

## 비정형 농업기상자료를 활용한 배추 도매가격 예측모형 연구<sup>†</sup>

장수희<sup>1</sup> · 전희주<sup>2</sup> · 조인호<sup>3</sup> · 김동환<sup>4</sup>

<sup>13</sup>THEIMC · <sup>2</sup>동덕여자대학교 정보통계학과 · <sup>4</sup>안양대학교 국제통상유통학과

접수 2017년 5월 2일, 수정 2017년 5월 23일, 게재확정 2017년 5월 25일

### 요약

주로 노지에서 재배되는 배추는 기상 여건에 따라 생산량의 변화가 크고, 대체 작물의 존재로 인해 가격 변동이 크게 나타난다. 기존의 연구에서는 실제 기상정보를 활용해 배추의 생산량을 예측하였으나, 본 연구에서는 실제 기상정보가 아닌 웹상의 비정형 농업기상 정보를 활용하여 도매가격을 예측하였다. 2009년 1월부터 2016년 10월까지 포털사이트에서 배추를 포함한 문서를 수집하여, 수집된 문서 내에 나타난 기상 관련 키워드를 추출하였다. 도매가격만을 이용해 자기회귀 (autoregressive; AR)모형으로 작형별 출하시기인 1, 5, 8, 11월을 예측한 단순모형과 비정형 농업기상 정보를 추가적으로 활용해 AR모형으로 예측한 농업기상모형을 비교하였다. 그 결과 비정형 농업기상 정보를 활용한 농업기상모형의 성능이 더 우수하고 예측력에 도움이 되는 것으로 나타났다.

주요용어: 도매가격, 배추, 비정형농업기상, 자기회귀모형.

### 1. 서론

우리나라는 가격안정을 위해 수급정책을 펴고 있으나, 노지에서 재배되는 채소류의 경우 기상 여건에 따라 생산량의 변화가 크고, 대체 작물의 존재로 인해 가격 변동이 크게 나타난다 (Nam과 Choe, 2015). 예를 들어, 2010년 10월 배추는 가락시장에서 한 포기에 12,410원으로 거래되었으나 2011년 5월에는 한 포기에 600 ~ 700원으로 폭락하였다. 정부는 가격 폭등으로 인한 재배면적 증가와 소비부진을 배추파동의 원인으로 분석하고, 1만 톤의 산지자율폐기를 추진하였다. 채소값 파동으로 2005년부터 2009년까지 배추, 대파, 마늘, 양파 등의 채소를 약 364,000톤 정도 산지 폐기하였으며, 이를 환산한 금액은 290억 원이 넘는 것으로 나타났다 (Kim과 Yoon, 2011).

또한 농산물유통정보에서 제공하는 월간 도매가격을 살펴보면 2015년 9월 배추의 경우 kg당 611원이었던 상품도매가격이 2016년 9월 2,104원으로 전년 대비 244% 상승하였고, 무 역시 2015년 9월 kg당 455원이었던 상품도매가격이 2016년 9월 1,186원으로 전년 대비 160% 상승하였다. 이와 같이 극심한 가격변동을 보이는 채소류의 경우 수급이나 가격 예측에 대한 연구는 이루어지고 있으나, 비정형데이터를 활용한 연구는 활발하지 않았다.

소셜미디어의 활용은 인터넷과 모바일의 확산으로 더욱 증대되고 있으며, 소셜미디어에서 생성되는 빅데이터는 새로운 분석을 가능하게 하고 있다. 웹에서 생성되는 데이터는 폭발적으로 증가하고 있는

<sup>†</sup> 본 연구는 농림축산식품부 농생명산업기술개발사업(과제번호: 514002-03)에 의해 지원됨.

<sup>1</sup> (04030) 서울시 마포구 동교로 142-16, 사원.

<sup>2</sup> 교신저자: (02748) 서울시 성북구 화랑로 13길, 동덕여자대학교 정보통계학과 부교수.

E-mail: hjchun@dongduk.ac.kr

<sup>3</sup> (04030) 서울시 마포구 동교로 142-16, 이사.

<sup>4</sup> (14028) 경기도 안양시 만안구 삼덕로 37번길 22.

