XI Congreso Ibérico de Agroingeniería
XI Congresso Ibérico de Agroengenharia
2021

Type of the Paper **(Oral)**

Virtual sensor for ventilation flux estimation in greenhouses

Mounir Guesbaya [1,*], Francisco García-Mañas [2], Francisco Rodríguez [2], Hassina Megherbi [3] and Mohamed Ridha Ouamane [4]

[1] University of Biskra, Department of Electrical Engineering, LI3CUB laboratory, BP 145 RP, Biskra, Algeria
[2] University of Almería, Department of Informatics, CIESOL, ceiA3, E04120 Almería, Spain; {francisco.gm; frrodrig}@ual.es
[3] University of Biskra, LARHYSS laboratory, BP 145 RP, Biskra, Algeria; h.megherbi@univ-biskra.dz
[4] INSA Centre Val De Loire, PRISME laboratory, 88 Bd Lahitolle, Bourges, France; mohamed.ouamane@insa-cvl.fr
* Correspondence: mounir.guesbaya@univ-biskra.dz

Abstract: Natural ventilation flux is an important variable to be measured or estimated to consider its significant effect on greenhouse microclimate modelling and control. It is commonly known that it can be mathematically estimated depending on the type and dimension of the greenhouse and its vents, and most importantly on the vents opening percentage. However, most of the commercial greenhouses are not equipped with an automatic vent opening system which obligates the grower to perform manual control, in addition to the lack of the vent position sensors, due to economic and management reasons. This leads to the absence of the control signal variable that represents the vents opening percentage necessary for ventilation flux estimation. To cope with this issue, the estimation of ventilation flux without using the vent opening percentage is proposed. In this paper, a virtual sensor of greenhouse ventilation flux is developed using a nonlinear autoregressive network with exogenous inputs based on principal component analysis using the available measured data and the evolutions of the heat fluxes representing the greenhouse energy balance. Preliminary results show an encouraging performance of the virtual sensor in estimating the ventilation flux with a mean absolute error of 0.41 $m^3 s^{-1}$.

Keywords: Protected agriculture; climate modelling; natural ventilation; principal component analysis; artificial neural networks; virtual sensor.

1. Introduction

Nowadays, the world is witnessing serious issues with the change of climate, the increase of population and the lack of resources. All the countries are investigating sustainable solutions in all the domains and especially in agriculture. Providing food for the growing population is one of the main issues for sustainable development in the United Nations' 2030 agenda [1]. To tackle the food necessity and achieve sustainable agriculture, the reliance on greenhouse systems is continuously increasing in the last years. Greenhouses have become a prominent mean in the agricultural field. They are constantly under development to reach the quantity and quality of crops that not only meet the world growing population but also the strict standards of the competitive markets today.

In order to sustain the best agricultural outcome, greenhouse ventilation flux is an important variable that should be studied and controlled according to its vital role in influencing the crop. To measure this variable, it is necessary to use special anemometers as ultrasound or thermal effect-based ones but since the installation of such sensors is not usual due to their high costs, it can be estimated using the proposed methods in the literature. In [2], the authors proposed several models for ventilation flux estimation for greenhouses with continuous roof vents. The models were fitted to the data by identifying two model parameters: coefficient of the pressure drop through the vents and coefficient of wind effect. In another paper [3], an air exchange model for a Mediterranean greenhouse with roof and side vents was developed. It proved that there are two main driving forces of the natural ventilation flux which are: wind effect and temperature effect (difference between inside and outside temperature), furthermore, that the total ventilation flux is linearly dependent on the total area of the openings of vents. In [4], both of the previous techniques were more investigated and integrated into a multi-input multi-output (MIMO) greenhouse model to estimate the ventilation flux and the obtained performance demonstrated their efficiency. However, the previous methods of estimating or predicting the greenhouse ventilation flux are all dependent on the total surface of the openings of vents. This could be considered as a problem of high complexity in greenhouses where the vents opening is performed manually by growers. Manual control means the absence of the control signal which in turn means the lack of the exact and continuous recording of vents opening percentage making it impossible to estimate the ventilation flux based on the mentioned explicit methods in hands which makes it a serious issue. As far as we know, this problem was not investigated in the literature and for this reason, the present work is focused on studying new possibilities to estimate the ventilation flux in greenhouses using alternative techniques as artificial neural networks.

An artificial neural network (ANN) is a machine learning algorithm inspired by the way how the human brain neuron network processes information. The ANN application in modelling greenhouse climate systems has been a subject of interest during many years. It has proven its efficiency in the literature as a very useful tool for these types of problems [5, 6]. An ANN type called the nonlinear autoregressive neural network with exogenous inputs (NARX) is used in this work. It is a recurrent dynamic network that allows relating the estimation of the output to the past inputs and outputs. In [7], a NARX has been used and its capability in estimating transpiration as a nonlinear phenomenon has been proven.

A useful statistical analysis tool was used for NARX modelling is the principal component analysis (PCA). It is the most used method for data exploration and data analysis across all fields of science. It is used with high-dimensional data to identify a reduced set of principal components (PCs) that represent the original dataset with a minimal loss of information leading to a low-dimension dataset [8]. Since the ventilation flux in greenhouses is affected by numerous and different variables, PCA could help to identify the most influential ones.

