Journal of the Korea Academia-Industrial cooperation Society
Vol. 19, No. 5 pp. 480-488, 2018

https://doi.org/10.5762/KAIS.2018.19.5.480
ISSN 1975-4701 / eISSN 2288-4688

Development of Examination Model of Weather Factors on Garlic Yield Using Big Data Analysis

Shinkon Kim
Division of Business, Kwangwoon University

빅데이터 분석을 활용한 마늘 생산에 미치는 날씨 요인에 관한 영향 조사 모형 개발

김신곤
광운대학교 경영학부

Abstract The development of information and communication technology has been carried out actively in the field of agriculture to generate valuable information from large amounts of data and apply big data technology to utilize it. Crops and their varieties are determined by the influence of the natural environment such as temperature, precipitation, and sunshine hours. This paper derives the climatic factors affecting the production of crops using the garlic growth process and daily meteorological variables. A prediction model was also developed for the production of garlic per unit area. A big data analysis technique considering the growth stage of garlic was used. In the exploratory data analysis process, various agricultural production data, such as the production volume, wholesale market load, and growth data were provided from the National Statistical Office, the Rural Development Administration, and Korea Rural Economic Institute. Various meteorological data, such as AWS, ASOS, and special status data, were collected and utilized from the Korea Meteorological Agency. The correlation analysis process was designed by comparing the prediction power of the models and fitness of models derived from the variable selection, candidate model derivation, model diagnosis, and scenario prediction. Numerous weather factor variables were selected as descriptive variables by factor analysis to reduce the dimensions. Using this method, it was possible to effectively control the multicollinearity and low degree of freedom that can occur in regression analysis and improve the fitness and predictive power of regression analysis.

요 약 정보통신 기술의 발전으로 농업분야에서도 다량의 데이터로부터 가치 있는 정보를 생성하고 그 활용을 위해 빅데이터 기술을 적용하는 연구가 활발히 진행되고 있다. 농업에서 재배 가능한 작물과 품종은 기온, 강수량, 일조시간 등의 자연환경의 영향에 따라 결정된다. 본 논문은 마늘의 생육과정과 일별로 측정되는 기상변수를 활용하여 농작물 생산에 영향을 미치는 기상기후 요인을 도출하고 마늘을 대상으로 단위면적당 생산량 예측(단수) 모형을 도출하였다. 기상변수는 마늘의 생육단계를 고려하여 빅데이터 분석 기법을 이용하였다. 탐색적 자료 분석과정에서는 통계청, 농촌진흥청, 농촌경제연구원으로부터 생산량, 도매시장 반입량, 생육 데이터 등 다양한 농산물 생산 데이터를 제공받아 활용하였다. 또한 기상청으로부터 AWS, ASOS, 특보현황 등 다양한 기상관측 데이터를 수집하여 활용하였다. 상관관계 분석 과정은 변수선택, 후보모형 도출, 모형진단, 시나리오 예측 등을 통해 도출한 모형의 모형 적합도와 생산량 예측력을 비교하여 마늘생산단수예측 모형을 설계하였다. 수많은 기상요인 변수는 요인분석을 이용하여 차원을 감소시키고 설명변수로 선정하였다. 이 방법을 이용함으로써 회귀분석에서 발생할 수 있는 다중공선성과 낮은 자유도의 문제를 효과적으로 통제할 수 있었으며 회귀분석의 적합도와 예측력을 높일 수 있었다.

Keywords : Big Data, Weather, Garlic Production, Prediction, Multicollinearity

This paper was conducted during the sabbatical year of Kwangwoon University in 2014.
*Corresponding Author : Shinkon Kim (Kwangwoon Univ.)
Tel: +82-10-3891-1900 email: shinkon@kw.ac.kr
Received April 4, 2018 Revised April 23, 2018
Accepted May 4, 2018 Published May 31, 2018

1. Introduction

The United Nations Framework Convention on Climate Change (UNFCCC) defines it as a change of climate that is attributed directly or indirectly to human activity, altering the composition of the global atmosphere. Climate change is a major concern in many areas, including society, economy and the environment, both at home and abroad. Climate change affects the production process of individual crops in agriculture [1]. Application policies are required to minimize the effects and damage caused by climate change, and studies need to continue on the crisis handling [2]. In the case of rice, there is a study to predict the effects on the water level caused by meteorological variables during the growth phase and long-term global warming [3]. For wheat, the effects of the meteorological factors on the growth and production rates for each region and breed are studied [4]. Some studies have used weather factors as explanation variable for the growth conditions of cabbages, pumpkins and cucumber [5].

