






Article

# Application of Random Forest and ICON Models Combined with Weather Forecasts to Predict Soil Temperature and Water Content in a Greenhouse

Yi-Zhih Tsai <sup>1</sup>, Kan-Sheng Hsu <sup>1</sup>, Hung-Yu Wu <sup>2</sup>, Shu-I Lin <sup>2</sup>, Hwa-Lung Yu <sup>1</sup>,  
Kuo-Tsang Huang <sup>1</sup>, Ming-Che Hu <sup>1</sup> and Shao-Yiu Hsu <sup>1,\*</sup>

<sup>1</sup> Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 10617, Taiwan; tsaiyz@ntu.edu.tw (Y.-Z.T.); carter840826@gmail.com (K.-S.H.); hlyu@ntu.edu.tw (H.-L.Y.); huangkt@ntu.edu.tw (K.-T.H.); mchu@ntu.edu.tw (M.-C.H.)

<sup>2</sup> Department of Horticulture and Landscape Architecture, National Taiwan University, Taipei 10617, Taiwan; r05628135@g.ntu.edu.tw (H.-Y.W.); shuilin@ntu.edu.tw (S.-I.L.)

\* Correspondence: syhsu@ntu.edu.tw

Received: 20 March 2020; Accepted: 16 April 2020; Published: 20 April 2020



**Abstract:** Climate change might potentially cause extreme weather events to become more frequent and intense. It could also enhance water scarcity and reduce food security. More efficient water management techniques are thus required to ensure a stable food supply and quality. Maintaining proper soil water content and soil temperature is necessary for efficient water management in agricultural practices. The usage of water and fertilizers can be significantly improved with a precise water content prediction tool. In this study, we proposed a new framework that combines weather forecast data, numerical models, and machine learning methods to simulate and predict the soil temperature and volumetric water content in a greenhouse. To test the framework, we performed greenhouse experiments with cherry tomatoes. The numerical models and machine learning methods we selected were Newton's law of cooling, HYDRUS-1D, the random forest model, and the ICON (inferring connections of networks) model. The measured air temperature, soil temperature, and volumetric water content during the cultivation period were used for model calibration and validation. We compared the performances of the models for soil temperature and volumetric water content predictions. The results showed that the random forest model performed a more accurate prediction than other methods under the limited information provided from greenhouse experiments. This approach provides a framework that can potentially learn best water management practices from experienced farmers and provide intelligent information for smart greenhouse management.

**Keywords:** cherry tomato; time-series; machine learning; inferring connections of networks (ICON), HYDRUS-1D

## 1. Introduction

Given the vulnerability of agricultural production to extreme weather events whose intensity and frequency will increase with climate change, proactively managing agricultural risks to sustain production is becoming increasingly important [1]. To ensure a stable food supply and quality, intensive farming has been used since the mid-1980s to monitor the food production chain and manage its supply and quality. Precision agriculture can optimize crop/fruit production while maintaining environmental quality to achieve a safe food supply [2]. Precision agriculture includes a range of technologies, including enhanced sensors, information systems, and mechanical equipment. This approach primarily optimizes production efficiency by considering and managing uncertainty and variability within

agricultural systems [3]. Compared with a field cultivation system, precision agriculture is relatively easy to achieve in a greenhouse system because it can control environmental factors more effectively.

Greenhouse systems can ensure the quality of crops by saving water, controlling the temperature and reducing the use of pesticides [4]. Placing the monitoring instruments in a greenhouse can also enhance management effectiveness [5]. Moreover, for plants, root development will affect future growth [6]. Soil temperature and water content are closely related to the root systems of plants. Water content can provide protection for plant root systems and reduce changes in soil temperature caused by changes in air temperature between day and night [7]. However, previous studies have noted that even if a greenhouse system compared to outdoors can control many environmental factors, there are still many uncontrollable conditions and interactions [8]. These factors cause difficulties in maintaining soil temperature and water content for efficient water management and plant root protection in agricultural practices. To achieve this aim, the management of optimized soil temperature and water content in a greenhouse system still needs to be improved. Therefore, it is necessary to develop a framework for intelligent prediction.

To understand the relationship between soil temperature and water content, the interactions between water content and heat in soil have been explored [9–11]. Later, HYDRUS-1D was developed as a simulation software that combined thermal diffusion and the Richards' equation [12]. However, the soil parameters required for simulation must be measured by various complex and time-consuming experiments, thus, data-based mathematical models have recently emerged. Machine learning techniques have become a popular data-based mathematical model in recent years. Machine learning is a method used to teach machines how to handle data more efficiently when it is difficult to interpret patterns or extract information from data [13,14]. Without knowing the actual physical mechanisms involved, using a data-based mathematical model to train data can produce predictions of unknown generation criteria [13]. Machine learning techniques have been used to simulate large-scale soil water content [15,16] and have been compared with numerical methods in the simulation differences of soil water content [17]. However, machine learning has not been used to predict the change trends of soil water content from temperature. Through the inseparable relationship between soil water content and soil temperature [9,12], we can apply machine learning to predict the change of water content over time. Moreover, compared with machine learning, dynamic topology provides an ICON (inferring connections of networks) model to simulate the trend of time series data, uses the existing data to build the system, and determines the relationship between various factors in the system [18]. The ICON model was proposed by Wang *et al.*, 2018 with the aim to extract complex interactions in natural systems in which multiple factors affect each other dynamically. ICON is also a data-driven approach of dynamic interactions for determining the network topology of oscillators with different coupling functions, periodicities, degree nodes, and time scales through solving nonlinear estimation problems as a linear inverse problem [18]. This model can depict the dynamic interactions of a large complex system with noisy data in various fields. Both ICON and machine learning techniques can make predictions without first obtaining physical parameters, which is highly suitable for the dynamic interactions of complex impact factors.

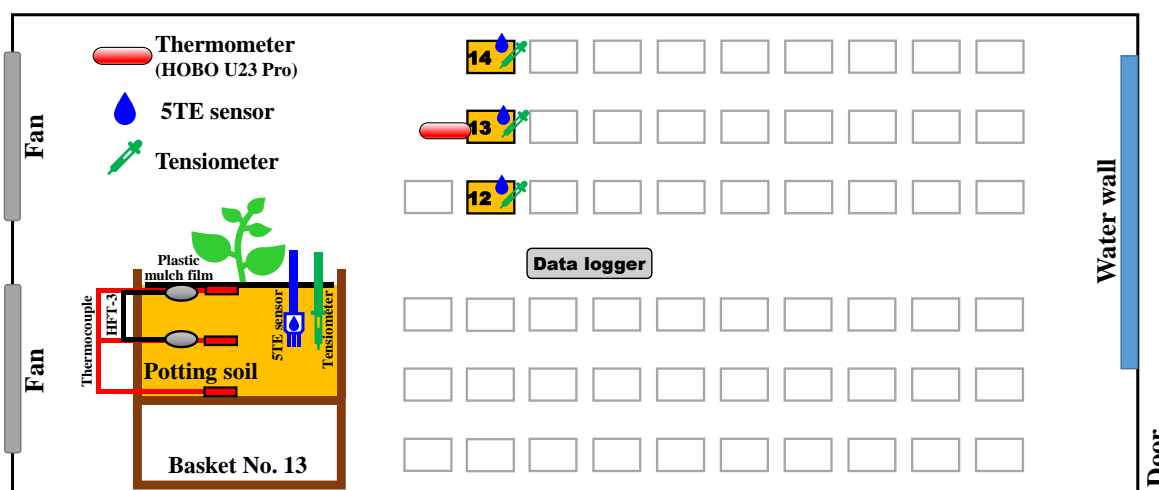
Previous studies focused on simulations but failed to provide future predictions [15–17]. In a greenhouse where the temperature cannot be controlled, the indoor temperature is related to the outdoor temperature and weather. Through this relationship, it is possible to use the weather forecast data to predict the soil temperature and water content in a greenhouse [4,19]. Therefore, the objective of this study is to provide a framework for predicting soil temperature and water content based on a weather forecast with limited measured information. In this study, we also compared the prediction performance of different types of models. We used the monitoring instruments and sensors to collect long-term time-series data for model calibration and validation. The soil temperature and water content were simulated and predicted by combining the weather forecast data. Our findings can improve the usage of water, accurately assess water requirements under various temperature effects, and further develop water content and temperature alarms for greenhouse management.

The greenhouse environment and the characteristics of the materials are described in Section 2. We also detail the methodology of the analyses' data processes. In Section 3, the related applied physics theories and numerical models are introduced. Section 4 is dedicated to the results of the temporal distribution of air and soil temperatures, and the volumetric water content (VWC) for cherry tomato growth. Moreover, we discuss the simulation results and the performances from the models and predict the future soil temperature and water content by combining the weather forecast data. The conclusion is presented in Section 5.

## 2. Materials and Methods

### 2.1. Greenhouse Setup

Our experiments were performed in a greenhouse (24 meters long, 9.6 meters wide, and 5.5 meters high) with a pad and fan system at NTU (National Taiwan University, Taipei, Taiwan). We used the HOBO U23 Pro v2 Temperature/Relative Humidity Data-logger (Onset Computer Corp., Bourne, MA, USA; accuracy: temperature  $\pm 0.2$  °C, relative humidity  $\pm 2.5\%$ ; range: temperature  $-40$  to  $70$  °C, relative humidity 0 to 100%) to collect air temperature and relative humidity in the greenhouse during the experiments. At the same time, the monitoring instruments and sensors were installed. These instruments included 5TE sensors (Decagon Devices, Inc., WA, USA; accuracy: VWC  $\pm 0.03$   $\text{m}^3 \text{m}^{-3}$ , soil temperature  $\pm 1$  °C; range: VWC 0 to  $1 \text{m}^3 \text{m}^{-3}$ , soil temperature  $-40$  to  $50$  °C), temperature probes (T-type Thermocouple, Nzing Co., Taiwan; accuracy: temperature  $\pm 1$  °C; range:  $-200$  to  $200$  °C), HFT-3 heat flux transducers (Campbell Scientific, Inc., UT, USA; accuracy: better than  $\pm 5\%$  of reading; range:  $-100$  to  $100 \text{W m}^{-2}$ ; thermal conductivity  $1.22 \text{W m}^{-1} \text{K}^{-1}$ ), 2100F tensiometers and Model 5301 current transducers (Soilmoisture Equipment Corp., CA, USA; accuracy: matric potential  $\pm 1\%$  span; range: 2 bar pressure difference), T5 pressure transducer tensiometers (UMS GmbH, München, Germany; accuracy: matric potential  $\pm 0.5$  kPa; range:  $-85$  to  $100$  kPa), and the CR1000 data-logger (Campbell Scientific, Inc., UT, USA) for real-time and long-term monitoring of soil water content, soil temperature, soil heat flow, electrical conductivity, and soil matric potential (suction) in the greenhouse. Moreover, before installing the sensors, an inspection process had been completed to ensure that each sensor satisfied the manufacturer's measurement specifications. The experimental instrument diagram of the greenhouse is shown in Figure 1.



**Figure 1.** Experimental instrument diagram of the greenhouse. The long-term cherry tomatoes experiment was performed in baskets No. 12, No. 13, and No.14.

We prepared three separate rectangular baskets (labeled No. 12, No. 13, and No. 14) that were filled with culture substrate about 20 cm deep for the experiments. They were placed in the middle of the greenhouse with a 100 cm spacing. The size of each basket was 60 cm long, 42 cm wide, and 23 cm

high. The entire surface of the culture substrate was mulched with a silver and black plastic mulch film to reduce evaporation of soil water and avoid the fluctuations of temperature in shallow soil, which resulted in a more uniform soil water content, contributed to plant root development, and promoted faster growth [20]. Within the root zone, we buried the aforementioned sensors in the middle of the culture substrate profile layer (about 10 cm deep) at 6–9 cm from the plant. Additionally, in basket No. 13, the soil temperature probes were placed in the upper, middle, and lower portions of the culture substrate profile at a 10 cm distance from each other. Then, the surface, the middle, and the bottom temperatures of the culture substrate were measured. Likewise, the heat flux transducers were placed on the surface surrounding the shallow and middle portions of the culture substrate profile, and the middle portion was 10 cm below the surface. We immediately and quickly collected the sensors' monitoring data via data-logger at 112 days after transplanting (DAT) the tomatoes. The monitoring data was a time series and was recorded in the data-logger at every minute.

