Original Paper

Environ. Control Biol., **59** (3), 107-115, 2021
DOI: 10.2525/ecb.59.107

Averaging Techniques in Processing the High Time-resolution Photosynthesis Data of Cherry Tomato Plants for Model Development

Yayu ROMDHONAH[1,2], Naomichi FUJIUCHI[3], Kota SHIMOMOTO[4],
Noriko TAKAHASHI[3], Hiroshige NISHINA[3], Kotaro TAKAYAMA[3,5]

[1] *The United Graduate School of Agricultural Sciences, Ehime University, 3−5−7, Tarumi, Matsuyama, Ehime 790−8566, Japan*
[2] *Faculty of Agriculture, Universitas Sultan Ageng Tirtayasa, Jl. Raya Jakarta km.04 Pakupatan, Serang, Banten 42121, Indonesia*
[3] *Graduate School of Agriculture, Ehime University, 3−5−7, Tarumi, Matsuyama, Ehime 790−8566, Japan*
[4] *Institute of Agricultural Machinery, National Agriculture and Food Research Organization, 1−40−2, Nisshin, Kita, Saitama, Saitama 331−8537, Japan*
[5] *Toyohashi University of Technology, Electronics-Inspired Interdisciplinary Research Institute (EIIRIS), 1−1 Hibarigaoka, Tempaku-Cho, Toyohashi, Aichi 441−8580, Japan*

(Received June 11, 2020; Accepted August 8, 2020)

We evaluated averaging techniques in data processing for the estimation of canopy net photosynthetic rates (P_n) of two cherry tomato plants using a multiple linear regression analysis with variables of aerial environmental factors. Whole canopy P_n and the environmental factors were measured in a high time resolution with a 5-minute interval under a commercial greenhouse by using a novel photosynthesis chamber. We processed the data by using a moving average (MA) and simple average (SA) with several time frames (30-minute, 1-hour, 2-hour). The canopy P_n was expressed as a general linear function of PAR irradiance (*I*), air temperature (*T*), relative humidity (*RH*), vapor pressure deficit (*VPD*), and CO_2 concentration (*C*). Model accuracy generally increased with longer time frames; however, it can be varied depending on the datasets and the variables used in the models. The 2-hour-SA datasets gave the best accuracy for both 5-variable model (*I*, *T*, *RH*, *VPD*, *C*) and 3-variable model (*I*, *VPD*, *C*) with R^2 of 0.81 and 0.67, respectively. This study indicates that datasets of 2-hour time frame with simple average are promising to make a practical general linear regression model for the estimation of P_n of cherry tomato by using the high time-resolution P_n data.

Keywords : moving average, multiple regression, open-bottom chamber, photosynthesis model, time frame

INTRODUCTION

Sensor-based plant diagnostic technology is an essential feature of the speaking plant approach (Hashimoto, 1989; Takayama, 2013). A recent plant diagnosis technology, which has been started to be installed in practical agricultural production in greenhouses, is the real-time photosynthesis and transpiration monitoring system (Shimomoto et al., 2020). The system allows to remotely and continuously measure photosynthesis of whole plants under greenhouse conditions without any contact or intrusive. Also, data are sampled with high time-resolution, which are recorded in 5-minute intervals. With these features, the system is desirable for precise quantification of plant responses to stimuli at the whole-plant level.

The numerous photosynthesis data produced by the photosynthesis monitoring system can be used for analysis and forecasting by way of modeling. Estimation models for the net photosynthetic rate (P_n) as a function of environmental factors for greenhouse tomato utilizing numerous data were limited to the whole greenhouse, such as the work of Nederhoff and Vegter (1994a). They used three variables of incident photosynthetically active radiation (PAR), CO_2 concentration, with double rectangular hyperbolic relation and leaf area index (LAI) to calculate the net canopy photosynthesis rate of tomato greenhouse with an R^2 of 0.892. In another work, they modified two established mechanical models of Acock (Acock et al., 1978) and Thornley (Thornley, 1976) to fit their data and gave R^2s of 0.893 and 0.817, respectively (Nederhoff and Vegter, 1994b).

On the other hand, the accuracy of sensor-based technology in modern greenhouses is jeopardized by disturbances, such as inaccuracy of the measurements by the sensor itself due to dynamic variations in the greenhouse climate (van Mourik et al., 2019), especially when measurements are performed in high time resolution. Therefore, it leads to an issue of how to address the high time-resolution data produced by the sensors for further use of model development. As most real-time data contain erroneous values and noise, such data required further processing for filtering noise and smoothing. Averaging techniques commonly used for smoothing data include the moving average (Čampulová, 2018) and the simple average (Yaffee and McGee, 2000). When the data are correctly prepared, the quality of the model can be reliable (Pyle, 1999).

The objective of the present study was to process the

Corresponding author : Kotaro Takayama, fax: ＋81−89−946−9916,
e-mail : takayama@tut.jp

high time-resolution photosynthesis data generated by the photosynthesis monitoring system for further use of model development. The focus intention in this study was on 1) preparing datasets for the estimation model of whole canopy net photosynthetic rate with averaging techniques, 2) determining an appropriate time frame decomposition for the dataset. We explored two averaging techniques, i.e., the moving average and simple average (Yaffee and McGee, 2000) with time frames of 30 minutes, 1 hour, 2 hours for the model development. The averaging techniques were then compared based on the accuracy in estimating net photosynthetic rates of cherry tomato plants by using a multiple linear regression analysis with variables of aerial environmental factors.

MATERIALS AND METHODS

Plant materials and the experimental greenhouse

The plant materials were two mature cherry tomato plants (*S. lycopersicum* var. *cerasiforme* cv. Scarlet) with 15 leaves planted on 24th August in 2017. The plants were grown hydroponically and maintained as a commercial crop with a plant density of 3.6 plants m^{-2} under a venlo-type semi-commercial greenhouse (65 m [W]×32 m [D]×6 m [H]) in Mie Prefecture, Japan. A commercial microcomputer (HortiMaX; Hortisystem Co., UK) controlled the greenhouse climate, irrigation and fertigation at a commercially optimum condition. CO_2 was supplied to maintain a set point of 500 μmol mol^{-1} per day. The average values of instantaneous PAR above the canopy, air temperature, relative humidity, CO_2 concentration, and vapor pressure deficit during the day time of study period were 175.1 W m^{-2}, 21.2℃, 78%, 424 μmol mol^{-1}, and 0.5 kPa respectively .