This paper uses weather factors that affect the garlic production, a crop that is easy to obtain growth information and is closely related to ordinary people's lives. Weather factors include average temperature, maximum temperature, minimum temperature, wind speed, precipitation, wind cooling index, daylight hours, and humidity. Using big data analytics, this study analyzes various weather factors that affect the garlic production and final prediction model for garlic production are proposed through model comparison method.

In Chapter 2, related studies were reviewed. Exploratory data analysis techniques were examined. Factor analysis and regression analysis were applied as main analysis methods. Chapter 3 analyzes and designs garlic yield prediction model. Data collection process, data cleansing and categorizing, data analysis and visualization are included. Chapter 4 developed a model to predict garlic yield and derived conclusion.

2. Related Works

2.1 Exploratory Data Analysis

In statistics, exploratory data analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell analyst beyond the formal modeling or hypothesis testing task.

The ultimate purpose of a exploratory data analysis is to provide all factors extracted from the data to help the analyst maximize his or her understanding of the underlying data structure. It can also determine the setting of the optimal parameter or test basic hypothesis by developing simple models after the detection of characteristics or anomalies by extracting important variables or hidden structures. It is possible to create a list and a ranking of the major variables that estimate the variables based on the statistical significance tests for each independent variable. The above sequence of activities can confirm the uncertainty of the estimate [6].

A descriptive data analysis has been used to analyse agricultural production data with climate and meteorological observations. For the analysis of agricultural production data, statistical office data are used to analyze the data characteristics and to determine the analysis goals. The meteorological data analysis collects a variety of meteorological data from the KMA's automated monitoring stations, the Automated Indicator Monitoring System (ASOS), or special weather reports. Definition of the region and utilization variables under analysis, equivalisation of the time and space resolution of the data, and definition of the standardized criteria of the meteorological variables are used to calculate the observed values depending on the characteristics of the meteorological data or crops.

2.2 Main Analysis Techniques

This section describes the correlation analysis, factor analysis, and regression analysis used in this study. Correlation analysis is used to correlate between variables through selection of variables, derivation of candidate models, diagnostics of models, and prediction of scenarios. This analysis utilizes the selection of variables in stages and all expected relocations for the selection of variables. The selection of step by step variables begins with the model, including all variables. It then deletes the variables that are not helpful for the baseline statistics or adds variables that can improve the baseline statistics between variables that are not included in the model. The most suitable variable would be found by adding and removing variables repeatedly. All possible regression models are calculated using all possible models, which are p independent variables, to estimate all 2 ^ p variables and to select the optimal model.

Factor analysis is applied to the monthly mean temperature, average minimum temperature, average maximum temperature, wind speed, wind speed, cooling index, daily temperature difference, and the prediction model of garlic yields from January to December using rainfall. All of the collected monthly meteorological factors are determined to be suitable for the analysis.

Regression analysis is to determine the regression equation best represents the sample collected to estimate the parameters, and to analyze the causal relationship between structures. This method uses the Minimum Square Method to search for regression equations that create a minimum square root between the measured and calculated values using the regression equation.

3. Design and Analysis of the Prediction Model of Garlic Yield

This study consists of the steps to prepare data, create analysis sets, predict models through exploratory data analysis, and to visualize knowledge. Overview of these process is described in Fig. 1.

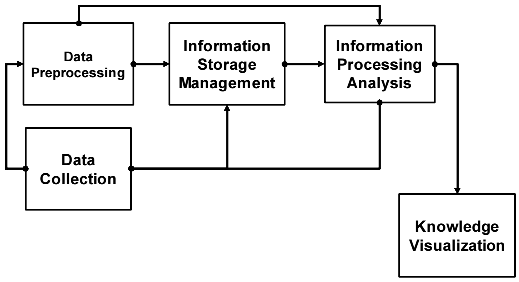

Fig. 1. Overview of the Process

In addition, we shall explore the data types and characteristics to utilize the exploratory data analysis step to select the models to analyze. The models are then complemented to enhance its performance to derive the optimal model.

