Comparison of neural network models with aerodynamic and empirical models toward real-time estimation of the number of air exchanges per hour of a naturally ventilated greenhouse

Ryo Matsuda*,†, Kota Hayano‡, Satoshi Yagi§ and Kazuhiro Fujiwara

*Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1–1–1 Yayoi, Bunkyo, Tokyo 113–8657, Japan
‡Faculty of Agriculture, The University of Tokyo, 1–1–1 Yayoi, Bunkyo, Tokyo 113–8657, Japan
§Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1–1–1 Yayoi, Bunkyo, Tokyo 113–8657, Japan

Abstract

The feasibility of estimating the number of air exchanges per hour (N) of a naturally ventilated greenhouse in real time using neural network (NN) models was evaluated. An aerodynamic (AD) model and an empirical model were also used to compare different types of model. The value of N for an eight-span Venlo-type greenhouse with roof vents containing no plants was measured using the tracer gas method with CO₂ for 17 d. An AD model derived from Bernoulli’s principle, an empirical model, and several NN models in which explanatory variables differed were trained and validated with 4,508 data points subjected to 10-fold cross-validation. We first compared the AD model, the empirical model, and two NN models that used the same explanatory variables in the empirical model, with and without wind direction. The mean accuracy (in terms of root mean square error (RMSE) and coefficient of determination (r²) in the relationship between measured and estimated N values) was the highest for the two NN models, followed by the empirical model and AD model. We next compared four NN models. The differences among them included that, as explanatory variable(s), only the difference between inside (Tᵢ) and outside (Tₒ) air temperatures or both Tᵢ and Tₒ was used and whether solar irradiance (I) was used or not. There was a slight improvement in accuracy when using I, irrespective of how air temperature was handled. The NN models thus tended to exhibit a higher degree of accuracy in estimating N of a naturally ventilated greenhouse than the AD and empirical models considered in this study. For the NN model that performed the best in our comparison, a mean RMSE of 1.08 h⁻¹ and a mean r² of 0.79 were observed.

Key words: Carbon dioxide, Environment control, Machine learning, Natural ventilation, Tracer gas method

1. Introduction

In general, greenhouses are occasionally ventilated so that inside air temperature and/or relative humidity are not too high or the inside CO₂ concentration is not too low. To control the levels of these environmental elements in a greenhouse as accurately and quickly as possible, the amount of air ventilated per unit of time, e.g., the number of air exchanges per hour (N), must be known, as well as the levels of the environmental elements both inside and outside the greenhouse. In particular, it is useful if the value of N at the moment can be estimated on-site in real time. Such real-time estimation of N of a greenhouse theoretically makes it possible to calculate in real time the amounts of CO₂, water vapor, and enthalpy transported through open ventilators (and also due to leakage) per unit of time and therefore contributes to performing cost-effective prompt environment control, such as CO₂ enrichment, during ventilation. There are two primary methods of ventilation, including natural ventilation and mechanical (or forced) ventilation. Compared with mechanical ventilation with fans, natural ventilation is complex, and real-time estimation of N for natural ventilation is challenging. To the best of our knowledge, a practical methodology for estimating N of a naturally ventilated greenhouse in real time with a sufficiently high accuracy is not available to commercial greenhouse operators.

There are several approaches to modeling natural ventilation (Boulard, 2006). Among them, four methods are considered to be applicable to commercial greenhouses with natural ventilation for calculating their N value, including the energy balance method, the tracer gas method, the method based on aerodynamic (AD) models, and the method based on statistical models. The energy balance method is based on measurements involved in the energy balance equation of the greenhouse. Although this method has been employed in many studies (e.g., Kozai et al., 1980; Fernandez and Bailey, 1993), it requires the measurement of a large number of variables, and a single inaccuracy can have a significant effect on the final result (Baptista et al., 1999). The tracer gas method exhibits a higher degree of accuracy than the energy balance method and is based on a mass balance of a tracer gas, such as CO₂, N₂O, SF₆, and H₂O, in the greenhouse air (e.g., Okada and Takakura, 1973; Nederhoff et al., 1985; Boulard and Draoui, 1995). However, CO₂ is not particularly reliable for real-time estimation of N of cropped greenhouses, and N₂O and SF₆ are too expensive to use in commercial greenhouses. By contrast, H₂O can be used for real-time estimation of N even though the greenhouse contains plants, provided that evapotranspiration rate therein is accurately measured (Takakura et al., 2009, 2017). Nevertheless, there appears to be room for improving accuracy, especially under the