In this paper, a virtual sensor for greenhouse ventilation flux has been developed based on PCA-NARX modelling following the methodology illustrated in Figure 1. A dataset has been generated from a Mediterranean multi-span greenhouse located in Almería, Spain including a combination of measured microclimate variables

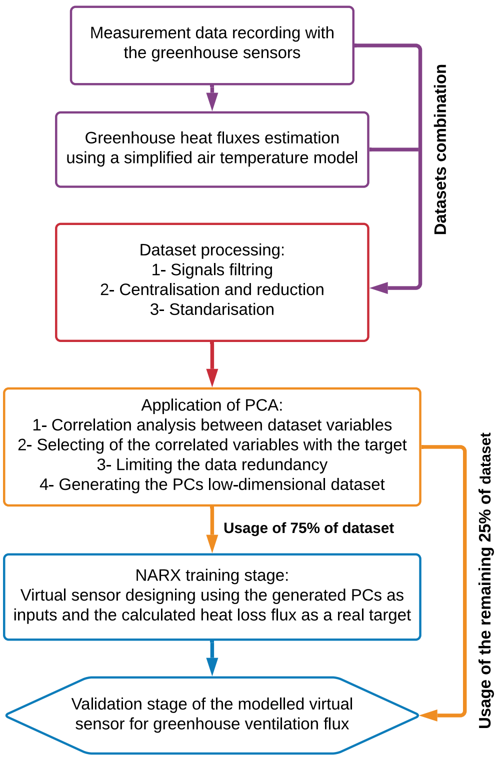

Figure 1. The methodology for developing the PCA-NARX-based virtual sensor for greenhouse ventilation flux.

and the evolutions of model heat fluxes. The model heat fluxes were calculated using an adaptive air temperature model due to its capability of providing their optimal estimations (error of <5% between a test and another under the same conditions) and reliable for estimating the greenhouse ventilation flux without installing expensive sensors. All the obtained variables were firstly processed by: signal filtring, centralization, reduction and standardisation. Secondly, the treated dataset was used to generate the PCs for data reduction using PCA. These PCs are then considered the new inputs of the neural network. Thus, the network was trained based on the PCs to fit the target which is the estimated heat loss using the opening percentage of roof and side vents based on the pre-mentioned explicit approach [3]. Finally, the estimated heat loss flux was used to inversely calculate the ventilation flux representing the ultimate objective in the proposed virtual sensing method. The validation of this developed virtual sensor has shown very promising preliminary results.

The rest of this paper is organized as follows. In section 2, the used methods and materials are described. Section 3 includes network training and validation results and their discussion. Section 4 concludes this paper.

2. Materials and Methods

2.1. Greenhouse description

Studies in this work were carried out using the greenhouse presented in Figure 2. It is a polyethene-covered traditional Mediterranean greenhouse, known as the “Almería-type” greenhouse. It is located at “Las Palmerillas” Experimental Station which is a property of the Cajamar Foundation (36.79316 latitude, -2.72014 longitude), in Almería, Spain at an altitude of 151 m. Its total surface is 877 m^2 (37.80 m x 23.20 m), and it is dedicated to tomato crops. The greenhouse is equipped with a natural ventilation system, including five zenithal windows (8.36 m × 0.73 m) that are installed on the roof of the structure and two lateral windows (32.75 m × 1.90 m) that are situated along the north and south sidewalls of the cover.

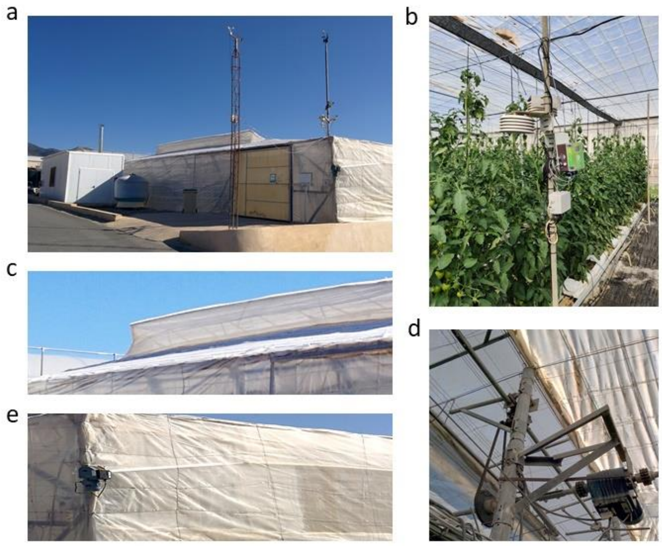

Figure 2. Greenhouse facilities used for the experimental tests. (a) exterior view; (b) interior view with an example of a data acquisition device; (c) exterior view of roof vent (d) interior view of roof vent; (e) exterior view of lateral vent.

2.2. Adaptive estimation of greenhouse heat fluxes

To estimate the evolutions of the greenhouse heat fluxes, an adaptive simplified model of greenhouse air temperature is used in this work. The structure of the model adaption method is presented in Figure 3, and it is entirely explained in [9]. The greenhouse temperature model (see Subsection 2.2.1) is adapted to estimate the greenhouse heat fluxes thanks to an online parameter estimation mechanism (see Subsection 2.2.3). The online parameter estimator changes the values of the parameters of the heat flux equations to reduce the error between the predicted and the real measured air temperature inside the greenhouse.

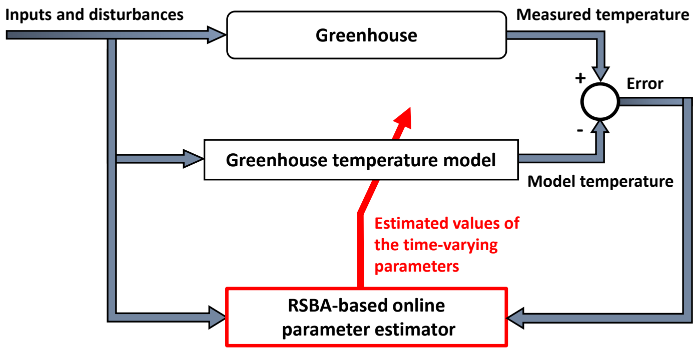

Figure 3. Adaptive model using the RSBA-based online parameter estimator [9].

