위성영상과 기후자료를 이용한 미중서부 옥수수과 콩의 수확량 추정

김나리¹ · 조재일² · 시바사키료스케³ · 이양원¹

¹부경대학교 공간정보시스템공학과 · ²국토연구원 국토정보연구본부 · ³동경대학교 공간정보과학연구센터
(2014년 11월 22일 접수, 2014년 12월 10일 수정, 2014년 12월 16일 게재 확정)

Estimation of Corn and Soybeans Yields of the US Midwest using Satellite Imagery and Climate Dataset

Nari Kim¹ · Jaeil Cho² · Ryosuke Shibasaki³ · Yang-Won Lee¹

¹Department of Spatial Information Engineering, Pukyong National University

²Geospatial Information Research Division, Korea Research Institute for Human Settlements

³Center for Spatial Information Science, The University of Tokyo

(Received 22 November 2014, Revised 10 December 2014, Accepted 16 December 2014)

Abstract : This paper described the estimation of corn and soybeans yields of four states in the US Midwest using time-series satellite imagery and climate dataset between 2001 and 2012. We first constructed a database for (1) satellite imagery acquired from Terra MODIS (Moderate Resolution Imaging Spectroradiometer) including NDVI (Normalized Difference Vegetation Index), EVI (Enhanced Vegetation Index), LAI (Leaf Area Index), FPAR (Fraction of Photosynthetically Active Radiation), and GPP (Gross Primary Productivity), (2) climate dataset created by PRISM (Parameter-Elevation Regressions on Independent Slopes Model) such as precipitation and mean temperature, and (3) US yield statistics of corn and soybeans. Then we built OLS (Ordinary Least Squares) regression models for corn and soybeans yields between 2001 and 2010 after correlation analyses and multicollinearity tests. These regression models were used in estimating the yields of 2011 and 2012. Comparisons with the US yield statistics showed the RMSEs (Root Mean Squared Errors) of 0.892 ton/ha and 1.095 ton/ha for corn yields in 2011 and 2012 respectively, and those of 0.320 ton/ha and 0.391 ton/ha for soybeans yields. This result can be thought of as a good agreement with the in-situ statistics because the RMSEs were approximately 10% of the usual yields: 9 ton/ha for corn and 3 ton/ha for soybeans. Our approach presented a possibility for extending to more advanced statistical modeling of crop yields using satellite imagery and climate dataset.

Key Words : satellite imagery, climate data, crop yield, corn, soybeans

주요어 : 위성영상, 기후자료, 곡물수확, 옥수수, 콩

1. 서론

곡물의 수확량 정보는 식량의 수급 조절 및 안보 유지, 가격 예측 등 농업 정책 수립에 중요한 판단 근거 자료로 사용된다. 특히 우리나라의 경우 곡물의 자급률은 2010년 기준 26.7%이며(Yoo *et al.*, 2012) 쌀을 제외한 다른 주요 곡물인 옥수수와 콩의 경우 0.5%, 8.7%의 매우 낮은 자급률을 보이며 대부분 수입에 의존하고 있어 수확량 정보의 중요성이 더욱 부각되고 있다(Ministry for Food, Agriculture, Forestry and Fisheries, 2001). 최근 세계적인 이상기후의 영향으로 국제 곡물시장은 불안정하게 변화하고 있으며, 2012년 미국에서 발생한 대규모 가뭄으로 인해 옥수수와 콩의 가격이 급격히 오르는 곡물파동을 겪는 등, 주요 곡물들을 수입에 의존하는 우리나라는 이와 같은 식량수급의 불안정에 대비하기 위해 주요 수입국 또는 관심국가의 곡물 수확량에 대한 장기적이고 지속적인 모니터링이 필요하다.

곡물 수확량의 시뮬레이션을 위하여 국외에서는 농작물 생육모형을 이용한 연구들이 꾸준히 수행되어 왔다. Liu *et al.*(2007)은 GIS(Geographic Information System) 기반의 GEPIC(GIS-based Environmental Policy Integrated Climate) 모델을 개발하고 전지구의 밀 생산량을 추정하였다. Wang and Li(2010)의 연구에서는 장기간의 실험 자료와 현장조사 자료를 기반으로 한 EPIC(Environmental Policy Integrated Climate) 모델을 이용하여 중국 황토고원에서 생산되는 주요 곡물의 수확량을 모의하였다. Liu *et al.*(2011)은 DSSAT(Decision Support System for Agrotechnology Transfer)의 서브모델인 CERES(Crop Environment Resource Synthesis)-Maize를 이용하여 캐나다 지역의 옥수수 수확량과 질소순환의 관계를 통해 장기간 옥수수 수확량을 시뮬레이션 하

였다. 이 외에도 WOFOST(World Food Studies), CROPSYST(Cropping Systems Simulation Model) 등 다양한 모형을 이용한 곡물 수확량 추정에 관한 연구들이 수행되었다(Vandiepen *et al.*, 1989; Vanevert and Campbell, 1994). 그러나 이러한 농작물 생육모형은 기후자료 이외에도 토양특성, 지표유출, 관개, 비료, 농약, 작물관리, 농사일정 등 다양한 입력자료를 필요로 하고 계산구조가 상당히 복잡한 것에 비해(Becker *et al.*, 2010) 아직까지는 곡물 수확량의 추정 정확도가 매우 만족스럽지는 않은 것으로 평가된다(Tan and Shibasaki, 2003; Wu *et al.*, 2007).