Measurements of net photosynthetic rate and environmental factors

Whole plant-level net photosynthetic rate (P_n, μmol s^{-1} per chamber) of two mature cherry tomato plants were measured from 7th to 19th May in 2018 by using the aforementioned photosynthetic chamber system developed in our previous study (Shimomoto et al., 2020). The system was composed of a transparent open-bottom chamber, which is similar to that of Takayama et al. (2012), and a sensing unit to measure CO_2 concentration of the inflow and outflow air of the chamber and relevant environmental factors.

The transparent open-bottom chamber (Vinyl sheet; SUS Co., Ltd., Japan) was 1 m [W]×0.5 m [D]×2.5 m [H], hung on a steel wire for hanging the tomato plants, and enclosed two adjacent plants on the same cultivation gutter (Fig. 1). The sensing unit (HaPPiMinder®; Shikoku Research Institute Inc., Takamatsu, Japan) recorded the CO_2 concentrations of the inflow and outflow air of the chamber and the environmental factors such as PAR at the top of the canopy (I, W m^{-2}), air temperature (T, ℃), relative humidity (RH, %), and the CO_2 concentrations of greenhouse air (C, μmol mol^{-1}), which is similar to that of inflow air of the chamber, at an interval of 5 minutes. The P_n was calculated by multiplying the difference in the CO_2 concentrations of inflow and outflow air and the airflow rate of the ventilation fan fixed at the top of the chamber. And, vapor pressure deficit (VPD, kPa) was also calculated with the recorded T and RH.

Data cleansing and dataset preparation

For the model development and validation, we used daytime data from 6:00 to 18:00 because the model focuses on to simulate the daytime photosynthesis, not dark respiration. The original dataset consisted of 1,912

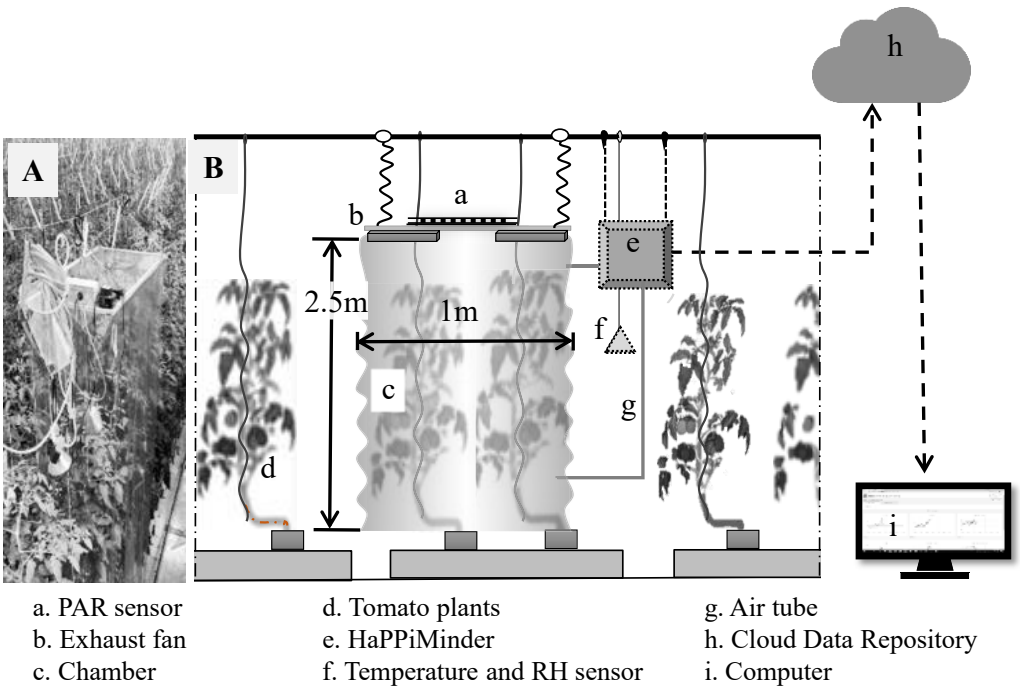

Fig. 1 Photo (A) and schematic diagram (B) of measurements of whole plant canopy photosynthesis and environmental factors in commercial greenhouse by using the photosynthesis chamber system.

data records. All data were checked to find out some error data by plotting the time course according to Takahashi et al. (2008). Some calculated P_n seem obviously incorrect, which might be caused by the fluctuating CO_2 concentration of greenhouse air or some human activities such as pruning and harvesting. Hence, such incorrect data were discarded manually. Through this data cleansing, 11.7% of the recorded data were deleted.

We then prepared some datasets by moving average (MA) and simple average (SA) (Yaffee and McGee, 2000) with the time spans of 30 minutes, 1 hour (1-hour), and 2 hours (2-hour). It meant that the datasets of MA had the time resolution at an interval of 5 minutes, which is similar to the original dataset, and the datasets of SA, they have different time resolutions for each averaging time span; 30-minute-SA is at an interval of 30 minutes, 1-hour-SA is at an interval of 1 hour, and 2-hour-SA is at an interval of 2 hours. Thus, we had sets of training and test data containing P_n, I, T, RH, VPD, and C of 30-minute-MA, 1-hour-MA, 2-hour-MA, 30-minute-SA, 1-hour-SA, and 2-hour-SA.

Linear regression model of net photosynthesis

By using the five environmental factors of I, T, RH, VPD, and C (5-var-model) as the predictor variables, P_n is characterized by the following linear equation:

$$P_n = \beta_0+\beta_1\times I+\beta_2\times T+\beta_3\times RH+\beta_4\times VPD+\beta_5\times C \quad (1)$$

where β_0 is a constant, and β_{1-5} are regression coefficients. We used two indices for humidity (i.e., RH and VPD) to investigate which variable contributed more to the P_n. Moreover, we also explored the modeling by using three essential predictor variables of I, VPD, and C (3-var-model) because multicollinearity problem can be expected between RH and VPD, as well as I and T. So, a further simple linear regression model of Eq. 2 with the predictor variables of I, VPD, and C was also investigated.

$$P_n = \gamma_0+\gamma_1\times I+\gamma_2\times VPD+\gamma_3\times C \quad (2)$$

where γ_0 is a constant, and γ_{1-3} are regression coefficients.

