

Exploring Environmental Factors Affecting Strawberry Yield Using Pattern Recognition Techniques[☆]

Wanhyun Cho¹ Yuha Park^{1*} Myung Hwan Na¹ Don-Woo Choi²

ABSTRACT

This paper investigates the importance of various environmental factors that have a strong influence on strawberry yields grown in greenhouse using the pattern recognition methods. The environmental factors influencing the production of strawberries were six factors such as average inside temperature, average inside humidity, average CO₂ level, average soil temperature, cumulative solar radiation, and average illumination. The results of analyzing the observed data using Dynamic Time Warping (DTW) showed that the most significant factor influencing the strawberry production was average soil temperature, average inside humidity, and cumulative solar radiation. Second, the results of analyzing the observed data using Multidimensional Scaling (MDS) showed that the most influential factors on the strawberry yields, such as average CO₂ level, average inside humidity, and average illumination were differently given for each farms. However, these results are based on the distance in 3D space and can be deduced from the fact that there is not a large difference between these distances. Therefore, in order to increase the harvest of strawberries cultivated in the farms, it is necessary to manage the environmental factors such as thoroughly controlling the humidity and maintaining the concentration of CO₂ constantly by ventilation of the greenhouse.

✉ keyword : Strawberry yields, Environmental factors, Graphic Analysis Method, Dynamic Time Warping, Multidimensional Scaling Method

1. Introduction

Strawberry is an important fruit crop in the Korea, and is a highly nutritious and very popular food source. In addition, strawberry is a fruit that can be utilized in various aspects such as cake making, dessert menu and juice making, so it is very high value added fruit. This trend is increasing the demand for strawberry. Therefore, strawberry cultivation has become a value-added business for farmers, and more attention has been paid to high-quality, high-yielding cultivation techniques.

For this reason, farmers are therefore very interested in developing smart farming technology that can improve strawberry yields by combining agriculture and internet of things (IoT) technology. There are a lot of researches using internet of things (IoT) technology in related with agriculture and other area. Chung, et al. [1] proposed a method for automatically checking cattle mounting using a side-view video camera and applied computer vision techniques to detect mounting behavior. Jalal, et al. [2] proposed a real-time activity recognition (Real-AR) system to identify the daily human activity routines and to make these surroundings an intelligent living space using real-time smart activity monitoring system.

Utilizing the IoT technology, we measure the growth of crops and environment information from various observation sensors in real time. In addition, by using the extracted information, an optimal growth management system is constructed, technologies that can increase the productivity and the quality of the crops by automatically managing the crops are developed dramatically.

Previous studies related to these technologies until now are given as follows. Using a generalized randomized block

¹ Dept. of Statistics, Chonnam National University, Gwanju, 61186, Korea

² Gyeongsangbuk-do Agricultural & Extension Services, 47, Chilgokjungangdaero 136gil, Buk-gu, Daegu, Korea

* Corresponding author (yuhapark0@gmail.com)

[Received 29 August 2018, Reviewed 19 September 2018(13 November 2018), Accepted 24 December 2018]

[☆] Acknowledgement: This work was partially supported by the Research Program of Rural Development Administration (Project No. PJ01386702), and the Korea National Research Foundation (Project No. 2017R1D1A1B03028808) of Korea Grant funded by the Korean Government.

[☆] A preliminary version of this paper was presented at ICONI 2017 seminar and was selected as an outstanding paper.

design, Estrada-Ortiz, et al. [3] evaluated the effect of different percentages of phosphite added to the nutrient solution on the concentration of total P in leaves and the activation of the antioxidant system, which determines the concentration of anthocyanin, yield, pH, electrical conductivity, and strawberry fruit size. They suggest that supplying 20% phosphite in the nutrient solution improved strawberry fruit performance and that supplying 30% phosphite activated dense mechanisms in the plants, which increased the concentration of anthocyanins and improved fruit quality.

Letourneau, et al. [4] had performed a field-scale experiment to simultaneously evaluate the impacts of three irrigation management scales and a pulsed water application method on strawberry yield and water use efficiency. The results of their experiment showed that spatial variability of the soil properties at the experimental site was important but most likely not enough to influence the crop response to irrigation practices.

