

조건부 자기회귀모형을 이용한 송이버섯 생산량 예측

이진희¹⁾ 신기일²⁾

요약

송이버섯 생산량과 기후인자와의 관계를 통계적으로 규명하기 위한 노력이 꾸준히 진행되어 왔다. 최근 박현 등(1998)은 송이버섯 생산량과 기후인자의 관계를 자기회귀모형을 이용하여 분석하였으나 예측력이 떨어지는 것으로 나타났다. 본 논문에서는 예측의 정확성을 높이기 위한 방법으로 송이버섯 생산이 있다는 조건을 이용한 조건부 자기회귀모형을 제안하였다. 두 모형의 예측력을 비교한 결과 조건부 자기회귀모형이 더 우수한 것으로 나타났다.

주요용어: 자기회귀분석법, 조건부 자기회귀분석법.

1. 서론

우리 나라 임업경제의 최고 소득원인 송이버섯의 인공재배를 위한 연구는 100 여 년에 걸쳐 이루어져 왔다. 그러나 그 동안의 꾸준한 연구노력에도 불구하고 송이버섯의 인공재배는 이루어지지 않고 있다. 더구나 최근의 기상이변 등으로 인한 연별 송이버섯 생산량의 차이와 지역별 편중현상의 심화는 보다 안정적이며 더 많은 송이버섯 생산량을 얻기 위한 연구를 필요로 하게 되었다. 최근 강안석 등(1989), 박현 등(1995), 조덕현 등(1995)은 기후특성이 생산에 미치는 영향과 관련성에 관하여 연구하였다. 그러나 지역적인 특성을 고려하지 않고 전국적인 수준에서 연구를 수행하였기 때문에 지역별 토양인자나 기후요건이 다름으로 인해 만족스러운 결과를 얻지 못했다. 더구나 기후인자나 송이버섯 생산량이 시계열 자료임에도 불구하고 임업시험장(1984)은 월별 단순회귀분석으로 박현 등(1989)은 지역별 입지환경을 이용한 권역별 회귀분석으로 분석하여 좋은 결과를 얻지 못하였다. 이후 박현 등(1998)은 각 지역별로 일별 송이 생산량과 일별 기후인자를 자기회귀모형을 이용하여 분석하였다. 9월과 10월의 일별 송이버섯 생산량 자료를 종속변수로(생산량이 없으면 "0"을 사용) 그리고 이에 해당되는 일별 기후자료를 독립변수로 자기회귀모형을 만들었다. 그러나 이 모형을 이용한 예측값은 실제값과 많은 차이를 보였다. 본 논문에서는 이를 극복하기 위한 하나의 방법으로 송이버섯 생산이 있다는 조건을 이용한 조건부 자기회귀모형을 제안하였다. 이는 송이버섯 생산량이 있을 확률이 예측에 많은 영향을 주는 것으로 나타났다기 때문이다.

1) (449-791) 경기도 용인시 모현면, 한국외국어대학교 통계학과, 대학원

E-mail: jhlee@stat.hufs.ac.kr

2) (449-791) 경기도 용인시 모현면, 한국외국어대학교 통계학과, 부교수

E-mail: keyshin@stat.hufs.ac.kr

2절에서는 박현 등(1998)에서 사용된 종속변수와 독립변수에 관하여 설명하였으며 이 자료는 본 논문에서 제안한 조건부 자기회귀분석법에서도 그대로 사용하였다. 3절에서는 박현 등(1998)에서 사용한 자기회귀분석법의 결과에 관하여 간단히 설명하였다. 4절에서는 자기회귀분석법과 조건부 자기회귀분석법에서 얻어진 모형을 이용한 예측결과를 비교하였으며 끝으로 결론은 5절에 있다.

2. 자료설명 및 독립변수의 생성

박현 등(1998)은 송이버섯 생산량과 기후인자와의 관계를 알아보기 위하여 28개 지역(송이 주산지)에서 생산된 일별 송이버섯 생산량과 각 지역의 일별 기후자료를 1990년부터 1995년까지 조사하여 자기회귀분석법을 이용하여 분석하였다. 조사된 기후 자료와 기호는 다음과 같으며 각 자료는 일별 자료이다. 생산량 (Y_t : 9월, 10월, kg), 최고온도 (X_1), 최저온도 (X_2), 평균온도 (X_3), 강수량 (X_4 , mm), 상대습도 (X_5 , %), 증발량 (X_6 , mm), 일조량 (X_7 , 시간), 지온 (X_8), 지중 5cm온도 (X_9), 지중 10cm온도 (X_{10}).

일부 지역의 경우 위에서 언급한 모든 기후자료가 얻어지지 않았으며 이 경우 얻어진 기후자료만이 독립변수로 사용되었다. 또한 독립변수로 사용되는 기후요인은 송이버섯 생산량에 영향을 미침에 있어 바로 당일이나 하루 전 뿐 아니라 어느 정도 시차를 두고 꾸준히 영향을 미치는 것으로 알려져 있으며 기존의 연구결과를 토대로 30일 전까지의 7일 평균 기후자료와 7일 평균기후를 제공한 자료의 시차자료를 생성하였다. 생성된 600여 개의 독립변수 중에서 Stepwise 방법을 이용하여 변수를 선택한 후 분석을 실시하였다. 이에 관한 자세한 내용은 박현 등(1998)을 참조하기 바란다.

