

기후재분석자료 시계열모델링을 통한 곡물수확량 예측실험: 1960-2009년 미국 아이오와주 사례*

김나리** · 이양원***

Experimental Predictions of Crop Yields Using Time-Series Modeling of Climate Reanalysis Data: A Case of Iowa, USA, 1960-2009*

Nari Kim** · Yang-Won Lee***

요약 : 온난화와 같은 전지구적 변화는 기온과 강수 등 기상요소에 직접적으로 반영되어 곡물 수확량의 변화를 가져온다. 기후변화 시나리오에 기초한 선행연구들에서는 GCM (general circulation model)의 공간해상도 문제로 인하여 상세한 모의가 어려웠고, 시계열통계법을 활용한 연구들에서는 기후요소를 통합하여 수확량을 예측한 사례가 매우 드물었다. 이에 본 연구에서는 상세화된 기후재분석자료의 시계열모델링을 통하여 옥수수과 콩의 수확량 예측실험을 수행하였다. 미국 아이오와 주의 99개 카운티를 대상으로 1960년부터 2009년까지 50년간의 고해상도 기후재분석자료와 정부통계 수확량 DB를 구축하고, 시계열통계법인 VAR (vector autoregression)와 ARIMA (autoregressive integrated moving average)를 이용하여 다음해 수확량 예측실험을 10개 연도에 대해 수행하여 예측력을 평가하였다. VAR는 16-18%, ARIMA는 11-14%의 오차율로 다음해의 수확량을 예측할 수 있는 것으로 집계되었으며, 옥수수의 경우 표토의 산성도, 심토의 점토와 나트륨 함유량 등의 토양특성이 실제 수확량 및 예측정확도에 영향을 미치는 것으로 나타났다.

주요어 : 곡물수확, 시계열모형, 기후변동, 토양특성

Abstract : Global warming can bring about changes in crop yield by directly affecting meteorological parameters such as temperature and precipitation. Previous studies based on the climate change scenarios had difficulties in detailed simulation owing to the problem of spatial resolution of GCM (general circulation model). The researches using time-series modeling rarely incorporated climate factors into the crop yield prediction. In this study, we conducted experimental predictions of corn and soybean yields by time-series modeling of downscaled climate reanalysis data. We built a database for the climate dataset and governmental yield statistics for the period of 1960-2009 for the 99 counties in Iowa State. Then we carried out 10 sets of the next-year prediction for corn and soybean yields using VAR (vector autoregression) and ARIMA (autoregressive integrated moving average) methods. The VAR and ARIMA were able to predict the next-year yields with the errors of 16-18% and 11-14%, respectively. In addition, soil properties such as topsoil pH, subsoil clay fraction and subsoil sodicity were closely related to the actual yields and the prediction accuracies.

Key Words : Crop yield, Time-series model, Climate variability, Soil properties

*이 논문은 한국연구재단의 글로벌연구실사업(NRF-2011-0021927)의 지원을 받아 수행된 연구임.

**부경대학교 지구환경시스템과학부 박사과정(Ph. D. Student, Division of Earth Environmental System Science, Pukyong National University, kim.nari13@gmail.com)

***부경대학교 공간정보시스템공학과 부교수(Associate Professor, Department of Spatial Information Engineering, Pukyong National University, moxconfi@pknu.ac.kr)

I. 서론

18세기 산업혁명 이후 지속된 산업화와 인간활동으로 인하여 온실가스 배출이 증가하면서 온난화를 비롯한 전지구적인 기후변화가 야기되고 있다. 이러한 변화는 기온과 강수 등 기상요소에 직접적으로 반영되어 곡물 수확량의 변화를 가져오게 되는데(송용호 등, 2013; Meza and Silva, 2009), 특히 기온은 곡물 수확량에 큰 영향을 미치는 것으로 알려져 있다. 곡물들은 저마다 생육시기에 적합한 기온이 있으며, 적정 기온 범위를 벗어난 저온 또는 고온이 지속될 경우에는 생육의 저조 및 결실 불량 등으로 인하여 곡물 수확이 상당히 감소할 수도 있다(심교문 등, 2013). 우리나라는 쌀을 제외한 주요 곡물에 대한 수입의존도가 높기 때문에, 국내의 곡물 가격은 주요 생산국의 수확량 증감에 민감하게 반응하므로 이에 대한 분석과 예측이 매우 중요하다.

곡물수확량 분석에 관한 연구는 (1)과거 또는 올해 수확량 추정, (2)다음해 또는 미래 수확량 예측이라는 두 가지 유형으로 나누어볼 수 있다. 대부분의 연구들은 위성영상과 기상관측자료를 활용하여 과거 또는 올해 수확량을 추정하는 부류에 속한다(홍석영 등, 2012; 김나리 등, 2014; 홍석영 등, 2015; Allen, 1990; Labus *et al.*, 2002; Ferencz *et al.*, 2004; Doraiswamy *et al.*, 2005; Prasad *et al.*, 2006; Ren *et al.*, 2008; Atzberger, 2013). 한편 다음해 또는 미래 수확량의 예측에 관한 연구는 미래 시점을 대상으로 하기 때문에 위성영상이나 기상관측자료를 사용할 수 없고, GCM (general circulation model)의 기후예측자료라는 제한된 정보에 의존할 수밖에 없다. 이 연구들은 크게 기후변화 시나리오에 기초한 유형과 시계열통계법을 활용한 유형으로 분류될 수 있다. 기후변화 시나리오에 기초한 연구들 중에서 Meza and Silva (2009)는 GCM 기후자료와 DSSAT (Decision Support System for Agrotechnology Transfer) 곡물모형을 이용하여 칠레의 기온과 강수 변화가 저울밀과 옥수수 수확량의 감소로 이어질 것으로 예측하였다. 송용호 등 (2013)의 연구에서는 과거 20년 자료와 미래 20년 모의를 토대로 중국과 미국의 곡물별 생산취약성을 분석한 결과, 과거에 밀은 미국이 유리하고 옥수수는 중국이 유리하였지만, 미래에는 두 곡물 모두 미국이 생산에 유리할 것으로 전망하였다. 또한 시계열통계법을 활용한 연구에서는 Pease *et al.* (1993), Amin *et al.* (2014), 그리고 Choudhury and Jones (2014)가 ARIMA(autoregressive integrated moving average)를 적용하였으나 기후요소를 설명변수로 사용하지 않는 단변량(univariate) 시계열분석에 그쳤다. 이후 Farook

and Kannan (2016)가 VAR (vector autoregression)를 이용하여 최고기온, 최저기온, 강수량을 설명변수로 하는 다변량(multivariate) 시계열분석을 통해 인도의 쌀 수확량을 모의하였고, Paul and Sinha (2016)가 기온을 설명변수로 하는 ARIMA 모형을 통하여 인도의 밀 수확을 예측하는 등, 최근에 야비로소 기후자료를 설명변수로 활용한 시계열모형이 곡물수확 예측에 도입되기 시작하였다.

