

CROP YIELD PREDICTION BASED ON GROWTH AND ENVIRONMENTAL FACTORS: A CASE STUDY OF ORIENTAL MELONS (*CUCUMIS MELO* L. VAR. *MAKUWA* MAKINO) IN SEONGJU REGION

MOON, S. H.¹ – NA, M. H.² – KIM, T. Y.³ – PARK, M. S.^{4*}

¹*Research Institute of Future Medicine, Samsung Medical Center, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Republic of Korea*

²*Department of Statistics, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju 61186, Republic of Korea*

³*Department of Statistics, Keimyung University, 1095 Dalgubeol-daero, Dalseo-gu, Daegu 42601, Republic of Korea*

⁴*Department of Information and Statistics, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea*

**Corresponding author*

e-mail: minsu.park51@gmail.com; phone: +82-42-821-5431; fax: +82-42-822-0260

(Received 23rd Aug 2021; accepted 23rd Nov 2021)

Abstract. In the Republic of Korea, Gyeongsangbuk-do accounts for 95% of the total oriental melon (*Cucumis melo* L. var. *makuwa* Makino) cultivation area and 91% of the yield. Recently, a smart farm system has been introduced that improves efficiency in farm management by collecting information on the growth and environmental factors for each facility farm. This study aimed to identify the factors that influence the increase in the crop yield of Seongju oriental melon through oriental melon information obtained from such a smart farm system, and to construct crop yield prediction models. The data used in the study was collected from four Seongju oriental melon facility farms measured from 2018 to 2019. To predict the crop yield of each farm per week, this study used principal component analysis, distributed-lag model, and auto-regressive integrated moving average with explanatory variables model (ARIMAX) based on growth and environmental variables. ARIMAX with principal components provided stable predictability. We expect the results of this study to be useful in developing optimal farming techniques that can improve the crop yield for farms cultivating Seongju oriental melon in facilities.

Keywords: *Seongju oriental melon, harvest, smart farm, ARIMAX model, principal component analysis*

Introduction

Agriculture was at the center of the first human settlements, enabling them to wait for harvest and cultivate nature to secure a stable food supply beyond hunting and gathering. Large quantities of high-quality agricultural products were harvested through efforts such as maintaining adequate solar radiation and temperature. In addition, the crop yield was increased by using appropriate fertilizers or by administering nutrients for specific agricultural products to remove harmful plants or insects around agricultural products and supply the nutrients necessary for growth.

However, climate change, manifesting in heat waves, abnormal temperatures, droughts, and heavy rains, acts as a factor that hinders the sustainable production of crops. Agriculture in particular is more affected by these environmental factors than other industries. Various studies and technological developments have been steadily

carried out to grow agricultural products under certain environmental conditions and thereby cope with such climate changes (Bindi and Olesen, 2011; Deschênes and Greenstone, 2007; Lee et al., 2012).

One such effort in response to these environmental changes is the use of smart farm systems. The smart farm system combines information and communications technologies (ICTs), such as the internet of things (IoT), big data, and artificial intelligence. The smart farm system improves yield and quality by maintaining and providing the environmental factors, such as temperature, solar radiation, humidity, and carbon dioxide, that are most suitable for the cultivation and growth of crops and livestock. In addition, the smart farm system enables automatic and remote management, and can be applied to reduce labor, increase productivity and convenience, and manage crops efficiently, leading to an increasing trend of farmers introducing smart farms (Shin and Jeon, 2020). Not only does it properly maintain the appropriate growing environment for agricultural products but it also reduces instability in agricultural product cultivation by detecting and reporting risks to users and reflecting important characteristics in agriculture. Through these smart farms, humans can cultivate and harvest plants stably regardless of the season by providing a suitable environment that is optimal for the growth and cultivation of each plant.

Table A1 (see *Appendix*) is the National Statistical Office's data on the area and yield of vegetable production (fruit) from 2017 to 2019. *Table A1* shows that more production per area is produced in facility farms than in outdoor farms. The crop yield of oriental melon (*Cucumis melo* L. var. *makuwa* Makino) was less than that of other fruits and vegetables. As of 2019, watermelons accounted for the largest harvest at 20.29%; tomato, 18.59%; cucumber, 16.84%; strawberry, 12.09%; pumpkin, 8.59%; and oriental melon, 7.62%. In the case of oriental melon, which has a lower yield than other fruits and vegetables, studies analyzing data collected from smart farms have been insignificant. Therefore, among fruits and vegetables, this study focused on oriental melon harvested in Seongju, Gyeongsangbuk-do, Republic of Korea, to identify the most suitable environment for the growth of Seongju oriental melon and predict crop yield.

Oriental melon is one of the *Cucurbitaceae* family in terms of total production and harvestable area worldwide (Shin et al., 2017). *Cucurbitaceae* family crops, such as oriental melons, cucumbers and pumpkins, mostly use grafting methods. Among them oriental melons can be grafted onto *Cucurbita* spp., bottle gourd, and *C. melo* rootstocks (King et al., 2010).

Grafting is to connect a scion to a rootstock and make it into a single plant. Grafting not only alleviates the damage caused by inadequate environmental factors and illnesses, but also increases growth speed, yields and affects quality improvements (Goldschmidt, 2014). *Figure 1* describes the growth factors of oriental melons associated with grafting.

In addition to growth factors, oriental melons with poor quality values such as cracking, fermentation, and deformation may be harvested depending on environmental factors, and yields will decrease. Fruit cracking is a phenomenon in which fruit cracks due to a large difference in growth velocity between the surface and the interior of the fruit. Fruit cracking usually occurs in dry soil and in humid environments in the air at the time of fruit formation (Hwang et al., 1999). Fermented fruit is a phenomenon in which pulp is browned and water is filled inside. It is caused by complex environmental factors and cultivation conditions, such as sudden supply of moisture while the soil is

dry or lack of sunlight. Therefore, sufficient solar radiation must be provided during periods of fruit growth, and the soil must maintain constant humidity and lower airborne humidity at night (Hwang et al., 1999). Deformation fruit is a fruit that grows in a different shape from normal, resulting from a lack of solar radiation at a time when the fruit grows. As such, growth factors and yield are subject to many environmental factors. Therefore, it is necessary to control through optimal environmental factors using smart farm systems.



Figure 1. Description of the growth factors of oriental melons (from 'nongsaro' homepage: <http://nongsaro.go.kr>)

Using advanced facilities, farms that cultivate fruits and vegetables, such as cucumber, pepper, tomato, and olive, have introduced smart farms to increase production; optimal environmental analysis has helped identify the factors influencing production (Galán et al., 2004; Kang et al., 2020; Lee et al., 2020, 2019). As such, smart farms are being actively introduced for producing fruits and vegetables with a high production weight and analysis studies are being actively conducted accordingly.

Early studies on crop yields using smart farm data in Republic of Korea have focused mainly on fruits and vegetables with a high proportion, such as watermelon, strawberry, and tomato. Lee et al. (2019) used environmental factor variables to explore the factors that affect the growth of plant cucumbers through multiple linear regression analysis. Their results showed that internal cumulative temperature, night average internal relative humidity, and maximum soil humidity, maximum soil moisture tension influence the production. Kang et al. (2020) predicted the production of pepper by region using multiple linear and nonlinear regression analyses using environmental factors, growth factors, and variables with or without irrigation facilities. In determining the yield of red pepper, the study revealed that plant height, main length, length, and branches are significant factors, and the presence or absence of irrigation facilities is a particularly important factor.