## 2.2. Crop Description and Planting

Tomatoes are one of the most common global greenhouse crops. The number of days a tomato plant grows depends on the variety and other environmental factors, such as air temperature, light conditions, soil conditions, and nutrients. The average duration to reach maturity is 65 to 100 days, depending on the variety, ripeness, and maturity [19]. The air temperature suitable for tomato growth is usually between 18.3 and 32.2 °C, and the soil temperature is between 16 and 29.5 °C [19]. Although some studies have discussed the suitable range of soil water content for tomato growth, there is still a lack of literature that clearly indicates predictions and precise control of soil water content during the tomato growth period [21]. Overall, the soil water content is one of the factors affecting the yield and quality of tomato fruits [21,22].

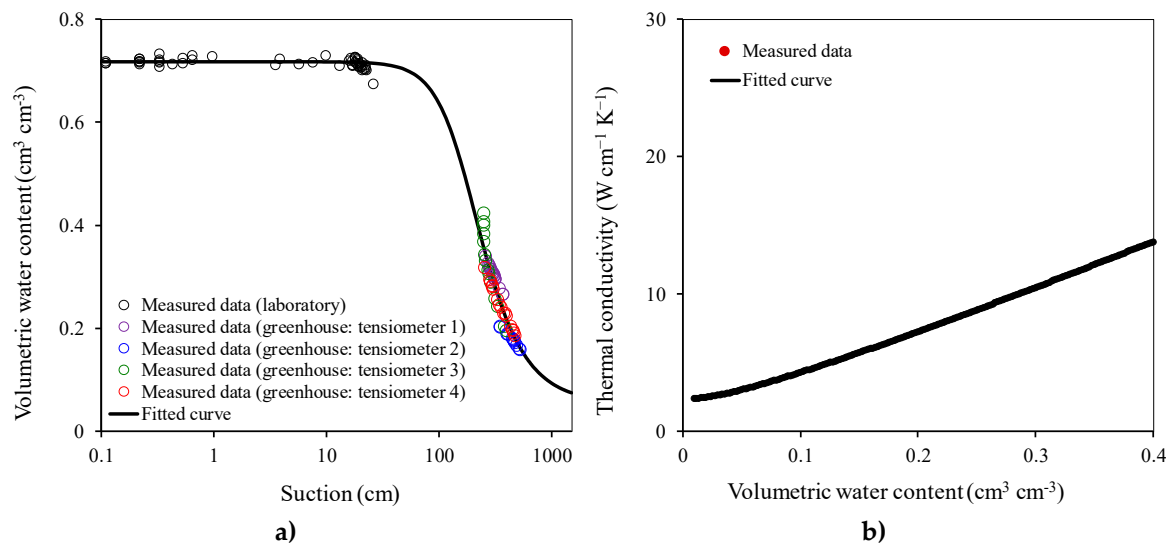
The soil most suitable for tomatoes is deep and rich in organic matter and well-drained materials. The growth of tomato plants is related to many factors, including variety, light, temperature, soil water, fertilizer management, and cultivation techniques. In our greenhouse, cherry tomatoes (*Solanum lycopersicum* cv. Rosada) were transplanted in baskets No. 12, No. 13, and No. 14 on September 20th, 2018. This approach was used to ensure that each tomato did not affect each other's growth, and especially to avoid competition over available water in the root zone. Each basket was evenly separated into four compartments by plastic sheets, and each compartment contained only one tomato plant. Each basket was installed with instruments and sensors in one compartment to monitor one tomato plant.

## 2.3. Culture Substrate Characteristics

The culture substrate is Sunshine® #5 Natural & Organic Mix (Sun Gro Horticulture Distribution Inc., MA, USA). The mix's appearance is fibrous, and the color is light brown to dark brown. Its relative density is between 100 and 400 g/L. The composition of the culture substrate contains dolomitic limestone, fine perlite, fine sphagnum peat moss, and silicon additive. It has the characteristics of fine particle size, low drainage, high water retention, and a higher soil air permeability than general soils. The pH is between 3.5 and 7.5 [23].

We used 2100F tensiometers and 5TE sensors to directly measure the soil water characteristic curve of the culture substrate in the greenhouse during the cultivation period. Figure 2a shows the result of soil water characteristics curve of the culture substrate during the drainage stage. The black circles in Figure 2a indicate the results that are measured in the laboratory under saturated conditions, and the other color circles (purple, blue, green, and red circles) are the results measured in the greenhouse under unsaturated conditions during the tomato growth period. In the greenhouse, water was drained through the holes in the bottom of the baskets, making it difficult for the culture substrate to achieve saturation. We used a container to fill the culture substrate and inject water to achieve a saturated situation in the laboratory. The black curve was fitted by van Genuchten's model (Equation (5)) with the parameters  $\alpha = 0.006 \text{ cm}^{-1}$ ,  $n = 2.62$ ,  $m = 0.62$ ,  $\theta_s = 0.72 \text{ cm}^3 \text{ cm}^{-3}$ , and  $\theta_r = 0.05 \text{ cm}^3 \text{ cm}^{-3}$ .

This curve illustrates the soil water retention characteristics of the culture substrate. A zero matric potential indicates that the soil is saturated. The culture substrate has a high saturated water content of  $0.72 \text{ cm}^3 \text{ cm}^{-3}$ , which means that it retains more water than ordinary soil in the saturated situation.



**Figure 2.** (a) Soil water characteristics curve of the culture substrate (Sunshine<sup>®</sup> #5 Natural & Organic Mix) during the drainage stage; (b) Relationship between thermal conductivity and volumetric water content of the culture substrate.

The saturated hydraulic conductivity ( $K_s$ ) of culture substrate was determined based on Darcy's law via the constant-head experiment. The experiment was carried out using glass filter columns (inner diameter = 2.6 cm; length = 30 cm). A Mariotte's bottle was connected through a silicone tube and a valve to the column to control the hydraulic head. The bottle was placed on an analytical balance (Practum 3102-1S, Sartorius AG, Göttingen, Germany). We converted the measured weight change into flux and calculated  $K_s$  from the known hydraulic head. The measured saturated hydraulic conductivity of the culture substrate was  $K_s = 57.02 \text{ cm day}^{-1}$ .

Furthermore, we used two heat flux transducers to measure the thermal conductivity of the culture substrate. Based on Fourier's law, the thermal conductivity was obtained from the measured heat flux and the temperature gradient. Figure 2b shows the change of thermal conductivity of the culture substrate with volumetric water content. The fitted curve was fitted by the thermal conductivity equation of *Chung and Horton, 1987*,  $\lambda_0(\theta) = b_1 + b_2\theta + b_3\theta^{0.5}$ , with the parameters  $b_1 = 2.78 \text{ W cm}^{-1} \text{ K}^{-1}$ ,  $b_2 = 40.34 \text{ W cm}^{-1} \text{ K}^{-1}$ , and  $b_3 = -8.15 \text{ W cm}^{-1} \text{ K}^{-1}$ .

#### 2.4. Processing and Analyzing Data

These soil temperature and volumetric water content data were collected every minute by the data-logger (CR1000) and averaged into hourly data. The air temperature per 5 minutes in the greenhouse was also averaged into hourly data. The outliers were directly removed before averaging. After conversion to the hourly average data, the first 2000 h of data were used as a training set (20 September 2018 to 12 December 2018), and the subsequent 200 h of data were used as a test set (12 December 2018 to 20 December 2018). Finally, we selected an additional 48 h of data (29 December 2018 to 30 December 2018) combined with the weather forecast for prediction analysis. Moreover, the parameters required for the physical models were obtained through various experiments.

### 3. Physical and Numerical Models

#### 3.1. Physical Models

##### 3.1.1. Newton’s Law of Cooling

In soil heat transport, we used a simple physical model for comparison with data-based or data-driven mathematical models in this study. We assumed a uniform temperature in the soil. Using Newton’s law of cooling [24], the soil temperature can be described by:

$$\frac{dT_s}{dt} = k(T_s - T_a) \tag{1}$$

where  $T_s$  (°C) is the soil temperature,  $T_a$  (°C) is the air temperature in the greenhouse,  $k$  [ $T^{-1}$ ] is a cooling constant, and  $t$  [T] is time.

We used the training set to fit the  $k$  of each basket with the least square method and validated  $k$  by using the test set. The simulated soil temperature value was calculated from the shift terms of Equation (1),  $T_{s(\xi+1)} = k(T_{s(\xi)} - T_{a(\xi)})(t_{(\xi+1)} - t_{(\xi)}) + T_{s(\xi)}$ , where  $\xi = 1, 2 \dots$  depending on the segmented time. The inputs and outputs of the equation are shown in Table 1.

**Table 1.** Inputs and outputs of the studied models.

Models	Parameters	Inputs	Outputs
Newton’s law of cooling	$k = -0.247 \pm 0.042 \text{ hr}^{-1}$	For temperature: $T_{s(\xi)}, T_{a(\xi)}$ , $\Delta t = t_{(\xi+1)} - t_{(\xi)}$ I.C.†: $T_{s0}, T_{a0}$	$T_{s(\xi+1)}$
HYDRUS-1D	$\alpha = 0.006 \text{ cm}^{-1}$ , $n = 2.62$ , $m = 0.62$ , $\theta_s = 0.72 \text{ cm}^3 \text{ cm}^{-3}$ , $\theta_r = 0.05 \text{ cm}^3 \text{ cm}^{-3}$ , $L = 0.5$ , $K_s = 57.02 \text{ cm day}^{-1}$ , $\Delta z = 10 \text{ cm}$ , $b_1 = 2.78 \text{ W cm}^{-1} \text{ K}^{-1}$ , $b_2 = 40.34 \text{ W cm}^{-1} \text{ K}^{-1}$ , $b_3 = -8.15 \text{ W cm}^{-1} \text{ K}^{-1}$ , $\beta = 0^\circ$ , $\beta_T = 5 \text{ cm}$ , $S = 0 \text{ cm}^3 \text{ cm}^{-3} \text{ s}^{-1}$ , $C_p = 1.43 \text{ J cm}^{-3} \text{ K}^{-1}$ , $C_w = 4.18 \text{ J cm}^{-3} \text{ K}^{-1}$	For temperature: $T_{s(\xi)}, \theta_{(\xi)}$ , $\Delta t = t_{(\xi+1)} - t_{(\xi)}$ I.C.: $T_{s0}, \theta_0$ B.C.‡: soil temperature and matric potential for upper and lower boundaries  For volumetric water content (VWC): $\theta_{(\xi)}$ , $\Delta t = t_{(\xi+1)} - t_{(\xi)}$ I.C.: $\theta_0$ B.C.: matric potential for upper and lower boundaries	$T_{s(\xi+1)}$          $\theta_{(\xi+1)}$
Random forest	$n\_estimators = 100$ , $max\_depth = \text{unlimited}$ , $min\_samples\_split = 2$	For temperature: $T_{a(\xi)}, T_{a(\xi+1)}, T_{s(\xi)}$ I.C.: $T_{s0}, T_{a0}$ For VWC: $T_{a(\xi)}, T_{a(\xi+1)}, T_{s(\xi)}, T_{s(\xi+1)}, \theta_{(\xi)}$ , $\theta_{(\xi-1)}$ I.C.: $T_{s0}, T_{a0}, \theta_0, \theta_{-1}$	$T_{s(\xi+1)}$      $\theta_{(\xi+1)}$
Inferring connections of networks (ICON)	$M = 2000$ , $N = 3$ (simulation) or 4 (prediction), $r = 312 \pm 20.1$ (for $T_s$ ); $313 \pm 22.4$ (for $T_a$ ); $38.7 \pm 7.0$ (for VWC)	For temperature and VWC: $T_{s(\xi)}, T_{a(\xi)}, \theta_{(\xi)}$ , forecasted outdoor air temperature as the fourth input factor for prediction stage I.C.: $T_{s0}, T_{a0}, \theta_0$	$T_{s(\xi+1)}$ , $T_{a(\xi+1)}$ , $\theta_{(\xi+1)}$

†: Initial conditions; ‡: Boundary conditions.

