

농지 휴경면적 예측 및 활용효과에 대한 연구

임청룡*, 최돈우**, 이항아**

*한국농어촌공사 농어촌연구원

**경상북도 농업기술원

e-mail: chrylims@ekr.or.kr

A Study on Idle Farmland Forecast and Utilization Effect

Cheong-Ryong Lim*, Don-Woo Choi**, Hang-Ah Lee**

*Rural Reserch Institute Korea Rural Community Corporation

**Gyeongsangbuk-do Agricultural Research & Extension Services.

요약

이 연구에서는 토양DB, 농업총조사 원자료 등으로 구성된 농업빅데이터로 회귀분석, 의사결정나무, 배깅, 부스팅, 랜덤 포레스트 등 모형들을 활용하여 읍면단위 휴경면적 예측모형을 추정하였으며, 분석결과를 요약하면 다음과 같다. 첫째, 논에 대한 휴경면적 예측모형 예측력 비교에서 랜덤포레스트 추정결과로부터 도출된 평균예측오차가 제일 작게 나타나 예측력이 가장 훌륭한 것을 알 수 있었다. 둘째, 밭에 대한 휴경면적 예측모형 예측력 비교에서 역시 랜덤포레스트 추정 결과로부터 도출된 평균예측오차가 제일 작게 나타나 예측력이 가장 훌륭한 것을 알 수 있었다. 셋째, 예측된 휴경면적에 지역별 유망작물 도입을 가정한 시나리오 하에서, 충주 고구마일 경우 총 2,438.6백만원, 제천 홉일 경우 총 2,159.3백만원, 영양 고추일 경우 총 2,942.7백만원, 문경 오미자일 경우 총 2,239.6백만원, 고흥 유자일 경우 총 6,698.1백만원, 보성 참다래일 경우 총 4,596.5백만원인 것으로 나타났다.

1. 서론

도시화로 인한 농촌지역 인구 감소 및 도농소득격차 증대 등 요인으로 휴경면적이 지속적으로 발생하고 이러한 휴경으로 인해 지역생산성 저하와 농가소득 감소 등의 주요 요인으로 작용하여 농촌의 지속가능성 유지에 적신호가 되고 있으며 궁극적으로는 토지 이용의 비효율성을 야기하고 있다.

농지 휴경 관련하여 자료의 유형과 특징에 따라 합동 OLS, 토빗모형, 변화할당효과, 다중로짓모형의 다양한 방법이 시도되었다. 종속변수가 휴경여부(박인환 외, 2006; Hualin et al, 2017)이거나 용도별 토지 이용 변화(Thomas et al, 1999; Ahn et al, 2001)인 경우, 농업인의 상대적인 토지 이용 의사결정에 관한 행태방정식을 확률로써 나타낼 수 있으며 주로 로짓모형으로 모형화하여 분석하였다. 경지면적과 같이 특정한 값 이상에서만 관찰되는 종속변수는 일반회귀모형의 정규분포 가정을 부합하지 않으므로 토빗 모형을 적용하였다(이항미 외, 2015). 패널 자료를 활용할 경우 지역별 특성을 고려하기 위하여 주로 고정효과모형을 사용하나, 패널 자료임에도 불구하고 자료의 변동이 완만하고 지역평균이므로 합동 OLS 모형으로 추정하여 이분산성 발생 소지를 해소하는 연구가 수행되었다(정준호 외, 2012). 또한 특정 지역 및 규모별 경지 면적 변화를 추정하기 위하여 변화할당효과를 고려할 수 있고(장우석 외, 2008) 사회경제적 변수로는 농업인의 연령, 성별, 교육수

준, 농업종사경력, 농기계 사용여부, 인구밀도, 1인당 지방세 등이 있음. 또한 정부의 관련 정책에 대한 농업인의 입장 및 태도를 조사하여 반영하기도 하였다(Hualin et al, 2017).

