



Article

A Soft Sensor to Estimate the Opening of Greenhouse Vents Based on an LSTM-RNN Neural Network

Mounir Guesbaya ¹, Francisco García-Mañas ², Francisco Rodríguez ^{2,*} and Hassina Megherbi ³

¹ LI3CUB Laboratory, Department of Electrical Engineering, University of Biskra, BP 145 RP, Biskra 07000, Algeria

² Department of Informatics, University of Almería, CIESOL, ceiA3, E04120 Almería, Spain

³ LARHYSS Laboratory, University of Biskra, BP 145 RP, Biskra 07000, Algeria

* Correspondence: frrodrig@ual.es

Abstract: In greenhouses, sensors are needed to measure the variables of interest. They help farmers and allow automatic controllers to determine control actions to regulate the environmental conditions that favor crop growth. This paper focuses on the problem of the lack of monitoring and control systems in traditional Mediterranean greenhouses. In such greenhouses, most farmers manually operate the opening of the vents to regulate the temperature during the daytime. Therefore, the state of vent opening is not recorded because control systems are not usually installed due to economic reasons. The solution presented in this paper consists of developing a Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) as a soft sensor to estimate vent opening using the measurements of different inside and outside greenhouse climate variables as input data. A dataset from a traditional greenhouse located in Almería (Spain) was used. The data were processed and analyzed to study the relationships between the measured climate variables and the state of vent opening, both statistically (using correlation coefficients) and graphically (with regression analysis). The dataset (with 81 recorded days) was then used to train, validate, and test a set of candidate LSTM-based networks for the soft sensor. The results show that the developed soft sensor can estimate the actual opening of the vents with a mean absolute error of 4.45%, which encourages integrating the soft sensor as part of decision support systems for farmers and using it to calculate other essential variables, such as greenhouse ventilation rate.

Keywords: protected agriculture; greenhouse ventilation; machine learning; long short-term memory; virtual sensor; climate modeling



Citation: Guesbaya, M.; García-Mañas, F.; Rodríguez, F.; Megherbi, H. A Soft Sensor to Estimate the Opening of Greenhouse Vents Based on an LSTM-RNN Neural Network. *Sensors* **2023**, *23*, 1250. <https://doi.org/10.3390/s23031250>

Academic Editor: Wonsuk (Daniel) Lee

Received: 8 December 2022

Revised: 8 January 2023

Accepted: 17 January 2023

Published: 21 January 2023



Copyright: © 2023 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Nowadays, agriculture faces numerous challenges, mainly, the growth of the world population, the effects of climate change, strict market regulations, and energy inefficiency [1]. Overcoming these challenges requires the application of robust and adaptive management strategies that involve the implementation of technology at all stages of the hierarchical agricultural production system. Greenhouses are prominent facilities that can help to address such challenges [2]. Greenhouses are designed to protect crops and provide suitable environmental conditions to favor their growth. The climate inside a greenhouse can be regulated using different systems such as ventilation, heating, humidification, and CO₂ injection, among others. These actuators are necessary to regulate essential variables like air temperature, humidity, or CO₂ concentration to improve the growth of plants and fruits [3]. This requires sensors that measure the evolution of the variables of interest to provide farmers and automatic controllers with the information needed to determine when to activate greenhouse actuators to improve crop conditions or protect them.

Traditionally, sensors and automatic controllers are not installed in greenhouses due to the associated costs. Farmers are used to manually activating the actuators based on their own experience and continuous supervision of the crop and the weather. In recent years,

with the availability of low-cost devices, monitoring and control systems are more common in commercial greenhouses, mainly due to the proliferation of emerging technologies such as wireless sensor networks (WSN) [4] and the Internet of Things (IoT) [5,6], which allow farmers to remotely monitor the state of the crop in real time through mobile apps or computing platforms [7,8]. Despite the proven advantages that monitoring and control systems can offer in terms of increased crop productivity and energy efficiency [9], it is still unusual to find a high level of technology in most traditional greenhouses [10].

In this context, this paper focuses on the problem of the lack of monitoring and control systems in traditional Mediterranean greenhouses. These greenhouses operate in warmer climatic conditions and with higher values of solar radiation. Therefore, not many actuators are typically used, as the control actions are mainly focused on air temperature regulation, which directly affects crop growth [3]. In greenhouses located in warm climate zones, natural ventilation is the most widely used system to regulate air temperature during the daytime [11] due to its low cost [12]. It consists of opening and closing the greenhouse vents up to a certain point depending on the desired air temperature for the crop, allowing an exchange of the hot air leaving the greenhouse with the cooler air entering it. Most farmers manually operate the opening of the vents in these greenhouses, which means that it is not regulated by a controller. For that reason, the state of vent opening is not usually recorded, although potentiometers could be used for this purpose [13]. In this regard, it is important to note that some farmers use monitoring devices, such as commercial meteorological stations. However, these stations are equipped and designed with a limited number of sensors that only measure the evolution of the main climate variables [14]. Therefore, even with such monitoring systems, the state of vent opening is often not recorded unless a control system is used for ventilation or specific sensors are installed in the vents.

The importance of knowing the state of vent opening (generally, the state of the actuators) lies in the possibility of performing a more detailed analysis of the climatic information measured in a greenhouse, such as for modeling tasks or climate prediction. It is essential for IoT platforms as decision support systems that provide recommendations to farmers after analyzing the data measured by the stations installed in greenhouses [8]. Such recommendations are based on predictive models of greenhouse climate and crop growth [3,15], which require as input the state of vent opening to quantify the relevant effect of the ventilation flux (airflow rate) on the inside climate and crop [16,17]. In addition, continuous recording of the state of vent opening could also be important for legal aspects. For example, greenhouse structures can be damaged on stormy days by heavy wind and rain, and the recorded information on vent opening may be necessary for insurance claims.

The solution proposed in this paper consists of developing a soft sensor to estimate the state of vent opening when it is not measured in a greenhouse. Soft sensors, also known as virtual sensors, are useful tools for estimating a variable from the measurements provided by other physical sensors installed in a given system [18]. There are numerous examples of soft sensors applied to different fields, and in particular, some studies were applied to greenhouses, such as to design irrigation controllers or estimate the leaf area index of the crop [19–21]. Soft sensors can be classified into two categories: model-based and data-based. Model-based soft sensors use mathematical expressions that represent the relationship between the variable to be estimated and the other variables that are measured with physical sensors. For data-based soft sensors, the relationships between the estimated and measured variables can be determined using identification techniques and statistical or machine learning (ML) methods [22].

To the best of our knowledge, no previous work has been published on the estimation of the vent opening of greenhouses because it is assumed to be a measurable variable. This work is the first attempt to implement a tool for this purpose. In this sense, there are no mathematical models that directly relate the greenhouse climate to the opening of the vents. Instead, this relationship has been studied in the literature for calculating the ventilation flux, which is the airflow that circulates through the vents when they are open [16,23]. There are some well-known models for the calculation of the ventilation flux [24–26] that use

mathematical expressions fed with the state of vent opening measured by physical sensors installed in the vents (e.g., potentiometers). These models consist of nonlinear equations with parameters that must be calibrated for different types of greenhouses. Therefore, using these models to develop a soft sensor for vent opening would be complicated. If the equations of the cited models are inverted, the vent opening could be estimated, but then actual measurements of the ventilation flux would be obligatory, which would require installing expensive sensors, such as sonic anemometers [27].

For the reasons explained above, this work aims to develop a data-driven soft sensor to estimate vent opening. To design the soft sensor, Deep Learning (DL) methods [28] are known to be a suitable option due to the satisfactory results presented in other applications for greenhouses [29]. In particular, neural networks based on Long Short-Term Memory (LSTM) are of particular interest for the described problem, considering the complexity and nonlinearity of the dynamics involved in greenhouses [30]. Consequently, an LSTM-based network has been selected for the implementation of the soft sensor due to its powerful advantages and ability to deal with the vanishing gradient problem. This capability is expected to be advantageous for successfully modeling some delayed and correlated dynamics of the greenhouse inside climate [31].

In summary, the main contribution of this work is the development of a soft sensor using an LSTM-based network to estimate the opening of greenhouse vents from climate variables commonly measured in medium-technology greenhouses. Figure 1 presents the concept of the developed soft sensor, which receives as inputs the measurements of the following climate variables recorded inside and outside a greenhouse: air temperature, air relative humidity, global solar radiation, CO₂ concentration in the air, and outside wind velocity. The estimation of vent opening with a data-based soft sensor is possible because the evolution of the inside climate variables is affected by the ventilation flux. Every time that the vents of a greenhouse are opened or closed, variations in all or some of the aforementioned variables are measured. A historical dataset from a traditional Mediterranean greenhouse was used to train and test a series of LSTM-based network architectures to reproduce the actual opening of the vents which caused the measured variations in the inside climate variables. The training and testing processes were performed using a progressive elimination procedure (PEP) before selecting the final soft sensor based on an LSTM-RNN neural network. The results show a satisfactory performance, demonstrating that the developed soft sensor can estimate the actual opening of the vents with a reduced error.