기후변화 시나리오에 기초한 선행연구들에서는 GCM의 공간해상도(100-200km) 문제로 인하여 상세한 모의가 어려웠고, 시계열통계법을 활용한 연구들에서는 기후요소를 통합하여 수확량을 예측한 사례가 매우 드물었다. 이에 본 연구에서는 상세화된 기후재분석자료의 시계열모델링을 통하여 옥수수와 콩의 수확량 예측실험을 수행하고자 한다. 연구지역으로는 미국 아이오와 주를 선정하였는데, 그 이유는 옥수수와 콩의 주요 생산국인 미국은 수확량의 정부통계와 기후재분석자료가 장기시계열로 구축되어 있기 때문이다. 아이오와 주를 대상으로 하여서는 과거시점의 수확량 추정은 많이 수행되었으나(김나리 등, 2014; Steinwand *et al.*, 1996; Prasad *et al.*, 2006), 시계열통계법으로 다음해 수확량을 예측 및 비교분석한 연구는 아직 보고된 바 없다. 이에 본 연구에서는 미국 아이오와 주의 99개 카운티를 대상으로 1960년부터 2009년까지 50년간의 고해상도(4km) 기후재분석자료와 정부통계 수확량 DB를 구축하고, 시계열통계법인 VAR와 ARIMA를 이용하여 다음해 수확량 예측실험을 10개 연도에 대해 수행하여 예측력을 평가하며, 그 결과를 비교 분석하고자 한다. 또한 카운티별 10년간의 수확량 예측에 있어 정확도의 패턴이 어떠한지 토양의 물리·화학적 특성과 재배적합도를 기준으로 분석하여 제시하고자 한다.

II. 자료와 방법

1. 기후 및 수확량 자료

대표적인 미국의 고해상도 기후재분석자료는 PRISM (Parameter-elevation Relationships on Independent Slopes Model) Climate Group에서 제공하고 있다(<http://www.prism.oregonstate.edu>). PRISM 내삽법은 지상관측과 수치표고모형(digital elevation model)을 활용하여 지형효과를 고려한 것으로서(Daly *et al.*, 2008), 미국 전역에 대해 4km 해상도의 강수, 최고기온, 최저기온, 평균기온, 평균이슬점온

도, 최소포화수증기압, 최대포화수증기압 등의 자료가 존재한다. 이중에서 옥수수과 콩의 수확량에 영향을 미치는 것으로 알려져 있는 강수, 최고기온, 최저기온, 평균기온, 평균이슬점온도(Peng, 1995)의 5개 변수에 대하여 옥수수와 콩의 생육기간(5-9월) 동안의 자료를 평균화하여 사용하였다. 곡물 수확량 자료는 미국농무성(United States Department of Agriculture: USDA) 농업통계청(National Agricultural Statistics Service: NASS)에서 연도별, 카운티 단위로 제공되기 때문에, 격자형태의 5개 기후변수는 구역연산(zonal operation)을 통해 카운티 단위로 재집계하여 수확량과 대응시켰다.

그림 1은 1960-2009년 옥수수와 콩의 수확량 및 기후변수들의 시계열 분포를 나타낸 것이다. 옥수수와 콩의 수확량은 전반적으로 증가하는 추세이지만, 이례적으로 1988년과 1993년에 수확량이 급격히 감소했음을 알 수 있다. 1988년은 강수가 매우 적고 최고기온과 평균기온이 평년에 비해 매우 높게 나타났는데, 이는 1988년 미국 중서부 지역에서 발생한 극심한 가뭄을 반영하는 것으로 보인다. 또한 1993년에는 강수량이 급격히 증가함을 볼 수 있는데, 실제로 1993년 중서부 지역에서는 대홍수가 발생한 바 있다. 이처럼 곡물 수확량은 기상이변의 직접적인 타격을 받기도 한다.

표 1은 1960-2009년의 기간 동안 옥수수 및 콩 수확량과 기후변수와의 상관계수를 보여준다. 옥수수 수확량은 최고기온에 대해 상당한 음의 상관관계를 가지는데, 이는 여름의 무더위가 심하면 옥수수 수확량에 부정적인 영향을 미칠 수 있는 것으로 해석된다. 그러나 전반적으로 기후변수들과 수확량의 선형상관이 매우 뚜렷하지는 않은 것으로 나타났는데, 이는 50년의 기간 동안 수확량은 꾸준히 증가해온 반면 기후변동은 일정하지 않은 패턴으로 나타났기 때문으로 사료된다. 이처럼 선형성이 뚜렷하지 않은 제한된 정보만으로 시계열 모델링을 수행할 경우에는 일반회귀법을 사용하기보다는 시간의 흐름에 따른 반응변수의 변화경향이 반영되는 자기회귀법을 다변량 설명변수와 결합함으로써 예측력을 보다 향상시킬 수 있다.

2. 시계열모형

시계열모형에서는 현재의 값은 과거 값들이 가지는 규칙성에 의해 재현되며, 이러한 규칙성이 미래에도 유지된다는 가정을 기반으로 미래를 예측하는 방법이다. VAR는 Sims (1980)가 제안한 것으로 다변량 자기회귀 모형을 다변량으로 확장시킨 모형이다. 전통적인 회귀모형은 독립변수 간의 상

표 1. 옥수수와 콩의 수확량과 기후변수와의 상관계수(1960-2009년)

기후변수	옥수수 수확량	콩 수확량
강수	0.021	0.014
최고기온	-0.289	-0.135
최저기온	-0.068	0.112
평균기온	-0.198	-0.019
평균이슬점온도	0.009	0.175

주: 모든 경우에서 p-value < 0.01

관성을 통해 종속변수를 설명하게 된다. 그러나 회귀모형에서는 독립변수의 영향력이 시간이 변하더라도 일정하다고 가정하기 때문에 시계열적 변화를 반영하지 못한다는 단점을 가지고 있다(문권순, 1997). VAR 모형은 회귀모형의 한계를 극복하기 위하여, 종속변수의 과거 관측치들을 설명변수에 포함시킨 k 개의 선형회귀식을 통하여 시계열의 확률과정을 추정하는 방법으로서(박현수·안지아, 2009), 다음과 같은 식으로 요약된다.

$$Z_t = \delta + \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + \epsilon_t, \quad t = 1, 2, \dots, n$$

여기에서 k 는 시계열, Z_t 는 $k \times 1$ 의 시계열 벡터, δ 는 $k \times 1$ 의 상수벡터, ϕ_i 는 $k \times k$ 의 계수행렬, 그리고 ϵ_t 는 오차항을 나타낸다. VAR 모형에서는 일반적으로 AIC (Akaike information criterion), SBC (Schwarz's Bayesian criterion)와 같은 적합도 통계량을 최소화하는 시차길이 p 를 결정할 수 있다(Judge *et al.*, 1988). VAR 모형에서는 변수들의 배열순서 및 시차 길이 등에 따라 결과가 달라질 수 있으므로 변수 선정 및 시차 설정에 신중을 기해야 한다(문권순, 1997).