3.1 Data Collection

The Automated Weather System (AWS) is a weather observation device designed to automatically monitor what people have done in the past, and it has about 660 AWS units installed throughout the country. ASOS (Automated Synoptic Observing System) is an automated meteorological observation device that is installed at a national meteorological station for the purpose of making meteorological observations. It is not provided with the city or military unit meteorological identification information. The data used in this paper is shown in the following Table 1.

Table 1. Summary of Data Set

Source	Item	Value
AWS-based Meteorological Observation Information	National/Time Unit Meteorological Observation	11,200,000
AWS-based Daily Observation Statistics	National/Day Unit Meteorological Observation Statistics	37,300,000
ASOS-based Meteorological Observation Information	National/Time Unit Meteorological Breaking News	97,700,000
Breaking News	National/Day Unit Meteorological Breaking News	3,472
Typhoon Track	National/Time Unit Meteorological Observation	8,958
	Sum	146,212,430

3.2 Data Cleansing & Categorizing

It goes through two stages of pre-processing to obtain the meteorological variables. The first is to convert the climate observation of Gun-unit values of automatic weather observation device (AWS) into Do-unit values using the weighted average of the growth-area baseline Expression (1). And then, weather variables use average temperature, lowest temperature, highest temperature, wind speed, wind cooling index, temperature range, and river water.

$$Weighted\,Avgx = W_1X_1 + W_2X_2 + \ldots + W_nX_n \qquad (1)$$

Second, the AWS ID and ASOS ID do not match, so the ASOS ID nearest to the AWS ID is used in order for matching using the distance weighted calculation method of Expression (2). The climate values observed in the Gun-unit of meteorological observation systems (ASOS) are converted into Do-unit ones. Humidity and daylight hours are also used as weather variables.

$$Z(x) = \frac{\sum w_i z_i}{\sum w_i} \qquad (2)$$

3.3 Analysis & Visualization

As shown in Table 2, garlic production of the main producing area, which are Jeollanam-do, Gyeongsangnam-do, and Jeju-do accounts for 65.5 percent of the nation's garlic production.

Table 2. Production Ratio by Region

Region	Cultivation Area Ratio	Production Ratio
Jeollanam-do	29.4%	27.3%
Gyeongsangbuk-do	22.1%	25.5%
Jeju-do	11.9%	12.7%
Total	63.4%	65.5%

In addition, garlic production in Jeollanam-do in Fig. 2 is almost the same for 14 years, and the production in Gyeongsangnam-do and Jeju-do showed a tendency to increase rapidly. The output increase Gyeongsangnam-do is very clear. The output increase of garlic in Jeju-do is high. Jeollanam-do, Gyeongsangnam-do, and Jeju-do are not major source of garlic production and is therefore excluded from this paper.

Since 2011, the volatility of the water supply has remained very high, ranging from 17 % to 21 %, and the volatility of the water supply dropped in 2015. In 2014, the price of garlic declined as the stock of garlic dropped to 10,000 tons. Since 2007, the number of water sources in Jeju-do and Gyeongsangnam-do has skyrocketed, maintaining a high productivity per unit area. Jeollanam-do recorded its lowest production per unit area in 2011-2013 since its highest point in 2010, but production has increased again in the past two years (Fig. 2).

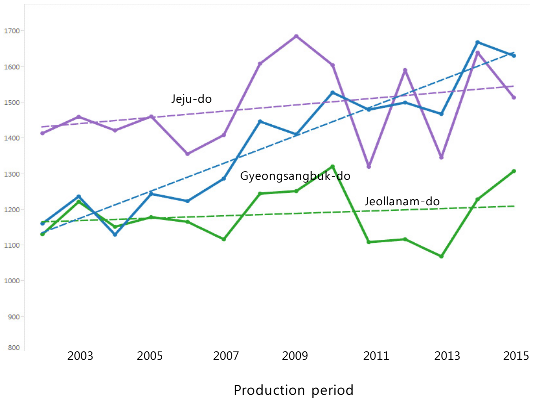

Fig. 2. Analysis of Production Trend per Unit Area of Garlic

Table 3. Characteristics of Growing Season

Name	Period
Sowing / Germination	The First Part of September ~ The Last Part of October
Leaf Elongation	The First Part of November ~ The Middle Part of December
Wintering	The Last Part of December ~ The Last Part of January
Dividing	The First Part of February ~ The Last Part of March
Auxesis	The First Part of April ~ The Last Part of June

