

Analysis of Predictions of Garlic Bulb Weight using LASSO Regression Model

Junki Kim^{1†}, Hyungyong Lee^{1†}, and Tongjoo Suh^{2*}

¹Center for Agricultural Outlook, Korea Rural Economic Institute, Naju 58321, Korea

²Department of Food and Resource Economics, Pusan National University, Miryang 50463, Korea

*Corresponding author: tjsuh@pusan.ac.kr

†These authors contributed equally to this work.

Abstract

As part of a preemptive response to ensure a stable supply and demand policy for garlic (*Allium sativum* L.), the Korean government strives to determine the final yield of garlic by observing its growth status by season. In particular, estimation of the bulb weight of garlic is one of the main steps in estimating garlic production, conducting observational work on spherical weight prediction is getting more important. The observation data on garlic can be viewed as data for predicting the final weight, but since root growth occurs intensively at a specific time, there is a limit to predicting the final bulb weight only by the bulb weight corresponding to each observation time point. In addition, since the factors affecting the final bulb weight of garlic may be different for each growing stage, an estimation is required for each stage to predict the final bulb weight at each time point. In this study, garlic bulb weight estimation analysis, which reflects the characteristics of Korean garlic grown in an outfield, is performed using growth survey data. A stage-by-stage model for predicting the bulb weight of garlic at each time point is established, based on the fact that the factors influencing the bulb weight of garlic may differ according to the growth stage. In the process, LASSO regression is used for variable selection and coefficient estimation of the bulb weight of garlic. As a result of the estimation, significant factors regarding growth, lot and environment were identified stage by stage, providing indicators for forecasting final bulb weight of garlic at each stage.

Received: November 29, 2022

Revised: April 24, 2023

Accepted: May 12, 2023

 OPEN ACCESS



HORTICULTURAL SCIENCE and TECHNOLOGY
41(4):437-447, 2023
URL: <http://www.hst-j.org>

pISSN : 1226-8763
eISSN : 2465-8588

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright©2023 Korean Society for Horticultural Science.

Additional key words: growth observation survey, LASSO regression model, outfield crop, stage-by-stage model, variable selection

Introduction

Garlic (*Allium sativum* L.) is one of the bulb crops that Koreans traditionally enjoy. Since Korea is one of the largest consumers of garlic (Choi et al., 2021a) and annual garlic consumption per capita in Korea is about 7 kg (Kim et al., 2021), it can be concluded that garlic is a vegetable that cannot be omitted from the Korean diet. However, compared to consumption, which maintains a stable level due to continuous demand, garlic production shows relatively large fluctuations. This is because, due to the characteristics of garlic grown in the outfield, the crop yield is greatly affected by weather conditions.

To control the supply and demand of garlic, the Korean government implements various measures, such as stockpiling, surplus disposal, and suspension of shipments (Kim et al., 2020). As part of a preemptive response to ensure a stable supply and demand policy for garlic, the government, in collaboration with the Korea Rural Economic Institute (KREI), strives to determine the final yield of garlic by observing its growth status by season. In particular, estimation of the bulb weight of garlic is one of the main steps in estimating garlic production, so under the leadership of the government, several institutions in Korea, including the KREI, are conducting observational work on bulb weight prediction.

With regard to the bulb weight of garlic, many studies have analyzed factors affecting growth and predictions of growth. Representative studies include analyses of the effects of irradiance on garlic growth (Oh et al., 2017; Atif et al., 2020), the effects of an increase in growing temperature on garlic growth (Kamenetsky et al., 2004; Portela and Cavagnaro, 2004; Oh et al., 2019), the effects of soil properties and soil management on garlic growth (Diriba-Shiferaw, 2016; Oh and Koh, 2020), the effects of diseases on garlic growth and bulb development (Lee et al., 2022), and the effects of high-temperature and humid environmental conditions on the growth and quantity of garlic (Rahim and Fordham, 1997). However, existing analyses explain the growth and development of garlic based on data obtained through controlled experiments in limited spaces. Since garlic is grown mainly in the outfield, there are limits to applying the results derived from controlled experimental environments to yield prediction as it is. For example, when drought occurs, farmers respond to the environment through water sprinkling or plastic mulching to prevent crops from withering. Therefore, even if the humidity controlled in the laboratory and the humidity observed in the outfield are the same, the actual effects may be different. In other words, the effect of weather conditions on crop productivity may be overestimated or underestimated due to the effects of the adaptation and learning of farmers, which diverge from the environmental controls used in experiments. Therefore, existing studies are of great help in identifying the growth characteristics of crops but are different from observations for estimating production, which means that to predict the yields of outfield crops, analysis is required that uses actual survey data rather than experimental data.