본 논문에서는 자기회귀분석법과 조건부 자기회귀분석법을 비교하기 위하여 위의 자기회귀분석법에서 사용한 종속변수와 독립변수를 그대로 사용하였으며 분석을 간단히 하기 위하여 28개 지역 중에서 강릉, 영덕, 거창, 청도 등 4개 지역을 선택하였다.

3. 자기회귀분석방법

2절에서 설명한 바와 같이 박현 등(1998)은 9월과 10월의 일별 송이생산량을 종속변수로, 같은 해의 일별 기후인자를 독립변수로 하는 자기회귀모형을 설정하였다. 박현 등(1998)의 자기회귀분석에서는 1996년도의 송이버섯 생산량을 예측하지 않았으나 본 논문에서는 조건부 자기회귀모형과의 예측력을 비교하기 위해서 박현 등(1998)의 방법을 토대로 송이 주산지 28개 지역 중 4개 지역을 대상으로 분석을 다시 하였다.

분석에 사용된 자기회귀모형은 다음과 같다.

$$Y_t = \beta_0 + \sum_{j=0}^{29} \sum_{k=1}^{10} \beta_{kj} X_{kt-j} + \sum_{j=0}^{29} \sum_{k=1}^{10} \alpha_{kj} X_{kt-j}^2 + \epsilon_t \quad (3.1)$$

여기서, Y_t 는 9월, 10월의 일별 송이버섯 생산량이고 X_{kt-j} 는 7일간의 평균기후자료와 이에 따른 시차(lag) 자료이며 X_{kt-j}^2 은 각 자료의 제곱을 의미한다. 이때, ϵ_t 는 오차 항으로 μ -차

자기회귀모형(AR(p))을 따른다고 가정한다.

자기회귀분석법에 의한 분석 결과는 표 3.1과 표 3.2 그리고 그림 3.1에 나와있다. 여기서 자기회귀모형에서 나오는 오차항의 차수는 SAS/AUTOREG 의 BACKSTEP문을 사용하여 결정하였다. 표 3.1은 박현 등(1998)의 결과표이고, 표 3.2는 조건부 자기회귀분석법과의 비교를 위하여 새롭게 만든 표이다.

예측의 정확성을 알아보기 위하여 표 3.2에서 사용한 통계량의 정의는 다음과 같다. 여기서 연별 MAPE, MSPE는 1996년도 송이버섯 총 생산량의 MAPE와 MSPE로 $n = 1$ 인 경우에 해당된다.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}, MSPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{y_i}\right)^2$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

표 3.1: 지역별 선택변수 수 및 결정계수

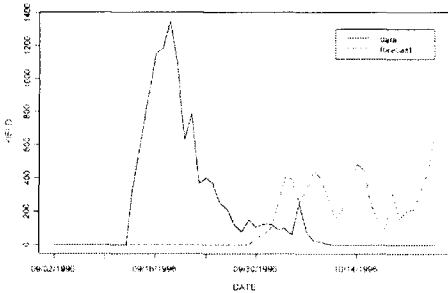
지역	사용 안된 독립변수	자기회귀 모형에서 선택된 변수 수	자기회귀 모형의 결정계수	사용된 생산량의 연도
강릉	없음	29	0.90	90 ~ 95년
영덕	X_6, X_8, X_9, X_{10}	22	0.89	90 ~ 95년 (94년*)
거창	$X_6, X_7, X_8, X_9, X_{10}$	37	0.92	90 ~ 93 95년
청도	없음	70	0.96	90 ~ 95년

*표가 붙어있는 해는 자료수가 적음

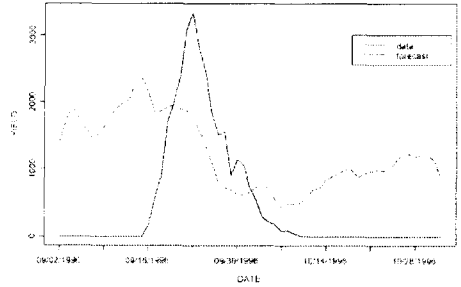
표 3.2: 지역별 자기회귀모형에 의한 예측결과

지역	MAPE (일별)	MSPE (일별)	MSE (일별)	MAE (일별)	MAPE (연별)	MSPE (연별)
강릉	4.886155	180.5971	1731.185	6.324393	0.73498	0.541099
영덕	10.46324	1348.610	17.336892	0.5331148	0.028676	0.0008224
거창	1.628136	8.879803	1406.96	93.56406	0.03836	0.001472
청도	0.860810	0.826527	925.589418	5.7495031	0.614731	0.377895

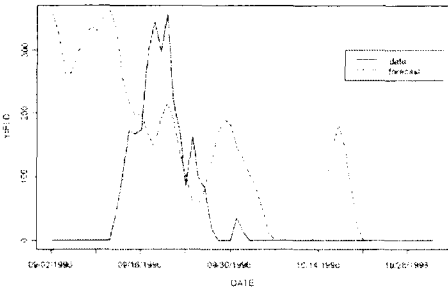
먼저 표 3.1을 살펴보면 자기회귀모형의 결정계수는 약 0.9를 상회하는 것으로 나타나 모형이 어느 정도 잘 적합 되었다고 할 수 있다. 그러나 표 3.2에서 청도지역을 제외한 3개 지역에서 MAPE와 MSPE값이 상당히 크게 나오며 또한 청도지역도 MSE와 MAE가 매우 크기 때문에 예측을 기준으로 한다면 자기회귀모형은 좋은 모형이라 할 수 없다. 또한 임업연구원에서 관심을 갖는 연별 총 생산량의 예측도 잘 맞지 않는 것으로 나타났다. 그림 3.1의 그래프(실선은 실제 생산량을, 점선은 예측량을 나타냄)에서도 실제생산량의 그래프와 예측 그래프를 비교해 볼 때 잘 맞지 않음을 알 수 있다.