Multiple regression (MR) analysis in the SPSS software program (IBM SPSS Statistics version 26) was used to establish the model. We trained the model by using the data from 7th to 14th May 2018 to get the regression coefficients. Thus, we performed a test on residuals, goodness of fit (R^2), and multicollinearity. Last, we validated the model by using the residual data from 15th to 19th May 2018, and calculated the R^2 and root mean squared error (RMSE) by comparing the measured and estimated P_n. Figure 2 summarizes the representative steps of the model development.

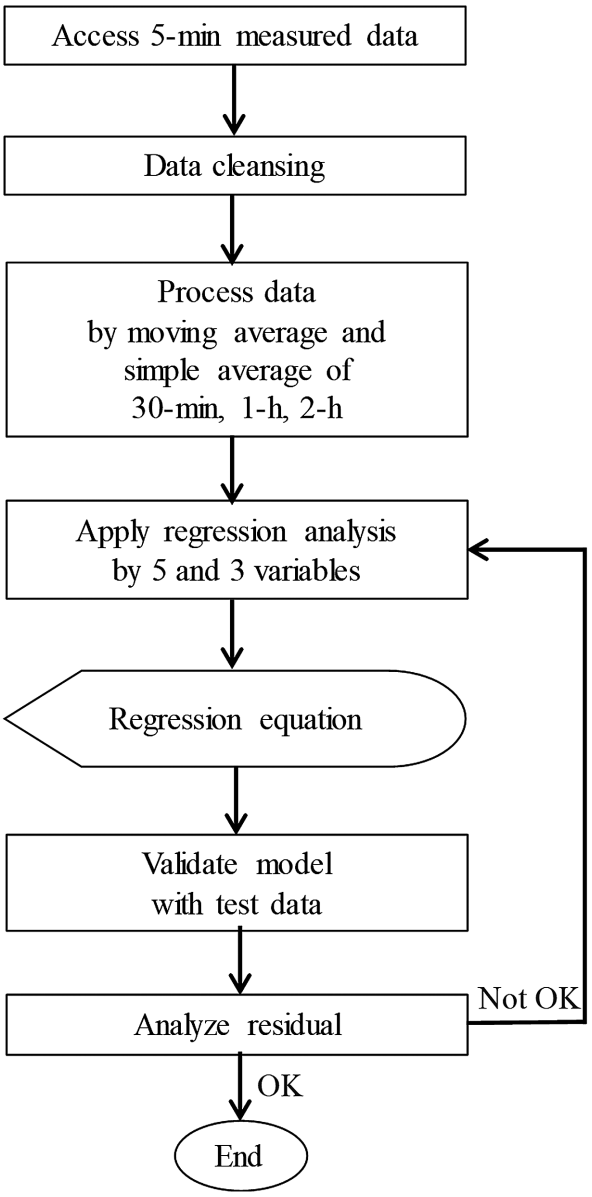

Fig. 2 Flowchart of the model development. Measured data were composed of daytime whole canopy P_n of two mature cherry tomato plants grown hydroponically and the corresponding I, T, RH, VPD, and C from 7th to 19th May 2018. The measured data at an interval of 5 minutes were processed by 30-minute, 1-hour, 2-hour moving average or simple average for the regression model. A general linear model was applied to the training data, and the results were validated with test data.

RESULTS AND DISCUSSION

Figure 3 shows the time course of P_n of two mature cherry tomato plants enclosed by the chamber and the environmental factors during consecutive 13 days after data cleansing. The time course of P_n shows a similar pattern to that of I, which suggests that I dominantly drove and limited photosynthesis activity on a daily basis. On rainy days like the 7th and 13th May, the maximum I were 95.1 W m^{-2} and 56.8 W m^{-2}, respectively. The low I was accompanied with low T and high RH. Even though VPD was low and C was above 500 μmol mol^{-1}, the P_n only reached a maximum of 10.7 μmol s^{-1} per chamber and 11.7 μmol s^{-1} per chamber on the 7th and 13th May, respectively. As for sunny days, high I increased the greenhouse air temperature and promoted the automatic ventilation. Together with CO_2 enrichment, the P_n on the sunny days were able to