2.2.1. Simplified air temperature model

A pseudo-physical model has been used in this work to obtain the evolution in time of the air temperature inside the greenhouse according to the following energy balance:

$$C_{sph}\, C_{den}\, \frac{C_{vol}}{C_{area}}\, \frac{dXT_{in}}{dt} = Q_{sol,a} + Q_{cnv,ss-a} - Q_{cnd-cnv,a-e} - Q_{trp,cr} - Q_{ven,a} \qquad (1)$$

where XT_{in} (K) is the inside air temperature of the greenhouse, Q represent the heat fluxes (W m^{-2}), where $Q_{sol,a}$ is the solar radiation absorbed by the air, $Q_{cnv,ss-a}$ is the convective flux with the soil surface, $Q_{cnd-cnv,a-e}$ is the convective and conduction flux (through the cover) between the inside and outside air, $Q_{trp,cr}$ is the latent heat effect due to crop transpiration, and $Q_{ven,a}$ is the heat lost due to natural ventilation. C_{sph} (J kg^{-1} K^{-1}) is the air specific heat, C_{den} (kg m^{-3}) is the air density, C_{vol} (m^3) is the greenhouse volume, and C_{area} (m^2) is the greenhouse surface. In the sequel, only the details of the natural ventilation heat flux are provided. A complete description of this model with a detailed explanation of the equations to calculate the different heat fluxes can be found in [4].

2.2.2. Ventilation flux model

Since the greenhouse used in this work has lateral and roof windows, the ventilation flux $V_{ven,flux}$ (m^3 s^{-1}) is estimated with the next equation:

$$V_{ven,flux} = c_{ven,d} \left[\left(\frac{V_{ven,area-lat}\, V_{ven,area-roof}}{\sqrt{V_{ven,area-lat}^2 + V_{ven,area-roof}^2}} \right)^2 \left(2\, c_g\, c_{ven,h}\, \frac{XT_{in} - DT_e}{DT_e} \right) + \right.$$
$$\left. + \left(\frac{V_{ven,area-lat}\, V_{ven,area-roof}}{2} \right)^2 c_{ven,wd}\, D_{ws,e}^2 \right]^{0,5} + V_{loss}\,, \qquad (2)$$

where $D_{ws,e}$ (m s^{-1}) is the external wind velocity, $c_{ven,d}$ (dimensionless) is the discharge coefficient, $c_{ven,wd}$ (dimensionless) is the wind effect coefficient, c_g (m s^{-2}) is the gravity constant, $c_{ven,h}$ (m) is the vertical distance between the midpoints of the lateral and roof vents, and V_{loss} (m^3 s^{-1}) is the leakage when the vents are closed (e.g., due to filtrations through small gaps in the cover), which is also dependent on external wind speed.

In Eq. (3), the areas of the lateral and roof windows, $V_{ven,area-lat}$ and $V_{ven,area-roof}$, respectively, are calculated with the following expressions:

$$V_{ven,area-lat} = c_{ven,n-lat}\ c_{ven,l-lat} c_{ven,w-lat} \left(\frac{U_{ven}}{100}\right) \quad (3)$$

$$V_{ven,area-roof} = 2\ c_{ven,n-roof}\ c_{ven,l-roof}\ c_{ven,w-roof}\ sin\left(\frac{U_{ven}}{100}\ \frac{c_{ven,max}}{2}\right) \quad (4)$$

where $c_{ven,n-[lat-roof]}$ are the number of each type of vent, $c_{ven,l-[lat-roof]}$ (m) is the length, and $c_{ven,w-[lat-roof]}$ is the width, all referred to as the lateral or roof vents. For the roof windows, their maximum opening is limited by an angle of $c_{ven,max}$ (deg). As it can be noticed in Eqs. (4) and (5), the opened area of the windows of the greenhouse depends on the percentage of the opening of the ventilation, expressed as U_{ven} (%), which can be a value provided by a controller to regulate the inside air temperature [10] or it can be manually set by the farmers.

The heat lost due to natural ventilation is estimated by the next expression:

$$Q_{ven,a} = \frac{c_{sph}\, c_{den}}{c_{area}}\ V_{ven,flux}\ (XT_{in} - DT_e) \quad (5)$$

where DT_e (K) is the external air temperature. The ventilation flux depends on the structure of the greenhouse, the external wind velocity, as well as on the difference between tempertures.

2.2.3. RSBA-based online estimation of model parameters

The random scaling-based bat algorithm (RSBA) is a variant of the bat algorithm, which is a nature-inspired optimization method that imitates bats' random walk technique and their searching on prey using echolocation. In the adaptive model utilized in this work, the RSBA is in charge of finding optimal values for the model parameters in order to minimize the error between the model output and the real air temperature measured inside the greenhouse. The mathematical formulation of this algorithm was described in previous work [11] and it was successfully tested in a real greenhouse for the first time in [9], in which an extensive analysis of its performance and its optimality was offered by studying the evolution of the estimated parameters and the heat fluxes affecting the inside air temperature according to Eq. (1).

The RSBA-based online parameter estimator was tuned in [9] with a search range restriction for the values of each time-varying parameter. Moreover, the RSBA was prepared to use a dynamical search range based on a variation ratio for each parameter. Specifically, the parameter value in each run is constrained to a neighbourhood of variation around the last optimal value (change every run). This strategy avoids the appearance of sudden changes in the value of certain parameters that physically evolve more slowly. The search ranges and variation ratios for the relevant time-varying parameters of the adaptive model are presented in Table 1. These limitations are needed to preserve the physical sense of the parameters and to ensure an adequate estimation of the related heat fluxes.

Table 1. Search ranges and variation ratios for some of the time-varying parameters of the air temperature model.

Parameter	$c_{asw,a}$	$c_{cnv,ss-a}$	$c_{cnd-cnv,a-e}$	$c_{ven,d}$	$c_{ven,wd}$	c_{loss}
Range	[0.1, 0.9]	[1, 100]	[1, 300]	$[15, 35]\cdot 10^{-4}$	[0.1, 1]	[0.1, 1]
Units	-	$W\ m^{-2}\ K^{-1}$	$W\ m^{-2}\ K^{-1}$	-	-	$m^3\ s^{-1}$
Variation ratio	±2%	±10%	±10%	±7%	±7%	±7%

2.3. Greenhouse microclimate dataset

The greenhouse microclimate dataset used in this work to design the virtual sensor of ventilation flux consists of two combined parts which are:

- An experimental dataset collected from the greenhouse containing measured variables is presented in Figure 4. It includes: the wind velocity $D_{ws,e}$, the difference between the internal and external measured air temperature $T_{air,diff}$ and as well as relative humidity $H_{air,diff}$.