Received: March 17, 2019
Accepted: June 7, 2019
*Corresponding Author: amatsuda@mail.ecc.u-tokyo.ac.jp
DOI: 10.2480/agrmet.D-19-00009
conditions where $N$ suddenly changes (Takakura, 2017).

The AD models of natural ventilation are based on Bernoulli’s principle. They consider that airflow through a vent is caused by the difference between the pressure inside and outside the vent, and that the pressure difference is induced by buoyancy forces (so-called stack effect or thermal effect) and wind forces (wind effect) (e.g., Kozai et al., 1980; Sase et al., 1980; Boulard and Baille, 1995; Boulard et al., 1996; Kittas et al., 1997). Using this method, it is possible to estimate $N$ in real time with measured levels of relevant environmental elements, including the difference between the temperature inside and outside the greenhouse (for stack effect) and wind speed (for wind effect), as well as the opening area of vents. In this method, as pointed out by Boulard (2006), identification of the discharge coefficient ($a$) and the wind pressure coefficient ($W$), which are specific to individual greenhouses, must be obtained prior to estimating the value of $N$.

The statistical models range from relatively simple empirical models (e.g., Nederhoff et al., 1985; Fernández and Bailey, 1992) to complex data-driven models such as neural network (NN) models (Seginer et al., 1994; Boulard et al., 1997). Similar to the AD models, this method generally uses levels of relevant environmental elements and vent opening area as explanatory variables, and all parameters/coefficients in the model must be determined before it is used for real-time $N$ estimation.

In this study, we focused on NN models. Early pioneering works on NN models used to estimate the value of $N$ (Seginer et al., 1994; Boulard et al., 1997) revealed relatively high coefficients of determination ($r^2$) of 0.8 or higher for the relationship between $N$ estimated with the models and that measured by the tracer gas method. This is indicative of their potential usefulness. However, in their research, the number of data points used for training and validation was considerably limited (less than 100, Seginer et al., 1994), or cross-validation was not performed (Boulard et al., 1997). Thus, the accuracy of estimated $N$ value using NN models should be re-evaluated more stringently. In addition, NN models employed in the previous studies had relatively simple structures with only one hidden layer. Recent rapid advances in NN technology have enabled us to train NN models that have more complex structures using an inexpensive personal computer for a reasonably short time. Therefore, this approach is now readily available to anyone involved in the greenhouse industry.

In this study, the feasibility of estimating $N$ of a naturally ventilated greenhouse in real time was evaluated by using NN models with an uncropped greenhouse. An AD model and an empirical model were used for comparison because some of these models have also exhibited relatively high degrees of accuracy (Nederhoff et al., 1985; Fernández and Bailey, 1992; Boulard and Baille, 1995; Kittas et al., 1997; Baptista et al., 1999). As a dataset for training and validation of the models, the $N$ values for the cropless greenhouse were determined by the tracer gas method using CO$_2$.

### 2. Materials and Methods

#### 2.1 Greenhouse specifications

The experimental greenhouse was located in the head office of Seiwa Co., Ltd. in Shimotsuke, Tochigi, Japan (36°22' N). The greenhouse was a Venlo type with eight spans, oriented northwest and southeast, and covered with an ethylene-tetrafluoroethylene film (F-CLEAN, AGC Green-Tech Co., Ltd., Tokyo, Japan). The length, width, eave height, and ridge height of the greenhouse were 27.0 m, 36.0 m, 4.5 m, and 5.8 m, respectively. The floor area and volume of the greenhouse were approximately 970 m$^2$ and 5,000 m$^3$, respectively. Each span had a roof vent that was continuous along the full length of the greenhouse. The vents either faced northeast or southwest, with every other vent facing the same direction. In this experiment, northeast-facing vents were closed throughout the experiment, and only southwest-facing vents were regulated. The extent of vent opening of the four vents was synchronized. The entire floor inside the greenhouse was covered with concrete or impermeable mulch.

