



제1회

농산물 수급 예측모형 경진대회

우수논문집



농림축산식품부

KREI

한국농촌경제연구원



한국농업경제학회

한국농업경제학회

우수논문집

Chapter 1 지정과제

★ 최우수상	① 생육지표와 기상정보를 이용한 양파 구중 예측	4
★ 우수상	② 가격 위험을 고려한 양파 재배면적 전망모형 개발	32
	③ SIMPLE 작물모형을 활용한 마늘·양파 단수 예측	56
★ 장려상	④ 기후변화가 마늘과 양파 주산지의 생산량에 미치는 영향 분석	88
	⑤ 덩러닝 기반 감성분석을 이용한 양파수급모형	106
	⑥ 실측기반 생육정보와 기상정보를 고려한 양파 생산 단수 예측 및 사회후생효과 추정	132
	⑦ 주산지 작황 실측 정보를 이용한 양파가격 예측모형 개발	156
	⑧ 단기 수급조정행위를 고려한 양파 수급예측모형 개발	194

Chapter 2 자유과제

★ 최우수상	① 고랭지 배추 단수예측 추정	218
★ 우수상	② 시설재배 오이 수확량 및 가격예측모형	246
	③ 생육 실측 자료와 비선형적 기후 효과를 반영한 배추 수급 예측모형 고도화 연구	274
★ 장려상	④ 영상 정보 기반 산지 출하량 정보와 도매시장 반입량 간의 관계를 이용한 고랭지 배추 출하량 산출 방법 연구	306
	⑤ 고랭지배추의 무게와 생육환경의 가변적 시차 관계분석	340
	⑥ 돼지고기 공급량 예측 모형 연구	362
	⑦ 조류인플루엔자의 영향을 고려한 계란가격 예측에 관한 연구	392
	⑧ 사육주기를 반영한 한육우 수급모형 구축	416



제1회

농산물 수급 예측모형 경진대회

우수논문집

Chapter 1
지정과제



최우수상

생육지표와 기상정보를 이용한 양파 구중 예측

김지수 · 정수연

생육지표와 기상정보를 이용한 양파 구중 예측

Forecasting Onion Bulb Weight using Growth Indicators and Meteorological Factors

김지수* 정수연*

Key words

양파(Onion), 생육지표(Growth indicators), 기상정보(Meteorological Factors), 다층모형(Multilevel model), 구중 예측(Forecasting Bulb Weight)

ABSTRACT

This study develops onion bulb weight forecasting model as a way to alleviate supply-demand imbalance. The actual measurement data in 2020/2021 are obtained from Korea Rural Economic Institute and a multilevel model is applied to reflect the hierarchical data characteristics consisting of growth indicators in individual units and meteorological factors in parcel units. The results are as below. The harvest timing should be considered when constructing the forecasting model since the weights are very different depending on that. The growth indicators, especially the number of leaves, at the early of bulb enlargement period and the number of optimum temperature days during the bulb enlargement period show the positive effects on the bulb weight at the harvest. The growth indicators at the early of bulb enlargement period can be forecasted by using those at the early of leaf elongation period. When applying the estimation results for 2020 data to 2021 data, the growth indicators at the early of bulb enlargement period can be explained as much as about 90% based on those at the early of leaf elongation period.

차례

- | | |
|----------|------------|
| 1. 서론 | 4. 분석 결과 |
| 2. 분석 모형 | 5. 요약 및 결론 |
| 3. 분석 자료 | |

* 충북대학교 농업경제학과 농식품 소비 및 소비자 행동 연구실
(지도교수: 김선웅, seonwoong@cbnu.ac.kr)

1. 서론

양파는 우리나라의 대표적인 양념채소 중 하나로, 주된 수확 시기는 5월말부터 6월 중순이며 저장 등을 통해 연중 꾸준히 소비된다. 더하여 연도별 소비량 역시 큰 변화 없이 안정적인 것으로 조사되고 있다. 예컨대 2010년 1인당 연간 양파 소비량은 약 28.6kg으로, 2020년에는 약 30kg 전후로 안정적인 소비세를 유지하고 있다(한국농촌경제연구원, 2021). 그러나 양파 생산량은 연도별로 큰 편차를 나타내고 있다. 실제로 통계청 농작물생산조사 자료를 이용하여 산출한 2010년에서 2019년 사이 양파 생산량의 연평균 변화율은 25.8%로 타 농산물 대비 큰 폭의 증감을 보였으며, 이에 따라 양파의 가격변동성 역시 크게 나타나고 있는 실정이다. 농산물유통정보(KAMIS)의 2018년에서 2020년 월별 양파가격(전국/상품/20kg)을 이용하여 연도별 가격변동성(annualized price volatility)을 산출한 결과, 12.04%(2018), 27.80%(2019), 34.28%(2020)로 각각 나타나 타 농산물 대비 매우 큰 폭의 가격변동성을 갖는 것으로 분석되었다¹⁾. 이러한 양파의 수급 불균형으로 인한 높은 가격변동성은 변동성을 위험으로 간주하는 농가들이 양파 생산을 축소하게 만드는 원인으로 작용할 수 있으며, 이는 농업 생산의 전반적인 위축과 사회적 후생 손실을 초래할 수 있다. 따라서 양파 수급 불균형 완화 방안이 필요한 시점이다.

현재 우리나라 양파 공급의 대부분은 국내 생산에 의존하고 있다. 한국농촌경제연구원에 따르면(2021) 2019년 국내 양파 공급량 1,644천 톤 중 국내 생산량은 1,594천 톤으로 약 96% 이상을 국내 생산이 담당하고 있는 것으로 나타났다. 전술(前述)한 상대적으로 안정적인 양파의 소비량과 높은 국내 생산 비중을 기준으로 볼 때, 국내 양파 공급량 조정이 선제적으로 이루어 질 수 있다면 일정 수준 이상의 양파 수급 안정을 달성할 수 있을 것으로 판단된다.

이를 위해 정부는 시장 공급량 조절을 통한 양파 가격 안정화를 목표로 산지폐기, 공공비축제도 등의 수급안정 대책을 마련하고 있다. 이중 산지폐기는 2018년과 2019년도에 시행되었으며, 과도한 공급 물량을 제한하여 안정적 시장 수급에 긍정적 영향을 미쳤다는 평가가 주

1) 동 기간 쌀의 해당 변동성은 6.92%(2018), 1.89%(2019), 9.02%(2020)였으며, 간마늘 역시 6.8%(2018), 18.2%(2019), 24.0%(2020)로 나타나 양파의 가격 변동이 타 작목에 비해 상대적으로 큰 것으로 나타났다.

를 이룬다. 그러나 일부 양파 생산 농가들은 좀 더 이른 시점에 해당 대책들이 시행되어야 함을 주장하는데 이를 세부적으로 살펴보면 생산량 추정 시기로 인해 발생하는 것을 알 수 있다. 통상 정부의 양파 수급안정 대책은 4월 말 통계청의 양파 재배면적 조사 결과와 한국농촌경제 연구원의 단수 예측치를 종합하여 총 생산량을 추정한 뒤 추정된 초과 수요량/공급량이 일정 수준 이상일 경우 5월 중순 이후, 즉 수확을 한 달가량 남겨둔 시점에서 발표 및 시행된다. 5월 중순은 양파의 구가 한참 비대해지는 시점으로 산지폐기 등의 수급안정 대책을 시행하기 위해서는 많은 정책 비용이 요구되며 농가 입장에서조차 작목으로의 전환 시기를 놓쳐 높은 기회비용 손실을 초래한다.

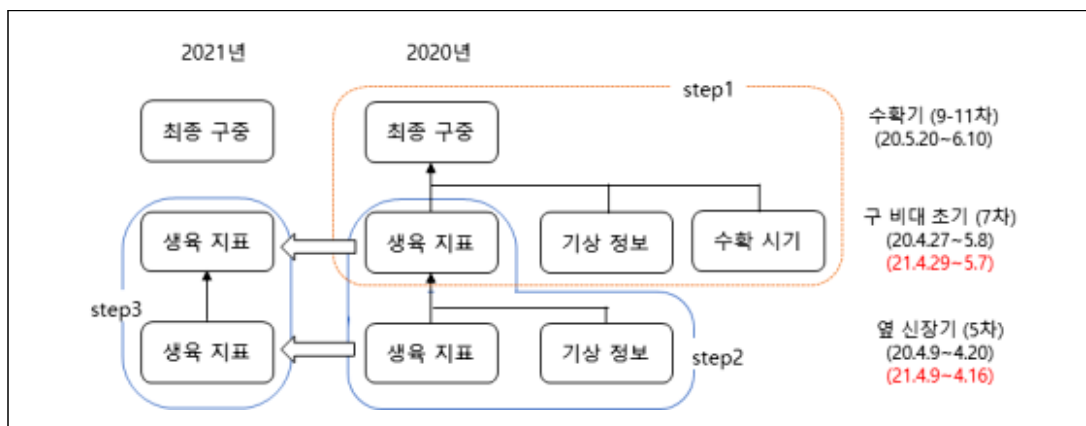
이러한 배경으로 양파 생산량을 예측하기 위한 많은 선행 연구들이 진행되어 왔다. 우선 공간 및 기상 요인을 통해 양파 단수를 추정한 오승원 외(2017)의 연구에서는 지역의 거리를 고려하여 공간 요인을 반영한 모형보다 기상 요인을 반영한 모형의 적합도가 더 높은 것을 확인하였으나, 외생 변수로 기상 요인만을 사용하였다는 한계점을 갖는다. 정인철(2019) 연구에서는 단계적 변수 선택법, LASSO(Least Absolute Shrinkage and Selection Operator: LASSO), 주성분분석(Primary Component Analysis: PCA), 부분최소제곱법(Partial Least Squares: PLS)의 4가지 통계 기법을 통해 재배 면적과 생산 단수에 대한 모형의 예측력을 비교하고 주성분 분석 모형이 가장 높은 정확도를 보임을 밝혔다. 그러나 재배면적과 생산단수의 범위가 달라 정확한 지역별 예측이 어렵다는 한계를 보였다. 이승인 외(2020) 연구에서는 기상 요인, 영농 투입재 비용 요인 사이의 교호작용을 베이지안 네트워크 분석을 통해 밝혔다. 다만 비료비, 농약비 등 영농 투입재 비용이 해당 필지의 데이터가 아닌 전국 평균 영농 투입재 비용 데이터를 사용하여 정확한 추정이 아니라는 한계를 가진다. 김준기 외(2020) 연구에서는 양파의 최종 구중(球重)을 추정하기 위해 생육 요인과 기상 요인을 활용한 다층(multilevel) 모형을 도입하였다. 과거 주로 사용했던 단층 구조 모형이 아닌 위계적 자료의 특성을 반영한 다층 구조 모형을 제시하여 생육 요인의 개체 수준과 기상 요인의 필지 수준을 모두 분석 단위로 받아들였다는 점에서 큰 차이점을 보인다. 그러나 분명히 존재하는 수확시기 차이에 따른 구중 변화를 통제할 수 있는 변수를 분석 모형에 포함하지 않았다는 한계가 존재한다. 강윤정 외(2021) 연구에서는 주성분분석과 부분최소제곱법을 이용하여 재배 기간 동안 양파의 구중에 유의한 영향을 주는 환경 요소를 도출하고, 재배 시 생육 관리에 관한 전략을 제시하였다.

다만 특정 지역의 데이터만을 사용하여 분석이 진행되어 해당 연구 결과를 일반화하기에는 어렵다는 한계가 존재한다.

이상의 과정과 2021년 양파·마늘 의무경작신고제²⁾ 시행을 근거로 본 연구의 목적을 양파의 안정적 수급을 위한 구중(球重) 예측모형 개발로 설정하였다. 간략한 연구 과정은 다음과 같다. 구중을 예측하기 위한 설명 변수로는 개별 양파의 생육지표와 조사 필지별 기상 자료를 활용한다. 이는 선행연구에서 제시한 바와 같이 양파의 단수 및 구중이 재배기간의 생육 요인 뿐만 아니라 기상 요인에도 유의한 영향을 받는다는 사실을 기반으로 설정하였으며, 특히 수확시기별 구중 차이를 통제하기 위한 변수를 도입하였다. 더하여 기상 요인이 필지별로 동일한 영향을 준다는 점을 고려하기 위하여 김준기 외(2020)와 같이 다층(multilevel)모형을 적용하였다. 분석 대상 품종은 생산량을 기준으로 중만생종을 선택하였다. 분석 절차는 다음과 같다. 우선 2020년 자료를 바탕으로 구 비대 초기(5월 초)의 생육지표, 수확시기 변수, 기상 자료를 이용하여 수확기 구중을 예측한다. 이후 엽 신장기(4월 초)의 생육지표를 바탕으로 구 비대 초기(5월 초)의 생육지표를 예측한다. 만일 해당 분석 결과의 정확도가 일정 수준 이상으로 실제 예측에 활용가능하다면 생산량 예측 시기를 한 달 가량 앞당김으로써 선제적인 수급 안정 대책의 수립에 도움이 될 수 있을 것으로 기대한다. 더하여 2020년 생육조사 자료에서 구성된 모형을 2021년 생육조사 자료에 그대로 적용하여 모형 구성의 타당성을 검증하고자 하였다. 현재 KREI에선 2021년 10차 생육조사 자료까지 공개하여 구비대 초기 지상부 추정만 실시하였고, 해당 단계에서는 기상정보를 제외하여 엽 신장기 하에서 지상부 추정이 가능하도록 하였다.

2) 2021년부터 국내 농산물 최초로 양파와 마늘에 도입된 의무경작신고제는 재배면적 1,000㎡ 이상의 농가가 재배품종과 재배면적, 파종·수확 예정시기 등을 의무적으로 신고해야 하는 제도이다. 해당 제도로 통해 과거에 비해 상대적으로 정확한 양파 재배면적을 사전적으로 알 수 있다.

그림 1. 본 연구의 분석 과정



2. 분석 모형

수확기 양파 구중은 품종, 생육지표 등 개별 요인에 더불어 온도, 햇빛, 토양 등 재배환경 요인에 의하여 영향을 받게 된다(농촌진흥청, 2018). 이는 동일한 필지(筆地)에서 재배되고 있는 양파의 구중을 정확하게 예측하기 위해서는 개별 양파의 생육지표에 추가적으로 동일 필지 내 양파 구중 간 상관관계를 고려해야 함을 의미한다. 현재 국내에서 양파 생육자료를 제공하고 있는 대표적인 기관은 한국농촌경제연구원과 농촌진흥청으로, 두 기관 모두 개별 양파의 구중을 필지별로 제공하고 있다. 이에 본 연구에서는 필지별 양파의 생육지표 간 상관관계를 고려할 수 있는 다층(multilevel)모형³⁾을 분석에 적용하였다.

다층모형(multilevel model)은 종속변수가 개체(수준-1)의 특성과 집단(수준-2)의 특성에 모두 영향을 받는 위계적(hierarchical) 자료를 분석하는데 특화된 모형이며, 시계열 데이터에 적용된 다층모형의 경우 성장곡선 모형(growth curve model)이라 구분되어 불리기도 한다. 본 연구의 분석 자료는 필지 단위로 구분되어 측정되며, 각 개체(수준-1)의 생육 변수와

3) 연구에 따라 다층모형은 위계적 선형모형(hierarchical linear model), 혼합모형(mixed model), 랜덤효과모형(random effect model)이라고도 불린다.

각 필지(수준-2)의 기상 변수 간 위계적 구조를 갖는다는 특징이 있다. 특히 동일한 필지에 속한 개체들은 기상 변수를 공유하므로 구중이 서로 독립적이지 않음을 가정하며 필지 간에는 서로 독립적임을 가정한다. 만일 다층구조의 자료에 전술한 종속성을 고려하지 않고 일반적인 단층모형으로 추정 시에는 오차분산 추정값이 과소추정(underestimated)되고, 1종 오류(type 1 error)가 증가하게 된다(강상진, 2016).

다층모형을 수준별로 나누어 설명하면 다음과 같다. 우선 수준-1에서 추정하고자 하는 대상은 각 필지의 양파 구중이며, 필지별 영향은 고려되지 않는다. 이는 일반적인 단층 회귀분석과 동일하며 아래 <식 1>과 같이 표현된다.

$$(1) \text{ 수준-1 모형: } Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \dots + \beta_{pj}X_{pj} + e_{ij}, \quad e_{ij} \sim N(0, \sigma^2) \\ i = 1, 2, \dots, n(j), \quad j = 1, 2, \dots, J.$$

여기서 Y_{ij} 는 필지 j 에 속한 개체 i 의 구중을, X_{ij} 는 엽 신장기 또는 구 비대 초기의 지상부 생육변수를, β_{pj} 는 필지 j 의 p 번째 생육변수의 회귀계수를, $e_{ij} \sim N(0, \sigma^2)$ 는 각 필지에서 잔차의 평균이 0이며 분산이 동일함을 의미한다.

위의 <식 1>에서 β_{pj} 는 필지 j 별로 다르게 추정될 수 있으며, 이는 필지별로 생육변수가 양파 구중에 미치는 영향이 서로 다를 수 있음을 의미한다. 예컨대 구 비대기의 초장이 수확기 최종 구중에 미치는 영향은 필지별로 다를 수 있다. 이는 아래 <식 2>와 같이 구성이 가능하다.

$$(2) \text{ 수준-2 모형: } \beta_{pj} = \gamma_{p0} + \gamma_{p1}W_{1j} + \gamma_{p2}W_{2j} + \dots + \gamma_{pQ}W_{Qj} + u_{pj} \\ p = 0, 1, 2, \dots, P.$$

여기서 W_{Qj} 는 필지 j 의 Q 번째 기상 요인을, γ_{pQ} 는 필지 j 에서 p 번째 생육지표의 회귀계수(β_{pj})에 대한 Q 번째 기상 요인의 효과를 반영하는 모수를 의미하며, 잔차들은 $(P+1) \times (P+1)$ 의 분산-공분산행렬을 갖는다.

이제 <식 2>를 <식 1>에 대입하면 단일식으로 표현되는 2-수준 다층모형의 일반식을 <식 3>과 같이 구할 수 있다.

(3) 일반화된 2-수준 다층모형:

$$Y_{ij} = \gamma_{00} + \sum_{q=1}^{Q_0} \gamma_{0q} W_{q(j)} + \sum_{p=1}^P \gamma_{p0} X_{Pij} \\ + \sum_{p=1}^P \sum_{q=1}^{Q(P)} \gamma_{pq} W_{q(p)} X_{\pi j} + u_{0j} + \sum_{p=1}^P u_{pj} X_{Pij} + e_{ij}$$

여기서 우변의 첫 항부터 순서대로 절편계수, 기상 요인(수준-2)의 고정효과, 생육지표(수준-1)의 고정 효과, 기상 요인과 생육지표간의 상호작용 고정효과를 나타내며, 마지막 세 개 항은 필지별 생육지표 효과를 기상 요인 변수로 설명한 이후 남은 수준-1과 수준-2 잔차의 합을 의미하는 무선효과를 나타낸다.

우선 다층모형을 이용한 분석에 앞서 해당 모형 적용의 타당성을 확보하기 위해 급내상관 계수(intraclass correlation coefficient: ICC)를 산출하고 유의성을 확인한다. 급내상관 계수는 관찰된 분산 전체에서 필지 간 분산이 차지하는 비율을 의미하며, 해당 수치가 유의하다는 것은 집단에 의해 설명되는 부분이 통계적으로 유의하다는 것이므로 단층모형이 아닌 다층모형을 적용하는 것이 타당함을 의미한다. 예컨대 ICC 값이 0.2로 산출되었고 통계적으로 유의하다면, 구중이 필지에 의해 설명되는 비율이 20%라고 해석 가능하다. ICC값은 위의 <식 3>에 의해 도출된 공분산 행렬을 바탕으로 산출되며 아래 <식 4>와 같이 표현될 수 있다.

$$(4) ICC = \frac{Var(\beta_{0j})}{Var(Y_{ij})} = \frac{\sigma_{field}^2}{\sigma_{field}^2 + \sigma_{error}^2}$$

여기서 σ_{field}^2 는 필지 간 공분산을, σ_{error}^2 는 잔차의 공분산을 의미한다.

3. 분석 자료

본 연구에서는 우선 생육지표로서 한국농촌경제연구원(KREI)에서 2020년 실시한 양파 생육조사 자료를 이용하였다. 해당 실측조사는 양파 표본농가 중 5년 이상 재배경력을 가지고

있는 농가들 중 평균 생육을 대표할 수 있는 106개의 필지를 선정하여 2020년 2월 15일부터 6월 11일까지 진행되었으며, 수확기로부터 상대적으로 먼 2~3월은 15일 주기로, 수확기가 근접한 4~6월은 10일 주기로 총 11회 진행되었다. 조사항목은 초장, 엽수, 엽초경, 엽초장 등 지상부 생육지표와 구중, 구직경 등 지하부 생육지표로 구성된다.

양파의 품종은 수확시기에 따라 크게 조생종과 중만생종으로 나누어 볼 수 있다. 통계청(2021)에 따르면 2021년 양파 생산면적은 약 18,014ha로, 이 중 중만생종 생산면적이 약 90%를 차지하는 것으로 나타났다. 정확한 구중 예측을 위해서는 가능한 품종을 세분화하는 것이 필요하므로 본 연구에서는 재배면적을 기준으로 중만생종 구중을 예측 대상으로 설정하였다. 이에 KREI에서 조사한 56개 필지 중 50개의 중·만생종 수확 필지를 선정하였고, 구중 측정치에 이상이 있다고 판단한 필지 8개를 제외하여 42개 필지, 총 563개 개체를 최종 대상으로 선정하였다.⁴⁾

양파의 생육 과정은 크게 유모기(8~10월), 활착기(11~12월), 월동기(12~2월), 경엽 신장기(2~4월), 구 비대기(4~6월)로 구분된다(농촌진흥청, 2018). 일반적으로 3월 중순부터 4월 중순까지 양파의 새 잎이 많이 나오는 시기인 엽 신장기로 불리고 있으며, 그 이후로는 새로운 잎이 나오는 경우는 드물다(이종태, 2015). 더하여 5월 초부터 중순까지는 잎의 신장과 함께 구가 비대해지는 초기로, 이 시기가 지나면 잎이 마르기 시작하는 것으로 알려져 있다. 이를 바탕으로 볼 때, 생육지표를 이용하여 구중을 직접 예측할 수 있고 해당 결과를 신뢰할 수 있는 가장 이른 시기는 5월 초로 판단되며, 해당 시기는 KREI 생육조사의 7회차(조사시기: 2020년 기준, 4월 27일~5월 8일)에 해당된다. 이에 본 연구에서는 구중을 예측하기 위한 자료로 KREI의 2020년 7차 생육조사 자료를 이용한다.

앞서 서론에서 제시한 바와 같이 통상 양파의 수급안정 대책은 수확을 한 달 정도 남겨둔 시점에서 계획이 수립 및 발표되며, 해당 시기에는 이미 구의 비대가 한창인 시점으로 대책 시행에 따른 정책 비용 및 농가 기회비용의 손실이 크다. 만약 해당 시기보다 좀 더 이른 시점에 양파 생산량 예측이 이루어진다면 양파 수급 안정이라는 목표를 달성하기가 좀 더 용이할 것으

4) 1차 분석대상 선정 기준은 모든 시점의 조사완료 여부, 표본교체 여부, 병해충 발생 여부 등이었으며, 해당 과정을 통해 확보된 577개의 자료 중 각 생육지표별 평균으로부터 ± 3 표준편차 구간을 벗어나 측정 및 코딩과정에 문제가 있었을 것으로 판단되는 14개의 자료를 제거한 후 분석을 수행하였다.

로 판단되며, 국가적 그리고 개인적 후생 손실 역시 일정 수준 감소할 수 있다. 이에 본 연구에서는 엽 신장기가 마무리 되어가는 시점인 KREI의 2020년 5회차(조사시기: 2020년 기준, 4월 9일~20일) 생육조사 자료를 이용하여 7차 생육지표들을 추정하고자 한다. 만약 5회차 자료를 이용하여 추정한 7회차 생육자료의 정확도가 일정 수준 이상이라면, 비록 7회차에 추정된 최종 단수에 비해 예측 정확도가 약간 낮더라도, 최종 단수 예측 시기를 약 한 달 정도 앞당길 수 있다.

더하여 양파 생육 전문가 대상 자문결과, 수확시기에 따라 구중이 크게 달라질 수 있으며, 특히 지역별로 약간씩 상이하나, 대략 5월 20일부터 6월 10일 사이에 최종 구중의 절반 정도가 증가한다고 밝혔다. 이는 KREI 생육조사의 수확기(9, 10, 11차)조사 시기와 거의 일치하므로 9차 조사를 기준으로 10차와 11차에 수확된 양파에 대한 더미변수를 통해 해당 효과를 통제하였다. 아래 <표 1>은 본 연구에서 사용된 생육지표 및 생산 시기 더미변수에 대한 기초 통계량이다.

표 1. 생육데이터 기초통계량

생육과정 (조사회차)	생육지표	N	Mean	Std. Dev.	Min	Max
엽 신장기 (5회차)	초장(cm)	563	65.80	12.05	27.00	100.60
	엽수(매)	563	7.82	1.29	4.00	14.00
	엽초장(cm)	563	9.26	3.06	2.10	19.80
	엽초경(mm)	563	18.72	4.07	5.00	29.63
구 비대 초기 (7회차)	초장(cm)	563	82.30	11.65	47.90	115.10
	엽수(매)	563	8.89	1.39	5.00	13.00
	엽초장(cm)	563	16.69	3.68	6.30	26.80
	엽초경(mm)	563	22.02	4.18	11.21	34.40
수확기 (9회차)	생구무게(g)	51	241.57	75.14	108.00	403.00
수확기 (10회차)		346	319.34	100.47	120.00	619.00
수확기 (11회차)		166	323.70	94.63	109.00	615.00

주: 잎마름병, 노균병, 무름병 등 병해 발생과 측정 초기 종료로 인한 생육 부진이 예상되는 필지 등을 제외함.

본 연구에서는 양파 구중 예측을 위해 생육지표와 생산시기 더미변수에 더하여 기상 자료를 예측변수로 적용하였다. 우선 필지의 생육 데이터와 기상 데이터의 불일치 문제를 해결하고자 국립농업과학원 소기후 모형에 의해 추정된 실측 농가별 기상 데이터를 사용하였다. 소기후 모형은 지역이 복잡하고, 소규모 농장이 많은 우리나라 농업 환경에서 재해 위험을 예측하고 대비하기 위한 모형으로, 우리나라 전역에 대해 필지 단위까지 정밀한 기상정보를 추정하도록 개발되었다(김대준 외, 2019). 소기후 모형에 적용된 기술은 아래 <표 2>와 같다.

표 2. 소기후 모형 적용기술 요약

기상 요인		적용기술
기온	최고	관측소별 최고기온 관측값을 IDW 방법으로 공간내삽한 후 동적 해발고도 보정, 일사효과를 적용(Kim et al., 2013)
	최저	관측소별 최저기온 관측값을 IDW 방법으로 공간내삽한 후 동적 해발고도 보정, 도시열섬효과, 온난대 보정, 냉기류 보정을 적용(Kim et al., 2012)
	평균	최고기온과 최저기온의 평균
강수량		PRISM(Parameter-elevation Regression on Independent Slopes Model) 모형을 이용(Kim and Yun, 2013)
일조시간		시간별 음영기복도를 이용하여 추정(Kim and Yun, 2016)
일사량		관측소별 일직선일사량 관측값을 IDW 방법으로 공간내삽한 후 수평면일사량의 지형보정치인 상대일사량지수 분포도(PRR, Potential Relative Radiation)를 이용하여 보정(Yun, 2009)

소기후 모형에서 제시된 실측 농가의 기상 요인들을 활용하여 분석에 필요할 것으로 판단되는 파생 기상 변수를 김준기 외(2020) 등을 참고하여 산출하였다. 파생 기상 변수 도출 근거가 되는 양파의 생리적 특성은 다음과 같다. 우선 양파 종자는 햇빛이 없는 조건에서 발아율이 높지만 구 비대기에는 일정 이상의 일조시간과 일사량이 요구된다. 일반적으로 중·만생종은 조생종과 비교하면 생육 과정에 요구되는 일사량이 많고, 어느 정도 높은 온도에서 구 비대가 이루어지며, 13~14.3시간의 일조시간이 필요하다(농촌진흥청, 2018). 더하여 김재휘 외(2021) 연구에 의하면 중만생종 구 비대에 요구되는 적당한 온도는 20℃ 내외인 것으로 나타나 이에 대한 고려가 필요하다. 엽초의 생육이나 구의 비대가 지속적으로 이루어지기 위해서는 탄소동화작용에 의해 탄수화물의 이동이 이루어져야 한다. 그러나 25℃ 이상의 고온에는 생육이 둔해지고, 고온이 계속되면 생육이 정지하게 된다. 양파는 구 비대가 시작되는 시기에 많은 물이 필요하며, 이 시기에 건조하면 구 비대가 정상적으로 이뤄지지 않아 수량이 현저히

감소한다. 구체적으로는 7~10일 간격으로 30~40mm의 강수량이 필요하다(농촌진흥청, 2018). 파생 기상 변수의 상세한 내용은 <표 3>과 같으며 강수량과 일사량은 누적값으로, 그 외 파생 기상 요인은 평균값으로 사용하였다.

표 3. 파생 기상 변수의 정의

파생 기상 변수	정의
생육도일(℃)	일평균 기온과 기본 기온와의 차를 합한 것. (평균기온 -4℃) 값의 합
최적기온일수(일)	최고기온이 20℃ 내외(15℃~25℃)인 날들의 수
고온일수(일)	최고기온이 25℃ 이상인 날들의 수
누적일사량(MJ/m ²)	일 일사량 누적
누적일조시간(시간)	일 일조시간 누적
누적강수량(mm)	일 강수량 누적
강우일수(일)	강수량이 5mm 이상인 날들의 수
수확 전 강수량	최종 조사 전 3일 간의 누적강수량

일반적으로 양파 수확은 토지가 충분히 건조된 상태에서 이루어진다. 건조되지 않은 상태에서 양파를 수확하면 줄기 절단면이나 구의 상처로 병원균이 침입하여 번식할 가능성이 커지기 때문에 기상 조건에 따라 수확 시기를 조절한다. 그러나 생육조사에서는 일정상의 한계로 수확 시기를 조절하는 것이 쉽지 않은 것으로 판단된다. 이에 최종 조사 전 3일간의 누적강수량을 설명변수로 도입하여 해당 효과를 통제하고자 하였다. 앞서 제시한 과정을 근거로 본 연구에서 채택한 기상 자료의 기초통계량은 <표 4>와 같다.

표 4. 기상 데이터 기초통계량

생육과정	기상 변수	Mean	Std. Dev.	Min	Max
엽 신장기 (5회차)	평균기온(℃)	12.34	1.62	9.65	15.99
	최고기온(℃)	18.66	2.24	14.66	21.73
	최저기온(℃)	5.51	1.49	2.25	8.42
	생육도일(℃)	166.99	39.14	109.70	238.17
	최적기온일수(일)	12.75	2.97	8.00	21.00
	고온일수(일)	2.37	1.91	0.00	5.00
	일사량(MJ/m ²)	21.11	1.35	17.20	23.15
	누적일사량(MJ/m ²)	421.08	76.12	151.52	552.10

생육과정	기상 변수	Mean	Std. Dev.	Min	Max
	누적일조시간(시간)	152.50	18.34	118.28	196.33
	강수량(mm)	2.49	1.14	0.84	5.18
	누적강수량(mm)	67.31	80.03	20.12	414.08
	강우일수(일)	2.73	0.96	1.00	5.00
	최적강수량(mm)	3.33	9.92	0.00	38.93
구 비대기 (7회차)	평균기온(°C)	17.79	0.76	16.61	19.76
	최고기온(°C)	23.79	1.28	21.48	26.45
	최저기온(°C)	11.80	0.94	9.13	13.41
	생육도일(°C)	430.97	84.66	290.01	627.11
	최적기온일수(일)	4.61	2.22	2.00	20.00
	고온일수(일)	11.38	6.50	0.00	24.00
	일사량(MJ/m ²)	20.23	1.01	17.51	21.58
	누적일사량(MJ/m ²)	630.62	112.57	444.30	877.61
	누적일조시간(시간)	192.81	35.81	131.07	264.18
	강수량(mm)	4.61	2.46	2.28	10.77
	누적강수량(mm)	138.98	67.55	73.03	295.16
	강우일수(일)	6.15	1.83	3.00	9.00
	최적강수량(mm)	40.06	56.27	0.00	212.33
	수확전강수량	7.41	9.48	0.00	31.85

4. 분석 결과

우선 2020년 KREI 생육조사 자료에 다층모형 적용의 타당성을 검증하였다. 이 때 모형의 종속변수는 최종 구중이고, 독립변수로 필지별 상수항에 대한 임의효과(random effect)만을 포함하며 이를 M1이라 칭한다. 도출된 ICC의 값은 1% 수준에서 통계적으로 유의한 것으로 나타나 개체 수준의 측정치가 필지라는 집단 속에 군집화(clustered)되어 있는 위계적 자료라는 것을 보여주며, 다층모형 적용의 타당성을 뒷받침하였다(표 5). 더하여 ICC가 27.995로 산출되어, 구중의 전체 분산 중 필지에 따라 설명되는 부분이 약 28%, 양파 개체간의 차이로 설명되는 부분이 약 72%임을 나타내었다.

표 5. 군집분석을 통한 다층모형 적용의 적합성 검정 결과

모형	ICC	Std. Err.	Chisq	Prob	[95% Conf. Interval]	
M1	27.995	0.055	108.52	0.0000	0.186	0.398

4.1. 구 비대 초기 생육지표와 기상 요인을 통한 수확기 구중 예측

본 연구에서 양파 구중 예측을 위한 다층모형의 설정은 Bethany et al.(2013)에서 제시한 과정을 바탕으로 수행되었다. 우선 모형 설정 시 설명변수로 고려된 각 수준별 요인들은 <표 6>과 같으며, 엽초장과 초장 간의 상관도를 고려하여 초장 중 엽초장의 비율인 ‘엽초장비율’ 변수를 생성하여 설명변수로 고려하였다.

표 6. 구중 예측모형 설정 시 고려된 수준별 설명변수

요인	예측 변수
생육지표 (수준-1)	초장 엽수 엽초장비율 엽초경
기상 요인 (수준-2)	평균기온 최고기온 최저기온 최적기온일수 고온일수 누적일사량 누적일조시간 누적강수량 강우일수 수확전강수량

본 연구의 모형 설정 과정을 간략히 요약하면 다음과 같다. 수확 시기에 따라 구조적으로 발생하는 구중의 차이를 모형에 반영하기 위해 앞서 제시한 M1에 10차, 11차 조사에 해당하는 수확시기 더미변수인 D10과 D11을 도입하였으며, 두 변수 모두 통계적 유의성을 확보하는 것으로 나타나 해당 변수들을 유지하고 이를 M2라 칭하였다. 여기서 특정 설명변수의 포함 여부는 AICc(Corrected Akaike Information Criterion)⁵⁾와 BIC(Bayesian Information Criterion)를 기준으로 결정하였으며, 정보기준지수(Information criterion Index)가 낮을수록 더 높은 모형 적합도를 가진다고 판단한다. 다음으로 통계적으로 유의한 생육지표만

5) AICc는 AIC가 상대적으로 적은 샘플에서 편향되는 것을 수정한 것으로, 두 정보기준의 차이는 추가적인 설명변수 도입에 대한 패널티 크기에 있으며, AIC보다 AICc에서 해당 패널티를 더 크게 부여한다. 단 샘플 크기가 증가하면 두 정보기준은 수렴하게 된다(SAS, 2017).

의 고정효과를 포함하고, 이를 M3라 칭한다. 이후 M3에 포함된 생육지표별 통계적으로 유의한 랜덤효과를 탐색한다. 만약 특정 생육지표에 랜덤효과가 통계적으로 유의하다면 이는 각 필지별로 해당 생육지표가 구중에 미치는 영향이 다르다는 것을 의미한다. 본 연구에서는 랜덤효과를 갖는 생육지표가 발견되지 않아 필지별로 영향을 주는 기상 요인의 고정효과를 다음 단계 모형(M4)에서 고려하였다. 이후 M4 다음 단계에서 고려될 수준-2 변수들(기상 요인)의 랜덤효과가 통계적으로 유의하지 않은 것으로 나타나 M4 모형을 최종 모형으로 선정하였으며, 각 단계별 최적으로 선정된 모형의 분석결과는 아래 <표 7>과 같다.

표 7. 구중 예측모형별 추정 결과

	M1	M2	M3	M4
Fixed Effects				
Intercept	313.295*** (8.849)	241.137*** (26.334)	-196.071*** (47.094)	-217.694*** (46.509)
D10		78.134*** (28.329)	78.712*** (26.250)	83.188*** (23.360)
D11		82.715*** (30.127)	86.415*** (27.750)	116.608*** (29.134)
초장(cm)			1.021** (0.433)	0.943** (0.421)
엽수(매)			17.717*** (3.225)	18.702*** (3.190)
엽초장비율(%)			2.063** (1.013)	1.886* (0.998)
엽초경(mm)			6.910*** (1.082)	6.838*** (1.060)
최적기온일수(일)				5.320** (2.699)
고온일수(일)				-11.274** (4.721)
Error Variance				
Level-1	7078.801*** (438.574)	7077.233*** (438.380)	5317.634*** (330.517)	5318.974*** (330.534)
Level-2 Intercept	2752.212*** (718.225)	2212.908*** (598.932)	1901.996*** (523.670)	1407.654*** (410.710)
ICC	27.995	23.820	26.345	20.927
Model Fit				
AICc	6671.079	6667.439	6519.106	6513.101
BIC	6676.078	6689.105	6558.106	6560.767

주: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$

최종 모형인 M4의 추정 결과를 살펴보면 다음과 같다. 우선 9차 생육조사 시점에 수확된 양파의 구중을 기준으로 다른 변수들의 값이 평균에 위치할 때, 10차, 11차 조사 시점에 수확된 양파는 각각 83.19g, 116.61g 정도의 무게가 더 나가는 것으로 분석되었다. 이는 본 연구에서 수행한 자문결과(대략 5월 20일부터 6월 10일 사이에 최종 구중의 절반 정도가 증가)를 일정 수준 뒷받침하는 근거가 되며, 구중 예측모형 구축 시 수확시기에 대한 통제가 반드시 필요함을 나타낸다. 더하여 생육지표인 초장, 엽수, 엽초장비율, 엽초경은 구중에 모두 모두 양(+)의 영향을 주는 것으로 나타났으며, 이는 일반적으로 예상되는 부호 방향과 일치한다고 볼 수 있다. 구체적으로는 초장, 엽수, 엽초장비율, 엽초경이 1단위 증가하면 최종 구중은 각각 0.94g, 18.70g, 1.89g, 6.84g 증가하는 것으로 추정되었다.

기상 요인 중 최적기온일수는 양(+)의 영향을, 반대로 고온일수는 음(-)의 영향을 주는 것으로 나타났으며, 세부적으로는 최적기온일수가 1일 증가하면 최종 구중은 5.32g 증가하고, 고온일수가 1일 증가하면 최종 구중은 11.27g 감소하는 것으로 추정되었다. 해당 기상 요인에 대한 추정 결과는 농촌진흥청(2018)에서 제시하고 있는 중만생종 양파의 최적 구 비대 온도가 20℃ 이내라는 점과, 25℃ 이상의 고온에서 탄소동화작용이 저해되어 탄수화물이 구로 내려가 축적되는 과정을 방해한다는 내용을 뒷받침하는 결과로 볼 수 있다.

4.2. 엽 신장기 생육지표와 기상 요인을 통한 구 비대 초기 지상부 예측

4.2.1. 엽 신장기 생육지표와 기상 요인을 통한 구 비대 초기 생육지표 예측

앞서 구 비대 초기의 생육지표와 기상 요인을 활용하여 최종 구중을 추정하였다. 본 절에서는 동일한 분석 방법을 이용하여 엽 신장기의 생육지표와 기상 요인을 바탕으로 구 비대 초기의 생육지표(초장, 엽수, 엽초장비율, 엽초경)를 추정하였다. 아래 <표 8>은 구 비대 초기의 생육지표 예측을 위해 고려된 엽 신장기의 생육지표와 기상 요인을 나타낸다.

표 8. 구 비대 초기 생육지표 예측모형 설정 시 고려된 수준별 설명변수

요인	예측 변수
생육지표 (수준-1)	초장 엽수 엽초장비율 엽초경
기상 요인 (수준-2)	평균기온 최고기온 최저기온 최적기온일수 고온일수 누적일사량 누적일조시간 누적강수량 강우일수 수확전강수량

우선 구 비대 초기의 각각의 생육지표를 종속변수로 두고 독립변수로 필지별 상수항에 대한 임의효과만을 포함하여 해당 생육지표 예측에 다층모형의 적용이 적합한지에 대하여 살펴 보았다. 분석 결과 구 비대 초기의 초장, 엽수, 엽초장비율, 엽초경을 종속변수로 구성한 모형의 ICC값은 각각 61.904, 38.378, 34.959, 40.032으로 산출되었으며(표 9), 모두 1% 수준에서 통계적으로 유의한 것으로 나타나 다층모형을 통한 분석이 필요함을 나타냈다.

표 9. 군집분석을 통한 다층모형 적용의 적합성 검정 결과(2020)

모형	ICC	Std. Err.	Chisq	Prob	[95% Conf. Interval]	
초장	61.904	0.056	173.92	0.000	0.505	0.721
엽수	38.378	0.060	51.82	0.000	0.275	0.506
엽초장비율	34.959	0.059	160.26	0.000	0.245	0.471
엽초경	40.032	0.061	86.32	0.000	0.289	0.523

아래 <표 10>은 각각의 구 비대 초기 생육지표별 최종 모형 분석 결과를 나타낸다. 구 비대 초기 초장은 엽 신장기의 초장이 1단위 증가할 때 0.56cm, 엽수가 1단위 증가할 때 0.93cm 증가하나 엽초장비율이 1단위 증가할 때에는 0.22cm 감소하는 것으로 분석되었다. 엽 신장기 동안의 최고기온과 강우일수의 증가는 구 비대 초기 초장에 양(+)의 영향을 주는 것으로 나타났는데 이는 해당 시기의 온도가 높을수록 지상부 생장이 빠르다는 기존의 연구와 일치하며, 4월 초 가뭄의 피해를 받지 않는다면 지상부 생장이 양호하다는 것을 의미한다.

구 비대 초기 엽수는 엽 신장기의 엽수가 1단위 증가할 때 0.31매, 엽초경이 1단위 증가할 때 0.09매 증가하는 것으로 나타났으나, 엽초장비율이 1단위 증가할 때 0.05매 감소하는 것으로 추정되었다. 엽 신장기 동안의 최고기온의 증가는 구 비대 초기 엽수에 양(+)의 영향을 주는 것으로 분석되었다. 이는 앞서 설명한 바와 같이 해당 시기 온도가 높을수록 지상부 생장이 빠르다는 것과 일치한다.

구 비대 초기 엽초장비율은 엽 신장기 초장이 1단위 증가할 때 0.04%, 엽초장비율이 1단위 증가할 때 0.57% 증가하는 것으로 나타났다. 이는 앞서 구 비대 초기 초장에 엽 신장기 엽초장비율이 미치는 영향과 상반되는 결과이다. 엽 신장기 동안의 최적기온일수의 증가는 구 비대 초기 엽초장비율에 양(+)의 영향을 주는 것으로 분석되었다.

구 비대 초기 엽초경은 엽 신장기 초장, 엽수, 엽초경이 1단위 증가할 때 각각 0.05mm, 0.45mm, 0.50mm 증가하는 것으로 분석되었으며, 엽 신장기 동안의 고온일수의 증가는 구 비대 초기 엽초경에 음(-)의 영향을 주는 것으로 나타났다. 이는 일정 온도 이상에서 고온에 의한 장애가 발생하여 생육이 둔해진다는 기존 연구 결과와 일치한다. 구 비대 초기 추정 모형 분석 결과를 통해 모든 지상부에서 온도와 관련된 기상 요인에 영향을 받는 것을 볼 수 있다.

표 10. 구 비대 초기 생육지표 예측모형 설정 시 고려된 수준별 설명변수

Fixed Effects		초장(cm)	엽수(매)	엽초장비율(%)	엽초경(mm)
Level-1	Intercept	21.996*** (6.857)	3.745*** (0.983)	6.490*** (1.837)	5.878*** (1.940)
	초장(cm)	0.558*** (0.039)		0.042*** (0.016)	0.051*** (0.019)
	엽수(매)	0.928*** (0.272)	0.312*** (0.048)		0.451*** (0.131)
	엽초장비율(%)	-0.219** (0.105)	-0.047*** (0.017)	0.565*** (0.046)	
	엽초경(mm)		0.091*** (0.017)		0.495*** (0.051)
Level-2	최고기온(°C)	0.673* (0.344)	0.089* (0.050)		
	강우일수(일)	2.567*** (0.792)			
	최적기온일수(일)			0.255** (0.101)	0.150 (0.113)
	고온일수(일)				-0.819*** (0.189)
Error Variance		Estimate	Estimate	Estimate	Estimate
Level-1		36.944*** (2.292)	0.952*** (0.059)	7.168*** (0.445)	7.099*** (0.441)
Level-2 Intercept		20.933*** (5.255)	0.447*** (0.115)	3.406*** (0.889)	4.271*** (1.071)

주: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$

4.2.2. 2020/2021 분석 결과 비교 및 예측 정확도 검정

2021년 6월 21일 현재 KREI에서는 10차 생육조사 자료까지 공개하고 있으므로, 2020년 생육조사 자료를 바탕으로 구성된 구 비대 초기 생육지표 예측모형을 2021년 생육조사 자료에 그대로 적용할 수 있다. 이에 본 절에서는 2020년 생육조사 자료를 이용하여 도출한 구 비대 초기 생육지표 예측모형의 회귀계수와 해당 모형을 2021년 생육조사 자료에 그대로 적용하였을 때 산출되는 회귀계수 사이에 통계적 유사성을 분석한다. 본 단계에서는 엽 신장기 동안의 기상정보를 제외하여 엽 신장기 초기에 구 비대기 초기 지상부 추정 가능성이 가능하도록 하였고, 이는 앞서 제시한 구중 예측모형의 M3와 유사한 형태이다.

전술한 분석들과 동일하게 ICC 도출 과정을 바탕으로 다층모형 적용의 적합성을 확인하였으며, 모든 모형에서 1% 유의수준에서 다층모형의 적용이 타당함을 보였다(표 11).

표 11. 군집분석을 통한 다층모형 적용의 적합성 검정 결과(2021)

모형	ICC	Std. Err.	Chisq	Prob	[95% Conf. Interval]	
초장	55.583	0.078	173.92	0.000	0.402	0.700
엽수	27.869	0.075	51.82	0.000	0.156	0.446
엽초장비율	40.682	0.083	160.26	0.000	0.259	0.573
엽초경	40.682	0.083	86.32	0.000	0.259	0.573

아래 <표 12~15>와 <그림 2>는 2020년 구 비대 초기 생육지표 예측모형의 분석 결과와 해당 모형을 2021년 자료에 그대로 적용하였을 경우의 분석 결과를 나타낸다. 구 비대 초기(7차 조사)의 엽수 예측모형에는 엽 신장기(5차 조사)의 엽수, 엽초장비율, 엽초경이 수준-1의 설명변수로 사용되었다. 2020년, 2021년의 회귀계수 비교 결과, 모든 회귀계수의 부호가 동일하며 그 크기 또한 5% 유의수준에서 통계적으로 무차별한 것으로 나타났다. 다만 2021년의 엽초장비율은 5% 수준에서 통계적 유의성을 확보하지 못하지 못하는 것으로 분석되었다. 엽 신장기의 초장, 엽수, 엽초장비율이 수준-1의 설명변수로 사용된 구 비대 초기의 초장 예측모형의 경우 역시 설명변수별 회귀계수가 5% 유의수준에서 통계적으로 무차별한 것으로 분석되었다. 다만 2021년 엽수와 엽초장비율이 구 비대 초기의 초장에 미치는 영향은 5% 수

준에서 통계적으로 유의하지 않은 것으로 나타났다. 구 비대 초기 엽초장비율 예측모형에는 엽 신장기의 초장과 엽초장비율이 수준-1의 설명변수로 사용되었으며, 2020년, 2021년의 회귀계수 비교 결과, 모든 회귀계수의 부호 및 크기가 5% 유의 수준에서 무차별한 것으로 나타났다. 끝으로 엽 신장기의 초장, 엽수, 엽초경이 수준-1의 설명변수로 사용된 구 비대 초기 엽초경 예측모형에서는 각 설명변수별 연도 간 회귀계수들의 부호가 동일하며 그 크기 또한 5% 유의수준에서 무차별 한 것으로 나타났다. 다만 2021년 구 비대 초기의 엽초경에 엽 신장기의 엽수가 미치는 영향은 5% 유의수준에서 통계적으로 유의하지 않은 것으로 분석되었다.

이상의 분석 결과를 종합하면 다음과 같다. 엽 신장기 생육지표를 바탕으로 구성된 구 비대 초기 생육변수 예측모형의 2020년과 2021년의 회귀계수들은 5% 유의수준에서 통계적으로 무차별한 것으로 나타나, 각각의 설명변수가 종속변수에 미치는 영향이 유사한 것으로 판단된다. 특히 구중 결정에 중요한 영향을 미치는 것으로 알려져 있는 엽수의 경우(University of New Hampshire, 2020), 2020년과 2021년 회귀계수 간 크기가 매우 유사한 것으로 나타나 해당 설명변수들을 예측에 활용하는 것이 타당함을 입증하였다. 이상의 결과는 본 연구의 설명변수 선정이 일정 수준 적합하였음을 의미하며, 특히 KREI 실측조사 자료의 신뢰성을 상당 부분 뒷받침하는 것으로 볼 수 있다.

표 12. 2020/2021 구 비대 초기 엽수 예측모형 분석 결과

		2020		2021	
Fixed Effects		Estimate	Std. Err.	Estimate	Std. Err.
Level-1	Intercept	5.340***	0.397	5.281***	0.489
	엽수(매)	0.323***	0.047	0.323***	0.068
	엽초장비율(%)	-0.046***	0.017	-0.014	0.022
	엽초경(mm)	0.089***	0.017	0.060**	0.024
Error Variance		Estimate	Std. Err.	Estimate	Std. Err.
Level-1		0.952***	0.059	0.795***	0.067
Level-2 Intercept		0.482***	0.123	0.245***	0.098

주: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$

표 13. 2020/2021 구 비대 초기 초장 예측모형 분석 결과

		2020		2021	
Fixed Effects		Estimate	Std. Err.	Estimate	Std. Err.
Level-1	Intercept	39.916***	3.133	45.692***	0.053
	초장(cm)	0.574***	0.039	0.513***	4.075
	엽수(매)	0.952***	0.274	0.444	0.492
	엽초장비율(%)	-0.186*	0.106	-0.163	0.162
Error Variance		Estimate	Std. Err.	Estimate	Std. Err.
Level-1		36.949***	2.292	42.933***	3.606
Level-2 Intercept		28.846***	7.009	21.438***	7.641

주: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$

표 14. 2020/2021 구 비대 초기 엽초장비율 예측모형 분석 결과

		2020		2021	
Fixed Effects		Estimate	Std. Err.	Estimate	Std. Err.
Level-1	Intercept	9.824***	1.230	11.680***	1.299
	초장(cm)	0.043***	0.016	0.048***	0.015
	엽초장비율(%)	0.556***	0.046	0.408***	0.054
Error Variance		Estimate	Std. Err.	Estimate	Std. Err.
Level-1		7.168***	0.445	4.838***	0.404
Level-2 Intercept		4.010***	1.025	2.899***	0.952

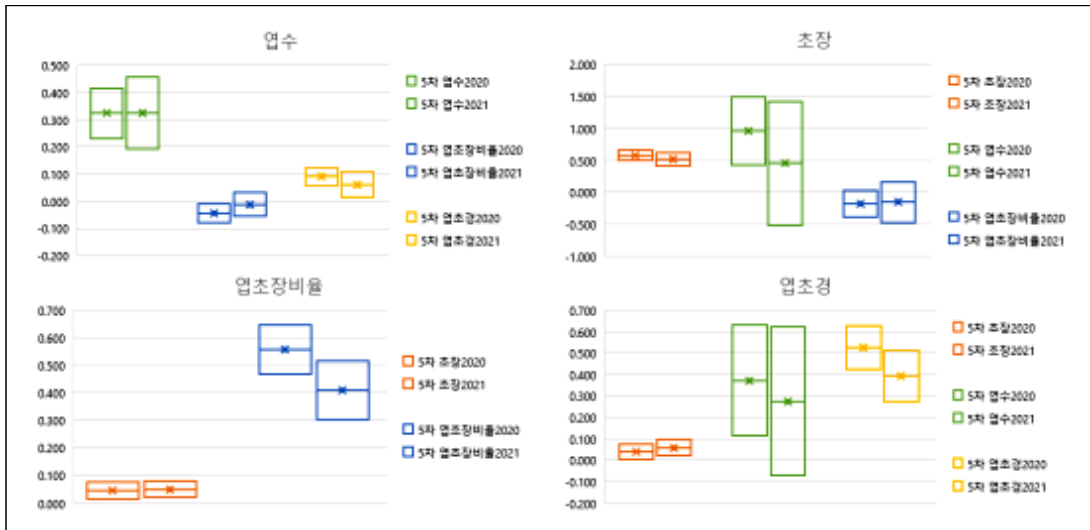
주: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$

표 15. 2020/2021 구 비대 초기 엽초경 예측모형 분석 결과

		2020		2021	
Fixed Effects		Estimate	Std. Err.	Estimate	Std. Err.
Level-1	Intercept	6.864***	1.300	7.230***	1.277
	초장(cm)	0.037*	0.019	0.055***	0.019
	엽수(매)	0.371**	0.131	0.273	0.177
	엽초경(mm)	0.524***	0.051	0.392***	0.061
Error Variance		Estimate	Std. Err.	Estimate	Std. Err.
Level-1		7.106***	0.441	4.531***	0.381
Level-2 Intercept		6.423***	1.567	4.892***	1.588

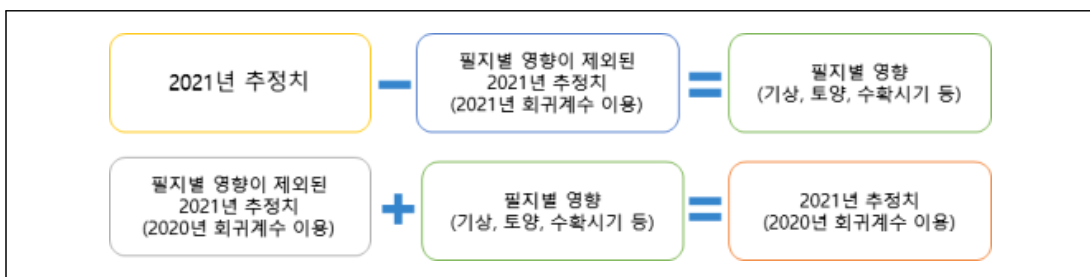
주: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$

그림 2. 2020/2021 구 비대 초기 생육지표 예측모형 분석 결과



마지막으로 구 비대 초기 생육지표 예측모형의 정확도를 확인하기 위해 2020년 회귀계수를 이용하여 2021년 추정치를 도출하고, 이를 2021년 실측치와 비교하여 평균절대비율오차 (Mean Absolute Percentage Error, MAPE)를 계산하였다. 2020년 회귀계수를 이용한 2021년 추정치 산출과정은 아래 <그림 3>과 같다.

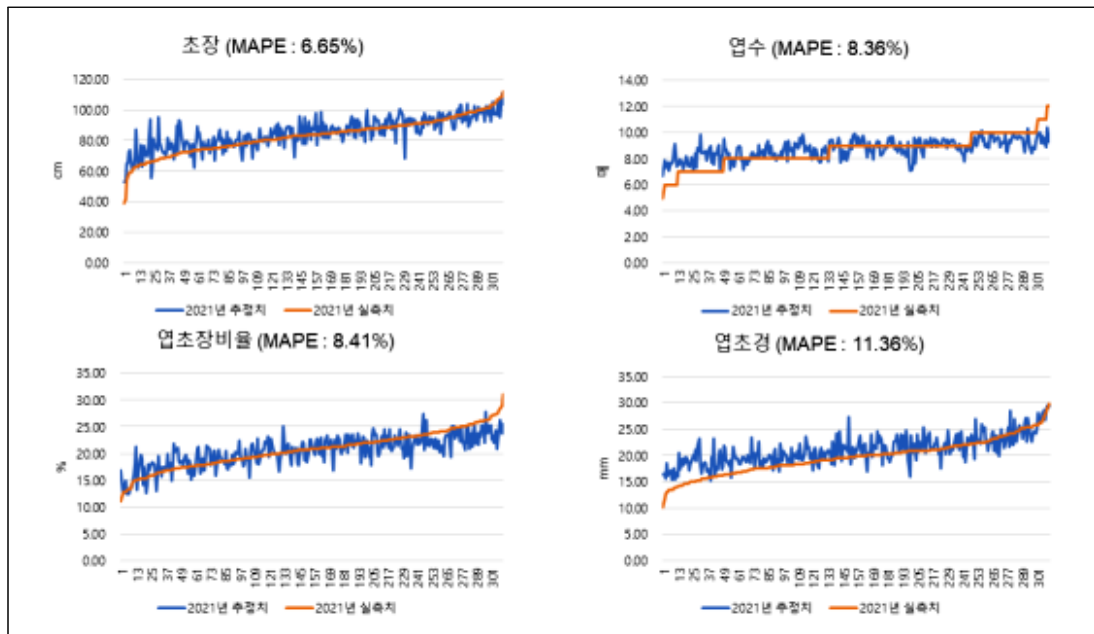
그림 3. 2021년 필지별 영향 도출 과정



구체적으로는 2021년 추정치는 2020년 자료를 이용하여 구성된 예측모형으로부터 산출되며, 해당 추정치에서 생육지표로 설명되는 부분을 제외하여 각 필지별 영향을 도출한다. 여기에 2021년 생육지표에 2020년 회귀계수를 적용하여 확보한 2021년 생육지표의 영향을

도출한 뒤 전 단계에서 산출한 필지별 영향을 합산하여 2020년 회귀계수를 이용한 2021년 추정치를 도출한다. 이를 2021년의 실측치와 비교하여 MAPE를 계산하였으며, <그림 4>는 해당 결과를 나타낸다. 분석 결과, 초장, 엽수, 엽초장비율, 엽초경에 대한 MAPE는 각각 6.65%, 8.36%, 8.41%, 11.36%로 나타나, 상대적으로 높은 예측 정확도를 보이는 것으로 분석되었다. 이는 엽 신장기 생육지표를 이용하여 구 비대 초기 생육지표를 일정 수준 이상 예측할 수 있다는 것을 의미하며, 구중 예측 시기를 현행보다 좀 더 앞당길 수 있다는 가능성을 시사한다.

그림 4. 2021년 구 비대 초기 생육지표 추정치와 실측치 비교 결과



5. 요약 및 결론

본 연구는 양파의 생육지표와 기상정보를 이용한 구중 예측모형 구축을 목적으로 수행되었다. 우선 구 비대 초기(5월 초)의 생육지표, 수확시기, 기상자료에 다층모형을 적용하여 구중

예측모형을 구성하였다. 분석 결과, 5월 20일과 6월 10일 사이에서는 수확시기가 늦어짐에 따라 양파 구중이 크게 증가하는 것으로 나타나 향후 생육자료를 이용한 구중 예측모형 구성 시, 수확시기 통제가 반드시 고려되어야하는 사항임을 나타냈다. 더하여 구 비대 초기의 생육 지표인 초장, 엽수, 엽초장비율, 엽초경은 모두 최종 구중에 양(+)의 관계를 갖는 것으로 나타났으며, 기상 요인 중 최적기온일수는 구중에 양(+)의 영향을, 반대로 고온일수는 음(-)의 영향을 미치는 것으로 분석되어 기존의 연구 결과와 부합하는 것으로 나타났다.

보다 앞선 구중 예측을 위한 사전 단계로 엽 신장기(4월 초)의 생육지표와 기상정보를 바탕으로 구 비대 초기(5월 초)의 생육지표 예측모형을 구성하여 구 비대 초기에 영향을 미치는 생육지표와 기상 요인을 도출하였다. 이후 해당 결과의 타당성을 확보하기 위하여 2020년 구 비대 초기 생육지표 예측모형을 2021년 자료에 그대로 적용하여 연도별 회귀계수를 비교하였다. 분석 결과 각 설명 변수의 연도별 회귀계수의 크기가 통계적으로 무차별한 것으로 나타났으며 특히 양파 구중 결정에 주요 요인으로 알려져 있는 엽수의 경우 각의 설명변수의 회귀계수의 부호 및 크기가 매우 유사한 것으로 분석되었다. 더하여 본 구성된 예측 모형의 정확도를 검증하기 위해 2020년 자료를 바탕으로 구성된 예측모형에 2021년에 자료를 적용한 결과, 약 90%정도의 설명력을 갖는 것으로 나타났다.

이상의 연구 결과를 2021년부터 시행되고 있는 양파 의무경작신고제 면적에 적용한다면 보다 정확한 양파 생산량을 예측할 수 있을 것이라 판단되며, 본 연구가 미약하게나마 안정적인 양파 수급 달성에 기여할 수 있기를 바란다. 본 연구의 한계점은 다음과 같다. 우선 예측 대상을 중만생종 양파의 구중으로 한정하여 시장 물량의 일정 수준을 담당하고 있는 조생종 양파에 대하여 다루지 못하였으며, 자료의 한계로 인해 중요 구중 결정 요인 중 하나인 토양에 대한 정보를 반영하지 못하였다. 이에 해당 한계는 차후의 연구 주제로 남긴다.

참고문헌

- 강운정, 나명환, 조완현, 고현석. 2021. “양파의 생구무게 예측을 위한 여러 가지 일반회귀모형의 성능 비교.” 『한국자료분석학회』 23(1): 135-148.
- 김대준, 김수옥, 김진희, 윤은정. 2019. “농장맞춤형 농업기상정보 생산을 위한 소기후 모형 구축.” 『한국농림기상학회지』 21(3): 146-157.
- 김대준, 김진희, 노재환, 윤진일. 2012. “신 기후변화시나리오에 따른 한반도 내 겨울보리 재배적지 이동.” 『한국농림기상학회지』 14(4): 161-169.
- 김대준, 윤진일. 2013. “기상청 동네예보의 영농활용도 증진을 위한 방안: 2. 강수량 분포 상세화.” 『한국농림기상학회지』 15(3): 171-177.
- 김승호, 윤진일. 2016. “하늘상태와 음영기복도에 근거한 복잡지형의 일조시간 분포 상세화.” 『한국농림기상학회지』 18(4): 233-241.
- 김재휘, 최성천, 김준기, 서홍석. 2021. “양파의 생육시기별 생육요인과 기상요인 간의 관계 탐색.” 『한국농림기상학회지』 23(1): 1-14.
- 김준기, 최성천, 김재휘, 서홍석. 2020. “다층모형을 활용한 양파 구중 추정 연구.” 『한국통계학회』 33(6): 763-776.
- 오승원, 김민수. 2017. “공간 및 기상정보 시계열 모형을 이용한 양파 생산량 예측.” 『한국자료분석학회』 19(5): 2447-2456.
- 윤진일. 2009. “지형보정계수를 이용한 복잡지형의 일 적산일사량 분포 추정.” 『한국농림기상학회지』 11(1): 13-18.
- 이승인, 손찬수, 이해림. 2020. “기상 및 영농투입재 비용이 양파 단수에 미치는 영향: 패널회귀분석과 베이지안네트워크모형을 중심으로.” 『농촌경제』 43(2): 1-28.
- 정인철. 2019. “양파 재배면적과 생산단수의 예측 모형 비교 Stepwise, LASSO, PCA, PLS 방법의 비교.” 전남대학교 이학석사학위논문.
- Bell, B., M. Ene, W. Smiley, J. Schoeneberger. 2013. “A Multilevel Model Primer Using SAS.” University of South Carolina.
- Emma Erler. 2020. “How do I get my onions to grow bigger?” University of New Hampshire.
- 통계청. 2021. 「2021년 마늘, 양파 재배면적조사 결과」.
- 한국농촌경제연구원. 2021. 「농업전망 2021」.
- 강상진. 2016. 다층모형. 학지사. 41-104.
- 농촌진흥청. 2018. 농업기술길잡이 양파. 내일날씨. 13-14.
- 백영민. 2018. R을 이용한 다층모형. 한나래. 83-86.
- 이종태. 2015. 양파: 한살이부터 유기농, 무경운 재배법까지. 들녘. 430-431.
- 농림축산식품부. <<https://www.mafra.go.kr/>>. 접속일: 2021. 5. 30.

- 농사로. <<https://www.nongsaro.go.kr/>>. 접속일: 2021. 5. 30.
- 농산물 유통정보(KAMIS). <https://www.kamis.or.kr>. 접속일: 2021. 6. 8.
- 농촌진흥청 국립농업과학원. <<http://www.naas.go.kr/>>. 접속일: 2021. 5. 24.
- 농촌진흥청. 농작업일정. <<http://www.nongsaro.go.kr/portal/ps/psb/psbl/workScheduleDtl.mo?menuId=PS00087&cntntsNo=30633>>. 접속일: 2021. 6. 7.
- 통계청. <<https://kostat.go.kr/>>. 접속일: 2021. 6. 8.
- 한국농촌경제연구원 농업관측 통계정보시스템. <<https://oasis.krei.re.kr/>>. 접속일: 2021. 6. 5.



제1회

농산물 수급 예측모형 경진대회

우수논문집

Chapter 1
지정과제



우수상

가격 위험을 고려한 양파 재배면적 전망모형 개발

안병일

가격 위험을 고려한 양파 재배면적 전망모형 개발 : EV 모형과 MGARCH 모형을 중심으로

안병일*

Key words

가격 위험(price risk), 변이계수(coefficient of variation), MGARCH 모형(Multivariate Generalized AutoRegressive Conditional Heteroskedasticity model), EV 모형(Expected Variance model)

ABSTRACT

Although the price and price volatility of an agricultural product are the most important factors to be considered in forecasting the cultivated area, previous studies that investigated the determinants of the cultivated area have not analyzed the dynamic relationship between price volatility and the cultivated area. In this study, in order to analyze the effect of price and price risk on the cultivated area, the expected-variance model (EV-Model), which is derived from the farmer's utility maximization problem and can be empirically analyzed, is applied. In addition, to compensate for the limitations of the EV model, a generalized autoregressive conditional heteroskedasticity time series model is estimated for reflecting the dynamic correlation between the cultivated area and price as well as price volatility. As a result of forecasting the cultivated area of onion using the analysis results of this study, the error between the actual cultivation area and the forecast for 2021 is estimated to be 5.07% when only the EV model estimation result is used. In the case of taking into account the changing trend the dynamic correlation coefficient derived from the MGARCH model, the error is calculated to be 4.82%, which implies that calculating the forecast by reflecting the dynamic correlation is a better approach that reduces the error. The area of onion in 2022 is estimated to reach 22,197ha when the EV model is used and 22,123ha when both the EV model and the MGARCH model are used.

차례

- | | |
|--------------------|------------|
| 1. 서론 | 4. 분석 결과 |
| 2. 이론적 배경 | 5. 요약 및 결론 |
| 3. 분석 자료 및 사전 분석결과 | |

* 고려대학교 식품자원경제학과 교수

1. 서론

농산물 재배면적은 생산자들의 가격 예측 여하에 따라 크게 변동하고 있는데, 이러한 재배면적 변동은 양파의 경우에도 예외는 아니다. 또한, 재배면적 결정은 생산자의 가격 예측에 의해 영향을 받기도 하지만, 동시에 한번 결정된 재배면적은 생산량과 직결되어 당년도의 가격에도 직접적인 영향을 주고 있다. 농산물 가격의 불안정성(변동성)은 농가 수입의 변동성으로 귀결되기 때문에, 농업경영 위험의 가장 대표적인 것이라 할 수 있다. 따라서 재배면적 결정에는 농산물 가격뿐만 아니라 가격의 변동성도 중요한 영향을 미치고 있다.

이와 같은 측면을 감안 한다면 재배면적을 전망하는 데에는 해당 품목의 가격과 가격 변동성이 가장 중요하게 고려되어야 할 요인들이라는 것을 알 수 있다. 하지만, 그간 재배면적을 전망하는 국내의 선행연구들에서는 가격수준 자체 만에 초점을 맞추어 왔지, 가격의 변동성이 미치는 동태적인 영향은 상대적으로 소홀히 다루어 왔다.

재배면적 결정요인을 연구한 대표적인 몇몇 선행연구를 보면, 예를 들어, 박지연, 박영구(2013)의 연구에서는 배추와 무의 재배면적을 추정하여 전년도의 동일 품목의 해당 작형과 금년도의 전작형 가격이 재배면적에 유의한 영향을 미치는 것을 밝혔으며, 홍승지 외(2015)에서는 시계열모형인 오차수정모형을 이용하여 고추의 재배면적 반응함수를 추정하였는데, 이들 연구에서는 가격과 재배면적 간의 관계에만 초점을 맞추어 연구를 진행하였다. 한편 정학균 외(2009) 및 조재환(2021)의 연구에서는 가격 변동성이 재배면적을 결정하는 중요한 설명변수로 삽입하여 연구를 진행해 가격 위험이 재배면적에 미치는 영향이 실존함을 입증하였다. 하지만 이들 연구에서는 가격변동성과 재배면적 간에 나타나는 동태적인 연관 관계까지는 분석하지 못하였다.

본 연구에서는 가격과 가격 위험이 재배면적에 미치는 영향을 분석하기 위해 농가들의 효용 극대화 문제로부터 도출된 것으로 실증분석이 가능한 평균-분산모형(Expected-Variance Model: EV-Model)을 적용하고자 한다. 또한, EV 모형이 갖는 한계를 보완하기 위해 재배면적과 가격, 재배면적과 가격 변동성을 동시에 고려하는 다변량 조건부 이분산 시계열 모형을 추정하여 재배면적과 가격 및 가격 변동성 간에 존재하는 동태적인 상관관계를 추정하고자

한다. 이와 같은 상관관계 추정을 통해 EV 모형에서 산출된 전망치가 어떻게 개선될 수 있는지도 보여주고자 한다.

본 고의 구성은 다음과 같다. 제2절에서는 분석모형의 이론적 배경을 설명하고 실증분석을 위한 두 가지 방법에 대해서 논의한다. 제3절은 분석에 사용된 자료에 대해 설명하고 본 분석을 위한 자료의 특징에 대한 사전 분석을 실시한다. 이어서 제4절은 분석 결과를 제시하고 양파 재배면적 전망치를 산출하는데 있어 중요한 이슈를 논의하고, 이를 고려한 전망치 산출결과를 제시한다. 마지막으로 제5절에서는 연구결과를 요약하고 결론을 제시한다.

2. 이론적 배경

2.1. Expected-Variance(EV) 모형

본 연구에서는 Markowitz(1952)가 제안한 예를 따라, 식 (1)과 같이 생산자들이 기대효용을 극대화하도록 재배면적에 대한 의사결정을 한다고 설정하였다(Freund, 1956; 김태후, 서상택, 2011).

$$(1) \text{Max } E[U(P_G X_G, P_O X_O)] \quad \text{st. } X_G + X_O \leq L, X_G \geq 0, X_O \geq 0$$

식 (1)에서 $EU(P_G X_G, P_O X_O)$ 는 대표적 생산자(농가)의 기대효용이며, P_G 양파 단위면적당 평균 수익(가격), P_O 는 여타 작물 단위면적당 평균 수익(가격)이며, X_G 대표적 생산자의 양파재배면적, X_O 는 대표적 생산자의 여타 작물 재배면적, L 은 대표적 생산자의 재배 가능 면적이다. 식 (1)의 기대효용 극대화 문제는 식 (2)와 같은 Expected-Variance(EV Model) 모형으로 변환할 수 있다(Markowitz, 1987).

$$(2) \text{ Maz } Z = P_G X_G + P_O X_O - \psi(\sigma_G^2 X_G^2 + \text{cov}(P_G, P_O) X_G X_O + \sigma_O^2 X_O^2) \\ \text{st. } X_G + X_O \leq L, X_G \geq 0, X_O \geq 0$$

식 (2)에서 ψ 생산자의 위험회피계수, σ_G^2 는 양파 재배면적당 수익의 분산, σ_O^2 는 여타 작물 재배면적당 수익의 분산, $\text{cov}(P_G, P_O)$ 는 양파와 여타 작물 재배면적당 수익의 공분산이다. 식 (2)로 표시된 EV 모형을 선택 변수인 X_G 와 X_O 로 미분하면 다음과 같은 1계 조건을 얻을 수 있다.¹⁾

$$(3) \frac{\partial Z}{\partial X_G} = P_G - 2\psi(\sigma_G^2 X_G + \text{cov}(P_G, P_O) X_O) = 0 \\ \frac{\partial Z}{\partial X_O} = P_O - 2\psi(\sigma_O^2 X_O + \text{cov}(P_G, P_O) X_G) = 0$$

식 (3)을 통해 X_G 는 $X_G = f(P_G, \sigma_G^2, \text{cov}(P_G, P_O), X_O)$ 의 함수형태라는 것을 유추할 수 있으며, $X_O = f(P_O, \sigma_O^2, \text{cov}(P_G, P_O), X_G)$ 의 함수형태라는 것을 유추할 수 있으므로, $X_G = f(P_G, \sigma_G^2, \text{cov}(P_G, P_O), X_O(P_O, \sigma_O^2, \text{cov}(P_G, P_O)))$ 로 표현할 수 있다. 따라서 X_G 는 외생변수인 $P_G, P_O, \sigma_G^2, \sigma_O^2$ 및 $\text{cov}(P_G, P_O)$ 의 함수로 표현할 수 있으므로 식 (4)와 같이 정리할 수 있다.

본 연구에서는 식 (4)를 최종적으로 양파재배면적 결정을 분석하는 실증분석 방정식으로 사용하였다. 식 (4)의 실증분석 방정식에는 가격 위험을 나타내는 변수가 설명 변수로 삽입되어 있어 해외 선행연구에서 제시한 위험과 생산 결정 간의 관계를 규명한 결과와 같은 맥락으로 이해할 수 있다(Just, 1974; Hurt and Garcia, 1982; Adesina and Brorsen, 1987; Nieuwoudt et al., 1988; Chavas and Matthew, 1990). 본 연구의 실증분석에서는 이전 작

1) 식 (2)에는 부등식 제약이 있으므로, 이를 반영한다면 1계 조건이 식 (3)과 같이 도출되는 것이 아니라, Khun-Tucker 조건으로 도출되어야 한다. 본 연구에서 Khun-Tucker 조건을 사용하지 않은 이유는 실증분석방정식으로 도출한 식 (4)를 개별농가 차원이 아니라 개별농가의 재배면적 결정식을 합산하여 우리나라 전체의 양파 재배면적 차원에서 분석하였기 때문이다. 즉, 이 경우 개별농가 차원에서의 토지제약과 비음 제약은 가격 및 가격 변동성에 대한 우리나라 전체 재배면적의 반응을 이끌어내는 데에는 큰 영향을 미치지 못한다.

물 연도의 평균가격(월별가격을 평균하여 연평균 가격으로 환산)을 평균 수익을 나타내는 변수로 사용하였으며, 분산을 나타내는 대리변수로는 이전 작물 연도의 변이계수를 사용하였고, 공분산으로는 이전 작물 연도의 양파 가격과 여타 작물가격의 공분산을 도출하여 사용하였다.

$$(4) X_G = f(P_G, \sigma_G^2, cov(P_G, P_O), \sigma_O^2, P_O)$$

2.1. 다변량 GARCH (MGARCH) 모형

식 (1)~(4)로 논의한 EV 모형에서는 농가들이 가격(수익)의 불확실성을 고려하여 재배면적을 결정한다는 메커니즘이 충실하게 반영되긴 하였지만, 재배면적의 변화가 반대로 가격의 변동성에 영향을 미치는 동태적 메커니즘은 반영할 수 없다. 따라서 본 연구에서는 가격 및 가격변동성과 재배면적의 변동성이 상호 영향을 주고받는 관계를 추가적으로 분석하기 위해 다변량 GARCH(Multivariate GARCH: MGARCH) 모형을 추정하였다.

MGARCH 모형의 각 시계열은 P차 자기회귀(즉, AR(P))구조를 가지고 있으면서, 각 시계열 간의 조건부 이분산(Conditional Heteroscedasticity)은 상호 연계되어 있는 구조이다. 본 연구에서는 분산구조를 나타내는 GARCH 항에서 분산의 비대칭성을 고려한 TGARCH 모형을 적용하여 식 (5)와 같은 형태로 실증분석을 수행하였다.

$$(5) \Sigma_t = M_0 + \sum_{i=1}^n A_i \odot (\epsilon_{t-i} \epsilon_{t-i}') + \sum_{j=1}^m B_j \odot [\epsilon_{t-j} (\epsilon_{t-j} < 0)] [\epsilon_{t-j} (\epsilon_{t-j} < 0)]'$$

식 (5)에서 M_0 는 상수항 계수 행렬이며, A_i 및 B_j 는 각각 이전 시점의 잔차 제곱과 잔차가 음인 경우에 대한 잔차 제곱에 대한 추정계수를 나타내는 대칭 행렬이며, $\odot$ 는 행렬 내의 각 원소별로 곱하는 Hadamard 곱을 말한다. 따라서 예를 들어, 만일 재배면적과 평균 가격 각각의 2개 시계열을 대상으로 n과 m이 1이라면(즉, 과거 시차가 1기까지만 경우), 식 (5)는 다음 식 (6)과 같이 풀어쓸 수 있다.

$$(6) \begin{bmatrix} \sigma_{11,t}^2 \\ \sigma_{21,t}^2 \sigma_{22,t}^2 \end{bmatrix} = \begin{bmatrix} \delta_{11,0} \\ \delta_{21,0} \delta_{22,0} \end{bmatrix} + \begin{bmatrix} \theta_{11,1} \\ \theta_{21,1} \theta_{22,1} \end{bmatrix} \odot \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{2,t-1} \epsilon_{1,t-1} \epsilon_{1,t-1}^2 \end{bmatrix} \\ + \begin{bmatrix} \psi_{11,1} \\ \psi_{21,1} \psi_{22,1} \end{bmatrix} \odot \begin{bmatrix} [\epsilon_{1,t-1} (\epsilon_{1,t-1} < 0)]^2 \\ [\epsilon_{1,t-1} (\epsilon_{1,t-1} < 0)] [\epsilon_{2,t-1} (\epsilon_{2,t-1} < 0)] [\epsilon_{2,t-1} (\epsilon_{2,t-1} < 0)]^2 \end{bmatrix}$$

이를 풀어 정리하면 식 (7)과 같이 쓸 수 있다.

$$\begin{aligned} \sigma_{11,t} &= \delta_{11,0} + \theta_{11,1} \epsilon_{1,t-1}^2 + \psi_{11,1} [\epsilon_{1,t-1} (\epsilon_{1,t-1} < 0)]^2 \\ (7) \sigma_{22,t} &= \delta_{22,0} + \theta_{22,1} \epsilon_{2,t-1}^2 + \psi_{22,1} [\epsilon_{2,t-1} (\epsilon_{2,t-1} < 0)]^2 \\ \sigma_{21,t} &= \delta_{21,0} + \theta_{21,1} \epsilon_{1,t-1} \epsilon_{2,t-1} + \psi_{21,1} [\epsilon_{1,t-1} (\epsilon_{1,t-1} < 0)] [\epsilon_{2,t-1} (\epsilon_{2,t-1} < 0)] \end{aligned}$$

식 (7)에서 $\sigma_{11,t}$ 는 재배면적 시계열에 대한 t시점에서의 조건부 분산, $\epsilon_{1,t-1}^2$ 는 재배면적 시계열의 t-1시점 평균 방정식에서 도출된 잔차 곱, $[\epsilon_{1,t-1} (\epsilon_{1,t-1} < 0)]^2$ 는 t-1시점 재배면적 시계열 평균 방정식에서 도출된 잔차가 음인 경우 이의 잔차 제곱, $\sigma_{22,t}$ 는 양파 가격 시계열에 대한 t시점 조건부 분산, $\epsilon_{2,t-1}^2$ 는 양파 가격 시계열에 대한 t-1시점 평균 방정식으로부터 도출된 잔차 제곱, $[\epsilon_{2,t-1} (\epsilon_{2,t-1} < 0)]^2$ 는 t-1시점 양파 가격 시계열 평균 방정식에서 도출된 잔차가 음인 경우 이의 잔차 제곱, $\sigma_{21,t}$ 는 t시점에서의 양파 재배면적 시계열과 양파 가격 시계열에 대한 조건부 공분산, $\epsilon_{1,t-1} \epsilon_{2,t-1}$ 는 t-1시점 양파 재배면적 시계열과 양파 가격 시계열의 평균 방정식에서 도출된 잔차의 곱, $[\epsilon_{1,t-1} (\epsilon_{1,t-1} < 0)] [\epsilon_{2,t-1} (\epsilon_{2,t-1} < 0)]$ 는 t-1시점 재배면적 시계열 평균 방정식에서 도출된 잔차가 음이고, 양파 가격 시계열 평균 방정식에서 도출된 잔차가 음인 경우 이들 두 잔차의 곱이다.

본 연구에서는 특히 두 시계열의 조건부 공분산을 나타내는 $\sigma_{21,t}$ 항의 추정결과를 상세히 제시하고 시간의 변화에 따라 조건부 공분산의 변화가 어떻게 변해왔는지를 논의하고자 한다.

3. 분석 자료 및 사전 분석결과

3.1. 분석 자료

실증분석을 위한 분석 자료로 본 연구에서는 통계청에서 발표하는 연도별 양파 재배면적 자료(1976년~2020년)와, 양파의 월별 가격(생산자가격지수), 농산물의 월별 가격(생산자가격지수)을 사용하였다.

〈그림 1〉은 양파의 재배면적 추이를 나타낸 것이다. 지난 45여 년간 양파 재배면적은 등락을 거듭하긴 했지만, 연평균 약 313ha씩 추세적으로 증가해 온 것을 알 수 있다. 한편, 〈그림 2〉의 오른쪽 그래프에서 볼 수 있는 바와 같이, 양파의 재배면적과 양파의 연평균 가격(생산자가격지수) 간에는 매우 밀접한 양의 상관관계가 있는 것으로 나타나, 식 (3)을 통해 알 수 있는 이론적 논의와 합치함을 알 수 있다.

그림 1. 양파 재배면적 추이(ha)

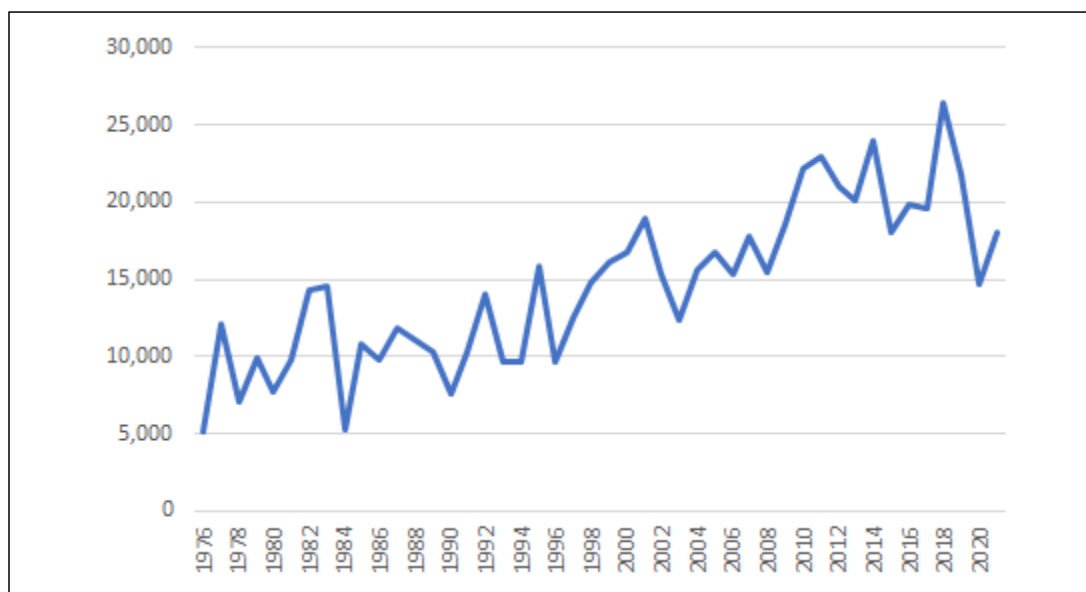
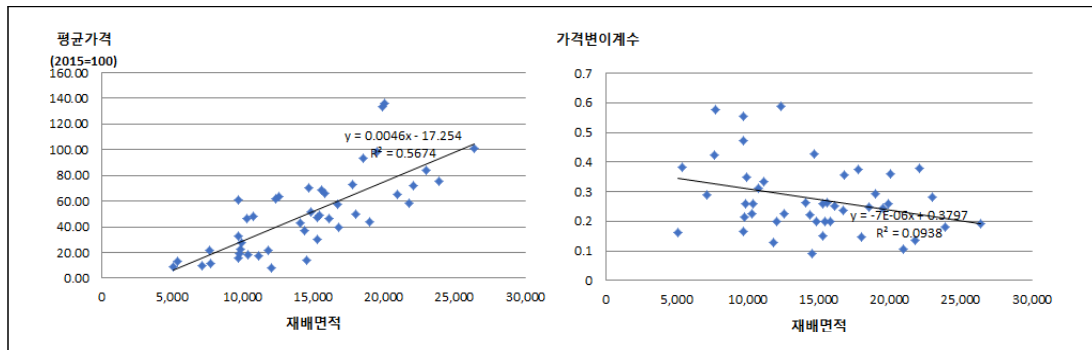


그림 2. 양파 재배면적과 양파 연평균 가격 및 연중 가격 변이계수 간의 상관관계



한편 양파의 재배면적과 양파 가격의 작물 연도 내의 변이계수 간에는 <그림 2>의 왼쪽 그래프에서 확인할 수 있는 바와 같이, 음의 상관관계를 보이는 것으로 나타난다. 이는 위험을 회피하고자 하는 농가들이 가격 변동성이 커질수록 재배면적을 줄이는 형태로 위험을 관리하고 있음을 나타내는 것이라 할 수 있다.

3.2. 실증분석을 위한 단위근 검정 결과

식 (4)와 식 (5)의 실증분석을 위해서는 분석대상 자료가 시계열적인 안정성(Stationarity)을 확보하고 있는지에 대한 검정이 필요하다. 안정적인 시계열이 아닐 경우 가성회귀 문제가 발생할 수 있기 때문이다.

<표 1>은 분석에 사용된 자료의 단위근 검정 결과이다. 표에서 확인할 수 있는 바와 같이, 대부분의 자료는 유의수준 1% 이상에서 안정적인 시계열인 것으로 판명되었으며, 농산물 연평균 가격은 유의수준 10%에서 안정적인 시계열인 것으로 판명되었다. 이와 같은 단위근 분석 결과를 바탕으로 하여 본 연구에서는 원자료를 이용하여 식 (4)와 식 (5)의 실증분석 방정식을 추정하였다.

표 1. 분석에 사용된 자료의 단위근 검정 결과

구분	검정통계량	유의확률
양파 재배면적	-5.6574	0.000
양파 연평균 가격(가격지수, 작물년도기준)	-7.3548	0.000
연중 양파 가격 변이계수(가격지수, 작물년도 기준)	-8.2136	0.000
농산물 연평균 가격(가격지수, 작물년도기준)	-3.1868	0.100
연중 양파 가격 변이계수(가격지수, 작물년도 기준)	-4.6549	0.003
Critical Value at 1% level	-4.1756	
Critical Value at 5% level	-3.5130	
Critical Value at 10% level	-3.1868	

주: 귀무가설(H0)은 해당 시계열 변수가 단위근을 가지고 있다고 설정된 것임. 작물년도는 양파의 수확기를 고려하여 금년 6월에서 익년 6월까지로 설정하였음.

4. 분석 결과

실증분석 방정식 (4)의 추정결과는 <표 2>와 <표 3>에 제시되어 있다. <표 2>는 종속변수를 양파의 재배면적으로 설정한 경우이며, <표 3>은 종속변수를 양파 재배면적의 증감률로 설정한 경우이다. 실증분석 방정식은 표에서 확인할 수 있는 바와 같이 두 가지 경우로 구분하였는데, <모형 1>은 생산자가 다른 작물의 가격 변동성은 양파 재배면적 결정에 반영하지 않는다는 전제를 바탕으로한 설명변수로 양파의 평균 가격과 가격 변이계수만을 삽입한 경우이며, <모형 2>는 생산자가 양파뿐만 아니라 여타 작물의 가격 변동성까지도 감안하여 양파의 재배면적을 결정한다는 전제를 바탕으로 추정된 것이다. 따라서 <모형 2>에는 실증분석 방정식 (4)에 나타난 모든 변수, 즉 여타 작물가격과 가격 변이계수 및 여타 작물과 양파 가격 간의 공분산(상관계수)이 포함되었다. 또한 실증분석에서는 추세변수를 포함하였는데, 이는 <그림 1>에서 볼 수 있는 바와 같이, 양파 재배면적이 추세적으로 증가해온 경향을 반영하기 위해서이다.

<표 2>의 추정결과를 보면, 이전 년도 양파의 평균가격이 높으면 금년도 양파 재배면적이 증가하는 관계에 있으며, 이전 년도 양파 가격 변동성(즉, 변이계수)이 높으면 금년도 양파 재

배면적은 감소하는 관계에 있는 것으로 나타난다. 이는 <모형 1>과 <모형 2> 양자에 걸쳐 이전 년도 양파의 평균가격과 변이계수에 대한 추정계수가 모두 90% 이상 통계적으로 유의하게 추정된 결과를 통해 잘 확인할 수 있다.

<모형 2>의 추정결과를 보면, 양파 재배면적에는 여타 작물의 가격수준 자체는 유의한 영향을 미치지 못하는 반면, 여타 작물가격의 변동성은 통계적으로 유의한 영향을 미치는 것으로 분석되었음을 알 수 있다. 즉, 생산자들은 양파 재배면적을 결정함에 있어 양파 가격의 변동성뿐만 아니라 여타 작물가격의 변동성까지도 심각하게 고려하고 있는 것으로 나타난다. 이는 생산자들이 가격 위험을 관리하는 형태로 양파 재배면적을 결정하고 있다는 의미로 풀이할 수 있다.

표 2. EV 모형에 바탕을 둔 양파 재배면적 결정함수 추정결과(재배면적 기준)

	모형 1 양파가격의 불확실성만 고려한 경우			모형 2 양파 및 농산물가격의 불확실성을 동시에 고려한 경우		
	계수	표준 오차	t 통계량	계수	표준 오차	t 통계량
상수항	10142.2***	1191.4	8.5124	7923.2***	1642.6	4.8200
이전 년도 양파 연평균 가격	48.09***	20.37	2.3600	36.48*	20.83	1.7500
이전 년도 연중 양파가격 변이계수	-11468***	3183.6	-3.6023	-9186.1**	4056.6	-2.2600
이전 년도 농산물 연평균 가격				17.27	82.55	0.2100
이전 년도 연중 농산물가격 변이계수				49302.7*	25291.7	1.9500
이전 년도 농산물 가격과 양파가격의 상관계수				-56822.8	68498.57	-0.8300
추세항	224.1***	48.8	4.5941	174.7***	199.7	0.8700
R ²	0.7815			0.8061		

주: *, **, ***는 각각 10%, 5%, 1%에서 유의하다는 것을 의미함.

<표 3>의 추정결과를 보면, 이전 년도 양파의 평균가격이 높으면 금년도 양파 재배면적은 이전 년도에 비해 증가하는 관계에 있으며, 이전 년도 양파 가격 변동성(즉, 변이계수)이 높으면 금년도 양파 재배면적은 이전 년도에 비해 감소하는 관계에 있는 것으로 나타난다. 이와 같은 추정결과 <표 2>에 나타난 추정결과와 일맥상통하는 것이라 할 수 있다. 한편 <표 3>의 <모형 2> 추정결과를 보면 통계적으로 유의미한 추정계수가 이전 년도 양파의 평균가격과 추세항에 국한되어, 특별한 의미있는 해석을 덧붙이기는 어려운 것으로 보인다.

표 3. EV 모형에 바탕을 둔 양파 재배면적 결정함수 추정결과(재배면적 증감을 기준)

	모형 1 양파가 격의 불확실성만 고려한 경우			모형 2 양파 및 농산물 가격의 불확실성을 동시에 고려한 경우		
	계수	표준 오차	t 통계량	계수	표준 오차	t 통계량
상수항	0.009***	0.002	3.950	0.009***	0.003	3.670
이전 년도 양파 연평균 가격	1.090***	0.373	-2.930	1.180***	0.501	-2.350
이전 년도 연중 양파가격 변이계수	-0.025***	0.006	-4.410	-0.004	0.010	-0.420
이전 년도 농산물 연평균 가격				0.755	3.125	0.240
이전 년도 연중 농산물가격 변이계수				1.470	8.465	0.170
이전 년도 농산물 가격과 양파가격의 상관계수				-0.015	0.025	-0.610
추세항	0.495***	0.139	3.550	0.489***	0.203	2.410
R ²	0.3783			0.3852		

주: *, **, ***는 각각 10%, 5%, 1%에서 유의하다는 것을 의미함.

〈표 4〉와 〈표 5〉는 양파 재배면적-양파 연평균 가격 조합 시계열의 MGARCH 모형을 추정
한 결과이다. 우선 〈표 4〉는 MGARCH 모형의 평균 방정식을 추정한 결과인데, 편자기상관
함수(Partial Autocorrelation Function: PACF)를 분석한 결과 양파 재배면적 시계열은
AR(2)까지를 포함하는 모형이, 양파 평균가격 시계열은 AR(3)까지를 포함하는 모형이 적절
한 것으로 분석되었다(부록 참조). 평균 방정식 추정결과를 보면 〈표 4〉에서 확인할 수 있는
바와 같이, 양파 재배면적 시계열에서는 상수항과 AR(1)항이, 양파 평균가격 시계열에서는
상수항과 AR(1), AR(3)항이 통계적으로 유의한 것으로 추정되었다. 식 (6)의 조건부 분산을
추정한 결과인 〈표 5〉를 보면, $\theta_{11,1}$ 과 $\psi_{11,1}$ 를 제외하고는 모든 계수들이 통계적으로 유의한
것으로 추정되었다.

표 4. 재배면적-연평균 가격 조합의 MGARCH 모형에서의 평균방정식 추정결과

변수	양파 재배면적			양파 연평균 가격		
	추정계수	z-Statistic	Prob.	추정계수	z-Statistic	Prob.
상수항	14728.5***	11.769	0.0000	51.4333***	4.4357	0.0000
AR(1)	1.1147***	5.5194	0.0000	0.5184***	4.0215	0.0000
AR(2)	-0.3617	-1.5898	0.1118	0.0472	0.4406	0.6594
AR(3)				0.2095**	2.3132	0.0207

주: *, **, ***는 각각 10%, 5%, 1%에서 유의하다는 것을 의미함.

표 5. 재배면적-연평균 가격 조합의 MGARCH 모형에서의 조건부 이분산식 추정결과

계수	추정 값	z-Statistic	유의확률
$\delta_{11,0}$	13057884***	2.8537	0.0043
$\delta_{22,0}$	43951*	1.9341	0.0531
$\delta_{12,0}$	199*	1.6560	0.0977
$\theta_{11,1}$	0.1899	0.6178	0.5367
$\theta_{22,1}$	0.7101***	4.0081	0.0001
$\psi_{11,1}$	0.4725	0.8297	0.4067
$\psi_{22,1}$	1.6007**	2.2491	0.0245

주: 하첨자에서 쉼표(.) 앞의 1, 2는 각각 양파 재배면적과 양파의 연평균 가격(작물년도 기준)을 가리키며, 하첨자에서 쉼표(.) 뒤의 0과 1은 시점을 가리킴(즉, 0은 t기를 나타내며, 1은 t-1기를 나타냄).

주: *, **, ***는 각각 10%, 5%, 1%에서 유의하다는 것을 의미함.

〈표 6〉과 〈표 7〉은 양파 재배면적-양파 연중 가격 변이계수 조합 시계열의 MGARCH 모형을 추정한 결과이다. 〈표 6〉은 MGARCH 모형의 평균 방정식을 추정한 결과인데, 편자기상관함수(Partial Autocorrelation Function: PACF)를 분석한 결과 양파 연중 가격 변이계수 시계열은 AR(0)까지를 포함하는 모형이 적절한 것으로 분석되었다(부록 참조). 평균 방정식 추정결과를 보면 〈표 6〉에서 확인할 수 있는 바와 같이, 양파 재배면적 시계열에서는 상수항과 AR(1)항과 AR(2)항이, 양파 연중 가격 시계열에서는 상수항이 통계적으로 유의한 것으로 추정되었다. 식 (7)의 조건부 분산을 추정한 결과인 〈표 5〉를 보면, $\delta_{11,0}$, $\delta_{22,0}$ 및 $\theta_{22,1}$ 계수가 통계적으로 유의한 것으로 추정되었다.

표 6. 재배면적-연중 가격 변이계수 조합의 MGARCH 모형에서의 평균방정식 추정결과

변수	양파재배면적			양파 연중 가격 변이계수		
	추정계수	z-Statistic	Prob.	추정계수	z-Statistic	Prob.
상수항	1978.2	3372.6	0.5866	0.2651***	0.0163	16.2465
AR(1)	0.5599***	0.0961	5.8292			
AR(2)	0.3557***	0.0965	3.6865			

주: *, **, ***는 각각 10%, 5%, 1%에서 유의하다는 것을 의미함.

표 7. 재배면적-연중 가격 변이계수 조합의 MGARCH 모형에서의 조건부 이분산식 추정결과

계수	추정 값	z-Statistic	유의확률
$\delta_{11,0}$	10389087**	4779967	2.1735
$\delta_{12,0}$	-117.427	120.408	-0.975
$\delta_{22,0}$	0.0125***	0.0029	4.2794
$\theta_{11,1}$	0.4193	0.6493	0.6458
$\theta_{12,1}$	0.2304	0.7015	0.3285
$\theta_{22,1}$	-1.20E-01***	0.025395	-4.72E+00
$\psi_{11,1}$	5.79E+02	509.4533	1.14E+00
$\psi_{12,1}$	9.84E-07	24.63917	3.99E-08
$\psi_{22,1}$	1.67E-06	0.23988	6.97E-06

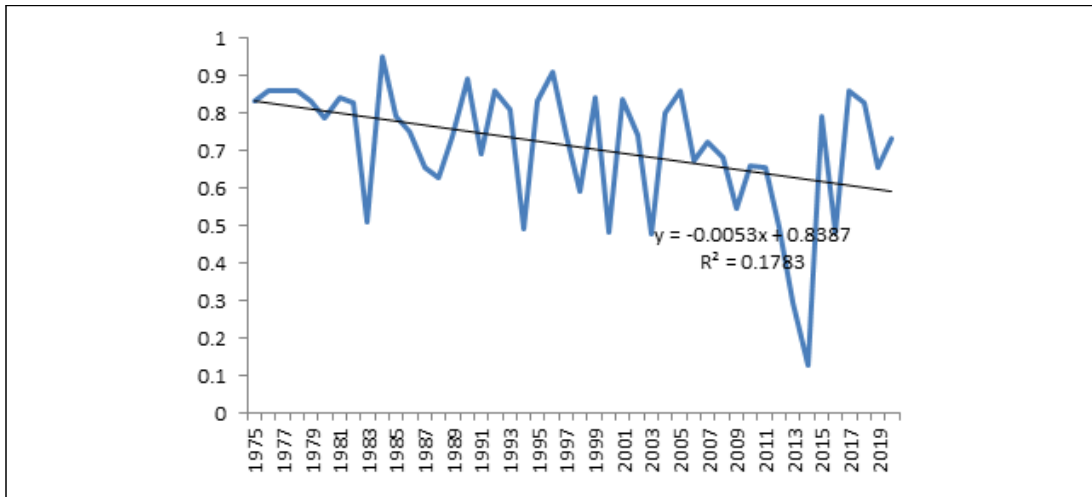
주: 하첨자에서 쉼표(.) 앞의 1, 2는 각각 양파 재배면적과 양파의 연중 변이계수(작물년도 기준)을 가리키며, 하첨자에서 쉼표(.) 뒤의 0과 1은 시점을 가리킴(즉, 0은 t기를 나타내며, 1은 t-1기를 나타냄).

주: *, **, *** 는 각각 10%, 5%, 1%에서 유의하다는 것을 의미함.

〈그림 3〉은 〈표 4〉와 〈표 5〉의 추정결과를 이용하여 도출한 양파 재배면적과 양파 연평균 가격 간 상관계수의 동태적 변화를 나타낸 것이다. 그림에서 볼 수 있는 바와 같이, 양파 재배면적과 연평균 가격의 양의 상관계수 값은 시간이 흐를수록 낮아지고 있는 것으로 나타난다. 이와 같은 분석 결과는 연평균 가격이 양파 재배면적에 미치는 영향이 최근으로 올수록 낮아지고 있다는 것을 의미한다.

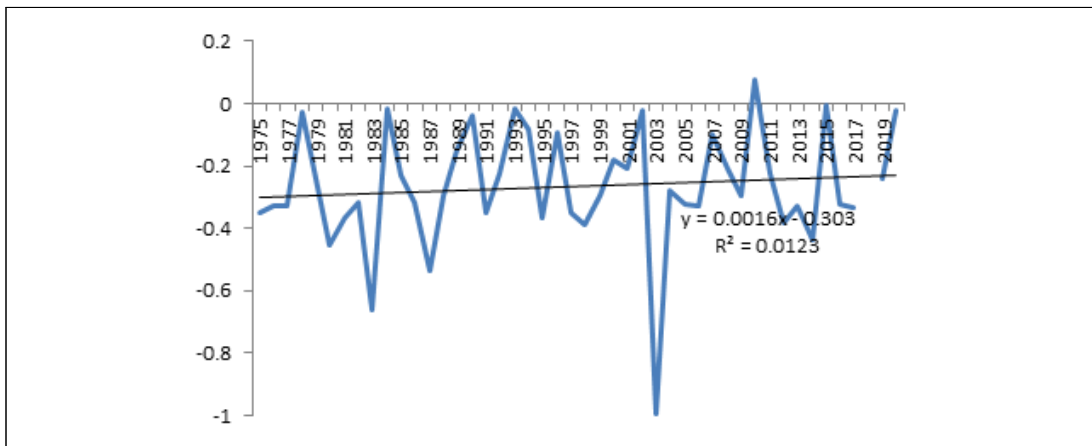
〈그림 3〉의 분석 결과는 EV 모형으로 추정한 〈표 2〉의 결과만을 이용해서는 확인할 수 없는 정보를 추가적으로 제공하고 있다. 즉, 〈표 2〉의 추정결과는 이전 년도의 연평균 가격은 양파 재배면적을 결정하는 매우 중요한 변수라는 것을 확인해 주고 분석 기간 전체에 걸쳐 〈표 2〉의 추정계수만큼의 영향력을 평균적으로 발휘해 왔다는 것을 증명해 주고 있다. 하지만 〈그림 3〉의 동태적 상관계수 추정결과는 연평균 가격의 영향력이 시간이 지남에 따라 낮아지고 있기 때문에 향후 양파 재배면적의 전망에서는 이와 같은 추세적인 영향력 약화 경향을 감안해야 한다는 점을 말하고 있다.

그림 3. 양파 재배면적과 양파 연평균 가격 간 상관계수의 동태적 변화 추이



〈그림 4〉는 〈표 6〉과 〈표 8〉의 추정결과를 이용하여 도출한 양파 재배면적과 양파 가격의 연중 변이계수 간 상관계수의 동태적 변화를 나타낸 것이다. 그림에서 볼 수 있는 바와 같이, 양파 재배면적과 연중 가격 변이계수 간 음의 상관계수 값은 시간이 흐를수록 절댓값 기준으로 낮아지고 있는 것으로 나타난다. 이와 같은 분석 결과는 연중 가격 변이계수가 양파 재배면적에 미치는 부정적인 영향이 최근으로 올수록 낮아지고 있다는 것을 의미한다.

그림 4. 양파 재배면적과 양파 연중 가격 변이계수 간 상관계수의 동태적 변화 추이



이러한 분석 결과 역시 EV 모형으로 추정된 <표 2>의 결과만을 이용해서는 확인할 수 없는 정보를 추가적으로 제공하고 있다. 즉, <표 2>의 추정결과는 분석 기간 전체에 걸쳐 <표 2>의 추정계수만큼 연중 가격 변이계수가 양파 재배면적 결정에 음의 영향력을 평균적으로 발휘해 왔다는 것을 증명해 주고 있지만, <그림 4>의 동태적 상관계수 추정결과는 연중 가격 변이계수의 부정적 영향력이 시간이 지남에 따라 낮아지고 있기 때문에 향후 양파 재배면적의 전망에서는 이와 같은 추세적인 영향력 약화 경향을 감안해야 한다는 점을 말하고 있다.

<표 8>은 EV 모형 추정 결과와 EV 모형 추정 결과에 MGARCH 모형에서 도출된 동태적 상관계수 변화 경향을 감안하여 2022년의 양파 재배면적을 전망한 결과이다. 전망에 앞서 표에서 확인할 수 있는 바와 같이 2021년 전망치를 산출한 결과 2021년 실제 재배면적과 전망치 간의 오차는 EV 모형 추정 결과만을 이용한 경우는 5.07%, EV모형 추정 결과에 MGARCH 모형에서 도출된 동태적 상관계수 변화 경향을 감안한 경우에는 4.82%인 것으로 계산된다. 즉, 동태적 상관계수 변화 경향을 반영하여 전망치를 산출하는 것이 보다 오차를 줄이는 접근법을 알 수 있다.

추정결과를 이용한 2022년 양파 재배면적 전망치는 EV 모형을 바탕으로 한 경우에는 22,197ha, EV모형과 MGARCH 모형을 모두 이용한 경우에는 22,123ha에 이르는 것으로 추정된다.²⁾

표 8. 추정결과를 이용한 2022년 양파 재배면적 전망(조생종 및 중만생종 재배면적 합산 기준)

(단위: ha)

구분	2021년	2022년
실제재배면적(A)	18,014	-
EV 모형을 이용한 전망치(B)	18,928	22,197
EV 모형에 MGARCH 모형에서 도출된 동태적 상관계수를 감안한 전망치(C)	18,881	22,123
예측오차: [B-A]/A	5.07%	
예측오차: [B-C]/C	4.82%	

주: <표 2> 추정결과 중 <모형 1>의 결과를 바탕으로 한 것임.

2) 2022년 재배면적 전망을 위한 양파 생산자 가격지수는 2021년 5월 통계치가 공표되지 않아 2020년 6월~2021년 4월 간의 자료만을 이용하여 전망치 산출에 사용하였다.

5. 요약 및 결론

재배면적을 전망하는 데에는 해당 품목의 가격과 가격 변동성이 가장 중요하게 고려되어야 할 요인들임에도 불구하고 재배면적 결정요인을 연구한 선행연구들에서는 가격변동성과 재배면적 간에 나타나는 동태적인 연관 관계까지는 분석하지 못하였다. 본 연구에서는 가격과 가격 위험이 재배면적에 미치는 영향을 분석하기 위해 농가들의 효용 극대화 문제를 설정하고 이를 실증분석이 가능한 평균-분산모형(Expected-Variance Model: EV-Model)으로 변환하였다. 또한, EV모형이 갖는 한계를 보완하기 위해 재배면적과 가격, 재배면적과 가격 변동성을 동시에 고려하는 다변량 조건부 이분산 시계열 모형을 추정하여 재배면적과 가격 및 가격 변동성 간에 존재하는 동태적인 상관관계를 추정하여 EV 모형에서 산출된 전망치가 어떻게 개선될 수 있는지도 보여주고자 하였다.

EV 모형에서 도출된 실증분석 방정식에는 가격 위험을 나타내는 변수가 설명변수로 삽입되어 있어 해외 선행연구에서 제시한 위험과 생산 결정 간의 관계를 규명한 결과와 일치하였다. 하지만 EV 모형에서는 재배면적의 변화가 반대로 가격의 변동성에 영향을 미치는 동태적 메커니즘은 반영할 수 없다. 따라서 본 연구에서는 가격 및 가격변동성과 재배면적의 변동성이 상호 영향을 주고받는 관계를 추가적으로 분석하기 위해 다변량 GARCH(Multivariate GARCH: MGARCH) 모형을 추정하였다. 본 연구에서는 특히 GARCH항에서 분산의 비대칭성을 고려한 TGARCH 모형을 적용하였다.

분석에 앞서 지난 45여 년간 양파 재배면적 등락 추이를 살펴본 결과, 양파의 재배면적과 양파의 연평균 가격(생산자가격지수) 간에는 매우 밀접한 양의 상관관계가 있는 것으로 나타났다. 양파의 재배면적과 양파 가격의 작물 연도 내의 변이계수 간에는 음의 상관관계를 보이는 것으로 나타나 위험을 회피하고자 하는 농가들이 가격 변동성이 커질수록 재배면적을 줄이는 형태로 위험을 관리하고 있음을 유추할 수 있었다.

EV 모형 실증분석 방정식 추정결과, 이전 년도 양파의 평균가격이 높으면 금년도 양파 재배면적이 증가하는 관계에 있으며, 이전 년도 양파 가격 변동성(즉, 변이계수)이 높으면 금년도 양파 재배면적은 감소하는 관계에 있는 것으로 나타났다. 또한, 양파 재배면적에는 여타

작물의 가격수준 자체는 유의한 영향을 미치지 못하는 반면, 여타 작물가격의 변동성은 통계적으로 유의한 영향을 미치는 것으로 분석되어, 생산자들은 양파 재배면적을 결정함에 있어 양파 가격의 변동성 뿐만 아니라 여타 작물가격의 변동성까지도 심각하게 고려하고 있는 것으로 나타났다.

양파 재배면적-양파 가격 조합의 MGARCH 모형의 평균 방정식을 추정한 결과, 양파 재배면적 시계열에서는 상수항과 AR(1)항이, 양파 평균가격 시계열에서는 상수항과 AR(1), AR(3)항이 통계적으로 유의한 것으로 추정되었으며, 조건부 분산을 추정한 결과에서는 대부분의 계수들이 통계적으로 유의한 것으로 추정되었다.

양파 재배면적-양파 연중 가격 변이계수 조합 시계열의 MGARCH 모형을 추정한 결과, 양파 재배면적 시계열에서는 상수항과 AR(1)항과 AR(2)항이, 양파 연중 가격 시계열에서는 상수항이 통계적으로 유의한 것으로 추정되었으며, 조건부 분산을 추정한 결과에서도 통계적인 유의성을 확보한 계수들이 확인되었다.

MGARCH 모형 추정결과를 이용하여 양파 재배면적과 양파 연평균 가격 간 상관계수의 동태적 변화를 분석한 결과, 양파 재배면적과 연평균 가격의 양의 상관계수 값은 시간이 흐를수록 낮아지고 있는 것으로 나타났으며, 양파 재배면적과 연중 가격 변이계수 간 음의 상관계수 값은 시간이 흐를수록 절댓값 기준으로 낮아지고 있는 것으로 나타났다. 이와 같은 분석결과는 가격과 연중 가격 변이계수가 양파 재배면적에 미치는 영향이 최근으로 올수록 낮아지고 있다는 것을 의미한다.

본 연구의 분석 결과를 이용하여 2021년 전망치를 산출한 결과 2021년 실제 재배면적과 전망치 간의 오차는 EV 모형 추정 결과만을 이용한 경우는 5.07%, EV 모형 추정에 MGARCH 모형에서 도출된 동태적 상관계수 변화 경향을 감안한 경우에는 4.82%인 것으로 계산되어 동태적 상관계수 변화 경향을 반영하여 전망치를 산출하는 것이 보다 오차를 줄이는 접근법인 것으로 분석되었다. 추정결과를 이용한 2022년 양파 재배면적 전망치는 EV 모형을 바탕으로 한 경우에는 22,197ha, EV모형과 MGARCH 모형을 모두 이용한 경우에는 22,123ha에 이르는 것으로 추정된다.

농산물의 재배면적 예측에는 생산자들의 의사결정 과정과 시장 내외적인 요인들을 동시에 고려해야 하기 때문에, 정확한 예측이 매우 어렵다. 그동안 한국농업관측본부에서도 다년간

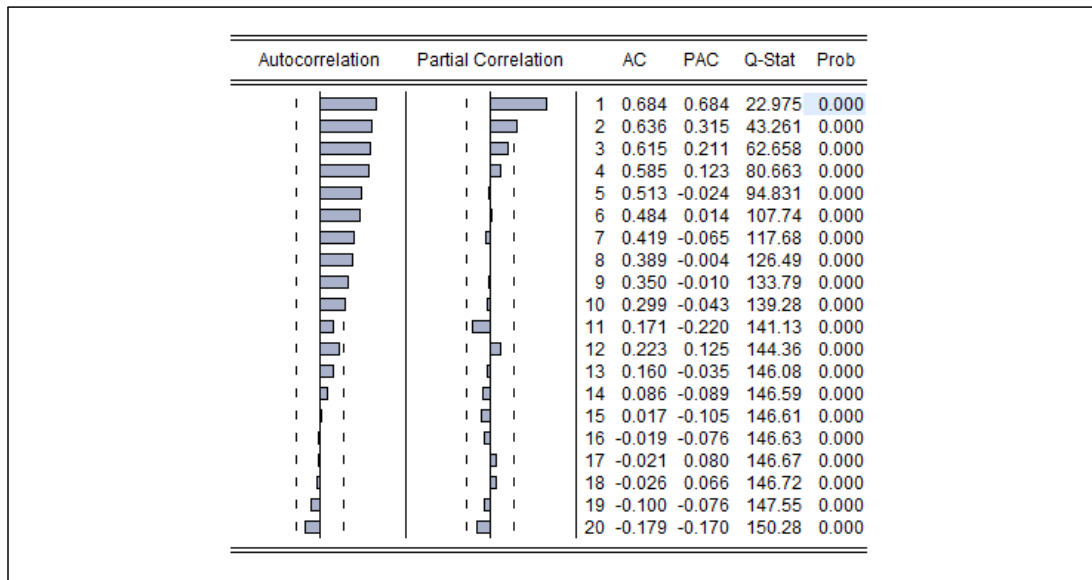
의 경험과 노하우를 통해 사전 재배면적 의향 조사, 가격변화 추이 등의 정보를 재배면적 전망에 다각도로 활용하여 왔으며 재배면적 반응함수에 대한 추정과 갱신도 주기적으로 이루어져 왔다. 본 연구는 이와 같은 다각적인 정보를 활용하는데 하나의 새로운 방향을 제시할 수 있을 것으로 기대된다. 재배면적에 영향을 미치는 가격과 가격 변동성이라는 요인은 그 영향력이 고정된 것이 아니라 시간에 따라 변화할 수 있다는 점을 전제로 한다면, 본 연구와 같이 동태적 상관관계를 전망모형에 삽입하여 활용하는 것은 예측의 정확도를 높이는 유용한 방법이 될 수 있을 것이다.

참고문헌

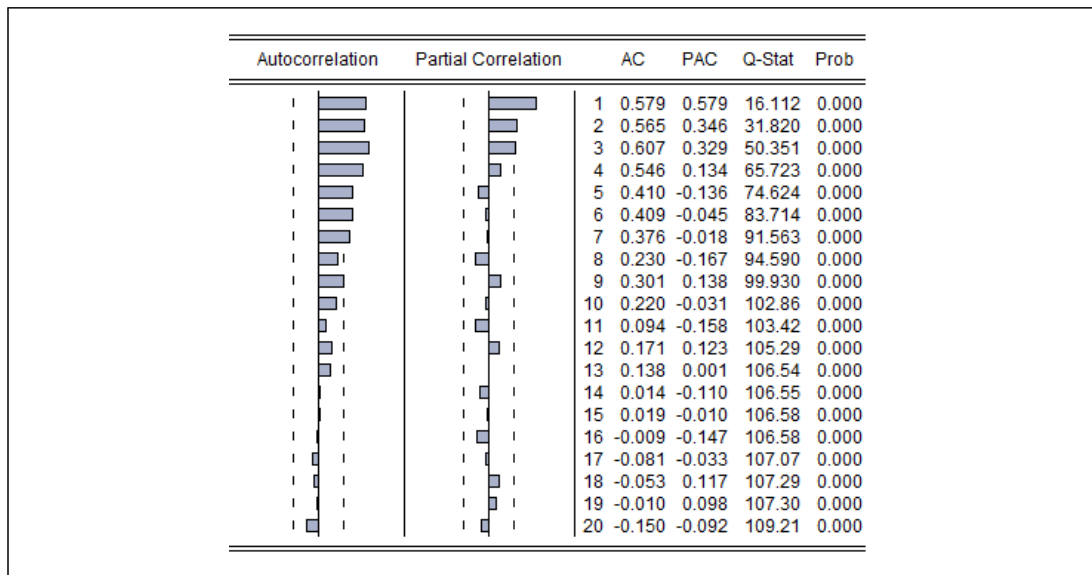
- 김태후, 서상택. 2011. “저장사과 최적 출하시기 결정 모형 개발-시뮬레이션을 결합한 EV 모델을 중심으로.” 『농업경제연구』 42(3): 29-48.
- 박지연, 박영구. 2013. 『배추·무 예측모형 고도화 방안』. 한국농촌경제연구원 기타연구보고서.
- 정학균, 이태호, 김관수. 2009. “위험하의 시설과채농가 재배면적변화 분석- 남부지방 오이사례를 중심으로.” 『농업경제연구』 50(3): 81-104.
- 조재환. 2021. “경남지역 주요 채소류 재배면적 반응함수 추정.” 『한국산학기술학회논문지』 22(2): 131-137.
- 홍승지, 김성훈, 이을경. 2015. “오차수정모형을 이용한 고추의 재배면적 반응함수 추정.” 『농업경영·정책연구』 52(1): 507-522.
- Adesina, A. A., Brorsen, B. W.(1987). “A risk responsive acreage response function for millet in Niger.” Journal of the International Association of Agricultural Economists 1(3): 229-239.
- Chavas, Jean-Paul, Holt, T. Matthew(1990). “Acreage Decisions under Risk: The Case of Corn and Soybeans.” American Journal of Agricultural Economics 72(3): 529-538.
- Freund, R(1956). “The Introduction of Risk into a Programming Model.” Econometrica 21: 253-263.
- Hurt, C. A. and Garcia, Philip(1982). “The Impact of Price Risk on Sow Farrowings, 1967-78.” American Journal of Agricultural Economics 64(3): 565-568.
- Just, R. E.(1974). “An Investigation of the Importance of Risk in Farmers' Decisions.” American Journal of Agricultural Economics 56(1): 14-25.
- Markowitz, H.M.(1987). Mean-Variance Analysis in Portfolio Choice and Capital Market. B. Blackwell, New York.
- Nieuwoudt, W. L., Womack, Abner W., Johnson and R. Stanley(1988). “Measurement of Importance of Risk on Supply Response of Corn and Soybeans.” North Central Journal of Agricultural Economics 10(2): 281-292.

부록

양파 재배면적 시계열에 대한 Autocorrelation 및 Partial Correlation 분석결과



양파 연평균 가격 시계열에 대한 Autocorrelation 및 Partial Correlation 분석결과



양파 연중 가격변이계수 시계열에 대한 Autocorrelation 및 Partial Correlation 분석결과

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.225	-0.225	2.4316	0.119
		2	-0.032	-0.087	2.4833	0.289
		3	-0.195	-0.237	4.4073	0.221
		4	0.203	0.105	6.5320	0.163
		5	-0.167	-0.144	8.0104	0.156
		6	0.148	0.079	9.1996	0.163
		7	0.108	0.215	9.8493	0.197
		8	-0.180	-0.201	11.708	0.165
		9	-0.102	-0.070	12.320	0.196
		10	0.255	0.246	16.236	0.093
		11	-0.035	-0.085	16.313	0.130
		12	-0.153	-0.111	17.814	0.121
		13	-0.026	-0.028	17.860	0.163
		14	0.021	-0.146	17.891	0.212
		15	-0.014	0.079	17.904	0.268
		16	0.007	-0.042	17.908	0.329
		17	0.170	0.041	20.089	0.270
		18	-0.229	-0.055	24.207	0.148
		19	0.047	0.049	24.384	0.182
		20	-0.101	-0.172	25.241	0.192



제1회

농산물 수급 예측모형 경진대회

우수논문집

Chapter 1
지정과제



우수상

SIMPLE 작물모형을 활용한 마늘·양파 단수 예측

현신우 · 김광수

SIMPLE 작물모형을 활용한 마늘·양파 단수 예측

현신우* 김광수*, **

Key words

작황예측(Crop yield forecast), 모수추정(Parameter calibration), 생구중(Fresh bulb weight), 월동작물(Overwintering crop), 과정기반 모형(Process based model)

ABSTRACT

The SIMPLE crop model has been developed to minimize the requirement of input data for crop growth simulations, which allows for prediction of crop yield under given crop management conditions as well as weather conditions. The objectives of this study were to improve the SIMPLE model for the growth simulations of onion and garlic and to assess the uncertainty of fresh bulb weight predicted upto three months in advance. Weather input files to the crop model were prepared compiling weather data obtained in a year of interest until the dates on which crop yield prediction was made, e.g., March 1st, April 1st and May 1st in 2020. Weather data in previous four years were also added to the weather input data for the period from the prediction date to the harvest date. As a result, four sets of weather data were prepared to predict crop yield on the given dates. The average fresh bulb weight predicted using these weather input data explained at least 89% of variation in the bulb weight estimated using weather data obtained after harvest. The degree of agreement between estimated and predicted crop yield was relatively high irrespective of the reliability of cultivar parameters. However, the difference between observed and predicted bulb weight was considerably large when crop yield was predicted using the cultivar parameters that resulted in a larger bias during the parameter calibration process. These results suggest that the fresh bulb weight of onion and garlic could be predicted with acceptable accuracy once the reliable sets of cultivar parameters would become available.

차례

- | | |
|------------|------------|
| 1. 서론 | 3. 결과 |
| 2. 재료 및 방법 | 4. 고찰 및 결론 |

* 서울대학교 농림생물자원학부

** 서울대학교 농업생명과학연구원

1. 서론

국내 자급률이 높은 양파와 마늘은 이상기상 조건의 발생으로 생산성이 급격하게 변동하여 소비자 경제에 부정적인 영향을 미칠 수 있다(Choi and Baek, 2016; Ha et al. 2019). 국내에서 재배되는 양파와 마늘 품종들은 월동기를 요구하기 때문에, 소비자 물가에 상당한 영향을 미치는 배추와 비교하여 상대적으로 긴 재배 기간이 요구된다. 특히, 월동 기간에 발생하는 이상기상 조건에 의해 이들 작물의 급격한 생산성 변동이 빈번하게 발생하고 있다(Song et al., 2018). 예를 들어, 2015년에 발생한 수확기 고온 및 가뭄의 영향으로 양파의 생산량이 전년 대비 31% 감소하였다. 반면, 겨울철 기온이 평년보다 높았던 2019년에는 해당 작물의 단위면적당 수량이 27% 증가하였다. 따라서 기상조건에 따른 생산성 변동을 수확기 이전에 예측할 경우, 작물의 재배관리뿐만 아니라 안정적인 농산물의 공급을 위한 의사결정을 지원할 수 있다.

양파와 마늘의 단위면적당 생산량을 예측하기 위한 경험적 모형들이 개발되어 왔다(Na et al., 2017). Oh and Kim(2017)은 공간자기상관 이동평균 모형을 기반으로 작물의 단위면적당 수확량을 예측하였다. Na et al.(2017)은 분광영상으로부터 얻어지는 식생지수를 기반으로 단위면적당 수량을 추정할 수 있는 경험식을 개발하였다. 그러나 경험적 작물모형들은 주로 월별 기상자료를 입력자료로 사용하고 있어, 단기간에 발생하는 이상기상 조건에 의한 생산성 변동을 예측하기 어렵다. 또한, 이러한 방식의 모형들은 농가에서 선택하는 재배관리가 고려되지 않고 있어, 동일한 기상조건이 형성되더라도 새로운 품종이나 재배관리 기법을 적용함에 따라 발생할 수 있는 생산성의 변화를 예측하기 어렵다.

기상조건뿐만 아니라 재배조건에 따른 작물의 생육을 모의하기 위해 생물리학적 작물모형들이 개발되어 왔다(Kim et al., 2015). 일반적으로 생물리학적 작물모형은 작물의 일별 생육량을 계산하고, 이를 누적하여 수확일의 단위면적당 생산량을 예측한다. 특히, 생물리학적 모형을 사용하여 일별 광합성량이나 발육단계 진전 속도 및 생육 단계에 따른 식물체 기관별 동화산물의 배분량을 계산할 수 있다. 이를 통해, 생물리학적 작물모형은 경험적 작물모형과 달리 작물의 재배 방식이나 품종 특성에 따른 생산성의 변화를 합리적으로 예측할 수 있다. 또

한, 생물리학적 모형을 개발할 때, 작물의 생육 과정을 나타내기 위한 수식들이 축적된 지식으로부터 도출되기 때문에, 환경조건과 생산성과의 경험적인 관계를 찾아가는 기계학습법에 비해 상대적으로 적은 양의 데이터가 요구된다.

작물의 생육을 현실과 가깝게 모의할 수 있는 생물리학적 작물모형은 자료 수집이 어려운 입력요소들에 의존한다. 예를 들어, 지표면에서의 에너지 수지를 기반으로 광합성량과 생산성을 예측하는 모형들은 이상기상 조건에서 작물의 생육 변화를 정밀하게 모의할 수 있다. 이러한 모형들은 시간 단위의 일사량, 온도, 습도, 풍속 및 강수가 필수적인 기상요소로 사용된다. 국내에서는 작물모형 구동에 필요한 기상요소를 대상으로 시간별 기상관측자료를 수집할 수 있는 종관기상 관측소가 적어, 작물이 재배되는 대부분의 지역에서 높은 수준의 생물리학적 작물모형을 사용하기 어렵다. 또한, 생물리학적 작물모형들을 사용할 때, 재배관리 조건을 설정하기 위해 품종이나 시비 및 관개와 같은 입력요소들이 요구된다. 예를 들어, 재배관리에 따른 단위면적당 생산량을 예측하기 위해 시비량 및 시비시기가 생물리학적 작물모형의 입력자료로 사용된다. 그러나, 일반 농가를 대상으로 상세한 재배관리 자료들을 확보하기 어려운 국내에서는 이러한 입력자료를 준비하기 어렵다(Kim et al., 2015).