Boyer, et al. [5] carried out studies to investigate whether arbuscular mycorrhizal fungi (AMF) could improve strawberry production in coir under low nitrogen input and regulated deficit irrigation. Application of AMF led to an appreciable increase in the size and number of class 1 fruit, especially under either deficient irrigation or low nitrogen input condition.

Fan, et al. [6] had evaluated the effects of plastic mulch (PM) and plastic mulch with row covers (PMRC) versus the conventional MRS, on total yield, yield per plant, average fruit weight, soluble solids content, titratable acidity, firmness, fruit postharvest quality, total phenolic content, total antioxidant content, oxygen radical absorbance capacity and phenolic composition analyzed by high-performance liquid chromatography in strawberry selection 'SJ8976-1' at different harvest times during the growing season.

Kurokura, et al. [7] had examined the responses of diploid wild strawberry and garden strawberry varieties to plant growth promoting rhizobacteria (PGPR) application. Application of PGPR during the flower induction increased the total yield and number of fruits in wild strawberry and garden strawberry. But sucrose content was altered in cultivar dependent manner.

Pham, et al. [8] proposed a Query-by-Singing/ Humming

system which is obtained a higher matching accuracy using MP3 files on mobile device using combining dynamic time warping matching algorithm and chroma-based DTW algorithm.

Therefore, based on previous studies, we proposed methods to identify the various environmental factors affecting strawberry production using two pattern recognition methods. First, we are roughly examined the relationship between production of strawberry and environmental factors through various graphs. Second, we use the dynamic time warping method (DTW) and the multidimensional scaling method (MDS) which are typically used to recognize the patterns in time series data to determine the interrelation of strawberry yield and environmental factors. Finally, based on the results of the analysis, we propose a new cultivation method to increase the strawberry production.

2. Dataset and Methods

2.1 Dataset

The data used in this study were based on data observed from three farms grown in the Gyeongbuk area in South Korea. Table 1 shows the list of variable names and measurement unit.

(Table 1) The list of variable names and measurement unit.

Variables	Name	Measure
Response variable		
3-day cumulative yield	y	kg/m ²
Environmental variables		
Average CO ₂ level	x_1	ppm
Average inside temperature	x_2	°C
Average inside humidity	x_3	%
Average illumination	x_4	lx
Cumulative solar radiation	x_5	mW/cm ²
Average soil temperature	x_6	°C

The observed data consist largely of both the production of strawberries and the measurements of six environmental factors, respectively, given at the three farmhouses. These six

environmental factors are 3-day average inside temperature, 3-day average inside humidity, 3-day average CO₂ level, 3-day average soil temperature, 3-day cumulative solar radiation, and 3-day average illumination.

2.2 Method

In this study, we used Dynamic Time Warping (DTW) and Multidimensional Scaling (MDS) as pattern recognition methods that can be used to correlate time series data.

First, DTW [6] has earned its popularity by being extremely efficient as the time-series similarity measure which minimizes the effects of shifting and distortion in time by allowing elastic transformation of time series in order to detect similar shapes with different phases. Given two time series $X = (x_1, \dots, x_N)$, $M \in \mathbb{N}$ and $Y = (y_1, \dots, y_N)$, $N \in \mathbb{N}$ respectively by the sequences of values DTW yields optimal solution in the $O(MN)$ time which could be improved further through different techniques such as multi-scaling. The only restriction placed on the data sequences is that they should be sampled at equidistant points in time.

If sequences are taking values from some feature space than in order to compare two different sequences $X, Y \in \Phi$, one needs to use the local distance measure which is defined to be a function:

$$d: \Phi \times \Phi \rightarrow \mathbb{R} \geq 0. \quad (1)$$

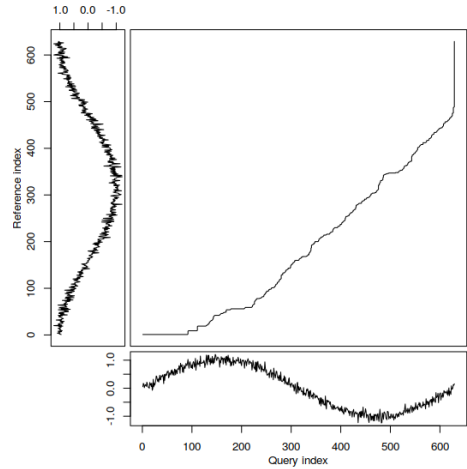
Intuitively d has a small value when sequences are similar and large value if they are different. Since the Dynamic Programming algorithm lies in the core of DTW it is common to call this distance function as the cost function and the task of optimal alignment of the sequences becoming the task of arranging all sequence points by minimizing the cost function.