추세지만, 이에 비해 비정형 데이터인 빅데이터를 활용하는 시도는 제한적으로 이루어지고 있다 (Das 등, 2013).

채소류 수급 및 가격 예측과 관련된 선행 연구로는 Mishra 등 (2013)의 ARIMA모형을 이용한 양파 생산량을 예측이 있다. 또한 Kim (2005)은 ARIMA 모형과 GARCH 모형을 이용하여 양파의 가락시장 도매가격을 예측했는데, 인공신경망보다 ARIMA와 GARCH모형의 예측력이 더 우수하게 나타났다. Park과 Park (2013)은 배추와 무의 작형별 재배면적반응함수와 단수함수를 추정해 총생산량을 계산하였다.

기상자료를 활용한 연구로 Nam과 Choe (2015)은 한국농수산식품유통공사의 도매가격 자료와 기상청의 기상자료, 통계청의 양파 생산자료로 자기회귀시차 (autoregressive and distributed lags; ARDL) 모형을 활용하여 재배면적과 단수를 예측하고, 양파의 출하시기 도매가격을 전망하였다. 또한 Choi와 Baek (2016)은 마늘 주산지 15곳의 기상자료를 활용하여 패널회귀모형으로 마늘의 생산량을 추정하였고, Lim 등 (2016)은 국내 채소작물 중 대표적으로 높은 생산 및 소비량을 보이는 다섯 가지 작물 (배추, 무, 고추, 마늘, 양파)을 주산지별 기상정보를 활용하여 회귀모형으로 단수예측모형을 개발하였다.

비정형 데이터를 활용한 연구로 Cho 등 (2016)은 웹상에서 나타나는 채소류 관련 비정형 데이터를 정량화하여 비정형지수를 구성하였으며, 비정형지수와 채소 구매량의 장기적인 관계를 연구하였으며 Hyun 등 (2015)은 기후변화와 식품관련 뉴스기사의 비정형 데이터 간의 관계를 연구하였다. Kang 등 (2015)은 텍스트마이닝을 이용한 감성분석을 이용하여 사회 이슈 찬반을 분류하였다.

기존의 연구에서는 정형데이터만을 이용하여 도매가격을 예측하거나 생산성을 예측하였으나, 본 연구에서는 비정형 농업기상 자료를 이용하여 채소류 중 배추의 상품 도매가격을 예측하고자 한다. 2009년 1월부터 2016년 10월까지의 포털사이트의 비정형 농업기상 데이터를 활용해 시계열모형을 구축하여, 기상자료의 변화에 따른 도매가격을 추정한다. 이는 빅데이터 상 비정형 자료가 예측변수로서의 활용이 가능하다는 것을 보이는 계기가 될 것이다.

## 2. 연구방법

### 2.1. 시계열 모형

시계열분석 (time-series analysis)은 시간의 흐름에 따라 동일한 관측주기로 관측된 시계열 자료를 분석하는 방법이다. 자료가 생성된 시스템을 이해하고 과거의 자료를 이용하여 시스템을 가장 잘 설명하는 모형을 찾아, 미래를 예측하는 것이 시계열분석의 목적이다 (Cho와 Lee, 2014). 본 연구에서는 현 시점의 변량이 과거의 변량들에 의해 영향을 받는 자기회귀모형 (autoregressive model, AR model)을 이용하였다.

$t$ 시점에서의 시계열  $Z_t$ 는  $p$ 개의 과거값들과 오차항  $\epsilon_t$ 의 선형결합으로 이루어진  $p$ 차 자기회귀모형식은 식 (2.1)과 같이 표현한다.

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \epsilon_t. \quad (2.1)$$

### 2.2. 연구 과정

본 연구에서 시행된 연구 과정은 다음과 같다.

- 1) 빅데이터 분석 솔루션인 텍스트롬(textom)을 이용하여 네이버 블로그, 카페, 뉴스, 웹 채널에서 ‘배추’ 키워드를 포함하고 있는 문서의 월별 데이터를 수집한다.