Considering the growing season, the meteorological variables in the sowing and germination stage are applied (Table 3).

Process for conducting analysis and model forecasting is described in Fig. 3.

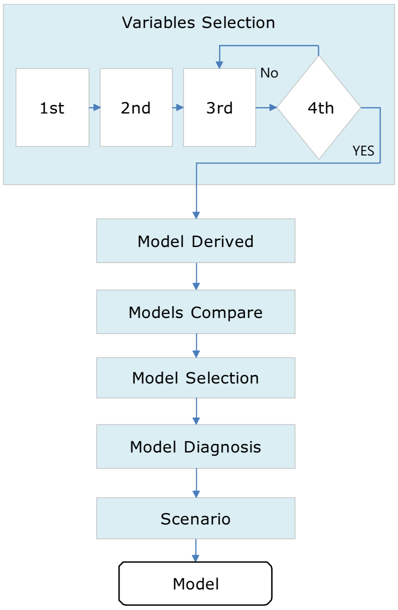

Fig. 3. Process of Model Estimation

(1) Selection of Variables

We select variables that may affect the garlic production and apply them to the model generation. Examples of weather variables and observation periods at the beginning stage are shown in Table 4.

Garlic's sowing begins in the early September and its harvest begins in the early June. The observation periods of "month" used in the analysis were limited from September to May. Fig. 4 is examples of weather climate variables.

Stepwise method is employed for variable selection. Stepwise regression is a method of fitting regression models in which the choice of predictive variables is carried out by an automatic procedure. In each step, a variable is considered for addition to or subtraction from the set of explanatory variables based on some prespecified criterion.

Step 1 applies Stepwise variable selection to the reference variable in Expression 3. For each reference variable in Table 5, one or three major "Month" variables are selected.

Table 4. Description of Meteorological Climate Variables

Variables	Period
Average temperature	January, March
Maximim temperature	December, March
Minimum temperature	January
Wind speed	September
Precipitation	October
Precipitation day	November, January
Daylight hours	November, June
Wind cooling index	March
Daytime difference	October, January
Humidity	November, April
Frost	February
Maximum Temperature of day >25℃ observations of Monthly	October
Minimum temperature of day > 20℃ observations of Monthly	January
Rainfall Observations of Monthly Over 50mm	October
Rainfall Observations of Monthly Over 20mm	October
Minimum temperature of day < 5℃ observations of Monthly	February
Maximum Temperature of day >15℃ observations of Monthly	October, March
Minimum temperature of day < 4℃ observations of Monthly	March

```
   SUup15do11        SUup15do12        SUup15do1         SUup15do2
 Min.   : 6.143   Min.   : 0.000   Min.   :0.0000   Min.   :0.0000
 1st Qu.:13.415   1st Qu.: 1.000   1st Qu.:0.0000   1st Qu.:0.5804
 Median :17.766   Median : 2.081   Median :0.0000   Median :2.6403
 Mean   :17.331   Mean   : 3.428   Mean   :0.8744   Mean   :3.2375
 3rd Qu.:20.683   3rd Qu.: 5.021   3rd Qu.:0.7904   3rd Qu.:4.7903
 Max.   :26.464   Max.   :14.232   Max.   :6.5581   Max.   :9.7903
      ID1               ID2               ID3          ID4          ID5
 Min.   :0.0000   Min.   :0.0000   Min.   :0    Min.   :0    Min.   :0
 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0    1st Qu.:0    1st Qu.:0
 Median :0.0000   Median :0.0000   Median :0    Median :0    Median :0
 Mean   :0.7722   Mean   :0.2900   Mean   :0    Mean   :0    Mean   :0
 3rd Qu.:1.3786   3rd Qu.:0.3894   3rd Qu.:0    3rd Qu.:0    3rd Qu.:0
 Max.   :4.2593   Max.   :1.6683   Max.   :0    Max.   :0    Max.   :0
   TAMIN_4do3         TAMIN_4do4        TAMIN_4do5          R50mm9
 Min.   : 0.8839   Min.   : 0.000   Min.   :0.00000   Min.   :0.0000
 1st Qu.: 7.1395   1st Qu.: 0.000   1st Qu.:0.00000   1st Qu.:0.1793
 Median :21.8133   Median : 6.905   Median :0.00000   Median :0.7263
 Mean   :17.6210   Mean   : 6.037   Mean   :0.15800   Mean   :1.0448
 3rd Qu.:24.5304   3rd Qu.: 9.473   3rd Qu.:0.03313   3rd Qu.:1.4104
 Max.   :28.3052   Max.   :14.808   Max.   :1.37486   Max.   :4.3703
```