After 2020 when KREI provided measurement data from the observational work on outfield crops, studies on quantity and growth models have been carried out, showing efforts to reflect the characteristics of outfield crop (Kim and Kim, 2021; Choi et al., 2021b; Moon et al., 2021). However, there is only one case in which garlic has been analyzed (Choi et al., 2021b), and even this has an ex post perspective due to the application of normal year values, and has limitations in that it is not easy to estimate the bulb weight in a timely manner at each time point, implicating more related research is necessary. Considering the phenology and morphology processes of bulbous crop growth, environments affect the development in garlic throughout the growth stage, but factors affecting the bulb weight of garlic should be considered differently over time (Hsiao et al., 2019).

The purpose of outfield vegetable observation is to predict what the final production will be beginning at the time of observation, not to explain an incident that has already occurred. Since the production of garlic is aggregated by weight, the final bulb weight of garlic is used as a direct indicator in calculating production. Therefore, the observation data on garlic can ultimately be viewed as data for predicting the final weight. However, since root growth occurs intensively at a specific time, there is a limit to predicting the final bulb weight only by the bulb weight corresponding to each observation time point. Since root growth is not proportional to the growth of leaves and stems, and overgrowth of leaves and stems adversely affects root growth, the final weight cannot be measured mechanically through leaf or stem growth. Therefore, to predict the final bulb weight, it is necessary to comprehensively consider the growth status for each stage.

Meanwhile, surveys on garlic in Korea are conducted regularly. Since the growth statuses of crops observed at the time of observation are different, and environmental factors considered optimal for each growth period are also different, at each point in time, the information available for the prediction of final bulb weight also varies. In addition, since the factors affecting the final bulb weight of garlic may be different for each growing stage, an estimation is required for each stage to predict the final bulb weight at each time point. However, with limited information on the growth state of garlic, it is difficult to know how much growth information at each stage affects bulb weight. Accordingly, there is a problem with variable selection stage by stage, and the effect thereof has the limitation that it is difficult to make intuitive estimations.

In this study, as part of a growth analysis to predict the yield of garlic, garlic bulb weight estimation analysis, which reflects the characteristics of Korean garlic grown in an outfield, is performed using growth survey data. In addition, based on the fact that the factors influencing the bulb weight of garlic may differ according to the growth stage, a stage-by-stage model for predicting the bulb weight of garlic at each time point is established. To make intuitive interpretation easier and reduce errors, a stage-by-stage model is used to perform variable selection and regularization, and to estimate the parameters necessary for predicting the bulb weight of garlic. This analysis makes it possible to perform variable selection and regression of the bulb weight of garlic, and to build an intuitive and simple framework for the prediction of the quantity of garlic.

Materials and Methods

Measurement and Climate Data

In this paper, to analyze the relationship between garlic growth and bulb weight, actual data from surveys conducted by KREI were used. To gauge the effects of weather conditions, topo-climatology model data provided by the National Institute of Agricultural Sciences (NAS) were used. KREI has conducted surveys on 100 farms every year since 2018 to confirm the growth process of garlic, and these growth survey data include growth information and lot information. In the

Table 1. Observation period by growth stage (survey of growth in 2020 and 2021)

	Stage	Round	Survey date
Stage A	Leaf elongation (early stage)	1	16–28 February
		2	1–15 March
Stage B	Leaf elongation (late stage)	3	16–31 March
		4	1–10 April
		5	11–20 April
Stage C	Bulb development	6	21–30 April
		7	1–10 May
		8	11–20 May
Stage D	Pre-harvesting	9	21–31 May
		10	8–10 June
		11	11–20 June

Note: The survey was conducted on the same day in 2020 and 2021, respectively.

survey, the representative lot of the target farm is divided into three survey areas of about 3.3 square meters each, and after selecting five individuals in each investigation area, the growth information on the above-ground and underground parts of these individuals is collected. The survey data used for the analysis were obtained through 22 observations from February 15 to June 11, 2020 and from the same period in 2021. The observation period and growth period in the survey are shown in Table 1.