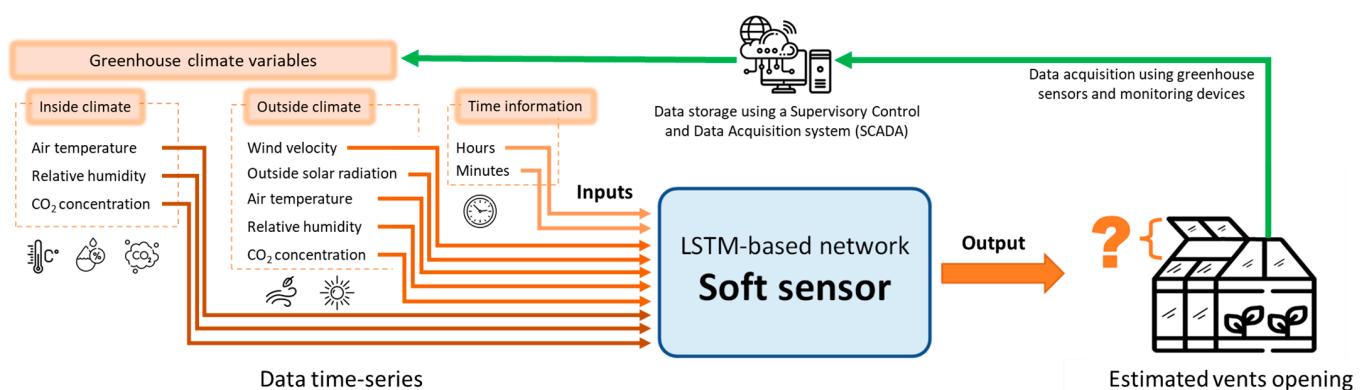


Figure 1. Conceptualization of the soft sensor to estimate the opening of greenhouse vents.

The soft sensor's contribution lies in providing an estimated signal of the opening of vents for its possible use by the existing models to calculate the ventilation flux and predict greenhouse climate variables [23,24]. In this sense, the soft sensor would allow the calculation of the ventilation flux when no physical sensors are available in a greenhouse to measure the opening of the vents. Calculating the ventilation flux is important for the predictive modeling of other greenhouse climate variables that are strongly affected by it,

such as relative humidity, CO₂ concentration, or air temperature. Other potential applications of the soft sensor may include its integration into IoT platforms, as discussed above, which would be useful for continuous estimation of vent opening from data measured with commercial weather stations installed in greenhouses.

The remainder of the paper is organized as follows. In Section 2, materials and methods are described. In Section 3, the results of training and testing the soft sensor using an LSTM-based network are presented and discussed. Finally, the conclusions of the work are summarized in Section 4.

2. Materials and Methods

In this section, an experimental greenhouse was used to obtain the dataset needed for this work. The data obtained include a set of inside and outside greenhouse climate variables to be analyzed and selectively used as inputs in the next section, as well as the target, which is the actual vent-opening signal generated by an automatic controller. As a result of changes in the opening of vents, the climate variables inside the greenhouse are affected by the ventilation flux. Thus, the theory of this work is that a soft sensor based on an LSTM neural network can estimate the vent opening that causes those changes in the inside climate variables. The components and equations that constitute the LSTM-based neural network are explained in this section. Furthermore, the potential network architectures and their hyperparameters are preselected.

2.1. Greenhouse Description

The greenhouse used in this study is located at “Las Palmerillas” Experimental Station of the Cajamar Foundation in Almería, Spain, at an altitude of 151 m. It is a traditional Mediterranean greenhouse (see Figure 2a) with a surface of 877 m² (37.80 m × 23.20 m). Tomato (*Lycopersicon esculentum* “Ramy”) is the crop grown inside this experimental greenhouse, with a plant density of 1.4 plants/m². The greenhouse is equipped with auxiliary systems, such as different actuators, to control the indoor climate. Particularly, the greenhouse has five roof vents (8.36 m × 0.73 m) and two lateral vents (32.75 m × 1.90 m) for natural ventilation, situated on the north and south sides. The roof vents have an angled opening, as shown in Figure 2d, while the lateral vents are opened by rolling up a plastic film, as presented in Figure 2c. All vents can be opened from 0 to 100% of their ventilation area with a resolution of 10% by means of three electric motors (see Figure 2b), which can be manually or automatically operated.

2.2. Experimental Dataset

A dataset with 81 recorded days (233,280 samples) was used to train, validate, and test the developed soft sensor. This dataset was acquired using a commercial data acquisition system called Compact FieldPoint (National Instruments, Austin, TX, USA). It contains 13 climate variables related to inside and outside air temperature, relative humidity, global solar radiation, CO₂ concentration in the air, wind velocity, vent opening, and time variables. The data were recorded during the growth cycle of a tomato crop using sensors installed inside and outside the experimental greenhouse (see Table 1). Their acronyms and units are presented in Table 2. To capture the rapid climate changes inside the greenhouse due to the effect caused by the opening of the vents, a sampling time of 30 s was selected for the data. The opening signal of vents was generated by a supervisory and control data acquisition (SCADA) system, in which a controller was executed to regulate the air temperature inside the greenhouse by natural ventilation.

The period for the selected dataset was from 10 October to 29 December 2020. Due to the large size of this dataset, Figure 3 shows an example of the data recorded between 24 November and 9 December 2020. Notice that the selected data represent the usual dynamics in the greenhouse, with a mix of sunny, cloudy, and windy days. The opening signal of the vents presents different amplitudes and changes due to the action of the automatic control

system, and also some days with a less variant behavior, similar to the manual operation performed by farmers, as can be observed from 3 to 9 December 2020.

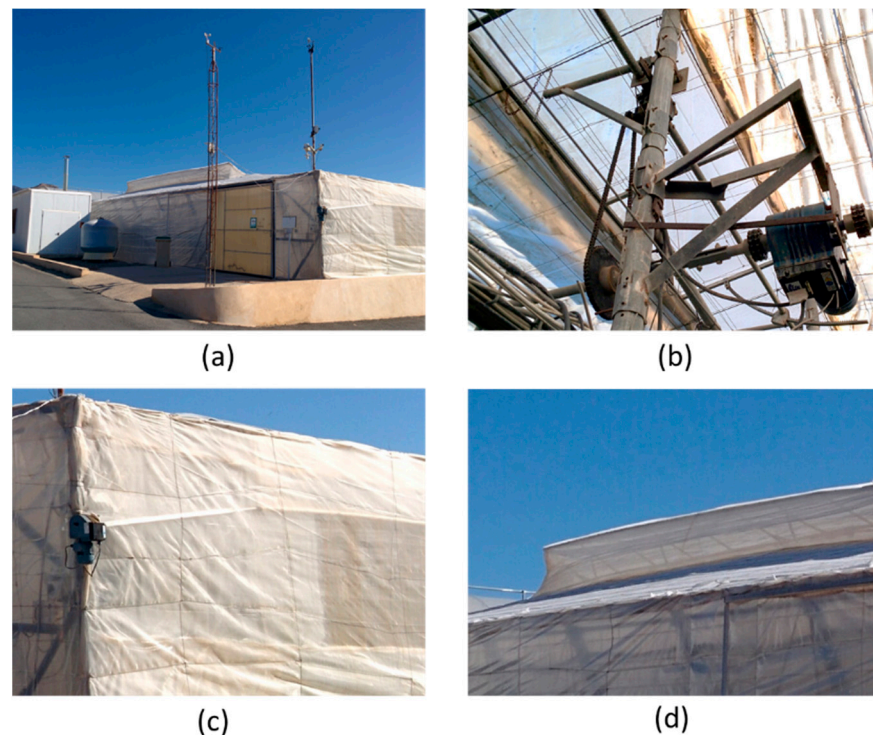


Figure 2. Experimental greenhouse: (a) Outside view; (b) Detail of a motor to open the roof vents; (c) Detail of a lateral vent with its motor to roll up the plastic; (d) Detail of a roof vent.

Table 1. Sensors installed outside and inside the experimental greenhouse.

Sensor	Brand	Model	Precision	Range
Air temperature	Campbell Scientific	HC2S3	± 0.1 °C	−40 to 60 °C
Air relative humidity	Campbell Scientific	HC2S3	± 0.1 %	0 to 100%
Global solar radiation	Hukseflux	LP02	$< \pm 1$ %	0 to 2000 W/m ²
CO ₂ concentration in air	E+E Elektronik	EE820-C2	$< \pm 50$ ppm	0 to 2000 ppm
Wind velocity	Vector Instruments	A100L2/PC3	< 2 %	0 to 75 m/s

Table 2. Dataset variable acronyms, descriptions, and units.

Acronyms	Description	Unit
T _{in}	Inside air temperature	°C
T _{out}	Outside air temperature	°C
H _{in}	Inside air relative humidity	%
H _{out}	Outside air relative humidity	%
CO _{2in}	Inside CO ₂ concentration	ppm
CO _{2out}	Outside CO ₂ concentration (it can be considered constant as 400 ppm)	ppm
RAD _{in}	Inside solar radiation	W/m ²
W _v	Outside wind velocity	m/s
CO _{2diff}	Difference between inside and outside CO ₂	ppm
H _{diff}	Difference between inside and outside relative humidity	%
T _{diff}	Difference between inside and outside air temperature	°C
UVENT _{roof}	Vent opening signal	%