Fig. 4. Example of Weather Climate Variables

$$Y = TAD9 + TAD10 + TAD11 + TAD12 + TAD1 + TAD2 + TAD3 + TAD4 + TAD5 + TAD6 \quad (3)$$

In step 1, TAD1 and TAD3 are selected.

Step 2 applies the stepwise sequence to the variable that has the analysis value in Fig.5. Until two or three optimal variables are left, the variable selection process is continued by applying Expression 4.

Table 5. Variables of 'Month'

Variable	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
TAD					o		o			
TAmax				o			o			
TAmin					o					
WSD	o									
RAIN_DAY		o								
TNdwnM5do						o				
SUup15do				o			o			
ID					o					
TAMIN_4do							o			
R50mm		o								
R20mm		o								
RAINDAY_COUNT								o		
WCI							o			
SS									o	
HM			o					o		
DTD		o			o					
SU		o								
FD						o				

$$YD = TAD1 + TAmax12 + TAmin1 + WSD9 + RAINDAY10 + SUup15do3 + ID1 + WCI3 + SS5 + HM4 + FD2 \tag{4}$$

```
                Df Sum of Sq    RSS    AIC
<none>                       194021 378.40
- SUup15do3      1     12904 206925 379.10
+ SUup15do12     1      5112 188908 379.28
+ TAMIN_4do3     1      4173 189848 379.48
- WSD9           1     17086 211107 379.94
+ TAD3           1      2008 192012 379.96
+ R20mm10        1      1948 192072 379.97
+ TAmax3         1      1621 192400 380.05
+ R50mm10        1       805 193216 380.22
+ DTD1           1       606 193415 380.27
+ DTD10          1       582 193438 380.27
+ TNdwnM5do2     1       452 193568 380.30
+ SU10           1       237 193783 380.35
+ HM11           1       225 193796 380.35
+ RAINDAY_COUNT4 1        82 193939 380.38
- WCI3           1     29583 223604 382.36
- HM4            1     43041 237062 384.81
- TAmin1         1     47560 241581 385.61
- RAIN_DAY10     1     75379 269399 390.18
- TAmax12        1     78886 272907 390.73
- TAD1           1     97039 291059 393.43
- SS5            1    105596 299617 394.65
- ID1            1    116747 310768 396.18
- FD2            1    228852 422872 409.12
```

Fig. 5. Analysis Values and Variables

Table 6. Stepwise Variable Selection of Step 2

Stepwise Variable Selection			
TAD1	O	R50mm10	
TAD3		R20mm10	
TAmax12	O	RAINDAY_COUNT4	
TAmax3		WCI3	O
TAmin1	O	SS5	O
WSD9	O	HM11	
RAIN_DAY10	O	HM4	O
TNdwnM5do2		DTD10	
SUup15do12		DTD1	
SUup15do3	O	SU10	
ID1	O	FD2	O
TAMIN_4do3			

Step 3 assumes that there are regional differences in the number of times for water-cutting due to the edaphic and technical properties, i.e., Adjusted R2, Schwark's Bic, and Mallow`s Cps shown in Fig.6 and Table 7. In order to apply all the possible regressions, a local dummy variable is added.

