

# 스마트 농업을 위한 생산량 예측 방법

이준구\* · 문애경\*

\*한국전자통신연구원

## Yield Forecasting Method for Smart Farming

Joon-goo Lee\* · Aekyung Moon\*

\*Electronics and Telecommunications Research Institute

E-mail : leejg01679@etri.re.kr

### 요 약

최근 심각한 기후변화로 인하여 농산물 생산성 및 농산물 가격의 변동성이 커지고 있으며, 이러한 문제를 해결하기 위하여 농산물의 생산량을 예측하는 방법이 연구되고 있다. 본 논문에서는 선형 방정식을 이용하여 생산 단수, 생산 면적, 기후 요소를 예측하였으며, 기후 요소의 가중치 합을 이용하여 보정된 생산 단수와 생산 면적을 곱하여 생산량을 계산하였다. 실험에서 예측한 생산량은 실제 생산량과 비교하여 약 90% 이상의 예측율을 보였다.

### ABSTRACT

Recently, there are growing fluctuations of productivity and price caused by severe weather conditions in the agriculture. Yield forecasting methods have been studied to solve the problems. This paper predicted yield per area, production area, and elements of weather based on the linear equation. A yield is calculated by multiplying the production area times the yield per area that is compensated using the weighted sum of the elements of weather. In experiments, proposed method shows that a forecasting precision is the more than 90%.

### 키워드

스마트 파밍, 생산량 예측, 기후 변화, 선형 방정식, 스마트 농업

### 1. 서 론

최근 이상 기온으로 인한 재배 환경의 변화, 특정 작물의 과잉 재배 및 수급(출하량 조절) 불균형으로 인한 가격 급등락 등의 이유로 작물의 생산량에 대한 관심이 높아지고 있다. 이러한 문제들은 생산량을 예측함으로써 재배 환경의 변화에 대응할 수 있고, 과잉 재배 시 대체 작물을 재배하거나 출하량을 조절함으로써 가격 급등락에 대처할 수 있다[1].

생산량 예측에는 크게 통계학적 방법과 생물리학적 방법을 이용한다. 통계학적 방법에는 결과에 영향을 미치는 독립 변수와 그에 대한 반응 및 결과로서의 종속 변수를 설정하여 계산하는 선형 또는 비선형 회귀 등의 방법, 과거 학습 자료를 통해 학습 모델을 만들고 이를 예측에 이용하는

기계학습 등의 방법이 있다. Becker-Reshef 등은 회귀를 기반으로 겨울철 밀의 생산량을 예측하였으며[2], Sujjaviriyasup 등은 SVM(Support Vector Machine)을 함께 고려하여 태국의 난초에 대하여 예측하였다[3]. 생물리학적 방법은 DSSAT, Ecocrop 등의 모델을 이용하여 작물이 재배되기 위한 최저, 적정, 최대 온도 구간 등의 재배 생리와 관련된 생육 정보들을 사용하여 계산한다. Shim 등은 지구 온난화에 따른 벼 생육 및 생산성 변화 예측에 DSSAT CERES-Rice 모델을 적용하였으며[4], Lane 등은 Ecocrop 모델을 사용하여 기후 변화에 따른 적정 재배지역의 변화와 대안에 대하여 연구하였다[5]. 하지만 회귀 및 기계학습은 과거 자료의 추세만을 보거나 학습 시간이 오래 걸리는 등의 문제가 있고, DSSAT 및 Ecocrop은 입력 파라미터가 너무 많이 필요하거

나 예측율이 떨어지는 등의 문제가 있다. 또한 참외의 경우 국내 생산량의 90%가 성주에서 재배되고 있어 수급 조정 정책 및 대구 경북지역의 온도 상승률에 대한 대책이 필요함에도 불구하고 생산량 예측에 대한 연구가 미진한 실정이다.

따라서 본 논문에서는 선형 방정식과 상관 계수를 기반으로 계산된 가중치를 이용하여 간단하면서도 예측율이 높은 참외의 생산량을 예측하는 방법을 제안한다.

## II. 본 론

본 논문에서는 선형 방정식을 이용해 생산 단수와 생산 면적의 선형 추세를 각각 예측하여 곱한 값으로 생산량을 예측한다[6]. 생산량을  $Y$ (Yield), 생산 단수를  $YPA$ (Yield Per Area), 생산 면적을  $PA$ (Production Area)라고 할 때, 각각의 관계는 다음과 같다.

$$Y = YPA \times PA \quad (1)$$

따라서 예측된 생산 단수와 생산 면적을 곱함으로써 생산량을 예측할 수 있다. 각 데이터들은 단위 및 위계가 맞지 않아 계산시 문제가 발생할 수 있기 때문에 0.0~1.0 사이의 값으로 정규화 하여 사용한다.