ARIMA는 모형의 확률적 성질이 시간에 따라 가변적인 비정상성을 전제로 하는(Ahn *et al.*, 2014) 점에서는 VAR와 유사하지만, 이동평균 항을 가지며 최우법(maximum likelihood)에 의한 추정을 수행한다는 점에서 VAR와 차이가 있다(이주호·박영환, 2002). 다변량 ARIMA 모형의 일반적인 표현은 $ARIMA(p, d, q)$ 인데, 여기에서 p 는 자기회귀성분, d 는 차분, q 는 이동평균 차수를 각각 의미하며, p, d, q 모두 음이 아닌 정수값을 가진다. 이 파라미터들은 ARIMA 모형의 예측력을 결정하므로 AIC나 BIC (Bayesian Information criterion)와 같은 기준을 이용하여 모형의 차수를 최적화할 수 있다. 일반적인 ARIMA 모형의 개념식은 다음과 같다(Box and Jenkins, 1976).

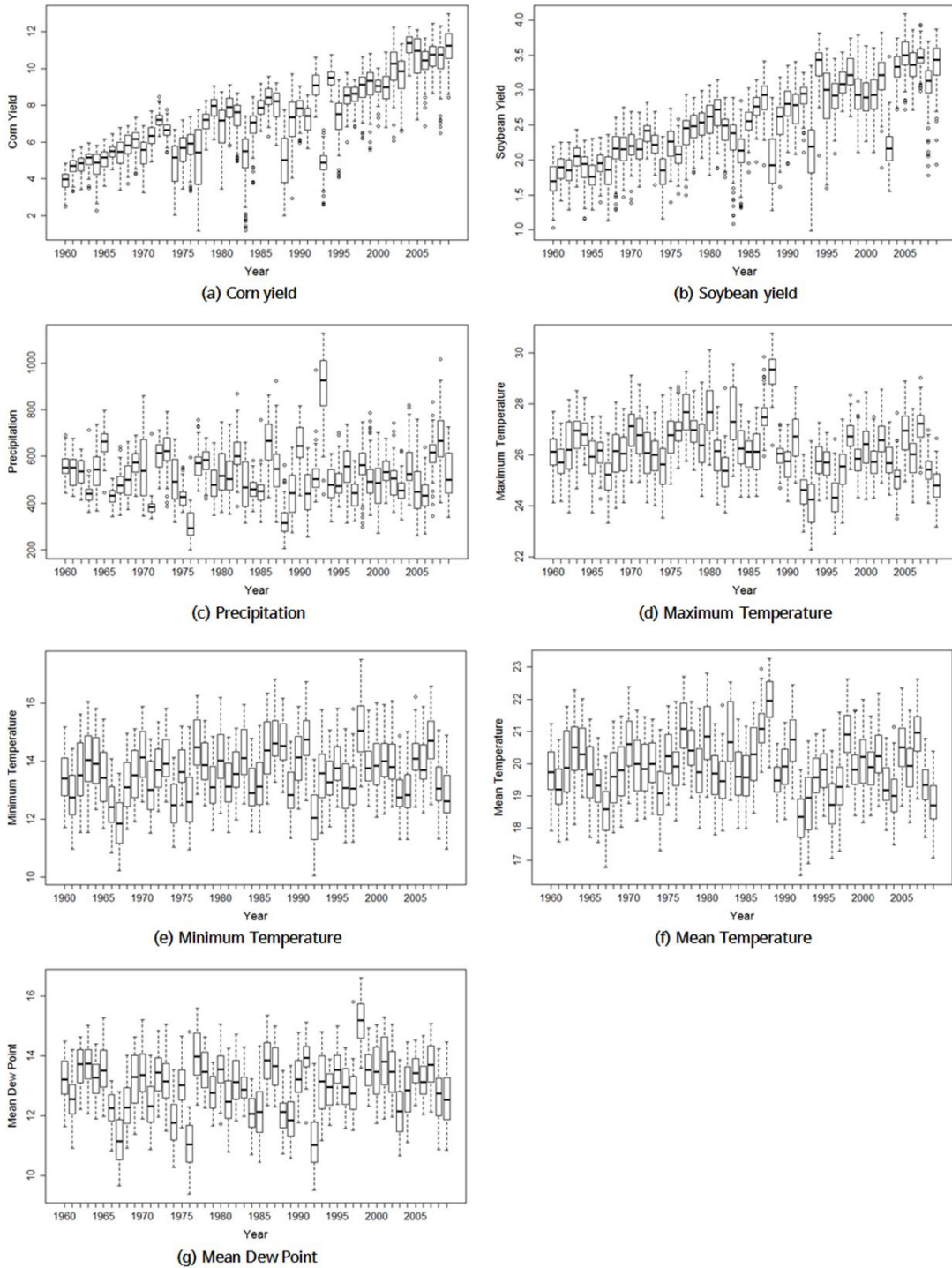


그림 1. 자료의 시계열 패턴(1960-2009년): (a) 옥수수 수확량, (b) 콩 수확량, (c) 강수량, (d) 최고기온, (e) 최저기온, (f) 평균기온, (g) 평균이슬점온도

$$(1-B)^d(1-B^*)^D Y_t = \mu + \frac{\theta(B)\theta_s(B^*)}{\phi(B)\phi_s(B^*)} a_t$$

여기에서 B 는 후향연산자, Y_t 는 시계열, $\phi(B)$ 는 비계절적 자기회귀성분, $\phi_s(B^*)$ 는 계절적 자기회귀성분, $\theta(B)$ 는 비계절적 이동평균성분, $\theta_s(B^*)$ 는 계절적 이동평균성분, 그리고 a_t 는 오차항을 나타낸다. 본 연구에서는 R 라이브러리를 이용하여 VAR와 ARIMA의 다변량 시계열모형을 작성하였다.

3. 검증방법

시계열모형의 예측정확도를 평가하기 위하여, 1960-1999년 VAR와 ARIMA 모형을 수립하여 2000년 수확량을 예측하고 1961-2000년 모형으로부터 2001년 수확량을 예측하는 등, 40년 자료의 시계열분석을 통해 그 다음해의 수확량을 예측하는 방식으로 2000-2009년의 10개 연도에 대한 예측을 수행하였다. 99개 카운티에 대한 모형예측치와 정부통계치를 비교하여 평균편이(mean bias), RMSE (root-mean-square error), MAPE (mean absolute percentage error)를 산출함으로써 정확도 평가의 지표로 사용하였다. 또한 예측정확도가 카운티에 따라 일정정도의 공간적 패턴을 보일 수 있으므로, 이를 알아보기 위하여 토양의 물리·화학적 특성에 따른 수확량 및 예측정확도 패턴을 조사하였다.