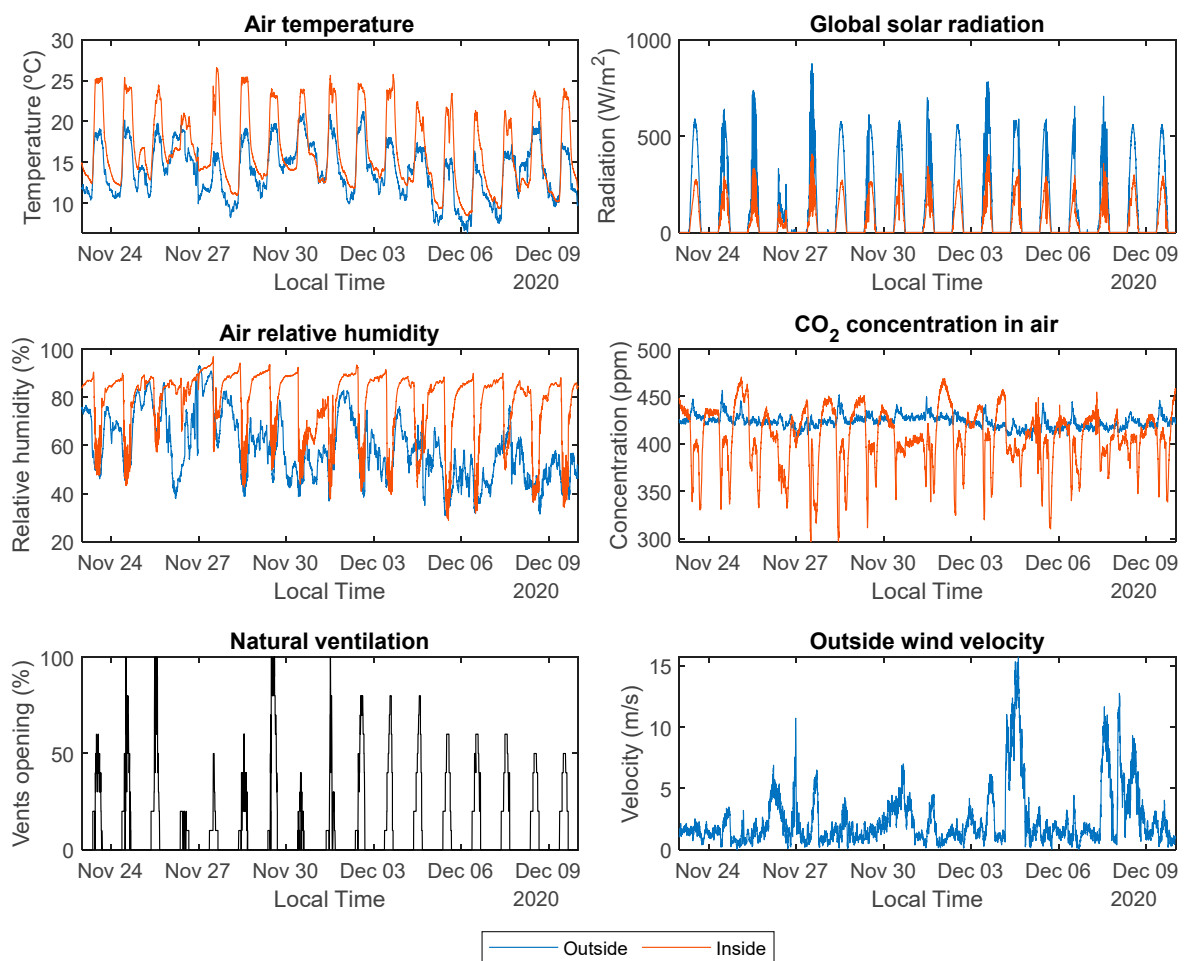


Figure 3. Example of some days contained in the experimental dataset from 10 October 2020 to 29 December 2020.

2.3. Long Short-Term Memory

In artificial neural networks, the LSTM cell is a powerful deep recurrent neural system developed specifically to deal with the vanishing gradient problems that often occur when learning long-term relationships between system inputs and target outputs [32]. This fact motivates the application of LSTM-based neural networks for greenhouse climate modeling, which involves short- and long-term dependencies for multiple inputs (e.g., different climate variables) and the idea that their numerical effects would gradually vanish over time during the training of a neural network if LSTM structures were not used. For this reason, the opening of greenhouse vents could be better estimated not only on the basis of the current states of the measured climate variables but also on the basis of stored information of the long-term past states. In this sense, as a recurrent network, the output of an LSTM cell is fed back as input, creating a recursive flow of information with increased capability for information storage.

An LSTM unit consists of four main components: a cell, an input gate, an output gate, and a forget gate. The cell remembers values over varying time intervals, and the cell gates control the flow of information. The LSTM structure consists of memory blocks, which are recurrently connected subnetworks. The memory block objective is to maintain its state over time while regulating information flow by means of nonlinear gate units. Figure 4 shows the architecture of an LSTM cell, involving an input signal $x(t)$, an output signal $y(t)$, the cell state $c(t)$, and different activation functions σ , g , and h . The components and the way in which an LSTM block processes the flow of information are briefly explained below [31]:

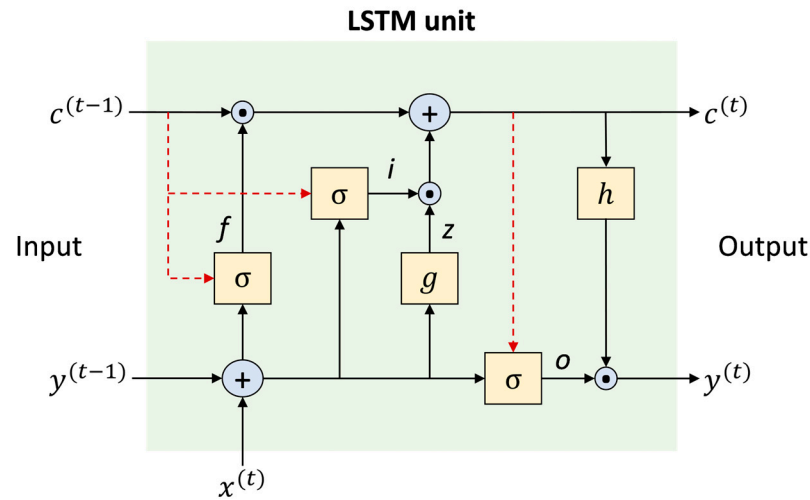


Figure 4. The architecture of a typical LSTM block [31].

Block input, z . It incorporates the current input $x^{(t)}$ and the previous value of the output $y^{(t-1)}$ of one LSTM unit. It is calculated as follows:

$$z^{(t)} = g\left(W_z x^{(t)} + R_z y^{(t-1)} + b_z\right) \quad (1)$$

where t is the time instant, W_z and R_z are the weights for $x^{(t)}$ and $y^{(t-1)}$, respectively, and b_z represents a bias weight vector.

Input gate, i . It is updated by merging $x^{(t)}$, $y^{(t-1)}$ and $c^{(t-1)}$ as follows:

$$i^{(t)} = \sigma\left(W_i x^{(t)} + R_i y^{(t-1)} + p_i \odot c^{(t-1)} + b_i\right) \quad (2)$$

where \odot is a point-wise multiplication of the weights W_i , R_i , and p_i , with $x^{(t)}$, $y^{(t-1)}$, and $c^{(t-1)}$, respectively, in which b_i is the bias vector associated with the input gate.

Forget gate, f . In this component, the LSTM unit decides which information from its previous cell states $c^{(t-1)}$ should be removed. Hence, the activation value of the forget gate at t is calculated with the following expression:

$$f^{(t)} = \sigma\left(W_f x^{(t)} + R_f y^{(t-1)} + p_f \odot c^{(t-1)} + b_f\right) \quad (3)$$

where W_f , R_f , and p_f are the weights associated with $x^{(t)}$, $y^{(t-1)}$, and $c^{(t-1)}$, respectively, and b_f is the bias of the forget gate.

Cell state, c . It is calculated by merging the previous value of the cell state $c^{(t-1)}$ with the block input $z^{(t)}$, the input gate $i^{(t)}$, and the forget gate $f^{(t)}$ values as follows:

$$c^{(t)} = z^{(t)} \odot i^{(t)} + c^{(t-1)} \odot f^{(t)} \quad (4)$$

Output gate, o . Its value is calculated with the following expression:

$$o^{(t)} = \sigma\left(W_o x^{(t)} + R_o y^{(t-1)} + p_o \odot c^{(t)} + b_o\right) \quad (5)$$

where W_o , R_o and p_o are the weights associated with $x^{(t)}$, $y^{(t-1)}$ and $c^{(t-1)}$, respectively, and b_o is a bias weight vector.

Block output, y . Finally, the block output is calculated as:

$$y^{(t)} = h\left(c^{(t)}\right) \odot o^{(t)} \quad (6)$$

In Equations (1)–(6), σ , g , and h refer to the point-wise nonlinear activation functions. The function used for gate activation is the logistic sigmoid, as presented in Equation (7).

$$\sigma(x) = \frac{1}{1 + e^{1-x}} \quad (7)$$

The hyperbolic tangent is often used as the block input and output activation functions, $g(x) = h(x) = \tanh(x)$.

Generally, the vanishing gradient problems can be overcome by using a constant error carousel (CEC), which preserves the error signal within each cell. In a neural network, the role of the LSTM cells is to abstract a meaningful representation of the input time series and then transmit them to the additional hidden layers. Although LSTM-based networks are already performing very well, the potential for improvements is still being explored, as indicated in the comprehensive state-of-art in [31].

2.4. Network Architecture and Hyperparameters Preselection

The LSTM-based network employs full gradient training to adapt the learnable network parameters (weights). The Backpropagation Through Time (BPTT) technique is used to calculate the weights that connect the network components. The LSTM-based network has a set of parameters, which are called hyperparameters. They are specifically determined to define the network architecture and control the learning process in the training phase before applying it to a dataset. These hyperparameters were selected as follows:

- The number of hidden layers. It is selected by trial and error between 4 or 5 layers. Three types of hidden layers constituted the initial network architectures that were tested: LSTM cells, Dense (feedforward ANN), and RNN layers, as shown in Table 3. The reader is encouraged to find more details about RNNs and their relationship with LSTM cells in [30].
- The number of neurons. The number of network weights, which depends on the number and type of the hidden layers and the number of their neurons, is recommended to be much smaller than the number of data samples to avoid overfitting the network to the training data and to favor the generalization of the network output [33]. Hence, the number of neurons was selected accordingly, as presented in Table 3. The number of network weights remains around 18,000, which is much smaller than the number of training data samples (186,705 samples) multiplied by the number of selected inputs (8–13 inputs).
- Historical input data. By trial and error, 40 samples (20 min) were chosen as historical input data to capture all the delayed dynamics of the greenhouse climate, knowing that it presents some slow responses to disturbances (i.e., external weather conditions) and control actions as time-dependent events. It is a fundamental feature of RNNs, specifically of LSTM-based networks, which allows the selective and meaningful mapping of historical input data to the final output.
- Activation function. As presented in Equation (7), sigmoid is the selected function for all the regular layers. It is proven to be significantly useful in the multinomial logistic regression method, which can model types where the discrete output can have more than two possible discrete outcomes [34]. This is particularly important considering that the vent opening is normally a signal restricted to 11 states as discrete values ranging from 0 to 100% with 10% jumps. These jumps are due to the resolution of the motors used to open the vents in greenhouses.
- Optimizer. Adam is the selected optimizer. It is used as a mini-batch gradient descent method. It is based on adaptive estimation of first- and second-order moments. It is computationally efficient, requires little memory, and is suitable for problems with noisy and sparse gradients [35].
- Learning rate. The default learning rate of 0.001 is used for Adam. Higher and lower values were tested, but 0.001 proved to be more efficient in terms of loss reduction and computation time.