생산 단수는 최대 온도, 평균 온도, 최저 온도, 강수량, 일조량, 토양, 질병 및 해충, 생육 정보 등 많은 요소들의 복잡한 관계를 통해 결정된다. 통계적으로 생산 단수의 추세와 상관관계가 있는 정보를 추출하기 위해 피어슨 적률상관계수(Pearson Product Moment Correlation)를 계산하여 비교한 후 생산 단수와 가장 상관 관계가 높은 최대 온도, 평균 온도, 최저 온도를 선별하였다. 선별된 요소들은 생산 단수를 예측하기 위한 입력 벡터를  $YPA = \{opa_{1980}, opa_{1981}, opa_{1982}, \dots, opa_{2012}\}$ , 최대 온도를  $MT$ (Max Temperature) =  $\{mt_{1980}, mt_{1981}, mt_{1982}, \dots, mt_{2012}\}$ , 평균 온도를  $AT$ (Average Temperature) =  $\{at_{1980}, at_{1981}, at_{1982}, \dots, at_{2012}\}$ , 최저 온도를  $LT$ (Low Temperature) =  $\{lt_{1980}, lt_{1981}, lt_{1982}, \dots, lt_{2012}\}$ 로 설정하였다.

선형 방정식을 이용하여 생산 단수를 예측하려면 과거 추세를 보기 위한 증감 변화도와 절편의 값을 계산해야 한다. 예측 년도가  $t$ 이고, 예측에 사용될 과거 기간이  $n$ 일 때,  $t$ 년도 생산단수를 예측하기 위한  $n$ 년간의 증감 변화도는 다음과 같이 계산한다.

$$gradient(t) = \frac{y_{pa_{t-1}} - y_{pa_{t-n}}}{(t-1) - (t-n)} \quad (2)$$

이때,  $n$ 이 너무 짧으면 전체적인 추세를 고려하지 못하게 되고, 너무 길면 갑작스럽게 변화하는 이상 패턴을 감지하지 못하기 때문에 실험을 통해

적절한 길이를 선택해야 한다. 생산 단수의 절편은 다음과 같이 계산한다.

$$intercept(t) = -gradient(t) \times (t-1) + y_{pa_{t-1}} \quad (3)$$

계산된 증감 변화도와 절편을 이용한 선형 연장을 통해 다음과 같이  $t$ 년의 생산 단수를 예측할 수 있다.

$$y_{pa_t} = gradient(t) \times t + intercept(t) \quad (4)$$

여기서, 식 (2) ~ 식 (4)에 요소를 변경하여 동일한 방법으로  $mt_t$ ,  $at_t$ ,  $lt_t$ 를 계산함으로써 과거 데이터의 추세를 이용해  $t$ 년도의 생산 단수와 보정을 위한  $t$ 년도의 최대, 평균, 최소 온도를 예측한다.

추세만을 이용하여 예측된  $t$ 년도의 생산 단수  $y_{pa_t}$ 는 밀접한 상관 관계를 갖는 최대, 평균, 최소 온도를 이용해 보정해야 한다. 이를 위해 앞서 계산한 상관 계수를 가중치로 하여 보정하며,  $w_1 \sim w_4$ 를 각각의 가중치라 할 때 보정된  $t$ 년도의 생산 단수는 다음과 같이 계산한다.

$$y_{pa}(t) = y_{pa_t} \times w_1 + mt_t \times w_2 + at_t \times w_3 + lt_t \times w_4 \quad (5)$$

여기서, 모든 가중치의 합은 1이 되며, 생산 단수의 추세와 관계가 높은 기온들의 추세가 가중치 별로 함께 고려되어 예측율에 대한 신뢰도를 높일 수 있다.

생산 면적의 경우 상관 관계를 갖는 다른 데이터가 없기 때문에 단독으로 선형 방정식을 이용하여  $t$ 년도를 예측한다. 예측된  $YPA(t)$ 는 역정규화를 이용해 원래의 위계로 변경한 후 식 (1)과 같이  $PA(t)$ 와 곱하여  $t$ 년도의 생산량을 계산한다.