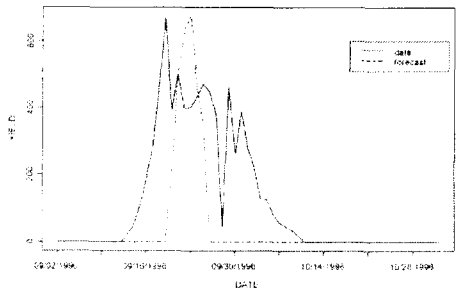
강릉 : 자기회귀모형



영덕 : 자기회귀모형



거창 : 자기회귀모형



청도 : 자기회귀모형

그림 3.1: 지역별 자기회귀모형에 의한 예측그래프

4. 조건부 자기회귀분석법

이 절에서는 자기회귀분석에서 발생하는 예측의 부정확성을 개선하기 위한 방법으로 송이버섯 생산이 있다는 조건을 이용한 조건부 자기회귀모형을 살펴보도록 하자. 일반적인 시계열 자료와 달리 송이버섯은 일년 중에 9월과 10월에 주로 생산되기 때문에 자기회귀분석에서는 이 두 달의 일별 송이버섯 생산량을 종속변수로 사용하였다. 그러나 종속변수로 사용되는 일별 송이버섯 생산량의 정확한 값을 구하는 것은 매우 어렵다. 예를 들어, 송이버섯이 자라 산에 있더라도 이를 채취하지 않으면 생산량이 없는 것으로 되고 오늘 송이버섯을 채취하였더라도 다음날 이를 수매하면 다음날의 생산량이 되는 등의 문제점이 있기 때문이다. 이에 반하여 송이버섯이 생산된다 또는 송이버섯 생산이 안 된다 하는 것은 기후요인과 밀접한 관계가 있고 어느 정도 정확하게 알아 낼 수 있다. 따라서 이러한 정보를 이용하면 기존의 자기회귀분석법에 비하여 더 좋은 예측을 할 수 있게 된다. 이 절에서 소개 될 조건부 자기회귀 모형은 다음의 두 부분으로 나누어진다. 먼저 송이버섯 생산량의 크기에 관한 부분으로 그 크기는 실제로 송이버섯이 생산된다는 조건하에서의 생산량의 크기를 나타낸다. 따라서 이 값은 언제나 양의 값을 갖으며 기후 조건이 맞으면 큰 값을 갖고

기후 조건이 맞지 않으면 작은 값을 갖게 된다. 다음으로 송이버섯이 생산된다 또는 생산이 안 된다 하는 부분으로 이는 베르누이 확률변수로 나타낼 수 있다. 이 베르누이 확률변수 역시 기후 조건과 관계가 있다. 이 두 결과를 이용하여 최종적인 송이 생산량은 송이버섯이 생산될 확률과 송이버섯이 생산된다는 조건 하에서 얻어진 생산량의 크기의 곱으로 얻어질 수 있다.

4.1. 조건부 자기회귀분석법을 위한 자료 구조

본 절에서는 조건부 자기회귀분석법을 위한 종속변수와 독립변수의 구조를 살펴본다. 먼저 1990년에서 1995년까지의 일별 기후자료와 일별 송이버섯 생산량 자료로부터 2절에서 설명한 바와 같이 종속변수와 독립변수로 사용될 7일 평균 기후자료를 생성한다. 그리고 얻어진 1996년 자료는 예측을 위하여 분석에서 제외한다. 그리고 앞 절에서 언급하였듯이 생산량의 크기를 위한 분석과 송이버섯 생산이 있다 또는 생산이 없다에 관한 분석에 맞게 종속변수와 독립변수에 사용될 자료를 생성한다.

먼저 생산량의 크기를 분석하기 위하여 다음의 두 가지 방법을 이용하여 자료를 만든다. 첫 번째 방법은 자기회귀분석법에서 사용한 종속변수와 독립변수 자료(모든 lag 변수와 이를 제공한 변수) 중에서 만약 종속변수 값이 "0" 이면, 주어진 자료 set에서 "0"인 종속변수와 이에 해당하는 모든 독립변수 자료를 제거한다. 그리고 남아있는 종속변수와 독립변수 자료를 분석에 사용하는 것이다. 이 방법은 식물의 특성상 전 연도의 송이 생산량이 많으면 올해도 송이버섯이 많이 발생하는 것을 전제로 한 방법이다. 두 번째 방법은 순수하게 기후요인과 생산량의 크기를 보기 위하여 전 연도의 영향을 제거하는 방법으로 첫 번째 방법의 모든 "0"인 종속변수와 이에 해당하는 독립변수를 제거하는 것과 달리 일부 "0"인 자료는 결측처리를 하여 사용하는 방법이다. 즉, 자기회귀분석법에서 사용된 종속변수 중 "0"의 값을 갖는 것 중에서 매년 생산량이 있는 후 일부에 "0" 대신에 "."을 찍어 결측처리를 한다. 그리고 매년 생산량이 있기 전과 결측점 후의 다른 "0"을 갖는 종속변수 그리고 이에 해당하는 모든 독립변수를 제거한 종속변수와 독립변수 자료를 분석에 사용하는 방법이다. 본 논문에서는 편의상 전자의 방법을 "결측처리 전 모형"이라 하고 후자를 "결측처리 후 모형"이라 부른다.