### 3.1.2. HYDRUS-1D

HYDRUS-1D has been widely applied to simulate the one-dimensional movement of water and heat in soil [12,25]. It simulates soil heat transport by using the soil heat conduction equation [12], which includes heat convection terms caused by liquid water:

$$\frac{\partial C_p T_s}{\partial t} = \frac{\partial}{\partial z} \left[ \lambda(\theta) \frac{\partial T_s}{\partial z} \right] - C_w \frac{\partial q_L T_s}{\partial z} - C_w S T_s \quad (2)$$

where  $\theta$  ( $L^3 L^{-3}$ ) (e.g.,  $m^3 m^{-3}$ ) is the soil volumetric water content,  $C_w$  and  $C_p$  ( $ML^{-1} T^{-2} K^{-1}$ ) (e.g.,  $J m^{-3} K^{-1}$ ) are the volumetric heat capacities of water and moist soil, respectively,  $q_L$  ( $L T^{-1}$ ) (e.g.,  $m s^{-1}$ ) is the flux density of liquid water,  $S$  ( $L^3 L^{-3} T^{-1}0$ ) (e.g.,  $m^3 m^{-3} s^{-1}$ ) is a sink term,  $z$  (L) is vertical distance, and  $\lambda(\theta)$  ( $ML T^{-3} K^{-1}$ ) (e.g.,  $Wm^{-1} K^{-1}$ ) is the apparent soil thermal conductivity.  $\lambda(\theta)$  can be estimated from:

$$\lambda(\theta) = \lambda_0(\theta) + \beta_T C_w |q_L| \quad (3)$$

where  $\beta_T$  [L] is the thermal dispersivity, and the thermal conductivity ( $\lambda_0$ ) can be set in HYDRUS-1D using either the Campbell model [26] or Chung and Horton model [27]. In this study, we used the Chung and Horton model to estimate  $\lambda_0$  of the culture substrate from the measured data for soil heat flux and temperature in the greenhouse, and the result is shown in Figure 2b.

In HYDRUS-1D, the soil water movement is solved using the Richards' equation for uniform water flow [12]:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z} \left[ K \left( \frac{\partial h}{\partial z} + \cos \beta \right) \right] - S \quad (4)$$

where  $h$  (L) is the matric potential,  $K$  ( $L T^{-1}$ ) is the unsaturated hydraulic conductivity, and  $\beta$  is an angle between the flow direction and the vertical axis, where  $\beta = 0^\circ$  for vertical flow. The soil water retention model [28] and soil hydraulic conductivity model [29] are presented as:

$$\theta(h) = \begin{cases} \theta_r + \frac{\theta_s - \theta_r}{[1 + |\alpha h|^n]^m} & h < 0 \\ \theta_s & h \geq 0 \end{cases} \quad (5)$$

$$K = K_s S_e^L \left[ 1 - \left( 1 - S_e^{\frac{1}{m}} \right)^m \right]^2 \quad (6)$$

where  $K_s$  ( $L T^{-1}$ ) is the saturated hydraulic conductivity,  $S_e$  is the effective saturation,  $S_e = (\theta - \theta_r) / (\theta_s - \theta_r)$ ,  $\theta_s$  and  $\theta_r$  ( $L^3 L^{-3}$ ) are the residual and saturated water content, respectively, and  $L$ ,  $\alpha$ ,  $n$ , and  $m$  are four independent parameters.

As shown in Equations (2) and (4), soil water movement and soil heat transport are described by physical models. We used the HYDRUS-1D model in the direct mode with the measured/empirical parameters, including  $\alpha$ ,  $n$ ,  $m$ ,  $\theta_s$ ,  $\theta_r$ ,  $L$ ,  $K_s$ ,  $\Delta z$ ,  $b_1$ ,  $b_2$ ,  $b_3$ ,  $\beta$ ,  $\beta_T$ ,  $S$ ,  $C_p$ ,  $C_w$ ,  $q_L$ , and the soil temperature data for upper and lower boundaries (as shown in Table 1), to simulate and predict the soil temperature and volumetric water content.

### 3.2. Random Forest

In machine learning, the ensemble learning method combined with decision trees comprises the famous random forest model [30]. Breiman, 1996 [31] conducted the bagging method, which is divided into the random selection, classifier training, and majority vote steps, to randomly select subsets by randomizing the training data sets. Breiman, 2001 [30] proposed a random forest model by establishing decision trees that minimized the variance for each subset. The random forest model via the bagging method can effectively improve accuracy and build a mathematical regression model of training data to predict or make decisions without realizing the real physical mechanism. More detailed information on the random forest model can be found in Breiman, 2001 [30].

The use of the random forest model includes two parts: random forest modeling and a simulation or prediction based on the random forest classifier from the modeling. A random forest regression model is an ensemble classifier composed of a set of decision tree classifiers [30]. We applied the bagging method to establish a set of trained classifiers and then classified new data points by taking a vote, which could be weighted on their prediction or decision [32]. By voting and randomly selecting and thus establishing powerful decision trees controlling correlations, the random forest model can thoroughly cover the trained set. The random forest classifier is then verified or calibrated by using the test set.

In this study, we used the scikit-learn package [33] in the Python language to establish the random forest regression model with the training set of 2000 h. There are three adjustable parameters that affect modeling and must be set in advance: the number of decision trees ( $n\_estimators$ ), the maximum depth of each tree ( $max\_depth$ ), and the minimum number of samples required to split an internal node ( $min\_samples\_split$ ). We fixed the number of trees to 100 as a compromise between accuracy and efficiency. The maximum depth of trees was set to unlimited, which meant that the nodes would expand until all leaves contained less than  $min\_samples\_split$  samples. The third parameter was set to 2 ( $min\_samples\_split = 2$ ), which controlled a subset that was the minimum number of samples for each split.

For simulating and predicting soil temperature, the current air temperature data, the air temperature at the previous time point (1 h earlier), and the soil temperature at the previous time point (1 h earlier) from the training set acted as the input layer and the current soil temperature data was the output layer. For soil volumetric water content, the current air and soil temperature, the air and soil temperature at the previous time point (1 h earlier), the volumetric water content at the previous time point (1 h earlier), and the volumetric water content at the previous two-time points (2 h earlier) acted as the input layer, while the current volumetric water content was the output layer. These inputs and outputs of the random forest model are shown in Table 1. The layers were calculated by using the hidden layer established by the regression tree. Then, the test set was used to verify the random forest classifier. We modeled the prediction method of soil temperature and volumetric water content for tomato growth by the random forest model with the training and test set. Furthermore, this prediction method was used to predict soil temperature and volumetric water content from weather forecast data.

### 3.3. Inferring Connections of Networks (ICON)

ICON assumes that there are dynamic interactions between units within a complex network. Each unit in the network follows the dynamical law, including the effects of self-dynamics and those from other units [18]:

$$\dot{x}_i(t) = f(x_i) + \sum_{\substack{j=1 \\ j \neq i}}^N K_{ij}(x_i, x_j), \quad i = 1, \dots, N \quad (7)$$

where  $x_i(t)$  is the vector of state of unit  $i$  at time  $t$ ,  $f$  represents baseline dynamics,  $K_{ij}$  is a coupling function between unit  $i$  and  $j$ , and  $N$  is the number of units. Both  $f$  and  $K_{ij}$  are time series functions, specifically, which can be constituted by orthonormal bases of the individual truncated series [18]. Thus, Equation (7) becomes:

$$\dot{x}_i(t) = \sum_{k=1}^M a_k Q_k(x_i) + \sum_{\substack{j=1 \\ j \neq i}}^N \sum_{k=1}^r \sum_{l=1}^r b_{ij}^{kl} P_k(x_i) P_l(x_j), \quad i = 1, \dots, N \quad (8)$$

where  $a_k$  and  $b_{ij}^{kl}$  are the scalar coefficients,  $\{Q_k(x_i)\}_{k=1}^{\infty}$  and  $\{P_k(x_i)\}_{k=1}^{\infty}$  constitute orthonormal bases of the respective function spaces containing  $f$  and  $K_{ij}$ ,  $M$  is the number of data points in the time series,



and  $r$  value is the largest order of the Fourier series. By using the orthonormal basis representation of Equation (8), the complex nonlinear topological estimation of each unit  $i$  can be converted into a typical linear inverse problem:

$$\min \|y^{(i)} - A^{(i)}z^{(i)}\| \tag{9}$$

where  $y^{(i)}$  is the data vector,  $A^{(i)}$  is a matrix composed of the orthonormal bases, and  $z^{(i)}$  is the coefficient vector.

In this study, we established an ICON framework without presumptions on the various impact factors to describe the complex interactions, which may be a nonlinear dynamical relationship consisting of air temperature, soil temperature, volumetric water content, and plants at different temporal scales. Then, the dynamic interactions between units and units were expressed as:

$$\begin{cases} \dot{x}_1(t) = f(x_1) + K_{12}(x_1, x_2) + K_{13}(x_1, x_3) \\ \dot{x}_2(t) = f(x_2) + K_{21}(x_2, x_1) + K_{23}(x_2, x_3) \\ \dot{x}_3(t) = f(x_3) + K_{31}(x_3, x_1) + K_{32}(x_3, x_2) \end{cases} \tag{10}$$

We applied the Fourier series to the orthogonal basis of the coupling function of the ICON. Then, Equation (8) is rewritten as:

$$\begin{aligned} \dot{x}_i(t) = & \sum_{k=1}^M [a_i^k \sin(kx_i) + b_i^k \cos(kx_i)] \\ & + \sum_{\substack{j=1 \\ j \neq i}}^N \sum_{k=1}^r \xi_{ij}^k [\sin(kx_i) \sin(kx_j)] + \eta_{ij}^k [\cos(kx_i) \cos(kx_j)], \quad i = 1, \dots, N \end{aligned} \tag{11}$$

where  $a_i^k, b_i^k, \xi_{ij}^k$ , and  $\eta_{ij}^k$  are the coefficients of the Fourier series. In the simulation stage, where  $N = 3$ ,  $x_1, x_2$ , and  $x_3$  are the soil temperature, air temperature, and volumetric water content, respectively. In the prediction stage, which is combined with the weather forecast,  $N = 4$ , and  $x_4$  is the outdoor air temperature of the weather forecast.  $A^{(i)}$  and  $z^{(i)}$  are expressed as:

$$A^{(i)} = \left[ L_1^{(i)} \mid \dots \mid L_m^{(i)} \mid G_1^{(i)} \mid \dots \mid G_m^{(i)} \mid S_1^{(i)} \mid \dots \mid S_r^{(i)} \mid C_1^{(i)} \mid \dots \mid C_r^{(i)} \right] \tag{12}$$

$$z^{(i)} = \left( z_{L_1}^{(i)}, \dots, z_{L_m}^{(i)}, z_{G_1}^{(i)}, \dots, z_{G_m}^{(i)}, z_{S_1}^{(i)}, \dots, z_{S_r}^{(i)}, z_{C_1}^{(i)}, \dots, z_{C_r}^{(i)} \right) \tag{13}$$

where:

$$\begin{aligned} L_k^{(i)} &= \begin{bmatrix} \sin(kx_i(t_1)) \\ \vdots \\ \sin(kx_i(t_M)) \end{bmatrix}, \quad G_k^{(i)} = \begin{bmatrix} \cos(kx_i(t_1)) \\ \vdots \\ \cos(kx_i(t_M)) \end{bmatrix}, \\ S_k^{(i)} &= \begin{bmatrix} \sin(kx_i(t_1)) \sin(kx_1(t_1)) & \dots & \sin(kx_i(t_1)) \sin(kx_N(t_1)) \\ \vdots & \ddots & \vdots \\ \sin(kx_i(t_M)) \sin(kx_1(t_M)) & \dots & \sin(kx_i(t_M)) \sin(kx_N(t_M)) \end{bmatrix}, \\ C_k^{(i)} &= \begin{bmatrix} \cos(kx_i(t_1)) \cos(kx_1(t_1)) & \dots & \cos(kx_i(t_1)) \cos(kx_N(t_1)) \\ \vdots & \ddots & \vdots \\ \cos(kx_i(t_M)) \cos(kx_1(t_M)) & \dots & \cos(kx_i(t_M)) \cos(kx_N(t_M)) \end{bmatrix}. \end{aligned}$$