이러한 이유로 최근에는 곡물 수확량 추정에 있어 위성원격탐사 자료의 사용이 확대되고 있다. 전통적으로 곡물 수확량 통계는 현지조사를 통해 이루어졌으나 이 방법은 자료의 신속성과 정확성이 떨어지고 국외 및 비접근 지역에 대한 객관적인 자료 취득에 어려움이 있는 반면, 위성원격탐사는 국내외 관측위성을 통해 넓은 지역에 대한 지속적이고 다양한 자료의 취득 및 분석이 용이한 장점이 있다(Hong *et al.*, 2012). 따라서 위성자료를 이용한 곡물 수확량 추정은 농업분야에서 중요한 역할을 하고 있으며 많은 국가에서 활발하게 연구되고 있다. 그 중에서도 광학 위성자료를 기반으로 산출된 식생지수를 활용한 통계분석을 통해 작물의 수확량을 추정하는 연구가 주류를 이루고 있다.

Allen(1990)은 1977년 아이오와 주에서 최초로 위성영상 자료를 이용하여 작물재배면적 및 수확량을 추정하였다. Lee *et al.*(1997)은 벼 군락의 반사특성을 이용하여 엽면적지수(Leaf Area Index: LAI)를 추정하고 이를 통해 수확량을 산출하였다. Doraiswamy *et al.*(2005)은 MODIS(Moderate Resolution Imaging Spectroradiometer) 엽면적지수를 작물 모형에 적용하여 미국 일리노이 주의 콩과 옥수수의 수확량을 추정하였고, 미국 통계

자료와 비교하여 정확도 분석을 수행한 결과 10% 이내의 오차를 나타내었다. Ferencz *et al.* (2004)은 위성영상 자료를 이용하여 일반수확량통합참고지수(General Yield Unified Reference Index: GYURI)를 개발하고 이를 이용한 작물의 수확량 연구를 수행하였다. Ren *et al.* (2008)은 MODIS NDVI(Normalized Difference Vegetation Index)를 이용하여 중국 산둥반도의 겨울 밀 수확량을 추정하였으며, 그 결과 농업기후모델보다 적은 오차를 나타내었다. Hong *et al.* (2012)은 MODIS NDVI와 기상자료를 기반으로 회귀모형을 통해 한국의 논벼 수확량을 추정하였으며, 이를 통해 기상관측지점의 평균자료보다 지형특성을 고려한 격자형 기상자료가 추정오차를 줄일 수 있음을 보였다. Na *et al.* (2012)는 기상요소와 MODIS NDVI를 이용하여 논벼 수확량 예측모형을 개발하였으며, 예측치는 다소 과소추정되기는 하나 NDVI와 기상요소에 따른 논벼 생산량 변화특성을 잘 반영하는 것으로 평가되었다. 또한 Na *et al.* (2013)의 연구에서는 시계열 MODIS 자료와 기상자료를 CASA(Carnegie-Ames-Stanford Approach) 모형에 적용하고 NPP(Net Primary Productivity)와 수확량과의 상관성 분석을 통해 우리나라의 논벼 수확량을 추정하였는데, 약간의 과대추정이 발생하지만 연간 누적 NPP와 9월의 NPP가 수확량과 높은 상관성이 있음을 밝혔다.

이와 같이 국내외의 많은 연구들이 위성원격탐사에 기초하여 곡물수확량 추정을 위한 통계모델링을 수행하였으며, 이를 통해 의미있는 성과를 거두었다. 그러나 위성기반의 선행연구들에 사용된 데이터의 종류와 용량은 비교적 제한적이어서, 샘플 크기에 따라 결과가 달라질 수 있는 통계모델링의 특성상 보다 다양한 장기간의 대용량 위성영상과 기후자료가 요구되는 것이 사실이다. 또한 곡물의 종류에 따라 생장특성에 차이가 존재하고,

식생지수들은 곡물에 반응하는 정도가 서로 상이하다. 따라서 기존의 연구에서 주로 사용되었던 NDVI뿐 아니라 작물의 종류에 따라 민감도가 달라지는 EVI(Enhanced Vegetation Index), LAI 등 수확량에 영향을 미치는 각종 식생지수들과 작물의 생장에 영향을 미치는 기온과 강수 등 기후자료를 이용한 분석이 필요하다. 그리고 기존의 연구에서는 주로 식생지수와 기후요소, 수확량간의 상관성을 통해 곡물 수확량을 추정하는 연구가 대부분이었으나, 수확량 예측 정확도를 보다 향상시키기 위하여 광합성유효복사율(Fraction of Photosynthetically Active Radiation: FPAR)과 총일차생산량(Gross Primary Product: GPP) 등 가용한 다양한 위성자료를 분석에 활용할 필요가 있다. 이러한 맥락에서 본 연구에서는 10년 이상의 비교적 장기간에 걸쳐 다양한 대용량의 위성영상과 기후자료를 이용하여 미국 중서부 지역의 주요 곡물인 옥수수와 콩의 수확량을 통계모델링을 수행하고, 미국의 곡물수확량 통계자료와의 비교 검증을 통해 본 연구의 수확량 추정이 적절한지에 대해 검토하고자 한다.