Galán et al. (2004) developed a multi-regression model to predict future olive fruit yields for effective and optimized agricultural management. At this study, rainfall five months before the forecast and pollen emissions eight months before the forecast had a significant impact on fruit yields. Choudhury and Jones (2014) predicted crops harvested over different periods in the same region using several time series methods to demonstrate a more robust time series model than smoothing techniques. In this study, auto-regressive moving average showed higher decision coefficients compared with

other time series models, such as simple, double, and attenuation trend linear exponential smoothing. The problem of variable crop yields within a region could thus be avoided over time. Bertin and Génard (2018) reviewed several models to simulate tomato growth and quality using Process-based simulation models and investigated the potential application of ideas design of ideotypes, such as conceptual plants, which are expected to be performed in a particular environment. Di Vittori et al. (2018) defines an objective concept of strawberry quality and summarizes the key pre-harvest factors affecting the quality of the fruit. It was analyzed that climate variables such as soil conditions, production cycles, and factors related to cultivation systems such as traditional or soil-free cultivation affect the final quality of strawberry fruits.

In this study, we aim to explore the environmental and growth factors with high yield obtained from smart farms for relatively unstudied oriental melons due to lower yields than other fruits and vegetables, and provide models for crop yield prediction.

Materials and methods

Data description

We collected crop yield, environmental, and growth data for oriental melon from four facility farms that cultivate oriental melons in Seongju. The GPS coordinates for four farms are 35°56'34.0"N 128°13'09.0"E, 35°55'16.0"N 128°16'12.0"E, 35°57'20.0"N 128°17'51.0"E, and 35°56'15.0"N 128°13'21.0"E, respectively. The data used for the analysis were provided by the Rural Development Administration, Republic of Korea, and the response variable was defined as oriental melon yield per 10a (kg/10a). Moreover, the growth and facility farm environmental variables were used as explanatory variables. We measured the oriental melon yield per 10a (kg/10a) and growth variables every week from December 19, 2018 to July 9, 2019, and the environmental variables of facility farms were collected in minutes, from January 24, 2019 to June 4, 2019. Therefore, we used common data obtained from January 24, 2019 to June 4, 2019, or a total of 19 weeks.

Initially, we considered 8 growth and 28 environmental variables. Some variables were created through averages or cumulates to expand specifically the growth and environmental variables. Growing degree day (GDD) variable was defined by taking the average of the daily maximum and minimum temperatures compared to the base temperature of the oriental melon. For factors related to temperature in environmental variables, we used the following: average, maximum, minimum, daily average, nighttime average, GDD, and daily temperature range.

Data in each facility farm showed no outlier. Therefore, the average value of 10 oriental melon trees in each facility farm was used to represent them as one data. *Table 1* describes the response and explanatory variables. For additional information on the measurement methods and experimental conditions of the listed variables in *Table 1* collected through smart farm technology, we can refer to sites of the Rural Development Administration (<http://www.nongsaro.go.kr>) and the Korea Agency of Education, Promotion and Information Service in Food, Agriculture, Forestry and Fisheries (<https://www.epis.or.kr/main/view>). Because similar variables exist, we separately marked the variables used in our model ("†" and "‡" in *Table 1*) by removing semantically redundant variables. Furthermore, we used the reduced variables marked "†" in the prediction models without principal components (PCs) in consideration of multicollinearity between explanatory variables (*Table 1*).

Table 1. Description of variables

Variable name	Explanation	Unit
Response variable		
yield 10a	Weekly cumulative yields area of 10a	kg/10a
Explanatory variables: growth		
ll	Leaf length	cm
lw	Leaf width	cm
sd rs †, ‡	Stem diameter of rootstock	cm
sd s †	Stem diameter of scion	cm
sd gp	Stem diameter of grafting point	cm
sd	Stem diameter	cm
crp	Chlorophyll	-
angel †, ‡	Angle of growing point	°
Explanatory variables: environment		
cumulate sun †	Cumulate of solar radiation	W/m ²
mean sun †, ‡	Average of solar radiation	W/m ²
mean ltemp †	Average of leaf temperature	°C
daymean ltemp †	Day average of leaf temperature	°C
nightmean ltemp	Night average of leaf temperature	°C
max ltemp	Highest leaf temperature	°C
min ltemp	Lowest leaf temperature	°C
mean 2h sunrise ltemp	Average leaf temperature at two hours before and after sunrise	°C
mean temp †	Average of temperature	°C
daymean temp †	Day average of temperature	°C
nightmean temp	Night average of temperature	°C
max temp	Highest temperature	°C
min temp	Lowest temperature	°C
mean 2h sunrise temp	Average temperature at two hours before and after sunrise	°C
GDD temp	Growing degree day temperature	°C
DIF temp †, ‡	Daily temperature range	°C
mean stemp †	Average of soil temperature	°C
daymean stemp	Day average of soil temperature	°C
nightmean stemp †	Night average of soil temperature	°C
max stemp †, ‡	Highest soil temperature	°C
min stemp †	Lowest soil temperature	°C
mean 2h sunrise stemp †	Average soil temperature at two hours before and after sunrise	°C
GDD stemp †, ‡	Growing degree day soil temperature	°C
mean mois	Average of moisture	%
daymean mois	Day average of moisture	%
nightmean mois †, ‡	Night average of moisture	%
max mois	Highest moisture	%
min mois	Lowest moisture	%

† = variables used to generate principal component; ‡ = variables considering multicollinearity

Workflow

This study aimed to identify the significant factors that affect the crop yield of oriental melon between farms and estimate prediction models for crop yield. The response variable was defined as the crop yield per 10a (kg/10a) of oriental melon calculated as the weekly cumulative yield. The explanatory variables consisted of the average growth and environmental variables for that week. To construct prediction models, we examined whether the explanatory variables of the past lag and response variable of the current time were correlated with Pearson's correlation coefficients. The

lag with the explanatory variable having the most influence on the response variable was then selected. The variables were expanded into eight-week lags from the past to the current time, and the correlation coefficients with the response variable were measured and compared. Finally, the optimal lag was selected using the largest average of the absolute correlation coefficient values of all farms. Prediction models were constructed according to the following procedure:

1. Select optimal lag for each explanatory variable, called a lagged variable.
2. Construct time-series models based on lagged variables or PCs of lagged variables.
 - A. In the lagged variable-based models, only nine lagged variables (marked “‡” in *Table 1*) are used, excluding variables with high similarity in consideration of multicollinearity.
 - B. In the PCs-based models, all lagged variables (marked “†” in *Table 1*) are used.
3. Compare the performance measures of estimated models.

All statistical analyses with a significance level of 0.05 for a two-sided test were performed using R version 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria).

Statistical methods

Principal component analysis

When the number of explanatory variables is greater than the number of observations in measured data and multicollinearity between explanatory variables is high, then the number of variables needs to be reduced using principal component analysis (PCA). In multivariate data of size, PCA is a method of extracting important information and generating new variables ($PC^T = (PC_1, \dots, PC_k)$), which are new variables independent of one another, through a linear combination of correlated $x^T = (x_1, \dots, x_k)$ variables (Hotelling, 1933). PCA can solve the multicollinearity problem and reduce the number of variables.