Zhao et al.(2019)는 작물 생육의 생물리학적 과정을 모의하더라도 입력자료 요구도를 상당히 경감시킨 SIMPLE 작물모형을 개발하였다. SIMPLE 작물모형은 생육 추정식에 사용되는 모수를 최소화할 수 있는 알고리즘을 사용한다. 예를 들어, 작물모형을 구성하는 수식을 단순화하기 위해 일사량과 흡광계수(fPAR; fractional Photosynthetically Active Radiation)를 사용하여 계산되는 일별 광합성량을 개별 기관별로 배분하는 과정을 생략하고, 식물체 전체 건물중의 변화를 모의한다. 또한, 일반 농가에서는 시비관리가 적절하게 수행되고 있다고 가정하여 재배관리에 대한 입력자료 요구도를 낮추었다. 따라서, SIMPLE 작물모형은 해당 작물의 일별 생육을 예측하기 위해 최소한의 입력자료가 사용된다. 특히, 작물모형의 입력자료를 구축하기 어려운 국내 여건을 고려할 때, 다양한 작물의 생산성을 예측하기 위해 SIMPLE 모형을 사용하는 것이 유리하다.

기상 예측 자료를 생물리학적 작물모형의 입력자료로 사용할 경우, 작물의 생산성을 사전에 예측할 수 있다. 예를 들어, 구 비대기 초기 시점까지는 현재까지의 기상자료를 사용하고, 수확기 시점까지는 기상 전망 자료를 활용하여, 수확기 이전에 작물의 생육 및 생산성을 예측

할 수 있다. Basso and Liu(2017)는 기존에 확보된 과거 기상자료들을 활용하여 작물 생산성을 예측할 수 있는 것으로 보고하였다. 그러나 이러한 방식으로 작물 생산성을 예측할 때, 예측 시점부터 수확기까지의 기상조건을 나타내는 기상입력자료의 신뢰도에 따라 상당한 수준의 불확도가 발생할 수 있다(Lee et al., 2015).

본 연구에서는 SIMPLE 작물모형을 마늘과 양파의 생육 모의가 가능하도록 개선하고, 현재와 과거 기상자료를 조합하여 이들 작물의 생구증을 예측하고자 하였다. 특히, 수확기 3개월 이전 시점부터 생산성을 예측하고, 이들 예측값의 불확도를 평가하고자 하였다. 수확 시점에 앞서 예측된 작물의 생산성 자료는 효과적인 수급 정책을 수립할 수 있도록 지원할 수 있다(Hur et al., 2019). 예를 들어, 작물 생산성이 예측된 자료를 활용하여 해당 농산물의 가격을 예측할 수 있다(Ha et al. 2019). 특히, 한 달 이상의 선행시간을 두고 양파나 마늘의 생구증 예측 자료가 생산될 경우, 수확량 변동에 따른 대책을 마련하고 이를 시행할 수 있는 시간 확보가 가능하다. 따라서, 본 연구의 결과들은 작물에 대한 생육 및 생산성 예측을 위한 작물모형 개선 방향을 제시할 수 있을 뿐만 아니라 장기 작황 예측 체계를 구축하기 위해 활용될 수 있을 것이다.

2. 재료 및 방법

2.1. SIMPLE 작물 모형

SIMPLE 작물모형은 단위면적당 생산량을 예측하기 위해 일별 광합성량을 계산하고 이 중 일부가 수확기관으로 배분되는 과정을 모의한다. 특정일 t 에 생산된 동화산물에 의해 근락의 일별 건물중인 B_t 가 증가하는 방식으로 생육이 모의된다:

$$B_{t+1} = B_t + \delta_t \text{ (Eq. 1)}$$

여기서 δ_t 는 일별 순광합성량을 나타낸다. 광합성에 관여하는 환경요인과 생물학적인 요인들을 고려하여 일단위로 δ_t 값을 계산하기 위해, 일사량(SR_t), 흡광계수($fPAR_t$), 광합성 변동 계수(V) 및 광 사용 효율계수(RUE , Radiation Use Efficiency)가 다음과 같이 사용된다:

$$\delta_t = RUE \cdot SR_t \cdot fPAR_t \cdot V_t \text{ (Eq. 2)}$$

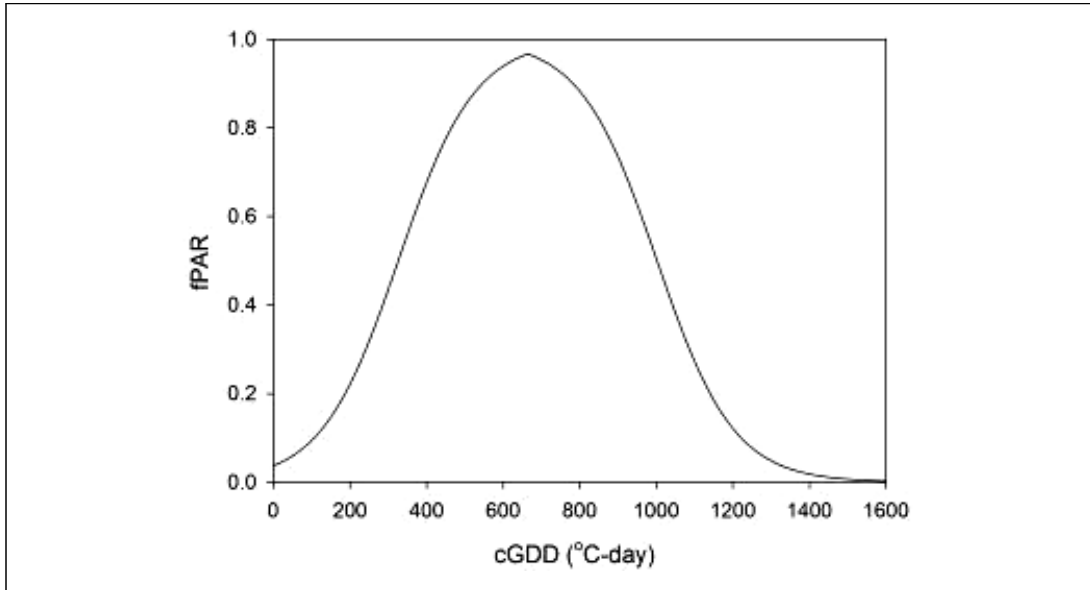
생육 단계에 따른 광합성량은 군락 조건에 영향을 받기 때문에 SIMPLE 작물모형은 군락단위의 흡광계수를 일별로 계산한다. 작물의 영양생장기와 생식생장기에 작물 군락에서 광이 흡수되는 정도를 계산하기 위한 함수인 $fPAR$ 가 다음과 같이 정의된다:

$$fPAR(cGDD, Phase, p) = 1 / [1 + e^{p(cGDD - Phase)}] \text{ (Eq. 3)}$$

여기서 $cGDD$ 와 $Phase$ 는 각각 정식일 이후 누적 생장도일과 생육 단계별 모수이다. 작물이 재배되는 기간 동안 개화기를 기점으로 광합성 유효복사를 흡수하는 특성이 변화하기 때문에, 영양생장기와 생식생장기를 대표하는 $Phase_{veg}$ 와 $Phase_{rep}$ 값이 사용된다. 또한, 품종의 생태형에 따라 개화기 시점에 차이가 있기 때문에, $Phase$ 값은 품종별로 설정된다. p 는 엽 생장에 따른 일사량 흡수 효과를 나타내기 위한 모수로, 엽 생장 시기에는 $-0.01(p_{gro})$ 을, 노화 시기에는 $0.01(p_{sen})$ 을 기본값으로 사용한다. 작물 재배 기간에 걸쳐 $fPAR_t$ 값은 다음과 같이 계산된다(Fig. 1):

$$fPAR_t = \max\{fPAR(cGDD, Phase_{veg}, p_{gro}), fPAR(cGDD, Phase_{rep}, p_{sen})\} \text{ (Eq. 4)}$$

Fig. 1. Example of dynamics of fractional Photosynthetically Active Radiation(fPAR). cGDD indicates cumulative growing degree days from the transplanting date.



SIMPLE 작물모형은 환경조건에 따라 광합성이 촉진되거나 제약받는 현상을 모의하기 위해 광합성 변동 계수를 일 단위로 계산한다. 이산화탄소, 온도 및 환경 스트레스가 광합성에 영향을 미치기 때문에, 이들 개별 요인들을 고려하여 V_t 값을 다음과 같이 계산한다:

$$V_t = (1 + E_{CO_2}) \cdot E_T \cdot E_{ST} \text{ (Eq. 5)}$$

여기서 E_{CO_2} 및 E_T 는 각각 이산화탄소 농도와 온도가 광합성에 영향을 미치는 정도를 나타내는 퍼지숫자(Fuzzy number)이다. 퍼지숫자는 해당 변수값을 0과 1 사이의 정도를 나타내는 멤버십 함수를 사용하여 정의된다. 이산화탄소와 온도가 증가함에 광합성량이 변동하는 정도를 나타내기 위해 E_{CO_2} 와 E_T 의 멤버십 함수는 Fig. 2와 같이 정의된다. E_{ST} 는 고온(ST_H)과 토양 수분(ST_W)에 의한 스트레스가 광합성에 미치는 영향을 모의하기 위해 다음과 같이 계산된다(Zhao et al. 2019):

$$E_{ST} = \min\{ST_H, ST_W\} \text{ (Eq. 6)}$$

여기서 ST_H 와 ST_W 는 각각 최고기온과 가뭄 조건을 나타내는 지수인 $ARID$ (Agricultural Reference Index for Drought) 지수를 활용하여 계산된다(Fig. 3). $ARID$ 지수는 Woli et al.(2012)에 의해 다음과 같이 제안되었다:

$$ARID_t = 1 - TR_t/ET_t \text{ (Eq. 7)}$$

여기서 TR 와 ET 는 각각 증산량(mm)과 잠재증발산량(mm)을 나타낸다. 만약 토양 수분이 부족하여 증산량이 0이 되는 경우 $ARID$ 지수값은 1이 되며, 이는 수분 스트레스의 최대치를 나타낸다. 반대로, 토양 수분이 충분한 조건에서 증산량이 잠재증발산량과 일치하게 될 경우 $ARID$ 값이 0이 될 수 있으며, 이때에는 수분 스트레스가 없다는 것을 의미한다.

Fig. 2. The membership functions of (A) E_{CO_2} and (B) E_T which represent the effects of CO_2 and temperature on photosynthesis, respectively.

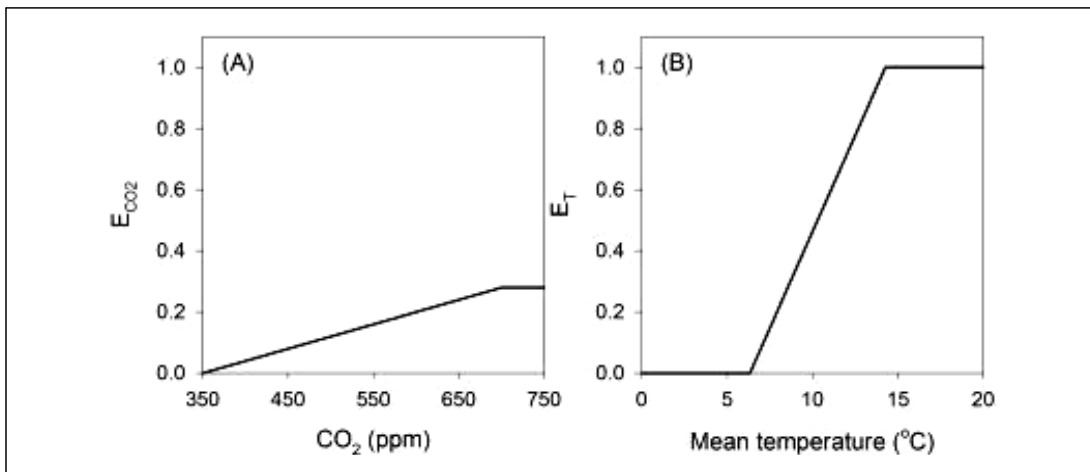
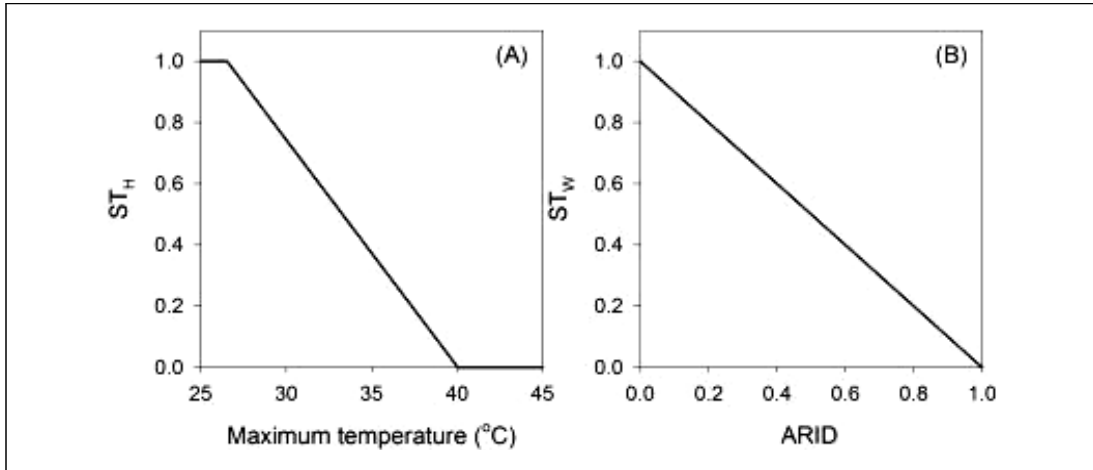


Fig. 3. The membership function of (A) ST_H and (B) ST_W which represent heat and water stress, respectively.



증산량은 기상조건과 토양 및 작물의 특성을 고려하여 계산된다. TR_t 값은 기상조건에 의해 결정되는 잠재증발산량과 토양의 물리적 특성 및 작물의 생리적 특성에 영향을 받는 토양 수분 흡수량을 사용하여 다음과 같이 계산된다(Zhao et al., 2019):

$$TR_t = \min\{ET_t, UP_t\} \quad (\text{Eq. 8})$$

여기서 UP 는 작물의 토양 수분 흡수량을 나타낸다. SIMPLE 작물모형은 일사량, 온도, 습도 및 풍속 자료를 입력자료로 하여 계산된 잠재증발산량으로 ET_t 값을 결정한다. 특히, Allen et al.(2006)이 제안한 FAO-56식으로부터 ET_t 값을 다음과 같이 계산한다:

$$ET_t = \frac{0.408\Delta(Rn - G) + \gamma\left(\frac{900}{T + 273}\right)\mu_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34\mu_2)} \quad (\text{Eq. 9})$$

여기서 Rn 과 G 는 각각 순복사량($\text{MJ m}^{-2} \text{ day}^{-1}$)과 토양 열 유속 밀도($\text{MJ m}^{-2} \text{ day}^{-1}$)를 나타낸다. 토양으로 열이 전달되는 정도를 나타내는 G 의 경우, 일사량에 비해 상대적으로 작은 값을 가지기 때문에, G 값을 무시하고 ET 값을 계산하였다. T 와 μ_2 는 각각 기온($^{\circ}\text{C}$)과 2m 높이

에서의 일 평균 풍속($m\ s^{-1}$)을 나타낸다. e_s 와 e_a 는 각각 포화 수증기압(kPa) 및 실제 수증기압(kPa)을 나타낸다. 또한, λ 와 γ 는 각각 증기압곡선의 기울기($kPa\ ^\circ C^{-1}$) 및 건습계 상수($kPa\ ^\circ C^{-1}$)이다. UP_t 값은 다음과 같이 계산된다(Woli et al., 2012):

$$UP_t = SW_t \cdot UC \text{ (Eq. 10)}$$

여기서 SW_t 와 UC 는 각각 일별 토양 수분과 작물의 토양 수분 흡수 효율을 나타내는 계수이다.

생물리학적 작물모형들은 일반적으로 층위별 토양 수분을 계산하는 반면, SIMPLE 작물모형에서는 전체 토양층의 물수지가 계산된다. 특정일의 토양수분 SW_t 은 기상 및 재배관리 조건과 토양의 물리적 특성을 고려하여 다음과 같이 정의된다:

$$SW_t = SW_{t-1} + P_t + I_t - TR_t - D_t - R_t \text{ (Eq. 11)}$$

여기서 P_t 와 I_t 는 각각 강수량 및 관개량을 나타낸다. SW_t 를 계산하기 위해 포장 내 수평적인 이동이 없는 것으로 가정된다. D_t 는 배수량(Drainage)을 나타내며, 다음과 같이 계산된다(Woli et al., 2012):

$$D_t = DDC \cdot RZD \cdot (SW_t^{bd} - PAWC) \text{ (Eq. 12)}$$

여기서 DDC , $PAWC$ 및 RZD 는 각각 배수계수(Deep Drainage Coefficient)와 토양 수분 보유능(Water Holding Capacity) 및 근권심(Root Zone Depth)을 나타낸다. 작물의 종류에 따라 RZD 값을 달리 사용될 수 있으나, 본 연구에서는 400mm를 기본값으로 설정하였다(Woli et al. 2012). SW_t^{bd} 는 배수되기 전 토양수분을 나타낸다. R_t 은 유출량(Run-off)을 나타내며 다음과 같이 계산된다(NRCS, 2004):

$$R_t = (P_t - I_a)^2 / (P_t - I_a + S) \text{ (Eq. 13)}$$

여기서 I_a , S 및 RCN 은 각각 초기감소(initial abstraction), 침투가능량(potential maximum

retention) 및 유출곡선지수(Runoff Curve Number)를 나타낸다. I_a 및 RCN 은 토양의 특성에 영향을 받는 모수이며, 각각 0.2와 65를 기본값으로 사용할 수 있다(Gijsman et al., 2007). S 값은 다음과 같이 계산된다(Hong et al., 2010):

$$S = 25400/RCN - 254 \text{ (Eq. 14)}$$

2.2. 마늘과 양파의 생구중 추정 모형

구 비대기 기간 동안의 일별 생구중값을 계산하기 위해 생물계절을 추정하는 모듈을 SIMPLE 작물모형에 추가하였다. 벼, 콩 및 옥수수과 같은 곡물 작물들은 수량이 형성되는 개화기 이후 등숙과정을 모의하기 위해 영양생장과 생식생장을 구분하여 생육단계를 추정한다. 마늘과 양파의 경우, 개화기 이전과 이후에 따라 형성되는 영양생장기와 생식생장기보다는 월동 단계 및 구 비대기를 구분하여 생육을 추정하는 것이 생산량 예측에 유리하다. 따라서 SIMPLE 작물모형에 구 비대기 단계에서 구의 생육 양상을 모의하기 위한 모듈이 추가되었다.

구 비대기의 시작 시점은 일장과 온도에 영향을 받는다. 이러한 요인들을 고려하여 구 비대기 단계를 모의할 수 있는 수식들로 생물계절 모듈을 구성하였다. 양파와 마늘 군락 내에서 구가 형성되는 단계로 전이되는 개체수는 환경조건에 따라 점진적으로 증가한다. 이러한 관계를 나타내기 위해 로지스틱 함수를 사용하여, 군락 내에서 상대적으로 구가 형성된 정도를 나타내는 F_t 를 다음과 같이 정의한다:

$$F_t = \kappa \cdot \exp[-\exp\{-\psi^*(GS_t - \tau)\}] \text{ (Eq. 15)}$$

여기서 κ , ψ , τ 는 경험적으로 얻어지는 모수이다. 구 발육단계를 나타내는 변수인 GS 는 다음과 같이 계산된다(De Visser, 1994):

$$GS_t = \sum GDD \cdot DL_t \text{ (Eq. 16)}$$

여기서 GDD 와 DL_t 는 각각 생장도일과 일장에 의한 구 발육단계 진전 정도를 나타낸다. DL_t 값은 다음과 같이 계산된다(De Visser, 1994):

$$DL_t = \max\{0, (L_t - TH) / TH\} \quad (\text{Eq. 17})$$

여기서 L_t 와 TH 는 각각 일장과 구 발육단계로 진전하기 위해 요구되는 최소 일장을 나타낸다.

작물 군락에서 구가 형성된 개체들의 상대적인 빈도와 구의 생체중이 증가하는 경향을 동시에 고려하여 단위면적당 수량이 계산된다. 단위면적당 생구중을 나타내는 $Yield(\text{kg ha}^{-1})$ 는 다음과 같이 정의된다:

$$Yield_t = B_t \cdot F_t \cdot WF^{-1} \quad (\text{Eq. 18})$$

여기서 WF 는 건물중에서 생체중으로 전환하기 위해 사용되는 생구의 수분함량지수이다.

2.3. 전역모수 추정기법을 활용한 농가별 모수 추정

작물 생육 모형의 불확도는 기상입력자료 또는 모수와 같은 입력요소에 의해 발생할 수 있다. 또한, 작물모형을 구성하는 수식이 현실을 충분히 현실을 반영하지 않을 경우에도 불확도가 발생할 수 있다. 이러한 요인들이 작물 생육을 일단위로 모의하는 과정에서 상호작용을 하게 되어 하나의 요인에서 유래된 불확도가 다른 요인에 의해 발생한 불확도를 증가시키거나 오히려 감소시킬 수 있다. 따라서 분석 대상 요인이 발생시킨 불확도 크기를 파악하기 위해 다른 요인들로부터 생겨나는 불확도를 최소화하는 것이 필요하다.

품종모수에 의한 불확도와 미래 기상 자료를 사용함에 따라 발생하는 불확도를 분리하여 분석하기 위해 농가별로 모수를 추정하였다. 작물모형을 구동할 때, 일반적으로 작물이나 품종별로 추정된 모수가 사용된다. 예를 들어, 관측된 생체중값과 모의값의 오차를 최소화할 수 있는 RUE값을 작물별로 추정한다. 이러한 모수 추정 과정에 상세한 생육 관측자료뿐만 아니

라 개별 농가별로 사용된 시비량이나 관개량과 같은 상세한 재배관리 자료가 필요하다. 그러나 국내에서 농가별로 수집된 재배관리 자료에는 구체적인 재배관리 정보가 포함되어 있지 않다. 이러한 제약 요건들을 감안하여, 생육 관측자료가 수집된 농가들을 대상으로 모수를 추정하여 재배관리에 따른 작물 생육의 변동 양상을 간접적으로 작물 생육 모의 과정에 반영하였다. 특히, 농가별 모수를 사용하여 생산성을 예측할 경우, 작물모형의 모수에 의해 발생할 수 있는 작물 생산성 예측값의 불확도를 최소화할 수 있다.

생물리화학적 모형들에 사용되는 모수의 수가 비교적 많기 때문에, 모수 추정을 위해 상당한 양의 관측자료가 사용되어야 한다. 예를 들어, 벼의 생육 모의에 사용되는 ORYZA2000 모형이나 CERES-Rice 모형은 각각 70개 및 17개 이상의 모수가 요구된다. SIMPLE 모형은 이러한 모형보다 적은 13개의 모수가 사용된다. 그러나 여전히 생육 단계별 관측된 건물중 자료를 사용하는 것이 모수 추정에 유리하다. 그러나 국내에서는 광범위한 지역에서 상당 기간 수집된 생육 관측자료가 부족한 실정이다(Hyun and Kim, 2019).

본 연구에서는 비교적 적은 양의 관측 자료가 사용되더라도 대량의 전산 자원을 사용하여 신뢰도 높은 모수를 추정할 수 있는 도구인 Quantification of Uncertainty Estimation, Simulation and Optimization(QUESO)를 사용하였다(Prudencio and Schulz, 2011). QUESO는 전역모수추정 기법인 Markov Chain Monte Carlo(MCMC) 기반으로 모수들을 추정하기 위해 개발된 소프트웨어 개발 도구이다. Hyun and Kim(2019)는 QUESO를 사용하여 ORYZA2000 모형의 모수를 추정하였다. 본 연구에서는 생구중의 시계열 관측값과 예측값을 비교하여 얻어진 Likelihood를 최대화하는 방식으로 SIMPLE 모형의 모수를 추정하였다.

모수 추정을 위한 생육 관측자료 및 기상자료는 한국농촌경제연구원에서 운영하는 농업관측통계시스템(<https://oasis.krei.re.kr>)에서 수집하였다. 농업관측통계시스템 홈페이지에는 개별 농가별로 수집된 생구중 시계열 자료가 제공된다. 본 연구에서는 2020년에 수확된 양파와 마늘들을 대상으로 개별 농가에서 얻어진 재배관리 정보와 생구중 관측값을 수집하였다. 기상자료의 경우, 국립농업과학원에서 운영하는 농업날씨 365(<https://weather.rda.go.kr>) 서비스에서 제공되는 공간자료로부터 생성된 일별 기상자료를 사용하였다. 이들 자료는 기상 관측소에서 측정된 기상요소들을 공간적으로 추정하는 소기후 모형을 사용하여 생산되었다.

본 연구에서는 생육 관측자료가 수집된 농가별로 지점별 기상자료를 농업관측통계시스템 홈페이지를 통해 수집하였다.

2.4. 과거와 현재 기상자료를 조합한 기상입력자료 생성

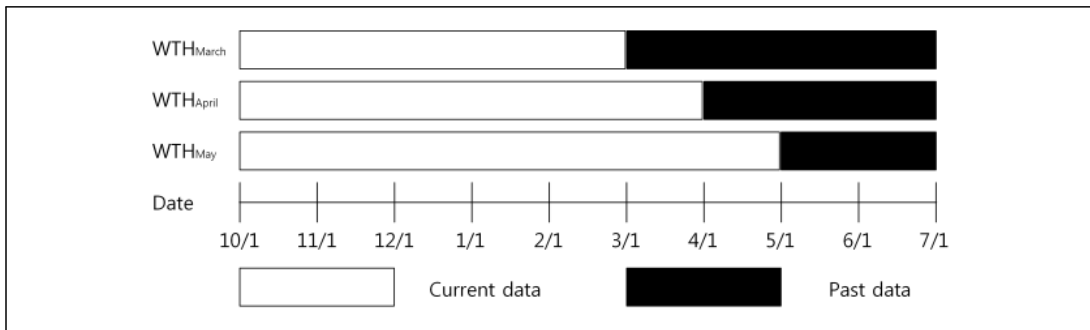
과거와 현재 기상자료를 조합하여 최대 3개월 이전에 작물 생산성을 예측하기 위한 작물모형의 기상입력자료를 생성하였다. 기상 예측 자료를 생산하기 위해 사용되는 수치기상모형의 경우, 대부분 3일 이내의 단기 예측 자료나 최대 10일 정도의 중기 예보자료가 생산된다(Moon et al., 2020; Lee et al., 2016). 전국 수준의 공간 규모에서 수치기상모형을 운영하여 생산되는 수개월 이상의 기상 전망자료들은 온도와 강수의 변동성만이 일반에게 공개된다(Hong et al., 2013). 따라서 대규모 수치기상모형을 운영하는 일부 기관을 제외하고 일사량을 포함하는 다양한 기상요소에 대한 장기 전망자료를 수집하는 것은 어렵다. 반면, Lee et al.(2015)는 장기 전망자료를 대신하여 비교적 손쉽게 확보할 수 있는 과거 기상자료를 예측 시점부터 수확 예상 시기까지의 기상자료로 사용할 경우, 비교적 정확하게 생산성을 전망할 수 있는 것으로 보고하였다.

본 연구에서는 구의 생체중 예측일 이전까지는 현재 기상자료를 사용하고, 그 이후에는 과거 기상자료를 사용하여 SIMPLE 작물모형을 구동하기 위한 기상입력자료를 생성하였다(Fig. 4). 예를 들어, 수급조절을 위해 양파나 마늘의 생산성 예측자료가 필요한 3월부터 5월까지 매월 1일에 생구중을 예측하기 위해, 정식일 이후 생구중 예측일 이전까지 수집된 현재 기상자료를 사용하여 기상입력자료를 생성하였다. 또한, 각 월의 1일인 생구중 예측일부터 수확 예상 시점까지 이전 년도에 얻어진 과거 기상자료를 생성된 기상입력자료에 추가하였다.

농가 단위의 과거 및 현재 기상자료를 수집하기 위해 농촌진흥청에서 운영하고 있는 농업기상재해 조기경보시스템(<http://agmet.kr>)을 사용하였다(Kim et al., 2020a). 이러한 조기경보시스템은 일부 지역을 대상으로 2016년부터 현재 시점까지 과거 4년 기간 동안 일 단위 기상자료들을 제공하고 있다. 본 연구에서는 농가별로 2019년도 정식일부터 2020년 6월의 수확 예상 시기까지 현재와 이전 년도의 기상자료를 조합하여 예측일별로 각각 4개의 기상

입력자료를 생성하였다. 또한, 과거 기상자료를 미래 기상자료를 대신하여 사용함에 따라 발생한 불확도를 분석하기 위해 정식일 이후 수확 시점까지 관측된 현재 기상자료를 사용하여 추가적으로 기상입력자료를 생성하였다.

Fig. 4. A simple representation of weather input data prepared using current and past observation data. WTH_{March} , WTH_{April} and WTH_{May} represent the weather input data used to perform crop growth simulations on given prediction dates including March 1st, April 1st and May 1st, respectively.



2.5. 작물모형 구동 및 평가

본 연구에서는 양파와 마늘의 주산지 중에서 생구중 관측자료와 함께 기상자료를 수집할 수 있는 농가들을 대상으로 SIMPLE 모형을 구동하였다(Table 1). 양파는 전라북도 고창군, 부안군, 완주군 및 경상남도 함양군에 위치한 농가들이 모의 대상에 포함되었다. 마늘의 경우, 전라북도 부안군과 완주군에 위치한 농가들이 모의 대상에 포함되었다. 품종 모수 사용에 따른 불확도를 고려하여 현재와 과거 자료가 조합된 기상입력자료를 사용하였을 때 발생할 수 있는 불확도를 평가하기 위해 모수 추정 과정에서 최대오차와 최소오차가 나타났던 농가들을 대상으로 분석을 수행하였다.

개별 농가별로 과거와 현재 기상자료를 조합하여 얻어진 4개의 기상입력자료를 사용하여 각 예측 기준일별로 SIMPLE 작물모형을 구동하였다. 예를 들어, 2020년 3월 1일, 4월 1일

및 5월 1일을 기준으로 하여 생성된 기상입력자료를 사용하여 생구증을 예측하였다. 기상입력자료 생성에 사용된 과거년도를 y 라고 할 때, 이들 생구증 예측값을 PP_y 으로 표기하였다. 모형 구동에 요구되는 토양 특성과 관련된 모수들은 Woli et al.(2012)에서 제시된 기본값을 사용하였다. 관개량인 I_t 값은 증산이 억제되지 않도록 일별로 토양 수분 부족분을 반영하여 설정되었다. 정식일은 관측된 재배관리 자료를 활용하여 설정하였다.

Table 1. The number(N) of farms where cultivar parameters of the SIMPLE crop model were calibrated

Onion		Garlic	
Region	N	Region	N
Gochang	4	Buan	3
Buan	3	Wanju	4
Wanju	4		
Hamyang	6		

현재 기상자료만을 사용하여 생구증을 추정하고 생구증 예측값과 관측값의 시계열 자료를 분석하였다. 현재 기상자료를 사용하여 얻어진 생구증 추정값은 PE 로 표기하였다. 과거와 현재 기상자료를 조합하여 생성된 기상입력자료를 사용하여 발생한 생구증 예측값의 불확도를 분석하기 위해, PP_y 값의 최대(PP_x), 최소(PP_n) 및 평균값(PP_m)으로 구성된 시계열 자료와 PE 값을 비교하였다. 또한, 생구증이 관측된 일자를 기준으로 PE 와 PP_m 값의 차이를 비교하였다.

2.6. 예측값의 일치도 분석

모수 추정 과정에서 최대오차와 최소오차가 되는 농가를 선정하기 위해 시기별 생구증 추정값과 관측값 사이의 오차를 계산하였다. 이를 위해 Root Mean Square Error(RMSE)를 다음과 같이 계산하였다:

$$RMSE = \sqrt{\frac{\sum (O_i - S_i)^2}{N}} \quad (\text{Eq. 19})$$

여기서, O 와 S 는 각각 i 번째 시기의 생구중에 대한 관측값과 모의값을 나타낸다. N 은 비교에 사용된 관측값의 개수를 나타낸다. 생구중은 시기별, 농가별로 상이한 값을 가지기 때문에 오차의 절대값을 비교하는 것에는 한계가 있다. 따라서 오차를 상대적으로 비교하기 위해 정규화된 지표인 Normalized Root Mean Square Error(NRMSE)를 다음과 같이 계산하였다:

$$NRMSE = \frac{RMSE}{\bar{O}} \quad (\text{Eq. 20})$$

여기서 $\bar{O}$ 는 생육 관측값의 평균을 나타낸다.

예측 시기에 따른 과거 기상자료를 사용하여 얻어진 생구중 추정값(PP)의 불확도를 평가하기 위하여 현재 기상자료만을 사용하여 예측된 생구중(PE)와의 일치도를 비교하였다. 이를 위해 관측값이 측정된 시기에 해당하는 PE 및 PP_m 값을 추출하고, 이를 활용하여 결정계수(R^2) 및 일치 지수(d)를 다음과 같이 계산하였다(Krause, 2005):

$$R^2 = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum (y_i - \bar{y})^2} \sqrt{\sum (x_i - \bar{x})^2}} \quad (\text{Eq. 21})$$

$$d = 1 - \frac{\sum (y_i - x_i)}{\sum (|x_i - \bar{y}| + |y_i - \bar{y}|)^2} \quad (\text{Eq. 22})$$

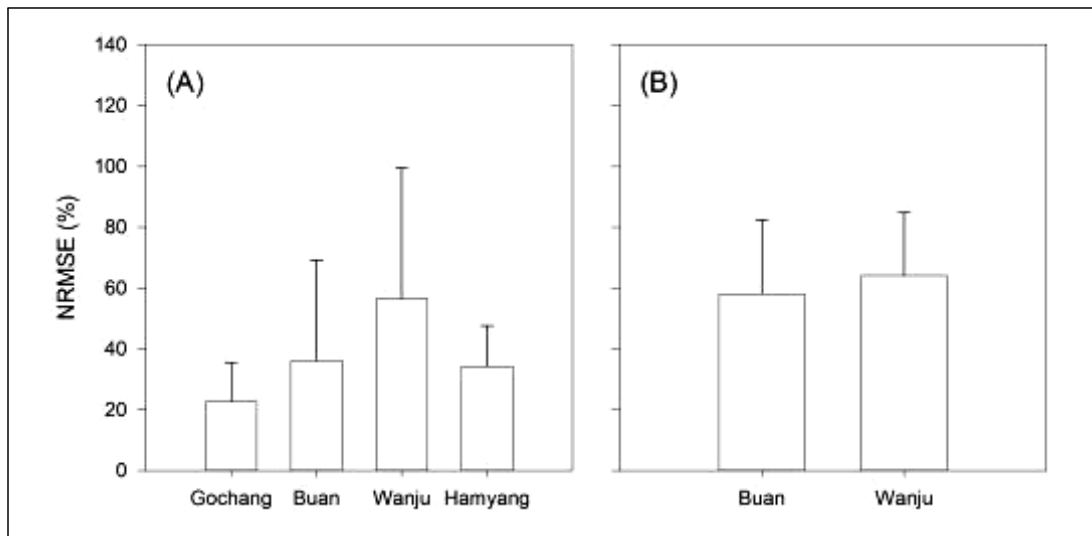
여기서 x_i 및 y_i 는 각각 관측값이 측정된 시기에 얻어진 PP_m 값과 PE 값을 나타낸다. x 및 y 는 각각 비교에 사용된 PP_m 값들과 PE 값들의 평균값을 나타낸다.

3. 결과

3.1. 모수 추정 결과

농가별로 모수를 추정한 결과, 농가가 위치한 지역에 따라 관측값과 모의값 사이의 오차가 상이하게 나타났다(Fig. 5). 예를 들어, 개별 농가별로 추정된 모수를 사용하여 얻어진 양파의 생구중에 대한 평균적인 NRMSE는 고창군과 완주군에서 각각 23%와 56%이었다. 마늘의 수량에 대한 NRMSE는 부안군과 완주군에서 각각 58% 및 64%로 지역별로 큰 차이는 없었으나 전반적으로 양파와 비교하여 오차가 큰 경향이 있었다. 이들 중 오차가 가장 작았던 농가는 양파와 마늘의 경우 각각 고창군과 부안군에 위치하였다. 반면, 두 작물 모두 완주군에 위치한 농가에서 모수추정과정에서 얻어진 생구중의 추정 오차가 가장 크게 나타났다.

Fig. 5. Central tendency and dispersion of the error obtained from the parameter calibration process by individual farms where (A) onion and (B) garlic were grown, respectively. NRMSE represents normalized root mean square error.



3.2. 양파의 생구중 예측

양파에 대하여 과거 기상 자료를 활용하여 수확 이전에 수량을 예측한 값(PP_i)은 생육 후반까지는 관측값과 유사한 경향이 있었다(Fig. 6). 특히, 모수 추정 과정에서 오차가 가장 작았던 농가에서는 수확기 시점까지 비교적 관측값과 유사한 생구중이 예측되었다. 반면, 모수 추정 과정에서 오차가 큰 지역에서는 5월 중순부터 생구중의 예측값과 관측값의 차이가 크게 발생하였다. 예를 들어, 고창군의 농가에서 5월에 예측된 생구중 값은 마지막 관측값과 비교하여 5% 정도의 오차가 발생하였다. 반면, 모수를 추정하였을 때 생구중 추정 오차가 가장 큰 것으로 나타났던 완주군의 농가에서는 생구중 예측값이 49% 정도의 비교적 큰 오차를 가졌다.

현재와 과거 기상자료를 조합하여 예측된 생구중 값은 수확기 시점까지 측정된 기상자료를 사용하여 추정된 값과 유사하였다(Fig. 7). 특히, 모수 추정 과정에서 발생한 오차 크기와 상관없이 이러한 경향이 나타났다. 예를 들어, 고창군의 농가에서 과거 자료를 사용하여 수확기 이전인 3월, 4월 및 5월에 예측된 수량은 현재 기상자료를 사용하여 추정된 수량의 변이를 각각 97.8, 97.7 및 99.5%를 설명하였다. 또한, 동일한 조건에서 완주군의 농가에서 예측된 값은 대략 99.9% 이상의 변이를 설명하였다.

Fig. 6. The outcome of onion growth simulations performed at a farm in (A, C, E) Gochang and (B, D, F) Wanju. It was assumed that the prediction was made on (A, B) March 1st, (C, D) April 1st and (E, F) May 1st in 2020. $[PP_n, PP_x]$ indicates a range of simulated fresh bulb weights using past weather data (PP_y). PP_n , PP_x and PP_m indicate minimum, maximum and average of PP_y , respectively. PE and OBS represent bulb weight estimates using current weather data and observed bulb weight, respectively.

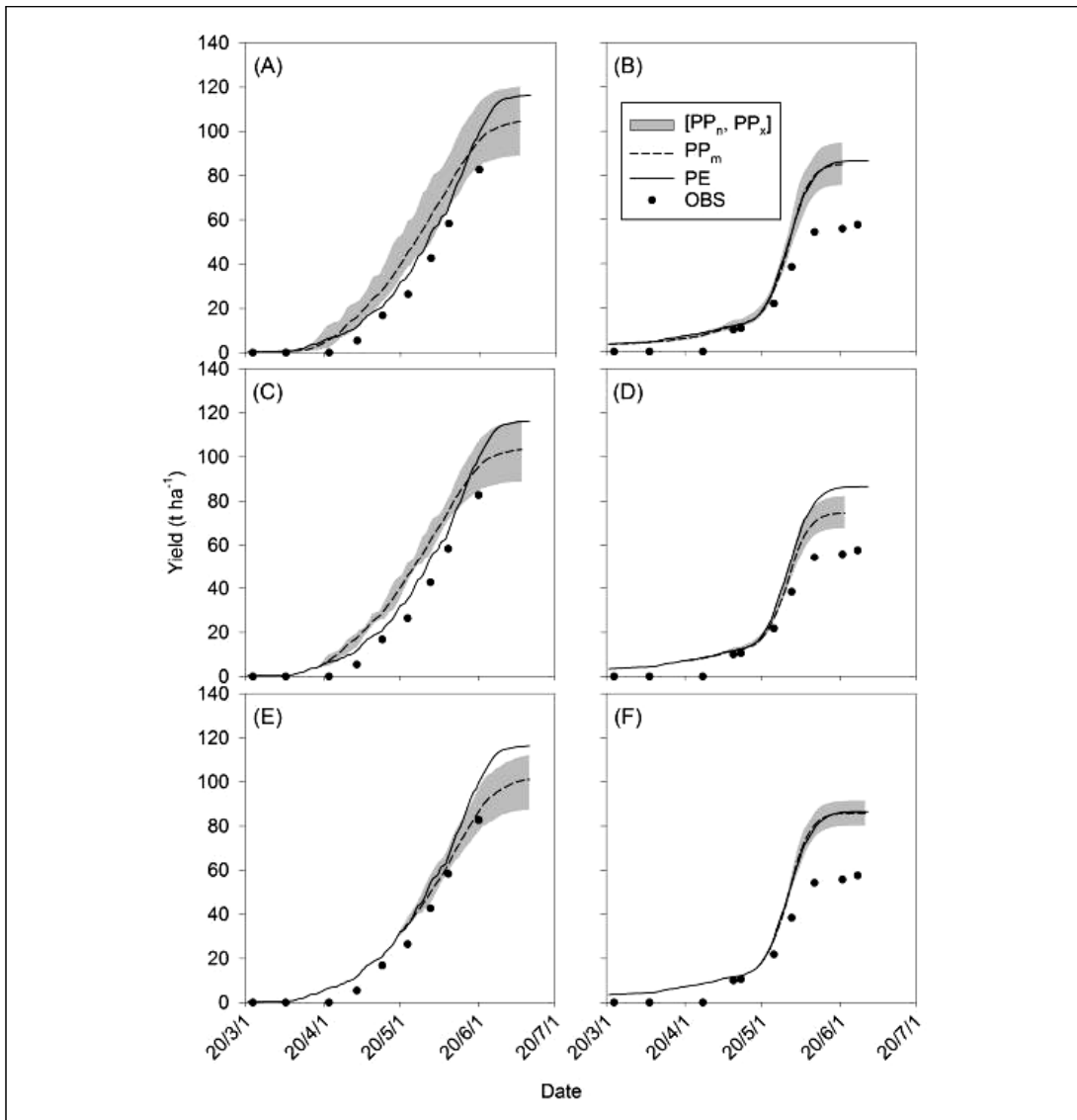
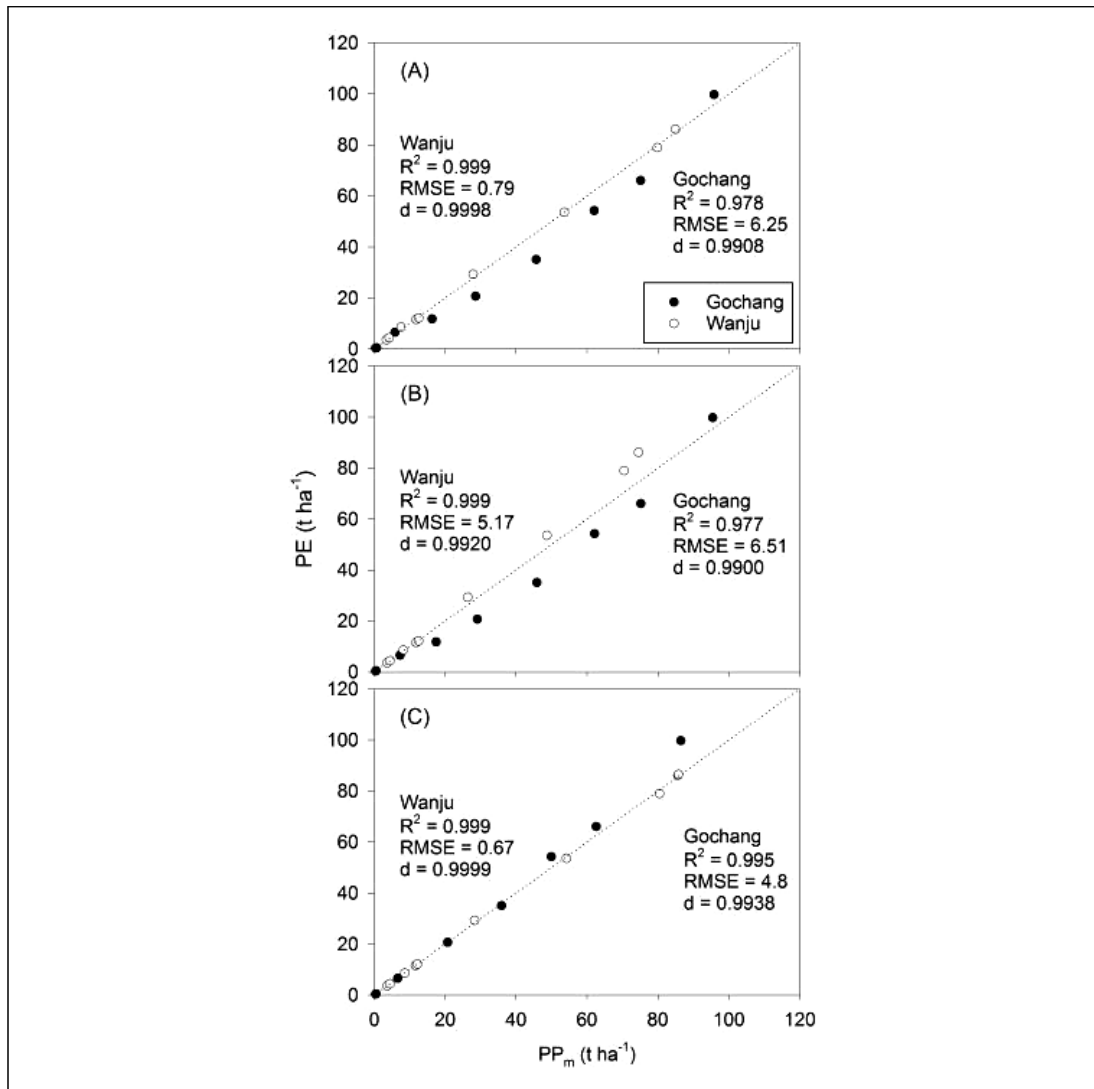


Fig. 7. Comparison between PP_m and PE of onion on the dates when the bulb weight was measured. It was assumed that prediction was made on (A) March 1st, (B) April 1st and (C) May 1st in 2020 for PP_m . PP_m represents the average of bulb weight predicted combining weather data in current and previous years. PE indicates bulb weight estimates using weather data obtained after harvest. R^2 , RMSE, and d represent coefficient of determination, root mean square error, and degree of agreement, respectively.



3.3. 마늘의 생구중 예측

마늘을 대상으로 현재와 과거 기상자료를 조합하여 수확기 이전에 생구중을 예측하였을 때, 구가 형성된 이후 기간 동안 예측값과 관측값이 유사한 경향이 있었다(Fig. 8). 예를 들어, 3월부터 과거자료를 사용하여 모의를 한 경우, 구가 형성되기 시작한 4월 중순의 예측값은 관측값의 20% 수준으로 오차가 발생하였다. 생구중 뿐만 아니라 수확기에서도 예측값과 관측값에 차이가 나타났다. 특히, 예측시점에 따라 수확시기의 오차에 차이가 있었다. 예를 들어, 부안군 농가에서는 관측시기보다 일찍 수확된 것으로 모의된 반면, 완주군 농가에서는 관측시기보다 상당히 늦게 수확되는 것으로 모의되었다. 또한, 수확시기의 차이는 수확기 직전의 생구중 예측값에 상당한 크기의 오차를 초래하였다. 예를 들어, 완주군 농가에서의 수량은 마지막 관측 시기에 60%의 오차가 발생하였으나, 생육 모의 과정에서 수확시기가 지연되어 최종 예측 수량과 관측된 수량의 차이는 260%에 달하였다.

수확기 시점까지 측정된 기상자료를 사용하여 추정한 마늘의 생구중(PE)과 과거와 현재 기상자료를 사용하여 예측한 생구중(PP_m)이 재배기간에 걸쳐 비슷한 경향이 나타났다(Fig. 9). 두 지역에서 마늘의 구 형성 이전 시기에 관측값과는 큰 차이를 보였으나, PE 값과 PP_m 값의 차이가 크지 않았다. 또한, 구 형성 이후에도 모수 추정에 큰 오차가 있었던 지점에서는 모의값과 관측값에 상당한 차이가 있었으나 PE 값과 PP_m 값은 유사하게 변동하였다. 반면, PP_m 값의 경우 생육이 일찍 종료되는 경향이 나타났다. 예를 들어, 부안 지역에 대해서 3월부터 과거 자료를 사용할 경우 5월 7일부터 5월 20일 사이에 생육이 종료되었다. 그러나, 예측시점이 늦어질수록 수확 시기가 지연되어, PE 값과 PP_m 값의 차이가 상당히 감소하였다.

Fig. 8. The outcome of garlic growth simulations performed at a farm in (A, C, E) Buan and (B, D, F) Wanju. It was assumed that the prediction was made on (A, B) March 1st, (C, D) April 1st and (E, F) May 1st in 2020. $[PP_n, PP_x]$ indicates a range of simulated fresh bulb weights using past weather data (PP_y). PP_n , PP_x and PP_m indicate minimum, maximum and average of PP_y , respectively. PE and OBS represent bulb weight estimates using current weather data and observed bulb weight, respectively.

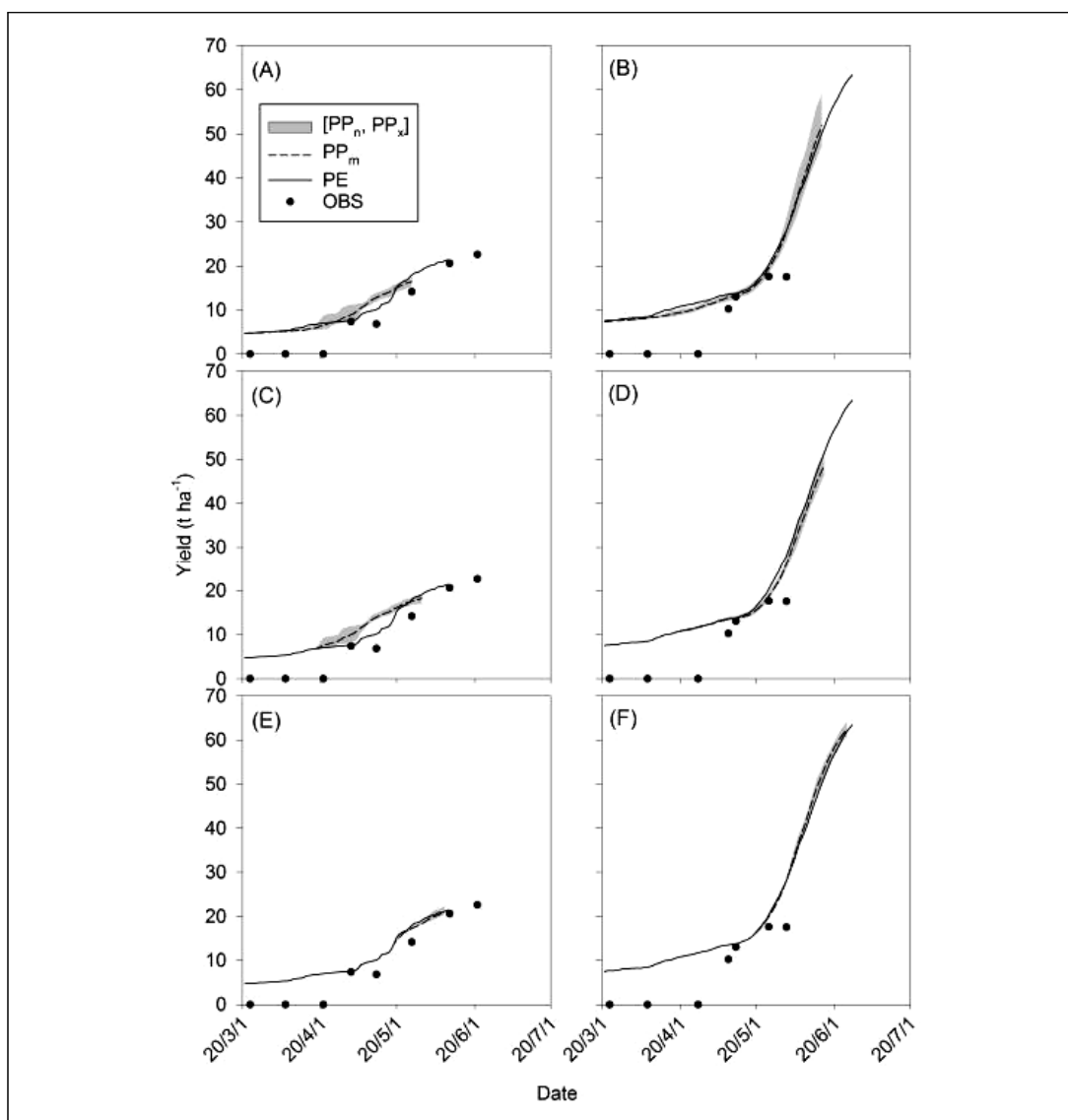
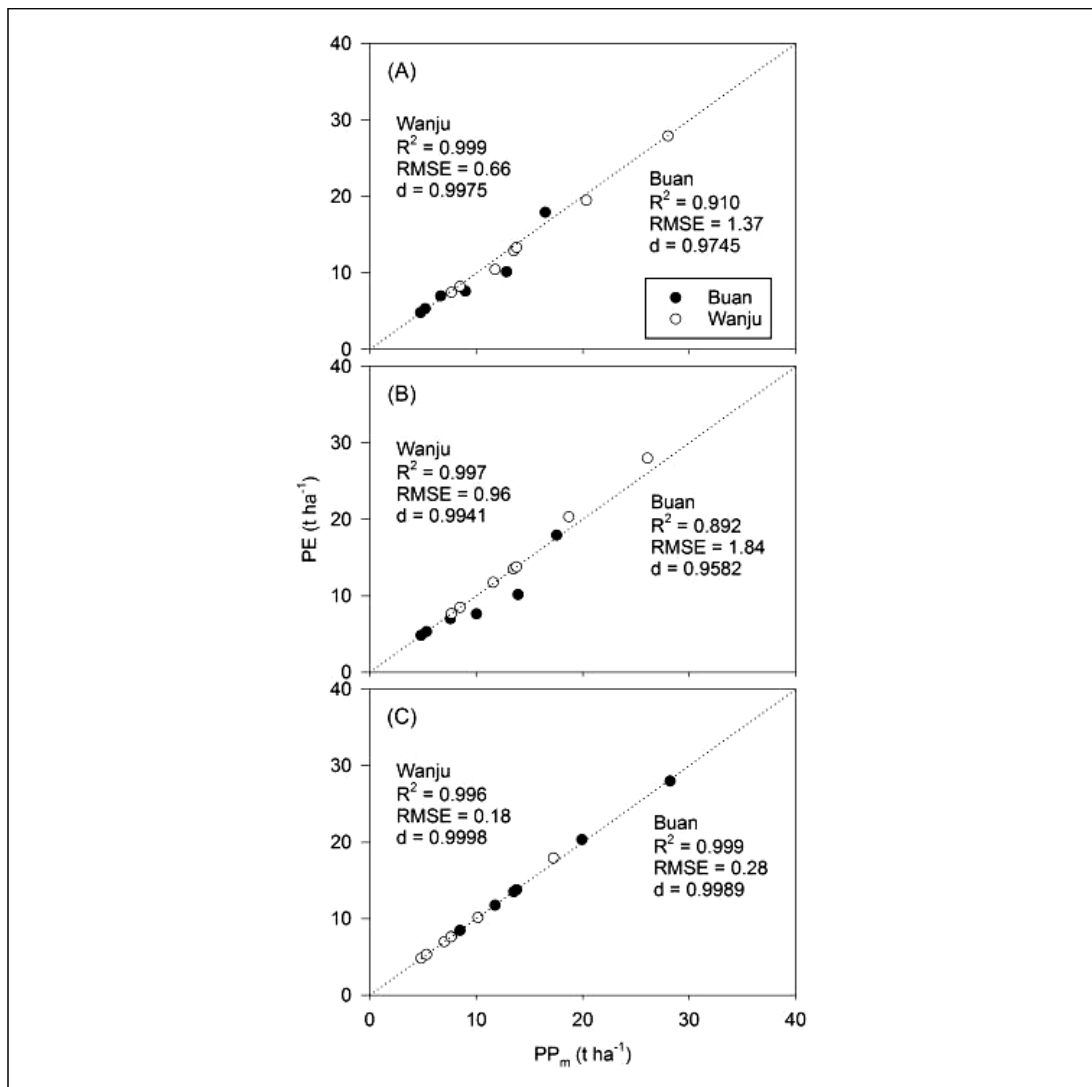


Fig. 9. Comparison between PP_m and PE of garlic on the dates when the bulb weight was measured. It was assumed that prediction was made on (A) March 1st, (B) April 1st and (C) May 1st in 2020 for PP_m . PP_m represents the average of bulb weight predicted combining weather data in current and previous years. PE indicates bulb weight estimates using weather data obtained after harvest. R^2 , RMSE, and d represent coefficient of determination, root mean square error, and degree of agreement, respectively.



4. 고찰 및 결론

본 연구에서는 현재와 과거 자료를 조합하여 생성된 기상자료를 사용하여 생구중을 예측할 경우, 관측된 기상자료만을 사용하여 생구중을 추정하는 것과 유사한 결과를 얻을 수 있다는 것을 보여주었다. 미래 시점의 생산성을 예측하기 위해 과거 기상조건을 기반으로 기상 입력 자료를 사용하였을 때, 불확도가 발생하게 되는 것을 피할 수 없다(Basso and Liu, 2019). 본 연구에서도 농가와 예측 시점에 따라 예측값의 불확도 범위에 차이가 있었다. 그럼에도 불구하고, 생구중을 최대 60일 정도의 선행기간을 두고 예측하였을 때, 관측 기상자료를 입력자료로 사용하여 추정된 생구중의 변이를 89% 이상 설명하였다. 이러한 결과는 신뢰도 높은 품종모수가 SIMPLE 작물모형에 사용된다면, 농가 단위에서 실용적으로 활용 가능한 생구중 전망자료를 생산할 수 있다는 것을 시사하였다. 예를 들어, Archontoulis et al.(2020)는 현재와 과거자료를 조합하여 옥수수과 수량을 3개월 이전에 예측하여 최종 수량을 20% 이내의 오차로 모의하였다. 다만, 본 연구 결과는 기상자료 및 작물의 생육 관측자료가 수집이 가능한 하나의 작기에 적용이 되었기 때문에, 추가적인 자료확보를 통해 과거 기상자료를 활용하여 예측된 생구중 값들의 신뢰도 한계를 탐색해야 할 것으로 판단된다.

양파와 마늘의 생산성을 예측하기 위해, SIMPLE 모형의 신뢰도를 향상시킬 수 있는 고품질 관측자료들이 지속적으로 수집되어야 할 것으로 보인다. 작물모형의 품종모수 추정을 위해 일반적으로 작물의 부위별 건물중과 단위면적당 생산량이 사용된다(Hyun et al., 2020). 반면, 본 연구에서는 일정 시기별로 관측된 생구중값만이 모수추정을 위해 사용되었다. 이로 인해, 모의된 생육 기간과 실제 생육 기간이 차이나는 경우가 발생하였다. 모수 추정을 위한 최적의 관측자료는 농가별로 정밀하게 수집되는 것이 유리하나, 이를 위해 상당한 노동력과 비용이 요구된다(Kim et al., 2018). 모수 추정을 지원하기 위한 관측자료 데이터베이스를 효율적으로 구축하기 위해, 해당 작물의 재배단지가 형성되어 있는 지역에서 일부 농가를 중심으로 고품질 생육 관측자료를 수집하는 체계가 고려되어야 할 것이다. 또한, 양파와 마늘과 같은 노지 채소류 작물의 생육은 기상, 토양과 같은 자연적인 조건뿐만 아니라 품종과 파종일 및 정식일 등의 재배관리 요소에 의해 결정된다. 따라서 재배관리 요소를 포함하는 생육 관측

프로토콜을 구축하여 관측자료의 품질을 향상시키는 노력이 수반되어야 할 것이다(Kim et al., 2020b). 관측값의 품질을 높일 수 있는 프로토콜에 기반하여 자료들이 수집된다면 생물리화적인 과정을 모의하는 모형 뿐만 아니라 인공지능을 활용한 작물모형 개발을 지원할 수 있을 것이다. 특히, 최근 작물 생육 예측을 개선하기 위해 다양한 기계학습을 활용한 연구가 진행됨에 따라(Shahhosseini et al., 2021; Chlingaryan et al., 2018; Kim et al., 2019), 이를 위한 학습자료 및 검증 자료로 사용할 수 있는 관측자료들이 수집되어야 할 것이다.

본 연구 결과는 SIMPLE 모형으로 예측되는 생구중 값의 신뢰도를 높이기 위해 자료동화 알고리즘이 개발되어야 한다는 것을 시사하였다. 작물 모형은 주어진 조건에서 작물이 반응하는 양상을 모의하기 위해 간단한 형식의 수식들로 구성하여 개발된다(Kim et al., 2018). 따라서 작물모형은 모형 구동시 사용되는 초기값이나 입력자료의 불확도로 인해 현실과 동일한 수준으로 작물 생육을 재현하는 것은 어렵다. 이러한 문제점들을 개선하기 위해 센서로부터 측정된 관측값들을 모형 구동 과정에 활용하는 자료동화 기법들이 개발되어왔다. 예를 들어, Ban et al.(2019)는 Moderate Resolution Imaging Spectroradiometer(MODIS) 자료를 활용한 자료동화 알고리즘을 개발하여 CERES-Maize 모형의 수량 예측 정확도를 개선하였다. 그러나 국내에서는 자동관측이 가능한 센서보다는 초장과 엽수와 같이 노동력에 의존하여 측정되는 생육 요소들이 생육 관측자료로 수집되고 있다(Kim et al., 2020b). 특히, 초장이나 엽수와 같은 요소들은 생체중의 직접적인 측정값이 아니기 때문에, 자료동화에 활용되기 어렵다. 따라서 이러한 생육 관측요소들을 사용하더라도 생구중 예측값의 불확도를 최소화할 수 있는 자료동화 알고리즘들이 개발되어야 할 것이다.

추후 연구에서는 과거 년도 및 예측 시기에 따른 생육 예측값의 불확도 뿐만 아니라 입력자료 생성을 위해 사용되는 기상자료의 종류에 따른 불확도에 대한 평가가 이루어져야 할 것이다. 작물모형을 활용하여 개별 농가별로 단위면적당 생산량을 예측하기 위해 인근 기상 관측소에서 관측된 지점별 자료나 특정 지역을 대상으로 생성된 격자형 기상자료들이 사용될 수 있다. 예를 들어, 기상청에서 운영하고 있는 종관기상관측망이나 방재기상관측망 자료를 활용하여 작물모형의 기상입력자료를 생산할 수 있다(Lee and Oh, 2019). 그러나 인근 관측소의 기상자료를 사용할 경우 농가와 기상관측소 사이의 거리 및 지형적인 원인으로 인해 농가의 실제 기상조건을 대표하기 어려울 수 있다(Hyun et al., 2020). 본 연구에서 사용된 농

업기상재해 조기경보시스템에서 제공하고 있는 격자형 기상자료들은 서비스가 제공되는 지역 내에 모든 농가들을 대상으로 기상자료 제공이 가능하다. 그러나 격자자료를 생성하는 방법에 따라 기상요소별로 신뢰도의 차이가 발생하게 된다(Kim et al., 2016). 특히, 일반적으로 작물모형으로 얻어지는 생육 예측 결과는 입력자료로 사용된 기상자료의 불확도와 선형적인 관계를 보이지 않는다. 따라서 생구종을 예측하기 위해 작물모형의 기상입력자료로 사용될 수 있는 기상자료별로 생산성 예측값의 불확도를 파악하는 것이 선행되어야 할 것이다. 국내에서는 국가농림기상센터나 기상청 및 농촌진흥청 등 여러 기관에서 작물모형을 구동하기 위해 사용될 수 있는 기상자료들이 생산 및 제공되고 있다. 이러한 자료들을 수집하고 이들이 생구종 예측값의 불확도에 기여하는 정도를 분석하여 현업에서 사용될 수 있는 기상자료를 파악하고, 이들의 신뢰도를 개선하여 생구종 예측의 불확도를 줄여 나가야 할 것이다.

참고문헌

- Allen, R. G., W. O. Pruitt, J. L. Wright, T. A. Howell, F. Ventura, R. Snyder, D. Itenfisu, P. Steduto, J. Berengena and J. B. Yrisarry. 2006. "A recommendation on standardized surface resistance for hourly calculation of reference ETo by the FAO56 Penman-Monteith method." *Agricultural Water Management* 81(1-2): 1-22.
- Archontoulis, S. V., M. J. Castellano, M. A. Licht, V. Nichols, M. Baum, I. Huber, R. Martinez-Feria, L. Puntel, R. A. Ordóñez and J. Iqbal. 2020. "Predicting crop yields and soil-plant nitrogen dynamics in the US Corn Belt." *Crop Science* 60(2): 721-738.
- Ban, H.-Y., J.-B. Ahn and B.-W. Lee. 2019. "Assimilating MODIS data-derived minimum input data set and water stress factors into CERES-Maize model improves regional corn yield predictions." *PloS one* 14(2): e0211874.
- Basso, B. and L. Liu. 2019. "Seasonal crop yield forecast: Methods, applications, and accuracies." *Advances in Agronomy* 154: 201-255.
- Chlingaryan, A., S. Sukkarieh and B. Whelan. 2018. "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review." *Computers and electronics in agriculture* 151: 61-69.
- Choi, S. and J. Baek. 2016. "Onion yield estimation using spatial panel regression model." *Korean Journal of Applied Statistics* 29(5): 873-885.
- De Visser, C. 1994. "ALCEPAS, an onion growth model based on SUCROS87. I. Development of the model." *Journal of Horticultural Science* 69(3): 501-518.
- Gijsman, A. J., P. K. Thornton and G. Hoogenboom. 2007. "Using the WISE database to parameterize soil inputs for crop simulation models." *Computers and Electronics in Agriculture* 56(2): 85-100.
- Ha, J.-H., S.-T. Seo and S.-W. Kim. 2019. "Improving Forecasting Performance for Onion and Garlic Prices." *Journal of Korean Society of Rural Planning* 25(4): 109-117.
- Hong, H., W. Kim, J. Kim and B.-J. Kim. 2013. "Analysis of demand characteristics for long-term forecasts." *Journal of Climate Research* 8(2): 117-126.
- Hong, S.-Y., K.-H. Jung, C.-U. Choi, M.-W. Jang, Y.-H. Kim, Y.-K. Sonn and S.-K. Ha. 2010. "Estimation of SCS runoff curve number and hydrograph by using highly detailed soil map (1: 5,000) in a small watershed, Sosu-myeon, Goesan-gun." *Korean Journal of Soil Science and Fertilizer* 43(3): 363-373.
- Hur, J., K.-M. Shim, Y. Kim and K.-K. Kang. 2019. "Estimation of Rice Yield by Province

- in South Korea based on Meteorological Variables.” *Journal of the Korean earth science society* 40(6): 599-605.
- Hyun, S. and K. S. Kim. 2019. “Calibration of cultivar parameters for cv. Shindongjin for a rice growth model using the observation data in a low quality.” *Korean Journal of Agricultural and Forest Meteorology* 21(1): 42-54.
- Hyun, S., B. H. Seo, S. Lee and K. S. Kim. 2020. “Quantitative Assessment of the Quality of Regional Adaptation Trial Data for Crop Model Improvement.” *Korean Journal of Agricultural and Forest Meteorology* 22(3): 194-204.
- Kim, D.-J., J.-H. Park, S.-O. Kim, J.-H. Kim, Y. Kim and K.-M. Shim. 2020a. “A System Displaying Real-time Meteorological Data Obtained from the Automated Observation Network for Verifying the Early Warning System for Agrometeorological Hazard.” *Korean Journal of Agricultural and Forest Meteorology* 22(3): 117-127.
- Kim, J., C. K. Lee, H. Kim, B. W. Lee and K. S. Kim. 2015. “Requirement Analysis of a System to Predict Crop Yield under Climate Change.” *Korean Journal of Agricultural and Forest Meteorology* 17(1): 1-14.
- Kim, J., J. Lee, W. Sang, P. Shin, H. Cho and M. Seo. 2019. “Rice yield prediction in South Korea by using random forest.” *Korean Journal of Agricultural and Forest Meteorology* 21(2): 75-84.
- Kim, K. S., J. Kim and S. Hyun. 2020b. “Analysis of Crop Survey Protocols to Support Parameter Calibration and Verification for Crop Models of Major Vegetables.” *Korean Journal of Agricultural and Forest Meteorology* 22(2): 68-78.
- Kim, K. S., S.-O. Kim, J. H. Kim, K. H. Moon, J. H. Shin and J. Cho. 2018. “Development and application of crop models in Korea.” *Korean Journal of Agricultural and Forest Meteorology* 20(2): 145-148.
- Kim, M.-K., S. Kim, J. Kim, J. Heo, J.-S. Park, W.-T. Kwon and M.-S. Suh. 2016. “Statistical downscaling for daily precipitation in Korea using combined PRISM, RCM, and quantile mapping: Part 1, methodology and evaluation in historical simulation.” *Asia-Pacific Journal of Atmospheric Sciences* 52(2): 79-89.
- Krause, P., D. Boyle and F. Bäse. 2005. Comparison of different efficiency criteria for hydrological model assessment.
- Lee, C. K., J. Kim and K. S. Kim. 2015. “Development and application of a weather data service client for preparation of weather input files to a crop model.” *Computers and Electronics in Agriculture* 114: 237-246.
- Lee, J.-S. and M.-K. Oh. 2019. “Distribution Analysis of Land Surface Temperature about

- Seoul Using Landsat 8 Satellite Images and AWS Data.” *Journal of the Korea Academia-Industrial cooperation Society* 20(1): 434-439.
- Lee, S.-J., J. Song and Y.-J. Kim. 2016. “The ncam land-atmosphere modeling package (Lamp) version 1: implementation and evaluation.” *Korean Journal of Agricultural and Forest Meteorology* 18(4): 307-319.
- Moon, J., S. Jung, J. Rew, S. Rho and E. Hwang. 2020. “Combination of short-term load forecasting models based on a stacking ensemble approach.” *Energy and Buildings* 216: 109921.
- Na, S.-i., C.-w. Park, K.-h. So, J.-m. Park and K.-d. Lee. 2017. “Development of garlic & onion yield prediction model on major cultivation regions considering MODIS NDVI and meteorological elements.” *Korean Journal of Remote Sensing* 33(5_2): 647-659.
- Prudencio, E. E. and K. W. Schulz. 2011. The parallel C++ statistical library ‘QUESO’: Quantification of Uncertainty for Estimation, Simulation and Optimization, European Conference on Parallel Processing, Springer, 398-407.
- Shahhosseini, M., G. Hu, I. Huber and S. V. Archontoulis. 2021. “Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt.” *Scientific reports* 11(1): 1-15.
- USDA-NRCS. 2004. *Chapter 10: Estimation of direct runoff from storm rainfall*. NRCS National Engineering Handbook, Part 630: Hydrology.
- Woli, P., J. W. Jones, K. T. Ingram and C. W. Fraisse. 2012. “Agricultural reference index for drought (ARID).” *Agronomy Journal* 104(2): 287-300.
- Yun, J. I. 2014. “Agrometeorological early warning system: a service infrastructure for climate-smart agriculture.” *Korean Journal of Agricultural and Forest Meteorology* 16(4): 403-417.
- Zhao, C., B. Liu, L. Xiao, G. Hoogenboom, K. J. Boote, B. T. Kassie, W. Pavan, V. Shelia, K. S. Kim and I. M. Hernandez-Ochoa. 2019. “A SIMPLE crop model.” *European Journal of Agronomy* 104: 97-106.



제1회

농산물 수급 예측모형 경진대회

우수논문집

Chapter 1
지정과제



장려상

기후변화가 마늘과 양파 주산지의 생산량에 미치는 영향 분석

강혜정

기후변화가 마늘과 양파 주산지의 생산량에 미치는 영향 분석 공간적 의존성을 고려한 공간패널모형을 이용하여

강혜정*

Key words

마늘 단수(garlic yield), 양파 단수(onion yield), 기후통계(climate statistics), 공간의존성(spatial dependence), 공간패널모형(spatial panel model)

ABSTRACT

This study analyzed the effect of climatic factors on the yield of garlic and onion, taking into account spatial dependence between regions, using statistical data by the main production area of garlic and onion. The spatial panel model, which is a panel model considering spatial dependence, was used.

As a result of the analysis, it was found that the average minimum temperature and the amount of sunlight had a positive effect on the yield change of garlic and onion. Other things being equal, if the average minimum temperature rises and the high temperature phenomenon of increasing the amount of sunlight continues, the yield of garlic and onions is expected to increase.

In addition, the production of garlic and onion in the main production area has a statistically significant effect on the production in the adjacent area, so it can be seen that there is a spatial dependence on the formation of the main garlic and onion production area.

차례

- | | |
|----------|------------|
| 1. 서론 | 4. 분석 결과 |
| 2. 분석 자료 | 5. 요약 및 결론 |
| 3. 분석 방법 | |

* 전남대학교 농업경제학과 교수

1. 서론

마늘과 양파는 우리나라의 대표적인 양념채소류로서 한국인 식생활의 필수 농산물이다. 2020년 기준 마늘과 양파 재배면적은 각각 25,372ha와 14,673ha로, 조미채소류 면적의 각각 36%와 21%를 차지하고 있다.

특히 마늘과 양파는 특정 지역을 중심으로 주산지가 형성되어 있다. 2020년 마늘 생산량은 363,432톤인데, 경남, 경북, 전남, 충남의 생산량이 전체의 79.4%를 차지하고 있다. 경남이 26.8%로 가장 많고, 경북 22.8%, 전남 19.4%, 충남 10.3% 등의 순이다. 2020년 양파 생산량은 1,168,227톤이며, 전남이 전체 생산량의 38.8%로 가장 많고, 경남이 21.2%, 경북이 16.4% 등을 차지하고 있다.

주산지는 타 지역보다 특정 작물의 생산량이 대량으로 집중되어 있고, 어떤 통합된 체제에 의해 생산되어 타 지역보다 생산력이 높고, 시장의 수요에 대응할 능력이 있는 생산지역을 의미한다. 주산지는 생산지역의 자연적 요인뿐만 아니라 경제적 요인에 의해 형성되었다(박현태 외, 2002).

그러나 자연적·사회적 여건변화에 따라 주산지는 변화할 수 있다. 특히 최근 기후변화와 관련한 주산지 이동현상은 미래 농업생산의 중요한 논제이다. 지구 온난화로 주요 농산물의 주산지가 남부지방에서 충북, 강원지역 등으로 북상하고 있다. 사과, 복숭아 등 주요 과수 작물의 주산지 이동이 관찰되고 있다.¹⁾ 마늘과 양파의 경우도 향후 기후변화에 따라 한지형 마늘은 재배적지가 점차 사라지고, 난지형 마늘 재배적지는 북상하여 적지 면적은 증가할 것이며, 양파는 고온조건에서 생산량이 증가할 것으로 예측되었다(환경부, 2020).²⁾

조미채소류 마늘과 양파는 아직은 과거의 주산지가 유지되고 있으나, 기후변화의 영향이 점차 나타나고 있다. 2018년에는 5월의 잦은 비와 고온현상 등 이상기후로 양파와 마늘 생산량이 급감했던 반면, 2019년에는 최적의 기상조건으로 중만생종 양파·마늘 생산량이 전망치를 크게 웃돌아 공급과잉으로 양파와 마늘 시장가격이 폭락한 바 있다. 이처럼 마늘과 양파는

1) 통계청(2018), 「기후변화에 따른 주요 농작물 주산지 이동현황」

2) 환경부(2020), 「한국 기후변화 평가보고서 2020 - 기후변화 영향 및 적응」

기상여건 변화에 따라 생산량 및 가격 변동이 심하여 정부의 수급안정 대책이 지속적으로 시행되고 있다.

마늘과 양파의 효과적인 수급안정화 정책을 위해서는 정확한 생산량 예측이 선결되어야 한다. 생산량 예측모형을 개발하기 전에 우선 마늘과 양파의 단수에 영향을 미치는 요인, 특히 기상요인의 영향에 대해서 분석할 필요가 있다. 따라서 본 연구는 마늘과 양파의 주산지 시군의 연도별 단위면적(10a)당 생산량과 기상자료를 이용하여, 기상변수들이 마늘과 양파의 단수에 미치는 영향을 분석하고자 한다. 양파는 주로 노지에서 재배되는 특성상 기온이나 강수량, 일조량 등 기상정보에 의한 영향이 다른 품목보다 크기 때문에 본 연구의 결과는 향후 기후변화에 따른 생산 예측에 시사점을 줄 것이다.

마늘과 양파와 같이 주산지를 중심으로 재배지역이 밀집되어 있는 품목의 경우, 인접해 있는 지역의 생산량에 의해 영향을 받을 가능성이 높다. Tobler(1970)의 지리학 제1법칙은 모든 것은 그 밖의 다른 모든 것과 관련되어 있지만, 서로 가까이 있는 것들이 멀리 있는 것들 보다는 더 높은 관련성을 보인다는 것이다. 이러한 현상은 공간적 의존성 또는 자기상관(spatial autocorrelation, spatial dependence)이라고 한다. 즉, 공간상의 한 위치에서 발생하는 사건과 그 주변 지역에서 발생하는 사건과는 높은 상관관계를 보이고 있는데, 이는 공간상에 인접함으로써 나타나는 파급효과로 볼 수 있다.

공간데이터가 갖고 있는 특성으로 인해 공간적 의존성이 나타나는 경우 통상적인 회귀모델(OLS)을 적용하면 편의가 발생하고 오류를 가져오게 된다. 따라서 본 논문은 마늘과 양파 주산지 지역 간 공간효과를 통제하기 위해 공간계량경제모형을 이용한다.³⁾ 특히 본 연구는 주산지 시군별 시계열자료로 구축된 패널자료를 이용하여, 지역의 고유한 특성을 통제하면서 공간효과를 동시에 고려할 수 있는 공간패널모형을 적용한다는 점에서 기존 선행연구와 차별성이 있다.

본고의 구성은 다음과 같다. 제2장과 제3장에서 이용한 자료와 분석방법에 대해 설명하고, 제4장에서 분석결과를 제시하고, 제5장은 요약 및 결론을 내린다.

3) 공간데이터가 갖고 있는 공간적 종속성과 공간적 이질성 문제를 해결하기 위해 사용하고 있는 모형을 일컬어 공간계량경제모형(spatial econometrics models)이라고 한다. 공간계량모형은 어떤 특정한 현상의 공간분포와 그러한 공간분포 패턴에 영향을 미치는 요인들을 분석하는데 있어서 공간데이터가 갖는 공간효과를 통제함으로써 모형 추정 결과에 대한 신뢰도를 높이고자하는 기법이다(이희연·노승철, 2013).

2. 분석 자료

본 논문이 이용한 자료는 2006~2015년 기간의 마늘과 양파의 주산지 시군(마늘 15개 시군, 양파 17개 시군)의 단위(kg/10a)와 기후통계 자료이다.

마늘의 주산지 시군은 충남 태안, 충남 서천, 전남 해남, 전남 신안, 전남 무안, 전남 함평, 전남 고흥, 경북 의성, 경북 영천, 경북 군위, 경북 합천, 경남 창녕, 경남 남해, 제주 제주시, 제주 서귀포시이다. 양파의 주산지 시군은 전남 무안, 전남 신안, 전남 함평, 전남 해남, 전남 고흥, 경북 의성, 경북 군위, 경북 영천, 경북 청도, 경북 합천, 경남 함양, 경남 창녕, 제주 제주시이다. 주산지 시군의 10a당 생산량 자료는 통계청의 「농작물생산조사」에서 2006~2015년까지 조사·발표하였다. 2015년 이후에는 시도별 통계자료만 발표하고 있어, 최근 연도의 통계자료를 이용하는데 한계가 있다.

2006~2015년 기간 동안 마늘과 양파의 주산지 시군별 10a당 생산량 통계는 각각 <표 1>, <표 2>와 같다. 두 품목 모두 연도별 단수의 변동성이 크며 지역마다 차이가 있으나, 대체로 단위면적당 생산량이 증가 추세를 보인다. 마늘과 양파의 주산지 시군이 약간 달라서, 본 논문은 두 품목의 모형을 구분하여 추정한다.

표 1. 마늘 주산지 시군별 10a당 생산량

(단위: kg)

도	시군	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
충남	태안	1,217	1,275	1,471	1,648	1,421	1,355	1,316	1,791	1,697	1,537
충남	서천	832	986	1,084	1,405	1,120	1,048	1,308	1,431	1,464	1,330
전남	무안	1,029	1,198	1,190	1,355	989	969	1,092	1,117	1,179	1,234
전남	함평	985	1,126	986	1,200	1,141	873	1,038	1,125	1,174	1,074
전남	해남	1,073	1,234	1,223	1,340	1,127	1,138	1,024	1,210	1,403	1,110
전남	신안	1,181	1,195	1,236	1,306	1,102	1,148	1,014	1,155	1,252	1,240
전남	고흥	1,164	1,312	1,368	1,356	1,157	1,253	1,229	1,418	1,361	1,393
경북	합천	1,345	1,371	1,406	1,503	1,536	1,565	1,538	1,490	1,582	1,358
경북	의성	1,049	1,270	1,300	1,122	1,063	1,023	1,138	1,206	1,259	1,103
경북	영천	1,787	1,965	1,986	2,163	2,015	1,964	2,104	2,166	2,183	1,977
경북	군위	1,020	1,034	1,153	1,096	1,100	1,065	1,401	1,349	1,274	1,188

도	시군	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
경남	창녕	1,463	1,662	1,533	1,962	1,690	1,631	1,636	1,964	1,778	1,580
경남	남해	1,222	1,442	1,378	1,325	1,405	1,415	1,306	1,499	1,424	1,323
제주	서귀포	1,486	1,759	1,838	1,670	1,513	1,736	1,425	1,786	1,605	1,526
제주	제주	1,345	1,439	1,517	1,526	1,094	1,400	1,218	1,447	1,396	1,369

자료: 통계청, 「농작물생산조사」

표 2. 양파 주산지 시군별 10a당 생산량

(단위: kg)

도	시군	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006
전남	무안	6,017	6,051	5,713	5,500	6,977	6,787	7,949	7,090	7,646	5,759
전남	신안	5,978	6,152	5,785	5,570	6,669	7,047	8,284	6,633	6,965	5,727
전남	함평	5,562	6,975	6,283	5,567	5,926	5,788	6,817	5,936	6,521	5,978
전남	해남	4,909	5,690	5,148	4,182	5,796	6,074	7,299	6,815	6,224	5,551
전남	고흥	6,808	6,638	5,862	5,419	6,964	6,509	6,542	6,168	6,247	5,016
경북	의성	6,711	7,086	7,716	5,988	7,827	7,264	8,548	6,882	7,255	6,042
경북	군위	6,645	7,577	8,317	7,588	8,611	7,911	7,982	6,915	6,860	6,577
경북	영천	5,685	7,519	8,321	6,012	7,060	6,941	8,581	6,701	7,014	5,787
경북	청도	6,552	7,523	7,975	6,992	7,608	7,492	8,727	7,560	8,447	7,979
경북	합천	5,270	6,075	7,363	5,709	7,381	7,529	6,487	7,416	6,824	6,306
경남	함양	5,329	6,894	7,585	7,410	7,350	6,094	6,744	6,657	6,837	6,298
경남	창녕	5,921	7,647	7,448	6,250	7,484	6,038	6,395	6,539	6,060	6,059
제주	제주	6,412	7,213	6,580	6,338	7,032	5,078	5,790	6,147	6,231	5,918

자료: 통계청, 「농작물생산조사」

기상변수는 기상청 기상자료개방포털(<https://data.kma.go.kr>)에서 발표하는 2006~2015년 기간의 지역별 기온, 강수량, 일조일사 등의 통계자료를 이용하였다. 양파와 마늘의 수확 시기가 5월~6월인 점을 고려하여 1월~4월은 해당연도를, 7월~12월까지는 전년 기상정보를 적용하여, 각 변수의 지역별 연도별 평균값을 구하였다. 또한 기상관측 자료가 없는 지역의 경우에는 가장 인접한 지역의 기상자료로 대체하여 사용하였다. 마늘과 양파의 생육에 영향을 미치는 기상요인으로 누적강수량(mm), 평균강수량(mm), 평균기온(℃), 평균최고기온(℃), 평균최저기온(℃), 평균습도(%rh), 일조율(%)⁴⁾의 총 7가지 변수를 검토하였다.

4) 일조율은 태양이 구름이나 안개에 차단되지 않고 지표를 쬔 시간인 일조시간을, 일출에서 일몰까지의 시간인 가조시간으로 나눈 후, 그 값을 퍼센트 비율(%)로 나타낸 것이다. 즉, 일조시간/가조시간×100(%)으로 계산된다.

〈표 3〉은 기상변수들 간의 상관관계를 나타낸다. 평균기온과 다른 두 기온 관련 변수간의 상관율은 50% 정도로 높았으며, 누적강수량과의 상관율도 높았다. 평균강수량과 평균습도는 각각 평균최저기온과 평균최고기온과의 상관율이 50% 이상을 상회하는 것으로 나타났다. 따라서 모형 추정 시 설명변수 간 높은 상관관계로 인해 발생하는 다중공선성(Multicollinearity) 문제를 방지하기 위해서, 모형에 사용할 기상변수는 누적강수량, 평균최고기온, 평균최저기온, 일조율로 설정하였다. 특히 지구온난화 현상으로 인한 마늘과 양파 생육변화를 좀 더 살펴보기 위해서, 기온 관련 세분화된 변수로 평균최고기온, 평균최저기온을 포함시켰다.

표 3. 기상변수들 간의 상관관계

구분	누적 강수량	평균 강수량	평균기온	평균 최고기온	평균 최저기온	평균 습도	일조율
누적강수량	1						
평균강수량	0.9991 (0.0000)	1					
평균기온	0.4968 (0.0000)	0.4946 (0.0000)	1				
평균최고기온	0.1256 (0.1257)	0.1271 (0.1212)	0.4773 (0.0000)	1			
평균최저기온	0.4971 (0.0000)	0.494 (0.0000)	0.906 (0.0000)	0.0686 (0.4045)	1		
평균습도	-0.0779 (0.3435)	-0.0827 (0.3146)	-0.1257 (0.1254)	-0.6257 (0.0000)	0.1611 (0.0489)	1	
일조율	-0.1864 (0.0224)	-0.1843 (0.0239)	-0.2559 (0.0016)	0.1157 (0.1585)	-0.3404 (0.0000)	0.0027 (0.9738)	1

주: ()은 p-value

이와 같이 구축된 분석 자료는 마늘과 양파의 10년(2006~2015년) 동안의 주산지 시군단위 패널자료(panel data)이다. 모형에 포함될 주요 변수들에 대한 기초통계량을 살펴보면, 〈표 4〉에서 마늘의 평균 단수는 1,363kg/10a이고, 표준편차는 지역 간(between) 차이가 시점 간(within) 차이보다 2배 정도 큰 것으로 나타났다. 누적강수량은 1,335mm이고, 지역 간 차이와 지역의 시점 간 차이의 정도가 유사하다. 평균최고기온과 평균최저기온은 각각

18.8℃와 9.2℃이고, 지역 간 차이가 지역의 시점 간 차이보다는 큰 특징을 보인다. 평균 일조율은 47.3%으로 지역 간 차이와 시점 간 차이가 유사하다.

〈표 5〉에서 양파의 평균 단수는 6,665kg/10a이고, 표준편차는 지역 간(between) 차이가 시점 간(within) 차이보다 2배 정도 큰 것으로 나타났다. 누적강수량은 1,239mm이고, 평균 최고기온과 평균최저기온은 각각 19.1℃와 8.7℃이고, 지역 간 차이가 지역의 시점 간 차이보다는 크게 나타났다. 평균 일조율은 46.9%이다.

표 4. 마늘 모형: 주요 변수의 기초 통계량

변수		평균	표준편차	최소	최대
마늘 단수(kg/10a)	overall	1,363.2	283.5	832.0	2,183.0
	between		262.1	1,072.2	2,031.0
	within		125.7	994.4	1,681.4
누적강수량(mm)	overall	1,335.2	415.5	596.5	2,700.8
	between		303.3	978.9	2,087.5
	within		293.6	334.4	2,185.2
평균최고기온(℃)	overall	18.8	1.2	15.7	21.2
	between		1.2	16.4	20.5
	within		0.4	17.9	19.8
평균최저기온(℃)	overall	9.2	2.3	4.8	14.3
	between		2.3	5.4	14.1
	within		0.4	8.4	11.5
일조율(%)	overall	47.3	4.0	37.0	57.2
	between		2.8	38.7	50.5
	within		3.0	41.1	55.9

표 5. 양파 모형: 주요 변수의 기초 통계량

변수		평균	표준편차	최소	최대
양파 단수(kg/10a)	overall	6,664.8	886.4	4,182.0	8,727.0
	between		542.9	5,768.8	7,685.5
	within		715.2	5,078.0	8,467.8
누적강수량(mm)	overall	1,239.3	300.0	596.5	2,248.3
	between		170.1	978.9	1,574.3
	within		251.2	524.1	1,913.3

변수		평균	표준편차	최소	최대
평균최고기온(°C)	overall	19.1	1.1	15.7	21.2
	between		1.1	16.4	20.4
	within		0.4	18.1	19.9
평균최저기온(°C)	overall	8.7	2.0	4.8	14.1
	between		2.0	5.4	13.4
	within		0.5	7.9	11.0
일조율	overall	46.9	3.9	37.0	57.2
	between		2.8	38.7	48.6
	within		2.8	40.7	55.6

패널데이터 분석을 실시하기 전에 패널데이터 자체의 안정성(stationarity) 여부를 확인하기 위하여 패널 단위근 검정(panel unit root test)검정을 실시하였다. 가장 많이 사용되는 패널 단위근 검정방법으로 Im-Pesaran-Shin의 W-test(이하 IPS 검정), ADF-Fisher의 Chi-square test(이하 ADF-Fisher 검정), Levin, Ln and Chu의 t-ratio test(이하 LLC 검정) 등이 있다. 이 세 가지 검정방법을 각 모형의 주요 변수들을 대상으로 적용한 결과는 <표 6>, <표 7>과 같다.

모형 분석에서 모든 변수들은 자연로그를 취하므로, 자연로그를 취한 값으로 패널단위근 검정을 하였다. 전술한 세 가지 방법을 적용하여 패널단위근 검정결과, 최소한 한 방법 이상에서 검정의 귀무가설이 통계적으로 유의하게 기각되어, 모든 변수들은 추세가 없는 안정적인 시계열임을 알 수 있다.

표 6. 마늘 모형: 패널 단위근 검정결과

구분	마늘 단수	누적 강수량	평균 최고기온	평균 최저기온	평균 일조율
IPS 검정	-1.626***	-2.109***	-0.540	-6.193***	0.125
LLC 검정	-4.427***	-4.329***	-0.617	-19.863***	-2.054
ADF-Fisher 검정	102.884***	114.790***	30.001*	40.756***	80.684***

주: *** p<0.01, ** p<0.05, * p<0.1

표 7. 양파 모형: 패널 단위근 검정결과

구분	양파 단수	누적 강수량	평균 최고기온	평균 최저기온	평균 일조율
IPS 검정	-1.508***	-3.049***	-1.310***	-5.187***	-0.700
LLC 검정	-4.116***	-5.250***	-2.923***	-13.990***	-2.386**
ADF-Fisher 검정	87.999***	102.034***	27.588	32.053	79.272***

주: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

3. 분석 방법

본 논문은 공간패널모형(Spatial Panel Model)을 이용하여, 지역 간 공간적 의존성을 고려하면서 기후요인이 마늘과 양파의 단수에 미치는 영향을 분석한다.

공간패널모형은 통상적인 단순 패널모형이 간과한 공간적 종속성을 고려할 수 있으며, 패널통계에 내재한 시공간적 종속성을 동시에 고려하여 통계적 추정의 오차를 줄일 수 있다. 특히, 이러한 공간적 종속성의 영향력은 지역단위 연구에서 반드시 고려해야 할 요소이다(Belotti et al., 2016).

본 연구에서는 정적 공간패널모형인 공간자기회귀모형(Spatial Auto-Regressive, SAR), 공간더빈모형(Spatial Durbin Model, SDM), 공간오차모형(Spatial Error Model, SEM)을 적용하고, 각 모형의 추정결과를 비교한다(김봉태 외, 2015).

가) 공간자기회귀모형(Spatial Auto-Regressive, SAR)

SAR 모형은 공간적 종속성이 종속변수에 내재되어 있다고 전제하고, 이를 공간자기회귀계수(ρ)를 통하여 모형에 반영한다. 즉, 특정시점(t) 및 특정지역(i)의 종속변수의 관측값(Y_{it})은 동일시점의 인접지역의 관측값(Y_{jt})에 의해 영향을 받으며, 이러한 영향력은 공간가중치행렬(spatial weight matrix)과 결합하여 일종의 독립변수로서 모형에 반영된다. SAR 모형은 식 (1)과 같다.

$$(1) Y_{it} = \rho \sum_{j=1}^N w_{ij} Y_{jt} + X_{it} \beta + \mu_i + \epsilon_{it}$$

여기서, ρ 는 공간자기회귀계수로서 특정지역의 관측값(Y_{it})이 인접지역의 관측값(Y_{jt})들의 공간가중된 평균값으로 설명되는 정도를 나타내며, μ_i 는 특정지역(i)의 관측되지 않은 변수들에 의한 오차이고, ϵ_{it} 는 일반적 오차이다. w_{ij} 는 공간가중치행렬로 공간적으로 더 인접한 지역에 더 높은 가중치를 부여하도록 설정한다. 공간가중치행렬을 구성하는 방식은 다양한데, 가장 많이 사용하는 방법으로 인접(contiguity)에 가중치를 부여하는 방식⁵⁾과 공간좌표(coordinates)를 이용하는 방법⁶⁾이 있다. 본 논문에서는 공간적 인접에 가중치를 부여하는 방식을 적용하여, 지역들이 인접해 있으면 '1', 인접하지 않으면 '0'의 가중치를 부여한 공간가중치 행렬을 구축하고, $\sum_j w_{ij} = 1$ 이 되도록 횡단표준화하였다.

나) 공간더빈모형(Spatial Durbin Model, SDM)

SDM 모형은 공간적 종속성이 종속변수뿐만 아니라 설명변수에도 내재되어 있다고 전제하고, 이를 모형에 반영한 것이다. SDM 모형은 식 (2)와 같다.

$$(2) Y_{it} = \rho \sum_{j=1}^N w_{ij} Y_{jt} + X_{it} \beta + \theta \sum_{j=1}^N w_{ij} Z_{jt} + \mu_i + \epsilon_{it}$$

여기서, ρ 는 공간자기회귀계수이며, θ 는 특정 설명변수의 공간가중된 평균값이 종속변수(Y_{it})를 설명하는 정도를 나타낸다.

5) 인접에 가중치를 부여하는 방식은 지역이 서로 인접하는 경우에는 '1', 그렇지 않으면 '0'으로 가중치를 부여한다. 일반적으로 가중치는 행을 기준으로 합이 1이 되도록 다시 표준화한다.

6) 공간좌표를 이용하는 방법은 지역의 중심을 기준으로 다른 지역의 중심까지 좌표거리를 계산하고, 이를 횡단표준화함으로써 행의 합이 1이 되도록 하는 방법이다.

다) 공간오차모형(Spatial Error Model, SEM)

SEM 모형은 공간적 종속성이 오차항에 내재되어 있다고 전제하고, 이를 공간자기회귀계수(λ)을 통하여 모형에 반영한다. SEM 모형은 식 (3)과 같다.

$$(3) Y_{it} = X_{it}\beta + \mu_i + \phi_{it}, \phi_{it} = \lambda \sum_{j=1}^N w_{ij}\phi_{jt} + \epsilon_{it}$$

여기서, ϕ_{it} 는 공간적 종속성을 갖는 오차항을 의미하며, λ 는 오차의 공간자기회귀계수이다.

위의 세 가지 모형을 이용하여 기후요인이 마늘과 양파의 단수에 미치는 영향을 추정하였다. 시계열 자료 측면에서 시간에 따른 생산기술 변화를 고려하기 위해서, 기후변수들 이외에 시간 또는 추세변수를 모형에 포함시켜서 추정하였다.

모형을 추정한 후 각 변수의 한계효과를 도출하였다. 종속변수 또는 설명변수에 공간적 종속성이 있는 경우 특정 변수의 한계적인 변화는 해당 지역에게 영향을 미칠 뿐만 아니라 다른 지역에도 영향을 미칠 수 있다. 이러한 상황을 파악하기 위해서 본 논문은 LeSage and Pace (2009)가 제안한 방법을 따라서 한계효과를 직접효과와 간접효과로 구분하여 도출한다. 직접효과는 특정 설명변수의 변화가 동일 지역의 종속변수에 미치는 효과이고, 간접효과는 해당 설명변수의 변화가 인접지역의 종속변수에 미치는 효과를 의미한다.

4. 분석 결과

공간패널모형을 이용하여 기후요인이 마늘과 양파 단수에 미치는 영향을 분석한 결과는 각각 <표 8>, <표 9>와 같다. 모형 간 비교를 위해서 앞에서 설명한 세 가지 공간패널모형들의 추정결과뿐만 아니라 패널고정효과모형(Fixed Effect Model, FE)의 결과도 함께 제시하였다. 모든 공간패널모형에서 공간자기회귀계수 ρ 와 λ 가 통계적으로 유의하게 나타나, 마늘과 양파의 생산량에 공간적 의존성이 영향을 미치고 있음을 알 수 있다. 시간변수는 모든 연도에서 유의하게 나타나, 시간에 따른 기술변화가 마늘과 양파 단수 증가에 기여하였음을 알 수 있다.

모든 공간패널모형에서 평균최저기온과 일조율은 통계적으로 유의한 양(+)의 값을 나타내었다. 평균최고기온의 추정계수값도 양(+)의 값이나, 통계적으로 유의하지는 않았다. 즉, 평균최저기온과 평균최고기온이 상승할 경우, 마늘과 양파의 단수는 모두 증가하였음을 알 수 있다. 일조율도 단위 증가에 긍정적으로 기여하였다. 따라서 다른 조건이 동일한 상황에서 평균최저기온이 상승하고 일조율이 늘어나는 고온 현상이 지속되면 마늘과 양파의 단위 증가로 인해 발생할 수 있는 관련 농가들의 피해를 예방하기 위해서 기후온난화에 대응한 마늘과 양파의 수급안정화 전략이 강구될 필요가 있다.

표 8. 마늘의 공간패널모형 분석결과

변수	FE		SAR		SDM		SEM	
	추정 계수	표준 오차	추정 계수	표준 오차	추정 계수	표준 오차	추정 계수	표준 오차
누적강수량	-0.018	0.041	-0.015	0.037	-0.036	0.039	-0.017	
평균최고기온	0.760	0.493	0.392	0.464	0.351	0.492	0.184	
평균최저기온	0.028	0.207	0.136*	0.191	0.258*	0.182	0.252*	
일조율	0.351*	0.149	0.346*	0.134	0.419*	0.149	0.417*	
W*누적강수량					-0.003	0.637		
W*평균최고기온					2.869	0.255		
W*평균최저기온					1.088	0.179		
W*일조율					0.374	0.003		
상수항	3.649	1.391						
ρ			0.198***	0.073	0.152***	0.074		
λ							0.216***	0.082
R2(within)	0.205		0.205		0.344		0.193	
AIC	-301.785		-306.567		-323.786		-306.154	

주: *** p<0.01, ** p<0.05, * p<0.1

표 9. 양파의 공간패널모형 분석결과

변수	FE		SAR		SDM		SEM	
	추정 계수	표준 오차	추정 계수	표준 오차	추정 계수	표준 오차	추정 계수	표준 오차
누적강수량	-0.055	0.045	-0.055	0.041	-0.037	0.065	-0.055	0.059
평균최고기온	0.093	0.477	0.097	0.515	0.188	0.646	0.099	0.638
평균최저기온	0.158	0.182	0.157*	0.172	0.209*	0.174	0.155	0.178
일조율	0.023	0.145	0.022*	0.148	0.147*	0.207	0.026*	0.198
W*누적강수량					0.038	0.119		
W*평균최고기온					0.901	1.120		
W*평균최저기온					0.156	0.321		
W*일조율					0.401	0.366		
상수항	9.064	1.924						
ρ			0.357***	0.078	0.351***	0.080		
λ							0.365***	0.080
R2(within)	0.219		0.219		0.271		0.4193	
AIC	-251.210		-249.220		-235.613		-249.241	

주: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

공간패널모형의 적합성 검증결과⁷⁾, SDM이 다른 모형에 비해 통계적으로 유의하게 나타났다. 따라서 마늘과 양파 모두 SDM 모형 추정결과에 대한 한계효과를 도출하였다. 그 결과는 <표 10>과 같다.

마늘의 경우 일조율이 1% 상승하면 마늘 단수는 약 1.04% 증가하는데, 이 중 해당 지역의 직접 영향은 0.859%, 인접 지역의 간접효과는 0.182%로 구분할 수 있다. 모든 변수들의 한계효과에서 직접효과의 크기가 간접효과보다 크게 나타났다. 그러나 인접지역의 간접효과도 일정부분 유의하게 기여하고 있었다. 즉, 마늘과 양파의 주산지에서 생산량은 인접지역의 생산량에도 영향을 미치는 공간적 의존성이 나타나고 있었다.

7) SDM을 기준으로 귀무가설 $\theta=0$ 과 $\theta + \rho\beta = 0$ 에 대해서 각각 검정한 결과, 귀무가설이 5% 유의수준에서 모두 기각되었다.

표 10. SDM모형의 한계효과

구분	변수	마늘 SDM모형	양파 SDM모형
직접효과	누적강수량	-0.014	-0.039
	평균최고기온	0.391	0.462
	평균최저기온	0.159	0.191
	일조율	0.859	0.735
간접효과	누적강수량	-0.003	-0.01
	평균최고기온	0.177	0.115
	평균최저기온	0.04	0.054
	일조율	0.182	0.147
총효과	누적강수량	-0.017	-0.049
	평균최고기온	0.568	0.577
	평균최저기온	0.199	0.245
	일조율	1.041	0.882

5. 요약 및 결론

본 연구는 마늘과 양파의 주산지 시군별 통계자료를 이용하여, 지역 간 공간적 의존성을 고려하면서 기후요인이 생산량에 미치는 영향을 분석하였다. 공간적 종속성을 고려한 패널모형인 공간패널모형을 이용하였다. 또한 공간패널모형 간 적합성 검정을 통해 선택된 SDM 모형을 이용하여, 추정계수의 한계효과를 직접효과와 간접효과로 구분하여 도출하였다.

분석 결과, 평균최저기온과 일조율은 마늘과 양파의 단수 변화에 긍정적인 영향을 미치는 것으로 나타났다. 다른 조건이 동일한 상황에서 평균최저기온이 상승하고 일조율이 늘어나는 고온 현상이 지속되면 마늘과 양파의 단수는 증가할 것으로 보인다. 기후온난화로 인한 마늘과 양파의 공급과잉 문제가 발생할 수 있으므로, 이에 대한 적절한 수급안정화 방안이 필요할 것이다.

또한 마늘과 양파의 주산지 지역의 생산량은 인접지역의 생산량에 통계적으로 유의하게 영향을 미치고 있어서, 마늘과 양파 주산지 형성에 공간적 의존성이 존재함을 알 수 있었다. 따

라서 마늘과 양파 생산 예측 모형을 구축할 때 주산지 인접지역의 영향도 함께 고려할 수 있는 공간계량경제모형이 적절할 것이다.

이 연구에서는 과거 기후통계자료의 시공간적 추세를 통해서 마늘과 양파의 생산량에 미치는 기후요인의 영향을 분석하였으나, 보다 객관적인 마늘과 양파 생육에 대한 기후요인의 영향에 관해서는 자연과학적 연구가 더 필요할 것이다.

본 연구는 기후변화가 진전됨에 따라 농작물 생산에 유의한 영향을 미치는 기상요인의 영향을 살펴보았다는 점과 마늘과 양파의 주산지 지역 간 공간적 의존성이 생산에 영향을 미치고 있음을 밝혀냈다는 점에서 관련 선행연구들과 차별성이 있다.

그러나 장기 시계열 통계의 부족으로 분석 결과의 통계적 유의성을 충분히 확보하지 못한 한계가 있다. 또한 마늘과 양파 단수에 영향을 미치는 요인으로 기상변수 이외에도 토양 정보, 재배정보 등 다양한 변수들을 모형에 포함시키지 못한 한계가 있다. 향후 농산물의 지역 단위 패널자료가 잘 구축된다면 기후요인과 공간의존성을 동시에 고려하면서 생산량을 보다 정확하게 예측할 수 있는 품목 예측모형 개발과 적용이 가능할 것으로 보인다.

참고문헌

- 김봉태·엄기혁·이준수·박혜진·육근형. 2015. “공간패널모형을 이용한 연가어업 생산량과 기후변화 요소의 관계에 대한 연구.” 『한국수산경영학회』 46(3): 63-72.
- 박현태·김연중·한석호. 2002. 「주요 과채의 주산지 구조와 지역간 경쟁력 분석」. 한국농촌경제연구원.
- 이희원·노승철. 2013. 「고급통계분석론-이론과 실습」. 문우사.
- 통계청(2018), 「기후변화에 따른 주요 농작물 주산지 이동현황」.
- 환경부(2020), 「한국 기후변화 평가보고서 2020 - 기후변화 영향 및 적응」.
- Belotti, F., Hughes, G., and Mortari, A. P. (2016). Spatial panel data models using Stata. The Stata Journal 17(1): 139-180.
- LeSage, J. and Pace, R. K. (2009). Introduction to Spatial Econometrics. New York, U.S.: Chapman and Hall.
- Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. Economic Geography 46(1): 234-240.



제1회

농산물 수급 예측모형 경진대회

우수논문집

Chapter 1
지정과제



장려상

딥러닝 기반 감성분석을 이용한 양파수급모형

조수민 · 오정은 · 백종현 · 순병민

딥러닝 기반 감성분석을 이용한 양파수급모형

조수민* 오정은** 백종현*** 순병민****

Key words

감성분석(sentiment analysis), 딥러닝(deep learning), 양파수급모형(onion model)

ABSTRACT

Our study examines how news related to onion affect producer behavior for the harvested area. To do that, we create sentiment index related to Korean onion using LSTM-CNN (Long Sequence Term Memory-Convolution Neural Network). We collect news in June, July, and August before planing onion to make sentiment index. Using these sentiment index during that periods, these sentiment index helps producers plant onion or other agricultural products. We then impose these sentiment index into the harvested area function for onion. Finally, we build the onion supply-demand model to examine the impact of change in sentiment index on supply and demand. To the best of our knowledge, our approach using sentiment index in the agricultural model is a first trial. Therefore, our study can introduce an approach to improve the accuracy of modeling for agriculture.

차례

- | | |
|------------|-------------|
| 1. 서론 | 4. 자료 |
| 2. 선행연구 분석 | 5. 분석 결과 |
| 3. 연구 방법 | 6. 결론 및 시사점 |

* 충남대학교 농업경제학과 학부생, e-mail: sumin_official@naver.com

** 충남대학교 농업경제학과 학부생

*** 충남대학교 컴퓨터공학과 학부생

**** 충남대학교 농업경제학과 조교수, e-mail: soonbm@cnu.ac.kr

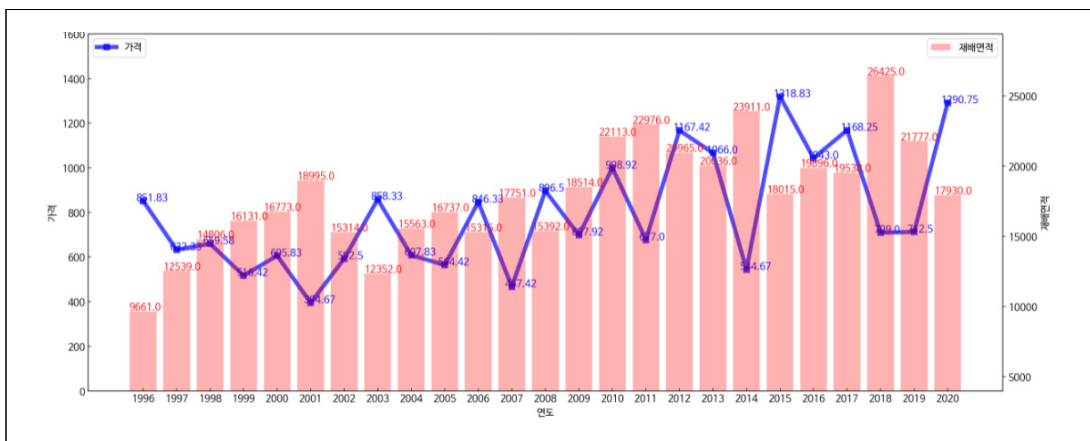
1. 서론

1.1. 연구배경 및 필요성

양파는 배추, 무 등과 함께 국내 5대 민감 채소로 지정된 수급 및 가격 불안정성이 큰 품목이다. 연중 생산이 불가능한 특성상 저장을 통해 수급량을 조절하며, 생산량 변동이 심한 해에는 부분적으로 수출과 수입을 통해 균형을 유지하고 있다.

〈그림 1〉은 1996년부터 2020년까지의 양파 가격과 재배면적 변동추이를 나타낸 그래프이다. 직전 해의 양파 가격이 상승하면 당해 양파 재배면적이 증가하고, 재배면적 증가로 양파 생산량이 증가하면 과잉공급으로 다음 기의 가격이 폭락하기를 반복하며 변동성이 커 전망이 어려운 양상을 보인다. 2018년에는 생산량 예측 실패로 과잉 생산된 양파를 산지 폐기한 규모는 95억 9,900만 원(3만 6,386톤)에 달했다. 농림축산식품부에서는 매년 한국농촌경제연구원과 통계청이 생산량 및 재배면적 전망치를 통해 수급 조절에 나서고 있는데, 농촌경제연구원에서는 양파 재배면적이 18.3% 증가할 것으로 전망한 반면 통계청은 35.2% 증가할 것으로 전망해 농림부의 수급 조절 대책이 차질을 빚은 것이다.

그림 1. 양파 재배면적, 가격 변동 추이



자료: 농산물유통정보(KAMIS), 농업관측통계시스템(OASIS)

양파 생산량 파악을 위해서는 재배면적과 단수 정보가 중요하나, 대체로 단수보다는 재배면적이 생산량에 더 큰 영향을 주는 것으로 알려져 있다(이승인, 2020). 따라서 재배면적 의사결정은 가격 안정과 생산량 전망의 중요한 요인이며, 이를 반영하기 위해 농업관측본부에서 매달 관측 월보를 통해 재배의향면적 조사를 발간하고 있다. 하지만 월 단위 조사로 실시간 정보 수집이 어려우며 비용 문제로 인해 표본농가와 지역 모니터 조사가 이루어진다는 한계로 재배면적 의사결정의 불확실성 문제를 완전히 해결하지 못하고 있다. 본 연구에서는 양파 수급모형에 농가의 재배면적 의사결정을 보다 정밀하게 반영하기 위해 감성분석을 활용한 감성지수를 재배면적 함수의 설명변수로 도입하고, 양파 가격 및 생산량 전망을 정교화하고자 한다.

1.2. 연구 목적

본 연구는 텍스트 마이닝 기법의 하나인 감성분석을 활용해 양파 관련 감성지수를 생성하고, 양파 재배면적을 함수의 설명변수로 도입해 양파 관련 감성 변화가 양파 수급에 미치는 영향을 분석한다. 양파 파종기(8월 말)에 근접할수록 감성지수가 양파 재배면적 결정에 미치는 영향이 크다는 가정 하에 6월, 7월, 8월, 6·7월, 7·8월, 6·7·8월로 기간을 나누어 재배면적 함수에 감성지수를 포함하여 추정한다. 포함된 감성지수에 따른 각 모형의 추정값과 실제값을 비교하기 위해 평균 제곱근 오차(RMSE)를 계산, 가장 RMSE가 작은 함수를 활용하여 양파 수급모형을 구축한 뒤, 감성지수 등락에 따른 시나리오 분석을 실시하여 감성지수 변화가 양파 재배면적에 미치는 영향을 분석한다. 본 연구는 양파 수급 모형을 정교화함으로써 농가와 정책 담당자의 합리적인 의사결정에 도움을 주고자 하며, 더 나아가 감성분석과 같은 빅데이터 분석의 농업 분야에서의 활용 가능성을 보이고자 한다.

2. 선행연구 분석

양파 재배면적과 관련된 국내 연구로 류재현(2020)은 OLS 분석을 통해 농업관측본부가 제공하는 재배면적 전망치가 농민의 재배면적 조절에 유의한 영향을 미친다는 사실을 보였지만, 농가 특성을 설명변수에 사용하지 않은 자료 제약의 문제로 양파 도매가격과 금년도 후기 마늘 재배면적이 비례하는 결과가 도출되었다는 한계가 있다. 하지희(2019)는 양파와 마늘 가격 예측을 위해 이분산성과 월별 더미, 물량자료를 반영하여 ARMGARCH(36개월)이 가장 적합한 시계열 모형임을 추정하였다. 연구결과, 마늘 수입량은 마늘 가격에 유의한 영향을 미쳤지만, 양파의 경우 양파 가격에 영향을 미치는 적합한 외생변수를 모색하지 못한 한계가 있다.

최근 스마트폰의 보급과 SNS의 활성화로 수많은 정보들이 생성되었고, 비정형 데이터 처리 및 활용의 중요성이 커지고 있는 흐름 속에서 금융, 부동산 등 다양한 분야에서 빅데이터를 활용하여 기존과는 다른 새로운 분석 방법과 연구결과를 제시하는 시도가 많이 이루어지고 있다. 차은정(2016)은 뉴스 정보를 수집, 감성사전(SentiWordNet)에 대입하여 뉴스가 가진 감성값을 확인하고 감성정보가 실제로 주식 시장에 미치는 영향을 SVM과 로짓분석을 이용해 추정하였고, 그 결과 S&P500의 예측 정확도가 4% 향상된다는 결론을 도출했다. 하지만 하나의 언론사 뉴스만을 이용해 다소 국부적이며, 여러 다양한 외부적 요소들을 포함하지 않아 복잡한 주식 시장을 반영하기 힘들다는 한계가 있다. 박재수(2019)는 주요 온라인 신문기사에서 추출한 아파트 매매와 관련된 긍정 감정지수와 아파트 매매가격과의 연관 관계를 알아보기 위해 교차상관 및 인과관계 분석을 시행하였고 서울시와 모든 권역에서 긍정 감성지수는 3개월 이내의 선행성을 보이며 소형 아파트 매매가격과 유의미한 교차 상관관계를 보임을 확인하였다. 하지만 2008년 이후의 자료들만을 대상으로 해 보합세와 상승기만 포함하였고, 심리지표인 감성지수만을 포함했기 때문에 경제변수를 포함시키지 못했다는 한계점이 있다.

또한 농업 분야에서도 빅데이터 활용의 중요성이 증가함에 따라 비정형 데이터와 기존에 사용되어온 경제변수 간의 관계를 분석하는 연구가 이루어지고 있다. 강산하(2017)는 웹 크

롤링 방식을 통해 텍스트 마이닝을 시행해 부정기사 수와 '양파 가격' 검색 빈도는 재배면적과 반비례하며 긍/부정 단어에 의해 양파 가격이 영향을 받음을 시사하였다. 하지만, 심도있는 모형에 적용한 것이 아닌 단순 상관 관계분석을 시행하였다는 한계점이 있다. 나형철(2018)은 방송 뉴스 및 소비자 패널 데이터 등과 같은 비정형 자료들을 변수로 도입하여 양파 소비와 일정 시차를 두고 양의 상관관계가 있음을 확인하였다. 하지만 수요 예측 모형과 예측력을 제시하지 않아 양파 관련 정보가 양파 소비에 미친 영향을 심도있게 분석하지 못하였으며, 양파 가격이 폭락하였던 2014년의 자료만을 사용했기 때문에 결과 검증을 위해 추가적인 연구가 필요하다는 한계가 있다.

양파 생산량 및 가격 전망의 정확도 향상을 위해서는 여러 경제적 요인과 더불어 양파 재배면적 의사결정에 영향을 미칠 수 있는 새로운 요인들을 반영한 확장된 연구가 필요하나 이와 관련한 연구는 아직 부족한 실정이다. 본 연구에서는 타 분야에서 비교적 활발하게 활용되고 있는 비정형 데이터인 감성지수를 농업 분야에 도입해 재배면적 함수를 도출하고, 수급 균형모형을 통해 생산량 및 가격 전망과 시나리오 분석을 시행하였다는 차별점이 있다.

3. 연구 방법

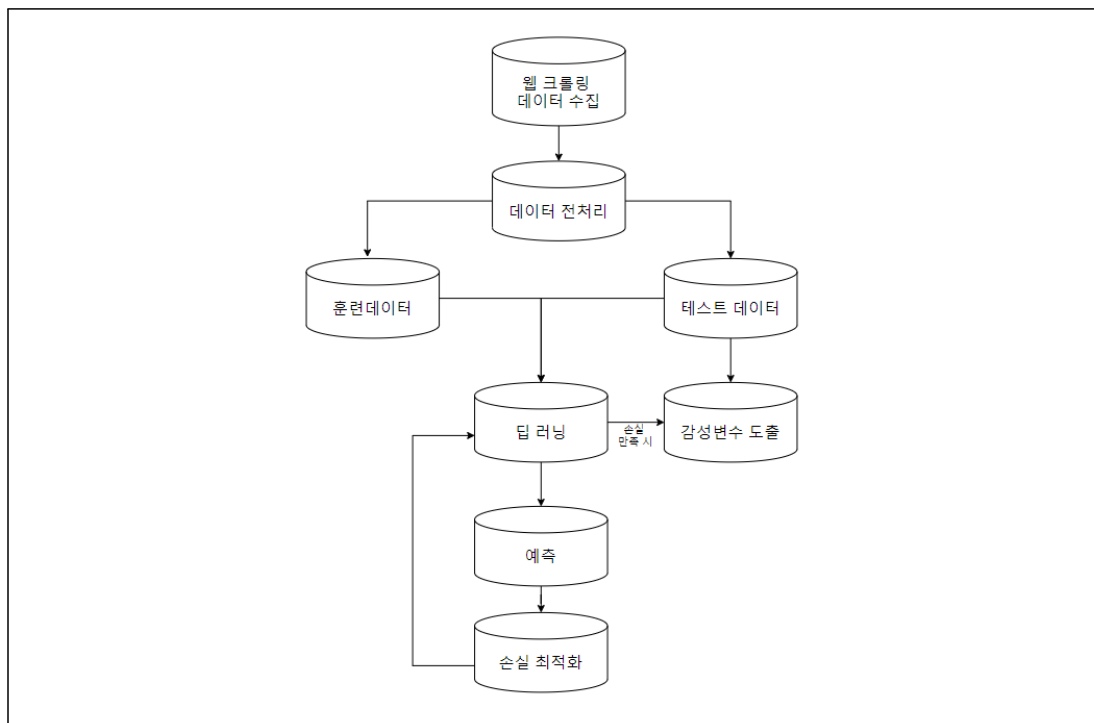
3.1. 감성분석

기존 선행연구에서는 군집 알고리즘을 이용한 감성사전 구축 방식을 활용하고, 웹 크롤링 기술로 문서 데이터를 수집하였다. 수집한 데이터를 바탕으로 TF-IDF(Term Frequency-Inverse Document Frequency)라는 문서별 단어 빈도수를 통하여 가격 문서, 재배면적 문서 등을 파악하여 빈도수를 계산한 뒤 각 빈도수 마다 유사도를 측정했다. 유사도 값들을 바탕으로 K-means, DBSCAN의 클러스터링 알고리즘을 적용했다. 적용한 알고리즘을 바탕으로 군집을 형성하여 긍정, 부정, 무정의 비율 값을 측정하여 감성사전을 구축하는 비지도 학습 방식이다.

본 연구에서는 기존 연구와의 차별성을 두기 위해 입력층, 은닉층, 결과층으로 나누어 각 층을 거쳐서 학습하는 지도학습 방식인 신경망 기반으로 감성지수를 도출한다. 입력층은 데이터를 넣는 역할을 하며, 은닉층은 가중치 벡터와 편향 벡터를 통해 층을 통과하면서 가중치와 편향을 학습하는 역할, 결과층은 학습된 가중치와 편향 벡터를 데이터에 적용하여 결과를 출력하는 역할을 한다. 신경망을 통해 나온 결과는 0~1 사이의 실수 형태로 적용하여 각 기사마다 감성지수를 도출한다.

감성분석은 <그림 2>와 같은 방식으로 감성의 수치화된 값을 얻는다. 인터넷 기사 데이터를 수집하는 웹 크롤링 방식으로 감성지수 데이터를 수집하고, 데이터 수집 기간은 연구 범위에 맞게 1996년부터 2020년까지 양파와 관련된 기사들로 선정하기로 한다. 수집한 기사들을 바탕으로 제목과 내용을 따로 저장하여 각각 데이터들의 신경망이 인식하지 못하는 문자나 다른 특수기호를 삭제하는 전처리 과정을 거친다.

그림 2. 감성지수 도출 과정

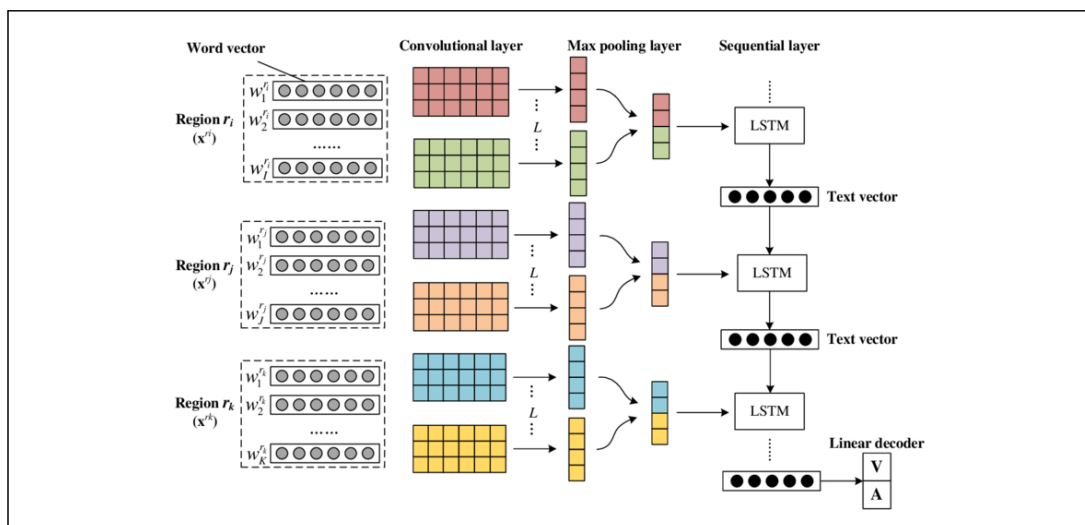


자료: 저자 작성

전처리 과정을 거친 제목과 기사는 신경망이 학습할 수 있게 숫자 형태의 데이터로 변환되며, 이를 위해 단어사전을 구축한다. 단어사전은 데이터에 포함된 단어를 기반으로 구축되며, 해당하는 단어의 키워드를 통해 문자열 형태의 데이터를 실수화한 숫자로 변환한다. 이때 신경망 훈련을 위해 전체 데이터를 훈련 데이터와 테스트 데이터로 분류하게 되는데 훈련 데이터는 모델의 가중치와 편향을 학습하는 역할을 하고, 테스트 데이터는 모델이 훈련 데이터의 과적합 현상을 막는 역할을 한다.

모델은 자연어 처리 딥러닝에서 많이 사용하는 LSTM-CNN(Long Sequence Term Memory-Convolution Neural Network) 모델을 이용한다. LSTM은 장기 기억 메모리 셀 계층으로, 모델의 학습시간이 길수록 처음에 학습했던 내용을 기억하지 못하는 경우가 잦지만, LSTM은 처음에 학습했던 내용까지 셀에 저장하여 가중치가 모든 학습의 경우를 반영하여 학습한다는 장점이 있다. CNN은 합성곱 신경망을 뜻하며, <그림 3>과 같이 각 계층마다 필터를 적용하여 범위를 축약한 값을 나타내는 형태이다. 신경망 모델을 구축하였으면 학습률과 훈련 횟수 등 세밀한 매개변수를 통해서 학습한 결과를 예측하며 예측한 값과 실제 값의 오차를 최적화하여 모델을 반복적으로 학습한다. 모델 학습 중 오차 범위 안에 들어오면 학습을 중지하고 모델을 통하여 감성지수 값을 도출한다.

그림 3. LSTM-CNN 모델



자료: Jin 외(2016)

3.2. 감성지수를 도입한 재배면적 함수의 추정

농가의 양파 재배면적 결정에 양파의 직전 해 실제 가격(P_{Ot-1}), 대체 작목인 마늘의 직전 해 실제 가격(P_{Gt-1}), 당해의 양파 농가 경영비(C_t), 당해의 데이터 감성분석으로 도출된 양파 관련 감성지수(SI_t)와 기술 발전을 대표하는 대리변수(proxy variable)인 추세변수($trend$)가 영향을 주는 것으로 가정했다. 이때 양파 기대가격, 마늘 기대가격은 직전 해의 실제 가격이 당해의 기대가격이라는 단순기대($P_t^* = P_{t-1}$)를 가정한다.