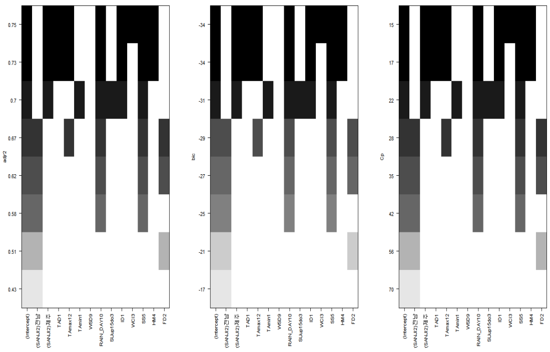

Fig. 6. Adjusted R2, Schwarz's Bic, Mallow's Cp

Table 7. Variable of Adjusted R2, Schwarz's Bic, Mallow's Cp

	Adj R^2	bic	Cp
TAD1	O	O	O
TAmax12	O	O	O
TAmin1			
WSD9			
RAIN_DAY10	O	O	O
SUup15do3			
ID1	O	O	O
WCI3	O	O	O
SS5	O	O	O
HM4	O	O	O
FD2			

With only two to three variants, Step 4 removes the resulting variables from Step 3. During the final parameter selection process in Table 8, the TAD1 has a multicollinearity issue with HM 4. WCI 3, TAmax 12, and ID1 were removed due to the high p value. The final selection variables are RAIN DAY 10, SS5.

Table 8. Process of the Final Variable Selection

	AAll Possible Regressions (Adding Regional Dummy Variable)			Deleting Variable	Selected
	Adj r^2	bic	Cp		
TAD1	O	O	O	Multicollinearity	
TAmax12	O	O	O	p value	
TAmin1					
WSD9					
RAIN_DAY10	O	O	O		O
SUup15do3					
ID1	O	O	O	p value	
WCI3	O	O	O	p value	
SS5	O	O	O		O
HM4	O	O	O	Multicollinearity	
FD2					

(2) Derivation of Candidate Model

Garlic farming requires at least 20 rainy days after sowing. Especially garlic produced in warm soil has a short dormant period and is rather sensitive to the low temperature. In addition to the finally selected variables, the inclusion of cold-related variables reduce significantly the model's Adjusted R2. The schedule models1 includes the finally selected variables and DTD1 (January temperature range). The schedule models2 includes the finally selected variables and WSD1 (January wind speed). The schedule model s3 includes variables selected in step 4. Consequently the candidate models are represented Expression 5, Expression 6, and Expression 7, respectively.

$$YD = f(RAINDAY10 + SS5 + (Region\,Dummy\,Parameter) + DTD1 \quad (5)$$

$$YD = f(RAINDAY10 + SS5 + (Region\,Dummy\,Parameter) + WSD1 \quad (6)$$

$$YD = f(RAINDAY10 + SS5 + (Region\,Dummy\,Parameter) \quad (7)$$

(3) Model Comparison

In order to compare the performance of the 3 candidate prediction models, the LOOCV (leave-one-out cross validation) method was employed. The LOOCV showed that the Model 3 has the lowest estimate of error rate with 7.3 %. The Model 3 is finally selected.

Table 9. Result of LOOCV

Division	Result
Model 1	[1] 0.07797016 0.07790081
Model 2	[1] 0.07747454 0.07740497
Model 3	[1] 0.07349335 0.07339579

(4) Model Diagnosis

The regression diagnosis in Model 3 and the diagnostics results for the parameter significance are as shown in Fig. 7 and 8. Residuals identifies the bias in the data as a representation of the difference between the expected value and the measured value, i.e. the distribution of the residual value in small fractions. The Coefficient lists the estimated values, standard errors, t values, and p values of each row to indicate how much the slope varies and is of statistical significance. Multiple R-squared, and Adjustment R-squared are the coefficient of determination and the coefficient of determination for adjusting the degree of freedom. As the coefficient of determination becomes closer to 1, the model is suitable for the given data.

```
Call:
lm(formula = YD ~ factor(SANJI2) + RAIN_DAY10 + SS5, data = df3)

Residuals:
     Min       1Q   Median       3Q      Max
-196.732  -78.165    7.368   78.734  241.654

Coefficients:
                    Estimate Std. Error t value     Pr(>|t|)
(Intercept)         954.9015   120.9923   7.892 0.0000000019 ***
factor(SANJI2)전남 -191.2247    42.5737  -4.492 0.0000669956 ***
factor(SANJI2)제주   84.6650    46.2588   1.830      0.07528 .
RAIN_DAY10            1.2605     0.4920   2.562      0.01462 *
SS5                   1.6793     0.4943   3.397      0.00164 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 112.3 on 37 degrees of freedom
Multiple R-squared:  0.6433,	Adjusted R-squared:  0.6047
F-statistic: 16.68 on 4 and 37 DF,  p-value: 0.00000006738
```