Distributed-lag model

The distributed-lag model (DLM) is a method for analyzing linear relationships with time-dependent variables by including past lags of explanatory variables as explanatory variables (Pesaran et al., 2001). The general form of DLM for the dependent variable is expressed as

$$Y_t = \alpha + \sum_{i=1}^k \sum_{j=1}^{h_i} \beta_{i,j} X_{i,t-j} + \varepsilon_t, \quad (\text{Eq.1})$$

where α is a constant term, k are the number of explanatory variables, h_i are the lag of the i -th explanatory variable X_i , $\beta_{i,j}$ are the coefficient of X_i at the j -th lag, and ε_t is an error term at time t . In this study, h_i were used as the optimal with the largest correlation coefficient for the response variable among lags of X_i .

Auto-regressive integrated moving average model

Auto-regressive integrated moving average with explanatory variables model (ARIMAX) is a linear combination of an auto-regressive (AR) model and a moving average (MA) model with explanatory variables (Box et al., 2015). The general form of ARIMAX for Y_t is represented as

$$\Delta^d Y_t = \alpha + \sum_{i=1}^p \phi_i \Delta^d Y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{m=1}^k \beta_m X_{m,t} + \varepsilon_t, \quad (\text{Eq.2})$$

where α is a constant term, p is the autoregressive order. ϕ 's are the autoregressive coefficients, q is the moving average order, θ 's are the moving average coefficients, ε_t is an error term at time t , k is the number of explanatory variables, β 's are the regression coefficients, and X 's are the lagged explanatory variables. In this study, we set $d=0$ for simplicity. Thus, the final model can be expressed as ARIMA($p,0,q$)-lagged X model. Here, the orders ($p,0,q$) are automatically determined using the `auto.arima` function of the R package `{forecast}`, which selects the order that minimizes Akaike information criterion (AIC).

Model performance

The predicted models were evaluated using root mean square error (RMSE), root mean absolute error (RMAE), and mean magnitude of relative error (MMRE) measures. These model performance measures are defined by

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}, \quad (\text{Eq.3})$$

$$RMAE = \sqrt{\frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|}, \quad (\text{Eq.4})$$

$$MMRE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|. \quad (\text{Eq.5})$$

Here, n is the number of measured data, t is the measurement time point, y_t is the actual value at time t , and \hat{y}_t is the predicted value at time t . As the error decreases, the values of the evaluation indicators RMSE, RMAE, and MMRE decrease. Therefore, the smaller the value of the model performance measures, the better the estimated model predicts the response variable.

Results

In this study, to find and predict the factors affecting the weekly yield per 10a (kg/10a) of oriental melon, we used the growth and environmental variables measured at four Seongju oriental melon facility farms in 2019.

Statistics of variables

Prior to the analysis, basic statistics on the yield, growth factors and environmental factors of the response variable oriental melon were measured on a farm-by-farm basis. Figure 2 showed basic statistics of the yield of oriental melon by farm. Figure 2 show that the highest yield is 124.8 kg/10a, which is confirmed in Farm2. Farm3 had a lower yield than other farms, with a peak yield of 63.7 kg/10a. Table 2 measures basic statistics for each facility farms growth factors and environmental factors.

Table 2. Statistics of explanatory variables

Variable	Mean ± SD	Min~Max	Median(Q1, Q3)
Farm1			
sd rs	1.2 ± 0.3	0.8~2.0	1.2(1.1, 1.4)
sd s	0.8 ± 0.2	0.5~1.4	0.7(0.7, 0.8)
angle	35.3 ± 5.1	22.9~53.2	34.1(33.4, 37.2)
mean temp	20.5 ± 4.4	7.6~25.8	21.1(19.3, 23.2)
daymean temp	25.2 ± 5.5	9.3~31.7	25.9(23.8, 28.9)
DIF temp	22.5 ± 5.7	8.2~32.0	23.1(20.5, 25.9)
mean stemp	19.7 ± 4.0	7.1~24.4	20.2(19.0, 21.9)
night mean stemp	19.9 ± 4.1	7.1~24.4	20.4(19.1, 22.3)
min stemp	18.4 ± 3.7	6.7~23.2	18.8(17.6, 20.5)
max stemp	21.2 ± 4.4	7.5~25.7	21.6(20.5, 23.8)
mean 2h	18.8 ± 3.9	6.8~23.6	19.2(18.0, 20.9)
sunrise stemp	18.9 ± 3.9	6.8~23.4	19.5(18.0, 21.2)
daymean ltemp	22.9 ± 4.8	8.3~27.8	23.4(21.8, 26.3)
nightmean mois	88.6 ± 18.6	24.7~112.7	91.1(84.2, 97.2)
mean sun	141.5 ± 47.9	60.0~226.4	137.1(104.6, 181.2)
cumulate sun	20237.0 ± 6921.0	7944.3~32597.5	19742.2(15041.7, 26095.3)
GDD stemp	84.9 ± 23.6	34.4~125.0	86.5(69.7, 99.0)
Farm2			
sd rs	1.1 ± 0.2	0.6~1.4	1.1(1.1, 1.2)
sd s	0.7 ± 0.1	0.4~1.0	0.7(0.7, 0.8)
angle	30.3 ± 7.0	18.5~46.2	30.3(23.8, 35.2)
mean temp	19.6 ± 4.2	9.6~25.1	19.5(18.6, 21.7)
daymean temp	24.7 ± 5.2	11.8~32.1	24.9(23.6, 28.5)
DIF temp	23.0 ± 5.3	10.0~32.5	23.3(21.5, 24.5)
mean stemp	20.1 ± 4.1	9.8~26.1	20.2(18.7, 22.3)
night mean stemp	20.3 ± 4.1	10.0~26.4	20.4(18.9, 22.7)
min stemp	19.1 ± 3.8	9.2~24.4	19.2(18.0, 20.9)
max stemp	21.2 ± 4.3	10.4~27.6	21.3(19.7, 23.7)
mean 2h	19.5 ± 3.9	9.4~25.1	19.6(18.3, 21.4)
sunrise stemp	18.3 ± 3.7	8.9~23.2	18.4(17.3, 20.5)
daymean ltemp	22.2 ± 4.5	10.8~28.8	22.3(21.3, 25.5)
nightmean mois	92.1 ± 18.5	33.2~114.6	95.0(92.0, 101.7)
mean sun	139.0 ± 49.2	46.3~221.2	134.2(98.6, 177.7)
cumulate sun	19898.6 ± 7126.0	6283.4~31852.1	19326.3(14204.0, 25588.3)
GDD stemp	86.9 ± 21.1	43.3~121.0	86.9(75.9, 97.2)
Farm3			
sd rs	1.1 ± 0.2	0.5~1.5	1.2(1.1, 1.2)
sd s	0.7 ± 0.2	0.3~1.4	0.7(0.6, 0.8)
angle	33.5 ± 6.8	13.5~42.4	35.3(30.5, 37.9)
mean temp	20.7 ± 2.8	12.6~24.3	20.8(20.4, 22.9)