III. 결과 및 고찰

1. 시계열분석 결과

표 2에서 옥수수 수확량의 경우, 99개 카운티의 10개 연도 평균편이가 VAR는 -1.665 ton/ha, ARIMA는 -0.630 ton/ha로서 두 기법 모두 약간 과소추정의 경향이 있는 것으로 보인다. RMSE는 VAR가 2,050 ton/ha, ARIMA가 1,395 ton/ha이며, MAPE는 VAR가 17.2%, ARIMA는 11.0%로서, 전체적으로 VAR보다 ARIMA의 예측 정확도가 높은 것으로 나타났다. 특히 과거 40년 자료로부터 다음해의 옥수수 수확량을 예측할 때 ARIMA 기법이 11.0%라는 작은 오차율을 나타냈다는 점은 주목할 만하다. 콩의 경우, VAR와 ARIMA의 평균편이는 각각 -0.356 ton/ha와 -0.165 ton/ha로서 약간 과소추정의 경향을 보였다. RMSE의 경우, VAR는 0.612 ton/ha, ARIMA는 0.517 ton/ha이며, MAPE는 17.2%와 13.9%로서, 옥수수

표 2. VAR와 ARIMA 모형에 대한 옥수수와 콩의 예측정확도 (2000-2009년)

예측모형	곡물	Mean Bias (ton/ha)	RMSE (ton/ha)	MAPE (%)
VAR	Corn	-1,665	2,050	17.2
	Soybean	-0,356	0,612	17.2
ARIMA	Corn	-0,630	1,395	11.0
	Soybean	-0,165	0,517	13.9

와 마찬가지로 ARIMA의 예측 정확도가 높게 나타났다.

그림 2는 2000-2009년의 각 연도별 99개 카운티의 수확량 예측오차를 표현한 것으로서, 이들 오차는 대략적으로 정상적인 분포를 보이는 것으로 나타났다. 옥수수와 콩 모두 VAR에 비해 ARIMA의 연도별 오차 변동이 적어, VAR보다 ARIMA의 수확량 예측이 좀 더 안정적이라고 할 수 있을 것이다. 그러나 2003년 콩의 경우, 이례적으로 과대 추정되어 정확도 패턴이 다른 연도와 다르게 나타났는데, 이는 2003년 발생한 콩 진딧물의 영향으로 추측되며 이 시기에 아이오와주에서는 평년에 비해 약 32%의 수확량 감소가 보고되었다(<http://www.extension.iastate.edu/Publications/SP247X2007.pdf>). 그림 3과 4는 VAR와 ARIMA에 대한 카운티별 평균편이와 MAPE를 나타낸 것으로, VAR에 비해 ARIMA의 평균편이와 MAPE가 전반적으로 양호한 것으로 보인다.

2. 수확량 예측정확도 패턴

그림 3과 4를 살펴보면 특히 옥수수 수확량의 경우, 카운티에 따라 예측정확도가 높은 경우와 낮은 경우에 일정정도의 공간적 패턴이 존재하는 것으로 추측된다. 이러한 예측정확도의 공간적 패턴은 수확량에 영향을 미치는 지면환경 요소와 관련되므로 이를 알아보기 위하여 토양특성도(soil property map)와 함께 CSR (Corn Suitability Rating)을 분석하였다. 토양특성도는 HWSD (Harmonized World Soil Database)에서 제공하는 표토(topsoil, 0-30cm)와 심토(subsoil, 30-100cm)의 물리·화학적 특성으로 구성되고, CSR은 토양 유형에 따른 옥수수의 재배적합도(그림 5)를 나타낸 것이다(Miller, 1988). 옥수수 수확량과 토양특성도의 상관성을 분석해본 결과(표 3), 표토의 pH, 심토의 점토비율(clay fraction) 및 나트륨 함량과의 상관성이 높은 것으로 나타났다. 이처럼 상관성이 높은 세 가지 토양특성과 CSR을 기준으로 옥수수 실제 수확량과 VAR와 ARIMA를 통해 예측된 값의 관계가 어떠한지

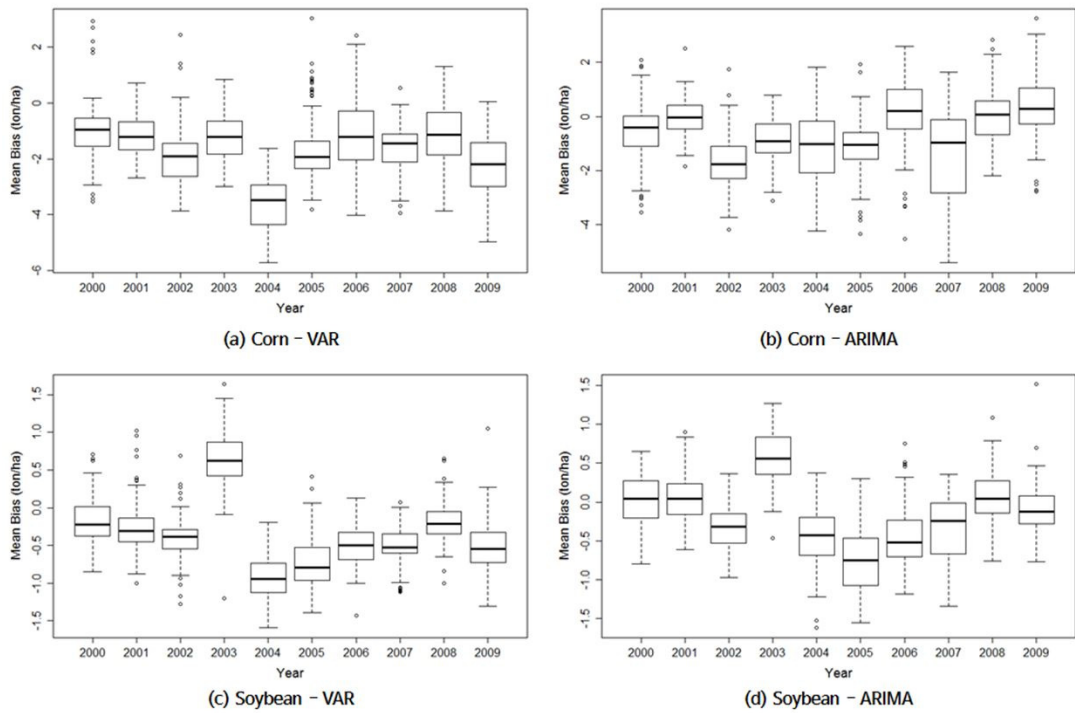


그림 2. 수확량 예측의 시계열 패턴(2000-2009년): (a) VAR 모형의 옥수수 수확량 예측오차, (b) ARIMA 모형의 옥수수 수확량 예측오차, (c) VAR 모형의 콩 수확량 예측오차, (d) ARIMA 모형의 콩 수확량 예측오차