- An simulated dataset consisting of the estimated evolutions of the greenhouse heat fluxes is presented in Figure 5. It includes: $Q_{sol,a}$, $Q_{cnv,ss-a}$, $Q_{cnd-cnv,a-e}$, $Q_{trp,cr}$ and $Q_{ven,a}$ that were estimated based on the equations of the adaptive air temperature model in which the measured variables were supplied as inputs. Substantially, this was performed by executing the model in two scenarios:

a. Using the available vents control signal as one of the model inputs to consider the explicit effect of ventilation flux calculated based on Eq. (2) on the heat loss flux calculated based on Eq. (5), and in turn, on the model heat balance. Thus, the calculated variables of internal ventilation and heat loss fluxes are considered the real targets in designing the virtual sensor of ventilation flux.

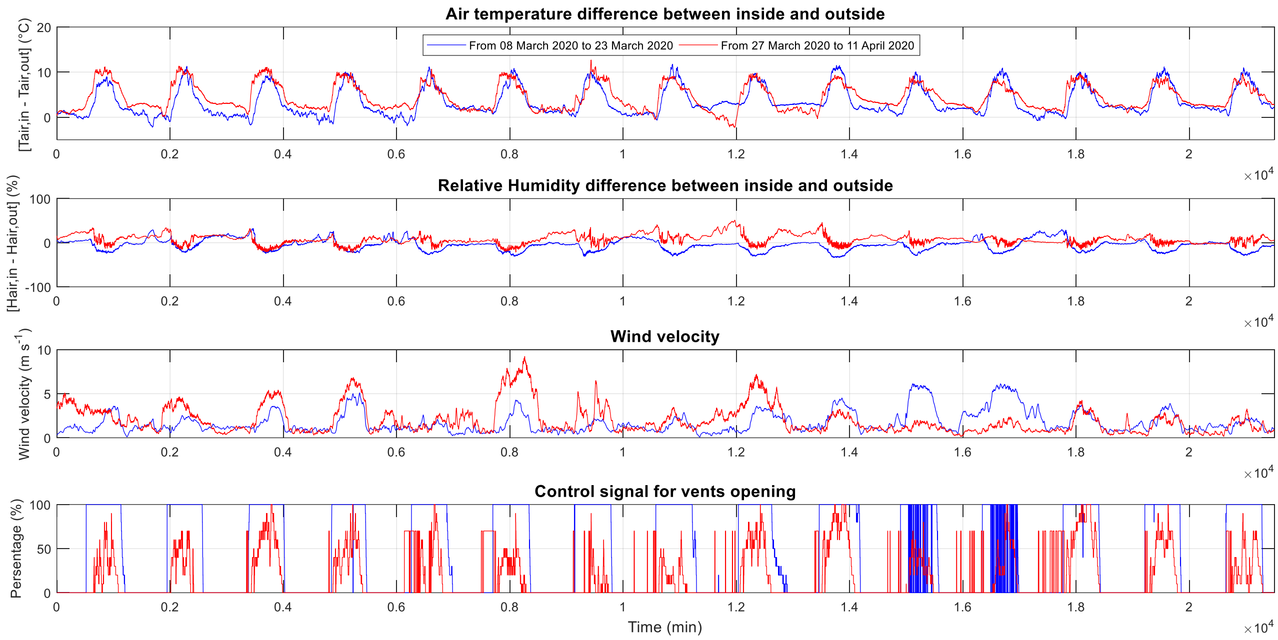

Figure 4. Dataset of the measured microclimate variables in the transitional period between winter and spring seasons.

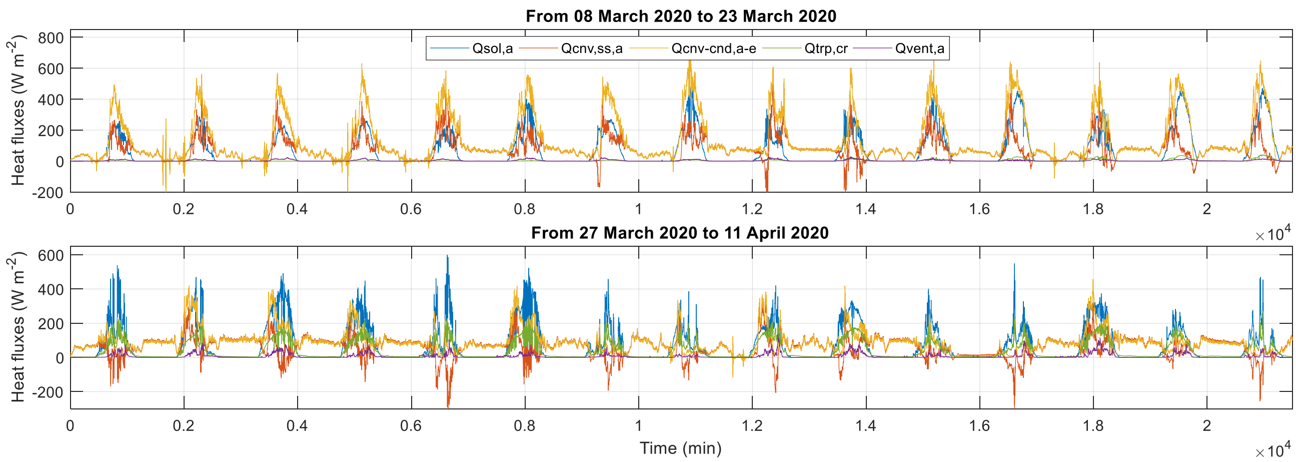

Figure 5. Dataset of the calculated greenhouse heat fluxes in the transitional period between winter and spring seasons.

b. Without using the vents control signal which means that the ventilation flux and the heat loss flux are not influencing explicitly (Indirect effect) heat balance of the air temperature model (using the same Eq. (1) but with $Q_{ven,a} = 0$). Since the used air temperature model is adaptive thanks to the RSBA-online parameter estimator, the effect of the heat loss flux will be adaptively compensated by certain behaviours in the evolutions of the other heat fluxes. This has led to an accurate prediction of internal air tempertaure which is considred the criterion for a successful estimation of heat fluxes. Thus, the estimated heat fluxes in fact hold implicitly the ventilation flux information (ventilation effect). Accordingly, the heat fluxes are considered as useful data to be used as inputs for designing the virtual sensor of ventilation flux.