Fig. 7. Coefficient Estimation of Multiple Regression Analysis

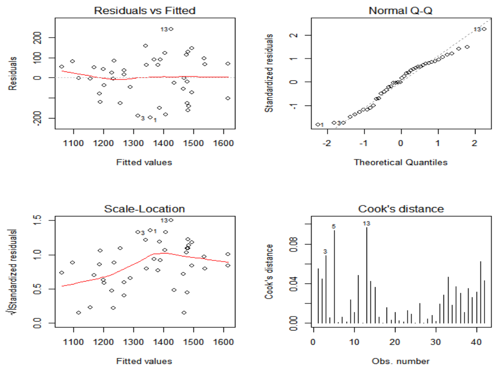

Fig. 8. Beta Weights of Regression diagnostics

(5) Scenario

For the establishment of scenarios in which the accumulated rainfall in October 2014 and the value of the daylight hours in May 2015 are made, Expression 8 is used.

$$Y = f(RAINDAY10 + (Region\ Dummy\ Parameter) + SS5 \tag{8}$$

To predict the case of water-cutting, three scenarios are made and used. Scenario 1 uses the mean value of the accumulated precipitation in October 2014 and the exclusion of the maximum and minimum daylight hours from 2002 to May 2014. Scenario 2 uses the accumulated rainfall in October 2014 and the first half (if the daylight hours are short) daylight hours from 2002 to May 2014. Scenario 3 uses the accumulated rainfall in October 2014 and the 4th percentile daylight hours (if the daylight hours are longer) from 2002 to May 2014.

Fig. 9 shows the illustration of comparison between the actual water-cutting and the expected value after assuming the model uses the variables observed in March 2015 to predict the production of garlic.

Table 10 shows the actual observations made during the daylight hours in May 2015. S1 is the meteorological scenario 1, and it applies the mean value of the daylight hours excluding maximum and minimum between May 2001 and May 2013. S2 estimates the first quartile, if the daylight hours are short, of the daylight hours between 2001 and May 2014. S3 predicts the 3rd quartile, if the daylight hours are longer, of the daylight hours between 2001 and May 2013.

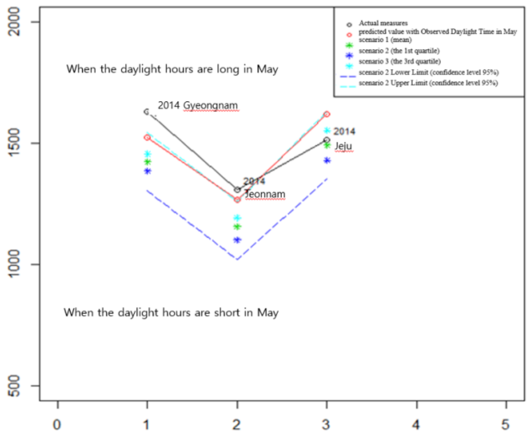

Fig. 9. Illustration of Comparison between the actual water-cutting and the expected value

Table 10. Garlic Yield Forecasting Error Analysis Example in 2015

Region	Real Water-Cutting in 2015	Substituted Observed Value	S1(%)	S2(%)	S3(%)
Gyeongsangnam-do	1630	6.6	12.7	15.0	10.7
Jeollanam-do	1307	3.1	11.6	15.7	8.8
Jeju-do	1513	7.1	1.4	5.5	2.5
MAPE		5.6	8.6	12.1	7.4