농가는 6월에는 주로 마늘 수확작업과 모내기 등에 노동력을 집중시키며, 여름작물 파종이 어느 정도 마무리된 다음인 8월 말에 양파 재배면적을 결정한다. 이는 9월 중/하순에 이루어지는 양파 파종을 위한 종자와 모판 등의 준비를 보름 정도 앞둔 시점이다(이중웅, 1995). 따라서 월별로 분석된 감성지수(SI)는 양파의 파종기를 기점으로 주된 의사결정이 이루어지는 8월에 근접할수록 양파 재배면적 의사결정에 더 유의미한 영향을 미칠 것으로 가정해 <표 1>과 같이 재배면적 함수에 반영되었다. 모형 1, 2, 3에서는 각각 6월(SI_6), 7월(SI_7), 8월(SI_8)의 감성지수가 포함되었으며, 모형 4에서 6·7월의 감성지수(SI_{67}), 모형 5에서 7·8월의 감성지수(SI_{78}), 모형 6에서 7·8·9월의 감성지수(SI_{678})가 포함되었다. 이렇게 월별 데이터가 둘 이상 포함된 변수는 파종기(8월 말)에 가까울수록 가중치를 두어 가까운 기간의 데이터가 재배면적 결정에 더 큰 영향을 미치는 점을 반영했다.

표 1. 도입된 감성지수 기간별 재배면적 함수

모형	다중회귀분석식
모형 1	(1.1) $A_t = \beta_0 + \beta_1 P_{Ot-1} + \beta_2 P_{Gt-1} + \beta_3 A_{t-1} + \beta_4 C_t + \beta_5 SI_6 + \beta_6 trend + e_t$
모형 2	(1.2) $A_t = \beta_0 + \beta_1 P_{Ot-1} + \beta_2 P_{Gt-1} + \beta_3 A_{t-1} + \beta_4 C_t + \beta_5 SI_7 + \beta_6 trend + e_t$
모형 3	(1.3) $A_t = \beta_0 + \beta_1 P_{Ot-1} + \beta_2 P_{Gt-1} + \beta_3 A_{t-1} + \beta_4 C_t + \beta_5 SI_8 + \beta_6 trend + e_t$
모형 4	(1.4) $A_t = \beta_0 + \beta_1 P_{Ot-1} + \beta_2 P_{Gt-1} + \beta_3 A_{t-1} + \beta_4 C_t + \beta_5 SI_{67} + \beta_6 trend + e_t$
모형 5	(1.5) $A_t = \beta_0 + \beta_1 P_{Ot-1} + \beta_2 P_{Gt-1} + \beta_3 A_{t-1} + \beta_4 C_t + \beta_5 SI_{78} + \beta_6 trend + e_t$
모형 6	(1.6) $A_t = \beta_0 + \beta_1 P_{Ot-1} + \beta_2 P_{Gt-1} + \beta_3 A_{t-1} + \beta_4 C_t + \beta_5 SI_{678} + \beta_6 trend + e_t$

주: $SI_{67} = SI_6 \times 0.4 + SI_7 \times 0.6$, $SI_{78} = SI_7 \times 0.4 + SI_8 \times 0.6$, $SI_{78} = SI_6 \times 0.1 + SI_7 \times 0.3 + SI_8 \times 0.6$

3.3. 시장 균형

본 연구에서는 양파 수급모형을 작성하기 위해서 <표 2>에 제시된 행태방정식을 활용했다. 일반적으로 단수함수의 추정에는 기상 여건(평균온도, 최저온도, 강수량, 일조시간, 풍속 등)과 생육 정보, 기술변수의 함수로 이루어지나, 모형의 단순화를 위해 식 (2)와 같이 기술 발전을 대표하는 대리변수(proxy variable)로 하여 추세변수(*trend*)만을 사용했다. 구축된 단수함수에서는 다른 내생변수를 포함하지 않아 수급 균형 모형 시스템에 영향을 주지 않는다. 단년 생 작물의 생산량은 식 (3)과 같이 재배면적과 단수의 곱으로 도출되며, 계산된 총생산량과 이 전재고량, 수입량의 합이 당해의 총공급량이 된다. 소비함수는 식 (5)와 같이 당해의 양파 가격, 마늘 가격을 포함했고 1인당 실질 GDP를 통해 1인당 소비량을 계산했다. 총소비량은 계산된 1인당 소비량에 해당 년도의 인구수를 곱함으로써 도출되며, 총수요량은 식 (6)과 같이 총소비량과 기말재고량, 수출량의 합이다.

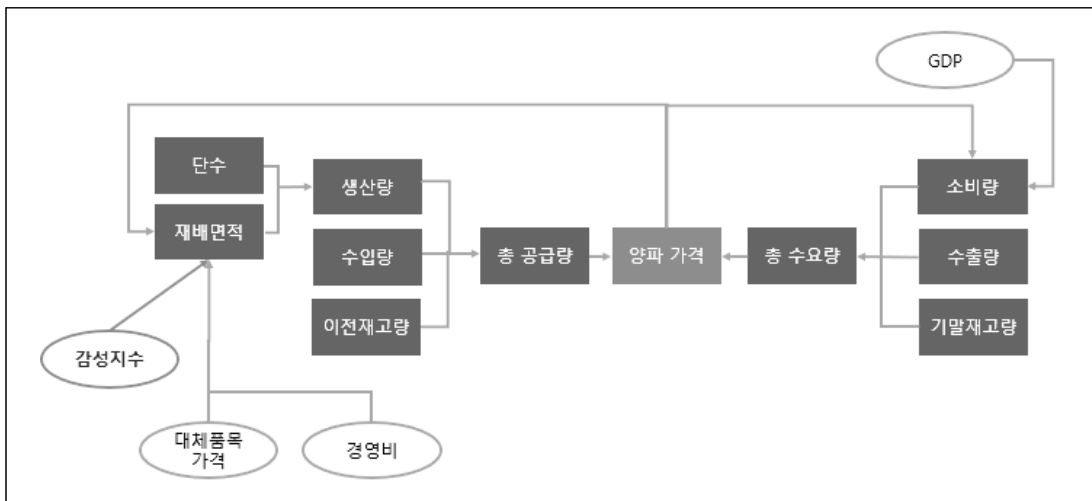
이렇게 구성된 균형 모형은 양파에 관한 단일품목 모형으로 마늘 도매가격, 이전재고량, 수입량, 기말재고량, 수출량, 1인당 GDP, 인구수는 모형 내에서 결정되지 않는 외생변수로 가정한다. 미래 전망을 위한 시나리오 분석 시에는 마늘 도매가격, 이전재고량, 수입량, 기말재고량, 수출량, 1인당 명목 GDP, GDP 디플레이터는 전년도 5년 이동평균으로 계산하고, 인구는 통계청의 장래인구추계 자료를 활용하기로 한다.

표 2. 수급 균형 모형 도출을 위한 함수

구분		행태방정식
공급	단수함수	(2) $YD = f(trend)$
	총생산함수	(3) $YD_t \times A_t$
	총공급량	(4) 총생산 + 이전재고량 + 수입량
수요	소비함수	(5) $C = f(P_{ot}, P_{gt}, GDP_t)$
	총수요량	(6) 총소비 + 기말재고량 + 수출량
시장청산		(7) 총공급(S) = 총수요(D)

〈그림 4〉는 양파의 수급 구조를 나타낸다. 총공급량(s)과 총수요량(D)이 일치할 때, 즉 초과공급($S - D$) 또는 초과수요($D - S$)가 0이 될 때 시장청산(market clearing)이 이루어지고 양파의 시장균형가격(equilibrium price)이 도출된다.

그림 4. 양파의 수급 구조



자료: 저자 작성

시장균형가격은 식 (8)처럼 P_{i-1} 이 높게 설정되었다면 공급은 증가하고 수요는 감소해 초과공급($S - D$)은 양의 값을 취하고, 조정계수(equilibrators) 앞의 음의 부호 때문에 P_i 는 P_{i-1} 보다 낮아져 초과공급($S - D$)이 0이 될 때까지의 반복(iteration)과정을 통해 도출된다.

$$(8) \quad P_i = P_{i-1} - \delta(S_i - D_i)$$

i : 반복횟수, P : price, S : supply, D : demand, δ : equilibrators

4. 자료

4.1. 정형데이터

재배면적 함수 분석 자료로 농산물유통정보(KAMIS)에서 제공하는 양파, 마늘 도매가격 자료와 통계청에서 제공하는 양파 재배면적, 농업경영종합정보시스템에서 제공하는 농축산물 소득자료집의 농가 경영비를 활용했다. 양파 가격은 양파 상품 도매가격(원/kg)을 사용하였고, 마늘 가격은 난지 마늘 상품 도매가격(원/kg)을 사용하였다. 단수는 농업관측통계시스템 자료를 활용했고, 수요함수 추정을 위해서는 농업관측본부에서 매년 발간되는 농업전망 자료집에서 확인된 양파의 1인당 소비량, 농산물유통정보(KAMIS)에서 제공하는 양파·마늘 도매가격 자료와 통계청의 1인당 GDP를 활용했다. 도매가격 자료는 재배면적함수와 소비함수에서 각각 통계청에서 제공하는 생산자물가지수, 소비자물가지수를 통해 실질가격으로 디플레이트했다. 활용된 자료의 기초통계량은 <표 3>에 제시되어 있다.

표 3. 분석모형 정형 데이터 기초통계량

변수(단위)	기간	평균	표준편차
양파 재배면적(ha)	1997~2020년	18,324	3522
양파 단수(kg/10a)	1997~2020년	6267.13	585.12
양파 도매가격(원/1kg)	1997년 4월~2020년 3월	752.76	348.28
마늘 도매가격(원/1kg)	1997년 6월~2020년 5월	3122.076	1583.55
양파 농가 경영비(원)	1997~2020년	1,055,424	513473.41
1인당 GDP(만 원)	1997~2020년	2440.63	862.05
1인당 소비량(kg)	1997~2020년	24.4	4.61

자료: 농업관측통계시스템(OASIS), 농산물유통정보(KAMIS), 농업경영종합정보시스템, 통계청, 농업전망 자료집

4.2. 비정형데이터

재배면적 함수에서 변수로 활용된 양파 관련 감성지수는 0에서 1 사이의 값을 가지며 소비자의 경기 전망 판단 정도를 분석한 경기심리지수와 유사한 역할을 한다(송민채, 2017). 언론

사는 국내 점유율 1위 포털 사이트인 네이버 뉴스에서 제공하는 기사로 설정하였으며, 기간은 1992년 1월부터 2020년 12월까지로 수집했으나, 1992년부터 1995년까지 기사는 데이터의 수가 많지 않고 직접적인 영향을 끼친 기사가 존재하지 않아 모델이 학습할 때 제외하였다. 기사 수집은 일차적으로 네이버 포털 사이트에서 양파와 관련된 모든 기사를 수집하였고, 2차적으로는 본 연구와 상관관계가 낮은 기사는 삭제하였다.

표 4. 감성지수 도출에 활용된 기사의 수

데이터	6월 기사의 수	7월 기사의 수	8월 기사의 수
1997 ~ 1999년도 기사	32개	25개	32개
2000 ~ 2004년도 기사	56개	41개	39개
2005 ~ 2009년도 기사	82개	76개	78개
2010 ~ 2014년도 기사	89개	95개	57개
2015 ~ 2020년도 기사	162개	117개	121개
합계	421개	354개	317개
총합계	1,092개		

자료: 1997~2020년 네이버 포털 양파 관련 기사

5. 분석 결과

5.1. 감성지수

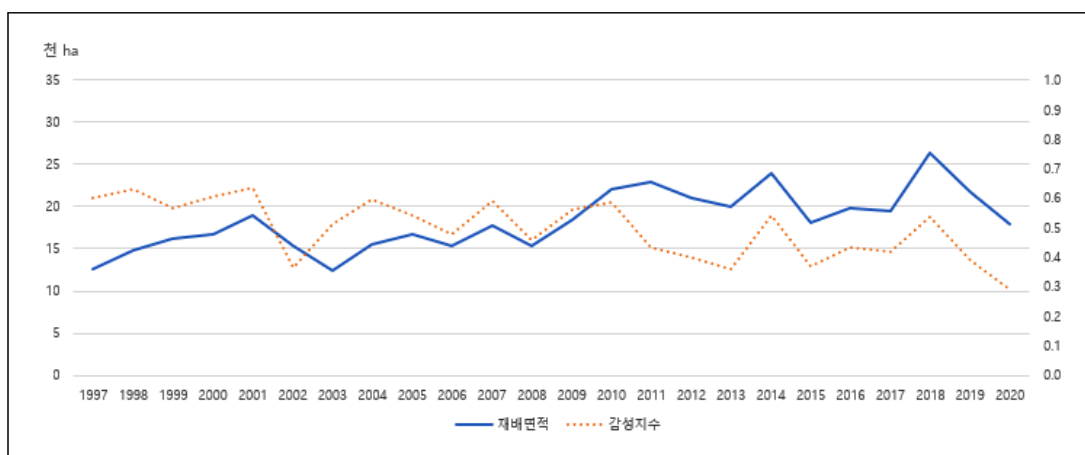
감성지수는 비정형 데이터로 일정한 규칙이 없고, 특정 사건에 단기적으로 영향을 받기 쉽다는 특성이 있으므로 양파의 파종 시기(8월 말)에 근접한 6, 7, 8월 감성지수를 채택하였다. 도출한 감성지수에 대한 기초통계량은 아래의 <표 5>와 같다.

표 5. 감성지수 기초통계량

변수(단위)	기간	평균	표준편차
감성지수(6월)	1997~2020년	0.47	0.11
감성지수(7월)	1997~2020년	0.50	0.13
감성지수(8월)	1997~2020년	0.50	0.12

아래 <그림 6>은 6월 감성지수 데이터에 0.1, 7월 감성지수 데이터에 0.3, 8월 감성지수 데이터에 0.6의 가중치를 준 자료와 재배면적의 추이를 비교한 그래프이다. 감성지수가 상승할 때 재배면적에 긍정적 영향을 미치는 것으로, 하락할 때 재배면적에 부정적 영향을 미치는 것으로 해석할 수 있으며, 감성지수와 재배면적의 추세는 선행 또는 후행하지 않고 동행하는 행태를 보인다. 감성지수가 가장 하락한 값을 보인 연도는 2002년과 2020년으로 2002년의 감성지수는 한-칠레 FTA 관련 기사 자료가 많아 자료가 많아, 양파 공급 과잉을 예상해 낮게 나타났음을 알 수 있다. 2020년의 경우 재배면적에 직접적인 영향을 끼치는 기사의 수는 적고 COVID-19의 영향으로 인해 유통 과정의 축소 및 소비자 소비 위축 심리가 반영된 기사가 많아 당해 감성지수가 하락하였고, 이에 따라 재배면적 또한 2019년도에 비해 감소한 것으로 나타났다.

그림 6. 감성지수와 재배면적 변동 추이



자료: 저자 작성(감성지수), 통계청(재배면적)

5.2. 양파 재배면적 반응함수 추정

<표 6>과 <표 7>은 감성지수를 도입한 log-log 형태의 재배면적 반응함수의 OLS(Ordinary Least Square) 추정결과이다. <표 6>의 모형 1, 모형 2, 모형 3은 각각 6월(SE_6), 7월(SE_7),

8월(SE_8)의 감성지수를 도입한 모형을 의미한다. 표에서 확인할 수 있는 바와 같이, 각 설명 변수 계수 값의 부호와 크기, 통계적 유의성은 다소 차이가 있다. 세 모형 모두 전년도 재배면적은 1보다 작은 계수 값을 가지며, 1% 유의수준에서 통계적으로 유의한 것으로 나타났다. 감성지수가 1% 증가했을 때, 모형 1, 2, 3은 각각 재배면적이 8%, 15%, 51% 증가하는 것으로 나타나 파종기에 근접할수록 감성지수가 재배면적 의사결정에 미치는 영향이 큰 것으로 보인다. 단, 감성지수는 모형 3에서는 1% 유의수준에서 통계적으로 유의한 것으로 나타났으나, 모형 1, 2의 감성지수 변수 유의성이 떨어지는 한계가 있다.

표 6. 재배면적 반응함수 추정 결과(모형 1, 2, 3)

변수	모형 1	모형 2	모형 3
intercept	1.636 (2.91)	0.45 (2.851)	-0.716 (2.162)
전년도 재배면적	0.822 (0.253)***	0.92 (0.247)***	0.766 (0.183)***
trend	0.358 (0.279)	0.276 (0.264)	0.054 (0.208)
경영비	-0.259 (0.143)*	-0.23 (0.134)	0.162 (0.113)
전기 양파 가격/농가물가지수	0.486 (0.137)***	0.481 (0.126)***	0.118 (0.12)
전기 마늘 가격/농가물가지수	-0.081 (0.092)	-0.066 (0.085)	0.025 (0.074)
감성지수 (모형 1: SI_6 , 모형 2: SI_7 , 모형 3: SI_8)	0.083 (0.097)	0.149 (0.086)	0.51 (0.126)***
R-squared	0.7827	0.8073	0.872

주: () 표준오차를 뜻함. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

〈표 7〉의 모형 4, 모형 5, 모형 6은 각각 6월 데이터에 0.4, 7월 데이터에 0.6의 가중치를 둔 감성지수(SE_{67}), 7월 데이터에 0.4, 8월 데이터에 0.6의 가중치를 둔 감성지수(SE_{78}), 6월 데이터에 0.1, 7월 데이터에 0.3, 8월 데이터에 0.6의 가중치를 둔 감성지수(SE_{678})를 포함한 모형이다. 표에서 확인할 수 있는 바와 같이 감성지수는 모형 4에서 10% 유의수준, 모형 5, 6에서는 1% 유의수준에서 통계적으로 유의한 것으로 나타났으며 감성지수가 1% 증가하면 농가들은 당해 양파재배면적을 각각 21%, 50%, 55% 증가시키고 있는 것으로 나타난다. 따라

서 파종기에 근접할수록 농가들은 양파와 관련된 정보에 민감하게 반응하는 것으로 분석된다.

표 7. 재배면적 반응함수 추정 결과(모형 4, 5, 6)

변수	모형 4	모형 5	모형 6
intercept	0.2 (2.9)	-1.442 (2.336)	-1.513 (2.313)
전년도 재배면적	0.954 (0.251)***	0.954 (0.19)***	0.956 (0.188)***
trend	0.29 (0.261)	0.115 (0.216)	0.117 (0.214)
생산비	-0.238 (0.133)*	-0.002 (0.124)	0.007 (0.123)
전기 양파 가격/농가물가지수	0.507 (0.127)***	0.295 (0.108)**	0.29 (0.107)**
전기 마늘 가격/농가물가지수	-0.08 (0.085)	-0.005 (0.07)	-0.009 (0.069)
감성지수 (모형 3: SI_{67} , 모형 4: SI_{78} , 모형 5: SI_{678})	0.21 (0.114)*	0.503 (0.131)***	0.545 (0.139)***
R-squared	0.8101	0.8782	0.8809

주: () 표준오차를 뜻함. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

아래의 <표 8>은 감성지수가 포함된 모형과 감성지수를 모형에 포함하지 않았을 때의 재배면적 추정 정확도를 비교하기 위해 추가한 log-log 형태로 추정된 다중회귀분석식이다.

표 8. 감성지수를 포함하지 않은 재배면적 함수

모형	다중회귀분석식
모형 7	(9) $A_t = \beta_0 + \beta_1 P_{A-1} + \beta_2 P_{A-1} + \beta_3 A_{t-1} + \beta_4 C_t + \beta_6 trend + e_t$

추정값과 실제값을 비교하는 척도로서 평균 제곱근 오차(RMSE)를 이용했으며, 모형 1~7의 RMSE는 다음 <표 9>와 같다. 감성지수를 포함하지 않은 모형 7의 RMSE가 가장 높게 나타났고, 파종기인 8월 말에서 상대적으로 먼 시기인 6월을 포함한 모형 1, 모형 4 순서대로 RMSE가 크게 나타났다. 하지만, 6, 7, 8월 감성지수를 포함한 모형 6의 RMSE가 7, 8월 감성지수만을 포함한 모형 5의 RMSE보다 더 낮은 값을 갖는 것으로 나타났다. 이는 파종기와 시

차가 거의 없는 기사만이 아닌 최대 3개월 전의 기사도 또한 낮은 가중치(0.1)를 부여해 함께 고려했을 때, 추정된 모형의 정확도가 올라감을 알 수 있다.

표 9. 모형별 RMSE 값 비교

	모형 1	모형 2	모형 3	모형 4	모형 5	모형 6	모형 7
RMSE	1,661	1,553	1,192	1,519	1,133	1,107	1,724

5.3. 수급 균형 모형 구축

양파 시장에 대한 수급 균형 모형 구성을 위해 식 (1.6)과 식 (2)~식 (7)의 방정식을 이용했다. 본 연구의 균형 모형에서 마늘 도매가격, 이전재고량, 수입량, 기말재고량, 수출량, 1인당 GDP, 인구수는 모형 내에서 결정되지 않는 외생변수로 가정한다. 재배면적 함수는 RMSE가 가장 낮았던 <표 7>의 모형 6 추정결과를 사용했으며, 단수함수의 추정 결과는 <표 10>, 소비함수의 추정 결과는 <표 11>에 정리되어 있다.

단수함수 추정결과, 기술진보에 따라 단수가 양(+)의 방향으로 증가하는 것으로 나타났다. 소비함수 추정결과, 소득의 대리변수인 1인당 GDP의 계수가 양(+)의 방향으로, 양파 가격이 음(-)의 방향으로 경제 이론과 부합하는 것으로 추정되었으며, 유의성도 각각 1%와 10%에서 유의한 것으로 높게 나타났다. 대체가격인 마늘 가격은 통계적으로 유의하게 추정되지 않았으며, 이는 소비자에게는 같은 양념채소더라도 마늘이 양파의 대체 품목이 되기 어렵기 때문으로 해석된다.

표 10. 단수함수 추정결과

변수	Coefficient
intercept	5121 (412.03)***
trend	43.23 (15.04)***

주: () 표준오차를 뜻함. *** p<0.01 ** p<0.05 * p<0.1

표 11. 소비함수 추정결과

변수	Coefficient
intercept	12.5368 (2.34)***
1인당 GDP	0.0063 (0.0008)***
양파 가격/소비자물가지수	-0.0079 (0.0021)***
마늘 가격/소비자물가지수	0.0006 (0.0004)

주: () 표준오차를 뜻함. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

소비에 대한 자체가격 탄력성, 소득 탄력성은 <표 12>와 같다. 자체가격 탄력성은 -0.29로 나타났으며, 이는 양파 가격이 1% 상승할 때 1인당 소비량은 0.29% 감소한다는 의미로 가격 변화에 대한 소비량 변화가 비탄력적임을 알 수 있다.

표 12. 수요의 자체가격 탄력성, 소득 탄력성

구분	탄성치 계산값
자체가격 탄력성	-0.28581
소득 탄력성	0.82409

5.4. 베이스라인

추정된 모형 6의 재배면적 함수, 단수함수, 소비함수를 <표 5>의 계산식을 사용해 총공급 함수, 총수요함수로 계산하고 식 (8)과 같은 방식으로 초과공급($S - D$) 혹은 초과수요($D - S$)를 줄이며 반복계산을 통해 시장 균형에 이를 수 있다. 모형에서 결정된 2021~2025년 전망치는 <표 13>에 정리되어 있다. 2021년 베이스라인은 한국농촌경제연구원의 농업전망 2021에서 제시하는 중장기 수급 전망 자료를 참고하여 결정되었다. 감성지수는 직전 3년 감성지수의 평균값이 2021~2025년에도 같다는 가정 하에 6월의 감성지수가 0.35, 7월이 0.48, 8월이 0.35로 계산되었다.

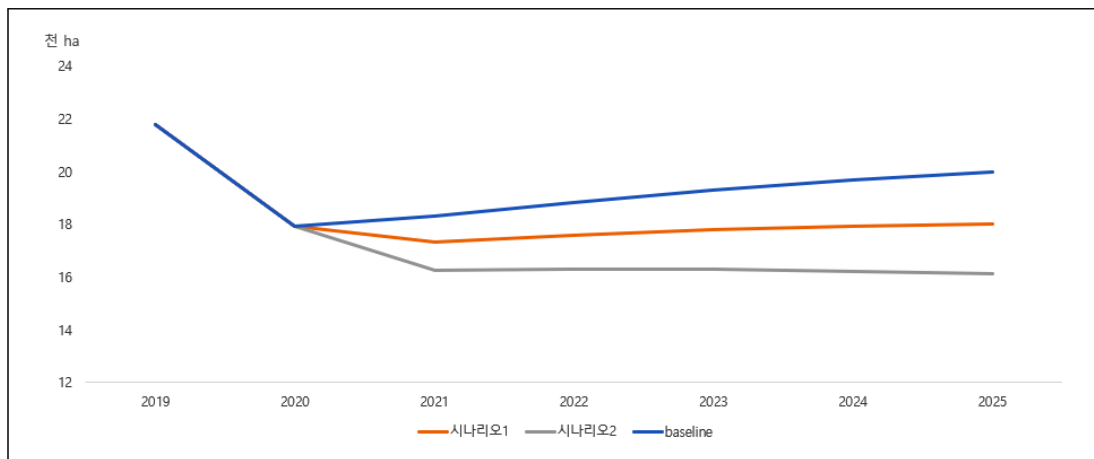
표 13. 양파 베이스라인(2021~2025년)

변수	2021	2022	2023	2024	2025
재배면적(ha)	18,336	18,843	19,292	19,673	19,986
생산량(천 톤)	1,247	1,282	1,314	1,340	1,362
소비량(천 톤)	1,318	1,359	1,314	1,340	1,394
도매가격(원/kg)	1,361	1,282	1,257	1,185	1,117
6월 감성지수	0.35	0.35	0.35	0.35	0.35
7월 감성지수	0.48	0.48	0.48	0.48	0.48
8월 감성지수	0.35	0.35	0.35	0.35	0.35

5.5. 시나리오 분석

시나리오 1은 2021년 7월의 감성지수가 0.1점 하락했을 때, 시나리오 2는 2021년 8월의 감성지수가 0.1점 하락했을 때의 상황을 가정한다. 이때 <그림 7>은 2021~2025년의 재배면적 전망을 나타내는 그래프이다. 재배면적은 시간이 지남에 따라 증가하는 양상을 보이며, 감성지수가 상승하면 재배면적도 증가하고 감성지수가 하락하면 재배면적도 감소한다.

그림 7. 양파 재배면적 시나리오 분석



2021년의 결과를 비교하면, 시나리오 1에서는 베이스라인에 비해 재배면적이 5.51% 하락했으며, 생산량 역시 5.51% 하락, 1인당 소비량은 5.21% 하락하고, 도매가격은 12.74% 증가하였다. 시나리오 2에서는 베이스라인에 비해 재배면적과 생산량이 11.3% 감소하였으며, 1인당 소비량은 10.70% 감소하고, 도매가격은 26.12% 증가하였다. 시나리오 분석 결과는 감성지수 가중치에 따라 달라질 수 있다(부록 참고).

표 14. 시나리오 분석 결과(2021~2025년)

	2021	2022	2023	2024	2025
재배면적(ha)					
베이스라인	18,336	18,843	19,292	19,673	19,986
시나리오 1	17,326	17,588	17,793	17,932	18,007
% 변화	-5.51%	-6.66%	-7.77%	-8.85%	-9.90%
시나리오 2	16,264	16,310	16,304	16,243	16,129
% 변화	-11.30%	-13.44%	-15.49%	-17.44%	-19.30%
생산량(천 톤)					
베이스라인	1,247	1,282	1,314	1,340	1,362
시나리오 1	1,179	1,197	1,212	1,222	1,227
% 변화	-5.51%	-6.66%	-7.77%	-8.85%	-9.90%
시나리오 2	1,106	1,110	1,110	1,107	1,099
% 변화	-11.30%	-13.44%	-15.49%	-17.44%	-19.30%
소비량(천 톤)					
베이스라인	1,318	1,359	1,314	1,340	1,394
시나리오 1	1,249	1,273	1,269	1,275	1,290
% 변화	-5.21%	-6.28%	-7.64%	-8.51%	-9.47%
시나리오 2	1,177	1,186	1,166	1,160	1,162
% 변화	-10.70%	-12.69%	-14.86%	-16.76%	-18.45%
도매가격(원/kg)					
베이스라인	1,361	1,282	1,257	1,185	1,117
시나리오 1	1,534	1,498	1,515	1,485	1,458
% 변화	12.74%	16.84%	20.55%	25.31%	30.52%
시나리오 2	1,716	1,718	1,772	1,776	1,781
% 변화	26.12%	34.00%	40.95%	49.86%	59.49%

6. 결론 및 시사점

본 연구는 감성분석을 통해 얻어지는 비정형 데이터인 양파의 관련 감성지수를 재배면적 반응함수의 변수로 활용함으로써 양파 가격 전망의 정확도를 높이고, 시나리오 분석을 통해 감성지수가 재배면적에 미치는 영향을 제시해 양파 재배 농민 및 정책 수립자들에게 보다 정확한 데이터를 제공하고자 수행되었다. 이와 같은 연구목적을 달성하기 위하여 본 연구에서는 국가통계포털과 KAMIS, 농축산물 소득자료집을 통해 양파 재배면적, 양파 단수, 양파 도매가격, 마늘 도매가격, 양파농가 경영비, 1인당 GDP, 1인당 소비량 자료를 수집하였다. 이후 웹 크롤링 방식으로 기사를 수집하고, LSTM-CNN 모델을 이용해 재배면적 함수에 활용할 감성지수를 도출하였다. 감성지수를 월별(6, 7, 8월), 가중치를 둔 기간별(6·7월, 7·8월, 6·7·8월)로 분류, OLS(Ordinary Least Square)로 추정한 양파 재배면적 분석모형에 도입하였다. 각 모형별 평균 제곱근 오차(RMSE)를 이용해 가장 적합한 모형을 선정한 후, 추정된 재배면적 함수를 활용하여 양파 수급예측 모형을 설계하였으며 시나리오 분석을 통해 양파 감성지수가 재배면적에 미치는 영향을 분석하였다.

본 연구에 대한 분석 결과를 요약하면 다음과 같다. 첫째, 양파 재배면적 함수에 감성지수를 도입하고, 파종기에 가까울수록 감성지수의 추정된 계수가 통계적으로 유의해진다는 점을 확인했다. RMSE를 통해 비교한 결과 6월, 7월, 8월에 각각 0.1, 0.3, 0.6의 가중치를 둔 모형이 가장 설명력이 높게 나타났으며, 이를 감성지수가 포함되지 않은 모형과 비교했을 때 감성지수를 변수로 포함한 경우에 실측값과의 괴리가 감소하는 것으로 밝혀졌다. 둘째, 양파 단수 함수와 소비함수를 추정하여 총공급함수와 총수요함수를 구축하고, 감성지수의 움직임에 따른 시장균형가격의 움직임을 살핀 뒤 감성지수의 변화에 따른 2021~2025년 재배면적 전망치를 살펴보았다. 그 결과, 7월의 감성지수가 0.1점 감소(시나리오 1)했을 때 2021년에 베이스라인에 비해 재배면적이 5.51% 감소하였으며, 8월의 감성지수가 0.1점 감소(시나리오 2)했을 때는 베이스라인에 비해 11.30% 감소하는 모습을 보였다.

이러한 분석 결과를 토대로 한 시사점은 다음과 같다. 본 연구에서는 부동산, 금융과 같은 경제학 분야에서 활용되고 있으나 농업 분야에서는 활용이 미진했던 뉴스의 감성분석을 통한

감성지수를 재배면적 함수에 설명변수로 추가하고, 비정형 데이터를 활용하여 재배면적 추정 및 가격 전망의 정교화가 가능함을 보였다. 이는 농업 부문 전망모형에서 기존에 활용되던 여러 경제적 변수 외에도 시장 동향 및 전망 예측에 감성분석과 같은 빅데이터 분석이 활용될 수 있다는 점을 환기하는 결과이다.

본 연구는 다음과 같은 몇 가지 한계를 지닌다. 첫째로, 가격자료 확보의 어려움으로 모든 데이터를 1996년부터 2020년까지 수집하여, 상대적으로 적은 기간의 자료가 활용되었다. 또, 품종이 조생종과 중만생종으로 나뉘는 양파와 생산지 기후에 따라 난지형과 한지형으로 나뉘는 마늘의 특성을 고려하지 못하고 양파는 두 품종을 한꺼번에, 마늘은 난지형 자료만 사용했다. 보다 정확한 추정을 위해서는 본 연구에서 사용된 자료보다 더 이전의 과거 자료를 확보해 데이터의 길이를 늘려야 하며, 확보하고 활용할 수 있어야 하며, 양파와 마늘의 품종이 다를 수 고려하여 두 품종에 따라 다른 변수로 포함하는 과정이 필요하다고 판단된다.

둘째, 수급 균형 모형의 단순화를 위하여 재배면적 함수 외에 다른 함수에서 설명변수가 적게 활용되었고, 수입수요 함수, 재고량 함수 등의 함수가 포함되지 않았다. 특히 양파 단수 함수에는 기술 발전의 대리변수인 추세변수(*trend*)만이 사용되어 모형 적합도가 낮은 문제가 있다. 선행연구에서 연구된 양파 단수 결정함수와 같이 기상변수, 생육 정보 등을 함수에 포함시켜 균형 모형 분석에 반영한다면 균형 모형의 체계화가 가능할 것으로 보인다.

그럼에도 불구하고 본 연구는 양파 관련 뉴스 감성지수를 산출하여 재배면적에 영향을 미치는지 살펴보고, 이를 통해 수급전망을 실시했다는 점에서 연구적 의의가 있다. 감성지수 산출을 통해 시간과 비용이 많이 드는 재배의향 설문조사를 대체할 수 있을 뿐만 아니라 최근 뉴스 감성지수 산출이 가능하여 즉각적인 수급대응에 용이할 것이다. 마지막으로 빅데이터를 활용한 분석기법을 농업 수급모형에 활용할 수 있는 기초연구로 활용될 것으로 기대한다.

참고문헌

- 강산하, 오승원, 강혜정. 2017. “농업분야 빅데이터의 활용 사례와 소셜 네트워크 데이터 분석을 통한 농산물 가격 분석.” 「한국농식품정책학회 2017년 하계학술대회 논문집」 한국농식품정책학회.
- 나형철, 오은화, 유도일, 조완섭, 이지스 나스리디노프, 박성호, 조용빈, 류관희. 2018. “정형 및 비정형 빅데이터를 이용한 양파 소비 예측.” 「한국콘텐츠학회논문지」 제18권 제11호. 한국콘텐츠학회.
- 류재현, 이진아, 안병일. 2020. “농업관측 정보가 농간의 재배면적 조절에 미치는 영향: 양파와 마늘을 중심으로.” 제61권 제1호. 한국농촌계획학회 「농업경제연구」.
- 박재수, 이재수. 2019. “아파트 매매가격과 부동산 온라인 뉴스의 교차상관관계와 인과관계 분석: 온라인 뉴스 기사의 비정형 빅데이터를 활용한 감성분석 기법의 적용.” 「國土計劃」 제54권 제1호. 대한 국토·도시계획학회.
- 서홍석, 순병민, 김충현, 김문희. 2019. 「농업부문 전망모형 KREI-KASMO 2019 운용·개발 연구」. 한국 농촌경제연구원.
- 송민채, 신경식. 2017. “뉴스기사를 이용한 소비자의 경기심리지수 생성.” 「지능정보연구」 제23권 3호. 한국지능정보시스템학회.
- 이승인, 손찬수, 이해림. 2020. “기상 및 영농투입재 비용이 양파 단수에 미치는 영향: 패널회귀분석과 베이지안네트워크모형을 중심으로.” 「농촌경제」 제43권 제2호. p. 3, 한국농촌경제연구원 보고서.
- 이중용. 1995. “마늘, 양파 주산지 육성을 위한 기본방향.” 「농촌경제」 제18권 제4호, 한국농촌경제연구원 보고서.
- 차은정, 홍태호. 2016. “뉴스 감성분석과 SVM을 이용한 S&P 500 주가지수 예측.” 「한국경영정보학회 학술대회논문집」 제6호. 한국경영정보학회.
- 하지희, 서상택, 김선웅. 2019. “양파와 마늘가격 예측모형의 예측력 고도화 방안.” 「농촌계획」 제25권 제4호. 한국농촌계획학회 「농업경제연구」.
- 한국농촌경제연구원. 2021. 『농업전망 2021: 코로나19 이후 농업·농촌의 변화와 미래』. pp. 442-455. 한국농촌경제연구원 보고서.
- Jin, Z, Y Yang, Y Liu - Neural Computing and Applications, 2019 - Springer Stock closing price prediction based on sentiment analysis and LSTM.
- Saura, JR , P Palos-Sanchez, A Grilo - Sustainability, 2019 Detecting indicators for startup business success: Sentiment analysis using text data mining.
- Tang, D, F Wei, N Yang, M Zhou, T Liu... - Proceedings of the 52nd ..., 2014 Learning sentiment-specific word embedding for twitter sentiment classification.
- Yenter, A, A Verma - 2017 IEEE 8th Annual Ubiquitous ..., 2017 Deep CNN-LSTM with combined kernels from multiple branches for IMDB review sentiment analysis.
- Ynkun Ma. and Haiyun Peng and Erik Cambria. 2018. “Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM.”

부록

부록 1. 양파 재배면적, 생산량, 1인당 소비량, 도매가격 전망치 표(시나리오 2 가중치 변경)

	2021	2022	2023	2024	2025
재배면적(ha)					
베이스라인	18,336	18,843	19,292	19,673	19,986
$SI_6 \times 0.3 + SI_7 \times 0.3 + SI_8 \times 0.4$	16,981	17,118	17,201	17,224	17,190
% 변화	-7.39%	-9.15%	-10.84%	-12.45%	-13.99%
$SI_6 \times 0.6 + SI_7 \times 0.3 + SI_8 \times 0.1$	17,891	18,173	18,399	18,562	18,664
% 변화	-2.43%	-3.56%	-4.63%	-5.65%	-6.62%
생산량(천 톤)					
베이스라인	1,247	1,282	1,314	1,340	1,362
$SI_6 \times 0.3 + SI_7 \times 0.3 + SI_8 \times 0.4$	1,155	1,165	1,171	1,173	1,172
% 변화	-7.39%	-9.15%	-10.84%	-12.45%	-13.99%
$SI_6 \times 0.6 + SI_7 \times 0.3 + SI_8 \times 0.1$	1,217	1,237	1,253	1,265	1,272
% 변화	-2.43%	-3.56%	-4.63%	-5.65%	-6.62%
소비량(천 톤)					
베이스라인	1,318	1,359	1,314	1,340	1,394
$SI_6 \times 0.3 + SI_7 \times 0.3 + SI_8 \times 0.4$	1,226	1,241	1,227	1,227	1,234
% 변화	-6.99%	-8.64%	-10.40%	-11.97%	-13.38%
$SI_6 \times 0.6 + SI_7 \times 0.3 + SI_8 \times 0.1$	1,288	1,313	1,308	1,318	1,335
% 변화	-2.30%	-3.36%	-4.44%	-5.43%	-6.33%
도매가격(원/kg)					
베이스라인	1,361	1,282	1,257	1,185	1,117
$SI_6 \times 0.3 + SI_7 \times 0.3 + SI_8 \times 0.4$	1,593	1,579	1,617	1,607	1,599
% 변화	17.08%	23.15%	28.66%	35.60%	43.13%
$SI_6 \times 0.6 + SI_7 \times 0.3 + SI_8 \times 0.1$	1,437	1,398	1,410	1,376	1,344
% 변화	5.61%	8.99%	12.24%	16.15%	20.40%



제1회

농산물 수급 예측모형 경진대회

우수논문집

Chapter 1
지정과제



장려상

실측기반 생육정보와 기상정보를 고려한
양파 생산 단수 예측 및 사회후생효과 추정

문지혜 · 정채은

실측기반 생육정보와 기상정보를 고려한 양파 생산 단수 예측 및 사회후생효과 추정

문지혜* 정채은**

Key words

단수 예측(yield estimation), 실측기반 생육요인(actual growth factor), 기상요인(meteorological factor), 사회후생효과(social welfare effects)

ABSTRACT

In order to stabilize the supply and demand of agricultural products, it is important to take a preemptive response through accurate production forecasting. Onions have been an item with high price volatility since the past. As one of the representative methods of stabilizing supply and demand, KREI's agricultural outlook service advanced agricultural outlook service by introducing an actual survey system. The purpose of this study is to predict onion yield and analyze the economic feasibility using onion growth data and meteorological data obtained through the actual survey. Accordingly, onion yield using actual growth data, onion yield using meteorological data, and onion yield using hybrid data (the actual growth data, meteorological data) were respectively estimated and the predictive power was compared. As a result of the analysis, the hybrid model using actual growth data and weather data had the best predictive power. Finally, the maximum social welfare effect of the predicted values using growth data and meteorological data was estimated to be about KRW 25 billion. This study is meaningful in that it verified the economic effect of the hybrid model by estimating the social welfare effect of predicted values using actual growth data and meteorological data.

차례

- | | |
|------------|--------------|
| 1. 서론 | 4. 추정 결과 |
| 2. 선행연구 검토 | 5. 사회후생효과 분석 |
| 3. 분석 모형 | 6. 요약 및 결론 |

* 전남대학교 농업경제학과 석사과정

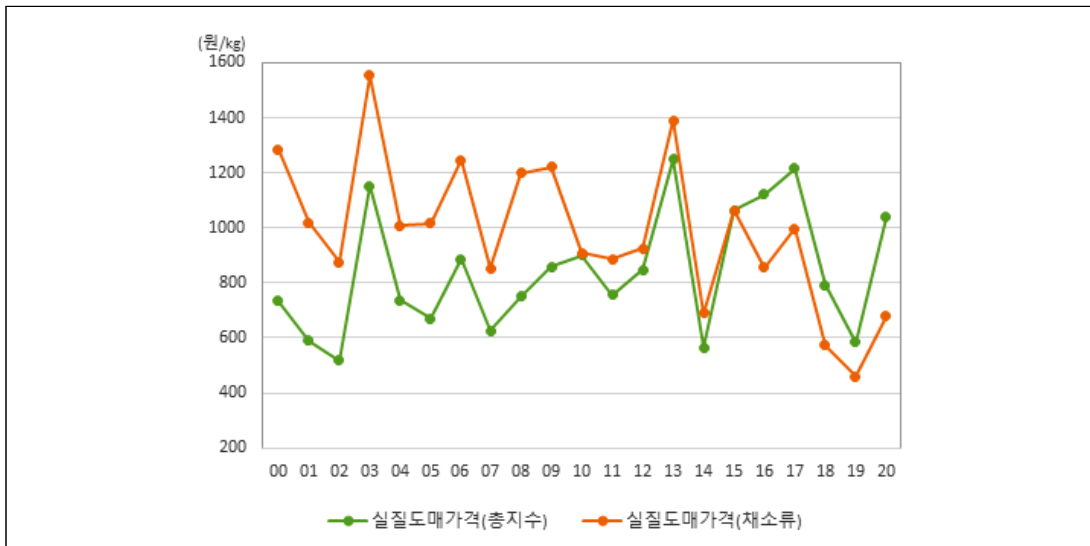
** 전남대학교 농업경제학과 석사과정

1. 서론

정부의 5대 수급관리 채소인 양파, 배추, 무, 고추, 마늘은 국민생활과 밀접한 관계가 있으며 수급과 가격 변동이 큰 품목으로 전체 채소류의 가격안정을 위해 필수적인 품목이다. 그 중 양파는 평균 자급률('15~'19)이 92%로 수입의 영향이 비교적 낮다. 농산물 특성상 수요가 일정하다고 가정한다면, 양파 수급 및 가격 변동의 원인은 주로 공급측면에서 발생한다고 볼 수 있다(남국현 외 2015).

양파는 재배면적 변동이 심하고 가격의 폭·등락 현상이 반복되고 있다(그림 1). 2019년 양파 도매가격은 평년 대비 45.1% 하락하며 '양파 가격 파동'을 겪었다. 주된 원인은 기상 호조로 인한 과잉생산이었다. 공급측면에서 정확한 단수예측이 이루어진다면 농산물의 선제적인 수급 안정대책 수립이 가능할 것으로 짐작할 수 있는 부분이다. 단수예측은 기상, 환경, 기타 인자들의 관계를 기반으로 개발된 모형을 통해 이루어지나, 기상이변으로 인한 생산량 증감 등 변수가 많아 단수예측이 어려운 실정이다.

그림 1. 양파 연도별 가격 변화



주: 실질도매가격은 도매가격을 총 생산자물가지수 및 채소류 생산자물가지수로 디플레이트 하였음(2015=100).

자료: 통계청: aT KAMIS

양파 수급 및 가격 예측의 어려움에 대응하여 정부는 정부수매, 산지폐기, 농산물 수급조절 매뉴얼 운영 등을 통해 안정화를 도모하고 있다. 대표적인 수급 안정화 방안 중 하나로 한국농촌경제연구원 농업관측본부가 주관하여 국내 농산물 생산 및 가격 예측을 실시하고 있다. 농업관측본부는 2020년부터 실측기반 산지정보 조사체계를 도입하여 농업관측사업을 고도화시켰으며, 실측조사 자료를 농업관측통계시스템에 공개하고 있다.

본 연구의 목적은 양파를 대상으로 실측기반 생육자료 및 기상자료를 활용하여 단수를 예측하는 것이다. 실측기반 생육자료를 활용하기 위해 양파의 생물학적 생육주기를 살펴보고자 한다.

그림 2. 농업관측통계시스템 생육 실측조사 결과 자료

년도	조사일자	필지ID	품목	NO	번목	재배번호	시도	시군구	품종	조사일
2020	2	41	양파	41	2	8	전라남도	무안군	조생종	2020-03-02
2020	2	41	양파	41	2	9	전라남도	무안군	조생종	2020-03-02
2020	2	41	양파	41	2	10	전라남도	무안군	조생종	2020-03-02
2020	2	41	양파	41	3	1	전라남도	무안군	조생종	2020-03-02
2020	2	41	양파	41	3	2	전라남도	무안군	조생종	2020-03-02
2020	2	41	양파	41	3	3	전라남도	무안군	조생종	2020-03-02
2020	2	41	양파	41	3	4	전라남도	무안군	조생종	2020-03-02
2020	2	41	양파	41	3	5	전라남도	무안군	조생종	2020-03-02
2020	2	41	양파	41	3	6	전라남도	무안군	조생종	2020-03-02
2020	2	41	양파	41	3	7	전라남도	무안군	조생종	2020-03-02
2020	2	41	양파	41	3	8	전라남도	무안군	조생종	2020-03-02

자료: 농업관측통계시스템(<https://oasis.krei.re.kr/>)

양파의 생육주기¹⁾는 크게 영양생장기와 생식생장기로 구분할 수 있다. 영양생장기는 다시 육묘기-활착기-월동기-경엽신장기로 구분되며, 생식생장기는 구 비대기-휴면기 등으로 구분된다.

육묘기는 씨앗이 발아해서 본 밭에 아주심기(정식)를 하기 전까지이며, 활착기는 정식 후 토양의 양분과 수분의 흡수기능이 발휘되는 시기, 월동기는 12월부터 2월 초까지이다. 경엽신장기는 월동기 이후의 뿌리와 잎의 생장이 빨라지는 시기이며, 특히 양파의 지상부와 뿌리가 자라는 시기이다. 양파 뿌리의 생육은 잎보다 약 20일 일찍 진행된다. 뿌리의 생육에 적합한 기온은 15~20℃ 이고, 그 이상 고온이 되면 성숙한 뿌리의 노화 속도가 어린뿌리의 발생보다

1) 이종태(2015), 나우현(2000)을 참고하여 요약 정리하였다.

빨라져 뿌리의 활력이 떨어진다. 잎이 눈에 띄게 신장을 시작하는 것은 3월 중순 이후이다. 온도가 높아지면 신장이 증가해서 엽장도 길어지는 경향이 있으나, 25℃ 이상이 되면 잎의 신장도 감소해서 엽장도 짧아진다. 엽수는 고온일수록 증가가 현저하나 25℃를 넘거나 저온에서는 억제된다. 잎 수와 구의 크기는 비례하며 잎에 난 상처는 구의 비대에 나쁜 영향을 미친다. 상품성 있는 구로 키우기 위해서는 최소한 6~8매의 잎 수를 확보해야 한다.

구 비대기는 비늘잎이 형성되면서 시작된다. 특히 중만생종 양파는 12.5~13.5시간의 해길 이과 20℃ 정도의 낮 온도가 충족되어야 비늘잎이 형성되는데, 비늘잎 형성과 그 후의 구 비대에는 해길이와 온도 조건이 모두 갖춰져야 한다. 구 비대기에 잦은 강우는 습해를 발생시켜 구 비대를 억제할 가능성이 있다. 반면에 너무 건조해도 발육을 억제하여 구의 비대가 불량하게 된다. 즉, 구 형성 전후부터는 수분이 많거나 너무 적어도 생육이 억제된다는 것이다. 구 비대기 초기에는 비늘잎이 형성되기 시작하고 잎의 수, 잎의 길이와 무게는 최대치를 기록한다. 그 때 밖으로 나온 잎의 수는 7~10개 정도이다. 그 후로 잎의 무게는 줄어들고, 수확기에 이르면 잎 수는 4~6개로 줄어든다.

이와 같은 양파의 생물학적 생육주기를 고려하여 본 연구에서는 양파의 지상부가 생장하기 시작하는 경엽신장기부터 구 비대기까지의 생육요인과 육묘기부터 구 비대기까지의 기후요인을 적극 고려하였다.

2. 선행연구 검토

그동안 대부분의 선행연구에서는 기상요인을 이용한 단수예측 연구가 지속적으로 진행되어 왔다. 한석호 외(2011)는 쌀, 콩, 고랭지 배추 3개 품목을 대상으로 확률추정 방법을 적용한 단수예측모형을 개발하였다. 고정효과 모형을 이용한 패널분석 방법으로 설명변수인 기상변수의 계수를 추정하고, 다양한 시나리오를 접목시켜 확률적 기법을 활용한 단수예측모형을 개발하였다. 또한 기상요인과 단수 간의 상관관계를 이차식 형태로 고려하여 기상요소에 대한 단수의 최적점을 도출하였다. 임철희 외(2016)는 주요 채소작물들을 대상으로 농경지 단

위의 기상요인과 생산량 간의 상관관계가 높은 기상인자를 추려 회귀분석을 통해 단수를 예측하였다. 양파 단수 예측 분석의 경우 상관성이 높은 최저기온, 상대습도, 강수량을 모형에 활용하였으며 양파 단수 추정에 기여수준이 가장 높은 기상요인은 최저기온으로 나타났다. 나상일 외(2017)는 기상자료 및 위성자료(MODIS NDVI)를 사용하여 지역별 마늘, 양파 단수예측 모형을 개발하고 예측 시기별 정확도를 측정하였다. 상관관계 분석 결과, 양파의 경우에는 잎이 출현하는 시기의 위성자료와 강수량, 일조시간 등의 기상정보가 단수에 영향을 미치는 요인으로 나타났다. Chmielewski 외(2000)는 기상요인이 겨울호밀의 예측단수에 미치는 영향을 분석하기 위해 패널모형을 이용하였다. 호밀단수에 영향을 미치는 요인으로는 고온, 가뭄 등으로 나타났다.

한편 빈번한 기상변화 발생에 대비하여 예측력을 높이기 위해 생육자료와 기상자료를 모두 활용한 단수 예측 또는 요인 분석이 진행되고 있다. 이춘수 외(2017)는 생육변수(초장, 엽수, 10a당 예상수량)와 기상변수(기온, 일교차, 강수량 등)를 설명변수로 하는 pooled OLS 모형을 이용하여 가을배추와 가을무 단수 예측모형을 구축하였다. 생육변수와 기상변수 시나리오를 바탕으로 총 24개의 모형을 추정하고 모형의 예측력을 비교하였다. 김준기 외(2020)는 생육자료와 기상자료를 바탕으로 양파의 최종 구중을 추정하는 다층모형을 구축하였다. 위계적 구조를 가진 자료 분석기법을 활용하여 다층모형 구축을 시도하였지만, 1년차 자료만이 분석에 활용되었기에 모형의 신뢰도가 다소 미흡하다는 점을 지적하였다. 김재휘 외(2021)는 양파의 생육조사 자료와 실측 대상 지역의 추정 기상자료를 이용하여 양파의 생산량을 결정하는 구중과 지상부 생장이 어떤 기상요인과 시기별로 상관관계를 갖는지 분석하였다. LANCASTER 외(1996)는 일조시간과 생육요인(잎의 수, 구중), 기타요인(파종시기, 구 비대시기)이 양파의 성숙 시기에 어떤 영향을 미치는지 분석하였다. 양파의 성숙시기에 큰 영향을 미치는 것은 구 비대가 시작한 이후의 잎의 수인 것으로 나타났다.

단수 예측 모형 개발 외에도 농산물 단수예측, 가격예측 등 농업관측의 경제적 효과에 대해 분석한 연구도 이루어지고 있다. 이용선 외(2002), 김관수 외(2009), 김배성(2014)은 실증자료를 바탕으로 거미집형 기대가설 하에 품목별 수요 및 공급곡선을 추정하고 이를 통해 관측정보의 제공에 따른 사회적 후생효과를 계측하였다. 추정 결과, 품목별로 사회후생효과 측정치의 차이가 보였으며, 그 이유를 생산액, 수요의 가격탄력성, 공급의 가격탄력성, 가격변동

른의 크기가 다르기 때문이라고 주장하였다. 김종규(2009)는 축산 관측 품목 중 육계의 산지
 가격전망 정보를 활용하였을 때 기대수익 및 예측정보의 가치를 베이지안 분석모형을 이용하
 여 계측하였다. 분석 결과 농가에서 가격전망 정보를 활용하였을 때 기대수익이 증가하는 것
 으로 나타났으며, 가격전망 정보가 완전할수록 가치는 배단위로 증가한다고 주장하였다. 유
 도일(2005)은 농업관측사업을 거시 및 미시적 관점으로 경제적 성과를 분석하였다. 거시적
 분석을 위해 가격의 월평균 표준편차를 이용한 연도 내 변동계수와 연도 간 변동률을 계측하
 였다. 미시적인 분석은 농가 입장에서 농업관측정보를 전달받기 위해 정보통신비 지출이 반
 영된다는 점을 가정하여 정보통신비용을 JP형 소득함수의 위험함수 부분에 관측사업 실시 전
 후의 위험과 관계를 분석하였다.

기존 연구들은 기상자료만을 이용한 분석모형을 적용하여 단수를 예측하거나, 기상과 생육
 자료를 고려하여 단수 및 구중을 예측하고 있다. 본 연구에서는 선행연구들과 차별성을 두기
 위해 1)실측기반 생육자료를 이용한 단수, 2)기상자료를 이용한 단수, 3)혼합자료(실측기반
 생육자료 및 기상자료)를 이용한 단수를 각각 추정한다. 더 나아가 예측력이 가장 높은 모형
 의 사회후생효과 계측을 통해 경제성을 논의하고자 한다.

3. 분석 모형

3.1. 분석자료

3.1.1. 실측기반 생육자료

한국농촌경제연구원 농업관측본부가 수행하는 양파 실측조사 원시자료를 이용하였다. 한
 국농촌경제연구원에 따르면 실측조사 대상 농가는 양파 표본농가 중 5년 이상 재배한 농가이
 며, 대상 농가의 평균생육을 대표할 수 있는 필지 100개를 선정한다. 실측조사는 양파 생육기
 간 중 총 11회 실시되며, 생육초기(2~3월)는 15일 주기로 조사되고 생육 중기이후(4~6월)는
 10일 주기로 조사된다.

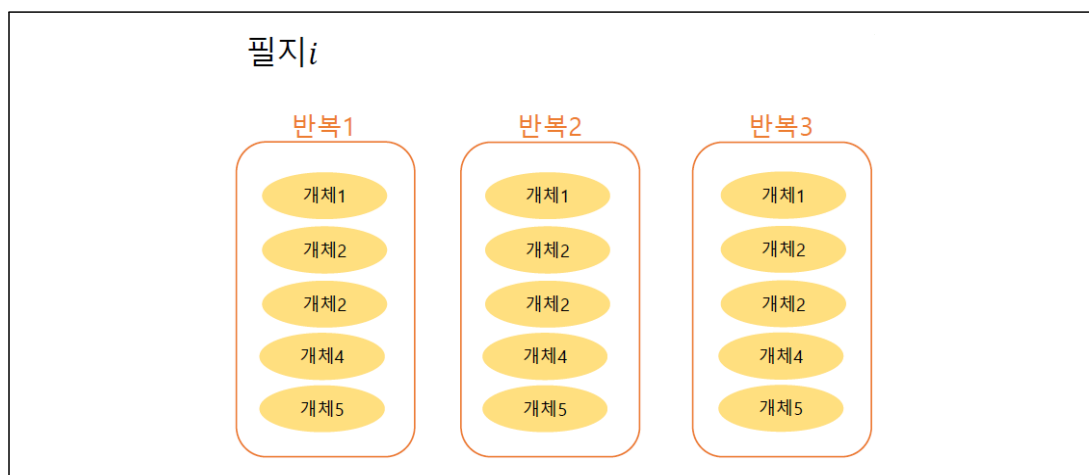
표 1. 실측 조사 차수 일자

조사 차수	조사 일자
1, 2, 3차수	2020-02-13 ~ 2020-03-23
4, 5, 6차수	2020-03-30 ~ 2020-04-28
7, 8, 9차수	2020-04-27 ~ 2020-05-26
10, 11차수	2020-06-01 ~ 2020-06-10

자료: 한국농촌경제연구원 생육 실측조사

〈그림 3〉과 같이 실측조사 대상 농가의 대표 필지 1곳을 3개 조사구역(반복)으로 구분한 후 각 조사구역에서 5개체씩 지상부와 지하부의 생육자료를 수집한다. 지상부의 생육요인은 고사엽장(cm), 초장(cm), 엽수(매), 엽초장(cm), 엽초경(mm)이며, 지하부의 생육요인은 구직경(mm), 생구무게(g)로 구분된다.

그림 3. 실측조사 대상 구역



자료: 저자 작성

본 연구는 2020년 양파 생육 실측조사 자료 중 병해발생, 측정오류 등에 따라 최종 생구 무게가 제시되지 않은 조사구역을 제외하고 모든 조사차수에서 조사가 완료된 필지의 조사구역을 한정하여 신뢰도를 높이고자 하였다. 국내 양파 생산량의 대다수를 차지하는 중만생종 양파를 분석대상으로 삼았으며, 중만생종의 총 100개 필지 중 60개 필지에서 740개 개체를 대

상으로 각 조사차수의 반복별 지상부 생육자료와 최종 수확기의 생구 무게 자료를 사용하였다.

3.1.2. 기상자료

국립농업과학원에서 소기후 모형을 통해 도출한 양파 농가의 필지 단위별 기상추정자료를 사용하였다. 소기후 모형은 우리나라 전역에 대해 농장필지 단위까지 공간적으로 정밀한 농업기상 및 기후 정보를 표현할 수 있는 모형이다. 일별 농업기상 관측자료를 기준으로 최고기온, 최저기온, 평균기온, 강수량, 일조시간, 일사량을 30~270m의 격자 해상도로 상세화하여 실측조사 필지 단위별 일별 기상정보를 추정한다(김수옥 외 2019).

3.2. 변수 및 분석모형

통계청에서 매년 발표하는 양파 단수는 생중량과 건조율을 이용해 식 (1)과 같이 추정하고 있다. ‘20개당 생중량(g)’은 표본구역별 3㎡이랑길이 내의 정상포기 중에서 20개의 조사용 재료를 계통 추출하여 수확하고, 흙을 제거한 후 양파줄기와 뿌리를 1cm씩 남기고 절단한 후 20개의 생중량을 측정한 것이다.

$$(1) \text{ 3㎡당 생중량}(g) = \text{3㎡당 정상포기수} \times \text{20개당 생중량}(g) \div 20$$

$$10a \text{당 생산량}(kg) = \text{3㎡당 생중량}(g) \div 3 \times \text{건조율}$$

한국농촌경제연구원 실측조사 농가필지정보 원시자료는 필지 단위별 3.3㎡당 주수 및 면적을 제공하고 있다. 또한 최종 생구무게를 생중량으로 이용할 수 있기 때문에 식 (2)와 같이 단수 추정이 가능하다고 판단하였다. 따라서 본 연구의 목적인 양파 생육자료를 이용하여 단수를 추정하기 위해 최종 생구 무게를 종속변수로 설정하였다.

$$(2) 3.3\text{m}^2 \text{당 생중량}(g) = 3.3\text{m}^2 \text{당 주수율} \times 1\text{개당 생중량}(g)$$

$$10a \text{당 생산량}(kg) = 3.3\text{m}^2 \text{당 생중량}(g) \div 3.3 \times \text{건조율}$$

생구무게와 관련된 실측기반 생육변수를 설정하기 위해 상관관계분석을 이용하였다. 상관관계분석 결과는 <표 2>와 같다. 엽수, 엽초경 및 초장은 모든 조사 차수에 생구무게와 양(+)의 상관, 엽초장은 3차를 제외한 모든 조사차수에 생구무게와 양(+)의 상관관계를 가지고 있는 것으로 나타났다. 고사엽장의 경우, 4차 조사에서 생구무게와 유의미한 음(-)의 상관관계를 나타냈다.

표 2. 양파 생구무게와 양파 생육자료의 상관관계

조사차수	엽수	엽초경	엽초장	초장	고사장더미
1차	0.2829*	0.2692*	0.1595*	0.3150*	0.0897
2차	0.3207*	0.2347*	0.1385*	0.2600*	0.0919
3차	0.2765*	0.3229*	0.1187	0.2472*	-0.0531
4차	0.3222*	0.4307*	0.1765*	0.2854*	-0.2386*
5차	0.4150*	0.4457*	0.2184*	0.2955*	-0.1140
6차	0.4893*	0.4990*	0.3019*	0.3562*	-0.0814
7차	0.4651*	0.5142*	0.3197*	0.3690*	-0.0355
8차	0.4998*	0.3927*	0.2974*	0.3378*	-0.0342
9차	0.4404*	0.4048*	0.2812*	0.4163*	-0.1197
10차	0.3754*	0.3970*	0.2914*	0.3812*	0.1012
11차	0.4256*	0.3565*	0.5089*	0.5459*	0.0547

주: 1) *p<0.1

2) 양파 생육주기에 의하면 고사엽장(cm)보다 상처가 없는 것이 더 중요하므로 고사엽장 변수는 고사엽장 유무를 나타내는 더미변수로 설정하였음.

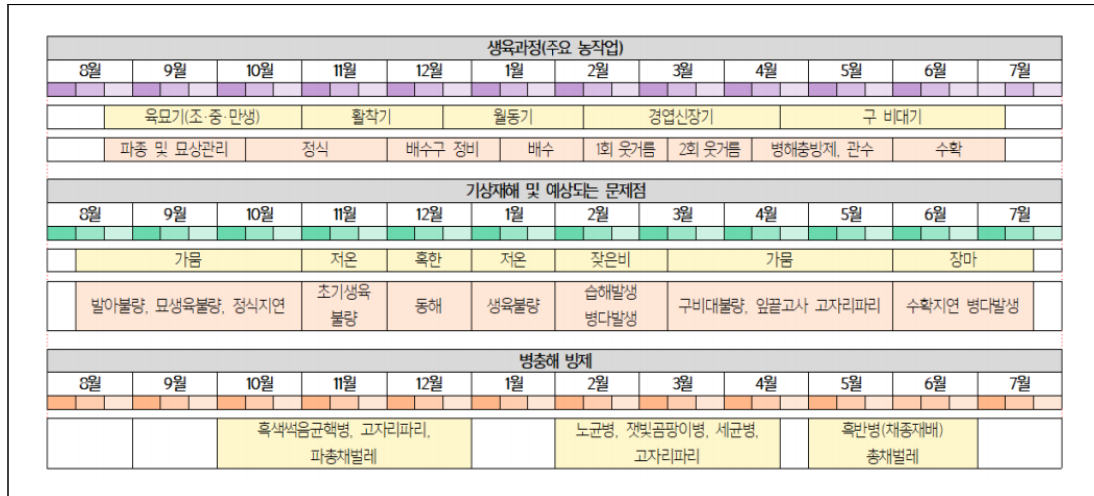
자료: 저자 작성

상관관계를 바탕으로 잎이 신장하는 시기와 구가 비대하는 시기에 해당하는 4차부터 8차까지의 조사기간을 대상으로 양파 생구무게에 영향을 미치는 생육자료를 설명변수로 선택하였으며, 각 표본의 지역을 더미변수로 설정하였다.

기상요인은 선행연구를 포함하여 농촌진흥청에서 제공하는 양파의 생육과정별 기상조건 자료를 참고하여 기상요인 설명변수를 선정하였다(그림 4). 또한 동일한 필지에 있는 양파 개

체들은 기상요인을 공유하므로 이들의 최종 생구 무게는 서로 영향을 미치지만, 다른 필지에 있는 양파와는 상관관계를 갖지 않는다. 따라서 필지별 기상요인을 도출하였다.

그림 4. 양파의 생육과정



자료: 농촌진흥청 농사로(www.nongsaro.go.kr)

신뢰도를 높이기 위해 각 기상변수는 육묘기인 9월 1일부터 최종 생구 조사일인 6월 10일 이전까지 15일 간격으로 나누었으며, 필지 단위로 평균값, 누적값, 최댓값, 최솟값을 추정하였다. 추정된 기상변수는 평균기온, 최고기온, 최저기온, 강수량, 일조시간, 일사량 등 기본정보 외에도 GDD, 최적기온일수, 고온일수, 강우일수, 탄소동화저해일수 등을 추가하였다.

표 3. 기상변수 정의

설명변수	정의	설명변수	정의
평균기온	일평균기온의 평균	GDD (생육도일)	평균기온이 4℃ 이상인 날들의 값의 합
최고기온	최고기온의 최댓값	최적기온일수	최고기온이 15℃~25℃인 날들의 수
최저기온	최저기온의 최솟값	고온일수	최고기온이 30℃ 이상의 날들의 수
누적강수량	일강수량의 합	강우일수	일강수량이 5mm 이상인 날들의 수
누적일조시간	일일조시간의 합	탄소동화저해일수	최고기온이 25℃ 이상인 날들의 수
누적일사량	일일사량의 합		

자료: 저자 작성

분석자료는 1개년도 횡단면의 자료인 것을 감안하여 생구무게와 여러 생육 및 기상변수를 이용하여 다중회귀분석을 통해 생구 예측치를 도출한 후 단수를 추정하고자 한다.

$$(3) y = f(\text{생육기반자료}, \text{기상요인}, \text{기타요인})$$

통상최소자승법(OLS)에 의한 일반적인 다중회귀분석 모형은 식 (4)로 표현할 수 있다.

$$(4) y = X\beta + \epsilon$$

여기서 y 는 생구무게인 종속변수로 $N \times 1$ 벡터, X 는 생구무게에 영향을 미치는 설명변수로 $N \times k$ 행렬이며, ϵ 는 오차항으로 $N \times 1$ 벡터이다.

4. 추정 결과

4.1. 실측기반 생육자료 단수예측 모형

실측기반 생육자료 단수예측 이전에 분산 팽창 요인(Variance Inflation Factor, VIF)을 이용하여 우선 변수 간 다중공선성을 진단하였다. 생육 설명변수의 평균 VIF값은 2.77이고, 각 설명변수의 VIF값은 모두 10 미만으로 다중공선성이 발생하지 않은 것으로 나타났다.

〈표 4〉는 양파의 생육자료와 지역변수를 이용한 회귀분석의 결과이다. 분석 결과, 모든 설명변수의 p값은 유의한 것으로 나타났다. 중만생종 양파 재배에서 잎이 신장하기 시작하는 때는 3월 중순 이후로 4차 및 5차 조사시기에 해당한다. 4차 초장 및 엽초장과 5차 엽수는 각 1.75, 11.10, 12.56으로 양의 계수를 가지고 있으며, 잎의 신장이 높을수록 최종 구의 무게가 큰 것으로 분석된다. 이 결과는 생육주기에서 언급한 잎이 양파 구의 크기에 비례한다는 내용과 일치하였다.

표 4. 실측기반 생육자료를 이용한 회귀분석 결과

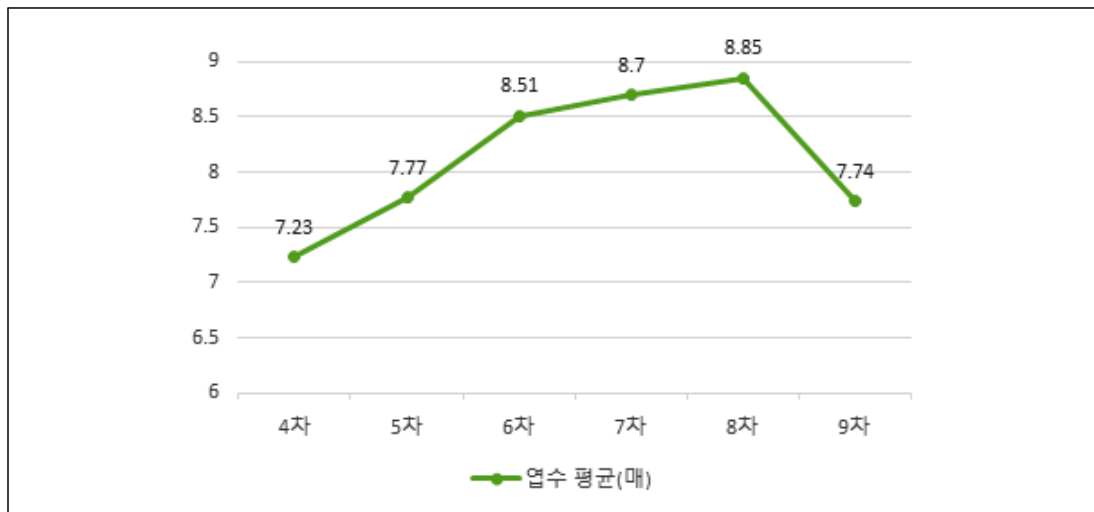
변수	β 계수	표준오차	t-value	p-value
4차 초장	1.751	0.856	2.04	0.043
4차 엽초장	11.101	4.027	2.76	0.007
4차 고사장 더미	-52.018	17.007	-3.06	0.003
5차 엽수	12.563	6.761	1.86	0.065
8차 엽초장	-4.441	1.620	-2.74	0.007
8차 엽수	36.257	4.870	7.45	0.000
지역변수(경북)	78.150	22.462	3.48	0.001
지역변수(전남)	62.645	15.960	3.93	0.000
지역변수(전북)	74.904	17.790	4.21	0.000
지역변수(충남)	149.993	20.611	7.28	0.000
_cons	-242.496	46.084	-5.26	0.000

F=19.92, p-value = 0.0000 Adj R-squared = 0.5223

자료: 저자 작성

4차 고사장 더미변수는 -52.02로 음의 계수를 가진다. 생육 주기에 따르면 잎은 양파 구의 형성과 비대에 필수적이다. 따라서 잎에 난 상처는 구의 비대에 나쁜 영향을 미치는 것으로 해석할 수 있다.

그림 5. 조사차수별 양파 엽수 변화



자료: 저자 작성

8차 조사 시기는 5월 중순으로 지하부의 생장이 빠르게 증가하는 구 비대기에 해당한다. 8차 엽초장은 -4.44로 음의 계수를 가졌고, 엽초장이 짧을수록 구의 무게가 크다고 해석할 수 있다. 8차 엽수는 36.26으로 양의 계수를 가지며 잎의 수가 많을수록 구의 무게가 크다고 분석된다.

생물학적 생육주기에 의하면 비늘잎이 형성되기 시작하여 구가 비대하는 시기에 잎의 수가 최대치를 기록한다. 이 때 잎의 개수는 양파 구의 크기를 결정하는 가장 중요한 요소이다. 실제로 조사차수별 엽수 통계량을 살펴보면, 엽수가 지속적으로 증가하는 추세를 보였으며 8차 조사 시기에 평균 엽수가 8.85개로 가장 높은 것으로 나타났다. 따라서 8차 시기의 잎의 수가 구의 크기를 결정하는 중요한 변수인 것으로 판단된다(그림 5).

4.2. 기상자료 단수예측 결과

기상변수는 10월 초 GDD, 11월 말 누적강수량, 12월 말 누적강수량, 2월 초 누적강수량, 6월 초 누적강수량, 2월 초 누적일사량으로, 우선 변수들 간에 다중공선성을 확인하기 위해 VIF값을 도출하였고, 10 이하로 나타나 다중공선성은 존재하지 않는 것으로 판단된다. 종속 변수인 생구무게와 설명변수인 기상요인들을 이용하여 분석한 결과를 <표 5>에 제시하였으며, 모든 변수는 통계적으로 유의한 값을 나타냈다.

표 5. 기상자료를 이용한 회귀분석 결과

변수	β 계수	표준오차	t-value	p-value
10월(1) GDD	1.799	0.795	2.26	0.028
11월(2) 누적강수량	5.218	1.180	4.42	0.000
12월(2) 누적강수량	-18.826	5.341	-3.52	0.001
2월(1) 누적강수량	-19.595	4.996	-3.92	0.000
2월(1) 누적일사량	1.304	0.615	2.12	0.039
6월(1) 누적강수량	-9.295	2.192	-4.24	0.000
상수항	144.780	208.597	0.69	0.491

F: 4.97, p-value: 0.0004, Adj R-squared: 0.2877

주: 기상변수의 (1)은 1일~15일을 의미하며, (2)는 16일~31일을 의미함.

자료: 저자 작성

10월 초 GDD는 양의 관계로 10월 1일부터 10월 15일까지 GDD가 높을수록 생구 무게는 증가하는 것으로 나타났다. 이는 양파가 발아할 때 4℃ 이상인 날이 많을수록 생구 무게가 비대에 지는 것을 의미한다. 4℃보다 낮은 온도에서 발아 시 시간이 오래 걸리고 발아율이 떨어진다. 11월 말 누적강수량은 11월 16일부터 30일까지 누적강수량이 많을수록 양파 최종 생구 무게가 큰 것으로 나타났다. 양파 정식 직후 토양의 수분이 많으면 뿌리 발아에 걸리는 시간이 단축되어 구 비대에 영향을 미치는 것으로 판단된다. 12월 말 누적강수량이 많을수록 생구 무게는 더 적은 것으로 나타났다. 양파 정식 직후 강수량은 양파 묘를 동해로부터 방지해 주지만 이후에도 강수가 많으면 오히려 토양의 수분과다로 인해 동해를 입을 수 있는 것으로 해석된다.

2월 초 누적강수량, 누적일사량의 회귀 계수는 각각 -19.595와 1.304으로 2월 1일부터 15일까지의 강수량이 많을수록 생구 무게는 적어지고, 일사량이 많을수록 생구무게는 커지는 것으로 나타났다. 2월에는 잦은 비로 인해 습해가 발생하기 쉬우므로 강수량이 적은 것이 좋고 일사량이 많으면 땅이 습해지는 것을 방지하여 생구의 비대에 도움을 주는 것으로 판단된다.

6월 초 누적강수량은 음의 관계로 6월 1일부터 최종 수확 날짜인 6월 10일까지의 강수량이 많을수록 생구무게는 작아지는 것으로 나타났다. 6월 초는 구 비대기로 구가 비대하는 데 많은 에너지를 소모하기 때문에 뿌리의 활력이 감소하며, 부패 및 병해충을 막기 위해서는 토양의 수분이 적은 상태여야 한다. 또한 강수로 인해 수확기가 늦춰진다면 결주율이 높아져 양파 최종생구에 부정적인 영향을 주는 것으로 짐작할 수 있다.

4.3. 혼합(생육 및 기상)자료 단수예측 결과

양파의 실측기반 생육자료 및 기상자료 모형을 모두 고려하여 혼합 단수예측모형을 추정하였다. 생육자료에서 선정된 변수들은 7차 초장, 8차 엽수, 8차 엽초장, 4차 고사장 더미변수이며, 기상자료에서 선정된 변수들은 12월 초 누적강수량, 2월 초 누적강수량, 4월 초 누적강수량, 5월 말 누적강수량, 6월 초 누적강수량, 10월 초 최고기온, 11월 말 최고기온, 12월 말

최고기온이다. 분석하기 이전에 변수들 간 VIF를 확인한 결과 수치가 10 이하로 나타나 다중 공선성은 존재하지 않는 것으로 판단된다. 실측기반 생육변수 및 기상변수를 혼합한 다중회귀모형 분석결과는 <표 6>에 제시되어 있다.

생육변수인 7차 초장, 8차 엽수, 8차 엽초장, 4차 고사장 더미변수는 통계적으로 유의하였다. 7차 초장이 길수록 8차 엽수가 많을수록 생구의 크기는 커지며, 8차 엽초장이 짧을수록 4차 고사엽장이 없을수록 생구의 크기는 증가한다. 양파 생구의 크기를 결정하는 가장 중요한 요소는 엽수이다. 구비대기에 병해충에 의한 피해나 물리적인 상처를 입어서 잎몸이 상하게 되면, 잎몸에 연결된 잎집 기부에는 엽록소의 양이 적기 때문에 잎집만으로는 생육을 계속할 수 없다. 따라서 상처 없는 잎의 수가 증가하는 만큼 구의 비대도 커진다.

표 6. 실측기반 생육자료 및 기상자료를 이용한 회귀분석 결과

변수	β 계수	표준오차	t-value	p-value
7차 초장	2.638	0.587	4.49	0.000
8차 엽수	29.996	4.197	7.15	0.000
8차 엽초장	-4.806	1.839	-2.61	0.010
4차 고사장 더미	-62.225	14.313	-4.35	0.000
12월(1) 누적강수량	21.493	4.674	-4.60	0.000
2월(1) 누적강수량	-15.996	3.234	-4.95	0.000
4월(1) 누적강수량	14.277	4.205	3.39	0.001
5월(2) 누적강수량	-6.881	1.614	-4.26	0.000
6월(1) 누적강수량	-2.374	1.080	-2.20	0.029
10월(1) 최고기온	107.951	16.197	6.66	0.000
11월(2) 최고기온	-76.531	13.016	-5.88	0.000
12월(2) 최고기온	30.989	10.314	3.00	0.003
상수항	-1253.892	412.181	-3.04	0.003

F: 19.94, p-value: 0.0000, Adj R-squared: 0.5410

주: 기상변수의 (1)은 1일~15일을 의미하며, (2)는 16일~31일을 의미함.

자료: 저자 작성

기상변수는 12월 초, 2월 초, 4월 초, 5월 말, 6월 초 누적강수량과 10월 초, 11월 말, 12월 말 최고기온이 통계적으로 모두 유의한 것으로 나타났다. 12월 초와 4월 초의 누적강수량은 양의 효과가 나타났으므로 누적강수량이 많을수록 생구무게는 증가한다. 12월 초는 양파의

생육과정 중 활착기에 해당하며, 이 시기의 양파는 물이 꼭 필요하기 때문에 물을 아무리 많이 주어도 별문제가 없다. 4월 초는 구 비대기 직전으로 수분이 많을수록 생육량이 많다.

2월 초, 5월 말, 6월초의 누적강수량이 적을수록 생구가 비대해지는 것으로 나타났다. 2월 초는 월동기에서 경엽신장기로 넘어가는 시기로 잦은 비는 습해발생과 병다발생이 일어날 수 있다. 이 시기의 양파는 건조한 토양 조건에 오히려 강하다(이종태 2015). 따라서 토양 수분이 적더라도 생육에 큰 지장을 받지 않는다. 또한 5월 말과 6월 초는 구 비대기이므로 구가 비대하는 데 많은 에너지를 소모하여 뿌리와 잎의 활력이 감소한다. 그렇기 때문에 토양의 수분이 많으면 양파의 뿌리는 습해를 받기 쉬워진다.

10월 초와 12월 말은 육묘기와 월동기로 최고기온이 높을수록 생구무게가 증가하며, 11월 초는 활착기로 최고기온이 낮을수록 생구무게가 증가한다. 육묘기에서 씨앗이 발아하기 위한 최고온도는 33℃로 기온이 높아야하며, 월동기에는 혹한으로 인한 동해 방지를 위해 기온이 역시 높아야 함을 미루어 짐작할 수 있다. 반면에 활착기에는 고온보다 저온에서 더 오랫동안 성장할 수 있다.

4.4. 예측력 검정 비교

본 연구의 분석자료는 1개년도 횡단면자료만으로 활용되었기에 정확한 추정치를 얻기보다는 예측력 있는 모형을 식별하는 것이 적합하다고 판단된다. 따라서 각 모형의 생구 예측치를 도출하고 이들의 예측력을 비교하여 경제성 분석을 하기 위한 최종모형을 선정하였다. 예측력을 평가하는 기준은 MAD(mean absolute deviation), MSE(mean square error), RMSE(root mean square error), MAPE(mean absolute percent error) 등이며 계측한 값이 낮을수록 모형의 예측력이 높다. 각 평가 기준의 수식은 식 (5)와 같다. 여기서 A_i 는 생구 무게 실측치, F_i 는 생구무게 예측치, i 는 표본수를 의미한다.

$$(5) \text{ MAD} = \frac{\sum_{i=1}^n |A_i - F_i|}{n}$$

$$\text{MSE} = \frac{\sum_{i=1}^n (A_i - F_i)^2}{n}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (A_i - F_i)^2}{n}}$$

$$\text{MAPE} = \frac{\sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right|}{n} \times 100$$

위의 식을 이용하여 모형별 예측력 검정 결과를 <표 7>에 제시하였다. 혼합모형의 예측 통계값은 MAD 42.41, MSE 3037.06, RMSE 55.11, MAPE 16.61로 나타났으며, 실측기반 모형 및 기상자료 모형과 비교하여 가장 낮게 추계되었다. 이는 혼합모형의 예측치가 다른 모형의 예측치보다 정교한 것을 보여주고 있다.

표 7. 모형별 예측력 비교

구분	MAD	MSE	RMSE	MAPE
실측기반 모형	46.31	3398.33	58.30	18.10
기상 모형	48.94	3550.30	59.58	17.52
혼합 모형	42.41	3037.06	55.11	16.61

자료: 저자 작성

또한 4가지 예측 기준 중 MAD 값을 이용하여 오차율을 구하면 실측기반 모형 15.3%, 기상기반 모형 16.9%, 혼합모형 13.9%로,²⁾ 혼합모형의 예측치가 가장 작은 오차율을 나타냈

2) MAD는 MSE에 비교하여 큰 오차에 대해 덜 민감하게 반응하며, MAPE는 실측치가 낮은 경우 오차가 작더라도 오차율이 상대적으로 크게 계산되는 문제점이 있어 해석에 주의할 필요가 있다. 따라서 일반적으로 가장 많이 쓰이는 MAD를 이용하였다.

다.3) 본 연구는 세 가지 모형 중 가장 우수한 예측력과 오차율을 보인 혼합모형을 최종 모형으로 선택하고 이를 바탕으로 사회후생효과를 계측하여 경제성을 분석하였다.

5. 사회후생효과 분석

농업관측의 경제적 효과는 관측정보 제공에 의해 자원배분이 효율적으로 이루어짐에 따른 사회적 후생의 증감으로 나타낼 수 있으며, 정보부족으로 인한 수급불균형을 줄이거나 제거됨으로써 감소되는 사회적 손실분이 곧 사회적 후생의 증가분에 해당된다(이용선 외 2002). 즉, 최대 사회적 후생 증대효과는 관측정보 제공에 의해 과잉생산 혹은 과소생산이 모두 해소되어 균형생산이 달성되었을 때를 의미한다.

따라서 혼합모형을 이용하여 양파 관측정보 제공에 의해 가장 효율적으로 자원이 배분되었을 때의 경제성 분석을 위하여 사회후생효과를 계측해보았다. 사회적 후생효과를 추정하는 방식은 거미집형 기대가설 하에서 기존의 선행연구에서 추정한 수요/공급 탄성치를 이용하였으며, 가격 변동분이 아주 적을 때 수요와 공급의 가격 변동률은 같다고 가정하였다(이용선 외 2002; 김관수 외 2009).

우선 혼합모형을 이용하여 추정한 예측단수는 8,669kg/10a로 나타났다. 예측 단수와 실제 양파재배면적 및 가격을 이용하여 산출된 총 생산액은 13,128억 원이다. 이를 바탕으로 추정한 최대 사회후생효과를 <표 8>에 제시하였다.

3) 한국농촌경제연구원 농업관측본부에서 제공한 실측조사 생육정보는 1개년도 자료로서, 자료의 제한으로 인한 오차는 불가피하다. 자료가 축적되면 모형의 예측력이 개선될 것이라고 여겨진다.

표 8. 혼합모형을 활용한 예측치의 최대 사회후생효과 추정

재배 면적 (ha)	예측 단수 (kg/10a)	생산량 (만 톤)	가격 (원/kg)	생산액 (억 원)	수요의 가격 탄성치	공급의 가격 탄성치	가격 변동률	사회후생효과 (백만 원)
14,673	8,669	127.2	1,032	13,128	0.45	0.36	0.44	25,416

주: 1) 가격은 2020년 서울특별시 농수산물공사의 양파 도매가격의 평균값임.

2) 가격변동률은 당해를 포함한 지난 3년간의 연간 가격 변동률의 평균값임. 가격변동률은 서울특별시 농수산물공사의 양파 도매가격을 기준으로 하였음.

3) 사회후생효과는 $\{(1/2) * (\text{가격변동률})^2 * (1/((1/\text{수요탄성치}) + (1/\text{공급탄성치}))) * \text{생산량}\}$ 으로 계측하였음.

4) 수요의 가격탄성치는 이용선 외, 「농업관측의 경제적 효과분석」, 2002에 제시된 수요의 가격탄성치를 사용하였음.

5) 공급의 가격탄성치는 이용선 외, 「농업관측의 경제적 효과분석」, 2002에 제시된 재배면적의 가격탄성치를 사용하였음.

자료: 저자 작성

혼합모형을 활용한 예측치의 최대 사회후생효과는 25,416백만 원으로 나타났다. 이 중 소비자 사회후생효과는 11,296백만 원, 생산자 사회후생효과는 14,120백만 원으로 생산자 후생효과가 소비자 후생효과보다 큰 것으로 나타났다. 또한 생육자료와 기상자료를 이용하여 예측한 단수의 사회후생효과 기여도는 총 생산액 대비 약 1.9% 수준으로 추정된다.

이상에서 계산된 것은 혼합모형의 단수 예측력에 따른 경제성을 확인하기 위함으로 혼합모형을 활용한 예측치에 의해 최적 균형이 달성될 때의 사회적 후생 최대치를 의미한다. 하지만 생산자의 관측정보에 대한 활용도 유무 및 예측치의 부정확성에 따라 최적 균형에 도달하지 않는다면 사회적 후생 증가분은 최대치보다 적을 수 있다. 그러나 생육자료와 기상자료를 이용한 단수 예측치가 생산자 후생 증대뿐만 아니라 소비자 후생 증대를 발생시켜 모형의 경제적 효과를 검증하였다는 것에 의의가 있다.

6. 요약 및 결론

농산물 수급 안정화를 위해서는 정확한 생산량 예측을 통한 선제적인 대처가 중요하다. 양파는 과거부터 꾸준히 가격 변동성이 높은 품목으로 이를 완화하기 위한 수요와 공급 관측의 중요성이 더욱 커지고 있다. 정부는 정부수매, 산지폐기, 농산물 수급조절 매뉴얼 운영 등을 통해 안정화를 도모하고 있다. 대표적인 수급 안정화 방안 중 하나로 한국농촌경제연구원의

농업관측사업은 기존 전화조사에서 직접 산지를 방문하는 실측조사 체계를 도입하여 면적, 작황 이외에도 생육정보를 활용하여 관측신뢰도를 높이고 있다.

본 연구는 양파를 대상으로 실측 기반 생육자료와 기상자료를 이용하여 단수를 추정하였다. 양파 단수는 생중량, 건조율 및 주수를 이용하여 산출된다. 단수 산출에 필요한 항목은 한국농촌경제연구원의 농가필지정보에서 제공된다. 따라서 생육자료의 최종 생구무게를 모형의 종속변수로 설정하였고 추정된 생구무게를 바탕으로 예측 단수를 도출하였다. 양파 생육 주기의 이론적 검토 및 상관관계 분석 결과를 바탕으로 양파의 최종 구중에 영향을 미치는 지상부 생육요인과 기상요인을 설명변수로 설정하였다.

생육자료만을 이용한 단수, 기상자료만을 이용한 단수, 생육 및 기상자료를 모두 이용한 단수를 각각 추정하였다. 세 가지 모형의 예측력을 비교 검토한 결과 생육 및 기상자료를 활용한 혼합모형의 예측력이 가장 높았다. 이는 실측조사 기반 생육자료의 활용이 단수 예측 정확도를 향상시킬 수 있음을 보여주고 있다. 예측력이 가장 높은 혼합모형에서는 생육요인인 7차 초장, 8차 엽수, 4차 및 8차 엽초장과 기상요인인 12월 초, 2월 초, 4월 초, 5월 말, 6월 초 누적 강수량과 10월 초, 11월 말, 12월 말 최고 기온이 생구무게에 영향을 미치는 설명변수로 활용되었다.

예측력이 가장 높은 혼합모형의 사회적 후생효과를 도출하고 경제성을 분석하였다. 생육자료 및 기상자료를 이용한 단수 예측치의 최대 사회후생효과는 약 254억 원으로, 그 중 소비자 사회후생효과는 약 113억 원, 생산자 사회후생효과는 약 141억 원으로 나타났다. 추정된 최대 사회후생효과에 대한 2020년 기준 양파의 총 생산액 대비 기여도는 1.9%로 추정된다. 이와 같은 추정결과는 최적 균형이 달성될 경우의 최대 사회적 후생효과로, 예측치의 정확도 등으로 인해 실제 사회적 후생 증가분은 최대치보다 적을 수 있다. 그러나 생육자료 및 기상자료를 이용한 예측치의 사회후생효과 추정으로 혼합모형이 가지는 경제적 효과를 검증했다는 것에 의의가 있다.

분석결과에 근거하여 관측의 정확도를 높이기 위한 방안에 대한 시사점을 다음과 같이 제시하고자 한다.

생육자료만을 이용한 모형과 혼합모형 추정 결과 모두 8차 엽수가 구 비대에 중요한 요인으로 분석되었다. 8차 엽수는 구가 비대하는 시기에 가장 많은 엽수를 가지고 있는 것으로 나타

났으며, 이는 구 비대를 위해 충분한 엽록소의 양이 필요하기 때문으로 여겨진다. 이와 같이 엽수를 포함한 실측기반 생육정보가 생산량 예측에 중요한 역할을 하고 있다. 따라서 기존의 실측 조사 항목뿐만 아니라 잎의 무게 등 지상부 실측 조사 항목을 추가하여 구중에 영향을 미치는 요인을 탐색할 필요가 있다.

기상자료만을 이용한 모형과 혼합모형을 추정한 결과에 근거하여 2월 초와 6월 초의 누적 강수량 모두 통계적으로 유의하였다. 2월 초에는 양파가 건조한 토양 조건에 강하기 때문에 토양 수분이 적더라도 생육에 큰 지장을 받지 않는다. 반면에 6월 초에는 구가 신장하는 데 많은 에너지를 소모하기 때문에 토양의 수분이 적을수록 구가 비대해진다. 이러한 결과는 생산자에게 수확기 전 작황 관리에 중요한 정보가 될 수 있으며, 시장 참여자의 합리적인 의사결정에 도움을 줄 수 있다.

마지막으로 본 연구의 결과와 같이 실측기반 생육자료와 기상자료는 정교한 관측을 위한 중요한 수단이다. 그러나 본 연구에서는 실측조사체계가 도입된 1개년도 자료만이 분석에 활용되어 단수 추정 방법에 대한 신뢰도가 다소 미흡한 부분이 있다. 예측력과 신뢰도를 높이기 위해서는 향후 축적된 생육자료를 활용하여 예측 모형을 개선할 필요가 있다. 또한 본 연구에서는 토양, 병해충 등 양파 생육과 관련된 환경적 요인은 고려하지 않았음을 한계로 두고 있다. 실측조사체계를 통해 양파 생육정보 뿐만 아니라 생육에 영향을 미치는 환경적 요인에 대한 정보를 구축한다면 보다 풍부한 자료를 활용하여 양파 생산량 예측도를 높일 수 있는 다양한 시도가 가능할 것으로 기대된다.

참고문헌

- 김관수, 안동환, 성재훈. 2009. “농업 관측정보의 사회적 후생효과 추정.” 『농촌경제』 32(5): 1-16.
- 김배성. 2014. “과일류 관측의 최대 사회후생 효과 계측.” 『한국산학기술학회논문지』 15.11: 6646- 6651.
- 김수옥, 김진희, 김대준, 윤은정. 2019. “농장맞춤형 농업기상정보 생산을 위한 소기후 모형 구축과 활용.” 『한국농림기상학회 학술발표논문집』 315-315.
- 김재휘, 최성천, 김준기, 서홍석. 2021. “양파의 생육시기별 생육요인과 기상요인 간의 관계 탐색.” 『한국 농림기상학회지』 23(1): 1-14.
- 김종규. 2009. “Bayesian 분석모형을 활용한 정보의 가치분석.” 전남대학교 경제학석사학위논문.
- 김준기, 최성천, 김재휘, 서홍석. 2020. “다층모형을 활용한 양파 구중 추정 연구.” 『응용통계연구』 33(6): 763-776.
- 나상일, 박찬원, 소규호, 박재문, 이경도. 2017. “MODIS NDVI와 기상요인을 고려한 마늘·양파 주산 단지 단수예측 모형 개발.” 『Korean Journal of Remote Sensing』 33호 5-2권: 647-659.
- 나우현. 2000. 『양파·파·양파』. 오성출판사.
- 남국현, 최영찬. 2015. “양파 출하시기 도매가격 예측모형 연구.” 『농촌지도와 개발』 22(4): 423-434.
- 유도일. 2005. “농업관측정보의 경제적 효과.” 서울대학교 경제학석사학위논문.
- 이용선, 김연중, 김윤희. 2002. “농업관측의 경제적 효과 분석.” 『농촌경제』 25(3): 1-15.
- 이종태. 2015. 『양파: 한살이부터 유기농·무경운 재배법까지』. 들녘.
- 이춘수, 양성범. 2017. “생육정보를 이용한 가을배추와 가을무 단수 예측 모형 개발.” 『한국유기농업 학회지』 25(2): 279-293.
- 임철희, 김강선, 이은정, 허성봉, 김태연, 김용석, 이우균. 2016. “주산지 기상정보를 활용한 주요 채소 작물의 단수 예측 모형 개발.” 『한국기후변화학회지』 7(2): 193-203.
- 통계청. 2020. 『농작물 생산조사 지침서』.
- 한석호, 이병훈, 박미성, 승준호, 양현석, 신성철, 이동필. 2011. 『기상요인을 고려한 단수예측모형 개발 연구』. 한국농촌경제연구원.
- Chmielewski, F.-M and Köhn, W. 2000. “Impact of weather on yield components of winter rye over 30 years.”, *Agricultural and Forest Meteorology* 102(4): 253-261.
- LANCASTER, J.E., TRIGGS, C.M., DE RUITER, J.M. and GANDAR, P.W. 1996. “Bulbing in Onions: Photoperiod and Temperature Requirements and Prediction of Bulb Size and Maturity.”, *Annals of Botany* 78(4): 423-430.
- 농촌진흥청 농사로. <www.nongsaro.go.kr>.
- 통계청. <<http://kosis.kr/>>.
- 한국농수산식품유통공사 농산물 유통정보(kamis). <<https://www.kamis.or.kr/customer/>>.
- 한국농촌경제연구원 농업관측 통계정보시스템. <<https://oasis.krei.re.kr/>>.



제1회

농산물 수급 예측모형 경진대회

우수논문집

Chapter 1
지정과제



장려상

주산지 작황 실측 정보를 이용한 양파가격 예측모형 개발

김건아 · 전기석 · 정소영

주산지 작황 실측 정보를 이용한 양파가격 예측모형 개발

김건아* 전기석** 정소영***

Key words

가격 변동성(price volatility), 양파 생육 정보(Onion growth information), 구조모형(structural model), 시계열모형(time series model), 인공신경망모형(artificial neural network model), 가격 예측(Price forecast)

ABSTRACT

Onion, the most important seasoning vegetable, has a very large price fluctuation, causing damage to farmers and hindering rational decision-making. Accurate forecasting of prices is important for onion supply and demand and price stabilization. In order to improve the forecast for onion supply and demand stabilization, the Korea Rural Economic Research Institute provides information by observing onion growth information. In this study, the onion price is predicted through structural model, time series model, and artificial neural network model to compare between models. In addition, a production forecast model was developed using growth information. As a result of the analysis, the structural model showed the highest forecast power of onion price than other models. However, it is difficult to predict the price of onion due to various external factors, so measures to improve the forecast are needed.

차례

- | | |
|-------------------|------------------|
| 1. 서론 | 4. 양파가격 예측 분석 결과 |
| 2. 양파 시장 현황 및 문제점 | 5. 결론 및 시사점 |
| 3. 예측 분석 방법 및 자료 | |

* 고려대학교 식품자원경제학과 박사수로

** 고려대학교 식품자원경제학과 석사과정

*** 고려대학교 식품자원경제학과 석사과정

1. 서론

양파는 우리나라의 중요한 양념 채소로서 국내 조미 채소 생산액 중 가장 높은 비중을 차지하며, 발작물 농가의 주 소득 작목이다. 양파의 재배면적은 우리나라 조미 채소 재배면적의 약 20%를 차지할 정도로 그 중요도가 높으며, 주산지는 전남과 경남 등 주로 남부지방에 분포되어 있다(통계청, 2020). 또한, 공영도매시장 청과 부류의 전체 거래 물량 중 점유율이 약 10%로 가장 많이 거래되고 있는 품목이며 발작물 농가 소득의 상당 부분을 차지하고 있다(서울시 농수산물공사, 2020). 또한, 물가 안정 5개 품목으로 지정되는 등 매우 중요한 품목이다.

그러나 양파는 노지 재배 특성상 기상 변화에 민감해 생산량 예측이 어렵고, 가격 변동이 심해 배추, 무 등과 함께 국내 채소류 5개 물가 민감 품목 중 하나로도 지정되어 있다(김성우 외, 2016; 최성천 외, 2016). 양파는 재배지의 기상 변화에 따라 생산량 변화가 심하고, 주산지가 전국으로 이동하기 때문에 전체적인 생산량 추정이 어렵다(나상일 외, 2017). 이러한 이유로 매년 양파 등 주요 채소류에 대한 수급 실패 사례는 반복되고 있으며, 가격 및 재배면적 예측 실패와 정부의 소극적 정책에 대한 비판 등으로 언론에서 주목을 받는 실정이다(한국농정, 2019. 7. 7).

양파가격의 급등락은 농민들의 혼란을 가중하고, 잘못된 의사결정을 유발하여 양파 농가의 피해를 더욱 악화시킨다. 국내 조미 채소의 중요한 부분을 차지하는 양파의 높은 가격 변동성은 국내 양파 농가의 큰 피해를 유발하기 때문에 정부는 총력을 다해 가격 안정을 위한 노력을 하고 있다(중앙일보, 2020. 8. 10). 특히, 정부는 양파의 수급 불안정성을 낮추기 위해 예측 정확성의 필요성을 강조하고 있다. 가격 예측은 수급 조절의 핵심인 생산량 및 재배면적에 영향을 미치므로 잘못된 예측은 잘못된 방향을 제시하여 가격 불안정성을 악화시킬 수 있다(하지희 외, 2019). 따라서 정확한 가격 예측은 합리적인 의사결정에 도움을 주어 양파 수급 안정과 가격 변동 완화에 기여할 수 있다(김동환, 2018).

정부는 정확한 예측을 위한 다양한 정보를 제공하기 위해 노력하고 있으며, 특히 농업 관측을 통해 재배면적, 작황, 생산량 및 생육 정보를 제공함으로써 농업인의 합리적인 의사결정을 지원하고 관측 정보를 바탕으로 선제적 수급 안정 대책을 추진하고 있다. 최근에는 정확한 실

측 정보를 확보하기 위해 조사 방법을 실측 조사로 전환하고, 생육 진행 상황과 생산량 실측 정보를 수집하기 시작하였다(농림부, 2020).

또한, 양파 예측의 정확도 향상을 위해 선행연구에서는 다양한 예측모형을 이용하여 관측 능력을 높이려 시도하고 있다. 양파가격 예측에 사용된 대표적인 구조모형은 양파의 재배 면적과 단수를 예측하고, 도매가격을 전망하는 한국농촌경제연구원의 KASMO(Korean Agricultural Simulation Model)이다. KASMO는 양파를 재배업 부문으로 분류하여, 획일적인 방법으로 추정하므로, 월별 가격 변동성이 높은 양파의 개별 특성을 반영하지 못했다는 한계점이 있다.

시계열모형은 양파의 가격 예측에 유용하게 사용될 수 있으며, 김배성(2005), 남국현·최영찬(2015), 하지희 외(2019)의 연구에서 다양한 변수와 모형을 통해 양파가격 예측을 수행하였다. 김배성(2005)은 가락시장 월별 도매가격 자료를 이용해 가격 패턴만을 고려한 ARIMA, GARCH 모형과 인공신경망모형으로 표본 외 예측을 수행하고, RMSPE, MAPE 등을 기준으로 예측력을 평가하였다. 그 결과 양파가격 예측은 인공신경망보다 시계열모형 특히 ARMA(1,2) 모형의 예측력이 우수한 것으로 분석되었다.

그러나 시계열모형이 30% 이상의 오차율을 가져 가격 패턴 이외의 변수를 고려한 모형 개발의 필요성을 언급하였다. 남국현, 최영찬(2015)은 ARDL(자기회귀시차모형)을 이용해 생산량을 예측하고, 2015년 양파 출하 시기 도매가격을 예측하였다. 김배성(2005) 연구와 비교했을 때 생산량을 모형에 반영해 예측력이 향상되었으며, 생산량이 출하 시기 도매가격에 영향을 미침을 시사하였다. 한편 출하 시기 이후 가격 예측 시 재고량과 그에 따른 도매시장 반입량 증감, 수요량을 고려하지 못했다는 한계점을 가진다. 하지희 외(2019)는 양파가격자료의 이분산성을 고려해 양파가격예측모형을 설계하였고, 시장 구조의 변화 가능성을 반영하기 위해 구간이동회귀법(Rolling Window Regression)을 적용해 예측력을 평가하였다. 분석 결과, 3년의 가격 자료를 바탕으로 월별 터미를 고려한 ARMGARCH 모형이 양파가격 예측에 가장 적합한 모형임이 분석되었다.

또한, 남국현 외(2015)는 기상 및 생산자료를 이용하여 자기회귀시차모형(ARDL)에 적용해 지역별 단위 함수를 추정 후 전국 단위 예측치를 산출한 결과, 김배성(2015)의 결과와는 달리 단위와 재배면적으로 고려할 경우, 예측력이 향상됨을 시사하였다. 또한, 하지희 외(2019)

는 이분산성(Heteroskedasticity)을 고려하여, AR-GARCH 모형을 통해 예측하였고, 한국 농촌경제연구원(KREI)의 예측치와 합성할 경우, 현재의 예측 수준을 향상할 수 있음을 시사하였다.

한편, 윤성주 외(2016)는 관측의 정확도가 떨어지는 이유가 예측하지 못한 큰 폭의 변동성이 가장 큰 문제로 지적하며, 단수 추정을 위해서는 전통적인 시계열모형보다 인공신경망모형이 적합할 수 있음을 강조한 바 있다. 인공신경망모형은 2000년대 딥러닝(deep structured learning)의 개발로 다시 주목받고 있으며, 최근 인공지능(AI)의 개발 등 다양하게 사용되고 있다. 양승룡 외(1999)는 축산물 가격 예측을 위해 인공신경망모형을 사용하여 분석하였고, 이 모형이 다른 시계열모형보다 우수함을 보인 바 있으며, 김배성(2005)은 양파가격에 대해 시계열(ARIMA, GARCH)모형과 인공신경망모형을 통해 예측을 시도하였으며, 분석 결과, 양파가격은 시계열모형과 인공신경망모형이 모두 예측력이 떨어져 추가적인 모형 개발을 시사한 바 있다.

기존 선행연구와 마찬가지로 본 연구는 가격 변동성이 큰 양파가격 예측을 위한 적합한 예측모형을 개발한다는 점에서 의의가 있다. 그러나 기존 선행연구는 다양한 분석 방법을 통해 가격을 예측하고 이를 비교, 검토하기보다는 단일 분석 방법 내에서만 예측을 수행한 경우가 다수 존재한다. 또한, 일부 선행연구의 결과 가격 예측력이 낮아 다양한 분석 방법과 모형, 변수를 고려해 양파가격 예측모형을 개선하는 것이 필요하다. 본 연구는 양파가격 예측을 하는데 수급을 고려한 구조모형과 단순 시계열모형, 인공신경망모형을 적용하여 각 분석 방법별 예측력을 비교하여 최적의 양파가격 예측모형을 설계하고자 한다. 더하여, 선행연구에서 한계점으로 지적했던 양파의 재고량과 수요량 등을 가격 예측모형에 반영했을 때, 예측력이 향상되는지 검토한다.

또한, 다양한 정보를 활용할수록 예측의 정확도를 높일 수 있으나, 현재 관측자료 정보가 부족하여 양파의 생육 정보를 이용한 분석은 거의 진행된 바 없다. 정부가 조사하는 양파의 생육 정보를 이용할 경우 양파가격의 예측력을 향상할 수 있을 것으로 기대되며, 생육 정보를 반영하기 위한 예측 방법론이 개발되어야 하며, 지역별로 기상 상황 및 생육환경의 차이가 존재하므로 주산지별로 구분하여 예측을 시도할 필요가 있다. 따라서 양파의 수급 안정을 위해서

는 보다 정확한 예측 결과가 필요하며, 예측력 향상을 위해 다양한 예측모형에 대한 비교분석이 필요하다.

본 연구의 목적은 구조모형, 시계열모형, 인공신경망모형을 이용하여 양파가격을 예측하고, 모형 간 비교를 통해 예측력 향상을 위한 함의를 제공하고자 한다. 본 연구의 세부 내용을 다음과 같다. 첫째, 양파 시장의 현황 조사를 통해 양파 시장의 문제점을 검토하여 양파가격 예측의 중요성을 제시하고자 한다. 둘째, 시계열모형, 인공신경망모형을 이용하여 양파가격을 예측 분석하며, 셋째, 구조모형의 세부항목별 예측을 분석하고자 한다. 넷째, 각 분석모형의 예측 효율성을 비교분석하고, 마지막으로 분석 결과를 이용하여 양파가격 예측력 향상을 위한 함의를 도출하고자 한다.

2. 양파 시장 현황 및 문제점

2.1. 양파 시장 현황 및 문제점

2.1.1. 양파 시장 현황

양파의 재배 작형은 파종 시기 및 재배 지형에 따라 구분되며, 우리나라에서 가장 일반화되고 수량을 많이 올릴 수 있는 작형은 추파 재배이다(농촌진흥청, 2018). 우리나라 양파 품종은 수확 시기에 따라 6종으로 나뉘고 있으나 크게 조생종과 중만생종으로 구분할 수 있다. 양파는 중만생종 재배비율이 약 90%이며, 조생종의 경우 저장이 어렵고 제주산 햇양파 출하 시기가 겹쳐 감소하는 추세를 나타낸다(aT KAIMS, 2021).

2020년 양파 생산량은 단수가 증가하였지만, 재배면적 감소로 전년 대비 감소하여 134만 톤을 기록하였다. KREI 농업전망(2021)에 따르면 2020년 양파 단수는 기후와 생육기 적절한 강우로 생육상황이 양호하여 전년 대비 2% 증가하였으며, 역대 가장 많은 단위당 수확량을 기록하였으며, 양파의 단수는 기상여건에 따라 증감을 반복하나 재배기술의 발달로 증가하는 추세를 보인다.

표 1. 연도별 양파 재배면적, 생산량, 단수

(단위: 천 ha, kg/10a, 천 톤, %)

구분	1975	1985	1995	2000	2005	2010	2015	2016	2017	2018	2019	2020
재배면적 (천 ha)	4	11	16	17	17	22	18	20	20	26	22	18
생산량 (천 톤)	95	440	975	878	1,023	1,412	1,094	1,299	1,144	1,521	1,595	1,340
단수 (kg/10a)	2,262	4,092	6,162	5,232	6,114	6,384	6,072	6,528	5,858	5,756	7,322	7,474

주: 1) 2020년 값은 2021년 KREI의 농업전망에서 인용함.

자료: 농림축산식품부(2020), KREI(2021)

우리나라 양파 주산지는 전남, 경남, 경북으로 이들이 전체의 77.4%(2020년 기준)를 차지할 정도로 주산지의 집중도가 매우 높다는 특징을 가지고 있다. 통계청(2021)에 따르면 최근 3년간 양파 재배면적이 지속해서 감소하고 있으며, 감소폭이 가장 큰 지역은 전북(-56.8%), 전남·광주(-49.3%), 충남·대전(-43.9%) 순인 것으로 나타났다. 양파는 중만생종 재배비율이 높지만, 최근 조생종 양파가격이 높게 형성되었으며 중만생종 대비 적은 생산비 등의 이유로 조생종 재배 비중이 증가한 것으로 나타났다(KREI, 2021).

양파 수입량은 다른 양념 채소류에 비해 많지 않으나, 국내 수급 상황과 최대 수입국인 중국산 양파의 수입 가능 가격에 따라 매년 증감을 반복하고 있다(KREI, 2021). 국내에서 생산된 양파는 대부분 국내에서 소비되어, 과일 생산되었던 2014년과 2019년을 제외하고 수출량은 미미한 수준이다.

표 2. 양파 품목별 수출입양 동향

(단위: 톤, %)

구분	2000	2005	2010	2015	2016	2017	2018	2019	2020	
수입	신선·냉장	4,869.5 (80.4)	38,348.6 (93.1)	17,667.4 (82.8)	146,285.5 (94.8)	45,236.2 (1.7)	148,406.4 (93.1)	58,015.4 (82.5)	22,751.6 (64.9)	28,485.0 (67.7)
	냉동	108.0 (1.8)	248.2 (0.6)	2,480.8 (11.6)	6,502.9 (4.2)	8,342.1 (15.1)	9,234.0 (5.8)	10,642.7 (15.1)	10,450.7 (29.8)	11,186.5 (26.6)
	건조	1,080.4 (17.8)	2,517.4 (6.1)	1,093.0 (5.1)	1,564.4 (1.0)	1,750.9 (3.2)	1,687.9 (1.1)	1,680.5 (2.4)	1,848.0 (5.3)	2,392.1 (5.7)
	초산제조	-	66.7 (0.2)	93.1 (0.4)	16.6 (0.0)	21.2 (0.0)	13.8 (0.0)	6.9 (0.0)	8.1 (0.0)	16.1 (0.0)
	합계	6,057.9 (100.0)	41,180.9 (100.0)	21,334.3 (100.0)	154,369.4 (100.0)	55,350.4 (100.0)	159,342.1 (100.0)	70,345.5 (100.0)	35,058.4 (100.0)	42,079.7 (100.0)

구분		2000	2005	2010	2015	2016	2017	2018	2019	2020
수출	신선·냉장	262.7 (100.0)	69.2 (81.9)	3,808.5 (99.5)	385.8 (100.0)	8,884.8 (99.7)	543.7 (99.5)	3,435.7 (99.2)	50,409.9 (100.0)	5,621.8 (100.0)
	냉동	-	-	12.1 (0.3)	-	24.0 (0.3)	-	24.0 (0.7)	-	-
	건조	0.1 (0.0)	14.5 (17.2)	8.2 (0.2)	-	-	-	-	10.0 (0.0)	0.1 (0.0)
	초산제조	-	0.8 (0.9)	0.5 (0.0)	-	-	3.0 (0.5)	2.0 (0.1)	-	-
	합계	262.8 (100.0)	84.5 (100.0)	3,829.3 (100.0)	385.8 (100.0)	8,909.3 (100.0)	546.7 (100.0)	3,461.7 (100.0)	50,419.9 (100.0)	5,621.9 (100.0)

주: 1) 각 품목에 대한 HS code는 신선·냉장(0703101000), 냉동(0710801000), 건조(0712200000), 초산제조(210190070)임.

2) 각 년은 1~12월의 수출입 실적임.

자료: 관세청(2021)

양파의 저온 저장고 입고량은 저온저장 시설 증가로 2010년 이후 전체 생산량의 48%를 저장하며, 65만 톤 내외를 유지하고 있다(KREI, 2021). 양파는 중만생종 위주로 저장되고 있으며 중만생종이 본격적으로 수확되는 6월부터 8월 사이에 이루어지며, 저장 기간은 약 8개월로 이듬해 4월까지 저장 양파가 소비된다(농촌진흥청, 2015; 김성우 외, 2016). 양파 재고량은 공식 통계가 없어 농업관측본부가 저장업체를 대상으로 입고량, 출고량, 부패율 등을 조사해 8월부터 이듬해 3월 말의 재고량을 추정하고 관측 월보를 통해 정보를 제공하고 있다.

표 3. 양파 생산량 및 저장량 추이

(단위: 천 톤, %)

구분	2005	2010	2015	2016	2017	2018	2019	2020
생산량(A)	1,023	1,412	1,094	1,299	1,144	1,521	1,594	1,340
저장량(B)	478	574	556	632	553	672	712	672
(B/A)*100	46.7	40.6	50.9	48.7	48.3	44.2	44.6	50.2

주: 1) 저장량은 농업관측본부 추정치이며 2020년산 생산량은 농업관측본부 실측 조사결과임.

자료: KREI(2021)

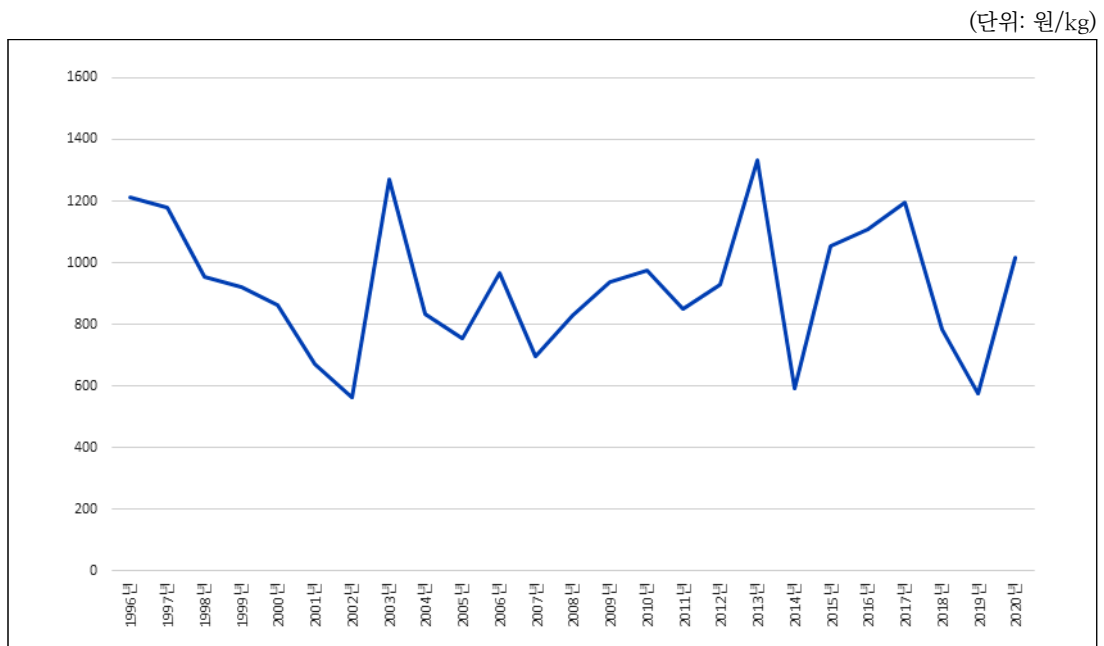
국내 양파 소비량은 수요 증가와 생산량 증가로 2000년 이후 지속해서 증가하였으며 2019년에는 1인당 30.9kg을 기록하고 있다(KREI, 2021). 국내 양파 중 96%는 가정용으로 소비되며, 나머지 4%는 가공용으로 소비된다(농촌진흥청, 2018).

양파의 도매시장 출하량은 매년 증가하고 있으며, 수확기 출하 비중은 전체의 30% 이상을

차지하며, 나머지는 간이저장 일부를 포함하여 저온 저장고에 저장되어 다음 해 새 양파가 생산될 때까지 소비되고 있다(농촌진흥청, 2018).

양파 도매가격은 국내산 생산량 변화에 따라 등락을 반복하며, 2018년과 2019년에는 생산량 과잉으로 실질 도매가격이 평년보다 낮게 형성되었으나, 2020년은 생산량 감소에 따라 전년 동기보다 높은 수준이다(KREI, 2021; aT KAMIS, 2021).

그림 1. 연평균 양파 실질 가격 동향



주: 1) 실질 도매가격은 도매가격을 생산자물가지수(2015=100)로 디플레이트 하였으며 상품 기준임.

자료: aT KAMIS(2021)

3.1.2. 양파 시장의 문제점

국내 양파의 수급은 전체적으로는 국내 생산에 의존하여 수급 균형을 유지하고 있으나 재배면적의 급격한 증감에 따른 생산량의 변동이 심한 해에는 부분적으로 수출과 수입을 통하여 수급의 균형을 유지하고 있다(농촌진흥청, 2018).

당해 연도 생산량과 입고량 수준에 따라 저장 출하기 가격 패턴이 다르게 나타나며 양파의

입고율은 이듬해 조생종 양파 재배와 가격에도 영향을 미친다. 양파 저장량이 마늘의 6~7배 수준이며, 저온 저장고 수 및 저장 주체(국내 저온 저장고 중 양파는 23%를 차지함)가 많으므로 저장 출하기에 수급 조절이 어려운 문제가 있다.

양파의 단수는 재배기술의 발달로 증가하는 추세를 보이거나 기상여건에 따라 증감을 반복하며 변동성이 큰 모습을 나타내고 있다(KREI, 2021).

양파의 10a당 실질 생산비는 2012년 이후 지속해서 상승하고 있으나, 실질 조수입은 양파 가격과 생산량 변동에 따라 증감을 반복하고 있어 소득의 불안정성이 나타나고 있다. 농업관측본부 실측 조사 결과(2020)에 따르면 가뭄이 있었던 2012~2013년 단수와 습해, 고온, 가뭄이 있었던 2017~2018년 단수가 감소하였다.

양파 수입량은 다른 양념 채소류에 비해 많지 않으나, 국내 수급 상황과 최대 수입국인 중국산 양파의 수입 가능 가격에 따라 매년 증감을 반복하여 수급 불안정 상황을 보여주고 있다(KREI, 2021). 실제 2014년, 2019년은 생산량 과잉으로 정부의 수출 촉진정책으로 수출이 증가하는 모습을 보였고, 2015년, 2017년은 국내산 양파 생산량 부족으로 TRQ 도입에 따라 18~21만 톤의 양파가 수입되는 모습을 보였고, 그다음 해(2016년, 2018년은)에는 7만 톤 내외로 수입량이 감소하였다.

노지 재배 특성으로 따른 기상 변화에 따른 생육환경 변화에 민감하여 생산량의 변동이 크며, 이로 인해 가격 변동성이 커져 농가 소득이 불안정할 뿐만 아니라, 합리적인 의사결정을 내리는데도 어려움을 겪고 있다. 그러나 양파가격은 생산뿐만 아니라 소비, 수출입, 재고 등의 영향을 받지만, 정보가 미흡하여 가격 변동성을 쉽게 예측하지 못한다는 한계가 있다.

2.2. 양파가격 예측의 중요성 및 필요성

매년 양파 등 주요 채소류에 대한 수급 실패 사례는 반복되고 있으며, 금값된 채소, 재배면적 예측 실패로 인한 수급 초비상, 정부의 소극적 정책에 대한 비판 등으로 언론에서 주목받고 있다. 특히 수급 전망 실패에 따른 지자체 재정부담과 농민 부담 증가를 감안하면 정확한 관측과 수급 전망은 중요하다. 특히 양파는 연중 가격 변동성이 높고, 재배면적의 고정성이 낮아

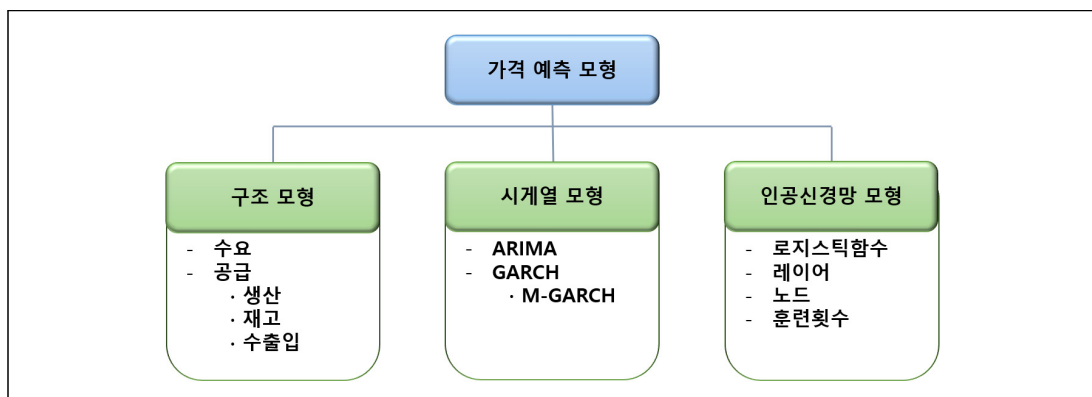
수급 전망이 더 어려운 것으로 판단된다(김동환 외, 2017). 양파는 대부분 노지에서 재배되므로 기온, 강수량, 일조량 등 기상요인에 의해 단수가 크게 영향을 받는 특징을 지니며, 최근 빈번한 기상이변과 농가의 불확실한 재배 의사결정으로 양파 생산량 예측의 어려움이 더욱 커지고 있다(최성천, 백장선, 2016; 오승원, 김민수, 2017; 이승인 외, 2020). 정확한 가격 예측은 생산량 및 재배면적, 재고량 등 농가의 의사결정 및 소득, 저장업체의 출하 행태에 영향을 미쳐 양파의 수급 조절에 가장 큰 영향을 미친다. 따라서 양파의 개별 품목적 특성에 맞는 변수를 모색하여 가격 예측모형에 반영하는 것이 필요하다.

3. 예측 분석 방법 및 자료

3.1. 가격 예측 분석의 개요

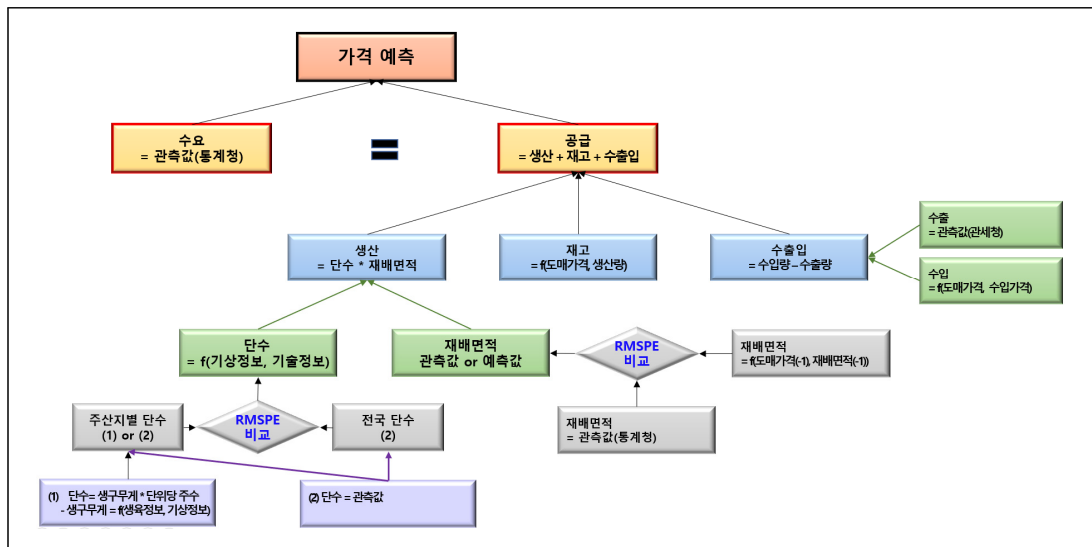
본 연구에서는 양파가격 예측의 정확도 향상을 위해 다양한 모형을 이용하여 가격 예측을 분석하며, 분석모형은 크게 구조모형, 시계열모형, 인공신경망모형으로 구분한다(그림 2).

그림 2. 가격 예측 모형의 구조



가격 예측모형 중 구조모형은 세부모형으로 구분될 수 있다. 가격은 수요와 공급이 일치하는 균형점에서 결정이 되며, 공급은 다시 생산, 재고, 수출입으로 구분할 수 있다. 또한, 생산은 단수와 재배면적에 의해 결정된다.

그림 3. 구조모형의 세부 항목



본 연구는 1월, 4월, 6월, 10월의 국내 양파가격 예측을 분석하였다. 4월의 경우, 조생종이 본격적으로 수확되어 도매시장에 출하되며 햇양파의 가격이 결정되는 중요한 시기이므로 4월의 양파가격을 예측하였다. 6월은 국내 양파 물량의 가장 많은 부분을 차지하는 중만생종이 수확되어 도매시장에 출하 시기로 국내 양파 농가의 소득이 결정되는 중요한 시기이므로 6월의 양파가격을 예측하였다. 일반적으로 국내 양파 농가는 10월 상순에서 11월 상순 사이에 정식을 시작하며, 이때 재배면적에 대한 의사결정을 하므로 농가들은 10월의 양파가격에 관심이 높은 것으로 파악하여 10월의 국내 양파가격을 예측하였다. 마지막으로 1월은 KREI 주최하는 농업전망 대회¹⁾에서 양파의 수급 동향과 당해 상반기의 가격 전망을 발표하는 시기로

1) KREI 농업전망(2021)에는 당해 상반기 전망으로 추정 재고량, 품종(조생종, 중만생종)별 재배면적, 단수, 생산량을 발표하며, 중장기 전망으로 KREI-KASMO 모형을 이용하여 당해, 2025년, 2030년의 재배면적, 총공급량(생산량, 수입량), 총수요량(소비량, 수출량) 전망을 발표함

서 동일한 기준을 적용하여 양파가격 예측 분석한 결과와 비교 분석하기 위해 1월의 양파가격을 예측하였다.