Growth information includes results of observations of plant height, leaf number, and leaf sheath length aboveground, and leaf sheath diameter and bulb weight in the underground part. The above-ground growth information was collected from the first survey, but since the growth information of the underground part was gathered through a destructive investigation that removed garlic to examine individual conditions, data were collected from the seventh round of the survey, which started with the bulb growth stage. The lot information includes information on the cultivation area, furrow width, ridge width, plant spacing, row spacing, number of plants, and number of missed plants related to garlic cultivation.

In this paper, data from topo-climatology models were used for the analysis of weather conditions. To examine the effect of weather conditions on garlic bulb weight, meteorological information was used based on the data provided by

Table 2. Data summary by growing period

Variables	Unit	Stage A		Stage B		Stage C		Stage D	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Growth State									
Plant height	cm	37.2	14.8	60.8	18.3	81.9	13.3	78.5	14.5
Leaf number	no. of leaf	7.0	1.3	8.2	1.2	8.1	1.7	6.5	1.6
Leaf sheath length	mm	8.6	5.8	19.1	9.8	32.6	8.1	37.5	6.8
Leaf sheath diameter	mm	13.8	4.1	17.8	4.6	19.0	4.2	16.8	3.7
Bulb weight	g					41.9	19.7	75.0	25.3
Lot Information									
Cultivation area	m ²	670.3	496.9	671.7	505.3	672.0	507.5	696.0	573.7
Furrow width	m	2.0	0.5	2.0	0.5	2.0	0.5	1.9	0.4
Ridge width	m	1.6	0.5	1.6	0.5	1.7	0.5	1.6	0.3
Plant spacing	m	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Row spacing	m	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0
Planting Density	no. of bulb	128.2	18.2	127.9	18.1	127.8	17.9	121.6	13.2
Missed plant	no. of bulb	5.5	6.4	5.4	6.3	5.5	6.3	4.6	6.0
Environment Information									
Average temperature	°C	7.0	1.8	10.9	1.6	14.2	2.5	17.9	1.0
Max temperature	°C	12.0	1.9	16.6	1.6	19.6	2.8	22.9	1.4
Min temperature	°C	2.0	2.5	5.2	2.6	8.7	2.9	12.8	1.4
Precipitation	mm/day	30.2	16.7	33.0	26.8	35.9	44.6	49.3	22.0
Irradiance ^z	MJ/m ²	193.7	37.4	244.6	60.0	201.9	48.0	188.3	30.9
GDD ^y	°C	46.9	21.9	92.6	34.8	100.3	36.6	148.0	27.7
Other									
Growth period ^x	day	147.4	14.2	176.3	14.8	204.2	14.7	222.8	13.4

^zThe daily cumulative irradiance during the round.

^yThe sum of values of days with an average temperature of 4°C higher.

^xThe difference between the observation date and the transplanting date.

topo-climatology models. The topo-climatology model is a model independently developed by the NAS that provides meteorological information for the entire area of Korea by lot level through estimation (Kim et al., 2019). These meteorological data include information on the average temperature, maximum temperature, minimum temperature, precipitation, irradiance, and growing degree day (GDD).

To integrate individual-level growth information and lot-level meteorological data, the units of analysis were unified by lot by calculating the growth data of individual levels as the average values for each farm. The growth period is indicated by the difference between the observation date and the transplanting date.

The growth stages were set to leaf elongation (early stage and late stage), bulb development, and pre-harvesting, and for estimation by growth stage, the data were constructed by reclassifying the observation period according to the growth stage. The variables used in the analysis and data summary are shown in [Table 2](#).

Methods

The relationship between garlic bulb weight, factors indicating the growth state of garlic, and factors affecting the growth of garlic can be represented as follows:

$$\text{Bulb weight} = f(\text{Growth; Lot, Climate}) \quad (1)$$

To estimate the final bulb weight, linear regression can be used, and the relationship between the dependent variable and the explanatory variable can be represented as follows:

$$y = \alpha + X'\beta + Z'\gamma + \varepsilon \quad (2)$$

where y is the bulb weight of garlic at the harvest, X is the vector of growth state variables, Z is the vector of factors that affect the growth of garlic, including lot and climate, α is the intercept, and ε is the disturbance.