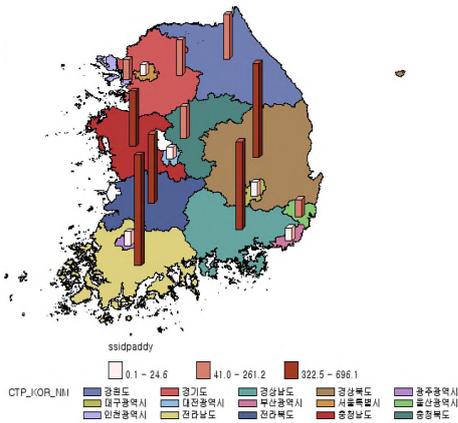
선행연구를 살펴본 결과, 도시 개발이나 영농 조건 변화 등으로 토지 이용 변화가 예상되는 시점에서 그 결정 요인을 파악하고자 하는 목적은 모두 비슷하나 연구 대상인 토지의 영역을 어느 정도로 한정하느냐에 따라 국내와 국외 간 차이가 있었으며, 자료의 유형과 특징에 따라 합동 OLS, 토빗모형, 변화할당효과, 다중로짓모형의 다양한 방법이 시도되었다. 따라서 이 연구에서는 농업 빅데이터를 활용하여 다양한 농지 휴경면적 예측모형을 분석하여 최적의 모형을 선택함과 동시에 시군 유망작물 재배 시 발생가능 효과에 대해 분석하고자 한다.

2. 지역별 휴경지 현황

2.1 시도별 논 휴경 현황

2015년 논에 대한 휴경면적은 0~40ha미만, 40~300ha미만 및 300ha이상 등 3개 그룹으로 구분하면 다음과 같다. 상위그룹에는 전라남도, 경상북도, 경상남도, 전라북도, 충청남도 등 5개 도가 포함되고, 중간그룹에는 강원도, 충청북도, 경기도, 인천광역시, 울산광역시가 포함되며, 하위그룹에는 대구광역

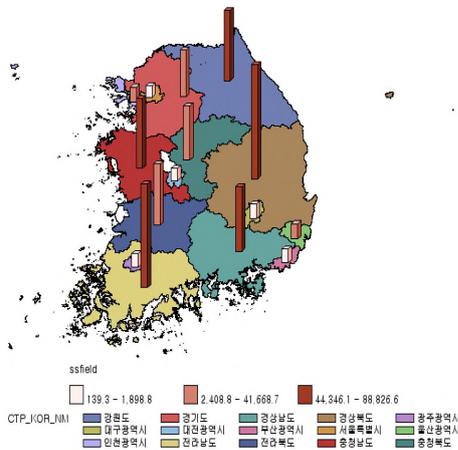
시, 부산광역시, 광주광역시, 대전광역시 및 서울특별시가 포함된다.



[그림 1] 시·도별 논 휴경면적

2.2 시도별 밭 휴경 현황

2015년 논에 대한 휴경면적은 0~2,000ha미만, 2,000~44,000ha미만 및 44,000ha이상 등 3개 그룹으로 구분하면 다음과 같이 나타났다. 상위그룹에는 경상북도, 전라남도, 충청남도, 강원도, 경상남도 등 5개 도가 포함되고, 중간그룹에는 전라북도, 충청북도, 경기도, 인천광역시, 울산광역시가 포함되며, 하위그룹에는 대구광역시, 부산광역시, 광주광역시, 대전광역시 및 서울특별시가 포함되었다.



[그림 3] 시·도별 밭 휴경면적

3. 분석모형 및 추정방법

3.1 분석모형

이 연구에서는 휴경농지 예측모형을 구축함에 있어서 방법론적 측면에서 다양한 연구들에 대한 고찰을 수행하였으며, 농가의 측면에서 내부적 요인과 외부적 요인들로 변수들을

구성하였다. 또한 분석에 활용될 데이터의 성격에 부합하고 논과 밭의 휴경 결정에 영향을 미칠 수 있는 변수들로 분석모형을 다음과 같은 함수형태로 각각 설정하였다.

$$\begin{aligned}
 PL_i &= F(land_i, agri_i, farm_i, region_i) + \epsilon_i \\
 FL_i &= F(land_i, agri_i, farm_i, region_i) + \theta_i
 \end{aligned}
 \tag{1}$$

식(1)에서 PL_i 는 i 읍·면의 논에 대한 휴경면적이고, FL_i 는 i 읍·면의 밭에 대한 휴경면적이다. $land_i$ 는 토양특성 변수들에 대한 벡터이고, $agri_i$ 는 농업특성 변수들에 대한 벡터이며, $farm_i$ 는 농가특성 변수들에 대한 벡터이다. $region_i$ 은 지역변수들에 대한 벡터이고, 논 휴경모형에 대한 잔차항 ϵ_i 와 밭 휴경모형에 대한 잔차항 θ_i 와 같다고 설정하였다.