- Batch size. The batch size defines the number of samples to work within one iteration before updating the internal weights of the network. By trial and error, the batch size was set as 32 samples to accelerate the training process of the network.
- The number of epochs. An epoch is when all data samples pass through the neural network. By trial and error, 150 epochs were deemed sufficient for this study. In addition, the early stop feature is used to automatically stop the training of the network if no improvement in the validation loss function is shown for more than 50 epochs. In this case, the network with the best weights until that moment is stored. Also, the training can be stopped manually when overfitting is graphically noticed, knowing that the best network is automatically saved after every epoch.

Table 3. Selection of network architecture: number and type of layers and number of neurons.

	Type of Layers		Number of Neurons	
	Architecture A (LSTM-ANN)	Architecture B (LSTM-RNN)	Architecture A (LSTM-ANN)	Architecture B (LSTM-RNN)
Input layer	LSTM	LSTM	Number of inputs	Number of inputs
First hidden layer	Dense	RNN	50	30
Second hidden layer	Dense	RNN	80	60
Third hidden layer	Dense	RNN	80	60
Fourth hidden layer	Dense	RNN	50	30
Fifth hidden layer	Dense	/	15	/
Output layer	Dense	Dense	1 neuron for 1 target	1 neuron for 1 target

In summary, the network architecture obtained in this work is presented in Figure 5.

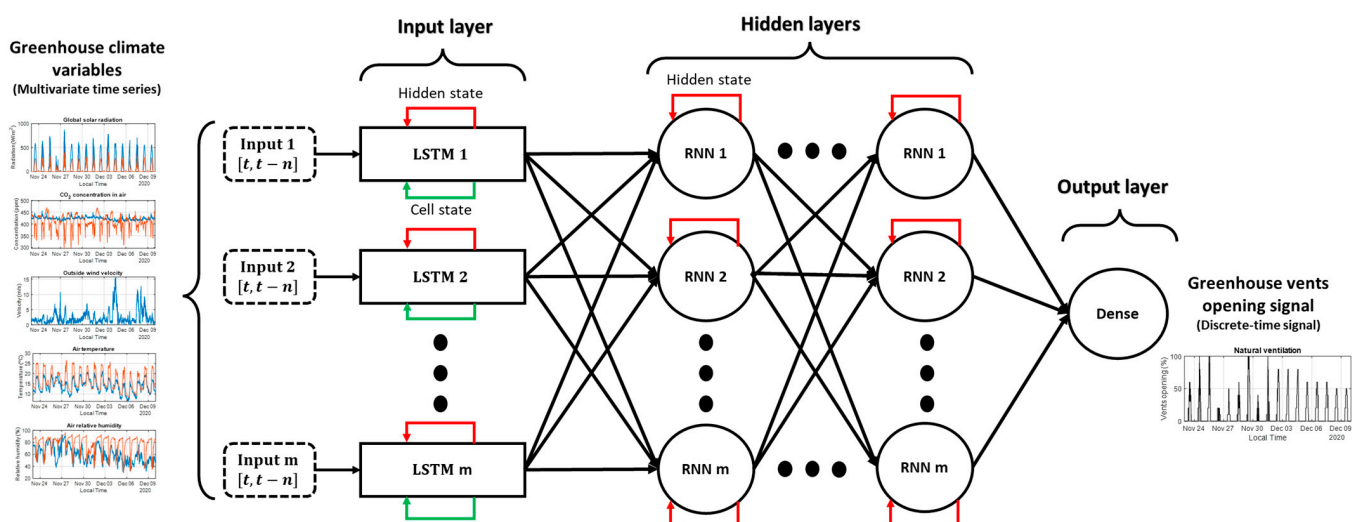


Figure 5. The architecture of the developed LSTM-RNN network.

3. Results and Discussion

In this section, a data analysis is performed to preselect network inputs. The preselected inputs are then used to test two possible LSTM-based network architectures using supervised learning. Finally, a network architecture is selected based on statistical and graphical results and using different sets of the selected inputs in a PEP procedure. The

methodology to develop the soft sensor is summarized in Figure 6. The development stages of the soft sensor were carried out using the machine learning platform called Tensorflow. The statistical evaluations are based on four loss functions which are the coefficient of determination (R^2), the mean absolute error (MAE), the maximum absolute error (MaxAE), and the root mean absolute error (RMAE). The loss function used in the training process is the mean square error (MSE). For the different tests, the computational unit used was a computer with an AMD Ryzen 5 3400G and Radeon Vega Graphics, eight cores, 3.7 GHz, and 8 GB RAM DDR4 1333 MHz. The developed soft sensor was coded and tested in Python 3.9 using the Anaconda software and Visual Studio Code editor.

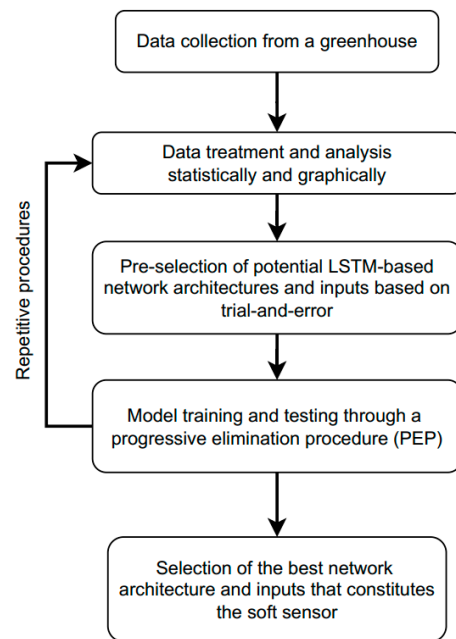


Figure 6. Soft sensor development methodology.

3.1. Data Analysis and Inputs Preselection

The available data were standardized and analyzed to study the relationships between the greenhouse climate variables and the vent opening signal for the preselection of the network inputs. The analysis was carried out in two phases: statistically, using different correlation coefficients, and graphically, using regression analysis.

The dataset includes the opening signals for roof vents and lateral vents ($UVENT_{roof}$ and $UVENT_{lat}$), which are almost identical (see their linear regression analysis in the upper right of Figure 7). In this sense, only the opening signal of the roof vents was used as the target to be estimated for simplicity. To extract additional information from the dataset and reduce the computational time of the training process, three variables T_{diff} , H_{diff} , and $CO2_{diff}$, representing differences between the inside and outside greenhouse environments, were added to the dataset as potential inputs after calculating them as follows:

$$T_{diff} = T_{in} - T_{out} \quad (8)$$

$$H_{diff} = H_{in} - H_{out} \quad (9)$$

$$CO2_{diff} = CO2_{in} - CO2_{out} \quad (10)$$

In the statistical analysis, two cases were studied. First, using the complete signal of the vent opening ($UVENT_{roof} \geq 0\%$), and second, using only the time intervals when the vents were open ($UVENT_{roof} > 0\%$), as presented in Table 4. In both cases, three correlation coefficients were used: Pearson's coefficient for the linear correlation analysis,

and Spearman’s and Kendall’s rank coefficients [36] for the analysis of linear and nonlinear relationships.