#### 2.2 Measurements

The experiment was conducted from 12:00 on December 10, 2017, to 18:00 on December 26, 2017, with the greenhouse containing no plants. The wind speed ($U$ [m s$^{-1}$]) and direction (in true north-based azimuth degree, $\theta$ [$^\circ$]) and solar irradiance ($I$ [W m$^{-2}$]) outside the greenhouse and air temperatures inside ($T_i$ [°C]) and outside ($T_o$ [°C]) the greenhouse were measured and recorded every minute with a greenhouse environment measurement system (ProFinder Next 80; Seiwa). CO$_2$ concentrations inside ($C_i$ [mol mol$^{-1}$]) and outside ($C_o$ [mol mol$^{-1}$]) the greenhouse were measured every minute using nondispersive infrared CO$_2$ analyzers (ZFP; Fuji Electric Co., Ltd., Kanagawa, Japan) that were previously calibrated, and a data logger (NR-1000; Keyence Corp., Osaka, Japan) was used to record data. All sensors for measuring the levels of environmental conditions inside the greenhouse were located near the center of the greenhouse. As the source of CO$_2$ tracer, the exhaust gas of a kerosene-burning boiler (SHB-310TK; NEPON Inc., Tokyo, Japan) located in the headhouse was continuously supplied to the greenhouse through an underground duct and released through the northeast edge of gutters. The rate of CO$_2$ supply (5 [m$^3$ h$^{-1}$]) was constant and estimated to be 5.32 m$^3$ h$^{-1}$, assuming 2.49 kg of CO$_2$ generated per 1 L of burned kerosene. The rate of CO$_2$ release derived from soil respiration was negligible. The extent of the roof vent opening ($R$) was evaluated as the proportion of the vertical height of the vent openings to the maximal height when fully opened (0.65 m) ($0 \leq R \leq 1$). The extent was regulated based on the levels of environmental elements and measured every minute, and 5-min averages were recorded, using a greenhouse environment control system (Maximizer; Priva B.V., De Lier, The Netherlands).

#### 2.3 Calculation of the number of air exchanges per hour

$N$ of the greenhouse is defined by the following differential equation:

$$\frac{dC_i}{dt} = N_{in}(C_o - C_i) + \frac{S}{V}$$

$$= -N_{ex}AC_{C_i} + \frac{S}{V}$$

(1)
where $N_{\text{m}}$ [h$^{-1}$] is $N$ measured by the tracer gas method with CO$_2$ at a given time of $t$, $C_{\text{o}}$, $C_t$, and $\Delta C_t$ [mol mol$^{-1}$] are $C_{\text{o}}$, $C_t$, and their difference ($C_t - C_{\text{o}}$) at $t$, respectively, and $V$ [m$^3$] is the volume of the greenhouse (= 5,000 m$^3$) (refer also to Table 1 for variables and coefficients). In this study, we solved Eq. 1 by assuming a steady state ($dC_t/dt = 0$) during a considered short period of time of 5 min, and $N_{\text{m}}$, was calculated using the following equation:

$$N_{\text{m}} = \frac{S}{V \Delta C_t}$$

(2)

2.4 AD model

The AD model used to estimate $N$ of a naturally ventilated greenhouse with only roof vents was based on the equation derived by Boulard and Baille (1995):

$$N = \frac{3,600}{V} \frac{A}{2} \alpha \left( 2G \cdot \frac{\Delta T}{T_v + 273.15} + \frac{H}{4} + WU \right)^{0.5}$$