3.2. 양파가격 예측모형의 분석 방법

3.2.1. 구조모형을 이용한 가격 분석 방법

가) 가격모형 분석 방법

본 연구에서는 구조모형을 통한 가격 예측 모형을 분석하였다. 본 연구에서는 김연중 외(2006) 연구와 같이 역수요 함수를 이용하여 가격예측모형을 구축하여 양파가격을 추정 및 예측하고자 하였다.

가격 예측을 위해 모형에 사용된 시계열 자료는 ADF(Augmented Dickey Fuller) unit root test를 통해 단위근을 검정하여 안정성(stationarity)을 확보하였다. 또한, 연평균, 4월, 6월 10월 가격예측모형별로 예측에 적합한 모형을 판단하기 위해 모형의 적합성(model validation)을 판단한 후, 모형 선정 기준을 통해 적정 모형을 식별하였다.

가격 예측에 적합한 모형을 판단하기 위해 첫째, 자기상관(autocorrelation) 존재 여부, 둘째, 이분산성(heteroskedasticity) 존재 여부, 셋째, 통계적으로 유의미한 변수의 존재 여부를 기준으로 모형의 적합성(validation)을 판단하였다.

이분산성 검정은 OLS를 이용한 회귀분과 오차항의 이분산성을 검정하기 위해 Breush-Godfrey Heteroskedasticity test를 이용하여 이분산성 존재 유무를 확인할 수 있으며, 오차항의 동분산(homoskedasticity) 가정이 기각되어 이분산성(heteroskedasticity)이 존재하면 Heteroskedasticity-Consistent Matrix Estimator를 이용하여 모형을 추정하였다. 자기상관은 시계열 자료를 이용하는 회귀분석에서 흔히 발생하는 문제이며, 고전적 선형회귀 모형은 오차항이 상관관계를 갖지 않는다는 것을 가정하고 있으므로(Damodar Gujarati, 2016), Breush-Godfrey Serial Correlation LM test를 이용하여 자기상관(AR) 구조를 확인하였다. 오차항의 자기상관이 있는 경우, GLS(Generalized Least Squares)를 이용한 회귀분석을 통해 모형을 추정하였다.

최종적으로 예측 모형은 RMSPE(Root of Mean Square Percent Simulation Error)와 R^2 을 기준으로 모형을 선택하였으며, RMSPE 계산식은 다음과 같다.

$$(1) \text{RMSPE}(\%) = \sqrt{\frac{1}{k+1} \sum_t^{t+k} \left(\frac{\text{예측치}_t - \text{실제치}_t}{\text{실제치}_t} \right)^2} * 100$$

국내 양파가격예측모형은 포함되는 설명변수에 따라 다양한 모형으로 구분하였으며, 모형 적합성(model validation) 기준과 모형 선택(model selection) 기준에 따라 연평균 수입량 예측치 추정에 가장 적합한 모형을 선택하였다. 국내 양파가격 추정을 위해 본 연구에서 사용된 자료는 <표 4>와 같다.

표 4. 국내 양파가격예측모형의 주요 변수

구분	변수명	변수	출처	연구
종속변수	$price_t^m$	t 년 m 월의 국내 실질가격 (상품 도매가격, 월, 원/kg)	명목가격(aT KAMIS)	김명환 외(2000), 김연중 외(2006), 안병일(2006)
독립변수	$demand_t^m$	t 년 m 월의 1인당 수요량 (월, kg/명)	수출입양(농넷, 관세청), 총인구(통계청), 월별 생산량(직접 계산)	
	di_t^{Q2}, di_t^{Q3}	t 년의 2, 3분기의 1인당 실질 가처분소득(분기, 원)	통계청	김연중 외(2006), 안병일(2006)

주: 1) 국내 양파가격과 수입가격은 CPI(통계청, 2015=100)으로 실질가격화 하였으며, 1인당 가처분소득은 GDP디플레이터(한국은행, 2015=100)를 이용하여 실질가격화 하였음.

국내 양파가격은 도매시장 양파 상품의 경락가격 기준이며, 국내 양파의 1인당 소비량은 t 기의 국내 양파 총공급량에 t 기의 총인구로 나눠 계산하였으며, 본 연구에서는 1인당 공급량은 모두 소비됨을 가정하였다.

$$(2) \text{1인당 수요량}_t = \frac{(\text{생산량}_t + \text{재고량}_t + \text{수입량}_t - \text{수출량}_t)}{\text{총인구}_t}$$

각 가격예측모형에 사용된 생산량은 <표 5>와 같이 계산하였다. 여기에서 도매시장 반입량은 전국 도매시장에 반입된 총량을 의미하며, 모든 양파를 대상으로 하였다. 또한, 생산, 재

고, 수입모형을 통해 산출된 예측값은 연간 데이터의 형태로 연평균 가격 예측 모형과 6월 가격 예측모형의 총공급량에 반영하여 산출하였다. 이용 가능한 자료의 한계로 연평균 가격예측모형의 자료 기간은 1996년~2020년이며, 4월 모형은 2009년~2020년, 6월 모형은 1997년~2020년, 10월 모형은 2009년~2020년이다.

표 5. 가격예측모형에 따른 생산량 계산 방법

모형	생산량	계산	근거
연평균 가격 예측모형	당해의 총공급량	연 생산량 = 생산량 _t + 수입량 _t - 수출량 _t	한 해 총공급된 양파의 총량
4월 가격 예측모형	4월 총공급량	4월 생산량 = 4월 조생종생산량 _t + 3월말 재고량 _t + 수입량 _t - 수출량 _t	4월 시장에는 당해 조생종 생산량과 함께 전년도 저장 양파가 함께 거래됨
6월 가격 예측모형	6월 총공급량	6월 생산량 = $\frac{6월도매시장반입량}{도매시장월별경유율}$ cf. 경유율 = $\frac{월별도매시장반입량}{총공급량}$	6월은 당해 중만생종 양파가 생산되어 출하되는 시기임
10월 가격 예측모형	10월 총공급량	10월 생산량 = 9월말 재고량 + 수입량 - 수출량	10월은 양파가 생산되지 않으며 저장된 양파가 출하되는 시기임

주: 1) 사용 가능한 자료의 한계에 따라 월별 도매시장 양파 경유율은 연간 경유율과 모두 동일함으로 가정

나) 단수모형 분석방법

본 연구에서는 생육정보를 이용하여 단수를 예측하는 방법과 주어진 관측값을 활용하여 예측하는 방법을 활용하였다. 또한, 전국적으로 단일화된 단수를 추정하여 예측하는 방법과 주산지별 특성을 고려하여 지역별 단수를 추정하여 전국 생산량을 추정하는 방법으로 구분하여 분석을 실시하였다.

우선 생육정보를 이용하여 추정하는 방법의 경우, 단수는 다양한 생육환경 및 기상환경에 민감하게 영향을 받으며, 지역별 특징에 따라 단수의 차이가 나므로 지역별 단수를 추정하기 위한 모형이 필요하다. 통계청에서 관측된 지역별·연도별 자료를 제공하고 있으나, 생육정보 및 기상정보를 활용하여 관측을 시도하기 위해서는 단수를 새로 설정할 필요가 있다.

단수모형의 추정은 한국농촌경제연구원 및 농촌진흥청에서 제공하는 생육자료를 이용한 단수 추정과 통계청에서 제공하는 지역별 단수를 이용하여 분석하였다.

생육자료를 이용한 단위 추정 방법은 통계청에서 제공하는 단위 대신 지역별 생육정보를 활용하여 단위를 계산하였으며, 단위를 산정하는 계산식은 아래의 식 (3)과 같다.

$$(3) \text{ 단위 } (kg/10a) = 1m^2 \text{ 당 주수} * \text{생구무게} * 1,000m^2$$

여기서 생구무게는 생육정보와 기상변수에 의해 결정되며, 생구무게는 아래와 같은 함수식으로 추정할 수 있다. 이때 단위추정에 활용된 생육 정보는 농촌진흥청과 KREI에서 제공하는 필지정보, 생육정보(지상부, 지하부)를 이용하였다.

$$(4) \text{ 생구무게} = f(\text{초장, 잎끝고사장, 엽초장, 엽초경, 기상변수})$$

사용되는 기상변수는 평균기온, 기온 차(최고기온-최저기온), 강수량, 일조시간, 풍속을 고려하였다. 그러나 분석하기에는 자료 기간이 2019년과 2020년 단 2개년도의 자료만 존재하여 일반적인 방법을 이용하여 분석이 어려우며, 설명변수 및 종속변수가 온전히 유지된 ID별 자료를 이용하여 분석하였다.

주어진 관측값을 이용하여 단위를 추정하는 방법은 통계청에서 제공하는 관측정보를 활용하여 분석하였다. 생육정보 모형과 동일하게 기상정보를 이용하여 단위를 예측하였으며, 전국의 평균 단위를 이용하여 생산량을 추정하는 방법과 주산지별 단위를 추정하여 전국의 생산량을 예측하는 방법을 비교 분석하였다. 단위 예측을 위한 관측 기간은 1980년부터 2020년까지이다. 분석 방법은 생육정보 모형과 동일하게 인공신경망모형을 이용하여 분석하였다.

또한, 단위는 외부의 여러 요인에 의해 영향을 받아 변동성이 심해 전통적인 시계열모형보다는 인공신경망을 통한 분석이 더 적합할 수 있다(양승룡, 1999). 인공신경망 중 가장 널리 사용되는 모형이 오류 역전파 알고리즘(error back-propagation algorithm)을 이용한 다층 퍼셉트론(multi-layer perceptron)으로 본 연구에서도 동일한 방법을 사용하였다. 인공신경망을 이용하여 단위를 예측하기 위해서는 활성화 함수의 형태, 알고리즘, 각 층의 노드 수, 은닉층의 수, 데이터 표준화 방식 등을 결정해야 한다. 본 연구는 활성화 함수로 로지스틱 함수(logistic function), 알고리즘으로 가중치의 역추적을 허용하는 '탄력적 역전파 방식(resilient-backpropagation with weight backtracking)'의 알고리즘을 적용한다.²⁾

입력층과 출력층의 노드 수는 각각 투입변수와 출력변수의 수이다. 은닉층과 은닉층별 노드 수는 시나리오별 예측력을 비교하여 결정해야 한다(윤성주 외, 2016). 본 연구는 분석의 단순화를 위해 3개 은닉층 시나리오(1개, 2개, 3개)와 노드 수 시나리오(5, 10, 15, 20, 30) 등 총 5개 시나리오를 설정하고, 예측력이 가장 우수한 시나리오의 은닉층과 노드 수를 단수 예측 모형에 적용하였다. 또한, 출력층의 활성화 함수로는 지수 함수를 사용하였다.

인공신경망모형을 효율적으로 이용하기 위해서는 데이터 표준화 작업이 필요하다. 본 연구에서는 윤성주 외(2016)에서 설명한 표준화를 위한 3가지 방식 중 평균과 표준편차를 이용한 z 값의 형태로 전환해 주는 방법이 가장 예측력이 우수한 것으로 도출되어 사용하였으며, 산출식은 식 (5)와 같다.

$$(5) z_i = \frac{(x - \mu)}{\sigma}$$

인공신경망모형의 예측력을 비교하기 위해 동일한 자료를 이용하여 패널 데이터 분석을 통해 예측력을 분석하였다. 패널 모형을 이용하여 단수를 분석하기 위해 일원 오차 성분 모형을 이용하였으며, 모형의 구조는 다음의 식 (6)과 같다.

$$(6) y_{i,t} = \alpha + \sum_{k=1}^K \beta_k X_{k,i,t} + u_t + \epsilon_{i,t}$$

여기에서, y 는 종속변수로서 t 시점 i 지역의 단수, X 는 시점 t 지역의 k 번째 설명변수, u 는 관찰되지 않은 시간 불변 오차항, e 는 지역 및 시간에 따라 달라지지 않는 순수 오차항을 의미한다. 단수 예측 모형의 주요 설명 변수인 기상 정보는 각 지역의 지형 등 모형에 반영되지 못한 지역별 특성의 영향을 받기 때문에 패널 고정효과 모형이 적합하므로(이춘수 외, 2017) 본 연구에서는 패널 고정효과 모형을 사용하여 분석하였다.

$t+1$ 시점의 예측 단수 $E(y_{i,t+1})$ 는 다음의 식 (7)을 이용하여 추정할 수 있다.

2) 탄력적 역전파 알고리즘의 내용에 대한 설명은 Gunther and Fritsch (2010)을 참고할 것

$$(7) E(y_{i,t+1}) = \hat{\alpha} + \sum_{k=1}^K \hat{\beta}_k X_{k,i,t+1}$$

다) 재고모형 추정방법

본 연구에서는 당월의 가격, 전월 말 재고량, 생산량을 이용하여 재고 예측모형을 구성하였다. 한 해 저장 입고량은 생산량이 많을수록 많으며 월별 재고량은 한해 저장 입고량에 영향을 받을 수 있으며, 재고와 가격은 반비례 관계이고, 전월 말 재고는 당월 초 재고로 이월되므로 전월 말 재고가 많을수록 당월 말 재고도 많아지는 경향이 있다.

재고모형 분석은 양파의 월별 가격을 예측하는 것이므로 재고함수도 동 기간에 맞춰 분석을 실시하였다. 월별 재고량의 특징을 보면 저장 출하기인 8월부터 익년 4월까지 재고는 가격에 영향을 끼치며, 3월 말 재고, 9월 말 재고는 각각 4월 가격, 10월 가격에 영향을 끼친다. 만약 양파의 연도별 재고량의 변동이 크지 않고, 월별 감소 행태가 같다면 재고량 추정이 의미가 없으나, KREI 농업관측본부에서 제공하는 재고량 자료를 검토한 결과, 2008년에서 2020년까지 연도별 생산량에서 재고가 차지하는 비중이 40~60%이며, 월별 감소 행태가 다르므로 재고모형 추정이 필요하다. 3월 말, 9월 말 재고모형은 다음 식 (8), 식 (9)와 같이 구성하였다.

$$(8) ES_{3,t} = \beta_o + \beta_1 ES_{2,t} + \beta_2 RP_{3,t} + \beta_{3,t} Q_{t-1} + \epsilon$$

$$(9) ES_{9,t} = \beta_o + \beta_1 ES_{8,t} + \beta_2 RP_{9,t} + \epsilon$$

$ES_{m,t}$: t년 m월 말 재고
 $RP_{m,t}$: t년 m월 실질가격
 Q_t : t년 생산량

3월 말과 9월 말 재고는 당월 실질 가격과 전월 말 재고량을 이용하여 추정하였으며, 9월 말 재고 추정 시 생산량과 8월 말 재고 변수 간 상관관계가 높아(0.81) 모형에 8월 말 재고만 반영하였다. 또한, 생산량이 많을수록 저장 입고량이 많아 8월과 9월 말 재고가 많은 경향이 나타나는 것으로 분석되었으며, 3월 말 재고 추정 시 생산량과 2월 말 재고는 변수 간 상관관

계가 높지 않아(0.14) 모형에 함께 반영하였다. 월별 재고모형 분석을 위해 필요한 자료를 정리하면 <표 6>과 같다.

표 6. 재고모형 분석자료 및 기초통계량

변수		평균	표준편차	최대	최소	변이계수	출처
실질 도매가격 (원/kg)	3월	935	430.82	1,940	432	46.06	aT KAMIS
	9월	777	246.34	1,221	362	31.70	
재고량(톤)	2월	137,185	56,869.21	259,149	76,380	41.45	농업관측본부
	3월	39,151	18,739.09	71,491	17,037	47.86	
	8월	601,857	69,727.27	729,372	508,612	11.59	
	9월	551,750	74,064.73	682,610	445,087	13.42	
생산량(톤)		1,350,375	170,822.61	1,594,450	1,093,932	12.65	통계청

주: 1) 실질도매가격은 한국농수산식품 유통공사(aT KAMIS)에서 제공하는 양파 상품 가격에 통계청에서 제공하는 소비자물가지수(2015=100)를 적용해 계산함.

라) 수출입 모형 추정방법

국내 양파 시장에서의 수입수요를 고려하기 위해 모형 수입 예측 모형을 설정하고 변수를 선정하였다. 박기환 외(2007)는 양파의 수입량(t), 국내 도매가격(t, t-1), 국내 수입 단가(t, t-1), 생산량(t), 저장량(t)를 변수로 설정하여 수입 예측모형을 개발하였으며, 수확기와 저장기를 구분하여 회귀분석(OLS)을 통해 추정하였다. 또한, 한석호 외(2010)은 월평균 및 연평균 도매가격, 소비자물가지수, 수입량, 수입단가, 관세율을 고려하여 수입 예측을 시도하였다.

본 연구에서는 양파 수입량 예측을 위한 회귀모형(OLS 혹은 GLS을 이용)을 설정한 뒤, 실측치(actual value)와 회귀분석을 통해 추정한 예측치(forecasted value)를 비교하여 예측력을 평가하였다. 수입량 예측을 위해 모형에 사용된 시계열 자료의 안정성(stationarity)을 확인하기 위해 ADF(Augmented Dickey Fuller) unit root test(단위근 검정)을 실시하였으며, 모형의 예측력과 설명력을 기준으로 모형별 적합한 모형을 선택하여 예측치를 추정하였다.

본 연구에서는 수입량, 국내 도매가격, 국내 수입가격, 1인당 실질 가처분소득 자료를 이용하여 수입 예측 모형을 구축하였으며, 분석자료는 아래의 <표 7>과 같다. 설명변수로서 사용

되는 국내 도매가격과 생산량 및 재고량은 다중공선성 문제를 일으키므로 분석모형에서 생산량과 재고량을 제외하였다.

표 7. 국내 양파 수입모형의 주요 변수

구분	변수명	변수	출처	연구
종속 변수	$import_t^m$	t 년 m 월의 국내 양파 수입량 (신선·냉장 환산량, 월, kg)	농넷('96~'99년), 관세청('00~'20년)	김연중 외(2006), 박기환 외(2007), 한석호 외(2010),
독립 변수	$price_t^m$	t 년 m 월의 국내 실질가격 (상품 도매가격, 월, 원/kg)	명목가격(aT KAMIS)	
	$w price_t^m$	t 년 m 월의 수입 실질가격 (월, 원/kg)	수입금액(농넷, 관세청), 관세율(무역협회), 환율(한국은행)	
	di_t^Q, di_t^B	t 년의 2, 3분기의 1인당 실질 가처분소득(분기, 원)	통계청	이상민·장철수(2006)

주: 1) 국내 양파가격 및 수입가격 CPI(통계청, 2015=100)으로 나눠 실질가격을 사용하였으며, 1인당 가처분소득은 GDP디플레이터(한국은행, 2015=100)를 이용하여 실질소득으로 사용함.

2) 자료 기간은 1996년부터 2020년임.

수입량의 경우, 국내로 수입되는 4종류의 양파(신선냉장, 건조, 냉동, 초산조제) 품목을 신선·냉장 양파로 환산한 신선·냉장 환산값을 이용하였으며, 환산식은 다음의 식 (10)과 같다.

$$(10) \text{ 양파 환산량} = \text{신선냉장 수입량} + (\text{건조양파}/0.072) \\ + (\text{냉동양파}/0.9) + (\text{초산조제양파}/0.85)$$

수입단가(\$/kg)는 환율(원/kg)을 고려하여 원/kg로 환산하여 사용하였으며, 세율은 국내 양파 수입국 중 가장 큰 비중을 차지하고 있는 중국과의 신선·냉장 양파의 관세율을 기준으로 계산하였다. 이를 이용하여 최종 수입가격 산정식은 다음의 식 (11)과 같다.

$$(11) \text{ 국내 수입가격(원/kg)} = \text{수입단가} * (1 + \text{관세율}) * (1 + \text{부가세율}) * \text{환율}$$

본 연구에서는 최근 3년(2018년~2020년)의 국내 양파 수입량을 예측하기 위해 연평균 수입모형 기준으로 1996년~2017년의 자료를 이용하여 모형을 추정한 후 2018년의 수입량을 예측하고, 2019년 수입량은 1996년~2018년의 자료를 이용하여 예측하였으며, 2020년 수입량은 1996년~2019년 자료를 이용하여 예측하였다.

3.2.2. 시계열모형을 이용한 가격 분석 방법

양파의 가격 예측을 위해 시계열 데이터 분석에서 가장 기본적이고 간단한 시계열모형을 설정한 뒤, 예측치와 실측치를 비교해 예측력을 평가하였다. 또한, 양파가격에 영향을 미치는 적합한 외생변수를 반영함으로써 예측력이 향상되는지도 검토하였다.

본 연구의 예측 대상은 연평균 가격과 4월, 6월, 10월 월별 가격이며, 대상별로 다른 모형을 적용하였다. 모형은 추세 시계열모형, ARDL(자기회귀시차) 모형, ARMA(자기회귀이동평균) 모형, ARMA에 외생변수를 도입한 자기회귀이동평균(ARMA with Exogenous Variables: ARMAX), 이분산이 있을 시 이를 고려한 GARCH 모형을 사용하였다.

추세 시계열모형은 시계열모형에서 추세를 고려한 모형을 의미한다. 자기회귀이동평균 모형은 예측하고자 하는 대상의 과거 정보만을 바탕으로 미래값을 예측하며, 식 (12)로 표현될 수 있다.

$$(12) \quad y_t = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{i=0}^q \beta_i \epsilon_{t-i}$$

외생변수를 고려한 자기회귀이동평균 모형은 식 (13)으로 표현되며, x 와 γ 는 외생변수와 회귀계수의 벡터를 나타낸다.

$$(13) \quad y_t = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{i=0}^q \beta_i \epsilon_{t-i} + x_t' \gamma$$

자기회귀시차 모형은 종속변수의 과거값과 설명변수의 현재값, 과거값이 모두 종속변수의 현재값에 영향을 주는 모형으로 식 (14)로 표현된다.

$$(14) \quad y_t = \alpha_0 + \gamma y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t$$

앞의 모형은 분산이 시간에 따라 변하지 않는다는 동분산(Homoskedasticity)을 가정하지만, 시계열 자료에서는 이분산(Heteroskedasticity) 문제가 빈번하게 나타나며, 이를 고려할 필요가 있다. AREA의 잔차항 ϵ_t 을 식 (15)와 같이 변형하여 정의할 수 있다.

$$(15) \epsilon_t = v_t \sqrt{h_t} \quad \text{where } \sigma_v^2 = 1, \\ h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i}$$

여기서, v_t 는 white-noise process이므로, ARMA 잔차항인 ϵ_t 의 조건부, 무조건부 평균은 0이며, 이는 결과적으로 $E\epsilon_t = Ev_t(h_t)^{1/2}$ 이다. GARCH(p,q) 모형은 GARCH term 시차가 p, ARCH term 시차가 q이며, 만약 β_i 가 0이면 ARCH(q)이다.

분석의 순서는 첫째, 시계열 자료의 안정성 확보 후 모형 식별(Model Identification), 둘째, 표본 외 예측 방식을 활용해 예측을 수행하고, RMSPE나 MAPE 등의 예측력 지표를 바탕으로 예측력을 평가 후 최종 가격예측 모형 설정, 셋째, 본 연구의 가격 예측력과 기타 선행연구의 예측력을 비교하는 순으로 진행하였다.

시계열 자료의 안정성을 확보하기 위해 Box-Jenkin's approach를 활용해 모형을 식별하였다. 우선 단위근 검정을 통해 시계열 자료의 안정성 여부 검정하였고, 안정적 시계열 자료로 변환 이후, 자기상관함수와 편자기상관함수를 계산하고, 모형의 최종 차수는 AIC, SBC 통계량을 기반으로 선택하였다. 마지막으로 최종 모형의 추정결과를 바탕으로, 자기상관이나 모수의 안정성, 이분산성 등을 진단하였다.

ARMA 등 동분산을 가정한 모형 외에 양파가격 시계열에 이분산이 있을 시, 이분산을 고려한 GARCH 모형 등을 적용해 예측력이 높은 모형을 설정하였다.

본 연구에 이용된 양파가격 자료는 한국 농수산물유통공사 농산물유통정보에서 제공하는 월별 상품 도매가격으로, 단위는 원/kg이다. 분석 기간은 1996년부터 2020년까지이며 모형의 식별 및 모수 추정을 위해 1996년부터 2017년까지의 자료를 내표본(In-Sample)으로 이용하였고, 예측 및 진단을 위해 2018년부터 2020년까지의 자료를 외표본(Out-of-sample)으로 이용하였다.

3.2.3. 인공신경망모형 및 가격 분석 방법

일반적으로 가격 예측에는 선형 시계열모형이 널리 사용되고 있으나, 예상치 못한 큰 폭의

가격 등락이 있고, 다양한 외부요인의 영향을 받는 경우에는 선형 시계열보다 인공신경망모형이 적합할 수 있다(양승룡 외, 1999; 윤성주 외, 2016). 다만, 시계열모형과는 달리 인공신경망모형에서는 시차 변수가 체계적으로 고려되지 않는다(김종진 외, 2016). 따라서 인공신경망모형의 적합도와 예측력을 높이기 위해 시차 변수 대신 3개월 이동평균 값을 사용하였다.

4. 양파가격 예측 분석 결과

4.1. 구조모형 분석 결과

4.1.1. 가격모형 분석 결과

분석 결과, 연평균 가격 및 10월의 예측력은 비교적 우수한 것으로 분석되었으나, 4월과 6월의 가격 예측력은 낮은 모습을 보였다. 이는 4월과 6월은 전국에서 양파가 출하됨으로 인해 가격이 혼란스러운 시기이므로 예측이 비교적 어렵다고 볼 수 있다.

표 8. 최근 3년의 국내가격 예측 결과

(단위: 원/kg)						
모형	구분	2018년	2019년	2020년	RMSPE (out-of-sample)	RMSPE (in-sample)
연평균 가격예측모형	예측치	845	688	1,012	12.09%	7.59%
	실제치	785	575	1,019		
	RMSPE	7.57%	19.50%	0.69%		
4월 가격예측모형	예측치	941	661	603	31.72%	14.06%
	실제치	752	826	1,088		
	RMSPE	25.18%	19.95%	44.57%		
6월 가격예측모형	예측치	745	564	512	19.56%	60.85%
	실제치	673	463	669		
	RMSPE	10.71%	21.99%	23.44%		
10월 가격예측모형	예측치	731	588	1,000	13.30%	3.66%
	실제치	683	495	1,125		
	RMSPE	7.01%	18.92%	11.13%		

4.1.2. 단수모형 분석 결과

생육 정보를 이용하여 인공신경망모형으로 단수를 추정하는 경우, 예측력 향상을 위해 훈련 횟수, 은닉층 수, 노드 수를 동일한 정보를 이용하여 비교 분석하였다. 훈련 횟수의 경우 100회를 학습시키는 경우의 RMSPE가 가장 높은 것으로 분석되었다.

예측력과 효율성을 높이기 위해 은닉층 수와 노드 수의 결합을 통해 RMSPE를 비교 분석한 결과 2개의 은닉층과 (10, 10)의 노드 수를 설정하는 것이 가장 예측력이 높은 것으로 분석되었다.

생육 정보를 이용한 인공신경망 분석 결과, 내표본 예측 오차 수준은 0.14%로 분석되었다. 그러나 예측력의 정확도 수준은 외표본을 도출하여 비교하는 것이 일반적이나, 생육자료의 기간이 짧아 시계열 자료 분석이 어려워 횡단면 자료를 이용하여 분석하였기 때문에 외표본을 도출하는 데 한계가 있었다. 다만, 인공신경망모형은 자료의 크기가 커질수록 예측의 정확도가 높아지는 특징이 있으므로 지속적인 생육 정보를 관측할 경우, 인공신경망모형을 이용하여 분석한다면 시간의 흐름에 따른 예측력을 향상할 수 있으며, 본 연구에서는 생육 정보를 활용하여 단수를 추정할 수 있다는 방법을 제안한다는 점에서 의미가 있다.

표 9. 생육정보를 이용한 인공신경망모형의 예측 결과

구분		실제값	예측값	RMSPE
in sample	제주1	0.82206	0.82094	0.0014
	제주2	1.46901	1.46759	
	제주3	-0.29816	-0.29316	

지역별 단수를 인공신경망모형으로 예측한 결과는 다음의 <표 10>과 같다. 4개의 주산지 모두 외표본을 이용하여 RMSPE를 분석한 결과 20% 이상의 높은 오차가 나타나는 모습을 나타내었다. 특히 가장 재배면적이 넓은 전남은 49% 수준으로 오차가 발생한 것으로 분석된 반면 초생종이 주를 이루는 제주지역의 RMSPE는 23% 수준으로 비교적 낮은 것으로 분석되었다.

표 10. 인공신경망을 이용한 지역별 단위 예측 결과

구분	지역	연도	실제값	예측값	RMSPE
내표본	전남	2018	4932.0	4941.3	0.0064
		2019	6873.0	6874.2	
		2020	7883.0	7851.8	
	경북	2018	6296.0	6291.8	0.0129
		2019	7471.0	7468.3	
		2020	8098.0	8086.3	
	경남	2018	6599.0	6599.0	0.0766
		2019	8174.5	8174.5	
		2020	8780.3	8780.3	
	제주	2018	6740.0	6740.7	0.0129
		2019	7875.0	7877.5	
		2020	7594.0	7592.4	
외표본	전남	2018	4932.0	4975.3	0.4906
		2019	6873.0	6795.6	
		2020	7883.0	5739.1	
	경북	2018	6296.0	6355.1	0.4481
		2019	7471.0	7272.7	
		2020	8098.0	7286.7	
	경남	2018	6601.0	6622.3	0.2891
		2019	8170.0	7544.0	
		2020	8783.0	7577.2	
	제주	2018	6740.0	6697.6	0.2315
		2019	7875.0	7187.7	
		2020	7594.0	7221.7	

전국 평균을 이용하여 생산량을 예측한 결과와 주산지별 단수를 추정한 예측한 결과는 아래의 <표 11>과 같다. 분석 결과, 전국 평균을 이용할 경우의 외표본 오차 수준은 약 9.1%인 반면, 주산지별 단수를 추정하여 예측하는 경우 외표본 오차 수준은 25.8% 수준인 것으로 분석되었다. 이는 지역별로 재배환경 및 기상환경, 기타 외부적 요인으로 차이로 단위 추정의 예측력이 떨어져, 전체적인 생산량의 예측이 낮아진 것으로 판단된다.

표 11. 인공신경망을 이용한 전국 평균 및 지역별 단수를 이용한 생산량 예측 결과

구분	연도	실제값	실제값	예측값	RMSPE
전국 평균	내표본	2018	152,102,300	151,974,250	0.0007
		2019	159,451,194	159,447,104	
		2020	116,826,426	116,838,377	
	외표본	2018	152,102,300	151,729,181	0.0908
		2019	159,451,194	147,996,348	
		2020	116,826,426	100,487,213	
주산지별	내표본	2018	152,102,300	149,373,652	0.075
		2019	159,451,194	153,707,033	
		2020	116,826,426	113,825,704	
	외표본	2018	152,102,300	122,394,760	0.2581
		2019	159,451,194	122,391,405	
		2020	116,826,426	78,495,014	

패널모형을 이용하여 주산지별 단수를 예측한 결과, 외표본의 오차 수준은 전남 22.7%, 경북 34.4%, 경남 39.2%, 제주 116.6% 수준으로 분석되었다(표 12). 인공신경망의 추정결과와 비교해보면 전남과 경북 지역을 패널모형을 이용한 예측력이 높은 것으로 분석된 반면, 경남과 제주지역은 인공신경망모형을 이용하는 경우가 예측력이 높은 것으로 분석되었다. 이러한 결과는 지역별 생육 환경이 다르며, 양파 생산에 영향을 미치는 외부요인의 차이가 존재하기 때문으로 사료된다. 따라서 지역별 특성을 고려하여 적합한 모형을 선택하여 예측하는 것이 예측력을 향상하는 데 도움을 줄 수 있을 것으로 판단된다.

표 12. 패널모형을 이용한 주산지별 단수 예측 결과

지역	연도	실제값	실제값	예측값	RMSPE
전남	내표본	2018	4932.0	5267.7	0.2129
		2019	6873.0	5549.8	
		2020	7883.0	5461.7	
	외표본	2018	4932.0	5479.3	0.2269
		2019	6873.0	5267.7	
		2020	7883.0	5549.8	
경북	내표본	2018	6296.0	4887.3	0.2819
		2019	7471.0	5725.9	
		2020	8098.0	5135.9	

지역	연도	실제값	실제값	예측값	RMSPE
	외표본	2018	5974.0	7636.0	0.3426
		2019	6296.0	8396.9	
		2020	7471.0	10490.8	
경남	내표본	2018	6601.0	5188.9	0.2857
		2019	8170.0	5894.5	
		2020	8783.0	5720.0	
	외표본	2018	5974.0	8172.5	0.3922
		2019	6296.0	8654.4	
		2020	7471.0	10689.6	
제주	내표본	2018	6740.0	8825.6	0.2309
		2019	7875.0	9030.0	
		2020	7594.0	9163.8	
	외표본	2018	5974.0	13675.5	1.1658
		2019	6296.0	13674.5	
		2020	7471.0	15097.9	

4.1.3 수입모형 분석 결과

수입모형의 예측 결과 전반적으로 29% 이상의 예측력이 낮은 것으로 분석되었다. 이는 생산작황 및 가격 변화에 따라 수입량 변화가 발생하므로 양파의 생산 및 가격의 높은 변동성으로 수입량의 변동이 크게 나타나기 때문이다. 또한, 수입량은 국내 생산이 결정된 후 공급되는 형태로 시차 변화에 따른 변동이 추가로 발생하기 때문으로 판단된다.

표 13. 최근 3년의 수입량 예측 결과

(단위: 톤)

모형	구분	2018년	2019년	2020년	RMSPE (out-of-sample)	RMSPE (in-sample)
연평균 수입모형	예측치	113,049	108,473	68,942	48.34%	10.88%
	실제치	98,189	60,040	74,157		
	RMSPE	21.31%	80.67%	7.03%		
4월 수입모형	예측치	3,679	4,409	5,389	39.70%	27.75%
	실제치	7,894	6,942	7,077		
	RMSPE	53.17%	36.49%	23.86%		

모형	구분	2018년	2019년	2020년	RMSPE (out-of-sample)	RMSPE (in-sample)
6월 수입모형	예측치	5,978	3,785	3,605	29.35%	4.80%
	실제치	5,419	2,534	3,837		
	RMSPE	10.31%	49.41%	6.04%		
10월 수입모형	예측치	2,760	-313	13,371	77.08%	57.1%
	실제치	5,930	2,993	8,763		
	RMSPE	53.45%	110.47%	52.59%		

4.1.4. 재고모형 분석 결과

재고모형을 추정결과 바탕으로 표본 외 예측을 한 결과, 재고 예측치와 예측력은 <표 14>와 같다. 재고모형을 통해 2018년~2020년 월말 재고를 추정한 결과, 내표본과 외표본의 RMSPE의 차이가 크며, 이는 재고량을 추정해 설명하는 것은 가능하나, 예측하기엔 어려움을 의미한다.

표 14. 재고모형 예측결과 및 예측력 비교

모형	연도	실측치(톤)	예측치(톤)	예측력(%) 비교		
				RMSPE		김성우 외 (2016)
				외표본	내표본	
3월 말 재고	2018	22,000	46,716	72.066 (62.6288)	33.8500 (27.2186)	(10.9)
	2019	37,282	54,082			
	2020	34,439	44,936			
9월 말 재고	2018	600,010	580,276	3.2857 (3.1321)	2.1191 (2.0682)	(4.3)
	2019	632,089	605,141			
	2020	595,855	584,857			

주: 1) 외표본과 내표본 RMSPE에서 괄호 안의 수는 김성우 외(2016) 연구결과와 비교하기 위한 MAPE임.

2) 김성우 외(2016)는 중기선행관측모형과 농업관측본부의 양념채소 저장량 추정모형 중 MAPE가 낮은 값을 정리하였으며, MAPE는 내표본(In-Sample) 추정에 따라 계산됨.

4.2. 시계열 분석을 이용한 가격모형 분석 결과

시계열 분석을 이용한 각 모형의 분석 결과는 다음의 <표 15>와 같다. 4월 모형은 내표본, 외표본 예측력이 높은 AR(2) 모형을 선택하였다. 4월은 단경기로서 가격수준이나 변동성이 커질 수 있고, 재고량이 굉장히 중요한 역할을 할 것으로 4월 가격예측 시 재고량 등을 고려할 필요가 있다. 실제로 2013년 4월에 가격 이상치가 존재하여, AR(2)의 모형 설명력은 ARDL 보다 현저히 떨어졌지만, 예측력은 AR(2) 모형이 더 높은 경향을 나타냈다. 6월 모형은 둘 중 외표본 예측 오차가 낮은 AR(5) 모형을 선택하였다.

표 15. 최종 모형 선정을 위한 설명력, 예측력 비교

구분	모형	Adjusted R-squared	RMSPE	
			내표본	외표본
연평균	추세 시계열	0.324	38.321	49.49
4월	AR(2)	0.18	14.1343	17.41
	ARDL	0.774	18.4618	28.42
6월	AR(5)	0.616	49.3559	41.65
	ARDL	0.645	38.0793	61.91
10월	ARX	0.982	6.5546	7.72

10월은 양파의 내년도 재배면적을 결정하는 시기로, 다수의 선행연구에서 재배면적 결정 시 기대가격으로 전년도 양파가격이나 전년도 수확기 가격 등을 대리변수로 활용한 바 있다(이중웅, 1996; 정인철, 2019; 정다운, 2020). 양파 농업 관측사와의 인터뷰에 따르면 정식기 때, 농민들은 당시의 가격을 보고 재배면적을 결정하므로, 10월의 가격을 예측하는 것은 내년도 재배면적을 결정한다는 점에서 중요하다.

그러나 대부분의 양파 농가는 10월의 가격 예측을 바탕으로 내년도 가격을 예상하고 재배면적을 결정하지만, 월별 가격 분석 결과, 수확기인 6월의 가격 예측력이 떨어져 양파 농가들이 가격 위험 즉, 높은 경영 리스크를 갖고 있다고 볼 수 있다.

또한, 매년 1월 농업전망을 통해 한 해의 양파가격을 예측하지만, 본 연구의 연평균 가격예측 모형의 예측력이 가장 낮으며, 연평균 가격을 예측하는 것이 어려운 것으로 분석되었다.

연평균 가격예측 모형은 설명변수가 없고, 상수항과 추세만 고려해 설명변수로 추가적인 변수를 모색 및 도입하는 것이 필요하다.

기존 선행연구와 예측대상이 다르지만, RMSPE를 통해 예측력을 비교한 결과는 <표 13>에 정리하였다. 본 연구의 6월 가격 예측력과 남국현, 최영천(2015) 연구의 6월 도매가격 예측력과 비교할 수 있으며, 당년과 전년 생산량을 고려한 남국현, 최영천(2015) 연구의 예측력이 더 좋은 것으로 분석되었다. 이는 양파의 수확기인 6월의 가격이 당해연도 생산량 변동에 크게 영향을 받기 때문일 것으로 사료된다.

표 16. 가격예측 시계열모형 예측력 비교

본 연구				선행연구			
구분	모형	예측력		연구	모형	예측대상	외표본 예측력
		내표본	외표본				
연평균	추세 시계열	38.321	49.49	김배성 외 (2005)	ARMA (1,2)	2015년 1~7월 월별 도매가격	32.96
4월	AR(2)	14.1343	17.41	남국현, 최영천 (2015)	ARDL	2015년 6월 도매가격	29.7
6월	AR(5)	49.3559	41.65	하지희 외 (2019)	ARGARCH	2014년 1월~2018년 12월 월별 도매가격	14.00
10월	ARX	6.5546	7.72	한국농촌경제연구원 예측치			16.68

주: 1) 한국농촌경제연구원 양파가격 예측력은 하지희 외(2019) 연구에서 정리됨.

선행연구와의 예측력 비교와 한계에 따라, 최종가격 예측모형에 재고량이나 수요량을 추가로 반영했을 때 예측력이 향상되는지 검정할 수 있다. 전반적으로 외생변수를 포함함으로써 모형의 설명력은 높아졌지만, 외생변수인 수요나 재고, 생산량 또한 표본 외 예측을 통해 예측된 값을 가격 예측에 활용함으로써, 예측된 가격만 활용했던 모형보단 예측력이 낮은 것으로 분석되었다. 4월 가격 예측의 경우, 3월 말 재고나 당월 수요를 고려했을 때 4월 가격을 잘 설명한 데 반해, 예측은 어려운 것으로 분석되었다. 6월 가격 예측의 경우, 당년 생산량을 모형에 반영 시 표본 외 예측력이 높아졌으며, 이는 수확기 가격 예측에 있어 생산량의 예측이 중요함을 의미한다. 10월 가격 예측의 경우 당년 생산량이나 10월 수요를 고려했을 때 예측력이 높아지는 것으로 나타났다.

표 17. 외생변수 반영 시 최종모형의 예측력 향상 여부

구분	추가 외생변수	Adjusted R-squared	외생변수 추가 후 RMSPE	기존 모형 RMSPE
연평균	연간 수요	0.391	54.38	49.49
4월	3월 말 재고	0.482	30.24	17.41
	4월 수요	0.542	23.84	
6월	6월 수요	0.594	47.87	41.65
	당년 생산량	0.676	39.89	
10월	9월 말 재고	0.982	8.37	7.72
	10월 수요	0.974	7.25	
	당년 생산량	0.983	7.15	

4.3. 인공신경망모형을 이용한 가격모형 분석 결과

인공신경망모형을 이용하여 양파가격을 예측한 결과 연평균 가격 예측의 오차 수준은 32.3%이며, 10월 가격의 오차 수준은 12.2%인 것으로 분석되었다. 4월과 6월의 예측 오차 수준은 매우 높아 인공신경망모형이 적절하지 않은 것으로 분석되었다.

표 18. 인공신경망을 이용한 가격 예측 분석 결과

구분		연도	실제값	예측값	RMSPE
연평균 가격	내표본	2018	785.3	786.6	0.0292
		2019	575.4	574.4	
		2020	1018.8	1018.5	
	외표본	2018	785.3	821.2	0.3230
		2019	575.4	717.6	
		2020	1018.8	1036.0	
4월 가격	내표본	2018	751.8	750.0	0.0426
		2019	825.6	824.8	
		2020	1088.3	1083.4	
	외표본	2018	751.8	654.0	4.6811
		2019	825.6	818.1	
		2020	1088.3	784.5	
6월 가격	내표본	2018	672.6	673.0	0.0412
		2019	462.7	463.8	
		2020	668.8	668.9	

구분		연도	실제값	예측값	RMSPE
10월 가격	외표본	2018	672.6	677.3	2.1131
		2019	462.7	500.8	
		2020	668.8	736.0	
	내표본	2018	683.5	683.7	0.0351
		2019	494.6	492.6	
		2020	1125.2	1125.0	
	외표본	2018	683.5	698.9	0.1223
		2019	494.6	518.9	
		2020	1125.2	1078.5	

4.4. 예측 모형의 비교

구조모형, 시계열모형, 인공신경망모형을 이용하여 양파가격의 최근 3년(2018년~2020년)을 외표본으로 예측한 결과, 연평균 가격과 6월 가격은 구조모형으로 추정된 결과가 예측 오차 수준이 가장 낮은 것으로 분석되었으며, 4월과 10월의 경우 시계열모형을 이용하는 것으로 예측 오차 수준이 가장 낮은 것으로 분석되었다.

양파는 생산, 재고, 수출입뿐만 아니라 생장 단계에서도 기상변화에 영향을 받기 때문에 가격 형성은 다양한 외부요인에 의해 결정된다. 따라서 모형을 단순화한 시계열모형이나 인공신경망모형 보다는 다양한 요인을 고려한 구조모형으로 예측하는 것이 가장 예측력이 높이는 방법이라고 할 수 있다. 다만 데이터의 가용성, 주요 변수의 선정 등 분석 결과를 개선한다면 시계열모형이나 인공신경망모형도 양파가격을 분석하는데 적절한 분석 방법이라고 사료된다.

표 19. 모형별 예측 가격의 RMSPE 비교 분석

구분	연평균	4월	6월	10월
가격예측모형	12.1%	31.7%	19.6%	13.3%
시계열모형	49.5%	17.4%	41.7%	7.7%
인공신경망모형	32.3%	468.1%	211.3%	12.2%

5. 결론 및 시사점

본 연구는 양파가격의 높은 변동성으로부터 발생하는 국내 양파 농가의 피해를 감소하고 합리적인 의사결정에 도움을 주고자 양파가격의 예측력을 향상을 위한 다양한 분석을 시도하였다.

구조모형, 시계열모형, 인공신경망모형을 이용하여 양파가격의 예측력 수준을 비교 분석하였다. 또한, 단수 추정을 위해 패널데이터 분석, 인공신경망 분석을 하였으며, 지역별, 가용자료(생육 정보 등)별 차이를 두어 예측력을 높일 방법을 모색하였다.

분석 결과, 지역별 기후환경, 생육환경 등의 차이가 존재하여 지역마다 오차의 차이도 크게 나타나는 것으로 분석되었다. 또한, 양파가격은 다양한 외부 환경의 영향을 받으므로 구조모형을 통해 예측하는 것이 가장 적절한 것으로 분석되었다.

본 연구는 관측 정보가 제시되는 1월과 양파가 본격적으로 수확되는 4월과 6월 그리고 재배 농가의 생산의향이 결정되는 10월의 가격모형을 예측함으로써 양파 농가의 합리적인 경영 의사결정에 도움을 주고자 하였다. 또한, 시기별 가격에 영향을 미치는 주요 변수를 활용하여 가격 예측의 오차를 최소화할 수 있는 방안을 제시하였다는 점에서 의미가 있다. 또한, 양파가격 예측에 관한 연구가 부족한 상황에서 패널모형, 인공신경망모형, 시계열모형, 구조모형 등 다양한 방법론에 대해 비교분석을 하여 예측에 효율적인 모형을 찾는 노력을 시도하였다는 점에서 본 연구의 시도는 중요하다고 볼 수 있다.

지역별 생장 환경 및 기상 조건의 차이로 인해 지역별 예측 오차가 크게 차이가 나고, 다양한 외생변수를 고려하지만 예측력이 떨어진다는 점은 그만큼 양파가격을 예측하기가 어렵다는 것을 방증하는 결과로 볼 수 있다. 특히, 연평균 가격과 4월 가격 예측모형의 예측력이 떨어졌는데, 이는 연평균 가격이 예측하기 어려우며, 4월, 즉 전년도의 저장 양파와 올해의 햇양파, 조생종 출하가 함께 되는 시기에 시장이 혼란스러움을 의미한다.

본 연구는 다루는 연구 범위가 넓지만, 자료의 가용성과 제한된 일정으로 인해 생산, 재고, 수출입, 가격 분야에서 세부적인 결정 요인을 상세하게 확인하지 못했다는 점에서 한계점을 가진다.

양파가격 및 생산의 예측력을 향상하기 위해 한국농촌경제연구원에서 운영하는 양파의 생육 정보에 대한 지속적인 관측이 유지되어 다양한 정보를 제공해야 한다. 또한, KREI에서 운영하는 관측 모형에 대한 접근을 개방하여 관측 모형을 개선하기 위한 연구가 지속할 수 있도록 협력 체계가 마련되어야 할 것이다.

참고문헌

- 강연실, 강기춘. 2012. “제주지역 거시계량경제모형 개발에 관한 연구.” 「산경논집」 29: 27-49.
- 김동환 외. 2017. 「과학기술기반 채소류 수급유통고도화 사업단 최종보고서」. 농림축산식품부.
- 김명환, 박재민, 박준기, 서대석, 허주옥. 2000. 「주요 채소·과일의 수급함수 추정」.
- 김배성. 2005. “채소가격 예측을 위한 응용기법별 예측력 비교.” 「농업경제연구」 46(4): 89-113.
- 김성우, 송성환, 이형용, 한은수, 김서영. 2016. 「양념채소 저장 실태와 효율적인 수급안정체계 구축 방안 - 마늘·양파를 중심으로」. 한국농촌경제연구원.
- 김연중, 김배성, 송성환, 박영구. 2006. 「배추 계절별 수급모형 개발」. 한국농촌경제연구원.
- 김인석, 김라이, 변승연, 김화영. 2013. 「건고추·마늘 중기선행관측 모형 개발」. 한국농촌경제연구원.
- 남경수, 안병일. 2020. “명절기간 주요 국내 과일의 수요함수 추정 및 생산자잉여 변화: 배와 사과를 중심으로.” 「농업경영정책연구」 47(2): 225-249.
- 남국현, 최영찬. 2015. “양파 출하시기 도매가격 예측모형 연구.” 「농촌지도와 개발」 22(4): 423-434.
- 농촌진흥청. 2015. 「농업경영관리 길잡이① 양파 경영관리」.
- 문권순. 1997. “백터자기회귀(VAR)모형의 이해.” 「통계분석연구」 2(1): 23-56.
- 박기환, 박양구, 김재한, 윤종열. 2007. 「주요 채소류의 월별 수입수요함수 추정」. 한국농촌경제연구원.
- 박환재. 2008. “역수요모형과 한국육류시장의 수요분석.” 「경제학연구」 56(2): 5-37.
- 서홍석, 순병민, 김충현, 김문희. 2019. 「농업부문 전망모형 KREI-KASMO 2019 운용·개발 연구」. 한국농촌경제연구원.
- 서홍석, 김충현, 김준호. 2020. 「농업부문 전망모형 KREI-KASMO 2020 운용, 개발 연구」. 한국농촌경제연구원.
- 안병일. 2006. “주요 유제품의 수요함수 추정.” 「농업경영정책연구」 33(4): 1068-1084.
- 양승룡, 박유신. 1999. 「인공 신경망을 이용한 돼지가격 예측 모형 개발」. 「농업경제연구」 40: 105-127.
- 오승원, 김민수. 2017. “공간 및 기상정보 시계열모형을 이용한 양파 생산량 예측.” 「Journal of the Korean Data Analysis Society」 19(5): 2447-2456.
- 윤성주, 이춘수, 양승룡. 2016. 「인공 신경망을 이용한 국제곡물 가격 예측 모형 개발」. 「농업경제연구」 57(2): 83-101.
- 이동현, 이병서, 강석주, 김창환, 양성범, 이춘수, 석다솜, 홍인기, 한별, 박성화. 2018. 「주요 채소 단수 예측시스템 개발 연구」. 농촌진흥청.
- 이상민, 장철수. 2006. 「밤의 수급 전망모형 구축에 관한 연구」. 한국농촌경제연구원.
- 이승인, 손찬수, 이혜림. 2020. “기상 및 영농투입재 비용이 양파 단수에 미치는 영향:패널회귀분석과 베이지안네트워크모형을 중심으로.” 「농촌경제」 43(2): 1-28.
- 이중웅. 1996. “고추·마늘·양파의 생산결정요인 분석.” 「농촌경제」 19(3): 27-50.
- 이진면, 변창욱, 최용재, 김진웅, 이상호. 2007. 「KIET 산업경제계량모형」. 산업연구원.

- 이창선, 오영균. 2000. 「시계열모형을 이용한 원화환율 예측」. LG경제연구원.
- 이춘수, 이동현, 양성범. 2017. 「채소 단수 예측 매뉴얼」.
- 정인철. 2019. “양파 재배면적과 생산단수의 예측모형 비교-Stepwise, LASSO, PCA, PLS방법의 비교.” 전남대학교 대학원 수학/통계학과 석사학위논문.
- 정다운. 2020. “농민 재배의향을 반영한 양파 재배면적 증감률 예측.” 전남대학교 수학/통계학과 석사학위논문.
- 정민제, 최영찬. 2018. 「인공신경망 가격 예측 모델」. 「신유통 리서치」.
- 최상천, 백장선. 2016. “공간 패널 회귀모형을 이용한 양파 생산량 추정.” 「응용통계연구」 29(5): 873-885.
- 하지희, 서상택, 김선웅. 2019. “양파와 마늘가격 예측모형의 예측력 고도화 방안.” 「농촌계획」 25(4): 109-118.
- 한석호, 이정민, 박미성, 박영구, 장석진. 2010. 「중기선행관측 기본모형 개발연구」. 한국농촌경제연구원.
- 홍성욱, 신현수, 박성근, 김정현. 2017. 「국가별·경제권별 수출결정요인 분석을 통한 수출전망모형 구축 연구」. 산업연구원.
- 농림축산식품부. 1998~2019. 「1998~2019년 농수산물 도매시장 통계연보」.
- 농림축산식품부. 1987~1996. 「1987~1996년 농림수산 주요통계」.
- 농림축산식품부. 1997~2007. 「1997~2007년 농림업 주요통계」.
- 농림축산식품부. 2008~2012. 「2008~2012년 농림수산식품 주요통계」.
- 농림축산식품부. 2013~2020. 「2013~2020년 농림축산식품 주요통계」.
- 한국농촌경제연구원 농업관측본부. 매월. 「양념채소(양파)」.
- 한국농촌경제연구원. 2021. 「농업전망 2021(2권) : 코로나19 이후 농업·농촌의 변화와 미래」.
- 농촌진흥청. 2018. 「농업기술길잡이 96(개정판) 양파」.
- 농촌진흥청. 2012. 「FTA 대응 품목별 경쟁력 제고 대책⑥(양파)」.
- Damodar Gujarati. 2016. 「예제를 통한 계량경제학 제2판」.
- aT KAMIS. <<https://www.kamis.or.kr>>.
- aT 도매유통 정보시스템. <<http://at.agromarket.kr/>>.
- FTA강국KOREA. <<https://www.fta.go.kr/>>.
- KREI OASIS. <<http://oasis.krei.re.kr>>.
- KOSIS 국가통계포털. <<https://kosis.kr/>>.
- KDI 경제정보센터. <<https://eiec.kdi.re.kr/>>.
- TRADLINX. <<https://www.tradlinx.com/>>.
- 관세청. <<https://www.customs.go.kr/>>.
- 관세정보법령포털(CLIP). <<https://unipass.customs.go.kr/>>.
- 농넷. <<https://www.nongnet.or.kr/>>.
- 농사로(농업정보포털). <<https://www.nongsaro.go.kr/>>.
- 무역협회 K-stat. <<https://stat.kita.net/>>.

수출입무역통계. <<https://unipass.customs.go.kr/>>.
위키피디아. <<http://en.wikipedia.org/>>.
통계청. <<https://kostat.go.kr/>>.
통합무역정보서비스(TradeNAVI). <<http://www.tradenavi.or.kr/>>.
한국은행 경제통계시스템. <<https://ecos.bok.or.kr/>>.
KREI 농업관측본부. <<http://aglook.krei.re.kr>>.
통계청aT KAMIS.



제1회

농산물 수급 예측모형 경진대회

우수논문집

Chapter 1
지정과제



장려상

단기 수급조정행위를 고려한 양파 수급예측모형 개발

이상현 · 홍성민

단기 수급조정행위를 고려한 양파 수급예측모형 개발

이상현* 홍성민**

Key words

머신러닝(Machine learning), 중기구조모형(Medium-term structural model),
유향 비순환 그래프(Directed acyclic graph)

ABSTRACT

This study developed a model to increase short-term predictability by reflecting the short-term behavior in the basic structural model. The medium-term forecasting structural model has the advantage of being able to predict the production of major crops early, but it has the disadvantage that it is difficult to consider short-term changes in climate and distribution. Conversely, time-series models and machine learning have relatively high short-term predictability for climate factors and price fluctuations, but have a disadvantage in that the theoretical basis is poor or cause analysis is difficult. This study complemented the disadvantage of the existing machine learning by applying the directed acyclic graph analysis method in the selection of data. Also, outputs of the medium-term forecasting structural model has used for short-term onion price prediction. As a result of model development and verification in this study, it was found that the model developed in this study is superior in predictability to measurement through the existing machine learning method. It is hoped that the results of this study will be usefully adopted in the development of agricultural product price prediction models in future.

차례

1. 서론
2. 양파 수급예측모형 개발
3. 결론

* 강원대학교 농업자원경제학과 교수

** 강원대학교 농업자원경제학과 석사과정

1. 서론

농산물은 비농산물과 비교하였을 때 높은 가격 변동성을 가지고 있는 것이 특징이다. 가장 큰 원인은 소비가 상대적으로 가격에 대하여 비탄력적이기 때문이며, 두 번째는 태풍, 홍수, 가뭄 및 냉해와 같은 이상 기후 등 생산과정에서 불확실성을 야기하는 요인들이 많기 때문이다. 이러한 농산물의 높은 가격 변동성은 소비자들에게는 구매 부담을 생산자에게는 소득안정에 대한 위협을 증가시킨다. 따라서 정확한 가격예측은 정부의 농산물 공급 조절을 위한 정책적 대응을 가능하게 한다는 점에서 중요하다고 할 수 있다.

양파는 대표적인 양념 채소 중 하나로 지속적인 월별 소비를 보이는 품목이나, 생산은 4~6월이 수확기에 이루어지고 노지에서 연중 생산은 불가능하므로 주로 저장을 하고 월별 가격 동향에 따라 출하량이 결정된다. 저장에 의한 출하 시기의 조정이 이루어지고 있음에도 양파의 가격 변동성은 지난 10년간 지속해서 증가하고 있으며, 이에 가격 변동이 큰 대표적인 품목 중 하나로 인식되고 있다(하지희 외, 2019). 가격 변동성을 완화하기 위한 대책이 강구되어 왔지만, 실효적인 효과를 거두고 있지 못하는 가운데 양파 가격을 보다 정확하게 예측하는 것에 대한 필요성은 커져 왔다.

높은 양파의 가격 변동성으로 인하여 농림축산식품부, 지자체, 유통업체, 생산 농가는 각기 다른 방법으로 가격 변동에 대응한다(하지희 외, 2019). 농림축산식품부는 수매, 산지폐기, 수입 결정 등을 통하여 가격 안정화를 달성하고자 하고 있으며, 일부 지자체에서는 양파 생산 농가를 보호하기 위하여 농산물 최저가격 보장사업을 실시하고 있다. 유통업체에서는 할인 판매를 하거나, 직거래장터를 통한 소비촉진을 유도한다. 생산농가는 생산자단체를 통한 작목전환, 판매 증대, 계약재배 증대 등으로 대응하고 있다.

그런데도 2018년부터 2020년까지 매년 2~4만 톤의 양파를 산지폐기 해왔다. 미국의 경우 주요 품목별로 생산자단체가 존재하고 전국 단위로 생산량을 사전적으로 조절하고 있지만, 우리나라의 경우에는 그러한 기능을 수행하는 단체가 부재하여 전국적인 물량 통제에 어려움을 겪고 있다. 따라서 전국 수준의 재배면적과 파종량을 측정한 이후에 가격을 예측하고 가격이 매우 높을 것으로 예상될 경우 시장격리 조치를 시행하고 있다. 따라서 재배면적과 파종량

에 대한 측정에 오류가 있을 경우에는 생산량을 잘못 예측할 가능성이 높다. 또한, 기후에 따른 작황 변화가 높아 재배면적과 파종량을 정확하게 예측하더라도 생산량이 잘못될 가능성이 높다. 특히 양파의 경우 여타 채소 작물에 비하여 저장성이 양호한 편이고 이에 4~6월에 수확되는 양파는 저온 저장고에 보관된 다음에 출하되기 때문에 월간 가격예측에 더욱 어려움을 겪고 있다. 전국에 다수의 저장창고가 흩어져있으며, 실시간 유통 현황도 파악이 어려워 시기별 저장량이 얼마나 되는지, 가격 변동에 따라 출하되는 양이 어떻게 될 것인지 예측하기가 상당히 어렵다.

따라서 양파의 가격예측 정확성 제고가 매우 중요하다. 한국농촌경제연구원에서는 양파 수급 동향을 분석하고 관측 월보를 통하여 매월 가격을 예측하고 있다. 하지만 이상 기후의 빈도 증가, 유통과 저장 단계에서의 복잡성으로 인하여 점차 가격예측이 어려워지고 있다. 따라서 빅데이터나 ICT 등의 첨단기술을 예측모형에 접목해 기상변화 등에 따른 작황을 보다 정확하게 예측할 수 있도록 하는 노력이 필요하다.

기존의 주요 농산물의 수급 및 가격예측과 관련 모형은 크게 구조모형과 시계열모형 두 가지로 분류할 수 있다. 구조모형이 주로 사용되는 중기선행관측모형은 대표적으로 한국농촌경제연구원에서 운용 중인 KREI-KASMO가 있으며, 시계열모형은 오승원 외(2017), 이형용 외(2017), 김배성(2005) 등에 의하여 개발 운영된 바 있다. 중기선행관측모형은 주요 작물의 생산량을 조기에 예측할 수 있다는 장점이 있으나, 기후, 유통 등에서의 단기 변화를 고려하기 어렵다는 단점이 있으며, 시계열모형은 기후요인, 가격 변동 등에 대한 단기 예측성은 상대적으로 높으나 이론적 근거가 빈약하고 원인 분석이 어렵다는 단점이 있다.

농산물 가격을 분석한 기존 연구들은 가격 변동 자체만을 분석하거나 생산량과 수입량 변수를 활용한 가격함수 추정이 이루어졌으나, 최근에는 도매시장과 반입량 변수를 이용한 연구들이 이루어지고 있다. 하지만 아직 농산물 가격 변동의 원인이 되는 반입량, 수입량, 수요량을 종합적으로 고려한 분석은 이루어지지 않고 있다(남국현 외, 2017).

강태훈(2007)은 저장성이 약한 배추와 저장성이 강한 감자, 양파 품목을 일일 도매가격과 반입량 간의 관계를 VAR 모형으로 추정한 바 있으며, 분석결과 양파는 출하자들이 1~2일 전까지 반입량에 관심을 두고 있고 배추와 다르게 반입량이 가격을 주도하는 것으로 보는데 이런 이유를 저장성과 관련이 있어 보인다고 주장하였다. 김라이(2015)는 기존의 모형과 다르

게 실제 시장 상황을 반영하는 월간 자료들을 새롭게 생성하여, 생산자, 유통/저장업자, 소매상의 월간 의사결정과정을 모형에 반영하여 마늘 중기선행관측모형을 추정하였다. 추정결과 정부의 마늘 비축물량 방출은 해당 월 가장 크게 나타났으며, 그 후 2~3개월간 영향을 받으나 장기적인 가격 안정 효과는 없다고 판단하였다. 남국현 외(2017)는 배추, 양파, 마늘의 도매가격과 반입량, 수입량 및 수요량의 상호관계를 분석하기 위해 VAR 모형을 이용하여 도매시장가격과 반입량, 수입량 및 수요량의 관계를 추정하였다.

최근 머신러닝의 발전으로 경제학에서도 머신러닝 기법을 활용하고자 하는 연구가 활발하게 진행되고 있다. 특히, 개인 컴퓨터에서도 머신러닝 작업을 구현할 수 있을 정도의 높은 컴퓨팅 성능과 사용자 친화적인 통계패키지가 일반화되었다(정재현, 2019). 신성호 외(2018)는 쌀, 시금치, 애호박, 대파, 양파 등 5가지 농산물을 대상으로 대도시에서 거래되는 소매가격을 자연재해 발생에 따른 농산물 가격 변화를 딥러닝 알고리즘을 통해 예측한 바 있다. LSTM 네트워크를 적용한 결과, 모델의 정확도는 백분율 기준 93%로 높은 정확도를 보이는 것으로 나타났다.

본 연구는 기본적인 구조모형에 (출하연도의) 반입량, 재고량 등을 이용하여 생산 이후 유통과정에서의 수급조정 행동 변화를 모형에 반영하여 이를 월별로 추정하고 단기 수급조정행위를 학습하는 머신러닝을 결합하여 단기 예측성을 높일 방안을 고안하였다. 본 연구는 유향비순환 그래프(DAG) 분석 방법으로 선정된 단수 예측 및 출하량 예측에 관련된 데이터를 학습시킴으로써 기존의 머신러닝의 단점으로 지적된 이론적 근거가 빈약하거나 원인 분석이 어렵다는 점을 보완하였으며, 중기선행관측모형의 예측결과 또한 단기 양파 가격예측을 위한 머신러닝에 활용하였다. 분석에 사용된 모델은 LSTM(Long short term memory)이다. LSTM(Long short term memory)란 RNN(Recurrent Neural Networks)의 단점을 보완한 모델로 시계열 분석과 과거 데이터를 통한 미래 예측에 많이 사용된다. 모형 개발 및 검증을 통하여 본 연구에서 개발된 모형이 기존의 머신러닝 모형 구축방법을 통한 예측보다 예측성에서 우월함이 있는지 확인하고자 하였다. 본 연구의 경우 시간과 예산의 제약으로 완성도 높은 모형을 개발하는 것에 집중하는 대신 기존 방법보다 개선된 모형 구축 방식을 제안하는 것에 초점을 두고 진행하였다.

2. 양파 수급예측모형 개발

본 장의 구성은 다음과 같다. 제2.1절은 중기선행관측모형을 설명한다. 제2.2절은 단기가격예측을 위한 머신러닝 모형에 대하여 설명한다.

2.1. 양파 수급 중기선행관측모형 개발

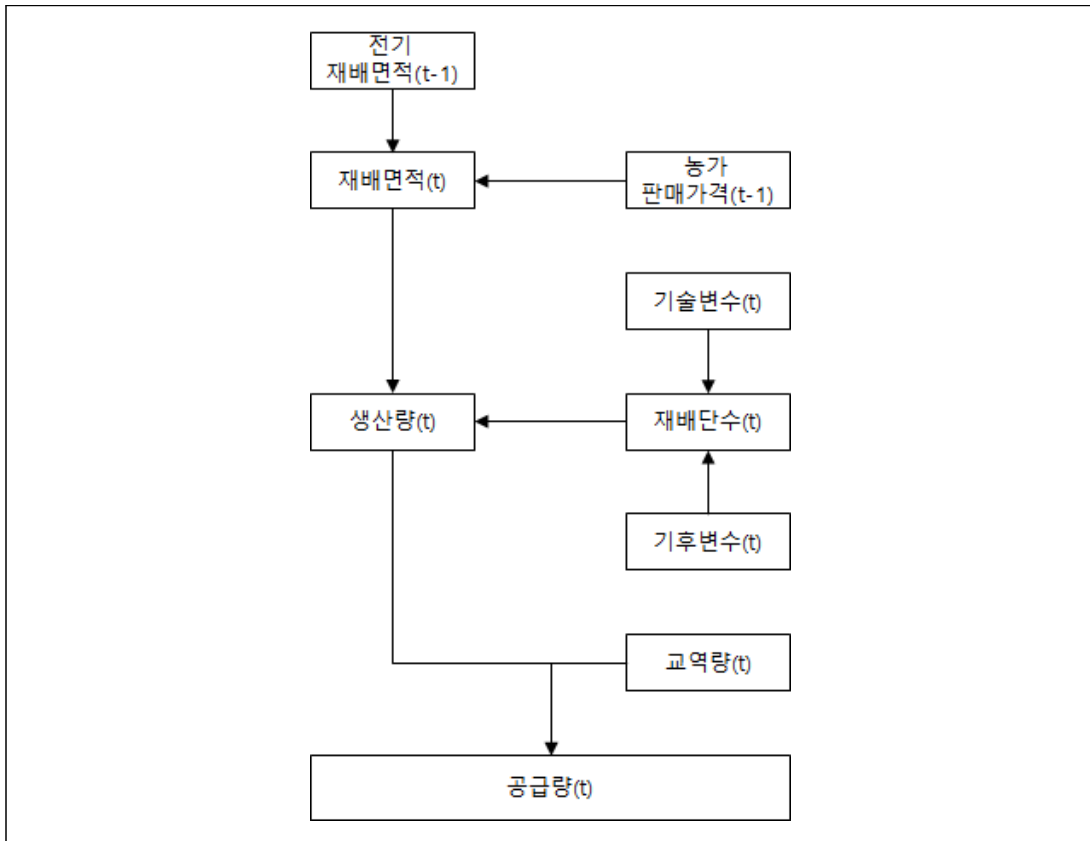
중기선행관측모형은 연간 자료를 활용하여 공급부문과 수요부문을 구축한 축차적 부분균형 구조모형으로 개발하였다. 주요 거시변수는 외생적으로 처리하였으며, 연간 단위의 중기예측을 위한 모형으로 설계하였다.

2.1.1. 공급부문 추정

양파의 생산량 추정은 재배면적 추정과 단위 추정의 두 가지로 수행할 수 있음. 양파의 경우 1년 단위로 경작이 이루어지는 작물로서 재배면적 추정에 전기 재배면적과 전기 양파 농가판매가격이 활용될 수 있음. 단수의 경우 기술개발 정도를 나타내는 변수와 기후변화를 반영한 변수를 활용하여 추정하였다. 따라서 생산량은 재배면적과 단수의 곱으로 산출할 수 있으며, 여기에 교역량이 반영되면 연간 양파 공급량이 도출될 수 있다(그림 1).

양파의 재배면적은 일반적으로 경지 반응함수(acreage response function)를 추정하며 가격보다는 기대 순수익(expected net returns)을 함수에 포함하는 것이 보다 바람직하나, 기대 수익에 대한 자료 확보가 어려울 경우, 기대 수익의 대리변수로 농가판매가격을 활용하기도 한다. 과일과 같이 식목 후 수확까지 다년의 기간이 필요한 품목은 식목하는 시점에서의 재배면적과 생육 기간이 고려된 다년간의 농가판매가격이 활용되나, 양파의 경우 1년 단위로 경작이 이루어지므로, 전기 재배면적과 전기 농가판매가격으로 경지 반응함수를 추정할 수 있다.

그림 1. 양파 공급 추정 단계 모식도



양파 재배면적에 대한 추정 식은 다음과 같다. 보다 설명력 높은 모형 구축을 위해서는 경작에 있어서 대체관계에 있는 작물의 농가판매가격도 포함되어야 하나, 본 연구의 목적이 완성도 있는 양파 수급구조모형을 구축하는 것이 아닌, 기존 방법에 비해 예측성을 개선하는 방식을 제안하는 데 있기에 이를 생략하였다.

$$PA_t = f(FP_{t-1}, PA_{t-1}, Others)$$

PA는 재배면적, FP은 농가판매가격, Others는 기타 변수, t는 연도를 의미한다. 추정결과는 다음과 같다.

$$PA = -431040 - 0.413PA(-1) + 5.909FP(-1) + 4851.7HAD1 - 3354.2HAD2 + 218.1HAD3$$

(198097) (0.187) (2.350) (1539.4) (2177.9) (100.4)

(Adjusted R-squared: 0.83)

주: 괄호 안의 숫자는 표준편차를 의미함.

단위 수확량 추정에서 기술 발전의 정도를 반영하는 대리변수로 추세를 사용하는 것이 일반적이며, 기후 변수는 이상 기후가 많았던 연도를 반영하여 생성 및 본 추정에 활용하였다.

$$YLD_t = f(Trend, WeatherIndicator, Others)$$

YLD는 단위 수확량, Trend는 추세, Weather Indicator은 기후 관련 변수를 의미한다. 순수입량을 나타내는 교역량 함수는 국내 양파 도매가격과 추세를 활용하여 추정하였다. 추정결과는 다음과 같다.

$$YLD = -671797.3 + 340.6Trend - 286.4YDD$$

(105533) (53.16) (58.05)

(Adjusted R-squared: 0.80)

주: 괄호 안의 숫자는 표준편차를 의미함.

$$M_t = f(WSP_t, Trend, Others)$$

M은 순수입량, WSP는 국내 양파 도매가격, Trend는 추세 변수를 의미한다. 추정결과는 다음과 같다.

$$M = -5049197743 + 93064.3WSP + 2501806Trend + 111860066MD$$

(1055780627) (16159.9) (524946) (17535520)

(Adjusted R-squared: 0.84)

주: 괄호 안의 숫자는 표준편차를 의미함.

2.1.2. 수요부문 추정

수요는 일반적으로 자체 가격, 대체재 가격, 소득, 인구수 등을 이용하여 추정한다. 본 연구에서는 보다 단순한 형태로 양파 수요를 역수요함수 형태로 추정하였으며, 자체 가격인 양파 도매가격, 추세만을 고려하였다.

$$WSP_t = f(CON_t, Trend, Others)$$

WSP는 국내 양파 도매가격, CON는 양파 소비량, Trend는 추세 변수를 의미한다. 추정결과를 다음과 같다.

$$WSP = -29432.6 - 0.00008445 CON + 15.505 Trend + 306.6 COND1 - 252.7 COND2$$

(11505) (0.00000147) (5.805) (56.81) (70.59)

(Adjusted R-squared: 0.82)

주: 괄호 안의 숫자는 표준편차를 의미함.

본 모형에서는 국내 양파 도매가격과 농가판매가격이 사용되었으며, 이에 두 가격 간 상관관계를 고려하는 가격연계함수 추정이 필요하다.

$$FP_t = f(WSP_t, Others)$$

추정결과는 다음과 같다.

$$FP = 129.03 + 0.553 WSP + 247.64 PLD$$

(40.52) (0.067) (102.04)

(Adjusted R-squared: 0.67)

주: 괄호 안의 숫자는 표준편차를 의미함.

2.1.3. 양파 월별재고량 도출

양파의 경우 저장성이 강해 중·만생종은 수확기 저온 저장하여 9월 하순부터 출하되며, 매년 전체 생산량의 약 50%~60%가 저장되어 출하되고 있다. 따라서 재고량이 가격에 많은 영향을 미칠 것으로 예상된다. 양파 월별재고량은 한국농촌경제연구원 양파 관측 월보 자료를 활용할 수 있지만 1월, 3월, 4월, 5월, 6월, 12월 자료는 기록되어 있지 않아 김라이(2015) 논문의 마늘 월별재고량 도출방법을 참고하여 월별 양파 재고량을 도출하였으며 도출과정은 다음과 같다.

가) 연산 연도 적용

양파의 경우 지역마다 날씨마다 조금씩 차이가 있지만 대부분 4~6월에 수확되어 즉시 판매되거나, 저장시설로 입고되는데 본 연구에서는 저장되는 물량이 다음연도 물량이 수확되기 직전인 5월까지 공급된다고 가정하였으며 본 연구에서 재고량을 도출하는 데 연산 연도를 당해 연도 6월에서 다음 해 5월까지로 정의하였다.

나) 월별 출하량

월별 출하량은 연간 생산량에 월별 가락시장 반입비율의 곱을 통하여 산출하였으며, 가락시장 반입비율은 연간 총 가락시장 반입물량을 월별도매시장 반입물량으로 나누어 비율을 바탕으로 월별 출하량을 산출하였다.

$$(1-1) \text{ 월별 출하량} = \text{월별 가락시장 반입비율} \times \text{연간 생산량}$$

다) 월말 재고량

본 연구에서 양파 월말 재고량 추정을 위해 농업관측본부의 농업관측 월보를 이용하였고 아래 <표 1>과 같이 자료의 연속성이 없어 빠진 자료에 대해서 같은 기준을 바탕으로 재고량을 추정하였다. 따라서 자료가 존재하지 않는 1월, 3월, 4월, 5월, 6월, 12월의 도출과정은 다음과 같다.

$$(1-2) 12월말 재고량 = 11월말 재고량 - 12월 출고량$$

$$(1-3) 12월 출고량 = 12, 1, 2월 출고량 합 \times \frac{12월 도매시장반입비율}{12, 1, 2월 도매시장 반입비율합}$$

식 (1-3)에서 도출된 12월 출고량을 식 (1-2)에 대입하여 각 월에 해당하는 재고량을 도출하였다.

여기서 5월 말 재고량은 당해 연도 재고 물량이 다음연도로 얼마나 이월될 것 인가에 초점을 맞추어 도출하였으며 3년 연산 연도 이동평균 생산량 대비 당해 연도 생산량 증가율을 기준으로 전월(4월 재고량)의 10%~50%까지 비율을 적용하였다. 또한, 도출된 재고량은 양파 조미 채소 관측 월보의 감모율 정보를 반영하여 최종 재고량을 도출하였다.

표 1. 양파 재고량(추정 전)

(단위: 천 톤)

	6월	7월	8월	9월	10월	11월	12월	1월	2월	3월	4월	5월
2000		420.4	386.9	380.0	277.0	235.0			49.0			
2001		446.0	446.0	450.0	450.0	359.0			61.2			
2002		425.0	425.0	396.0	262.9	211.2			31.4			
2003		391.9	387.3	357.5	239.3	200.7			40.5			
2004		441.0	441.0	401.5	256.9	225.0			44.7			
2005		474.3	478.2	442.0	287.3	252.1			74.0			
2006		473.4	473.4	428.4	298.8	230.3			57.3			
2007		537.9	522.2	480.6	317.5	249.4			93.1			
2008		503.5	508.6	445.1	368.2	298.4			79.1			
2009		570.1	578.8	506.6	351.6	272.0			90.9			
2010		565.5	553.7	458.3	382.5	289.8			94.1			
2011		733.1	710.0	593.5	503.4	399.4			146.3			
2012		428.0	567.5	529.1	458.5	379.1			87.8			
2013		621.5	537.0	478.0	395.0	395.0			133.0			
2014		770.5	729.4	625.0	550.0	452.0			147.4			
2015		556.4	518.3	435.0	365.0	277.0			64.0			
2016		629.0	596.0	498.0	438.0	327.0			78.0			
2017		626.0	523.0	425.0	389.0	304.0			81.0			
2018		672.0	637.0	549.0	477.0	372.0			94.0			
2019		737.0	666.6	566.7	490.5	374.1			87.0			
2020		673.0	631.6	530.5	450.4	344.4						

주: 1) 2004년 이전 자료는 엽근채소 관측 월보를 이용하였음.

자료: 농업관측본부(<https://aglook.krei.re.kr/>). 조미 채소 관측 월보

표 2. 양파 재고량(추정 후)

(단위: 천 톤)

	6월	7월	8월	9월	10월	11월	12월	1월	2월	3월	4월	5월
2000	421.8	420.4	386.9	380.0	277.0	235.0	146.3	97.7	49.2	37.1	14.3	1.4
2001	447.4	446.0	446.0	450.0	450.0	359.0	197.6	137.6	61.4	49.0	21.9	6.1
2002	431.1	425.0	425.0	396.0	262.9	211.2	124.1	77.4	31.5	25.0	8.0	3.4
2003	395.3	391.9	387.3	357.5	239.3	200.7	137.4	87.9	40.6	31.4	7.8	3.3
2004	444.3	441.0	441.0	401.5	256.9	225.0	160.7	89.4	44.8	37.7	11.4	6.1
2005	480.4	474.3	478.2	442.0	287.3	252.1	194.2	128.0	74.2	60.1	16.3	1.6
2006	475.0	473.4	473.4	428.4	298.8	230.3	166.1	121.6	57.5	45.2	8.7	2.0
2007	539.9	537.9	522.2	480.6	317.5	249.4	197.5	154.1	93.3	73.8	5.7	0.8
2008	504.3	503.5	508.6	445.1	368.2	298.4	219.1	157.7	79.4	59.1	10.9	5.9
2009	575.9	570.1	578.8	506.6	351.6	272.0	208.4	154.8	91.2	67.3	21.7	3.0
2010	568.5	565.5	553.7	458.3	382.5	289.8	219.5	160.5	94.3	66.2	19.6	5.5
2011	738.6	733.1	710.0	593.5	503.4	399.4	295.7	228.5	146.5	94.5	25.2	2.5
2012	430.5	428.0	567.5	529.1	458.5	379.1	237.0	173.3	88.0	67.5	26.9	2.7
2013	624.2	621.5	537.0	478.0	395.0	395.0	259.0	204.8	133.2	103.2	49.7	5.0
2014	775.5	770.5	729.4	625.0	550.0	452.0	306.8	227.2	147.6	114.7	50.1	5.0
2015	561.4	556.4	518.3	435.0	365.0	277.0	196.2	118.8	64.1	50.6	15.3	8.2
2016	637.2	629.0	596.0	498.0	438.0	327.0	240.3	127.6	78.2	62.4	18.5	10.0
2017	636.0	626.0	523.0	425.0	389.0	304.0	215.6	140.7	81.1	71.3	23.3	7.7
2018	679.7	672.0	637.0	549.0	477.0	372.0	271.8	189.5	94.2	80.5	22.2	8.4
2019	745.4	737.0	666.6	566.7	490.5	374.1	267.8	186.5	87.3	73.7	19.6	3.5
2020	676.5	673.0	631.6	530.5	450.4	344.4	216.9	117.3	42.0	31.8	8.3	0.8

주: 1) 2004년 이전 자료는 엽근채소 관측 월보를 이용하였음.

자료: 농업관측본부(<https://aglook.krei.re.kr/>). 조미 채소 관측 월보

2.1.4. 양파 반입량 추정

양파의 도매시장 가격은 수요와 공급의 균형가격으로 형성이 되는데 공급량에 해당하는 양파의 반입량은 가격이나 가격 등락에 직접적인 영향을 준다. 본 연구에서는 양파 월별 가격을 도출하는 것이 목적이므로 반입량 추정변수는 전기 재고량, 전월 양파 도매시장가격, 전전월 양파 도매시장가격, 전월 반입량을 고려하였다.

$$Q_t = f(S_{m-1}, WSP_{m-1}, WSP_{m-2}, Q_{t-1})$$

Q 는 반입량, S 는 재고량, WSP 는 도매시장 실질 가격 변수를 의미하며 t 는 년도, m 은 월을 의미한다. 반입량의 추정결과는 다음과 같다.

$$Q_t = 6153955.1 - 2.20S_{m-1} - 90.45WSP_{m-1} - 90.99WSP_{m-2} + 0.66Q_{m-1}$$

(6153955) (0.83644) (32.31) (31.66) (0.05)

(Adjusted R-squared: 0.49)

주: 괄호 안의 숫자는 표준편차를 의미함.

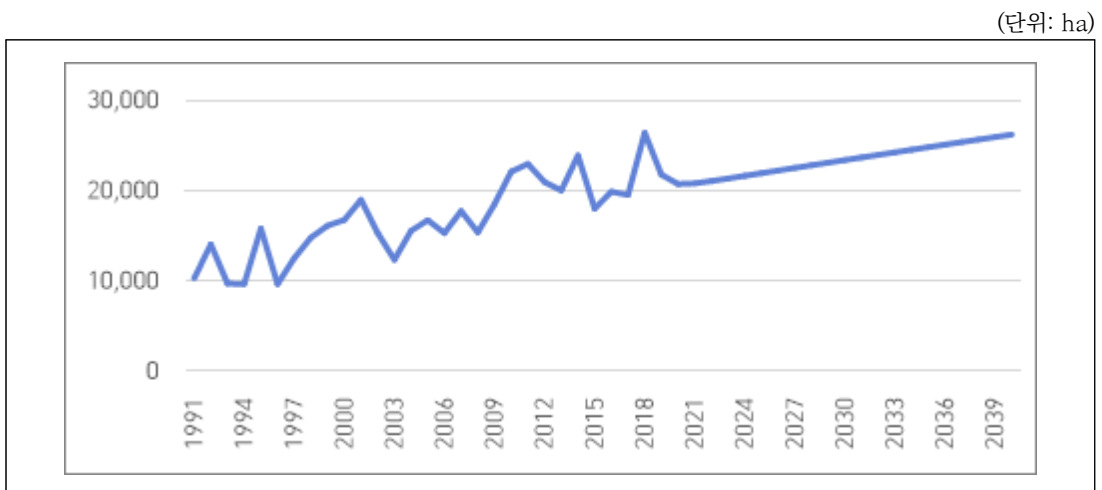
2.1.5. 시뮬레이션 결과

양파 중기모형을 통해 예측된 값을 머신러닝에 변수로 활용하기 위하여 재배면적, 단위당 수확량, 순수입량, 양파 도매가격을 추정하였다.

가) 양파 재배면적 추정결과

양파 중기모형의 시뮬레이션 결과는 다음 그래프와 같다. 2040년까지 재배면적을 중기모형을 통해 예측한 결과 전체적으로 우상향하여 2040년의 재배면적을 26,233ha로 추정하였다.

그림 2. 양파 재배면적 추정결과

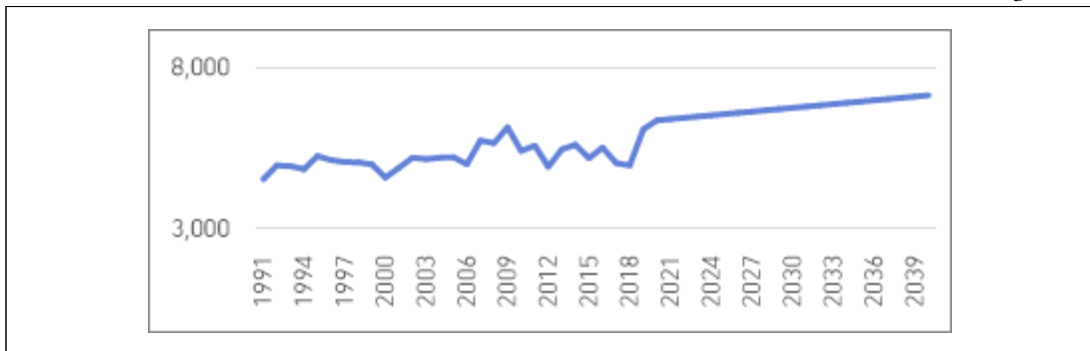


나) 양파 단수 추정결과

양파의 단수 추정결과 우상향하는 것으로 나타났으면서 이는 기술의 발전으로 인하여 수확량이 늘어날 것으로 판단된다.

그림 3. 양파 단수 추정결과

(단위: kg/10ha)

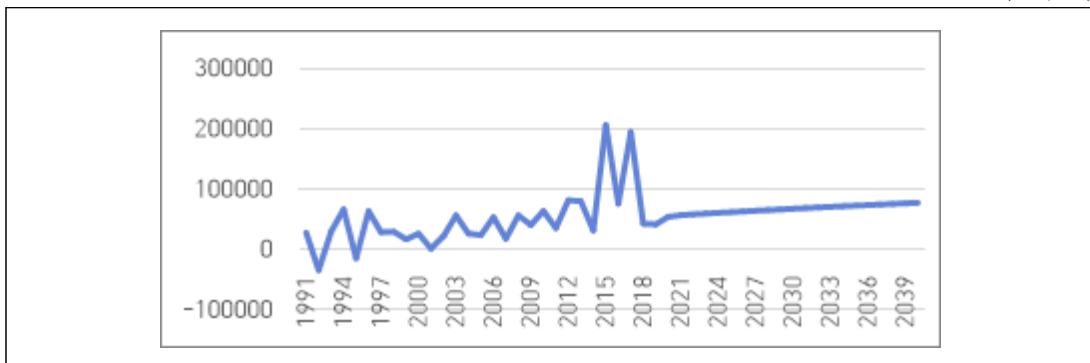


다) 양파 순수입량 추정결과

양파의 순수입량 추정결과 과거 1992년과 1995년을 제외한 모든 년도에서는 수출량보다 수입량이 더 많았으며 2020년 이후에도 우상향하는 것으로 보아 국내의 양파 수입량은 증가할 것으로 예상된다.

그림 4. 양파 순수입량 추정결과

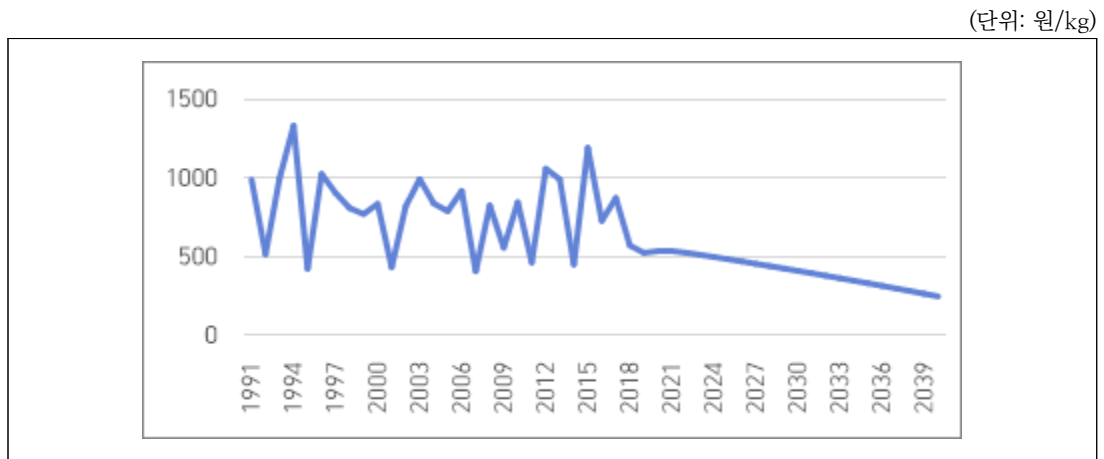
(단위: 톤)



라) 양파 도매가격 추정결과

양파의 도매가격 추정결과 우하향하는 것으로 나타났는데 이는 GDP 디플레이터를 반영한 실질 가격이고 도매가격이 우하향함에 따라 농가 순이익은 감소하고 있는 것으로 추정되었다.

그림 5. 양파 도매가격 추정결과



2.2. 양파 단기 가격예측 머신러닝모형 개발

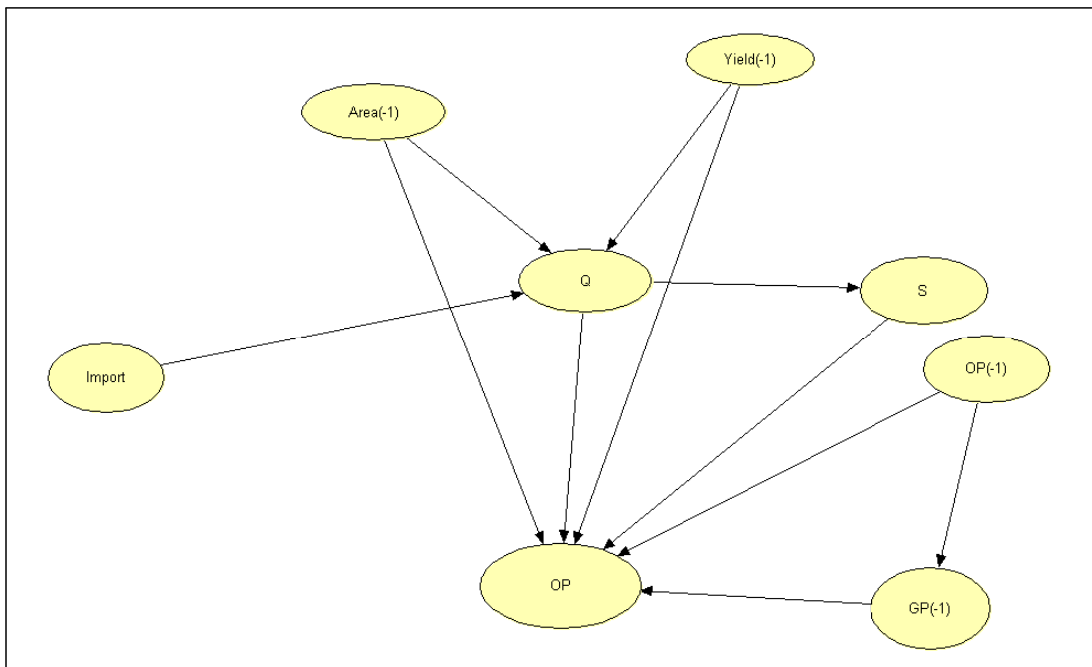
중기선행관측모형은 연간 자료를 활용하여 공급부문과 수요부문을 구축한 축차적 부분균형 구조모형으로 개발하였다. 주요 거시변수는 외생적으로 처리하였으며, 연간 단위의 중기예측을 위한 모형으로 설계하였다. 본 장에서는 양파 가격예측을 위한 변수와 머신러닝 알고리즘을 선정하고 1달 후 양파 가격을 예측하는 것을 목적으로 한다.

예측하는 데 앞서 추정하였던 단수, 재배면적의 값을 적용한 것과 적용하지 않은 경우를 비교함으로써 중기선행관측모형 변수를 머신러닝에 적용하고 예측물의 차이를 살펴보고자 한다.

2.2.1. 변수 선정

DAG 분석을 통해 파악한 인과관계는 시간의 흐름을 따르지 않는 비대칭적 인과관계(Non-time sequence asymmetry in causal relations)로서 동시적 인과관계(Contemporaneous causality)를 의미한다(권동건 외 1, 2017). DAG 그래프이론은 기본적으로 변수 간의 상관 계수(correlation) 및 편상관계수(partial correlation)를 이용하여 인과 흐름(casual flow)을 그림을 통해 간결하게 보여주는 방법이다(Pearl, 2000). 이번 장에서는 양파 가격에 영향을 미치는 변수들의 인과관계를 파악하기 위해 이용하는 유향 비순환 그래프(DAG)를 이용하여 변수 간 인과순위를 결정할 수 있게 한다. <그림 6>은 Hugin Expert Lite 8.9 통계패키지를 통해 양파 도매시장 가격이 결정되는 과정을 도식화한 것이다. 전년도 재배면적(Area(-1))과 전년도 단수(Yield(-1)), 수입량(import)은 반입량(Q)에 영향을 미치며 반입량(Q)은 양파 가격(OP)과 재고량(S)에 영향을 주게 된다. 재고량(S)은 양파 가격에 영향을 주게 되며 전월 가격(OP(-1))은 마늘가격(GP(-1))과 양파 가격(OP)에 영향을 준다.

그림 6. 양파 도매시장가격 결정과정



본 연구에서 양파 가격예측에 사용된 변수들은 앞선 연구들을 참고하여 변수에 이용하였다. 인공신경망 가격 예측모델(정민제, 2017), 주요 채소류 가격 단수 예측모형(류상모, 2019), LSTM 네트워크를 활용한 농산물 가격 예측모델(신성호 외, 2018) 등의 선행연구를 기반으로 변수를 설정하였다. 또한, 앞서 중기모형으로 추정한 재배면적, 단수 함수와 반입량 함수를 변수로 선정하였다.

분석에 사용된 변수의 기간은 1999년 1월~2021년 3월까지의 변수가 이용되었으며, 데이터 전처리 과정을 거쳐 실제로 분석에 사용된 변수는 2000년 1월~ 2020년 12월까지의 변수를 활용하여 분석을 진행하였다. 분석에 선정된 변수 정보는 다음과 같다.

표 3. 학습 데이터 및 출처

	세부변수	단위	출처
종속변수	가락 도매시장가격	원/kg, 상품	농산물 유통정보(KAMIS) <www.kamis.or.kr>
설명변수	전월 도매시장가격	원/kg, 상품	직접계산
	월평균 대비 편차 가격	원/kg, 상품	직접계산
	전월 가격 차이	원/kg, 상품	직접계산
	전월 반입량 차이	kg	직접계산
	가격이 동 평균 곡선	6개월 기준	직접계산
	전월 양파 수입량	톤	농업관측본부 관측 월보
	대체재(마늘) 가격	원/kg, 상품	농산물 유통정보(KAMIS) <www.kamis.or.kr>
	반입량(추정치)	kg	농업관측본부 자료를 활용하여 직접도출
	재고량(추정치)	kg	농업관측본부 관측 월보자료를 활용하여 직접도출
	재배면적(추정치)	ha	국가통계포털(KOSIS) 자료를 통해 직접도출
	단수(추정치)	kg/10a	국가통계포털(KOSIS) 자료를 통해 직접도출

〈표 3〉은 학습에 필요한 데이터들을 수집하여 나타낸 표이다. 수집된 데이터의 기간이 모두 다를 뿐만 아니라 과거의 데이터를 현재에 적용하는 경우가 많아 효과적인 분석을 위하여 공통으로 존재하는 기간의 데이터만 사용하였다. 따라서 2000년 1월~2020년 12월까지의 월별자료로 총 253개의 관측치를 이용하였으며 비어있는 값들은 최근 날들의 평균으로 채워 주었다. 재배면적과 단수와 같은 연간 데이터는 양파 가격에 영향을 미치는 시기를 고려하여 월별로 갖게 적용해 주었다.

2.2.2. 알고리즘 선정

양파를 포함한 농산물의 경우 가격은 수확기에 많은 영향을 받는다. 따라서 계절에 따른 예측이 이루어져야 하지만 한 계절 단위로 예측을 하기엔 너무 많은 기간을 가지고 있기에 한 달 단위 데이터를 가지고 예측을 진행하였다.

농산물을 포함한 대부분의 판매 제품의 가격은 기본적으로 과거의 가격 변화와 무관하지 않고 특별한 경우를 제외하고는 과거의 가격 변화에서 심하게 벗어나지 않는다(신성호 외, 2018). 머신러닝 모형의 경우 ANN, RNN, LSTM 등 예측을 위한 다양한 모델이 있으며, 예측률 또한 뛰어나다고 평가받고 있다. LSTM은 RNN의 한 종류로 90년대 중반 처음 등장하였으며 RNN에 비해 과거의 일을 기억하기 쉽게 만들어준다는 장점이 있다. 또한, LSTM은 시계열 데이터 분석에서 뛰어난 성능을 보인다. 따라서 본 장에서는 LSTM 알고리즘으로 단기 양파 가격을 예측하고자 한다.

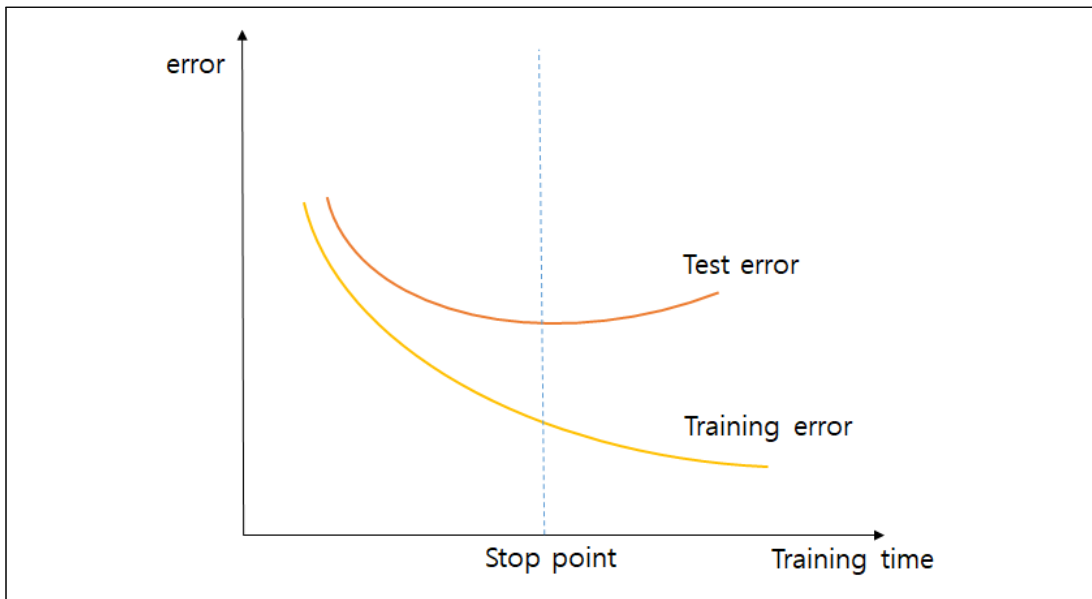
2.2.3. 분석과정 및 결과

머신러닝을 사용하여 데이터 학습을 통해 미래의 데이터를 예측해보았다. 분석은 과거 1년(12개) 데이터를 기반으로 한 달 후를 예측하도록 하였다. 인공지능이 잘 작동하는지 확인하기 위해서는 학습용 데이터(Training Set)와 검증용 데이터(Test Set)로 나누어 학습을 진행한다. 보통의 선행연구들에서는 두 데이터 간의 비율을 80%와 20%로 나누어 학습을 진행한다. 따라서 총 253개의 데이터 중 80%에 해당하는 200개 데이터를 학습용으로 20%에 해당하는 53개 데이터를 검증용으로 분류하였다.

보통의 경우 한 번의 계산으로 최적화된 값을 찾는 것은 어렵기에 머신러닝에서 최적화를 할 때는 많은 반복 학습 과정을 거치게 된다. 이때 epoch라는 용어가 사용되는데 epoch란 학습의 횟수를 의미한다. 특히, epoch의 수가 너무 많아지면 과적합(overfitting)의 문제가 생기기 때문에 이를 방지하기 위하여 <그림 7>의 그래프와 같이 test error가 최저가 되는 점에서 학습을 중단시키도록 하였다. 본 연구에서 epoch 수는 100으로 설정하였고 test error의 개선이 30번 이상 되지 않으면 학습을 중단하도록 설정하였다.

본 장에서는 중기선행관측모형의 추정된 값들을 머신러닝에 적용할 때와 적용하지 않을 때의 차이를 비교해 보고 모델의 예측값과 실제 값을 비교할 수 있는 평균 제곱근 오차(RMSE)에 로그값을 취한 RMSLE를 통하여 모델의 정확도를 측정해 보았다. RMSLE는 특정 수치에 대한 예측의 정확도를 표현할 때 이용되는 방법으로 아웃라이어에 덜 민감하고 상대적 오류를 측정해주며 예측값이 실제값보다 작을 경우 더 큰 패널티를 부여한다는 특징이 있다. 또한, 일반적으로 값이 0에 가까울수록 정확도가 높다고 판단된다.

그림 7. epoch와 과적합(overfitting)



가) 중기모형을 반영하지 않은 결과

우선 중기모형 값을 반영하지 않은 분석결과 <그림 8>과 같은 결과가 나왔으며 파랑 선이 실제 값이고 주황색 선이 머신러닝을 통해 예측한 결과값이다. 분석결과 예측을 잘하지 못하는 구간이 있는 것으로 나타났으며 검증손실 값(val_loss)은 0.1975로 나타났다.

그림 8. 중기모형을 반영하지 않은 예측가격

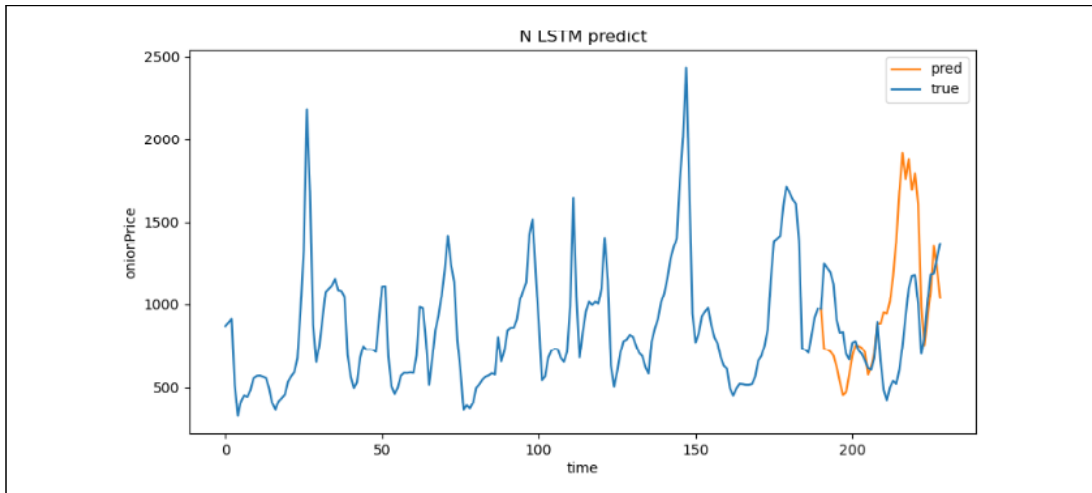
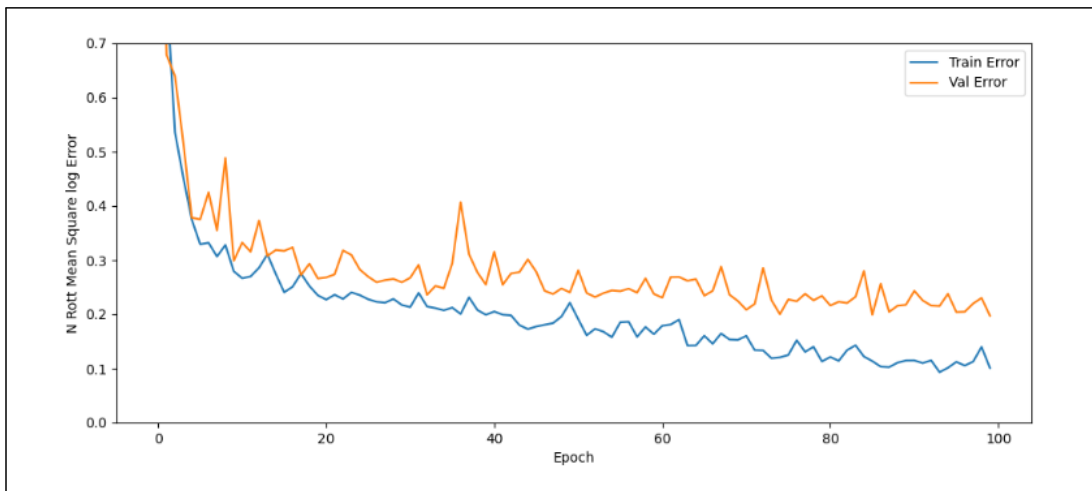


그림 9. 중기모형을 반영하지 않은 RMSLE



나) 중기모형을 반영한 결과

중기모형 값을 반영한 결과는 <그림 10>과 같다. 처음에는 예측을 잘하지 못하지만, 시간이 지날수록 실제 값과 비슷해지는 추세를 볼 수 있다. 또한, 검증손실 값(val_loss)도 0.0877로 나타났다.

그림 10. 증기모형을 반영한 예측가격

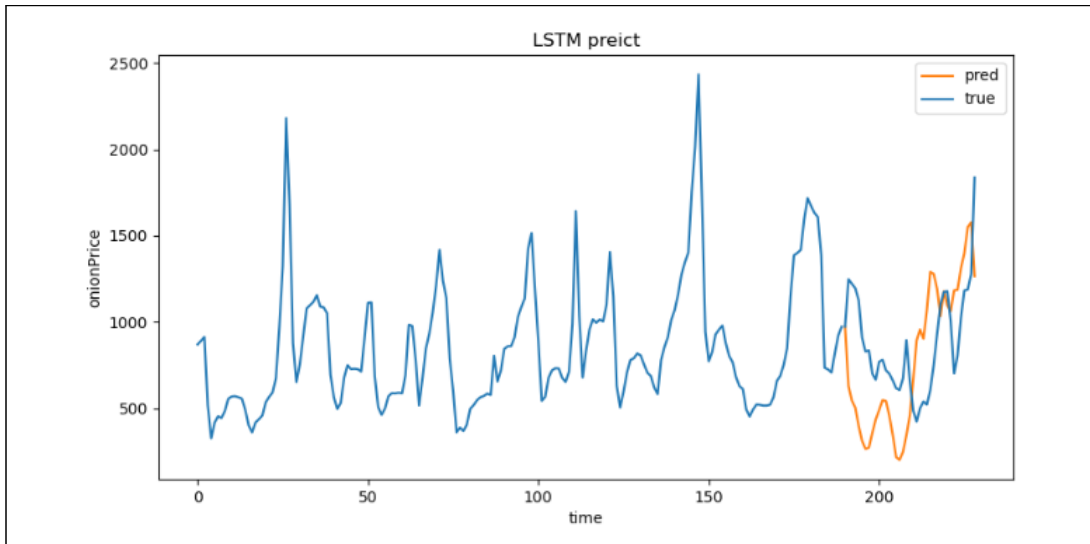
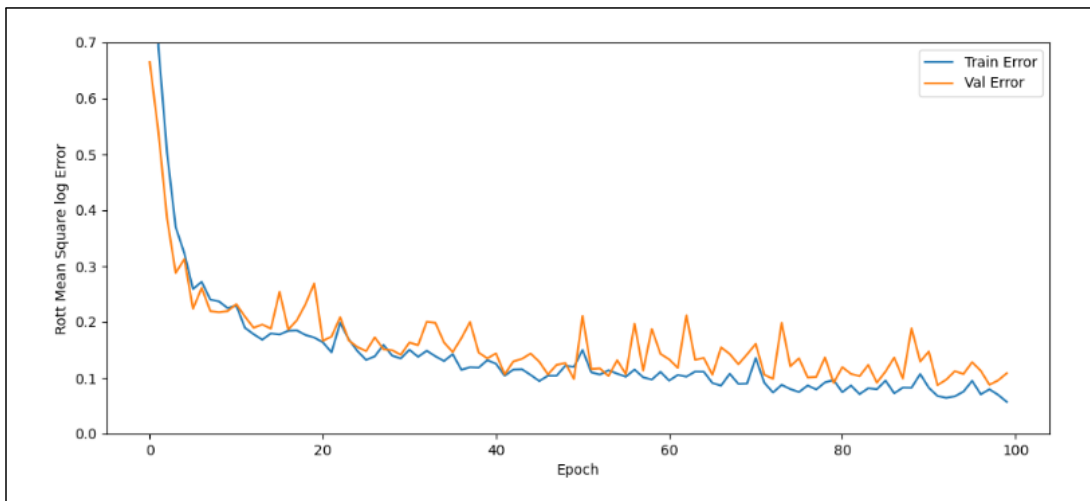


그림 11. 증기모형을 반영한 RMSLE



분석결과 증기모형을 반영한 머신러닝 방법이 그렇지 않은 방법보다 정확도가 더 높게 나타났다. 이는 증기선행관측모형에 단기 예측모형을 결합함으로써 예측력을 높일 수 있다는 것을 보여주고 있다.

3. 결론

본 연구에서는 농산물유통정보(KAMIS)에서 도매가격 자료와 통계청의 재배면적, 단수 농업관측본부의 재고량 농촌진흥청 농업통계연보의 자료를 이용하여 중기모형을 도출하였다. 중기선행관측모형의 시물레이션 결과 재배면적과 단수, 순수 수입량은 우상향하는 반면 실질 도매시장 가격은 우하향하는 것으로 나타났다. 또한, 월별 양파 가격에 영향을 미치는 요소인 재고량을 월별로 추정하여 이를 가격예측에 활용하였다. 양파 중기선행 관측모형을 구축하고 이를 월별로 재생산하여 단기 예측모형을 결합함으로써 중기 예측값을 기반으로 일반 단기모형과 중기선행관측모형을 결합한 모형의 예측성을 분석하였다. 분석결과 중기선행관측모형 추정치를 넣은 예측모형의 예측력이 더 높게 나타났다.

또한, 예측력을 높이려는 방안으로 기존의 계량모형을 대체할 수 있는 머신러닝 모델을 이용하였다. 향후 예측력을 높이기 위한 과제로는 양파의 품목 특성상 수확기인 4~6월 가격 하락 폭이 크게 나타나는데 수확기와 단정기를 구분하지 못했다는 점이 연구의 한계로 남아있다. 또한, 분석을 진행할 때 LSTM 알고리즘으로 분석을 진행하였는데 ANN이나 RNN 등 다양한 머신러닝 알고리즘을 적용하여 분석을 진행하는 시도도 필요할 것이다.

본 연구의 경우 시간과 예산의 제약으로 완성도 높은 모형을 개발하는 것에 집중하는 대신 기존 방법보다 개선된 모형 구축 방식을 제안하는 것에 초점을 두고 진행하였다. 본 연구 결과가 향후 품목별 예측성 높은 모형 개발에 기초자료로 활용됨으로써 농가소득 안정에 기여할 수 있기를 바란다.

참고문헌

- 강태훈. 2007. “도매시장 경락가격과 반입량의 상호관계 분석-배추, 감자, 양파를 중심으로.” 『농업경제연구』 48(4): 45-67.
- 권동진, 장우환. 2017. “DAG 모형을 활용한 농촌현장포럼 효과분석.” 『동북아관광연구』 13(1): 81-103.
- 김라이. 2015. “마늘 정부비축물량 방출이중·단기 마늘 시장에 미치는 영향.” 전남대학교 경제학석사 학위논문.
- 김배성. 2015. “채소가격 예측을 위한 응용기법별 예측력 비교.” 『농업경제연구』 46(4): 89-113.
- 남국현, 최영찬. 2017. “VAR을 이용한 도매가격, 반입량, 수입량 및 수요량의 동태적 상관분석.” 『농촌지도와 개발』 24(1): 9-19.
- 류상모. 2019. 「주요 채소류 가격 단수 예측 모형」. 농식품신유통연구원.
- 신성호, 이미경, 송사광. 2018. “LSTM 네트워크를 활용한 농산물 가격 예측 모델.” 『한국콘텐츠학회논문지』 18(11): 416-429.
- 오승원, 김민수. 2017. “공간 및 기상정보 시계열 모형을 이용한 양파 생산량 예측.” 『Journal of The Korean Data Analysis Society』 19(5): 2447-2456.
- 이형용, 여민수. 2017. “마늘 도매가격 시계열 예측 모형 비교.” 『농촌경제』 40(2): 55-74.
- 정민재, 최영찬. 2017. 「인공신경망 가격 예측 모델」. 농식품신유통연구원.
- 하지희, 서상택, 김선웅. 2019. “양파와 마늘가격 예측모형의 예측력고도화 방안.” 『농촌계획』 25(4): 109-117.
- Pearl. J. 2000. “Causal inference without counterfactuals: Comment.” Journal of the American Statistical Association 95(450): 428-431.



제1회

농산물 수급 예측모형 경진대회

우수논문집

Chapter 2
자유과제



최우수상

고랭지 배추 단수예측 추정

이병훈 · 김경덕 · 김명배 · 강형준 · 안용진

고랭지 배추 단수예측 추정 토양수분균형모형과 공간회귀모형 중심으로

이병훈* 김경덕* 김명배** 강형준** 안용진**

Key words

토양수분균형모형(Soil water balance model), 공간효과(Spatial effect),
유실곡선지수(Runoff Curve Number), 공간회귀모형(Spatial regression model)

ABSTRACT

This study established and estimated a singularity prediction model to predict the national production change of high-altitude Chinese cabbage by considering meteorological factors, soil factors, and spatially dependent factors of 16 main producing regions of high-altitude cabbage. In particular, the growth of Chinese cabbage in high altitudes is greatly affected by not only climatic environmental factors but also soil environmental factors, thereby affecting yield. In particular, it can be said that the moisture content of the soil among the factors of the soil environment is very important as it is a singular determinant. In this study, the soil moisture content was measured by using a soil water balance model to measure the moisture content of soil moisture during the growing period after planting of Chinese cabbage in high cold regions, and this was applied when constructing the singular model.

차례

- | | |
|------------------|------------|
| 1. 서론 | 4. 분석 결과 |
| 2. 연구의 이론적 배경 | 5. 요약 및 결론 |
| 3. 분석 모형 구조 및 자료 | |

* 강원대학교 교수

** 대아청과 기획팀장

*** 강원대학교 석사

**** 강원도농업기술원 연구관

1. 서론

고랭지 배추는 연도별 수확면적과 단수의 변동 폭이 가장 큰 품목 중 하나이다. 이는 공급량의 불안정성이 크기 때문에 배추 작기 중에서 가격의 변동성도 가장 큰 것으로 알려져 있다. 실례로서, 지난 2020년도에는 54일이라는 유례없이 긴 장마와 세 차례 잇따른 태풍의 영향으로 배추의 수확량 급감으로 고랭지 배추 가격은 사상초유의 폭등으로 이어진바 있다.

배추는 일반적으로 기온이 선선한 환경에서 잘 자라는 채소로써 여름철 고온 환경에서는 재배가 어려운 채소이다. 그러나 고랭지는 고도가 높을수록 기온이 차츰 낮아지는 기후 환경적 특징을 가지고 있어 여름철에도 배추 재배가 가능한 지역이다(농진청, 2019). 고랭지 배추는 말 뜻 그대로 고도가 높고(高), 온도가 찬(冷) 곳에서 생산되는 배추이고, 그래서 초여름부터 늦은 여름까지 3~4개월 짧은 생육기간 동안 강원도 등 400m 이상의 고원이나 산지 등의 고산 지역에서 생산된다. 이와 같이 고랭지 배추는 봄배추, 가을배추, 월동배추가 생산되는 저지대와 비교했을 때 재배 기간과 지역이 극히 일부 기간과 지역에 국한되어 있다. 따라서 이상 기상요인, 병충해, 가뭄, 폭우 등 기상이변에 의해 배추의 단위당 수확량이 변동했을 경우 여름철 배추 수급 및 가격 불안이 매우 클 수밖에 없는 특성을 보이고 있는 것이다. 특히, 최근 들어 이상기후가 빈번하고 심해짐에 따라 고랭지 배추 생산 변동성에 미치는 영향이 점차 확대되면서 과거보다 여름철 배추 수급 및 가격 불안에 대한 우려가 더욱 커지고 있다.

이에 따라 배추 생육에 지대한 영향을 미치는 기상 요인을 고려한 보다 신속하고 정확한 고랭지 배추 단수 예측의 필요성이 그 어느 때보다도 절실해 지고 있다. 만일, 배추 생산 예측이 신속 정확하게 이루어진다면, 특히, 배추 생산자, 산지유통인의 경우 생산 예측 정보를 활용하여 출하량 및 출하시기 의사결정시 적절히 활용될 수 있을 것이다. 아울러, 정부는 여름 배추 출하기 수급안정대책의 긴급 기초자료로서 활용가치가 매우 높을 것으로 보인다.

본 연구의 목적은 전국 고랭지 배추 주산지 16개 지역의 기상요인, 토양환경 요인, 지형 공간적 상관요인 등을 고려하여, 고랭지 배추의 전국 생산변화를 예측하는 단수 예측모형을 추정하는데 있다. 특히, 모형의 구조를 직관적으로 단순화하고 자료 갱신도 손쉽게 구성하여 관측센터 배추 관측 담당자 또는 주산지 시군 담당자들도 쉽게 모형을 다룰 수 있게끔 하고자 하

였다. 연구의 구성은 다음과 같다. 2장에서는 연구의 이론적 배경으로써 배추의 생육 특성과 기상요인 및 공간적 종속요인, 그리고 토양 내 수분 환경에 대해 살펴보고, 3장은 분석 모형 및 자료에 관하여 정리하며, 4장에서는 추정결과를 제시하고, 5장에서는 요약 및 결론으로 마무리 한다.

2. 연구의 이론적 배경

본 장의 제1절은 배추의 생육 관련 선행연구를 통해 생물학적 생육주기의 특징과 해당 주기별 최적의 생육 환경에 대해 살펴본다. 특히, 온도, 일조, 강수, 토양 환경 등 여건과 생육 간 관계를 면밀히 살펴본다. 이어서 제2절은 배추 주산지 간 상호 지정학적 연계성을 고려하기 위해 공간적 상관성에 대해 논의하고, 제3절은 여름철 수분을 많이 요구하는 고랭지 배추 재배의 특성상 관수가 매우 중요하기 때문에 토양의 생장유효 수분과 측정방법에 대하여 정리한다.

2.1. 고랭지 배추의 생육¹⁾

저온성 작물인 배추는 우리나라 여름의 경우 배추의 생육적기가 아니기 때문에 평지 재배 시 고온으로 인해 배추의 결구에 문제가 생긴다. 또한 강우량이 많으면 흑부병이 확산되고, 강우량이 부족하면 바이러스 발생하며, 특히 고온 환경이 지속되면 무름병이 극심하고 결구 불량률이 대량으로 발생하여 생산이 거의 불가능한 수준에 이른다. 따라서 고랭지에서 재배하는 가장 큰 이유는 평지에 비해 비교적 서늘하기 때문이다. 이에 고랭지 배추는 고랭지 여름배추라고도 부른다. 고랭지의 작형은 높이에 따라 여름배추를 다시 해발 400~600m의 준고랭

1) 농업기술길잡이 128, 농촌진흥청(2019)의 생육관련 내용을 참조함.

지 작형과 해발 600m 이상에서 재배하는 작형으로 나눈다. 따라서 여름배추의 주산단지는 해발 400m 이상의 강원도 평창, 정선, 태백 등의 지역과 전북 무주, 경북의 고랭지 일부 지역에 국한되어 있다. 고랭지 배추의 출하기는 봄배추 수확에 이어 7~9월에 집중적으로 이루어지며 재배 시기 및 지역의 한계 때문에 단수가 적고 시장출하가 불안정하여 가격 등락이 매우 심하다. 고랭지 배추의 생산현황을 지역별로 살펴보면, 강원도의 경우 전체 고랭지 배추 재배 면적의 95%를 차지하고 있으며 경상북도 및 충청북도의 고랭지에서 일부 재배되고 있다. 최근 기후 온난화로 인해 경북 및 충북 지역의 고랭지 배추 재배면적은 급속히 위축되었으며, 심지어 강원도 지역의 고랭지 배추 면적도 감소하는 경향을 보이고 있다.

2.1.1. 재배 환경

일반적으로 배추는 서늘한 기후를 좋아하는 채소로써 성장에 적합한 온도는 약 20℃ 내외이다. 결구시기에는 이보다 더욱 낮은 15~18℃ 정도가 적절하며, 최저 온도는 4~5℃ 정도로 알려져 있다. 생육초기의 경우 높은 온도에서 생장이 촉진되며, 이후 결구가 시작되면 오히려 고온이 결구 불량 생육이 낳게 된다. 배추는 강렬한 태양 아래에서 광합성 활동이 왕성해 지고 생육에 필요한 물질의 생성도 활발해 진다. 생육 초기 광을 최대한 충분히 받을 수 있도록 하고, 이후 결구가 시작되면 강한 광보다 약한 광이 생육에 유리하게 작용한다.

수분의 경우 배추의 구성성분은 대부분이 수분이고 단기간에 왕성한 발육이 되므로 절대적으로 많은 수분을 요구한다. 따라서 배추는 건조한 기후환경에 약하여 생육 초기에 물이 부족하면 잎 발생과 생육이 억제되어 수량이 급격히 감소하게 된다. 배추의 가장 좋은 생장기에는 하루 10a당 200kg 이상 체중이 증가하고, 수분이 가장 많이 필요한시기는 파종 후 40~50일 정도의 초기 단계이다. 또한 뿌리가 너무 젖으면 뿌리에 충분한 물을 유지하면서 배수가 잘되는 토양을 선택해야 한다. 배추의 생육 초기인 8~9월에는 수년 동안 계속되는 가뭄이 있으며, 이 경우 관개를 통해 물 관리를 한다.

표 1. 고랭지 배추 지대별 작형

작형	파종기	수확기	지대
초여름 재배	4월 하순 ~ 5월 중순	6월 하순 ~ 7월 하순	400~600m
한여름 재배	5월 중순 ~ 7월 상순	7월 하순 ~ 9월 중순	600m 이상
늦여름 재배	6월 하순 ~ 8월 상순	9월 중순 ~ 10월 중순	400~600m

자료: 농촌진흥청(2019)

토양의 경우 배추는 뿌리가 넓고 잔뿌리가 많기 때문에 흙이 깊고 배수가 잘되는 토양을 선호한다. 따라서 배추의 주산지인 보수력이 좋고 배수가 좋은 충적토이며 토양층이 깊은 사질양토 지역이다. 충적토에서는 배추가 잘 자라고 배추의 빠른 성장을 가능하게 하는 반면, 모래 토양에서는 초기 생장은 빠르지만 늦은 생장은 좋지 않아 잎에 황변이 나타난다.

2.1.2. 주요 기상 및 병해충 피해

배추의 구성 성분은 90~95%가 대부분 수분이며 단시간에 성장하는 특징이 있어 비교적 많은 물을 필요로 한다. 특히 정식 후 약 20~30일이 지나면 성장이 정점인 시기인데, 1일 10a당 200kg 이상의 무게가 증가하는데 이를 늘리기 위해서는 10a당 200kg 이상의 물이 필요하며, 이를 증발량 기준으로 10a당 하루에 3,000~4,000L에 이른다.

30℃ 이상의 고온과 가뭄이 2주 이상 지속되면 배추의 생체중이 현저히 떨어지고 결구 불량, 석회 결핍, 무름병이 발생한다. 결구기에는 성장을 위해 약물과 영양분을 뿌린다. 특히 밤에 드립 및 스프링클러와 같은 관개 시설을 사용하여 관리하여 물을 주면 토양에서 칼슘을 흡수하는 데 도움이 된다. 가뭄에 대처 방법으로 비닐 멀칭으로 토양 수분을 보존하고 토양을 개량하여 보수력을 높이고 뿌리를 관리하여야 한다.

2.2. 주산지 지역 간 공간적 연계성

행정구역 단위로 집계되는 고랭지 배추 단수와 기상, 병충해, 토양환경 등 지역 요인과 지역 농민의 성품, 언행, 정보력, 농업기술 등 지역 특색은 지형학적으로 긴밀히 연계되어 있다

(윤성은, 2013). 따라서 단수모형 추정 시 이와 같은 지역 간 공간적 연계성을 고려하여야 하나, 일반적으로 지역 간 자료는 독립적으로 간주하고 추정함으로써, 지역 간 공간적 인접 정보를 단수모형 추정 시 활용하지 못하고 있다. 따라서 공간적으로 연계되어있는 지역 간 자료를 고려하여 단수 모형 추정의 고도화를 통한 예측력 개선이 필요하다.

즉, 특정 지역의 공간에서의 지역요인 및 지역 특색이 해당 공간에만 영향을 미치지 않고 인접 지역이나 근거리에 위치한 생산자, 주민 등 의사결정자에게도 영향을 미칠 수 있는데 이를 ‘공간적 인접성, 종속성 또는 공간적 자기상관성이 발생한다’고 한다. 이 같은 공간적 인접 내지 종속 또는 자기상관 특성 고려하기 위해 Anselin(1988)이 제안한 공간회귀모형을 차용할 필요가 있다. 본 연구에서는 주산지 간 거리 기준의 공간가중치행렬을 이용한 공간회귀모형(Spatial regression model) 3가지 형태²⁾를 고려할 수 있다. 공간가중치 행렬(W)을 포함한 공간회귀모형 기본 식은 하여 다음과 같다.

$$(1) y = \lambda Wy + X\beta + \epsilon, \epsilon = \rho W_2\epsilon + \varepsilon, \varepsilon \sim N(0, \sigma^2 I) \quad W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1j} \\ w_{21} & w_{22} & \cdots & w_{2j} \\ \vdots & & \ddots & \vdots \\ w_{i1} & w_{i2} & \cdots & w_{ij} \end{bmatrix}$$

여기서 y 는 종속변수, X 는 설명변수의 행렬 그리고 $\varepsilon \sim N(0, \sigma^2 I)$ 는 오차항이다. λ, ρ 는 공간회귀계수를 의미하며 지역의 공간적 연계성 정도를 의미하며, 만일 $\lambda, \rho > 0$ 이라면 각 지역의 종속변수 및 오차항이 다른 지역의 종속변수 및 오차항들과 정(+)의 관계로 연계되어 있음을 의미한다. 추정방법 등 보다 자세한 사항은 이성우 외(2006)를 참조할 수 있다.