All variables were processed by centralization, reduction and standardisation and some of them were filtered.

2.4. Nonlinear autoregressive network with exogenous inputs

A discrete-time NARX model is used in this work. It consists of a recurrent dynamic network with feedback connections including numerous layers of the network [7]. This model is defined by the following function:

$$y(k) = f(y(k-1), \dots, y(k-d_{output}), \mathrm{u}(k-1), \dots, u(k-d_{input})) \quad (6)$$

where $y(k)$ is the greenhouse ventilation flux estimated by the model at the discrete-time step k, u is the column vector of inputs, d_{input} and d_{output} are the orders of the past inputs and outputs, respectively, to be used for producing $y(k)$. The NARX was chosen for its advantage in relating the output to the past inputs and outputs.

2.5. Principle components analysis

PCA is a multivariate technique that transforms a set of correlated variables into a smaller set of uncorrelated variables, called principal components (PCs). The power of PCA is more apparent for a larger number of variables. Several PCA applications in the field of greenhouse data analysis have been investigated. In [12] the PCA was used to simplify the data samples and to optimize the model learning speed for internal air humidity modelling. In [13], neural networks were combined with PCA to explain mathematically the redundancy between variables in order to simplify the complex model by keeping the efficient one. In this paper, the PCA is used to choose the relevant information from the greenhouse microclimate variables of the available dataset.

The starting point for PCA is the sample covariance matrix S. The covariance matrix for the p-variable is calculated as follows:

$$S = \frac{1}{1-n} X_c^T X_c \ \in \mathbb{R} \quad (7)$$

$$S = \begin{pmatrix} s_1^2 & \cdots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{1p} & \cdots & s_p^2 \end{pmatrix} \quad (8)$$

where n is the number of samples and X_c is the centred data around the mean value and it is calculated as follows:

$$X_c = X - X_{mean} \quad (9)$$

Secondly, the obtained covariance matrix S is reduced to a diagonal matrix L by pre-multiplying it and post-multiplying it by a particular orthonormal matrix U. It is calculated as follows:

$$U^T S U = L \quad (10)$$

where s_i^2 is the variance of the ith variable and s_{ij} is the covariance between the ith and j^{th} variables.

The diagonal elements of L are called the eigenvalues of S. The cumulative variance contribution of each principal component is calculated with the next expression:

$$W_{PC}(i) = \frac{l_i}{\sum_{j=1}^{m} l_j} \tag{11}$$

The columns of U are called the eigenvectors, which represent the direction of maximal variability for each variable. These vectors are used for analyzing the correlation between variables. Finally, the uncorrelated components Z are calculated as follows:

$$Z = U^T X_c \tag{12}$$

The PCA correlation circle is also used to study the relation between the inputs and the target and to reveal any possible redundancy of data and avoid it by eliminating the variables that approximately hold the same information.

3. Results and Discussion

This section presents the results and observations obtained from each stage of the proposed methodology (see Figure 1) and their corresponding discussion toward designing and validating a virtual sensor for greenhouse ventilation flux.

3.1. PCA application

The application of the PCA on the available dataset was performed for two purposes as described in the following sub-section.

3.1.1. Data analysing

Initial PCs were generated based on the full dataset including the inputs (generated without considering $V_{ven,flux}$ effect) and the target (generated with considering $V_{ven,flux}$ effect). Figure 6 shows the correlations between the PC1 and PC2 that hold the largest amount of information from the original data via

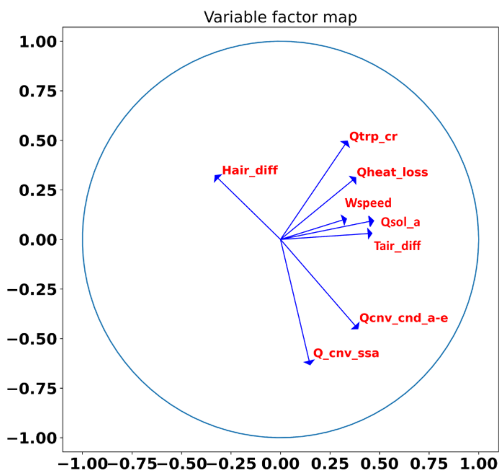

Figure 6. Correlation circle between PC1 and PC2.

coordinates in a 2-dimensional plot. Based on the fact that the correlations are proportional to the angle between vectors, it can be noticed that: (I) $Q_{ven,a}$ which represents the target has a strong positive correlation with $Q_{trp,cr}$, $D_{ws,e}$, $Q_{sol,a}$ and $T_{air,diff}$ since their vectors are grouped in the neutral quadrant; (II) $Q_{cnv-cnd-a-e}$ and $H_{air,diff}$ have much lower correlation since they are orthogonal with the target; (III) $Q_{cnv,ss-a}$ has a low but noticeable correlation with the target which is negative since it is in the opposing quadrant.

The relation between the inputs and the target was also analysed with the Pearson correlation coefficient using datasets of different climate conditions as illustrated in Table 2. It can be noticed that: (I) $Q_{trp,cr}$ appears to be the most correlated variable in the different calm, windy and cloudy days. This could be physically explained as a result of the direct effect of internal air temperature and the indirect effect of relative humidity through vapour pressure deficit on the crop transpiration. Alos, both $T_{air,diff}$ and $H_{air,diff}$ also have high positive and negative correlations, respectively, with heat loss since ventilation flux affects them directly. $Q_{cnv-cnd,a-e}$ could also be considered as one of the most correlated variables with the target based on the shown correlation values because it is calculated based on $T_{air,diff}$. $D_{ws,e}$ is known to be the main driving force of natural ventilation in greenhouses but it does show the only strong correlation with heat loss except for the calm day, however, it is still considered a useful input. $Q_{sol,a}$ has an interesting noticeable positive correlation especially on the calm day. This could be considered an effect as a consequence of the model adaptation using the RSBA-based online parameter estimator. $Q_{sol,a}$ has probably highly compensated the implicit effect of heat loss due to ventilation, because of its relatively fast ratio of variation (±7%/1min) leading to its ease of manipulation by the online estimator (with respect to its physical constraint) and its strong effect on the model output. $Q_{\mathrm{cnv,ss-a}}$ has the weakest correlation due to its relation with the soil surface temperature which changes slowly far from the effect of the rapidly varying ventilation flux, thus, it could be eliminated to lower the computational cost.