- 2) 수집된 문서 중 작물의 생장과 관련이 있는 농업기상 키워드를 선정하여 해당 키워드의 출현빈도를 월별 데이터로 구성한다.
- 3) 농업기상 키워드의 출현빈도는 연간 문서량을 이용하여, 2016년 문서량을 기준으로 표준화한다.
- 4) 농산물유통정보의 배추 상품도매가격과 비정형 농업기상데이터를 시계열자료로 병합한다.
- 5) 도매가격의 그래프를 통해 추세를 확인하고, DW (Durbin-Watson) 검정을 통해 자기상관이 있는지 확인한다.
- 6) DF 단위근검정 (Dickey-Fuller unit root tests)을 시행하여 정상성을 확인한다.
- 7) 도매가격 자료만을 이용하여 자기회귀모형을 적합한다.
- 8) 도매가격과 비정형 농업기상 자료의 교차상관분석을 통해 선행하는 농업기상을 선별하고, 선행하는 시차의 비정형 농업기상을 결합하여 자기회귀모형을 적합한다.
- 9) 단순모형과 농업기상모형의 AIC, 예측오차 등을 통해 모형의 성능을 비교한다.

### 3. 실증 분석

#### 3.1. 데이터

도매가격 자료는 농산물유통정보 (Kamis)에서 제공하는 2009년 1월부터 2016년 11월까지의 월간 상품도매가격 (kg/원)을 이용하였다. 배추는 일 년에 4번 수확되기 때문에 각각의 작형에 따라 1월에서 4월은 월동, 5월에서 7월은 봄, 8월에서 10월은 고랭지, 11월에서 12월은 가을 작물의 상품 도매가격을 사용하였다.

비정형 자료는 빅데이터 분석 솔루션인 텍스트롬 (textom)을 이용하여 포털사이트 (네이버)의 2009년 1월부터 2016년 10월까지 '배추' 키워드가 포함된 문서를 수집하였으며, 수집된 문서에서 작물 생육에 영향을 미치는 풍수해, 냉해, 동상해, 습윤해 등 농업기상재해 관련 키워드를 선정하였다.

#### 3.2. 단순모형

Figure 3.1은 2009년 1월부터 2016년 10월까지 월별 배추 상품 도매가격이며, 시간에 따른 추세는 없는 것으로 보인다. 또한 배추 도매가격은 월에 따라 다르게 나타나는데, Figure 3.2를 보면 3, 4, 8, 9월에 가격이 높은 것을 알 수 있다. 이를 모형에 반영하기 위해 모형에 3, 4, 8, 9월의 가변수를 적용하였다.

DW검정 결과 배추 상품 도매가격은 자기상관이 있으며 ( $DW = 0.6954$ ,  $p\text{-value} < .0001$ ), DF 단위근검정 결과 배추 상품 도매가격은 유의수준 0.1에서 단위근이 없는 것으로 나타나( $p\text{-value} = 0.0817$ ), 정상화 과정을 거치지 않고 AR모형의 적합이 가능하는 것으로 나타났다. 2016년 10월까지의 데이터를 사용하여 자기회귀모형에 적합한 결과 (Table 3.1), AR모형 ( $p = 1, 5, 12$ )이 선택되었으며, 잔차는 백색잡음 과정을 만족하였다. 추정된 모형은 식 (3.1)과 같으며, 여기서 March, April, August, September는 해당 월의 가변수이다. 추정된 모형을 이용한 2016년 11월 예측값은 972.7원이며, 실제 가격은 847원이다.

$$\begin{aligned} \hat{Z}_t = & 695.0646 + 167.8650 * \text{March} + 260.2382 * \text{April} + 266.9272 * \text{August} \\ & + 442.3177 * \text{September} - 0.6244Z_{t-1} - 0.1866Z_{t-5} + 0.2438Z_{t-12}. \end{aligned} \quad (3.1)$$