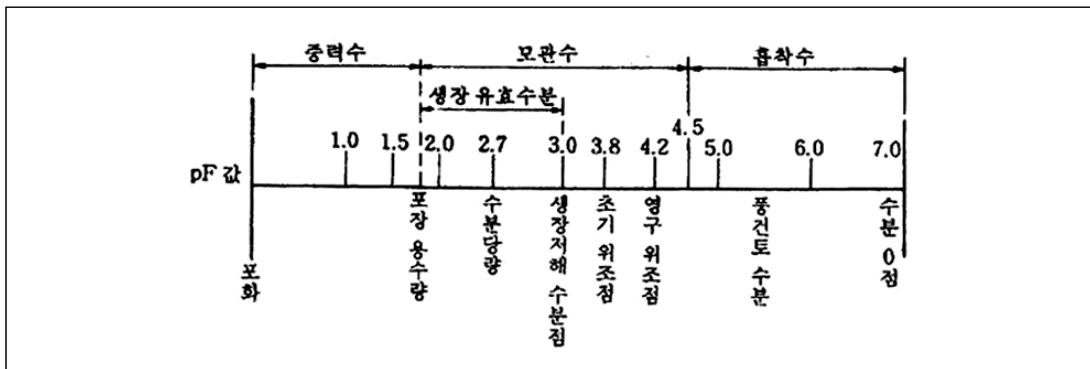
2) 공간자기회귀모형(SAR: spatial autoregressive model), 공간오차모형(SEM: spatial error model), 공간회귀오차모형(SAE: spatial autoregressive-error model) 등이며 공간가중치행렬(W)이 $W_2 = 0, W_1 = 0, (W_1, W_2) \neq 0$ 일 경우 각각 SAR, SEM, SAC모형이 적용된다. 공간가중치행렬은 각 지역의 공간상관관계를 나타내며 모든 지역이 인접성(근접성) 정도에 따라 가중치를 가지되 행 합 1이 되도록 하여 활용된다(이성우 외(2006)).

2.3. 토양의 생장유효 수분

고랭지 배추의 생육은 기후환경 요인뿐만 아니라 토양환경 요인에도 지대한 영향을 받아 단수에 영향을 주게 된다. 토양환경은 일반적으로 토양의 구조, 토성, 토양의 수분, 양분을 말한다. 배추는 단년생 작물이지만 상대적으로 뿌리가 길고, 잔뿌리가 많아 생육과정에서 표토층 수분 함유가 매우 중요하다. 앞서 살펴본 바와 같이 배추는 구성 성분이 대부분 물이고 짧은 기간 동안 왕성하게 자라며, 정식 후 20~30일 경과한 후부터는 배추가 흡수할 물 요구량이 급격히 증가한다.

배추 생육에 있어서 유효수분은 모관수의 포장용수량과 영구 위조점 차이의 수분을 의미하며 배추 생육에 중요한 수분을 말한다. 배추의 정상적인 생육을 보장하고 배추가 쉽게 이용할 수 있는 수분량을 '생장유효수분'이라고 하며 모관수의 포장용수량과 생장저해수분점 사이의 수분함량을 의미한다. 토양수분 측정기 기준 pF 1.8~4.2 사이가 생장유효수분에 해당된다. 참고로 배추는 pF 2 내외(포장용수량 75% 수준)에서 생육에 가장 적합하다.

그림 1. 토양수분의 분류 및 작물 이용성³⁾



자료: 농촌진흥청(2019)

3) 중력수는 토양 공극을 통해 밑으로 흐르는 물을 의미하고, 모관수는 토양 입자 사이에 존재하는 공극수를 일컬으며, 포장용수량은 최적 식물생육 물(水)량이고 이는 비온후 중력수가 빠져나간 다음 남은 물(水)량임(농진청, 2019).

2.3.1. 토양수분균형 모형(Soil water balance model)

고랭지 배추의 생육과정과 단수 결정에 있어서 토양환경 중 토양의 함수량은 매우 중요하다고 할 수 있다. 그러나 그동안 대부분 선행연구(한석호 외, 2011; 정호중 외, 2016; 최병옥 외, 2020)의 작물의 단수함수 구성은 최고온도, 최저온도, 강수량, 일조량 등 기후요인을 중심으로 단수 추정을 시도해 왔다. 그러나 고랭지 토양의 경우 토성이 보수력이 좋을 지라도 급격한 경사지라서 강수가 발생해도 모두 침투되지 않고, 그대로 흘러내리는 표면유실로 실제 강수량과 토양 수분의 함수량 간의 차이가 크게 발생해 왔다. 또한 해발고도가 높아 안개와 구름으로 일조량이 그대로 토양에 영향을 미치지 않아 토양의 증발량을 시산할 때도 곤란함을 겪는 것으로 알려져 있다.

따라서 고랭지 배추의 정식 이후 생육기간동안 토양 수분의 함수량 측정은 단수 예측 추정을 위한 핵심 과정이라고 할 수 있으며, 다음의 절차에 따라 토양 수분의 함수량을 계측하고자 한다. 토양수분균형모형을 이용하여 고랭지 배추 주산지의 표토층(top soil layer)⁴⁾ 토성 기준 하에서 토양 수분함량을 추정할 수 있다(Seonghuyk, H et al., 2009).

$$(2) \quad SW_t = SW_{t-1} + P + I - ET - R - D,$$

여기서 SW_t 는 월별 토양 수분 함량(mm)을 나타내며, P 는 월 누적강수량(mm), I 는 관개수(mm), R 는 표면유실량(mm), ET 는 실제 증발량(mm), D 는 심층배수량(mm)을 나타낸다. 관개수는 생산자가 기후환경 및 토양환경에 인위적으로 조정하므로 측정이 불가하여 실제 토양 수분함량 추정 시 생략할 수 있다.

$$(3) \quad R = \frac{(P - 0.2S)^2}{(P + 0.8S)}, \quad P \geq 0.2S,$$

4) 식물이 성장하는 맨위 토양층을 표토라 하며 동생물 유해 또는 부식물로 구성된 부식토가 포함되어 식물생장에 있어 최적토라 함(농진청 2019).

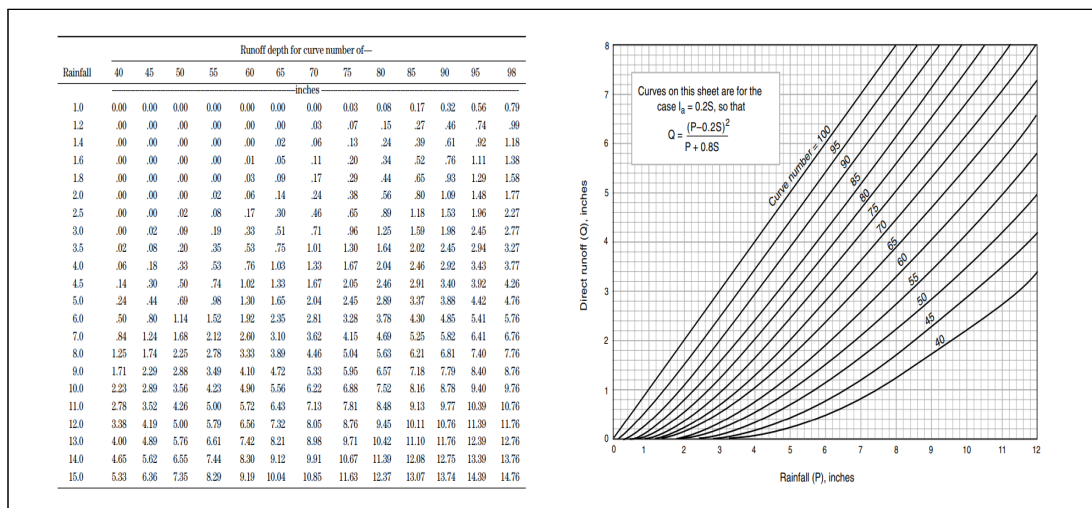
R 은 표면유실량(surface runoff)이며 누적강수량(P)과 보수력 계수(S)를 이용하여 계산된다. 누적강수량은 $0.2S$ 보다 큰 조건이 있다.

$$(4) S = \frac{25400}{CN} - 254$$

보수력 계수(S)식의 CN 은 유실곡선지수(Runoff Curve Number)을 나타내며 미국 토양보호국(U.S. Soil Conservation Service)의 토성별 기준 값을 적용하여 계산된다. CN 값의 범위는 누적강수량과 유실량 수준에 따라 40~100까지 분포하게끔 정해져 있다.

식 (4)와 식 (3)을 토대로 살펴보면, 유실곡선지수(CN) 값이 작은(큰) 경우 토양보수력계수(S)은 증가(감소)하고, 표면유실량(R)은 감소(증가)하게 된다. 다시 정리하면, CN 값 40~100 분포하며, 토양보수력이 좋은(나쁜) 경우 낮은(높은)수치 값을 갖게 된다. 실제 본 연구에서 실제 CN 값은 주산지별 토양보수력 환산값(표 3)을 역수 범위로 정하여 시산한 결과 평균 63.7, 최대 100 최소 50 수준의 값으로 시산되었다.

그림 2. 유실곡선지수 산출 기준표



자료: U.S. Soil Conservation Service(1986)

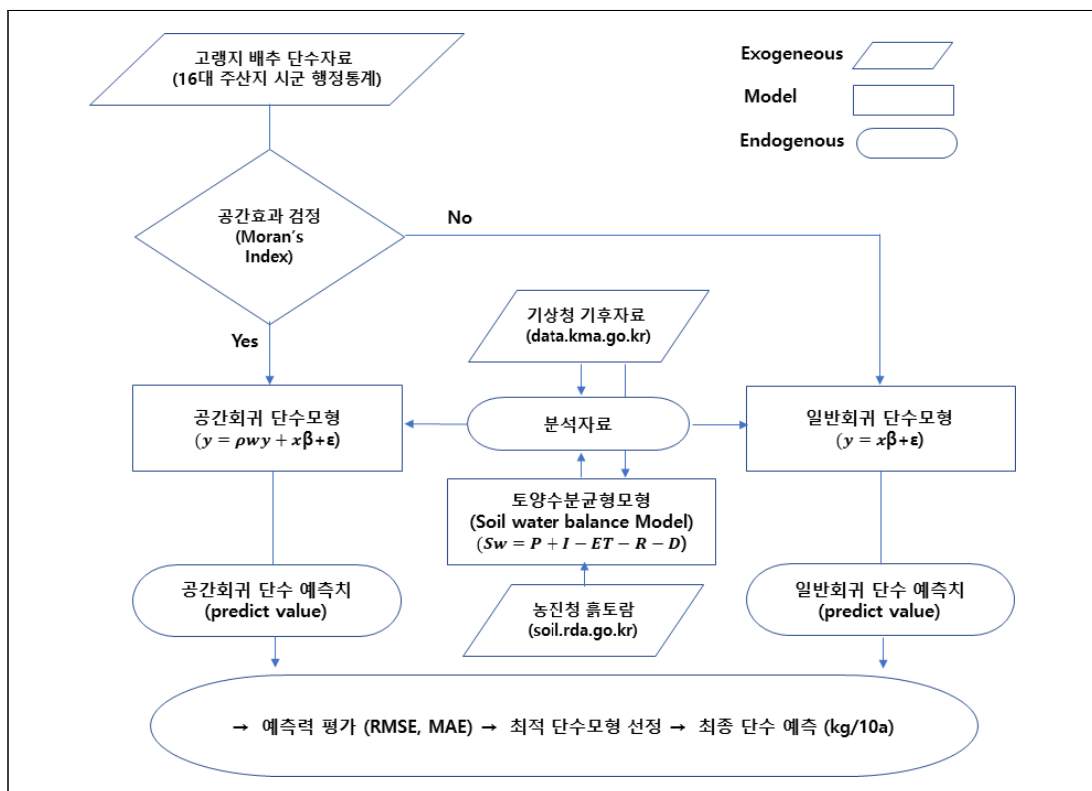
$$(5) ET = (\text{상대습도} \times \text{평균온도} \div 100) \times \text{토양보수력}$$

증발량(ET , mm)은 주산지 지역의 월평균 상대습도(%)와 평균온도(℃)을 기준으로 산출하며, 해당 지역의 양토비율을 기준으로 토양 보수력을 산정하여 활용한다.

$$(6) D = 0.5(P - R) \times \text{토양배수력}$$

D 는 침출배수량(mm)으로 지역의 월평균 누적강수량(mm)과 유실량(mm) 그리고, 토양의 배수력(≒토양 보수력 역산)을 산정하여 활용한다. 0.5는 배수계수(drainage coefficient)로써 일반적으로 0~1사이 값을 사용한다.

그림 3. 연구의 흐름도



자료: 저자 작성

3. 분석 모형 구조 및 자료

이 장에서는 고랭지 배추의 단수 자료를 이용하여 지역 간 공간적 연계성 여부를 Moran's I 값을 이용하여 살펴보고, 연계성 여부에 따라 일반회귀 단수 모형과 공간회귀 단수 모형을 선택하게 된다. 이어서 토양수분균형모형을 통해 주산지별 월별 토지수분 함수량을 추계하고, 기상청 기후자료와 함께 단수모형 추정 시 활용할 수 있도록 구축한다. 마지막으로 추정된 단수 예측치를 토대로 예측력 평가 하에 최적단수모형을 선정한다.

3.1. 분석 모형 구조 및 절차

3.1.1. 단수 모형

고랭지 배추 단수와 기후 및 토양 환경 간 관계를 식 (7)과 같이 행렬식 형태로 나타낼 수 있고, 이를 다시 패널형태인 실증 일반 단수 함수식으로 정리하면 식 (8)과 같으며, 여기에 공간 연계성 효과를 고려한 공간회귀 함수식은 식 (9)로 나타낼 수 있다.

$$(7) \quad y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma_\varepsilon^2)$$

$$(8) \quad y_{it} = \sum_j^n \beta_j X_{jit} + \mu_i + \varepsilon_{it}$$

$$(9) \quad y_{it} = \lambda W_N y_{it} + \sum_j^n \beta_j X_{jit} + \mu_i + \varepsilon_{it}$$

여기서 y_{it} 는 i 시군, t 년도의 고랭지 배추 단수, X_{ijt} 는 j 개의 기후변수 및 토양변수, 그리고 W_N 은 $N \times N$ (16×16) 16대 주산지 지역 간 거리 기준 공간가중치행렬을 의미한다. 한편, μ_i 는 개별지역효과를 나타낸다. 만일 이 개별지역효과 여부에 따라 확률효과모형을 적용할지 Pooled OLS 모형을 적용할지 결정된다. 다시 말하면, 고랭지 배추 단수모형은 최고온도, 일

조량, 강수량, 토양수분함수량 등 기후변수와 토양변수로 구성한다. 여기서 모형의 최적 형태는 사전 추정결과를 토대로 적합성이 제일 높은 로그-선형 방식을 선택하였으며, 아울러 변수 구성에 있어서도 단수함수의 특성상 일반 선형이 아닌 비선형 형태를 고려하기 위해 최고 온도 변수의 경우 2차항 변수로써 고려하였다. 분석절차를 다시정리하면, 우선 종속변수인 배추 단수자료 대상으로 Moran'I 검정을 통해 공간효과 여부를 검정하여 일반회귀모형 또는 공간회귀모형을 이용할지 결론짓고, 패널자료이므로 Pesaran(2004)의 Pesaran CD($H_0: \sigma_\mu^2 = 0$) 검정을 통해 개별지역효과 여부를 결정하여 확률효과모형을 이용할 것인지, Pooled OLS 모형을 적용할 것인지 정하였다. 만일 개별지역효과가 존재하는 것으로 판명된다면 이 개별지역효과를 다음단계에서 공간 하우스만 검정을 통해 고정효과 또는 확률효과 모형을 적용할지 결정하였다. 분석결과를 토대로 단수 예측치를 각각 도출하고, 표본외(out of sample) 검정에 의한 예측력을 평가하여 최종 단수모형을 선정하였다.

3.2. 분석 자료

3.2.1. 기상자료

본 연구에서는 16개 고랭지 배추 주산지과 1997년부터 2018년까지 패널자료를 구축하여 이용하였다. 고랭지 배추 단수자료(kg/10a)는 해당 시군 각년도 행정통계를 이용하였으며, 지역별 기상자료는 기상청 기상자료 개방포털(www.data.kma.go.kr)에서 제공하는 전국 102개 지점에서 수집된 자료 중 해당 지역과 가장 가깝게 수집된 기상자료를 적용하였다. 1997년 4월~2018년 9월까지 일일자료, 순자료, 월자료 모두 구축하였으나, 지역마다 생육 시기가 모두 상이하여 월자료를 최종 사용하였다. 수집 자료는 평균온도($^{\circ}\text{C}$), 최고온도($^{\circ}\text{C}$), 최저온도($^{\circ}\text{C}$), 누적강수량(mm), 상대습도(% rh), 평균풍속(m/s), 일조시간(hr), 황사일수(일) 등이다. 이들 기상자료 중 평균풍속, 상대습도 자료는 이하 토양수분균형모형 추계 시 활용되어 공선성 문제 방지 차원에서 제외하였고, 황사일수는 통계적으로 유의하지 않아 실증분석 시 제외하였다(부표 3 참조).

고랭지 배추가 집중 출하되는 시기를 보통 7월~9월(부록 참조)로 보면 지역에 따라 다를 수 있지만 정식 후 평균적인 생육기간을 전반부(6월~7월)와 후반부(8월~9월)로 나눌 수 있다. 이 두 시기 모두 정식 후 기후요인이 제일 중요한 시점임을 고려하여 이 기간의 기후변수들을 적극 고려하였다. 아울러, 최고온도의 경우 생육의 최적온도가 임계점을 지나게 되면 다시 단수 증산(감산)이 감소(증가)하는 형태의 함수식으로 고려하기 위해 2차 함수형태를 적용하였다.

표 2. 고랭지 배추 단수, 기상자료, 토양수분 함수량 기초통계량(1997~2018)

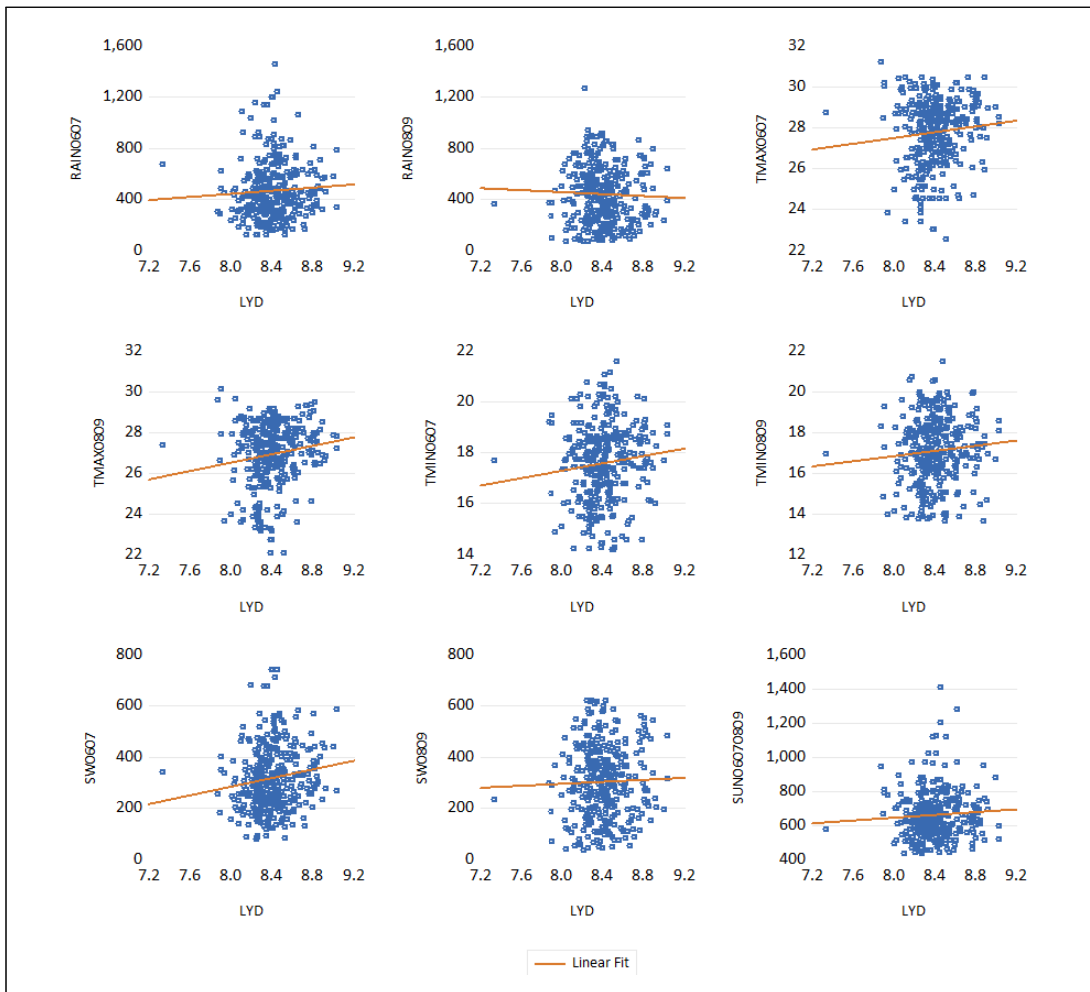
변수		Mean	Maximum	Minimum	Std.	Obs.
ln 단수(kg/10a)	ln(yd)	8.41	9.03	7.34	0.22	336
최고온도 6~7월(C°)	TMAX0607	27.77	31.20	22.55	1.63	336
최고온도 8~9월(C°)	TMAX0809	26.92	30.15	22.10	1.56	336
최고온도 6~7월(C°) ²	TMAX0607S	773.64	973.44	508.50	88.53	336
최고온도 8~9월(C°) ²	TMAX0809S	726.86	909.02	488.41	82.20	336
최저온도 6~7월(C°)	TMIN0607	17.57	21.60	14.20	1.50	336
최저온도 8~9월(C°)	TMIN0809	17.07	21.50	13.65	1.61	336
토양수분6~7월(mm)	SW0607	317.47	742.16	78.63	125.17	336
토양수분8~9월(mm)	SW0809	301.30	622.85	36.91	141.01	336
누적강수량6~7월(mm)	RAIN0607	464.66	1457.50	123.20	217.77	336
누적강수량8~9월(mm)	RAIN0809	436.89	1265.50	73.60	222.59	336
누적일조량6~9월(hr)	SUN06070809	658.26	1408.50	438.70	139.18	336
평균풍속 6~7월(m/s)	WIN06	1.49	2.50	0.40	0.35	336
평균풍속 8~9월(m/s)	WIN08	1.36	2.70	0.30	0.36	336
황사발생(일)	SAND	3.25	27	0	5.11	336

자료: 기상청, 농진청 원자료 이용 저자 작성

상기 기상자료와 고랭지 배추 단수와와의 기본적 관계구조를 살펴보기 위해 <그림 3>과 같이 산포도를 작성하였다. 단수(LYD) 기준으로 보면, 전반부 6월~7월 기상자료의 경우 최고기온도(TMAX0607), 최저온도(TMIN0607), 토양수분함수량(SW0607) 회귀선의 기울기가 상대적으로 가파른 우상향 양상을 보이고 있다. 따라서 전반부 6월~7월 단수와 최고온도, 최저온도, 그리고 토양수분 함수량은 정(+)의 연관성을 보이는 것으로 나타났다. 후반부 8월~9월의 경우 최고온도(TMAX0809)와 최저온도(TMIN0809) 정(+)의 연관성을 가지는 것으로 판단된다.

한편, 물 흡수가 많은 고령지 배추의 생육특성상 단수(LYD)와 강수량(RAIN0607, RAIN0809)이 높은 정(+)의 연관성을 보일 것으로 예상하였으나, 연관성이 미미하거나 또는 오히려 부(-)의 연관성을 보였다. 이는 경사도가 가파른 고령지의 특성상 강수의 대부분이 표토유실되어 토양의 수분공급에 크게 기여하지 못하는 것으로 추론된다. 이에 반해 토양수분 함수량(SW0607, SW0809) 변수의 경우 단수와 정(+)의 상관관계를 보이고 있어 모형 형태(Model specification) 수립 시 강수량의 대체변수로써 고려할 필요가 있다.

그림 3. 기상자료, 토양수분함수량, 단수 산포도



자료: 저자 작성

3.2.2. 토양자료

농촌진흥청의 『흙토람』 토양환경지도 시스템을 활용하여 16대 고랭지 배추 주산지의 토성을 검색하여 해당 지역 양토 비중을 토대로 토양 보수력을 0.5~1사이 범위로 설정하였다(부표 4 참조). 보수력 상위 주산지는 태백(1.0), 정선(0.9), 평창·영양(0.8)이며, 중위 주산지는 영월·단양(0.7), 하위 주산지는 강릉·삼척(0.6), 양구·인제·무주(0.5)의 순으로 나타났다. 이를 토대로 유실곡선지수(CN)를 구해보면, 100이하의 범위 내에서 토양 보수력 순위와 같게 집계되었다. 이어서 보수력 계수(S) → 표면유실량(R) → 침침배수량(D) → 증발량(ET) 순으로 시산하고, 토양수분균형모형(Soil water balance model)을 통해 주산지 토양 수분의 함수량(SW)을 구축하였다.

표 3. 주산지별 토양 보수력 및 배수량 시산 결과

주산지	토양 보수력	토양 배수량	양토면적	면적	양토비
강릉	0.6	0.9	56	103	54.4%
봉화	0.5	1	34	120	28.3%
삼척	0.6	0.9	69	118	58.5%
양구	0.5	1	12	55	21.8%
영월	0.7	0.8	72	112	64.3%
인제	0.5	1	58	147	39.5%
정선	0.9	0.6	106	121	87.6%
태백	1	0.5	27	30	90.0%
평창	0.8	0.7	109	146	74.7%
홍천	0.5	1	86	181	47.5%
횡성	0.5	1	28	99	28.3%
무주	0.5	1	23	63	36.5%
문경	0.5	1	38	91	41.8%
영양	0.8	0.7	64	82	78.0%
단양	0.7	0.8	52	78	66.7%
제천	0.5	1	39	88	44.3%

자료: 농촌진흥청 『흙토람』 자료 이용 시산

한편, 토양수분균형모형(soil water balance model)을 통해 추계된 토양수분함수량과 고랭지 배추 단수와의 관계를 산포도로 정밀하게 나타내면 다음과 <그림 4>와 같다. 상기 <그림 3>과 마찬가지로 단수는 강수량(rain0607, rain0809)보다 토양수분함수량(sw0607, sw0809)와 정(+)의 연관관계가 큰 것으로 나타났다.

표 4. 토양수분균형모형을 이용한 토양수분 함수량(SW) 시산결과

(단위: mm, %, index(100))

변수		평균	최대값	최소값	표준편차	관측치
심층배수 4월	D04	8.459	21.575	0.600	4.617	336
심층배수 5월	D05	9.209	29.511	0.210	5.183	336
심층배수 6월	D06	12.906	34.995	2.510	6.927	336
심층배수 7월	D07	24.665	60.123	6.190	9.929	336
심층배수 8월	D08	21.753	49.318	3.260	10.088	336
심층배수 9월	D09	14.038	41.376	1.260	8.716	336
표면유실곡선치	CN	63.750	100.0	50.000	16.233	336
증발량 4월	E04	3.946	8.942	1.841	1.122	336
증발량 5월	E05	6.453	11.410	4.134	1.609	336
증발량 6월	E06	8.973	15.228	5.340	2.289	336
증발량 7월	E07	11.685	19.035	6.935	2.990	336
증발량 8월	E08	11.743	20.003	7.025	3.012	336
증발량 9월	E09	9.077	15.192	5.198	2.345	336
표면유실수 4월	R04	2.806	58.250	0.000	6.222	336
표면유실수 5월	R05	3.620	59.479	0.000	7.729	336
표면유실수 6월	R06	12.382	285.382	0.006	30.734	336
표면유실수 7월	R07	76.573	643.268	0.032	95.904	336
표면유실수 8월	R08	57.667	643.817	0.000	82.867	336
표면유실수 9월	R09	21.316	260.177	0.000	41.331	336
보유력 계수	S	589.091	762.000	254.000	185.070	336
토양수분 함수량 4월	SW04	72.189	190.416	1.128	41.429	336
토양수분 함수량 5월	SW05	76.447	261.047	2.031	46.640	336
토양수분 함수량 6월	SW06	107.177	308.132	12.535	62.531	336
토양수분 함수량 7월	SW07	210.295	531.647	45.714	90.427	336
토양수분 함수량 8월	SW08	184.036	432.914	13.356	91.689	336
토양수분 함수량 9월	SW09	117.260	364.365	3.362	78.386	336

자료: 토양수분균형 모형 추계식

그림 4. 토양수분함수량, 강수량, 단수 산포도



자료: 저자 작성

3.2.3. 공간가중치행렬

고령지 배추 주산지 지역 간 연계성을 고려하기 위하여 16개 주산지 지역 간 거리를 이용하여 시산된 공간가중치행렬 <표 5>와 같다. 일반적으로 가중치 행들의 합이 1이 되도록 표준화 (standardization) 과정을 거친 후 이용된다. $w_{ij}s = w_{ij} / \sum_{j=1}^n w_{ij}$ 여기서 $w_{ij} = 1/\text{distance}_{ij}$ 는 행

렬 원소이며 w_{ij} s는 표준화된 원소를 의미한다. <표 5> 내 표준화된 원소 값이 클수록(적을수록) 지역 간 상대적으로 근거리(원거리)에 위치하고 있어 공간적 연계성이 높다(작다)는 것을 나타낸다.

표 5. 표준화된 공간가중치행렬(16대 주산지×16대 주산지)

	강릉	봉화	삼척	양구	영월	인제	정선	태백	평창	홍천	횡성	무주	문경	영양	단양	제천
강릉	0.0000	0.0436	0.1221	0.0596	0.0816	0.0704	0.0849	0.0666	0.0599	0.0523	0.0503	0.0479	0.0421	0.0514	0.0402	0.0421
봉화	0.0545	0.0000	0.0703	0.0343	0.1415	0.0351	0.0691	0.1182	0.0608	0.0394	0.0455	0.0788	0.1042	0.1550	0.1167	0.0721
삼척	0.1268	0.0584	0.0000	0.0337	0.0850	0.0441	0.0860	0.1273	0.0533	0.0396	0.0416	0.0514	0.0478	0.0740	0.0459	0.0432
양구	0.0605	0.0279	0.0330	0.0000	0.0542	0.3107	0.0384	0.0330	0.0414	0.0969	0.0622	0.0455	0.0361	0.0342	0.0309	0.0360
영월	0.0735	0.1019	0.0737	0.0480	0.0000	0.0498	0.0430	0.0363	0.0459	0.0887	0.0633	0.0457	0.0371	0.0360	0.0324	0.0376
인제	0.0752	0.0299	0.0453	0.3266	0.0591	0.0000	0.0440	0.0364	0.0454	0.0915	0.0629	0.0456	0.0368	0.0358	0.0322	0.0373
정선	0.1160	0.0756	0.1131	0.0517	0.0653	0.0563	0.0000	0.1190	0.1552	0.0603	0.0707	0.0610	0.0632	0.0736	0.0793	0.0798
태백	0.0773	0.1098	0.1422	0.0377	0.0468	0.0396	0.1010	0.0000	0.0653	0.0408	0.0450	0.0616	0.0641	0.1149	0.0687	0.0557
평창	0.0845	0.0686	0.0723	0.0575	0.0719	0.0600	0.1602	0.0793	0.0000	0.0818	0.1138	0.0652	0.0690	0.0634	0.0931	0.1270
홍천	0.0596	0.0359	0.0434	0.1086	0.1123	0.0976	0.0503	0.0400	0.0661	0.0000	0.1814	0.0573	0.0492	0.0411	0.0453	0.0596
횡성	0.0617	0.0447	0.0491	0.0751	0.0863	0.0722	0.0635	0.0476	0.0989	0.1954	0.0000	0.0640	0.0603	0.0471	0.0608	0.0914
무주	0.0235	0.0310	0.0243	0.0220	0.0249	0.0209	0.0219	0.0260	0.0227	0.0247	0.0256	0.0000	0.0774	0.0432	0.0324	0.0302
문경	0.0361	0.0715	0.0395	0.0305	0.0354	0.0296	0.0397	0.0474	0.0419	0.0371	0.0421	0.1354	0.0000	0.0782	0.0873	0.0671
영양	0.0421	0.1017	0.0584	0.0276	0.0328	0.0275	0.0442	0.0811	0.0368	0.0296	0.0315	0.0722	0.0747	0.0000	0.0533	0.0411
단양	0.0530	0.1229	0.0582	0.0400	0.0475	0.0397	0.0764	0.0779	0.0869	0.0523	0.0652	0.0869	0.1340	0.0856	0.0000	0.1799
제천	0.0559	0.0766	0.0551	0.0470	0.0554	0.0464	0.0775	0.0637	0.1195	0.0694	0.0988	0.0817	0.1039	0.0665	0.1814	0.0000

자료: 저자 작성

4. 분석 결과

고랭지 배추의 단수의 지역 간 공간적 연계성을 검정한 결과, Moran's I 값이 0.894이고, P값은 <.0001로 나타나 공간적으로 연계성이 존재하는 것으로 분석되었다. 따라서 3장의 분석 절차에 따라 구축된 공간회귀 단수모형을 활용하고, μ_i 의 개별지역효과 검정결과, P값이 0.02393로 나타나 귀무가설($\sigma_\mu^2 = 0$)을 기각하여 개별지역효과가 존재하는 것으로 확률효과 모형(random effect model)을 적용하고, 다음 단계에서 공간하우스만검정(spatial hausman test)결과, χ^2 값이 365.7이고, p값이 <.0001로 나타나 개별지역효과는 고정적 효과로 취급하는 고정효과모형인 공간회귀 단수 모형으로 추정하였다.

4.1. 공간회귀 단수모형 추정결과

고랭지 배추의 공간적 연계성이 존재하는 것으로 판명되어, 실제 <표 6>의 일반회귀단수모형 ①과 ② 추정 결과는 불필요하나, 공간회귀 단수모형과의 모형적합도 비교를 위해 함께 제시하였다. 모형 ①~⑤ 모두 패널분석이므로 절대적 설명력은 매우 낮은 수준이나, 일반회귀 단수모형 ①, ② 보다 공간회귀 단수모형 ③, ④, ⑤ 모형의 적합도가 우수한 것으로 나타났다. 이는 지역 간 상호 연계성을 의미하는 공간회귀계수가 통계적으로 유의($|λ| > 0$, $|ρ| > 0$)하게 단수 결정요인으로써 역할을 한 것으로 보인다. 한편, 공간회귀 단수모형 중 ④ SEM 모형이 13.2% 설명력으로 모형적합도가 가장 우수한 것으로 나타났으며, ⑤ SAE 모형 12.8%, ③ SAR 모형 12.7% 순으로 Pseudo R^2 값이 높은 것으로 분석되었다.

추정된 계수 중심으로 결과를 살펴보면, ‘최고온도(6월~7월)’은 모든 모형에서 부(-)의 영향으로 단수에 영향을 미치는 것으로 나타났으며, ‘최고온도(8월~9월)’도 통계적으로 유의하진 않았으나 추정 부호가 동일하여 고랭지 배추 단수에 부정적인 영향을 주는 것으로 판단되었다. 이와 같은 결과는 결구가 시작되면 고온에 약해져 결구가 불량하거나 정상적 생육이 불가능한 것에 기인한 것으로 보인다. ‘토양수분’ 변수의 경우 전반부(6~7월), 후반부(8~9월) 모두 단수에 양(+)의 영향을 미치는 것으로 나타났는데, 이는 고랭지 배추의 생육과정에서 수분 공급의 중요성을 반증하는 결과라고 할 수 있다. 한편, 일반회귀 단수모형 중 전통적인 기본 모형 형태의 ① 일반회귀 단수모형은 최고온도와 누적강수량 변수를 포함하여 추정한 것으로서 최고온도만 통계적 유의성이 있는 것으로 추정되었다. 강수량 추정계수는 통계적으로 유의하진 않았으나 양(+)의 부호를 보이며 미미하게 단수 증가에 영향을 미치는 것으로 추론할 수 있다. 반면, 강수량 변수 대신 토양수분 변수를 포함시킨 ② 일반회귀 단수모형의 추정 결과를 보면, ‘토양수분’ 변수가 유의수준 5%에서 통계적으로 유의하게 나타나 단수 증장에 지대한 영향을 미치는 것으로 판단된다. 그러나 본 연구의 목적은 단수 결정요인 변수의 계수 추정에 있는 것이 아니라 정확한 단수 예측에 있으므로 추정된 계수를 바탕으로 단수 예측을 실시하여 예측력에 대한 평가를 시도하고자 한다.

표 6. 일반 및 공간단수모형 추정 결과

변수	① 일반회귀 단수모형 (강수량)			② 일반회귀 단수모형 (토양수분)			공간회귀 단수모형(토양수분)																	
	Coeff	t값	P값	Coeff	t값	P값	③ SAR(spatial lag)			④ SEM(spatial error)			⑤ SAE(spatial lag-error)											
상수	-24.93300 *	-1.89	0.060	-20.9950	-1.60	0.111	Coeff	-14.2510	t값	-1.13	P값	0.257	Coeff	-21.3980 *	t값	-1.65	P값	0.099	Coeff	-14.19000	t값	-1.13	P값	0.260
최고온도 (6~7월)	-0.96909 *	-1.95	0.053	-0.98624 **	-1.99	0.047	Coeff	-0.85014 *	t값	-1.79	P값	0.073	Coeff	-1.04450 **	t값	-2.10	P값	0.036	Coeff	-0.90753 *	t값	-1.88	P값	0.060
최고온도 (8~9월)	-0.21599	-0.43	0.670	-0.04824	-0.10	0.924	Coeff	0.07627	t값	0.16	P값	0.875	Coeff	-0.00169	t값	0.00	P값	0.997	Coeff	0.13991	t값	0.28	P값	0.777
토양수분 (6~7월)				0.00021 **	2.09	0.037	Coeff	0.00022 **	t값	2.31	P값	0.021	Coeff	0.00024 **	t값	2.31	P값	0.021	Coeff	0.00026 **	t값	2.48	P값	0.013
토양수분 (8~9월)				0.00019 **	2.12	0.035	Coeff	0.00013 **	t값	2.10	P값	0.036	Coeff	0.00022 **	t값	2.29	P값	0.022	Coeff	0.00021 **	t값	2.22	P값	0.026
누적일조량	0.00011	1.18	0.238	0.00012	1.25	0.211	Coeff	0.00008	t값	0.83	P값	0.404	Coeff	0.00014	t값	1.39	P값	0.164	Coeff	0.00009	t값	0.95	P값	0.341
최고온도 (6~7월) ²	10.15400 *	1.95	0.052	10.36000 **	2.00	0.046	Coeff	9.00140 *	t값	1.81	P값	0.070	Coeff	10.97600 **	t값	2.11	P값	0.035	Coeff	9.61240 *	t값	1.90	P값	0.057
최고온도 (8~9월) ²	2.40190	0.46	0.644	0.64068	0.12	0.902	Coeff	-0.59999	t값	-0.12	P값	0.904	Coeff	0.15769	t값	0.03	P값	0.976	Coeff	-1.26060	t값	-0.25	P값	0.804
강수량 (6~7월)	0.00005	0.87	0.384				Coeff		t값		P값		Coeff		t값		P값		Coeff		t값		P값	
강수량 (8~9월)	0.00007	1.20	0.230				Coeff		t값		P값		Coeff		t값		P값		Coeff		t값		P값	
lambda(λ)							Coeff	-0.0305 ***	t값	-4.59	P값	0.000	Coeff		t값		P값		Coeff	-0.0307 ***	t값	-4.62	P값	0.000
rho(ρ)							Coeff		t값		P값		Coeff	0.1233 ***	t값	3.51	P값	0.000	Coeff	0.1337	t값	1.07	P값	0.284
Pseudo R ²	0.0483			0.06854			Coeff	0.12663	t값		P값		Coeff	0.13267	t값		P값		Coeff	0.128137	t값		P값	

주: 1) ***, **, *은 각각 1%, 5%, 10%에서 유의함.

2) 공간 모형의 경우 최우추정법(MLE)을 활용하였기 때문에 log-likelihood 통계량을 Pseudo R^2 로 환산한 것임.

4.2. 단수모형의 예측력 평가

토양수분 변수가 포함된 공간회귀 단수모형 ③, ④, ⑤를 대상으로 표본 외(out of sample test) 예측검정을 실시하기 위해 추정기간을 기존보다 2년(1997~2016년) 짧게 설정하고, 나머지 2년간(2017~2018년) 단수 예측치와 실측치를 이용하였다(표 6). 예측오차 통계량 RMSE를 보면 ④ SEM 모형의 경우 각각 0.215 값으로 예측력이 가장 뛰어난 것으로 평가된다. 이어서 ③ SAR 모형 0.3358, ⑤ SAE 모형 0.3369 순으로 예측력이 뛰어난 것으로 나타났다. RMSE 기준으로 볼 때 ④ SEM 모형의 RMSE 값이 나머지 모형의 RMSE 값보다 약 56% 작은 것으로 나타나, 공간회귀 단수모형 ④의 예측능력이 여타 모형 대비 월등한 것으로 평가할 수 있다.

표 6. 표본 외 예측력 평가 결과

예측오차 통계량	③ SAR(spatial lag)	④ SEM(spatial error)	⑤ SAE(spatial lag-error)
RMSE	0.3358114	0.215374	0.3369259
MAE	0.2870383	0.1650536	0.2878926

자료: 추정기간 1997~2016년, 표본 외 기간 2017~2018년

5. 요약 및 결론

최근 들어 이상기후 현상에 따른 고랭지 배추 생산 변동성이 점차 확대되면서 과거보다 여름철 배추 수급 및 가격 불안에 대한 우려가 더욱 커지고 있다. 이에 본 연구는 고랭지 배추 주산지 16개 지역의 기상요인, 토양 및 지형요인, 공간적 종속 요인 등을 고려하여, 고랭지 배추의 전국 생산변화를 예측하는 단수 예측 모형을 수립하여 추정하였다. 특히, 고랭지 배추의 생육은 기후환경 요인뿐만 아니라 토양환경 요인에도 지대한 영향을 받아 단수에 영향을 주게 된다. 특히, 토양환경 요인 중 토양의 함수량은 단수 결정요인으로써 있어 매우 중요하다고 할 수 있다. 그러나 그동안 대부분 선행연구들의 작물의 단수함수 구성은 최고온도, 최저

온도, 강수량, 일조량 등 기후요인을 중심으로 단수 추정을 시도해 왔다. 따라서 본 연구에서는 고랭지 배추의 정식 이후 생육기간동안 토양 수분의 함수량 측정을 위해 토양수분균형모형(soil water balance model)을 활용하여 토양 수분의 함수량을 계측하였고, 이를 단수모형 구축 시 적용하였다. 더하여, 공간적으로 연계되어있는 지역 간 자료를 고려하여 단수 모형 추정의 고도화를 시도하여 단수모형의 예측력을 개선하고자 하였다. 이에 따라 일반회귀 단수모형 보다 공간회귀 단수모형 모형의 적합도가 우수한 것으로 추정되었고, 이를 토대로 개별 변수의 통계적 유의성을 살펴본 결과, 토양수분균형모형으로부터 추계된 토양수분 함수량 변수의 경우 단수에 양(+)의 영향을 미치는 것으로 나타나 고랭지 배추의 생육과정에서 수분 공급의 중요성을 재확인되었다. 그러나 이 연구의 궁극적 목적은 추정된 계수를 바탕으로 정교한 단수 예측에 있다고 할 수 있으며, 이를 위해 예측력에 대한 평가를 실시하였다. 선정된 단수 모형의 예측력 평가 결과, 지역 간 공간 연계성을 고려하고 토양수분 함수량을 모형형태에 포함시킨 공간회귀 단수모형이 예측능력이 월등한 것으로 평가되었다. 이와 같이 새로이 개발된 고랭지 배추 단수 모형을 이용하여 배추 생산 예측이 정확하게 이루어진다면, 배추 생산자, 산지유통인의 경우 생산 예측 정보를 활용하여 출하량 및 출하시기 의사결정 시 적절히 활용될 수 있을 것이다. 나아가 정부는 배추 출하기 수급안정대책 수립 시 긴급 기초자료로서 활용가치 제고를 기대할 수 있다.

참고문헌

- 농촌진흥청. 2002. 「배추 재배기술」.
- 농촌진흥청. 2019. 「농업기술길잡이 128」.
- 이성우·윤성도·박지영·민성희. 2006. 「공간계량모형응용」. 박영사.
- 윤성은. 2013. 「농업경영체 유형화 및 농업생산 공간분포 요인분석」. 서울대학교 대학원 석사학위 논문.
- 정호중, 정원호. 2016. “배추수급모형추정을 통한 가격변동시나리오분석.” 농업경영·정책연구 43(4): 897-920.
- 최병옥 외. 2020. “기상환경 변화가 배추·무 단수와 가격 변동성에 미치는 영향 분석.” 농촌경제 4(1): 21-47.
- 한석호 외. 2011. 「기상요인을 고려한 단수예측모형 개발연구」. 한국농촌경제연구원.
- 국립농산물품질관리원. 각 연도. 「작물통계」.
- 강릉시. 각 연도. 「강릉시 통계연보」.
- 봉화군. 각 연도. 「봉화군 통계연보」.
- 평창군. 각 연도. 「평창군 통계연보」.
- 삼척시. 각 연도. 「삼척시 통계연보」.
- 양구군. 각 연도. 「양구군 통계연보」.
- 영월군. 각 연도. 「영월군 통계연보」.
- 인제군. 각 연도. 「인제군 통계연보」.
- 정선군. 각 연도. 「정선군 통계연보」.
- 태백시. 각 연도. 「태백시 통계연보」.
- 홍천군. 각 연도. 「홍천군 통계연보」.
- 횡성군. 각 연도. 「횡성군 통계연보」.
- 무주군. 각 연도. 「무주군 통계연보」.
- 문경시. 각 연도. 「문경시 통계연보」.
- 영양군. 각 연도. 「영양군 통계연보」.
- 단양군. 각 연도. 「단양군 통계연보」.
- 제천시. 각 연도. 「제천시 통계연보」.
- KOSIS 국가통계포털. <<http://kosis.kr>>.
- 기상청 관측자료. <<http://www.kma.go.kr>>.
- 농촌진흥청 흙토람. <<http://soil.rda.go.kr>>.
- Anselin, L. 1988. Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic Publishers.
- Pesaran, M. H. 2004. General diagnostic tests for cross section dependence in panels.

- University of Cambridge, Faculty of Economics, Cambridge Working Papers in Economics No. 0435.
- Seonghuyk, H., Epplin, F.M., Byoung-hoon L., Raymond H. 2009. "A probabilistic estimate of the frequency of mowing and baling days available in Oklahoma USA for the harvest of switchgrass for use in biorefineries." *Biomass & Bioenergy*. 33: 1037-1045.
- U.S. Soil Conservation Service. 1986. *Urban hydrology for small watersheds*. Technical Release 55. Washington DC: U.S.Department of Agriculture.
- Wooldridge, J.M. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

부표 1. 최근 5개년 월평균 배추 반입량(2016~2020)

산지	7월	산지	8월	산지	9월	산지	10월
강원평창군	158,975	강원평창군	123,973	강원강릉시	282,759	강원평창군	140,622
강원횡성군	57,058	강원태백시	107,520	강원평창군	96,404	강원강릉시	79,050
강원강릉시	43,353	강원강릉시	99,566	강원정선군	65,611	강원횡성군	70,640
강원정선군	43,347	경북문경시	50,783	강원태백시	62,740	충북단양군	58,488
경북영양군	43,262	강원영월군	33,921	강원삼척시	42,726	강원홍천군	48,058
경북봉화군	40,796	강원정선군	32,726	경북문경시	36,100	강원영월군	47,858
경북문경시	35,112	강원양구군	30,580	강원횡성군	35,187	강원정선군	43,843
강원태백시	33,229	강원횡성군	28,750	강원홍천군	31,114	충북제천시	35,987
충북제천시	32,488	강원삼척시	27,654	충북제천시	29,565	강원태백시	29,911
강원삼척시	31,082	경북영양군	25,405	강원영월군	21,242	경북문경시	29,032
강원홍천군	30,863	경북봉화군	15,211	충북단양군	15,637	경북영양군	28,944
강원화천군	25,800	강원홍천군	15,144	경북봉화군	15,005	경북봉화군	28,146
강원인제군	8,900	전북무주군	9,375	강원양구군	14,995	강원삼척시	21,413
강원영월군	21,681	강원인제군	8,960	전북무주군	11,510	강원인제군	16,957
전북무주군	18,115	기타	73,235	강원인제군	9,826	전북무주군	12,500
기타	135,902			경북영양군	5,000	강원양구군	10,598
				기타	47,195	기타	493,466
주산지계	624,062 (82.1%)		609,569 (89.3%)		784,890 (94.3%)		714,887 (59.2%)
합계	759,964		682,804		832,085		1,208,353

자료: 서울가락동 농수산물도매시장 반입량 기준

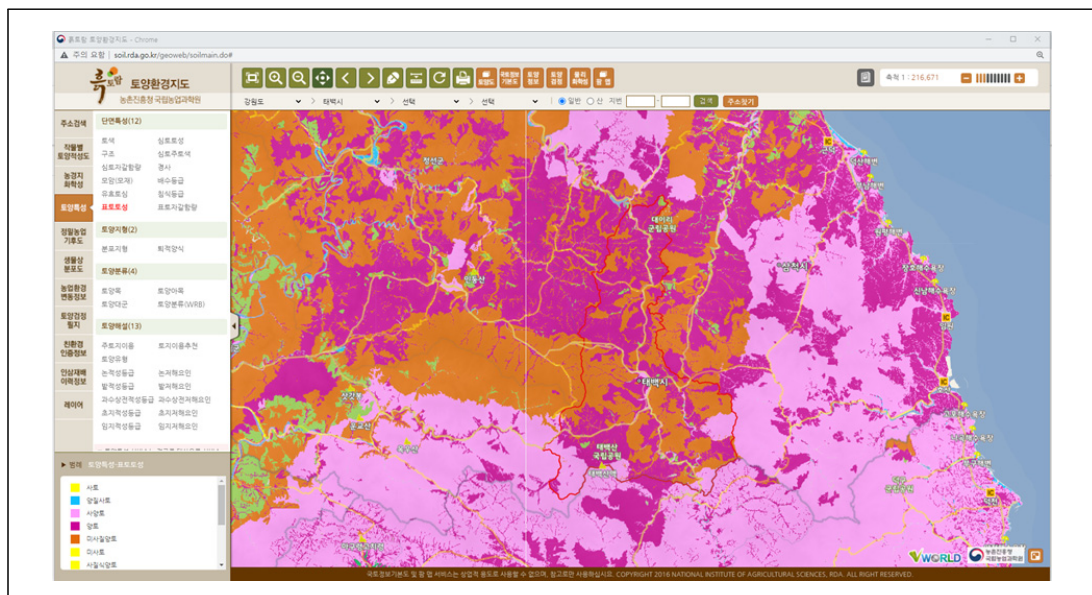
부표 2. 배추 정식 및 수확기 순기표

구분	작형	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월	주 출하지
봄	시설		정식			수확								충남(예산), 전남(나주) 경기(평택)
	터널													경남(의령, 외성) 충남(아산, 홍성)
	노지													경북(문경, 영양) 강원(영월)
고랭지	준고랭지 1기작													강원(평창, 횡성)
	고랭지배추													강원(강릉, 태백, 삼척, 정선, 평창 등)
	준고랭지 2기작													강원(평창, 횡성)
가을	-													전국(해남, 아산, 고창, 강릉, 춘천)
겨울	-													전남(해남)

출처: 한국농촌경제연구원 농업관측본부 담당자 의견

부표 3. 기상청 기상자료 개방포털(www.data.kma.go.kr)

부표 4. 농촌진흥청 흙도람(http://soil.rda.go.kr)





제1회

농산물 수급 예측모형 경진대회

우수논문집

Chapter 2
자유과제



우수상

시설재배 오이 수확량 및 가격예측모형

이브람

시설재배 오이 수확량 및 가격예측모형 스마트팜 빅데이터를 활용한 딥러닝 모델링

이브람

Key words

수확량예측(Crop Yield Prediction), 가격예측(Price Prediction),
LSTM순환신경망(LSTM)

ABSTRACT

Cucumbers are mainly cultivated in the greenhouse throughout the year and the production is relatively stable. However, prices fluctuate largely due to supply and demand in high or low temperature seasons. In the harvest soaring season, prices may drop by less than half, and in winter season, prices may soar more than tripled. This study predicts the crop yield of cucumbers using the Long Short-Term Memory (LSTM) recurrent neural network, and predicts the trading volume and price of the wholesale market using time series prediction machine learning models so that the farmer can predict the revenue of a crop period. Farmers may use this information to make agricultural decisions such as planting time, end of cropping season, and alternative crop cultivation. All data used for machine learning are collected from real farms in Gyeonggi-do, Korea from 2018 to 2020, and the price prediction model was trained with actual wholesale market transaction price. Prediction results are presented using comparative graphs and performance evaluation metrics, the mean absolute error and the coefficient of determination, R-Squared.

차례

- | | |
|----------|---------------|
| 1. 서론 | 3. 연구 내용 및 결과 |
| 2. 연구 방법 | 4. 결론 |

1. 서론

주로시설에서연중재배되는오이는생산량이안정적인편이나고온기와저온기의수급에따른 가격변동이큰작물이다. 경기도의 경우 2~6월 봄작기와 9~11월 가을작기 재배가 일반적이며 시세가 좋은 1~2월에 출하하기 위해 겨울철 가온 재배를 하거나 봄작기와 가을작기 사이 7~8월에 짧은 여름작기 재배를 하기도 한다. 농가에 따라서는 토마토나 방울토마토 등의 대체 작물 재배로 수익증대를 꾀한다. 비슷한 시기에 작기가 진행되기 때문에 수확량이 집중되는 시기에 도매시장 가격이 연평균 거래가격의 50% 이하로 떨어져 농가 수익이 비용에 못 미치는 경우가 발생하기도 하여 농가에서는 정식을 앞당겨서 출하시기를 조정하기도 한다. 이 경우 난방비 등 추가 비용이 발생하게 되므로 비용상승에 대한 수익구조를 잘 살펴보아야 한다.

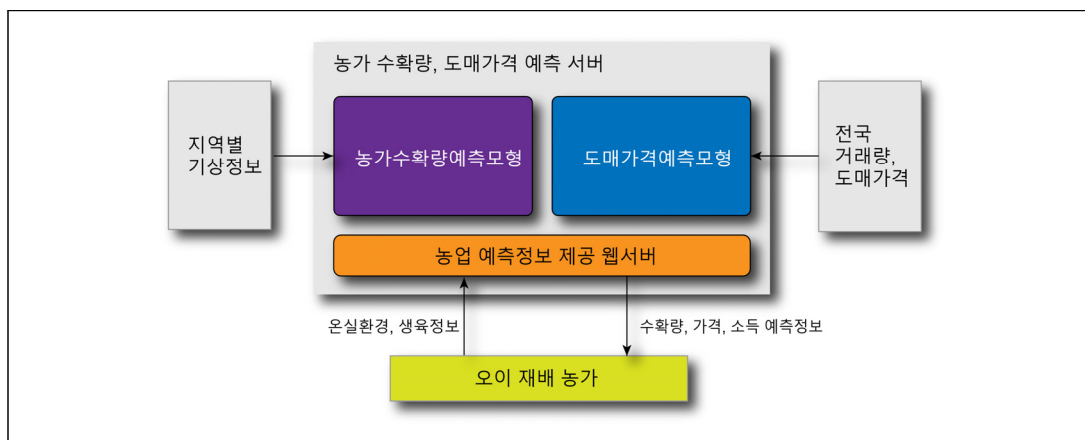
본 연구의 목적은 시설재배 오이의 수확량과 도매가격을 예측하여 농가의 정식시기나 작기 종료시기, 대체 작물 재배 등의 영농 의사결정에 활용하는 것이다.

2. 연구 방법

오이 스마트팜 시설재배 농가의 실측 생육조사 데이터와 온실 환경데이터를 학습데이터로 활용한 인공지능망 딥러닝 모델로 작기 주차별 수확량을 예측하고 농업관측통계시스템에서 제공하는 도매시장 가격정보를 활용한 가격예측 모델로 주차별 거래가격을 예측한다. 예측된 정보는 오이재배 농가에서 웹사이트를 통해 제공받을 수 있으며 예측을 위한 입력정보는 스마트팜 환경센서 수집정보 서버에서 전송되거나 직접입력 혹은 지역별 외부 기상정보를 활용할 수 있도록 한다. 가격예측모형은 전일까지의 도매시장 거래량과 가격으로 예측할 수 있도록 실시간 업데이트 될 수 있어야 한다.

위에서 제시한 서비스 모형 <그림 1>이 가능하도록 수확량예측모형과 가격예측모형을 머신러닝 기술을 통해 구현하였으며 먼저 연구에 사용된 자료와 모델링 방법을 살펴보고 예측 결과를 실제 결과와 비교해 보도록 하겠다.

그림 1. 수급예측 서비스 모형



2.1. 연구에 사용된 자료

머신러닝 학습을 위해 스마트팜 시설재배 오이의 생육조사 데이터와 온실 환경데이터 그리고 전국 도매시장 경락가격 데이터를 전처리, 가공 및 표준화(평균이 0이고, 표준편차가 1이 되도록 변환)하여 사용하였다. 본 연구결과에서는 머신러닝 “학습”데이터와 “검증”데이터로 표현하기로 한다.

2.1.1. 수확량 예측모형에 사용된 자료

2018~2020년 경기도 오산, 평택, 안성 지역 9개 농가 33개 작기의 온실 환경데이터와 생육조사(경기도 농업기술원) 데이터를 사용하였다(표 1).

표 1. 연구에 사용된 작기 데이터

연도		2018			2019			2020		
작기		봄	여름	가을	봄	여름	가을	봄	여름	가을
봄합계	14	2			5			7		
여름합계	4		-			2			2	
가을합계	15			2			5			8

봄, 가을 각 3개 작기와 여름 1개 작기, 총 7개의 작기 데이터는 검증데이터로 활용하였고 나머지 26개 작기 데이터를 학습데이터로 사용하였다. 검증에 사용된 작기 데이터는 학습에 노출되지 않았다.

가) 온실 환경 데이터

온실 환경데이터는 시설 내·외부의 환경센서에 의해 분 단위로 측정된 데이터 중에서 광합성에 주로 관여하는 항목을 가공하여 활용하였으며 가공 항목과 기준을 <표 2>에 정리하였다. 상대습도는 온도에 따른 절대습도와 포화습도를 구한 후 수분부족분으로 변환하여 사용하였다. 1~5분 단위로 측정된 온실 환경데이터는 1주일 단위로 변환하였고 주 1회 진행한 생육조사일 기준으로 이전 7일간의 데이터를 평균 및 합산하여 주단위 환경데이터로 데이터셋을 구성하였다.

표 2. 온실환경 데이터 항목

	센서수집데이터	학습데이터가공	데이터 가공 기준
온실내부환경	내부온도 내부상대습도 CO2농도 내부일사	주간내부온도(℃) 야간내부온도(℃) 주야간온도차(℃) 주간수분부족분(g/m ³) 야간수분부족분(g/m ³) 주간CO2농도(ppm) 내부일사평균(W/m ²) 누적내부일사(W/m ²)	주간 온실 내부온도 평균 야간 온실 내부온도 평균 일출 전 후 4시간 평균 온도 차이 주간 포화수분과 절대습도 차이 야간 포화수분과 절대습도 차이 일출 후 일몰 전 누적 CO2 일출 후 일몰 전 일사량 5분단위 일사량의 1주일간 누적
온실외부센서	외부온도 외부상대습도 외부일사	주간외부온도(℃) 야간외부온도(℃) 외부일사평균(W/m ²) 누적외부일사(W/m ²)	주간 온실 외부온도 평균 야간 온실 외부온도 평균 일출 후 일몰 전 일사량 5분단위 일사량의 1주일간 누적

나) 생육조사 데이터

주 1회 농가별 3반복 각 3개체의 총 9개 개체에서 수집된 데이터를 가공하여 학습에 사용하였고 항목 및 가공 기준을 <표 3>에 정리하였다. 작기진행주차로 정식일 기준으로 주차를 표현하였으며 줄기직경과 엽장/엽폭은 생장점 아래 10번째 마디에서 측정된 것이다.

표 3. 생육조사 데이터 항목

	생육조사데이터	학습데이터가공	데이터 가공 기준
영양생장	정식일 마디수 줄기직경 엽장/엽폭 엽수	작기진행주차(주) 마디수(개) 줄기직경(cm) 엽면적(cm ²) 엽수(개)	정식일 기준으로 주차로 표현 생장점 아래 10번째 마디 경경 측정 마디의 앞으로 면적 근사 ¹⁾ 적엽 개수 반영
생식생장	착과수 수확량 적과수 열매길이	착과수(개) 수확량(개)	생육조사 당시의 착과 개수 전 주 열매길이 15cm 이상 수확으로 간주

주: 1) 엽면적 = $0.88 \times \text{엽장} \times \text{엽폭} - 4.27$ Blanco, F. "Estimation Of Leaf Area for Greenhouse Cucumber"

2.1.2. 가격예측모형에 사용된 자료

농업경제연구원 농업관측 통계정보시스템에서 제공하는 전국도매시장 경락가격자료 raw 데이터를 사용하였으며 '06~'19년 자료를 학습데이터로, '20년 자료를 검증데이터로 활용하였다. 전체 거래량의 99%를 차지하는('20년 기준) 백다다기, 가시오이, 취청 3개 품종의 전국 도매시장 거래량과 거래가격 데이터를 가공하여 사용하였다. <표 4>는 가격예측모델 학습에 사용된 항목이다. 거래가격은 단위기간 거래금액의 합을 거래량의 합으로 나누어 산출하였으며 물가상승분은 별도로 학습에 반영하지 않았다.

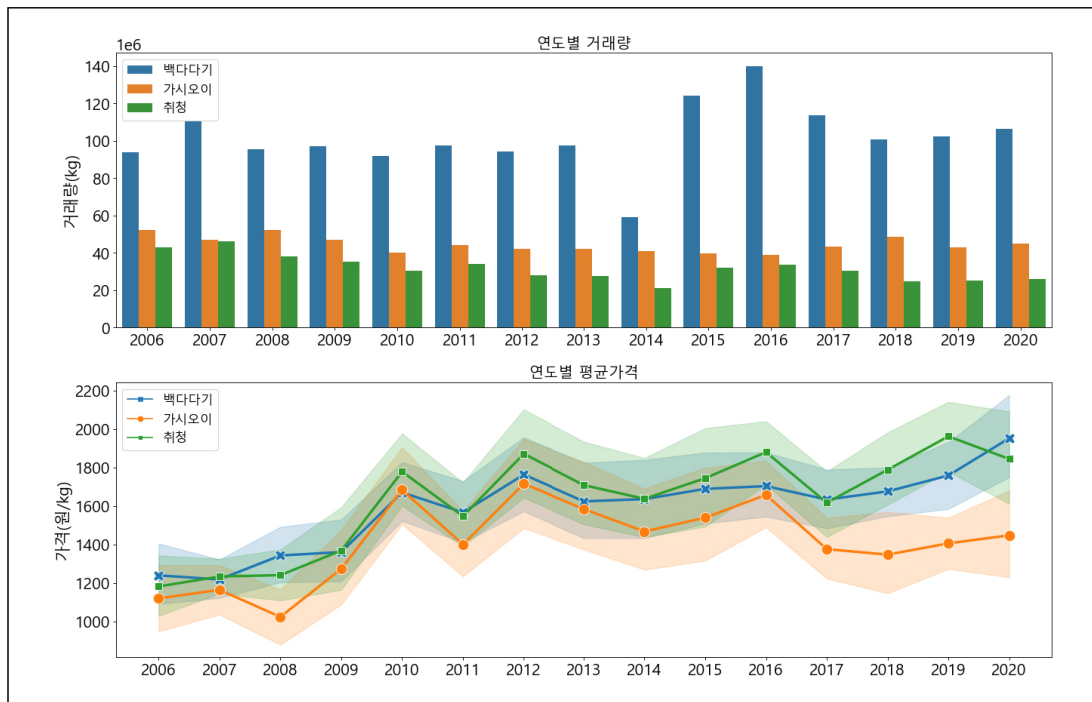
표 4. 가격예측모델 데이터 항목

	도매가격정보	학습데이터가공	데이터 가공 기준
전국도매시장 경락가격정보	거래년월일 품목명 출하지명 평균가격 거래량 단위코드	거래년월일 품목명 출하지역 거래가격(원/kg) 거래량(kg) 거래금액(원) 수요량지표	일일 거래량으로 모델 훈련 및 검증 백다다기, 가시오이, 취청 3개 품목 전국 13개 시도로 변환 일거래금액 ÷ 일거래량 일합계[거래량 × 평균가격] 일거래금액 ÷ 연평균거래가격

수요량지표는 수요가 예측에 반영되도록 하는 지표로써 거래금액을 연평균 거래가격으로 나누어 가격에 의한 기대수요량을 학습에 반영될 수 있도록 하였다(그림 8). 품종별 연간 거래

량은 백다다기는 증가하는 추세이고 취청은 감소하는 추세임을 <그림 2>에서 확인할 수 있으며 2012년 이후로 가격 변동은 크지 않음을 알 수 있다. 물가상승분은 거래량에 의한 가격예측에 왜곡이 있을 수 있어 배제하였다.

그림 2. 오이 품종별 연간 거래량, 가격 추이



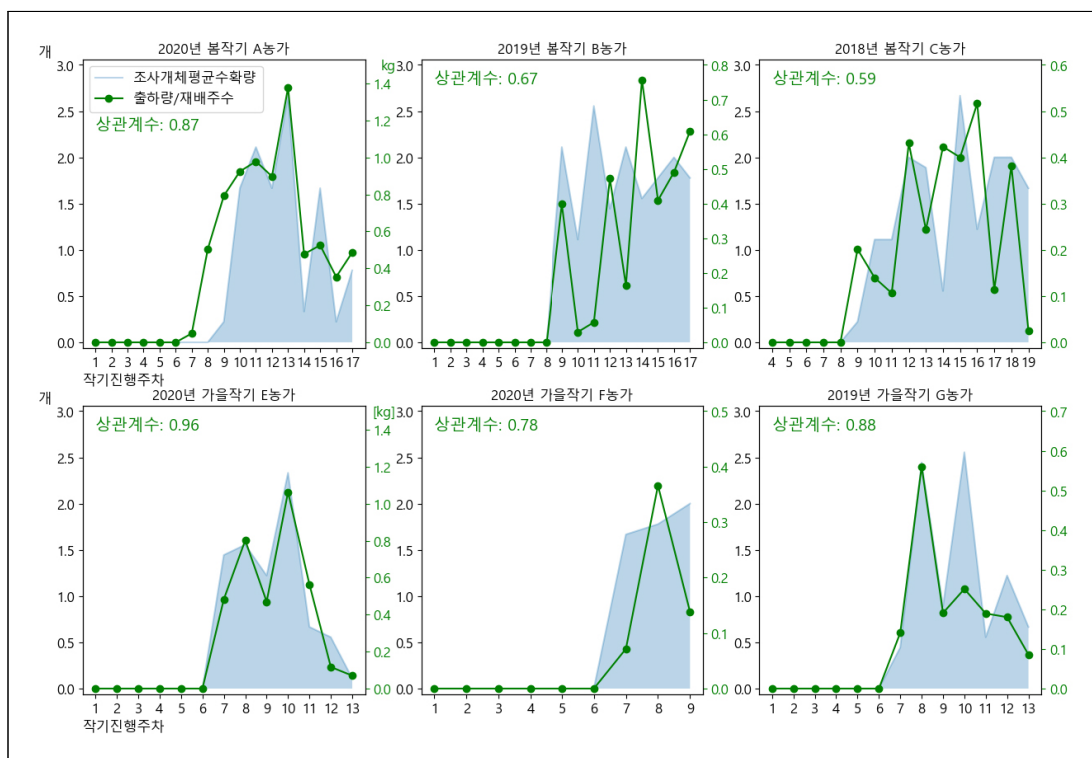
2.2. 모델링 방법

수확량과 가격예측 모델링은 파이썬 라이브러리를 활용한 머신러닝으로 진행하였고 딥러닝 신경망 모델과 Random Forest, SVR(Support Vector Regressor) 모델 결과를 비교하였다. 신경망 모델링 환경으로는 파이썬 케라스(Keras)-텐서플로(TensorFlow)엔진을 사용하였고, SVR(Support Vector Regressor)과 RF(Random Forest) 모델은 사이킷런(Scikit-learn) 라이브러리를 사용하였다.

2.2.1. 오이 수확량 예측 모형

학습에 사용된 수확량 데이터는 조사 개체로부터 수집한 샘플 데이터이며 실제 출하량과의 상관관계를 먼저 확인해보았다. <그림 3>의 왼쪽 y축은 1개 조사 개체당 수확량을 나타내고, 오른쪽 y축은 1개체당 출하중량(kg)을 나타내며 농가의 주별 출하중량(kg)을 재배주수로 나타낸다. 출하내역은 관할 농협에 등록된 출하량만 보여주기 때문에 직거래 등의 미등록 내용이 있을 수 있으며 재배주수에 포함되지 않은 동일 농장주의 내역이 함께 포함되어 있을 수 있다. 따라서 그래프 스케일을 조사개체 수확량과 비슷하게 맞추어 주차별 추세를 확인할 수 있도록 하였다. 작기별 조사개체 샘플 수확량과 출하량의 상관계수는 평균 0.692이며 상관관계가 있다고 판단하고 학습을 진행하였다.

그림 3. 조사개체 수확량과 농가 출하량 비교



수확량 예측모형은 재배환경이 상이한 봄작기, 여름작기, 가을작기 모형을 각각 학습하였으며 작기주차에 따른 착과수와 수확량을 예측하도록 학습되었다. <표 5>에서와 같이 입력변수를 구분하여 세 가지 상황을 가정하였다. 먼저 온실내부 환경정보와 생육정보를 모두 입력변수로 활용하는 경우로 작기 진행 중에 수확량을 예측하는 상황을 가정하였다. 두 번째는 온실내부 환경정보만을 입력변수로 활용하는 경우로 과거의 데이터나 온실의 평균적인 환경데이터를 활용하여 수확량을 예측하는 경우이다. 마지막으로 외부환경정보만을 입력변수로 사용하는 경우는 온실에 환경 센서가 없어 외부 기상조건만을 변수로 사용할 수 있는 경우이다. 두 번째와 세 번째의 경우는 예측서비스 제공 시 따로 입력변수를 입력하지 않고 서버로 기상정보나 환경데이터를 시스템을 연동하여 데이터를 전송하는 방식으로 구현 가능할 것이다.

표 5. 입력변수 항목 구분

구분	입력변수
1. 온실내부환경과 생육정보를 입력변수로 사용	주간내부온도, 야간내부온도, 주야간온도차, 주간수분부족분, 야간수분부족분, 주간CO2농도, 내부일사평균, 누적내부일사, 작기진행주차, 마디수, 줄기직경, 엽면적, 엽수
2. 온실내부환경만 입력변수로 사용	주간내부온도, 야간내부온도, 주야간온도차, 주간수분부족분, 야간수분부족분, 주간CO2농도, 내부일사평균, 누적내부일사, 작기진행주차
3. 외부환경만 입력변수로 사용	주간외부온도, 야간외부온도, 주간/야간 외부상대습도 외부일사평균, 누적외부일사, 작기진행주차

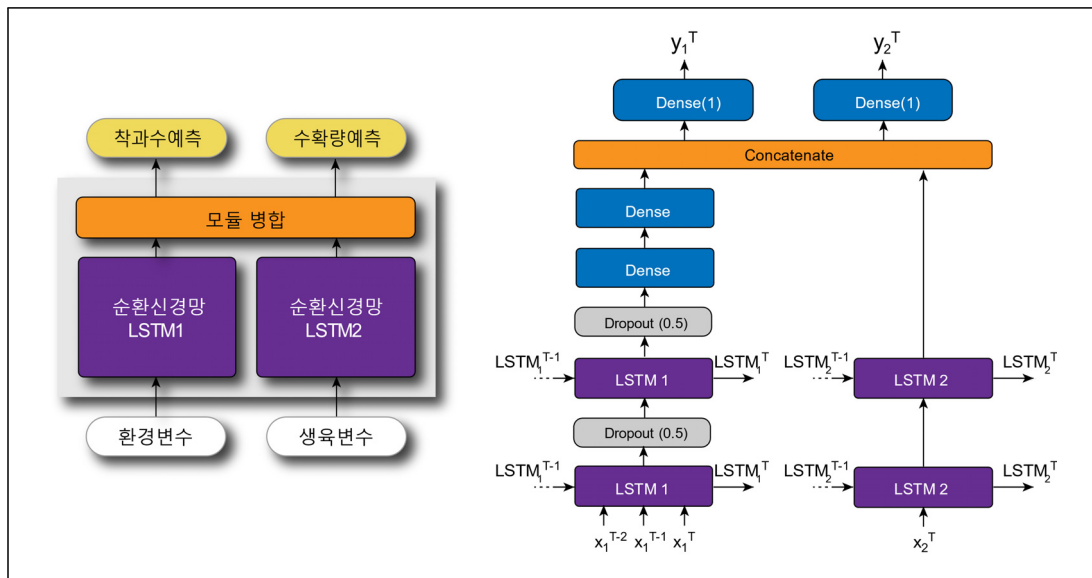
모델링은 시계열 분석을 위해 순환신경망 LSTM(Long Short-Term Memory)을 사용하였으며 <그림 4(a)>에서와 같이 환경변수를 처리하는 순환신경망 모듈과 생육변수를 처리하는 순환신경망 모듈을 따로 구성하여 출력 전 단계에서 병합하였다. 순환신경망의 layer 구성은 <그림 4(b)>에 표현하였다. 환경변수 순환신경망 모듈은 2개의 LSTM층과 2개의 Dense층으로 구성되어 있으며 각 LSTM층에 Dropout(=0.5)층을 추가하여 과대적합 감소를 꾀하였다. 생육변수 순환신경망 모듈은 입력변수로 생육변수와 작기진행주차 정보를 받는다. LSTM 2개층으로 구성되며 Dropout층은 추가하지 않았다.

두 순환신경망 모듈의 출력은 Concatenate층에서 병합되어 2개의 Dense출력층 -착과수와 수확량- 을 거치게 된다. 모델의 모든 Dense층은 활성화 함수로 'Relu' activation을 사용하였고 출력층 Dense도 'Relu' activation을 통해 음수값은 0이 되도록 하였다. 비용함수

(Cost Function)를 최소화하기 위한 옵티마이저로 Adam optimizer를, loss는 MSE(Mean Squared Error, 평균제곱오차)를 사용하였다.

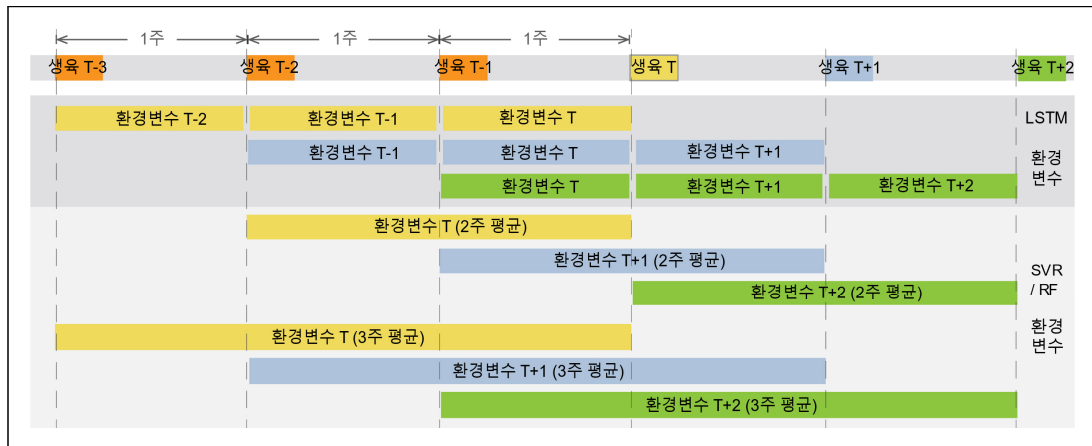
〈그림 4(b)〉에서 입력 $x_1^{T-2}, x_1^{T-1}, x_1^T$ 는 환경입력변수이며 각각 관측 시점의 3주전부터의 1주평균 데이터이고, x_2^T 는 생육입력변수, y_1^T 와 y_2^T 는 출력인 착과수와 수확수를 의미한다.

그림 4. 수확량 예측 순환신경망 모델



〈그림 5〉는 환경입력변수와 생육입력변수의 시계열 데이터 시퀀스를 표현한 것이다. 환경변수 순환신경망 모델은 시계열 데이터 시퀀스로 3개의 타임스텝 $x_1^{T-2}, x_1^{T-1}, x_1^T$ (하나의 타임스텝은 1주)를 하나의 학습데이터로 입력하게 되며 다음 학습데이터는 한 스텝씩 Shift한 $x_1^{T-1}, x_1^T, x_1^{T+1}$ 를 입력하게 된다. 비교모델로 사용한 SVR과 RF모델의 입력변수는 시계열 시퀀스가 아니므로 환경변수를 1주, 2주평균, 3주평균으로 가공하여 학습한 후 예측 결과를 비교하였다.

그림 5. 입력변수의 시계열 시퀀스



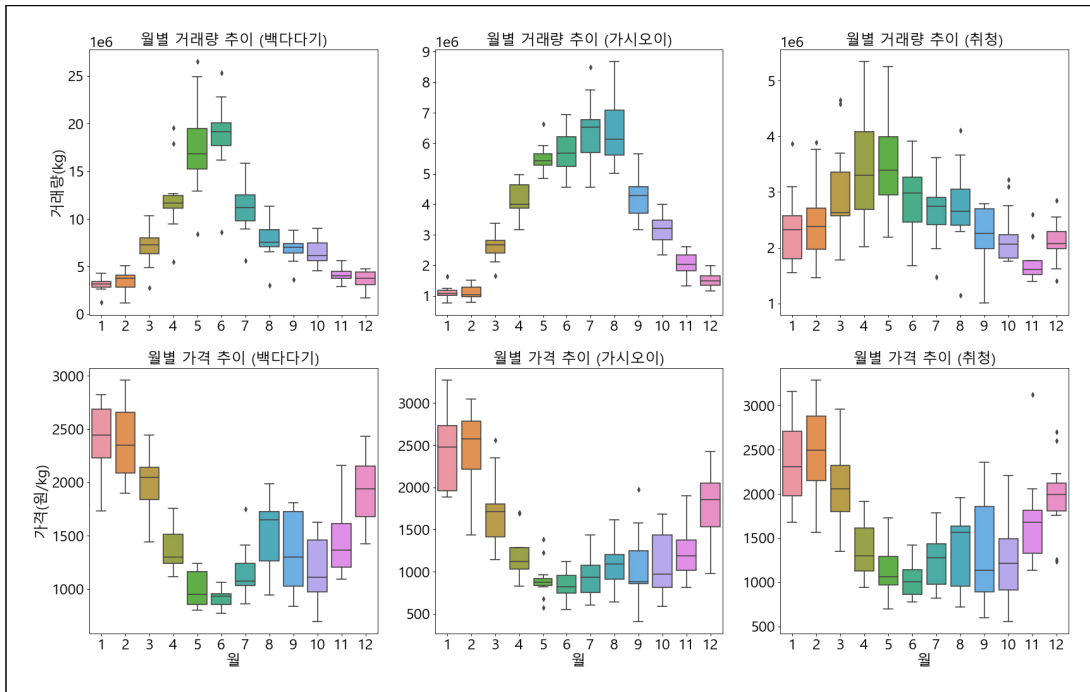
2.2.2. 가격예측 모형

농산물 가격예측은 ARIMA모형 등의 시계열예측으로 주로 진행되어 왔고 최근에는 순환 신경망(Recurrent Neural Network)을 이용한 시계열예측 연구가 활발히 이루어지고 있다. 기상변수와, 유가, 전년 수확량 등의 변수를 활용한 LSTM 순환신경망 모델로 향후 7일 가격을 예측한 연구가 있으며, 기상변수와 도매시장 가격을 변수로 RNN 순환신경망을 적용하여 배추 가격을 예측한 연구가 있다.

본 연구에서는 기상변수를 사용하지 않고 도매시장 거래량과 가격 자료만으로 중장기 가격을 예측할 수 있는 모델을 적용하여 검증해보았다. 입력 변수를 최소화하고 학습 비용이 낮은 모델을 사용하여 요청이 들어왔을 때 실시간으로 최신정보를 학습하여 웹서비스를 통해 가격 예측결과를 제공할 수 있는 모델을 연구 방향으로 설정하였다.

〈그림 6〉의 '06~'20년 월별 거래량과 가격 추이를 보면 뚜렷한 계절성과 품종별로 거래량 추이가 상이한 것을 확인할 수 있다. 본 연구에서는 이러한 품종별 거래량과 가격의 계절성을 예측에 적용하였다.

그림 6. 오이 품종에 따른 월별 거래량 및 가격 추이('06~'20년)



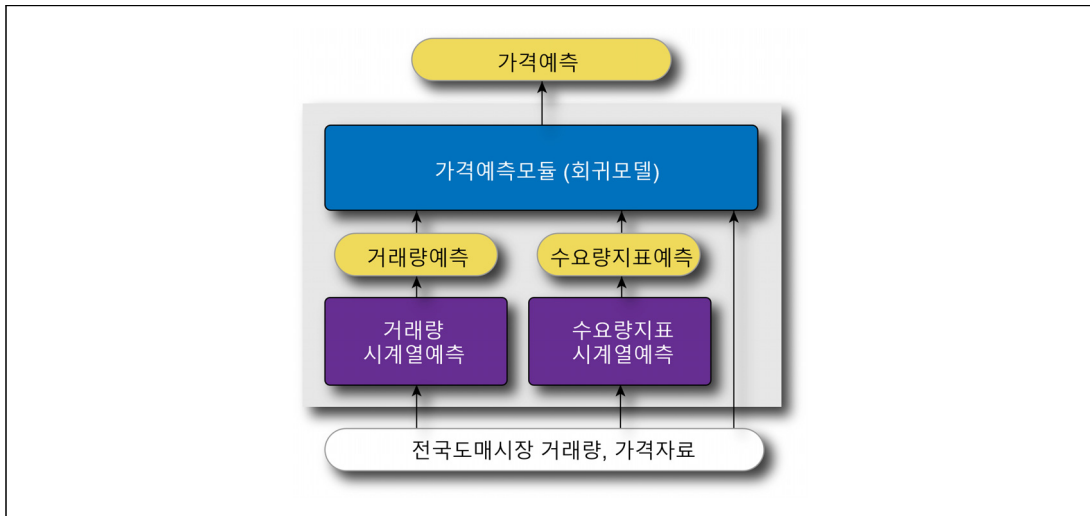
가격예측모형 <그림 7>은 거래량과 가격, 그리고 계절성 반영을 위한 거래일자를 입력변수로 받아 가격을 예측하며 거래량 시계열예측, 수요량지표 시계열예측, 가격예측의 3개 모듈로 구성되어 있다. 시계열예측 모듈은 파이썬 Prophet 라이브러리를 활용하였다. Prophet은 페이스북에서 개발한 시계열예측 라이브러리로 학습속도가 빠르고 계절성이 있는 모델 예측에 높은 성능을 보이며 계절성, 트렌드 등의 파라미터로 모델 최적화가 가능하다. 가격예측 모듈은 회귀모델이며 시계열예측 모듈에서 예측된 거래량과 수요량지표, 그리고 거래일자를 입력변수로 최종 가격을 예측한다. 본 연구에서는 SVR, RF, Lasso회귀모델의 결과를 비교하였다. 각각의 회귀모델은 사이킷런의 GridSearchCV를 활용하여 최적의 하이퍼파라미터¹⁾를 적용하였다.

1) SVR: kernel="rbf", C=1000, gamma=1

RF: n_estimator=100

Lasso Regressor: Polynomial degree=5, alpha=0.01(백다다기), alpha=1(가시오이, 취청)

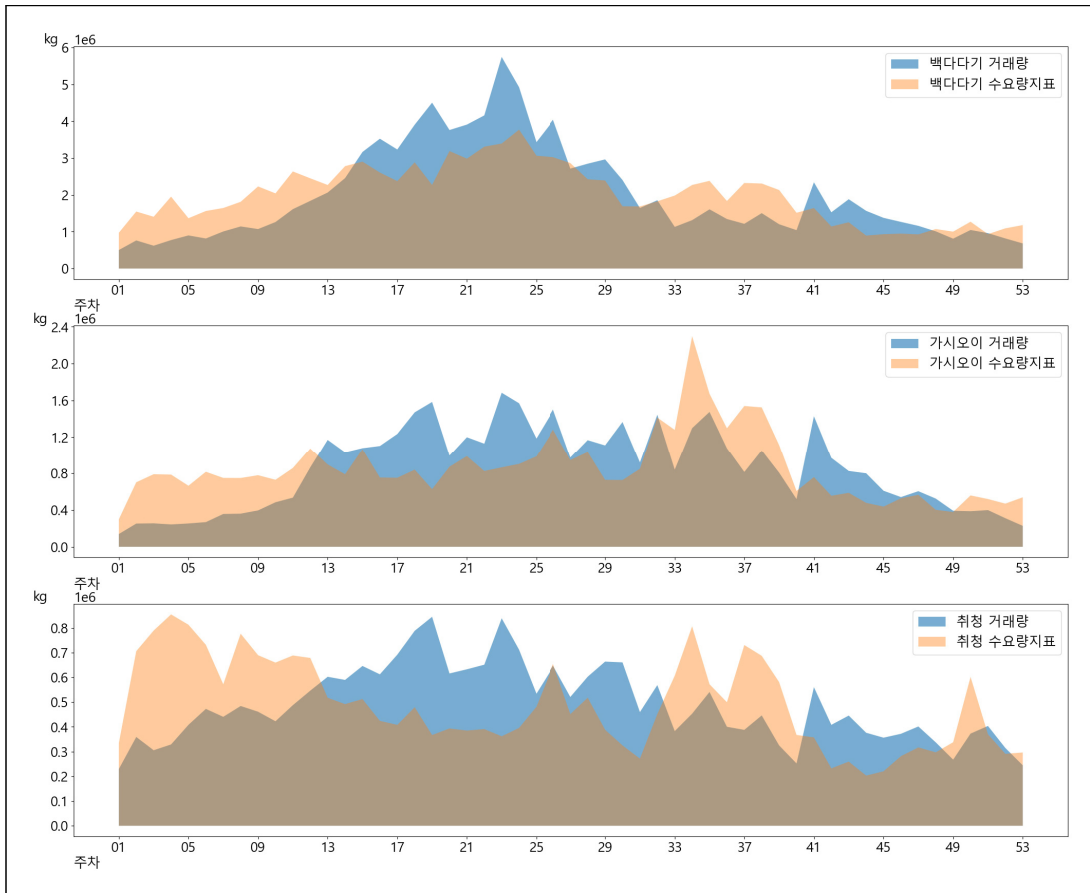
그림 7. 가격예측 모형



가격예측에 수요를 반영하기 위해 ‘2.1.2. 가격예측모형에 사용된 자료’에서 언급된 수요량지표를 활용하였다. 수요량지표 (1)은 기준가격(연평균거래가격)에서의 기대거래량을 예상해보는 방법을 사용하였다. <그림 8>에서 파란색 영역이 실제 거래량(공급물량)이며 수요량지표를 나타내는 주황색 영역이 더 높은 경우는 수요에 비해 공급물량이 부족하다고 볼 수 있고, 반대로 파란색 영역이 더 높은 경우는 수요보다 공급이 많아 가격이 기준가격 보다 낮은 것으로 추정할 수 있다.

$$(1) \text{수요량지표} = \text{거래량} \times \frac{\text{거래가격}}{\text{연평균거래가격}}$$

그림 8. 주차별 거래량과 수요량지표 비교(2020년)



연구 결과에서는 거래량과 수요량지표의 시계열예측결과를 오이 품종별로 확인해보고 예측된 거래량과 수요량지표를 통해 가격예측모듈에서 최종 예측된 결과를 회귀모델 별로 비교해보았다.

3. 연구 내용 및 결과

수확량예측모형과 가격예측모형은 시계열예측 모형이므로 머신러닝을 위한 학습세트와 검증세트를 무작위 추출하지 않고 작기와 기간을 선택하여 데이터의 시퀀스를 유지하였다. 수확량 예측을 위한 검증세트는 작기별로 연도와 농가가 골고루 검증에 포함될 수 있도록 구성하였고, 가격예측을 위한 데이터세트는 '06~'19년 데이터를 학습세트로, '20년 데이터를 검증세트로 구성하여 결과를 확인하였다.

검증결과의 성능 지표는 (2) MAE(절대오차평균, Mean Absolute Error)와 (3) 결정계수 R^2 (결정계수, R-Squared)을 사용하였다.

$$(2) MAE = \frac{\sum |y - \hat{y}|}{n}$$

$$(3) R^2 = \frac{\sum (\hat{y} - \bar{y})^2}{\sum (y - \bar{y})^2}$$

y 는 관측값, $\bar{y}$ 는 관측값의 평균, $\hat{y}$ 는 예측값을 나타낸다. 성능 지표는 검증세트의 지표만 표기하였다.

3.1. 오이 수확량 예측 결과

수확량 예측모형으로 작기진행주차별 착과수와 수확량을 예측하였으며, 결과의 값은 오이 1개체에서 관측되는 열매의 개수이다. 예측 모델은 봄작기, 여름작기, 가을작기로 나누어 학습하였으며 학습에 노출되지 않은 작기 데이터로 검증을 진행하였다. 검증은 봄 3작기, 여름 1작기, 가을 3작기 데이터를 사용하였다.

33개 작기의 총 4,488개의 데이터세트를 사용하였으며 학습세트 80%, 검증세트 20%로 구성되어있다(표 6).

표 6. 수확량 예측에 사용된 학습세트와 검증세트 개수

구분	봄작기	여름작기	가을작기
학습세트	2,021	240	1,379
검증세트	462	65	321
합계	2,483	305	1,700

수확량 예측 테스트는 아래의 항목으로 진행하였고 결과를 그래프와 예측성능으로 정리하였다.

가) 계절별 작과수와 수확량 예측

환경조건이 상이한 봄, 여름, 가을 작기를 구분하여 예측 결과를 확인하였다.

- 봄 3작기(A, B, C농가): 작과수 <그림 9>, <표 7>, 수확량 <그림 10>, <표 8>
- 여름 1작기(D 농가): 작과수, 수확량 <그림 11>, <표 9>
- 가을 3작기(E, F, G농가): 작과수 <그림 12>, <표 10>, 수확량 <그림 13>, <표 11>

LSTM 순환신경망 모델이 작기별로 안정적인 예측결과를 보여주는 반면 SVR, RF 모델은 검증세트에 따라 편차가 큰 것을 확인할 수 있다.

나) 예측모형에 따른 결과 비교

LSTM 순환신경망(그래프의 1행), SVR(2행), RF(3행) 모델의 예측결과를 비교하였다. SVR과 RF 모델은 사이킷런의 GridSearchCV를 활용하여 최적의 하이퍼파라미터를 적용하였다.

LSTM 순환신경망 모델은 1주일 간격의 3개의 환경데이터를 하나의 시퀀스 데이터로. SVR, RF 모델은 입력 환경변수를 1주일, 2주평균, 3주평균으로 각각 학습 및 검증하였고, 그 중 결과가 가장 좋은 3주평균의 예측 결과를 비교하였다.

SVR, RF 모델의 경우 전제 데이터세트에서 검증세트를 무작위 추출하여 검증하는 교차검증(Cross Validation) 결과의 성능 수치는 LSTM 순환신경망 모델과 비슷한 수준으로 나왔으나 시퀀셜 검증 데이터의 경우 낮은 예측 성능을 보였다.

다) 입력변수에 따른 결과 비교

결과 그래프의 파란색 영역은 실제 값이며 주황색은 온실 내부환경변수와 생육변수를 모두 입력변수로 예측한 결과, 초록색은 내부환경변수로만 예측한 결과이며, 빨간색은 외부환경변수로만 예측한 결과이다. 내부환경변수와 생육변수를 모두 사용한 결과의 예측 성능이 높게 나타났으며 SVR, RF 모델은 입력변수에 따른 편차가 매우 큰 것을 확인할 수 있다.

그림 9. 봄작기 착과수 예측 결과

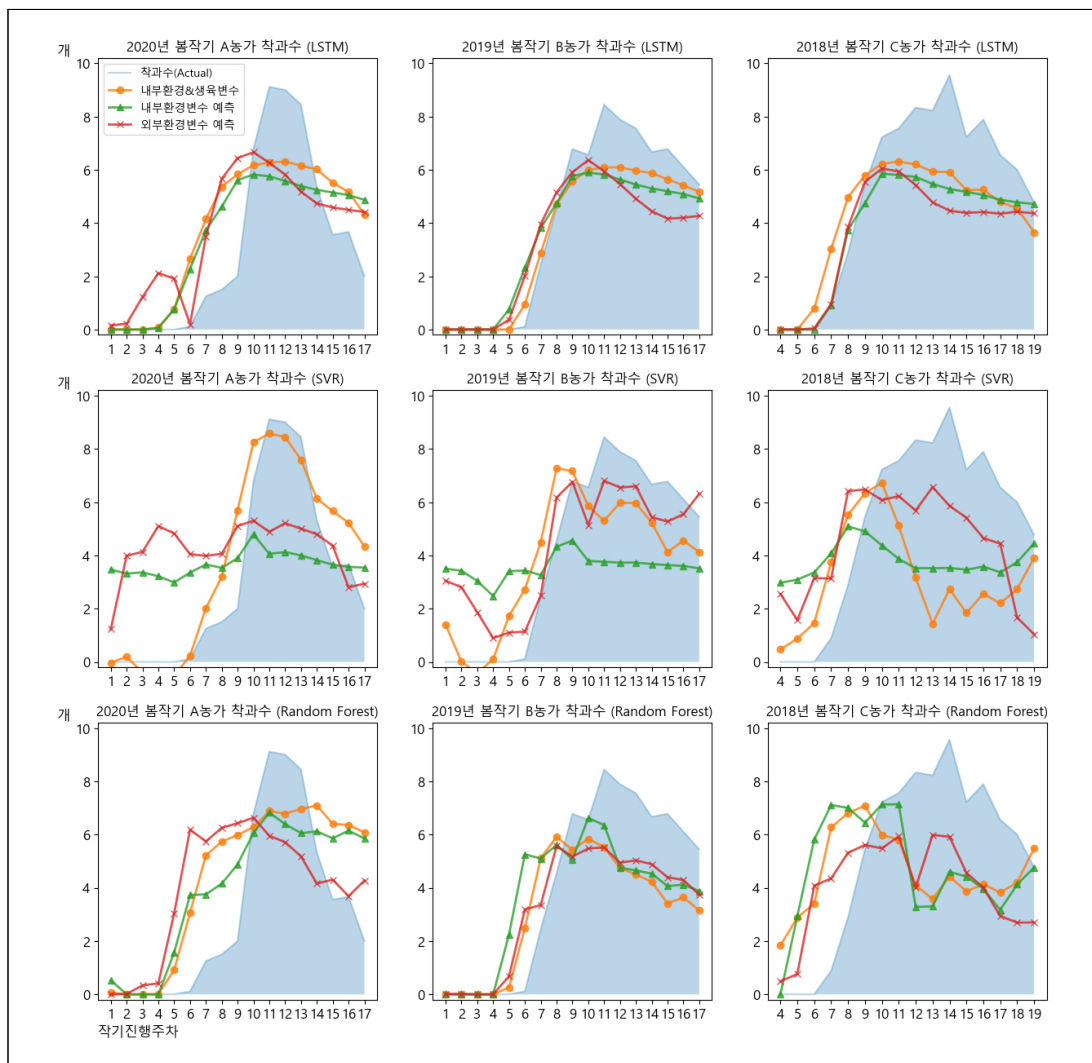


표 7. 봄작기 착과수 예측성능 비교

봄작기 착과수 예측	LSTM		SVR		Random Forest	
	MAE	R ²	MAE	R ²	MAE	R ²
내부환경&생육변수	1.39	0.69	0.58	0.45	0.61	0.38
내부환경변수	1.52	0.64	0.85	0.11	0.71	0.14
외부환경변수	1.82	0.53	0.64	0.39	0.68	0.29

그림 10. 봄작기 수확량 예측 결과

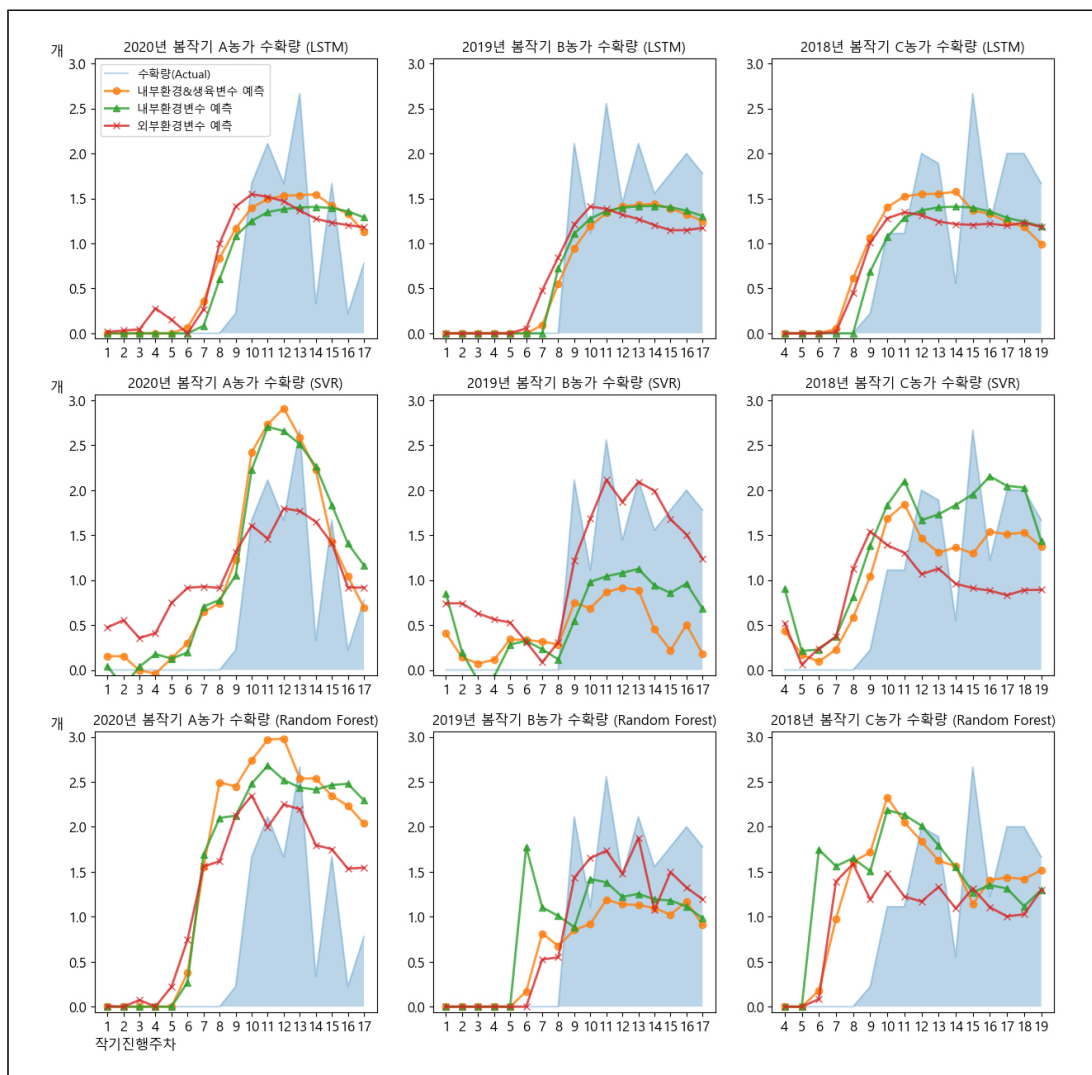


표 8. 봄작기 수확량 예측성능 비교

봄작기 수확량 예측	LSTM		SVR		Random Forest	
	MAE	R ²	MAE	R ²	MAE	R ²
내부환경&생육변수	0.58	0.43	0.64	0.21	0.77	-0.09
내부환경변수	0.57	0.43	0.65	0.27	0.88	-0.17
외부환경변수	0.64	0.37	0.70	0.30	0.67	0.20

그림 11. 여름작기 착과수, 수확량 예측 결과

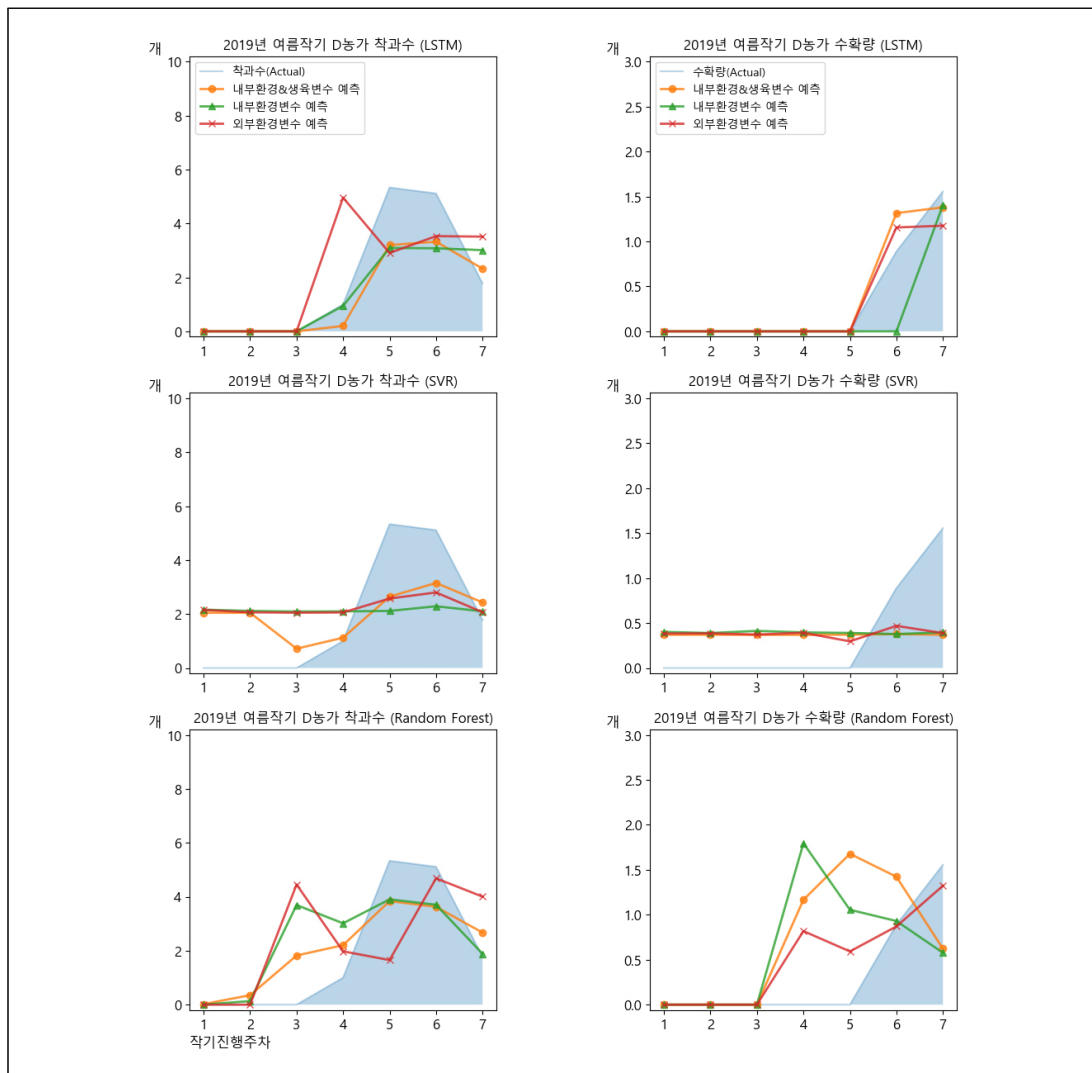


표 9. 여름작기 착과수, 수확량 예측성능 비교

여름작기 예측		LSTM		SVR		Random Forest	
		MAE	R ²	MAE	R ²	MAE	R ²
착과수	내부환경&생육변수	0.93	0.61	0.63	0.45	0.52	0.54
	내부환경변수	0.99	0.55	0.96	0.23	0.90	0.33
	외부환경변수	1.54	0.16	1.86	0.09	0.57	0.55
수확량	내부환경&생육변수	0.20	0.56	0.73	-0.07	0.73	-0.01
	내부환경변수	0.23	0.39	0.81	0.28	0.78	0.06
	외부환경변수	0.23	0.52	0.72	-0.03	0.65	0.22

그림 12. 가을작기 착과수 예측 결과

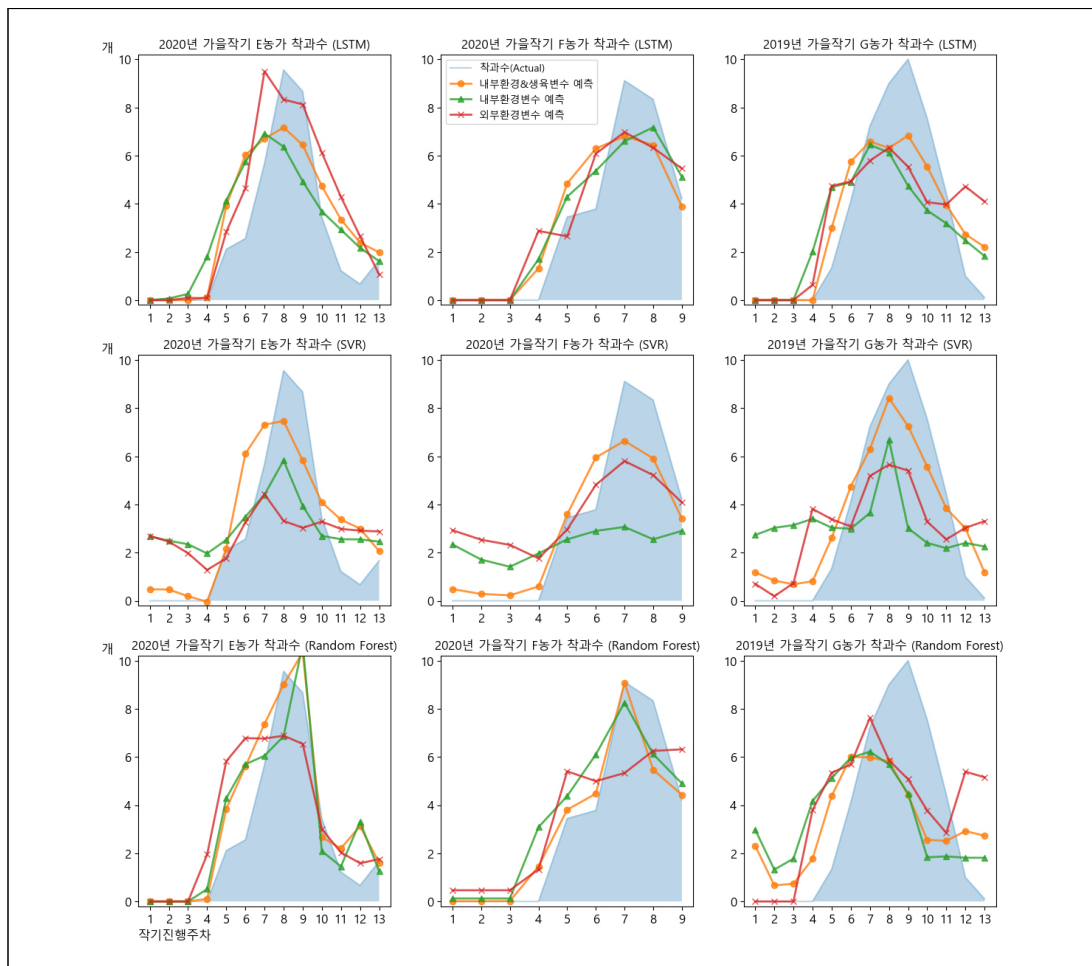


표 10. 가을작기 착과수 예측성능 비교

가을작기 착과수 예측	LSTM		SVR		Random Forest	
	MAE	R ²	MAE	R ²	MAE	R ²
내부환경&생육변수	1.50	0.64	0.41	0.62	0.49	0.53
내부환경변수	1.70	0.54	0.69	0.19	0.55	0.42
외부환경변수	1.79	0.51	0.64	0.27	0.55	0.40

그림 13. 가을작기 수확량 예측 결과

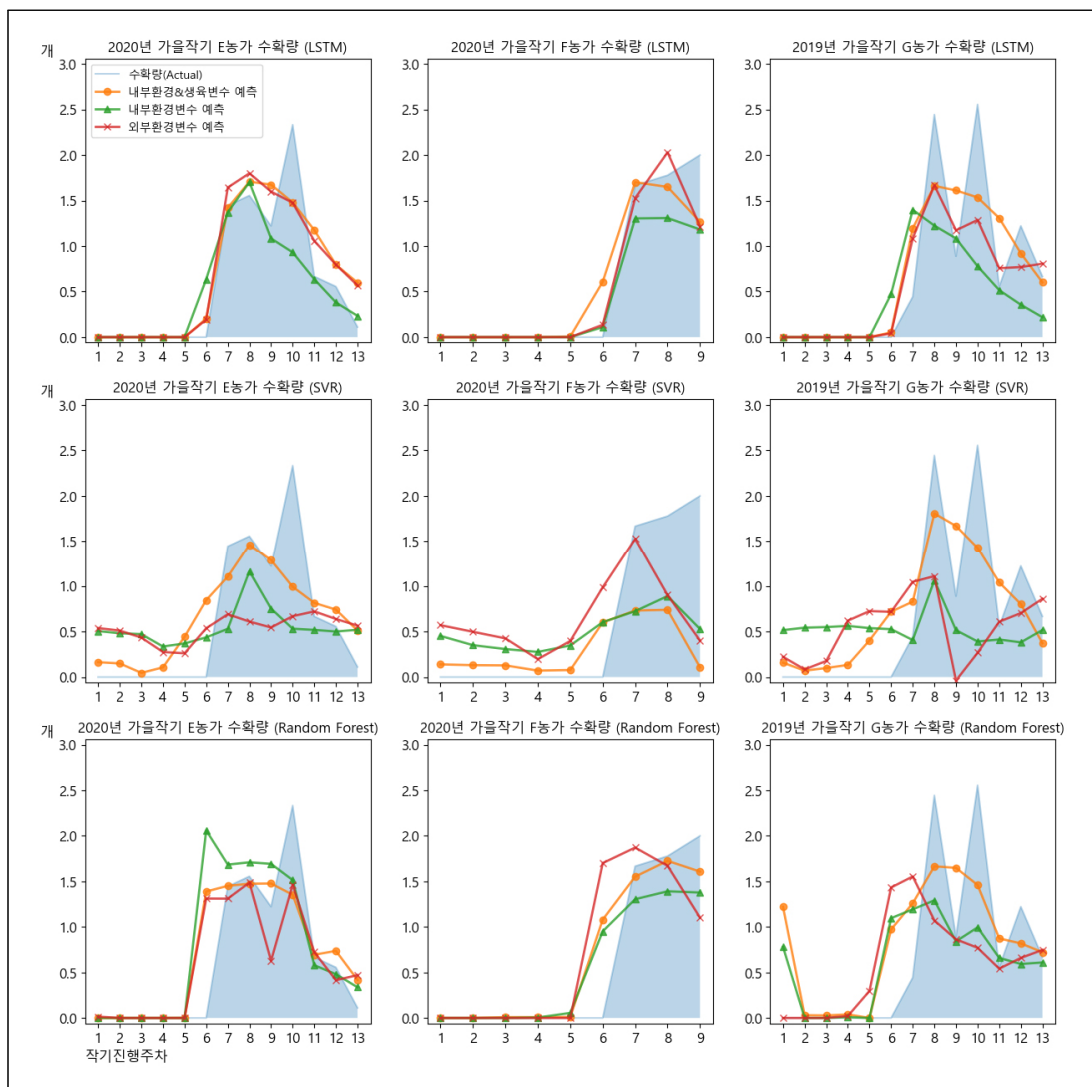


표 11. 가을작기 수확량 예측성능 비교

가을작기 수확량 예측	LSTM		SVR		Random Forest	
	MAE	R ²	MAE	R ²	MAE	R ²
내부환경&생육변수	0.47	0.46	0.56	0.28	0.57	0.29
내부환경변수	0.51	0.34	0.70	0.10	0.56	0.24
외부환경변수	0.45	0.45	0.71	0.06	0.58	0.20

3.2. 오이 가격예측 결과

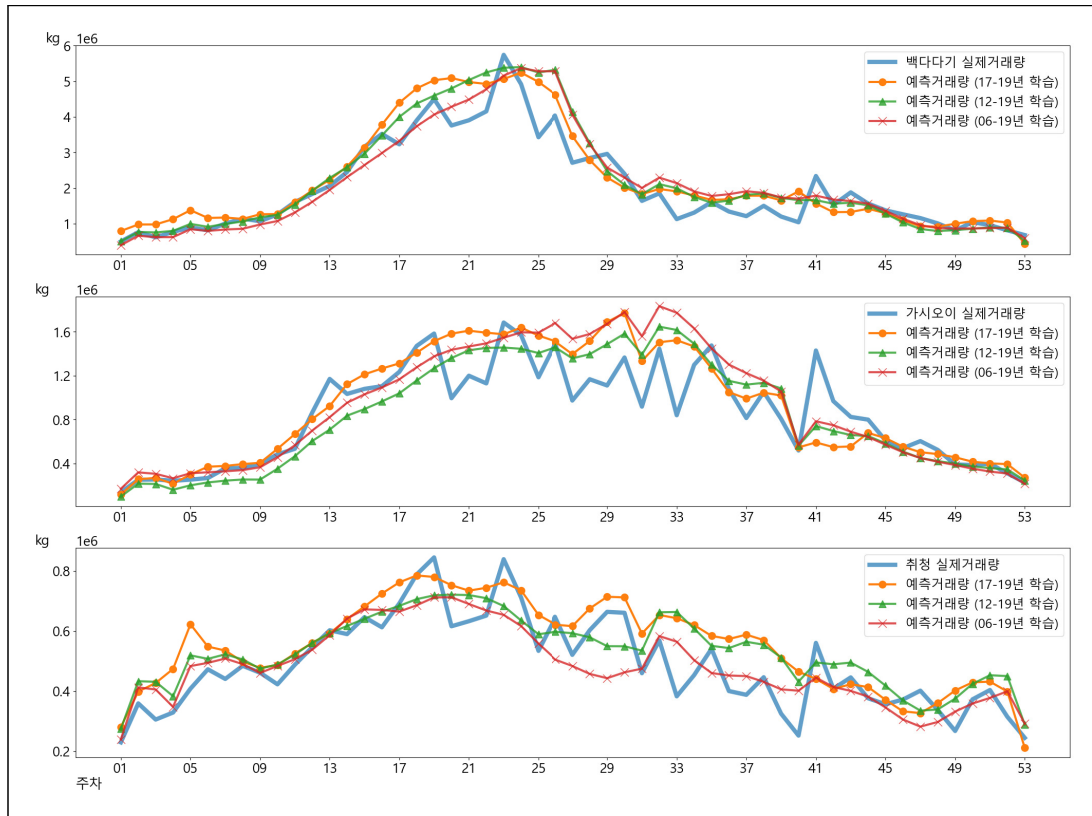
가격예측은 품종별로 결과를 확인하였다. <표 12>는 가격예측에 사용된 '06~'20년 데이터 세트의 개수이다.

표 12. 가격예측에 사용된 학습세트와 검증세트 개수

구분	백다다기	가시오이	취청
학습세트	4,594	4,601	4,599
검증세트	366	366	366
합계	4,960	4,967	4,965

거래량과 수요량지표의 시계열예측 결과를 그래프로 확인하고 최종 가격예측 결과를 그래프와 예측 성능수치로 확인하였다.

그림 14. '20년 거래량 시계열예측 결과



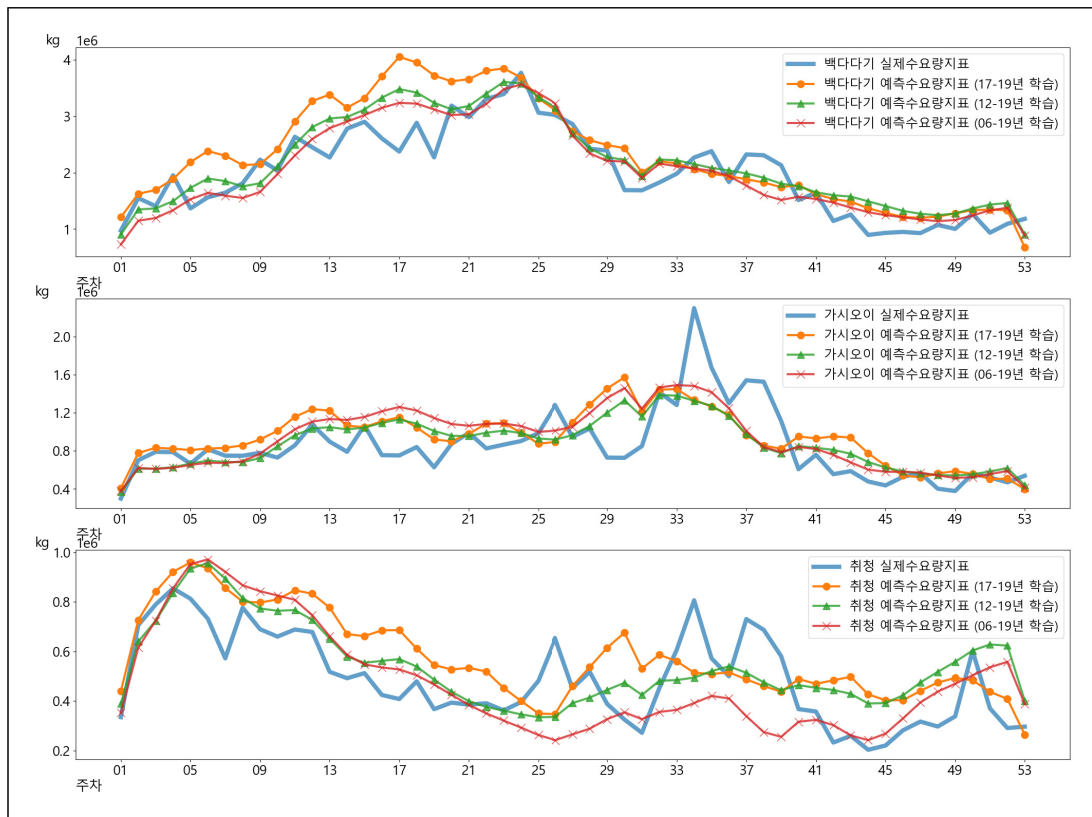
가) 거래량과 수요량지표 시계열예측

거래량과 수요량지표는 30일 단위로 학습데이터를 업데이트하여 예측하였다. 시계열예측은 과거 데이터의 계절성과 트렌드를 반영하므로 학습데이터의 기간을 '06~'19년, '12~'19년, '17~'19년으로 변경하여 예측 결과를 비교하였다. <그림 14>는 거래량 시계열예측 결과이며 취청오이의 경우 학습데이터의 기간에 따라 예측 결과의 차이가 크게 나타나는 것을 확인할 수 있다. 예를 들어 '17~'19년 여름철(27~30주차) 취청오이의 생산량이 이전보다 많았다고 해석이 가능하다.

<그림 15>의 수요량지표도 마찬가지로 학습데이터의 기간을 변경하여 확인하였으며 백다다기의 경우 '17~'19년 학습데이터의 경우 상반기 지표가 실제보다 높게 나오는 것을 확인할

수 있으며 이는 '17~'19년 상반기에 백다다기의 가격이 다소 높았음을 예측할 수 있다. 자기 상관을 이용하는 시계열예측의 경우 결과의 정확도를 높이기 위해 학습데이터의 기간 설정 시 트렌드를 감안한 최적화가 필요하다는 것을 알 수 있다.

그림 15. '20년 수요량지표 시계열예측 결과



나) 가격예측 결과

앞서 시계열예측 모듈에서 예측된 거래량과 수요량지표를 입력변수로 사용하여 가격예측 모듈에서 예측한 결과를 회귀모델별(SVR, RF, Lasso회귀)로 비교하였다. <그림 16>은 '17~'19년 데이터로 학습하여 '20년 데이터로 검증한 결과이며 전체적인 가격추세를 제대로 예측하고 있으나 가격 피크구간의 예측은 미흡한 것으로 나타났다.

〈표 13〉은 모델별 예측 성능 비교이며 SVR모델의 성능이 비교적 높게 나왔다. 품종별 수요와 공급의 시계열예측 최적화에 대한 추가적인 연구가 필요하다.

그림 16. '20년 수요량지표 시계열예측 결과

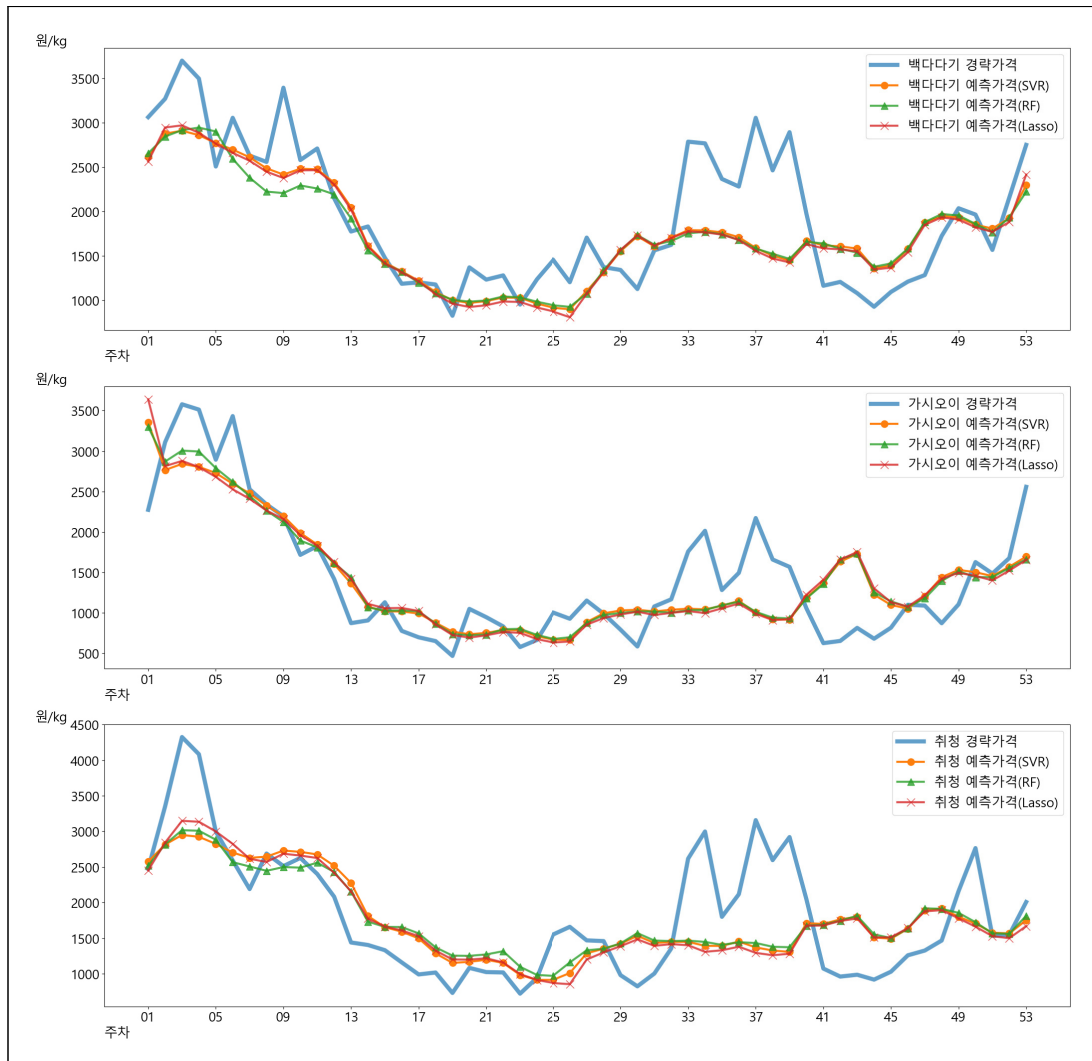


표 13. 가격예측 성능 비교

품종	SVR		Random Forest		Lasso Regressor	
	MAE	R ²	MAE	R ²	MAE	R ²
백다다기	457.08	0.44	511.47	0.26	518.85	0.28
가시오이	568.23	0.30	594.29	0.23	639.82	0.07
취청	455.16	0.50	426.17	0.56	487.85	0.42

4. 결론

머신러닝을 통해 시험 재배가 아닌 실제 재배 환경에서의 데이터 학습으로 비교적 높은 예측 성능을 얻을 수 있는 것을 확인하였다. 연구에 사용된 생육데이터는 병충해, 기상 악조건 등 현실의 모든 환경 변수가 녹아 있는 결과이며 관측데이터가 축적된다면 더 정확한 예측이 가능할 것이다. 다만 일관된 기준으로 정확하게 수집된 데이터가 머신러닝의 성능을 효과적으로 끌어올릴 수 있으므로 실측데이터 수집 계획 단계에서 장기적인 활용 방안이 고려되어야 하겠다.