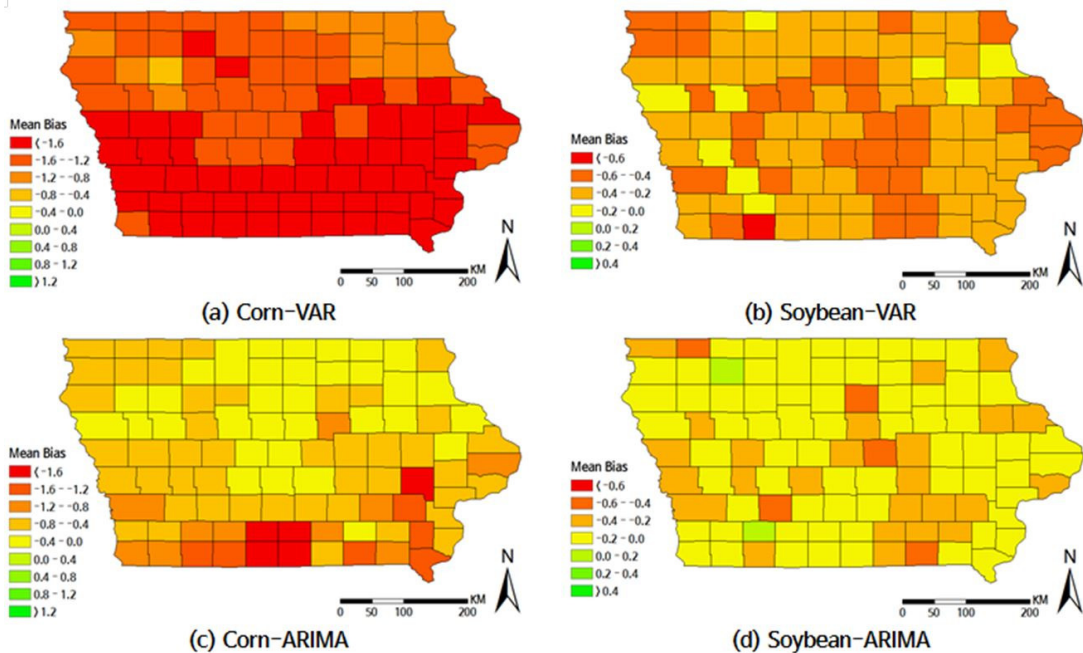


그림 3. VAR와 ARIMA 모형을 통해 예측된 수확량과 실제 수확량과의 차이(2000-2009년 평균편이)

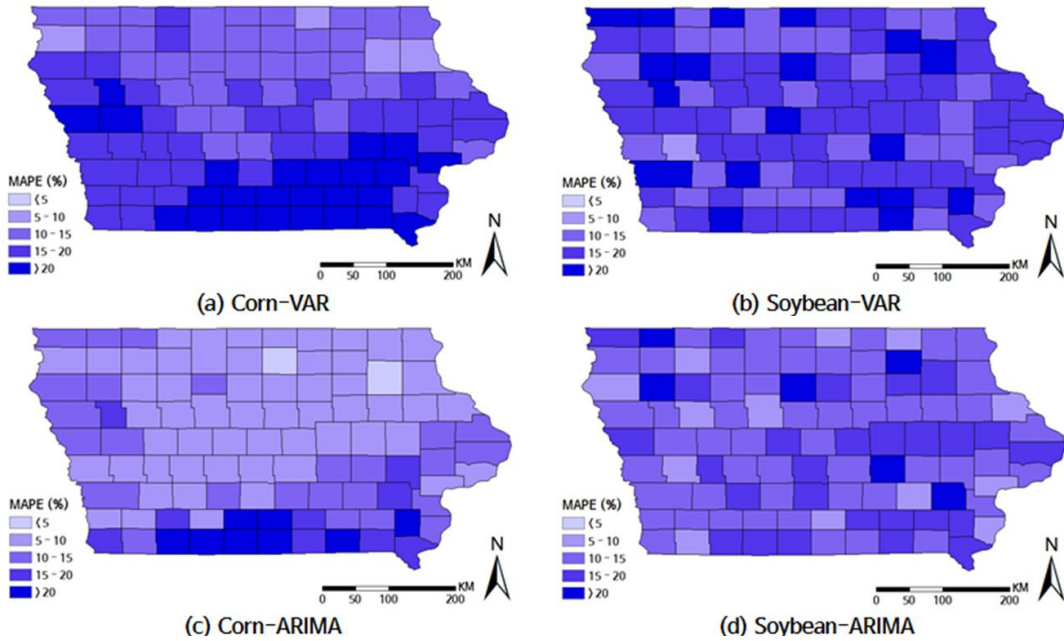


그림 4. VAR와 ARIMA 모형을 통해 예측된 수확량과 실제 수확량과의 차이(2000-2009년 MAPE)

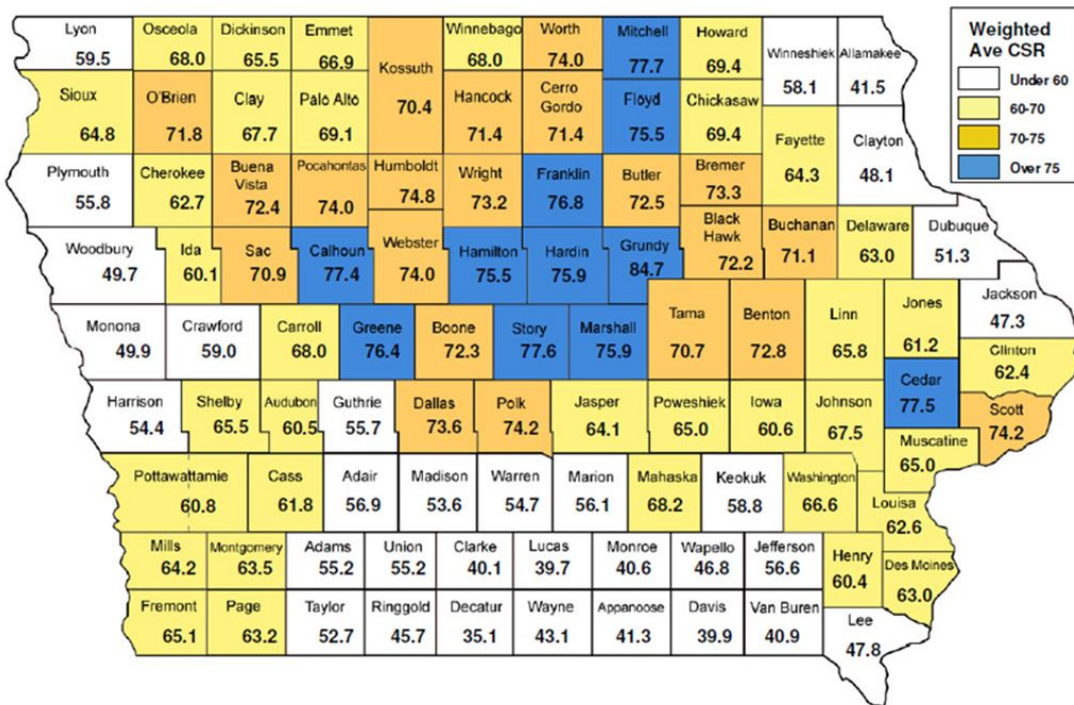


그림 5. 아이오와 주의 카운티별 옥수수 재배적합도(2010년)