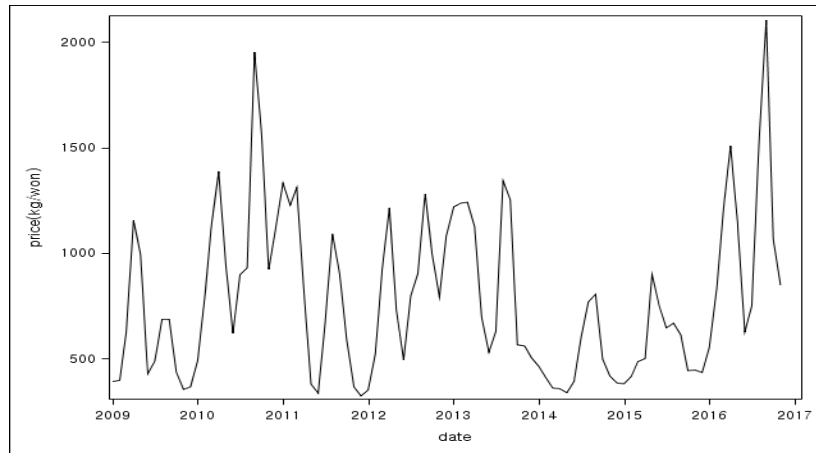


Figure 3.1 Time series data of cabbage wholesale price

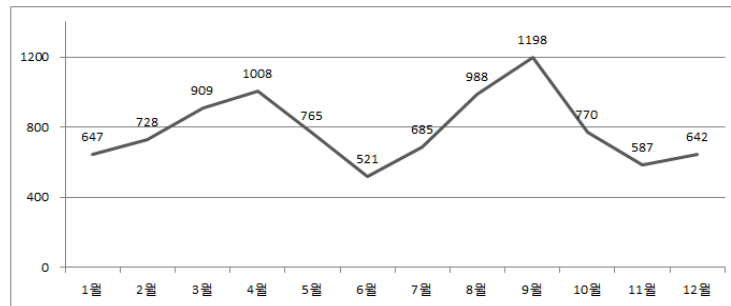


Figure 3.2 Monthly average wholesale price of cabbage

Table 3.1 The simple model

Variable	Df	Estimate	S.E.	t value	p-value
intercept	1	695.0646	58.8896	11.8	<.0001
March	1	167.865	66.1478	2.54	0.013
April	1	260.2382	65.9206	3.95	0.0002
August	1	266.9272	66.0233	4.04	0.0001
September	1	442.3177	66.2062	6.68	<.0001
AR1	1	-0.6244	0.0735	-8.5	<.0001
AR5	1	-0.1866	0.0793	-2.35	0.0209
AR12	1	0.2438	0.0795	3.07	0.0029
Obs.				94	
AIC				1298.67428	
AICC				1300.3684	

### 3.3. 농업기상모형

농업기상모형은 단순모형 절차를 따르되, 선정된 비정형 농업기상 자료를 활용하였다. 각 키워드의 빈도는 Table 3.2의 연간 문서량을 이용하여 2016년 문서량을 기준으로 표준화한 값을 사용하였다. Table 3.3은 표준화된 농업기상 키워드의 출현빈도와 배추 상품 도매가격의 교차상관분석 결과 선행하는 상관성을 가지는 키워드이며, 선행하는 시차의 농업기상 키워드 빈도를 설명변수로 사용하였다.

**Table 3.2** Documents collected by year

year	2009	2010	2011	2012	2013	2014	2015	2016
documents	106,965	157,559	192,839	199,425	260,619	393,358	508,674	558,233

**Table 3.3** Keywords and preceding time lags prior to price

keywords	cold-weather damage	low temperature	frost damage	cold wave
lags	5	1	1	2
keywords	heat wave	heavy rain	downpour	strong wind
lags	1	1	1	4

2016년 10월까지의 데이터를 사용하여 자기회귀모형에 적합한 결과 (Table 3.4), AR(1)모형이 선택되었으며, 잔차는 백색잡음 과정을 만족하였다. 추정된 모형은 식 (3.2)와 같으며, 여기서 March, April, August, September는 해당 월의 가변수, V1은 frost damage, V2는 cold wave, V3는 heat wave이다. 추정된 모형식을 이용한 2016년 11월 예측값은 782.1원이며, 실제 가격은 847원이다. 농업기상모형의 AIC와 AICC는 각각 1272.65, 1275.06로 단순모형의 AIC = 1298.67와 AICC = 1300.37보다 더 낮게 나타나 우수한 성능을 보였다. 예측오차의 크기 역시 농업기상모형은 64.9로 단순모형의 125.7보다 낮아 예측력 역시 더 우수했다.

