IJAIR

International Journal of Agriculture Innovations and Research
Volume 4, Issue 1, ISSN (Online) 2319-1473
Manuscript Processing Details (dd/mm/yyyy) :
Received : 28/05/2015 | Accepted on : 04/06/2015 | Published : 18/07/2015

A Multivariable Computational Fluid Dynamics Validation Method Based in Bayesian Networks Applied in a Greenhouse

Tlaloc Daniel Espinoza-Huerta, Irving Cesar Ortíz-Vázquez, Gabriela García-Manzo, *Guillermo Alfonso De la torre-Gea
Universidad Tecnológica de Corregidora, Querétaro, México.
*Corresponding author. E-mail: gtorre@abanet.mx .Tel: (52) (442) 4830070.

***Abstract* — In recent years, studies have been conducted to determine gradients expressing climate conditions inside greenhouses using computational fluid dynamics, which has been able to increase the precision and realism. Although these models provide numerical solution of the energy balance, these models need to be validated to establish its accuracy with respect to real data measured by sensors inside the greenhouse. Many methods have been employed validation, however at this time, neither limit has been reported to determine whether a model is acceptable or not. Bayesian networks are numerical techniques of uncertainty and heuristic method that could be used as validation CFD multivariable methods. We validated a Computational Fluid Dynamic model using a Bayesian Network analysis. BN have shown the relationships between variables CO2 concentration and humidity with respect to temperature and it has also been possible to quantify these relationships by calculating inferences in dependent probability distributions**.

***Keywords* – Greenhouse, Bayesian Networks, Natural Ventilation, Climate Distribution.**

I. Introduction

Computational Fluid Dynamic (CFD) models have been developing in the last decade, and one of the major problems is the validation methods. Some CFD studies have focused on simulate with more realism the steady state models to define the conditions, using different validation methods. The small-scale validation method can be useful to examine the complex problems of ventilation if similarity conditions of flow can be maintained between the model and reality. The large-scale validation method is the most reliable in predicting the efficiency of ventilation, but is expensive and time consuming [5].

Once a CFD model has been validated, it can be used to study the effect of temperature on air motion inside greenhouses (Rico-García et al., 2008). The quality of the results is often deduced from the agreement with experimental data. Nevertheless, no standard procedure exists to really assess the accuracy of the simulations, and the type of comparison often differs from one study to the next (Bournet and Boulard, 2010). Chen [3], evaluated seven types of models to predict the ventilation rate in crops, obtaining that CFD model provides more detailed information on the performance of ventilation and is the most sophisticated. However, the model must be validated by corresponding experimental data and the user must have solid knowledge of fluid mechanics and numerical techniques.

Model accuracy is generally assessed by just graphically comparing both numerical and experimental distributions and no criterion was used to quantify the adequacy of models with experiments. At this time, neither limit has been reported to determine whether a model is acceptable or not [1]. Therefore, the choice of an appropriate validation model depends on the problem to be solved.

CFD models provide a numerical solution of an energy balance in a controlled volume, that in comparison with other methods and expensive technologies allow for an effective climate within the greenhouse. However, CFD techniques take into account the values of the main independent variables as unknowns, in a finite number of locations within the domain. Due to the above, it is necessary to determine the correlation between the values of the variables calculated by this method.

Determining the relationships between climatic variables inside a greenhouse with natural ventilation is difficult, due to the inherent stochastic nature of the airflow. Bayesian networks (BN) are uncertain numerical techniques, which using Bayesian inference can be helpful to describe the relationships between the variables that define the ventilation conditions [4]. The aim of this work is to test a BN model as a validation method of CFD model.

II. Bayesian Networks Theory

BN are types of knowledge representation developed in the field of Artificial Intelligence for approximate reasoning [7,10,11]. A BN is a direct acyclic graph whose nodes correspond to concepts or variables and whose links correspond to relationships or functions [2]. Variables are defined in a discrete or qualitative domain, and functional relationships describe causal inferences expressed in terms of conditional probabilities that are shown in equation (1).

$$P(x_1,\ldots,x_n) = \prod_{i=1}^{n} P(x_i \mid parents\ (x_i)) \qquad (1)$$

BN can be used to identify previously undetermined relationships among variables or to describe and quantify these relationships even with an incomplete data set [12]. The solution algorithm of BN allows the computation of the expected probability distribution of output variables. The result of this calculation is dependent on the probabilities distribution of input variables. Globally, BN can be perceived as a joint probabilities distribution of a collection of discrete random variables [8].