daymean temp	25.1 ± 3.4	15.3~29.3	25.2(24.4, 27.1)
DIF temp	22.3 ± 3.9	12.9~31.4	22.4(20.4, 24.0)
mean stemp	22.2 ± 3.0	13.5~25.9	22.8(22.1, 23.8)
night mean stemp	22.1 ± 3.0	13.6~25.7	22.4(21.9, 23.9)
min stemp	20.3 ± 2.8	12.4~23.5	21.1(20.2, 21.6)
max stemp	24.4 ± 3.3	14.6~29.5	25.1(24.0, 25.6)
mean 2h	20.7 ± 2.9	12.7~24.0	21.4(20.6, 22.1)
sunrise stemp			
mean ltemp	20.3 ± 2.8	11.4~24.4	20.7(19.8, 21.8)
daymean ltemp	23.7 ± 3.2	13.6~27.1	23.9(23.0, 26.5)
nightmean mois	87.5 ± 14.6	43.3~110.8	89.5(84.6, 95.5)
mean sun	125 ± 35.9	49.1~197.1	121.1(98.6, 150.1)
cumulate sun	17930.6 ± 5120.0	7075.5~27670.1	17436.0(14199.1, 21613.2)
GDD stemp	102.5 ± 15.8	66.5~128.6	105.7(96.6, 112.7)
Farm4			
sd rs	1.3 ± 0.3	0.6~2.1	1.2(1.0, 1.5)
sd s	0.9 ± 0.2	0.5~1.5	0.9(0.8, 1.0)
angle	34.6 ± 4.9	15.7~41.9	35.3(33.2, 36.7)
mean temp	22.9 ± 3.5	18.1~32.2	22.2(20.4, 24.6)
daymean temp	25.6 ± 3.5	19.5~35.0	25.4(23.7, 27.9)
DIF temp	14.0 ± 3.8	8.5~22.9	13.2(11.2, 16.3)
mean stemp	23.0 ± 3.2	17.9~31.6	22.9(20.9, 24.9)
night mean stemp	22.7 ± 3.3	17.5~31.6	22.2(20.5, 24.4)
min stemp	20.8 ± 3.4	15.6~29.5	20.0(18.5, 21.8)
max stemp	25.6 ± 3.0	21.0~33.4	24.9(23.5, 27.2)
mean 2h	21.2 ± 3.4	15.9~30.1	20.5(18.8, 22.2)
sunrise stemp			
mean ltemp	22.8 ± 3.5	18.0~32.5	22.2(20.4, 24.5)
daymean ltemp	26.7 ± 3.3	22.0~36.5	26.4(24.4, 28.7)
nightmean mois	78.6 ± 8.7	60.2~97.3	80.8(71.4, 83.9)
mean sun	132.3 ± 43.4	49.1~203.4	134.2(101.7, 165.4)
cumulate sun	18943.0 ± 6245.6	7075.5~28288.8	19331.9(14648.7, 23813.1)
GDD stemp	106.0 ± 21.3	70.4~156.0	102.5(91.2, 118)

SD = standard deviation; Q1 = 1st quartile; Q3 = 3rd quartile

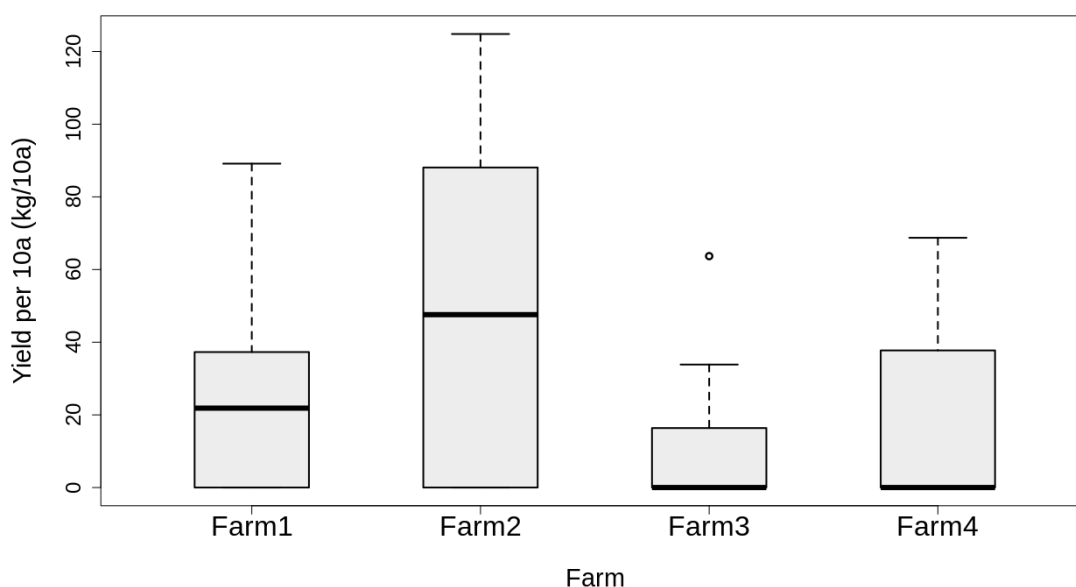


Figure 2. Statistics of response variable over the entire period

Correlation analysis

To confirm the relation between the explanatory (growth and environmental) variables with lags and crop yield, we conducted a correlation analysis. *Table A2* represents the average correlation coefficient of the farm between lagged covariates and yield of oriental melon per 10a (kg/10a). For each selected lag, the lagged covariates affecting crop yield were mostly 2- or 6-week lags (*Table A2*). In the growth variables, the stem diameters of rootstock and stem diameters of scion at the time of lag 2 were strongly correlated with crop yield. In addition, angle of growing point, a growth variable, and temperature related variables, which were environmental variables, showed a strong correlation with crop yield at the time of lag 6. Therefore, as above, we used the covariate at the selected time point where the correlation with the crop yield was large to construct the model. The lags of night average of moisture, average of solar radiation, cumulate of solar radiation, and GDD soil temperatures, all environmental variables, were determined to be 2, 3, 0, and 7, respectively.

Next, the correlation analysis between the explanatory variables of the four farms, shown in *Table A3*, confirmed potential multicollinearity issues. Among the growth variables, in all farms, reception diameter and stock tree diameter showed a high correlation. Among the environmental variables, most of those related to solar radiation, soil temperature, and night average humidity were highly correlated. In addition, the average soil temperature had correlation coefficient values close to 1 with other temperature-related variables. Finally, we determined the explanatory variables that could be used in the prediction model with covariates, marked with “‡” in *Table 1*.

Prediction models

We performed PCA to solve the problem of multicollinearity in lagged explanatory variables. First of all, All variables used were standardized in the PCA process. The number of PCs was determined to include a dimension that accounted for up to 99% of the cumulative variance over the total data. We found that eight PCs could be used for each farm. Finally, based on the lagged variables and selected PCs, we identified and predicted the factors affecting oriental melon yield per 10a (kg/10a) using DLM and ARIMAX, respectively. When explanatory variables are used directly in the model without PCA, some variables with multicollinearity were excluded. In addition, in the case of using PCs, the prediction model was constructed using all explanatory variables, and the explanatory variables used in each approach are represented (*Table 1*).

As a result of selecting the order of the model that minimizes AIC, in ARIMA with covariates, the auto-regressive and moving average orders were selected as 1, respectively, and thus the ARIMA(1,0,1) model was used. Similarly, in ARIMA with PCs, the orders were selected in the same way, and the ARIMA(1,0,3) model was used. To test whether the residuals from the ARIMA model had no autocorrelation, we used the Ljung–Box test. The Ljung–Box test results confirmed that the significance probability value of all models in the four farms was over 0.05, indicating the absence of serial autocorrelation (*Table A4*).