2. 자료 및 방법

1) 연구지역

미국은 세계 최대의 곡물 수출국 중 하나로 2010년과 2011년에 세계 옥수수 수출량의 49.2% 이상을 수출하였다(USDA/FAS, 2012). 특히 미국 중서부 지역은 세계 최대의 곡창지대로서 12개의 주(Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin)로 이루어져 있다. 본 연구에서는 미국 중서부 지역 중에서 농지면적의 비율이 주 면적의 50% 이상이고 옥수수와 콩을 주

Figure 1. Study area.

요 작물로 재배하고 있는 Iowa, Minnesota, North Dakota, South Dakota 4개 주를 선정하여 연구를 수행하였다(Figure 1). 연구 기간은 옥수수와 콩의 생육시기인 5월부터 10월까지로 설정하였고 위성 자료의 제공시기에 맞춰 2001~2012년의 자료를 이용하여 수확량의 통계모델링과 검증을 수행하였다.

2) 자료처리

본 연구에서는 위성영상과 기후자료를 이용하여 옥수수와 콩의 수확량을 추정하기 위하여, Terra MODIS 위성영상, PRISM(Parameter-Elevation Regressions on Independent Slopes Model) 기후자료 및 옥수수와 콩의 수확량 통계자료를 수집하였고, 위성자료가 존재하는 시기를 기준으로 2001년부터 2012년까지의 NDVI, EVI, LAI, FPAR, GPP, 강수, 평균기온의 7개 변수들로 데이터베이스를 구축하였다. 2001~2010년의 10년 자료를 이용하여 통계모델링을 수행하였고 2011년과 2012년 자료를 이용하여 검증을 수행하였다. 연구에서 사용된 위성 및 기후 영상은 총 3,640장으로서 각

자료의 세부사항은 Table 1과 같으며, Figure 2는 이러한 자료들의 처리 및 분석 흐름을 나타낸 것이다.

(1) 수확량 통계자료

미국 농무성(United States Department of Agriculture: USDA) 농업통계청(National Agricultural Statistics Service: NASS)에서는 미국에서 생산되는 곡물들에 대한 연도별 통계자료를 카운티(county) 단위로 제공하고 있다(<http://quickstats.nass.usda.gov>). 이 중 4개 주의 305개 카운티(Iowa 99개, Minnesota 87개, North Dakota 53개, South Dakota 66개)에 대한 2001년부터 2012년까지의 통계자료를 수집하였으며, bushel/acre 단위의 수확량은 편의상 ton/ha로 변환하였다.

(2) MODIS 위성영상

본 연구는 육상생태계에 초점을 두고 있으므로 Terra 위성의 MODIS 자료를 주로 사용하였으며, 3가지 식생지수 NDVI, EVI, LAI와 함께, 광합성유효복사율과 총일차생산량 자료를 사용하였

Table 1. Summary of data used.

Data		Spatial Resolution	Temporal Resolution	Source	No. of Scenes	Volume(GB)
Satellite Products	NDVI	1 km	Monthly	NASA EOSDIS	288	5.13
	EVI				288	5.13
	LAI		8-day		960	3.11
	FPAR				960	3.11
	GPP	960			6.51	
	Land Cover	500 m	Yearly		40	3.43
Climate Data	Precipitation	4 km	Monthly	PRISM Climate Group	72	0.47
	Tmean				72	0.47
Statistical Data	Crop Yield	County	Yearly	USDA		

다. NDVI는 가장 널리 사용되는 식생지수로 식생의 활력도를 나타내며 근적외 채널과 적색 파장대의 반사도를 통해 계산된다. EVI는 NDVI의 신호포화(saturation)문제를 해결하기 위해 청색 파장대를 추가로 적용하여 산출된 식생지수로서 대기 에어로솔과 토양반사의 영향을 감소시켜 식생밀도가 조밀한 곳에서 NDVI보다 식생변화에 더 민감하

게 반응하는 장점을 가지고 있기 때문에 전지구 규모의 식생 모니터링에 널리 사용된다(Townshend *et al.*, 1991; Huete *et al.*, 2002). LAI는 단위 면적당 잎 면적의 비율을 나타내는 것으로 식물의 캐노피로부터 에너지, CO₂, 수분의 교환을 나타내는 식물의 중요한 구조학적 변수로 사용된다(Sellers *et al.*, 1986). 광합성유효복사율은 식물이 흡수한 PAR(Photosynthetically Active Radiation)의 비율로서 식물의 캐노피에서 흡수되는 0.4–0.7 μm 파장대의 광에너지가 광합성에 사용되는 비율을 말한다(Sellers, 1985). 육상식생의 총일차생산성을 나타내는 GPP는 NPP에 식물의 호흡량을 더한 바이오매스 화학에너지의 총량으로서 지구 탄소순환에서도 중요한 요소이다.