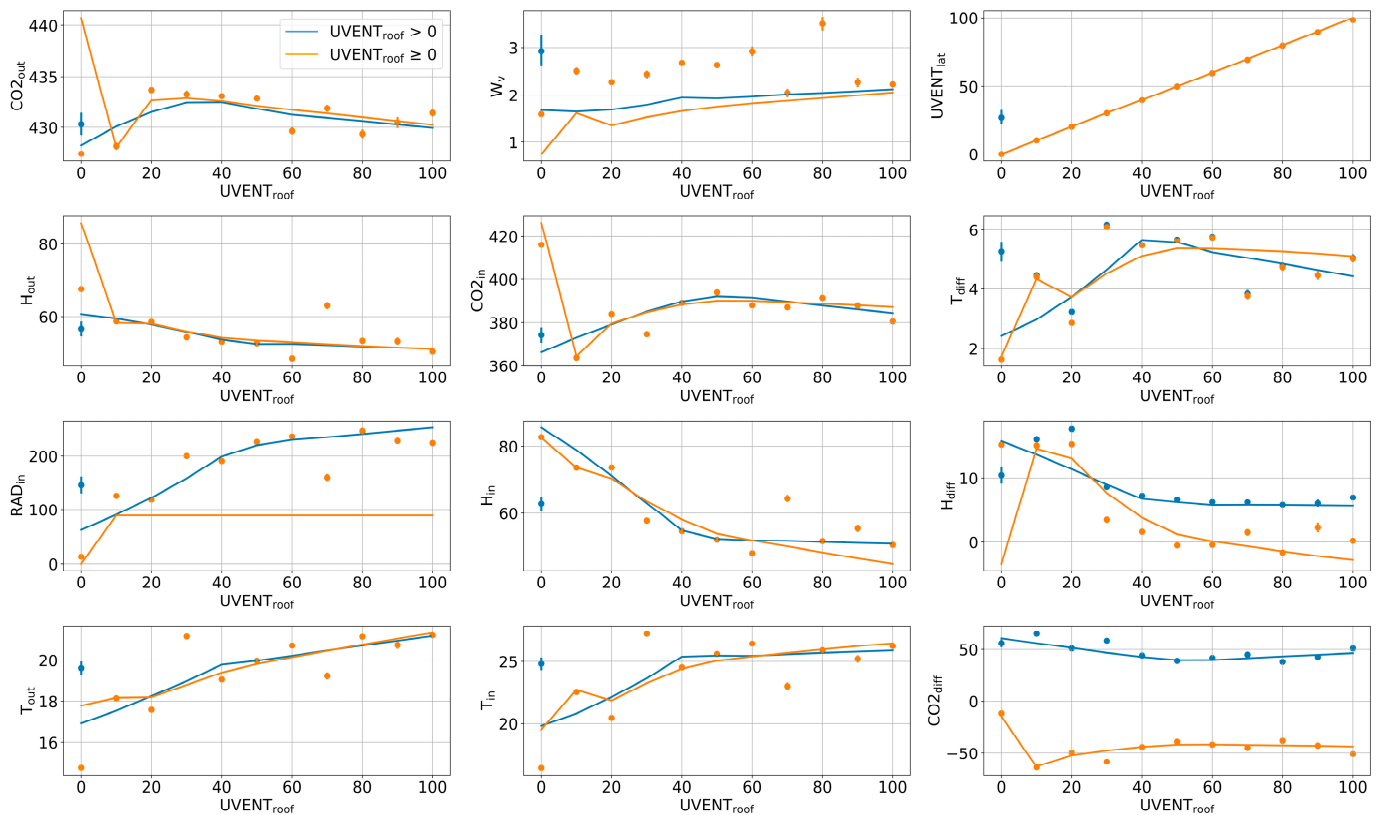


Figure 7. Regression analysis of the greenhouse climate variables with the vent opening in two cases: using the complete variable of vent opening signal $UVENT_{roof} \geq 0\%$ (orange) and using only time intervals when vents are open $UVENT_{roof} > 0\%$ (blue).

Table 4. Statistical analysis of the linear and nonlinear correlation.

	Pearson’s Coefficient		Spearman’s Rank Coefficient		Kendall’s Rank Coefficient	
	$UVENT_{roof}$	$UVENT_{roof}$	$UVENT_{roof}$	$UVENT_{roof}$	$UVENT_{roof}$	$UVENT_{roof}$
	$\geq 0\%$	$> 0\%$	$\geq 0\%$	$> 0\%$	$\geq 0\%$	$> 0\%$
CO2 _{out}	0.22	−0.05	0.09	−0.02	0.07	−0.02
H _{out}	−0.34	−0.15	−0.03	−0.18	−0.02	−0.13
RAD _{in}	0.75	0.37	0.07	0.45	0.05	0.34
T _{out}	0.5	0.3	0.17	0.32	0.14	0.23
W _v	0.2	0.03	−0.01	0.11	−0.01	0.08
CO2 _{in}	−0.3	0.14	−0.03	0.2	−0.02	0.14
H _{in}	−0.07	−0.46	−0.03	−0.51	−0.02	−0.38
T _{in}	0.6	0.32	0.08	0.32	0.07	0.24
CO2 _{diff}	−0.36	0.14	0.1	−0.19	0.07	−0.13
H _{diff}	−0.37	−0.4	−0.05	−0.36	−0.04	−0.26
T _{diff}	0.5	0.19	−0.08	0.21	−0.06	0.14
UVENT _{roof}	1	1	1	1	1	1

As for the graphical analysis, all the data variables were graphically represented, as previously shown in Figure 3. Normally, it is expected that the vents are closed mostly at night, so the inside solar radiation measurements could be useful as an indication of daytime and nighttime periods. In addition, regression analysis to study the linear/nonlinear and monotonic/non-monotonic relationships in the data was also performed for two cases ($UVENT_{roof} \geq 0\%$ and $UVENT_{roof} > 0\%$), as shown in Figure 7. The presented

curves were obtained using the “regplot” function of a Python data visualization library called Seaborn [37]. This function has been developed as a practical tool to graphically demonstrate linear and nonlinear data relationships and obtain the best-fit curve. It has a useful feature called “x_estimator” for regression analysis when discrete variables are involved. It can calculate and plot the mean of the y-axis samples corresponding to each repeatable discrete value on the x-axis (a data category), which in most cases helps to demonstrate how the best-fit curve was fitted to the data distribution.

The relationships represented by the values of correlation coefficients and the results of the regression analysis are briefly discussed, and the potential inputs are initially selected accordingly in Table 5. It is commonly known that, when manually controlling the opening of greenhouse vents, farmers usually follow a predetermined time-based schedule to know when the vents should be opened or closed and what their opening percentage should be depending on the conditions of the crop, the season, and the greenhouse geographical location. For these reasons, two time-related variables, X_{hours} and X_{minutes} , were also considered as inputs to the soft sensor, which may be helpful to take into account specific changes in the vent opening and climate evolutions that repeatedly occur at a given time. In summary, based on the findings of the data analyses, 10 variables were selected as inputs to the LSTM-based network: X_{hours} , X_{minutes} , RAD_{in} , W_v , CO_2_{in} , H_{in} , T_{in} , T_{diff} , H_{diff} , and $\text{CO}_2_{\text{diff}}$.

3.2. Training and Testing the LSTM-Based Network

The described LSTM-based network was trained and tested using different architectures and inputs in a PEP procedure to obtain the final network for the soft sensor, knowing that the target to be estimated is the greenhouse vent opening ($\text{UVENT}_{\text{roof}}$). The training process consists of identifying the network weights by minimizing a loss function (MSE) that indicates the error between the real measured signal of the vent opening and the output of the trained network (i.e., the vent opening estimated by the soft sensor).

3.2.1. Dataset Splitting

The development of any ANN requires dividing the available data into three sets (for training, validation, and testing processes) manually or automatically, depending on different techniques. In this work, the time series data were manually divided based on the evolution of the target variable as one of the main factors when manually splitting data. Hence, the data were divided as presented in Figure 8, consisting of the following parts:

- A training dataset is used for the network learning process to adjust its parameters. The complete dataset includes two different control methods for the opening of the vents. One is an automatic control showing rapid changes (before sample 160,000), and the other is a time-dependent control showing fewer changes (after sample 160,000). The training dataset was selected to include both types of control for the vent opening to enhance the training process with sufficient information. Moreover, this dataset was shuffled to ensure generalization during the training process. It contains 64 days representing 80% of the total dataset, from 19 October 2020 to 22 December 2020.
- A validation dataset was also used during the training process to provide an unbiased evaluation of the network while being fitted to the training dataset. The validation dataset is also involved in other forms of network preparation, such as feature and threshold selection. The validation dataset was selected to contain 8 days representing 10% of the complete dataset: 4 days from the start of the complete dataset (from 14 October 2020 to 17 October 2020) and another 4 days from the end (from 26 December 2020 to 29 December 2020). These days were selected because they present the different types of control for the opening of the vents.
- A test dataset is used to perform an unbiased evaluation of the final network. The test dataset was also selected to contain 8 days representing 10% of the complete dataset: 4 days from the start of the dataset (from 10 October 2020 to 13 October 2020) and another 4 days from the end (from 22 December 2020 to 25 December 2020).

Table 5. Statistical analysis of the linear and nonlinear correlation between the potential inputs and the target vent opening signal.

	UVENT _{roof}					
	Pearson's Coefficient	Spearman's Rank Coefficient	Kendall's Rank Coefficient	Regression Analysis	Prior Knowledge about Physical Interactions	Inputs Pre-Selection
CO ₂ _{out}	Weak correlation	No correlation	No correlation	Weak non-monotonic negative correlation	Not influenced by greenhouse vent opening. The network could implicitly consider it a reference for the difference between inside and outside climate.	Eliminated
H _{out}	Weak negative linear correlation	Weak negative correlation when UVENT _{roof} > 0%	Weak negative correlation when UVENT _{roof} > 0%	Monotonic negative nonlinear correlation	Not influenced by greenhouse vent opening. The network could implicitly consider it a reference for the difference between inside and outside climate.	Eliminated
RAD _{in}	Significant positive linear correlation	Moderate positive correlation when UVENT _{roof} > 0%	Moderate positive correlation when UVENT _{roof} > 0%	Monotonic positive nonlinear correlation	Not influenced by greenhouse vent opening. Considered as a reference for the diurnal and nocturnal periods, knowing that vents are mostly closed at night.	Selected
T _{out}	Moderate positive linear correlation	Moderate positive linear correlation	Moderate positive linear correlation	Monotonic positive linear correlation	Not influenced by greenhouse vent opening. Probably correlated because it increases in the midday by solar radiation, which makes its evolution somehow similar to UVENT _{roof} .	Eliminated
W _v	Weak correlation	Weak correlation when UVENT _{roof} > 0%	Weak correlation	Weak correlationAssociated non-monotonic variations	Not influenced by greenhouse vent opening. The network could consider its effect on the ventilation rate, in turn, on the greenhouse air variables implicitly through vent opening.	Selected
CO ₂ _{in}	Weak negativelinear correlation	Weak positive correlation when UVENT _{roof} > 0%	Weak positive correlation when UVENT _{roof} > 0%	Non-monotonic nonlinear correlation when UVENT _{roof} > 0%	Rapidly influenced by vent opening. Considered to have the fastest reaction to UVENT _{roof} .	Selected
H _{in}	Moderate negative linear correlation when UVENT _{roof} > 0%	Moderate negative correlation when UVENT _{roof} > 0%	Moderate negative correlation when UVENT _{roof} > 0%	Negative monotonic nonlinear correlation	Rapidly influenced by vent opening. Considered to have a high sensitivity to UVENT _{roof} .	Selected
T _{in}	Moderate positive linear correlation	Moderate positive correlation when UVENT _{roof} > 0%	Weak positive correlation when UVENT _{roof} > 0%	Non-monotonic nonlinear correlation	Influenced by vent opening. Considered to have a high sensitivity to UVENT _{roof} .	Selected
CO ₂ _{diff}	Moderate negative linear correlation	Weak negative correlation when UVENT _{roof} > 0%	Weak negative correlation when UVENT _{roof} > 0%	Non-monotonic nonlinear correlation	Rapidly influenced by vent opening. Considered to have a high sensitivity to UVENT _{roof} .	Selected
H _{diff}	Moderate negative linear correlation	Moderate negative correlation when UVENT _{roof} > 0%	Weak negative correlation when UVENT _{roof} > 0%	Monotonic nonlinear correlation	Rapidly influenced by vent opening. Considered to have a high sensitivity to UVENT _{roof} .	Selected
T _{diff}	Moderate positive linear correlation	Weak positive correlation when UVENT _{roof} > 0%	Weak positive correlation when UVENT _{roof} > 0%	Non-monotonic nonlinear correlation	Influenced by vent opening. Considered to have a high sensitivity to UVENT _{roof} .	Selected