This paper constructs a model using actual measurement data rather than experimental data, which entails exploring the necessary variables in the model for each observation period. To forecast the final bulb weight of garlic at the point of each growth period in a situation where a large amount of information is given, a process of selecting appropriate regressors for the bulb weight is required. In this study, the least absolute shrinkage and selection operation (LASSO) was used to estimate the final bulb weight of garlic. LASSO is a regression analysis model that carries out variable selection and regularization, shrinking coefficients (Tibshirani, 1996). The model has the characteristic that some variables are removed as certain coefficients converge to zero, improving interpretability and preventing problems, such as multicollinearity (Tibshirani, 1996). The model extracts the appropriate variables for estimating the weight through a combination of various information indicating the growth state of the garlic and affecting the growth. Accordingly, in the sense that it enables farmers to identify the factors to focus on for each stage, it is useful for understanding garlic growth and forecasting the final growth state of garlic. In addition, since unnecessary variables or variables with little influence are reduced in the estimation of the bulb weight with LASSO analysis, discussion on figuring out the relationship between explanatory variables and the bulb weight becomes clearer, and farmers are able to intuitively understand the factors which affect the growth of garlic.

LASSO solves the optimization problem that minimizes the absolute value of the regression coefficient as well as the error. The optimization problem is represented as follows:

$$\min \sum_i^N (Y_i - X_i' \theta)^2 \text{ s.t. } \sum_j^K |\theta_j| \leq t \quad (3)$$

where y_i is the response, $X = (1, X_{i1}, \dots, X_{iK})$ is the covariate vector, $\theta = (\theta_0, \theta_1, \dots, \theta_K)$ is the coefficient vector, and t is a nonnegative tuning parameter that determines the degree of regularization (Tibshirani, 1996; Hastie et al., 2009). The optimization problem can be expressed as a regression formula for the estimation of the final garlic bulb weight, with regressors that indicate the growth state and factors affecting the growth, as follows:

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = \min \sum_i^N \left(y_i - \alpha - \sum_j^K \beta_j x_{ij} + \sum_l^M \gamma_l Z_{il} \right)^2 + \lambda \left[\sum_j^K |\beta_j| + \sum_l^M |\gamma_l| \right] \quad (4)$$

where α is the intercept, β_j are the coefficients of growth variables, γ_j are the coefficients of factors affecting growth, such as lot and environment variables, and λ is the coefficient of the LASSO penalty term that forces each β_j and γ_l to zero (Wu and Lange, 2008).

Since the growth state of garlic is affected by weather conditions, an interdependence between growth variables and environmental variables is formed, which may cause collinearity problems (Greene, 2012). In this study, considering this interdependence, growth variables were constructed based on information at the time of the survey, while environmental variables were constructed based on information from the time of the survey to the time of the next survey. Growth variables were predetermined at the point of the survey, while environmental variables reflected information during the growth period, so that effects and feedback effects were excluded (Halcoussis, 2005).

Results

In this paper, variable selection and coefficient estimation are performed using LASSO to forecast the final bulb weight of garlic for each growth stage. In LASSO analysis, different estimation results can be derived depending on the size of λ , a coefficient of the penalty term (Shojaie and Michailidis, 2010), so it is necessary to find the appropriate λ . Adjusting λ stands for the penalty controls, which means setting the number of selected variables through constraints in the model. As λ increases, the number of selected variables in the model decreases, and when λ becomes zero, the penalty term disappears, and the estimation results obtained from the model show the same results as those estimated by ordinary least squares regression (Konzen and Ziegelmann, 2016).