각 MODIS 자료들은 편의상 Sinusoidal Grid 좌표계에서 평면직각의 WGS84(World Geodetic System 1984) 경위도 좌표계로 변환하였다. 또한 수확량과의 비교를 위하여 위성자료를 월 단위 및 연도 단위로 집계하였으며, 옥수수와 콩의 생육기간을 고려하여 매년 5월부터 10월까지의 자료를 합성하였다.

Figure 2. Flowchart of data analysis.

(3) PRISM 기후자료

미국의 PRISM Climate Group은 미국 전역의 시공간적으로 연속된 기후자료를 생산하기 위해 만들어진 연구그룹으로서, 지형효과를 고려한 PRISM 내삽법(Daly *et al.*, 1994; Daly *et al.*, 2008)을 개발하여 800 m 및 4 km 격자의 기후자료를 산출하고 있다. PRISM 데이터셋은 강수, 평균기온, 최고기온, 최저기온, 평균이슬점온도, 그리고 수증기압 등의 변수로 구성되는데, 본 연구에서는 월별 4 km 공간해상도의 강수, 평균기온을 사용하였다. 위성영상과의 좌표계 및 공간해상도 일치를 위하여 NAD1983 측지좌표계를 WGS84 경위도 좌표계로 변환하고 4 km 해상도를 1 km 해상도로 조정하였으며, 시간해상도의 일치를 위하여 생육기간인 5월에서 10월까지의 자료를 집계하여 연자료로 합성하였다.

(4) 농지 마스크

MODIS 토지피복 자료는 월주기로 합성된 식생지수에 무감독 분류기법을 적용한 것으로서 IGBP(International Geosphere-Biosphere Programme)에서 제공하는 17개 분류법을 채택한다(Friedl *et al.*, 2010). 미국 중서부 지역의 농지만을 추출하기 위하여 17개 분류 중 “Croplands”인 12번 클래스를 사용하였다. MODIS 영상에서 2001~2010년의 10년 기간 동안 계속 “Croplands”를 유지한 화소만을 추출하여 농지 마스크를 작성하고 앞서 가공처리한 위성자료와 기후자료에 농지 마스크를 적용하였다. 또한 전체 면적에서 농지가 차지하는 비율이 50% 이상인 카운티만을 선별하여(North Dakota 36개, South Dakota 30개, Minnesota 46개, Iowa 75개) 카운티 단위로 데이터베이스를 구축하였다. 이때 폴리곤(카운티) 단위로 래스터(위성영상 및 기후자료) 화소값을 재집계하기 위하여 Figure 3과 같이 구역평균통계량(zonal

Table 2. IGBP classifications of land cover (Friedl *et al.*, 2010).

Land Cover	Code	IGBP Classifications
Forests	1	Evergreen Needleleaf Forest
	2	Evergreen Broadleaf Forest
	3	Deciduous Needleleaf Forest
	4	Deciduous Broadleaf Forest
	5	Mixed Forest
Shrublands	6	Closed Shrublands
	7	Open Shrublands
Woodlands	8	Woody Savannas
	9	Savannas
Grasses/cereals	10	Grasslands
Seasonally or permanently inundated	11	Permanent Wetlands
Croplands and mosaics	12	Croplands
	14	Cropland/Natural Vegetation Mosaic
Unvegetated	13	Urban and Built-up
	15	Snow and Ice
	16	Barren or Sparsely Vegetated
	17	Water Bodies

Figure 3. Calculation of zonal mean statistic using MODIS imagery and PRISM climate dataset.

mean statistic)을 계산하였다.

3) Ordinary Least Squares 회귀모형

일반적으로 사용되는 선형회귀모형인 OLS (Ordinary Least Squares) 회귀모형은 잔차의 제곱의 합이 최소가 되게 하는 모형으로 다음의 식으로 표현된다.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_i x_i + \varepsilon$$

여기에서 y 는 반응변수로서 설명변수에 의해 영향을 받는 변수이고 $x_1 \sim x_i$ 는 모형에 영향을 주는 설명변수이다. $\beta_0 \sim \beta_i$ 는 추정되는 회귀계수로서 β_0 는 절편이고 β_i 는 i 번째 설명변수의 영향력을 나타내며, ε 은 오차항을 의미한다. 본 연구에서 y 에는 옥수수와 콩의 수확량 통계자료를 사용하였고,

$x_1 \sim x_i$ 는 앞서 위성영상 및 기후자료로부터 도출된 수확량에 영향을 미치는 요소들로 구성하였으며, 회귀모형의 설명력을 높이기 위하여 4개주 187개 카운티의 10년간 자료를 사용하였다.

반응변수를 설명하기 위한 설명변수 후보들이 많은 경우에 어떤 설명변수들을 조합하여 회귀모형을 만들지가 중요하다. 식생지수인 NDVI, EVI, LAI, 광합성과 관련되는 FPAR, 바이오매스와 관련되는 GPP, 그리고 기후변수로서 강수, 평균기온 등 7개 설명변수 후보 중에서 반응변수인 수확량과의 상관계수가 일정수준 이상인 변수들을 일차 선택하였고, 설명변수 간의 선형종속성 여부를 확인하기 위하여 분산팽창계수(Variation Index Factor: VIF)를 이용한 다중공선성(multicollinearity) 검토를 거쳐 최종적으로 설명변수를 선정하였다.