3.2 추정방법

이 연구에서 읍면별 휴경면적 분석 및 예측을 위해 빅데이터 활용하여 수행된 농업 및 비농업분야 추정방식을 검토하여 최종적으로 회귀모형, 의사결정나무 및 예측력 향상을 위해 앙상블모형을 활용하였다.

3.2.1 의사결정나무(Decision Tree)

의사결정나무는 주어진 입력 값에 대하여 출력 값을 예측하는 모형으로 분류나무와 회귀나무모형이 있다. 의사결정나무는 지도학습기법으로 각 변수의 영역을 반복적으로 분할함으로써 전체 영역에서의 규칙을 생성하는 것이다. 훈련자료를 (x_i, y_i) , $i = 1, \dots, n$ 로 나타내고 여기서 $x_i = (x_{i1}, \dots, x_{ip})^T$ 이다. 전체 영역을 M 개의 영역 R_1, \dots, R_M 으로 나누고 각 영역에서 상수 값 c_m 으로 예측하는 다음과 같은 나무모형을 고려할 수 있다.(Breiman et al., 1984).

3.2.2 배깅(Bagging)

배깅은 불안정한 예측모형에서 불안정성을 제거함으로써 예측력을 향상시키기 위하여 개발되었으며, 학습 자료의 작은 변화에 의해 예측모형이 크게 변하는 경우 그 학습방법은 불안정하다고 할 수 있다. 불안정성은 통계학의 로버스트(robust)성질과 유사하며 배깅은 주어진 자료에 대하여 여러 개의 붓스트랩(bootstrap) 자료를 생성하고, 각 붓스트랩 자료에 예측모형을 만든 후 결합하여 최종 예측모형을 만드는 방법이다(Breiman, 1996).

3.2.3 부스팅(Boosting)

부스팅의 기본 아이디어는 예측력이 약한 모형(weak learner)들을 결합하여 강한 예측모형을 만드는 것이며, 여기서 약한 예측모형이란 랜덤하게 예측하는 것보다 약간 좋은 예측력을 지닌 모형을 말하고 강한 예측모형이란 예측력이 최적에 가까운 예측모형을 가리킨다. 자료분석을 위해 제안된 최초의 부스팅 알고리즘은 이진 분류문제에서 Freund와 Schapire(1997)에 의해서 개발된 AdaBoost(Adaptive Boost) 알고리즘이다. AdaBoost 알고리즘은 매 반복마다 오분류된 관측치의 가중치는 증가시키고 정분류된 관측치의 가중치는 감소시키면서 예측모형을 만들어 간다.

	시설기타	9.901	34.864	8.360	27.117
	주요과수	81.017	143.638	48.038	100.507
	특작	22.514	42.061	25.345	37.541
가구	1인가구	104.845	72.145	104.046	72.135
	다인가구	407.804	246.767	408.175	249.206
연령대	40대	100.742	78.825	100.974	78.798
	50대	230.790	145.433	231.025	146.508
	60대	278.320	167.286	278.423	169.009
	70대이상	345.123	209.422	344.872	211.905
주요업종	농업	807.095	483.316	806.603	488.565
	임업	0.963	3.982	0.949	3.929
	어업	7.908	30.050	7.963	30.295
	제조	17.790	27.908	17.943	27.731
	건설	12.504	13.780	12.589	13.785
	도소매	14.445	18.272	14.389	18.090
	숙박음식	17.949	19.536	17.885	19.375
	가타	113.974	116.392	115.113	116.589
	1인당 경지면적	1.212	0.649	1.213	0.649
	시·군GRDP	2,296.74	2.79	2,884.90	2.78

3.2.4 랜덤 포레스트(Random Forest)

Breiman(2001)에 의해서 개발된 랜덤 포레스트는 의사결정 나무의 특징인 분산이 크다는 점을 고려한 것으로서 배깅과 부스팅보다 더 많은 무작위성을 주어 약한 학습기들을 생성한 후 이를 선형결합하여 최종 학습기를 만든다. 랜덤 포레스트는 무작위성을 최대로 주기 위해 붓스트랩과 더불어 입력변수들에 대한 무작위 추출을 결합한다. 따라서 서로 연관성이 약한 학습기를 여러 개 만들어 내는 기법이라고 할 수 있다.