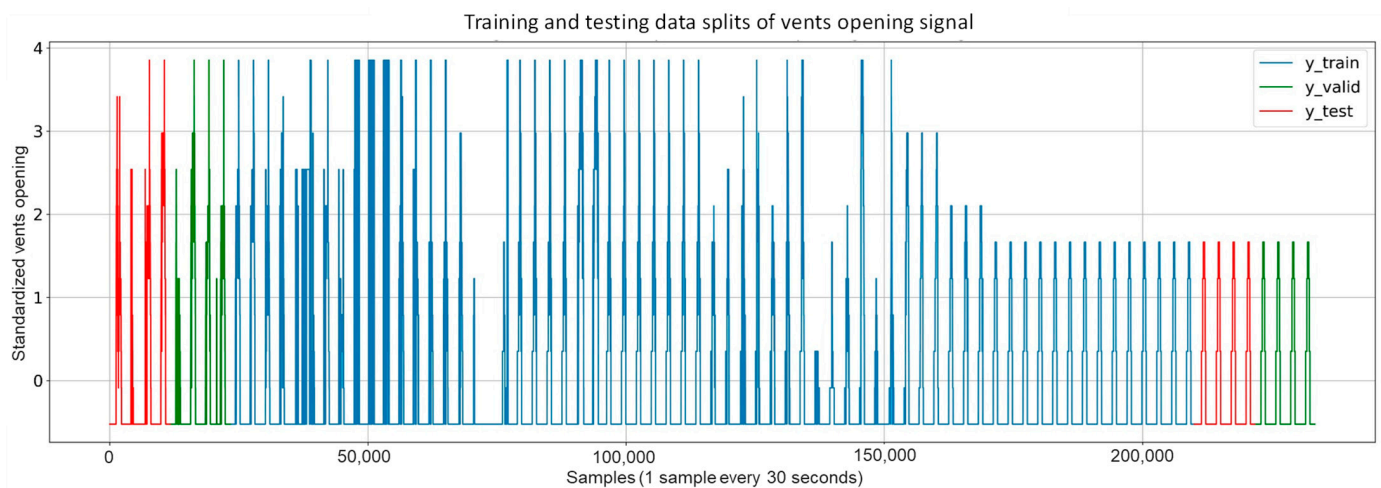


Figure 8. Splitting the target data into three parts: Training target (blue), Validation target (Green), and Test target (red).

3.2.2. Network Training and Progressive Elimination Procedure for Input Selection

The network training process was performed using different architectures and inputs. As presented in Table 3, two architectures, “A” and “B”, were preselected based on multiple tests. The first tests focused on evaluating a simple architecture (LSTM-ANN) consisting of an LSTM layer as an input layer, five hidden layers, and an output layer of a dense type. Secondly, a deep network (LSTM-RNN) was tested consisting of an LSTM layer as an input layer, four RNN hidden layers to increase the capability of the resulting network, and a dense output layer. The LSTM layer was used in both cases as an input layer to take advantage of its ability to abstract a meaningful representation of the input time series, and then the extracted higher-level information was transmitted to the hidden layers in order to produce the output, which is the estimated vent opening signal.

The preselected architectures were trained, tested, and statistically evaluated with different inputs in a PEP procedure, as presented in Table 6. According to the input preselection in Section 3.1 and the greenhouse climate dynamics, the first PEP procedure consisted of eliminating the input W_v because it was not correlated with the opening signal of vents. It is a very noisy variable that was graphically observed to cause undesirable fluctuations in the evolution of the output of the networks (i.e., the estimated vent opening). The second PEP procedure consisted of preserving W_v and eliminating T_{out} , H_{out} , and CO_{2out} because these variables do not change when the vents are opened or closed, and their physical effects on the greenhouse climate are already taken into account in the calculated climate differences, T_{diff} , H_{diff} , and CO_{2diff} . It was concluded that the elimination of these inputs resulted in a decrease in the error values; thus, the PEP is an efficient procedure for improving the estimation and reducing the size of the data.

Table 6. Statistical evaluation of the obtained LSTM-based networks using different cost functions.

	All the Available Inputs				Eliminating Only W_v				Eliminating T_{out} , H_{out} and CO_{2out}			
	R^2	RMSE (%)	MAE (%)	MaxAE (%)	R^2	RMSE (%)	MAE (%)	MaxAE (%)	R^2	RMSE (%)	MAE (%)	MaxAE (%)
Architecture A (LSTM-ANN)	0.77	9.8	3.94	77.66	0.74	10.04	4.04	99.03	0.77	9.64	5.33	65.76
Architecture B (LSTM-RNN)	0.61	12.76	5.98	93.05	0.75	10.32	5.43	63.42	0.8	9.13	4.45	62.94

The time consumed for the training processes of the LSTM-ANN network is around 6 h, and for the LSTM-RNN network, it is around 9 h, which is considered an acceptable time consumption with a moderate computational cost. According to the results presented in Table 6, architecture “B” with an LSTM-RNN network outperforms architecture “A” with an LSTM-ANN network. Moreover, based on the PEP procedure, the best results for the LSTM-RNN network are obtained using only 10 selected inputs which are: X_{hours} , X_{minutes} , RAD_{in} , W_v , CO_2_{in} , H_{in} , T_{in} , T_{diff} , H_{diff} , and $\text{CO}_2_{\text{diff}}$.

Concerning the graphical evaluation, the training process of the best LSTM-RNN network presented an adequate convergence for the evolution of the training and validation cost function, as shown in Figure 9. The training process was manually stopped when the onset of divergence (see the red box in Figure 9) was observed as a sign of network overfitting. An example of estimation results using the training data is shown in Figure 10, which presents a satisfactory fit between the actual and estimated vents opening, avoiding overfitting to the training data. The results using the test dataset with the 10 selected inputs are shown in Figure 11. In addition, Figures 12 and 13 show other results using the test dataset with 12 inputs and 13 inputs, respectively. The estimated opening of vent results were filtered to present a less noisy signal. A first-order filter was used for the output of the LSTM-RNN network, with a time constant of 250 s. In Figure 11, the results obtained with the LSTM-RNN network show a satisfactory fit to the real vent opening by reproducing the time intervals in which the vents are opened and closed, as well as the maximum opening amplitudes, and by estimating the main changes in the signal. It can be noticed that the fit is better in one part (see samples after 12,000) than in another due to the different evolution of the opening signal of vents. It is less challenging for the network to estimate the part of the signal with fewer changes per day because it is a repetitive dynamic, and the changes in the opening values occur more slowly and far apart over time. These two factors allow the network to learn more about this part of the signal than the variant part. In other words, the fewer changes per day in the opening of the vents, the easier it is for the network to interpret the corresponding change in the greenhouse climate and the more accurate the estimation provided by the soft sensor. This fact is particularly interesting for most traditional greenhouses, in which farmers manually open and close the vents in a similar way, so these results confirm the usefulness of the developed soft sensor in that context. However, it can be concluded that it will be necessary to train the LSTM-RNN network with larger datasets to increase the accuracy of estimating the rapid changes in vent opening.

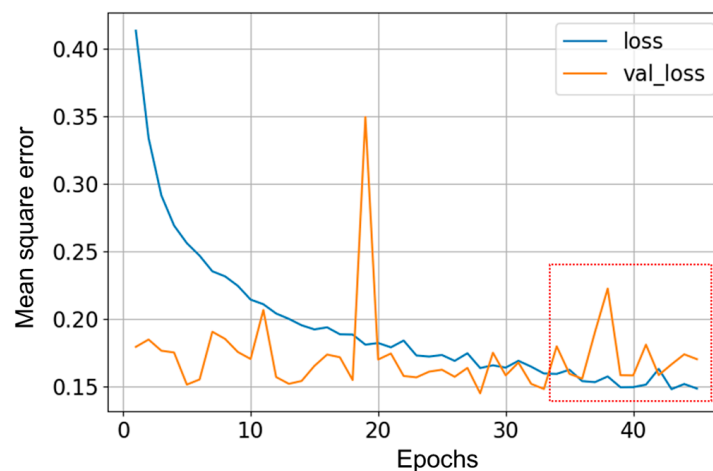


Figure 9. Evolution of the loss function during training (loss) and validation (val_loss) processes using the standardized data.