The training set, where  $M = 2000$ , was used to build  $x_i(t_1)$  to  $x_i(t_M)$  and fit the optimum  $r$  in our ICON model. The inputs and outputs of the ICON model are shown in Table 1. The test set was used to verify the simulation results from the built ICON model. Because the ICON model was based on the dynamic interactions between factors that affected each other, the simulation for the soil temperature and volumetric water content will be discussed individually in Section 4.3. In the

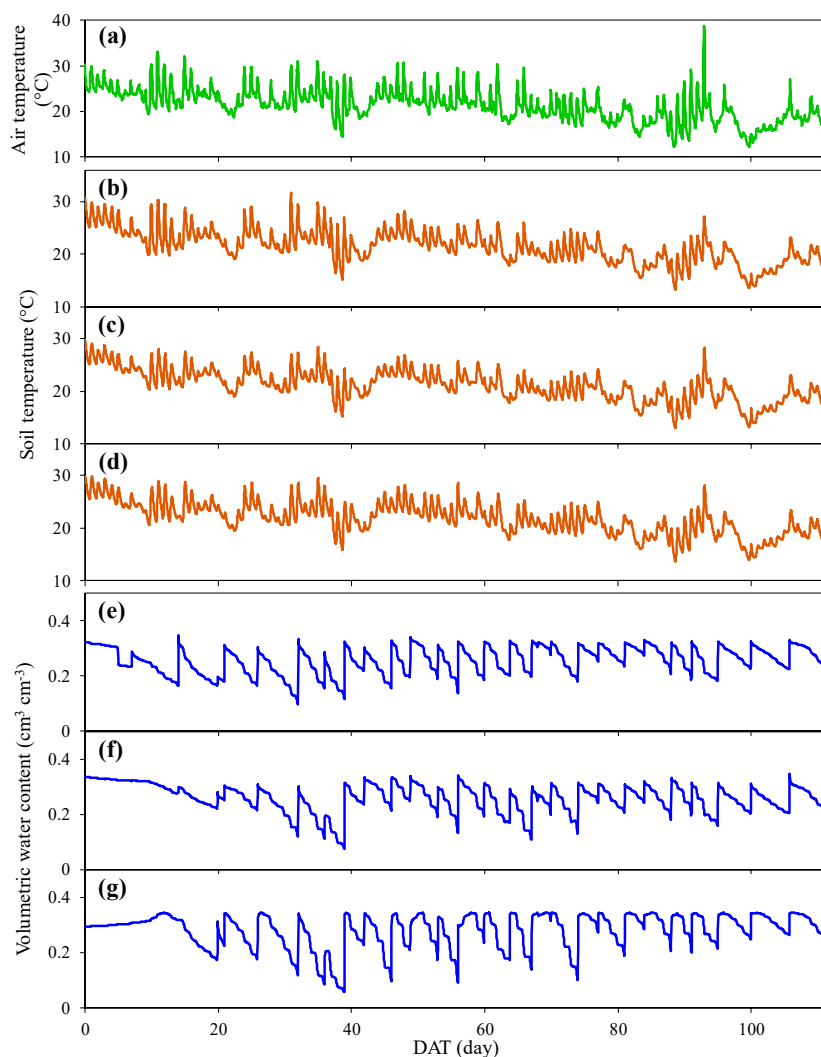
prediction stage, the weather forecast data were added as a unit, and then the ICON model was rebuilt. Finally, we predicted the air temperature, soil temperature, and volumetric water content based on the 1–48 h future weather forecast.

## 4. Results and Discussion

### 4.1. Temperature

#### 4.1.1. Temporal Distributions of Air Temperature ( $T_a$ ) and Soil Temperature ( $T_s$ )

Figure 3a shows the temporal distribution of measured air temperature in the greenhouse during the cultivation period. The observation period of Figure 3a was from September 20, 2018 to January 9, 2019, and the days after transplanting totaled 112. The average air temperature was 25 °C during the cultivation period. The air temperature showed a tendency of daily temperature oscillation. The maximum air temperature observed was 38.8 °C, which occurred at noon on 93rd DAT, and the minimum temperature was 12.1 °C, which occurred at night on 100th DAT. The air temperatures of the greenhouse are easily affected by outdoor temperatures. In this study, these measured air temperatures were used as input data for simulating and predicting soil temperatures and volumetric water contents.



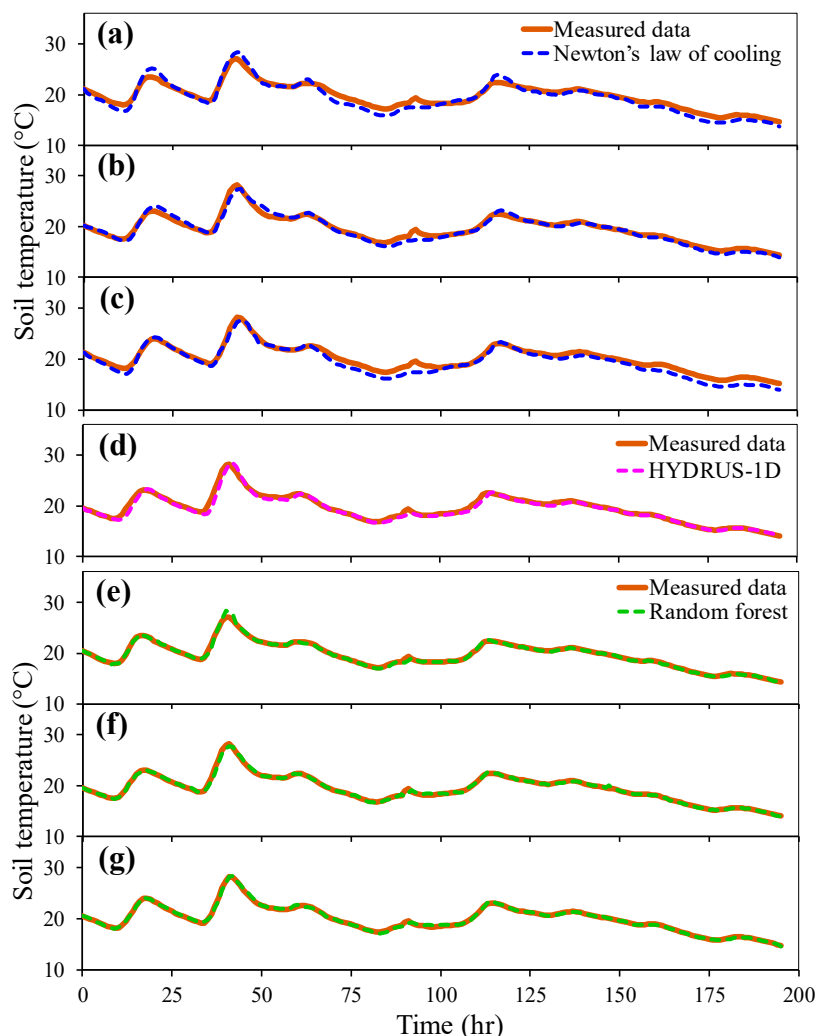
**Figure 3.** Temporal distributions of measured (a) air temperature, (b–d) soil temperature, and (e–g) volumetric water content for baskets (b,e) No. 12, (c,f) No. 13, and (d,g) No. 14 in the greenhouse during the cultivation period (20 September 2018 to 9 January 2019). The DAT of the X-axis represents the days after transplanting.

Figure 3b–d shows the temporal distribution of measured soil temperatures of the three culture substrates (basket No. 12, No. 13, and No. 14) in the greenhouse during the cultivation period. Similar trends were observed between these three culture substrates. The soil temperatures were between 13 °C and 31.7 °C, and the average soil temperature was 21.5 °C. The maximum soil temperature of the three culture substrates occurred on the 31st DAT, and the air temperature also reached a relatively high temperature of 30.1 °C. Moreover, the minimum soil temperature occurred on the 89th DAT, and the air temperature was a relatively low temperature of 12.3 °C. The maximum or minimum soil and air temperatures occurred on similar days. This indicates a clear relationship between the soil and the air temperatures [34]. We established the relationship between the air and soil temperature by using Newton’s law of cooling, HYDRUS-1D with measured/empirical parameters, the random forest model with training data, and the ICON model, respectively.

#### 4.1.2. Simulation and Verification of Soil Temperature

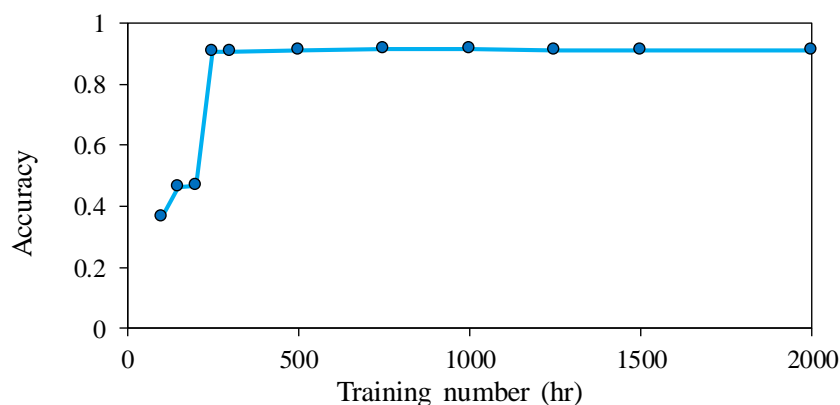
Figure 4a–c shows the soil temperature simulation by Equation (1) with  $k = -0.306 \text{ h}^{-1}$ ,  $k = -0.214 \text{ h}^{-1}$ , and  $k = -0.221 \text{ h}^{-1}$  for baskets No. 12, No. 13, and No. 14, respectively. The air and soil in the greenhouse were regarded as two individuals that transferred heat interactively, and we assumed that the temperature of soil was represented by the central temperature at the soil profile and did not change with depth. In Figure 4a–c, the orange lines are the test set of the measured soil temperature. The values observed in different baskets were slightly different, presumably because of the different volumetric water contents and the spatial distribution of the baskets. Nonetheless, the trends for the soil temperatures between the three baskets were the same. The blue dashed lines were simulated by Newton’s law of cooling (Equation (1)) with  $k$ . The simulation results were roughly consistent with the measured values, but there was still a slight difference in that the peaks deviated by approximately 1 h. The peak deviation of 1 h comes from the differentiation of the input data. The advantage of Newton’s law of cooling is that it uses a single parameter to describe the relationship between air temperature and soil temperature; however, it cannot accurately represent the influence of other factors on the heat transfer, such as water content, soil structure, or soil particle arrangement.