3.3 자료

휴경 농지 예측을 위해 경사도, 배수등급, 농작형태, 가구, 연령대, 주요업종, 1인당 평균재배면적 및 시군별 GRDP 관련 변수들을 중심으로 추정하였으며, 2010년과 2015년 변수들에 대한 기초통계량은 다음과 같이 나타났다.

[표 3] 분석 자료에 대한 기초통계량
(단위: ha, 십억원)

구분	2010년		2015년		
	평균	표준편차	평균	표준편차	
경사도	2-7%	293.799	265.742	298.088	283.796
	7-15%	264.779	250.984	263.933	252.021
	15-30%	111.862	329.439	112.187	325.862
	30-60%	105.433	295.037	103.418	291.143
	60-100%	218.929	1846.850	212.705	1820.470
배수등급	매우불량	53.239	144.846	56.747	155.999
	불량	29.834	119.279	32.137	125.552
	약간불량	299.230	372.344	302.264	376.100
	약간양호	256.174	245.060	253.856	245.780
	양호	431.701	857.774	431.805	849.492
농작형태	논벼	433.180	432.747	387.625	415.920
	식량	84.414	119.439	78.701	113.558
	노지채소	72.751	130.148	65.784	118.422
	기타노지	7287.280	84823.67	9211.250	76765.83
	시설채소	24.076	62.255	19.404	49.545

4. 휴경지 예측결과 및 활용효과

4.1 예측모형별 예측결과 비교

이 연구에서는 2010년 농업총조사 원자료를 활용하여 예측모형을 추정하고, 2015년 농업총조사 원자료를 활용하여 예측력을 점검해 보았다.

논 휴경면적에 예측에 있어서 그림을 활용한 예측력 비교보다 객관적인 통계량인 평균오차제곱에 대한 비교분석결과 역시 랜덤포레스트가 3.743로 가장 작게 나타나 예측력이 가장 좋은 것을 알 수 있었다.

밭 휴경면적에 예측에 있어서 그림을 활용한 예측력 비교보다 객관적인 통계량인 평균오차제곱에 대한 비교분석결과 역시 랜덤포레스트가 16.428로 가장 작게 나타나 예측력이 가장 좋은 것을 알 수 있었다.

4.2 휴경지 활용

이 연구에서는 지역별 휴경면적을 예측하고 충청북도, 경상북도, 전라남도의 2개 시군 총 6개 지역유망작물 조사결과를 활용하여 휴경농지에 지역별 유망작물을 도입할 경우 밭 생활 수 있는 효과를 추정해보았다.

이러한 지역별 유망작물 경영성과를 지역별 휴경지에 적용한 유망작물 도입효과를 살펴보면, 충주 고구마일 경우 총 2,438.6백만원, 제천 홉일 경우 총 2,159.3백만원, 영양 고추일 경우 총 2,942.7백만원, 문경 오미자일 총 2,239.6백만원, 고흥 유자일 경우 총 6,698.1백만원, 보성 참다래일 경우 총 4,596.5백만원인 것으로 나타났다.

5. 요약 및 결론

이 연구에서는 토양DB, 농업총조사 원자료 등으로 구성된 농업빅데이터로 회귀분석, 의사결정나무, 배깅, 부스팅, 랜덤포레스트 등 모형들을 활용하여 읍면단위 휴경면적 예측모형을 추정하였으며, 분석결과를 요약하면 다음과 같다.

첫째, 논에 대한 휴경면적 예측모형 예측력 비교에서 랜덤포레스트 추정결과로부터 도출된 평균예측오차가 제일 작게 나타나 예측력이 가장 훌륭한 것을 알 수 있었다.

둘째, 밭에 대한 휴경면적 예측모형 예측력 비교에서 역시 랜덤포레스트 추정결과로부터 도출된 평균예측오차가 제일 작게 나타나 예측력이 가장 훌륭한 것을 알 수 있었다.

셋째, 예측된 휴경면적에 지역별 유망작물 도입을 가정한 시나리오 하에서, 충주 고구마일 경우 총 2,438.6백만원, 제천 홉일 경우 총 2,159.3백만원, 영양 고추일 경우 총 2,942.7백만원, 문경 오미자일 경우 총 2,239.6백만원, 고흥 유자일 경우 총 6,698.1백만원, 보성 참다래일 경우 총 4,596.5백만원인 것으로 나타났다.