출처: Department of Agronomy, Iowa State University

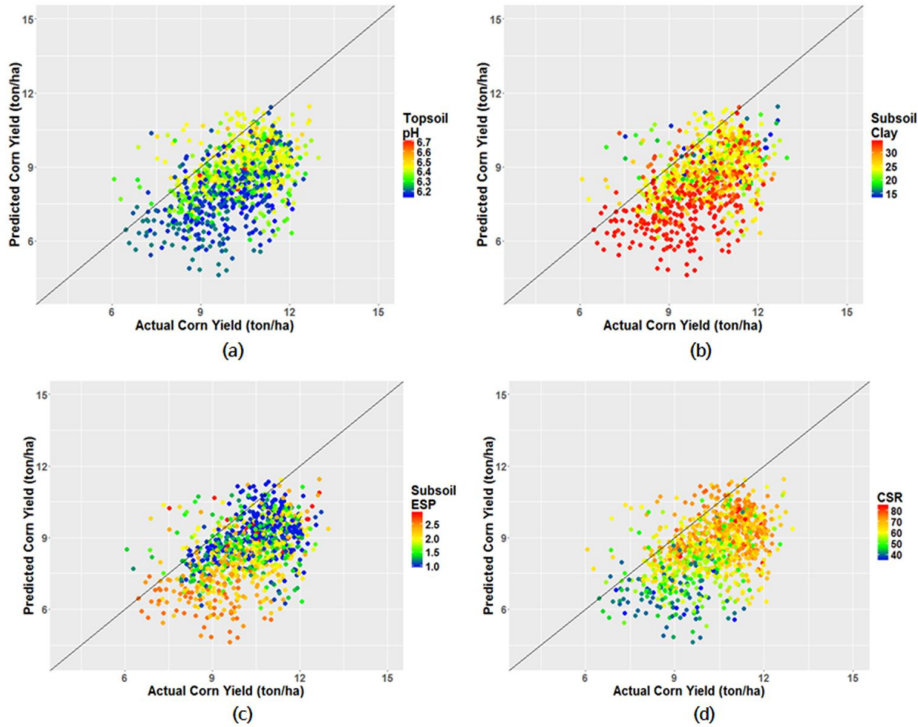


그림 6. 토양특성에 따른 VAR 모형의 옥수수 수확량 예측정확도 패턴: (a) 표토의 pH에 따른 정확도 패턴, (b) 심토의 점토비율에 따른 정확도 패턴, (c) 심토의 나트륨 함량에 따른 정확도 패턴, (d) CSR에 따른 정확도 패턴

표 3. 옥수수와 콩의 수확량과 토양의 물리·화학적 특성과의 상관계수(1960-2009)

구분	토양특성	옥수수 수확량	콩 수확량
표토	사토 비율 (sand fraction)	0.252	0.318
	미사토 비율 (silt fraction)	-0.281	-0.336
	점토 비율 (clay fraction)	-0.186	-0.273
	가비중 (bulk density)	0.048	0.025
	유기탄소	0.079	-0.162
	pH	0.452	0.171
	탄산칼슘	0.057	0.053
	나트륨 함량	-0.390	-0.131
심토	사토 비율 (sand fraction)	0.357	0.296
	미사토 비율 (silt fraction)	-0.192	-0.285
	점토 비율 (clay fraction)	-0.455	-0.271
	가비중 (bulk density)	-0.129	-0.115
	유기탄소	0.045	-0.112
	pH	-0.073	-0.117
	탄산 칼슘	-0.255	-0.084
	나트륨 함량	-0.538	-0.168

그림 6과 7에 나타내었다.

옥수수는 약산성 또는 알칼리성 토양환경에서 잘 자라는데, 그림 6(a)와 7(a)를 살펴보면, pH의 값이 큰 쪽에서(약산성) 실제 수확량이 많고 예측정확도가 높은 것으로 나타났다. 점토가 과다할 경우 옥수수 뿌리의 배수에 좋지 않은 영향을 줄 수 있는데, 심토의 점토비율(그림 6(b), 7(b))이 큰 쪽에서 실제 수확량이 적고 예측정확도가 낮은 것을 알 수 있다. 나트륨 성분은 작물 스트레스를 유발할 수 있는데, 심토의 나트륨 함량(그림 6(c), 7(c))이 적은 쪽에서 실제 수확량이 많고 예측정확도가 높은 것으로 나타났다. CSR의 경우(그림 6(d), 그림 7(d)), 재배적합도가 클수록 옥수수 수확량이 높은 반면 재배적합도가 낮을수록 수확량의 분포가 퍼져있고 예측정확도가 낮은 것으로 보인다. 표 4는 토양의 물리·화학적 특성에 따른 옥수수 수확량의 예측정확도 패턴을 나타낸 것으로, 표토의 pH가 약산성일수록, 심토의 점토비율과 나트륨함량이 적을수록 예측정확도가 높아짐을 알 수 있다. 또한 그림 4(a)와 (c)에서 VAR와 ARIMA 모형의 MAPE 경향을 살펴보았을 때, CSR이 낮은 지역이 높은 지역에 비해 예측정확도가 상대적으로 낮게 나타났다.

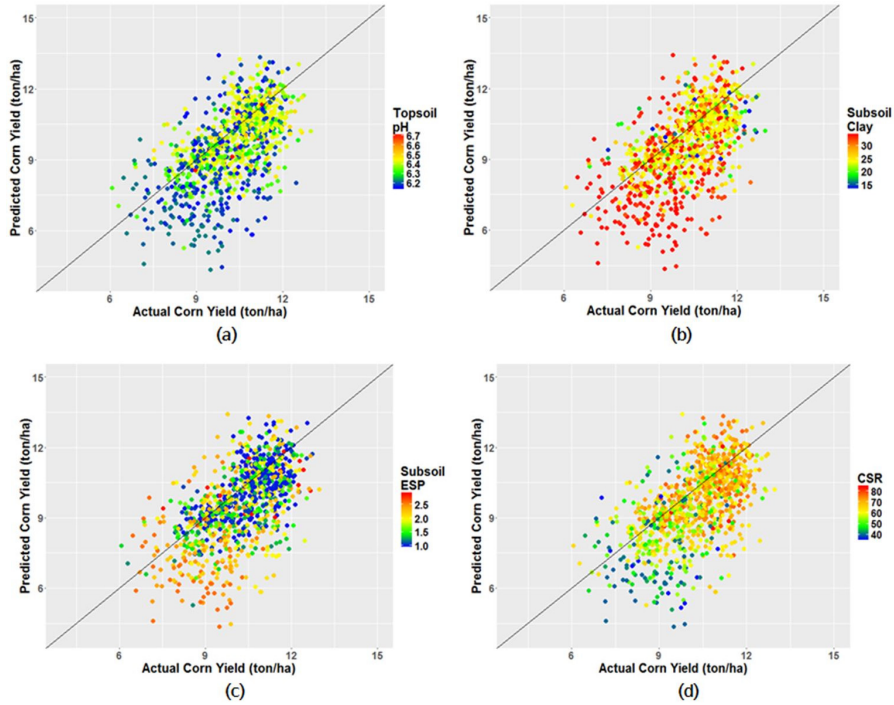


그림 7. 토양특성에 따른 ARIMA 모형의 옥수수 수확량 예측정확도 패턴: (a) 표토의 pH에 따른 정확도 패턴, (b) 심토의 점토비율에 따른 정확도 패턴, (c) 심토의 나트륨 함량에 따른 정확도 패턴, (d) CSR에 따른 정확도 패턴