The results of the comparison between the actual and predicted values for each farm using the four models are shown in *Figure 3*. In this figure, the black solid lines are the actual yield values across weeks. The green, blue, yellow, and green dotted lines represent the fitted values for DLM and ARIMAX(1,0,1) with covariates and DLM and

ARIMAX(1,0,3) with PCs, respectively. Based on these results, the predicted values (red dotted line) obtained from ARIMAX(1,0,3) with PCs were closest to the actual values, compared with other models. Thus, yield predictability improved when all covariates were used and their multicollinearity was controlled.

Estimated models that could predict oriental melon yield were compared through the following evaluation measurements: RMSE, RMAE, and MMRE. The comparison results of the model performance confirmed that the RMSE, RMAE, and MMRE of ARIMAX with PCs had the lowest values in all farms (Table 3). Therefore, ARIMAX(1,0,3) with PCs performed better than other prediction models.

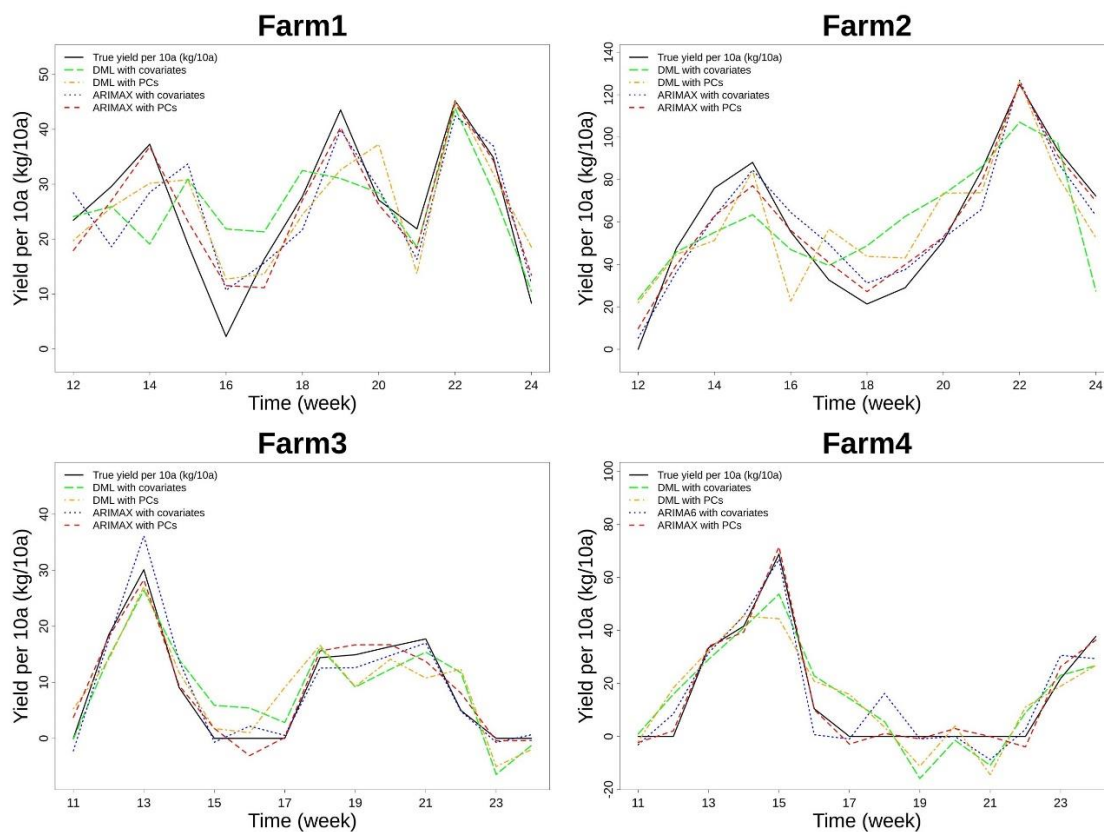


Figure 3. Prediction results of yield per 10a (kg/10a) in four farms (1 to 4)

Table 3. Comparison of predictability between estimated models

Farm	Model	Measurement		
		RMSE	RMAE	MMRE
Farm1	DLM with covariates	9.326	2.640	0.8732
	DLM with PCs	7.569	2.560	0.6385
	ARIMAX(1,0,1) with covariates	6.894	2.389	0.5021
	ARIMAX(1,0,3) with PCs	4.093	1.788	0.4538
Farm2	DLM with covariates	22.207	4.264	2.0864
	DLM with PCs	18.859	4.049	1.9467
	ARIMAX(1,0,1) with covariates	10.187	2.974	0.7916
	ARIMAX(1,0,3) with PCs	7.423	2.486	0.5081

Farm3	DLM with covariates	4.419	1.988	1.7927
	DLM with PCs	4.817	2.049	1.8119
	ARIMAX(1,0,1) with covariates	2.450	1.347	0.8906
	ARIMAX(1,0,3) with PCs	2.080	1.266	0.7061
Farm4	DLM with covariates	10.237	2.911	5.286
	DLM with PCs	11.717	3.091	5.724
	ARIMAX(1,0,1) with covariates	7.060	2.332	3.207
	ARIMAX(1,0,3) with PCs	2.384	1.413	1.325

RMSE = root mean square error; RMAE = root mean absolute error; MMRE = mean magnitude of relative error; DLM = distributed-lag model; PC = principal component; ARIMAX = auto-regressive integrated moving average with explanatory variables

Table 4 shows the ARIMAX model for each farm using the PCs in the models. In these models, standardized angle of growing point, standardized stem diameters of scion and standardized day average of leaf temperature showed relatively higher coefficients compared with other variables. These coefficients were recalculated using the weights for each PC.

Table 4. Estimated model equation for ARIMAX with PCs in each farm

Estimated models for ARIMAX(1,0,3) with PCs	
Farm1	$\hat{Y}_t = -197.15 - 0.56 \times ar_t - 0.11 \times ma_1 + 0.11 \times ma_2 - 1.0 \times ma_3$ $+ 14.73 \times sd_{rs_{t-2}} + 24.22 \times sd_{s_{t-2}} + 3.74 \times angle_{t-6} - 1.02 \times mean\ temp_{t-6}$ $- 3.69 \times daymean\ temp_{t-6} + 0.14 \times DIF\ temp_{t-6} + 0.79 \times mean\ stemp_{t-6}$ $+ 0.95 \times night\ mean\ stemp_{t-6} + 4.76 \times min\ stemp_{t-6} - 2.60 \times max\ stemp_{t-6}$ $+ 4.66 \times mean\ 2h\ sunrise\ stemp_{t-6} - 2.19 \times mean\ ltemp_{t-6} - 2.83 \times daymean\ ltemp_{t-6}$ $+ 0.79 \times nightmean\ mois_{t-2} - 0.09 \times mean\ sun_{t-3} + 0.001 \times cummlate\ sun_t$ $+ 0.38 \times GDD\ stemp_{t-7}$
Farm2	$\hat{Y}_t = 21.33 - 0.71 \times ar_t + 0.91 \times ma_1 - 0.91 \times ma_2 - 1.0 \times ma_3$ $+ 0.04 \times sd_{rs_{t-2}} + 1.39 \times sd_{s_{t-2}} + 1.62 \times angle_{t-6} - 0.83 \times mean\ temp_{t-6}$ $+ 1.78 \times daymean\ temp_{t-6} + 4.79 \times DIF\ temp_{t-6} - 1.45 \times mean\ stemp_{t-6}$ $- 0.67 \times night\ mean\ stemp_{t-6} - 2.54 \times min\ stemp_{t-6} - 0.66 \times max\ stemp_{t-6}$ $- 2.17 \times mean\ 2h\ sunrise\ stemp_{t-6} - 3.35 \times mean\ ltemp_{t-6} + 1.27 \times daymean\ ltemp_{t-6}$ $+ 0.002 \times nightmean\ mois_{t-2} + 0.021 \times mean\ sun_{t-3} - 0.001 \times cummlate\ sun_t$ $+ 0.498 \times GDD\ stemp_{t-7}$
Farm3	$\hat{Y}_t = -86.55 - 0.87 \times ar_t - 0.91 \times ma_1 - 0.91 \times ma_2 + 1.0 \times ma_3$ $- 0.80 \times sd_{rs_{t-2}} - 1.96 \times sd_{s_{t-2}} - 0.77 \times angle_{t-6} - 0.91 \times mean\ temp_{t-6}$ $- 0.30 \times daymean\ temp_{t-6} + 1.47 \times DIF\ temp_{t-6} + 0.55 \times mean\ stemp_{t-6}$ $+ 0.40 \times night\ mean\ stemp_{t-6} + 0.41 \times min\ stemp_{t-6} + 0.80 \times max\ stemp_{t-6}$ $+ 0.25 \times mean\ 2h\ sunrise\ stemp_{t-6} + 0.46 \times mean\ ltemp_{t-6} + 0.94 \times daymean\ ltemp_{t-6}$ $+ 0.18 \times nightmean\ mois_{t-2} + 0.07 \times mean\ sun_{t-3} - 0.001 \times cummlate\ sun_t$ $+ 0.33 \times GDD\ stemp_{t-7}$