In this analysis, to set the appropriate λ of the lasso model, the Akaike information criterion (AIC), Bayesian information criterion (BIC), root mean squared error (RMSE), and mean absolute error (MAE) were considered selection criteria. In general, the coefficient of determination is often used as a measure of fit, which indicates improvement in predictions using the relationship between variations in response variables based on variations in regressors (Barreto and Howland, 2006). However, since the value tends to increase as the number of variables increases, it has dimensionality

reduction limitations in the context of variable selection (Halcoussis, 2005). Thus, it is necessary to consider trading off goodness of fit and parsimony when selecting an appropriate model for a given set of data. As alternatives, the widely used model selection criteria are the AIC and BIC (Kennedy, 2008). This task is performed by asymptotically selecting the model that minimizes AIC and BIC, remaining consistent in selection as the population grows (Kennedy, 2008; Vrieze, 2012). AIC and BIC are represented as follows:

$$\text{AIC} = \ln(\text{SSE}/N) + 2K/N \quad (5)$$

$$\text{BIC} = \ln(\text{SSE}/N) + K \ln(N)/N \quad (6)$$

where SSE is the sum of squared error, for N populations and K regressors. Lower AIC and BIC indicate the better model. Another alternative is to use the criteria to evaluate the model's performance. Typical ways of doing this are RMSE and

Table 3. Comparison results of model selection and number of variables via magnitude of penalty terms λ

Stage	λ	AIC	BIC	RMSE	MAE	Number of selected variables		
						Growth	Lot	Climate
Stage A	0.1	10,466	10,551	28.89	22.59	4	6	4
	0.5	10,463	10,536	28.70	22.42	4	4	4
	1.0	10,465	10,531	28.66	22.37	4	4	3
	1.5	10,469	10,536	28.68	22.38	4	4	3
	2.0	10,478	10,551	28.72	22.41	4	4	4
Stage B	0.1	10,779	10,877	28.11	21.81	4	7	5
	0.5	10,790	10,875	28.15	21.83	4	5	5
	1.0	10,791	10,871	28.11	21.79	4	4	5
	1.5	10,796	10,875	28.09	21.76	4	4	5
	2.0	10,800	10,873	28.09	21.74	4	3	5
Stage C	0.1	15,618	15,722	25.79	19.94	5	6	5
	0.5	15,634	15,724	25.83	19.97	5	4	5
	1.0	15,638	15,729	25.80	19.94	5	4	5
	1.5	15,645	15,736	25.81	19.94	5	4	5
	2.0	15,655	15,746	25.78	19.91	5	4	5
Stage D	0.1	4,278	4,357	27.69	21.09	5	5	5
	0.5	4,281	4,354	25.77	20.10	4	5	5
	1.0	4,284	4,341	25.58	19.93	4	3	4
	1.5	4,290	4,352	25.54	19.93	4	4	4
	2.0	4,295	4,357	25.55	19.98	4	4	4
Average values (the whole period)	0.1	10,285	10,377	27.62	21.36	4.5	6.0	4.8
	0.5	10,292	10,372	27.11	21.08	4.3	4.5	4.8
	1.0	10,295	10,368	27.04	21.01	4.3	3.8	4.3
	1.5	10,300	10,375	27.03	21.00	4.3	4.0	4.3
	2.0	10,307	10,382	27.04	21.01	4.3	3.8	4.5

AIC: Akaike information criterion; BIC: Bayesian information criterion; RMSE: root mean squared error; MAE: mean absolute error.

MAE, which are widely used to measure predictive accuracy based on the proximity of predicted and actual values¹⁾ (Greene, 2012).

Table 3 represents the selection criterion calculation results according to the variation in λ . Examining the statistics regarding selection criteria shows that the relative sizes and rank of the values are different for each stage. Examining average values over the whole period, RMSE and MAE show a relatively low value, where λ is between 1.0 and 2.0, while AIC and BIC show a relatively low value, where λ is between 0.5 and 1.0. In the model, λ was applied as 1.0 since it resulted in minimum values in many cases in terms of average values in the whole period. Moreover, considering that the number of selection variables does not change significantly compared to when λ is greater than or less than 1.0, it was judged that $\lambda = 1.0$ has stability in selecting the number of variables.

As a result of applying LASSO to each growth stage, the numbers of selection variables were 13 in Stage A, 14 in Stage B, 15 in Stage C, and 13 in Stage D (Table 4). Looking at the growth variables, in all growth stages, the bulb weight during the observation period, plant height, leaf number, and leaf sheath diameter were selected as variables, and the sign of each variable was shown consistently over the growth stages.