LSTM 순환신경망이 SVR, RF 모델보다 모든 검증세트에서 안정적인 예측 결과를 보여주었으며 앞으로 순환신경망 기술을 활용한 농업 분야의 시계열예측 연구가 활발히 진행될 것이라 생각한다.

본 연구의 수확량 예측모형은 경기도 시설재배 농가의 데이터를 사용하였으며 향후 전국에서 수집된 데이터를 적용한다면 더 정확한 결과와 다양한 인사이트를 얻을 수 있을 것이다. 센서의 유지보수 등을 통해 정확한 환경데이터가 수집되도록 하는 것도 농업 수급예측 기술 발전의 중요한 요인이 될 것이다.

가격예측모형의 경우 수요와 공급의 시계열 예측 부분을 최적화한다면 실시간 업데이트가 가능한 점과 최소한의 정보로 증장기 예측이 가능하다는 장점으로 예측 정보 제공 서비스에 활용도가 높을 것으로 생각되며 오이 외 다른 농작물에 적용하는 것도 어렵지 않을 것이다.

스마트팜 농가가 확대되고 데이터가 축적되어 연동된 시스템에서 빅데이터를 스스로 학습시키고 활용하는 농업 수급예측 기술의 비약적인 발전이 기대된다.

참고문헌

- 신성호, 이미경, 2018. “LSTM 네트워크를 활용한 농산물 가격예측 모델.” 「한국콘텐츠학회논문지」. 한국 과학기술정보연구원.
- 김진, 이정일, 2019. “딥러닝 기반의 농산물 가격예측 시스템에 대한 연구.” 「한국정보기술학회논문지」 제17권 제6호(JKIIT, Vol.17, No.6).
- 임진모, 김월용, 2018. “인공지능을 이용한 과일 가격예측 모델 연구.” The Journal of the Convergence on Culture Technology(JCCT, 2018.5.26). 한세대학교.
- 채명수, 정성관, 2017. “딥러닝 기반의 기상정보를 반영한 배추 가격예측.” 한국정보통신학회 2017 추계 종합학술대회. 한국과학기술원(KAIST) IT융합연구소.
- Johnathon, S and Tryambak, G. 2020. “Crop Yield Prediction Integrating Genotype and Weather Variables Using Deep Learning.” arXiv:2006.13847.
- Gangopadhyay, T and Shook, J. 2020. “Interpreting the Impact of Weather on Crop Yield Using Attention.” *AI for Earth Sciences Workshop at NeurIPS 2020*.
- Khaki, S and Wang, L. 2020. “A CNN-RNN Framework for Crop Yield Prediction.” *Front. Plant Sci., 24 January 2020*.
- Yin, H. and Jin, D. 2020. “STL-ATT LSTM: Vegetable Price Forecasting Using STL and Attention Mechanism-Based LSTM.” *Agriculture 2020, 10(12), 612*.
- Alhnaity, B and Pearson, S. 2019. “Using Deep Learning to Predict Plant Growth and Yield in Greenhouse Environments.” *International Symposium on Advanced Technologies and Management for Innovative Greenhouses* 10.17660/ActaHortic.2020.1296.55.
- Blanco, F and Folegatti, M. 2005. “Estimation of Leaf Area For Greenhouse Cucumber by Linear Measurements Under Salinity and Grafting.” *Sci. agric.(Piracicaba, Braz.)*.



제1회

농산물 수급 예측모형 경진대회

우수논문집

Chapter 2
자유과제



우수상

생육 실측 자료와 비선형적 기후 효과를 반영한
배추 수급 예측모형 고도화 연구

김기환 · 서병선 · 문진언 · 김정현 · 박지원

생육 실측 자료와 비선형적 기후 효과를 반영한 배추 수급 예측모형 고도화 연구

- 딥러닝과 혼합주기 모형을 활용한 단수예측을 중심으로 -

김기환* 서병선** 문진언*** 김정현**** 박지원*****

Key words

생육 자료(Growth Measurement Data), 기후 변화(Climate Change),
혼합주기 모형(Mixed Data Sampling), 딥러닝 알고리즘(Deep Learning)

ABSTRACT

This paper focuses on improving the forecast accuracy of supply and demand of Korean cabbage, especially focusing on yield estimation of fall and summer(highland) types. The study compares various models which encompass basic models with contemporary meteorological factors to ones with novel methodologies and new source of data. In terms of methodology, Mixed Data Sampling (MIDAS) estimation and Deep Learning algorithm are introduced to incorporate the mixed frequency of growth measurement data and to explore the nonlinear relationship between yield and climate data. While in data, actual growth measurement variables such as leaf length, number of leaves, and presence of pests are collected from Korea Rural Development Administration (RDA) and Korea Rural Economic Institute (KREI) are used. The empirical results show that our proposed methods are effective in the improvement of forecasting accuracy, and the effectiveness is expected to increase in the extension of the sample period.

차례

- | | |
|-------------|------------|
| 1. 서론 | 4. 분석 자료 |
| 2. 배추 생육 특성 | 5. 분석 결과 |
| 3. 분석 모형 | 6. 요약 및 결론 |

* NH농협금융지주 NH금융연구소 책임연구원, 경제학 박사(kgh0120@korea.ac.kr)

** 교신저자. 고려대학교 식품자원경제학과 교수

*** 고려대학교 식품자원경제학과 대학원 석사과정

**** 산업연구원 연구원, 경제학 박사 수료

***** 산업연구원 연구원, 경제학 석사

1. 서론

본 연구는 배추 수급 불안정 해소를 위한 연구로서 생육 실측 자료와 비선형적 기후 효과를 반영하고 방법론에 있어 비선형성을 포함하고 있는 딥러닝 알고리즘과 혼합주기 모형을 활용하여 수급 예측 모형을 고도화 하고자 한다.

인간이 살아가는데 있어 의식주는 매우 중요하다. 그 중 ‘식’은 삶을 살아가는 에너지를 제공해 주는 등 생명 유지에 있어 필수적 요소이다. 이러한 ‘식’은 자연환경, 문화, 인종, 성별, 가치관 등에 의해 차이가 있으며 각기 다른 식습관을 가질 수 있다. 그리고 이러한 식습관은 시간의 흐름에 의해 구성원들에게 내재되어 식문화로서 자리 잡게 된다. 이렇듯 식문화는 해당 국가 구성원에게 내재된 기초 먹거리를 포함한다. 이에 가장 대표적인 것으로 우리나라 전통 식문화에서 빼 놓을 수 없는 것이 바로 ‘김치’이다. 예로부터 김치는 ‘김장’이라는 생명 유지를 위해 시작되었으며 전통적 먹거리 문화로 발전하였다. 그리고 이러한 문화 행위에 의해 각 가정에서 제조되어 왔으며 상업적으로도 작게는 음식점 단위, 크게는 공장에서의 생산에 이르며 나아가 한 국가의 산업으로 구성되기에 이르렀다.

이처럼 김치는 우리의 식문화에서 없어서는 안 되는 중요한 요소이며 이를 제조하기 위해 쓰이는 핵심 원재료인 배추는 식문화뿐만 아니라 영양학적 관점에서도 매우 중요한 품목이다. 이에 배추는 현재 우리나라에서 생산량이 많은 대표적인 엽채류이며 국민의 기초 먹거리 안전성 유지에 중요한 역할을 하고 있다.

하지만 최근 문제가 심화되고 있는 급격한 기후 변화는 우리가 생활을 위해 필요로 하는 원물의 생산 및 공급 측면에서 불안정성을 높이는 요인이 되고 있다. 이러한 생산의 불안정성 증대는 해당 농산물의 출하량, 저장량, 수입물량 등의 변화를 일으키고 가격의 변동을 불려옴으로서 수요의 변화까지 이어질 수 있다. 그러나 앞서 언급한 바와 같이 식습관은 식문화로 발전하며 국가 및 국민 생활 속에 내재되게 된다. 다시 말해 기후 변화로 인해 생산에 문제가 생겼다고 해서 간편하게 다른 품목으로 대체되기 어려운 특성이 있다는 것이다. 그리고 국내산 농산물을 수입산으로 완전 대체하기에도 식품은 안전성, 소비자 기호, 해당 생산업 종사자 소득 등을 고려할 때 어려움이 존재한다.

결국 이러한 여러 요인들로 인해 수많은 농산품 중 기후 변화로 커지고 있는 생산 변동성에도 불구하고 우리가 배추 수급 안정을 위한 노력을 포기할 수 없는 이유이다.

이에 해당 주제와 관련하여 많은 연구자들이 연구를 행하였다. 선행 연구들을 살펴보면 허신행 외(1984)는 기상 요인을 포함한 배추 공급 함수를 추정하였다. 이용선 외(2004, 2005), 한석호 외(2004, 2011), 박지연 외(2013), 최병옥 외(2020) 등에서 기상 요인을 포함한 배추 공급 부문 예측 연구를 진행하였다. 그리고 품목의 범위를 타 농산물로 확대하면 주식인 쌀과 관련한 연구가 활발히 진행되었다. 이를 살펴보면 권오상 외(2008), 노재선 외(2012), 심교문 외(2016), 한석호(2016) 등의 연구가 있었다.

앞서 설명한 기존 연구들의 대부분은 선형 모형을 중심으로 분석이 진행되었으며 기온, 강수량 등을 기상 변수로 활용하였다. 하지만 최근의 기후 변화는 과거에 비해 급격한 변동을 보여주고 있고 기후 변화의 누적된 영향에 의해 주산지의 이동도 발생하고 있다. 이러한 변동성 확대는 결국 공급 불안정을 증폭시키고 나아가 영양학적 측면에서도 문제를 발생시킬 수 있다. 이에 과거보다 고도화된 수급 예측이 필요하다.

고도화된 수급 예측을 위해서는 앞서 언급한 일반적인 기후와 이상 기후 요인과 함께 생육 실측 자료가 추가로 요구된다. 이에 농촌경제연구원과 농촌진흥청은 생육 실측 자료에 대한 수집 및 작성에 많은 노력을 기울이고 있다.

생육 실측 자료를 포함한 수급 예측 분석은 이를 포함하지 않았던 분석에 비해 상대적으로 농작물 생산에 보다 정밀한 정보 반영으로 고도화에 기여할 수 있다. 생육 실측 자료는 조사에 따라 작물의 생애 주기를 포함할 수 있다. 이에 현재 구축되고 있는 생육 실측 조사 또한 작물의 생애 주기를 일부 반영한 형태로 진행되고 있다. 이를 살펴보면 크게는 육묘 이후 생장기와 수확 직전에 시기를 포함하고 있다. 이를 활용한다면 기존의 수급 예측이 농작물의 최종 생산량을 기반으로 했던 것에 비해 많은 장점을 가질 수 있다. 생육 실측 자료의 활용은 시간에 따른 작물의 성장 단계를 포함하여 분석할 수 있다. 결국 생육 단계별로 미치는 기후의 변화, 병충해 등의 영향도 분석할 수 있다는 장점을 가진다. 이러한 생산 예측의 고도화는 결국 수요의 예측에도 긍정적인 영향을 미칠 수 있다.

이에 본 연구는 자료로서 생육 실측 자료와 기후 변수를 포함하고 모형에서는 비선형적 특성을 반영하여 수급모형을 고도화 하고자 한다. 또한 이를 통해 해당 품목 및 기타 농산물의 안정적 생산과 종사자의 소득 안정, 나아가 소비자 효용 증대에 기여 하고자 한다.

본고의 구성은 다음과 같다. 제2절은 배추 생육 특성을 설명한다. 제3절은 분석에 사용된 모형에 대해 설명하며, 이어서 제4절은 자료를 설명한다. 제5절은 분석 결과를 정리하고 의미를 해석한다. 마지막 제6절은 연구결과를 요약하고 결론 및 제언을 한다.

2. 배추 생육 특성

배추는 서늘한 기후를 선호하는 호냉성 작물 중 하나이다. 농촌진흥청에 따르면 최적 성장 적온은 18~20도이다. 그리고 동해는 일반적으로 -8도로 알려져 있으나 기온 변동 폭에 따라 -3~-4도에서도 발생하기도 한다. 재배의 경우 재배 방식에 따라 크게 노지 재배와 시설 재배, 재배 시기에 따라 봄 배추, 고랭지(여름) 배추, 가을 배추, 시설(겨울) 배추로 나눌 수 있다.

표 1. 배추 재배 방식 및 시기별 생산면적 및 생산량

재배유형별		계		노지		시설	
항목		면적(천 ha)	생산량(톤)	면적(천 ha)	생산량(톤)	면적(천 ha)	생산량(톤)
2017	소계	32.42	2,395.69	30.02	2,286.00	2.40	109.68
	봄	9.00	428.86	6.60	319.18	2.40	109.68
	고랭지	-	-	5.39	255.26	-	-
	가을	-	-	13.67	1,362.62	-	-
	겨울	-	-	4.36	348.94	-	-
2018	소계	31.14	2,391.95	29.44	2,320.21	1.71	71.74
	봄	7.37	350.91	5.66	279.17	1.71	71.74
	고랭지	-	-	5.73	278.38	-	-
	가을	-	-	13.31	1,404.15	-	-
	겨울	-	-	4.74	358.50	-	-
2019	소계	25.84	1,859.71	24.20	1,788.23	1.63	71.48
	봄	6.53	338.57	4.90	267.09	1.63	71.48
	고랭지	-	-	4.98	233.96	-	-
	가을	-	-	10.97	1,059.93	-	-
	겨울	-	-	3.36	227.25	-	-

자료: 통계청

앞선 <표 1>은 배추의 재배 방식별 재배 시기별 생산면적과 생산량을 나타낸 표이다. 이를 살펴보면 최근 스마트팜 등 농작물의 전반적인 생산에 있어 시설 재배의 확대되고 있는 추세 이기는 하나 배추의 경우에는 현재까지는 노지 재배가 절대적 생산 규모를 보이고 있다. 이에 해당 농작물은 국내 생산에 있어 기후 변화에 의한 영향이 타 작물에 비해 상대적으로 크게 나타날 수 있다는 것을 의미한다.

또한 배추는 기후 변화 요인 중 기온에 대한 영향이 큰 부분을 차지한다. 배추의 생애를 간략히 정리하면 육묘 - 정식 - 결구 - 수확으로 크게 구분할 수 있다. 육묘단계에서는 정식 이후와는 다르게 대부분 온도를 인위적으로 조절할 수 있는 육묘장에서 보낸다. 이에 본격적으로 기후 변화에 노출되는 시기는 정식 후이다. 이를 크게 나누면 정식 후부터 결구까지의 시기를 생육기라고 하며 결구 개시 후부터 수확 직전까지를 결구기라고 할 수 있다(김진희 외, 2015). 배추는 생애의 각 시기별로 적정 기온과 한계 기온에 차이를 보인다. 그리고 기온은 배추 호흡량에 영향을 미치고 이는 실질 광합성량에 영향을 주게 되며 기온의 변화는 배추 생육 주기에 따라 일부 상이한 영향을 발생시킬 수 있다.

표 2. 배추 생애 주기별 적정 기온과 한계 기온

재배유형별	최소 한계 기온 (Min)	임계 한계 기온 (Low)	적정 기온	임계 한계 기온 (High)	최대 한계 기온 (Max)
생육기	0	2~5	20~22	25	30
결구기	0	5	15~18	23	30

주: 1) 생육기: 정식 후부터 결구까지의 시기

2) 결구기: 결구 개시 후부터 수확 직전까지의 시기

자료: 농사로: 농촌진흥청, 김진희 외(2015)

또한 재배 시기별 생육 불안정을 가져오는 재해와 병충해 종류에도 차이가 있다. 재해는 일반적으로 태풍, 가뭄 등을 의미하여 병충해는 해당 농작물에 피해를 주는 생물학적 바이러스 및 벌레 등에 의해 나타나는 피해를 의미한다.

이에 <표 3>을 살펴보면 다음과 같다. 봄 배추의 경우 가뭄에 큰 영향을 받을 수 있으며, 뿌리혹병과 잎마름병의 발생 가능성이 높다. 여름(고랭지) 배추는 해발 고도가 높은 지역이 재배지역으로 선호되며 폭염과 호우, 태풍과 같은 재해 피해를 유의하여야 한다. 가을 배추는 타 시기에 비해 상대적으로 자연재해 및 병충해 피해가 적은 편이기는 하나 이상 기후 영향으

로 태풍에 대한 피해를 주의해야 한다. 또한 병충해로는 뿌리, 잎 모두에 있어 마름병의 발생 가능성이 높다. 겨울 배추는 해당 작물 특성에 따라 해당시기에 상대적으로 기온이 온화한 남부해안 지역을 주 재배 지역으로 하며 계절적 특성상 한파에 피해 가능성이 가장 크다.

표 3. 배추 재배 시기별 재해 및 병충해 종류

항목	봄	여름	가을	겨울
파종 및 생육기	1~7월	6~10월	8~11월	9월~익년 4월
주재배지역	전국	해발 600~800m	전국	남부해안
자연재해	가뭄, 호우	폭염, 호우, 태풍	태풍, 우박	한파
병충해	뿌리혹병, 잎마름병	무름병, 뿌리혹병	(뿌리, 잎)마름병	무름병

자료: 농사로: 농촌진흥청, 최병옥 외(2020)

이처럼 배추는 재배 시기에 따라 영향을 미치는 자연재해 및 병충해 종류가 상이할 수 있다. 이에 본고 분석의 경우, 가을 배추와 여름(고랭지) 배추를 중심으로 분석함으로서 일반적인 기후 요인과 함께 이상 기후로서 ‘태풍’이 공통적으로 현상으로 파악되어 해당 영향을 확인하기 위해 태풍의 기준을 적용한 풍속을 포함하여 분석한다.

3. 분석 모형

본고의 분석은 가을 배추에 대한 분석과 여름(고랭지) 배추 분석으로 구분되며 각 분석에 따라 적용 모형이 상이하다. 이는 사용 가능 자료 구조의 상이함에 따라 구분 적용하였다.

가을 배추의 생산부문 분석에서는 크게 패널자료 분석 모형과 혼합주기 모형 그리고 기계학습이라고 할 수 있는 딥러닝 알고리즘을 분석에 적용한다. 여름(고랭지) 배추의 생산부문 분석에서는 생육 기간 구분에 따른 횡단면자료 분석 모형을 사용하고 추가적으로 고차원 분석 모형을 적용한다. 수요에 있어서 배추는 저장량과 저장기간이 거의 없이 생산과 소비가 동시에 일어나는 특성을 지닌다. 이에 생산량과 소비량의 차이가 크지 않다. 이에 역수요함수인 가격 함수를 포함하고 고도화된 생산부문과 수요부문의 균형을 부분 균형 모형을 활용하여 예측 모형 분석을 한다.

3.1. 가을 배추 생산부문 분석

3.1.1. 패널자료 분석 모형

본 연구에서는 가을 배추 분석에 있어는 기본적으로 패널 자료를 사용하게 된다. 이에 패널 자료 분석 모형을 사용하였다. 패널자료 분석 모형은 아래 식 (1)과 같다.

$$(1) Y_{i,t} = \alpha_0 + \beta_1 X_{i,t} + u_{i,t}$$

$$(2) i = 1, 2, \dots, n, \quad t = 1, 2, \dots, T$$

$$u_{i,t} = \eta_i + \epsilon_{i,t}$$

생산 모형에 있어 Y 는 배추의 단수를 나타내고 X 는 생육 실측 변수, 기상 변수, 이상 기후 변수를 포함한다. 식 (2)에 나타나 있는 오차 $u_{i,t}$ 는 패널 개체에 따라 변하는 고유의 특성을 나타내는 η_i 와 패널 개체의 시간에 따라 변화하는 순수한 확률과정을 나타내는 $\epsilon_{i,t}$ 로 구성된다. 또한 분석 시 패널자료의 이용은 각 패널자료 간 이질성의 제어 여부에 따라 추정량의 일치성에 영향을 줄 수 있다. 이에 본 연구는 패널자료 분석에서 고정효과를 반영함으로써 지역 간 이질성을 모형에 적용하였다.

3.1.2. 혼합주기 모형(MIXed DATA Sampling, MIDAS)

혼합주기 모형(MIDAS)은 종속변수의 관측 빈도가 설명변수에 비해 낮아 고빈도 설명변수에 포함된 정보를 충분히 활용하지 못한다는 한계를 극복하기 위해 고안된 모형이다. Ghysels et al.(2006)의 연구에서 분기별 경제성장률의 단기예측에 월별 자료를 사용하여 예측 정확도를 개선할 수 있음을 보였으며, 이후 거시경제 및 금융 분야의 예측 정확도 향상 및 조기 예측에 널리 사용되고 있다. 본 연구에서는 종속변수인 배추의 단수 통계가 연 단위로 공표되는 데 반해 기상 자료는 일별로 관측 가능한 변수라는 점에서 혼합주기 모형의 도입을 통한 예측 정확도 향상 가능성을 검토하였다. 혼합주기 모형은 식 (3)과 같다.

$$(3) Y_t = \beta_0 + \beta_1 B(X_t^{(m)}) + Z_t' \gamma + \epsilon_t$$

Y_t 는 저빈도 종속변수, $X_t^{(m)}$ 는 고빈도 설명변수, Z_t 는 종속변수와 빈도가 일치하는 저빈도 설명변수이다. 고빈도 설명변수의 상첨자 m 은 종속변수와 빈도 비율을 의미한다. 예를 들어 Y_t 가 배추의 연도별 단수, $X_t^{(m)}$ 이 일별 기온이라면 연간 일별 기온을 모두 사용한다면 m 은 365가 된다. 본 연구의 가을 배추 분석에서는 생육 기간을 8월에서 11월로 설정하였으므로 m 은 8월 1일부터 11월 30일까지의 일수인 122가 된다. 함수 B 는 고빈도 변수에 적용되는 가중치 함수를 의미하며 함수 설정에 따라 추정하는 모수의 개수 및 추정 결과가 달라질 수 있다. 대표적인 가중치 함수로는 Almon 함수가 있으며 식 (4)와 같다.

$$\begin{aligned} (4) \text{ Almon}(k) : B(L^{\frac{1}{m}}; \theta) X_t^{(m)} &= \sum_{i=0}^{m-1} w_i X_t^{\left(\frac{i}{m}\right)} \\ &= \sum_{i=0}^{m-1} \left(\theta_0 + i \theta_1 + i^2 \theta_2 + \dots + i^k \theta_k \right) X_t^{\left(\frac{i}{m}\right)} \\ &= \theta_0 \sum_{i=0}^{m-1} X_t^{\left(\frac{i}{m}\right)} + \theta_1 \sum_{i=0}^{m-1} i X_t^{\left(\frac{i}{m}\right)} + \dots + \theta_k \sum_{i=0}^{m-1} i^k X_t^{\left(\frac{i}{m}\right)} \end{aligned}$$

본래는 종속변수의 한 시점에 대응하는 고빈도 설명변수의 각 시차에 대하여 가중치 모수를 추정해야 하므로 고빈도 설명변수 하나에 m 개의 모수를 추정해야 한다. 그러나 혼합주기 모형에서는 일반적으로 가중치 함수의 형태를 가정함으로써 고빈도 설명변수의 시차 간 가중치를 적용하면서 추정해야 하는 모수의 개수를 축소할 수 있다. 예를 들어 Almon 함수의 차수 k 를 3으로 설정하면 고빈도 시차에 따른 가중치는 이차함수 형태를 가지게 되고 이를 위해 3개의 모수를 추정하게 된다. 이에 본고의 분석에서는 가중치 함수로 차수가 3인 Almon 함수를 사용하였다.

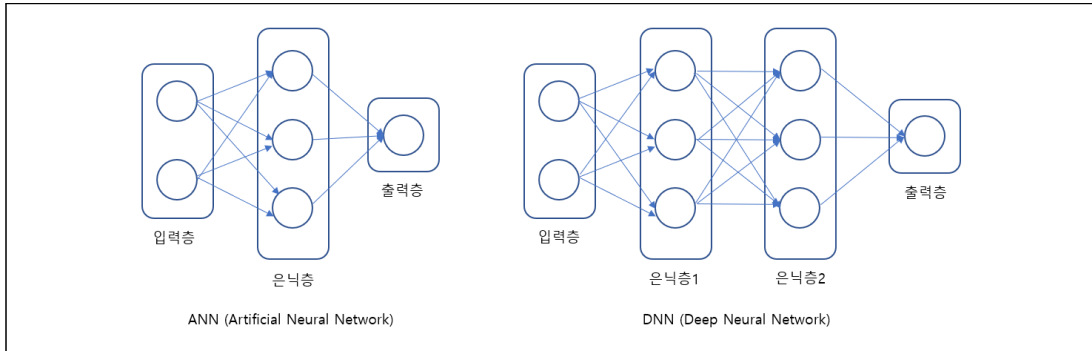
3.1.3 딥러닝 학습 모형

딥러닝 학습 모형은 비선형성을 포함한 분석 형태로 일반적으로 인공지능 알고리즘을 활용하여 예측력이 뛰어난 비선형 함수를 추정하는 방식이다. 딥러닝은 기계 학습 모형의 하위 개념이다. 기본적으로 두 방식 모두 인공지능 알고리즘을 사용하게 되는데 일반적인 기계 학습 모형은 인공지능 알고리즘에서 부정확한 예측을 반환하면 엔지니어가 개입하여 조정해야 하는 방식을 따르지만 딥러닝의 경우 알고리즘이 자체 신경망을 통해 예측의 정확성 여부를 스스로 판단하는 형태이다. 이에 딥러닝 알고리즘은 인공 신경망이라는 계층화된 알고리즘 구조를 사용하고 이러한 인공 신경망의 설계를 통한 분석은 일반적인 기계 학습 모형보다 상대적으로 뛰어난 학습 프로세스를 제공한다. 이러한 딥러닝 알고리즘은 현재 계속 발전 중이며 발전의 단계는 다음과 같다.

기본적인 형태는 ANN(Artificial Neural Network)이며 사람의 신경망 원리와 구조를 모방하여 만든 기계학습 알고리즘이다. 인간의 뇌에서 뉴런들이 어떤 신호나 자극 등을 받고 그 자극이 임계치를 넘어서게 되면 결과 신호를 전달하는 과정을 구현한 것이다. 통상 들어오는 자극을 Input data, 임계치는 Weight, 이에 대한 결과 행동을 Output data라고 할 수 있다. 이러한 ANN의 구동 과정은 입력층(input)에 다수의 데이터를 투입하면 출력층(output) 사이에 있는 은닉층(Layer)에서 노드라 불리는 인공뉴런이 학습을 통해 시냅스의 결합 세기 즉, 임계치를 변화시켜 문제 해결을 하는 과정을 거치게 된다. 하지만 앞서 설명한 ANN의 경우, 학습과정에서 일부 인식 문제 등의 발생으로 파라미터의 최적값 산출에 오류를 보일 수 있고 이에 따른 Over-fitting 문제를 발생시킬 수 있다. 또한 다량의 입력값을 필요로 하며 학습 시간이 느린 문제가 단점으로 지적되고 있다.

다음은 본고에서 사용한 DNN(Deep Neural Network)이다. 이는 앞선 ANN에 비해 모델 내 은닉층을 많이 늘려서 학습의 결과를 향상시키는 방법이다. 일반적으로 은닉층의 개수를 2개 이상 지닌 학습 방법을 뜻한다.

그림 1. ANN(Artificial Neural Network)과 DNN(Deep Neural Network)의 비교



최근 DNN을 응용한 방법으로 RNN(Recurrent Neural Network)와 LSTM(Long Short-Term Memory)이 자료 형태적 특성에 따라 상이하게 사용되고 있다. RNN의 경우, 순차적인 정보를 활용하여 분석하는데 적합한 알고리즘으로 시계열 분석에 주로 활용된다. LSTM도 DNN의 형태에서 응용된 것으로 기존의 알고리즘이 장기간에 걸쳐 반복되는 패턴을 인식하지 못하는 문제를 일부 개선한 형태의 알고리즘이다.

이에 현재 알려져 있는 딥러닝 알고리즘 중에는 LSTM이 일반적으로 예측력에 있어 우월함을 보이고 있지만 본고의 분석에서는 자료의 제약으로 DNN 알고리즘을 사용하였다. 이는 짧은 자료 주기, 연도별 해당 지역 관측 표본 수의 상이함, 결측치 문제 등이 원인으로 판단된다. 이에 고도화된 딥러닝 알고리즘의 적용을 위해서는 대규모 자료의 수집 및 축적과 함께 표본 증가의 체계적 선정을 통해 농가 수의 편차뿐만 아니라 수집 자료의 질을 향상시켜야 할 것이다. 이에 본고에서는 딥러닝 알고리즘 중 DNN 알고리즘을 중심으로 분석된 결과를 제시한다.

3.2. 여름(고랭지) 배추 생산부문 분석

3.2.1. 횡단면 분석 모형

여름(고랭지) 배추의 경우 모형에 활용되는 생육 실측 자료가 단일 연도 자료이므로 횡단면 모형을 사용한다. 분석 모형은 식 (5)와 같다.

$$(5) Y_i = \alpha_0 + X_{1i}'\beta_1 + X_{2i}'\beta_2 + \sum_{j=1}^J \gamma_j D_{ij} + u_i$$

$$i = 1, 2, \dots, n$$

식 (5)를 사용한 분석의 목적은 배추의 생산성을 예측하는 것으로 가을 배추 분석과 동일하지만, 자료의 특성 상 종속변수 Y 로는 지역별 단수를 대신하여 개별 표본의 무계인 구중을 사용한다. 설명변수로는 생육 실측 변수 X_1 과 기상 변수 X_2 를 활용한다. 이 때 사용하는 생육 실측 변수의 실측 단계에 따라 분석 모형을 구분하여 사용한다.

첫 번째 모형(모형 6)은 3차 실측 변수를 사용하며 이는 수확 약 5일 전에 조사된 자료를 사용한다. 이와 함께 해당 기간의 기후 변수를 활용하여 각 변수가 구중에 미치는 효과를 추정한다. 두 번째 모형(모형 7)은 1차 및 2차 실측 변수를 사용하며 이는 정식 후 20일~정식 후 40일에 조사된 자료를 사용한다. 또한 해당 기간의 기후 변수를 포함하고 모형 6에 비해 상대적으로 조기에 이루어지는 예측의 정확도를 분석한다. 세 번째 모형(모형 8)에서는 활용 가능한 모든 실측 변수 및 기후 변수를 포함하여 분석한다.

3.2.2. LASSO

앞선 세 번째 모형(모형 8)의 경우, 고전적인 선형모형 분석 시 과적합(Over-fitting) 문제를 일으킬 가능성이 있다. 이러한 단점을 보완할 수 있는 추가적인 방법으로 고차원 선형모형에서의 모형선택방법인 LASSO(Least Absolute Shrinkage and Selection Operator)를 사용한다.

$$(6) \hat{y} = \hat{\beta}_0 + x_1\hat{\beta}_1 + \dots + x_p\hat{\beta}_p$$

상기 식 (6)은 반응변수를 $y = (y_1, \dots, y_n)^t$, 설명변수를 $x_1, \dots, x_p$ 으로 주어진 경우이다. 이러한 선형회귀모형에서 설명변수 개수인 p 가 반응변수 관측치수 n 보다 훨씬 클 때, 이것을 고차원(High Dimensional) 문제라고 한다. 이때, 일반적으로 행하는 OLS 방법을 사용하면 자

유도가 부족하거나 추정이 제대로 이루어지지 못한다. 또한 다중공선성이 있는 경우 불안하게 나타난다. 고차원 문제는 고전적인 모형 선택방법들이 고차원 선형 모형에서 필요로 하는 변수들보다 많은 변수를 선택하는 과적합(Over-fit)문제를 일으키는 경향이 있다. 결국 이러한 과적합(Over-fit) 문제는 고차원 모형에서 심각한 편향을 발생시킬 수 있으며, 이 같은 단점을 보완할 수 있는 방법을 사용하여야 한다.

이에 본 분석에서는 Tibshirani(1996)에 의해 제안된 LASSO(Least Absolute Shrinkage and Selection Operator)를 사용한다. $x_i^t = (x_{i1}, \dots, x_{ip})$ 가 설명변수이고 y_i 가 반응변수인 경우, 자료를 (x_i, y_i) 로 표기한다. 일반적으로 회귀설정에서는 관측치들이 독립이거나 x_i 가 주어진 상태에서 조건적으로 독립을 가정한다. 그리고 x_i 가 표준화되어 있어 $\sum_i x_i/N=0, \sum_i x_i^2/N=1$ 이 된다. $\hat{\beta}$ 를 벡터형태로 나타내면 $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)^t$ 과 같다. 이 때 LASSO 추정치 $\hat{\beta}$ 는 아래 식 (7)과 같이 정의된다.

$$(7) \quad (\hat{\beta}) = \underset{\beta}{\text{Min}} \sum_{i=1}^n (y_i - \sum_j \beta_j x_{ij})^2$$

$$(8) \quad \text{s.t.} \quad \sum_j |\beta_j| \leq t$$

식 (8)의 $t(\geq 0)$ 는 조절모수이다. 이에 모수 t 를 LASSO 추정치에 적용시켜 응축(shrinkage)의 양을 조절한다. $\hat{\beta}_j^0$ 를 완전 최소제곱 추정치로 두고 $t_0 = \sum |\hat{\beta}_j^0|$ 로 둔다면 $t(\leq t_0)$ 의 값은 0으로 가며 응축(shrinkage)이 일어난다. 그리고 몇 개의 계수들이 정확하게 0의 값을 갖는다. 예를 들어 $t = t_0/2$ 라면, 효과는 사이즈 $p/2$ 의 최고의 부분집합(best subset)을 찾는 것과 유사하게 된다.

이에 본 분석에서는 상기 방법을 추가적으로 사용하여 모형 8에서 발생 가능한 문제를 보완하고자 한다.

3.3. 수급 및 가격 모형

3.3.1. 수급 균형 모형

수급 균형 모형의 분석을 위해 앞서 제시한 생산부문 모형 이외에 수요를 포함하는 모형이 필요하다. 이를 위해 먼저 배추의 특성을 확인하고 적합한 수요 모형을 적용하고자 하였다. 배추 수요의 특성은 다음과 같다. 수요에 있어서 배추는 저장량과 저장기간이 거의 없는 것이 특징이다. 이는 생산과 소비가 동시에 일어남을 의미하며 결국 생산량과 소비량의 차이가 크지 않은 특성을 지니고 있다. 이에 역수요함수인 가격 함수를 포함하여 분석하여 고도화된 생산부문과 수요부문의 균형을 고려한 예측 모형 분석을 한다. 모형 분석은 앞서 패널 단수 모형 고도화를 실시한 가을 배추를 대상으로 한다.

$$(9) \ln(ACR_t) = \beta_0 + \beta_1 \ln(ACR_{t-1}) + \beta_2 \ln(P_{t-1}) + \beta_3 Trend_t + \epsilon_t$$

$$(10) \ln(P_t) = \gamma_0 + \gamma_1 \ln(P_{s,t}) + \gamma_2 \ln(ACR_t Y_t) + \gamma_3 \ln(DINC_t) + \gamma_4 Trend_t + \nu_t$$

식 (9)는 재배면적함수, 식 (10)은 역수요함수를 나타내며 균형 가격 추정을 목표로 할 때 역수요함수에 포함되는 단수는 앞서 추정한 단수 모형을 통해 도출한 추정치를 사용한다. 즉 단수 함수와 식 (9)의 재배면적 함수가 곱해지는 형태로 공급함수가 결정되고, 이를 역수요함수 식 (10)과 동시에 추정하여 균형가격을 도출하는 것이다. ACR은 가을 배추 재배면적, P는 가을 배추 도매가격, P_s 는 직전 작기인 여름(고랭지) 배추 도매가격, Trend는 추세, DINC는 가처분소득, 그리고 Y는 단수를 의미한다. 이 때 외생변수인 단수는 앞서 고도화된 모형의 추정 결과를 사용하여 기본적인 단수 모형으로 분석한 결과를 사용한 경우와 비교한다.

3.3.2. 부분 조정 모형

실제 가격이 항상 균형가격에 머물러 있다고 보기는 어려우나 대부분은 충격으로 인한 가격 변동 시 시차를 두고 수급 균형가격으로 수렴하고자 하는 모습을 보인다(김기환·서병선, 2017). 이를 반영한 Nerlove(1956)의 부분 조정 모형은 식 (11)과 같다.

$$(11) \Delta P_t = \rho(P_t^* - P_{t-1}) + u_t, (0 < \rho \leq 1)$$

앞선 수급 균형 모형을 통해 추정한 가격 P_t^* 을 균형 가격으로 하여 실제 가격 변동 시 균형 가격으로 다시금 수렴하고자 하는 모습을 보인다. 해당 모형에 대한 예측 정확도를 비교하면 단수 모형의 차이가 가격 예측력 향상에 미치는 영향을 파악할 수 있다.

4. 분석 자료

본 연구에서 사용되는 자료는 크게 가을 배추와 여름(고랭지) 배추로 구분된다. 각 구분에 따라 자료의 형태가 상이하다. 가을 배추는 2001년부터 2020년까지의 기간과 8개 도(강원, 경기, 충북, 충남, 경북, 경남, 전북, 전남)를 중심으로 패널자료를 구축하여 분석에 사용하였다. 반면 여름(고랭지) 배추에 대한 분석은 2020년 단일년도의 기간과 강원지역의 8개 시군(강릉, 태백, 평창, 홍천, 횡성, 정선, 삼척, 영월)을 포함하여 횡단면 자료를 구축하여 분석에 사용하였다. 생육 실측 자료는 가을 배추는 농촌진흥청, 여름(고랭지) 배추는 농촌경제연구원 농업관측본부에서 실측한 자료를 사용하였다.

4.1. 가을 배추

가을 배추 분석의 주요 변수는 다음과 같다.

단수는 단위 면적(10a) 당 생산량을 나타낸 것으로, 통계청 농산물생산조사 자료를 이용하

였다. 생육 실측 변수는 표본 농가에 대한 농촌진흥청의 시도별 배추 생육 정보를 사용하였다. 평당 주수는 단위 면적(평) 당 정식한 배추 주수를 의미하며, 초장은 잎의 길이, 엽수는 잎의 개수를 나타낸다. 생육 실측 변수의 경우 자료 부족으로 2006년부터 2010년까지의 값이 결측되어 있다. 아래 <표 4>는 가을 배추 분석의 주요 변수에 대한 기초 통계량을 나타낸다.

표 4. 가을 배추 단수, 생육 기간 기후 및 생육 실측 자료 기초 통계

	단수 (kg/10a)	기온 (℃)	일조시간 (시간)	가뭄지수 (SPI3)	강풍일수 (일)	평당 주수 (주/평)	초장 (cm)	엽수 (매)
평균	9900.32	16.86	702.89	0.11	0.88	13.09	28.93	20.03
중앙값	9903.50	16.89	696.90	0.11	0.00	12.75	29.10	18.90
최대값	12986.00	18.97	1089.06	1.62	11.00	29.90	43.90	47.70
최소값	6873.00	13.27	512.82	-1.69	0.00	9.00	15.90	10.20
표준편차	1274.48	1.09	74.99	0.70	1.69	2.19	4.72	6.58
왜도	-0.15	-0.22	0.86	-0.21	3.31	4.09	-0.04	1.88
첨도	2.66	2.96	6.80	2.41	16.05	31.07	3.71	7.99
표본 수	160	160	160	160	160	120	120	120

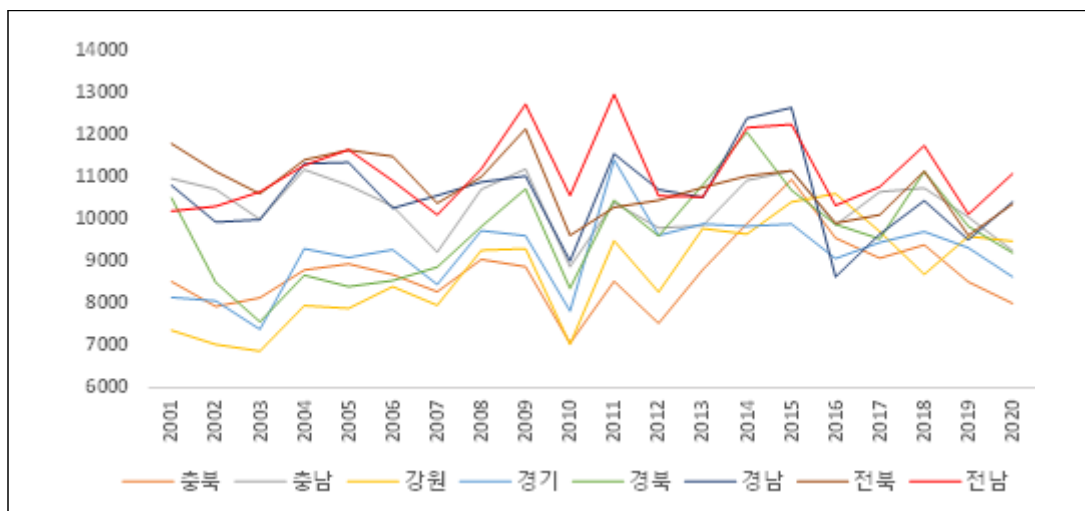
주: 1) 기후 변수는 연도 및 지역 별 생육 기간(8월~11월) 중 일 평균 값(일조시간 및 강풍일수는 누적 값)을 나타냄.

2) 생육 실측 변수는 3차 실측조사(수확 약 5일 전 측정) 결과를 나타냄.

자료: 통계청(농산물생산조사), 기상청, 농촌진흥청

기후 변수는 기상청의 일별 자료를 사용하여 생육 기간의 평균 또는 누적값을 구하여 설명 변수로 사용하였다. 기온은 관측지점의 일 평균 기온을 나타내며 일조시간은 태양이 구름이나 안개로 가려지지 않고 실제로 비친 시간을 뜻한다. 가뭄지수(SPI3)는 3개월 누적 강수량이 해당 지역의 평균 강수량보다 높거나 낮은 정도를 정규화한 지수이다. 값이 1보다 크면 습함, -1보다 작으면 가뭄을 나타낸다. 강풍일수는 일 최대 풍속이 17m/s 이상인 일수를 의미한다. 이는 기상청의 태풍 분류 기준인 ‘중심 부분 최대 풍속이 17m/s 이상인 열대 저기압’을 고려한 것이다. 다음 <그림 2>는 가을 배추의 지역별 단수 추이를 나타낸 것이다. 이를 살펴보면 배추 단수에 있어 지역별 차이가 두드러지면서도 해마다 증가 또는 감소하는 방향이 대부분 일치한다. 지역별로는 주로 남부지방의 단수가 중부지방의 단수보다 지속적으로 높으며, 시간적으로는 2009년, 2011년, 2015년의 풍작 및 2003년, 2010년, 2012년의 흉작 등이 전국적으로 발생한 것을 확인할 수 있다.

그림 2. 가을 배추 지역별 단위 추이



4.2. 여름(고랭지) 배추

여름(고랭지) 배추는 활용 가능한 생육 실측 변수가 단일년도(2020년)이다. 가을 배추에서와 같은 패널 데이터 분석을 위해서는 다년도 자료의 축적이 요구되므로 여름(고랭지) 배추에 대해서는 각 표본의 실측변수 값, 그리고 생육 기간 및 지역에 맞는 기후 조건을 변수로 하여 횡단면 분석을 실시하였다. 한편, 모든 생육 자료는 주산지인 강원도에 위치한 농가에서 실측하였으므로 지역 구분은 가을 배추에서와 같은 전국 8개 도가 아닌 강원도 내에서 생육 자료 표본이 존재하는 8개 시군으로 분류하였다. 다음 <표 5>는 분석에 사용한 변수의 기초통계량을 나타낸다.

여름(고랭지) 배추 생육 실측은 총 3차에 걸쳐 이루어졌으며 1차 조사는 정식 후 약 20일째, 2차 조사는 정식 후 약 40일째, 3차 조사는 수확 약 5일 전에 이루어지는 것으로 되어있으나 실제 관측 시점은 표본마다 조금씩 차이가 존재한다. 총 표본 수는 조사 차수마다 다르나 3차 실측을 기준으로 588개이다. 그러나 이 중 병충해의 영향으로 인한 결측치, 1차 또는 2차 실측 자료가 존재하지 않아 비교가 불가능한 표본, 그리고 일부 생육 변수에서 1차 또는 2차

실측값이 3차 실측값보다 더 높은 경우 등은 분석 대상에서 제외하였다. 결국 분석에 사용된 생육 실측 표본 수는 167개 이다.

표 5. 고랭지배추 생육 실측 변수 및 생육 기간 내 기후 변수 기초 통계

	구중 (g)	초장 1차 (cm)	초장 2차 (cm)	초장 3차 (cm)	엽수 1차 (매)	엽수 3차 (매)	엽폭 1차 (cm)	엽폭 2차 (cm)	엽폭 3차 (cm)
평균	1772.21	16.72	26.97	33.07	16.47	93.59	12.08	18.58	21.09
중앙값	1699.23	17.00	27.00	33.00	16.00	92.00	12.00	19.00	21.00
최댓값	3961.47	25.00	41.00	43.00	27.00	181.00	20.00	29.00	28.00
최솟값	413.71	2.00	14.00	21.00	7.00	39.00	2.00	9.00	12.00
표준편차	681.33	4.34	5.18	4.03	4.25	19.11	3.29	3.82	3.06
왜도	0.66	-0.29	0.05	-0.20	0.27	1.12	0.03	-0.18	-0.38
첨도	3.59	2.70	2.90	3.45	2.77	6.77	2.72	2.57	2.84
표본 수	167	167	167	167	167	167	167	167	167

	구폭 (cm)	구고 (cm)	병충해	기온 (℃)	강수량 (mm)	일조시간 (시간)	최대풍속 (m/s)
평균	18.03	26.38	0.15	20.42	770.99	339.04	8.72
중앙값	18.00	27.00	0.00	19.96	673.90	327.50	7.94
최댓값	30.00	35.00	1.00	24.09	1275.30	470.00	10.63
최솟값	9.00	13.00	0.00	17.67	270.00	221.70	7.31
표준편차	3.93	4.34	0.36	2.04	287.02	66.26	1.29
왜도	0.31	-0.44	1.96	0.53	0.39	0.08	0.39
첨도	2.71	3.18	4.86	2.05	2.11	2.32	1.34
표본 수	167	167	167	167	167	167	167

주: 1) 기후 변수는 각 표본의 생육 기간 중 일 평균 값(강수량 및 일조시간은 누적 값)을 나타냄.

2) 생육 실측 변수 중 구폭, 구고, 병충해는 3차 실측에만 측정되었으며 병충해 변수는 더미변수

자료: KREI 농업관측본부, 기상청

5. 분석 결과

5.1. 가을 배추 분석 결과

5.1.1. 생산부문 단수 모형 분석 결과

가을 배추의 단수 예측에 활용할 주요 모형의 추정 결과는 <표 6>에 나타나 있다.

표 6. 생육 실측변수 유무에 따른 단수 추정결과

변수	모형 1 : Pooled OLS (생육변수 미포함)		모형 2 : Pooled OLS (생육변수 포함)	
	계수	표준오차	계수	표준오차
상수항	-21782.70	14048.70	-28976.50*	14680.03
기온	2873.53*	1690.03	4120.27**	1797.22
기온 ²	-66.43*	50.50	-105.24*	54.07
일조시간	3.30***	1.09	1.58	1.27
가뭄지수	-194.57	206.02	-472.52*	243.96
강풍일수	-4.24	49.97	-20.60	60.80
평당 주수			-23.29	42.18
초장			-42.85	29.10
엽수			14.27	19.46
R^2	0.40		0.43	
Adj. R^2	0.38		0.39	
AIC	16.70		16.69	
RMSE	983.55		942.78	
D-W	1.64		1.38	

주: 1) *, **, ***는 추정된 계수가 각각 10%, 5%, 1% 수준에서 유의하다는 것을 의미함.

상기 <표 6>은 생육변수 포함 유무에 따라 구분하여 분석한 결과이다. 결과를 살펴보면 공통적으로 들어간 기후 변수 중 기온의 영향이 통계적으로 유의함을 보였다. 또한 일조시간과 가뭄지수도 모형에 따라 각각 통계적으로 유의함을 보였다. 다만 생육변수의 경우 모형 2에 포함하여 분석하였으나 통계적 유의성은 확인하지 못하였다. 그러나 모형의 적합성 및 예측력을 나타내는 AIC, RMSE 등은 해당 변수를 포함할 경우 개선되는 것으로 나타났다.

다음 <표 7>은 생육변수 추가 이후 고정효과 도입 및 패널 자기상관을 해결하기 위한 GMM 추정을 도입한 결과를 나타낸다.

표 7. 생육 실측변수 유무에 따른 단수 추정결과

변수	모형 3 : GMM		모형 4 : GMM (혼합주기: MIDAS)		모형 5 : GMM (딥러닝: DNN)	
	계수	표준오차	계수	표준오차	계수	표준오차
상수항	-113529***	26039.67	-111233***	24268.50	-3799.60	2480.70
기온	13690.18***	3070.82	0.90***	0.19	0.73***	0.19
기온 ²	-398.09***	90.90				
일조시간	3.91	3.45	6.99***	2.92	4.61	3.11
가뭄지수	-422.39	298.24	28.74	282.51	-278.91	288.76
강풍일수	-47.45	78.34	-74.72	77.84	-40.51	78.87
평당 주수	-35.15	36.36	-17.64	35.48	-39.19	36.75
엽수	11.61	18.02	2.84	17.43	-3.18	18.13
초장	-8.51	25.72	-9.96	24.23	-10.86	25.29
단수(-1)	0.44**	0.17	0.36**	0.15	0.43**	0.17
R^2	0.82		0.82		0.81	
Adj. R^2	0.75		0.76		0.74	
RMSE	506.09		502.21		520.87	
D-W	2.05		1.84		2.21	
J-Stat	0.76		0.75		0.13	

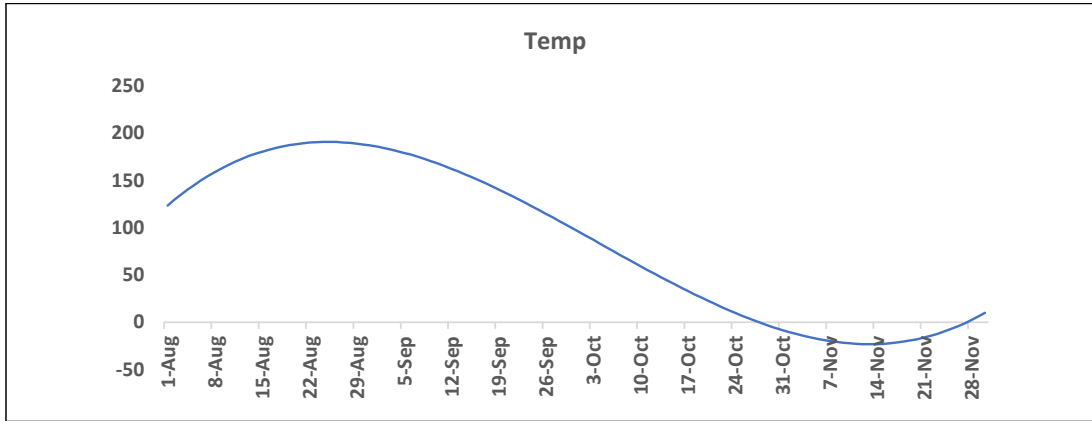
주: 1) *, **, *** 는 추정한 계수가 각각 10%, 5%, 1% 수준에서 유의하다는 것을 의미함.

2) 기온의 경우 모형 4에서는 생육 기간 동안의 일별 자료 가중합(MIDAS), 모형 5에서는 기온을 설명변수로 단수를 딥러닝 추정된 값을 의미함.

3) 가뭄지수는 절댓값을 사용

분석 결과 모든 형태의 분석에서 기온이 통계적으로 유의함을 보였다. 다만 생육변수는 통계적으로 유의하지 않았으나 해당 변수의 추가와 함께 고정효과 도입 및 패널 자기상관을 해결하기 위한 GMM 추정을 도입한 결과 모형의 적합도를 나타내는 지표가 크게 개선됨을 보였다. 또한 고빈도의 설명변수로 기온 변수를 포함하는 혼합주기 모형으로 분석한 결과는 앞선 분석과의 상대적 비교를 위해 도출된 일별 가중치를 활용한 가중합으로 기존 모형의 기온 변수를 대체하는 2단계 추정법을 사용하여 모형 4를 분석하였다. 분석 결과 앞선 모형에 비해 모형의 적합도가 개선됨을 보였다. 다음 <그림 3>은 일일 기온을 고빈도 설명변수로 사용하여 단수를 추정된 결과 도출된 MIDAS 가중치를 나타낸다.

그림 3. 일일 기온에 대한 MIDAS 가중치



상기 <그림 3>에 나타나 있듯이 혼합주기 모형을 사용하는 즉, 고빈도 기후 변수의 활용은 생육 단계에 따라 기후 변수가 단수에 미치는 영향이 상이하게 반영됨을 나타내어 준다. 결국 혼합주기 모형의 사용은 기존 저빈도 변수만을 사용한 모형에 비해 상대적으로 고빈도 적용 변수의 영향을 더욱 세밀하게 분석 가능함을 나타내 준다고 하겠다. 또한 생애 주기 측면에서 보면 생육 후반의 결구기보다는 전반기인 생육기에 가중치가 높게 적용되어 결국 생육기 기온이 단수에 보다 중요한 영향을 미치는 것으로 나타났다. 그리고 예측력도 일부 개선되어 향후 생육 변수 자료의 축적 시에도 적용하여 분석할 수 있는 가능성이 있다고 하겠다.

<표 7>의 분석 결과 중 모형 5는 앞선 모형과의 비교를 위해 딥러닝 모형에 설명변수로 평균 기온만을 투입하여 In-Sample로 예측한 결과를 반영한 것으로 딥러닝 모형에서 도출된 결과 변수를 기존의 기온 변수 대신 적용하여 분석한 결과를 나타낸 것이다. 이에 결과를 비교해 보면 RMSE 값의 비교 시 혼합주기 모형보다는 예측력이 일부 낮아짐을 보였으나 비교적 큰 차이를 보이지는 않았다.

5.1.2. 가을 배추 모형 : 예측력 비교

In-Sample 예측의 경우 전체 기간(2001~2020년)을 대상으로 모형을 추정된 뒤 단수의 맞춘값(Fitted value)과 실제 단수를 비교하여 예측 정확도 지표(RMSE, MAPE)를 도출하였

다. Out-of-Sample 예측의 경우 2001년부터 2019년까지의 자료로 모형을 추정한 뒤 2020년 지역별 단수에 대한 예측을 실시하고 실제 값과 비교하여 예측 정확도를 산출하였다. 만약 In-Sample에서의 예측력이 좋음에도 불구하고 Out-of-Sample에서의 예측력이 저조하다면 해당 모형은 과대 적합(Over-fitting) 되었다고 볼 수 있다.

표 8. 추정 모형 간 예측력 비교(In-Sample)

	모형 1	모형 2	모형 3	모형 4	딥러닝 모형
RMSE	983.54 (100.00%)	942.78 (95.85%)	506.09 (51.46%)	502.21 (51.06%)	349.90 (35.58%)
MAPE	8.20 (100.00%)	7.94 (96.80%)	3.96 (48.30%)	3.92 (47.84%)	2.43 (29.61%)

주: 1) 괄호 안의 숫자는 모형 1의 값을 100%로 두었을 때 상대적인 값을 나타낸 것

표 9. 추정 모형 간 예측력 비교(Out-of-Sample)

	모형 1	모형 2	모형 3	모형 4	딥러닝 모형
RMSE	727.32 (100.00%)	727.08 (99.97%)	609.42 (83.79%)	457.49 (62.90%)	840.77 (115.60%)
MAPE	6.53 (100.00%)	6.59 (100.85%)	5.72 (87.60%)	3.64 (55.74%)	7.59 (116.21%)

주: 1) 괄호 안의 숫자는 모형 1의 값을 100%로 두었을 때 상대적인 값을 나타낸 것

비교 항목 중 딥러닝 모형은 앞선 모형 5와 차이가 있다. 앞선 모형 5에서는 기후 요인에만 딥러닝 알고리즘을 적용한 반면 본 모형에서는 기온, 일조시간, 강풍일수, 평당 주수, 초장, 엽수 등 사용할 수 있는 모든 설명변수로 포함한 형태이다. 은닉층의 개수는 2개, 각 은닉층의 노드 수는 5개로 하였으며 오차 함수는 오차제곱합(sse), 학습 시 목표 오차는 0.01, 최대 반복횟수는 100,000번으로 설정하였다. 또한 활성화 함수로 로지스틱 함수를, 학습 알고리즘으로 탄력적 역전파 방식을 사용하였다. 이렇게 분석한 딥러닝 모형의 경우 In-Sample 예측 결과에서는 가장 우수함을 보였다. 하지만 Out-of-Sample 예측에서는 예측력이 가장 낮음을 보였다. 결국 이는 딥러닝 모형의 과적합 문제가 발생한 것으로 보인다. 이에 대한 원인은 모형 자체가 내포한 한계 혹은 사용 가능한 학습 데이터의 부족 문제로 사료된다. 이에 분석 결과 Out-of-Sample 예측 결과가 가장 좋은 모형은 혼합주기 모형인 MIDAS로 모형 4에 결과를 제시하였다.

5.2. 여름(고랭지) 배추 분석

5.2.1 생산부문 단수 분석 결과

표 10. 생육 실측변수 및 기상변수를 이용한 생산부문 단수 추정결과

변수	모형 6 : 3차 실측변수 포함		모형 7 : 1,2차 실측변수 포함		모형 8 : 모든 실측변수 포함	
	계수	표준오차	t 통계량	계수	표준오차	t 통계량
상수항	-27454.3***	7276.00	-43885.3***	10329.92	-26351.0***	8190.98
기온	2149.13***	763.17	4505.75***	1096.30	1981.79**	864.26
기온 ²	-50.17**	20.12	-100.51***	26.49	-45.19**	22.71
강수량	-0.56	0.39	-0.37	0.54	-0.50	0.42
일조시간	1.42	0.96	0.13	1.68	1.62	1.04
풍속	299.28	275.65	-725.66***	204.84	269.14	284.16
구고	57.31***	10.51			57.42***	10.52
구폭	54.31***	9.90			51.39***	10.00
병충해	-196.53**	86.65			-156.21*	91.33
초장_3차	21.81**	10.75			25.29***	11.24
엽수_3차	8.62***	2.02			8.82***	2.05
엽폭_3차	16.49	12.86			6.60	13.57
초장_2차			-18.36	17.29	-16.12	11.42
엽폭_2차			74.21***	25.40	39.27**	16.66
초장_1차			54.71***	20.97	24.80*	13.60
엽수_1차			0.67	16.32	-6.45	11.72
엽폭_1차			-52.84**	26.75	-26.32	18.13
R^2	0.77		0.44		0.78	
Adj. R^2	0.74		0.39		0.75	
AIC	14.62		15.48		14.63	

주: 1) *, **, ***는 추정한 계수가 각각 10%, 5%, 1% 수준에서 유의하다는 것을 의미함.

2) 기후변수의 경우 모형 6 및 모형 8은 정식일~3차 실측일, 모형 7는 정식일~2차 실측일을 대상으로 함.

〈표 10〉은 여름(고랭지) 배추의 생산부문 단수 예측 모형의 분석 결과이다. 본 분석에서는 앞선 가을 배추의 분석과는 달리 종속변수가 단수 대신 무게를 나타내는 구중을 사용하였다.

이는 해당 분석의 자료 제약에서 오는 한계에 의한 것으로 여름(고랭지) 배추의 생산 지역이 강원도로 한정되어 생긴 결과이다. 해당 분석의 경우, 강원도에 속해 있는 각 8개 시군의 자료를 사용하게 되는데 시군 범위의 자료 한계에 의한 것이다.

이에 분석 결과를 살펴보면 모든 분석 모형에서 기온과 기온²는 통계적으로 유의함을 나타냈다. 이는 기온과 해당 농작물의 생산은 비선형성이 있음을 나타낸다.

수확기 5일 전으로 측정된 3차 실측변수를 포함한 모형 6의 분석결과 구고, 구폭, 초장, 엽폭이 모두 통계적으로 유의하며 생산과 정(+)의 영향을 보이는 것으로 나타났다. 다만 병충해의 경우, 통계적으로 유의하나 부(-)의 영향을 보이는 것으로 나타났는데 이는 병충해가 생산에 부정적 영향을 주는 것으로 나타났다.

다음 정식 후 20일부터 40일까지의 실측변수를 포함한 모형 7의 분석 결과를 살펴보면 풍속의 영향을 많이 받은 것으로 나타났다. 해당 변수는 태풍을 의미하는 최대 풍속과 같은 기준을 가지는 변수로서 해당 측정 기간에 배추는 강한 풍속에 대해 부(-)의 영향을 받는 것으로 나타났다.

전체 생육기간이라고 볼 수 있는 정식 20일 후부터 수확 5일 전까지의 생육 변수를 포함한 모형 8의 분석 결과를 살펴보면 구고, 구폭, 수확 직전 초장, 수확 직전 엽수, 정식 후 40일째 엽폭, 정식 후 20일째 초장이 생산과 통계적 유의하며 정(+)의 관계를 가지는 것으로 나타났다. 또한 병충해는 앞선 모형 6과 마찬가지로 생산에 있어 부(-)의 관계를 보였다.

5.2.2. 생산부문 단수의 LASSO 모형 분석 결과

본 분석은 앞선 분석에서 전체 생육 기간에 측정한 모든 차수의 결과를 포함한 모형의 형태인 모형 8을 중심으로 해당 분석 모형과 같은 전통적 모형에서 나타날 수 있는 과대추정(Over-fitting) 문제를 고려하여 적용하고자 한 분석이다.

다음 <표 11>은 LASSO 추정 결과를 나타낸다.

표 11. LASSO 추정 결과

변수	모형 9 : Min. MSE	모형 10 : Min MSE +1 SE
	계수	계수
상수항	-9796.14	-1864.35
기온	215.43	15.66
기온 ²	1.45	0.08

변수	모형 9 : Min. MSE	모형 10 : Min MSE +1 SE
	계수	계수
강수량	-0.49	0.00
일조시간	1.07	0.01
풍속	817.31	-60.66
구고	55.15	46.11
구폭	53.39	34.95
병충해	-155.31	-1.86
초장_3차	21.82	19.36
엽수_3차	8.63	7.59
엽폭_3차	6.58	0.00
초장_2차	-15.18	0.63
엽폭_2차	43.11	17.54
초장_1차	25.30	0.00
엽수_1차	-10.94	0.00
엽폭_1차	-27.04	0.00
지역 더미_삼척	985.81	0.00
지역 더미_정선	887.77	0.61
지역 더미_태백	1262.66	1.31
지역 더미_평창	381.13	-86.24
자유도	20	15
L1 norm	14771.69	2156.98
R^2	0.77	0.68
Lambda	0.00059	0.63

LASSO를 적용하여 과대추정 문제를 고려하여 분석한 결과, MSE +1 SE에 상응하는 페널티 모수인 lambda를 적용하여 추정 시 강수량, 엽폭_3차, 그리고 모든 1차 생육 실측변수가 설명변수로 부적절하여 일부 제외되는 것으로 나타났다. 그럼에도 불구하고 10개의 생육변수 중 6개가 적합한 것으로 나타나 해당 부분의 적용 가능성이 있다고 할 수 있다.

5.3. 수급 균형 및 부분 조정 가격 모형

본 분석은 앞서 분석한 생산부문 단수 모형을 포함하고 재배면적 함수와 가격 함수를 사용하여 수급 균형 모형을 분석하였다. 사용 자료는 가을 배추이며 앞서 설명한 해당 방정식 모두를 연립방정식 형태로 Full Information Maximum Likelihood 방법을 사용하여 분석하였다. 결과는 <표 12>와 <표 13>에 나타나 있다.

표 12. 수급 균형 모형 추정 결과 : 재배면적 함수

변수	모형 11-1. 재배면적 (5.1.1.의 모형 1 단수 추정치 활용)		모형 12-1. 재배면적 (5.1.1.의 모형 4 단수 추정치 활용)	
	계수	표준오차	계수	표준오차
상수항	0.96	0.77	1.00	0.91
재배면적(-1)	0.73***	0.07	0.73***	0.09
가격(-1)	0.36***	0.05	0.35***	0.05
추세	-0.01**	0.00	-0.01**	0.00
R^2	0.74		0.74	
Adj. R^2	0.64		0.64	
잔차 제곱합	0.05		0.05	

주: 1) 재배면적 및 가격은 자연로그 취한 값을 사용함.

표 13. 수급 균형 모형 추정 결과 : 가격(역수요) 함수

변수	모형 11-2. 가격 (5.1.1.의 모형 1 단수 추정치 활용)		모형 12-2. 가격 (5.1.1.의 모형 4 단수 추정치 활용)	
	계수	표준오차	계수	표준오차
상수항	10.97***	2.27	2.00	2.31
여름배추 가격	0.36	0.22	0.08	0.21
재배면적*단수	-2.07***	0.12	-1.88***	0.12
가처분소득	2.88***	0.18	3.94***	0.12
추세	-0.11***	0.02	-0.16***	0.01
R^2	0.68		0.77	
Adj. R^2	0.50		0.63	
잔차 제곱합	0.56		0.41	

주: 1) 재배면적, 가격, 단수 및 가처분소득은 자연로그 취한 값을 사용함.

모형 11에서는 가을 배추 단수 모형 중 생육 실측변수를 도입하지 않았던 가장 기본적인 모형(5.1.1.의 모형 1)에서 추정된 단수를 사용하였다. 그리고 모형 12에서는 예측력이 가장 좋았던 고도화된 단수 모형(5.1.1.의 모형 4)에서 추정된 단수를 포함하여 분석을 실시하였다.

분석 결과를 살펴보면 생육 실측 변수를 도입하고 모형을 고도화하여 추정된 단수를 사용했을 때가 상대적으로 기본 단수 모형을 사용한 경우와 비교 시 수급 모형의 적합성이 개선되는 것으로 나타났다. 결국 생산부문의 단수 함수 및 재배면적 함수와 수요 함수를 사용한 수급 균형 모형의 분석에서 생육변수를 포함과 함께 고도화된 단수 함수의 반영은 수급 균형 모형의 적합성을 높일 수 있다고 판단된다.

표 14. 부분 조정 모형 추정 결과

변수	모형 11-3. Δ 가격 (모형 11-1,2의 균형 가격 활용)		모형 12-3. Δ 가격 (모형 12-1,2의 균형 가격 활용)	
	계수	표준오차	계수	표준오차
상수항	0.01	0.07	0.01	0.06
균형 가격 - 가격(-1)	0.95***	0.13	0.96***	0.11
R^2	0.88		0.92	
Adj. R^2	0.87		0.91	
AIC	0.04		-0.32	
잔차 제곱합	0.35		0.25	

주: 1) 가격 변수는 모두 자연로그 취한 값을 사용함.

상기 <표 14>는 부분 조정 모형을 분석한 결과이다.

실제 가격은 항상 균형가격에 머물러 있기 어렵다. 그리고 대부분은 충격으로 인한 가격 변동 시 시차를 두고 수급 균형 가격으로 수렴하고자 하는 모습을 보인다. 이에 부분 조정 모형을 활용한 분석을 추가로 실시하였다. 분석은 수급 균형 모형 추정 결과에 의해 도출된 균형 가격과 실제 가격 자료를 활용하였다. 부분 조정 모형의 분석 결과, 두 모형 모두 조정계수가 1에 가까워 균형 가격으로의 수렴속도가 매우 빠른 것으로 나타났다. 한편 고도화된 단수 함수로부터 추정된 모형 12-3은 기본 단수 모형으로부터 출발한 모형 11-3에 비해 가격 변화의 예측 정확도가 상대적으로 높은 것으로 나타났다.

6. 요약 및 결론

본 연구는 생육 실측 자료와 비선형적 기후 요인을 반영하고 모형의 고도화를 위해 비선형성을 포함하고 있는 딥러닝 알고리즘과 혼합주기 모형(MIDAS)의 적용을 통해 수급 예측모형을 분석하였다.

이에 분석 결과를 요약하면 다음과 같다.

첫째, 생육 실측 자료의 반영으로 예측력이 개선되었다. 기존 생산부문의 단수 모형에 생육 실측 자료의 반영은 미반영 모형에 비해 상대적으로 예측력을 향상시킴을 보였다. 다만 생육 실측 변수의 통계적 유의성은 확인하지 못하였다. 이는 일부 생육 자료에서 오는 한계로 판단되며 추후 자료의 축적 및 취득 시 표준화 등에 의해 개선될 것으로 판단된다.

둘째, 생산부문 단수 모형의 고도화는 해당 모형의 예측력을 개선시켰다. 고빈도 자료를 포함하는 혼합주기 모형을 활용한 분석과 딥러닝 알고리즘을 사용한 분석을 활용하여 모형 고도화를 할 경우, 기존 생산부문 단수 모형의 예측력이 개선됨을 보였다. 다만 딥러닝 알고리즘을 포함한 모형 분석의 경우, In-sample에서의 예측력은 개선된 반면 Out-of-sample에서의 예측력은 안정적이지 않아 과적합 문제가 발생한 것으로 판단된다. 이에 대한 원인으로 딥러닝 알고리즘 모형 분석 시 통상 대규모 자료의 사용이 요구되는데 이를 충족시키지 못해 발생하는 모형 자체 한계 또는 학습 자료 부족에 의한 문제로 판단된다. 그러나 In-sample 분석에서 딥러닝 알고리즘을 활용한 분석 결과의 예측력이 개선됨을 확인하였고 이는 해당 분석 방법론 적용의 가능성을 확인한 것이라 할 수 있겠다. 결국 대규모 학습 자료의 축적 후 해당 방법론을 적용한다면 Out-of-sample의 예측 시에도 안정성이 일부 개선될 여지가 있는 것으로 판단된다.

셋째, 생육변수 반영과 함께 고도화한 단수 모형의 활용은 수급 균형 분석 시 적합도를 향상시킨다. 수급 균형 분석을 위해 생산부문에 단수 모형, 재배 면적 함수, 수요 함수를 포함하여 분석할 경우, 고도화된 단수 모형의 활용은 해당 분석의 적합도를 향상시키는 것을 확인하였다. 이는 결국 단수 모형의 고도화가 전체 수급 균형 분석 모형의 적합도 개선에 긍정적인 영향을 미쳤음을 의미한다.

이상으로 자료로는 생육 실측 자료와 비선형적 기후 요인들의 활용, 방법으로는 모형의 예측력 향상을 위해 비선형성을 포함하는 딥러닝 알고리즘 및 혼합주기 모형(MIDAS)의 적용을 통해 향후 수급 예측 모형의 예측력 향상의 방향성을 확인하였다.

다만 포함할 수 있는 자료의 크기, 정확도 등의 한계는 분석 모형의 고도화만으로 예측력 개선에 일부 한계가 있는 것으로 확인되었다. 또한 딥러닝 알고리즘을 이용한 기계 모형은 학습을 할 자료의 양이 충분치 못하면 오히려 전망에 있어 불안정함을 보일 수 있다는 것으로 확인하였다. 이에 앞으로 생육 실측 자료의 충분한 축적과 축적 시 발생할 수 있는 오류의 통제가 반드시 필요하며 그렇지 않을 경우 예측력에 있어 기존 계량 모형의 고도화가 상대적으로 우위를 보일 것으로 판단한다.

향후 본 연구의 적용 모형과 같은 고도화된 모형의 도입, 정보의 정확성이 향상된 자료의 활용을 통해 수급 예측 정확도를 향상해야 한다. 결국 수급 예측 모형의 전반적인 예측력 향상을 위한 이와 같은 노력은 해당 품목 및 기타 농산물의 안정적 생산, 종사자의 소득 안정, 나아가 소비자 효용 증대에 기여할 수 있을 것이다. 이러한 점을 감안하여 추후 연구가 지속되어야 하겠다.

참고문헌

- 권오상, 김창길. 2008. “기후변화가 쌀 단수변화에 미치는 영향: 비모수적 및 준모수적 분석.” 농업경제연구 49(4): 45-64.
- 김기환, 서병선. 2017. “사과의 산지도매가격과 소매가격 간 동태적 인과성 및 비대칭적 가격 조정 연구.” 농업경제연구 58(1): 97-116.
- 김철현. 2014. 「혼합주기 자료를 이용한 전력수요 예측 모형 구축」. 에너지경제연구원.
- 노재선, 권오상, 조승현. 2012. “기후변수가 농업부문에 미치는 경제적 영향 분석.” 농업경제연구 50(2): 1-24.
- 박지연, 박영구. 2013. 「배추·무 예측모형 고도화 방안」. 한국농촌경제연구원.
- 박지연, 김태우. 2014. 「농업부문 전망모형 KREI-KASMO 2014 운영개발 연구」. 한국농촌경제연구원.
- 심교문, 정명표, 김용석, 최인태, 김호정, 강기경. 2016. “엘니뇨 발생연도의 우리나라의 이상기상 특징과 쌀 수량과의 관계.” 한국농림기상학회지 18(3): 143-150.
- 이승윤, 이은경, 한진현. 2017. “기계학습과 혼합주기 모형을 활용한 단기전망시스템.” 한국은행 조사통계연보 5: 16-41.
- 한석호. 2016. “쌀 단수 예측모형 개발연구.” 농업생명과학연구. 50(3): 219-229.
- 한석호, 이병훈, 박미성, 승준호, 양현석, 신성철. 2011. 「기상요인을 고려한 단수예측모형 개발 연구」. 한국농촌경제연구원.
- 허신행, 황연수. 1984. “가격형성의 시차성을 이용한 김장채소생산의 예측가능성.” 농촌경제 7(4): 11-25.
- 이용선, 정학균, 김원태, 최익창. 2004. 「과채류의 작형별 단수함수 추정」. 한국농촌경제연구원.
- 이용선 외. 2005. 「기상요인이 청과물 수급에 미치는 영향」. 한국농촌경제연구원.
- 한석호, 김병률. 2004 “시장개방하의 배추 수급모형과 전망.” 농촌경제 27(3): 35-53.
- 최병욱, 최선우, 임효빈. 2020. “기상환경 변화가 배추·무 단수와 가격 변동성에 미치는 영향 분석.” 농촌경제 43(1): 21-47.
- Ghysels, E., Santa-Clara, P., Valkanov, R. 2006. “Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies.” Journal of Econometrics. 131: 59-95.
- Nerlove, M. 1956. “Estimates of the Elasticities of Supply of Selected Agricultural Commodities.” Journal of Farm Economics. 38: 496-509.
- Tibshirani, R. 1996. “Regression Shrinkage and Selection Via the Lasso.” Journal of the Royal Statistical Society. 58(1): 267-288.
- KOSIS 국가통계포털. <<http://kosis.kr>>.
- 기상청 관측자료. <<http://www.kma.go.kr>>.
- 농촌진흥청 농업기술포털 농사로. <<http://www.nongsaro.go.kr>>.
- 국가농작물병해충관리시스템. <<https://ncpms.rda.go.kr>>.



제1회

농산물 수급 예측모형 경진대회

우수논문집

Chapter 2
자유과제



장려상

영상 정보 기반 산지 출하량 정보와 도매시장 반입량 간의
관계를 이용한 고랭지 배추 출하량 산출 방법 연구

이재서 · 이광수 · 김혁진

영상 정보 기반 산지 출하량 정보와 도매시장 반입량 간의 관계를 이용한 고랭지 배추 출하량 산출 방법 연구

이재서* 이광수** 김혁진***

Key words

드론 이미지(drone image), 도매시장 반입량(wholesale market imports), 출하량(shipment), 초과녹색채지수(Excess Green Vegetation Index)

ABSTRACT

The purpose of this study is to contribute to the observation of supply and demand by studying a method of calculating cabbage shipments using drone images. The difference of the study is that it is possible to monitor and utilize the highland cabbage cultivation area or the situation of the field, which is an important factor in production through drone images.

To do this, the cultivation area is extracted from the drone image and multiplied by the number of units of production to calculate the shipment. The number of units of production compares and verifies the data provided by KOSTAT and KREI, and confirms the relationship by comparing it with the Gangwon-do wholesale market transaction volume data. Meanwhile, ExG is used to compare the changes in the RGB value of the drone image and the characteristics at the time of shipment.

As a result of the study, the number of units of production of KREI was more rational, and the shipments calculated from the information extracted from drone images showed meaningful results compared to the imports in the wholesale market. And it was confirmed that the ExG calculated using the RGB value of the drone image had a negative correlation with the harvested area, and the harvest time was determined by the amount of change in ExG and ExG.

This study has a limitation that the data used is one year. This limitation is expected to be resolved by accumulating data of improved quality in consideration of the collection method proposed by this study.

차례

- | | |
|------------------|---------|
| 1. 연구 필요성 및 목적 | 3. 연구내용 |
| 2. 선행연구 검토 및 차별점 | 4. 결론 |

*, **, *** FarmAI Lab 연구원

1. 연구 필요성 및 목적

1.1. 배추 가격의 변동 특징

2020년 배추 가격은 기록적인 최장 장마와 태풍의 영향으로 가격이 치솟았고, 2021년 초 겨울배추는 냉해 피해를 입기도 하였다. 2020년 배추 도매가격은 1월부터 9월까지 평년대비 22%에서 84%의 상승폭을 보이고, 10월은 206%까지 상승하였다(농림축산식품부, 2020). 해당 월의 배추 한 포기 소매가격은 11,284원이고, 2015년에서 2020년 사이 평균 배추 한 포기 소매가격은 3,890원이었다.

배추는 가격 변동이 매우 큰 채소 중 하나이며, 농가 조사 결과, 가격이 농가의 재배면적 증감을 결정하는 중요한 요인임이 확인되었다(김명환 외, 2015). 농산물은 기상 환경에 영향을 많이 받기 때문에 공급이 불안정적이고 이를 해결하기 위해 정부의 수급 안정 정책이 실시되고 있다. 하지만 생산자의 재배면적 결정에 영향을 미치지 못하면서 정책의 효과가 낮다는 견해가 있다(류상모, 2017). 농가가 이미 심은 이후에 제공되는 정보는 활용이 어렵기 때문에 농가의 의사결정에 필요한 정보를 적시에 제공할 수 있는 관측과 단수 예측이 중요하다(김명환 외, 2015).

1.2. 기존의 수급 예측 방법의 한계

배추 생산에 영향을 주는 기상요인과의 관계를 규명하고 단수를 예측하기 위한 다양한 연구들이 존재하였다. 류상모(2019)에서는 기상 정보를 가공하여 기상 위험정보를 연구에 도입하였고, 한석호 외(2011)는 작물이 성장하기에 적합한 최적의 환경 조건을 고려하기 위해 작물과 기상요인의 상관관계분석을 단순한 선형 방정식이 아니라 비선형 방정식 중 2차 함수 형태를 이용하여 통계적으로 단수의 최적점을 추정한 연구를 하였다. 하지만 기후온난화 등으로 인한 기상이변이 농산물 수급의 불안정성을 키우고 있는 상황에서, 수치적인 기상 자료만으로는 현장 상황을 대변하는 예측 모형을 구축하기 어려운 점이 있다.

1.3. 연구 목적

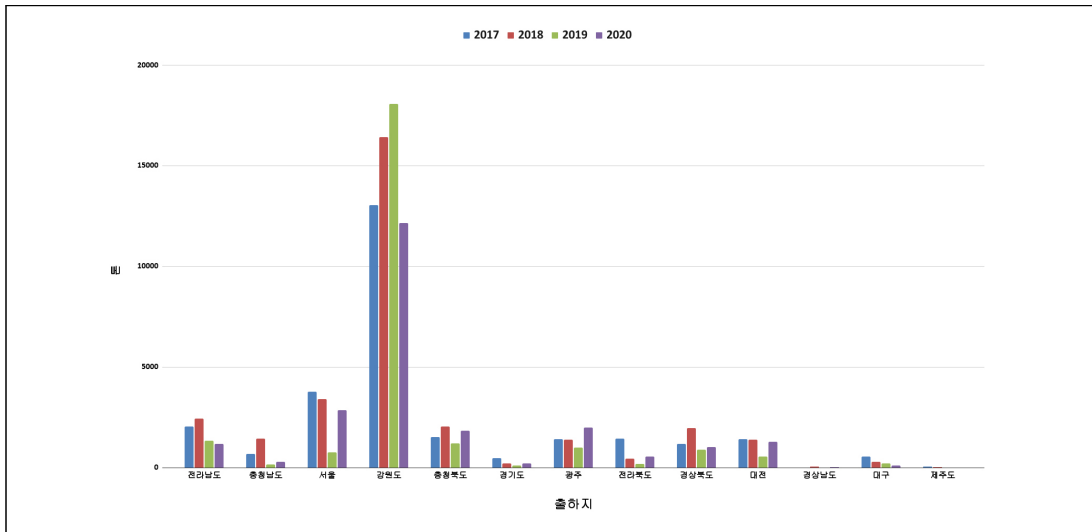
본 연구는 한국농촌경제연구원(이하 '농경연')에서 수집 중인 데이터(드론 이미지, 고랭지 배추 출하용 CCTV 집계)를 중점으로 산지 출하량과 도매시장 반입량 관계를 확인하고 드론 이미지가 관측에 적합하게 활용될 수 있는 방안을 제시하여 관측 정확도 제고 및 수급 예측에 기여하고자 한다.

2. 선행연구 검토 및 차별점

2.1. 강원도 고랭지 배추의 출하비중 검토

본 연구에서 활용하는 드론 이미지 데이터와 고랭지 배추 출하용 CCTV 집계 데이터는 강원도 지역에 해당하기 때문에 강원도 고랭지 배추 출하량이 전체 고랭지 배추의 출하량을 대표할 수 있는지 검토가 필요하였다. 이동현 외(2018)은 강원도 지역의 고랭지 배추 생산량은 전체의 90% 이상을 차지한다고 언급하였고, 농림축산식품부(2020)에 따르면 2019년도 강원도의 고랭지 배추 생산량은 전체의 93.4%이었다. 또한 김명환 외(2015)에서는 주산지의 경우 출하량 변동률 자체는 낮지만 비중이 높아 전체 출하량 변동에 미치는 기여도가 높기 때문에 주산지 재배면적 안정에 중점을 두는 것이 효과적이라고 언급함에 따라 본 연구는 강원도의 고랭지 배추 출하량을 분석하였다.

그림 1. 출하지별 전국 도매시장 반입량



주: 1) 2017~2020년 출하지별 전국 도매시장 반입량. 강원도가 가장 많은 비중을 차지하고 있음.

자료: 농경연 농업관측통계시스템 전국도매시장 경락가격자료

2.2. 고랭지 배추의 특성

배추는 서늘한 기후를 좋아하는 저온성 채소로서 생육기간은 50~90일이다. 생육 초기에는 높은 온도에 잘 견디나 생육에 가장 알맞은 온도는 20℃ 전후이고 결구의 최적온도는 15~16℃이다. 온도가 천천히 내려갈 경우에는 영하 8℃까지 견디나 갑자기 추워지면 영하 3~4℃에서 동해를 입는다. 종자는 젖은 상태에서 10℃ 이하로 일주일만 경과되면 꽃눈 분화를 하는 종자춘화형(種子春化形)이다. 또 어린 식물도 어느 정도 자라기 전에 13℃ 이하의 저온을 받으면 꽃눈이 분화되어 결구가 되지 않는다(한국민족문화대백과사전). 여름철 고온으로 재배가 어려운 채소작물이고, 해발고도가 높아질수록 기온이 점차 낮아지는 고랭지에서 단경기 고품질 생산이 가능하다(한석호 외, 2011). 고랭지 배추는 기상 영향이 가장 큰 작형으로 정식기 가뭄과 생육기 장마 및 태풍의 영향이 가장 크다. 이에 따라 습해와 병충해에 매우 취약한 편이고, 태풍과 집중 호우가 발생할 경우 생리적인 문제뿐만 아니라 포전 자체가 유실되는 경우도 많다(이동현 외, 2018).

표 1. 배추 생육시기별 온도조건

(단위: °C)

구분	최저온도	최적온도	최고온도
발아	4	18 ~ 22	35
생육	2 ~ 5	20 ~ 22	25
결구	5	15 ~ 18	23

자료: 농협종묘개발센터

표 2. 배추 재배시기 및 주산지

		육묘	정식 및 파종	수확	주산지
배추	겨울배추	8월 중순 ~ 9월 상순	9월 중순 ~ 9월 하순	12월 하순 ~ 3월 중순	해남, 진도, 무안
	봄배추	12월 하순 ~ 2월 하순	1월 하순 ~ 4월 하순	4월 상순 ~ 7월 상순	나주, 예산, 아산, 영월
	고랭지배추	3월 중순 ~ 4월 상순	4월 중순~ 6월 상순	6월 중순 ~ 10월 하순	강릉, 태백, 정선, 평창
	가을배추	7월 하순 ~ 8월 중순	8월 하순 ~ 9월 중순	10월 중순 ~ 12월 하순	아산, 춘천, 문경, 해남

자료: 농림축산식품부, 2020. 배추 수급 동향 및 전망

2.3. 선행연구 검토

본 연구에 앞서, 배추 가격 및 단수 예측에 관한 선행연구를 검토하였으며 주요 연구 방법은 배추 수급에 영향을 미치는 변수와의 상관관계 분석 및 예측을 실시하는 것이었다. 설명변수는 재배면적, 기온, 습도, 강수량 등 작물 생육에 영향을 미치는 변수였다.

표 3. 선행연구 비교표

연구명	연구내용	주요 사용 변수
인공신경망 채소류 가격·단수 예측모형 연구	농산물 가격 불안정성에 대응하는 정보 제공을 위해 인공신경망 기반 채소류 가격도매시장 가격 예측 모형을 개선하고 단수 예측 모형을 개발하는 것을 목적으로 함	기온, 습도, 강수량, 풍향, 풍속, 일사량, 일조시수, 강수확률, 폭염연중일수, 폭염지속일수, 호우일수, 한파연중일수, 한파지속일수, 강수지속일수, 무강수 지속일수, 누적강수량, 적산온도, 가뭄해, 냉해, 일조부족, 흑서, 가뭄, 일조부족, 동해, 다우, 강풍, 월평균기온, 최고 및 최저기온, 일조량, 습도 80% 이상 일수, 월평균 기온 30°C 이상 발생일수, 0°C 이하 발생일수, 강원도 전체 전년도 고랭지 배추 단수 등

연구명	연구내용	주요 사용 변수
주요 채소 단수예측 시스템 개발 연구	주요 양념채소류의 단수 예측 모형 개발을 위해 선행연구에서 검토한 기상변수와 함께 일교차, 일조시간, 상대습도, 생장도일(GDD) 등 채소의 생육에 영향을 미치는 주요 기상변수를 단수예측을 위한 설명변수로 포함하고 롤오버(roll-over) 방식으로 단수 추정	월평균기온, 월평균 누적강수량, 월평균 일교차, 월평균 상대습도, 월평균 일조시간, 생장도일
주산지 기상정보를 활용한 주요 채소작물의 단수예측모형 개발	다섯 가지 주요 채소작물을 각 작물 주산지별로 구축된 수량 및 기상정보를 활용하여 각 작물의 단수를 예측하는 회귀식 기반 단수 예측 모형으로써, 각 작물과 지역별 상관성이 높은 세 가지 기상인자를 선정함	최저기온, 일사량, 유효 강수량
기상요인을 고려한 단수 예측 모형 개발연구	선행연구의 점추정의 문제점을 극복하기 위해 확률 추정 방법을 적용하고 단수와 기상요인과의 선형성을 이차 형식으로 변경함으로써 개별 기상요소에 대한 최적점을 도출하였으며 패널 분석을 통해 자유도가 늘어나 추정의 효율성을 향상시킴	강수량(순별합계), 전반기 최고기온(기간 평균), 후반기 최고기온(기간평균)
시장개방하의 배추 수급모형과 전망	김치의 수입 수요 모형에서 국내의 작형별 출하시기별로 구분하여 작형별 배추의 파급영향을 분석하고 현실성 높은 예상 시나리오를 구성하여 수입 예상 가격, 국내 농가판매 가격, 국내소비자가격, 수출입량, 재배면적, 생산량, 소비량, 자급률, 생산액, 농업소득 등 생산, 소비, 수출입, 소득에 대한 구체적인 파급영향을 분석	재배면적, 소득 1인당 소비량, 소비자가격, 농가판매가격, 김치수입단가, 김치수입량, 김치수출량, 김치소비자가격, 김치 수출단가, 일본김치소비자가격, 장래인구추계, 시간변수, 단수

2.4. 차별점

기존의 선행연구는 배추의 수급 예측 및 가격 불안정성에 영향을 미치는 요인을 규명하기 위하여 도매시장 가격, 단수 예측 등을 실시하였고 이와 관련하여 중요한 설명변수로 기상 변수를 사용하였다. 날씨의 배추 생산량에 중요한 요소인데, 2020년 최장 장마 기록 경신이 말해주듯 기상변수는 계속해서 불확실하게 변화하고 있다. 기상변수를 평균으로 사용할 경우에 이러한 불확실한 변화를 반영하는데 취약하고, 기상변수에 대한 파생 변수가 다양해짐에 따라 복잡성은 증가하여 강건한 모형을 기대하기 어려워진다. 또한 날씨와 가격 이외에도 생산량에 중요한 요소인 재배면적이나 현장 상황의 변화를 정확하게 반영하지 못하여 연구의 정확성이 떨어지는 한계가 있다.

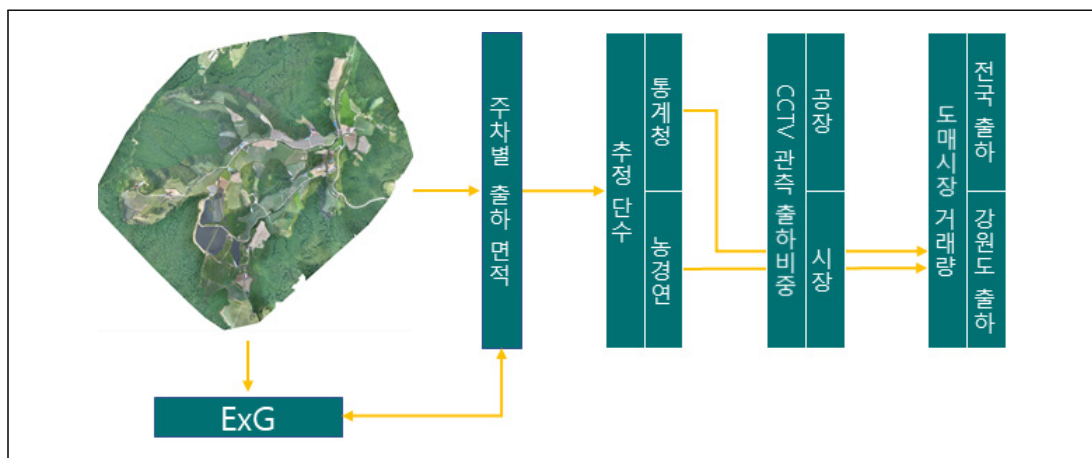
본 연구는 농경지에서 관측사업으로 4개월 동안 수집한 주차별 강원도 고랭지 배추 재배지

의 드론 이미지를 활용한다. 드론 이미지는 현장의 상황을 설명하는 데이터로써, 수급 예측에 필요한 정보를 추출하여 수급 안정화에 유용하게 사용할 수 있다.

이번 연구에서는 드론 이미지에서 추출한 데이터로 주차별 출하량을 도출하여 도매시장 반입량과의 관계를 분석한다. 결과적으로 드론 이미지를 활용한 수급 예측 체계에 대한 아이디어를 제시한다.

3. 연구내용

그림 2. 연구 흐름도



주: 1) ExG : Excess Green Vegetation Index

3.1. 전체 연구 결과

1. 농경면에서 추정하는 단수와 통계청의 단수를 비교한 결과, 농경면의 단수가 더 합리적임.
2. 드론 이미지 출하면적 라벨링에 대한 개선사항을 도출함.

- 가. 앞서 수확되었다고 기록한 지역을 다음 기간 출하면적에 포함하여 면적이 중복 계산되거나 파일이 중복되는 경우가 있음.
 - 나. 배추밭 주변의 숲, 산, 도로 등을 포함하여 출하면적이 과대평가되는 경우가 있음.
 - 다. 재배 중인 배추밭을 수확 면적에 포함한 경우가 있음.
3. 드론 이미지에서 추출한 출하면적으로 출하량을 산출하는 것은 도매시장 반입량과 비교하여 의미 있는 결과를 냄.
- 가. 강원도 도매시장 누적 반입량과 드론 이미지로 산출한 누적 출하량의 상관계수는 0.99임.
 - 나. 드론 이미지에서 산출한 누적 출하량은 강원도 도매시장 누적 반입량의 92% 수준임.
 - 다. 강원도 도매시장 누적 출하량과 드론 이미지 누적 출하량은 약 160만 kg의 차이가 있음. 이 차이를 줄일 수 있도록 드론 이미지의 품질 제고와 관리 방안을 갖추면 드론 이미지로 출하량을 모니터링할 수 있음.
4. 드론 이미지의 RGB 값을 활용하여 산출한 ExG는 수확 면적과 -0.71의 음의 상관관계를 가짐.
5. ExG와 ExG 변화량으로 수확 시점을 알 수 있음.

3.2. 드론 이미지에서 산출한 출하량으로 통계청과 농경연 단수 검토 및 도매시장 반입량과의 통계 패턴 분석

3.2.1. 가설

해당 연구의 가설은 농경연에서 추정하는 단수와 통계청의 단수 중 더 정확하게 출하량을 산출할 수 있는 단수가 있다는 것이다. 전체 연구를 진행하기 위하여 중요한 변수는 재배면적과 단수인데, 고려할 수 있는 단수는 농경연과 통계청의 두 가지이다. 따라서 이 두 단수를 검토하여 본 연구에 적용하는 한편, 검토 과정에서 드론 이미지에서 추출한 생산량과 도매시장 반입량의 통계 패턴을 비교한다.

3.2.2. 연구 환경 설정

1. 검증 대상인 도매시장 반입량 데이터는 정확한 것으로 간주한다.
2. 도매시장 반입량 데이터와 비교하기 위하여 드론 이미지에서 산출한 출하량에 시장 출하 비중을 반영하며, 시장 출하 비중은 고랭지 배추 출하용 CCTV 집계 자료를 활용하여 평균값인 75%로 적용한다.

3.2.3. 사용데이터 및 변수설명

해당 연구를 위해 드론 이미지 등 3개의 데이터를 활용하였으며, 각 데이터의 기간은 드론 촬영 일자를 기준으로 조정하였다.

표 4. 사용데이터 및 변수 설명

사용 데이터	변수명	내용
농경연 제공 드론 이미지 데이터	'level'	드론 촬영일자로 기간의 구분을 설정하였음. 데이터의 유무를 고려하여 2차~9차로 범위를 설정하였으며, 2020년 7월 20일~9월 25일 촬영분에 해당
	'sum_area'	농경연 드론 이미지 shp 파일의 고랭지 배추 재배면적 합계
농경연 도매시장 반입물량 자료	'gangwon_all'	강원 출하 전국 반입 도매시장 거래량 드론 촬영 기간과 동일한 기간의 데이터를 사용
농경연 농업관측본부 고랭지 배추 출하용 CCTV 집계 결과	'cctv_market'	시장으로 출하된 차량 대수로 산출한 출하량
	'cctv_factory'	공장으로 출하된 차량 대수로 산출한 출하량
	'cctv_all'	시장, 공장 합산 출하량
파생변수	'농경연(kg)'	농경연의 단수와 드론 이미지의 재배면적을 곱하여 산출한 출하량
	'통계청(kg)'	통계청의 단수('17~'19년 3년 평균)와 드론 이미지의 재배면적을 곱하여 산출한 출하량
	'krei_market'	CCTV 관측자료의 시장 평균 출하 비율(75%)을 적용하고 농경연 단수와 드론 재배면적을 곱한 값(출하량)
	'kostat_market'	CCTV 관측자료의 시장 평균 출하 비율(75%)을 적용하고 통계청 단수와 드론 재배면적을 곱한 값(출하량)
	'gangwon_itersum'	강원도 출하 전국 반입 도매시장 누적 거래량
	'krei_market_itersum'	농경연 단수와 시장 출하 비율을 적용한 누적 출하량
	'kostat_market_itersum'	통계청 단수와 시장 출하 비율을 적용한 누적 출하량

주: 1) shp 파일 : 배추밭을 특정할 수 있는 id, 배추밭의 좌표, 고도 등에 대한 정보 기록

2) tif 파일 : 드론으로 강원도 배추 재배지를 촬영한 사진을 가공해 지도처럼 만든 영상

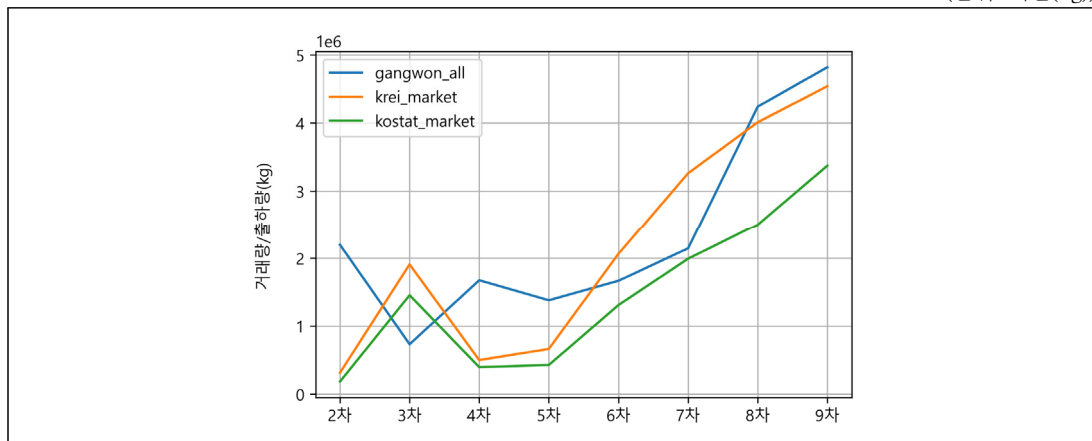
3.2.4. 절차

1. 드론 이미지 shp 파일별 고랭지 배추 면적 추출
2. 촬영 주차별 면적 합산
3. CCTV 집계 자료에서 드론 이미지 촬영 주차별 평균 시장 출하 비중 산출
4. 시장 출하 비율과 농경연, 통계청 단수를 반영한 출하량을 도매시장 반입량과 비교

3.2.5. 연구 결과

그림 3. 강원도 출하 도매시장 반입량과 농경연, 통계청 시장 출하량 비교

(단위: 백만(kg))



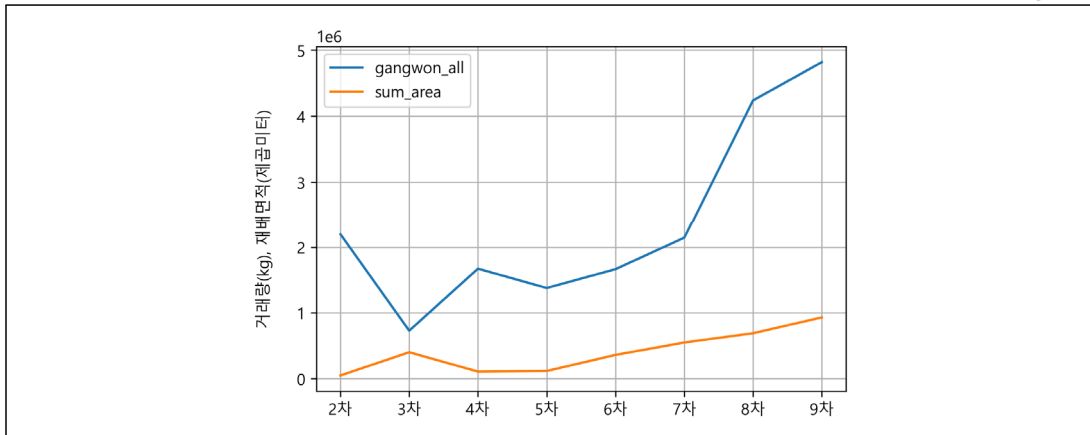
- 주: 1) gangwon_all : 강원 출하 전국 반입 도매시장 거래량
 2) kreimarket : 평균 시장 출하 비율(75%)을 적용하고 농경연 단수와 재배면적을 곱한 값(출하량)
 3) kostatmarket : 평균 시장 출하 비율(75%)을 적용하고 통계청 단수와 재배면적을 곱한 값(출하량)

표 5. 강원도 출하 도매시장 반입량과 농경연, 통계청 단수를 적용한 출하량의 상관관계

구분	상관계수
농경연 단수 적용 출하량	0.762889
통계청 단수 적용 출하량	0.765946

〈그림 3〉에서는 농경연의 단수를 적용한 결과가 그래프 상으로 나온 결과를 보였고 상관관계 수에서는 통계청의 단수가 약 0.003 높았다(표 5). 하지만 두 단수를 적용한 출하량 모두 도매시장 반입량과 서로 다른 변화 패턴을 보였다.

그림 4. 강원도 출하 도매시장 반입량과 드론 이미지에서 추출한 출하면적 비교

(단위: 백만(kg, m²))

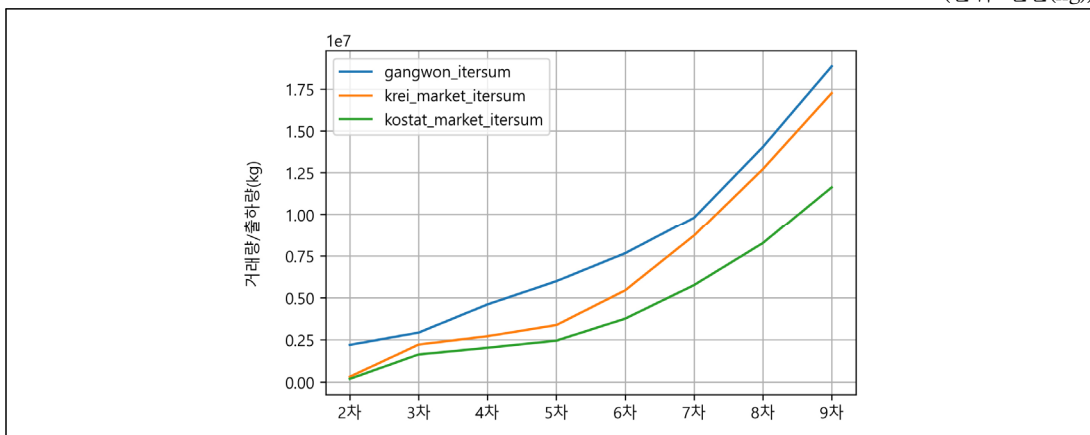
주: 1) gangwon_all : 강원 출하 전국 반입 도매시장 거래량

2) sum_area : 농경면 드론 이미지 shp 파일의 고랭지 배추 재배면적 합계

두 단수를 적용한 패턴이 도매시장 반입량과 다른 원인을 드론 이미지에서 추출한 면적 합의 패턴에서 확인할 수 있었다(그림 4). 따라서 출하면적에 대한 드론 이미지의 라벨링이 정확한지 확인하는 과정이 필요하였다.

그림 5. 강원도 누적 출하량과 농경면 단수 누적 출하량 검토

(단위: 천만(kg))



주: 1) gangwon_itersum : 강원도 출하 전국 반입 도매시장 누적 거래량

2) krei_market_itersum : 농경면 단수 적용 시장 누적 출하량

3) kostat_market_itersum : 통계청 단수 적용 시장 누적 출하량

표 6. 도매시장 누적 반입량과 농경연, 통계청 단위 적용 누적 출하량 비교

구분	농경연 단위 적용	통계청 단위 적용
누적출하량(kg)	17,248,600	11,623,926
강원도 도매시장 누적 반입량과의 차이(kg)	1,600,247	7,224,921
강원도 출하 도매시장 반입량 비교 비율(%)	91.51	61.67
누적 출하량 상관계수	0.994588	0.995190

주: 1) 강원도 출하 도매시장 누적 반입량 : 18,848,847kg

누적 출하량 그래프에서 농경연의 단수를 적용한 출하량이 상대적으로 적은 차이로 유지되는 패턴을 확인할 수 있다(그림 5). 또한, 누적 출하량 차이에서 통계청의 단수를 적용한 출하량이 농경연의 단위보다 약 4.5배 정도 큰 오차를 보였다(표 6). 따라서 본 연구에서는 농경연의 단수를 사용하는 것이 합리적이라고 판단하였다.

농경연 단수를 적용한 출하량은 강원도 출하 도매시장 반입량의 약 92% 수준이었다. 시장 출하 비율, 드론 이미지의 출하면적 오차, 드론 이미지의 촬영 면적(강릉, 삼척, 정선 지역에 해당)을 고려하였을 때 이 수치는 드론 이미지를 통한 출하량 산출 방법론이 의미 있는 결과를 나타냄을 시사한다.

3.3. 드론 이미지에서 추출한 EXG와 출하시기 관계

3.3.1. 가설

필지별 드론 이미지의 색감은 배추의 생육단계에 따라 어떠한 특성을 가질 것이다. 이미지의 RGB 값을 활용하여 특성을 추출하고 출하시기와와의 관계를 분석한다.

3.3.2. 연구 환경 설정

1. ExG(Excess Green Vegetation Index)를 이용해 생육단계에 따른 특성을 파악한다.

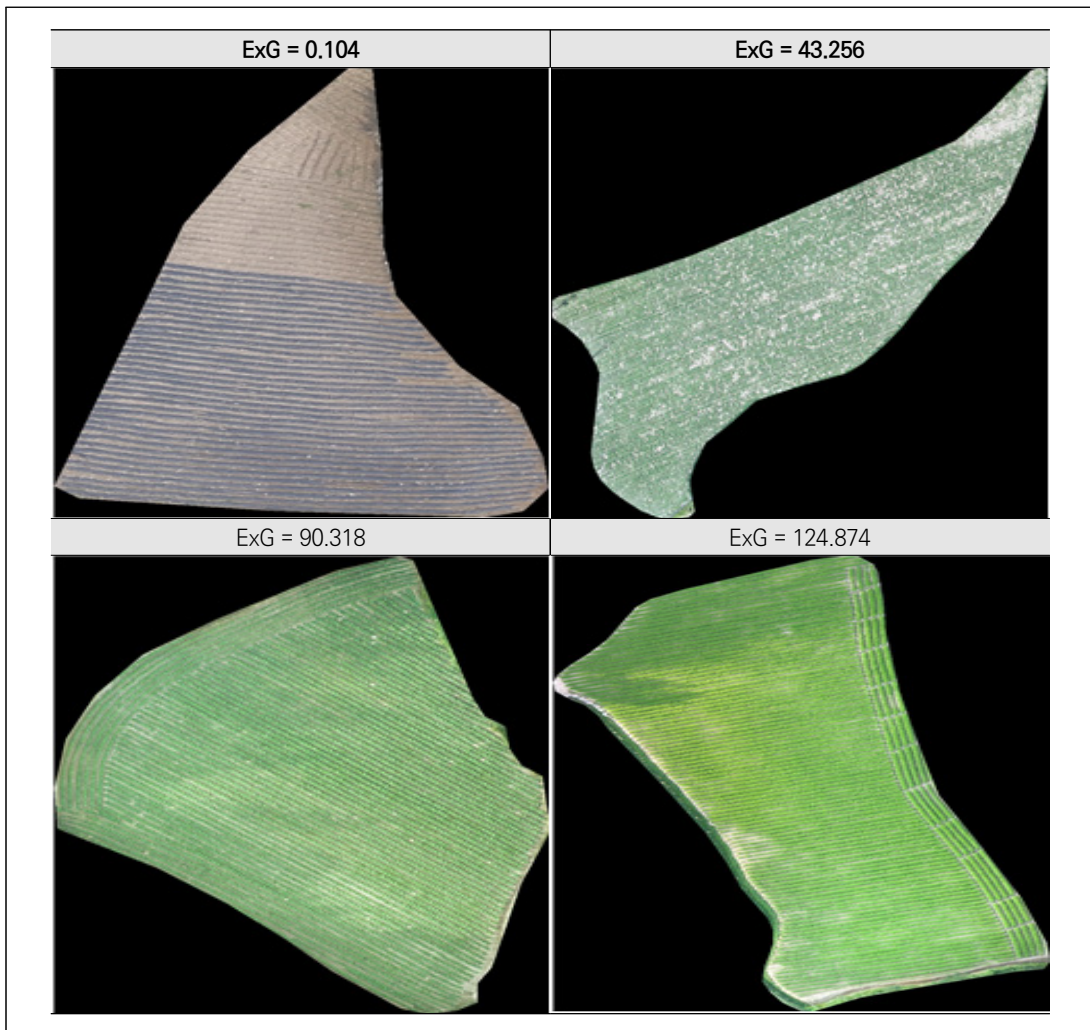
가. ExG는 초과녹색채지수로 이미지의 RGB를 이용해 픽셀당 녹색 정도를 수치화한 것이며 산출식은 다음과 같다.

$$\text{ExG} = (G * 2) - R - B$$

나. 해당 연구에서 ExG 값이 음수이면 0으로 변환한다.

2. 드론 이미지 촬영 상태, 날씨가 RGB에 영향을 준 것으로 간주되는 이미지는 이상치로 설정한다.

그림 6. 이미지별 ExG 산출



3.3.3. 사용데이터

농경연에서 제공하는 792개 필지의 드론 이미지를 필지별, 회차별로 추출하였다.

표 7. 결측 및 이상필지를 제거한 사용 데이터

촬영 회차	고단리(남)	고단리(북)	대기리	안반데기	어리	번천리	장전리	판문리	새비재
기존	2~9	2~9	2, 3, 5~9	2, 3, 5~9	3~9	3~9	3~9	3~9	3~9
제거 후	2~5, 7~9	2~5, 7~9	2, 3, 5, 6, 8, 9	5~9	3~9	4~9	3~9	3~9	3~9

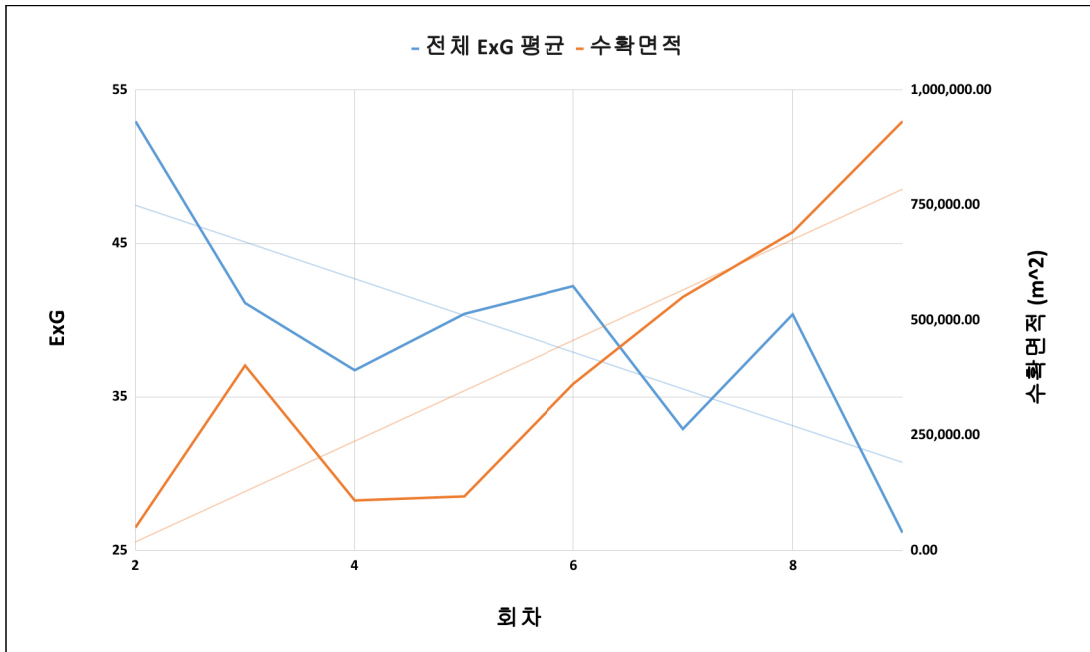
3.3.4. 절차

1. 데이터 전처리
2. 이미지별 ExG 추출 및 이상치 제거
3. 필지별 ExG 변화량 추출
4. 필지별 수확 시점 추출
5. 변화량과 ExG의 시사점 도출

3.3.5. 연구 결과

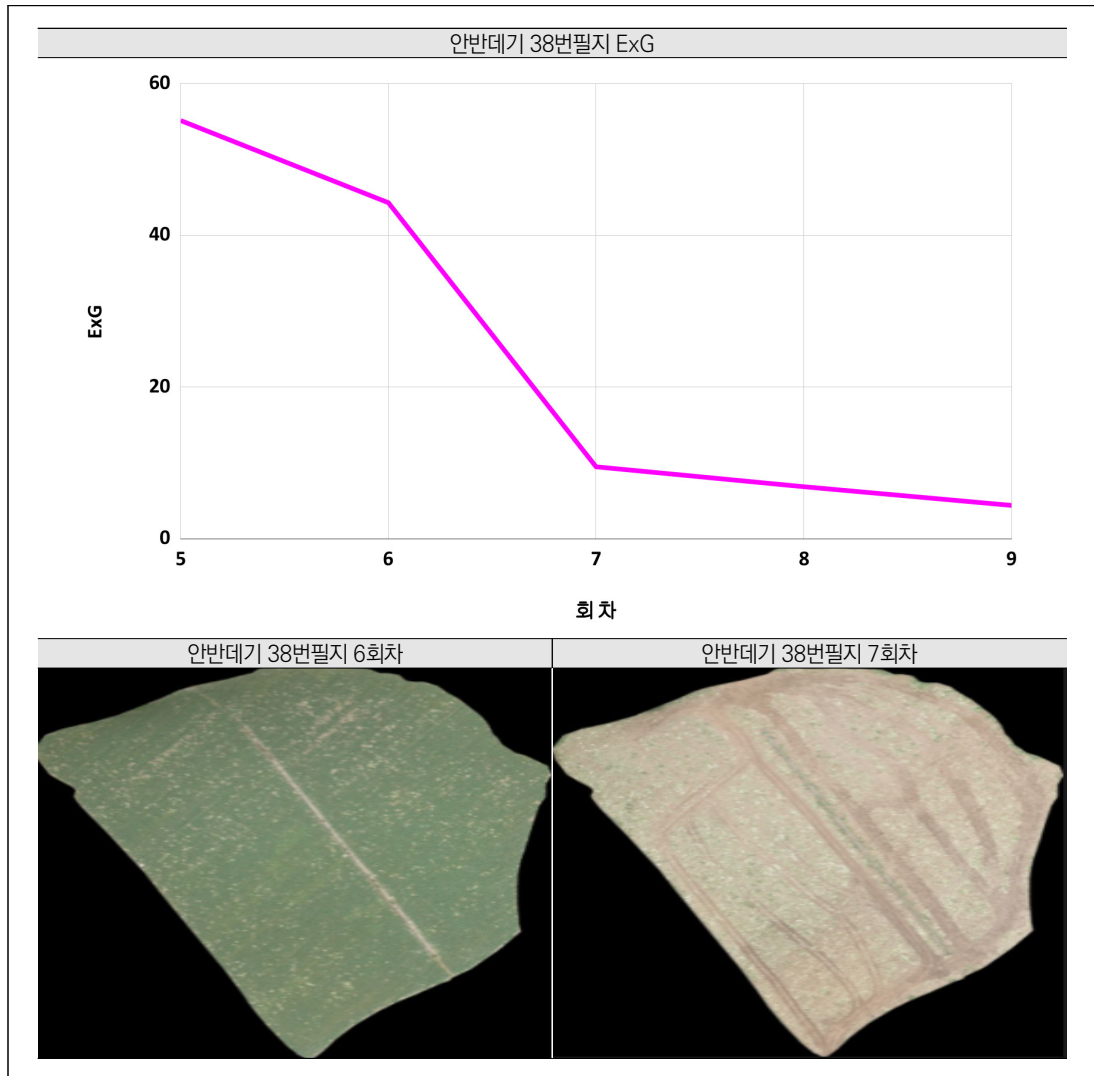
모든 필지의 ExG 평균과 전체 수확 면적 합계를 통해 ExG가 감소하면서 수확 면적은 증가하는 것을 확인할 수 있었다(그림 7). 두 선의 상관계수는 -0.71로 ExG와 수확 면적은 음의 상관관계를 가지고 있다.

그림 7. 모든 필지의 회차별 ExG평균, 전체 수확면적 그래프



수확 시점의 ExG는 큰 폭으로 감소한다(그림 8). 전체 필지를 대상으로 수확될 때의 ExG 변화량을 확인해본 결과, 가장 큰 폭으로 감소했을 때 대부분이 실제로 수확되었음을 확인하였다.

그림 8. ExG세부 결과



추가적인 검토를 위해 수확될 때 낮은 ExG를 가지는 특성을 포함하도록 다음과 같은 수식을 만들었다.

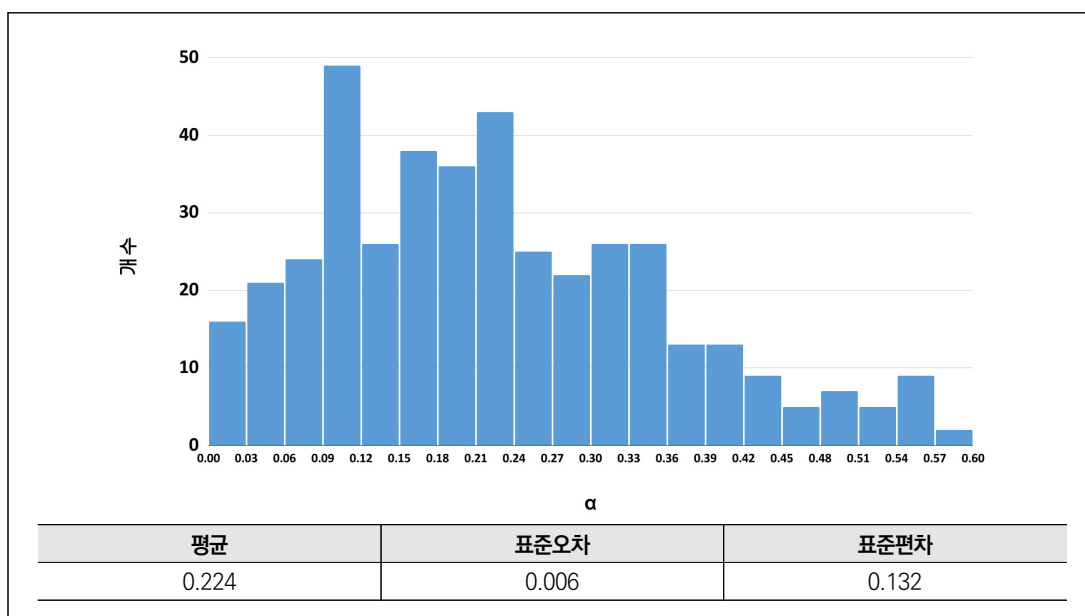
$$n \text{ 회차 } \alpha_n = E_n / (E_{n-1} + |V_n|)$$

* E_n : n 회차 ExG, E_{n-1} : $n-1$ 회차 ExG, V_n : $E_n - E_{n-1}$ (즉, 변화량)

〈그림 9〉는 수확 시점 α 의 히스토그램이다. 신뢰구간을 95%로 가정했을 때, α 값이 0.483보다 낮다면 수확 시점으로 간주할 수 있다.

해당 연구를 통해 산출한 ExG 값은 드론 이미지의 촬영 당시 날씨나 그림자의 영향을 받아 RGB 값이 일부 왜곡되어 오차를 내포할 수 있는 한계가 있다. 하지만 이를 극복할 수 있는 수집 방안과 알고리즘을 도입한다면 ExG를 활용한 수확 시점 확인이 가능함을 검증하였다.

그림 9. 수확시점 α 히스토그램



4. 결론

4.1. 기대효과

수급을 예측하기 위하여 통계나 경제학적 방법으로 모형을 밝혀내는 것도 의미가 있으나, 대부분의 모형은 자연이라는 불확실성과 외부 변수들로 무용지물이 될 수 있는 위험성을 가

지고 있다. 본 경진대회 목적인 농업관측 고도화에 기여할 수 있는 예측 모형을 개발하려면, 양질의 데이터를 수집하는 것이 선행되어야 한다. 양질의 데이터는 현장의 상황을 정확히 반영할 수 있으면서도 최대한 단순하여 모형의 복잡도를 높이지 않는 것이어야 한다. 현재 농경연에서는 농가의 실황을 반영하기 위하여 여러 데이터를 수집하고 있다. 그중 하나인 드론 이미지에는 실제 농가의 촬영 기간별 재배면적, 생육, 출하여부 등의 정보가 내포되어 있다. 때문에 본 연구는 드론 이미지를 활용하는 것에 초점을 두었다.

드론 이미지에서 추출한 재배면적으로 산출한 출하량을 도매시장 반입량과 비교하여 유의미한 관계를 검증함으로써 출하량 관측에 드론 데이터 활용이 가능함을 밝혀내었으며, 데이터 분석 실무자의 관점에서 데이터의 개선사항을 도출하여 데이터 수집에 대한 방향을 제시한다.

드론 이미지로 재배 실황을 정확하게 모니터링하고 이를 데이터화한다면 이미지의 RGB 값으로 수확 시점을 알 수 있고, 딥러닝을 활용한 배추 재배면적 판별을 시도할 수 있으며, 드론 이미지를 활용한 수급 예측 모델링이 가능할 것이다.

4.2. 연구의 한계점

가) 생육조사 데이터의 한계

생육조사 데이터에서 조사를 실시한 필지의 구체적인 지역 정보가 부족하고, 조사일별 조사 필지가 변동되어 연속적인 변화를 확인하기 어렵기 때문에 드론 이미지와 연계한 분석을 실시할 수 없었다.

나) 드론 이미지의 한계

1. 농경연에서 제공한 드론 이미지 데이터는 2020년 7월부터 10월까지이므로 다른 연도의 결과와 비교 검증은 제한된다.
2. 정사영상, shp 출하면적 파일의 오류로 인한 연구의 오차

아래와 같은 오류의 영향으로 드론 이미지에서 추출한 출하면적 추세가 도매시장 반입량과 달랐던 것으로 추정된다.