표 4. 토양의 물리·화학적 특성에 따른 옥수수 수확량의 예측정확도 패턴(ARIMA 모형)

표토의 pH			심토의 점토 비율		
범위	평균편이	RMSE	범위	평균편이	RMSE
< 6.2	-0.821	1.757	< 15	-0.351	1.308
6.2 - 6.3	-1.011	1.907	15 - 20	-0.456	1.555
6.3 - 6.4	-0.622	1.327	20 - 25	-0.436	1.130
6.4 - 6.5	-0.391	1.167	25 - 30	-0.535	1.267
≥ 6.5	-0.327	0.834	≥ 30	-0.917	1.856
심토의 나트륨 함량			CSR		
범위	평균편이	RMSE	범위	평균편이	RMSE
< 1.0	-0.449	0.728	< 45	-1.310	2.226
1.0 - 1.5	-0.455	1.163	45 - 55	-0.674	1.581
1.5 - 2.0	-0.644	1.505	55 - 65	-0.657	1.529
2.0 - 2.5	-0.804	1.733	65 - 75	-0.499	1.235
≥ 2.5	-1.179	2.240	≥ 75	-0.407	1.230

한편 콩 수확량의 경우, 표토와 심토의 사토, 미사토, 점토의 구성비가 영향을 미치는 것으로 나타났다(표 3). 그러나 그림 3과 4의 (b), (d)를 보았을 때 일정한 예측정확도 패턴을 찾기는 어려웠고, 아이오와 주의 토양분류도

(그림 8)와 비교하였을 때에도 수확량과 토양성분과의 유의한 패턴을 특정할 수는 없었다.

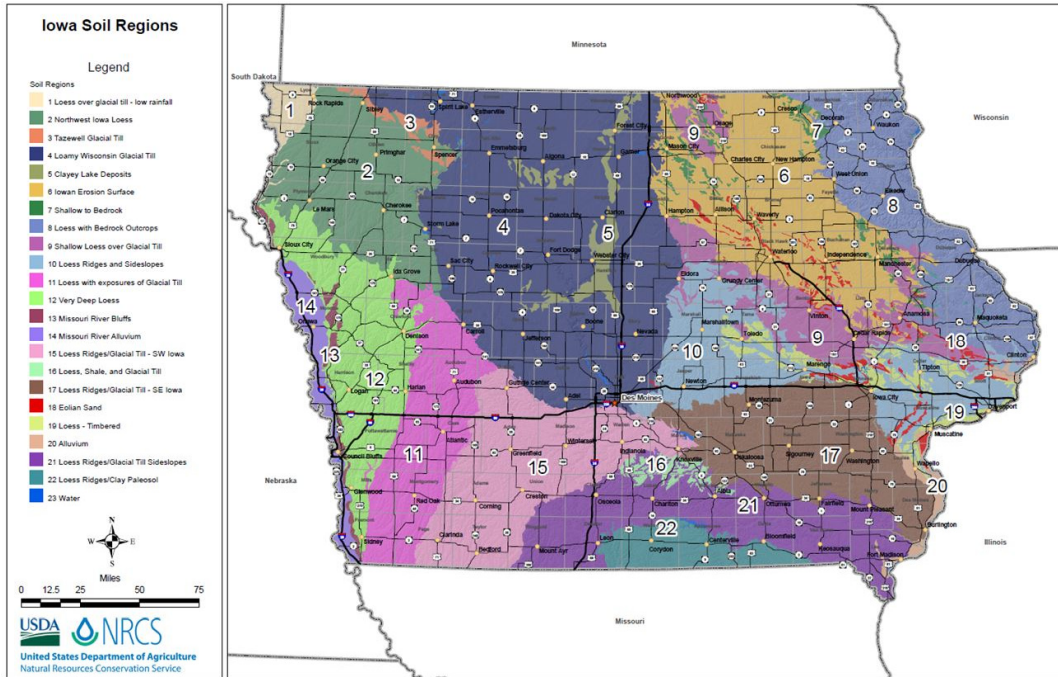


그림 8. 미국 아이오와 주의 토양분류도

출처: United States Department of Agriculture

IV. 결론

본 연구에서는 미국 아이오와 주의 옥수수와 콩의 수확량 예측을 위하여 1960-2009년의 장기간 기후자료를 사용하였고, 시계열통계법인 VAR와 ARIMA 모형을 이용하여 예측력의 비교분석을 수행하였다. 99개 카운티의 10개 연도에 대한 실험에서, VAR는 16-18%, ARIMA는 11-14%의 오차율로 다음해의 수확량을 예측할 수 있는 것으로 나타났다. 이러한 실험을 통하여 곡물수확의 미래예측에 있어 시계열모형의 가용성을 확인하였고, VAR와 ARIMA 기법의 성능을 비교평가할 수 있었다. 그러나 VAR과 ARIMA의 성능 차이에 대한 통계학적 접근을 위해서는 보다 심화된 향후연구가 필요할 것이다.

옥수수와 콩의 수확량과 기후요소 간의 패턴을 살펴보았을 때, 1988년과 1993년의 경우, 강수량의 급증과 고온 현상으로 인해 수확량이 급감한 것을 확인하였는데, 이는 기후변동이 수확량에 직접적인 영향을 미치는 것을 보여주는 단적인 예이다. 그러나 2003년의 경우, 기후요소가 평년과 다름없음에도 불구하고 콩의 수확량이 급감한 것을 볼 수 있는데, 이는 2003년 발생한 콩 진딧물에 의한 것으로 확인되었다. 이처럼 병해충에 의한 수확량의 감소는 기후요소로 설명할 수 없는

현상으로서, 향후 연구에서 고려할 필요가 있음을 시사한다.

토양특성이 실제 수확량 및 예측정확도에 미치는 영향을 알아본 결과, 옥수수의 경우 표토의 산성도, 심토의 점토와 나트륨 함량 등의 영향을 받는 것으로 나타났다. CSR이 높은 지역은 실제 수확량이 많고 예측정확도가 높으며, CSR이 낮은 지역에서는 실제 수확량이 적고 예측정확도가 낮은 것으로 보인다. 그러나 일부 카운티에서는 토양특성이 부적합하고 CSR이 낮음에도 불구하고 실제 수확량이 많은 경우도 있었는데 이는 농업기술의 발전(Tolhurst and Ker, 2014) 및 비료, 관개 등의 농업관리를 효율적으로 수행함으로써 기후요소의 영향을 극복한 것이라고 해석할 수 있다. 그러나 농업기술이나 농업관리에 대한 장기시계열 자료가 부재한 상황이므로, 그 영향에 대한 분석은 향후연구에서 시도되어야 할 것으로 사료된다.