다음으로 송이버섯 생산이 있다 또는 없다에 관한 분석을 위하여 다음의 두 가지 방법으로 새로운 자료를 만든다. 먼저 첫 번째 방법은 자기회귀분석에서 사용하였던 종속변수 중에서 종속변수가 "0" 이상이면 "1"로 바꾸고 그 나머지 모든 종속변수와 독립변수에 관한 자료는 그대로 사용하는 방법이다. 두 번째 방법은 첫 번째 방법을 그대로 사용하되 이 방법에서 얻어진 송이버섯 생산이 있을 확률을 그대로 사용하는 것이 아니라 기준 통계량으로 사용될 일별 MAPE와 MSPE 등을 최소로 하는 최적의 확률 값을 정한 후, 이 값보다 작은 송이버섯이 생산될 확률을 "0"으로 처리하는 방법이다. 이 두 방법도 편의상 본 논문에서는 첫 번째 방법을 "원확률 적용"으로 두 번째 방법을 "기준확률적용"이라 부르게 한다. 기준확률 적용의 경우 본 논문에서는 1996년도 자료에서 기준확률을 구하였기 때문에 기존의 자기회귀분석법은 물론 기타 다른 분석법과 직접 비교하는 것은 큰 의미가 없다는 것을 알아두기 바란다.

4.2. 모형과 예측

이 절에서는 앞에서 설명한 조건부 자기회귀분석법을 모형화하는 방법을 살펴보기로 하자. 먼저 모형은 다음과 같다.

$$\begin{cases} Y_t | (Z_t = 1) = \beta_0 + \sum_{j=0}^{29} \sum_{k=1}^{10} \beta_{kj} X_{kt-j} + \sum_{j=0}^{29} \sum_{k=1}^{10} \alpha_{kj} X_{kt-j}^2 + \epsilon_t \\ Y_t | (Z_t = 0) = 0 \end{cases} \quad (4.1)$$

여기서 Y_t 는 일별 송이버섯 생산량이고 X_{kt-j} 는 기후요인을 나타내는 독립변수이다. 그리고 오차항 ϵ_t 는 AR(p)모형을 따른다고 가정한다. 또한 Z_t 는 송이버섯 생산량이 있으면 "1"을 취하고 없으면 "0"을 취하는 베르누이 확률변수이며 $E(Z_t) = p_t$ 이다. 전술한 바와 같이 Z_t 또한 기후요인과 관계가 있으며 다음의 모형을 따른다고 가정한다.

$$Z_t = \beta_0^* + \sum_{j=0}^{29} \sum_{k=1}^{10} \beta_{kj}^* X_{kt-j} + \sum_{j=0}^{29} \sum_{k=1}^{10} \alpha_{kj}^* X_{kt-j}^2 + \epsilon_t^* \quad (4.2)$$

이때 모수 α_{kj}^* , β_{kj}^* 는 (4.1)에서 정의한 모수와 다르며 ϵ_t^* 는 독립이라 가정한다. 모형에 의한 예측은 조건부 기대값을 사용하여 다음과 같이 구해진다.

$$\begin{aligned} E(Y_t) &= E(E(Y_t | Z_t)) = E(Y_t | Z_t = 1) \cdot P(Z_t = 1) + E(Y_t | Z_t = 0) \cdot P(Z_t = 0) \\ &= E(Y_t | Z_t = 1) p_t \end{aligned} \quad (4.3)$$

주어진 모형에 관한 분석방법은 다음과 같다. 먼저 결측치 처리 전 모형을 살펴보자. 종속변수와 독립변수에 관한 내용은 전술한 바와 같다. 분석은 자기회귀분석법에서와 같이 Stepwise 변수선택법을 이용하여 독립변수를 선택하였고 SAS/PROC AUTOREG를 사용하여 모수 α_{kj} , β_{kj} 를 추정하며 이를 이용하여 적합값 $\hat{Y}_{t|Z_t=1}$ 을 구한다. 여기서 $\hat{Y}_{t|Z_t=1}$ 는 송이버섯 생산이 있다는 조건하에서의 일별 생산량의 크기에 대한 추정값이다. 다음으로 모형 (4.2)을 이용하여 같은 방법으로 송이버섯 생산이 있을 확률의 추정치 \hat{p}_t 를 구한다. 이제 (4.3) 식을 이용하면 송이버섯 생산량의 추정치를 \hat{Y}_t 라 할 때

$$\hat{Y}_t = \hat{E}(Y_t) = \hat{Y}_{t|Z_t=1} \hat{p}_t \quad (4.4)$$

가 된다. 다음으로 결측치 처리 후 모형의 분석은 결측치 처리 전 모형을 분석하는 방법과 일치하며 다만 사용되는 종속변수와 독립변수가 전술한 바와 같이 다를 뿐이다. 또한 원확률을 적용하는 분석은 (4.4)식을 사용하는 방법을 뜻하며 기준확률을 적용할 경우는 다음의 식을 사용하여 구한다.