Algorithm starts by building the distance matrix $C \in \mathbb{R}^{M \times N}$ representing all pairwise distances between X and Y . This distance matrix called the local cost matrix for the alignment of two sequences X and Y :

$$C \in \mathbb{R}^{M \times N}: d_{ij} = \|x_i - y_j\|, i \in [1: M], j \in [1: N]. \quad (2)$$

Once the local cost matrix built, the algorithm finds the alignment path which runs through the low-cost areas -

valleys on the cost matrix. This alignment path (or warping path, or warping function) defines the correspondence of an element $x_i \in X$ to $y_j \in Y$ following the boundary condition which assign first and last elements of X and Y to each other. Figure 1 shows an example of one optimal warping path alignment time series.



(Figure 1) The optimal warping path alignment time series

Second, MDS [7] can be considered to be an alternative to factor analysis. In general, the goal of the analysis is to detect meaningful underlying dimensions that allow the researcher to explain observed similarities or dissimilarities (distances) between the investigated objects. In factor analysis, the similarities between objects are expressed in the correlation matrix. With MDS, we can analyze any kind of similarity or dissimilarity matrix, in addition to correlation matrices.

In general, MDS attempts to arrange objects in a space with a particular number of dimensions (two-dimension or three - dimension) so as to reproduce the observed distances. As a result, we can explain the distances between objects in terms of underlying dimensions. MDS is a way to rearrange objects in an efficient manner, so as to arrive at a configuration that best approximates the observed distances. It actually moves objects around in the space defined by the requested number of dimensions, and checks how well the distances between objects can be reproduced by the new

configuration.

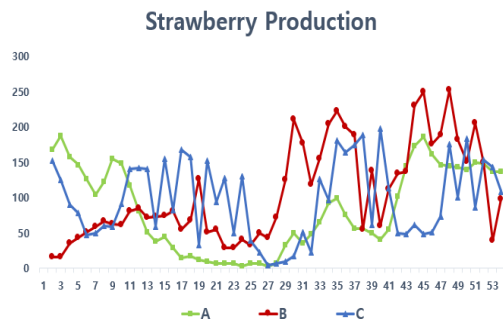
In more technical terms, it uses a function minimization algorithm that evaluates different configurations with the goal of maximizing the goodness-of-fit. The most common measure that is used to evaluate how well (or poorly) a particular configuration reproduces the observed distance matrix is the stress measure. The raw stress value PHI of a configuration is defined by:

$$\Phi = \sum (d_{ij} - f(\delta_{ij}))^2 \quad (3)$$

In this formula, d_{ij} stands for the reproduced distances, given the respective number of dimensions, and δ_{ij} stands for the input data (i.e. observed distances). The expression $f(\delta_{ij})$ indicates a nonmetric, monotone transformation of the observed input data (distances). Thus, it will attempt to reproduce the general rank-ordering of distances between the objects in the analysis. Here, the smaller the stress value, the better is the fit of the reproduced distance matrix to the observed distance matrix.

3. Experimental Results

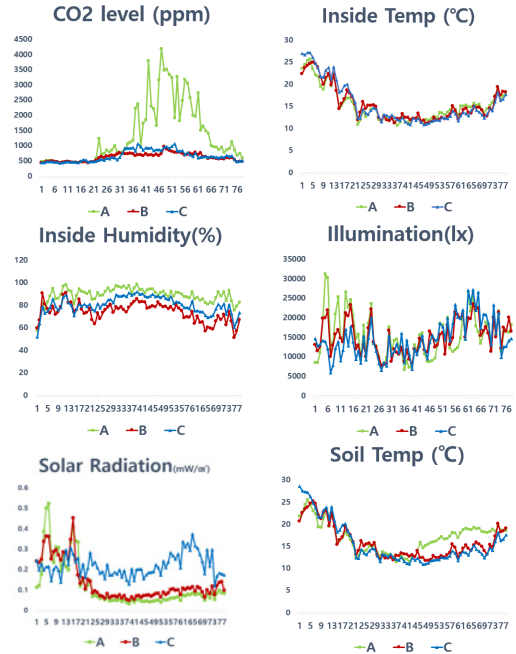
First, Figure 2 and Figure 3 below are line graphs to roughly determine how the strawberry yields produced in the three farms A, B, and C are affected by the six environmental variables. From Figure 2 we can see that the yields of farms B and C are similar, while the yield of farm A is different from that of the other two farms.