In all growth stage models, it was found that the longer the plant height and the leaf sheath diameter, the more the number of leaves had a positive effect on the final bulb weight. This shows that the growth environment of garlic was good or the input of production factors was appropriate, so that the growth state of the individual garlic was favorable for the growth of the root. However, as the length of the leaf sheath length increased, it was found to act as a negative indicator of the increase in the weight of the green bulb. The energy that should have been supplied to the roots contributed to the growth of the leaves, reducing the live bulb weights.

In the proposed model, the variables selected as lot variables were the cultivation area, the number of plants, and the number of missed plants; the effects of the lot variables were shown to be negative in all growth stages. In all growth stages, as the cultivation area increases, the final bulb weight decreases, and it is thought that the larger the cultivation area, the wider the farmhouse's management range, which negatively affects the growth of individual bulbs. The number of plants has a negative effect on final bulb weight, since higher planting densities negatively affect bulb development. As the number of individuals sharing nutrients within a certain space increases, it may not be good for the growth of a single individual because each needs enough space for its roots to grow. The final bulb weight of garlic is inversely proportional to the missed plants. The occurrence of missed plants means that the seed itself was poor or that the growth environment was not good enough.

In the proposed model, the variables selected as environmental variables are minimum temperature, precipitation, and irradiance. Maximum temperature and GDD were selected in Stages B and C, while maximum temperature and GDD were dropped out in Stage A, and maximum temperature was dropped out in Stage C. Average temperature was not chosen as an important variable in any growth stage. It is considered that if the minimum temperature is high, overgrowth phenomena, which often occur in production areas, appear. This can also be found in the GDD variable, which refers to the environment in which the growth energy of plants can be accumulated. GDD shows a negative relationship with the final bulb weight in Stage B, and it was found that the accumulated energy was transferred to the aboveground part of the

1) They can be calculated as $RMSE = \sqrt{\frac{1}{T} \sum_t (y_t - \hat{y}_t)^2}$ and $MAE = \frac{1}{T} \sum_t |(y_t - \hat{y}_t)|$, where t is the periods being predicted (Greene, 2012).

garlic, such as the scape, rather than the root, so that it had a negative effect on root growth. The maximum temperature was selected as a significant variable in Stages B and C, indicating that the warmer temperature in spring, when garlic grows rapidly, is more beneficial to the growth of garlic above and below the ground.

Precipitation was found to have a positive effect on garlic growth in Stages A and B, but it was shown to have a negative effect after Stage C. This is thought to be because the supply of moisture can help garlic growth in dry environments in winter and spring, but after Stage C, when the temperature rises and farmers are able to control the temperature, excessive moisture supply acts as a factor that hinders the growth of garlic. The irradiance has a negative effect on the bulb weight, except in Stage B, since it is considered that the growing environment may become dry or an environment with excessively high temperature may be created if the amount of irradiance is high. However, in Stage B, irradiance is selected as a variable that helps to generate energy, showing that environmental variables act differently on garlic growth at different times. Bulb weights increased in proportion to the growing period. It is thought that timely transplantation or a sufficient period for root growth have positive effects on growth.

The predicted value of the bulb weight of garlic can be specifically calculated, based on the estimation result of the Lasso regression model. The expected bulb weight of garlic can be estimated, by obtaining the average values of the factors that affect the bulb weight of garlic and substituting the values into the coefficients and variables presented in Table 4. Estimating the expected bulb weight of garlic by applying the 2022 data values, the predicted values for each stage were shown as 76.2, 68.4, 63.4, and 55.0, respectively. Considering that the actual value of the bulb weight of garlic is 59.1 g, the figures calculated from the model show error rates of about 6.8 – 22.4%, and the error rate tends decrease as the stages progress.

