



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(Climatic 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 불러옴으로서 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: 농사료: 농촌진흥청, 최병욱 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 분석함으로써 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 기계 학습이라고 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 종속변수와의 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.

$$\begin{aligned} (4) \text{ Almon}(k) : B(L^{\frac{1}{m}}; \theta) X_t^{(m)} &= \sum_{i=0}^{m-1} w_i X_t^{(m)} \\ &= \sum_{i=0}^{m-1} (\theta_0 + i \theta_1 + i^2 \theta_2 + \dots + i^k \theta_k) X_t^{(m)} \\ &= \theta_0 \sum_{i=0}^{m-1} X_t^{(m)} + \theta_1 \sum_{i=0}^{m-1} i X_t^{(m)} + \dots + \theta_k \sum_{i=0}^{m-1} i^k X_t^{(m)} \end{aligned}$$

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.

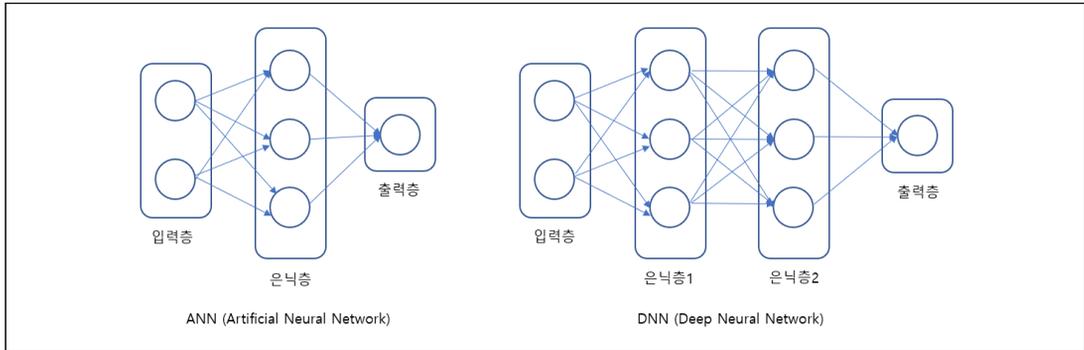
3.1.3 딥러닝 학습 모형

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 증가 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 가치분소득, 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 추정된 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 (°C)	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
최댓값	12986.00	18.97	1089.06	1.62	11.00	29.90	43.90	47.70
최솟값	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

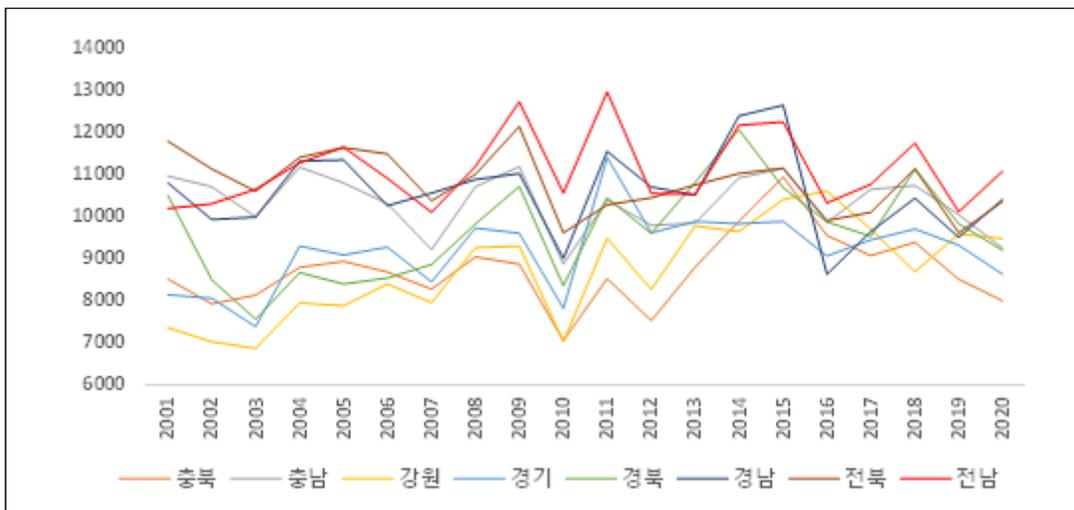
주: 1) 기후 변수는 연도 및 지역 별 생육 기간(8월~11월) 중 일 평균 값(일조시간 및 강풍일수는 누적 값)을 나타냄.

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 단수 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 되어 있으나 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 (°C)	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 (딤러닝: 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	

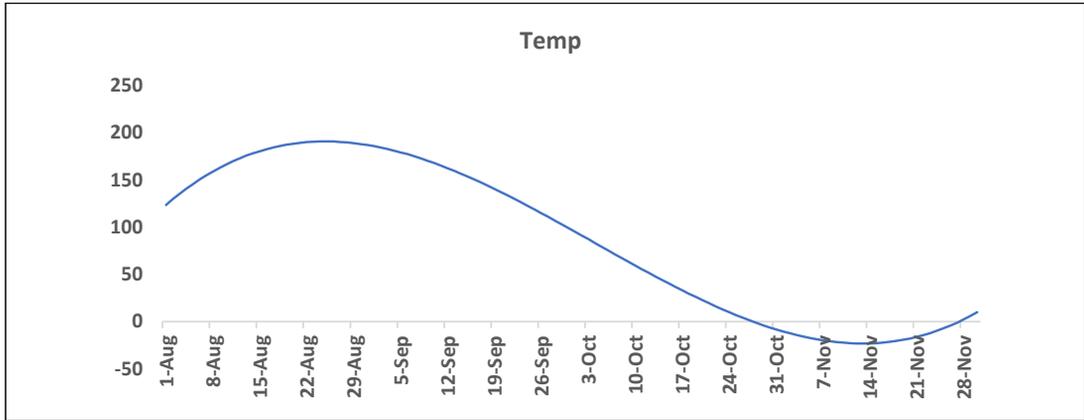
주: 1) *, **, ***는 추정된 계수가 각각 10%, 5%, 1% 수준에서 유의하다는 것을 의미함.

2) 기온의 경우 모형 4에서는 생육 기간 동안의 일별 자료 가중합(MIDAS), 모형 5에서는 기온을 설명변수로 단수를 딤러닝 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 추정한 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 추정된 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 구증을 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 추정한 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. △가격 (모형 11-1,2의 균형 가격 활용)		모형 12-3. △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 조사 통계연보 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>>.