$$\hat{Y}_t = \hat{E}(Y_t) = \hat{Y}_{t|Z_t=1} \hat{p}_t^* \quad (4.5)$$

여기서 \hat{p}_t^* 는 $\hat{p}_t \geq p_c$ 인 경우 $\hat{p}_t^* = \hat{p}_t$ 이 되고 $\hat{p}_t < p_c$ 인 경우 $\hat{p}_t^* = 0$ 이 되며 p_c 는 최적의 일별 MAPE 또는 일별 MSPE를 주는 확률을 의미한다. 다시 한번 언급하지만 p_c 의 값은 예측하려는 연도 전에 미리 정해져야 하나 4.1절에서 언급한대로 본 논문에서는 1996년의 자료의 예측에서 p_c 를 구했기 때문에 다른 분석법과의 비교는 큰 의미가 없을 수 있다. 그러나 1997년도 예측에는 사용이 가능하리라 생각된다.

4.3. 분석결과

표 4.1과 표 4.2는 모형 (4.1)에서 얻어진 결과로 표 4.1은 결측처리 전 결과이며 표 4.2는 결측처리 후 결과이다. 조건부 자기회귀분석법에서도 오차항의 차수는 자기회귀분석법과 같은 방법을 사용하였으며 전체적으로 조건부 자기회귀모형과 자기회귀모형의 차이점을 찾아보기 어렵다. 단지 모형 (4.1)의 결과가 자기회귀모형에서 얻어진 결과에 비해 선택된 변수의 수가 줄고 결정계수의 값도 약간 줄어들었다. 또한 결측처리 전과 후의 변화의 차이는 크게 나지 않는다.

표 4.1: 조건부 자기회귀모형에서 선택된 변수 수 및 결정계수(결측처리 전)

지역	사용 안된 독립변수	자기회귀 모형에서 선택된 변수 수	자기회귀 모형의 결정계수	사용된 생산량의 연도
강릉	없음	9	0.8896	90 ~ 95년
영덕	X_6, X_8, X_9, X_{10}	25	0.8884	90 ~ 95년 (94년*)
거창	$X_6, X_7, X_8, X_9, X_{10}$	6	0.8350	90 ~ 93 95년
청도	없음	22	0.9156	90 ~ 95년

*표가 붙어있는 해는 자료수가 적음

표 4.2: 조건부 자기회귀모형에서 선택된 변수 수 및 결정계수(결측처리 후)

지역	사용 안된 독립변수	자기회귀 모형에서 선택된 변수 수	자기회귀 모형의 결정계수	사용된 생산량의 연도
강릉	없음	16	0.8300	90 ~ 95년
영덕	X_6, X_8, X_9, X_{10}	4	0.8519	90 ~ 95년 (94년*)
거창	$X_6, X_7, X_8, X_9, X_{10}$	13	0.8863	90 ~ 93 95년
청도	없음	20	0.9039	90 ~ 95년

*표가 붙어있는 해는 자료수가 적음

다음절에서는 조건부 자기회귀모형에서 얻은 예측 결과를 MSPE, MAPE, MAE 그리고 MSE를 이용하여 자기회귀모형에서 얻은 예측결과와 비교하였다. 예측 결과에서 일별

MSPE, MAPE, MAE 그리고 MSE 등의 비교에 있어서는 송이 생산량 자체가 갖고 있는 여러 요인에 의하여 그 의미를 많이 축소하여야 하며 오히려 연별 MAPE, MSPE에 중점을 두고 비교하는 것이 타당할 것이다. 최적의 기준확률을 구하기 위하여 0.40부터 0.70까지 크기를 0.01로 하여 31개의 확률 중에서 가장 작은 MAPE와 MSPE를 주는 값을 기준 확률로 정하였으며(기준확률이 0.70인 경우에는 0.80까지의 확률을 더 구했음) \hat{p}_i 의 값이 기준 확률보다 작을 경우 \hat{p}_i 를 0으로 처리하였다.

4.3.1. 결측처리 전 조건부 자기회귀모형에 의한 예측결과

표 4.3와 표 4.4, 그림 4.1에서 그림 4.4가 결측처리를 하지 않았을 때의 조건부 자기회귀분석에서의 예측 결과이다. 원 확률을 적용한 그래프에서는 최고점이나 그에 가까운 날짜에서의 예측은 잘 되었으나 최고점에서 멀어지면서 예측의 정도가 떨어진다. 그러나 자기회귀모형의 예측 그래프와 비교하면 많은 향상이 있음을 알 수 있다. 이에 반해 기준확률을 적용한 예측 그래프는 구해진 원 확률을 그대로 사용하였을 때에 비해 실제 그래프와 상당히 많이 가까워졌으며 일별 MAPE와 MSPE도 많이 작아졌다. 그러나 일별 MSE 나 일별 MAE를 비교해보면 오히려 나쁜 결과를 주며 연별 MAPE, MSPE를 비교해 보아도 뚜렷하게 좋은 결과를 주지 않는다. 이는 어느 특정 기준값 이하를 모두 "0"으로 처리함으로써 실제 생산량에 비해 과소 예측됨으로써 생기는 오차에 의해서 발생하는 현상이라 볼 수 있다.