Farm4	$\hat{Y}_t = -390.2 - 0.87 \times ar_1 - 0.77 \times ma_1 + 2.73 \times ma_2 - 0.99 \times ma_3$ $- 1.55 \times sd\ rs_{t-2} - 0.94 \times sd\ s_{t-2} - 0.50 \times angle_{t-6} + 0.11 \times mean\ temp_{t-6}$ $+ 0.78 \times daymean\ temp_{t-6} + 0.83 \times DIF\ temp_{t-6} - 0.01 \times mean\ stemp_{t-6}$ $+ 0.15 \times night\ mean\ stemp_{t-6} + 0.22 \times min\ stemp_{t-6} - 0.14 \times max\ stemp_{t-6}$ $+ 0.12 \times mean\ 2h\ sunrise\ stemp_{t-6} - 0.44 \times mean\ ltemp_{t-6} + 0.92 \times daymean\ ltemp_{t-6}$ $+ 0.31 \times nightmean\ mois_{t-2} + 0.79 \times mean\ sun_{t-3} - 1.99 \times cummlate\ sun_t$ $+ 0.21 \times GDD\ stemp_{t-7}$
-------	--

ARIMAX = auto-regressive integrated moving average with explanatory variables; PC = principal component

Discussion

Oriental melon is mainly cultivated in Republic of Korea, Japan, and northern China in Central Asia (Hendryanti et al., 2020). The Republic of Korea, it is cultivated mainly in Seongju, in Gyeongsangbuk-do. Using smart farm data obtained from the Rural Development Administration, factors affecting the yield of Seongju oriental melon were identified and the predictive models were established. Through the results of this study, it is shown that the growth variables angle of growing point, stem diameters of scion and the day average of leaf temperature have a major effect on the yield of oriental melon in common in four farms. Also, the ARIMAX model with PCs has been proposed as an alternative model for Seongju oriental melon.

Growing points form organs such as root and stem through cell division. Clark and Burge (2002) tested plants by tubers size, and found that growing points had a significant relationship between the weights of tubers. The larger the weight of the tubers, the greater the size of the growing point. In the case of tubers of the same size, the larger the growing point, the better the plant grows. Mngrsquo et al. (2012) reviewed that plant growth is active when the gripping system diameters of scion is thicker. Stem diameters of scion found that thick trees grew faster and larger than other trees and produced much more. When the temperature is high, the stomata of the leaves close and the surface temperature of the leaves increases. Glenn et al. (2001) confirmed that the daytime air temperature was below 25 °C and that the daytime leaf surface temperature was low, increasing Carbon assimilation and statistical conductance, and increasing plant yields. Environmental variables such as day average of temperature and highest soil temperature were also identified as factors that increase the crop yield.

In addition, Liu et al. (2017) and Kim and Ferris (2002) found that the environment in which oriental melons are grown is dry, and soil moisture is grown in a relatively constant environment throughout the growth period. It has a better impact on oriental melons growth and therefore need to manage environmental factors. Shin et al. (2005) showed that when oriental melon is grown at a low temperature, the initial growth is delayed and the early yield decreases, which affects the production quantity and quality. In addition, the higher the heat retention, the faster the flowering and the shorter the number of harvesting days, the higher the fruit weight and sugar content.

In terms of lag time, the yield showed a high correlation with the environmental variables temperature, soil temperature, and leaf temperature 6 weeks before harvesting. In the oriental melon cultivation process, as shown in *Figure 4*, 6 weeks before harvesting is the fruiting setting season.

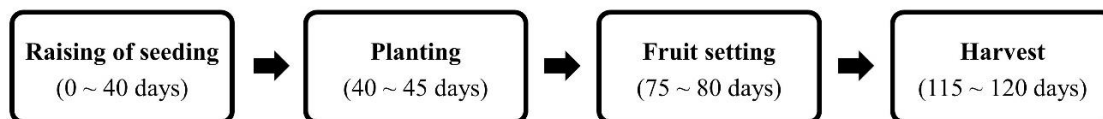


Figure 4. The process and duration of melon cultivation (from Seongju-gun homepage: <https://sj.go.kr/S0007/>)

The fruiting season of oriental melon is the time when the fruit is opened, and it is necessary to reduce the fall of the oriental melon and provide an environment suitable for the fruit to open, such as arranging vines and leaves, removing pests, and controlling environmental factors to improve the yield at harvest time. Also, excessively high temperatures interfere with fruiting and must be controlled. When the humidity is too high, the fruit becomes soft, and when the humidity is too low, the fruit may become malformed. Therefore, it is necessary to manage environmental factors during the fruiting season, 6 weeks before the oriental melon is harvested.

When the optimum temperature, humidity, soil temperature, and plant growth factors are identified for each crop and applied to the smart farm, it will be possible to increase the yield of not only oriental melon but also other crops and harvest high-quality crops. In particular, if the angle of growing point and day average of leaf temperature are managed 6 weeks before harvesting and stem diameters of scion 2 weeks before harvesting, high yields with high sugar content and high quality are expected to be obtained.

Conclusion

This study constructed models for predicting the yield of oriental melon produced from Seongju. We examined what growth and environmental variables affected the crop yield of oriental melon and modeled them using these covariates based on a comparison of the predictability of the models. When checking the performance of models based on three measurements, ARIMAX with PCs showed the best results, which demonstrated effectiveness in utilizing highly correlated covariates, such as growth and environmental variables. Therefore, the ARIMAX with PCs model could be a new alternative for predicting crop yield. In addition, we confirmed that angle of growing point, stem diameters of scion, and day average of leaf temperature had a significant effect on the prediction model of the Seongju oriental melon yield.

The operation of the smart farm system in farms calls for the collection of various growth and environmental variables, which in turn necessitates the construction of a precise model that takes into account these variable characteristics. Therefore, studies such as ours can be expected to contribute to progressive changes in the improvement of crop yield in the future.