We also used HYDRUS-1D and the measured/empirical parameters to simulate soil temperatures and compare them with the measured data. Since measured data for soil temperature in the upper and lower boundaries were only available in basket No. 13, it will be the only basket discussed for the simulation by HYDRUS-1D in this section. Figure 4d shows the soil temperature simulation by HYDRUS-1D with the measured/empirical parameters. The magenta dashed line was simulated by HYDRUS-1D. The result of the simulation was more accurate than that of Newton’s law of cooling. Specifically, in the interval of 80 h to 100 h, the deviation was smaller than Newton’s law of cooling. When all the required parameters of the HYDRUS-1D model can be provided, the physical model should become efficient enough to predict soil temperature and water content without having to collect long-term monitoring data. Moreover, the setup of the HYDRUS-1D model can also be easily transferred or extrapolated to represent other farms with different environmental conditions. In fact, without considering the effect of crops, the prediction from HYDRUS-1D was already close to the observation initially. However, the HYDRUS-1D simulation results still had a peak deviation of 1 h. Using HYDRUS-1D requires many measured parameters that must be prepared in advance, which means that we would need further experimental analysis and additional instruments or measurements to obtain parameters. Moreover, the complicated interaction between crops and the soil temperature and water content is not fully considered in most of the physical models.



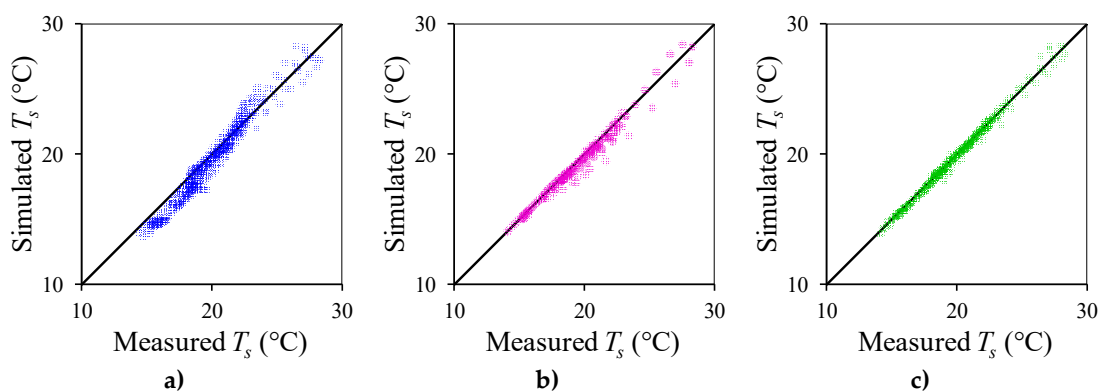
**Figure 4.** Soil temperature simulation of Newton's law of cooling for baskets (a) No. 12  $k = -0.306 \text{ h}^{-1}$ , (b) No. 13  $k = -0.214 \text{ h}^{-1}$ , and (c) No. 14  $k = -0.221 \text{ h}^{-1}$ . Soil temperature simulation by HYDRUS-1D with the measured/empirical parameters in basket (d) No. 13. Soil temperature simulation of the random forest model for baskets (e) No. 12, (f) No. 13, and (g) No. 14. The orange lines are the test set of the measured soil temperature.

For the random forest model, we first tested the relationship between the number of trainings and simulation accuracy, as shown in Figure 5. The accuracy is calculated by dividing the correct number of the random forest model into the total number of trainings. The accuracy exceeded 0.9 with more than 250 training numbers in our case. More accurate simulations can be achieved when the training numbers cover all possible scenarios (e.g., irrigation events), although this accuracy depends on the training numbers themselves [13]. To cover all the scenarios during the cultivation period and compare them with the other models, we used the same 2000-h training set to establish the random forest regression model. Figure 4e–g shows the simulation results (green dashed lines) of soil temperature by the random forest model. The result is a useful description for the measured data (orange lines). Compared with Figure 4b,d, the simulation result (Figure 4f) of the random forest model is more accurate, and the peak deviation is negligible. The random forest model can automatically learn the relationship between the air and soil temperatures through these training data, including the time lag; thus, the simulated soil temperature can be in close accordance with the measured value, which is different from the aforementioned physical models. However, the weakness of the machine learning technique is that it is impossible to simulate or predict the soil temperature directly from the air temperature without training data. A large amount of training data is required to support the model's simulations or predictions [13].



**Figure 5.** Simulation accuracy of the random forest model with various training numbers in our study. The accuracy was calculated by dividing the correct number of random forest model into the total number of trainings. The accuracy exceeded 0.9 with more than 250 training numbers.

Figure 6 shows a comparison of the measured and the simulated soil temperature as determined by Newton’s law of cooling, HYDRUS-1D, and the random forest model. The circles in Figure 6c overlap on the straight line better than Figure 6a,b, to verify that the random forest model simulations are more accurate than those determined by Newton’s law of cooling and HYDRUS-1D. The root mean square error (RMSE) and Nash-Sutcliffe model efficiency coefficient (NSE) between the measured soil temperature and the simulated soil temperature was also calculated to compare the simulation performances of these models. The RMSEs of soil temperatures for Newton’s law of cooling, HYDRUS-1D, and the random forest model were  $0.763 \pm 0.133$  °C,  $0.469$  °C, and  $0.201 \pm 0.020$  °C, respectively. The NSEs of soil temperatures as per Newton’s law of cooling, HYDRUS-1D, and the random forest model were  $0.905 \pm 0.033$  °C,  $0.970$  °C, and  $0.994 \pm 0.001$  °C, respectively. The results are shown in Table 2.



**Figure 6.** Comparison of the measured and the simulated soil temperature by (a) Newton’s law of cooling, (b) HYDRUS-1D, and (c) the random forest model. The straight slash is a 1:1 line.

**Table 2.** Root mean square errors (RMSEs) and Nash-Sutcliffe model efficiency coefficients (NSEs) of the model simulations.

Models	RMSE		NSE	
	Soil Temperature (°C)	VWC (cm <sup>3</sup> cm <sup>-3</sup> )	Soil Temperature (°C)	VWC (cm <sup>3</sup> cm <sup>-3</sup> )
Newton’s law of cooling	$0.763 \pm 0.133$	-	$0.905 \pm 0.033$	-
HYDRUS-1D	0.469	0.024	0.970	0.626
Random forest	$0.201 \pm 0.020$	$0.008 \pm 0.001$	$0.994 \pm 0.001$	$0.961 \pm 0.014$
ICON	$0.206 \pm 0.006$	$0.008 \pm 0.001$	$0.994 \pm 0.001$	$0.962 \pm 0.004$

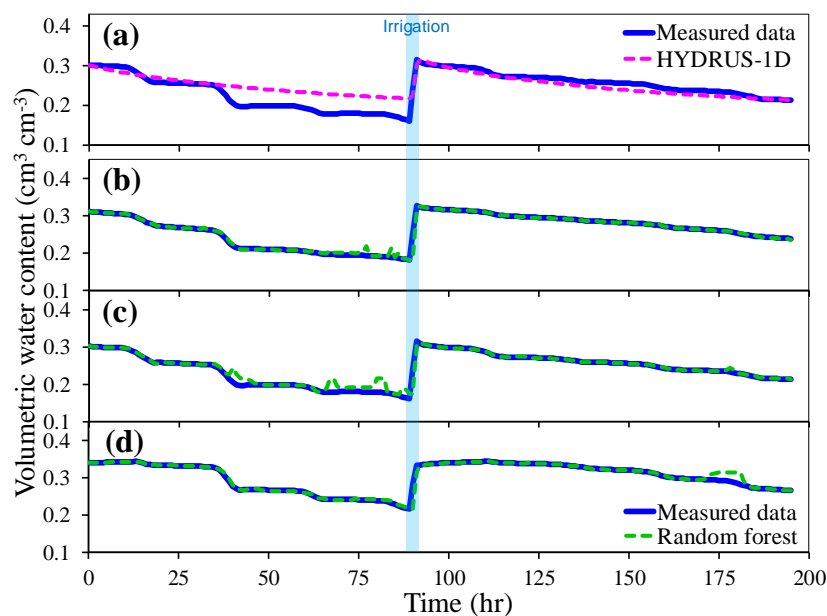
## 4.2. Volumetric Water Content

### 4.2.1. Volumetric Water Content During the Cultivation Period

Figure 3e–g shows the temporal distribution of measured soil volumetric water contents in three baskets during the cultivation period, which was measured from September 20, 2018 to January 9, 2019 (DAT = 112 days). These three baskets did not undergo water-saving treatments but were manually irrigated with the same frequency. In Figure 3e, the measured data was abnormal during the DAT period of 6 to 7 due to a problem with the moisture sensor in basket No. 12, and the sensor was immediately rechecked and relocated.

### 4.2.2. Simulation and Verification of Volumetric Water Content

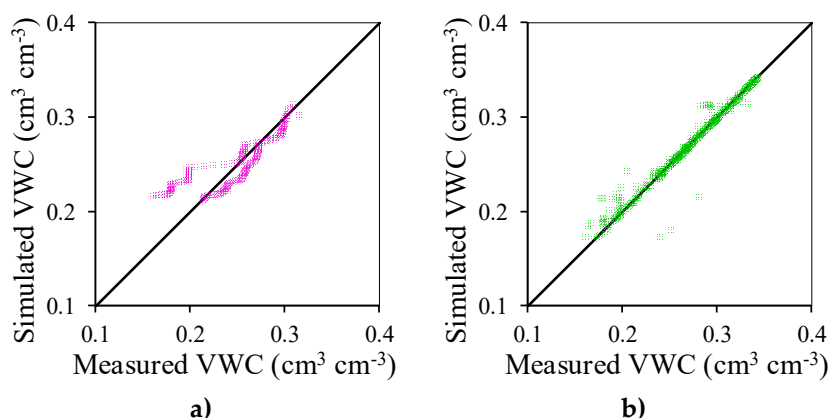
Figure 7a shows the simulation result of soil volumetric water content by HYDRUS-1D with the measured/empirical parameters in basket No. 13. The blue line was the test set of the measured volumetric water content, and the magenta dashed line was simulated by HYDRUS-1D. At the 89th hour of the irrigation event, the sudden increase in the simulated water content was due to the boundary conditions containing the information about the changes in matric potential. Overall, the simulation result after 89 h was closer to the measured values than before 89 h, and the measured volumetric water contents before 89 h were much lower than the simulated values. Due to the effect of high temperature from the 38th to the 41st hour (air temperature = 28.5–30.8 °C, soil temperature = 25.2–28.3 °C), the accelerated evapotranspiration caused the soil water content to decrease rapidly, which deviated from the ideal simulation by the model, and the water content did not increase until the irrigation event at the 89th hour. Moreover, the measured volumetric water contents (blue line) had a significant one-step declining trend with time. Due to the rotation between day and night, plant evapotranspiration was more pronounced during the day, causing a steeper slope; at night, photosynthesis stopped, thus, water consumption decreased and the decline in the volumetric water content curve was gentler. This situation cannot be simulated by HYDRUS-1D. As a result, any influencing factors will reflect the observed soil temperature and soil water content, which increases the difficulty of the model's ability to simulate soil temperature and soil water content.



**Figure 7.** Soil volumetric water content simulation by HYDRUS-1D with the measured/empirical parameters in basket (a) No. 13. Soil volumetric water content simulation of the random forest model for baskets (b) No. 12, (c) No. 13, and (d) No. 14. The blue lines are the test set of the measured soil volumetric water content.

Figure 7b–d shows the simulation results of soil volumetric water content by the random forest model in the three baskets. The blue lines are the test set of the measured volumetric water content. The green dashed lines were simulated by the random forest model. The simulation accuracy of the random forest model was higher than that of HYDRUS-1D. Specifically, the increase in volumetric water content caused by irrigation events during the training stage was included in the training set. These events were learned by the random forest model with the input features, especially for the changes in the volumetric water content of the previous time point (1 h earlier) and the volumetric water content at the previous two-time points (2 h earlier). Therefore, the simulated values (green dashed lines) show a sudden increase in the figure (e.g., Figure 7b 70–89 h; 7c 35–89 h; and 7d 170–180 h). This means that the random forest model has the potential to suggest irrigation needs when the volumetric water content continues to decrease.