**Table 3.4** The unstructured agricultural weather model

Variable	Df	Estimate	S.E.	t value	p-value
intercept	1	836.3145	103.5127	8.08	<.0001
March	1	100.2882	86.5432	1.16	0.2499
April	1	244.3836	79.4724	3.08	0.0028
August	1	274.8712	78.5987	3.5	0.0008
September	1	332.9209	95.4446	3.49	0.0008
lag1(frost damage)	1	-2.67	1.0246	-2.61	0.0109
lag2(cold wave)	1	0.5119	0.2449	2.09	0.0396
lag1(heat wave)	1	0.6344	0.2877	2.2	0.0302
AR1	1	-0.6435	0.09	-7.15	<.0001
Obs.	92				
AIC					1272.86619
AICC					1275.06131

$$\hat{Z}_t = 836.3145 + 100.2882 * \text{March} + 244.3836 * \text{April} + 274.8712 * \text{August} + 332.9209 * \text{September} - 2.6700 * V1_{t-1} + 0.5119 * V2_{t-2} + 0.6344 * V3_{t-1} - 0.6435Z_{t-1}. \tag{3.2}$$

### 3.4. 단순모형과 농업기상모형 비교

앞 절과 같은 절차로 전월까지의 관측치만을 사용하여 단순모형과 농업기상모형을 적합한 후, 월동배추, 봄배추, 고랭지배추, 가을배추의 출하시기인 2016년 1월, 5월, 8월, 11월의 가격을 예측하였다. Table 3.5는 단순모형과 농업기상모형의 모형적합도를 비교한 결과이다. 출하시기의 배추 상품 도매가격을 예측하기 위한 단순모형과 농업기상 모형을 비교한 결과, 농업기상모형의 AIC와 AICC가 단순모형보다 낮아 더 우수한 성능을 보인다. 또한 예측값 역시 농업기상모형의 예측오차의 크기가 더 작아 단순모형보다 예측력이 더 우수하게 나타났다.

**Table 3.5** Comparison of the simple and unstructured agricultural weather model

		AIC	AICC	Actual price	Forecast price	Prediction error
Forecast January	simple model	1153.9	1155.8	552	527.7	-24.3
	unstructured agricultural weather model	1077.0	1079.1		553.0	1.0
Forecast May	simple model	1210.9	1212.3	1143	1035.8	-107.2
	unstructured agricultural weather model	1132.7	1134.7		1092.8	-50.2
Forecast August	simple model	1252.0	1253.4	1524	1003.4	-520.6
	unstructured agricultural weather model	1173.4	1175.3		1016.2	-507.8
Forecast November	simple model	1298.7	1300.4	847	972.7	125.7
	unstructured agricultural weather model	1272.9	1275.1		782.1	-64.9

#### 4. 결론

주로 노지에서 재배되는 배추는 기상 여건에 영향을 많이 받기 때문에, 생산량이나 도매가격 추정 시 기상을 고려해야 한다. 실제 농업기상 자료를 활용한 기존의 연구와는 다르게 본 연구에서는 웹상의 비정형 농업기상 정보를 활용하였다. 도매가격 자료만을 이용하여 2016년 작형별 출하시기 도매가격을 예측한 단순모형과 비정형 데이터를 활용하여 예측한 농업기상모형을 비교한 결과, 농업기상모형이 더 우수하였다.

기상의 경우 지역마다 다르게 나타나지만, 본 연구에서 사용된 비정형 농업기상자료는 지역성을 반영하지 못하는 한계점이 존재한다. 비록 지역성은 반영하지 못하였지만, 지역별로 데이터를 수집·적용해야 하는 번거로운 실제 기상자료에 비해 비교적 데이터 수집과 적용이 용이한 비정형 농업기상자료의 효용성을 입증할 수 있었다.