Acknowledgments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MIST) (No. 2019R1C1C100315412 and No. 2021R1C1C1009976) and the Research Program of Rural Development Administration (No. PJ0153372021).

REFERENCES

- [1] Bertin, N., Génard, M. (2018): Tomato quality as influenced by preharvest factors. – *Scientia Horticulturae* 233: 264-276.

- [2] Bindi, M., Olesen, J. E. (2011): The responses of agriculture in Europe to climate change. – *Regional Environmental Change* 11(1): 151-158.
- [3] Box, G. E., Jenkins, G. M., Reinsel, G. C., Ljung, G. M. (2015): *Time Series Analysis: Forecasting and Control*. – John Wiley & Sons, Hoboken, NJ.
- [4] Choudhury, A., Jones, J. (2014): Crop yield prediction using time series models. – *Journal of Economics and Economic Education Research* 15(3): 53-68.
- [5] Clark, G. E., Burge, G. K. (2002): Effects of apical growing point size and tuber weight on production in *Sandersonia aurantiaca*. – *Scientia Horticulturae* 94(3-4): 323-332.
- [6] Deschênes, O., Greenstone, M. (2007): The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather. – *American Economic Review* 97(1): 354-385.
- [7] Di Vittori, L., Mazzoni, L., Battino, M., Mezzetti, B. (2018): Pre-harvest factors influencing the quality of berries. – *Scientia Horticulturae* 233: 310-322.
- [8] Galán, C., Vázquez, L., García-Mozo, H., Domínguez, E. (2004): Forecasting olive (*olea europaea*) crop yield based on pollen emission. – *Field Crops Research* 86(1): 43-51.
- [9] Glenn, D. M., Puterka, G. J., Drake, S. R., Unruh, T. R., Knight, A. L., Baherle, P., Prado, E., Baugher, T. A. (2001): Particle film application influences apple leaf physiology, fruit yield, and fruit quality. – *Journal of the American Society for Horticultural Science* 126(2): 175-181.
- [10] Goldschmidt, E. E. (2014): Plant grafting: new mechanisms, evolutionary implications. – *Frontiers in Plant Science* 5: 727.
- [11] Hwang, J. M., Um, J. S., Yi, Y. K. (1999): Survey of physiological disorders in greenhouse fruit vegetables in Kyungbuk province. – *Horticultural Science & Technology* 17(6): 737-741.
- [12] Hendryanti, D. N., Jeong, H., Kim, J. Y., Kwon, O. (2020): Serine protease in a bred variety of oriental melon (*Cucumis melo* L. var. *makuwa*) curtails vascular thrombosis by balancing hemostasis and fibrinolysis in a rodent model. – *Journal of Functional Foods* 68: 103925.
- [13] Hotelling, H. (1933): Analysis of a complex of statistical variables into principal components. – *Journal of Educational Psychology* 24(6): 498-520.
- [14] Kang, S. R., Cho, W. H., Na, M. H., Choi, D. W. (2020): Consideration of environmental factors and growth factors and watering facilities for prediction of pepper production. – *Journal of the Korean Data Analysis Society* 22(1): 177-188.
- [15] Kim, D. G., Ferris, H. (2002): Relationship between crop losses and initial population densities of *Meloidogyne arenaria* in winter-grown oriental melon in Korea. – *Journal of Nematology* 34(1): 43-49.
- [16] King, S. R., Davis, A. R., Zhang, X., Crosby, K. (2010): Genetics, breeding and selection of rootstocks for Solanaceae and Cucurbitaceae. – *Scientia Horticulturae* 127(2): 106-111.
- [17] Lee, K. K., Ko, K. K., Lee, J. W. (2012): Correlation analysis between meteorological factors and crop products. – *Journal of Environmental Science International* 21(4): 461-470.
- [18] Lee, J. E., Kang, S. R., Ok, Y. J., Chun, M. H., Na, M. H. (2019): A study on the optimal environmental factors affecting the growth of facility cucumbers. – *Journal of the Korean Data Analysis Society* 21(6): 2913-2920.
- [19] Lee, H., Hong, K. H., Kwon, D. H., Cho, M. C., Lee, J. G., Hwang, I., Ahn, Y. K. (2020): Changes of growth and yield by using rootstocks in tomato. – *Protected Horticulture and Plant Factory* 29(4): 456-463.
- [20] Liu, S., Liu, Y., Huang, X., Yang, W., Hu, W., Pan, S. (2017): Effect of ultrasonic processing on the changes in activity, aggregation and the secondary and tertiary structure of polyphenol oxidase in oriental sweet melon (*Cucumis melo* var. *makuwa* Makino). – *Journal of the Science of Food and Agriculture* 97: 1326-1334.

- [21] Mngersquo, S. A., Sileshi, G. W., Jamnadass, R., Akinnifesi, F. K., Mhango, J. (2012): Scion and stock diameter size effect on growth and fruit production of *Sclerocarya birrea* (Marula) trees. – *Journal of Horticulture and Forestry* 4(9): 153-160.
- [22] Pesaran, M. H., Shin, Y., Smith, R. J. (2001): Bounds testing approaches to the analysis of level relationships. – *Journal of Applied Econometrics* 16(3): 289-326.
- [23] Shin, B. H., Jeon, H. K. (2020): ICT-based smart farm design. – *Journal of Convergence for Information Technology* 10(2): 15-20.
- [24] Shin, Y. S., Park, S. D., Do Han, W., Bae, S. G., Kim, J. H., Kim, B. S. (2005): Effect of double layer nonwoven fabrics on the growth, quality and yield of oriental melon (*Cucumis melo* L. Var. *makuwa* Mak.) under vinyl house. – *Protected Horticulture and Plant Factory* 14(1): 22-28.
- [25] Shin, A. Y., Kim, Y. M., Koo, N., Lee, S. M., Nahm, S., Kwon, S. Y. (2017): Transcriptome analysis of the oriental melon (*Cucumis melo* L. var. *makuwa*) during fruit development. – *PeerJ* 5: e2834.