Table 2. Pearson-based correlation coefficients between the input variables and the target in different climate conditions.

	Climate type	$D_{ws,e}$	$T_{air,diff}$	$H_{air,diff}$	$Q_{sol,a}$	$Q_{cnv,ss-a}$	$Q_{cnv-cnd,a-e}$	$Q_{trp,cr}$
Qheat,loss	Calm day	0.656	0.663	- 0.627	0.736	0.069	0.476	0.809
	Windy day	0.367	0.397	- 0.325	0.472	- 0.256	0.264	0.516
	Cloudy day	0.381	0.040	- 0.138	0.475	0.025	0.325	0.708
	All days	0.357	0.669	- 0.610	0.547	0.299	0.729	0.342

3.1.2. Generating of PCs

Four new essential PCs are generated using the full dataset combining the two datasets from the winter and spring seasons. In this case, the target $Q_{ven,a}$ has not been included because these PCs are the ones to be used as inputs of the developed PCA-NARX model. Figure 7 shows the specific amount of information maintained by

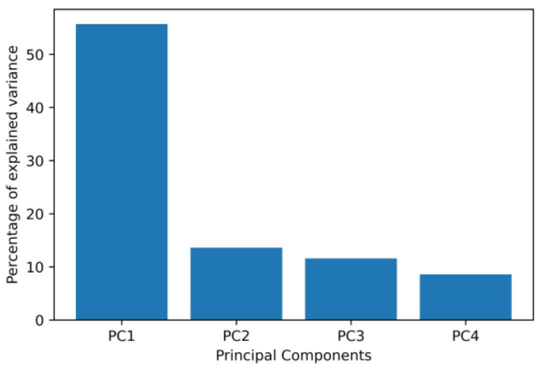

Figure 7. Amount of information maintained by the principal components constituting a low-dimensional dataset.

each PC and the total amount of information maintained by all the PCs is 93.8% which is considered sufficient. It can be noticed that PC1 include the largest amount of maintained information. The dimension of the dataset has been decreased from 7 to 4 variables

3.2. Virtual sensor design

The designed NARX model was trained using the generated PCs as inputs and the calculated heat loss due to ventilation as a target. Several training processes were performed using 75 % of the dataset for training and the remaining 25% for validation. The NARX model was trained using different orders of feedback output parameters using the Scaled conjugate gradient backpropagation (trainscg) as a training function and a tapped delay line with a delay from 0 to 5 samples at the input and also from 1 to 5 samples at the output. The best structure of the NARX model included: 4 layers in total and 2 hidden layers, they consist of 30, 50, 20 and 1 neuron, respectively, as it is shown in Figure 8. The simulation results after training are very satisfactory where the estimated variable follows the variations efficiently as it is qualitatively shown in Figure 9. Settings of the NARX model design were chosen in a way that helps to avoid the overtraining of the network and enhances the generalisation in its dynamics, especially because only a short dataset was used in this preliminary study of this problem. The quantitative result of this training process present a mean absolute error MAE = 1.88 W m^{-2} which represents a percentage of 1.34 % in a variation interval of [-2.98, 137.26], which is considered sufficient although the variations of the target are not fully captured by the trained model.

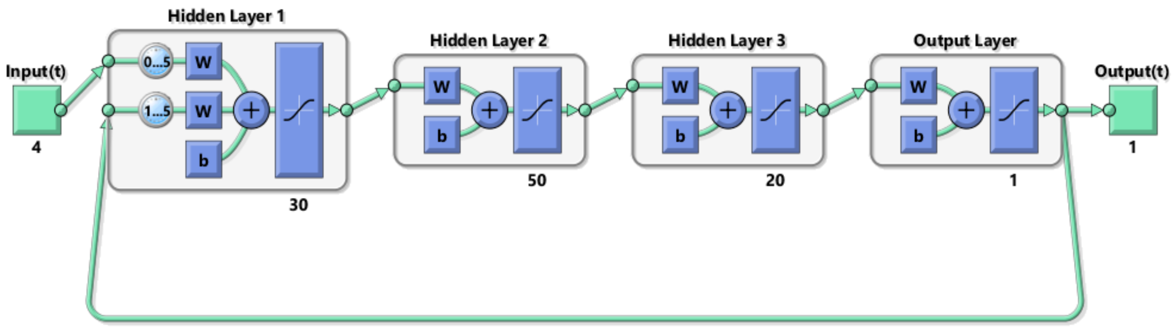

Figure 8. NARX network structure.

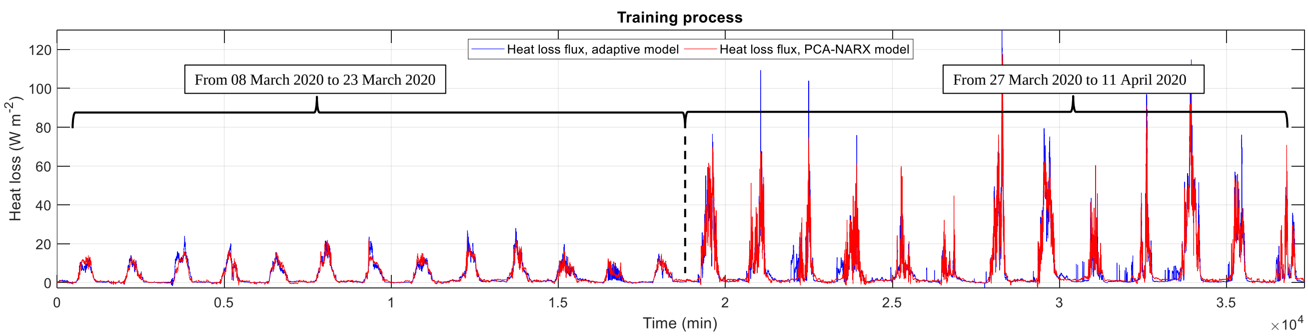

Figure 9. Traning of the PCA-NARX model using the obtained PCs as inputs and the calculated heat loss flux as a target.