International Journal of Agriculture Innovations and Research
Volume 4, Issue 1, ISSN (Online) 2319-1473

$$P(c_j \mid xi_{)} = P(x_i \mid c_j)\, P(c_j) / \sum_k P(x_i \mid c_k)\, P(c_k) \qquad (2)$$

A priori probability $P(c_j)$ is the probability that a sample belongs to class c_j, which is given no information on their characteristic values, as shown in equation (2). Machine learning in artificial intelligence is closely related to data mining, classification or clustering methods in statistics, inductive reasoning, and pattern recognition. Statistical machine learning methods can apply the framework of Bayesian statistics; however, machine learning can employ a variety of classification techniques to produce models other than BN. The objective of BN structure learning is to find a configuration that best describes the observed data. The number of possible structures of direct acyclic graph for searching is exponential in the number of variables in the domain, defined in (3):

$$f(n) = \sum_{i=1}^{n} (-1)^{i+1} C_i^n\, 2^{i(n-i)} f(n-i) \qquad (3)$$

The most representative method of the score-and-search-based approach is the K2 algorithm. The algorithm starts by assigning each variable without parents. It then incrementally adds a parent to the current variable which mostly increases the score of the resulting structure. When any addition of a single parent cannot increase the score, it stops adding parents to the variable. Since an ordering of the variables is known beforehand, the search space under this constraint is much smaller than the entire structure space, and there is no need to check cycles in the learning process. If the ordering of the variables is unknown, we can search over orderings [9].

III. Materials And Methods

The experimental greenhouse is located at the Queretaro University campus whose coordinates are: longitude, 100° 24’ W; latitude, 20° 36’ N; altitude, 1820 m. The covered area is 135 m^2 (9 m wide and 15 m long). The greenhouse is 5.49 m high with the gutter at 4.2 m. The ridge is orientated north-south. It has one roof windows (0.9 m wide and 13 m long), and four side windows.

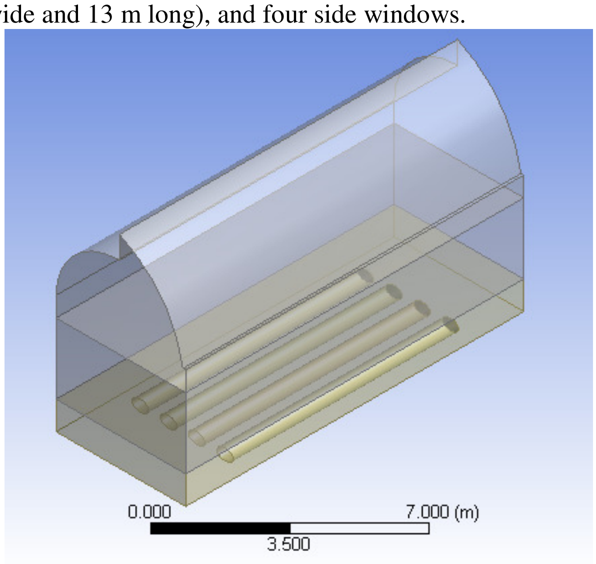


Fig.1. Experimental greenhouse geometry.

The north and south windows are 2.5 m wide and 8 m long, the east and west windows are 2.5 m wide and 11 m long. All the windows are of the roll-up type. The roof and side ventilation areas are 10% and 24% of the covered ground area, respectively (Figure. 1). We use *Lycopersicum pimpinellifolim* crop, with 2.7 plants/m^2 of density in harvest.