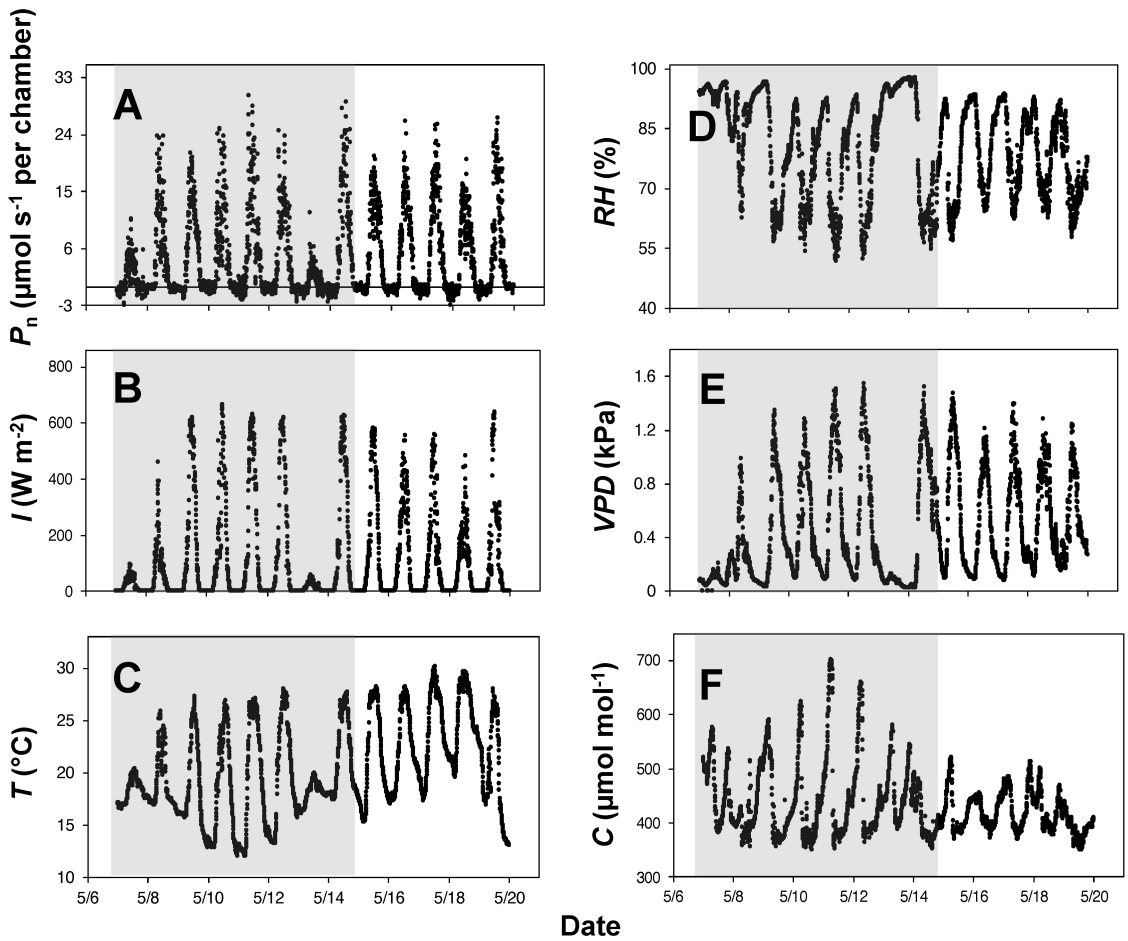

Fig. 3 Whole canopy P_n (A) of two mature cherry tomato plants grown hydroponically with the corresponding *I* (B), *T* (C), *RH* (D), *VPD* (E), and *C* (F) from 7th to 19th May 2018 in a commercial greenhouse in Japan after data cleansing. Each point represents the data measured at an interval of 5 minutes. The data within the period in grey were used for training and other data used for validation of the model.

reach three times (30.3 µmol s^{-1} per chamber on 11th May) than that of the rainy days.

Figure 4 shows the time course of measured and estimated P_n by the 5-var-model (Eq. 1) developed with the datasets of original (A), 30-minute-MA (B), 1-hour-MA (C), 2-hour-MA (D), 30-minute-SA (E), 1-hour-SA (F), and 2-hour-SA (G). The maximum value of recorded P_n was 26.7 µmol $chamber^{-1}$ s^{-1} as shown in Fig. 4A on the 19th May, whereas lower values were found in the other datasets with MA and SA due to averaging data within the time spans. In case of the original test dataset (Fig. 4A), the 5-var-model ($P_n = 12.622+0.022\times I+0.750\times T-0.299\times RH-1.309\times VPD+0.013\times C$) was able to moderately trace the curve pattern of the original P_n indicating that the model might be appropriate to explain the net photosynthetic rates. On the other hand, the datasets of 30-minute-MA (Fig. 4B: $P_n = 25.561+0.27\times I+0.765\times T-0.437\times RH-1.964\times VPD+0.014\times C$), 1-hour-MA (Fig. 4C: $P_n = 29.583+0.028\times I+0.679\times T-0.464\times RH-2.036\times VPD+0.015\times C$), and 2-hour-MA (Fig. 4D: $P_n = 34.249+0.029\times I+0.662\times T-0.514\times RH-2.252\times VPD+0.016\times C$) have the feature of smoothed time course and this feature increased the fit of the models. The datasets of 30-minute-SA (Fig. 4E: $P_n = 30.152+0.028\times I+0.835\times T-0.495\times RH-2.236\times VPD+0.013\times C$), 1-hour-SA (Fig. 4F: $P_n = -3.963+0.019\times I+0.279\times T-0.004\times VPD+0.006\times C$) and 2-hour-SA (Fig. 4G: $P_n = 29.074+0.03\times I+0.63\times T-0.452\times RH-2.041\times VPD+0.015\times C$) have similar features, however SA datasets lost many data points compared to MA datasets.

Figure 5 shows the time course of measured and estimated P_n by the 3-var-model (Eq. 2) developed with the datasets of original (A), 30-minute-MA (B), 1-hour-MA (C), 2-hour-MA (D), 30-minute-SA (E), 1-hour-SA (F), and 2-hour-SA (G). The 3-var-model could not trace the time course of P_n at most of the time in datasets of original (Fig. 5A: $P_n = 2.647+0.021\times I+0.110\times VPD+0.001\times C$). However, it fitted better in the 30-minute-MA (Fig. 5B: $P_n = 2.912+0.022\times I+0.032\times VPD+0.002\times C$), but the fit did not improve in the datasets of 1-hour-MA (Fig. 5C: $P_n = 2.491+0.021\times I+0.037\times VPD+0.003\times C$), and 2-hour-MA (Fig. 5D: $P_n = 2.727+0.021\times I+0.009\times VPD+0.003\times C$). The datasets of 30-minute-SA (Fig. 5E: $P_n = 4.114+0.022\times I-0.001\times VPD+0.001\times C$), 1-hour-SA (Fig. 5F: $P_n = 4.959+0.021\times I-0.028\times VPD+0.002\times C$), and 2-hour-SA (Fig. 5G: $P_n = 1.921+0.023\times I-0.023\times VPD+0.004\times C$) resulted similar to the MAs.