(Figure 2) Strawberry production in three farms

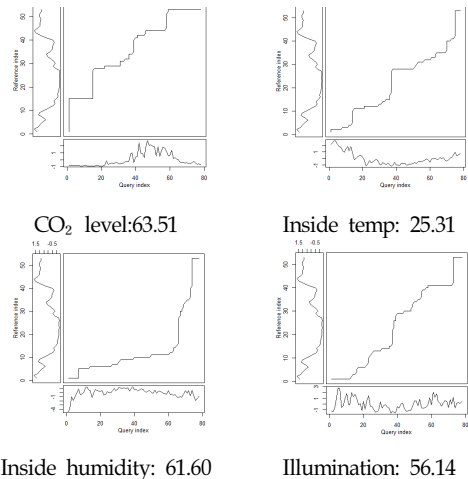
Also, from Figure 3, we can see that the level of CO₂ in farm B is very different from that in other farms, and that

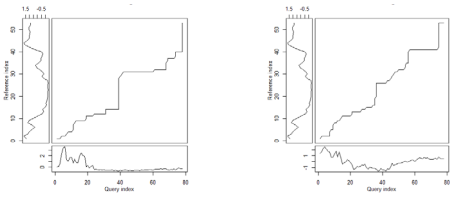
farm A has a remarkable difference from two farms with cumulative solar radiance, inside humidity.



(Figure 3) Values of six environmental variables

Second, the results of applying the DTW to obtain the relationship between the yields of strawberries produced and the six environmental variables in farm A, B and C are given as in Figure 4, 5 and 6.



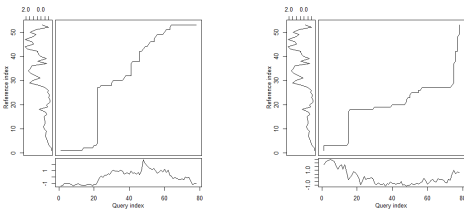


Solar radiation: 47.86

Soil temp: 24.83

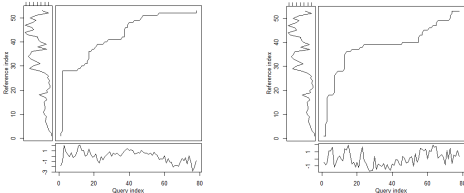
(Figure 4) Effect of environmental variables on farm A

From the results in Figure 4, the most relevant environmental variables of farm A were average soil temperature, average inside temperature, cumulative solar radiation, average illumination amount, average inside humidity, and CO₂ level.



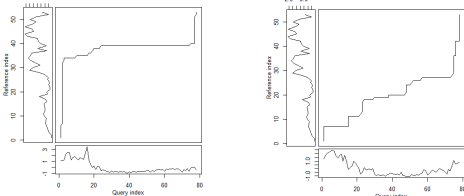
CO₂ level: 33.67

Inside temp: 81.47



Inside humidity: 44.75

Illumination: 49.44



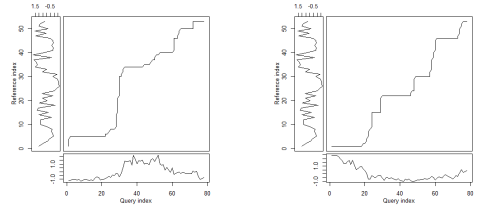
Solar radiation: 88.71

Soil temp: 81.24

(Figure 5) Effect of environmental variables on farm B

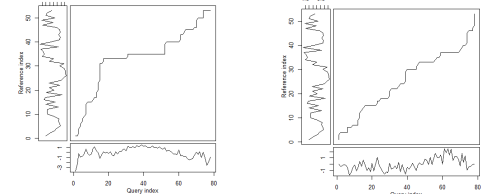
From the results in Figure 5, the most relevant environmental variables of yields of farm B were average CO₂ level, average inside humidity, average illumination, average inside temperature, average soil temperature, and cumulative solar radiation.

average inside temperature, average soil temperature, and cumulative solar radiation.