3.3. Virtual sensor validation

The best PCA-NARX model among the tested ones is selected to be the preliminary developed virtual sensor for greenhouse ventilation flux in this work. The estimation of ventilation flux was based on the calculated heat loss flux using the inverse formula of $Q_{ven,a}$ (inverse of Eq. (2)). The best PCA-NARX model was resulted after training and validating the network through a set of trial and error processes in two cases:

1- Using a dataset combining both datasets obtained from different periods in winter and spring seasons.
2- Using only one dataset of the first period recorded in the winter season.

As presented in Table 3, the preliminary quantitative results using the combined dataset show better performance than only using a separated dataset of one period presenting a $\text{MAE} = 3.26\ \text{W m}^{-2}$ for simulating the heat loss flux and leading to a $\text{MAE} = 0.41\ \text{m}^3\text{s}^{-1}$ which are both considered very promising as preliminary results. Figure 9 shows the qualitative results and a positive performance can be observed in both diurnal and nocturnal periods and for both different datasets affected by the different vents control signals. This supports the fact that the more diverse and large the dataset used for training is, the more the dynamics of the resulting model are driven by the supplied data rather than the bias parameters added by the ANN, and the more accurate the output is in terms of capturing the variations in the target evolution. Simulating such a strongly nonlinear phenomenon with such fast dynamics requires more than the provided information in this work. However, it is still considered as a preliminary result to help in leading to a better version of this solution.

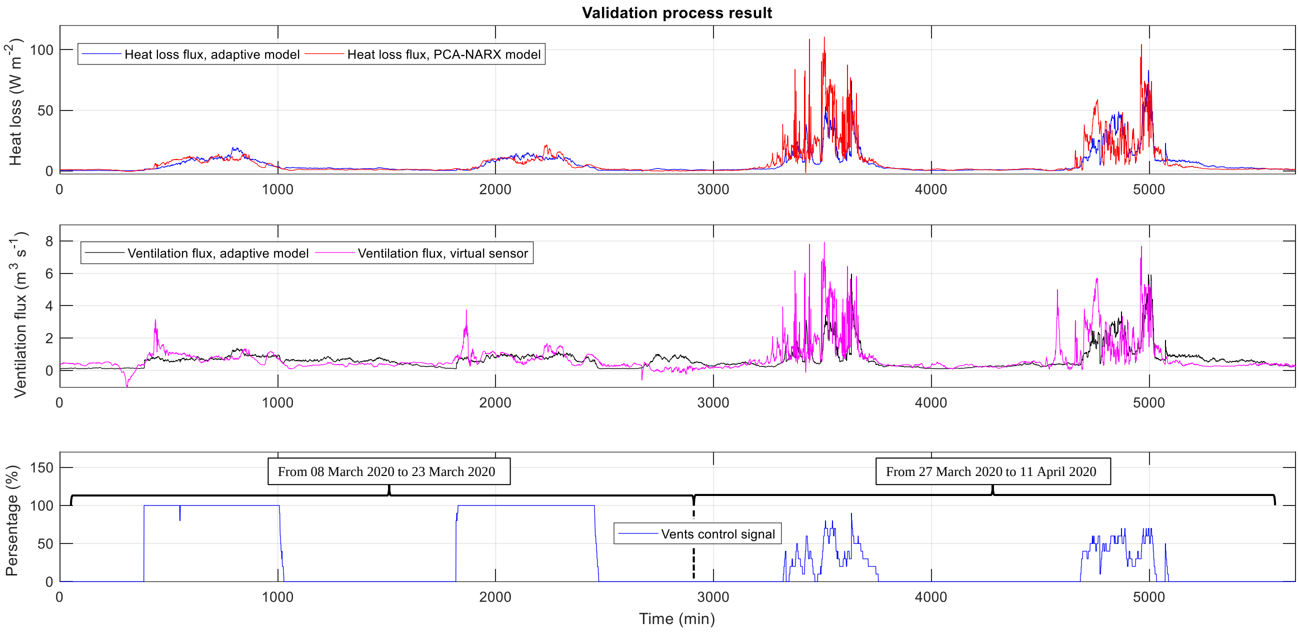

Figure 9. Validation of he virtual sensor for greenhouse ventilation flux estimation based on the PCA-NARX model

Table 3. Quantitative results of the validation process of the PCA-NARX virtual sensor

		MAE	Error persentage	Variation interval
Using the combined dataset	$Q_{\text{ven,a}}\ (W\ m^{-2})$	3.26	3.93 %	[0.08, 83.11]
	$V_{ven,flux}\ (m^3 s^{-1})$	0.41	4.57 %	[0.1, 5.96]
Only one dataset: from 27 March 2020 to 11 April 2020	$Q_{\text{ven,a}}\ (W\ m^{-2})$	6.27	5.47 %	[0.14, 114.75]
	$V_{ven,flux}\ (m^3 s^{-1})$	0.95	14.23 %	[0.1, 12.74]

4. Conclusions

In greenhouses, the absence of measurements for the opening percentage of vents (vents control signal) complicates the calculation of the ventilation flux. For this reason, a virtual sensor for greenhouse ventilation flux has been proposed in this work. It is designed based on a PCA-NARX model following the methodology illustrated in Figure 1. A measurements dataset generated from a Mediterranean multi-span greenhouse located in Almería, Spain, has been used. A complete dataset including a combination of the measured microclimate variables and the calculated evolutions of the heat fluxes of an air temperature model, all in from different periods of winter and spring seasons. A set of PCs was generated using PCA with the dataset, leading to a significant data dimensional reduction. The resulted components are considered the new inputs of the neural network and the calculated heat loss flux due to natural ventilation was considered a real target. Thus, the network was trained using the prepared inputs to fit the real target to be used to estimate the ventilation flux as the main objective.