## III. 실험 결과

제안하는 방법에 대한 분석 및 실험을 위해 통계청의 '농작물 생산조사' 데이터와 기상청의 '전국 년월별 평균' 데이터를 사용하였다. '농작물 생산조사' 데이터는 1980년부터 2012년까지의 전국 및 16개 시도별 생산 면적과 생산량이 기록되어 있다. '전국 년월별 평균' 데이터는 1980년 1월부터 2012년 12월까지 최대 온도, 평균 온도, 최저 온도, 일조량, 강우량의 전국 월별 평균치가 기록되어 있다. 통계 자료 분석 및 프로그램을 위해 R 통계 분석 언어와 R Studio 개발 환경을 사용하였다.

먼저 사용할 데이터를 선택하기 위해 수집 데이터들 간의 상관 관계를 분석하여 표 1과 같이 의미 있는 계수들을 정리하였다. 생산 단수와 상관 관계가 높은 최고, 평균, 최저 온도 데이터 생산 단수 예측을 위해 사용하였으며, 생산 면적은 상관관계가 뚜렷하게 나타나지 않아 단독으로 사용하였다.

$t$ 년을 예측하기 위한 과거  $n$ 년의 길이를 설정하기 위해 데이터를 분석한 결과 생산 단수는 평

군 3년을 주기로 값이 크게 변했다. 따라서 평년 3년을 기준으로 하고, 변화폭이 커질 수 있는 앞 뒤 1년씩을 추가하여 5년으로 개산하였다. 생산 면적의 경우 7년을 주기로 값이 크게 변했으므로 앞 뒤 1년씩을 추가하여 9년으로 계산하였다. 즉,  $t$ 년을 예측하기 위해 생산 단수는 직전년도로부터 과거 5년의 추세를 이용하고, 생산 면적의 경우 직전년도로부터 과거 9년의 추세를 이용한다.

표 1. 주요 변수들 사이의 피어슨 적률 상관계수

	생산량	생산면적	생산단수	최대온도	평균온도	최저온도
생산량	1.00	0.67	0.69	0.36	0.25	0.08
생산면적	0.67	1.00	-0.05	0.05	-0.12	-0.28
생산단수	0.9	-0.05	1.00	0.45	0.49	0.42
최대온도	0.36	0.05	0.45	1.00	0.89	0.62
평균온도	0.25	-0.12	0.49	0.89	1.00	0.90
최저온도	0.08	-0.28	0.42	0.62	0.90	1.00

선형 방정식을 이용해 생산단수의 추세를 예측하여 생산 단수에 대한 실제 값, 생산 단수 변수만을 사용한 예측 1, 예측된 최대, 평균, 최저 온도 변수들을 함께 고려한 예측 2, 실제 값과 예측 1에 대한 예측율 1, 실제 값과 예측 2에 대한 예측율 2를 표 2에 나타냈다. 2003년부터 2012년 까지 최근 10년의 실제 생산 단수와 예측 1의 결과 값에 대한 예측율1은 평균 93.22%로 나타났으며, 특히 2009년의 경우 실제 생산 단수와 비교하여 98.67%의 매우 높은 예측율을 보였다. 하지만 2007년의 경우 과거 실제 생산 단수 보다 오버 피팅 되어 약 97.69%의 오차율을 보였다. 이러한 오버 피팅은 상관성이 있는 데이터들과의 가중치 합을 이용해 완화시킬 수 있다. 예측 2와 예측율 2를 보면 평균 예측율은 약 93.72%로 선형 방정식만 이용했을 때보다 약 0.5%의 예측율을 증가시켰다.

표 2. 생산 단수에 대한 예측 결과

연도	실제 값	예측1	예측2	예측율1	예측율2
2003	30.96	31.61	31.25	97.91	99.05
2004	33.17	31.35	31.36	94.51	94.56
2005	28.23	33.31	32.78	82.01	83.88
2006	32.18	27.48	27.85	85.39	86.53
2007	31.74	32.47	31.47	97.69	99.16
2008	33.36	31.93	32.03	95.74	96.02
2009	33.85	33.40	31.90	98.67	94.24
2010	33.43	35.26	34.76	94.52	96.00
2011	30.76	33.74	32.47	90.32	94.45
2012	31.97	30.52	29.82	95.46	93.28
평균	31.96	32.11	32.57	93.22	93.72

생산 면적의 경우 표 3과 같이 변화 폭이 크지 않은 감소 패턴을 보이며 과거 추세가 비교적 단

순하여 9년간의 데이터 추세만 이용 하여도 평균 약 96.18%의 높은 예측율을 보였다.