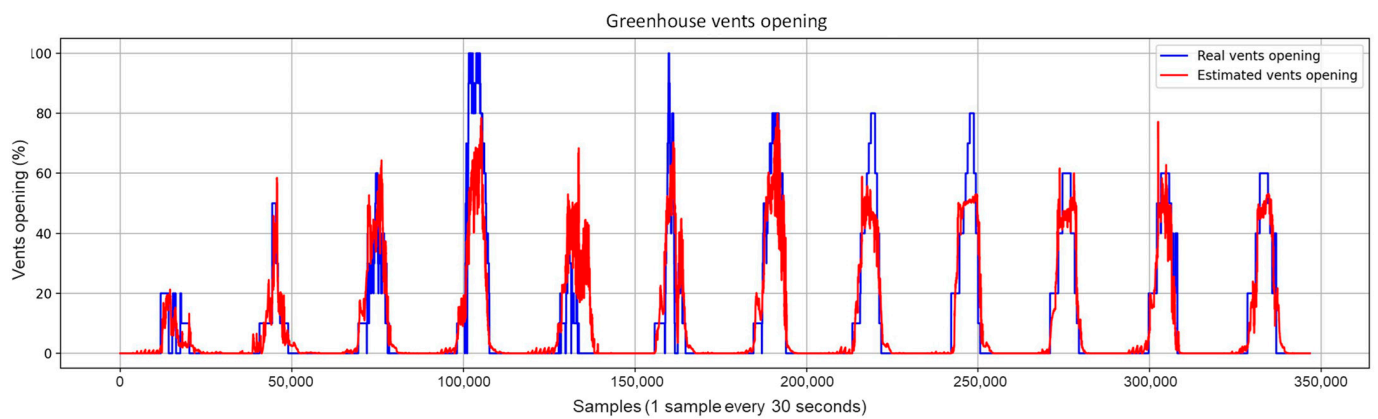


Figure 10. A sample from the results of the training process of the LSTM-RNN network.

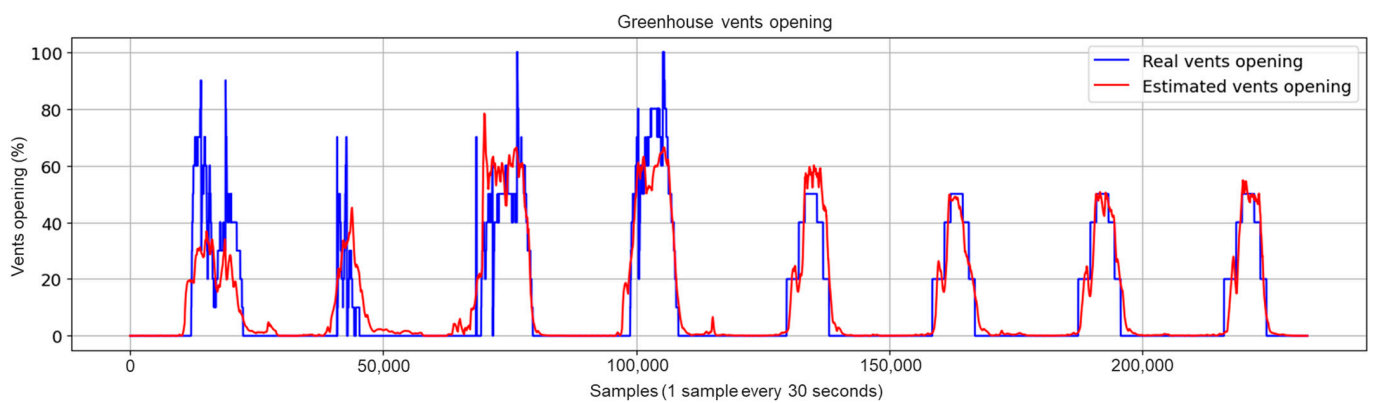


Figure 11. Test results for the LSTM-RNN network with the final selected 10 inputs (after eliminating T_{out} , H_{out} and CO_{2out}).

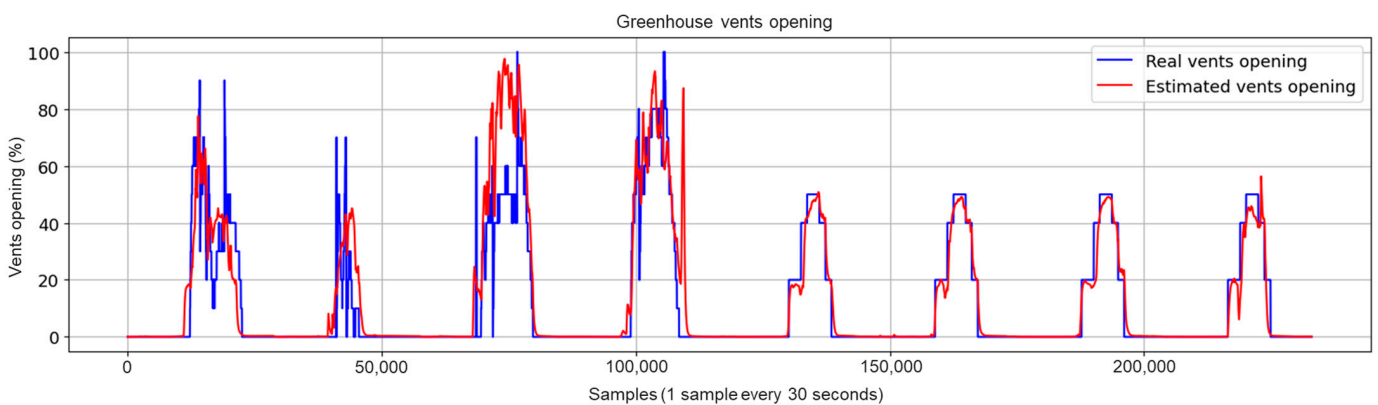


Figure 12. Test results for the LSTM-RNN network with 12 inputs (after eliminating W_v).

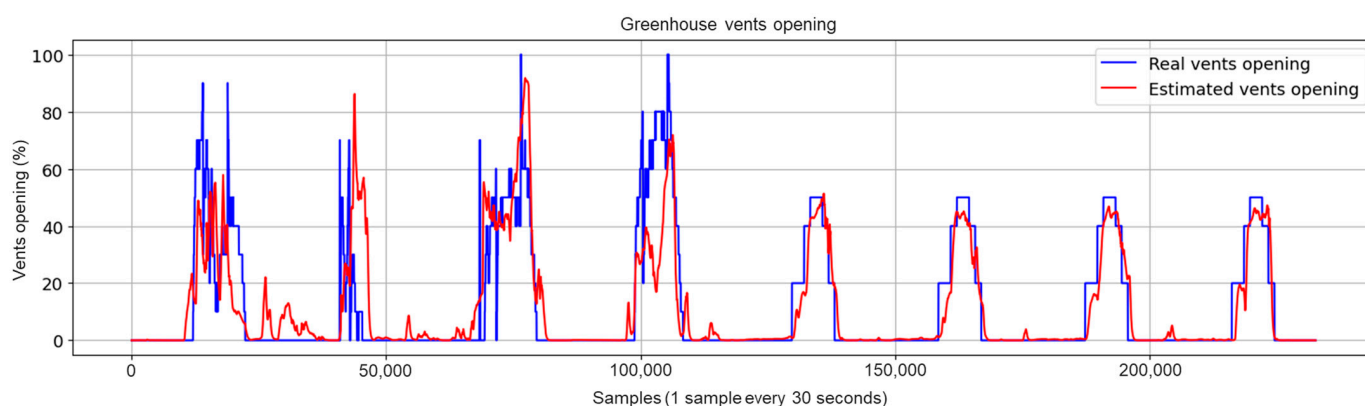


Figure 13. Test results for the LSTM-RNN network using all the available 13 inputs.

4. Conclusions

A soft sensor based on an LSTM-RNN neural network has been developed to estimate the opening of greenhouse vents using a set of measurable climate variables. A comprehensive statistical and graphical data study was performed using different linear and nonlinear correlation coefficients and regression analyses. Based on the results of this data analysis and trial-and-error training processes, two possible network architectures (LSTM-ANN and LSTM-RNN) and ten inputs were preselected for the soft sensor design. In addition, a series of training and testing processes were carried out in a PEP procedure. It has been shown that the external climate variables T_{out} , H_{out} , and CO_{2out} are necessary to calculate the corresponding differences between the inside and outside greenhouse climate T_{diff} , H_{diff} , and CO_{2diff} to be used as inputs to the LSTM-based network. It has also been found that H_{diff} is the most correlated input presenting a negative monotonic nonlinear correlation with the opening signal of vents.

The best network architecture is the LSTM-RNN, due to its performance in estimating the actually recorded opening of vents with reduced error values: $R^2 = 0.8$, $RMSE = 9.13\%$, $MAE = 4.45\%$ and $MaxAE = 62.94\%$. As for the graphical results, the soft sensor developed using the LSTM-RNN network provides a good fit between the estimated and the real vent opening in both daytime and nighttime. However, the estimation is more accurate when there are fewer changes in the opening of vents. Moreover, it has not been possible to compare the obtained network and results with other works since this study is the first attempt to estimate the opening of greenhouse vents.

Consequently, the results confirm that the soft sensor is suitable for use in greenhouses where farmers manually operate the opening of the vents. In this context, the soft sensor could be applied to:

- Estimate and monitor the evolution of the natural ventilation flux.
- Develop predictive models for greenhouse climate evolution as a function of the estimated vent opening.
- IoT platforms and decision support systems to provide recommendations to farmers after analyzing the measured data, and for example, alert them to close the vents whenever a high wind velocity is detected to avoid any risk of damage to the greenhouse and crop.

In conclusion, the contribution of this work to the field of greenhouse agriculture lies in the possibility of offering a tool that can be applied to estimate the opening signal of vents without the need to have installed specific sensors on the vents of a greenhouse or control systems for natural ventilation.

In future works, the use of larger datasets to improve the performance of the soft sensor will be studied. The focus will be on improving the estimation results when fast changes in the vent opening occur. The soft sensor could also be tested in different greenhouses with different shapes and geographical locations.