#### References

- Cho, I., Kim, D. H., and Chun, C. N., (2016). Web-based consumer involvement indices and vegetable purchase: The quantification of unstructured information and its use in predictive modeling. *Journal of the Korean Data Analysis Society* 18, **3(A)**, 1259-1270.
- Cho, S. and Lee, J. H. (2014). *Using SAS/ETS for economic time series analysis*. Freeacademy.
- Choi, S. and Baek, J. (2016). Garlic yields estimation using climate data. *Journal of the Korean Data & Information Science Society*, **27**, 969-977.
- Das, A., Lumezanu, C., Zhang, Y., Singh, V. K., Jiang, G., and Yu, C. (2013, June). *Transparent and flexible network management for big data processing in the cloud*. In HotCloud.
- Hyun, Y., Kim, Y. J. S., Jeong J., Yun, S. and Lee, M. (2015). Text mining on internet-news regarding climate change and food. *Journal of the Korean Data & Information Science Society*, **26**, 419-427.
- Kang, S., Kim, Y. S. and Choi, S. H. (2015). Study on the social issue sentiment classification using text mining. *Journal of the Korean Data & Information Science Society*, **26**, 1167-1173.
- Kim, B. S. (2005). A comparison on forecasting performance of the application models for forecasting of vegetable prices. *Korean Agricultural Economic Association*, **46**, 89-113.
- Kim, S. W. and Yoon, B. S. (2011). Development of vegetable price index and analysis of hedging effectiveness. *Korean Agricultural Economic Association*, **52**, 25-48.
- Lim, C. H., Kim, G. S., Lee, E. J., Heo, S., Kim, Y., Kim, Y. S., and Lee, W. K. (2016). Development on crop yield forecasting model for major vegetable crops using meteorological information of main

- production area. *Journal of Climate Change Research*, **7**, 193-203.
- Mishra, P., Sarkar, C., Vishwajith, K. P., Dhekale, B. S., and Sahu, P. K. (2013). Instability and forecasting using ARIMA model in area, production and productivity of onion in India. *Journal of Crop and Weed*, **9**, 96-101.
- Nam, K. H. and Choe, Y. C. (2015). A study on onion wholesale price forecasting model. *Journal of Agricultural Extension & Community Development*, **22**, 423-434.
- Park, J. Y. and Park, Y. G. (2013). The development of chinese cabbage and radish forecast models. *Korea Rural Economic Institute*, 1-80.

## A study on cabbage wholesale price forecasting model using unstructured agricultural meteorological data<sup>†</sup>

SooHee Jang<sup>1</sup> · Heuiju Chun<sup>2</sup> · Inho Cho<sup>3</sup> · DongHwan Kim<sup>4</sup>

<sup>13</sup>TheIMC

<sup>2</sup>Department of Statistics and Information, Dongduk Women's University

<sup>4</sup>Department of International Trade and Distribution, Anyang University

Received 2 May 2017, revised 23 May 2017, accepted 25 May 2017

### Abstract

The production of cabbage, which is mainly cultivated in open field, varies greatly depending on weather conditions, and the price fluctuation is largely due to the presence of a substitute crop. Previous studies predicted the production of cabbage using actual weather data, but in this study, we predicted the wholesale price using unstructured agricultural meteorological data on the web. From January 2009 to October 2016, we collected documents including the cabbage on the portal site, and extracted keywords related to weather in the collected documents. We compared the forecast wholesale prices of simple models and unstructured agricultural weather models at the time of shipment. The simple model is AR model using only wholesale price, and the unstructured agricultural weather model is AR model using unstructured agricultural weather data additionally. As a result, the performance of unstructured agricultural weather model was has been found to be more accurate prediction ability.

*Keywords:* Cabbage, wholesale price, unstructured agricultural meteorological, AR model.

---

<sup>†</sup> This research was supported by 'Agricultural Biotechnology Development Program', Ministry of Agriculture, Food and Rural Affairs.

<sup>1</sup> Employee, TheIMC, Seoul 04030, Korea.

<sup>2</sup> Corresponding author: Associate professor, Department of Statistics and Information Science, Dongduk Women's University, Seoul 02748, Korea. Email: [hjchun@dongduk.ac.kr](mailto:hjchun@dongduk.ac.kr).

<sup>3</sup> Director, TheIMC, Seoul 04030, Korea.

<sup>4</sup> Department of International Trade and Distribution, Anyang University, Gyeonggi-do 14028, Korea.