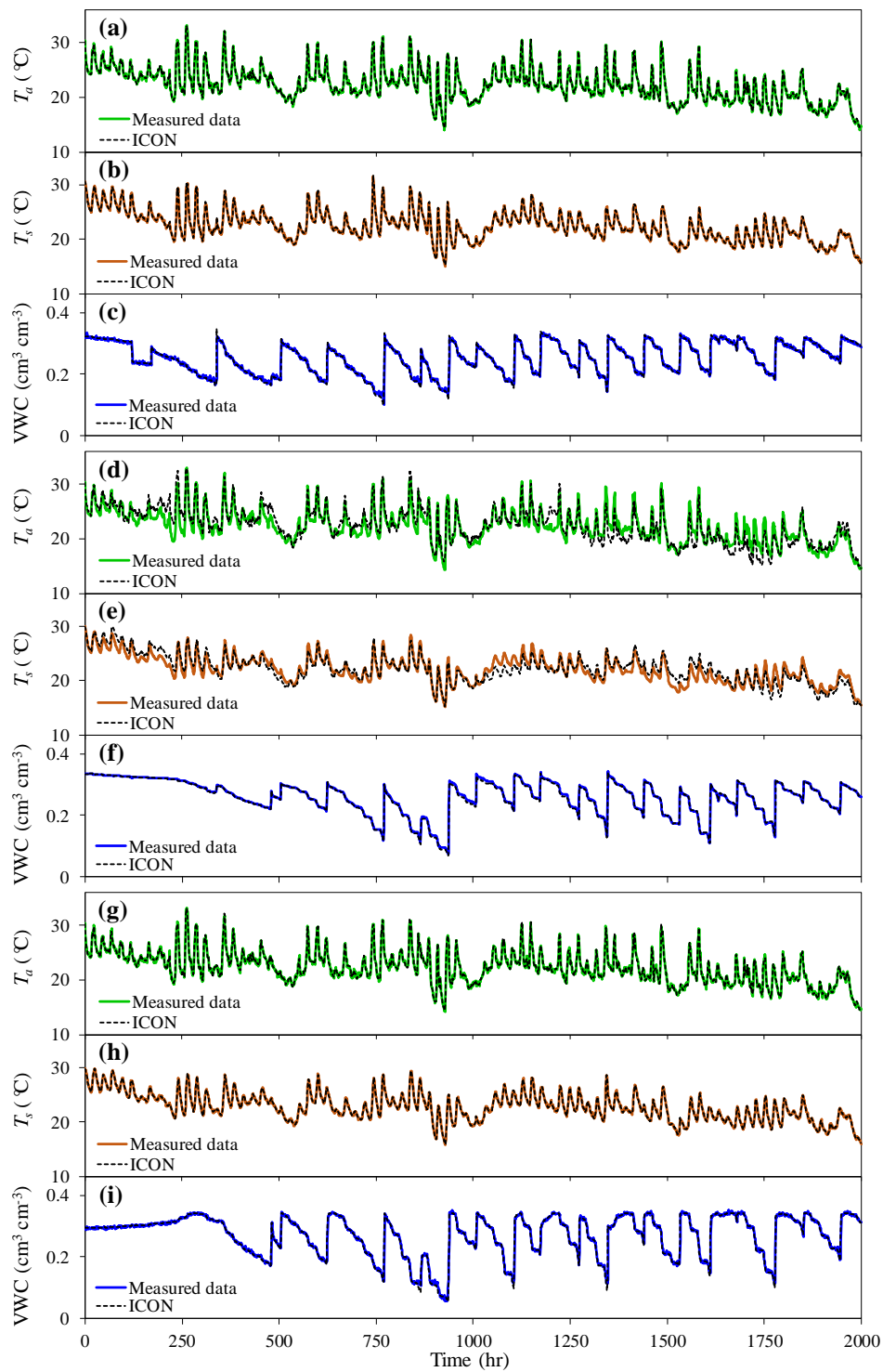
Figure 8 shows the comparison of the measured and simulated soil volumetric water content by HYDRUS-1D and the random forest model. Compared with Figure 8a, the green circles of the random forest model in Figure 8b overlap better on the straight line. Moreover, Table 2 shows that the RMSEs of soil volumetric water content for HYDRUS-1D and the random forest model are  $0.024 \text{ cm}^3 \text{ cm}^{-3}$  and  $0.008 \pm 0.001 \text{ cm}^3 \text{ cm}^{-3}$ , along with the NSEs of  $0.626 \text{ cm}^3 \text{ cm}^{-3}$  and  $0.961 \pm 0.014 \text{ cm}^3 \text{ cm}^{-3}$ , respectively. The simulation performance of the random forest model on the soil volumetric water content is more accurate than that of HYDRUS-1D.



**Figure 8.** Comparison of the measured and the simulated soil volumetric water content by (a) HYDRUS-1D and (b) the random forest model.

#### 4.3. ICON Simulation Based on Interactions Between Air Temperature, Soil Temperature, and Volumetric Water Content

Figure 9 shows the establishment of the ICON model for baskets No. 12, No. 13, and No. 14 via the interactions between the soil temperature, air temperature, and volumetric water content, which affect each other dynamically. The green, orange, and blue lines are the measured data of the training set. The fitted curves (black dashed lines) with the largest orders of Fourier series were as follows:  $r = 333, 285,$  and  $318$  for the soil temperature,  $r = 333, 282,$  and  $325$  for the air temperature, and  $r = 37, 48,$  and  $31$  for the volumetric water contents in baskets No. 12, No. 13, and No. 14, respectively. Larger  $r$  values more accurately depict real dynamic interactions [18]. Depending on the complexity of the patterns in the training set, the  $r$  values of air temperature and soil temperature were much larger than that of the volumetric water content.

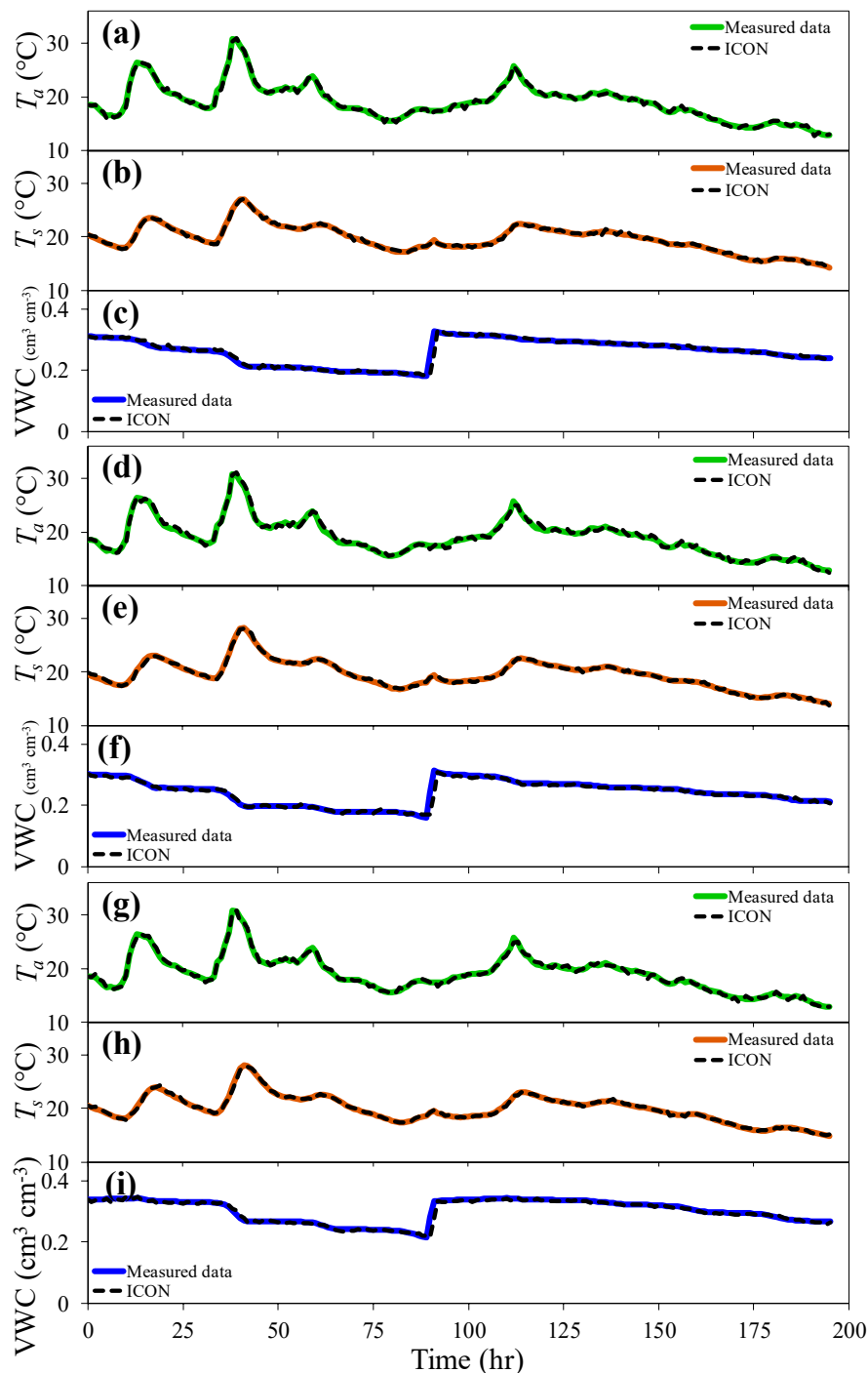


**Figure 9.** The interaction factors used to establish the ICON model with the training set: (a,d,g) air temperature ( $T_a$ ), (b,e,h) soil temperature ( $T_s$ ), and (c,f,i) volumetric water content (VWC) in baskets (a–c) No. 12, (d–f) No. 13, and (g–i) No. 14.

Figure 10a–i shows that the simulation results (black dashed lines) of the ICON model, which was established by the aforementioned fitting, compared with the test set (green, orange, and blue lines) of the measured soil temperature, air temperature, and volumetric water content. Among these three factors, the volumetric water content has the most accurate simulation result, and the soil temperature and air temperature are also in accordance. Although the simulation results have



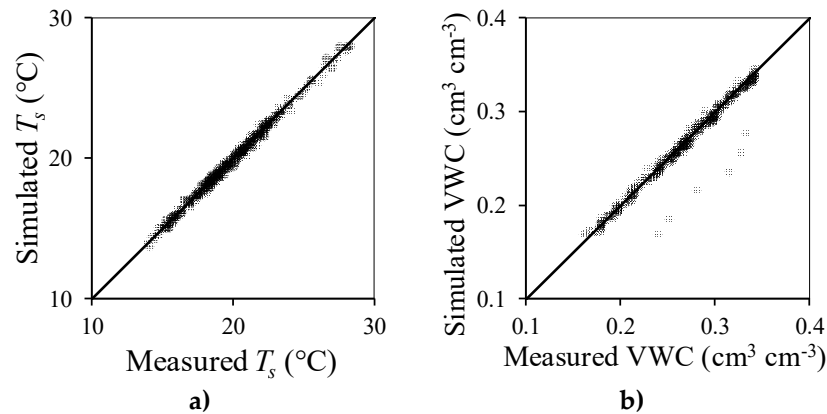
some small fluctuations, especially in the simulation result of air temperature, these small fluctuations fortunately do not affect the simulation trend.



**Figure 10.** Simulation results of the ICON model in baskets (a–c) No. 12, (d–f) No. 13, and (g–i) No. 14 for (a,d,g) air temperature ( $T_a$ ), (b,e,h) soil temperature ( $T_s$ ), and (c,f,i) volumetric water content (VWC).