그림 10. 초기 촬영지와 수확지가 다른 경우



그림 11. 수확되지 않았지만, 수확되었다고 표기한 경우

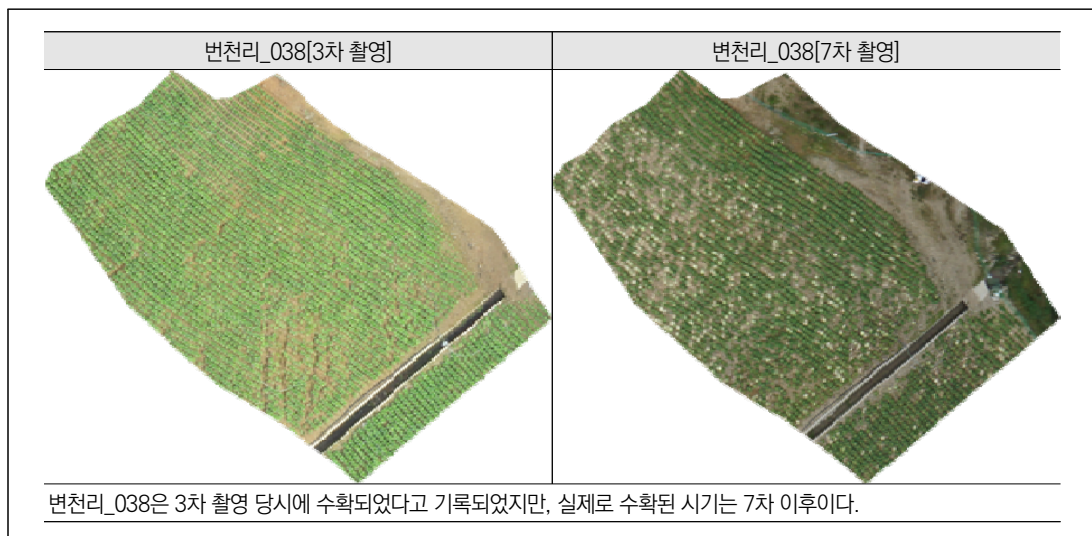


그림 12. 수확되었다고 기록했지만 실제로는 일부분만 수확된 경우

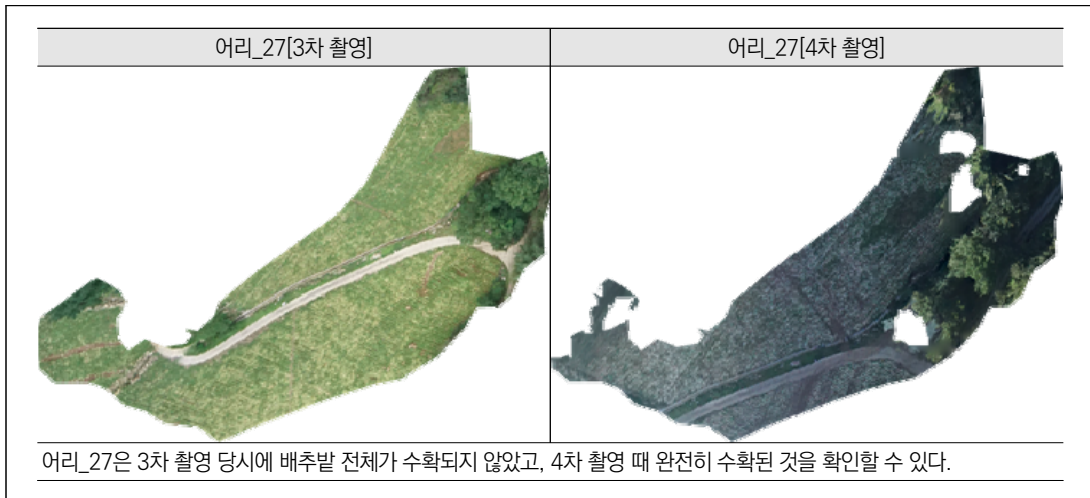
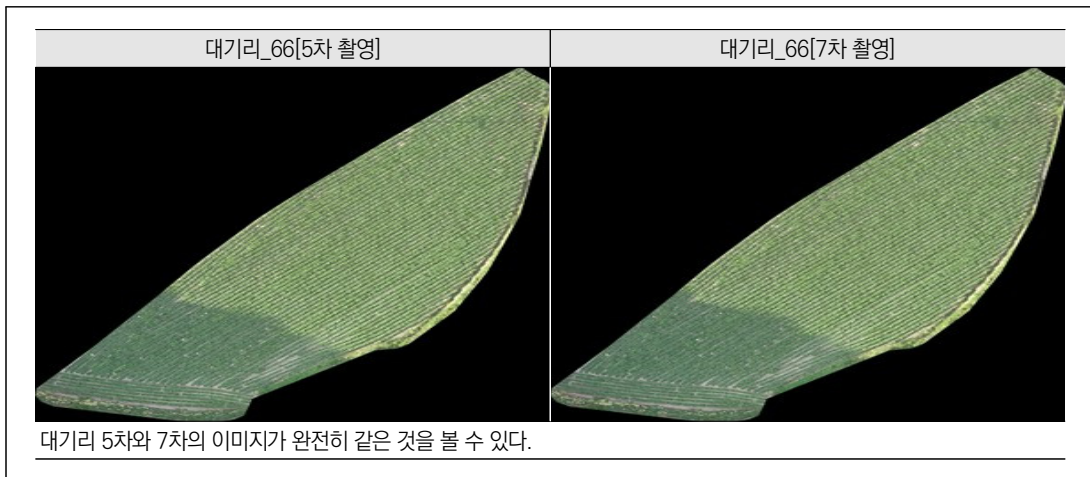


그림 13. 다른 차수의 tif파일이 중복된 경우

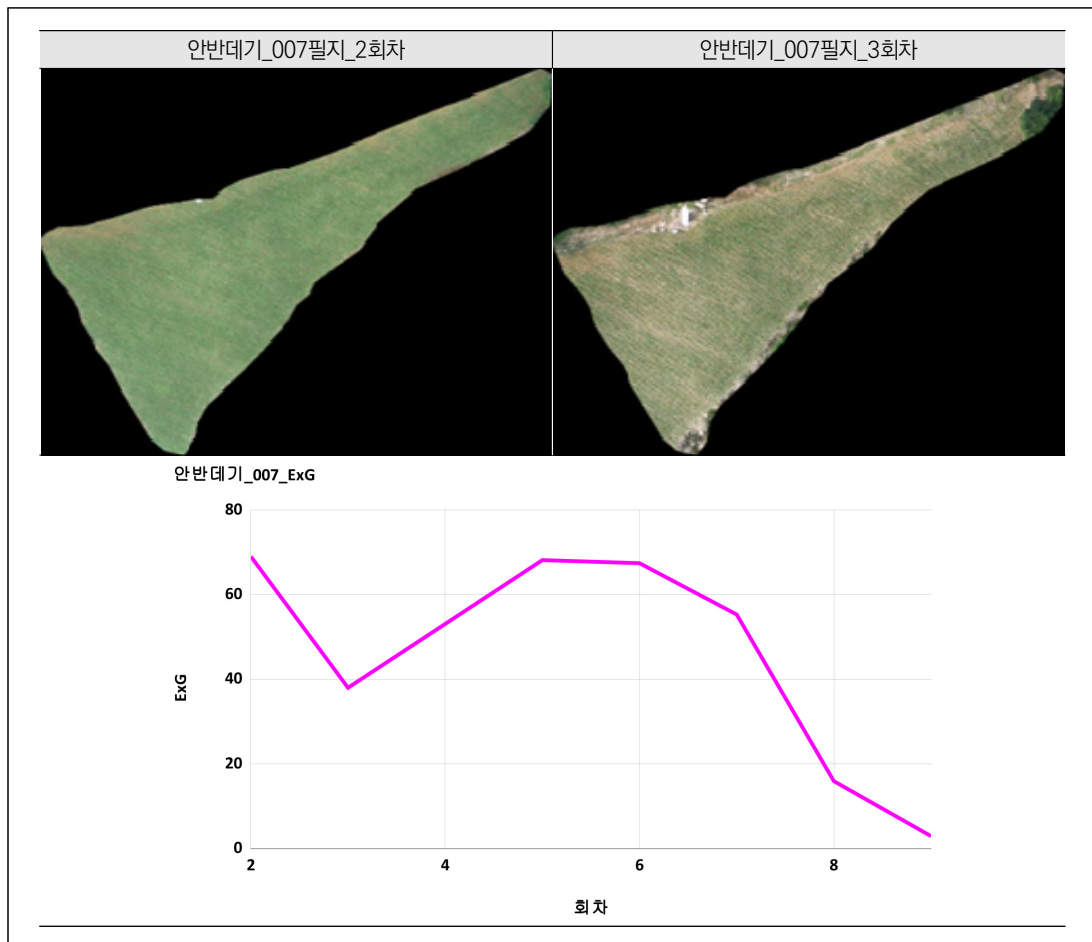


3. 딥러닝 학습에 부적합한 데이터 품질

드론 이미지에서 정확한 배추 필지를 추출하는 것은 본 연구에서 중요한 과업이며, 많은 이미지를 효율적으로 처리하기 위해 딥러닝을 활용한 배추 필지 판별을 시도하였다. 하지만 높은 고도에서 촬영되었기 때문에 이미지의 해상도가 딥러닝 모델 학습에 적합하지 않았다.

4. ExG는 촬영 당시 날씨나 그림자가 일부 RGB 값을 왜곡시켜 영향을 받는 한계가 있다. 안반데기 7번 필지를 보면 2회차 이미지를 찍을 당시 상황의 영향을 받아 ExG 값이 과도하게 높고, 3회차는 낮은 것을 볼 수 있다(그림 14). 따라서 본 연구는 이와 같은 경우를 이상치로 간주하였다.

그림 14. ExG 이상치 예시



다) 외부요인

병충해 및 기상재해가 생육에 영향을 미치는 요인은 본 연구에서 배제하였다.

4.3. 발전 방향

4.3.1. 데이터 수집

가) 생육조사데이터

드론 이미지 촬영지와 CCTV 출하 관측지, 생육 조사 필지, 조사 주기를 동일하게 설정한다면 본 연구의 방법론을 고도화하는 분석이 가능할 것으로 기대된다.

나) 드론 이미지

1. 딥러닝을 위한 품질 요구 사항

농경면에서 제공한 드론 이미지는 해상도가 높지 않아 눈으로 작물을 판별하기 어렵고, shp 파일의 레이블이 없었다면 배추 필지만 구별하여 연구를 실시할 수 없었을 것이다. 딥러닝 모델의 학습 및 입력 데이터로 사용하기 위해서는 25~30m의 저고도에서 촬영하는 것이 필요하다.

2. ExG를 활용하기 위한 품질 요구 사항

기상조건, 고도, 해상도를 일정하게 유지하고 2~3일의 짧은 주기로 촬영하는 것이 필요하다.

4.3.2. 딥러닝 판별모델 적용

현재의 드론 이미지 수집 과정은 정사영상과 실제 재배작물 현황을 비교하는 현장 전수조사가 포함되어 있고, 연구의 한계점에서 검토한 바와 같이 shp 파일의 라벨링 오류 위험이 있다.

드론 이미지의 품질을 개선한다면 작물 판별 딥러닝 학습이 가능할 것이고, 매년 드론 이미지가 대량으로 수집되더라도 딥러닝 모델을 통해 단시간 안에 판별이 가능할 것이다. 작물 판별 결과는 shp 파일 생성과정에 활용할 수 있을 것이며, 모델의 성능에 따라 현장 전수조사 소요는 줄어들어 드론 이미지의 활용도를 높일 수 있을 것으로 기대된다.

참고문헌

- 농림축산식품부. 2020. 「배추 수급 동향 및 전망」.
- 농림축산식품부. 2020. 「시설채소온실현황 및 생산실적」.
- 최석근 외. 2020. “무인항공기와 딥러닝을 이용한 소규모 농지의 발작물 분류.” 한국측량학회.
- 류상모. 2019. “인공신경망 채소류 가격·단수 예측 모형 연구.” (사)농식품신유통연구원.
- 이동현 외. 2018. “주요 채소 단수 예측시스템 개발 연구.” 농촌진흥청.
- 류상모. 2017. “농산물 수급안정정책에 대한 실증 연구.” 서울대학교 경제학박사학위논문.
- 한국민족문화대백과사전. 「배추」.
- 임철희 외. 2016. “주산지 기상정보를 활용한 주요 채소작물의 단수 예측 모형 개발.” (사)한국기후변화학회.
- 김명환 외. 2015. “배추·무 수급불안의 원인과 대응방안.” 농촌진흥청.
- 한석호 외. 2011. “기상요인을 고려한 단수예측모형 개발 연구.” 한국농촌경제연구원.
- 한석호, 김병률. 2004. “시장개방하의 배추 수급모형과 전망.” 한국농촌경제연구원.

부록

1. 단수예측에 대한 검토

수급 예측에는 단수가 중요한 요소이기 때문에 단수 예측을 위해 날씨, 재배면적, 도매시장 반입량, 도매시장 가격 등에 대한 ARCH, GARCH 분석을 고려하였다. 하지만 단수 데이터의 수집 방법, 신뢰성의 한계가 있기 때문에 단수 예측을 비롯한 수급 예측을 본 연구에서는 다루지 않았다. 미래의 농촌 경제 성장에 수급 예측이 중요한 만큼 단수 데이터를 정교화하고 표본이 되는 측정법이 필요하다.

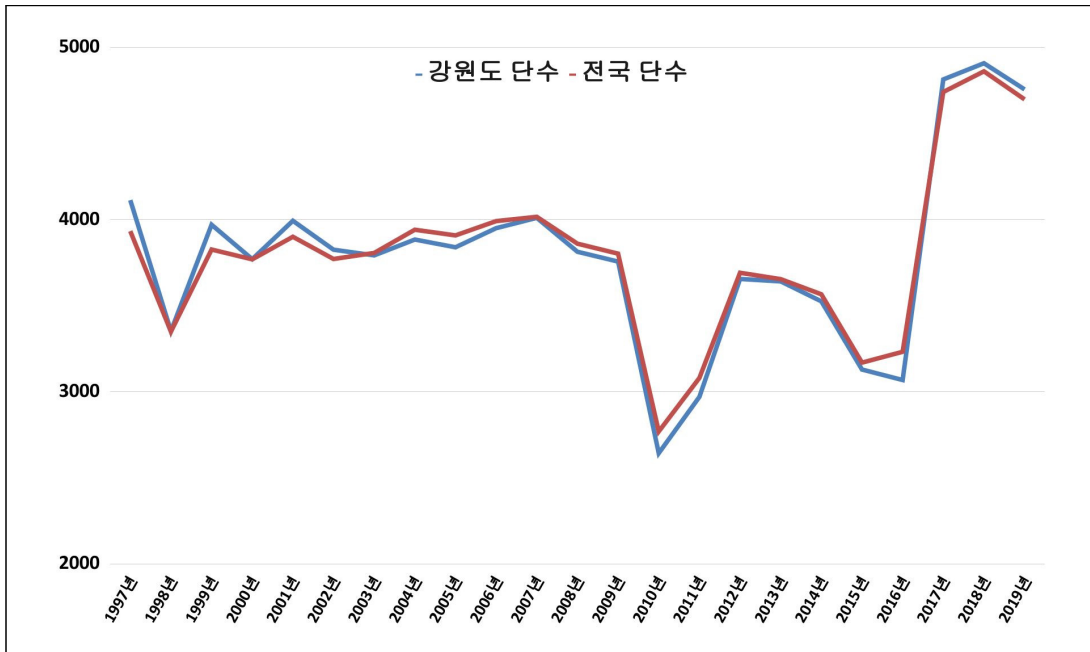
가) 단수 데이터 수집 방법

현재 통계청에서는 고랭지 배추 단수를 연도별, 지역별로 제공하고 있다. 단수 정보를 정확하게 활용하기 위해서는 수확 시점의 필지별로 단수가 산출되어야 한다. 또한 단수는 산출되는 시점의 외부요인에 영향을 받는데 연 단위 단수는 중요 요인들을 규명하기 어렵다.

나) 데이터의 신뢰성

통계청 자료에 따르면 1997년부터 2008년까지 3,500~4,000kg/10a의 단수가 기록되었고 2009년부터 2016년까지는 평균 3,100kg/10a의 낮은 단수가 기록되었다. 2017년부터 2019년까지는 다시 평균 4,900 kg/10a의 높은 단수를 기록하면서 심한 격차를 보였다(그림 15). 또한, 농경연의 2020년 고랭지 배추 주산지(매봉산 외 3)의 단수는 평균 5,960.25kg/10a를 기록하면서(표 8) 통계청의 기록과 차이가 있어, 현재 제공되고 있는 단수 데이터에 대한 검토가 필요함을 발견하였다.

그림 15. 통계청 단위자료



자료: 통계청

표 8. 농경연 단위 자료

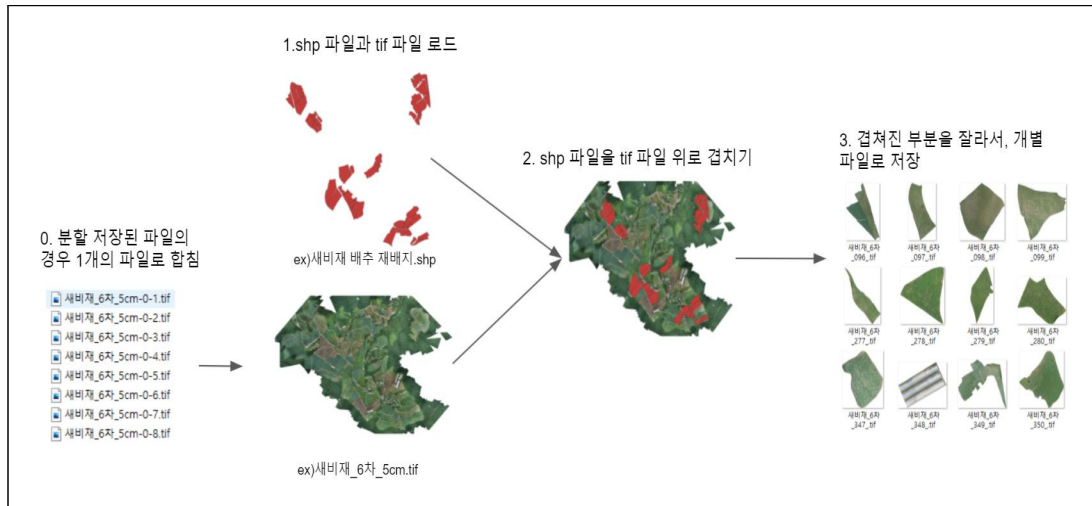
지역	단수(kg/10a)
매봉산	4,782
안반덕	8,318
새비재	4,704
귀네미	6,037

자료: 농경연 2020년 농업관측본부 고랭지 배추 출하용 CCTV 집계 결과

2. 드론 이미지 배추필지 추출 과정

실제 배추를 재배 중인 필지의 특성을 확인할 수 있도록 드론 이미지에서 배추 필지의 tif 파일을 추출하였다.

그림 16. 드론 이미지 필지 추출 가공 과정



1. 드론 정사영상이 완전한 지도 형태로 저장되도록 가공
2. 배추 재배지의 좌표 정보가 저장된 shp 파일과 배추 재배지의 정사영상 tif 파일을 적재
3. 각각의 파일이 중첩되도록, 좌표계를 일치
4. shp 파일의 좌표 정보와 배추밭 id를 기준으로 tif 파일 추출

3. 고랭지 배추 필지 판별 딥러닝 모델

고랭지 배추 필지를 딥러닝을 기반으로 판별하기 위하여 필지 전체를 입력 데이터로 사용하는 방법과 관심부분을 추출하여 사용하는 방식을 적용하였다. 두 방식 모두 데이터 세트 내의 가장 비중이 높은 클래스로 예측하는 과적합된 결과를 내었다.

그림 17. 입력데이터 방식에 따른 클래스 구분

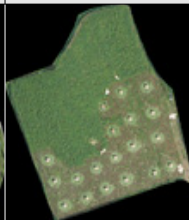
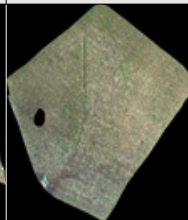
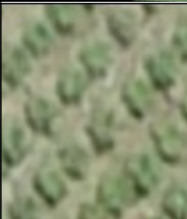
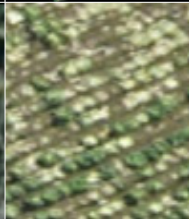
구분	배추	배추 외 작물	수확 중	수확 종료	휴경지
필지 전체					
관심 영역					

표 9. 딥러닝 모델 결과

사용 데이터	구조	학습 결과	테스트 결과
필지 전체	합성곱층 - 풀링층 - 합성곱층 - 풀링층 -	10 epoch 결과, 훈련 데이터에 대한 정확도 80%	데이터의 양이 많은 클래스로 과적합됨
관심 영역 추출	합성곱층 - 전연결층 - 예측	20 epoch 결과, 훈련 데이터에 대해 정확도 90%	정확도 35%로 성능이 낮으며, 훈련세트에 과적합됨

4. 노지작물 판별 딥러닝 모델

본 연구팀에서 과거 구축했던 노지 드론 이미지 작물 판별 모델(그림 18)에 사용한 드론 이미지의 촬영고도는 약 25m이었으며, 배추, 양배추, 무, 마늘, 양파 5개 품목의 분류 결과 CNN 모델은 98%, YOLO 모델은 98.5%의 정확도로 배추를 분류하면서 학습이 가능함을 확인했다.

그림 18. 자체 보유 25m 고도 드론 데이터로 학습한 딥러닝 배추판별 결과 예시



5. 드론 이미지 재배면적 산출방법

5.1. 산출

배추 경작지 shp 파일에 있는 좌표 정보에서 POLYGON 좌표를 추출하고 넓이를 계산하였다.

표 10. 새비재_배추경작지 shp POLYGON 좌표 예시

NAME	geometry
새비재-096	POLYGON((172280.849 511729.648, 172271.491 51...
새비재-097	POLYGON((172391.613 511674.110, 172389.460 51...
...	...

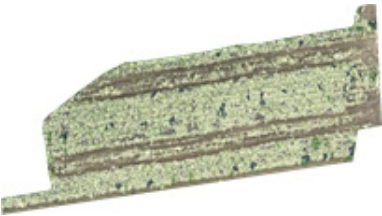
표 11. 추출 POLYGON을 이용한 넓이 계산 예시

지역	필지	면적(m2)	2차 출하면적	3차 출하면적	...	9차 출하면적	총 출하면적
새비재	096	3249.99	-	-	...	3249.29	3249.29
고단리_남	001	806.29	-	805.95	...	43.21	849.16
...		...	...	...	...	...	...

5.2. 조정

재배면적 산출 결과, 총 출하면적이 재배면적보다 큰 필지가 있었으며, 이를 조정하였다. 출하 면적이 재배 면적을 초과한 것을 확인한 필지를 shp 파일 내에서 추출하였다. 파이썬의 GIS 라이브러리 중 하나인 GeoPandas의 intersection 함수를 이용하여 각 조사 기간의 POLYGON 간 중복되는 POLYGON의 넓이를 확인하고 그 차이를 제거하였다.

표 12. 재배면적 조정과정

	대기리_062(2차 조사)	대기리_062(3차 조사)
실제 필지		
넓이 및 폴리곤 형태	668.2477033344436 	1810.9114872807747 
중복 계산 결과	<pre>daegiri_03_62.intersection(daegiri_02_62).area</pre> 668.24770333411 2차 조사 재배면적이 3차 조사 재배면적에 포함되었으므로 3차 조사 재배면적에서 2차 조사 재배면적을 빼는 방법으로 조정한다.	

6. 농경면과 통계청의 단수 적용 출하량 산출 방법

그림 19. 드론 촬영지 근접 CCTV 관측지

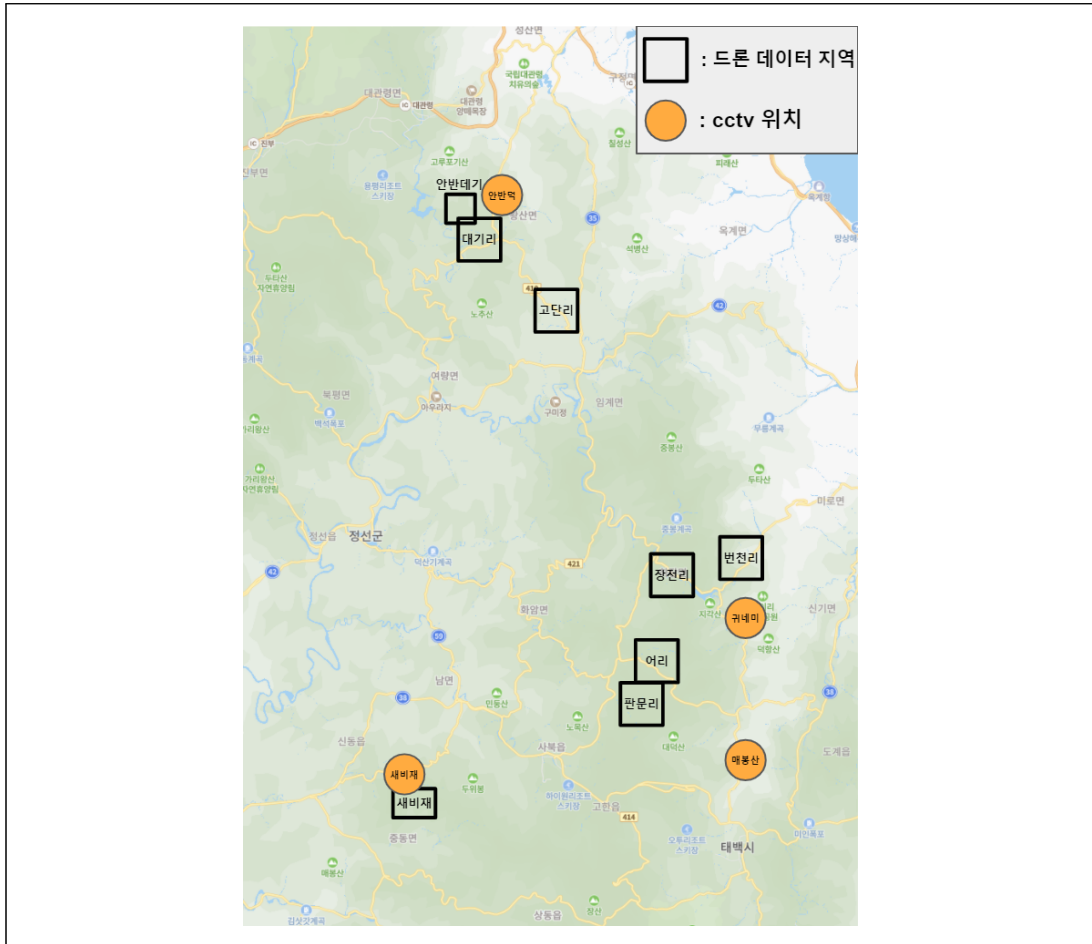


표 13. 드론 촬영지별 근접 CCTV 관측지 단수

드론 이미지 촬영 지역	근접 CCTV 관측지	단수(kg/10a)
안반데기, 대기리, 고단리	안반덕	8318
번천리, 장전리	귀네미	6037
어리, 판문리	매봉산	4782
새비재	새비재	4704

주: 1) 단수는 농경면 농업관측 본부 고랭지 배추 출하용 CCTV 집계 데이터의 자료

농경연의 CCTV 관측 데이터는 안반덕, 귀네미, 새비재, 매봉산 4개 지역에 설치된 CCTV 데이터이다. <그림 19>와 같이 지도에 각 데이터 생성 지역을 도식하고 거리가 가장 가까운 지역의 드론 이미지 촬영지와 CCTV 관측지를 묶어 단수를 지정하였다(표 13). 실제 CCTV 데이터는 각 지역을 대표하는 하나의 지역 데이터만 가지고 있지만 본 연구에서는 인접한 지역을 모두 포함한다고 가정하였다. 농경연의 단수를 적용한 출하량은 <표 14>와 같다.

통계청의 단수를 적용한 출하량(표 16)은 최근 3개년(2017년부터 2019년)의 평균 단수를 사용하였다(표 15).

표 14. 농경연 단수 적용 출하량

지역	2차_출하량(kg)	3차_출하량(kg)	...	9차_출하량(kg)	총 출하량(kg)
고단리_남	224,959.48	220,501.32	...	378,290.49	1,911,009.94
안반데기	0	0	...	2,190,873.55	9,459,157.02
장전리	0	887,333.46	...	117,936.87	1,236,810.39
...	...	...	...	...	...
새비재	0	0	...	338,817.22	577,748.32
총출하량(kg)	412,472.91	2,544,118.22	...	6,051,271.85	22,998,133.49

주: 1) 드론 이미지에서 산출한 재배면적(m²)에 농경연 단수를 적용하기 위해 재배면적을 10a로 환산하여 단위를 조정하고 단수를 곱하였음.

표 15. 통계청 단수 자료

강원도	2017년	2018년	2019년
단수(kg/10a)	4814	4908	4756

주: 1) 통계청 단수자료는 2020년이 아직 산출되기 전이므로 2017~2019년 평균값인 4826kg/10a 적용

2) 통계청 단수 적용 시 드론 이미지 촬영지 모두 동일한 단수를 갖는다고 가정함.

자료: 통계청

표 16. 통계청 단수 적용 출하량

지역	2차_출하량(kg)	3차_출하량(kg)	...	9차_출하량(kg)	총 출하량(kg)
고단리_남	130,518.69	127,932.12	...	219,479.43	1,108,744.17
안반데기	0	0	...	1,271,117.55	5,488,085.09
장전리	0	709,337.63	...	94,279.17	988,710.77
...	...	...	...	...	...
새비재	0	0	...	347,604.57	592,732.44
총출하량(kg)	239,311.65	1,937,809.86	...	4,496,762.89	15,498,567.69

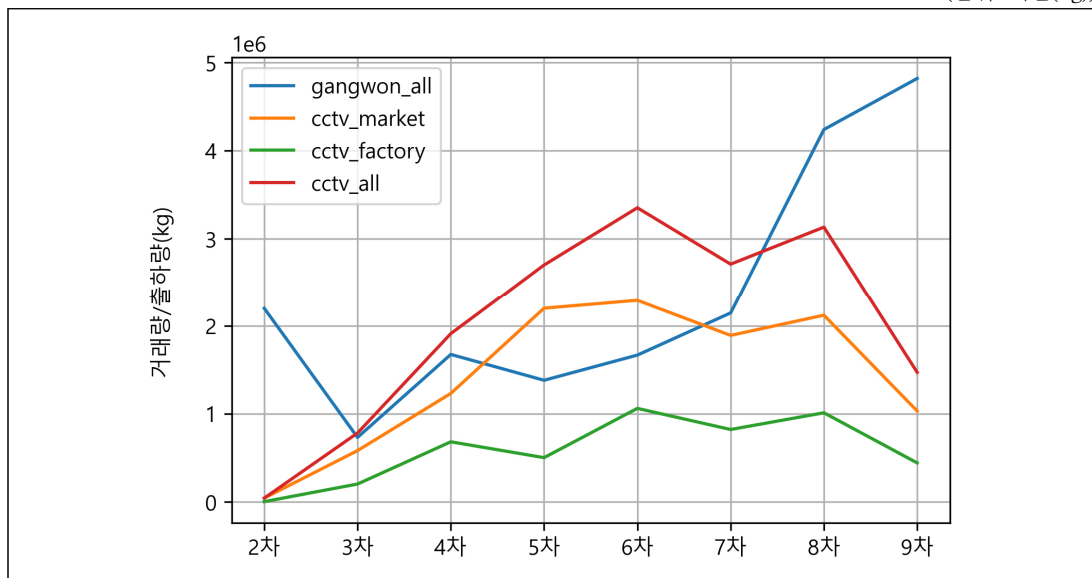
주: 1) 드론 이미지에서 산출한 재배면적(m²)에 통계청 단수를 적용하기 위해 재배면적을 10a로 환산하여 단위를 조정하고 단수를 곱하였음.

7. 농경연 농업관측본부 고랭지 배추 출하용 CCTV 집계 결과 데이터 검토

위 데이터에서 출하대수에 따른 출하 진행률을 활용, 드론 이미지 촬영 주차별 출하량을 산출하여 도매시장 반입량 데이터를 비교하였으나 수치와 변화 패턴이 상이하여 연구에 적용하지 않았다. 하지만 전체 출하 중 시장과 공장에서의 출하량이 비슷한 패턴을 보이며 등락하였고, 시장 출하량은 전체 출하와 0.98의 상관관계를 보였다. 따라서 시장과 공장으로 출하되는 비율에 대한 정보를 얻을 수 있는 데이터로써 본 연구에 활용되었다.

그림 20. 강원도 출하 도매시장 반입량과 CCTV로 관측한 시장, 공장 출하량 비교

(단위: 백만(kg))



- 주: 1) cctv_market: 시장으로 출하된 차량 대수로 산출한 출하량
 2) cctv_factory: 공장으로 출하된 차량 대수로 산출한 출하량
 3) cctv_all: 시장, 공장 합산 출하량
 4) gangwon_all: 강원 출하 전국 반입 도매시장 거래량



제1회

농산물 수급 예측모형 경진대회

우수논문집

Chapter 2
자유과제



장려상

고랭지배추의 무게와 생육환경의 가변적 시차 관계분석

이윤숙

고랭지배추의 무게와 생육환경의 가변적 시차 관계분석

이윤숙*

Key words

고랭지배추(highland chinese cabbage), 임의의 시차변수(arbitrary-time lag), 동적시간 왜곡(dynamic time warping), 가변 시차 인과관계(Variable-lag Granger causality)

ABSTRACT

The study analyzes dynamic time effects between weight of highland chinese cabbage and growth environmental variables and suggests an alternative approach to modelling in order to improve the existing production models. The weight of agricultural products is affected by arbitrary-time lags of growth and weather conditions in a growth period. However, general regression models used in most studies about cabbage production modelling assume fixed and constant time lags. The study applies for dynamic time warping algorithm to find arbitrary-time lags between weight and growth environmental variables. Based on the findings, we compare with well-known Granger causality and variable-lag Granger causality. The growth environmental variables are collected from the public data system operated by Rural Development Administration (wds.agdcm.kr). The daily data start from May 1, 2019 to September 31, 2019 and May 1, 2020 to September 31, 2020. We find the existence of arbitrary-time lags between weight and growth and weather variables. Based on the existence of arbitray-time lags, we find variable-lag Granger causality analysis perform better than Granger causality.

차례

- | | |
|---------------|------------|
| 1. 서론 | 4. 분석 결과 |
| 2. 선행연구 검토 | 5. 요약 및 결론 |
| 3. 분석 방법 및 자료 | |

* 강원대학교 농업자원경제학과 조교수

1. 서론

농업에서도 정보통신기술(Information and Communication Technology, ICT)과 사물인터넷(Internet of Things, IoT)을 융합하여 생산에서부터 소비까지 데이터를 기반으로 한 관리 및 예측 시스템을 구현을 위한 다양한 연구가 진행되고 있다. 정보시스템을 기반으로 축적된 데이터는 농업의 생산성 향상, 기후변화 및 병충해 등의 외부환경에 따른 작황 변화에 대응 및 균등한 품질의 농산물 안정적 공급 등에 활용되어 농업발전에 이바지할 수 있다. 농업에서 데이터의 역할은 첨단기술과의 융합으로 더욱 그 중요성이 강조되고 있다. 최근 시설뿐 아니라 노지 농산물의 생산성과 직접적으로 연관된 작물의 생육 및 환경요인들의 정보도 주기와 형식에 따라 축적되고 있다. 이렇게 축적되고 있는 데이터의 활용이 새로운 농업의 발전을 위해 필요한 시점이다.

본 연구에서는 다양한 실측조사 농산물 중에서 배추를 선택하였다. 배추는 김치의 주요 재료로 연중 소비되기 때문에 안정적 생산과 공급이 매우 중요한 채소이다(이진형 등 2017). 저온성 채소인 배추는 서늘한 기후를 좋아하기 때문에 기온이 높아지는 여름철에는 고랭지 이외의 지역에서는 재배가 어렵다(이승호 등 2014; 김기덕 등 2015). 고랭지배추는 여름철 김치의 주원료로 중요한 위치를 차지하고 있지만, 고온에 약한 특성이 있어 30℃ 이상의 기온과 가뭄이 2주일 이상 지속되면 생육이 떨어지고, 결구¹⁾불량, 석회결핍증, 무름병 등의 질병 발생률도 높아진다.²⁾ 지구온난화로 인한 빈번한 기후변화 문제는 고랭지배추의 불안정한 작황을 야기하고 농가의 소득감소로 이어진다. 농가의 소득감소는 다시 국내 고랭지배추의 재배면적과 생산량 감소에 영향을 미친다. 게다가 저렴한 중국산 김치의 수입 증가도 고랭지배추 생산 감소에 영향을 미치고 있다.

고랭지배추의 불안정한 생산에 대응하고 예측 가능한 생산모델 구축을 통해 영농계획과 출하시기조절 등의 의사결정에 도움을 줄 수 있는 적절한 생산모델이 필요하다. 여러 연구에서

1) 결구란 배춧잎이 여러 겹으로 겹쳐져 둥글게 속이 드는 상태를 말함(출처: <https://www.korea.kr/news/pressReleaseView.do?newsId=155971079>).

2) 출처: 농사로 고랭지배추 주요 재배기술(<http://www.nongsaro.go.kr/portal/ps/psv/psvr/psvre/curationDtl.mo?menuId=PS03352&srchCurationNo=1313>)

생산모형을 제안하고 있지만, 대부분은 기상요인이 배추의 단수에 미치는 영향을 회귀분석을 통해 추정하고 예측하였다(이진희 등 2015; 이승호 등 2018). 이러한 연구는 농산물의 생육 과정에서 영향을 미치는 다양한 요인들 사이에는 시간 차이가 존재하고 이러한 시차의 영향은 매우 가변적이라는 사실은 간과하고 있다. 나명환 등(2017)은 토마토의 최적생육환경을 찾기 위해 생장과 환경요인 사이의 시간 차이는 고려하고 있지만, 고정된 시차를 가정하여 분석하고 있어 빈번한 기후변화에 대응할 수 있는 생산모델로는 한계가 있다. 본 연구에서는 농산물 생육과정에서 발생하는 생육환경요인들의 역동적 시간 차이를 반영할 수 있는 분석 방법을 제안하고자 한다.

2. 선행연구 검토

배추의 생산 및 김치 수요의 변화를 반영하고 예측할 수 있는 수급모형 구축과 연관된 다양한 연구가 진행되어 왔다. 김연중 등(2006)은 기존의 배추 수급모형을 보완하기 위해 계절별 배추의 상호 연관성을 반영하였다. 또한, 계절에 따라 변화는 김치 수입량을 국내 공급량에 영향을 미치는 내생변수로 간주하여 생산부문의 모델을 구체화하고 김치 소비변화에 대응할 수 있는 수급모형을 구축하였다. 그러나 그들의 모형은 국내 배추 생산에 가장 큰 영향을 미치는 기후변화를 직접적으로 반영하지 못하는 한계가 있었다. 박지연 등(2013)은 이러한 단점을 보완하기 위해 생육단계별 기상요인을 연계한 개선된 모델을 제안하였다. 이승호 등(2018)은 기상요인이 지역별 고랭지배추 단위면적당 생산량에 미치는 영향을 생육단계별로 추정하였고, 이를 바탕으로 미래 기후변화(온실가스 배출 변화) 시나리오를 바탕으로 단위면적당 생산량 변화율을 전망하였다. 이러한 연구는 생육단계별 나타나는 차이는 고려하였지만, 생육 및 기상요인이 배추의 생장에 영향을 미치는 가변적인 시간 차이를 고려하지 않았다는 점에서 국내 생산량을 예측하는 데 한계가 있다. 나명환 등(2017)은 토마토 생장에 영향을 주는 요인들은 시차를 갖는 변수로 전환하여 시계열 분석을 수행하였다. 이 연구에서는 시차를 고려하여 분석을 시도하였지만 각 변수에 적용된 지연변수는 1주 전 평균온도, 7주 전 일사량,

5주 전 물흡수량, 7주 전 공급 pH농도 등 고정된 시차를 가정하고 있어 가변적인 시차의 영향을 분석하기에는 한계가 있다.

생육 및 환경의 특성을 중점적으로 반영한 생산모델 구축에 관한 연구도 다양하게 수행되고 있다. 김진희 등(2015)은 배추 성장에 온도의 중요성을 강조하였다. 그들은 온도 구간에 따라 배추의 생육을 정량화하였고, 이를 바탕으로 정식 후 경과기온만으로 수확시기를 추정하는 성장모형을 구축하였다. 김기덕 등(2015) 고랭지배추 생산량 예측모델을 구축하기 위해 비파괴 측정 생육량 변수에 생육도일을 포함한 구중³⁾추정 모델을 도출하였다. 이 구중추정식을 바탕으로 엽생장량, 실측 생장량의 비, 토양수분에 따른 생육속도와 생장단계 및 기간에 따른 상대생장률을 적용하여 고랭지배추의 총생산량을 예측할 수 있는 기반을 구축하였다. 이진형 등(2017)은 김기덕 등(2015)이 제안한 모델을 활용하여 봄배추와 가을배추의 정식시기에 따라 실시간으로 측정되는 생육지표 값과 재배기간 중의 기상요소를 기반으로 배추의 생산량을 추정하였다. 농산물의 생육 및 환경 특성을 반영한 생육모델 구축과 연관된 연구들은 실험용 데이터를 사용하였거나 데이터에 대한 접근이 제한적이어서 일반적인 수급예측 모델으로 활용되기 어려운 점이 있다.

선행연구에서 제안하고 있는 생산량 예측모델은 농산물의 생육단계에서 다양한 생육 및 환경요인들이 임의의 시점으로부터 영향을 받을 수 있다는 점을 간과하였거나, 접근성이 제한적인 데이터를 사용하여 모델을 일반화하기 어려운 점이 있었다. 본 연구에서는 공공데이터를 사용하여, 생육 및 환경요인 사이에서 발생하는 가변적 시차의 영향을 반영할 수 있는 관계를 분석을 수행하였다.

3) 외엽과 외엽이 붙었던 줄기 부위를 제외한 무게임(출처: 농촌진흥청 농사로).

3. 분석 방법 및 자료

3.1. 분석 방법

농산물은 파종 때부터 수확 전까지 임의의 시점(arbitrary-time lags)에서 나타나는 생육 변수와 기상요인으로부터 영향을 받는다. 배추의 생육과정은 정식(아주심기) 이후부터 결구가 발생하기 직전까지의 기간을 생육기, 결구 발생 이후부터 수확 전까지를 결구기라고 한다(김진희 등 2015). 배추가 결구의 형태를 보이기까지는 정식 이후 대략 30일 정도가 소요되고, 결구 형성을 경계로 생육 초기에는 20~22°C에서 생육이 촉진되고, 결구기는 7~8월로 15~16°C에서 생육 촉진된다(안재훈 등 2008; 김진희 등 2015; 농촌진흥청 2007). 배추의 상품성에 핵심적인 영향을 미치는 결구의 형성은 초기생육환경과 밀접한 연관이 있다. 이러한 배추의 성장과 환경요인의 동적인 관계를 설명하기 위해서는 기존의 시계열 분석에서 활용되고 있는 시차 간 반응 경로의 일대일 매칭이 아니라 가변적인 요소를 고려해야 한다.

본 연구에서는 고랭지배추의 실무게(g)와 가변적인 시차의 존재를 확인하기 위해 동적시간 왜곡(Dynamic time warping) 알고리즘이 반영된 유사도 추정(Similarity measure)을 수행하였고, 이 결과를 바탕으로 생육 및 환경변수 사이의 동적인 시차 관계를 분석하기 위해 가변 시차 인과관계(Variable-lag Granger causality) 방법을 활용하였다. 유사도 추정을 통해 배추의 무게와 설명변수 사이에는 가변적 시차에 의한 동적인 관계가 존재함을 시각적으로 제시하였고, 유사도 추정 결과를 바탕으로 전통적인 그랜저 인과관계(Granger causality) 검증과 가변적 시차 인과관계 검정의 비교를 통해서 고랭지배추의 무게와 생육 및 환경변수의 동적인 시차 관계를 분석하였다.

3.1.1. 유사도⁴⁾ 추정(Similarity measure)

시계열 데이터는 일정한 시간 간격으로 배치된 순서를 가지는 데이터를 의미한다. 농산물

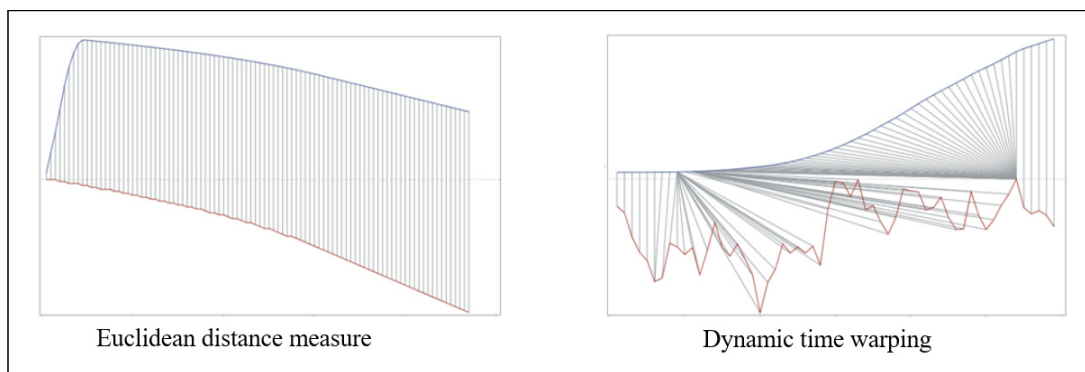
4) 두 데이터 사이의 비슷한 정도를 나타내는 척도로 유사도(similarity)는 n 차 상의 공간에서 두 벡터 사이의 거리를 의미함(출처: Wikipedia).

실측조사 자료는 1주 혹은 1일 간격으로 측정되고 있고, 기상변수는 실시간으로 측정되어 데이터가 조사되는 시간 간격이 다르다. 이러한 실측조사 자료를 통해 두 변수 관계를 분석하기 위해서는 시간 단위를 맞추는 과정을 거치는 것이 일반적이다. 그러나 데이터 변환과정 데이터의 특성이 변형될 수 있다(Batista et al. 2001). 그래서 이러한 생육환경 정보 사이의 가변적인 시차의 영향을 반영하기 위해 동적시간왜곡(Dynamic Time Warping, DTW) 알고리즘을 적용하여 유사도(거리)를 추정해 볼 수 있다.

DTW는 두 시계열 데이터의 길이 변환을 비교하는 알고리즘이다(Coelho 2012). DTW 알고리즘은 음성 인식 분야에서 말하는 속도가 다른 피실험자의 음성과 표준화된 음성의 유사도 측정을 목적으로 개발되었다(이환철 등 2020). 그러나 최근에는 시차 사이에 가변적인 특성이 있는 시계열 데이터를 분석하기 위해 다양한 분야에서 활용하고 있다. 이석준 등(2011)은 선물시장에서 최적의 매매 타이밍을 도출하여 거래 시 적합한 전략을 찾기 위해 DTW 알고리즘을 반영한 트레이딩시뮬레이션 모델을 제안하였다. Franses et al.(2020)와 Wang et al.(2012)은 시차 간의 가변성이 높은 금융시장 정보를 분석하기 위하여 DTW 알고리즘을 시계열 모델에 적용하여 분석하였다. Csillik et al.(2019)는 기후변화에 따른 농산물 지도를 구축하기 위해 경작지 위성 이미지 자료와 기상요인 변수의 관계를 분석하기 위해 DTW 알고리즘을 적용하였다.

본 연구는 DTW 알고리즘이 반영된 유사도(거리) 추정방법을 사용하여 고령지배추의 생육 및 환경변수와 무게 사이에 시차 간의 가변적인 관계가 존재하는 것을 분석하고자 한다. 이석준 등(2011)은 유클리디안 거리 측정(Euclidean distance measure)과 DTW 알고리즘을 비교하여 두 시계열의 반응거리를 분석에 하였다. 그 결과 DTW 알고리즘을 적용한 거리 측정 방법이 유클리디안 거리 측정 방법보다 가변적인 시차 관계를 더 적절하게 분석한다는 것을 보여주었다. <그림 1>은 가변적 시차의 특성을 갖는 실제 시계열 데이터에 유클리디안 거리 측정 방법과 DTW 알고리즘을 적용한 분석한 사례를 보여주고 있다. 두 변수 사이에 가변적 시차의 존재함에도 불구하고 일정한 패턴과 순서에 의존하는 일부 시계열 모델을 사용한 추정값은 잘못된 결과를 도출할 수 있다(Giorgino et al. 2009).

그림 1. 시계열 데이터 시차 관계 비교방법



주: 1) DTW 그래프는 2019년 5월 1일부터 6월 30일 동안의 안반데기 고랭지배추 엽수와 무게 실측 데이터를 활용함.
2) SAS 프로그램을 사용함.

3.1.2. 가변 시차 인과관계(Variable-lag Granger causality)

그랜저 인과관계(Granger causality) 분석은 변수 사이 및 변수의 과거와의 원인과 결과 관계를 분석할 때 일반적으로 사용되고 있는 기본적인 시계열 분석 방법이다(Granger 1969). 그러나 흔히 사용되고 있는 그랜저 인과관계 분석은 일정하게 고정된 시차(fixed-lag)를 기반으로 변수 간의 인과관계를 분석한다(McCrorie et al. 2006). 그러나 이러한 그랜저 인과관계 분석은 농산물의 생육과정 및 자연현상의 가변적인 시간차 관계를 분석하기에 한계가 있다. 본 연구에서는 DTW 알고리즘을 기반으로 임의의 시간 차이를 반영한 가변 시차 인과관계(Variable-lag Granger causality) 분석을 활용하였다. 본 연구에서는 이러한 분석 접근법 제시를 통해 배추 생산량 예측 모델을 고안하기 위해서는 가변적 시차가 반드시 고려되어야 함을 강조하였다.

전형적인 그랜저 인과관계와 가변적 시차 인과관계 검정의 차이를 설명하기 위해 $X(t)$ 는 농산물의 생육 및 환경변수로, $Y(t)$ 는 무게로 가정하였다. 먼저 일반적으로 사용되는 고정 시차 인과관계 회귀식을 살펴보았다.

$$(1) \gamma_Y(t) = Y(t) - \sum_{i=1}^{\delta_{\max}} a_i Y(t-i)$$

$$(2) \gamma_{YX}(t) = Y(t) - \sum_{i=1}^{\delta_{\max}} (a_i Y(t-i) + b_i X(t-i))$$

단, $\delta_{\max}$ 는 최대 시차, r_Y 는 Y의 회귀식의 잔차, γ_{YX} 는 X와 Y의 회귀식의 잔차, a_i 와 b_i 는 i 번째 시차에서 잔차를 최소화하는 상수를 의미한다. (1)과 (2) 식으로부터 γ_{YX} 의 분산값이 r_Y 의 분산값 보다 작으면 X와 Y는 그랜저 인과관계가 존재한다고 표현할 수 있다.

X(t)와 Y(t) 사이의 가변적 시차가 인과관계를 분석하기 위해 아래와 같은 회귀식을 적용하였다.

$$(3) \gamma_{YX}^*(t) = Y(t) - \sum_{i=1}^{\delta_{\max}} (a_i Y(t-i) + b_i X(t-i) + c_i X^*(t-i)),$$

$$X^*(t-i) = X(t-i+1 - \Delta_{t-i+1})$$

단, $\Delta_t > 0$.

(1), (2)와 (3) 식으로부터 $\gamma_{YX}^*(t)$ 의 분산값이 r_Y 와 γ_{YX} 각각의 분산값 보다 작으면 X와 Y는 가변적 시차가 인과관계가 존재한다고 표현할 수 있다.

3.2. 분석 자료

본 연구에서는 고랭지배추의 생육과정에 따른 동적 관계 분석하기 위하여 농촌진흥청(Rural Development Administration, RDA)에서 운영하는 배추, 마늘 주산지 생육정보 시스템(wds.agdcn.kr)이 제공하는 배추의 생육 및 기상정보를 활용하였다.

한국농촌경제연구원 농업관측통계시스템(Outlook & Agricultural Statistics Information System, OASIS)에서 제공하는 작물의 실측정보는 지역 및 품목별로 생육정보가 세분화되어 있어 횡단자료는 충분히 제공되고 있었다. 그러나 시간 이동에 따른 생육과정을 분석할 수 있는 시계열자료는 부족하였다. 고랭지배추의 경우 농가에 따라 2주 혹은 3주 정도의 데이터를

제공하고 있어 생육과정에서 시간에 따른 동적 변화를 분석할 수 있을 만큼의 시계열 정보가 충분하지 않았다. 농촌진흥청의 배추, 마늘 주산지 생육정보 시스템에서는 제공하는 생육정보는 엽수(매)와 실무게(g/포기당) 제공하고 있어, 횡단자료 측면에서는 OASIS 자료보다는 부족한 면이 있다. 그러나 생육과정 정보가 1일 기준으로 연속적으로 제공되고 있어 고랭지 배추의 동적인 생육변화를 추정하기에 더 적절하였다.

본 연구에서는 강원도 영월군 매봉산과 강릉시 안반데기 지역의 고랭지배추의 생육자료와 기상자료를 사용하였고, 정식부터 수확기까지를 고려하여 2019년 5월 1일부터 9월 30일, 2020년 5월 1일부터 9월 30일까지 일별 자료를 사용하였다. 고랭지배추의 실무게(g/포기당)는 생산량을 대체하여 사용하였고, 무게에 영향을 미치는 생육변수로는 엽수(매), 기상변수로는 최저기온(°C), 최고기온(°C), 강수량(mm), 일사량(MJ/m²), 일조시간(h)을 사용하였다. 김기덕 등(2015)과 이진형 등(2017)은 온도를 배추의 생산에 중요한 영향을 미치는 기상요인으로 간주하였고, 안재훈 등(2014)과 이인혜 등(2018)은 평균기온, 최저기온, 최고기온, 최대연속건조일수, 최대연속강우일수, 호우일수 등을 고랭지배추의 생장에 중요한 영향을 미치는 기상요인으로 사용하였다. 본 연구에서는 데이터의 접근성을 고려하여 선행연구에서 사용한 유사한 생육 및 환경변수를 선정하였다.

〈표 1〉은 연구에 사용된 지역별 고랭지배추의 생육 및 환경변수의 기초 통계량을 보여주고 있다. 매봉산 고랭지배추 포기당 실무게는 2019년에 비해 2020년 평균 무게가 더 무거웠다. 평균 엽수도 2020년이 더 많았다. 평균 최저기온은 2019년이 더 높았고, 평균 최고기온은 2020년이 더 낮았다. 강수량은 2020년이 더 많았고, 일사량과 일조시간은 2020년보다 더 낮았다. 안반데기의 경우 포기당 실무게는 2019년 평균 무게가 2020년보다 무거웠다. 평균 엽수도 2019년이 더 많았다. 평균 최저기온과 평균 최고기온의 차이는 뚜렷하지 않았다. 평균 강수량은 2020년이 더 많았다. 평균 일사량과 평균 일조시간의 차이는 뚜렷하지 않았다.

표 1. 고령지배추 생육 및 환경변수의 기초 통계량

영월군 매봉산(2019)						
변수명	관측치 수	단위	평균	표준편차	최솟값	최대값
실무게	153	g/포기당	1758.4	901.7	0.9	2432.6
엽수	153	매	45.4	14.5	3.5	56.6
최저기온	153	℃	12.3	4.0	0.0	21.3
최고기온	153	℃	20.0	3.6	9.9	26.8
강수량	153	mm	6.9	17.4	0.0	114.5
일사량	153	MJ/m ²	17.5	8.3	5.5	29.7
일조시간	153	h	5.6	2.9	1.1	9.7
영월군 매봉산(2020)						
실무게	153	g/포기당	2151.9	1160.9	1.1	3024.2
엽수	153	매	52.4	18.1	3.8	65.9
최저기온	153	℃	13.2	3.7	1.1	22.0
최고기온	153	℃	18.6	3.8	7.2	26.4
강수량	153	mm	10.3	20.4	0.0	126.9
일사량	153	MJ/m ²	16.2	8.9	5.4	29.8
일조시간	153	h	4.9	3.2	1.1	9.6
강릉시 안반데기(2019)						
실무게	153	g/포기당	2151.8	1160.8	1.1	3024.1
엽수	153	매	52.4	18.1	3.8	65.9
최저기온	153	℃	13.2	3.7	1.1	22.0
최고기온	153	℃	18.6	3.8	7.2	26.4
강수량	153	mm	10.3	20.4	0.0	126.9
일사량	153	MJ/m ²	16.2	8.9	5.4	29.8
일조시간	153	h	4.9	3.2	1.1	9.6
강릉시 안반데기(2020)						
실무게	153	g/포기당	1719.9	897.8	1.1	2385.6
엽수	153	매	45.6	14.7	3.8	57.1
최저기온	153	℃	13.2	4.0	0.6	20.9
최고기온	153	℃	18.4	4.0	8.5	27.0
강수량	153	mm	11.8	26.2	0.0	150.6
일사량	153	MJ/m ²	16.1	8.9	5.2	30.0
일조시간	153	h	5.0	3.3	1.1	10.1

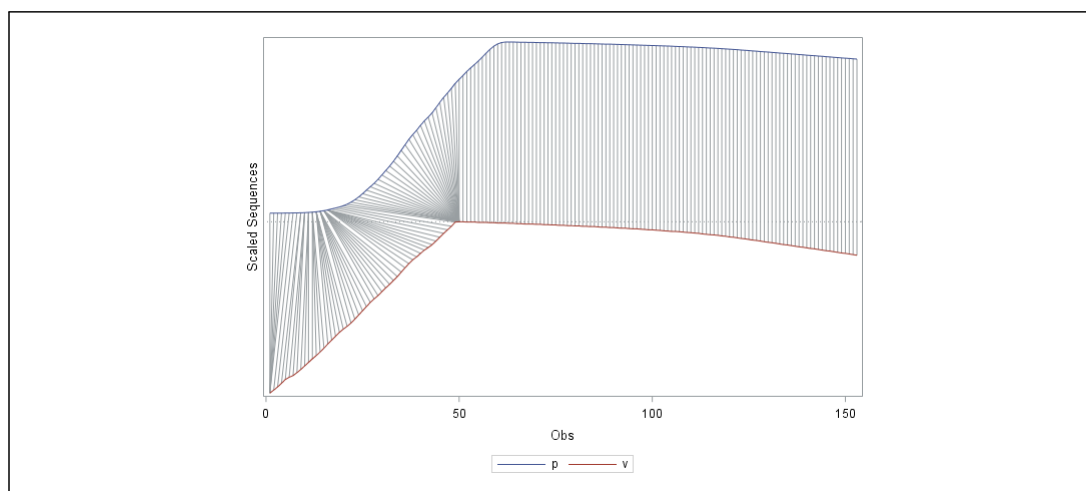
자료: 배추, 마늘 주산지 생육정보시스템(wds.agdcm.kr)

4. 분석 결과

4.1. 유사도 추정 결과

본 연구에서는 고랭지배추 무게(g/포기당)에 영향을 미치는 각각의 설명변수와 짝을 이루어 DTW 유사도 추정이 수행되었다. <그림 2>는 2019년 5월 1일부터 9월 30일 강원 영월군 매봉산 고랭지배추의 무게와 생육환경변수를 분석한 결과이다. <그림 2>는 매봉산 고랭지배추의 무게와 엽수의 DTW 유사도 추정 결과를 보여주고 있다. 생육초기(관측점 0~60일) 동안에 무게와 엽수의 시차 사이에서 나타나는 역동적인 대응 관계를 나타나고 있었다. 60일 이후로 엽수와 무게의 시차 간격은 일대일로 매칭이 되는 것을 볼 수 있었다.

그림 2. DTW를 적용한 엽수와 무게와의 관계(영월군 매봉산)



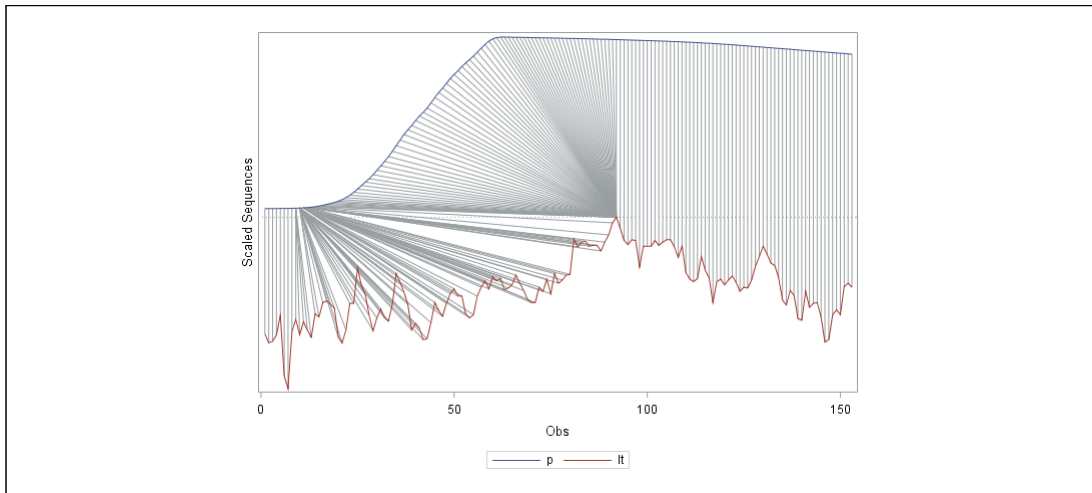
주: 1) p는 무게, v는 엽수를 의미함.

2) 태백시 매봉산 지역에서 2019년 5월 1일부터 9월 30일까지 수집된 데이터임.

자료: 배추, 마늘 주산지 생육정보시스템(wds.agdcm.kr)

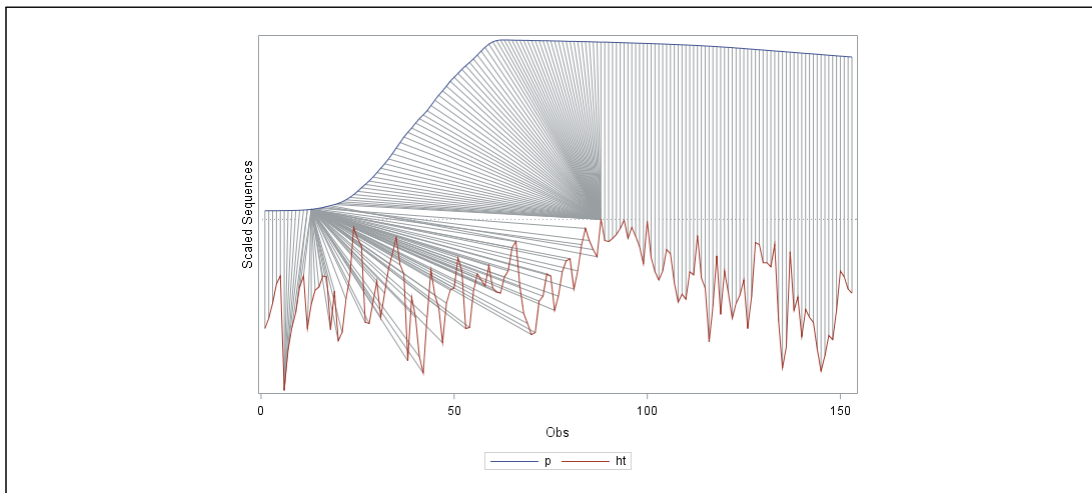
<그림 3>과 <그림 4>는 최저기온, 최고기온과 무게의 사이의 시차 간격이 가변적으로 변화하는 것을 보여주고 있다. 최저기온의 경우 시차의 가변적 특성은 생육초기 뿐 아니라 결구기에도 나타나고 있다.

그림 3. DTW를 적용한 최저기온과 무게와의 관계(영월군 매봉산)



주: p는 무게, lt는 최저기온을 의미함.

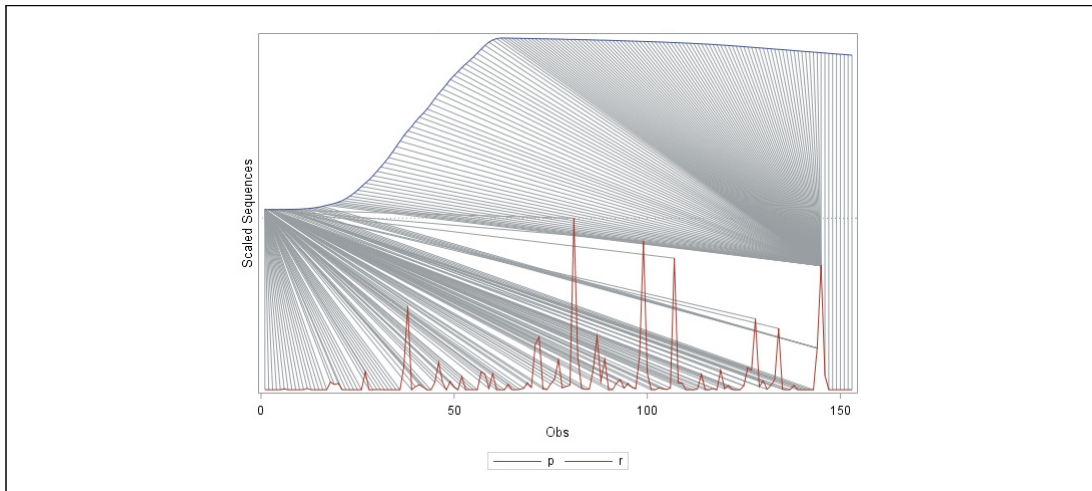
그림 4. DTW를 적용한 최고기온과 무게와의 관계(영월군 매봉산)



주: p는 무게, ht는 최고기온을 의미함.

〈그림 5〉는 고랭지배추의 무게와 강수량 사이에도 가변적 시차의 관계가 나타나고 있음을 보여주고 있다. 강수량의 경우 생육초기, 결구기, 수확 전까지 고랭지배추의 무게에 미치는 시차의 영향은 가변적임을 볼 수 있다.

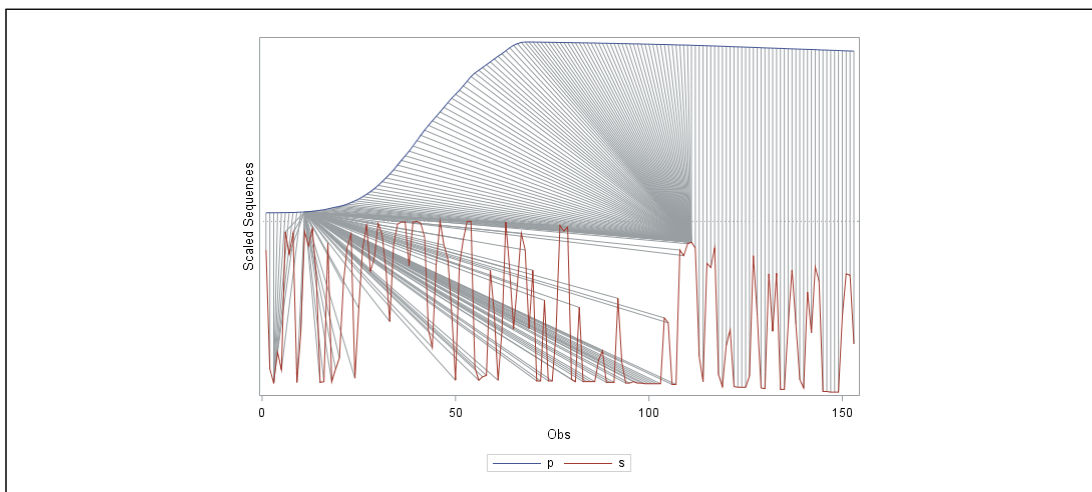
그림 5. DTW를 적용한 강수량과 무게의 관계(영월군 매봉산)



주: p는 무게, r는 강수량을 의미함.

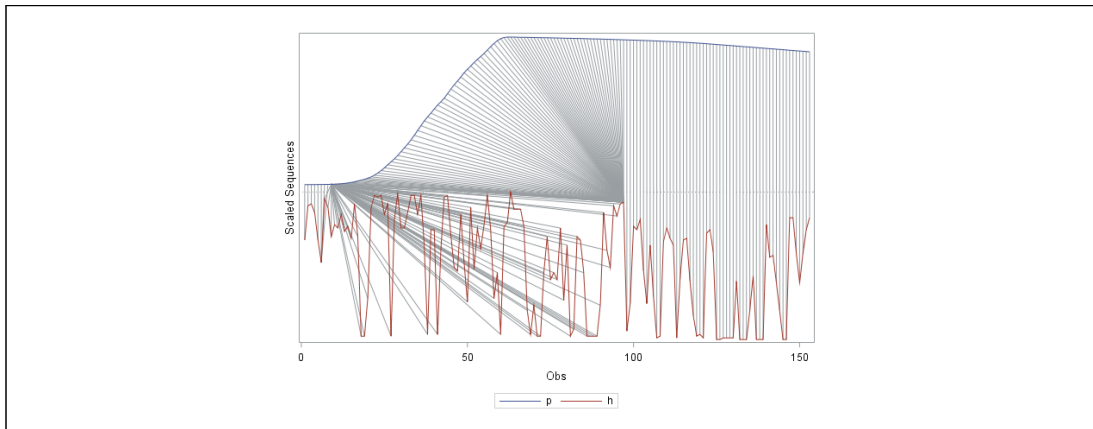
〈그림 6〉과 〈그림 7〉은 일조량, 일조시간과 무게의 가변적 시차 관계를 보여주고 있다. 각각의 기상변수가 고랭지배추 무게에 미치는 시차의 변동성이 다른 기상변수보다 큰 것을 볼 수 있다.

그림 6. DTW를 적용한 일조량과 무게의 관계(영월군 매봉산)



주: p는 무게, s는 일조량을 의미함.

그림 7. DTW를 적용한 일조시간과 무게의 관계(영월군 매봉산)

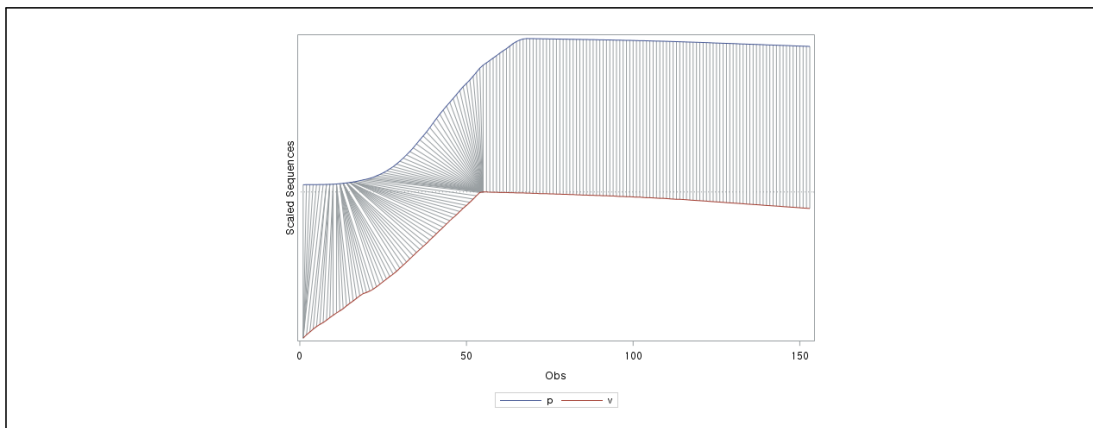


주: p는 무게, h는 일조시간을 의미함.

2020년 5월 1일부터 9월 30일 강원 영월군 매봉산 고랭지배추 자료에서도 무게와 생육 환경변수 사이에 가변적 시차의 특성이 존재하였다. 그래프는 2019년 유사도 결과와 유사하여 생략하였다.

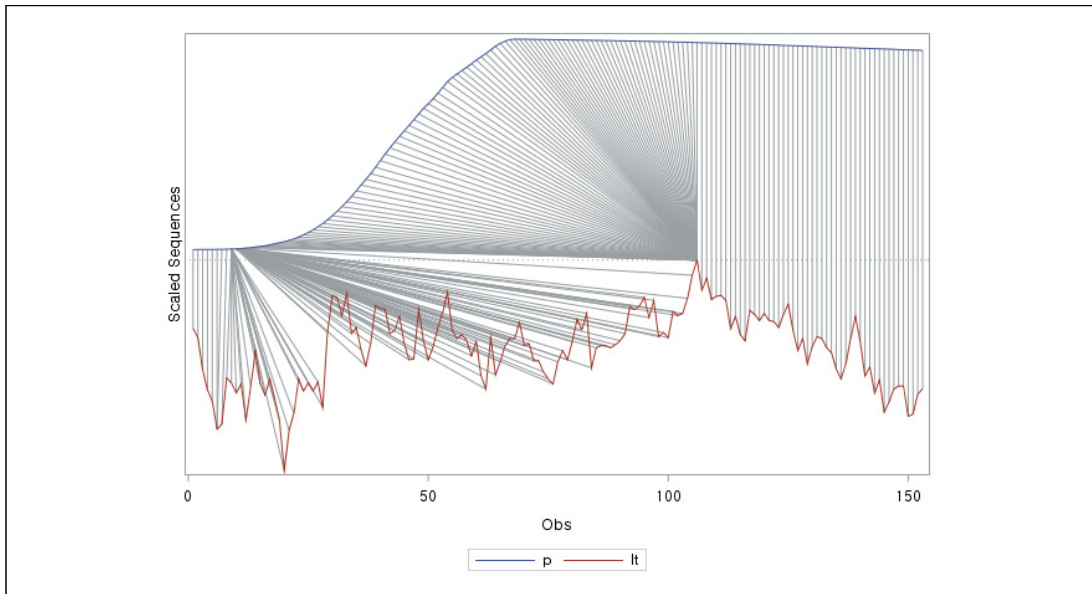
〈그림 8〉은 강릉시 안반데기 지역의 고랭지배추 자료도 사용하여 유사성 추정한 결과를 나타내고 있다. 안반데기 고랭지배추 역시 생육 및 환경변수와 무게 사이에 가변적 시차의 특성이 존재하였다.

그림 8. DTW를 적용한 엽수와 무게의 관계(강릉시 안반데기)



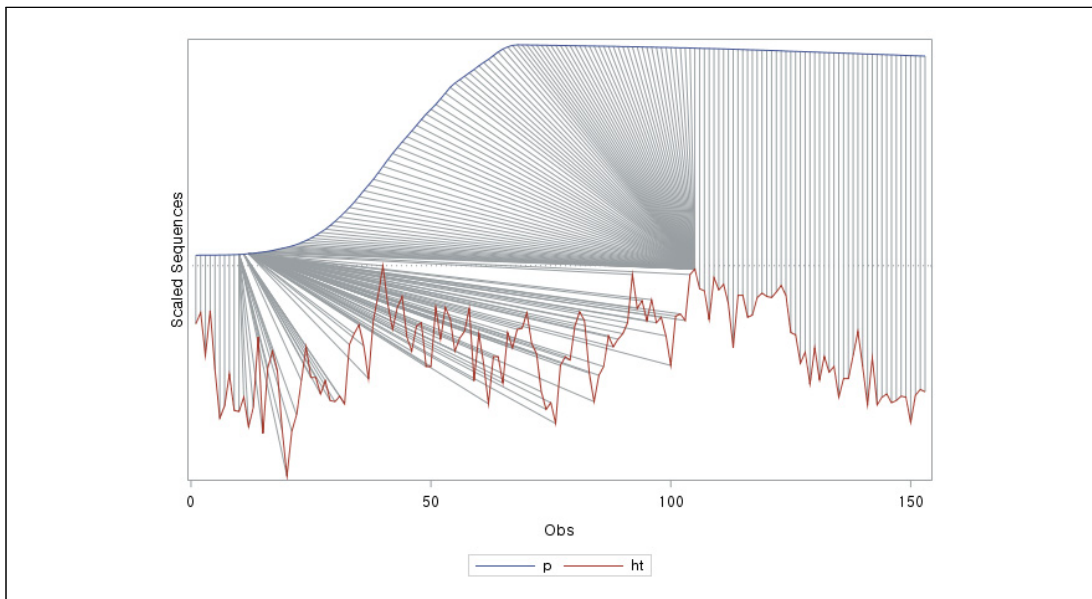
주: p는 무게, v는 엽수를 의미함.

그림 9. DTW를 적용한 최저기온과 무게의 관계(강릉시 안반데기)



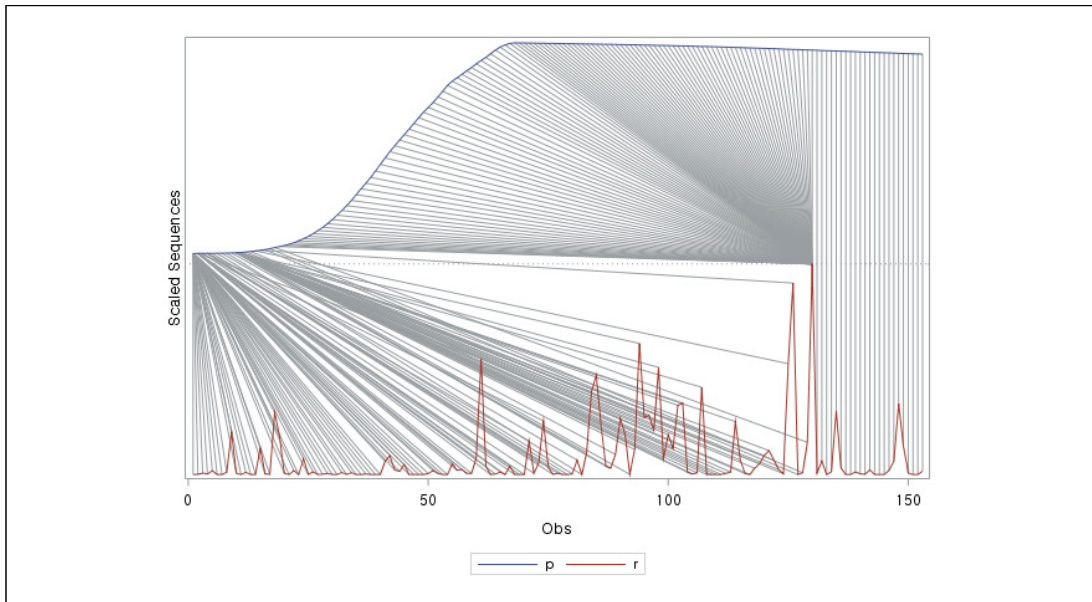
주: p는 무게, lt는 최저기온을 의미함.

그림 10. DTW를 적용한 최고기온과 무게의 관계(강릉시 안반데기)



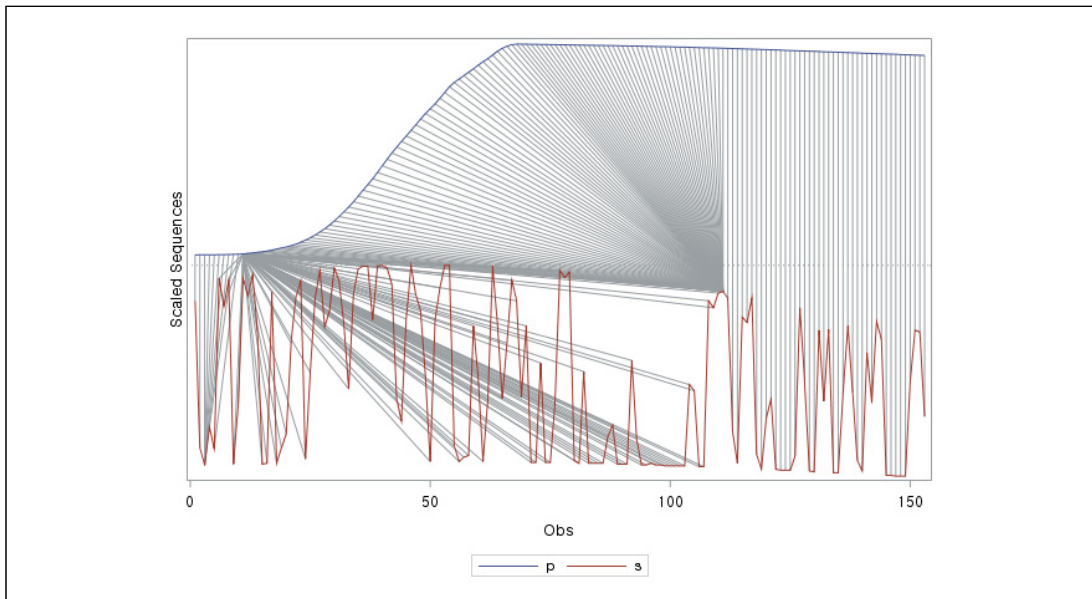
주: p는 무게, ht는 최고기온을 의미함.

그림 11. DTW를 적용한 강수량과 무게의 관계(강릉시 안반데기)



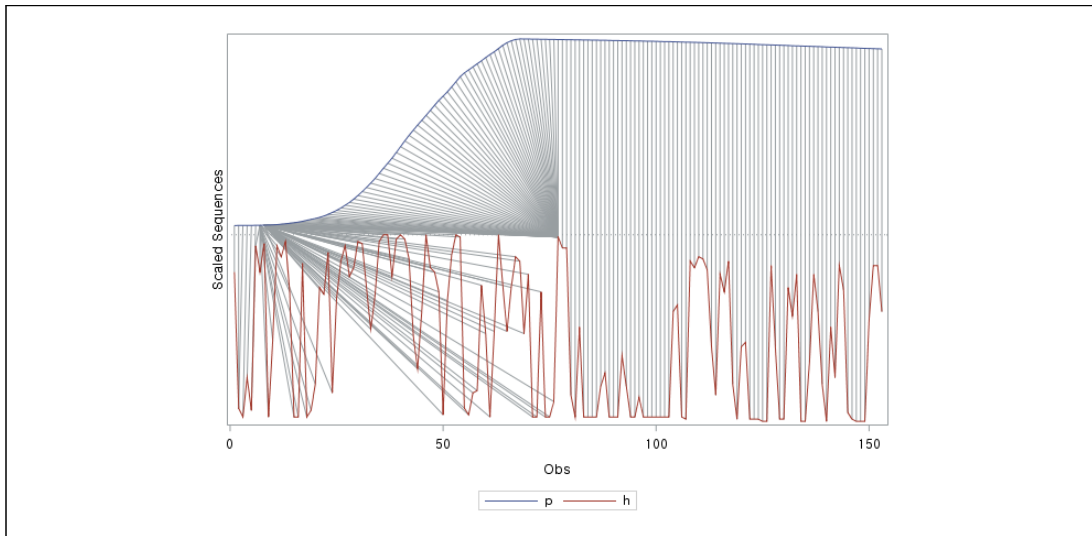
주: p는 무게, r는 강수량을 의미함.

그림 12. DTW를 적용한 일조량과 무게의 관계(강릉시 안반데기)



주: p는 무게, s는 일조량을 의미함.

그림 13. DTW를 적용한 일조시간과 무게의 관계(강릉시 안반데기)



주: p는 무게, h는 일조시간을 의미함.

4.2. 가변 시차 인과관계 결과

DTW 알고리즘이 반영된 유사성 추정 결과를 바탕으로 가변적 시차 인과관계 검정을 수행하였다. <표 2>와 <표 3>은 전형적인 그랜저 인과관계와 가변적 시차 인과관계 비교 결과를 보여주고 있다. False는 두 변수의 인과관계 분석에는 고정된 시차가 반영된 그랜저 인과관계가 더 적절함을 의미하고, True는 두 변수의 인과관계 분석에는 가변적 시차가 반영된 그랜저 인과관계가 더 적절함을 의미한다.

2019년 2020년 매봉산 고랭지배추의 무게와 과거 무게의 인과관계 분석에는 고정 시차 인과관계가 더 적절하였다. 엽수와 무게, 최저기온과 무게, 최고기온과 무게, 강수량과 무게, 일조량과 무게, 일조시간과 무게 사이에서 인과관계 분석에는 가변 시차 인과관계가 더 적절하였다. <표 3>은 2019년 2020년 안반데기 고랭지배추의 그랜저 인과관계와 가변 시차 인과관계 결과를 보여주고 있다. 안반데기 고랭지배추의 무게와 생육환경변수 관계 역시 매봉산 고랭지배추와 유사한 결과를 보여주고 있다.

고정 시차 인과관계와 가변 시차 인과관계 비교 결과는 고랭지배추 무게에 영향을 미치는 생육 및 환경변수의 사이에는 가변적 시차에 의한 인과관계가 존재하였다. 본 연구에서는 간단한 인과관계 분석만을 수행하였지만 보다 정확한 배추의 생산량 추정을 위해서는 시차의 가변성을 고려해야 할 필요성을 보여주고 있다.

표 2. 고정 시차와 가변 시차 인과관계 비교(영월군 매봉산)

		생육 및 기온 변수						
		무게	엽수	최저기온	최고기온	강수량	일조량	일조시간
2019	무게	False	True	True	True	True	True	True
2020	무게	False	True	True	True	True	True	True

주: 1) True는 설명변수와 반응변수 사이에는 variable-lag Granger cause가 적절함 의미함.

2) False는 설명변수와 반응변수 사이에는 fixed-lag Granger cause가 적절함 의미함.

3) R 프로그램을 사용함.

자료: 배추, 마늘 주산지 생육정보시스템(wds.agdcm.kr)

표 3. 고정 시차와 가변 시차 인과관계 비교(강릉시 안반데기)

		생육 및 기온 변수						
		무게	엽수	최저기온	최고기온	강수량	일조량	일조시간
2019	무게	False	True	True	True	True	True	True
2020	무게	False	True	True	True	True	True	True

5. 요약 및 결론

본 연구에서는 고랭지배추 생산성에 직접적인 영향을 미치는 생육 및 기상자료를 활용하여 두 변수 사이에 시차의 영향은 가변적으로 나타남을 분석하였다. 반응변수로는 고랭지배추의 실무게(g/포기당)를 사용하였고, 설명변수로는 엽수(매), 최저기온(°C), 최고기온(°C), 강수량(mm), 일사량(MJ/m²), 일조시간(h)을 사용하였다. 두 변수 사이의 가변적인 시차를 분석하기 위해 DTW 유사도 추정과 가변 시차 인과관계를 적용하였다.

DTW 알고리즘을 기반으로 하는 유사도 추정은 고랭지배추의 무게와 각각의 생육 및 기온

변수 사이에 가변적인 시차가 존재함을 보여주었다. 유사도 추정 결과를 바탕으로 DTW 알고리즘이 적용된 가변 시차 인과관계 검정을 수행하였다. 가변 시차 인과관계 검증에서도 고랭지배추의 무게에 영향을 미치는 생육 및 기온변수 사이에는 가변적인 시차에 의한 인과관계가 존재함을 보여주었다. 본 연구에서는 단순한 분석방법을 활용하여 생육변수들 간의 가변적 시차 관계가 형성되어 있음을 보여주었지만, 이러한 결과는 고정 시차를 가정하는 시계열 모델을 활용한 분석은 잘못된 예측값을 생산할 수 있는 위험이 있음을 시사하고 있다. 기존의 배추 수급모형의 예측력을 높이기 위해서는 배추생육환경을 포함할 수 있고, 생육환경의 가변적인 시간 차이뿐 아니라 가변적 시차의 매칭시점까지도 반영할 수 있는 생산모델의 개선이 필요하다.

참고문헌

- 국립식량과학원. 「고랭지배추 주요 재배기술」.
- 김연중, 김배성, 송성환, 박영구. 2006. 「배추 계절별 수급모형 개발」. 한국농촌경제연구원.
- 김진희, 윤진일. 2015. “온도시간 기반의 배추 생육단계 추정.” 「한국농림기상학회지」 16(4): 333-339.
- 나명환, 박유하, 조완형. 2017. “스마트팜 데이터를 이용한 토마토 최적인자에 관한 연구.” 「한국데이터정보과학회지」 28(6): 1427-1435.
- 농촌진흥청. 2007. 「배추 유기재배 매뉴얼」.
- 박지연, 박영구. 2013. 「배추·무 예측모형 고도화 방안」. 한국농촌경제연구원.
- 안재훈, 김기덕, 이정태. 2014. “고랭지배추의 생장모의.” 「한국농림기상학회지」 16(4): 309-315.
- 이석준, 오경주. 2011. “선물시장의 시스템트레이딩에서 동적시간와핑 알고리즘을 이용한 최적매매 빈도의 탐색 및 거래전략의 개발.” 「한국데이터정보과학회지」 22(2): 255-267.
- 이승호, 허인혜. 2018. “기후가 고랭지배추 생산에 미치는 영향: 강원도를 사례로.” 「대한지리학회지」 53(3): 265-282.
- 이진형, 이희주, 김성겸, 이상규, 이희수, 최장선. 2017. “재배시기, 정식일에 따른 배추의 생육 모델 개발 및 생산량 예측 평가.” 「시설원·예식물공장」 26(4): 235-241.
- Amornbunchornvej, C., E. Zheleva and T. Y. Berger-Wolf. “Variable-lag Granger Causality and Transfer Entropy for Time Series Analysis.” *ACM Trans. Knowl. Discov. Data.* 1(1): 1-30.
- Batista, G. E., D. Gada and V. Puttagunta. 2001. “Distance Measures for Effective Clustering of ARIMA time-series.” Proceedings 2001 IEEE International Conference on Data Mining.
- Csillik, O., M. Belgiu, G. P. Asner and M. Kelly. 2019. “Object-Based Time-Constrained Dynamic Time Warping Classification of Crops Using Sentinel-2.” *Remote Sensing.* 11: 1257-1283.
- Franses, H. P. and T. Wiemann. 2020. “Intertemporal Similarity of Economic Time Series: An Application of Dynamic Time Warping.” *Computational Economics.* 56: 59-75.
- Giorgino, T. 2009. “Computing and Visualizing Dynamic Time Warping Alignment in R: the dtw package.” *Journal of Statistical Software.* 31(7): 1-24.
- Granger, C. W. J. 1969. “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods.” *Econometrica.* 37(3): 424-438.
- McCrorie, J. R. and M. J. Chambers. 2006. “Granger Causality and the Sampling of Economic Processes.” *Journal of Econometrics.* 132: 311-336.
- Wang, G. J., C. Xie, F. Han and B. Sun. 2012. “Similarity Measure and Topology Evolution of Foreign Exchange Markets Using Dynamic Time Warping Method: Evidence from Minimal Spanning Tree.” *Statistical Mechanics and its Applications.* 391(16): 4136-4146.



제1회

농산물 수급 예측모형 경진대회

우수논문집

Chapter 2
자유과제



장려상

돼지고기 공급량 예측 모형 연구

양유경 · 부수빈 · 남주현 · 김솔희

돼지고기 공급량 예측 모형 연구

의사결정나무, 인공신경망, 서포트 벡터 회귀 활용을 중심으로

양유경* 부수빈* 남주현** 김솔희*

Key words

돼지고기 공급량(Pork supply), 기계학습(Machine learning), 예측 모형(Prediction model), 의사결정나무(Decision tree), 인공신경망(Artificial neural network), 서포트 벡터 회귀(Support vector regression)

ABSTRACT

The aim of this study is to develop a prediction model on pork supply in Korea using several machine learning methods. Factors affecting the amount of pork supply were classified into two groups: stock and production. Data collection and feature selection were performed based on this classification. Instead of deriving an optimal set of features for each model, stepwise selection was applied to a multiple linear regression model to unify the subset of features that were to be used in training for each model. As a result, 3 out of 8 features that were classified in the stock group, and 5 out of 18 features that were classified in the production group were selected. A total of 8 features were used to estimate the amount of pork supply using decision tree, artificial neural network(ANN), and support vector regression(SVR) methods. The performance of each optimal model was compared, and the results indicate that the feature selection method used in this study was not significantly effective. However, results showed that the decision tree model using the data set after feature selection showed the better performance than when using the data set before. Also, the model with the best performance was found to be the SVR model to which data before feature selection was applied.

차례

- | | |
|----------|----------|
| 1. 서론 | 3. 분석 결과 |
| 2. 분석 모형 | 4. 결론 |

* 건국대학교 동물자원과학과 소속

** 한남대학교 경영학과 소속

1. 서론

2018년 기준 축산물 생산액은 지난 54년간 꾸준히 증가해왔으며 그중 돼지가 높은 비율을 차지하고 있다. 국산 돼지고기 소비량은 2000년 67만 4,017톤에서 2018년 91만 9,931톤으로 연평균 1.7% 증가하였으며, 같은 기간 수입 돼지고기 소비량은 10만 5,891톤에서 41만 4,539톤으로 연평균 7.9% 증가하였다(통계청, 2020).

소비량과 생산액이 꾸준히 증가함에도 불구하고 돼지의 산지가격은 지난 21년간('99~'19년) 연평균 1.3% 하락하였고, 불안정한 비육돈 두당 수익성은 결과적으로 지나친 공급을 발생시켜 지난 39년간(1981~2019년) 비육돈 농가의 축산소득이 연평균 1.8% 감소하는 결과를 가져왔다(통계청, 2020). 실제로 2020년 5월까지만 해도 2019년 하반기 관측 당시보다 10만두 정도 많이 출하됐다. 한돈팜스의 월간 성적 보고서에 따르면, 11개월 전 번식돈 사료를 지표로 하여 추정한 2020년 한해 돼지 출하두수 또한 당초 전망치를 넘어 1천 800만두 이상이 될 수도 있다는 분석 결과를 통해 공급량 조절을 시행하기도 하였다.

최근에는 해외 수급상황과 코로나19 등의 외부적인 요인들이 국내 돼지가격에 큰 영향을 미치고 있는 것으로 나타났다. 코로나 사태에 영향을 받았던 미국의 육류생산 및 공급체계가 예상보다 빠르게 회복세를 보이고 있는 와중에, 중국의 수요가 이를 뒷받침하지 못하면 국제시장의 혼란은 불가피할 것으로 보인다. 결국, 남는 수요가 한국 시장으로 밀려드는 상황이 전개될 수 있다. 뿐만 아니라, 향후 국제시장이 안정되고 수입량이 증가하게 된다면, 국내 양돈산업 시장은 또다시 불안정해질 위험이 있다.

또한, 앞으로 돼지고기 소비는 가격에 따라 민감하게 반응할 것으로 예상된다고 한다. 돼지고기의 소비자 가격이 일정 수준 이상으로 오르게 되면 소비 방향은 닭고기를 비롯한 대체육으로 전환될 가능성이 크다. 실제로 국내산 삼겹살 소비자 가격이 kg당 2만 5천 원을 넘는 즉시 구매저항으로 이어지는 것이 그간의 추세다. 농촌경제연구원에서 주최한 '농업전망 2020'에서 발표한 소비자 대상 조사 내용에 따르면, 국산 돼지고기 소비를 줄일 경우 대체 육류로 많이 꼽은 것이 닭고기(47.2%)였으며, 수입 돼지고기의 대체 육류로 가장 많이 꼽은 것도 닭고기(40.3%)인 것으로 조사되었다.

한편 연도별 가정 내 돼지고기 구매량 추이를 살펴보면, 구매량이 지속적으로 감소하고 있는 것으로 나타나고 있다. 2014년 4주 평균 구매량이 2.28kg인 것에 반해, 2019년은 1.83kg으로 19.7% 감소했다(한국농촌경제연구원, 2020). 축산물가격과 생산성, 수익성 등은 단기적 수급 여건뿐만 아니라 중장기적인 양돈산업 변화 전망에서 중요한 요소이다. 그렇기 때문에 미래의 양돈산업 변화를 예측하기 위해서는 수요, 공급, 가격의 유기적 관계를 종합적으로 고려하여야 한다.

돼지고기 과잉공급과 가격 불안 상황에 직면해 있으면서도 그동안 돼지고기의 수급구조와 관련한 연구는 돼지 단독으로 진행된 경우가 적으며, 유통구조와 소비 동향을 중심으로 분석하는 연구가 대부분이기에 분명한 한계가 존재한다. 이는 합리적인 수급조절과 출하조절을 위하여 공급량에 대한 변수들을 함께 고려한 돼지 수급모형 구축 필요성을 시사한다.

따라서 본 연구는 돼지고기 총공급량의 변화를 초래하는 원인을 생산량과 재고량의 측면으로 구분하여 분석하고 공급되는 돼지고기의 총량을 예측하는 기계학습 기반의 모델을 설계하고자 한다. 또한, 돼지의 생산과정 및 사육 특성을 고려하여 돼지고기 수급에 장·단기적으로 영향을 미치는 변수 중 생산량과 관련된 변수를 새로 발견함으로써 보다 높은 예측력의 모델을 설계하고자 한다. 이러한 분석 결과는 양돈농가뿐만 아니라 유통업자, 소비자에게 직간접적으로 도움이 될 것으로 예상된다.

1.1. 선행연구 검토

전술한 바와 같이, 그동안 돼지고기의 수급구조에 관한 연구는 다양한 방법론을 사용하여 이루어져 왔지만, 주로 쇠고기 외 기타 축산물과 함께 연구가 되었거나 유통구조와 가격을 중심으로 연구가 진행되었다.

이정희(2001)는 우루과이라운드 협상 이후 육류시장의 유통단계별 인과관계를 확실히 규명하고 유통정책의 효율성을 개선하기 위하여 1995년부터 2000년까지 월별자료와 VAR 모형을 이용한 유통단계별 가격의 인과관계를 분석하였다. 그 결과 돼지고기의 산지가격과 도매가격은 연계성이 높지만, 그에 비하여 소매시장은 연계성이 낮은 것으로 분석하였다.

최정섭(2001)은 육류의 유통단계별 가격 연동성을 개선하고 소비자에게 육류 구매 기준가격을 제공하기 위하여 육류 소비자 권장가격 산정모형을 개발하였다. 육류 소비자 권장가격이 실효성과 공정성을 가지기 위해서는 가격 조사방법의 체계화, 지역별 권장가격의 차별화 그리고 자료제공자와 권장가격 공표가 간의 지속적인 조사 협조 등이 필요하다는 것을 제안하였다.

김혜영 등(2003)은 ‘육류 수요체계의 내생성과 구조변화’ 연구에서 소비자 가격과 분기별 소비량을 이용하여 LA/AIDS(준이상수요체계)와 LA/IAIDS(역준이상수요체계)로 한국의 육류 수요체계를 분석하였다.

심춘수, 정경수, 김민정(2006)은 가격과 관련된 데이터가 시계열 데이터임을 고려하여, 시계열의 비정상성을 고려하는 비대칭적 오차정수모형을 이용하여 돼지고기의 경우에는 산지와 도매 사이 단계에서 가격전달의 비대칭성이 관찰된다는 결과를 제시하였다.

김민정 등(2007)은 ‘돼지고기 일반광고에 대한 가격효과 분석’에서 돼지고기 일반광고가 돼지가격 상승으로 이어지기 때문에 양돈농가의 수익을 향상시키는 것에 기여하는 것으로 추정하였다.

윤병삼(2008)은 2000년부터 2006년까지 농협서울공판장의 일일 지육경락가격을 사용하여 쇠고기 및 돼지고기의 가격변동성을 분석하였다. 그 결과 요일효과와 계절효과가 쇠고기와 돼지고기의 가격변동성에 영향을 미치고 큰 변동 뒤에는 한동안 큰 변동이 이어지고 작은 변동 뒤에는 작은 변동이 이어지는 변동성 집적현상이 있음을 밝히고, 가격위험을 해소하기 위한 수단이 필요함을 제시하였다.

정민국과 김현중(2011)은 주로 공급 측면의 요인으로 인해 돼지고기 가격에 계절성이 나타남을 제시하였다. 돼지는 여름철 더위로 인한 스트레스로 수태율이 하락하기 때문에 10개월 후인 5~6월 출하물량을 감소시켜 산지 및 도매가격을 상승시키며, 겨울철 돼지의 수태율 상승은 다음해 10~11월 출하물량을 증가시켜 연중 산지가격의 최저 수준을 형성한다. 그랜저의 동태적 인과성 분석결과, 돼지산지가격은 돼지 도매가격과 소매가격에 많은 영향을 받으며, 돼지 소매가격은 돼지 산지가격과 쇠고기 소매가격에 많은 영향을 받는다. 공적분 관계의 구조변화에 대한 검정결과 산지와 도매 사이 단계에서 유통가격 간 공적분 관계가 구조변화를 겪은 것으로 나타났다.

지인배, 황운재, 이형우, 한봉희(2015)는 돼지의 경우 캠핑문화가 확산되고 일본의 방사능 오염수가 유출된 상황, HPAI 발생 등이 돼지고기 수요 증가에 영향을 주는 것으로 추정하였다. 육류 수요변화 요인인 캠핑, 마트, 정육점, 할인행사, 육가공품 구매 등이 육류 소비형태에 미친 영향을 조사한 결과 마트 육류 할인 행사가 육류소비에 가장 큰 영향을 미쳤으며, 육가공품 구매, 캠핑, 정육점형 식당 순으로 나타났다.

남국현과 최영찬(2016)은 돼지고기의 부위별 수요함수를 추정하고, 남편의 직업이 돼지고기의 부위별 소비량에 미치는 영향을 살펴보았다. 또한, 구입처에 따른 부위별 소비량 차이를 확인하였으며 재래시장이 마트에서 구입한 경우에 비해 소비량이 더 많은 것으로 분석하였다. 또한, 준이상수요체계 모형을 이용한 돼지고기 부위별 가격탄력성과 지출탄력성을 추정하였으며, 그 결과 삼겹살은 갈비를 제외한 모든 부위와 대체관계를 보였다.

2. 분석 모형

2장의 구성은 다음과 같다. 제1절은 본 연구에서 사용하는 3가지 머신러닝 방법의 이론을 간단하게 설명한다. 제2절은 분석에 사용된 자료에 대해 설명하며, 이어서 제4절은 분석 방법을 설명한다. 이후 이어지는 3장과 4장에서 분석 결과를 정리하여 보여주고 의미를 해석 및 요약하여 결론을 내린다.

2.1. 머신러닝 방법론

본 연구는 돼지고기 공급량 예측 모형을 제시하기 위해 의사결정나무(decision tree), 인공신경망(artificial neural network, ANN), 서포트 벡터 회귀(support vector regression, SVR) 알고리즘을 사용하였으며, 각 머신러닝 방법론을 다음과 같이 정리하였다.

2.1.1. 의사결정나무(Decision Tree)

의사결정나무는 의사결정 규칙을 나무 구조로 나타내어 전체 자료를 여러 소집단으로 분류하거나 예측을 수행하는 방법이다(최종후, 1999). 상위 노드에서 하위 노드로 이동하면서 트리구조를 형성하는데, 노드가 분기되는 모든 단계마다 하나의 변수와 값을 분류의 기준으로 선택한다. 분류변수와 기준값은 해당 기준에 의해 분기되는 노드에서 노드 내 동질성과 노드 간의 이질성이 최대화되도록 선택된다. 의사결정나무는 목표변수가 이산형인 경우의 분류나무와 목표변수가 연속형인 경우의 회귀나무로 구분되며, 회귀나무의 경우에는 분류변수와 분류 기준값의 선택 방법으로 F-통계량의 F-값, 분산의 감소량이 사용된다(김규범, 2019).

의사결정나무는 화이트박스 모델의 대표적인 예로, 분석 과정이 직관적이고 이해하기 쉽다는 장점이 있는 반면에 가지가 나뉘질 때 분류변수 한 개만 고려하기 때문에 예측력이 떨어지는 편이다. 또한, 데이터가 조금만 변화해도 나무의 구성이 달라져 모형 불완전성 문제가 제기될 수 있다(장제훈, 2021).

2.1.2. 인공신경망(Artificial Neural Network, ANN)

인공신경망 모형은 특정 자극이 발생하면 신경세포에서 신경 전달물질에 의한 신경세포 간의 자극 전달과정을 거쳐 학습과 의사결정 등이 이루어지는 인간의 신경세포 연결 구조와 유사하게 만들어졌다. 생물학적인 뇌의 감각 입력에 대한 처리와 출력 원리를 모사하여 인공 뉴런과 이러한 뉴런을 연결시켜주는 매개체로 구성된 모형이다. 뇌가 뉴런이라고 하는 상호 연결된 세포 네트워크를 사용하여 정보를 처리하는 것처럼, 인공신경망은 인공 뉴런(노드) 네트워크를 사용하여 데이터를 분석한다(류상모, 2020).

인공신경망의 구조는 입력층, 은닉층, 출력층 3개의 layer로 구성된다. 입력층은 각 입력변수에 대응되는 노드로 구성되며, 입력노드의 수는 입력변수의 수와 같다. 입력노드에 전달된 입력 신호를 출력노드에 전달되는 출력신호로 변환하는 활성화함수를 통해 출력신호의 범위가 결정된다. 출력신호는 입력층으로부터 전달되는 변수값을 작은 범위의 출력값으로 압축하기 때문에 출력신호에 영향을 미치는 입력층으로부터 전달되는 변수값의 범위가 상대적으로 좁

다. 이때 입력값에 대한 왜곡 또는 오류가 발생할 수 있다. 이런 문제를 해결하기 위해 입력값을 '0' 근처의 작은 범위 안으로 들어오도록 변환하는 표준화 혹은 정규화를 동원할 수 있다.

은닉층은 입력층으로부터 전달되는 변수값들의 활성화함수로 처리하고 이를 출력층 또는 다른 은닉층에 전달하는 기능을 수행한다. 이때, 은닉층이 2개 이상인 경우를 다층 퍼셉트론 모형이라고 한다. 출력층은 목표변수에 대응되는 노드로 구성된 층으로 예측 또는 분류된 결과를 나타낸다(이슬기, 2011).

인공신경망 모형은 블랙박스 프로세스의 예로, 입력을 출력으로 변환하는 메커니즘이 불투명하여 왜 그러한 결과가 나오는지에 대한 연관 관계를 설명하기 어렵다. 또한, 뉴런의 수가 많으면 훈련 데이터를 더욱 긴밀하게 반영하는 모델이 생성되지만 과적합화 가능성의 문제가 가장 큰 단점이다(남영우, 2006). 하지만 은닉층의 존재로 자료의 타입에 상관없이 처리 가능하며 구조적인 변화 또는 외부 충격으로 시장이 급변하는 경우에 복잡한 시장 추세를 유사하게 예측할 수 있다는 점에서 분명한 장점이 있다(김태훈, 2004).

2.1.3. 서포트 벡터 회귀(Support vector regression)

서포트 벡터 머신(support vector machine, SVM)은 훈련 데이터를 고차원의 형상 공간으로 대응시킨 후, 해당 공간에서 최적의 의사결정 영역을 찾는 학습기법이다(Hearst et al., 1998). 이때 최적의 의사결정 영역이라 함은, 데이터를 분류하는 경계인 초평면 중에서 두 클래스 간의 거리를 최대화시키는 최대 마진 초평면을 의미한다. 본 개념은 통계학자인 Vapnik에 의해 선형 분류기로 처음 등장하였다(Vapnik, 1992). 이후 변수 간의 관계가 비선형적인 경우에 원래 데이터를 고차원 공간으로 사상시킴으로써 변수 간의 관계를 선형적으로 나타내는 커널 함수가 소개되면서 다양한 경계면을 만들 수 있게 된다. 일반적인 커널의 예로는 다항식 커널, 가우시안 방사 기저 함수 커널, 라플라스 방사 기저 함수 커널 등이 있다(정성훈, 2020).

SVM은 구조적 위험에 대하여 최소화 원칙을 구현함으로써 과적합 문제에서 자유롭다는 점에서 장점을 갖는다. 또한, SVM은 블록함수를 최소화하는 학습을 진행하여 전역 최적해를 구할 수 있는데, 이런 점으로 인해 인공신경망보다 우수한 성능을 보이는 모형으로 주목받고

있다(Son et al., 2004; Ahn and Lee, 2009). 이에 따라, 최근에는 SVM을 활용한 다양한 연구가 진행되었으며 문서분류, 문자인식, 주가예측 등에서 뛰어난 일반화 성능을 보인 것으로 나타났다.

위와 같이 SVM은 분류문제에 주로 활용되지만, 임의의 실수를 예측할 수 있도록 SVM의 회귀 모형에 ϵ -무감도 손실함수(ϵ -insensitive loss function)를 도입한 서포트 벡터 회귀(support vector regression, SVR)가 등장하였다(Vapnik, 1998). SVR에서는 비선형 회귀 문제를 해결하기 위해 데이터를 고차원의 형상 공간으로 대응되고, 결과값과 연관된 선형 함수를 찾는 학습기법이다(Lu et al., 2009). 즉, SVR도 SVM과 마찬가지로 커널함수를 사용하여 훈련 데이터를 특징공간의 점으로 대응시키고 학습을 수행하지만, SVR은 클래스가 아닌 임의의 실수를 예측한다는 점에서 차별화된다.

2.2. 분석 자료

본 연구에서는 머신러닝 방법을 사용하여 돼지고기의 공급량을 예측하기 위해 통계청, 농협축산물정보센터, 축산물품질평가원, 축산과학원, 농림축산검역본부, 한국육류유통수출협회 등에서 보유하고 있는 돼지고기 생산 및 소비 관련 데이터를 확보하였다. 먼저 종속변수인 돼지고기의 공급량은 재고량, 생산량, 수입량의 합이다. 재고량은 국내산 재고량과 수입산 재고량의 합으로, 한국육류유통수출협회의 월간 육류유통실태조사와 농림축산검역본부를 통해 구득하였다. 생산량은 축산물품질평가원에서 제공하는 도축두수 데이터와 농촌진흥청 축산과학원에서 2004년에 발표한 돼지 도체 수율조사의 부위별 생산율을 통해 추정하였다. 수입량은 농림축산검역본부에서 제공하는 돼지고기 수입 데이터를 사용하였다. 본 연구는 돼지고기 공급량의 변동 요인을 재고량과 생산량의 측면으로 구분하여 변수를 선정하기 위해 관련 변수들의 데이터를 두 개의 그룹으로 나누어 구득하였다.

재고량 관련 속성은 총 8개이며, 변수 구성 및 정의는 <표 1>과 같다. 이용된 데이터는 2006년 5월부터 2014년 3월, 2015년 3월, 2015년 6월, 2015년 9월, 2015년 12월부터 2019년 9월까지의 자료이다. 국내 재고량의 경우 한국육류유통수출협회에서 회원사를 대상

으로 조사한 재고수치를 근거로 전체 국내산 재고량을 추정한 값이며, 수입 재고량의 경우 농림축산검역본부에서 제공한 전국 축산물검역시장 창고에 보관 중인 재고물량 조사 데이터를 사용하였다. 산지가격은 돼지 산지가격 조사 중단으로 2009년 3월부터는 전국도매시장 지육 평균경락가격을 기초로 환산한 수치이다. 도매가격은 탕박가격을 기준으로 한 값이며, 소비자 가격은 삼겹살 가격을 기준으로 산정한 값이다. 돼지고기 대체재는 닭고기로 선정하였다. 이는 농촌경제연구원이 주최한 ‘농업전망 2020’에서 발표된 조사결과를 참고하였다. 소비자 물가지수(costumer price index, CPI)는 소비자가 구입하는 상품과 서비스의 가격변동을 측정하기 위한 자료로서, 소비지출목적에 따라 분류한 지수 중 식료품 물가지수(food price index, FPI) 데이터를 추가하여 정확성을 높이고자 하였다. 변수 GNI(gross national income)는 1인당 국민총소득 데이터로서 명목 GNI를 인구수로 나누어 구한 값을 나타낸다. 재고량 관련 변수 중 가격과 관련된 변수는 수요에 영향을 미치는 변수임을 고려하여 데이터 기간을 조정하였다. 예를 들어 2021년 5월 재고량은 2021년 4월의 공급량에서 4월의 수요량(소비량)을 제외한 값이기 때문에 5월의 재고량을 구할 때에는 4월의 수요에 영향을 미친 변수와 대응시켜야 한다.

표 1. 재고량 관련 변수 구성 및 정의

변수명	단위	변수 설명	자료 출처
공급량	톤	돼지고기 재고량, 생산량, 수입량의 합	-
재고량	톤	국내산 재고량과 수입산 재고량의 합	한국육류유통수출협회 농림축산검역본부
월	월	월	-
산지가격	1,000원/110kg	1개월 전 농가가 받은 수취가격	농협축산물정보센터
도매가격	원/kg, 지육	1개월 전 전국 돼지고기 도매가격	
소비자가격	원/100g, 삼겹살	1개월 전 돼지고기 소비자가격	
대체재가격	원/kg, 지육	1개월 전 닭고기 소비자가격	
CPI	2015=100.0	소비자 물가지수	통계청
FPI	2015=100.0	소비자 식료품 물가지수	
GNI	명목, 만 원	1인당 국민 총소득	

생산량 관련 속성은 총 18개이며, 변수 구성 및 정의는 <표 2>와 같다. 이용된 데이터는 2003년 7월부터 2020년 12월까지의 자료이다. 돼지고기 생산량은 전술한 바와 같이, 축산

물품질평가원에서 제공하는 등급판정 두수 데이터를 성별과 도체중 구간별로 나누고, 각 구간의 평균 도체중에 농촌진흥청 축산과학원에서 2004년에 발표한 돼지 도체수율조사의 부위별 생산율을 곱하여 추정하였다. 사육두수 관련 변수 데이터는 2003년 7월부터 2016년 12월까지 농림어업총조사를 기반으로 한 통계 결과이며, 2017년 1월 이후의 자료는 돼지이력제를 기반으로 한 통계 결과이다. 총사육두수와 월령별 사육두수 변수들은 양돈산업에서 돼지의 사육기간이 일반적으로 180일인 것을 고려하여 데이터의 기간을 조정하였다. 예를 들어, 2021년 5월에 국내에서 사육한 2~4개월령 돼지는 당월에 도축되지 않고 일반적으로 2~4개월 이후에 도축되기 때문에 2021년 5월의 2개월령 돼지 사육두수는 2021년 9월의 공급량과 생산량과 대응시켜야 한다.

표 2. 생산량 관련 변수 구성 및 정의

변수명	단위	변수 설명	자료 출처
공급량	톤	돼지고기 생산량, 재고량, 수입량의 합	-
생산량	톤	(도체중 × (도체중 대비 부위별 생산율/100))의 합	축산물품질평가원, 축산과학원
월	월	월	-
총사육두수	마리	국내에서 사육하는 모든 돼지의 수	통계청
총사육두수.3		3개월 전 국내에서 사육한 모든 돼지의 수	
총사육두수.6		6개월 전 국내에서 사육한 모든 돼지의 수	
두수A.4		4개월 전 국내에서 사육한 2개월 미만 월령 돼지의 수	
두수A.6		6개월 전 국내에서 사육한 2개월 미만 월령 돼지의 수	
두수B.2		2개월 전 국내에서 사육한 2~4개월령 돼지의 수	
두수B.4		4개월 전 국내에서 사육한 2~4개월령 돼지의 수	
두수C		국내에서 사육한 4~6개월령 돼지의 수	
두수C.2		2개월 전 국내에서 사육한 4~6개월령 돼지의 수	
두수D.전체		국내에서 사육하는 6~8개월령 돼지의 수	
두수D.암컷		국내에서 사육하는 6~8개월령 돼지 중 암돼지의 수	
두수D.수컷		국내에서 사육하는 6~8개월령 돼지 중 수돼지의 수	
두수E.전체		국내에서 사육하는 8개월 이상 월령 돼지의 수	
두수E.암컷		국내에서 사육하는 8개월 이상 월령 돼지 중 암돼지의 수	
두수E.수컷		국내에서 사육하는 8개월 이상 월령 돼지 중 수돼지의 수	
방역치료비	원	비육돈 두당 사육비 중 평균 방역치료비	통계청
질병발생	0 또는 1	돼지열병, 아프리카돼지열병, 구제역 중 발생 보고 있음 = 1, 발생 보고 없음 = 0	농림축산검역본부

표 3. 전체 데이터 기초통계량

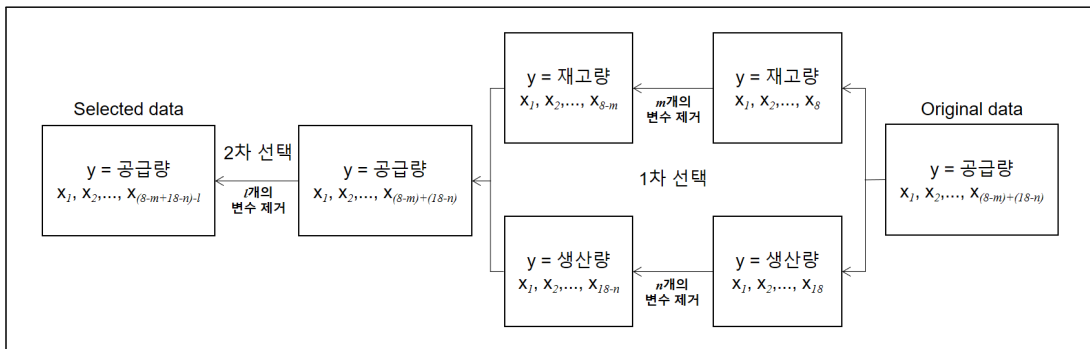
그룹	변수명	관측수	평균값	표준편차	최솟값	최댓값
-	공급량	144	189788.93	62850.76	99803.17	379520.33
-	재고량	144	64336.41	36631.68	18473.9	170730
재고량	월	144	6.45	3.43	1	12
	산지가격	144	341.75	77.71	185	581
	도매가격	144	4163.19	892.18	2538	7165
	소비자가격	144	1782.55	248.54	1292	2460
	대체재가격	144	5155.57	823.89	3302	6983
	CPI	144	94.50	7.85	80.013	105.65
	FPI	144	92	13.01	68.772	115.2
	GNI	144	2927.76	534.37	2069.984	3743.5
-	생산량	210	68784.33	13592.45	40684.28	100450.28
생산량	월	210	6.59	3.45	1	12
	총사육두수	210	9843409.97	1012073.7	7036116	11712607
	총사육두수.3	210	9814453.16	1005309.92	7036116	11712607
	총사육두수.6	210	9781053.31	992005.06	7036116	11712607
	두수A.4	210	3028770.32	317092.08	2149439	3660284
	두수A.6	210	3021473.67	3022738.45	2149439	3660284
	두수B.2	210	3045382.74	285300.37	2079354	3598490
	두수B.4	210	3040849.96	283332.43	2079354	3598490
	두수C	210	2740878.50	378759.76	2003077	3374192
	두수C.2	210	2731527.80	378014.97	2003077	3374192
	두수D.전체	210	103650.17	19096.62	70769	144370
	두수D.암컷	210	93766.53	16501.98	65885	127731
	두수D.수컷	210	9883.63	5287.35	2890	24950
	두수E.전체	210	908447.64	56037.64	688023	980577
	두수E.암컷	210	887688.74	54678.73	672289	961839
	두수E.수컷	210	20758.90	4732.90	15698	34355
	방역치료비	210	8148.77	2018.27	4217	10984
	질병발생	210	0.24	0.43	0	1

2.3. 분석 방법

본 연구에서는 의사결정나무, 인공신경망, 서포트 벡터 회귀 방법 각각에 대한 최적의 돼지고기의 공급량 예측모형을 구축하고, 각 최적 모형들의 성능을 비교한다.

본 연구에서는 모형별로 최적의 속성 집합을 도출하는 대신에, 다중 선행 회귀모델에 단계별 선택(stepwise selection)을 적용하여 각 모형에 학습시킬 변수들의 조합을 하나로 통일한다. 이는 변수들의 조합을 다르게 하면서 각 조합을 학습시켜야 하는 래퍼 기법(wrapper method)의 특성상 시간이 오래 걸리고 과적합(overfitting)의 위험성이 증가한다는 단점을 보완하기 위함이다. 예측하고자 하는 돼지고기의 공급량이 생산량, 재고량, 수입량의 합인 것을 고려하여, 생산량과 재고량을 종속변수로 하여 각각의 속성 집합을 1차적으로 구한다. 이후 공급량을 종속변수로 하고 두 속성 집합에 속한 속성들과의 상관관계를 2차적으로 분석하여 의사결정나무, 인공신경망, 서포트 벡터 회귀 모형에 적용할 변수 조합을 최종적으로 선택한다. 변수 조합을 선택하는 기준은 AIC(Akaike Information Criteria) 값으로 한다.

그림 1. 변수 선택 과정



머신러닝에 사용되는 변수들의 단위가 상이할 경우 학습되는 모형의 성능이 저하될 수 있으므로, 수치형 변수를 0과 1 사이의 값으로 정규화(normalization)하고 범주형 변수와 병합한다. 또한, 과적합 현상을 방지하기 위해 데이터는 7:3의 비율로 훈련 데이터(train data)와 검증 데이터(validation data)로 나눈다.

이어, 방법별 초모수를 다양하게 변화시켜가며 훈련 데이터를 학습에 사용하고, 검증 데이터에 대한 평균제곱근오차(root mean absolute error, RMSE)와 평균절댓값오차(mean absolute error, MAE)가 가장 작은 최적의 모형을 구축한다. 의사결정나무 모형은 교차 검증 수(xval)와 복잡성 매개지수(cp)를, 서포트 벡터 회귀 모형은 C(cost)와 입실론(epsilon)을 조정한다. 인공신경망의 경우는 은닉층의 개수를 3개로 고정하고 각 층의 노드 개수를 차례대로 조정한다.

3. 분석 결과

3장의 구성은 다음과 같다. 제1절은 속성 선택 결과를 정리하여 보여준다. 제2절은 알고리즘별 분석 결과를 정리한 것으로, 의사결정나무, 인공신경망, 서포트 벡터 회귀 순으로 설명한다. 이후 이어지는 4장에서 본 장에 정리된 내용에 대한 해석을 하며, 이를 요약하여 결론을 내린다.

3.1. 속성 선택

〈표 4〉와 〈표 5〉는 재고량과 생산량을 종속변수로 하여 각각의 속성 집합을 1차적으로 구한 결과이다. 재고량 그룹에 속한 변수들 중 선택된 변수들은 월, 산지가격, 도매가격, GNI이며 해당 조합의 ACI는 -663.52이다. 생산량 그룹에 속한 변수들 중 선택된 변수들은 월, 총사육두수, 총사육두수.6, 두수A.6, 두수B.4, 두수C, 두수D.전체, 두수E.전체, 두수E.암컷이며 해당 조합의 ACI는 -993.66이다.

〈표 6〉은 1차적으로 선정된 2개의 조합을 합하고 공급량을 종속변수로 하여 상관관계를 2차적으로 분석한 결과이다. 이때, 재고량 그룹의 1차 속성 집합 중 산지가격과 도매가격은 1개월 전의 가격으로 기간을 조정한다. 최종적으로 선정된 변수 조합의 ACI는 -761.81로, 산

지가격, 도매가격, GNI, 총사육두수, 두수A.6, 두수C, 두수E.전체, 두수E.암컷이 포함되었다. 이렇게 최종적으로 선택된 변수 조합 데이터의 구성과 기초통계량은 <표 7>과 <표 8>에 정리되어 있다.

표 4. 재고량 그룹의 1차 속성 선택 결과

변수 조합	ACI
산지가격, 도매가격, 소비자가격, 대체재가격, CPI, FPI, GNI, 월	-657.84
산지가격, 도매가격, 대체재가격, CPI, FPI, GNI, 월	-659.6
산지가격, 도매가격, 대체재가격, CPI, GNI, 월	-661.33
산지가격, 도매가격, 대체재가격, GNI, 월	-663.16
산지가격, 도매가격, GNI, 월	-663.52

표 5. 생산량 그룹의 1차 속성 선택 결과

변수 조합	ACI
총사육두수, 두수C, 두수D.전체, 두수D.암컷, 두수D.수컷, 두수E.전체, 두수E.암컷, 두수E.수컷, 총사육두수.3, 총사육두수.6, 두수A.4, 두수A.6, 두수B.2, 두수B.4, 두수C.2, 방역치료비, 월, 질병발생	-980.07
총사육두수, 두수C, 두수D.전체, 두수D.암컷, 두수D.수컷, 두수E.전체, 두수E.암컷, 두수E.수컷, 총사육두수.3, 총사육두수.6, 두수A.4, 두수A.6, 두수B.4, 두수C.2, 방역치료비, 월, 질병발생	-982.06
총사육두수, 두수C, 두수D.전체, 두수D.암컷, 두수D.수컷, 두수E.전체, 두수E.암컷, 두수E.수컷, 총사육두수.3, 총사육두수.6, 두수A.4, 두수A.6, 두수B.4, 두수C.2, 방역치료비, 월	-984.05
총사육두수, 두수C, 두수D.전체, 두수D.암컷, 두수D.수컷, 두수E.전체, 두수E.암컷, 두수E.수컷, 총사육두수.3, 총사육두수.6, 두수A.6, 두수B.4, 두수C.2, 방역치료비, 월	-985.77
총사육두수, 두수C, 두수D.전체, 두수D.암컷, 두수D.수컷, 두수E.전체, 두수E.암컷, 두수E.수컷, 총사육두수.3, 총사육두수.6, 두수A.6, 두수B.4, 방역치료비, 월	-987.49
총사육두수, 두수C, 두수D.전체, 두수D.암컷, 두수D.수컷, 두수E.전체, 두수E.암컷, 두수E.수컷, 총사육두수.3, 총사육두수.6, 두수A.6, 두수B.4, 월	-989.05
총사육두수, 두수C, 두수D.전체, 두수D.암컷, 두수D.수컷, 두수E.전체, 두수E.암컷, 두수E.수컷, 총사육두수.6, 두수A.6, 두수B.4, 월	-990.53
총사육두수, 두수C, 두수D.전체, 두수D.암컷, 두수D.수컷, 두수E.전체, 두수E.암컷, 총사육두수.6, 두수A.6, 두수B.4, 월	-991.51
총사육두수, 두수C, 두수D.전체, 두수D.수컷, 두수E.전체, 두수E.암컷, 총사육두수.6, 두수A.6, 두수B.4, 월	-992.17
총사육두수, 두수C, 두수D.전체, 두수E.전체, 두수E.암컷, 총사육두수.6, 두수A.6, 두수B.4, 월	-993.66

표 6. 2차 속성 선택 결과

변수 조합	ACI
산지가격.2, 도매가격.2, GNI, 총사육두수, 두수C개월, 두수D.전체, 두수E.전체, 두수E.암컷, 총사육두수.6, 두수A.6, 두수B.4, 월	-752.37
산지가격.2, 도매가격.2, GNI, 총사육두수, 두수C개월, 두수D.전체, 두수E.전체, 두수E.암컷, 총사육두수.6, 두수A.6, 두수B.4	-756.26
산지가격.2, 도매가격.2, GNI, 총사육두수, 두수C개월, 두수D.전체, 두수E.전체, 두수E.암컷, 두수A.6, 두수B.4	-758.25
산지가격.2, 도매가격.2, GNI, 총사육두수, 두수C개월, 두수D.전체, 두수E.전체, 두수E.암컷, 두수A.6	-760.5
산지가격.2, 도매가격.2, GNI, 총사육두수, 두수C개월, 두수E.전체, 두수E.암컷, 두수A.6	-761.81

표 7. 속성 선택 후 전체 데이터 구성

변수명	단위	변수 설명
공급량	톤	돼지고기 생산량, 재고량, 수입량의 합
산지가격.2	1,000원/110kg	2개월 전 농가가 받은 평균 수취가격
도매가격.2	원/kg, 지육	2개월 전 전국 돼지고기 평균 도매가격
GNI	명목, 만 원	1인당 국민 총소득
총사육두수	마리	국내에서 사육하는 모든 돼지의 수
두수A.6		6개월 전 국내에서 사육한 2개월 미만 월령 돼지의 수
두수C		국내에서 사육한 4~6개월령 돼지의 수
두수E.전체		국내에서 사육하는 8개월 이상 월령 돼지의 수
두수E.암컷		국내에서 사육하는 8개월 이상 월령 돼지 중 암돼지의 수

표 8. 속성 선택 후 전체 데이터 기초통계량

변수명	관측수	평균값	표준편차	최솟값	최댓값
공급량	144	189022.49	62395.93	99803.17	379520.33
산지가격.2	144	341.67	77.98	185	581
도매가격.2	144	4163.08	895.30	2538	7165
GNI	144	2932.61	530.85	2069.98	3743.50
총사육두수	144	9892431.41	1063066.39	7036116	11712607
두수A.6	144	3029770.29	319883.32	2149439	3618510
두수C	144	2771251.04	371810.39	2003077	3374192
두수E.전체	144	906049.98	64566.98	688023	980577
두수E.암컷	144	886539.80	63471.40	672289	961839

3.2. 모형별 학습 및 예측

3.2.1. 의사결정나무(Decision Tree)

의사결정나무 방법을 이용하여 최적의 모형을 도출하기 위해 xval(교차 검증 수)를 1부터 10까지, cp(복잡성 매개지수)를 0.001, 0.01, 0.1, 1로 조절하였고 minsplit(노드에 존재해야 하는 최소 관측치)은 (행의 길이)*0.05로 고정하였다. 그 결과로 xval이 10, cp가 0.001인 경우에 가장 오차가 낮은 것으로 나타났다. 속성 선택 전 데이터와 속성 선택 후 데이터를 각각 모형에 적용한 결과, RMSE와 MAE 모두 속성 선택 후 데이터를 적용한 모형이 더 낮은 것으로 나타났다.