Table 4. The proposed model: results of LASSO estimation for each stage

Variables	Stage A	Stage B	Stage C	Stage D
Growth State				
Plant height	0.09	0.41	0.62	0.30
Leaf number	7.10	4.79	1.57	0.88
Leaf sheath length	-1.16	-1.30	-0.94	
Leaf sheath diameter	0.85	1.85	2.34	3.07
Bulb weight			0.36	0.44
Lot Information				
Cultivation area	-0.00	-0.00	-0.00	-0.00
Planting Density	-0.35	-0.35	-0.32	-0.42
Missed plant	-0.49	-0.46	-0.27	
Environment Information				
Max temperature		0.89	0.62	
Min temperature	-2.26	-0.57	-0.68	-3.57
Precipitation	0.13	0.13	-0.03	-0.06
Irradiance	-0.01	0.05	-0.05	-0.20
GDD		-0.12	0.07	0.34
Other factors				
Growth period	0.05	0.02	0.08	0.24
Constant	62.87	30.31	9.04	0.11

The results of estimating the bulb weight of garlic based on LASSO imply three main conclusions. First, most of the variables representing the growth status of garlic were selected as significant variables for estimating bulb weight. This means that growth information has a greater influence on forecasting bulb weight than lot information or environmental information, and that growth status can be used as the most intuitive and fundamental indicator to predict the weight of garlic. Second, the variables of growth information and lot information maintain the same sign even as the growth stage progresses. This means that the growth information represented at each stage explicitly reflects the growth status of the bulb, and lot information represents the basis for the growth of garlic, implying that the growth conditions established from the planting stage are maintained throughout all growing stages. Third, environmental information can be meaningful for estimating the weight of garlic, but the signs of the related variables are not constant. This implies that environmental factors may have different effects at different stages of growth, and that the effects on bulb growth should not be interpreted piecemeal.

Conclusion

In this study, the indicators and factors affecting the estimation of garlic bulb weight were identified, and a prediction model was constructed. In the process, using LASSO analysis, which enables variable selection and coefficient estimation, variables to be used as indicators for estimating the final bulb weight of garlic were examined by growth stage. This study has significance in that it supplemented the limitations of existing experimental research using actual survey data and estimated the bulb weight of garlic by taking into account the characteristics of outfield crops. In addition, through variable selection, an optimal model was constructed for forecasting the final bulb weight of garlic by stage so that observers could intuitively apprehend which variables were important at each stage. In this respect, this study has an advantage over previous studies that do not consider models by point of time, providing indicators for forecasting final production at each stage.

This study has the limitation that management information or farmhouse learning effects were not included in the model-building process. In the actual cultivation process of outfield crops, there may be various efforts (for example, adaptation to the environment or pest control) to increase crop productivity, but in the models in this paper, the effects were reflected as errors, so the effects on the related information were not represented explicitly. In future studies, it will be necessary to improve the forecast model for estimating the bulb weight of garlic by additionally considering the variables affecting outfield crops. This paper is significant in that it deals with the important environmental factors, for garlic growth improving the predictive power of estimating garlic bulb weight. However, since the environment of outfield vegetables such as garlic is not constantly repeated every year, it is necessary to continuously develop the model, updating the related survey information. In addition, it is expected that the accuracy of estimation can be further improved through a hybrid model applying the environmental factors selected through the LASSO regression to forecasting models such as machine learning and deep learning.