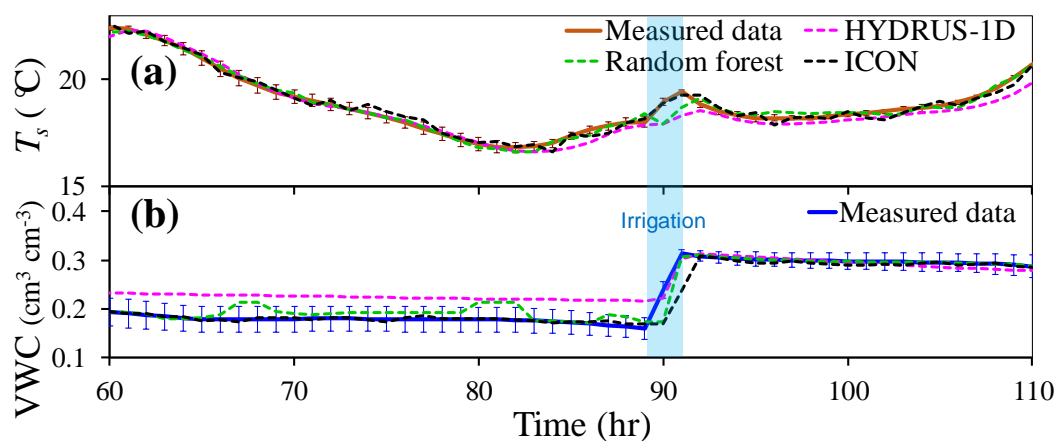
Figure 11 shows the comparison of the measured and the simulated values by the ICON model for soil temperature and volumetric water content. The RMSE of soil temperature for the ICON model is  $0.206 \pm 0.006$  °C, and the NSE is  $0.994 \pm 0.001$  °C (as shown in Table 2). The simulation performance of the ICON model for soil temperature is more accurate than that for Newton’s law of cooling (RMSE =  $0.763 \pm 0.133$  °C, NSE =  $0.905 \pm 0.033$  °C) and HYDRUS-1D (RMSE =  $0.469$  °C, NSE =  $0.970$  °C); similar to the random forest model (RMSE =  $0.201 \pm 0.020$  °C, NSE =  $0.994 \pm 0.001$

°C). The RMSE of the volumetric water content for the ICON model is  $0.008 \pm 0.001 \text{ cm}^3 \text{ cm}^{-3}$ , and the NSE is  $0.962 \pm 0.004 \text{ cm}^3 \text{ cm}^{-3}$ . The simulation performance of the ICON model for volumetric water content is more accurate than HYDRUS-1D (RMSE =  $0.024 \text{ cm}^3 \text{ cm}^{-3}$ , NSE =  $0.626 \text{ cm}^3 \text{ cm}^{-3}$ ) and similar to the random forest model (RMSE =  $0.008 \pm 0.001 \text{ cm}^3 \text{ cm}^{-3}$ , NSE =  $0.961 \pm 0.014 \text{ cm}^3 \text{ cm}^{-3}$ ).



**Figure 11.** Comparison of the measured and the simulated values by the ICON model for (a) soil temperature and (b) volumetric water content.

Figure 12 shows the simulation performance of the HYDRUS-1D, random forest model, and ICON model on the measured data with error bars, which represent the standard deviation of the measured data. There was an irrigation event at the 89th hour, as represented in the figure via blue highlights. The simulation results of the random forest model and ICON model for the soil temperature were consistent with the measured values (as shown in Figure 12a). However, during the irrigation event, only the ICON model was observed to respond to the changes in soil temperature that were caused by irrigation. In Figure 12b, both the random forest model and ICON model showed good simulation performance for volumetric water content and were consistent with the measured values. Contrastingly, the prediction of VWC from HYDRUS-1D before the 90th hour was higher than the measured data since HYDRUS-1D was not able to simulate the strong evapotranspiration due to high air temperatures.



**Figure 12.** Simulation performance of models (HYDRUS-1D, random forest model, and ICON model) as compared to the measured data with error bars for (a) soil temperature and (b) volumetric water content. Error bars represent the standard deviation of the measured data.

#### 4.4. Prediction of Soil Temperature and Volumetric Water Content from the Air Temperature of the Weather Forecast

In this study, we proposed a framework to predict the soil temperature and volumetric water content in a non-temperature-controlled greenhouse via applying HYDRUS-1D, the random forest

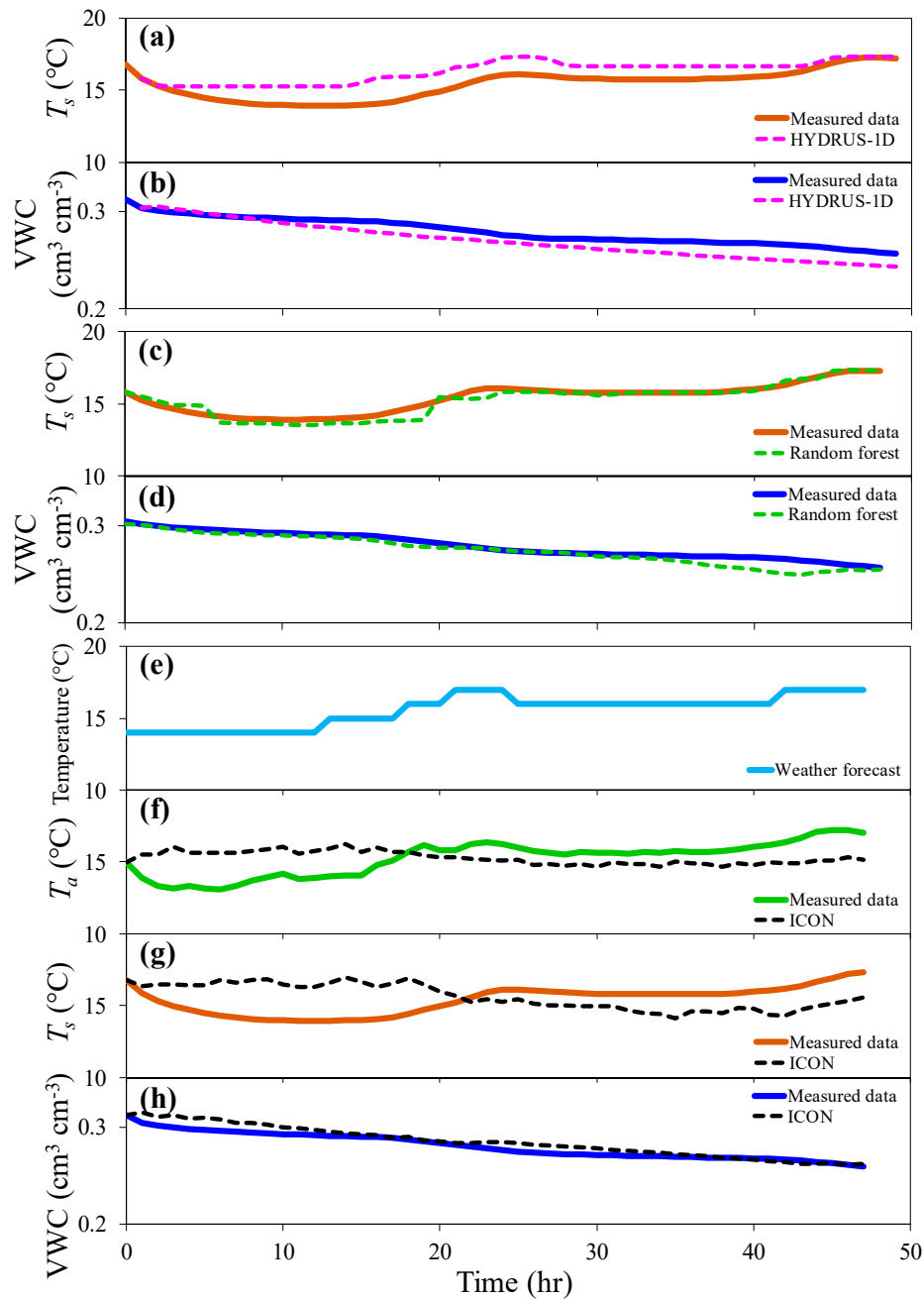
model, and the ICON model with the weather forecast data. The combined forecast data were used in these prediction models to obtain the changes in soil temperature and volumetric water content over time. The source of weather forecast data (from the Central Weather Bureau, Taipei, Taiwan) is a rolling forecast of hourly outdoor air temperature for the future 48 h, thus, the prediction stage was limited to 1–48 h.

Since the physical equations in HYDRUS-1D were not related to the forecasted outdoor air temperature, we performed a linear regression on 2667 forecasted outdoor air temperatures to establish a conversion equation,  $T_a^{convert} = 0.91T^{forecast} + 1.20$  ( $R^2 = 0.80$ ), for the indoor air temperature. In addition, the soil temperatures of the upper and lower boundaries required for the model were converted by the linear regression equations,  $T_s^{upper} = 0.83T_a^{convert} + 3.18$  ( $R^2 = 0.86$ ) and  $T_s^{lower} = 0.75T_a^{convert} + 5.19$  ( $R^2 = 0.84$ ), with the aforementioned converted indoor air temperature from the soil temperature data. Table 3 lists the inputs and outputs of the conversion. Figure 13a,b shows the prediction results (magenta dashed lines) from the outdoor air temperature data of the weather forecast by HYDRUS-1D with the measured parameters in basket No. 13. The green, orange, and blue lines are the measured data of 48 h. Overall, the prediction results overestimated the soil temperature and underestimated the volumetric water content, but the predictions were generally in accordance with actual trends. The results deviated from the measured values because of restrictions on the model parameters. The RMSEs of the predicted soil temperature and volumetric water content were 1.006 °C and 0.011 cm<sup>3</sup> cm<sup>-3</sup>, respectively, and the NSEs were -0.020 °C and 0.342 cm<sup>3</sup> cm<sup>-3</sup>. The results are shown in Table 4.

**Table 3.** Inputs and outputs for converting the forecasted outdoor air temperature to the indoor air temperature for the studied models.

Models	Inputs	Outputs
HYDRUS-1D	For air temperature: (linear regression) forecasted outdoor air temperature: $T^{forecast}$ For soil temperature: (linear regression) $T_a^{convert}$	converted indoor air temperature: $T_a^{convert}$ soil temperature of upper and lower boundaries: $T_s^{upper}$ and $T_s^{lower}$
Random forest	For air temperature and VWC: forecasted outdoor air temperature, 24-h time, ultraviolet index (UVI)	$T_a^{convert}$
ICON	The forecasted outdoor air temperature as the fourth input factor	

For machine learning, during the prediction stage, we directly applied the random forest model to train the indoor air temperature with the forecasted outdoor air temperature. The 24-h time, forecasted outdoor air temperature, and ultraviolet index (UVI) were used as the input layer, and the indoor air temperature data was the output layer (as shown in Table 3). After the training processes, we established a conversion relationship between the weather forecast data and indoor air temperature. When the converted indoor air temperature,  $T_a^{convert}$ , was available, we applied the established random forest model to predict the soil temperature and volumetric water content. Figure 13c,d shows the prediction results (green dashed lines) from the random forest model via training data with the converted indoor air temperature from the forecasted outdoor air temperature in basket No. 13. The prediction results are in accordance with the actual measurement data, and the prediction performance of the random forest model is more accurate than HYDRUS-1D. The RMSEs of soil temperature and volumetric water content predicted by the random forest model were 0.333 °C and 0.006 cm<sup>3</sup> cm<sup>-3</sup>, respectively, and the NSEs were 0.889 °C and 0.795 cm<sup>3</sup> cm<sup>-3</sup>. The results are shown in Table 4.



**Figure 13.** Prediction results from the outdoor air temperature of the weather forecast by (a,b) HYDRUS-1D with the measured parameters, (c,d) the random forest model via the training data, and (e–h) ICON model re-established with four factors in basket No. 13 for (a,c,g) soil temperature and (b,d,h) volumetric water content.