3. 결과 및 고찰

MODIS 위성자료와 PRISM 기후자료를 이용하여 옥수수와 콩의 수확량을 추정하는 OLS 회귀모형을 구축하기 위해서 2001~2010년 10년간의 위성영상과 기후자료를 카운티 단위로 집계하고 설명변수 선정을 수행하였다. EVI, FPAR, GPP, LAI, NDVI, PPT(강수), Tmean(평균기온) 등 7개 설명변수와 수확량과의 상관계수는 모두 일정정도(0.5) 이상이었다(Table 3, 4). EVI와 NDVI의 VIF는 10~12 정도로서 다중공선성 여부를 판단하는 $VIF > 10$ 의 기준(Kutner *et al.*, 2004)을 매우 근소하게 상회하였으므로, 이들을 설명변수에 포함하는 모형과 설명변수에서 제외하는 모형을 병행하여 분석하기로 하였다. 각 회귀모형에서 설명변수

들은 모두 표준점수(z-score)로 정규화함으로써 각 설명변수가 반응변수에 미치는 영향력이 상호비교 가능하도록 하였다. Table 5는 7개 변수를 모두 사용한 OLS 회귀분석 결과로서 설명변수들의 회귀계수 및 결정계수(R^2)를 나타낸 것이다.

결정계수 R^2 은 옥수수에서 0.790, 콩에서 0.775로 산출됨으로써 회귀모형의 설명력이 양호한 것으로 판단된다. 표준화된 설명변수들 중에서 가장 높은 회귀계수를 가진 EVI는 식생의 활력도를 나타내는 식생지수로서 식물의 성장과 밀접한 관계가 있으며, 식생밀도가 조밀한 곳에서는 NDVI나 LAI보다 EVI가 더 민감하게 반응하기 때문에 옥수수와 콩 수확량에 대해 가장 큰 양의 관계를 나타낸 것으로 보인다. 선행연구를 통해 볼 때 (Bala, 2008; Fu *et al.*, 2014; Zhang *et al.*, 2014)

Table 3. Correlation coefficient and VIF(gray cells) for variables of the corn yield model.

	Yield	EVI	FPAR	GPP	LAI	NDVI	PPT	Tmean
Yield		11.911	7.339	3.270	3.420	10.765	2.354	1.625
EVI	0.854							
FPAR	0.731	0.872						
GPP	0.577	0.632	0.721					
LAI	0.740	0.766	0.804	0.567				
NDVI	0.839	0.936	0.863	0.695	0.702			
PPT	0.566	0.532	0.496	0.695	0.433	0.597		
Tmean	0.424	0.535	0.470	0.304	0.298	0.479	0.385	

Table 4. Correlation coefficient and VIF(gray cells) for variables of the soybeans yield model.

	Yield	EVI	FPAR	GPP	LAI	NDVI	PPT	Tmean
Yield		12.262	7.445	3.279	3.448	10.898	2.342	1.674
EVI	0.842							
FPAR	0.728	0.875						
GPP	0.561	0.637	0.722					
LAI	0.686	0.767	0.803	0.568				
NDVI	0.799	0.937	0.862	0.698	0.698			
PPT	0.567	0.532	0.495	0.694	0.431	0.598		
Tmean	0.620	0.555	0.498	0.325	0.315	0.502	0.389	

Table 5. Regression coefficients and R-squared of the regression model.

	Corn	Soybeans
β_0	9.366	2.871
β_{EVI}	0.914	0.338
β_{FPAR}	-0.650	-0.141
β_{GPP}	-0.108	-0.017
β_{LAI}	0.717	0.166
β_{NDVI}	0.844	0.097
β_{PPT}	0.257	0.085
β_{Tmean}	0.004	0.202
R^2	0.790	0.775

($p < 0.05$ except for GPP in soybeans)

FPAR과 GPP는 일반적으로 곡물수확량과 양의 관계를 가지지만 이 모형에서 회귀계수가 음으로 나온 이유에 대해서는 향후 추가적인 조사가 필요할 것이나, 우선적으로는 여러 식생지수 등 다른 설명변수의 영향에 의한 것으로 파악된다. 수확량과 FPAR의 상관계수는 0.73 정도였고, 수확량과 GPP는 0.57 내외의 상관계수를 나타내었으며 (Table 3, 4), FPAR와 기후요소들의 조합에 의한 다중회귀식($Y = \beta_0 + \beta_{FPAR}X_{FPAR} + \beta_{PPT}X_{PPT} + \beta_{Tmean}X_{Tmean}$)에서 $\beta_{FPAR} = 1.195$, GPP와 기후요소들의 조합에 의한 다중회귀식($Y = \beta_0 + \beta_{GPP}X_{GPP} + \beta_{PPT}X_{PPT} + \beta_{Tmean}X_{Tmean}$)에서 $\beta_{GPP} = 0.937$ 로서 강한 양의 관계가 나타난 것으로 볼 때, 본 연구의 분석결과 역시 선행연구들의 연장선상에 있는 것으로 사료된다. 추정컨대 세 종류의 식생지수(NDVI, EVI, LAI)가 다중공선성 여부의 애매한 경계에 있으나 실제로는 유사한 변수들이므로 다중회귀식에서 상호작용을 일으키면서, 상대적으로 보다 독립적인 FPAR나 GPP의 회귀계수에 영향을 주었을 것으로 보인다. 이러한 영향을 최소화하기 위하여 주성분회귀(Principal Component Regression: PCR)가 시도되기도 하므로 향후 연구에서 참고할 필요가 있겠

다.