(3)

where $A$ [m$^2$] is the opening area of roof vents, $\alpha$ is the discharge coefficient of the roof vent opening, $G$ [m s$^{-2}$] is the gravitational acceleration, $\Delta T$ [°C] is the difference between inside and outside air temperatures ($T_i - T_o$), $H$ [m] is the vertical height of roof vent opening, and $W$ is the global wind pressure coefficient. The numerical value of 3,600 is a conversion factor used to convert seconds to hours. The first and second terms in the bracket correspond to the stack and wind effects, respectively. In this study, we assumed that respective $A$ and $H$ were linearly correlated with $R$ and that $\alpha$ and $W$ can be considered virtually constant. By integrating constants and numerical values into coefficients that were regarded specific to the greenhouse used and considering the leakage, Eq. 3 was modified as follows:

$$N_{\text{m}} = \left( c_1 R_i + c_9 \right) \left( c_1 R_i + c_9 \frac{\Delta T_i}{T_{\text{v}i} + 273.15} + c_5 U_{\text{ii}}^2 \right)^{0.5} + c_8$$

(4)

where $N_{\text{m}}$ [h$^{-1}$] is estimated $N$ at $t$, $R_i$, $\Delta T_i$, $T_{\text{v}i}$, $T_{\text{oi}}$, $U_{\text{ii}}$, and $U_i$, [m s$^{-1}$] are $R$, $\Delta T$, $T_v$, and $U$ at $t$, respectively, being explanatory variables of this model. $c_1$ to $c_9$ are coefficients.

2.5 Empirical model

We formulated several kinds of empirical model equations with $U$, $T_i$, $\Delta T$, and $R$ as explanatory variables by repeating trial and error. Among evaluated, the following equation gave the smallest root mean square error (RMSE) for the relationship between $N_{\text{m}}$ and $N_{\text{e}}$, and was therefore included in the model comparison:

$$N_{\text{m}} = \left( c_1 R_i + c_9 \right) \Delta T_i^{0.5} + c_8 R_i \exp(c_{10} U_{\text{ii}}) + c_{11} \exp(c_{11} U_{\text{ii}}) + c_{12}$$

(5)

where $\Delta T_i$ [°C] and $U_{\text{ii}}$ [m s$^{-1}$] are the averaged $\Delta T$ and $U$ at $t$ and 1 min before $t$, respectively. $c_1$ to $c_{13}$ are coefficients. Consequently, $\theta$ was not used.

2.6 NN models

NN models to estimate $N$ were built using a programming language (Python 3.6.1) with a machine learning library (TensorFlow 1.7.0). The output layer of the NN models was $N$, and several compositions of the input layer (i.e., combinations of explanatory variables) were tested (Table 2; see also Results and Discussion). Construction and hyperparameters of the NN models are presented in Table 3. In preliminary evaluation, we examined some of the hyperparameters and selected one for each hyperparameter that gave a low RMSE in a relatively short period of time required for learning. For the number of units in each hidden layer, 10, 20, ⋯, and 100 layers were tested for each of the two hidden layers (a total of 100 combinations) for all NN models with different input layers, and those which gave the lowest RMSE were selected for each NN.

### Table 1. Explanation of variables/coefficients and their units.

| Variable/coefficient | Explanation and unit |
|----------------------|----------------------|
| $A$                  | Opening area of roof vents, m$^2$ |
| $\alpha$             | Discharge coefficient, dimensionless |
| $C_{\text{o}}$        | CO$_2$ concentration inside the greenhouse, mol mol$^{-1}$ |
| $c_i$                | Coefficient ($i = 1, 2, ⋯, 13$) |
| $C_t$                | CO$_2$ concentration outside the greenhouse, mol mol$^{-1}$ |
| $\Delta C$           | $C_{\text{o}} - C_t$, mol mol$^{-1}$ |
| $G$                  | Gravitational acceleration, m s$^{-2}$ |
| $H$                  | Vertical height of the roof vent opening, m |
| $I$                  | Solar irradiance, W m$^{-2}$ |
| $N$                  | Number of air exchanges per hour, h$^{-1}$ |
| $N_{\text{e}}$       | Estimated $N$, h$^{-1}$ |
| $N_{\text{m}}$       | Measured $N$, h$^{-1}$ |
| $R$                  | Extent of roof vent opening, dimensionless |
| $S$                  | CO$_2$ supply rate, m$^3$ h$^{-1}$ |
| $T_i$                | Air temperature inside the greenhouse, °C |
| $T_{\text{v}i}$      | Air temperature outside the greenhouse, °C |
| $\Delta T_i$         | $T_i - T_{\text{v}i}$, °C |
| $\theta$             | Wind direction at time $t$, ° |
| $U$                  | Wind speed, m s$^{-1}$ |
| $V$                  | Greenhouse volume, m$^3$ |
| $W$                  | Wind pressure coefficient, dimensionless |
| $X_i$                | An arbitrary variable $X$ at time $t$ |
| $\overline{X}$       | Mean of an arbitrary variable $X$ at time $t$ and 1 min before $t$ |
| $X_{i-5}$            | An arbitrary variable $X$ at 5 min before time $t$ |