Figure 6 shows the correlation between the measured

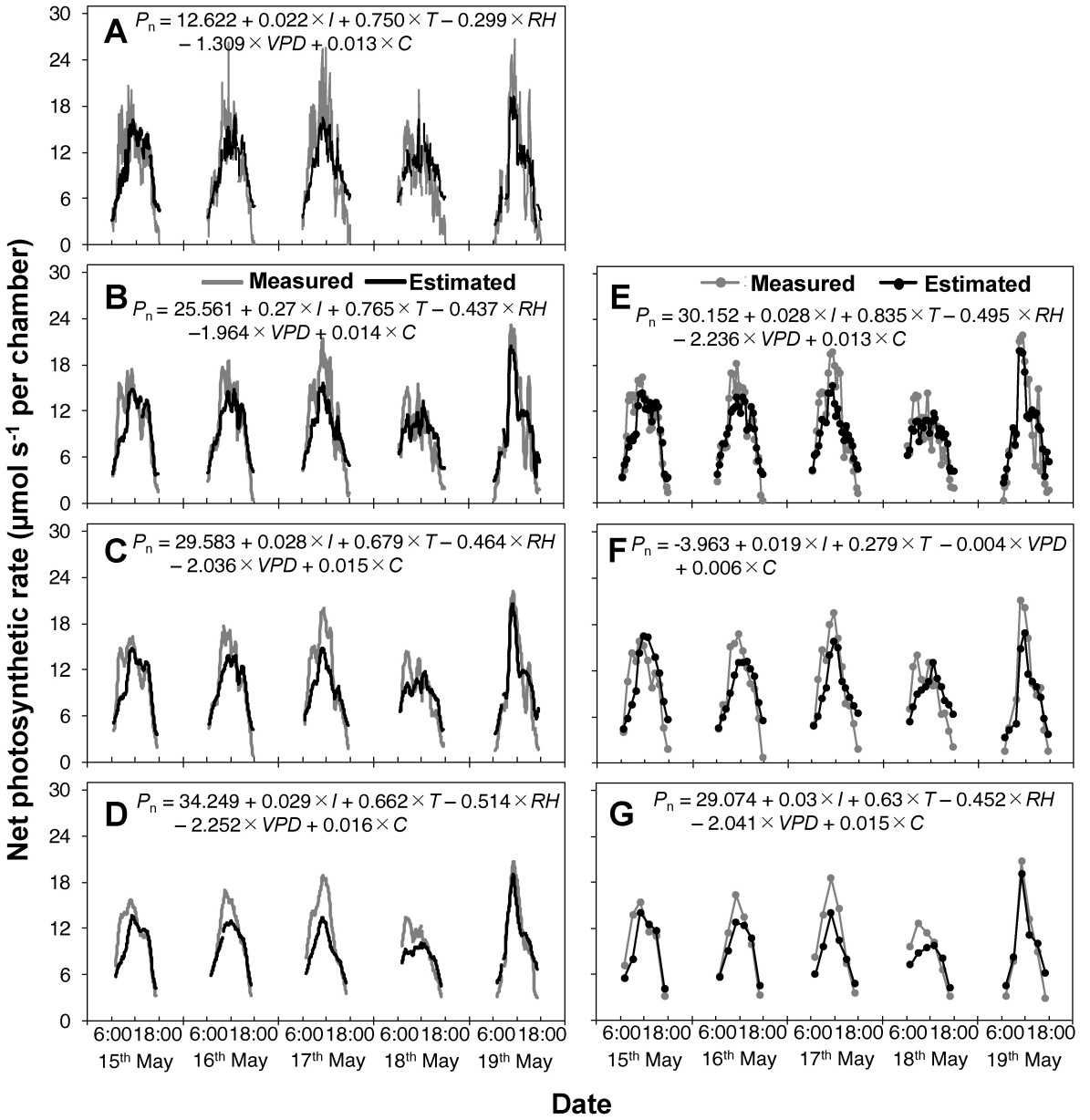

Fig. 4 Time course of measured and estimated P_n by the 5-var-model with the predictor variables of *I*, *T*, *RH*, *VPD* and *C* with the datasets of original (A), 30-minute-MA (B), 1-hour-MA (C), 2-hour-MA (D), 30-minute-SA (E), 1-hour-SA (F), and 2-hour-SA (G).

and estimated P_n by 5-var-model (Eq. 1) developed with the prepared datasets such as original dataset (A), 30-min-MA (B), 1-hour-MA (C), 2-hour-MA (D), 30-minute-SA (E), 1-hour-SA (F), 2-hour-SA (G). The 5-var-model of original dataset overestimated in the lower range of P_n less than 10 μmol s^{-1} per chamber and underestimated in the higher range of P_n more than 10 μmol s^{-1} per chamber (Fig. 6A). A part of these deviations might be attributed to that the linear regression model of Eq. 1 does not take into account a curvilinear response of photosynthesis to the absorbed PAR (Castilla, 2013). Furthermore, the estimated P_n showed a large deviation (Fig. 6A); hence the R^2 was relatively low (0.58). The 5-var-model of 30-minute-MA (Fig. 6B) showed significantly higher R^2 of 0.72 and lower RMSE of 2.9 compared with those of the 5-minute model of original dataset, which is mainly attributed to the dramatic improvement of the overestimation in lower P_n range and underestimation of in higher P_n range (Fig. 6A). The 5-var-models of 1-hour-MA (Fig. 6C) and 2-hour-MA (Fig. 6D) showed some improvements in the same way of the model of 30-minute-MA; however, the R^2s (0.73) and RMSEs (2.7) kept similar values those of the model of 30-minute-MA. In addition, a series of several plots, which is probably due to the time course, became apparent in the longer time-spans of MA, especially in 1-hour-MA (Fig. 6C) and 2-hour-MA (Fig. 6D). As MA takes the average of the sliding time-span, one characteristic data point would affect the moving averages for the time-spans before and after the data point. The 5-var-models of SA (Fig. 6E-G) showed a different feature from the models of MA (Fig. 6B-D). Similar to the 30-minute-MA (Fig. 6B), the 30-minute-SA (Fig. 6E) showed significantly higher R^2 of 0.74 and lower RMSE of 2.9 compared with those of the 5-minute model of original dataset (Fig.