위의 7개 설명변수로 구성된 모형은 설명력 향상에 주력한 것으로서, 이를 이용하여 2011년과 2012년의 옥수수와 콩의 수확량을 추정하고 실제 통계치와 비교 검증하였다(Table 6). 옥수수의 2011년과 2012년의 RMSE(Root Mean Squared Error, ton/ha)는 각각 0.892와 1.095로서 추정치 평균에 대하여 9.4%와 12.9% 정도이고, 콩의 RMSE는 2011년 0.320, 2012년 0.391로서 추정치 평균에 대하여 11.1%와 14.2% 정도로 나타나 오차가 크지 않음을 알 수 있다. 또한 실제 통계치와 모형 추정치와의 상관계수는 옥수수의 경우 2011년과 2012년에 0.930과 0.884, 콩의 경우 0.906과 0.904로서 높은 상관성을 보였다. Figure 4와 5는 2011년과 2012년의 옥수수와 콩의 통계치, 추정치 및 그 차이를 나타낸 것이다. Table 6의 평균편이(mean bias) 및 Figure 4와 5의 (e)와 (f), Figure 6을 통해 볼 때 약간 과소추정의 경향이 모형에 존재함을 알 수 있다.

또한 전술한 바와 같이 유사한 세 가지 식생지수(NDVI, EVI, LAI)가 독립적으로 수확량에 미치는 영향을 살펴보기 위하여 각 식생지수와 FPAR 및 기후요소들로 설명변수를 구성하여 회귀분석을 수행하였고(Table 7) 그 결과는 Table 8, 9, 10에 나타내었다. 이때 GPP는 7개 설명변수를 모두 사용한 모형에서 p-value가 0.05보다 크거나 그에 가까운 값(0.046)을 보였기 때문에 통계적 유의성을 위하여 설명변수에서 제외하였다. NDVI, EVI, LAI를 사용한 회귀모형의 설명력은 각각 0.711, 0.751, 0.655로서 다소 상이한 값을 보였으며, NDVI와 LAI를 사용한 모형에서 FPAR와 평균기온의 영향력이 양의 값으로 나타났다. 회귀모형의 설명력과 각 설명변수의 영향력 분석이라는 두 측면에서 보았을 때, 세 가지 식생지수 중에서는 NDVI를 사용하는 것이 추천되어지며, 이에 대

Table 6. Validation results of the corn and soybeans yield (ton/ha) for 2011 and 2012.

	No. of Counties	Min Estimation	Max Estimation	Mean Estimation	Mean Bias	RMSE	R
Corn(2011)	180	4,871	12,334	9,456	-0,560	0,892	0,930
Corn(2012)	182	1,883	12,208	8,463	-0,527	1,095	0,884
Soybeans(2011)	181	1,715	4,304	2,871	-0,171	0,320	0,906
Soybeans(2012)	177	0,780	3,766	2,762	-0,241	0,391	0,904

Figure 4. Observed corn yields(a, b), predicted corn yields(c, d), and differences between prediction and observation(e, f) in 2011 and 2012.

Figure 5. Observed soybeans yields(a, b), predicted soybeans yields(c, d), and differences between prediction and observation(e, f) in 2011 and 2012.

해서는 향후 보다 심도 있는 민감도 분석 등이 수
반되어야 할 것이다.

4. 결론

본 연구에서는 2001~2010년의 MODIS 위성자
료와 PRISM 기후자료를 이용하여 OLS 회귀모형
을 세우고 미국 중서부 4개 주에 대한 옥수수와 콩

Figure 6. Observed vs. predicted corn and soybeans yield in 2011(a, c) and 2012(b, d).

Table 7. Regression models using simply one vegetation index (NDVI, EVI, or LAI).

	Regression Model
Eq.1	$Y = \beta_0 + \beta_{NDVI} X_{NDVI} + \beta_{PPT} X_{PPT} + \beta_{Tmean} X_{Tmean}$
Eq.2	$Y = \beta_0 + \beta_{EVI} X_{EVI} + \beta_{PPT} X_{PPT} + \beta_{Tmean} X_{Tmean}$
Eq.3	$Y = \beta_0 + \beta_{LAI} X_{LAI} + \beta_{PPT} X_{PPT} + \beta_{Tmean} X_{Tmean}$

의 수확량을 추정하였다. 대부분의 선행 연구들은 단일 식생지수와 기후요소를 이용하여 수확량 추정하였지만, 본 연구에서는 곡물 수확량에 영향을 미치는 여러 식생지수들과 함께 광합성유효복사율, 총일차생산량 및 기후변수 등 다양한 대

용량의 자료를 활용함으로써 보다 다각도의 분석을 수행하였다. 작성된 회귀모형을 이용하여 2011년과 2012년의 옥수수과 콩의 수확량을 추정한 결과, 옥수수의 RMSE는 0.892ton/ha와 1.095ton/ha, 콩의 RMSE는 0.320ton/ha와 0.391ton/ha를

Table 8. Regression coefficients and R-squared of the regression model using simply one vegetation index(NDVI).