### Table 2. Exploratory variables (or input layers) in the aerodynamic (AD) model, empirical model, and neural network (NN) models.

| Model              | Exploratory variables$^a$ |
|--------------------|---------------------------|
| AD model           | $R_t$, $T_{\text{v}i}$, $\Delta T_i$, $U_t$ |
| Empirical model    | $R_t$, $\Delta T_i$, $U_t$ |
| NN #1-1            | $R_t$, $\Delta T_i$, $U_t$ |
| NN #1-2            | $R_t$, $\Delta T_i$, $U_t$ |
| NN #2-1            | $R_t$, $\Delta T_i$, $U_t$, $\theta_{t-5}$, $I_{t-5}$ |
| NN #2-2            | $R_t$, $\Delta T_i$, $U_t$, $\theta_{t-5}$, $I_{t-5}$ |
| NN #2-3            | $R_t$, $\Delta T_i$, $U_t$, $\theta_{t-5}$, $I_{t-5}$ |
| NN #2-4            | $R_t$, $\Delta T_i$, $U_t$, $\theta_{t-5}$, $I_{t-5}$ |

$^a$ See Table 1 for the explanatory variables.
2.7 Comparison of the models

We obtained time series of levels of environmental elements, \( R \), and measured \( N \) at 5-min intervals. Data at a given time in which \( C_i \) exceeded 2,000 \( \mu \text{mol mol}^{-1} \), which was outside the range of the \( \text{CO}_2 \) analyzers, were excluded, and the remaining 4,508 data points were used for analysis. Every model was trained and validated according to 10-fold cross-validation. The 4,508 data were randomly partitioned into 10 almost equal-sized subdata. Nine were used for training and the remaining one was used for validation, and this process was repeated 10 times. For the AD and empirical models, coefficients were determined with the least-squares method using each training dataset. NN models were trained with mini-batch gradient descent using backpropagation. The RMSE and \( r^2 \) for the 1:1 relationship between \( N_{m} \) and \( N_{e} \) were calculated in each validation of a model, and the mean RMSE and \( r^2 \) values of 10 validations were used to compare the models.

3. Results and Discussion

Daily mean \( U \) was relatively high in the first three days of the experimental period, then decreased and was 1 \( \text{m s}^{-1} \) or lower until December 24, and thereafter increased to 2 \( \text{m s}^{-1} \) or higher (Fig. 1a). Daily \( I \) integral was 8–10 \( \text{MJ m}^{-2} \) and remained virtually constant, except for December 17 and 24, when it was relatively low at 6.8 and 2.3 \( \text{MJ m}^{-2} \), respectively (Fig. 1b). Daily mean \( T_o \) ranged between 1\(^\circ\)C and 9\(^\circ\)C, and the daily mean \( T_i \) was maintained at higher than 11\(^\circ\)C (Fig. 1c). Although daily mean \( C_o \) was 540 \( \pm \) 40 \( \mu \text{mol mol}^{-1} \) throughout the period, the daily mean \( C_i \) increased up to 1,300–1,600 \( \mu \text{mol mol}^{-1} \) (Fig. 1d) because of the continuous supply of \( \text{CO}_2 \).

Figure 2 shows an example of the time course of levels of environmental elements, \( R \), and measured \( N \) on a sunny day (December 13). Roof vents were sometimes open to some extent in the morning and almost fully open in the afternoon until sunset. \( C_i \) and \( \Delta C \) appeared to decrease depending on the extent of the roof vent opening. The pattern of change in \( N \) was apparently similar to that in \( R \).