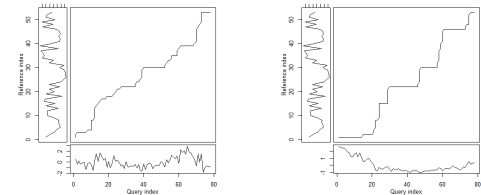
CO₂ level: 54.92

Inside temp: 55.11



Inside humidity: 55.98

Illumination: 48.18



Solar radiation: 47.95

Soil temp: 55.49

(Figure 6) Effect of environmental variables on farm C

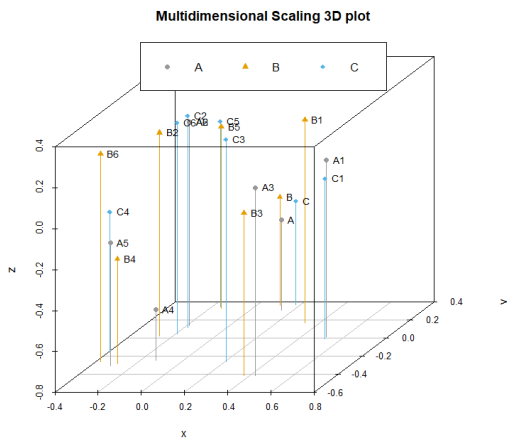
From the results in Figure 6, the most relevant environmental variables of yields of farm C were cumulative solar radiation, average illumination, average CO₂ level, average inside temperature, average soil temperature, and average inside humidity.

Therefore, when we compare the six environmental variables that affect the yield of the three farm, average inside humidity is the most important environmental variable, while the least influential environmental variable is average illumination.

Third, the results of applying the MDS to determine the influence of each of the strawberry yields produced by the three farms A, B, and C on the six environmental variables were given as in Figure 7, and the distance between them was calculated as shown in Table 2, 3 and 4.

First, from Figure 7 of the three-dimensional plot of the

MDS, we can see that there is a difference between the yields of farm A, B, and C. This result is consistent with the result in Figure 2. And we can find that there is a difference between average CO₂ level of farm B, and average CO₂ level of farm A and farm C. This result is consistent with the previous result of Figure 3. This is a similar result that the cumulative solar radiation of farm A is different from the cumulative solar radiation of farm B and farm C, which is consistent with the previous Figure 3. Finally, there is little difference between the three farms for every environmental factor, except for CO₂ level and cumulative solar radiation. The result is consistent with the previous Figure 3.



(Figure 7) 3D plot by MDS of production and environmental variables of three farms

Second, the environmental variables affecting the yield of farm A from the distance matrix by the multidimensional scaling method are average CO₂ level (A1), average illumination (A4), mean inside humidity (A3), cumulative solar radiation (A5), average inside temperature (A2) and average soil temperature (A6). The results are generally consistent with the previous results of the dynamic time warping method, but the difference is that the average illumination is a priority.

Third, the environmental variables affecting the yield of farm B from the distance matrix by the MDS are average inside humidity (B3), average soil temperature (B6), average illumination (B4), average CO₂ level (B1) and cumulative solar radiation (B5). In particular, it was found that average

inside humidity had the greatest effect on the yield and average CO₂ level caused the biggest change in yield.