Literature Cited

- Atif MJ, Amin B, Ghani MI, Ali M, Cheng Z** (2020) Variation in morphological and quality parameters in garlic (*Allium sativum* L.) bulb influenced by different photoperiod, temperature, sowing and harvesting time. *Plants (Basel)* 9:155. doi:10.3390/plants9020155
- Barreto H, Howland FM** (2006) *Introductory econometrics*. Cambridge University Press, Cambridge, New York
- Choi S, Kim J, Seo H** (2021b) Estimation of garlic's bulb weight at harvest using a multi-level model based on growth and meteorological data. *Hortic Sci Technol* 39:521-529. doi:10.7235/HORT.20210047
- Choi SH, Shin WJ, Bong YS, Lee KS** (2021a) Determination of the geographic origin of garlic using the bioelement content and isotope signatures. *Food Control* 130:108339. doi:10.1016/j.foodcont.2021.108339
- Diriba-Shiferaw G** (2016) Review of management strategies of constraints in Garlic (*Allium sativum* L.) production. *J Agric Sci* 11:186-207. doi:10.4038/jas.v11i3.8172
- Greene WH** (2012) *Econometric analysis*. Pearson Education, Essex, UK
- Halcoussis D** (2005) *Understanding econometrics*. South-Western, OH, USA
- Hastie T, Tibshirani R, Friedman JH, Friedman JH** (2009) *The elements of statistical learning: data mining, inference, and prediction*. Springer, NY, USA
- Hsiao J, Yun K, Moon KH, Kim SH** (2019) A process-based model for leaf development and growth in hardneck garlic (*Allium sativum*). *Ann Bot* 124:1143-1160. doi:10.1093/aob/mcz060
- Kamenetsky R, Shafir IL, Zemah H, Barzilay A, Rabinowitch HD** (2004) Environmental control of garlic growth and florigenesis. *J Am Soc Hortic Sci J* 129:144-151. doi:10.21273/JASHS.129.2.0144
- Kennedy P** (2008) *A guide to econometrics*. Blackwell Publishing, MA, USA
- Kim DJ, Kim SO, Kim JH, Yun EJ** (2019) Establishment of geospatial schemes based on topo-climatology for farm-specific agrometeorological information. *Korean J Agric For Meteorol* 21:146-157. doi:10.5532/KJAFM.2019.21.3.146
- Kim J, Kim S** (2021) Forecasting bulb weight of median-late variety onion using growth indicators and meteorological factors. *J Rural Dev* 44:1-26. <https://library.krei.re.kr/pyxis-api/1/digital-files/2114809d-bb01-4870-9d6f-e7f8b65cd2ed>
- Kim W, Han E, Shin S, Gouk S, Seo H** (2020) A study on the setting of the standard supply and demand for the efficient operation of the vegetable price stability system. Korea Rural Economic Institute, Naju, Korea
- Kim W, Jo N, Kim C, Shin S, Yang J** (2021) Trends and projection of supply of spice plants. *Agricultural outlook 2021: Changes and future of agricultural and rural areas after COVID-19*. Korea Rural Economic Institute, Naju, Korea
- Konzen E, Ziegelmann FA** (2016) LASSO-Type Penalties for Covariate Selection and Forecasting in Time Series. *J Forecast* 35:592-612. doi:10.1002/for.2403
- Lee JJ, Kim H, Kim JC, Choi GJ** (2022) A bioassay for testing garlic resistance to leaf blight caused by *Stemphylium vesicarium*. *Hortic Sci Technol* 40:324-337. doi:10.7235/HORT.20220030
- Moon J, Kim G, Seo B** (2021) A study of forecasting cabbage yields using a mixed data sampling model. *J Rural Dev* 44:29-51. <https://library.krei.re.kr/pyxis-api/1/digital-files/ba4ee63f-ac0b-49ca-aebb-5a68e43f1e29>
- Oh S, Moon KH, Koh SC** (2017) Effects of different day/night temperature regimes on growth and clove development in cool-type garlic (*Allium sativum* L.). *Hortic Sci Technol* 35:1-10. doi:10.12972/kjst.20170001
- Oh SY, Koh SC** (2020) Growth and mineral content of southern-type garlic grown in volcanic and non-volcanic ash soils on jeju island. *Hortic Sci Technol* 38:56-65. doi:10.7235/HORT.20200006
- Oh SY, Moon KH, Song EY, Koh SC** (2019) Photosynthesis, growth, and clove formation of southern-type garlic in response to different day/night temperature regimes. *Hortic Sci Technol* 37:696-707. doi:10.7235/HORT.20190070
- Portela JA, Cavagnaro JB** (2004) Growing phases of the white garlic (*Allium sativum*) plant in relation to field temperature and day length. IV International Symposium on Edible Alliaceae 688, Beijing, China, pp 239-246. doi:10.17660/ActaHortic.2005.688.32
- Rahim MA, Fordham R** (1997) Environmental manipulation for controlling bulbing in garlic. II International Symposium on Edible Alliaceae 555, Adelaide, Australia, pp 181-188. doi:10.17660/ActaHortic.2001.555.27
- Shojaie A, Michailidis G** (2010) Discovering graphical Granger causality using the truncating lasso penalty. *Bioinformatics* 26:i517-i523. doi:10.1093/bioinformatics/btq377
- Tibshirani R** (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc, B: Stat Methodol* 58:267-288. doi:10.1111/j.2517-6161.1996.tb02080.x
- Vrieze SI** (2012) Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods* 17:228-243. doi:10.1037/a0027127
- Wu TT, Lange K** (2008) Coordinate descent algorithms for lasso penalized regression. *Ann Appl Stat*, 2:224-244. doi:10.1214/07-AOAS147