The AD model (Eq. 4), the empirical model (Eq. 5) and two NN models (NN \#1-1 and \#1-2) were first compared. The

---

Table 3. Construction and hyperparameters of the neural network (NN) models tested.

| Item                        | Number, parameter, or function |
|-----------------------------|--------------------------------|
| Number of hidden layers     | 1, 2, 3                        |
| Number of units in each hidden layer | 10, 20, · · · , 100 \(^{\dagger}\) |
| Weight initializer          | Truncated normal distribution, uniform distribution |
| Activation function         | Rectifier                      |
| Loss function               | Mean square error function, cross entropy function |
| Regularizer                 | None, L2                        |
| Learning rate optimizer     | Adaptive moment estimation      |
| Batch sizes                 | 100, 300, 600                   |
| Number of epochs            | 10,000                         |

\(^{\dagger}\) Those employed for analysis are shown in bold type.

\(^{\dagger}\) Numbers that gave the lowest RMSE were selected for each model.

---

Fig. 1. Daily mean wind speed (a) and daily integral of solar irradiance (b) outside the greenhouse, and daily mean air temperatures (c) and \( \text{CO}_2 \) concentrations (d) inside and outside the greenhouse from December 11 to 26, 2017.
explanatory variables of the two NN models were those used for the empirical model, with (NN #1-2) and without (NN #1-1) the averaged $\theta$ at time $t$ and 1 min before $t$ (see Table 2). The mean RMSE was the lowest for NN #1-1 and #1-2, followed in order by the empirical model and AD model (Table 4). Mean $r^2$ values had a similar result, where the values for NN #1-1 and #1-2 were the highest. NN #1-1 and #1-2 were of almost equal mean RMSE and $r^2$, suggesting that the effect of $\theta$ on $N$ may not be large at

---

Fig. 2. Time course of the levels of environmental elements, the extent of roof vent opening, and the number of air exchanges per hour measured by the tracer gas method with CO$_2$ on December 13. Wind speed ($U$) (a), wind direction (in true north-based azimuth degree, $\theta$) (b), air temperature inside and outside the greenhouse ($T_i$ and $T_o$, respectively) and their difference ($T_i - T_o$) (c), the relative extent of roof vent opening ($R$) (d), solar irradiance ($I$) (e), CO$_2$ concentration inside and outside the greenhouse ($C_i$ and $C_o$, respectively) and their difference ($C_i - C_o$) (f), and the number of air exchanges per hour ($N$) (g). A part of the data are excluded in (e) and (g) at which time $C_i$ exceeded 2,000 µmol mol$^{-1}$. 
least under certain conditions, as previously reported by several studies (Fernández and Bailey, 1992; Fernández and Bailey, 1993; Bouard and Draoui, 1995; Kittas et al., 1997; Seginer et al., 1994). NN #1-1 showed a slightly higher accuracy than the empirical model, although it essentially used the same explanatory variables. This suggests that the empirical model was slightly too simplified to express the effects of $R$, $\Delta T$ and $U$ on $N$ under the conditions of this study. The AD model exhibited a relatively low degree of accuracy, although it has shown high levels of accuracy in previous studies (Bouard and Baille, 1995; Kittas et al., 1997; Baptista et al., 1999). There have been some studies reporting that $\alpha$ and $W$ are dependent more or less on wind speed and/or wind direction (Sase et al., 1980; Bouard and Baille, 1995; Bouard and Draoui, 1995; Baptista et al., 1999).

In this study, we assumed these values to be constant, and this assumption might affect in part the accuracy of estimation using the AD model.