Author Contributions: Conceptualization, F.R., M.G. and F.G.-M.; methodology, M.G., F.G.-M. and F.R.; visualization, M.G., F.G.-M.; software, M.G.; validation, M.G.; investigation, M.G., F.G.-M. and F.R.; Data, F.G.-M. and F.R.; data curation, M.G., F.G.-M.; writing—original draft preparation and editing, M.G., F.G.-M.; writing—review, F.R., H.M. and F.G.-M.; supervision, F.R. and H.M.; project administration, F.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work is a result of Project PID2021-122560OB-I00 funded by MCIN/AEI/10.13039/501100011033 and by ERDF A way to make Europe. Author Francisco García-Mañas is supported by an FPU grant of the Spanish Ministry of Science, Innovation and Universities.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Gomez-Gomez, J.D.; Pulido-Velazquez, D.; Collados-Lara, A.J.; Fernandez-Chacon, F. The Impact of Climate Change Scenarios on Droughts and Their Propagation in an Arid Mediterranean Basin. A Useful Approach for Planning Adaptation Strategies. *Sci. Total Environ.* **2022**, *820*, 153128. [[CrossRef](#)] [[PubMed](#)]
2. Katzin, D.; van Henten, E.J.; van Mourik, S. Process-Based Greenhouse Climate Models: Genealogy, Current Status, and Future Directions. *Agric. Syst.* **2022**, *198*, 103388. [[CrossRef](#)]
3. Rodríguez, F.; Berenguel, M.; Guzmán, J.L.; Ramírez-Arias, A. *Modeling and Control of Greenhouse Crop Growth*; Springer: Cham, Switzerland, 2015.
4. Kochhar, A.; Kumar, N. Wireless Sensor Networks for Greenhouses: An End-to-End Review. *Comput. Electron. Agric.* **2019**, *163*, 104877. [[CrossRef](#)]
5. Bersani, C.; Ruggiero, C.; Sacile, R.; Soussi, A.; Zero, E. Internet of Things Approaches for Monitoring and Control of Smart Greenhouses in Industry 4.0. *Energies* **2022**, *15*, 3834. [[CrossRef](#)]
6. Tzounis, A.; Katsoulas, N.; Bartzanas, T.; Kittas, C. Internet of Things in Agriculture, Recent Advances and Future Challenges. *Biosyst. Eng.* **2017**, *164*, 31–48. [[CrossRef](#)]
7. Rezvani, S.M.; Abyaneh, H.Z.; Shamshiri, R.R.; Balasundram, S.K.; Dworak, V.; Goodarzi, M.; Sultan, M.; Mahns, B. IoT-Based Sensor Data Fusion for Determining Optimality Degrees of Microclimate Parameters in Commercial Greenhouse Production of Tomato. *Sensors* **2020**, *20*, 6474. [[CrossRef](#)]
8. Muñoz, M.; Guzmán, J.L.; Sánchez-Molina, J.A.; Rodríguez, F.; Torres, M.; Berenguel, M. A New IoT-Based Platform for Greenhouse Crop Production. *IEEE Internet Things J.* **2022**, *9*, 6325–6334. [[CrossRef](#)]
9. van Mourik, S.; van der Tol, R.; Linker, R.; Reyes-Lastiri, D.; Kootstra, G.; Koerkamp, P.G.; van Henten, E.J. Introductory Overview: Systems and Control Methods for Operational Management Support in Agricultural Production Systems. *Environ. Model. Softw.* **2021**, *139*, 105031. [[CrossRef](#)]
10. Garcia-Martinez, M.C.; Balasch, S.; Alcon, F.; Fernandez-Zamudio, M.A. Characterization of Technological Levels in Mediterranean Horticultural Greenhouses. *Span. J. Agric. Res.* **2010**, *8*, 509–525. [[CrossRef](#)]
11. Mendoza-Fernández, A.J.; Peña-Fernández, A.; Molina, L.; Aguilera, P.A. The Role of Technology in Greenhouse Agriculture: Towards a Sustainable Intensification in Campo de Dalías (Almería, Spain). *Agronomy* **2021**, *11*, 101. [[CrossRef](#)]
12. Valera, D.L.; Belmonte, L.J.; Molina-Aiz, F.D.; López, A.; Camacho, F. The Greenhouses of Almería, Spain: Technological Analysis and Profitability. *Acta Hort.* **2017**, *1170*, 219–226. [[CrossRef](#)]
13. Kittas, C.; Papadakis, G.; Boulard, T. Greenhouse Ventilation Rates through Combined Roof and Side Openings: An Experimental Study. *Acta Hort.* **1997**, *443*, 31–38. [[CrossRef](#)]
14. Bhujel, A.; Basak, J.K.; Khan, F.; Arulmozhi, E.; Jaihuni, M.; Sihalath, T.; Lee, D.; Park, J.; Kim, H.T. Sensor Systems for Greenhouse Microclimate Monitoring and Control: A Review. *J. Biosyst. Eng.* **2020**, *45*, 341–361. [[CrossRef](#)]
15. Muñoz, M.; Guzmán, J.L.; Sánchez, J.A.; Rodríguez, F.; Torres, M. Greenhouse Models as a Service (GMaaS) for Simulation and Control. *IFAC-PapersOnLine* **2019**, *52*, 190–195. [[CrossRef](#)]
16. Guesbaya, M.; García-Mañas, F.; Rodríguez, F.; Megherbi, H.; Ouamane, M.R. Virtual Sensor for Ventilation Flux Estimation in Greenhouses. In Proceedings of the XI Congreso Ibérico de Agroingeniería, Valladolid, Spain, 11–12 November 2021.
17. Guesbaya, M.; García-Mañas, F.; Megherbi, H.; Rodríguez, F. Real-Time Adaptation of a Greenhouse Microclimate Model Using an Online Parameter Estimator Based on a Bat Algorithm Variant. *Comput. Electron. Agric.* **2022**, *192*, 106627. [[CrossRef](#)]
18. Fortuna, L.; Graziani, S.; Rizzo, A.; Xibilia, M.G. *Soft Sensors for Monitoring and Control of Industrial Processes*; Springer: Berlin/Heidelberg, Germany, 2007.
19. Sánchez, J.A.; Rodríguez, F.; Guzmán, J.L.; Arahal, M.R. Virtual Sensors for Designing Irrigation Controllers in Greenhouses. *Sensors* **2012**, *12*, 15244–15266. [[CrossRef](#)]
20. Guzmán, C.H.; Carrera, J.L.; Durán, H.A.; Berumen, J.; Ortiz, A.A.; Guirette, O.A.; Arroyo, A.; Brizuela, J.A.; Gómez, F.; Blanco, A.; et al. Implementation of Virtual Sensors for Monitoring Temperature in Greenhouses Using CFD and Control. *Sensors* **2019**, *19*, 60. [[CrossRef](#)]

21. García-Mañas, F.; Rodríguez, F.; Berenguel, M. Leaf Area Index Soft Sensor for Tomato Crops in Greenhouses. *IFAC-Papers OnLine* **2020**, *53*, 15796–15803. [[CrossRef](#)]
22. Shang, C.; Yang, F.; Huang, D.; Lyu, W. Data-Driven Soft Sensor Development Based on Deep Learning Technique. *J. Process Control* **2014**, *24*, 223–233. [[CrossRef](#)]
23. Bontsema, J.; van Henten, E.J.; Kornet, J.G.; Budding, J.; Rieswijk, T. On-Line Estimation of the Ventilation Rate of Greenhouses. *IFAC Proc. Vol.* **2005**, *38*, 125–130. [[CrossRef](#)]
24. Kittas, C.; Boulard, T.; Papadakis, G. Natural Ventilation of a Greenhouse with Ridge and Side Openings: Sensitivity to Temperature and Wind Effects. *Trans. ASAE* **1997**, *40*, 415–425. [[CrossRef](#)]
25. Teitel, M.; Tanny, J. Natural Ventilation of Greenhouses: Experiments and Model. *Agric. For. Meteorol.* **1999**, *96*, 59–70. [[CrossRef](#)]
26. Pérez-Parra, J.; Berenguel, M.; Rodríguez, F.; Ramírez-Arias, A. Ventilation Rate Models of Mediterranean Greenhouses for Control Purposes. *Acta Hort.* **2006**, *719*, 197–204. [[CrossRef](#)]
27. López, A.; Valera, D.L.; Molina-Aiz, F. Sonic Anemometry to Measure Natural Ventilation in Greenhouses. *Sensors* **2011**, *11*, 9820–9838. [[CrossRef](#)]
28. Graziani, S.; Xibilia, M.G. Deep Learning for Soft Sensor Design. In *Development and Analysis of Deep Learning Architectures*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 31–59.
29. Ojo, M.O.; Zahid, A. Deep Learning in Controlled Environment Agriculture: A Review of Recent Advancements, Challenges and Prospects. *Sensors* **2022**, *22*, 7965. [[CrossRef](#)]
30. Liu, Y.; Li, D.; Wan, S.; Wang, F.; Dou, W.; Xu, X.; Li, S.; Ma, R.; Qi, L. A Long Short-Term Memory-Based Model for Greenhouse Climate Prediction. *Int. J. Intell. Syst.* **2022**, *37*, 135–151. [[CrossRef](#)]
31. Van Houdt, G.; Mosquera, C.; Nápoles, G. A review on the Long Short-Term Memory Model. *Artif. Intell. Rev.* **2020**, *53*, 5929–5955. [[CrossRef](#)]
32. Sepp, H.; Jurgen, S. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
33. Arahál, M.R.; Berenguel, M.; Rodríguez, F. *Técnicas de Predicción Con Aplicaciones En Ingeniería*; Secretariado de Publicaciones de la Universidad de Sevilla: Sevilla, Spain, 2006.
34. Edgar, T.W.; Manz, D.O. Exploratory Study. In *Research Methods for Cyber Security*; Elsevier: Amsterdam, The Netherlands, 2017.
35. Kingma, D.P.; Ba, J.L. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
36. Chok, N.S. Pearson's Versus Spearman's and Kendall's Correlation Coefficients for Continuous Data. Doctoral Thesis, University of Pittsburgh, Pittsburgh, PA, USA, 2010.
37. Waskom, M. Seaborn: Statistical Data Visualization. *J. Open Source Softw.* **2021**, *6*, 3021. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.