시계열모형 적용에 있어서는 훈련기간과 모의기간을 다양하게 설정하거나 자료의 이동평균을 활용함으로써 성능개선이 가능할 것으로 보인다. 이와 함께 GTWR (geographically and temporally weighted regression)과 같은 시공간모형, BSTS (Bayesian structural time series) 등의 베이지언 접근법, 그리고 RNN (recurrent neural network)과 같은 기계학습법

등 최근에 개발된 시계열분석법을 추가 적용함으로써 예측정확도를 보다 향상시킬 수 있을 것으로 기대된다. 또한 GCM과 RCM (regional climate model)의 미래 기후자료를 사용함으로써, 다음의 예측에서 더 나아가 근미래(2020-2030년) 수확량 예측도 가능할 것으로 전망한다.

참고문헌

- 김나라·조재일·시바사키료스케·이양월, 2014, “위성영상과 기후자료를 이용한 미중서부 옥수수와 콩의 수확량 추정,” *기후연구*, 9(4), 315-329.
- 문권순, 1997, “벡터자기회귀모형(VAR)의 이해,” *통계분석연구*, 2(1), 23-56.
- 박현수·안지아, 2009, “VAR 모형을 이용한 부동산가격 변동요인에 관한 연구,” *부동산연구*, 19(1), 27-49.
- 송용호·이우균·곽한빈·김문일·양승룡, 2013, “기온변화에 대한 옥수수와 밀 생산량 취약성 평가 - 미국과 중국을 사례로,” *한국기후변화학회지*, 4(4), 371-384.
- 심교문·김용석·정명표·김석철·민성현·소규호, 2013, “한국의 농업기후지대별 이상기온 출현 특성 평가,” *한국기후변화학회지*, 4(2), 189-199.
- 이주호·박영환, 2002, “ARIMA 모형을 이용한 모든 교통량 예측에 관한 연구,” *산업기술연구*, 14(1), 81-96.
- 홍석영·허지나·안중배·이지민·민병걸·이충근·김이현·이경도·김선화·김진엽·심교문, 2012, “MODIS NDVI와 기상자료를 이용한 우리나라 벼 수량 추정,” *대한원격탐사학회지*, 28(5), 509-520.
- 홍석영·나상일·이경도·김용석·백신철, 2015, “MODIS NDVI와 강수량 자료를 이용한 북한의 벼 수량 추정 연구,” *대한원격탐사학회지*, 31(5), 441-448.
- Ahn, J., Hong, S., Cho, J., Lee, Y., and Lee, H., 2014, Statistical modeling of sea ice concentration using satellite imagery and climate reanalysis data in the Barents and Kara Seas, 1979-2012, *Remote Sensing*, 6(6), 5520-5540.
- Allen, J.D., 1990, A look at the remote sensing applications program of the National Agricultural Statistics Service, *Journal of Official Statistics*, 6, 393-409.
- Amin, M., Amanullah, M., and Akbar, A., 2014, Time series modeling for forecasting wheat production of Pakistan, *The Journal of Animal & Plant Sciences*, 24(5), 1444-1451.
- Atzberger, C., 2013, Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs, *Remote Sensing*, 5(8), 949-981.
- Box, G.E.P. and Jenkins, G.M., 1976, *Time series analysis: Forecasting and control*, San Francisco, CA: Holden-Day Inc.
- Choudhury, A. and Jones, J., 2014, Crop yield prediction using time series models, *Journal of Economic and Economic Education Research*, 15(3), 53-67.
- Daly, C., Halbleid, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., and Pasteris, P.P., 2008, Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States, *International Journal of Climatology*, 28(15), 2031-2064.
- Doraiswamy, P.C., Sinclair, T.R., Hollinger, S., Akhmedov, B., Stern, A., and Prueger, J., 2005, Application of MODIS derived parameters for regional crop yield assessment, *Remote Sensing of Environment*, 97(2), 192-202.
- Farook, A.J. and Kanna, K.S., 2016, Climate change impact on rice yield in India-Vector autoregression approach, *Sri Lankan Journal of Applied Statistics*, 16(3), 161-178.
- Ferencz, C., Bognár, P., Lichtenberger, J., Hamar, D., Tarcsai, G., Timár, G., Molnár, G., Pásztor, S., Steinbach, P., Székely, B., Ferencz, O.E., and Ferencz-Árkos, I., 2004, Crop yield estimation by satellite remote sensing, *International Journal of Remote Sensing*, 25(20), 4113-4149.
- Judge, G.G., Hill, R.C., Griffiths, W.E., Lutkepohl, H., and Lee, T.S., 1988, *Introduction to the Theory and Practice of Econometrics*, 2nd ed., New York: John Wiley & Sons.
- Labus, M.P., Nielsen, G.A., Lawrence, R.L., Engel, R., and Long, D.S., 2002, Wheat yield estimates using

- multi-temporal NDVI satellite imagery, *International Journal of Remote Sensing*, 23(20), 4169-4180.
- Meza, F.J. and Silva, D., 2009, Dynamic adaptation of maize and wheat production to climate change, *Climatic Change*, 94(1-2), 143-156.
- Miller, G.A. ed., 1988, *Corn Suitability Ratings-An Index to Soil Productivity*, Cooperative Extension Service, Iowa State University.
- Paul, R.K. and Sinha, K., 2016, Forecasting crop yield: A comparative assessment of ARIMAX and NARX model, *Journal of the Society for Application of Statistics in Agriculture and Allied Sciences*, 1(1), 77-85.
- Pease, J.W., Wade, E.W., Skees, J.S., and Shrestha, C.M., 1993, Comparisons between subjective and statistical forecasts of crop yields, *Review of Agricultural Economics*, 15(2), 339-350.
- Peng, S., 1995, *Climate Change and Rice*, International Rice Research Institute.
- Prasad, A.K., Chai, L., Singh, R.P., and Kafatos, M., 2006, Crop yield estimation model for Iowa using remote sensing and surface parameters, *International Journal of Applied Earth Observation and Geoinformation*, 8(1), 26-33.
- Ren, J., Chen, Z., Zhou, Q., and Tang, H., 2008, Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China, *International Journal of Applied Earth Observation and Geoinformation*, 10(4), 403-413.
- Sims, C., 1980, Macroeconomics and reality, *Econometrica*, 48, 1-49.
- Steinwand, A.L., Karlen, D.L., and Fenton, T.E., 1996, An evaluation of soil survey crop yield interpretations for two central Iowa farms, *Journal of Soil and Water Conservation*, 51(1), 66-71.
- Tolhurst, T.N. and Ker, A.P., 2014, On technological change in crop yields, *American Journal of Agricultural Economics*, doi: 10.1093/ajae/aau082.
- Department of Agronomy, Iowa State University, <http://www.agron.iastate.edu>.
- United States Department of Agriculture, <http://www.usda.gov>.
- 교신: 이양원, 48513, 부산광역시 남구 용소로 45, 부경대학교
공간정보시스템공학과(이메일: modconfi@pknu.ac.kr)
- Correspondence: Yang-Won Lee, Department of Spatial Information Engineering, Pukyong National University, 45 Yongso-ro, Nam-gu, Busan 48513, Republic of Korea (Email: modconfi@pknu.ac.kr)
- 투 고 일: 2016년 8월 9일**
심사완료일: 2016년 8월 14일
투고확정일: 2016년 8월 19일