We next compared four NN models, in which $R$ and levels of environmental elements at time $t$ and those at 5 min before $t$ were independently used as explanatory variables (Table 2), assuming the possible delayed effects of environmental conditions on $N$. The reason why data at 5 min before $t$, not 1 min before $t$ for example, were used was that $R$ was recorded at 5-min intervals. The differences among the four models were that, as an explanatory variable(s), only $\Delta T$ (NN #2-1 and #2-3) or both $T_r$ and $T_s$ (NN #2-2 and #2-4) were used and whether $I$ was used (NN #2-3 and #2-4) or not (NN #2-1 and #2-2). Only slight decreases in the mean RMSE were observed when using $I$, irrespective of how air temperature was handled ($\Delta T$ or $T_r$ and $T_s$) (Table 5). On the other hand, no consistent effect of using $T_r$ and $T_s$ instead of $\Delta T$ on mean RMSE and $r^2$ was observed. The highest mean $r^2$ of 0.79 was obtained with NN #2-4, of which 10 validations are shown in Fig. 3. This $r^2$ value was slightly lower than those previously obtained with NN models (0.8 or higher, Seginer et al., 1994; Bouard et al., 1997), although in their research, the

### Table 4. RMSEs and the coefficients of determination ($r^2$) for the relationship between $N_{m,t}$ and $N_{e,t}$ estimated using the aerodynamic (AD) model, empirical model, and neural network (NN) models.

| Model | Number of units in the 1st/2nd hidden layers | RMSE [h$^{-1}$] | $r^2$ |
|-------|---------------------------------------------|-----------------|-------|
| AD model | – | 1.47 ± 0.27 | 0.62 ± 0.07 |
| Empirical model | – | 1.32 ± 0.29 | 0.68 ± 0.09 |
| NN #1-1 | 40/20 | 1.25 ± 0.31 | 0.71 ± 0.09 |
| NN #1-2 | 10/100 | 1.25 ± 0.25 | 0.71 ± 0.10 |

*See Table 2 for the explanatory variables of the models

* Means = standard deviations (10-fold cross-validation)

### Table 5. RMSEs and the coefficients of determination ($r^2$) for the relationship between $N_{m,t}$ and $N_{e,t}$ estimated using neural network (NN) models with $X_r$ and $X_s$, as explanatory variables.

| Model | Number of units in the 1st/2nd hidden layers | RMSE [h$^{-1}$] | $r^2$ |
|-------|---------------------------------------------|-----------------|-------|
| NN #2-1 | 90/80 | 1.16 ± 0.38 | 0.76 ± 0.08 |
| NN #2-2 | 30/20 | 1.21 ± 0.19 | 0.74 ± 0.03 |
| NN #2-3 | 10/100 | 1.14 ± 0.20 | 0.75 ± 0.11 |
| NN #2-4 | 80/10 | 1.08 ± 0.21 | 0.79 ± 0.06 |

*See Table 2 for the explanatory variables of the models

* Means = standard deviations (10-fold cross-validation)

---

**Fig. 3.** Relationships between the measured ($N_{m,t}$) and estimated ($N_{e,t}$) numbers of air exchanges per hour ($N$) for validation of NN model #2-4 according to 10-fold cross-validation. Dashed lines are the 1:1 lines. See Table 2 and 3 for the explanatory variables and mean RMSE and $r^2$ of the model, respectively.
number of data points used for training and validation was small (≤ 100) or cross-validation was not performed. Our result thus demonstrates the applicability of NN models after more strict validation using a large number of data points (≈ 4,500).

A few points remain undetermined and are therefore important objectives that will be addressed in a future research. One is the optimum number of units in hidden layers of NN models, which largely differed among the models in this study (Tables 4, 5). It may vary depending on the composition of the input layer. Another is that how and what past environmental information is important for the current \( N \). In addition, if this estimation method is intended to be used all year round, it is important to verify the applicability of an NN model that is validated in a short period of time in a specific season (i.e., approximately two weeks in winter in this study) to other seasons.

In summary, NN models tended to exhibit a higher degree of accuracy (low RMSE and high \( r^2 \)) in estimating \( N \) of a naturally ventilated greenhouse than the AD and empirical models tested in this study. Under the conditions of this study, the NN model with \( R, T_e, T_a, U, \theta, \) and \( I \) at the time and 5 min before the time of the estimation of \( N \) performed the best (mean RMSE = 1.08 h\(^{-1}\); mean \( r^2 = 0.79 \)). An advantage of an NN model is that it can be trained and validated for a target greenhouse on-site without any detailed manual adjustments, which suggests that it may potentially be used for the practical estimation of \( N \). To perform more cost-effective prompt environment control including CO\(_2\) enrichment, we believe that such real-time estimation of \( N \) and hence the energy and mass budget of the greenhouse must be one of key technologies. We are conducting a subsequent study in which data are newly collected from a different greenhouse.