표 4.3: 결측처리 전 조건부 자기회귀모형에 의한 예측 결과(원 확률 적용)

지역	MAPE (일별)	MSPE (일별)	MSE (일별)	MAE (일별)	MAPE (연별)	MSPE (연별)
강릉	0.810019	4.176475	157.2563	2.369873	0.066000	0.004356
영덕	0.901349	3.585437	258.2548	1.634638	0.164435	0.027040
거창	0.580484	1.215444	0.884572	0.152570	0.096472	0.009307
청도	0.766655	1.478057	49.35368	1.110784	0.129577	0.016790

표 4.4: 결측처리 전 조건부 자기회귀모형에 의한 예측결과(기준확률적용)

지역	MAPE (일별)	MSPE (일별)	MSE (일별)	MAE (일별)	MAPE (연별)	MSPE (연별)	기준 확률
강릉	0.500701	0.344373	311.6746	3.681178	0.169152	0.028613	0.67
영덕	0.614774	0.877025	1680.017	8.738674	0.053152	0.002825	0.65
거창	0.488172	0.394666	403.5747	20.08917	0.054329	0.002952	0.70
청도	0.666195	0.971969	212.1349	4.391468	0.184459	0.034025	0.57

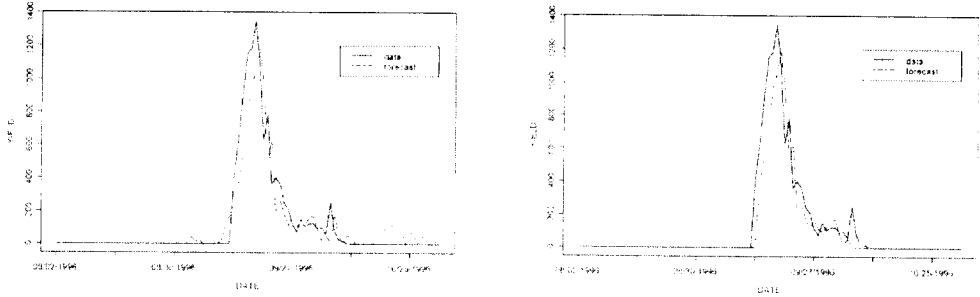


그림 4.1: 강릉 : 결측처리 전의 원확률 적용(왼쪽)과 기준확률 적용(오른쪽)

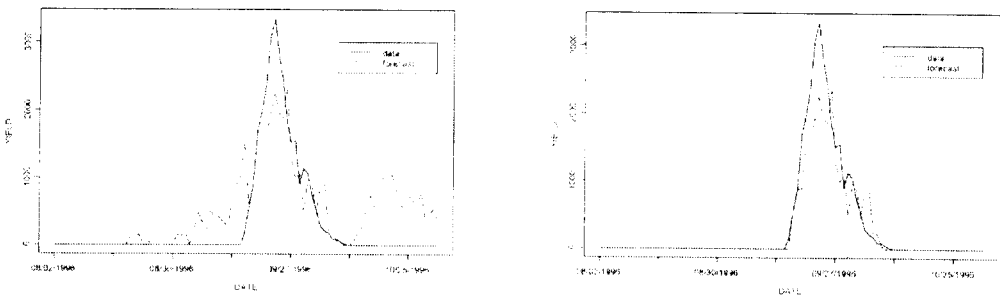


그림 4.2: 영덕 : 결측처리 전의 원확률 적용(왼쪽)과 기준확률 적용(오른쪽)

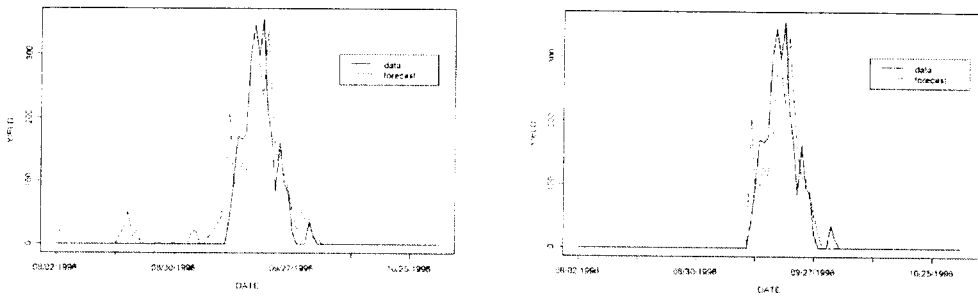


그림 4.3: 거창 : 결측처리 전의 원확률 적용(왼쪽)과 기준확률 적용(오른쪽)