(Table 2) The distance matrix by MDS of the yield and environmental variables of farm A

	A	A1	A2	A3	A4	A5	A6
A	0.000	5.965	8.957	7.173	6.757	7.726	8.953
A1	5.965	0.000	7.459	3.394	6.940	6.939	7.464
A2	8.957	7.459	0.000	5.209	6.082	4.943	0.163
A3	7.173	3.394	5.209	0.000	6.297	5.617	5.248
A4	6.757	6.940	6.082	6.297	0.000	2.119	6.064
A5	7.726	6.939	4.943	5.617	2.119	0.000	4.925
A6	8.953	7.464	0.163	5.248	6.064	4.925	0.000

(Table 3) The distance matrix by MDS of the yield and environmental variables of farm B

	B	B1	B2	B3	B4	B5	B6
B	0.000	8.679	8.173	7.266	7.891	9.926	7.768
B1	8.679	0.000	8.451	6.361	7.795	9.851	7.662
B2	8.173	8.451	0.000	4.638	4.855	4.928	1.694
B3	7.266	6.361	4.638	0.000	5.685	7.501	3.979
B4	7.891	7.795	4.855	5.685	0.000	5.856	4.379
B5	9.926	9.851	4.928	7.501	5.856	0.000	5.753
B6	7.768	7.662	1.694	3.979	4.379	5.753	0.000

Fourth, from the distance matrix by the MDS, the environmental variables affecting the yield of farm C are average illumination (C4), average CO₂ level (C1), average soil temperature (C6), cumulative solar radiation (C5), and average inside humidity (C3). This result is not generally consistent with the previous results of the dynamic time warping method, but we think it might be due to the fact that the distances between the yield and the environmental variables by the MDS do not differ greatly.

To summarize the results obtained so far, the environmental variables that have the greatest influence on the strawberry yields grown in the greenhouse are average inside humidity, average inside temperature and cumulative solar radiation. In particular, the measured value of average CO₂ level greatly influenced the increase and decrease of yield. Therefore, in order to improve the strawberry yield of the greenhouse, it is necessary to thoroughly manage the

humidity and to keep the CO₂ concentration constant by ventilation of the greenhouse from time to time.

(Table 4) The distance matrix by MDS of the yield and environmental variables of farm C

	C	C1	C2	C3	C4	C5	C6
C	0.00	7.75	9.77	10.19	7.63	10.02	9.76
C1	7.75	0.00	11.61	7.09	10.47	11.13	11.65
C2	9.77	11.61	0.00	9.13	6.06	4.29	0.59
C3	10.11	7.09	9.13	0.00	10.61	7.74	9.11
C4	7.63	10.47	6.06	10.61	0.00	5.92	5.96
C5	10.02	11.13	4.29	7.74	5.92	0.00	4.11
C6	9.77	11.65	0.59	9.11	5.96	4.11	0.00

4. Conclusion

In this paper, we analyzed the effects of six environmental variables affecting strawberry yields using real data collected from farms. From analyzing the observed data from the IoT sensors, it is helpful for farmers, researcher, and other the relatives to find the most influential environmental factors in order to improve the strawberry yield grown in the greenhouse. Since we used the pattern recognition technique in the paper, it is possible to find the influential environmental factors to the strawberry yield solving the limitations and problems what the model like the regression model nor the time-series model have.

Among the various factors, we select the most important factors based on the three farms, and from the results, we present cultivation techniques. So it can be called a data-driven cultivation technique. The facts from the experimental results are as follows.

First, the environmental variables that have the greatest influence on the strawberry yield are average inside humidity, average soil temperature, and cumulative solar radiation. Second, average CO₂ level greatly influence the increase and decrease of yield. Third, from the experimental results by comparing the strawberry yields in three farms, it is proper environmental variable management methods to maintain the concentration of CO₂ constantly by ventilating greenhouse thoroughly, and to thoroughly control the humidity in order to improve the strawberry yield of the greenhouse.

It is the limitation of our study that dataset used in the study only included the environmental factors after the transplanting date. Also, the collected various factors are limited. The environment is very important factors to strawberry yield in the raising seeding stage. Therefore, in further study, we collect dataset including the environmental factors and growth factors in the raising seeding stage and analyze the relationships between those factors.