	Corn	Soybeans
β_0	9,311	2,851
β_{NDVI}	1,509	0,338
β_{FPAR}	0,064	0,060
β_{PPT}	0,215	0,071
β_{Tmean}	0,036	0,221
R^2	0,711	0,712

Table 9. Regression coefficients and R-squared of the regression model using simply one vegetation index(EVI).

	Corn	Soybeans
β_0	9,372	2,869
β_{EVI}	1,782	0,467
β_{FPAR}	-0,165	-0,043
β_{PPT}	0,361	0,098
β_{Tmean}	-0,176	0,165
R^2	0,751	0,712

Table 10. Regression coefficients and R-squared of the OLS regression model using one vegetation index(LAI) except for GPP.

	Corn	Soybeans
β_0	9,337	2,859
β_{LAI}	0,947	0,245
β_{FPAR}	0,457	0,119
β_{PPT}	0,491	0,133
β_{Tmean}	0,263	0,280
R^2	0,655	0,695

나타냄으로써, 미국 중서부의 일반적인 수확량이 옥수수 9 ton/ha, 콩 3 ton/ha 정도임을 감안할 때 RMSE가 약 10%에 해당하여 비교적 양호한 검증 정확도를 보였다. 2012년의 옥수수 수확량 추정에서 RMSE와 상관계수가 상대적으로 좋지 않은 것은 그해 미국 중서부 지역에서 발생한 극심한 가

뭍의 영향 때문으로 사료된다. 옥수수와 콩은 주변 환경에 잘 적응하여 성장하지만, 옥수수는 키가 크고 잎이 넓은 작물로서 물을 많이 필요로 하기 때문에 콩에 비해 가뭄에 강하지 못하고 수분에 더욱 민감하게 반응한다(Classen and Shaw, 1970). 따라서 본 연구에서 도출된 OLS 회귀모형은 가뭄이나 홍수 등의 영향을 받은 특정 년도보다는 일반적인 경향에 의한 수확량 추정에 용이할 것으로 판단된다.

본 연구에서 사용된 식생지수인 MODIS NDVI, EVI, LAI 산출물은 NASA에서 복잡한 전처리 과정과 함께 자료에 영향을 주는 구름, 에어로졸, 수증기 등과 같은 잡음을 제거한 후 사용자에게 제공하고 있으나 일부 자료에는 오염된 화소가 존재하기 때문에 향후에는 이에 대한 추가적인 보정이 선행되어야 할 것이다. 또한 NDVI, EVI, LAI 등 유사한 식생지수의 존재로 인한 다중공선성 효과가 발생하지 않도록 설명변수로부터 주성분을 추출하여 회귀분석하는 주성분회귀를 도입함으로써 보다 신뢰성 있는 모형 구축이 가능할 것이다. 곡물 수확량은 식생의 상태나 기후요소 등의 영향을 받지만, 토양의 종류나 특성, 비료 및 관개 등의 양상에 따라 상당히 달라질 수 있다. 따라서 향후 곡물수확량 추정을 위한 연구에서는 위성영상과 기후자료뿐 아니라 재배곡물 유형(Cropland Data Layer), 토양 정보, 비료 투입, 관개 등의 보조자료들도 필수적으로 추가함으로써 곡물 성장에 미치는 영향력들에 대한 종합적인 평가를 수행함과 동시에 검증정확도를 보다 더 향상시킬 수 있을 것으로 기대한다.

사사

이 연구는 기상청 기상기술개발사업(CATER 2012-2069)의 지원으로 수행되었습니다.

참고문헌

- Allen, J.D., 1990, A look at the remote sensing applications program of the National Agricultural Statistics Service, *Journal of official Statistics*, 6(4), 393-409.
- Bala, S.K., 2008, *Estimation of potato yield in and around munshiganj using remote sensing NDVI data*, MA, Thesis, Bangladesh university of engineering and technology, Bangladesh, 40-59.
- Becker, R., Vermote, E., Lindeman, M., and Justice, C., 2010, A generalized regression-based model for forecasting winter wheat yield in 94 Kansas and Ukraine using MODIS data, *Remote sensing of environment*, 114(6), 1312-1323.
- Classen, M.M. and Shaw, R.H., 1970, Water deficit effects on corn. I. grain components, *Agronomy Journal*, 62, 652-655.
- Daly, C., Halbleib, 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.
- Daly, C., Neilson, R.P., and Phillips, D.L., 1994, A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *Journal of Applied Meteorology*, 33(2), 140-158.
- 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, 192-202.
- Duchemin, B., Maisongrande, P., Boulet, G., and Benhadj, I., 2008, A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat monitored with green leaf area index, *Environmental Modelling, Software*, 23(7), 876-892.
- Ferencz, C., Bongnár, P., Lichtenberger, J., Hamar, D., Tarcsai, G., Timár, G., Molnár, G., Pásztor, S.Z., 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.
- Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X., 2010, MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, *Remote Sensing of Environment*, 114, 168-182.
- Fu, X., Tang, C., Zhang, X., Fu, J., and Jiang, D., 2014, An improved indicator of simulated grassland production based on MODIS NDVI and GPP data: A case study in the Sichuan province, China, *Ecological Indicators*, 40, 102-108.
- Hong, S.Y., Hur, J., Ahn, J.-B., Lee, J.-M., Min, B.-K., Lee, C.-K., Kim, Y., Lee, K.D., Kim, S.-H., Kim, G.Y., and Shim, K.M., 2012, Estimating rice yield using MODIS NDVI and meteorological data in Korea, *Korean Journal of Remote Sensing*, 28(5), 509-520 (in Korean).
- Huete, A.R., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., and Ferreria, L.G., 2002, Overview of the radiometric and biophysical performance of the MODIS vegetation indices, *Remote Sensing of Environment*, 83, 195-213.
- Kastens, J.H., Kastens, T.L., Kasten, D.L.A., Price, K.P., Martinko, E.A., and Lee, R., 2005, Image masking for crop yield forecasting using AVHRR NDVI time series imagery, *Remote Sensing of Environment*, 99(3), 341-356.
- Kutner, M.H., Nachtsheim, C.J., and Neter, J., 2004, *Applied Linear Regression Models (4th ed.)*, McGraw-Hill Irwin.
- Lee, J.T., Lee, C.W., Joo, M.K., and Hong, S.Y., 1997, Estimation of rice canopy leaf area index(LAI) by spectral reflectance of solar radiation in pad-