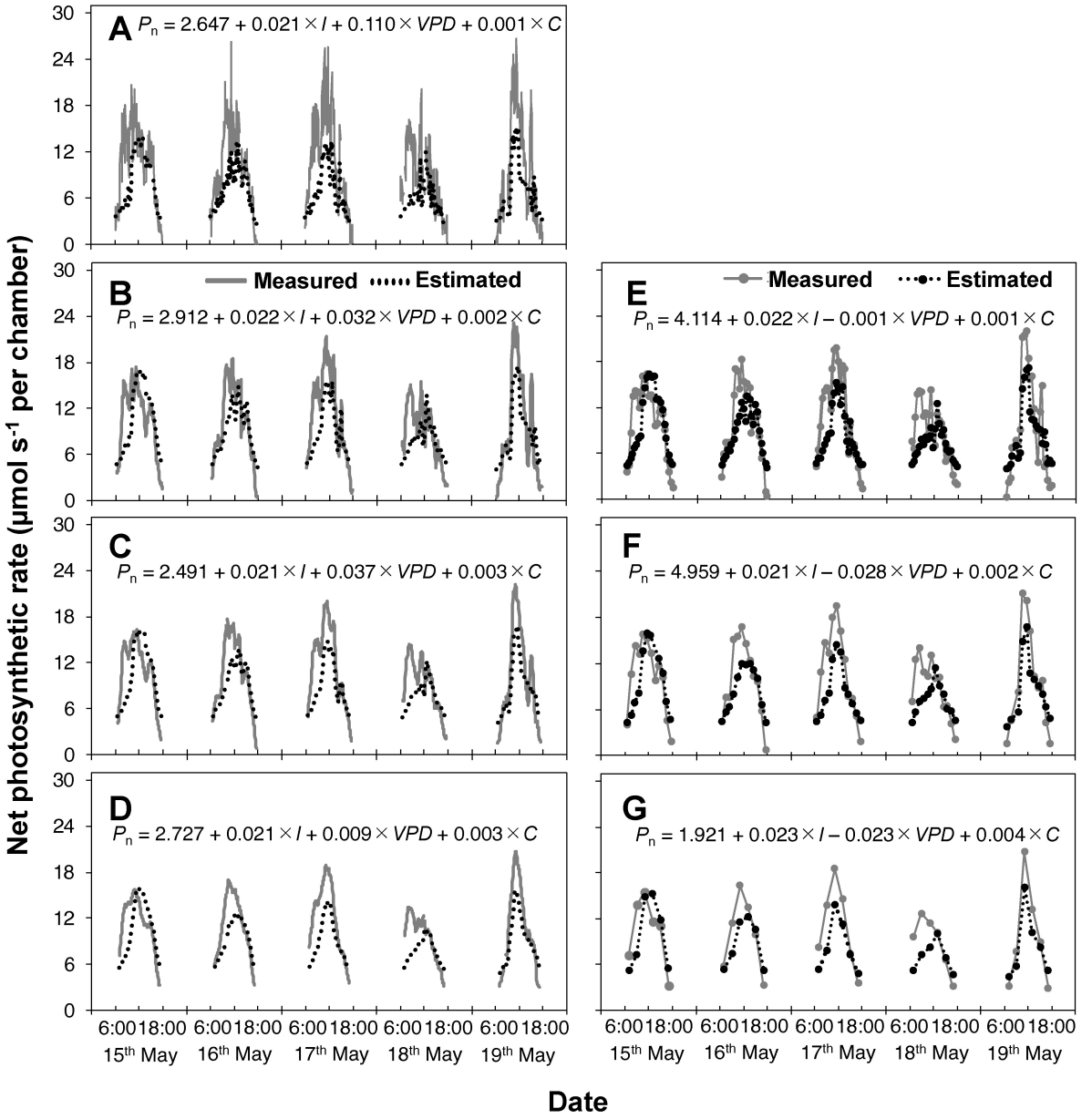

Fig. 5 Time course of measured and estimated P_n by the 3-var-model with the predictor variables of *I*, *VPD* and *C* with the datasets of original (A), 30-minute-MA (B), 1-hour-MA (C), 2-hour-MA (D), 30-minute-SA (E), 1-hour-SA (F), and 2-hour-SA (G).

6E); however, the 1-hour-SA did not contribute to improving R^2 (0.64) nor RMSE (3.2) (Fig. 6F). On the other hand, the 2-hour-SA showed the highest R^2 of 0.81 and the lowest RMSE of 2.4 among the developed 5-var-models (Fig. 6A-G). Furthermore, the series of several plots of time courses recognized in 1-hour-MA (Fig. 6C) and 2-hour-MA (Fig. 6D) were not apparent in 1-hour-SA (Fig. 6F) and 2-hour-SA (Fig. 6G)

Figure 7 shows the correlation between the measured and estimated P_n by using 3-var-model (Eq. 2) developed with the prepared datasets such as original dataset (A), 30-minute-MA (B), 1-hour-MA (C), 2-hour-MA (D), 30-minute-SA (E), 1-hour-SA (F), 2-hour-SA (G). The general features recognized in the results of 3-var-models (Fig. 7A-G) are quite different from those of 5-var-models (Fig. 6A-G) without the over (under) estimation in the lower (higher) range of P_n and the series of several plots of time courses in 1-hour-MA (Fig. 7C) and 2-hour-MA (Fig. 7D). Though 30-minute-MA (Fig. 7B) improved the values of R^2 and RMSE compared with those of the 5-minute model of the original dataset (Fig. 7A), 1-hour-MA (Fig. 7C), and 2-hour-MA (Fig. 7D) did not contribute to the improvements of R^2 and RSME. On the other hand, the longer the time-spans of SA contributed to the improvement of R^2 and RSME (Fig. 7E-G), which is also different from the effect of time-spans of SA in the 5-var-models (Fig. 6E-G). Moreover, the 3-var-models tended to underestimate the P_n more than the 5-var-models, i.e., the slopes of the linear regression between the measured and estimated P_n of 3-var-models were smaller than those of 5-var-models. In addition, the R^2s of the 3-var-models were smaller than those of the 5-var-models, and the RMSEs of the 3-var-models were larger than those of the 5-var-models.