Acknowledgments

We would like to thank Seiwa Co., Ltd for their technical support for measurements in the greenhouse. This work was financially supported in part by a grant from commissioned project study on “the research project for the future agricultural production utilizing artificial intelligence”, Ministry of Agriculture, Forestry and Fisheries and by JSPS KAKENHI (grant number 17K19306).

References

Baptista FJ, Bailey BJ, Randall JM, Meneses JF, 1999: Greenhouse ventilation rate: theory and measurement with tracer gas techniques. *Journal of Agricultural Engineering Research* 72, 363–374.

Boulard T, 2006: Greenhouse natural ventilation modelling: a survey of the different approaches. *Acta Horticulturae* 719, 29–40.

Boulard T, Baille A, 1995: Modelling of air exchange rate in a greenhouse equipped with continuous roof vents. *Journal of Agricultural Engineering Research* 61, 39–48.

Boulard T, Draoui B, 1995: Natural ventilation of a greenhouse with continuous roof vents: measurements and data analysis. *Journal of Agricultural Engineering Research* 61, 27–36.

Boulard T, Feuilloley P, Kittas C, 1997: Natural ventilation performance of six greenhouse and tunnel types. *Journal of Agricultural Engineering Research* 67, 249–266.

Boulard T, Meneses JF, Mermier M, Papadakis G, 1996: The mechanisms involved in the natural ventilation of greenhouses. *Agricultural and Forest Meteorology* 79, 61–77.

Fernández JE, Bailey BJ, 1992: Measurement and prediction of greenhouse ventilation rates. *Agricultural and Forest Meteorology* 58, 229–245.

Fernandez JE, Bailey BJ, 1993: Predicting greenhouse ventilation rates. *Acta Horticulturae* 328, 107–114.

Kittas C, Boulard T, Papadakis G, 1997: Natural ventilation of a greenhouse with ridge and side openings: sensitivity to temperature and wind effects. *Transactions of the ASAE* 40, 415–425.

Kozai T, Sase S, Nara M, 1980: A modelling approach to greenhouse ventilation control. *Acta Horticulturae* 106, 125–136.

Nederhoff EM, van de Vooren J, Udink ten Cate AJ, 1985: A practical tracer gas method to determine ventilation in greenhouses. *Journal of Agricultural Engineering Research* 31, 309–319.

Okada M, Takakura T, 1973: Guide and data for greenhouse air conditioning. 3. Heat loss due to air infiltration of heated greenhouse. *Journal of Agricultural Meteorology* 28, 11–18.

Sase S, Kozai T, Nara M, Negishi H, 1980: Ventilation of greenhouses. I. Wind tunnel measurements of pressure and discharge coefficients for a single-span greenhouse. *Journal of Agricultural Meteorology* 36, 3–12 (in Japanese with English abstract and captions).

Seginer I, Boulard T, Bailey BJ, 1994: Neural network models of the greenhouse climate. *Journal of Agricultural Engineering Research* 59, 203–216.

Takakura T, 2017: Calculation of plant canopy photosynthesis in a greenhouse. *Agriculture and Horticulture* 92, 849–852 (in Japanese).

Takakura T, Kubota C, Sase S, Hayashi M, Ishii M, Takayama K, Nishina H, Kurata K, Giacomelli GA, 2009: Measurement of evapotranspiration rate in a single-span greenhouse using the energy-balance equation. *Biiosystems Engineering* 102, 298–304.

Takakura T, Sunagawa H, Tamaki M, Usui T, Taniai N, 2017: In site net photosynthesis measurement of a plant canopy in a single-span greenhouse. *Journal of Advances in Agriculture* 2, 1015–1020.