이러한 분석결과로부터 다음과 같은 시사점을 제시하고자 한다.

첫째, 다양한 요인으로 인해 발생하는 휴경농지에 대한 정확한 예측을 통해 휴경농지 규모를 객관적이고 과학적으로 파악할 필요가 있다.

둘째, 지역별 휴경농지를 적정하게 활용할 경우 상당히 큰 경제효과를 유발하여 지역경제 활성화를 도모할 수 있다.

셋째, 휴경 규모에 대한 보다 정확한 파악을 위해 휴경과 관련된 보다 많은 변수들의 정보를 구축이 필요하다.

이 연구에서는 농업빅데이터를 활용하여 읍면단위 휴경면적을 예측하고 휴경농지 활용 시 효과를 특정 시나리오 하에서 추정하였지만 자료의 한계로 정확도가 다소 저하될 수 있다. 보다 정확한 휴경면적 예측을 위해서는 농가단위 또는 필지단위로 구성된 빅데이터 분석이 필요할 것으로 여겨진다.

이 연구는 농촌진흥청 연구과제(과제번호: PJ015331062020)의 지원으로 수행되었음.

참고문헌

[1] 김이환·이지은·김현서·장종선, 「2017 맞춤형 기상기후 빅데이터 서비스 기반 구축사업」, 기상청, 2017.
 [2] 김한호·이태호·남대희·우아미·민진경, 「농업경영체 등록 정보를 활용한 농가 유형별 특성 분석 및 DB 구축 개선 방안」, 농림축산식품부, 2015.
 [3] 박인환·장갑수·서동조, 「다중 로지스틱 모형에 의한 농경지 휴경잠재성 분석」, <환경영향평가>, 15(1):

pp53-65, 2006.
 [4] 부천시청, 「감염병 발생 위험도 예측분석」, 2018.
 [5] 안경아·정근오, 「기상빅데이터를 활용한 콩 수확량 모델 개발과 농업수입보장보험개선방안」, 제주연구원, 2017.
 [6] 이석민, 「인문사회과학을 위한 빅데이터 분석방법론」, 윤성사, 2019.
 [7] 이향미·고종태, 「중소도시 근교 농촌지역 경지면적 결정 요인 분석」, <농촌계획>, 21(2): pp1-10, 2015.
 [8] 임청룡·임상봉·이향미, 「농업 빅데이터 기반 작부체계 적지적작의 생산성 예측」, 한국농어촌공사, 2018.
 [9] 장우석·정남수·박기욱, 「변화할당효과를 고려한 논 면적 예측 모형의 개발」, <한국농공학회논문집>, 50(3):pp83-89, 2008
 [10] 정준호·김광배, 「기후변화가 강원도의 농업에 미친 경제적 효과 분석」, <한국경제지리학회지>, 15(2): pp192-205, 2012.
 [11] 한국수자원공사, 「머신러닝 기반 정수장 원수 냄새물질 예측을 통한 수돗물 품질 향상」, 2018.
 [12] 행정안전부·한국정보화진흥원, 「상수도 누수 예측 모델 매뉴얼」, 2018.
 [13] 행정안전부·한국정보화진흥원, 「지방세 체납 징수 매뉴얼」, 2018.
 [14] Breiman L, Fredman J, Olshen R and Stone C., Classification and regression trees, Wadsworth, 1984.
 [15] Breiman L., Bagging predictors. Machine Learning, 26: 123-140, 1996.
 [16] Breiman L., Random Forests. Machine Learning. 45: 5-32, 2001.
 [17] Freund Y. and Schapire R, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and System Science, 55: 119-139, 1997.
 [18] Hualin Xie·Lingjuan Cheng·Tianguai Lv., Factors Influencing Farmer Willingness to Fallow Winter Wheat and Ecological Compensation Standards in a Groundwater Funnel Area in Hengshui, Hebei Province, China, Sustainability, 9:1-18, 2017.
 [19] SoEun Ahn, Andrew J. Plantinga, Ralph J. Alig., Historical Trends and Projections of Land Use for the South-Central United States, U.S. Department of Agriculture Forest Service, Pacific Northwest Research Station, 2001.
 [20] Tomas E. Mauldin, Andrew J. Plantinga, Ralph J. Alig., Determinant of Land Use in Maine with Projections to 2050, NJAF, 16(2):82-88, 1999.