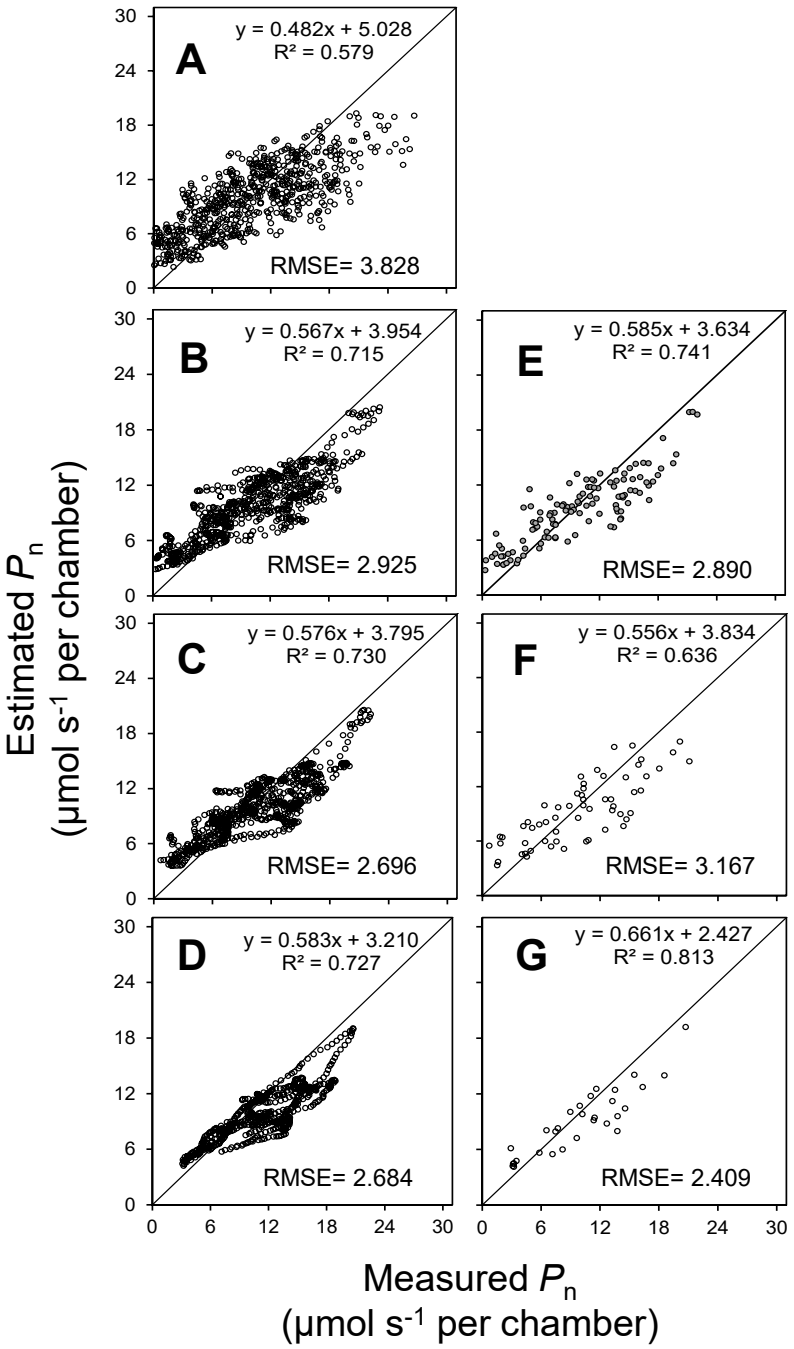

Fig. 6 Correlation between the measured and estimated P_n by 5-var-model with the predictor variables of *I*, *T*, *RH*, *VPD* and *C* developed with the datasets of original dataset (A), 30-minute-MA (B), 1-hour-MA (C), 2-hour-MA (D), 30-minute-SA (E), 1-hour-SA (F), 2-hour-SA (G).

We think that the results in this study, i.e., the 2-hour-SA dataset will give higher accuracy, are not limited to spring condition because the absolute P_n and the relevant environmental factors in other seasons only vary slightly under the prevailing greenhouse condition. However, further model development should opt for *VPD* than *RH* for the model (Grossiord et al., 2020) as the 5-var model resulted in a multicollinearity problem between them (variance inflation factors/VIFs were more than 10 for both variables) due to high correlation. Also, if high time resolution is necessary, the 30-minute time frame with moving average processing is appropriate (i.e., keeping data points; Fig 6B) and will result in a reasonable accuracy for P_n estimation; otherwise, the 2-hour time frame with simple average processing is used for precise estimation.

This study well-demonstrated how to address the high time-resolution photosynthesis data for model development. The results discussed above suggest that the high time-resolution P_n data measured at an interval of 5 minutes with the photosynthetic chamber system allows us to develop several simple linear regression models for P_n estimation by preparation of some datasets through the moving average (MA) or simple average (SA) with the appropriate time-spans. The longer time-spans of moving/