4. Model Results and Conclusions

The final garlic yield prediction for each region employs the function in Expression 9.

$$Y = f(RAINDAY10 + SS5 + (Region\ Dummy\ Parameter) \tag{9}$$

RAINDAY 10 is the cumulative rainfall in October, so proper precipitation is essential within 20 days after planting in October, and droughts in autumn do not have a good effect on growing garlic. SS5 is a cumulative sunlight in May, and it is greatly influenced by the hot patches and temperature during the growing season. Including a coefficient estimate (Fig. 8) for estimating the garlic production for each region obtained from the model diagnosis of multiple regression analysis, it is summarized as a prediction model of Expression 10.

$$Y = 954.90 + 1.26 * RAINDAY10 + 1.67 * SS5 - 191.22 * Jeonnam + 84.66 * Jeju \quad (10)$$

This paper analyzed only the meteorological factors that affect the production of garlic and developed the prediction model of garlic yields. In future research, climate variables such as temporary typhoons and hail should be used later. Further analysis should be made of pest and farmer technological changes in addition to climate changes. In future research, it is also necessary to verify the predicted results of the developed garlic yield prediction model using actual data.

REFERENCES

[1] Korea Institute for Industrial Economics & Trade, http://www.kiet.go.kr/servlet/isearch, 2014.

[2] J. K. Koh, J. H. Kim, "Have local officials recognized the importance of adaptive policy," Journal of the Korean Urban Management Association, vol. 24, no. 3, pp. 51-72, 2011.

[3] O. S. Kwon, H. K. Cho, E. B. Cho, K. S. Roh, "Climate Variables and Rice Productivity: A Semiparametric Analysis Using Panel Regional Date," Korean Journal of Agricultural Economics, vol. 54, no. 3, pp. 71-94, 2013.

[4] L. You, M. W. Rosegrant, S. Wood, D. Sun, "Impact of growing season temperature on wheat productivity in china," *Agricultural for Meteorology*, vol. 149, pp. 1009-1014, 2009.
DOI: https://doi.org/10.1016/j.agrformet.2008.12.004

[5] S. H. Han, B. H. Lee, M. S. Park, J. h. Seoung, H. S. Yang, S. C. Shin, "A Study of building Crop Yield Forecasting Model Considering Meteorological elements," Korea Rural Economic Institute, p.152, 2011.

[6] T. B. John, C. H. Yu, "Exploratory Data Analysis," Wiley, 2003.

[7] Y. Kano, H. Akira, "Stepwise variable selection in factor analysis," *Psychometrika*, vol. 65, no. 1, pp. 7-22, 2000.
DOI: https://doi.org/10.1007/BF02294182

[8] T. A. Brown, "Confirmatory Factor Analysis for Applied Research", 2nd Edi. (Methodology in the Social Sciences), Jan 8, 2015.

[9] R. L. Gorsuch, "Factor Analysis: Classic Edition", Psychology Press, Dec 24, 2014.

[10] G. Ciaburro, "Regression Analysis with R: Design and develop statistical nodes to identify unique relationships within data at scale", Jan 31, 2018.

[11] L. D. Schroeder and D. L. Sjoquist, "Understanding Regression Analysis: An Introductory Guide", Quantitative Applications in the Social Sciences, Nov 24, 2016.

[12] D. E. Farrar and R. R Glauber, "Multicollinearity in Regression Analysis; the Problem Revisited", Aug 24, 2017.

[13] Wikipedia contributors, "Focus On: Regression Analysis: Dependent and independent Variables, Multicollinearity, Simple linear Regression, Heteroscedasticity, Lasso (statistics), ... Estimation, Errors and Residuals, etc.", Feb 22, 2018.

Shinkon Kim [Regular member]

- Feb. 1980 : B.S. in Business from Yonsei University, Korea
- Feb. 1982 : M.S. in Finance from Seoul National University, Korea
- May. 1984 : M.S. in Computer Information Systems from Georgia State University, U.S.A.
- May. 1989 : Ph.D. in MIS from Georgia State University, U.S.A.
- March. 1992 ~ present : Professor in College of Business, Kwangwoon University

<Research Interests>
Management Information Systems, Big Data Analytics, De-identification