는 문제로 굴 생산량에 영향을 주는 자연·환경적 변수 및 시설량 변수 등의 유의미한 변수들의 발굴 및 이용이 자료 확보의 한계로 시도되지 못하면서 예측오차가 크게 나타난 것으로 판단된다.

해 다중회귀분석을 다양하게 접근해 볼 필요가 있다고 판단된다. 추가적으로 본 연구는 추후 out-of sample 분석에 기초하여 한 달 앞선 전망(one-month-ahead forecast)에서 적어도 12개월 앞선 전망(twelve-month ahead forecast)을 시도하여 전망치와 실제치 간에 DM 검정(Diebold and Mariano test)을 통해 모형의 예측력을 진단해 볼 필요가 있다(Diebold and Mariano 1995).

참고문헌

- 강창완, 김대학 (2000) 쌀 예상 생산량 추정방법에 대한 연구. 응용통계연구 13(2):329-341
- 고성보 (2003) 생산형태별 감자의 수급예측모형 개발. 농업경제연구 44(4):221-222
- 국립해양조사원 (2011) 실시간 연안정보. <http://info.khoa.go.kr/app/realtimeData/tideReal.asp?sgrp=D03&siteCmsCd=CM0016&topCmsCd=CM0190&cmsCd=CM0228&pnum=2&cnum=11> Accessed 10 Oct 2011
- 기상청 (2011) 관측자료. http://www.kma.go.kr/weather/observation/past_table.jsp Accessed 10 Oct 2011
- 김명직, 장국현 (2008) 금융시계열분석 제2판. 경문사, 822 p
- 김배성, 박미성, 조재환, 김태균 (2010) 중기선행관측을 위한 농축산물 작형별 수급모형 및 예측평가시스템 개발 연구. 한국농촌경제연구원, M103, 77 p
- 김현용 (2000) WTO 관세인하가 수산물 수급에 미치는 영향과 대책. 경제학박사 학위논문, 부경대학교, 134 p
- 남해안신문 (2012) 한파에 따른 저수온 우려·양식장 관리 비상 <http://www.nhanews.com/news/articleView.html?idxno=45566> Accessed 14 Jan 2012
- 박철형, 이광남 (1997) 우리나라 연근해산 수산물의 생산량 예측에 관한 조사 연구. 수산업협동중앙회 수산경제연구원, 연구보고 97-1, 128 p
- 박해훈, 윤갑동 (1996) 시계열분석을 이용한 한국 명태어업의 어획량 예측. 어업기술연구 32(3):235-240
- 수산업관측센터 (2011) 굴 수산물관측 244호, 257호. 한국해양수산개발원, 8 p
- 연합뉴스 (2011) 한파 장기화 양식장 저수온 피해 우려 <http://www.yonhapnews.co.kr/bulletin/2011/01/13/0200000000AKR20110113107800054.HTML?did=1179m> Accessed 14 Jan 2012
- 이진희, 신기일 (2000) 조건부 자기회귀모형을 이용한 송이 버섯 생산량 예측. 응용통계연구 13(2):307-320
- 장석환 (2000) 주요 식량작물의 생산량 예측 모형에 관한 연구. 한국데이터정보과학회 11(1):47-48
- 전남일보 (2008) 굴 생산 급감...어민 “어떡하나” <http://jnilbo.com/read.php3?aid=1227106800281788153> Accessed 16 Sep 2012
- 조용준 (2005) 수산물 생산량 예측모형 연구 : 시계열 모형을 중심으로. 수산업협동중앙회 수산경제연구원, 351 p
- 통계청 (2011) 어업생산동향조사. http://kosis.kr/abroad/abroad_01List.jsp?parentId=F Accessed 2 Oct 2011
- 한국농촌경제연구원 (1988) 수산재해에 따른 양식공제시험사업 설계 : 굴 수하식양식을 중심으로. 한국농촌경제연구원, 연구보고 176, 125 p
- 한국수산업 (2008) 남해안 가을가뭄에 어심도 탄다 http://fisheco.com/sub_page.html?newsNum=C20084467 Accessed 8 Jan 2012
- Chesoh S, Lim A (2008) Forecasting fish catches in the Songkhla Lake basin. Sci Asia 34:335-340
- Czerwinski I, Gutierrez-Estrada J, Hernando-Casal J (2007) Short-term forecasting of halibut CPUE : Linear and non-linear univariate approaches. Fish Res 86:120-128
- Diebold X, Mariano S (1995) Comparing predictive accuracy. J Bus Econ Statist 13:253-263
- Koutroumanidis T, Iliadis L, Sylaios G (2006) Time-series modeling of fishery landings using ARIMA models and Fuzzy Expected Intervals software. Environ Modell Softw 21:1711-1721
- Mendelsohn R (1981) Using Box-Jenkins Model to Forecast Fishery Dynamics : Identification, Estimation, and Checking. Fish B-NOAA 78(4):887-896
- Noble A, Sathianandan TV (1991) Trend analysis in all-India mackerel catches using ARIMA models. Indian J Fish 38:119-122
- Potter ECE, Crozier WW, Schön PJ, Nicholson MD, Maxwell DL, Prévost E, Erkinaro J, Gudbergsson G, Karlsson L, Hansen LP, Maclean JC, Maoiléidigh Ó, Prusov S (2004) Estimating and forecasting pre-fishery abundance of Atlantic salmon (salmo salar L.) in the Northeast Atlantic for the management of mixed-stock fisheries. ICES J Mar Sci 61:1359-1369
- Stergiou K, Christou E, Petrakis G (1997) Modelling and forecasting monthly fisheries catches : Comparison of regression, univariate and multivariate time series methods. Fish Res 29:55-95
- Wertheimer A, Orsi J, Fergusson E, Sturdevant M (2009) Forecasting Pink Salmon Harvest in Southeast Alaska from Juvenile Salmon Abundance and Associated Environmental Parameters : 2008 Returns and 2009 Forecast. NPAFC, 20 p

Received Jan. 3, 2012

Revised Feb. 20, 2012

Accepted Mar. 24, 2012