- dy field, *Korean Journal of Crop Science*, 42(2), 173-181 (in Korean).
- Liu, H.L., Yang, J.Y., Drury, C.F., Reynolds, W.D., Tan, C.S., Bai, Y.L., He, P., Jin, J., and Hoogenboom, G., 2011, Using the DSSAT-CERES-Maize model to simulate crop yield and nitrogen cycling in fields under long-term continuous maize production, *Nutrient Cycling in Agroecosystems*, 89, 313-328.
- Liu, J., Williams, J.R., Zehnder, A.J.B., and Yang, H., 2007, GEPIC-modeling wheat yield and crop water productivity with high resolution on a global scale, *Agricultural systems*, 94, 478-493.
- Ministry for Food, *Agriculture, Forestry and Fisheries*, 2011, Statistical Yearbook.
- Na, S.I., Hong, S.Y., Kim, Y.H., Lee, K.D., and Jang, S.Y., 2013, Prediction of rice yield in Korea using paddy rice NPP index, *Korean Journal of Remote Sensing*, 29(5), 461-476 (in Korean).
- Na, S.I., Park, J.H., and Park, J.K., 2012, Development of Korean Paddy Rice yield Prediction Model (KRPM) using meteorological element and MODIS NDVI, *Journal of the Korean Society of Agricultural Engineers*, 54(3), 141-148 (in Korean).
- Ren J.Q., Chen, Z.X., Zhou, Q.B., and Tang, H.J., 2008, Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China, *International Journal of Applied Earth observation and Geoinformation*, 10, 403-413.
- Sellers, P.J., Mintz, Y., Sud, Y.C., and Dalcher, A., 1986, A simple biosphere model(SiB) for use within general circulation models, *Journal of Atmospheric Science*, 43, 505-531.
- Tan, G. and Shibasaki, R., 2003, Global estimation of crop productivity and the impacts of global warming by GIS and EPIC integration, *Ecological Modelling*, 168(3), 357-370.
- Townshend, J.R.G., Justice, C.O., Li, W., Gurney, C., and McManus, J., 1991, Global land cover classification by remote sensing: present capabilities and future capabilities, *Remote Sensing of Environment*, 35, 243-255.
- USDA/FAS, 2012, *Grains: World Markets and Trade Archives*, USDA/PAS.
- Vandiepen, C.A., Wolf, J., Vankeulen, H., and Rappoldt, C., 1989, WOFOST: A simulation model of crop production, *Soil Use and Management*, 5(1), 16-24.
- Vanevert, F.K. and Campbell, G.S., 1994, CropSyst: A collection of object-oriented simulation models of agricultural systems, *Agronomy Journal*, 86(2), 325-331.
- Wang, X.C. and Li, J. 2010, Evaluation of crop yield and soil water estimates using the EPIC model for the Loess Plateau of China, *Mathematical and Computer Modelling*, 51, 1390-1397.
- Wu, W., Shibasaki, R., Yang, P., Tan, G., Matsumura, K., and Sugimoto, K., 2007, Global-scale modelling of future changes in sown areas of major crops, *Ecological Modelling*, 208(2), 378-390.
- Yoo, S.-H., Im, J.-B., Choi, J.-Y., and Lee, S.-H., 2012, Estimation of agricultural water and land required to substitute the import of feed-grain for domestic production, *Journal of the Korean Society of International Agriculture*, 24(3), 259-264 (in Korean).
- Zhang, Q., Cheng, Y.-B., Lyapustin, A.I., Wang, Y., Gao, F., Suyker, A., Verma, S., and Middleton, E.M., 2014, Estimation of crop gross primary production(GPP): fAPARchl versus MOD15A2 FPAR, *Remote Sensing of Environment*, 153, 1-6.

책임 편집: 장동호

영문 교열: 조창현