참고문헌(Reference)

- [1] Y. Chung, D. Choi, H. Choi, D. Park, H.-H. Chang and S. Kim, "Automated Detection of Cattle Mounting using Side-View Camera," *KSII Transactions on Internet and Information Systems*, vol. 9, no. 8, pp. 3151-3168, 2015.
[https://doi.org/ 10.3837/tiis.2015.08.024](https://doi.org/10.3837/tiis.2015.08.024).
- [2] A. Jalal, S. K. and D.-S. Kim, "Detecting Complex 3D Human Motions with Body Model Low-Rank Representation for Real-Time Smart Activity Monitoring System," *KSII Transactions on Internet and Information Systems*, vol. 12, no. 3, pp. 1189-1204, 2018.
<https://doi.org/10.3837/tiis.2018.03.012>.
- [3] E. Estrada-Ortiz, L. I. Trejo-Tellez, F. C. Gomez-Merino, R. Nunez-Escobar and M. Sandoval-Villa, "The effects of phosphite on strawberry yield and fruit quality," *Journal of Soil Science and Plant Nutrition*, Vol. 13, No. 3, 2013, pp. 612-620.
- [4] G. Letourneau, J. Caron, L. Anderson, and J. Cormier, "Matric potential-based irrigation management of field-grown strawberry: effects on yield and water use efficiency", *Agricultural Water management*, Vol.161, 2015, pp. 102-113.
- [5] L. R. Boyer, W. Feng, N. Gulbis, K. Hajdu, R. J. Harrison, P. Jeffries, and X. Xu, "The use of Arbuscular Mycorrhizal Fungi to improve strawberry production in coir substrate", *Frontiers in Plant Science*, Vol. 7, No. 1237, pp. 1-9. 2016.
<https://doi.org/10.3389/fpls.2016.01237>.
- [6] L. Fan, C. Dube, and S. Khanizadeh, "Chapter 7: The effect of production systems on strawberry quality", *INTECH*. 2017, pp. 197-213.

- <http://dx.doi.org/10.5772/67233>.
- [7] T. Kurokura, S. Hiraide, Y. Shimamura, and K. Yamane, "PGPR improves yield of strawberry species under less-fertilized conditions", *Environ. Control Biol.*, Vol. 55, No. 3, pp. 121-128, 2017. <https://doi.org/10.2525/ecb.54.121>.
- [8] T. D. Pham, G. P. Nam, K. Y. Shin and K. R. Park, "A Novel Query-by-Singing/Humming Method by Estimating Matching Positions Based on Multi-layered Perceptron," *KSII Transactions on Internet and Information Systems*, vol. 7, no. 7, pp. 1657-1670, 2013. <https://doi.org/10.3837/tiis.2013.07.008>.
- [9] P. Senin, "Dynamic Time Warping Algorithm Review", Technical Report (Information and Computer Science Department, University of Hawaii at Manoa, USA), 2008.
- [10] T. Hill and P. Lewicki, "Statistics: Methods and Applications", StatSoft (Tulsa, Okla.), 2006.

● 저 자 소 개 ●



조 완 현(Wanhyun Cho)

1977년 전남대학교 수학과(이학사)
1981년 전남대학교 대학원 수학과(이학석사)
1988년 고려대학교 대학원 통계학과(이학박사)
1994년~현재 전남대학교 통계학과 교수
관심분야 : statistical modeling, pattern recognition, image processing and machine learning.
E-mail : whcho@chonnam.ac.kr



박 유 하(Yuha Park)

2015년 전남대학교 통계학과(이학사)
2017년 전남대학교 대학원 통계학과(이학석사)
2017년~현재 전남대학교 농업빅데이터연구실 연구원
관심분야 : statistical modelling, machine learning and statistical learning for big data.
E-mail : yuhapark0@gmail.com



나 명 환(Myung Hwan Na)

1983년 서울대학교 수학교육학과(이학사)
1990년 서울대학교 대학원 통계학과(이학석사)
1993년 서울대학교 대학원 통계학과(이학박사)
2003년~현재 전남대학교 통계학과 교수
관심분야 : SQC, machine learning, big-data analysis.
E-mail : nmh@chonnam.ac.kr



최 돈 우(Don-Woo Choi)

1995년 경북대학교 농업경제학과(경제학학사)
1998년 경북대학교 대학원 농업경제학과(경제학석사)
2014년 경북대학교 대학원 농업경제학과(경제학박사)
1996년~현재 경상북도농업기술원 농업연조사
관심분야 : 농업경영, 농업회계, 빅데이터.
E-mail : gogemaroo@korea.kr