표 8. 의사결정나무 모형의 초모수 설정별 성능 비교(속성 선택 후 데이터 적용)

초모수		RMSE	MAE
xval	cp		
1	0.001	0.06570	0.06570
⋮	⋮	⋮	⋮
10	0.001	0.04965	0.03920
10	0.01	0.05418	0.04211
10	0.1	0.06042	0.04464
10	1	0.08333	0.06981

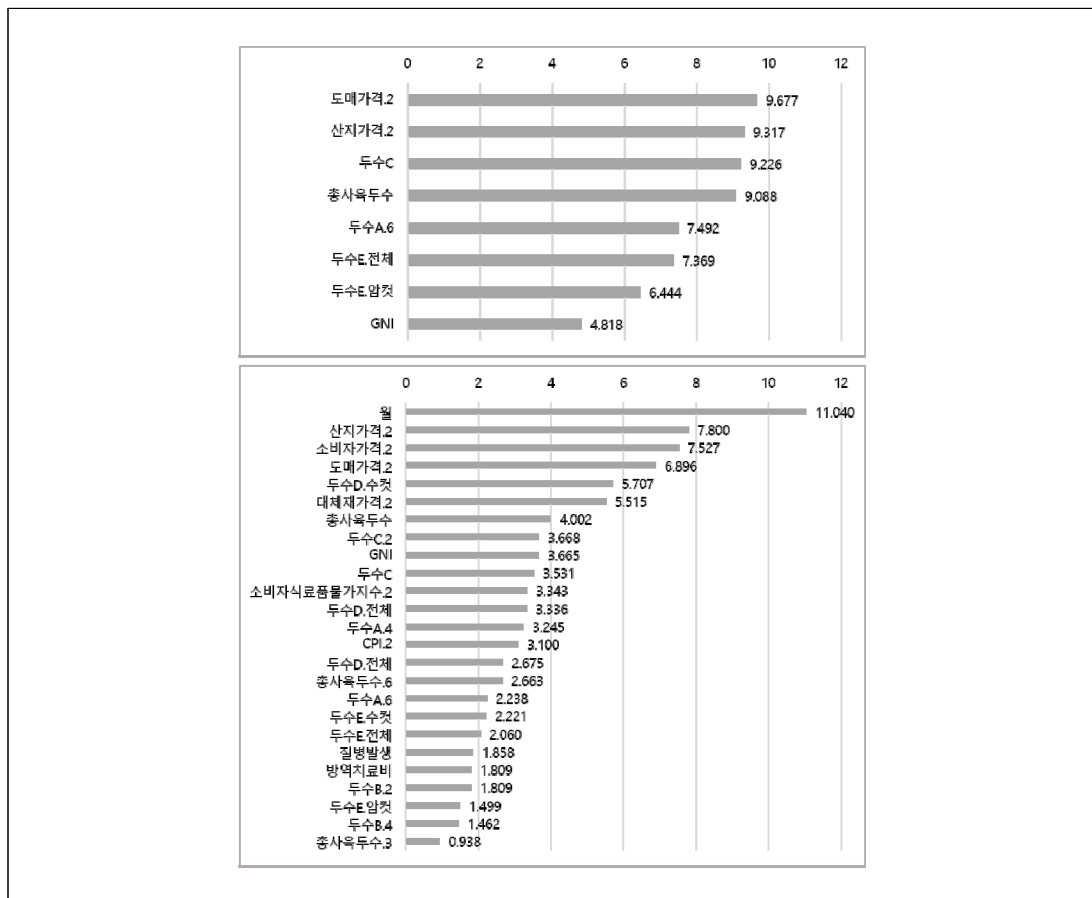
표 9. 의사결정나무 모형의 초모수 설정별 성능 비교(속성 선택 전 데이터 적용)

초모수		RMSE	MAE
xval	cp		
1	0.001	0.05158	0.04080
⋮	⋮	⋮	⋮
10	0.001	0.05093	0.04053
10	0.01	0.06042	0.04464
10	0.1	0.08333	0.06981
10	1	0.18930	0.16258

속성 선택 전 경우의 변수중요도는 월이 가장 큰 중요도를 나타내며 산지가격, 소비자가격, 도매가격, 두수D.수컷, 총사육두수, 두수C.2, GNI, 두수C, FPI, 두수D.전체, 두수A.4, CPI, 두수D.전체, 총사육두수.6, 두수A.6, 두수E.수컷, 두수E.전체, 질병발생, 방역치료비, 두수B.2, 두수E.암컷, 두수B.4, 총사육두수.3 순으로 중요성을 갖는다. 속성 선택 후 경우의 변수 중요도는 도매가격이 가장 큰 중요도를 나타내며 산지가격, 두수C, 총사육두수, 두수A.6, 두수E.전체, 두수E.암컷, GNI 순으로 중요도를 보인다.

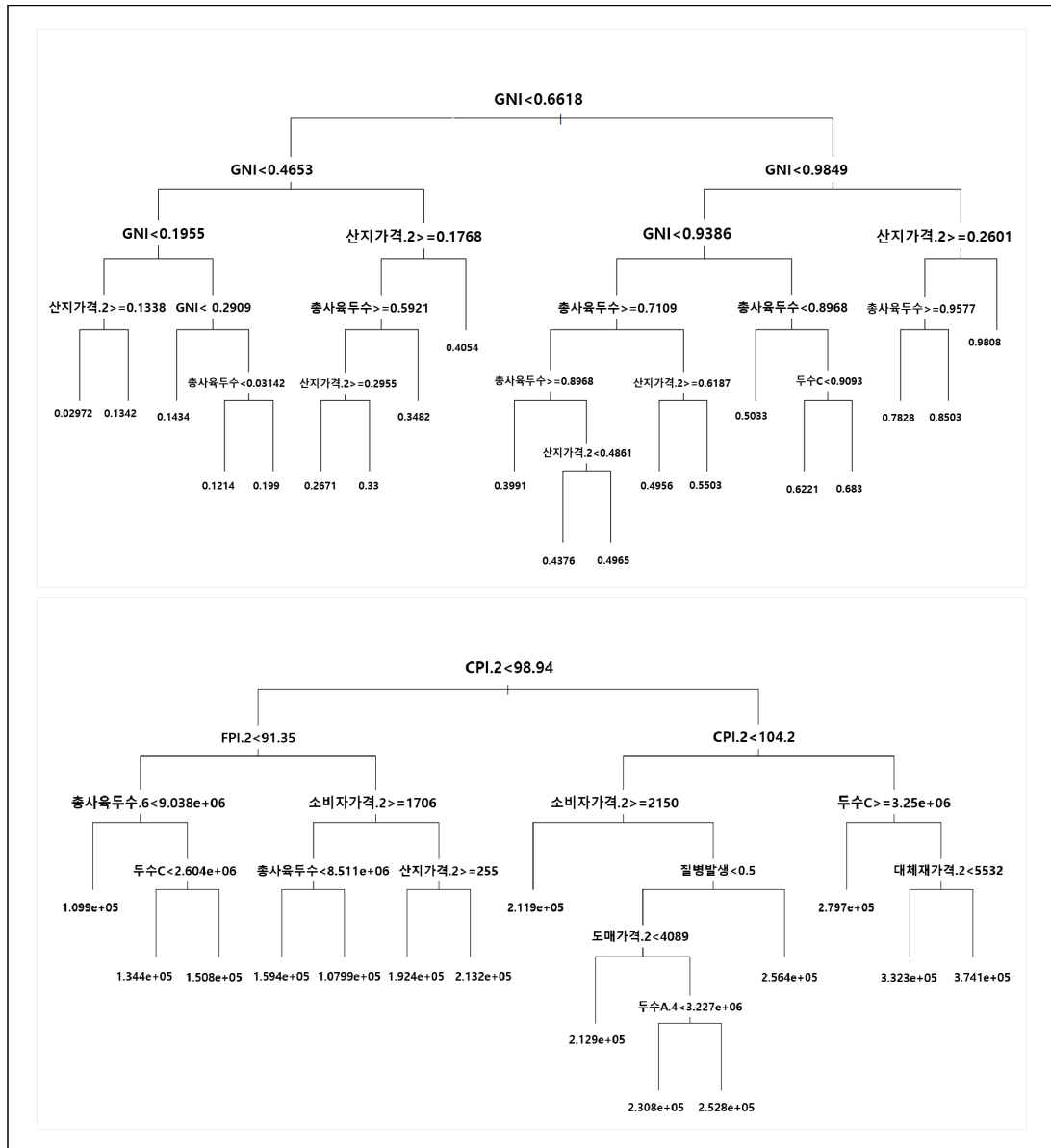
이때 속성 선택 후 변수의 중요도 순위가 달라진 것을 확인할 수 있으며, 이는 의사결정나무 모형이 데이터가 조금만 달라져도 나무의 전체적인 구성이 달라지기 때문이다.

그림 2. 의사결정나무 모형의 변수 중요도



주: 속성 선택 후 데이터 적용(위), 속성 선택 전 데이터 적용(아래)

그림 3. 의사결정나무 모형 시각화



주: 속성 선택 후 데이터 적용(위), 속성 선택 전 데이터 적용(아래)

3.2.2. 인공신경망(Artificial Neural Net, ANN)

인공신경망 방법을 이용하여 최적의 모형을 도출하기 위해 은닉층의 개수를 3개로 고정하고 각 층의 노드 개수를 1개부터 10개까지 차례대로 조절하였다. 또한, stepmax(최대 훈련 단계 수)를 1×10^6 , threshold(다음 뉴런으로 전파되는 역치)를 0.1로 고정하였다. 그 결과로 각 은닉층의 노드 개수가 5, 3, 1개인 경우에 오차가 가장 낮은 것으로 나타났다. 속성 선택 전 데이터와 속성 선택 전 데이터를 각각 모형에 적용한 결과, RMSE와 MAE 모두 속성 선택 전 데이터를 적용한 모형이 더 낮은 것으로 나타났다.

표 10. 인공신경망 모형의 은닉층 노드 개수별 성능 비교(속성 선택 후 데이터 적용)

노드 개수			RMSE	MAE
제 1 은닉층	제 2 은닉층	제 3 은닉층		
1	3	1	0.09084	0.06904
⋮	3	1	⋮	⋮
5	3	1	0.05252	0.04036
⋮	3	1	⋮	⋮
10	3	1	0.08909	0.06552
5	1	1	0.06879	0.05047
5	2	1	0.05789	0.04601
5	3	1	0.05252	0.04036
5	⋮	1	⋮	⋮
5	10	1	0.08240	0.05661
5	3	1	0.05252	0.04036
5	3	⋮	⋮	⋮
5	3	10	0.09275	0.07459

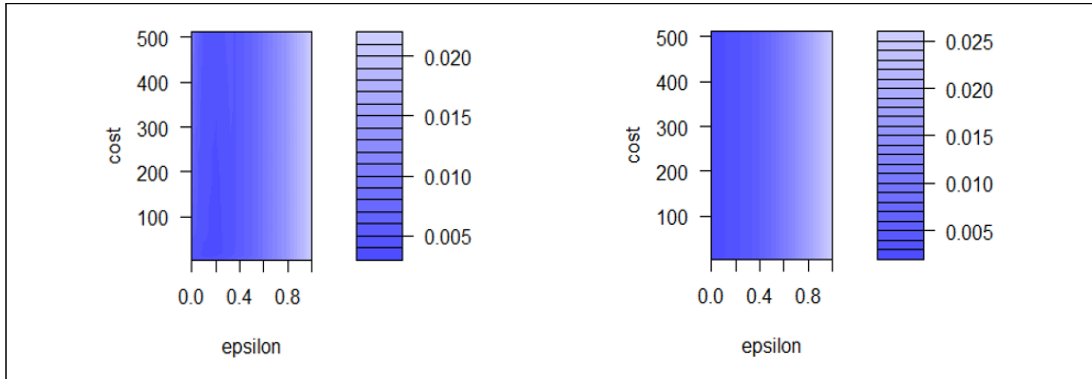
표 10. 인공신경망 모형의 은닉층 노드 개수별 성능 비교(속성 선택 전 데이터 적용)

노드 개수			RMSE	MAE
제 1 은닉층	제 2 은닉층	제 3 은닉층		
1	3	1	0.06185	0.04864
⋮	3	1	⋮	⋮
5	3	1	0.04988	0.04050
⋮	3	1	⋮	⋮
10	3	1	0.07811	0.05502
5	1	1	0.08373	0.06273
5	2	1	0.06766	0.05070
5	3	1	0.04988	0.04050
5	⋮	1	⋮	⋮
5	10	1	0.07870	0.06501
5	3	1	0.04988	0.04050
5	3	⋮	⋮	⋮
5	3	10	0.07107	0.05619

3.2.2. 서포트 벡터 회귀(Support Vector Regression, SVR)

서포트 벡터 회귀 방법을 이용하여 최적의 모형을 도출하기 위해 cost(과적합을 방지하기 위한 모수)를 0부터 500까지, epsilon(모수)을 0부터 1까지 조정하였으며, gamma(선형을 제외한 모든 커널에 요구되는 모수)는 변수의 개수에 따라 설정되는 기본값을 사용하였다. 그 결과로 속성 선택 전 데이터를 적용한 모형에서는 ‘gamma=0.0278, cost=32, epsilon=0’인 경우에 오차가 가장 낮은 것으로 나타났으며, 속성 선택 후 데이터를 적용한 모형에서는 ‘gamma=0.125, cost=32, epsilon=0.2’인 경우에 오차가 가장 낮은 것으로 나타났다. 또한, RMSE와 MAE 모두 후자의 경우가 더 낮은 것으로 나타났다.

그림 4. 서포트 벡터 회귀 모형의 초모수 설정별 성능

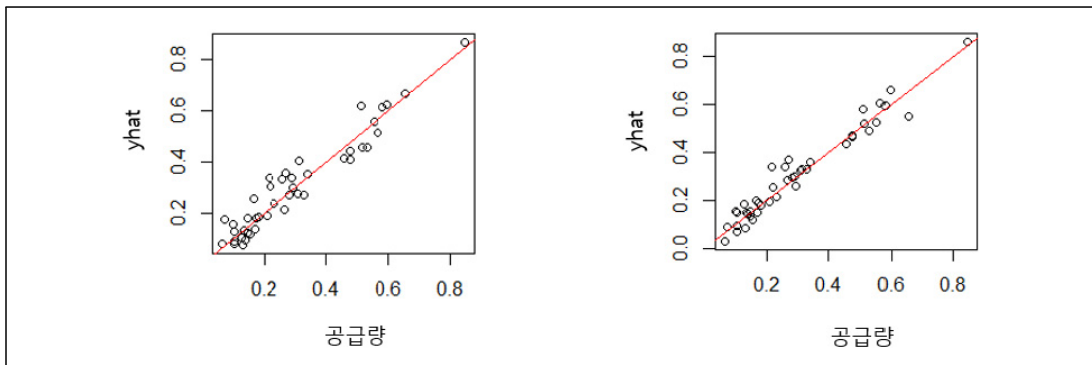


주: 속성 선택 후 데이터 적용(왼), 속성 선택 전 데이터 적용(오)

표 10. 서포트 벡터 회귀 모형의 초모수 설정별 성능 비교

데이터 구분	초모수			RMSE	MAE
	gamma	cost	epsilon		
속성 선택 후 데이터	0.125	32	0.2	0.05305	0.04220
속성 선택 전 데이터	0.0278	32	0	0.04305	0.03221

그림 5. 적용 데이터에 따른 서포트 벡터 회귀 최적 모형의 성능 비교

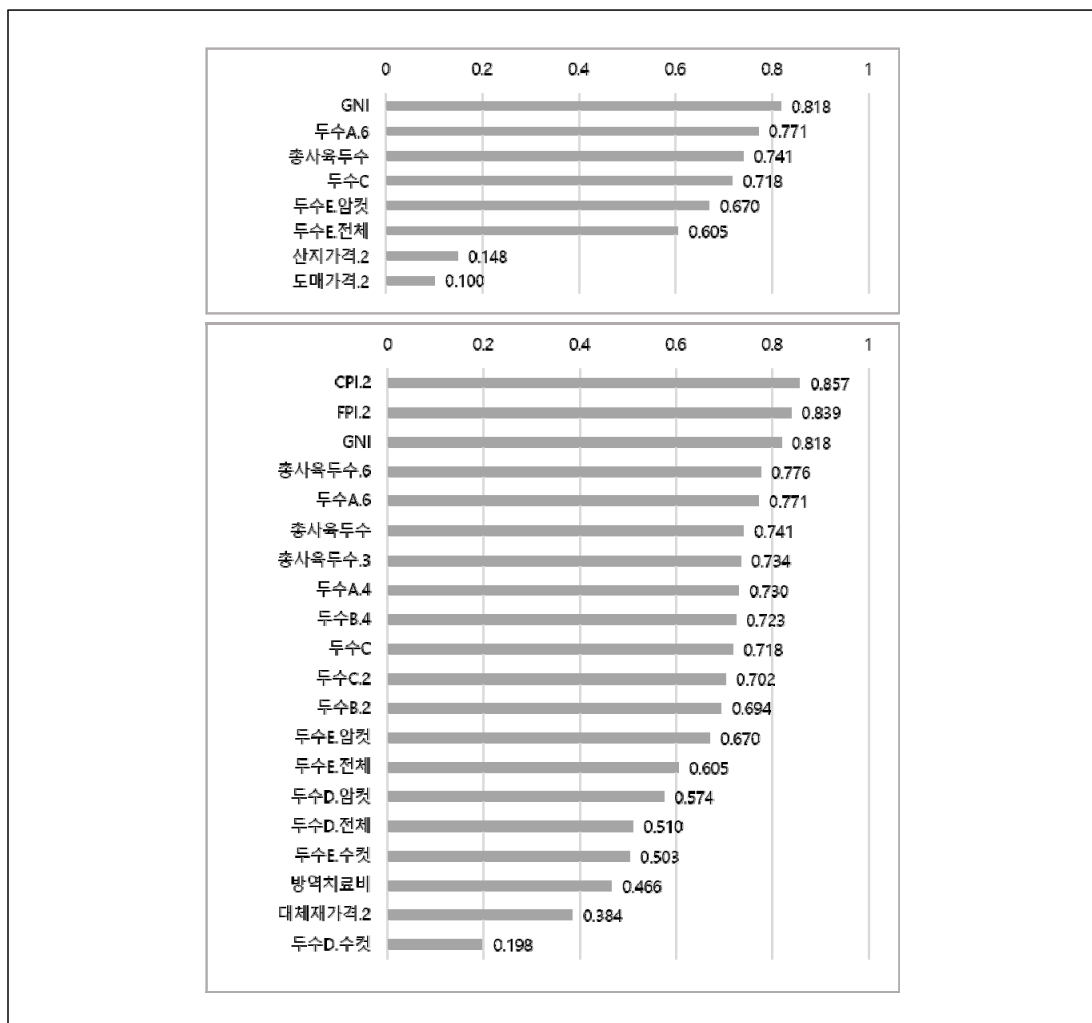


주: 속성 선택 후 데이터 적용, gamma=0.125, cost=32, epsilon=0.2(왼)
속성 선택 전 데이터 적용, gamma=0.0278, cost=32, epsilon=0(오)

속성 선택 전 경우의 변수중요도는 CPI가 가장 큰 중요도를 나타내며 FPI, GNI, 총사육두수.6, 두수A.6, 총사육두수, 총사육두수.3, 두수A.4, 두수B.4, 두수C, 두수C.2, 두수B.2, 두

수E.암컷, 두수E.전체, 두수D.암컷, 두수D.전체, 두수E.수컷, 방역치료비, 대체재가격, 두수D.수컷 순으로 중요성을 갖는다. 속성 선택 후 경우의 변수중요도는 GNI가 가장 큰 중요도를 나타내며 두수A.6, 총사육두수, 두수C, 두수E.암컷, 두수E.전체, 산지가격, 도매가격 순으로 중요도를 보인다. 이때 속성 선택 전과 후 데이터를 각각 적용한 모형에 모두 포함되어 있는 변수들의 순위가 변하지 않았으며, 그 중요도 값 또한 동일한 것을 확인할 수 있다.

그림 6. 서포트 벡터 회귀 모형의 변수 중요도



주: 속성 선택 후 데이터 적용(위), 속성 선택 전 데이터 적용(아래)

4. 결론

본 연구는 다양한 머신러닝 기법을 활용하여 돼지고기 공급량 예측 모형을 구축하고, 이들의 적용 가능성을 검토하였다. 돼지의 생산과정 및 사육 특성을 고려하여 변수를 선정하였으며, 이를 위하여 통계청, 농협축산물정보센터, 축산물품질평가원, 축산과학원, 농림축산검역본부, 한국육류유통수출협회 등에서 보유하고 있는 데이터를 확보하였다.

예측 모형을 구축하기 전에 다중 선형 회귀 모델에 단계별 선택(stepwise selection)을 적용하여 각 모형에 학습시킬 변수들의 조합을 사전에 선정하였다. 예측하고자 하는 돼지고기의 공급량이 생산량, 재고량, 수입량의 합인 것을 고려하여, 생산량과 재고량을 종속변수로 하여 각각의 속성 집합을 1차적으로 구하였다. 그 결과로 재고량 그룹에 속한 변수들 중 ‘월, 산지가격, 도매가격, GNI(국민총소득)’가 선정되었고, 생산량 그룹에 속한 변수들 중 ‘월, 총사육두수, 6개월 전 총사육두수, 6개월 전 2개월 미만 월령 사육두수, 4개월 전 2~4개월령 사육두수, 4~6개월령 사육두수, 6~8개월령 사육두수, 8개월 이상 월령 사육두수, 8개월령 이상 암돼지 사육두수’가 선정되었다. 두 속성 집합에 속한 속성들과 공급량 간의 상관관계를 2차적으로 분석한 결과, ‘산지가격, 도매가격, GNI, 총사육두수, 6개월 전 2개월 미만 월령 사육두수, 4~6개월령 사육두수, 8개월 이상 월령 사육두수, 8개월령 이상 암돼지 사육두수’가 최종적으로 선정되었다. 1차적으로 선택된 재고량과 생산량 속성 집합에 ‘월’ 변수가 모두 포함된 것으로 보아, 재고량은 돼지고기가 소비되는 계절적 패턴이 있고, 생산량은 계절에 따른 질병 발생률, 분만 및 폐사율로부터 기인한 패턴이 있는 것으로 추정된다. 또한, 2차적으로 선택된 속성 집합에 ‘월’ 변수가 제거된 것은 돼지고기 수급 안정화를 위해 수입량을 조절하면서 계절에 따른 공급량의 변동이 둔화된 것으로 판단된다.

속성 선택 후의 데이터를 적용하여 의사결정나무, 인공신경망, 서포트 벡터 회귀 방법 각각의 최적 모형을 구축한 결과, 최적 모형별 성능의 순위는 의사결정나무, 서포트 벡터 회귀, 인공신경망 모형 순으로 나타났다. 의사결정나무 모형의 경우, 속성을 학습 이전에 선택함에 따라 RMSE가 0.05093에서 0.04965로, MAE가 0.04053에서 0.0392로 오차가 낮아져 성능이 개선된 것으로 나타났다. 속성 선택 후 데이터를 적용한 의사결정나무의 RMSE와 MAE가

속성 선택 전 데이터를 적용한 인공신경망 모형보다 낮은 것으로 나타났지만, 서포트 벡터 회귀 모형보다 높은 것으로 나타났다. 따라서 성능이 가장 좋은 모형은 속성 선택 전 데이터를 적용한 서포트 벡터 회귀 모형인 것으로 나타났다. 다만, 속성 선택 이전 데이터를 적용한 경우 세 가지 모형 중에서 성능이 가장 낮았던 의사결정나무 모형이 속성 선택 이후 데이터를 적용한 경우에 성능이 가장 높은 것으로 보아, 속성 선택에 따라 의사결정나무의 성능이 크게 개선되는 것으로 판단할 수 있다.

본 연구의 의의는 다음과 같다. 첫째, 다양한 머신러닝 방법을 사용하여 돼지고기 공급량 예측 모형을 구축하고자 했다는 것이다. 돼지고기 수급구조에 관한 연구는 돼지 단독으로 진행된 경우가 드물고, 돼지고기의 유통구조와 소비 동향을 분석하여 돼지고기의 가격을 예측하는 것을 중심으로 이루어졌다. 반면에, 돼지의 사육 및 도축 과정을 중심으로 분석하여 국내 시장에 공급되는 돼지고기의 양을 산정하는 연구는 찾아보기 힘든 실정이다. 이에 본 연구는 돼지고기 수급 조절에 활용할 수 있는 돼지고기 공급량 예측치를 산정할 방법으로 의사결정나무, 인공신경망, 서포트 벡터 회귀의 적용 가능성을 검토하고자 했다는 점에서 의의를 갖는다. 다양한 머신러닝 방법들을 통해 돼지고기 공급량을 예측하고, 이를 가격 예측에 활용하고 수입량 조절 혹은 입식조절을 하는 등의 다양한 의사결정을 내릴 수 있을 것이다.

둘째, 각 모형별로 최적의 속성 집합을 도출하는 대신에, 다중 선형 회귀 모델에 단계별 선택을 적용하여 각 모형에 학습시킬 변수들의 조합을 학습 이전에 선택했다는 것이다. 속성 선택 전 데이터와 속성 선택 후 데이터를 적용한 모형들의 성능을 비교하여, 직관적으로 이해하기 쉬운 종속변수와 그 외 변수들 간의 선형관계를 기반으로 선정한 변수들을 다른 모형에 적용했을 때의 효과를 검증해볼 수 있다. 이는 변수들의 조합을 다르게 하면서 각 조합을 학습시켜야 하는 래퍼 기법(wrapper method)의 특성상 시간이 오래 걸리고 과적합(overfitting)의 위험성이 증가한다는 단점을 보완하기 위함이다.

본 연구는 의사결정나무, 인공신경망, 서포트 벡터 회귀 알고리즘을 활용하여 기계학습 기반의 수급예측 모형을 구축하고자 하였다. 하지만 최근 MARS, DNN, XGboost 등의 새로운 머신러닝 알고리즘들이 등장하고 있는 것을 감안하면, 이러한 신규 기계학습 방법론들의 적용 가능성을 검토하는 연구가 추가적으로 필요하다. 또한, 본 연구에서 학습 단계 이전에 사용한 속성 선택 방법은 다른 모형의 성능을 개선하는 유의한 효과가 없다. 속성 선택 후 데이

터를 적용한 의사결정나무 모형의 성능이 속성 선택 전 데이터를 적용한 인공신경망 모형보다 높은 것으로 나타났지만, 서포트 벡터 회귀 모형보다 낮은 것으로 나타났다. 따라서 속성 선택에 따른 성능 개선 효과가 모든 모형에게 적용되는 것으로 해석할 수 없다. 또한, 속성 선택 후 데이터를 적용한 모형의 변수 중요도에서 높은 순위를 차지하는 변수들이 속성 선택 과정을 통해 최종 선정된 변수 조합에 포함되지 않은 것으로 보아, 선택된 변수 조합이 각 모형의 성능을 최대화한 조합이 아닐 가능성이 높다. 그러므로 학습 단계 이전에 속성 선택을 위해 사용하는 알고리즘의 종류를 다양하게 검증하고 학습 및 예측에 사용하는 모형에 래퍼 기법을 적용한 변수 선정 과정을 거친 결과와 비교를 하는 후속 연구 또한 필요하다.

참고문헌

- 김규범, 김지옥, 신현정. 2019. “회귀나무(Regression tree)를 이용한 생활용(가정용 및 일반용) 지하수 이용량 추정 방법.” 지질학회지 55(6): 683-691.
- 김태훈, 홍한국. 2004. “회귀 모형과 신경망모형을 이용한 아파트 가격 모형에 관한 연구.” 국토연구: 183-200.
- 김혜영, 김배성. 2003. “육류 수요체계의 내생성과 구조변화.” 『농촌경제』 26(3): 39-53.
- 남국현, 최영찬. 2016. “돼지고기 부위별 수요함수 추정-수도권 소비자 중심으로.” Journal of Agricultural Extension & Community Development Vol.23 No.1. 27-37.
- 남영우, 이정민. 2006. “아파트시장예측을 위한 신경망분석 적응가능성에 대한 연구.” 한국건설관리학회 논문집 7.2: 162-170.
- 대한한돈협회. 2020. 「월간한돈」.
- 류상모. 2020. “인공신경망 채소류 가격·단수 예측 모형 연구.” (사)농식품신유통연구원. 농산물 가격* 단수 예측모형 고도화를 위한 조사용역.
- 심춘수, 정경수, 김민경. 2006. “한국 축산물 가격의 비대칭성.” 『농업경영 정책연구』 33(4): 1113-1126.
- 윤병삼. 2008. “육류시장의 가격변동성, 계절성 및 요일 효과.” 『농업경영 정책연구』 35(1): 21-38.
- 이슬기, 정성관, 이우성, 박경훈. 2011. “인공신경망을 이용한 도시기온 예측모형 구축.” 『국토계획』 46(1): 129-142.
- 이정희. 2001. “육류시장의 유통단계별 연계성 분석.” 『한국유통학회 동계학술대회 발표문』.
- 장재춘. 2021. “머신러닝 기법을 활용한 자영업자 폐업 예측 모형 연구: 서울시 25개 자치구를 중심으로.” (사)아시아문화학술원. 『인문사회21』 12(1).
- 정경수, 박종수, 김민경. 2007. “돼지고기 기초광고에 대한 가격효과 분석.” 『농업경영 정책연구』 34(2): 389-401.
- 정경수, 박종수, 김민경. 2007. “돼지고기 수요에 대한 일반광고 효과.” 『농업경영 정책연구』 34(3): 566-580.
- 정민국, 김현중. 2011. “쇠고기, 돼지고기 수급구조 분석 및 정책 시뮬레이션 모형 개발.” 『한국농촌경제연구원 기본보고서』 1-251.
- 정성훈, 진창하. 2020. “머신 러닝 방법을 이용한 오피스 임대료 산정-랜덤 포레스트, 인공신경망, 서포트 벡터 머신 활용을 중심으로.” 부동산학연구 26.2: 23-53.
- 지인배, 황운재, 이형우, 한봉희. 2015. “한우와 돼지고기 수요변화 요인 분석.” 『한국농촌경제연구원 정책연구보고서』 1-117.
- 최종후, 서두성. 1999. “데이터마이닝 의사결정나무의 응용.” 『통계분석연구』 4(1): 61-83.
- 통계청. 2020. 「통계로 본 축산업 구조 변화」.
- 한국농촌경제연구원. 2020. 「돼지고기 수급 및 유통체계 분석」.

- Ahn, H., and H. Y. Lee. 2009. "A combination model of multiple artificial intelligence techniques based on genetic algorithms for investment decision support aid: An application to KOSPI." *The e-Business Studies* 10.1: 215-236.
- Hearst, Marti A., et al. 1998. "Support vector machines." *IEEE Intelligent Systems and their applications* 13.4: 18-28.
- Lu, Chi-Jie, Tian-Shyug Lee, and Chih-Chou Chiu. 2009. "Financial time series forecasting using independent component analysis and support vector regression." *Decision support systems* 47.2: 115-125.
- Vapnik, Vladimir. 1992. "Principles of risk minimization for learning theory." *Advances in neural information processing systems*.
- Vapnik, V. 1998. "Statistical learning theory new york." NY: Wiley.
- Son, T. S., et al. 2004. "A Study on the Covert Channel Detection in the TCP/IP Header based on the Support Vector Machine." *Journal of The Korea Institute of Information Security and Cryptology* 14.1: 35-45.



제1회

농산물 수급 예측모형 경진대회

우수논문집

Chapter 2
자유과제



장려상

조류인플루엔자의 영향을 고려한 계란가격 예측에 관한 연구

강윤정 · 강소라 · 강수람 · 이하은

조류인플루엔자의 영향을 고려한 계란가격 예측에 관한 연구 온라인 매체의 활용을 중심으로

강윤정* 강소라** 강수람* 이하은***

Key words

계란가격(price of egg), 조류인플루엔자(AI), 외생변수자기회귀누적이동평균모형(ARIMAX), 벡터오차수정모형(VECM), 장단기메모리모형(LSTM)

ABSTRACT

Since it first broke out in 1996 in Korea, avian influenza continues to cause domestic egg market distortion. Especially supply-demand imbalance and massive killing due to quarantine measures are leading to social-economic problems. For this reason, it is needed to study the way of predicting egg price which can reflect effects of avian influenza immediately. This study suggests the model for predicting egg price using unstructured data to reflect latest information. Retail price of eggs between January, 2012 and April, 2021 was analyzed with media articles searched for bird flu or avian influenza as an exogenous variable. In addition, avian influenza quarantine history and fresh egg import and export history were used together to enhance the learning power of the model. For prediction, we used statistics models and deep learning model such as ARIMAX, VECM and LSTM. Also, predictive accuracy for each model is presented and compared. LSTM showed best performance between models which recorded lowest MAPE and RMSE. Furthermore the model was able to predict rapid price rising caused by avian influenza.

차례

- | | |
|------------------|------------|
| 1. 서론 | 4. 분석 결과 |
| 2. 관련 연구 | 5. 요약 및 결론 |
| 3. 자료의 전처리와 분석방법 | |

* 전남대학교 수학/통계학과 석사과정

** 전남대학교 수학/통계학과 석사

*** 전남대학교 통계학과 학사과정

1. 서론

2019년 기준으로 우리나라 농업 생산액의 약 40%를 축산업이 차지하고 있으며, 그 중 계란은 필수적인 식자재로 이용되고 있다. 전염성이 강한 조류인플루엔자의 발생은 농축산업에 부정적인 영향을 미치고 있다. 조류인플루엔자(Avian Influenza, AI)는 닭, 칠면조, 오리 등 조류에서 주로 발생하는 급성 전염병으로 바이러스의 병원성 정도에 따라 저병원성과 고병원성으로 분류된다. 고병원성 조류인플루엔자(Highly Pathogenic Avian Influenza, HPAI)는 위험성이 높은 질병으로 세계동물보건기구에서 이에 대한 발생 사건을 관리하고 있다. 1996년 국내에서 저병원성 조류인플루엔자가 처음 발생하였고 주로 종계의 20%를 폐사하거나 산란율이 감소하는 피해를 야기했다(모인필, 2007). 2003년 HPAI 바이러스는 아시아 전역에 확산되어 국내 최초로 유입되었고, 일부에서는 인체에 감염되어 사망한 사례가 보고되었다. 2006년 HPAI는 육용종계장에서 시작하여 여러 농장에 잇따라 발생하였으나 전파 경로에 대한 정확한 분석이 이루어지지 않았으며, 2009년에는 국내에서 처음 인체에 감염되는 사례가 나타났다. 2014년에서 2015년에는 조류인플루엔자의 발생으로 많은 양의 가금류가 매몰 처리되었고 이로 인한 경제적인 비용이 크게 소요되었다. 특히, 2016년 조류인플루엔자는 대부분 산란계 농장에서 발생하였고 모두 살처분하는 것으로 대응하였으며, 이로 인해 계란의 생산이 급격히 감소하였다.

이와 같이 조류인플루엔자는 국내외적으로 상당한 영향을 미치고 있으나 현재에는 계란 시장에 대한 통계적인 자료와 연구가 많지 않다. 따라서 매년 불안정한 수급과 그에 따른 가격의 급격한 변동을 피할 수 없는 실정이며, 조류인플루엔자의 영향을 고려한 가격 예측 연구의 필요성이 더욱 커지고 있다. 또한, 소비위축의 영향, COVID-19, 예방적 차원의 살처분 등 심리적, 사회적, 정책적인 요인들이 가격에 영향을 미치고 있으며, 이로 인해 소비자 가격을 정확히 예측하는 것에 어려움이 따른다. 종합적으로, 전염성이 있는 가축 질병과 가격에 영향을 미칠 수 있는 요인들을 동시에 고려하여 축산물의 가격을 예측하기 위한 연구들이 필요하다. 기존의 연구방법에서 정형적인 데이터만을 분석하여 빠른 속도로 발전하는 사회를 반영하지 못한다는 한계를 가진다. 정형적인 데이터는 모든 정보를 수집하고 수치화하기까지 시간이 필요하며, 최근 정보를 실시간으로 반영할 수 없다. 최근에는 이러한 한계점을 보완하기 위해

텍스트, 음성, 영상과 같이 비정형 데이터를 정형화하여 활용하는 연구가 진행되고 있다. 특히, 온라인 뉴스 기사는 사회, 경제, 문화, 정치 등 종합적인 이슈의 변화를 빠르게 반영할 수 있다는 점에서 의미 있는 정보를 얻을 수 있다.

지금까지 축산물 전체 또는 일부를 비교 분석한 연구는 많았으나 계란가격만을 집중적으로 분석한 연구는 많지 않았다. 본 연구에서는 조류인플루엔자와 관련한 온라인 뉴스 기사들을 수치화하여 이슈를 실시간으로 반영하고자 온라인 기사 건수와 가격에 영향을 미치는 요인들을 설명변수로 사용하였다. 분석모형으로는 시계열 자료 분석에 주로 이용하는 통계적 모형인 ARIMAX, VECM과 딥러닝 모형인 LSTM을 활용하였다. 결과적으로 조류인플루엔자로 인한 계란가격을 예측함으로써 소비자 물가의 흐름을 파악하고 대응책을 마련하는데 유용한 정보를 제공할 것으로 기대된다. 또한, 농축산물 가격 예측에서 온라인 뉴스 기사와 실측 자료의 결합이 주는 새로운 정보의 의미와 딥러닝 기법의 유용성을 고찰할 수 있다.

2. 관련 연구

먼저, 조류인플루엔자의 위험성에 대해 고찰하기 위해 가축 질병과 농산물 가격의 관계에 관한 연구들을 살펴보았다. 박문수 등(2012)은 축산물 가격 변동의 주요 요인이 가축 질병으로 인해 발생한 것으로 보고, 쇠고기 및 돼지고기와 같은 대체재와 비교하여 닭고기와 계란 시장 중심으로 실증분석을 하였다. 쇠고기는 소비위축과 더불어 돼지 살처분의 영향으로 인해 쇠고기 가격이 하락하였고, 돼지고기는 살처분의 마릿수가 증가함으로써 공급이 감소하여 돼지고기의 가격이 상승한 것으로 분석하였다. 김정현 등(2014)은 사건 연구 모형을 통해 구제역 발생이 국내 돼지고기 가격에 미치는 영향을 분석하였다. 3건의 구제역 발생 사건에 대한 일별 자료를 사용하여 구제역이 국내 돼지고기 가격을 하락시키고 하락 추세가 계속된다는 것을 확인하였다. 또한, 대규모 살처분과 같은 정책적 조치는 급격한 공급량의 감소로 이어져 가격이 재상승하도록 유도하여 J 곡선 현상이 나타날 수 있다는 것을 시사하였다. 서대교 등(2014)은 벡터오차수정모형을 통해 구제역과 조류독감 등 가축 질병이 쇠고기, 닭고기, 돼지

고기 가격에 미치는 장·단기적인 영향을 파악하였다. 충격반응함수를 통해 축산물 가격을 예측하였고 다른 축산물 가격과 비교하여 닭고기 가격은 상대적으로 가축 질병 발생에 대한 장기적인 영향을 받았다.

다음으로, 농산물의 가격을 예측하기 위해 활용하는 여러 가지 알고리즘에 대한 선행 연구들을 고찰하였다. 송대식(2001)은 ARIMA 모형을 통해 육계의 가격 예측 모형을 추정하고 가까운 미래를 예측하였다. 향후 1년 동안 갑작스러운 외부충격이 존재하지 않는다면 육계 가격이 전년 대비 하락할 것으로 전망하였다. 명광식(2005)은 Box-Jenkins 시계열 분석 방법을 통해 미래 육계 가격을 예측하였다. 과거의 가격 변화가 미래에도 반복된다는 가정을 두고 적용한다는 점에서 예상치 못한 변화가 일어났을 때 정확성이 떨어질 것이라는 한계점을 제시하였으나 단기와 중기에는 문제가 없을 것으로 판단하여 다른 농산물에 대한 적용이 가능할 것으로 제안하였다. 강태훈(2007)은 AR-GARCH 모형을 통해 축산물의 가격을 예측하고 비선형적 변동성 구조를 살펴보았다. 돼지고기의 가격은 80년대 초반 이후에 감소하는 양상을 보였으나 쇠고기, 닭고기, 계란의 가격은 시간이 지날수록 불안정한 것으로 분석되었다. 여러 부류의 평균 가격을 사용하고 월별 데이터를 사용하였다는 연구의 한계를 지적하고 각 부류에 따른 분석과 함께 주별 데이터나 일별 데이터를 사용할 것을 권장하였다. 최병옥 등(2007)은 VAR, ARIMA, GARCH 모형을 적용하여 과채류의 품종에 적합한 단기 가격을 예측하는 모형을 연구하였다. 생산 대체 관계가 중요한 경우는 VAR 모형이 효율적이었으며, 가격변동의 패턴을 반영하는 것이 중요한 경우는 ARMA 모형이 효율적임을 보여주었다.

마지막으로, 딥러닝 알고리즘을 적용하여 선행된 연구와 농축산물과 관련한 연구들을 살펴보았다. Gim 등(2015)은 광고효과 분석기법을 활용하여 구제역 등 가축 질병에 대한 뉴스가 돼지고기의 수요에 미치는 영향을 농업경제학 입장에서 분석하였다. 김양석 등(2018)은 조류인플루엔자와 관련된 온라인 뉴스 기사 내용에서 키워드 변화를 분석하여 달걀 가격변동과의 상관관계를 알아보았다. 텍스트 마이닝을 통해 키워드의 빈도를 추출하여 시계열에 따른 변화를 분석하고 다중 선형회귀모형을 통해 상승과 하락에 영향을 미치는 키워드를 파악하였다. 신성호 등(2018)은 LSTM 네트워크를 적용하여 자연재해가 농산물의 가격에 미치는 영향을 고려한 예측 모델을 생성하였다. 기상 데이터와 가격 인상률, 수확량, 재배 면적 등의 농산물 자료를 이용하여 구축한 예측모형이 높은 설명력을 보여주었다.

3. 자료의 전처리와 분석방법

3.1. 자료의 설명과 전처리

본 절에서는 계란의 가격에 영향을 미치는 요인들을 탐색하였다. 고병원성 조류인플루엔자 발생 시 정부는 방역대를 설정하여 이동을 통제하고 발생 농장에서 사육 중인 가금류를 전부 살처분하는 조치를 한다. 이러한 과도한 살처분으로 공급이 줄어들어 가격이 상승하게 된다. 따라서 계란가격을 예측하고자 할 때 조류인플루엔자의 발생을 가장 중요한 요소로 보고, 이에 대한 심각성을 반영하는 자료를 통해 가격의 변동성을 예측할 수 있다. 다음으로, 온라인 뉴스 기사는 실시간으로 이슈를 반영하기 때문에 조류인플루엔자의 심각성이 높아질수록 언론에서 언급하는 횟수가 증가하게 된다. 방역 건수에 대한 문제가 있더라도 언론을 통한 언급만으로 조류인플루엔자의 발생 예측이 가능하다고 볼 수 있다. 마지막으로, 조류인플루엔자가 심각한 수준에 이르면 정부에서는 수입란을 들여오게 되는데 이 과정에서 관세를 낮추거나 폐지하여 일부로나마 수입란을 들여오려 한다. 정부가 시장에 개입함으로써 가격에 영향을 미치기 때문에 계란의 수입·수출 내역도 가격에 영향을 미치는 요인으로서 역할을 한다고 본다.

따라서 모형에 사용한 자료는 계란가격, 조류인플루엔자 확진 건수·기사 건수, 신선란 수입 무게, 신선란 수출 무게 등이며, 목표변수는 계란가격(Price)이고 독립변수는 확진 건수(Count), 기사 건수(Articles), 신선란 수입 무게(Import), 신선란 수출 무게(Export)로 정의하였다. 다음 <표 1>은 데이터의 수집 기간, 기간 단위, 단위, 출처를 보여준다. 계란가격은 서울지역 대형유통소매점의 일별 특란 가격(30개 기준)이고, 신선란 수입·출입 무게는 월별 신선란 수출입 무게를 수집하였다. 확진 건수와 기사 건수는 웹스크래핑을 통하여 수집하였다. 웹스크래핑은 웹상의 정보를 프로그램을 이용하여 수집하거나 추출하는 기법을 나타내고, 우리는 Python 프로그램의 'Selenium'과 'BeautifulSoup' 패키지를 이용하였다. 모든 자료는 수집 단위가 상이하였으며, 다음과 같은 이유로 주별 단위로 데이터를 가공하였다. 월별 단위로 제공되는 데이터의 수가 많지 않았고 월별 기간의 데이터 정보의 양을 가지고 정확한 정보를

제공하기 어려울 것으로 보았다. 실제 시장에서 월별 예측 가격보다 주별 예측 가격이 소비자들에게 더 좋은 정보를 제공할 수 있을 것으로 판단하였다. 계란가격은 주별 평균으로, 신선란 수출입 무게는 각 월의 주 개수로 나누어 주별 평균으로 가공하였다. 확진 건수와 기사 건수는 주별 합계를 계산하였다.

표 1. 수집 데이터 정보

변수명	수집 기간	기간 단위	단위	출처	비고
Price	2001~2021	일별	원/30개	KAMIS	
Count	2003~2021	일별	건	KAHIS	
Articles	2001~2021	일별	건	네이버 (연합뉴스)	“조류인플루엔자” 또는 “조류독감”을 포함하는 기사
Import	2012~2021	월별	kg	TRASS	
Export	2012~2021	월별	kg	TRASS	

주: 1) KAMIS : 한국농수산물유통공사의 농산물 유통정보 시스템

2) KAHIS : 농림축산검역본부의 국가가축방역통합시스템

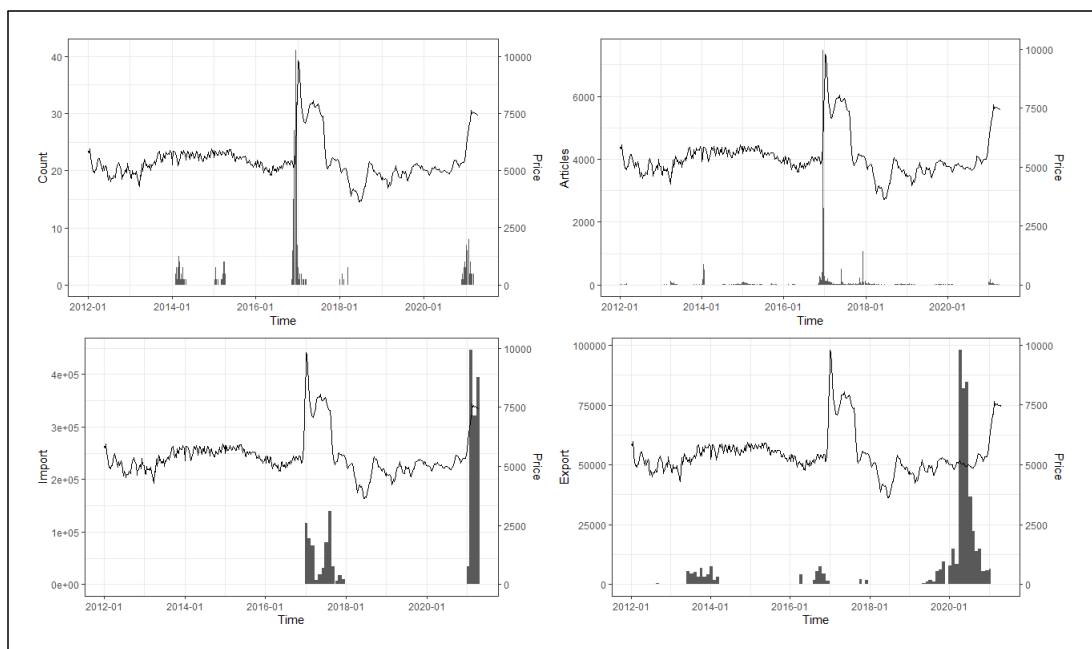
3) TRASS : 한국무역통계진흥원의 한국무역통계정보포털

다음 <표 2>는 계란가격, 확진 건수, 기사 건수, 신선란 수입 무게, 신선란 수출 무게 자료의 기초적인 통계 수치를 보여주고 있다. 또한, <그림 1>은 종속변수인 계란가격을 선그래프로, 각 설명변수를 막대그래프로 나타낸 것이다. <그림 1>의 왼쪽 위는 계란가격과 확진 건수를 함께 시각화한 것으로, 계란가격이 오르기 시작할 때 확진 건수가 확연히 많아지는 것을 확인할 수 있다. 이는 <그림 1>의 오른쪽 위에 있는 기사 건수도 같은 모습을 보인다. 특히 조류인플루엔자 확산에 따른 피해가 심각했던 2017년의 경우 계란가격과 확진 건수, 기사 건수가 모두 최고치를 기록하였다. <그림 1>의 왼쪽 아래는 신선란 수입 무게를, <그림 1>의 오른쪽 아래는 신선란 수출 무게를 기록한 그래프이다. 신선란 수입 무게의 경우, 조류인플루엔자가 확산되었던 2017년과 2021년에 가격 안정을 위해 정부가 수입량을 급격히 늘렸다. 신선란 수출 무게의 경우 2020년부터 급격히 늘었다가 최근 조류인플루엔자 확산에 따라 다시 줄어들었다.

표 2. 주별 데이터의 기초 통계

구분	계란가격	확진 건수	기사 건수	수입 무게	수출 무게
개수	490	490	490	490	490
평균	5447	0.51	44.06	15816	4261
표준편차	894.90	2.88	360.87	64759.92	14267.26
최솟값	3634	0	0	0	0
최댓값	9799	41	7461	446393	97932

그림 1. 계란가격과 설명변수별 시계열 그래프



3.2. 분석방법

계란가격을 예측하기 위해 외생변수를 이용한 자기회귀누적이동평균모형(ARIMAX), 벡터 오차수정모형(VECM)과 같은 통계적 시계열모형과 딥러닝모형 중 하나인 장단기 메모리모형(LSTM)을 사용하였다. 일반선형회귀모형은 설명변수의 영향이 시간에 불변한다는 가정을

하고 있어 설명변수의 영향이 변하는 경우 적절하게 반영하지 못한다. 이런 단점을 극복할 수 있는 방법이 ARIMA이다. ARIMA는 현재 관측된 값이 과거의 규칙에 따라 재현된 것으로 보고, 미래에도 이런 규칙이 유지된다고 가정한다. 이 방법은 모형 설정이 편리하지만, 변수 간의 상호작용을 무시하게 된다. 일반선형회귀모형과 ARIMA의 단점과 한계점을 보완하여 제시된 것이 벡터자기회귀모형(VAR)이다. 시계열의 정상성이 만족되지 않으면 VAR을 사용할 수 없지만, 공적분 관계를 가지면 VECM을 적합할 수 있다. 앞서 언급한 시계열 분석 방법의 경우 원 시계열의 길이가 길어질수록 예측의 정확도가 떨어지는 단점이 있다. 딥러닝모형 중 가장 대표적인 시계열 분석 모형인 LSTM의 경우 긴 길이의 데이터에 대해서도 상대적으로 예측 성능이 좋다고 알려져 있다.

3.2.1. 시계열 자료의 정상성과 공적분 검정

시계열 자료는 시간이 변화함에 따라 자료에 그 특성이 반영된다. 이 시계열 자료가 불안정할 경우 안정적인 시계열 자료로 전환 후 모형을 추정해야 한다. 불안정한 시계열 자료를 그대로 모형 추정에 사용하면, 아무 상관관계가 없는데도 의미가 있는 것처럼 보이는 허구적 회귀(spurious regression)와 같은 잘못된 결론이 도출되기 때문이다.

시간 $t = 1, 2, \dots, T$ 에 대해서 다변량 시계열이 아래 조건을 만족하는 경우 정상성(stationarity)을 만족한다고 한다.

$$(1) E(X_t) = \mu < \infty$$

$$(2) Var(X_t) = E[(X_t - \mu)(X_t - \mu)'] = \Sigma_X < \infty$$

$$(3) Cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)'] = T_X(k)$$

이는 평균과 분산, 공분산이 시간에 따라 불변함을 의미하는데, 정상시계열은 일시적인 충격에 의해 추세치를 이탈하더라도 결국 추세에 회귀하는 경향을 가진다. 시계열의 정상성을

확인하기 위해 단위근 검정을 시행한다. 본 연구에서는 일반적으로 널리 사용하고 있는 ADF (Augmented Dickey-Fuller test) 검정 방법을 사용하였다.

$$(4) \Delta y_t = \alpha + \delta y_{t-1} + \sum_{j=1}^p \gamma_j \Delta y_{t-j} + \epsilon_t$$

위 식에서 y_t 는 검정할 시계열 변수, δ 는 검정계수, p 는 시차, ϵ_t 은 백색잡음인 오차항을 나타낸다. ADF 검정의 귀무가설 H_0 은 $\Delta y_t \sim I(0)$ 이다. δ 가 0이 되면 귀무가설을 기각하는데, 이는 단위근이 존재하지 않는 안정적인 시계열임을 의미한다. 귀무가설을 기각하지 않는 반대의 경우, 시계열이 불안정함을 의미한다.

개별 시계열이 비정상(불안정)하여 변동이 있더라도 이들 사이에 장기적으로 안정적인 시계열을 생성하는 선형결합이 존재한다면 이 선형결합은 정상이며, 이들 시계열은 공적분(cointegration) 관계에 있다고 한다. 비정상시계열 사이에 공적분이 존재한다면 이를 나타내는 벡터오차수정모형을 이용해 모형을 추정할 수 있다. 공적분 검정을 하는 방법은 앵글&그레인저 공적분 검정(Engle and Granger Cointegration test)과 요한슨 공적분 검정(Johansen test)이 있다. 본 연구에서는 앵글&그레인저 공적분 검정의 한계점을 극복하고 다변량 시차의 구조를 설명하기에 적합한 요한슨 공적분 검정을 수행하였다.

3.2.2. 외생변수를 포함한 자기회귀누적이동평균모형(ARIMAX)

회귀모형은 시간의 흐름과 관계없이 설명변수의 영향력이 일정하다는 한계를 가지며, 이를 보완한 방법이 시계열 모형이다. 자기회귀모형(AR)은 임의의 변수에 대해 $t-p$ 시점 관측값이 t 시점 관측값에 영향을 미친다는 가정을 활용하고, 이동평균모형(MA)은 $t-q$ 시점 백색잡음과 t 시점 백색잡음으로 추세를 파악하여 미래를 예측한다. 자기회귀이동평균모형(ARMA)은 AR과 MA이 결합된 형태로 정상성을 만족하는 자료에 적합한 방법이다. ARIMA는 비정상적인 시계열 자료를 분석하는 기법으로 자기 자신의 추세만을 고려하는 일변량 자기회귀모형이다. 따라서 장·단기적인 예측력을 높이기 위해서 시간의 흐름에 따른 여러 가지 외부정보를 반영할 필요가 있다. ARIMAX(Auto Regressive Integrated Moving Average

with Exogenous)는 ARIMA에 독립변수를 추가하여 외생변수 또는 내생변수의 영향을 고려하는 다변량 자기회귀모형이다. ARIMAX(p,d,q,r) 모형의 식은 다음과 같다.

$$(5) \quad y_t = c + \epsilon_t + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \beta X$$

여기서 c 는 상수, ϵ_t 와 ϵ_{t-1} 는 동일한 두 기간에 대한 오류항, p 는 자기회귀모형의 차수, d 는 적분차수, q 는 이동평균 모형의 차수, r 은 독립변수의 지연 차수를 의미한다.

3.2.3. 벡터오차수정모형(VECM)

벡터오차수정모형(Vector Error Correction Model, VECM)은 오차수정모형을 VAR 모형으로 확장한 것이다. 공적분 검정을 통해 변수 간 공적분이 존재가 확인되면 장기적인 예측에 앞서 오차수정모형을 구성하여 추정할 필요가 있다. VECM은 특정 시점에서 장기적인 균형의 괴리가 발생하면 균형의 오차가 시간의 흐름에 따라 조정된다는 개념으로부터 설계된 모형이다. 앵글&그레인저(1987)는 두 변수가 비정상적인 시계열이면서 공적분 관계에 있을 때, 식 (6)을 추정하기 위해 식 (7)과 같은 오차수정모형으로 바꿔 쓸 수 있으며, 여기서 EC_{t-1} 은 오차수정항이 된다.

$$(6) \quad y_t = \alpha + \lambda y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t$$

$$(7) \quad (y_t - y_{t-1}) = \alpha - (1 - \lambda)y_{t-1} + \beta_0 \Delta x_t + (\beta_0 + \beta_1)x_{t-1} + \epsilon_t$$

$$\begin{aligned} \Delta y_t &= \beta_0 \Delta x_t - (1 - \lambda) \left[y_{t-1} - \frac{\alpha}{1 - \lambda} - \frac{(\beta_0 + \beta_1)}{1 - \lambda} x_{t-1} \right] + \epsilon_t \\ &= \beta_0 \Delta x_t - (1 - \lambda) EC_{t-1} + \epsilon_t \end{aligned}$$

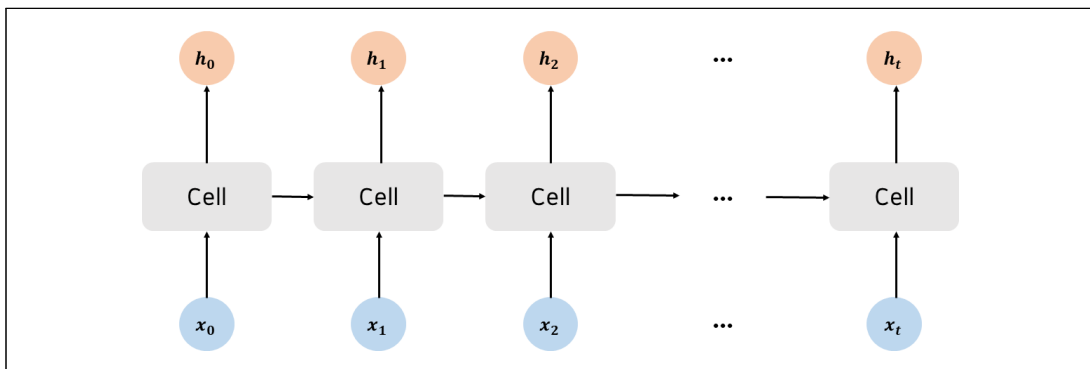
공적분 관계를 가지는 y_t, x_t 의 오차수정항은 정상시계열이다. 식 (7)에 대한 OLS 추정량은 일치추정량이 되고 t 검정이 가능하다. y_{t-1} 이 x_{t-1} 의 균형값에 비해 크면 균형 오차는 항상

양이 되고 다음 시점인 y_t 에 대해 음의 조정이 작용된다. 오차수정항의 계수인 $(1 - \lambda)$ 는 장기 균형으로 접근하는 조정속도이다. 안정적인 장기 균형을 위해 $(1 - \lambda)$ 는 0보다 커야 한다.

3.2.4. 장단기메모리모형(LSTM)

순환신경망모형(Recurrent Neural Network, RNN)은 연속성을 가지면서 시점 간의 연관성이 있는 데이터를 분석하는데 유용하다. 신경망 내부의 재귀적 특성으로 과거 시점의 데이터와 현재 시점의 데이터의 관계를 기억하여 순서 혹은 시간에 따른 데이터의 변화를 분석하는데 유용하다. <그림 2>는 RNN 구조로 새로운 정보를 입력받는 동시에 각 셀에서 다음 셀로 정보가 전달된다. $x_0, x_1, \dots, x_t$ 는 각 시점에 대한 새로운 입력값이고 $h_0, h_1, \dots, h_t$ 는 그에 대한 출력값이다.

그림 2. RNN 구조



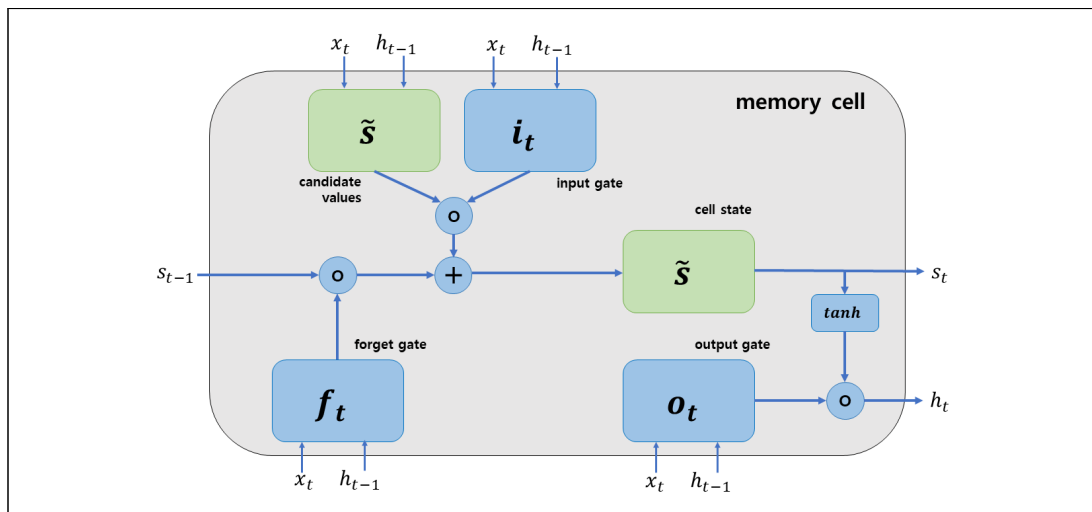
RNN에서 가장 중요한 점은 새로운 입력을 학습하면서 입력된 정보가 출력값의 형태로 다음 셀에 전달되어 신경망 내부에서 순차적 흐름의 학습이 동시에 이루어진다는 점이다. 이러한 구조로 시간에 따른 데이터를 학습하고 분석하는데 유용하게 이용되고 있다. t 시점의 출력값인 h_t 는 식 (8)과 같이 나타낼 수 있다.

$$(8) \quad h_t = \tanh(h_{t-1} W_h + x_t W_x + b)$$

RNN에서 t 시점의 출력값 h_t 는 두 개의 가중치가 있다. Wx 는 입력값인 x_t 를 h_t 로 변환시키는 가중치이고 W_h 는 이전 단계의 출력값인 h_{t-1} 을 다음 단계로 전달하는데 변환시키는 가중치로 학습을 통해 업데이트가 이루어진다. b 는 바이어스를 나타내며 이들은 활성화함수에 의해 변환되어 출력값인 h_t 가 생성된다. 그러나 RNN은 입력 데이터가 길어질수록 과거의 데이터의 영향이 감소하는 기울기 소실(vanishing gradient)이 발생하여 가중치 갱신이 일어나지 않고 학습의 문제가 생기는 장기 의존성 문제가 나타난다. LSTM은 이러한 문제에 대한 대안으로 나타났다.

LSTM은 RNN과 같이 순차적으로 정보가 전달되어 재귀적 특성을 갖지만, 신경망 내부에 메모리 셀이라는 개념이 추가되었다. 메모리 셀은 각 시점의 cell state를 통해 일정 비율로 기억하거나 폐기함으로써 과거의 정보를 얼마나 반영할지 제어할 수 있다. 이를 통해 RNN의 장기 의존성 문제를 해결하여 먼 시점 간의 관계를 학습하거나 그 정도를 조절할 수 있다. 메모리 셀 내부에는 forget gate, input gate, output gate라는 요소가 이전 정보와 입력 정보, 출력정보를 유지 및 조절한다. 여기서 forget gate는 cell state에서 정보의 망각 정도, input gate는 정보의 반영 정도, output gate는 출력을 결정한다. 각 게이트는 sigmoid 연산을 통해 0에서 1 사이의 값을 출력한다.

그림 3. LSTM Cell 구조



아래의 식 (9)의 forget gate에서는 과거의 정보를 얼마나 반영할 것인지를 결정하고 식 (10)의 input gate에서는 현재의 정보를 얼마나 반영할 것인가를 결정한다. 식 (11)은 활성화 함수를 통해 cell state에 반영할 값을 나타내고 식 (12)는 앞의 과정이 반영된 현재 시점의 cell state이다. 식 (13)의 output gate는 출력값을 나타내며 식 (14)는 output과 cell state의 조합으로 다음 셀에 전달할 값이다.

$$(9) f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$(10) i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$(11) \tilde{s}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

$$(12) s_t = f_t \cdot s_{t-1} + i_t \cdot \tilde{s}_t$$

$$(13) o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$(14) h_t = o_t \cdot \tanh(s_t)$$

3.2.5. 성능지표

예측 모델 간의 성능 비교를 위한 기준으로 회귀문제에서 대표적으로 쓰이는 두 가지 지표인 MAPE(Mean Absolute Percentage Error)와 RMSE(Root Mean Square Error)를 사용하였다. MAPE는 잔차에 대한 실제값의 비율을 의미하며 백분율로 나타낸다. 실제값이 1보다 작거나 0인 경우에는 사용하기 어렵지만 본 연구의 데이터는 값이 비교적 크고 0을 갖지 않기 때문에 적용하기에 알맞다. RMSE는 잔차의 제곱값을 평균한 MSE에 제곱근을 취한 것으로 예측변수와 같은 단위로 해석할 수 있다. 잔차를 제곱하기 때문에 이상치에 민감한 편이다. 두 지표의 공식은 다음과 같다.

$$(15) MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|, RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

4. 분석 결과

본 연구에서는 2012년부터 2018년까지의 데이터는 모형의 학습에 사용하였고, 2019년부터 2021년까지의 데이터는 학습된 모형의 성능을 평가하는데 사용하였다. 설명변수의 단위가 다르고, 값의 크기가 달라 절대적 크기의 차이로 인한 모형의 학습이 왜곡되는 것을 방지하고자 표준화하여 사용하였다. 각 설명변수의 평균과 표준편차를 이용하여 절대적 크기의 차이를 줄였다.

4.1. 시계열 자료의 정상성 및 공적분 검정 결과

시계열 자료의 정상성 확인을 위해 수행한 ADF 검정 결과는 아래 <표 3>과 같다. 검정 결과 계란가격과 신선란 수입 무게는 귀무가설을 기각하지 못해 단위근을 가지는 불안정 시계열이었다. 원 데이터를 1차 차분하면 유의수준 1%에서 귀무가설을 기각하여 정상적인 시계열임을 확인하였다.

표 3. ADF 검정 결과

구분	원데이터	1차 차분 데이터
Price	-1.85(0.63)	-11.18(0.01)***
Count	-6.7016(0.01)***	-15.902(0.01)***
Articles	-15.329(0.01)***	-24.353(0.01)***
Import	-1.8891(0.6251)	-15.657(0.01)***
Export	-3.964(0.0108)***	-15.541(0.01)***

주: *, **, ***은 각각 유의수준 10%, 5%, 1%에서 유의함을 의미함.

또한, 앞에서 언급한 바와 같이 개별 시계열 자료들이 비정상적이더라도 변수 간의 선형 결합이 안정적이면 그 변수 간에는 장기 균형 관계가 존재하게 된다. 요한슨 공적분 검정은 시계열 변수들 사이에 존재하는 공적분 벡터의 수를 우도비(likelihood ratio) 검정을 통해 결정하고 이후 VECM을 추정한다. 공적분 검정을 하기에 앞서 적절한 VAR의 차수를 결정해야 한

다. 차수를 결정하는 기준은 AIC(Akaike), SC(Schwarz) 등의 기초하여 결정하는데, AIC는 시차가 크게 선택되는 경향이 있어 SC를 기준으로 적정 시차를 2로 설정하였다. 아래 <표 4>에서 공적분 검정 결과를 확인할 수 있고, 유의수준 5%에서 4개 이상의 공적분 관계가 있으므로 VECM으로 분석이 가능하다.

표 4. 공적분 검정 결과

H_0	Eigenvalue	Trace Statistic	5% Critical Value
$r = 0$	0.4618	417.60	70.60***
$r \geq 1$	0.3187	190.85	48.28***
$r \geq 2$	0.0585	50.36	31.52***
$r \geq 3$	0.0467	28.29	17.95***
$r \geq 4$	0.0289	10.75	8.18***

주: *, **, ***은 각각 유의수준 10%, 5%, 1%에서 유의함을 의미함.

4.2. ARIMAX 추정 결과

ARIMAX를 적합하기 위해서는 적절한 계수를 선택해야 한다. 이를 위해 R 프로그램의 forecast 패키지를 이용하였다. AIC, BIC 등을 이용해 적절한 계수를 선택할 수 있는데, 계란가격을 추정하기 위한 적절한 모형으로는 ARIMAX(2,1,0)가 선택되었다. 추정된 모형은 아래와 같으며 <표 5>에서 각 계수별 표준편차와 유의확률을 확인할 수 있다. 각 계수를 확인하면 AR(1)과 AR(2)가 유의수준 1%에서 유의하며 시점에서 계란가격에 양의 영향을 주고 있다. 확진 건수 역시 유의한 변수이나 자기회귀항과는 다르게 음의 영향을 가지는 것을 확인할 수 있다. 이외의 기사 건수, 신선란 수입 무게, 신선란 수출 무게는 유의하지 않았다.

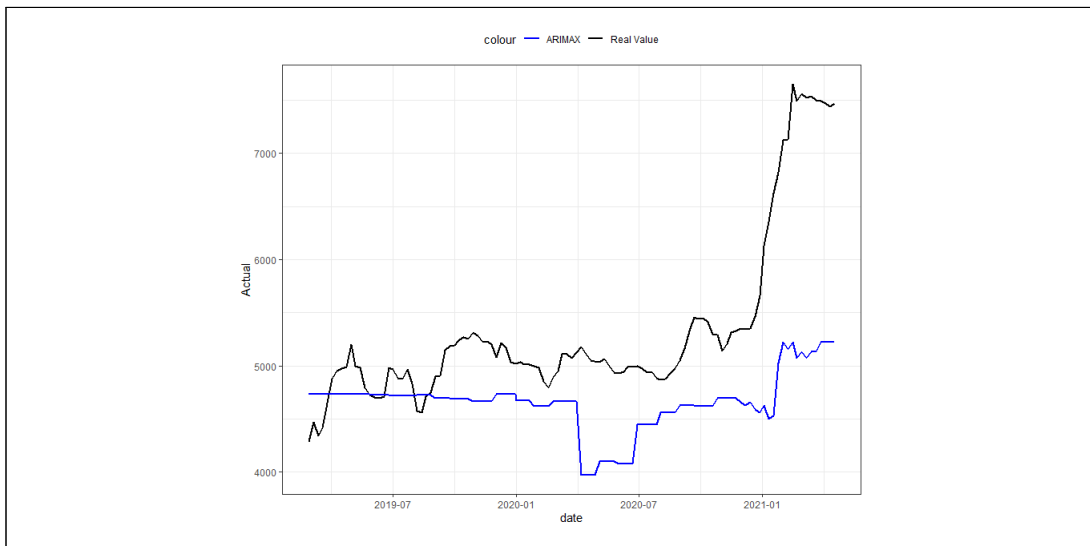
표 5. ARIMAX 모형 추정 결과

	Estimate	std.Error	z-value	Pr(> z)
y_{t-1}	0.2194	0.0556	3.94	8.148e-05***
y_{t-2}	0.1573	0.0528	2.97	0.002896**

	Estimate	std.Error	z-value	Pr(> z)
Count	-96.6296	15.4355	-6.26	3.845e-10***
Articles	11.3210	6.9668	1.625	0.1041
Import	80.5537	66.3087	1.2148	0.2244
Export	-110.8116	180.3068	-0.6146	0.5388

아래 <그림 4>는 추정된 모형을 이용해 예측한 값을 그린 것이다. 추정된 모형인 ARIMAX (2,1,0)는 평균적인 가격을 낮게 예측하였으며 변동을 정확히 포착하지 못했다. 예측된 계란 가격은 2020년 4월 초에 수직으로 하락하고 이후 10주가량 하락한 가격으로 유지되다가 2020년 6월 초에 다시 원래의 가격으로 급격히 상승하는 모습을 보인다. 이는 이전까지 꾸준히 조류인플루엔자 관련 기사가 보도되었지만 2020년 2월부터 급격히 퍼지기 시작한 COVID-19가 사회적으로 이슈화되어 조류인플루엔자에 대한 관심이 줄어들어 기사 건수가 0인 것의 영향으로 볼 수 있다. 이후 2020년 말부터 서서히 조류인플루엔자가 발병하며 2021년에 들어서면서 본격적으로 확산되어 계란가격이 급격히 상승하는 모습을 볼 수 있다. 실제로는 가격이 3,000원 정도 상승하였으나 예측값은 500원 정도에 그치는 것을 확인하였다. 마지막으로 추정된 모형을 이용해 계산한 MAPE와 RMSE는 각각 0.1213, 961.8527이었다.

그림 4. 실제 계란가격과 ARIMAX 예측값 비교



4.3. VECM 추정 결과

VECM은 예측하고자 하는 계란가격과 다른 변수를 함께 모형에서 사용하므로 따로 표준화나 정규화를 하지 않은 원 데이터를 사용하였다. VECM 추정 결과는 아래 <표 6>과 오차수정항은 아래 식 (17)과 같다.

표 6. VECM 추정 결과

	$\Delta Price$	$\Delta Count$	$\Delta Articles$	$\Delta Import$	$\Delta Export$
ECT_{t-1}	-0.002(0.0043)	0.000(0.00)	-0.0022(0.0066)	0.3967(0.185)*	0.0574(0.016)***
<i>Intercept</i>	7.1079(19.29)	-0.069(0.136)	9.283(29.529)	-1474.47(832.4)	-222.36(75.67)**
$\Delta Price_{t-1}$	0.26(0.054)***	-0.001(0.0004)**	0.0998(0.082)	10.617(2.335)***	-0.2576(0.2123)
$\Delta Count_{t-1}$	10.3573(8.279)	0.4207(0.061)***	124.45(13.35)***	-637.022(376.6)	-70.939(34.23)*
$\Delta Articles_{t-1}$	0.0682(0.0369)	-0.002(0.000)***	-0.801(0.056)***	-3.867(1.5945)*	-0.1756(0.1449)
$\Delta Export_{t-1}$	0.0149(0.0134)	-0.0002(0.00)	0.0473(0.0205)*	0.3048(0.577)	0.0202(0.0525)
$\Delta Import_{t-1}$	0.0035(0.001)**	0.000(0.000)	-0.0015(0.0019)	0.0184(0.0525)	0.0051(0.0048)
$\Delta Price_{t-2}$	0.1812(0.055)**	-0.0009(0.000)*	0.1966(0.0843)*	5.0292(2.3772)*	-0.3543(0.2161)
$\Delta Count_{t-2}$	13.3292(10.90)	0.0435(0.0769)	135.44(16.69)***	-81.485(470.64)	87.313(42.782)*
$\Delta Articles_{t-2}$	0.0099(0.040)	0.0005(0.0003)	-0.312(0.061)***	6.4895(1.724)***	-0.379(0.1568)*
$\Delta Export_{t-2}$	-0.0032(0.013)	0.000(0.000)	0.0676(0.0206)	-0.0437(0.5805)	0.0269(0.0528)
$\Delta Import_{t-2}$	-0.0024(0.001)*	0.000(0.000)	-0.018(0.0017)	-0.013(0.0471)	0.000(0.0043)

$$(17) ECT_{t-1} = Price - 155.86*Count - 0.70*Articles - 0.08*Import - 1.07*Export$$

식 (17)에서 계란가격, 확진 건수, 기사 건수, 신선란 수출입 무게의 장기균형 관계를 확인할 수 있다. 오차수정항에서 오차항이 0이면,

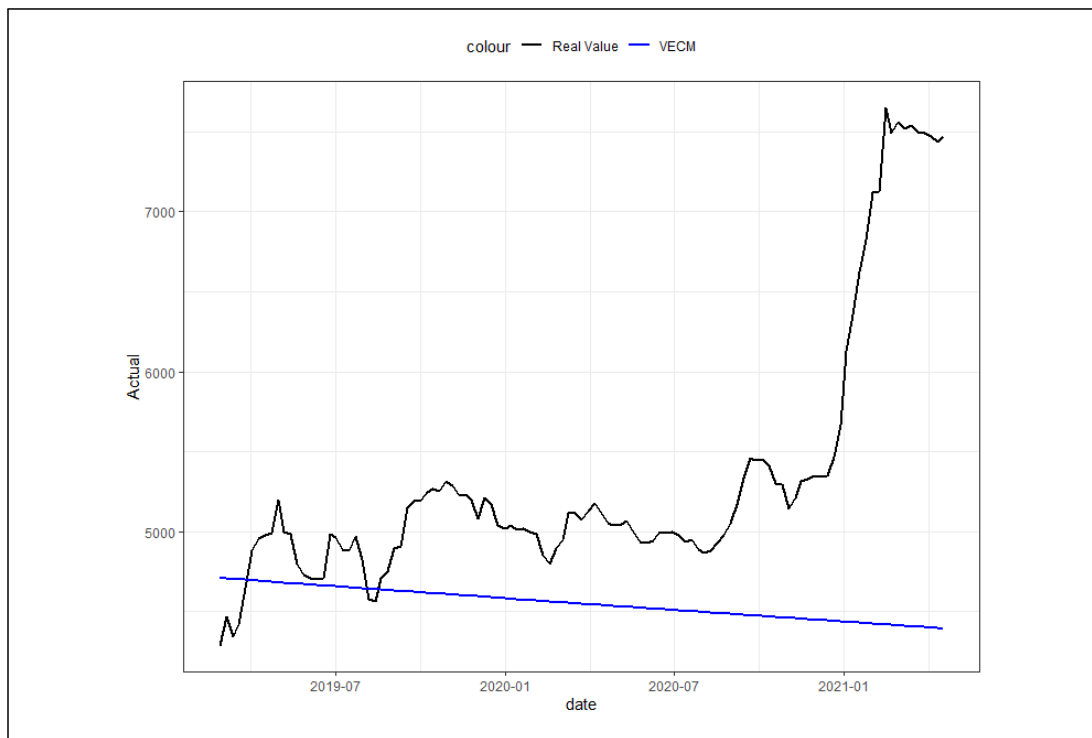
$$(18) Price = 155.86*Count + 0.70*Articles + 0.08*Import + 1.07*Export$$

식 (18)과 같이 정리할 수 있다. 이는 확진 건수가 1% 증가하면 계란가격은 155.86% 증가하고, 기사 건수, 신선란 수입 무게, 신선란 수출 무게가 각각 1%씩 증가하는 경우 계란가격

은 0.7%, 0.08%, 1.07% 증가함을 의미한다. 확진 건수가 가장 영향력이 컸으며 신선란 수입 무게는 다른 변수에 비해 미미한 것으로 나타났다. 오차수정항(ECT_{t-1})의 계수를 살펴보면 다른 변수에 비해 신선란 수입 무게(0.3967)의 절댓값이 가장 큰 것을 볼 수 있는데, 이는 만약 장기 균형 관계에 문제가 생긴다면 신선란 수입을 통해 계란가격의 조정이 이루어짐을 알 수 있다.

추정된 모형을 이용해 2019년부터 2021년까지의 주별 계란가격을 예상한 결과는 <그림 5>와 같다. 2019년까지는 평균적인 선을 잘 따라가는 것처럼 보이나 2021년으로 갈수록 점점 예측력이 떨어지는 것을 볼 수 있다. 특히 조류인플루엔자가 발생해 계란가격이 상승하기 시작한 2020년 말부터는 치솟는 계란가격을 따라가지 못했다. 마지막으로 추정된 모형을 이용해 계산한 MAPE와 RMSE는 각각 0.1352, 1175.4238이었다.

그림 5. 실제 계란가격과 VECM 예측값 비교



4.4. LSTM 추정 결과

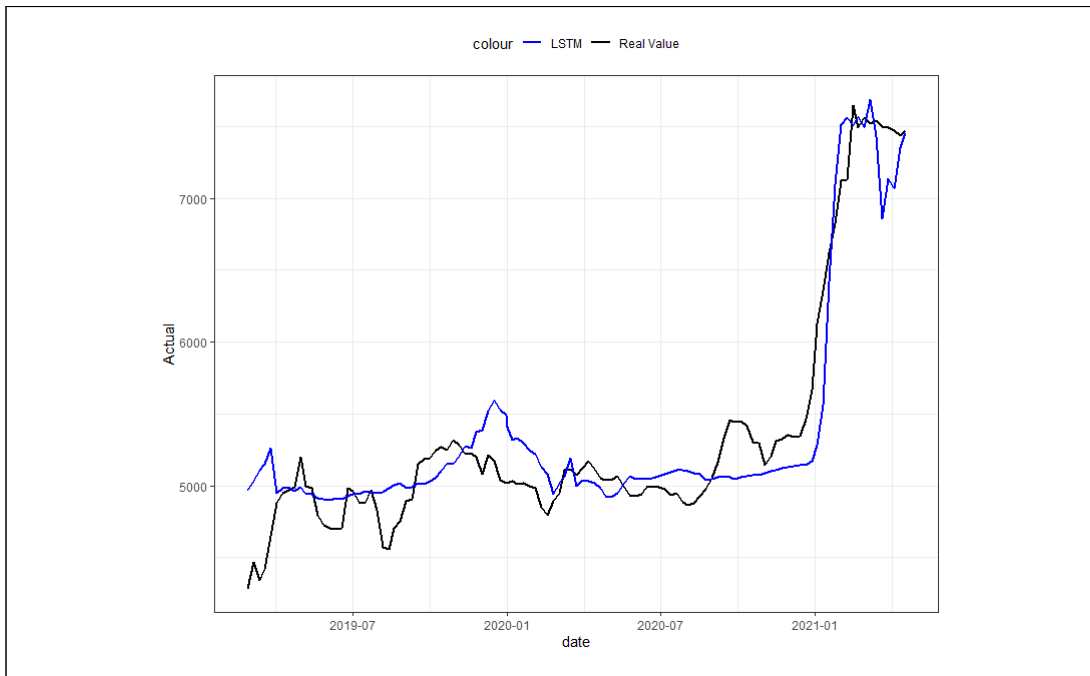
LSTM을 이용하여 시계열 분석을 하기 위해서는 연속된 형태의 데이터 구조를 시퀀스 형태로 변환하는 과정이 필요하다. 분석에 사용되는 데이터는 2012년 1월에서 2021년 4월까지 주 단위로 연속적인 형태이다. 이를 목적과 모델의 성능에 따라 학습 sampling window와 target의 형태로 데이터를 구조화해준다. LSTM은 기간의 데이터를 학습하여 target의 데이터를 예측하는 형태가 된다. 본 연구에서는 sampling window를 6주, 8주, 10주, 12주로 하였고 target을 1주로 하여 최소 6주에서 최대 12주의 데이터를 학습하여 다음 한 주의 가격을 예측하는 네 개의 데이터셋으로 구조화하였다. 또한, 계란가격의 연간 주기성을 반영하기 위하여 매년의 주 번호(weeknumber)를 변수로 추가하여 주 번호, 확진 건수, 기사 건수, 신선란 수입 무게, 신선란 수출 무게 등 다섯 가지의 설명변수를 이용하였다.

LSTM의 하이퍼 파라미터를 설정하기 위하여 그리드 탐색을 하였고 예측 성능이 가장 좋은 값으로 최적화하였다. 설정한 하이퍼 파라미터는 은닉층과 은닉노드의 개수, dropout과 recurrent dropout 비율, 최적화 알고리즘, 학습률, 에포크, 배치 크기 등이다. 분석 결과 가장 좋은 모델은 sampling window가 8주이고 각각 32개와 16개의 은닉노드를 갖는 2개의 은닉층을 갖는 모델로 나타났다. 과적합을 방지하는 방법으로써 LSTM에 적용되는 dropout은 각 cell의 input에 대해 수직 방향으로 이루어지는 dropout과 이전 셀에서 현재 셀로 전달되는 hidden state에 대해 수평 방향으로 이루어지는 recurrent dropout이 있다. recurrent dropout의 경우 본 연구 모델에 적용 시 성능을 하락시키므로 적용하지 않았고, dropout의 경우 input의 영향을 다소 감소시켜 성능을 향상시키므로 두 개의 은닉층에서 각각 0.2의 비율로 적용하였다. 최적화 알고리즘은 adam을 사용하였으며 학습률은 0.01이고 에포크는 100번, 배치 크기는 16으로 학습하였다.

〈그림 6〉은 추정된 모형을 이용해 2019년부터 2021년까지의 주별 계란가격을 예상한 결과이다. 예측 초기에는 실제값과 예측값의 방향과 값의 차이가 다소 크게 나타났지만, 이후에는 값과 방향의 차이가 크지 않는 것으로 나타났다. 2021년의 경우 실제 계란가격이 최고점에 다다른 뒤 횡보하고 있지만 예측값은 한번 떨어졌다가 복귀하는 모습을 보이고 있다. 이는 앞서 설명했던 ARIMAX 추정 결과와 같이 기사 건수의 왜곡에서 비롯되는 것으로 추측해볼

수 있다. 마지막으로 추정된 모형을 이용해 계산한 MAPE와 RMSE는 각각 0.0412, 282.7135 이었다.

그림 6. 실제 계란가격과 LSTM 예측값 비교



5. 요약 및 결론

계란은 다양한 용도로 활용되고 있는 필수 식자재 중 하나이다. 전염성이 강한 조류인플루엔자의 발생과 가금류 농장에 대한 방역 조치는 안정적인 계란 수급과 생산에 어려움을 불러 일으킨다. 고병원성 조류인플루엔자는 급성 전염병으로 폐사하거나 산란율이 감소하는 등의 피해를 야기한다. 전염성이 높기 때문에 예방적 차원에서 주변 농장의 가금류를 살처분함으로써 대응하는데, 이는 계란가격의 상승으로 이어진다. 따라서 조류인플루엔자의 영향을 고려한 계란가격 예측 연구가 필요하다. 정형적 정보는 정보 수집과 수치화까지 상대적으로 긴 시

간이 필요한데, 온라인 뉴스 기사는 종합적인 이슈와 변화를 빠르게 반영할 수 있다는 장점이 있다. 본 연구에서는 조류인플루엔자와 관련한 뉴스 기사를 수치화하고, 계란가격에 영향을 미칠 수 있는 요인들을 설명변수로 사용하여 계란 주별 소매 가격의 변동을 예측하고자 하였다.

분석에 사용한 데이터는 계란가격, 확진 건수, 기사 건수, 신선란 수입 무게, 신선란 수출 무게 등이며 계란가격을 종속변수로 사용하였다. 월별 단위로 제공되는 데이터 분석의 한계와 실제 소비자에게 정보 효용성을 고려하여 주별 단위로 데이터를 가공하고 분석에 사용하였다.

분석 방법으로는 ARIMAX, VECM 그리고 LSTM을 사용하였다. 분석에 앞서 원데이터의 정상성과 공적분 검정을 수행하였다. 계란가격과 신선란 수입 무게는 불안정 시계열이었으며 이외 확진 건수나 기사 건수, 신선란 수출 무게는 정상시계열이었다. 본 연구의 분석 데이터와 같이 비정상시계열이면서 공적분 관계를 가지면 VECM을 사용할 수 있다. 2012년부터 2018년까지의 데이터를 이용해 모형을 추정하고 2019년부터 2021년까지 데이터를 이용해 추정된 모형의 성능을 평가하였다. 가장 좋은 성능을 보인 것은 LSTM이었으며, 다른 모형에 비해 뛰어난 결과(MAPE 0.042, RMSE 282.7135)를 보였다. 전반적인 추세와 가격의 급격한 변동 모두 포착하였으나 본 연구의 설명변수로 확인할 수 없는 주기적 변동은 예측이 불가능하였다. ARIMAX의 경우, 전체적인 추세는 잘 예측하였지만 조류인플루엔자로 인한 급격한 가격 변화는 예측하지 못하였다. 이는 설명변수의 영향이 잘 반영되지 못한 것으로 보인다. VECM의 경우, 전반적인 추세와 방향을 모두 예측하지 못하므로 계란가격 예측을 위한 모형으로 적합하지 않았다. 통계적 시계열 모형은 장기적 예측과 급격한 변동을 가지는 데이터를 분석하는데 한계를 가진다. LSTM은 상대적으로 좋은 성능을 보였으나 계란가격을 예측하기 위한 변수가 충분하지 않았다. 따라서 이를 보완할 수 있는 다른 변수를 찾는 것이 향후 연구에 중요한 과제가 될 것이다.

참고문헌

- 모인필. 2007. “고병원성 조류인플루엔자의 역학적 특성과 국내 발생 동향.” 『Crisisonomy』 3(2): 30-37.
- 박문수, 황선웅, 서상택. 2012. “축산물 시장의 가격전달채널 분석 및 가격안정화 방안: 육계 및 계란 시장을 중심으로.” 『한국농촌경제연구원 연구자료』 1-158.
- 김정현, 서병선. 2014. “구제역 발생이 돈육가격에 미치는 영향: 사건 연구(Event Study)의 적용.” 『농촌경제』 37(1): 97-114.
- 서대교, 신종협. 2014. “가축질병이 축산물가격에 미친 영향 분석.” 『산업경제연구』 27(6): 2431-2450.
- 송대식. 2001. “육계의 가격예측과 가격변화에 관한 연구.” 『식품유통연구』 18(1): 117-130.
- 명광식. 2005. “Box-Jenkins 모형을 이용한 육계가격 예측.” 『농촌경제』 28(2): 73-83.
- 강태훈. 2007. “주요 축산물 가격변동의 시계열적 특성.” 『농업경영·정책연구』 34: 369-388.
- 최병옥, 최익창. 2007. “시계열 분석방법을 이용한 과채류 월별가격 예측.” 『농촌경제』 30(1): 129-148.
- Uhn-Soon Gim, Se-Hyun Choi, and Jae-Hwan Cho, 2015. “An impact analysis of FMD news on pork demand in korea.” The Korean Journal of Community Living Science 26(1): 75-85.
- 김양석, 이충권. 2018. “뉴스기사 분석을 통한 사회이슈와 가격에 관한 연구-조류인플루엔자와 달걀 가격 중심으로.” 『스마트미디어저널』 7(1): 45-51.
- 신성호, 이미경, 송사광. 2018. “LSTM 네트워크를 활용한 농산물 가격 예측 모델.” 『한국콘텐츠학회 논문지』 18(11): 416-429.
- 문권순. 1997. “벡터자기회귀(VAR) 모형의 이해.” 『통계분석연구』 2(1): 23-56.
- 이형용, 여민수, 홍승지. 2017. “마늘 도매가격 시계열 예측 모형 비교. FTA 이행에 따른 낙농·유가공 산업의 경제적 영향 분석.” 『농업경제연구』 40(2): 59-63.
- 홍승지, 김성훈, 이을경. 2015. “오차수정모형을 이용한 고추의 재배면적반응함수 추정.” 『농업경영·정책연구』 42(3): 507-522.
- Alibabaei K., Gaspar P. D., Lima T. M. 2021. “Crop Yield Estimation Using Deep Learning Based on Climate Big Data and Irrigation Scheduling.” Energies 14(11): 3004.
- Tran, Q., and Song, S. (2017). “Water level forecasting based on deep learning: A use case of trinity river-Texas-The United States.” The Journal of Korean Institute of Information Scientists and Engineers 44(6): 607-612.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. 2014. “Dropout: a simple way to prevent neural networks from overfitting.” The Journal of Machine Learning Research 15: 1929-1958.
- Hochreiter, S., Schmidhuber, J. 1996. “Lstm can solve hard long time lag problems.” In Proceedings of the 9th International Conference on Neural Information Processing Systems(NIPS'96), MIT Press. Cambridge, MA, USA, 473-479.



제1회

농산물 수급 예측모형 경진대회

우수논문집

Chapter 2
자유과제



장려상

사육주기를 반영한 한육우 수급모형 구축

안미란 · 지인배

사육주기를 반영한 한육우 수급모형 구축

안미란* 지인배**

Key words

한육우(Korean cattle), 수급모형(Supply and demand model), 사육주기(Breeding cycle), 구조변화(Structural change), 사인함수(Sine function)

ABSTRACT

Due to the biological characteristics of Korean beef cattle, the breeding period is long, so the number of breeding heads repeats increase and decrease. So they have a certain breeding cycle. In this study, we develop a model using a sine function that can reflect these characteristics of the Korean breeding cycle. Recently, the Korean beef cattle industry is undergoing structural changes such as scale-up, specialization, and increase in integrated breeding, and this has been reflected in the model. The explanatory power of the model was 0.92, which was very high, and the MAPE was 2.38%, indicating that the model had excellent predictive power. As a result of forecasting the number of Korean beef cattle, it is predicted that it will gradually increase from 3.4 million in 2021, reach the maximum level of about 3.85 million in 2026, and then decrease. The results of this study are expected to be useful for resolving the supply and demand instability in the Korean beef cattle market and the management risk of farms.

차례

- | | |
|------------------|------------------|
| 1. 서론 | 4. 한육우 사육두수 전망결과 |
| 2. 모형설계 및 분석자료 | 5. 요약 및 결론 |
| 3. 분석결과 및 예측력 검증 | |

* 동국대학교 푸드시스템연구소

** 동국대학교 식품산업관리학과 교수, 교신저자

1. 서론

한육우는 사육 특성상 임신부터 송아지 출생 그리고 출하단계에 이르기까지 비교적 긴 생산기간이 소요된다. 이에 한육우의 공급과 수요 간에는 괴리가 발생하여 사육두수는 상승과 하락을 주기적으로 반복하고 있다. 예컨대 한육우 사육두수는 1989년 약 153만두에서 1996년 약 284만두 수준에 이르러 8년 사이에 2배 가까이 증가하였다. 그 이후 감소세가 이어지면서 2001년에는 약 140만두 저점을 기준으로 다시 증가하여 2012년에는 300만두를 넘어, 12년 동안 2배 이상 증가하였다(통계청, 「가축동향조사」). 이러한 현상으로 한육우 시장가격 불안정, 수급 불안정 사태가 발생하고 있다. 이에 사육농가는 생산 의사결정에 한계를 보여 경영 리스크에 직면하고 있으며, 시장과 정부는 한육우 수급조절에 어려움을 겪고 있다. 이러한 측면에서 한육우 시장의 안정과 농가의 합리적인 의사결정을 위해 한육우 공급물량을 예측할 수 있는 수급전망모형의 구축이 필요하다.

최근 10년간, 번식과 비육의 이원화 체계로 구분되었던 한육우 사육형태는 큰 변화를 가져왔는데, 대부분 소규모농가는 번식과 비육을 동시에 겸하는 일관사육형태로 변화되고 있으며 점차 규모화·전문화되어 가고 있다. 이와 같은 한우산업의 사육구조 변화로 한육우 사육두수의 변동 폭은 점차 완화되었으며, 약 10년을 주기로 하던 한육우 사이클은 점차 줄어들고 있다(전상곤 외 2020년). 따라서 보다 정밀한 한육우 수급전망을 위해 사육두수의 변화패턴을 살펴보고 기존 사육주기와 더불어 한육우 산업의 구조변화로 인한 사육주기의 변화를 고려할 필요가 있다.

한육우 수급모형 개발에 관한 선행연구는 KREI-KASMO(KREI-Agricultural Simulation Model)모형, 한육우 단기관측모형(조재성 외 2013), 한우 분기관측모형(조재성 외 2017) 등이 대표적이다. KREI-KASMO는 농축산물의 품목별 수급을 연간단위로 전망하기 위한 모형으로 한육우 포함한 22개 품목군이 포함되어 있어, 한육우의 독립적인 전망이 어렵다는 한계가 있다. 그리고 한육우 단기관측모형과 한우 분기관측모형은 한육우(한우)시장에 최적화된 전망모형으로서 한육우(한우)의 구제역 발생과 소비시장의 계절성 등을 반영하였으며, 수급 전망 단위를 분기별로 세분화하였다는 점에서 KREI-KASMO모형과 차별성을 두고 있다. 그러나 이러한 모형은 과거 「가축동향조사」 자료를 토대로 구축된 모형인 만큼 쇠고기이력제

자료가 대체된 현시점에서 더 이상 활용할 수 없다는 한계가 있다(김충현 외 2018). 이에 김충현 외(2018)는 쇠고기이력제 자료의 특징과 장점을 살펴보고, 이를 토대로 한우 수급모형을 구축하여 암·수 출생두수, 사육두수 등의 분기별 전망치를 도출하였다. 이들 모형은 분석대상을 한우와 육우로 세분화하였는지, 동질화 하였는지 그리고 전망 단위를 연간으로 하였는지, 분기별로 하였는지에 따라 서로 차이를 보이고 있으며, 대체로 수요부문, 공급부문, 가격부문을 포함한 연립방정식체계를 활용하였으며, 성별·월령별 사육정보, 가격정보 등 자료를 바탕으로 한우(한육우) 사육두수를 성별·월령별로 전망하였다는 점은 유사하다. 이렇듯, 기존의 수급전망모형은 분석자료와 복잡한 수식에 기반하였기 때문에 활용도가 다소 떨어질 뿐만 아니라 한육우의 사육주기 변화를 반영하지 못하여 수급전망의 정확성에 한계를 지닌다. 한편, 전상곤(2012)은 사인함수를 통해 한우 사육주기를 분석하고, 이를 설명변수로 하는 회귀모형을 통해 사육두수를 전망하였다. 이 연구는 비록 사육주기를 사인함수형태로 표현하였으나 한육우 산업의 구조변화를 반영하지 못하였으며, 「가축동향조사」 자료를 활용하였다는 한계가 있다.

따라서 이 연구는 선행연구의 한계점을 보완하여 「가축동향조사」 자료와 쇠고기이력제 자료의 차이를 반영한 쇠고기이력제 자료를 활용하고자 한다. 또한, 한육우 산업의 구조변화 전과 후 사육주기의 변화패턴을 분석하고, 이를 반영한 수급모형을 개발하여 향후 한육우 사육두수를 전망하고자 한다. 이러한 수급모형의 개발은 한육우의 수급조절과 농가의 경영계획 수립에 유용한 기초자료가 될 것으로 기대된다.

2. 모형설계 및 분석자료

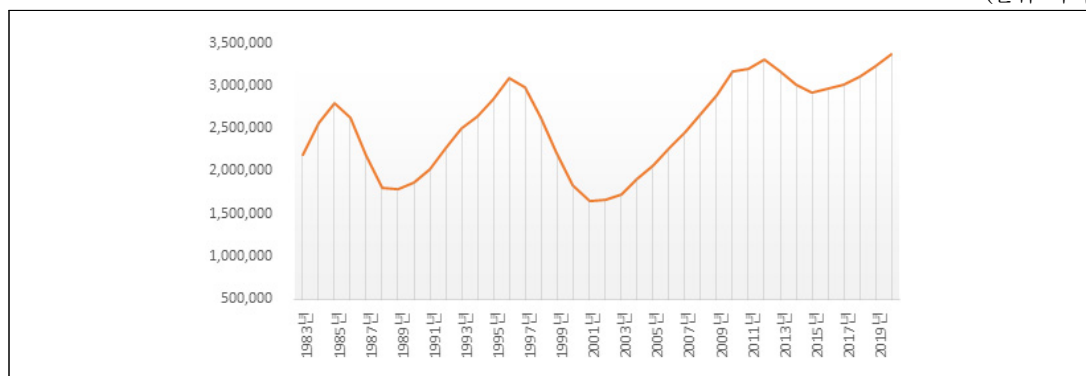
이 연구에서 구축한 수급모형과 기존 모형의 차이점은 한육우 사육두수에 내재된 사육주기의 변화특성을 고려한다는 점이다. 이에 분석모형 설정에 앞서 한육우 사육두수의 변화패턴을 살펴보고자 한다. 한육우 사육두수 관련 통계자료는 「가축동향조사」가 있는데 2017년부터 해당 자료가 쇠고기이력제 자료로 대체되었는데, 이력제 자료의 한육우 사육두수가 「가축

동향조사」의 사육두수보다 2016년 12월 기준 24만 9천 마리 더 많은 것으로 나타났다. 이는 조사 및 추계방법의 차이로 인해 초래된 것이다. 즉, 「가축동향조사」 자료는 소 사육농가를 대상으로 현장 표본농가 조사 방식으로 집계된 자료이나 쇠고기이력제 자료는 농가와 법인을 포함하여 소 개체에 대한 전수신고를 토대로 취합된 자료이기 때문이다. 이러한 측면에서 쇠고기이력제 자료의 신뢰도가 보다 높아 쇠고기이력제 자료를 활용한 수급모형은 기존 자료를 기반으로 한 수급모형 보다 더욱 정밀하다는 평가를 받고 있다(김현충 외 2018). 그러나 쇠고기이력제 자료의 누적기간이 짧아 수급모형을 구축하기 위한 연도별 자료가 불충분하다. 따라서 기존의 통계청 「가축동향조사」와 쇠고기이력제 자료를 연계하고자 두 자료 간의 사육두수 차이분을 「가축동향조사」에 더하여 한육우 사육두수를 재산출함으로써 1983년부터 2020년까지의 연도별 사육두수 자료를 구축하였다.

한육우 사육두수의 변화는 <그림 1>과 같이, 하나의 정점에서 저점으로 떨어졌다가 다시 정점에 오르는 패턴을 가지는데, 그에 따른 사육주기는 1985년부터 2020년까지 3개의 구간으로 구분된다. 첫 번째 구간은 1985년부터 1995년으로 10년이며, 두 번째 구간은 1996년부터 2012년으로서 첫 번째 구간보다 더 길어진 16년으로 나타나, 한육우 사육주기는 점차 길어진다는 것을 알 수 있다. 세 번째 구간은 2013년부터 시작되는데 2009년 축사시설현대화 사업의 추진에 따른 생산성 향상, 한우농가의 규모화, 전업화, 일관사육, 비육기간 증가 등의 구조변화로 인해 사육주기는 급격히 변화된 것으로 판단된다.

그림 1. 연도별 한육우 사육두수 현황

(단위: 마리)



자료: 축산물품질평가원

한육우 사육두수 변화에 따른 사육주기의 변화패턴은 사인함수의 주기와 유사한 형태를 가진다. 따라서 이 연구는 기존의 수급모형과 달리 사인함수식을 이용하여 한육우의 사육주기 변화를 반영하고, 이를 설명변수로 활용한 회귀모형을 통해 한육우의 수급을 전망하고자 한다. 그러나 사인함수는 라디안이 2π 를 주기로 하는 한편 사육주기(*cycle*)는 연도에 따라 달라지므로 시간(*time*)을 주기로 한다. 따라서 여기서 가장 중요한 것은 시간의 개념을 라디안의 개념으로 전환해 주는 것인데, 이에 대한 부분은 전상곤(2012)의 연구를 참고하여 시간의 변화에 따른 사육주기의 변화를 반영한 함수식을 다음의 수식 (1)과 같이 구축하였다.¹⁾

$$(1) \sin\left(\frac{time}{cycle} \times 2\pi\right) \quad (time = 1, 2, 3, \dots)$$

〈그림 1〉에서 제시한 사육두수의 변화패턴을 살펴보면, 사육두수가 점차 상승하는 기간은 하락하는 기간보다 더 긴 것으로 나타났다. 이와 같이 사육두수의 상승세가 하락세보다 완만한 것은 한육우의 도축보다 생산과 사육에 걸리는 시간이 더 길기 때문인 것으로 판단된다. 이러한 사육주기의 특성을 수급모형에 반영하기 위해 1983년의 사육주기를 10년으로 설정하고 그 이후부터 상승추세가 이어질 경우의 사육주기는 전년 대비 0.1, 하락추세가 이어질 경우의 사육주기는 전년 대비 0.05씩 늘어나는 값을 취하여 사육두수의 상승구간과 하락구간을 차별화하였다.

한육우 시장가격이 하락할 경우, 농가의 사육의향이 다소 떨어져 사육두수가 감소하는 반면, 시장가격이 상승할 경우, 농가는 사육의향이 높아져 사육두수는 증가할 것이다. 이렇듯, 한육우의 사육두수는 시장가격의 영향을 받는데 시장가격은 공급과 수요에 의해 결정된다. 궁극적으로 한육우의 사육두수의 증감은 한육우에 대한 수요(쇠고기 소비량)의 변화에 의해 결정된다. 국내 1인당 쇠고기 소비량은 2005년 6.74kg에서 지속적으로 증가하여 2019년 13kg에 이르고 있다(한국육류유통수출협회). 한육우의 소비가 꾸준히 증가추세에 있는 것은 우리나라 국민소득이 향상된 결과라고 볼 수 있다. 〈그림 1〉에서 보았듯이, 한육우의 사육두

1) 전상곤(2012)은 시간(*time*)을 라디안으로 전환하기 위한 수식은 $-1 \leq \sin\left(\frac{time}{사육주기} \times 2\pi\right) \leq 1$ 이라고 제시하였으며, 해당 수식으로 한육우의 사육주기를 반영하였다.

수는 등락을 반복하고 있지만 대체로 우상향하는 증가추세를 보이고 있다. 그러나 위의 식 (1)은 사육두수가 증감을 반복하는 현상을 반영하였으나 우상향하는 증가추세를 반영하지 못하였다. 이에 사인함수식에 1인당 쇠고기 소비량의 증가로 인해 사육두수가 우상향하는 추세를 반영하기 위해 1인당 실질 국민소득을 대리변수로 활용하여, 식 (2)를 구축하였다.

$$(2) \sin\left(\frac{time}{cycle} \times 2\pi\right) + \log(RGIN)$$

여기서 $RGIN$ 은 1인당 실질 국민소득을 의미한다.

한육우 가구당 사육두수는 1983년 2마리에 불과하였으나 2020년 무려 36마리로 급격히 늘어났으며, 한육우 사육 가구 수는 971천호에서 2020년 93천호로, 소규모 농가들의 비중이 급격히 감소하였다. 반면, 100두 이상 규모의 한우 농가 수는 2005년 1,234호에서 2020년 7,818호로 점차 늘어나고 있다. 또한, 최근 10년간 비육농가와 번식농가의 비중이 점차 줄어들고 있으나 일관사육농가의 비중은 늘어나고 있다. 이에 2020년 기준 100두 이상의 일관사육농가는 82.5%를 차지하는 것으로 나타났다(통계청 「가축동향조사」, 이력제 자료). 이는 2009년 축사시설현대화사업에 따른 생산성 향상으로 한육우 농가의 전업화와 규모화가 빠르게 진행되고 있다는 것을 보여준다. 이러한 한육우 산업의 구조변화가 나타나면서 과거처럼 소규모 농가들의 한우 사육에 자유로운 진입과 탈퇴가 어려워져 사육두수의 등락 폭은 낮아지며 한우 사육 사이클도 크게 변화될 것으로 예상된다(전상곤 외 2020년).

이러한 측면에서 한육우의 사육두수를 보다 정밀하게 예측하기 위해 한육우 산업의 구조변화 전과 후로 변화된 사육주기를 수급모형에 반영할 필요가 있다. 따라서 이 연구는 축사시설현대화사업이 시작된 2009년을 기준으로 구조변화의 전과 후를 반영하는 변수를 생성하였다. 한육우 산업의 구조변화 변수는 2009년 이전은 0의 값을 가지고 그 이후부터는 1의 값을 기준으로 해마다 0~1 사이 값을 더해준 다음, 그에 대해 로그값을 취하였다. 이는 사인함수의 치역은 -1과 1 사이에 있으며, 대다수 농가가 전문화·규모화, 일관사육농가로 전환하면서 한육우 사육구조의 변화는 급진적으로 나타나는 것이 아니라, 시간의 흐름에 따라 점차적으로 나타난 것을 고려하여, 2009년 1값을 기준으로 전연도보다 점진적으로 증가하는 값을 부여하였다. 그리고 그에 대해 로그를 취한 이유는 축사시설현대화사업의 추진으로 인한 산업의

구조변화 효과는 전년도보다 다소 증가할 것이지만 수확체감의 법칙(the law of diminishing returns)에 따라 그 증가 폭은 점차 감소되는 현상을 반영하기 위함이다.

$$(3) \sin\left(\frac{time}{cycle} \times 2\pi\right) + \log(RGIN) + \log(S)$$

여기서 S 는 한육우 산업의 구조변화를 나타내는 변수이다.

위의 식 (1)~(3)은 한육우 사육두수의 예측력을 제고시키기 위해 사육주기를 다양하게 나타낸 함수식이며, 이제 이를 설명변수로 활용하여 다음의 세 가지 모형을 통해 향후 한육우 사육두수를 전망하고자 한다. 이와 같이 구축된 수급전망 모형은 소개체별 다양한 자료와 복잡한 연립방정식모형을 활용하지 않아도 한육우의 사육두수를 예측할 수 있다는 장점이 있다.

$$(\text{model 1}) Y = \alpha + \beta_1 \cdot \sin\left(\frac{time}{cycle} \times 2\pi\right) + \epsilon$$

$$(\text{model 2}) Y = \alpha + \beta_1 \cdot \left[\sin\left(\frac{time}{cycle} \times 2\pi\right) + \log(RGIN) \right] + \epsilon$$

$$(\text{model 3}) Y = \alpha + \beta_1 \cdot \left[\sin\left(\frac{time}{cycle} \times 2\pi\right) + \log(RGIN) + \log(S) \right] + \epsilon$$

3. 분석결과 및 예측력 검증

한육우 산업의 구조변화에 따른 사육주기의 변화를 보다 정교하게 분석하기 위해 세 가지 모형을 구축하였으며, 통계 패키지 Stata 13을 이용하여 각 모형을 추정하였다. 또한, 모형별 한육우 사육두수의 예측값의 정밀도를 검증하기 위해 각 모형의 예측력을 평가하였다. 예측력 평가방법에는 평균오차(ME: Mean of Error), 평균 절대 오차(MAE: Mean of Absolute Error), 평균 절대 백분율 오차(MAPE: Mean of Absolute Percentage Error), 평균 자승 오

차(MSE: Mean of Squared Error), 평균 자승 백분율 오차(RMSPE: Root Mean of Squared Percent Errors) 등이 있다. 이중 평균 절대 백분율오차(MAPE)가 측정단위에 영향을 받지 않기 때문에 실측값에 대한 추정 오차의 상대 크기를 측정하는데 흔히 사용된다(서홍석 외 2017). 따라서 이 연구는 MAPE 방법을 활용하여 2010년부터 2020년에 대한 모형별 예측력을 검토하였다.

$$(4) \text{MAPE} = \frac{100}{n} \sum_{t=1}^n \left| \frac{\epsilon_t}{Y_t} \right|$$

여기서 n 은 예측기간 수(표본 수), ϵ_t 는 실측값과 예측값의 차이인 오차 또는 잔차, Y_t 는 실측값을 의미한다.

모형별 회귀분석 결과는 다음의 <표 1>과 같다. 모형의 설명력(R^2 값)을 살펴보면, 사인함수만을 이용한 모형 1은 0.1512로 가장 낮은 것으로 나타났으며, 사인함수에 쇠고기 소비량의 대리변수인 1인당 실질 국민소득을 반영한 모형 2는 0.3115로 단순 사인함수만을 이용한 모형보다 다소 높은 것으로 나타났다. 사인함수에 1인당 실질 국민소득뿐만 아니라 한육우 산업의 구조변화를 반영한 모형 3의 설명력은 더욱 높아져 0.9230으로 나타나, 가장 우수한 것으로 나타났다.

세 가지 모형에서 사육주기를 나타내는 설명변수는 모두 통계적으로 유의미한 영향을 미치는 것으로 나타났다. 그 중, 사인함수에 1인당 실질 국민소득과 한육우 산업의 구조변화를 반영한 모형의 계수값은 0.5674로 가장 높은 것으로 나타나, 해당 변수가 한육우 사육두수에 미치는 영향력이 가장 높다는 것을 의미한다.

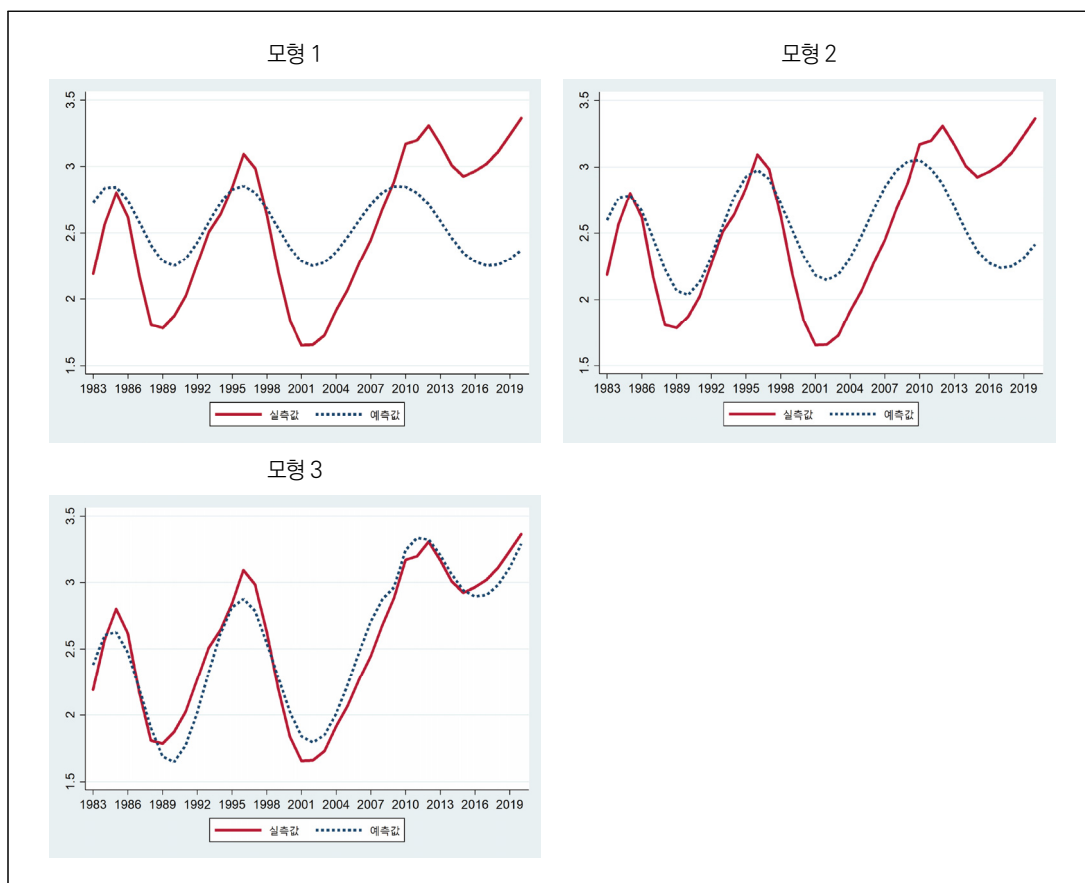
표 1. 모형별 분석결과

구분	model 1			model 2			model 3		
	Coef.	S.E	t-value	Coef.	S.E	t-value	Coef.	S.E	t-value
α	2.5511***	0.0800	31.8812	1.1323***	0.3424	3.3064	0.4648***	0.1015	4.5800
β	0.3030***	0.1010	2.7550	0.4340***	0.1030	4.2124	0.5674***	0.0269	21.0800
R^2	0.1741			0.3301			0.9250		
Adj R^2	0.1512			0.3115			0.9230		

주: * $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$

이러한 분석결과를 토대로 모형별 한육우 사육두수의 예측값을 추정하였으며, 실제 사육두수와 모형별 예측두수의 차이를 보다 직관적으로 살펴보기 위해 <그림 2>를 제시하였다. 실측값과 예측값의 오차는 모형의 설정에 따라 비교적 큰 차이를 보이는데, 모형 3에 의해 예측된 사육두수가 실제 사육두수와 가장 유사한 것으로 나타났다. 이는 사인함수에 1인당 쇠고기 소비의 동향을 반영하기 위한 1인당 실질 국민소득과 한육우 산업의 구조변화 변수를 고려한 모형 3의 예측력이 가장 우수하다는 것을 보여준다.

그림 2. 모형별 사육두수 실측값과 예측값 비교



모형별 예측값의 정밀도를 보다 자세히 살펴보기 위해 2010년~2020년 11개 표본에 대한 예측력을 검토하였다. 실제 관측값과 모형의 예측값 간 오차율을 나타내는 MAPE를 추정한

결과는 다음의 <표 2>와 같으며, 그 값은 작을수록 예측치의 정밀도가 우수한 것을 나타낸다. 2010년~2020년의 MAPE는 한육우 사육주기를 사인함수만으로 표현한 모형 1에서 0.58%~7.62%로 추정되었으며 평균 3.20%로 나타났다. 모형 2의 MAPE는 0.07%~7.58% 범위에 있으며 평균 3.19%로 추정되어, 모형 1보다 예측력이 높아진 것으로 확인되었다. 그리고 모형 3의 MAPE는 0.65%~4.49%로 추정되어 평균 2.38%로 모형 1과 모형 2에 비해 우수한 예측력을 가지는 것으로 분석되었다. 이는 향후 한육우 사육두수는 1인당 쇠고기의 소비추세와 한육우 산업의 구조변화 영향을 고려해야 한다는 것을 의미한다.

표 2. 모형별 MAPE 추정결과

연도	model 1	model 2	model 3
2010	1.85%	1.97%	2.58%
2011	0.58%	0.67%	2.68%
2012	3.38%	3.32%	1.34%
2013	0.07%	0.07%	0.71%
2014	4.06%	4.01%	3.01%
2015	6.03%	5.96%	3.45%
2016	4.03%	3.95%	1.06%
2017	1.88%	1.82%	0.65%
2018	1.11%	1.15%	2.37%
2019	4.56%	4.57%	3.86%
2020	7.62%	7.58%	4.49%
전체(2010~2020)	3.20%	3.19%	2.38%

4. 한육우 사육두수 전망결과

모형별 예측력 검증결과 모형 3의 예측력이 가장 우수한 것으로 나타나, 모형 3에 기반하여 향후 10년의 한육우 사육두수를 전망하고자 한다. 이를 위해 한육우 사육주기에 대한 가정이 필요하다. 우선 한육우 사육주기에 대한 가정을 총 3가지로 구분하였다. 첫 번째는 현재의 증가추세가 2021년에 이어 2023년까지 지속되어 향후 3년 증가한 후, 2024년부터 감소한다

고 가정하였다. 두 번째 가정은 사육두수의 증가 기간이 5년간 이어지고(2025년까지), 그 다음해인 2026년부터 감소세로 반전한다는 것이다. 다음으로 사육두수의 증가기간이 비교적 길어 7년 간 증가(2027년까지)한 후, 2028년부터 감소한다는 가정을 세웠다.

표 3. 한육우 사육주기에 대한 가정

구분	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
가정 1	증가 기간			감소 기간						
가정 2	증가 기간					감소 기간				
가정 3	증가 기간							감소 기간		

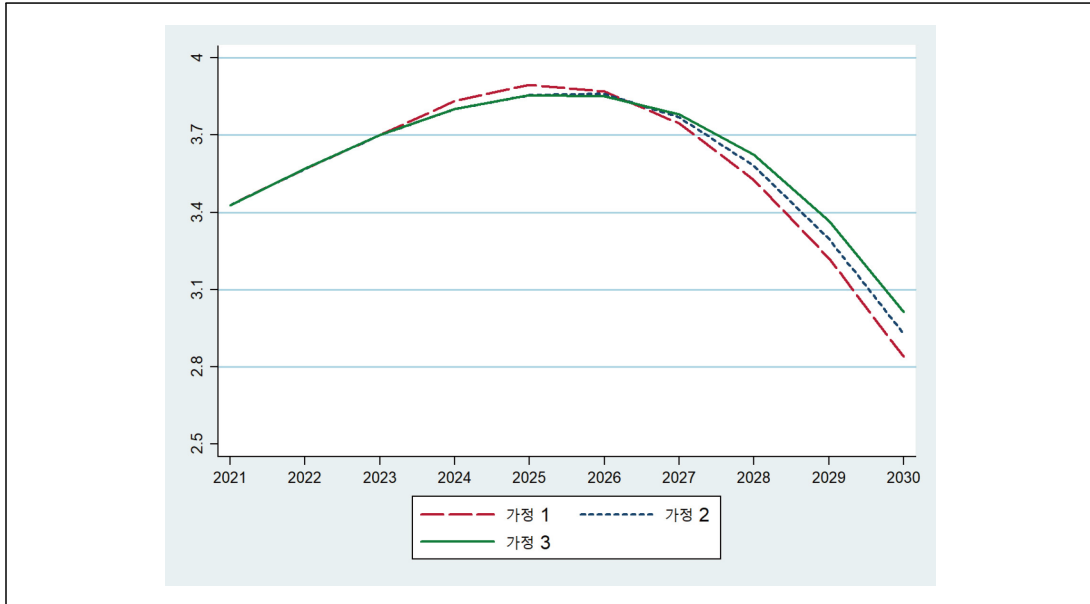
그리고 국내 1인당 쇠고기의 소비량의 대리변수인 1인당 실질 국민소득자료는 1983년~2020년의 실제 자료를 토대로 하여 단순회귀분석을 통해 미래 10년을 예측한 값을 활용하였다.

한편, 다수의 소규모 번식농가와 비육농가가 일관사육농가로 전환되고 산업의 전문화·규모화 발전으로 인한 사육구조의 변화가 한육우 산업에 미치는 영향은 한계효과의 감소로 인해 초기단계에 비해 일정 기간 이후에는 점차 감소할 것이다. 따라서 구조변화의 한계효과를 모형에 반영하기 위해 모형 3의 설정과 같은 방법을 활용하였다. 즉 사인함수의 치역범위를 고려하여 2021년부터 -1~0의 값을 부여하여 순차적으로 감소하는 추세를 나타냈으며, 이에 대한 로그값을 취하였다.

이러한 가정에 의해 미래의 사육두수를 전망한 결과는 다음의 <그림 3>과 같이 나타났다. 한육우 사육두수는 사육주기의 변화에 따라 약간의 차이가 나지만 대체로 2021년부터 점진적으로 증가하다가 2025년~2027년 사이에 점차 감소하는 추세를 보인다. 사육두수의 증가 추세는 가정 1(3년 증가, 7년 감소)에서 가장 급격한 것으로 나타났다. 반면, 사육두수가 감소할 경우에는 가정 3(7년 증가, 3년 감소)에서 가장 급격히 하락하며, 다음으로 가정 2(5년 증가, 5년 감소), 가정 1(3년 증가, 7년 감소)의 순으로 나타났다. 보다 구체적으로 설명하자면, 향후 10년간 한육우 사육두수는 전반적으로 2021년 340만두 정도에서 점차 증가하며, 2025년~2027년에 약 385만두로 최대수준에 이른 후(사육주기에 대한 가정이 1일 경우에는 2025년에 최대에 이르며, 가정 2일 경우는 2026년, 가정 3일 경우는 2027년), 다시 점진적으로 감소할 것으로 예측된다.

그림 3. 향후 한육우 사육두수 전망결과

(단위: 백만)



5. 요약 및 결론

한육우는 생태적 특성상 타 축종보다 사육기간이 비교적 길어 공급변화에 따른 가격에 반응하며 사육두수는 등락을 반복하면서 일정한 사육주기를 가진다. 한편, 최근 한우농가의 규모화·전업화, 일관사육농가의 증가 등 산업의 구조변화로 인해 기존의 사육주기는 크게 변화되고 있다. 이로 인해 한육우 사육농가는 합리적인 의사결정에 어려움을 겪고 있으며, 시장과 정부는 한육우 수급조절에 한계를 지니고 있다. 따라서 농가의 합리적인 의사결정과 효율적인 수급조절을 위해 한육우 사육두수에 대한 중장기 전망이 필요하며, 보다 정밀한 수급전망을 위해 한육우 산업의 구조변화에 따른 사육주기의 변화를 고려할 필요가 있다.

이에 이 연구는 기존의 수급모형과 달리, 한육우 산업의 구조변화 전과 후의 사육주기의 변화패턴을 사인함수를 통해 살펴보았으며, 한육우 수급모형을 단순 사인함수뿐만 아니라 사인

함수에 1인당 쇠고기 소비동향 또는 산업의 구조변화를 반영한 등의 세 가지 형태로 구축하였다. 그리고 각 모형에 의해 추정된 사육두수의 예측값과 실측값의 오차율을 분석하였으며, 모형의 예측력 검증을 통해 가장 정밀한 수급모형을 제시하여 향후 10년간의 한육우 사육두수를 예측하였다. 이를 위해 기존의 「가축동향조사」 자료와 쇠고기이력제 자료의 차이를 살펴보고 통계청의 「가축동향조사」를 재산출하여 쇠고기이력제 자료와 연계하였다.

모형의 예측력 검증결과, 사인함수에 1인당 국민소득과 한육우 산업의 구조변화를 고려한 모형의 MAPE가 평균 2.38%로 추정되어 예측력이 우수한 것으로 나타났다.

따라서 해당 모형을 활용하여 한육우 사육주기에 대한 세 가지 시나리오를 가지고 2021년~2030년의 사육두수를 예측하였다. 그 결과, 한육우 사육두수는 사육주기의 변화에 따라 약간의 차이를 보이고 있지만, 대체로 2021년 340만두에서 점진적으로 증가하여 2026년쯤 약 385만두로 정점에 이르며, 그 후부터 다시 점차 감소할 것으로 전망된다.

이 연구는 「가축동향조사」 자료가 쇠고기이력제 자료로 대체된 상황에서 이력제 자료를 기반으로 하였다는 점에서도 선행연구와 차별화된다. 또한, 이 연구에서 구축된 한육우 수급모형은 기존의 수급모형과 달리, 사육주기를 반영하는 사인함수에 1인당 쇠고기 소비량의 동향과 한육우 산업의 구조변화가 발생한 전과 후의 사육주기의 변화를 고려하였다는 점에서 연구의 가치를 지닌다. 이와 같이 구축된 수급모형은 소개체별 다양한 자료와 복잡한 계량기법을 활용하지 않아도 되기 때문에 활용도가 높을 것으로 예상되며, 이를 통해 전망한 사육두수는 한육우 시장의 수급조절에 유용한 정보가 될 것으로 기대된다.

한편, 한육우 산업의 구조변화 효과를 정밀하게 예측하기 어려워 이에 대한 시나리오가 그다지 충분하지 못한 한계가 있다. 또한, 연립방정식이 아닌 단순회귀모형을 이용하여 수급모형을 단순화한 장점이 있는 반면, 이로 인해 일부 변수를 누락할 수 있다는 단점도 존재한다. 따라서 향후 연구에서 이러한 한계점을 보완하여 보다 정밀하고 고도화된 한육우 수급모형을 개발할 필요가 있다.

참고문헌

- 김충현, 서홍석, 이형우, 김진년. 2018. “쇠고기이력제 자료를 이용한 한우 수급모형 구축.” 『농촌경제』 41(3호): 25-50.
- 서홍석, 김충현, 이형우, 김진년. 2017. 「쇠고기이력제를 활용한 축산관측 고도화 연구」 한국농촌경제연구원.
- 전상곤, 2012. “싸인함수를 이용한 한우 사육두수 예측.” 『농업생명과학연구』 46(6): 197-205.
- 전상곤, 신영식, 박동주. 2020. 「한우 사육구조 변화 및 수급영향 분석」 한우자조금관리위원회.
- 조재성, 송우진, 김태우. 2013. 「한육우 단기관측모형 개발 연구」 한국농촌경제연구원.
- 조재성, 이형우, 서강철. 2017. “한우 수급 및 가격 전망모형 구축 연구.” 『농업경영·경영정책연구』 44(2): 210-235.
- 축산물품질평가원. <<https://www.ekapepia.com>>.
- 통계청. 가축동향조사. <<https://kosis.kr>>.
- 한국육류유통수출협회. 통계자료실. <http://www.kmta.or.kr/kr/data/stats_spend.php>.