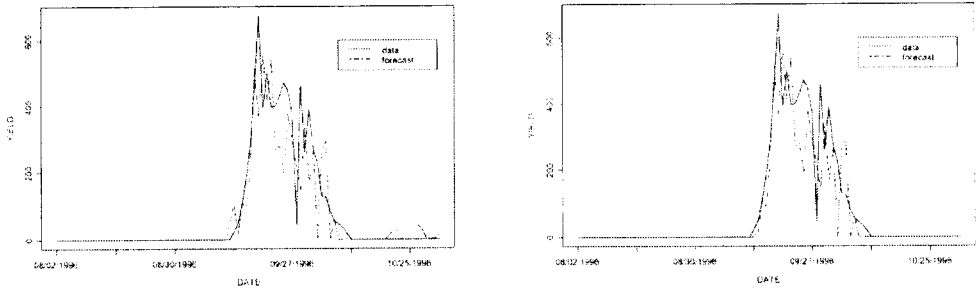


그림 4.4: 청도 : 결측처리 전의 원확률 적용(왼쪽)과 기준확률 적용(오른쪽)

4.3.2. 결측처리 후의 조건부 자기회귀모형에 의한 예측결과

전 절에서 언급하였듯이 전년도의 영향력을 제거하여 오로지 생산량과 기후조건과의 관계만을 알아보기 위해 “0”인 생산량의 뒤에 결측점을 찍어 그 영향력을 제거한 결측처리 후 결과가 표 4.5와 표 4.6, 그리고 그림 4.5 에서 그림 4.8이다. 결측처리 전과 결측처리 후의 결과를 비교해 보면 큰 차이를 보이고 있지 않으나 연별 MAPE와 MSPE를 비교하면 결측처리 전 모형의 연별 MAPE와 MSPE가 보다 안정적임을 알 수 있다.

표 4.5: 결측처리 후 조건부 자기회귀모형에 의한 예측결과표(원 확률 적용)

지역	MAPE (일별)	MSPE (일별)	MSE (일별)	MAE (일별)	MAPE (연별)	MSPE (연별)
강릉	0.688809	2.600634	76.83389	1.171339	0.103530	0.010718
영덕	1.393706	12.02255	2575.110	2.787787	0.062001	0.003844
거창	0.908587	2.681720	86.26765	1.450511	0.378666	0.143388
청도	0.751034	2.551727	131.0878	1.857331	0.112568	0.012672

표 4.6: 결측처리 후 조건부 자기회귀모형에 의한 예측결과표(기준확률 적용)

지역	MAPE (일별)	MSPE (일별)	MSE (일별)	MAE (일별)	MAPE (연별)	MSPE (연별)	기준 확률
강릉	0.454698	0.303887	147.6872	2.148304	0.206204	0.042520	0.67
영덕	0.653454	0.777628	2589.311	10.848777	0.183805	0.033784	0.65
거창	0.612668	0.513646	337.5526	4.456011	0.323497	0.104650	0.70
청도	0.506168	0.383140	260.6452	3.366363	0.144129	0.020773	0.57

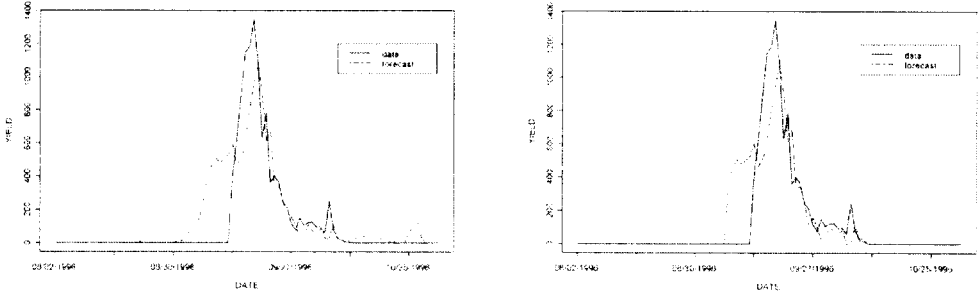


그림 4.5: 강릉 : 결측처리 후의 원확률 적용(왼쪽)과 기준확률 적용(오른쪽)

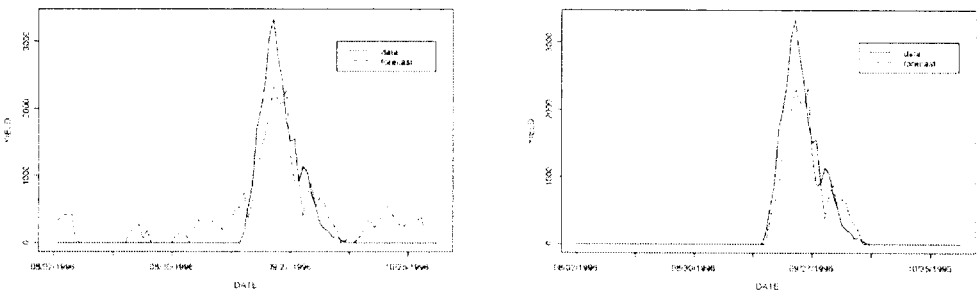


그림 4.6: 영덕 : 결측처리 후의 원확률 적용(왼쪽)과 기준확률 적용(오른쪽)