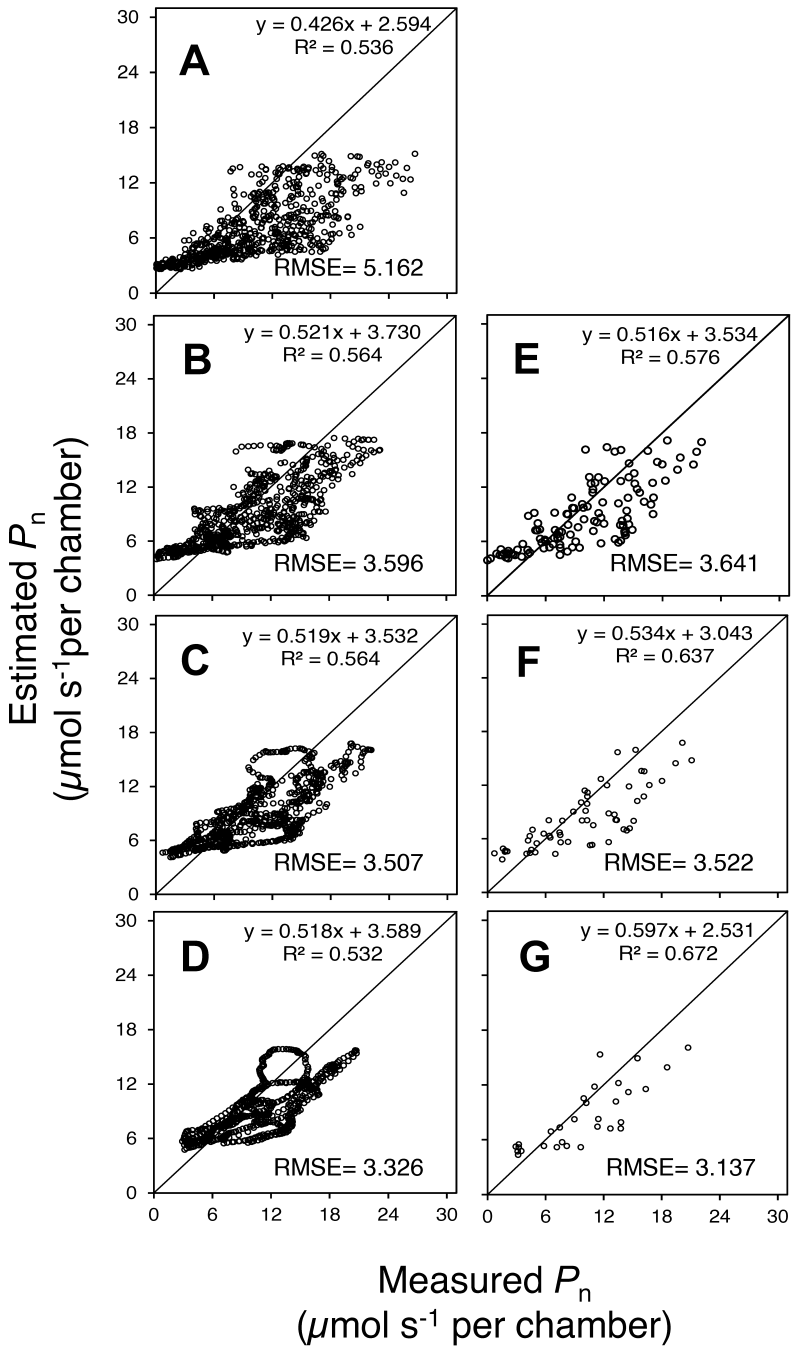

Fig. 7 Correlation between the measured and estimated P_n by 3-var-model with the predictor variables of *I*, *VPD* and *C* developed with the datasets of original dataset (A), 30-minute-MA (B), 1-hour-MA (C), 2-hour-MA (D), 30-minute-SA (E), 1-hour-SA (F), 2-hour-SA (G).

simple averages generally improves the accuracy of the estimation models; however, it can be varied depending on the datasets and the variables used in the models. The averaging techniques with longer time frames may contribute to a better distribution of P_n values, in terms of skewness, for the regression analysis. The 2-hour-SA datasets gave the best accuracy for both 5-var- and 3-var-models by R^2 of 0.81 (Fig. 6G) and 0.67 (Fig. 7G), respectively. These values of R^2 are high enough as P_n estimation model compared to some previous empirical studies that used more complex models such as Nederhoff and Vegter (1994a) (double rectangular hyperbolic relations; R^2 = 0.892) and Lootens and Vandecasteele (1998) (polynomial function; R^2 = 0.89). Thus, we can conclude that datasets of a 2-hour time frame with a simple average are promising to make a practical general linear regression model with high accuracy for the estimation of P_n of cherry tomato canopy grown in the commercial greenhouse by using the high time-resolution P_n measured with the photosynthetic chamber system.

ACKNOWLEDGMENTS

This work was supported by a grant from a commis-

sioned project study on "AI-based optimization of environmental control and labor management for large-scale greenhouse production", Ministry of Agriculture, Forestry and Fisheries, Japan. The authors acknowledge all students and secretaries of the laboratory of Physiological Green System, Ehime University for assistance in this study. The Ph.D. of Y.R. is funded by the Islamic Development Bank through the Ministry of Education and Culture, Republic of Indonesia.

REFERENCES

Acock, B., Charles-Edwards, D. A., Fitter, D. J., Hand, D. W., Ludwig, L. J., Wilson, W. J., Withers, A. C. 1978. The contribution of leaves from different levels within a tomato crop to canopy net photosynthesis: an experimental examination of two canopy models. J. Exp. Bot. **29**: 815−827.

Čampulová, M. 2018. Comparison of methods for smoothing environmental data with an application to particulate matter PM10. Acta Univ. Agric. Silvic. Mendel. Brun. **66**: 453−463.

Castilla, N. 2013. Greenhouse Technology and Management (2nd ed.). CABI, Oxfordshire, pp 335.

Grossiord, C., Buckley, T. N., Cernusak, L. A., Novick, K. A., Poulter, B., Siegwolf, R. T. W., Sperry, J. S., McDowell, N. G. 2020. Plant responses to rising vapor pressur deficit. New Phytol. **226**: 1550−1566.

Hashimoto, Y. 1989. Recent strategies of optimal growth regulation by the speaking plant concept. Acta Hortic. **260**: 115−122.

Khun, M., Johnson, K. 2013. Applied Predictive Modeling. Springer, New York, pp 600.

Lootens, P., Vandecasteele, P. 1998. Whole-plant net photosynthesis, an indication of actual growth, a tool for greenhouse climate control: a case study with *Ficus benjamina* 'Natasja'. Acta Hortic. **421**: 265−270.

Nederhoff, E. M., Vegter, J. 1994a. Photosynthesis of stands of tomato, cucumber and sweet pepper measured in greenhouses under various CO_2-concentrations. Ann. Bot. **73**: 353−361.

Nederhoff, E. M., Vegter, J. 1994b. Canopy photosynthesis of tomato, cucumber and sweet pepper in greenhouse: measurements compared to models. Ann. Bot. **73**: 421−427.

Pyle, D. 1999. Data Preparation for Data Mining. Morgan Kaufmann Publishers, San Francisco, pp 466.

Shimomoto, K., Takayama, K., Takahashi, N., Nishina, N., Inaba, K., Isoyama, Y., Oh, S. 2020. Real-time monitoring of photosynthesis and transpiration of a fully-grown tomato plant in greenhouse. Environ. Control Biol. **58**: 65−70.

Takahashi, N., Ling, P. P., Frantz, J. M. 2008. Considerations for accurate whole-plant photosynthetic measurement. Environ. Control Biol. **46**: 91−101.

Takayama, K., Jansen, R. M., Henten, E. J., Verstappe, F. W., Bouwmeester, H. J., Nishina, H. 2012. Emission index for evaluation of volatile organic compounds emitted from tomato plants in greenhouses. Biosyst. Eng. **113**: 220−228.

Takayama, K. 2013. Second generation speaking plant approach: practical application. (in Japanese) J. SHITA **25**: 165−174.

Thornley, J. H. M. 1976. Mathematical Models in Plant Physiology: A Quantitative Approach to Problems in Plant Crop Physiology. Academic Press, London, pp 318.

van Mourik, S., van Beveren, P. J. M., López-Cruz, I. L. 2019. Improving climate monitoring in greenhouse cultivation via model based filtering. Biosyst. Eng. **181**: 40−51.

Yaffee, R. A., McGee, M. 2000. An Introduction to Time Series Analysis and Forcasting: With Application of SAS and SPSS. Academic Press, New York, pp 582.