**Table 4.** RMSEs and NSEs of model predictions.

Models	RMSE		NSE	
	Soil Temperature (°C)	VWC (cm <sup>3</sup> cm <sup>-3</sup> )	Soil Temperature (°C)	VWC (cm <sup>3</sup> cm <sup>-3</sup> )
HYDRUS-1D	1.006	0.011	-0.020	0.342
Random forest	0.333	0.006	0.889	0.795
ICON	1.701	0.006	-2.813	0.850

Finally, we collected the outdoor air temperatures of weather forecasts as an influencing factor and attached them to the ICON model. The ICON model was re-established for tomato planting with the training set via the interactions between four factors, which were soil temperature, indoor air temperature, volumetric water content, and forecasted outdoor air temperature; these factors affected each other dynamically. The largest orders of Fourier series were fitted as follows:  $r = 250$  for the soil temperature,  $r = 250$  for the air temperature, and  $r = 49$  for the volumetric water content. Figure 13e–h shows the prediction results (black dashed lines) from the forecasted outdoor air temperature of as an input factor by the re-established ICON model in basket No. 13. The RMSEs of the predicted soil temperature and volumetric water content were  $1.701\text{ }^{\circ}\text{C}$  and  $0.006\text{ cm}^3\text{ cm}^{-3}$ , respectively, and the NSEs were  $-2.813\text{ }^{\circ}\text{C}$  and  $0.850\text{ cm}^3\text{ cm}^{-3}$  (as shown in Table 4). The prediction result of the volumetric water content was consistent with the actual measurement data and showed an accurate prediction result. Unfortunately, the partial predictions of soil temperature were inferior to the physical models and random forest model. Regardless, the advantage of the ICON model is that it can extract the dynamic interactions of a large complex system with multiple factors that affect each other dynamically. The reason for the somewhat inaccurate prediction results for soil temperature may be because we used only four factors for the short-term (48 h) prediction, which may have limited the performance of this model.

## 5. Conclusions

We proposed a novel framework that applied physical models, machine learning methods, and dynamic topology to simulate and predict soil temperature and volumetric water content in a greenhouse in combination with weather forecast data. We used Newton's law of cooling, HYDRUS-1D, the random forest model, and the ICON model, to simulate and verify the measured soil temperature. HYDRUS-1D, the random forest model, and the ICON model were used to simulate and verify the measured volumetric water content. Moreover, the simulation performances of these models were compared with RMSE. The random forest model was more accurate than the other methods with limited information provided from the greenhouse experiments; this approach also has the potential to suggest irrigation regimes. Additionally, the random forest model and ICON model can use historical data to effectively simulate soil temperature and volumetric water content without physical parameters.

Our study demonstrated the capability of the proposed framework with HYDRUS-1D, the random forest model, and the ICON model to predict soil temperature and volumetric water content based on a future weather forecast of 1–48 h. With limited information, predictions on soil temperature and volumetric water content by the random forest model were more accurate than the other models. Using the three models from our proposed framework should help farmers in choosing a suitable model for their agricultural practices.

Our proposed models can assess the water requirements for agricultural practices and develop water content and temperature alarms for greenhouse management. Moreover, our approach can collect soil and water information of the best practices determined by experienced farmers. By setting the collected data as a training set, our prediction framework can not only learn best management practices from experienced farmers, but it can also provide intelligent information for smart greenhouse management.

**Author Contributions:** Conceptualization, Y.-Z.T. and S.-Y.H.; Data curation, Y.-Z.T., K.-S.H., and S.-Y.H.; Formal analysis, Y.-Z.T. and K.-S.H.; Funding acquisition, H.-L.Y., K.-T.H., M.-C.H., and S.-Y.H.; Investigation, Y.-Z.T., K.-S.H., H.-Y.W., and S.-I.L.; Methodology, Y.-Z.T., K.-S.H., and S.-Y.H.; Project administration, S.-Y.H.; Resources, H.-Y.W., S.-I.L., K.-T.H., M.-C.H., and S.-Y.H.; Software, K.-S.H.; Supervision, H.-L.Y. and S.-Y.H.; Validation, Y.-Z.T. and S.-Y.H.; Visualization, Y.-Z.T.; Writing—original draft, Y.-Z.T.; Writing—review and editing, Y.-Z.T. and S.-Y.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Taiwan University (NTU-CC-107L892607) and Ministry of Science and Technology, Taiwan (106-2628-M-002-009-MY3).

**Acknowledgments:** The authors would like to thank Yu-Pin Lin and Hsiao-Feng Lo from National Taiwan University for the experimental design and thoughtful discussions. This research was funded by National Taiwan University, R.O.C. (Grant No., NTU-CC-107L892607) and the Ministry of Science and Technology, Taiwan, R.O.C. (MOST; Grant No., 106-2628-M-002-009-MY3). Additionally, the authors thank the editors and anonymous referees for their thoughtful comments and suggestions. We gratefully acknowledge the contribution of Yu-Che Lin, a student of National Taiwan University, for assisting in planting tomatoes. The data presented in this study are available through the corresponding author (Shao-Yiu Hsu) upon request (syhsu@ntu.edu.tw).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Shannon, H.D.; Motha, R.P. Managing weather and climate risks to agriculture in North America, Central America and the Caribbean. *Weather Clim. Extrem.* **2015**, *10*, 50–56. [\[CrossRef\]](#)
- Gebbers, R.; Adamchuk, V.I. Precision agriculture and food security. *Science* **2010**, *327*, 828–831. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhang, N.Q.; Wang, M.H.; Wang, N. Precision agriculture—A worldwide overview. *Comput. Electron. Agric.* **2002**, *36*, 113–132. [\[CrossRef\]](#)
- Duarte-Galvan, C.; Torres-Pacheco, I.; Guevara-Gonzalez, R.G.; Romero-Troncoso, R.J.; Contreras-Medina, L.M.; Rios-Alcaraz, M.A.; Mian-Amaraz, J.R. Review. Advantages and disadvantages of control theories applied in greenhouse climate control systems. *Span. J. Agric. Res.* **2012**, *10*, 926–938. [\[CrossRef\]](#)
- Ehret, D.L.; Lau, A.; Bittman, S.; Lin, W.; Shelford, T. Automated monitoring of greenhouse crops. *Agronomie* **2001**, *21*, 403–414. [\[CrossRef\]](#)
- Hodge, A.; Berta, G.; Doussan, C.; Merchan, F.; Crespi, M. Plant root growth, architecture and function. *Plant. Soil* **2009**, *321*, 153–187. [\[CrossRef\]](#)
- Al-Kayssi, A.W.; Al-Karaghoul, A.A.; Hasson, A.M.; Beker, S.A. Influence of soil moisture content on soil temperature and heat storage under greenhouse conditions. *J. Agric. Eng. Res.* **1990**, *45*, 241–252. [\[CrossRef\]](#)
- Hu, H.; Xu, L.; Wei, R.; Zhu, B. Multi-objective control optimization for greenhouse environment using evolutionary algorithms. *Sensors* **2011**, *11*, 5792–5807. [\[CrossRef\]](#)
- Philip, J.R. Evaporation, and Moisture and Heat Fields in the Soil. *J. Meteorol.* **1957**, *14*, 354–366. [\[CrossRef\]](#)
- Nassar, I.N.; Horton, R. Simultaneous transfer of heat, water, and solute in porous media: I. Theoretical development. *Soil Sci. Soc. Am. J.* **1992**, *56*, 1350–1356. [\[CrossRef\]](#)
- Saito, H.; Šimůnek, J.; Mohanty, B. Numerical analyses of coupled water, vapor and heat transport in the vadose zone. *Vadose Zone J.* **2006**, *5*, 784–800. [\[CrossRef\]](#)
- Šimůnek, J.; Šejna, M.; Saito, H.; Sakai, M.; van Genuchten, M.T. *The HYDRUS-1D Software Package for Simulating the One-Dimensional Movement of Water, Heat, and Multiple Solutes in Variably-Saturated Media*; University of California Riverside: Riverside, CA, USA, 2009; p. 296.
- Mohri, M.; Rostamizadeh, A.; Talwalkar, A. *Foundations of Machine Learning*; The MIT Press: Cambridge, MA, USA, 2012; p. 480.
- Alzubi, J.; Nayyar, A.; Kumar, A. Machine Learning from Theory to Algorithms: An Overview. *J. Phys. Conf. Series* **2018**, *1142*. [\[CrossRef\]](#)
- Ahmad, S.; Kalra, A.; Stephen, H. Estimating soil moisture using remote sensing data: A machine learning approach. *Adv. Water Resour* **2010**, *33*, 69–80. [\[CrossRef\]](#)
- Oroza, C.A.; Bales, R.C.; Stacy, E.M.; Zheng, Z.; Glaser, S.D. Long-Term Variability of Soil Moisture in the Southern Sierra: Measurement and Prediction. *Vadose Zone J.* **2018**, *17*. [\[CrossRef\]](#)
- Karandish, F.; Šimůnek, J. A comparison of numerical and machine-learning modeling of soil water content with limited input data. *J. Hydrol.* **2016**, *543*, 892–909. [\[CrossRef\]](#)
- Wang, S.; Herzog, E.D.; Kiss, I.Z.; Schwartz, W.J.; Bloch, G.; Sebek, M.; Granados-Fuentes, D.; Wang, L.; Li, J.S. Inferring dynamic topology for decoding spatiotemporal structures in complex heterogeneous networks. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 9300–9305. [\[CrossRef\]](#)
- Shamshiri, R.R.; Jones, J.W.; Thorp, K.R.; Ahmad, D.; Man, H.C.; Taheri, S. Review of optimum temperature, humidity, and vapour pressure deficit for microclimate evaluation and control in greenhouse cultivation of tomato: A review. *Int. Agrophys.* **2018**, *32*, 287–302. [\[CrossRef\]](#)
- Moreno, M.M.; Moreno, A. Effect of different biodegradable and polyethylene mulches on soil properties and production in a tomato crop. *Sci. Hortic. Amst.* **2008**, *116*, 256–263. [\[CrossRef\]](#)

21. Li, Y.; Xue, X.; Guo, W.; Wang, L.; Duan, M.; Chen, H.; Chen, F. Soil moisture and nitrate-nitrogen dynamics and economic yield in the greenhouse cultivation of tomato and cucumber under negative pressure irrigation in the North China Plain. *Sci. Rep.* **2019**, *9*, 4439. [[CrossRef](#)]
22. Chen, R.Q.; Kang, S.Z.; Hao, X.M.; Li, F.S.; Du, T.S.; Qiu, R.J.; Chen, J.L. Variations in tomato yield and quality in relation to soil properties and evapotranspiration under greenhouse condition. *Sci. Hortic. Amst.* **2015**, *197*, 318–328. [[CrossRef](#)]
23. Sun Gro Horticulture Distribution Inc. Safety Data Sheet: Sunshine® Professional Peat Moss and Peat-Based Mixes. 2015. Available online: <http://www.sungro.com/professional-product/sunshine-5-natural-organic> (accessed on 28 August 2018).
24. Besson, U. The History of the Cooling Law: When the Search for Simplicity can be an Obstacle. *Sci. Educ.* **2010**, *21*, 1085–1110. [[CrossRef](#)]
25. Šimůnek, J.; van Genuchten, M.T.; Šejna, M. Recent developments and applications of the HYDRUS computer software packages. *Vadose Zone J.* **2016**, *15*, 1–25. [[CrossRef](#)]
26. Campbell, G.S. *Soil Physics With BASIC: Transport. Models for Soil-Plant. Systems*; Elsevier: Amsterdam, The Netherlands, 1985; p. 150.
27. Chung, S.O.; Horton, R. Soil Heat and Water-Flow with a Partial Surface Mulch. *Water Resour. Res.* **1987**, *23*, 2175–2186. [[CrossRef](#)]
28. van Genuchten, M.T. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.* **1980**, *44*, 892–898. [[CrossRef](#)]
29. Mualem, Y. A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resour. Res.* **1976**, *12*, 513–522. [[CrossRef](#)]
30. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
31. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
32. Dietterich, T.G. Ensemble Methods in Machine Learning. In *Multiple Classifier Systems, MCS 2000*; Kittler, J., Roli, F., Eds.; Springer: Berlin/Heidelberg, Germany; Cagliari, Italy, 2000; pp. 1–15.
33. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
34. Perreault, S.; Chokmani, K.; Nolin, M.C.; Bourgeois, G. Validation of a soil temperature and moisture model in southern Quebec, Canada. *Soil Sci. Soc. Am. J.* **2013**, *77*, 606–617. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).