Preliminary quantitative and qualitative results have been obtained in this work. The validation process of the developed virtual sensor has presented a $\mathrm{MAE} = 0.41\ \mathrm{m^3 s^{-1}}$ which represents an error percentage of 4.57% between the estimated ventilation flux without considering the vents control signal and the calculated ventilation flux using the vents control signal. The qualitative result shows that the variations of ventilation flux are captured acceptably by the virtual sensor even though the vents control signal included different patterns.

As conclusions and future perspectives, the preliminary results obtained in this work support the fact that developing a superior and reliable virtual sensor for long-term applications calls some very important requirements and potential procedures that could be highlighted as follows:

- A large dataset of at least one year or several datasets of different seasons have to be provided, consisting of the needed measured and calculated variables representing the inputs and the target for the proposed PCA-NARX model.
- More greenhouse environment dynamics (e.g., the difference between internal and external CO_2 concentration) should be analysed to obtain more correlated variables to the greenhouse ventilation flux if they exist.
- Correlation between the target and the provided measured and calculated data has to be more investigated from different perspectives using other data analysis techniques.
- Different structures of the NARX model and other data-driven modelling methods have to be investigated.
- Based on the fact that the correlation between the inputs and the target can change depending on the climate conditions (calm, windy and cloudy days … etc), a classification technique and multiple data-driven models can be obtained for each climate scenario for more accurate performance.

Author Contributions: Conceptualization, Rodríguez, F., Guesbaya, M and García-Mañas, F; methodology, Guesbaya, M and Rodríguez, F; visualization, Guesbaya, M and Rodríguez, F; software, Guesbaya, M; validation, Guesbaya, M; investigation, Guesbaya, M; resources, García-Mañas, F and Ouamane, M, R; data curation, Guesbaya, M., García-Mañas, F and Ouamane, M, R; writing—original draft preparation, Guesbaya, M, García-Mañas, F and Ouamane, M, R; writing—review and editing, Guesbaya, M, García-Mañas, F, Rodríguez, F and Megherbi, H; supervision, Rodríguez, F and Megherbi, H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National R+D+i Plan Project DPI2017-85007-R of the Spanish Ministry of Science, Innovation and Universities and ERDF funds.

Acknowledgements: Author M. Guesbaya is supported by the Exceptional National Program (PNE) grant of the Algerian Ministry of High Education and Scientific Research. Author F. García-Mañas is supported by an FPU grant from the Spanish Ministry of Science, Innovation and Universities. The authors appreciate the collaboration of Cajamar Foundation. The authors would like to thank Pr. José Luis Guzmán Sánchez (Full Professor, University of Almería) for his assistance.

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

References

[1] Akrami, M.; Salah, A.H.; Javadi, A.A.; Fath, H.E.; Hassanein, M.J.; Farmani, R.; Dibaj, M.; Negm, A. Towards a sustainable greenhouse: Review of trends and emerging practices in analysing greenhouse ventilation requirements to sustain maximum agricultural yield. *Sustainability.* 2020. vol. 12. no. 7. pp. 2794.

[2] Boulard, T.; Baille, A. Modelling of air exchange rate in a greenhouse equipped with continuous roof vents. *Journal of Agricultural Engineering Research.* 1995. vol. 61. no. 1. pp. 37-47.

[3] Kittas, C.; Boulard, T.; Papadakis, G. Natural ventilation of a greenhouse with ridge and side openings: sensitivity to temperature and wind effects. *Transactions of the ASAE.* 1997. vol. 40. no. 2. pp. 415-425.

[4] Rodríguez, F.; Berenguel, M.; Guzmán, J.L.; Ramírez-Arias. R. *Modeling and control of greenhouse crop growth*. 2015. Springer International Publishing Switzerland.

[5] Fourati, F. Multiple neural control of a greenhouse. *Neurocomputing.* 2014. vol. 139. pp. 138-144.

[6] Taki, M.; Mehdizadeh, S.A.; Rohani, A.; Rahnama, M.; Rahmati-Joneidabad, M. Applied machine learning in greenhouse simulation; new application and analysis. *Information processing in agriculture.* 2018. vol. 5. no. 2. pp. 253-268.

[7] Wang, H.; Sánchez-Molina, J.A.; Li, M.; Berenguel, M.; Yang, X.T.; Bienvenido, J.F. Leaf area index estimation for a greenhouse transpiration model using external climate conditions based on genetics algorithms, back-propagation neural networks and nonlinear autoregressive exogenous models. *Agricultural Water Management.* 2017. vol. 183. pp. 107-115.

[8] Kherif, F. and Latypova, *Principal component analysis*. Machine Learning: Methods and Applications to Brain Disorders, Academic Press, 2019, pp. 209-225.

[9] 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. *Computers and electronics in agriculture.* (under review).

[10] Montoya-Ríos, A.P.; García-Mañas, F.; Guzmán, J.L.; Rodríguez, F. Simple tuning rules for feedforward compensators applied to greenhouse daytime temperature control using natural ventilation. *Agronomy.* 2020. vol. 1327. pp. 10.

[11] Guesbaya, M.; Megherbi, H.; Megherbi, A.C. Random Scaling-Based Bat Algorithm for Greenhouse Thermal Model Identification and Experimental Validation. *International Conference on Electrical Engineering and Control Applications.* 2019. Singapore.

[12] He, F.; Ma, C. Modeling greenhouse air humidity by means of artificial neural network and principal component analysis. *Computers and Electronics in Agriculture*. 2010. vol. 71. pp. 19-23.

[13] Pessel, N.; Balmat, J.F. Principal component analysis for greenhouse modelling. *WSEAS Transactions on Systems.* 2008. vol. 1. no. 7. pp. 24-30.

[14] Ma, D.; Carpenter, N.; Maki, H.; Rehman, T.U.; Tuinstra M.R.; Jin J. Greenhouse environment modeling and simulation for microclimate control. *Computers and Electronics in Agriculture*. 2019. vol. 162. pp. 134-142.