표 3. 생산 면적에 대한 예측 결과

연도	실제 값	예측	예측율
2003	7,731.00	7,682.63	99.37
2004	7,329.00	7,197.50	98.21
2005	7,077.00	6,910.25	97.64
2006	6,827.00	6,662.25	97.59
2007	6,472.00	6,378.88	98.56
2008	6,607.00	5,923.63	89.66
2009	6,730.00	6,157.50	91.49
2010	6,215.00	6,489.38	65.59
2011	5,852.00	5,995.88	97.54
2012	5,840.00	5,617.13	96.18
평균	6,668.00	6,501.50	96.18

생산 단수와 생산 면적을 곱하여 최종 예측된 생산량과의 비교를 위해 생산량에 대한 실제 값, 예측된 각 온도 변수를 함께 고려한 예측 1, 예측된 온도 변수가 아닌 그 해의 실제 온도를 고려한 예측 2, 실제 값과 예측 1 사이의 예측율 1, 실제 값과 예측 2 사이의 예측율 2를 표 4에 나타냈다. 예측 1의 경우 실제 값과의 예측율은 평균 90.66%로 높게 나타났지만 특정 연도에 대하여 예측율이 낮게 나타나는 경향을 보였다. 이러한 이유는 해당 연도의 실제 온도가 선형 방정식으로 예측한 온도와 큰 차이 값을 보였기 때문이다. 예측된 온도 대신 실제 온도를 적용한 예측 2와 그 예측율 2를 보면 2005, 2006년의 경우 모두 10% 이상 예측율이 증가하였으며, 평균 약 5%가 증가하여 95.75%의 예측율을 보였다.

표 4. 생산량에 대한 예측 결과

연도	실제 값	예측1	예측2	예측율1	예측율2
2003	239,360.00	240,110.60	238,087.00	99.69	99.47
2004	243,098.00	225,741.80	233,604.00	92.86	96.09
2005	199,785.00	226,530.30	196,551.10	86.61	98.38
2006	219,712.00	185,538.20	209,088.10	84.45	95.16
2007	205,416.00	200,749.70	202,983.70	97.73	98.82
2008	220,385.00	189,727.60	189,641.90	86.09	86.05
2009	227,832.00	196,449.20	206,275.60	86.23	90.54
2010	207,747.00	225,596.70	209,292.00	91.41	99.26
2011	180,013.00	194,676.20	181,202.20	91.85	99.34
2012	186,693.00	167,509.00	176,160.40	89.72	94.36
평균	213,004.10	205,262.93	204,288.60	90.66	95.75

#### IV. 결 론

본 논문에서 제안한 선형 방정식은 생산 단수 또는 생산 면적과 같이 일정한 증감 패턴이 있는 경우 용이하게 사용될 수 있다. 하지만 최대, 평균, 최소 온도의 경우 선형 방정식을 이용하기에

는 패턴의 주기가 너무 짧고 변화 폭이 매우 크기 때문에 적절하지 않았다. 따라서 예측율을 높이기 위해 기온을 예측하기 위한 적절한 방법의 연구가 추가되어야 할 것이다.

### 참고문헌

- [1] S. H. Chang, "Study on the prediction models for the productions of major food crops.," Journal of the Korean Data and Information Science Society, Vol. 11, No. 1, pp. 47-55, 2000.
- [2] I. Becker-Reshef, E. Vermote, M. Lindeman, Justice C., "A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data.," Remote Sensing of Environment, Vol. 114, No. 6, pp. 1312-1323, 2010.
- [3] T. Sujjaviriyasup, K. Pitiruek, Hybrid ARIMA-Support vector machine model for agricultural production planning.," Applied Mathematical Sciences, Vol. 7, No. 57, pp. 2833-2840, 2013.
- [4] K. M. Shim, K. A. Roh, K. H. So, G. Y. Kim, H. C. Jeong, D. B. Lee, "Assessing impacts of global warming on rice growth and production in Korea.," Climate Change Research, Vol. 1, No. 2, pp. 121-131, 2010.
- [5] A. Lane, A. Jarvis, "Changes in climate will modify the geography of crop suitability: Agricultural biodiversity can help with adaptation.," SAT eJournal, Vol. 4, No. 1, pp. 1-12, 2007.
- [6] Y. S. Lee, H. K. Jeong, W. T. Kim, I. C. Choi, "An estimation of yield functions of Korean fruit-vegetables.," Korean Rural Economic Institute, pp. 61, 2004.