APPENDIX

Table A1. Area and yield of vegetable production (fruit) from 2017 to 2019

Item		2017		2018		2019	
		Outdoors	Facility	Outdoors	Facility	Outdoors	Facility
Watermelon	Area (ha)	2,726 (6.50%)	9,935 (23.69%)	2,367 (5.63%)	9,447 (22.45%)	2,648 (6.21%)	9,325 (21.88%)
	Yield (ton)	92,736 (4.90%)	413,735 (21.88%)	76,543 (4.1%)	400,091 (21.3%)	84,469 (4.38%)	391,346 (20.29%)
	Yield/Area	34.02	41.64	32.34	42.35	31.90	41.97
Strawberry	Yield (ton)	124 (0.30%)	5,783 (13.79%)	93 (0.22%)	5,969 (14.19%)	41 (0.10%)	6,421 (15.06%)
	Area (ha)	2,473 (0.13%)	206,226 (10.91%)	1,745 (0.09%)	181,894 (9.67%)	934 (0.05%)	233,291 (12.09%)
	Yield/Area	19.95	35.66	18.85	30.47	22.83	36.33
Tomato	Area (ha)	0 (0.00%)	5,782 (13.78%)	0 (0.00%)	6,058 (14.40%)	0 (0.00%)	5,706 (13.39%)
	Yield (ton)	0 (0.00%)	355,107 (18.78%)	0 (0.00%)	388,657 (20.66%)	0 (0.00%)	358,580 (18.59%)
	Yield/Area	0.00	61.42	0.00	64.15	0.00	62.84
Cucumber	Yield (ton)	1,113 (2.65%)	3,805 (9.07%)	1,160 (2.76%)	4,164 (9.90%)	999 (2.34%)	3,963 (9.30%)
	Area (ha)	44,888 (2.37%)	296,476 (15.68%)	48,849 (2.59%)	342,365 (18.20%)	41,250 (2.13%)	324,815 (16.84%)
	Yield/Area	40.33	77.92	42.12	82.22	41.29	81.96
Oriental melon	Area (ha)	127 (0.30%)	3,454 (8.23%)	145 (0.34%)	3,469 (8.24%)	159 (0.37%)	3,488 (8.18%)
	Yield (ton)	2,298 (0.12%)	163,983 (8.67%)	3,104 (0.16%)	127,424 (6.77%)	3,194 (0.16%)	147,040 (7.62%)
	Yield/Area	18.09	47.48	21.38	36.74	20.07	42.15
Pumpkin	Yield (ton)	6,177 (14.73%)	2,919 (6.96%)	6,299 (14.97%)	2,907 (6.91%)	6,814 (15.99%)	3,060 (7.18%)
	Area (ha)	158,508 (8.38%)	154,182 (8.15%)	172,052 (9.15%)	138,166 (7.34%)	177,795 (9.22%)	165,716 (8.59%)
	Yield/Area	25.66	52.83	27.31	47.53	26.09	54.16

This data is from National Statistical Office of Republic of Korea

Table A2. Average correlation coefficients of farms between lagged covariates and yield of oriental melon per 10a (kg/10a)

Variable		lag0	lag1	lag2	lag3	lag4	lag5	lag6	lag7	lag8
Growth	sd rs	0.470	0.573	0.610	0.490	0.390	0.440	0.533	0.483	0.273
	sd s	0.553	0.630	0.633	0.495	0.380	0.425	0.560	0.528	0.350
	angle	0.158	0.295	0.315	0.185	0.270	0.348	0.450	0.440	0.233
Environment	mean temp	0.433	0.305	0.335	0.223	0.258	0.473	0.593	0.470	0.263
	daymean temp	0.438	0.295	0.310	0.213	0.203	0.400	0.568	0.463	0.260
	DIF temp	0.235	0.133	0.178	0.153	0.115	0.305	0.480	0.370	0.238
	mean stemp	0.365	0.303	0.328	0.175	0.208	0.440	0.563	0.475	0.325
	nightmean stemp	0.373	0.303	0.328	0.195	0.225	0.453	0.580	0.478	0.315
	min stemp	0.385	0.303	0.350	0.210	0.250	0.483	0.563	0.470	0.275
	max stemp	0.325	0.278	0.285	0.135	0.128	0.353	0.518	0.448	0.353
	mean 2h sunrise stemp	0.380	0.300	0.345	0.210	0.250	0.478	0.565	0.473	0.283
	mean ltemp	0.345	0.338	0.378	0.205	0.235	0.473	0.535	0.410	0.283
	daymean ltemp	0.375	0.320	0.323	0.158	0.183	0.415	0.533	0.403	0.268
	nightmean mois	0.108	0.240	0.353	0.203	0.198	0.085	0.178	0.213	0.208
	mean sun	0.455	0.265	0.263	0.455	0.365	0.438	0.444	0.410	0.318
	cumulate sun	0.453	0.265	0.245	0.438	0.363	0.440	0.445	0.410	0.315
	GDD stemp	0.400	0.350	0.260	0.323	0.328	0.375	0.415	0.445	0.295

Table A3. Average correlation coefficients of farms between selected lagged covariates and the yield of oriental melon per 10a (kg/10a), and significance of the correlation coefficient

	Yield per 10a	sd rs_2	sd s_2	angle1_6	mean temp_6	daymean temp_6	DIF temp_6	mean stemp_6	nightmean stemp_6	min stemp_6	max stemp_6	mean 2h sunrise stemp_6	mean ltemp_6	daymean ltemp_6	nightmean mois_2	mean sun_3	cumulate sun_0	GDD stemp_7	
Yield per 10a	1	0.61	0.63	0.45	0.59	0.57	0.48	0.56	0.58	0.56	0.52	0.57	0.54	0.53	0.35	0.46	0.45	0.44	
sd rs_2	*	1	0.94	0.32	0.38	0.45	0.45	0.36	0.38	0.36	0.28	0.36	0.32	0.35	0.27	0.43	0.43	0.38	
sd s_2	*	*	1	0.33	0.47	0.52	0.52	0.41	0.43	0.41	0.34	0.42	0.40	0.42	0.18	0.41	0.43	0.52	
angle1_6	*		*	1	0.67	0.71	0.57	0.71	0.71	0.72	0.65	0.72	0.60	0.65	0.45	0.35	0.27	0.22	
mean temp_6	*	*	*	*	1	0.94	0.54	0.96	0.96	0.95	0.93	0.95	0.96	0.94	0.42	0.25	0.22	0.43	
daymean temp_6	*	*	*	*	*	1	0.78	0.90	0.92	0.87	0.87	0.87	0.87	0.96	0.46	0.26	0.28	0.38	
DIF temp_6	*	*	*	*	*	*	1	0.57	0.58	0.50	0.55	0.51	0.49	0.70	0.36	0.48	0.30	0.30	
mean stemp_6	*	*	*	*	*	*	*	1	1	0.99	0.98	0.99	0.96	0.93	0.37	0.16	0.16	0.40	
nightmean stemp_6	*	*	*	*	*	*	*	*	1	0.98	0.97	0.99	0.96	0.94	0.38	0.16	0.17	0.41	
min stemp_6	*	*	*	*	*	*	*	*	*	1	0.94	1	0.94	0.90	0.38	0.17	0.14	0.39	
max stemp_6	*	*	*	*	*	*	*	*	*	*	1	0.94	0.94	0.92	0.31	0.18	0.15	0.38	
mean 2h sunrise stemp_6	*	*	*	*	*	*	*	*	*	*	*	1	0.95	0.90	0.38	0.18	0.14	0.40	
mean ltemp_6	*	*	*	*	*	*	*	*	*	*	*	*	1	0.92	0.32	0.13	0.20	0.38	
daymean ltemp_6	*	*	*	*	*	*	*	*	*	*	*	*	*	1	0.36	0.20	0.30	0.38	
nightmean mois_2	*			*	*	*	*	*	*	*	*	*	*	*	1	0.48	0.06	0.33	
mean sun_3	*	*	*	*		*	*							*	*	1	0.36	0.19	
cumulate sun_0	*	*	*													*	1	0.33	
GDD stemp_7	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*				1

upper right = average correlation coefficients; lower left = significance of the correlation coefficient; * = $p < .05$; covariate name = covariate name_lag

Table A4. Results of the Ljung-Box Test for prediction models

Model	DLM with covariates	DLM with PCs	ARIMAX(1,0,1) with covariates	ARIMAX(1,0,3) with PCs
Farm1	0.696	0.539	0.759	0.828
Farm2	0.193	0.644	0.177	0.147
Farm3	0.496	0.605	0.194	0.252
Farm4	0.756	0.346	0.455	0.051

DLM = distributed-lag model; PC = principal component; ARIMAX = Auto-regressive integrated moving average with explanatory