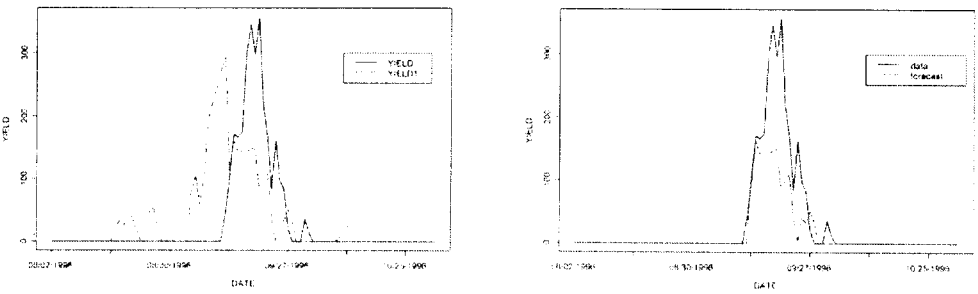


그림 4.7: 거창 : 결측처리 후의 원확률 적용(왼쪽)과 기준확률 적용(오른쪽)

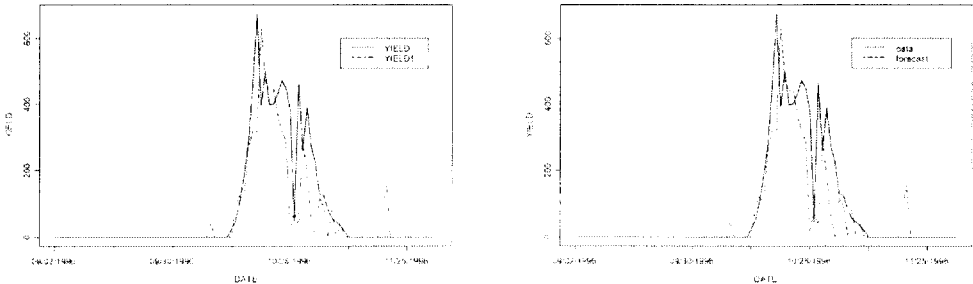


그림 4.8: 청도 : 결측처리 후의 원확률 적용(왼쪽)과 기준확률 적용(오른쪽)

5. 결론

송이버섯 생산량에 미치는 기후인자를 찾아내고 모형을 설정, 예측을 한 결과 위 지역(강릉, 영덕, 거창, 청도)에서 예측에 관한 통계량을 기준으로 하였을 때 자기회귀모형을 이용한 예측결과에 비해 조건부 자기회귀모형을 이용한 예측결과가 좋게 나왔다. 뿐만 아니라 예측 그래프의 비교에 있어서 훨씬 더 좋은 예측결과를 얻을 수 있었다. 기준확률을 적용할 경우 일별 MAPE, MSPE 등은 당연히 좋아졌으나 연별 MAPE, MSPE 등이 나빠지는 등 과소 예측되는 경향이 나타나므로 사용에 주의를 할 필요가 있으며 결측처리 후의 모형도 결측처리 전의 모형에 비해 크게 향상되었다고 볼 수는 없을 것이다.

조건부 자기회귀모형은 송이버섯이 생산될 확률을 근간으로 한다. 따라서 송이버섯이 생산될 확률의 정확한 추정이 매우 중요하다 하겠다. 이와 관련된 식 (4.2)에서 우리는 오차가 독립이라고 가정하였다. 그러나 이러한 가정은 약간의 무리가 따른다고 할 수 있으며 이에 관한 더 많은 연구가 필요하다고 하겠다.

참고문헌

- [1] 강안석, 차동렬, 김양섭, 박용환, 유창현 (1989). 송이(*Tricholoma matsutake*)생산과 관련된 기후특성 분석. 한국 균 학회지 17(2) : 51-56.
- [2] 박현, 가강현, 김교수, 김현중 (1996). 최근 6년간의 일자별 송이 발생동향 분석에 의한 송이 주산지 권역 구분. 산림과학 논문집 53 : 163-172.
- [3] 박현, 김교수, 구창덕 (1995). 한국에서 9월의 기후인자가 송이발생에 미치는 영향과 극복방안. 한국 임학회지 84(4) : 479-488.
- [4] 박현, 신기일, 김현중 (1998). 자기회귀 모형을 이용한 송이 생산 제한 기후인자 파악. 산림과학 논문집 57

- [5] 임업시험장 (1984). 송이 인공증식시험. 시험연구보고서. 산림청. p.939-1019.
- [6] 조덕현, 이경준 (1995). 29개 지역의 10년간 송이 발생림의 기상인자와 송이 발생량과의 상관관계. 한국 임학회지 84(3) : 277-285.

[1999년 2월 접수, 2000년 7월 채택]

Forecasting of Pine-Mushroom Yield Using the Conditional Autoregressive Model

Jin-Hee Lee¹⁾ Key-Il Shin²⁾

ABSTRACT

It has been studied to find relationships between pine-mushroom yield and climatic factors. Recently, Hyun-Park, Key-Il shin and Hyun-Joong Kim(1998) investigated relationships between pine-mushroom yield and climatic factors by autoregression model. In this paper, to improve the forecast we suggest the conditional autoregression model using probability of existing pine-mushroom production.

Keywords: Autoregression Analysis; Conditional Autoregression Analysis.

1) Graduate Student, Department of Statistics, Hankuk University of Foreign Studies.

E-mail: jhlee@stat.hufs.ac.kr

2) Associate Professor, Department of Statistics, Hankuk University of Foreign Studies.

E-mail: keyshin@stat.hufs.ac.kr