3.1 Experimental Measure

We obtained an experimental data set which was composed for the variables: air temperature, humidity, air flow speed and CO_2 concentration. The measures have been obtained at two distances above to the ground: 1 m and 3 m, in the canopy space and aisles between the rows of crop. The measurements of temperatures and humidity were recorded every four minutes by means of a CMOS Sensor Type FHAD462, with a measuring range 0 – 100% RH and accuracy ±1.8 %, hysteresis 1% RH, nominal temperature 25°C± 2k, with a range -20 ±60°C and accuracy ±0.3k at 25°C. The CO_2 concentration was measured with a two channel infra-red absorption principle Carbon Dioxide Sensor, hand-held Type FYA600CO2H, with measuring range 0 – 10,000 ppm and accuracy ± 2% and resolution 1 ppm. The measurements of speed air were taken using an omni-directional anemometer whose range of operation is 0 m s-1 to 20 m s-1 with an accuracy of +0.03 m s-1. The set of measurements for the greenhouse were taken on 21 to 25 of August 2011. The data set was discreted and used to develop a BN model that describes the relationships among all variables. The model shows differences that allow us to identify the independent and dependent variables as well as to quantify the degree of influence between them.

3.2 CFD Model Development

We performed numerical simulation development and CFD model with ANSYS v.12.1 software which was executed on a 32 bit machine under the Windows Vista operating system, thereby solved the equations of energy, continuity and momentum under 3 different values of solar radiation (500 w, 1000 w, 1350 w) and air flow velocity (0 m / s, 45 m / s and 0.9 m / s). The configuration CFD model is showed in Table 1.

Table 1: Model CFD configuration

	Description	Value	
Solver	3D simulation		
Condition time	Steady state		
Viscocity model	k-ε floatability effect active		
Energy equation	active		
Crop simulation	Porous media		
Dominion entrance	Inlet velocity	1 m s^{-1}	
Dominion exit	Pressure outlet		
Radiation	1400 W		
Physics properties			
	Air	Soil	Polyetilen
Density (kg m^3)	1.22	1400	920
Specific heat (J K^{-1} $°K^{-1}$)	1006.43	1738	1900
Termal conductivity (W m^{-2} $k°^{-1}$)	.0242	1.5	0.3
Thermal expantion coefficient $°k^{-1}$	.003389		

The methodology for developing the CFD model was performed according to the model proposed by Rico-

Garcia *et al.* (2006), which are the following steps: 1) Discretization of continuous flow: the field variables approached a finite number of points called nodes values 2) Discretization of the equations of motion according to the values of the nodes
3) Solving the system of algebraic equations and obtain the values of the variables at all nodes.

2.1 BN analysis

The analysis of BN was performed by the ELVIRA system version 0.162 in three stages suggested by De la torre-Gea and Rico-García [6].

1) Pre-processing.

We used the average imputation algorithm to complete the partial data sets. This algorithm replaces missing values / unknown to the average of each variable. This method does not need parameters. Massive data for each variable were discreted using the equal frequency algorithm with two intervals.

2) Machine learning.

According to Wang [13], for best regulate of Bayesian network structure we used the K2 algorithm learning method with Bayesian estimation and the maximum number of parents equal to 3, with no restrictions.

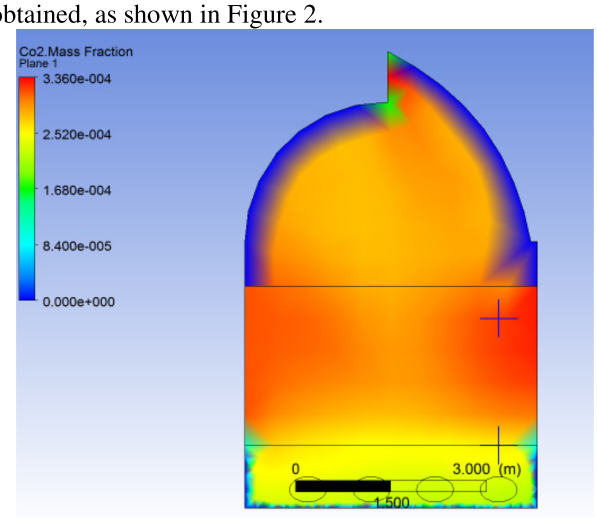
3) Post-processing.

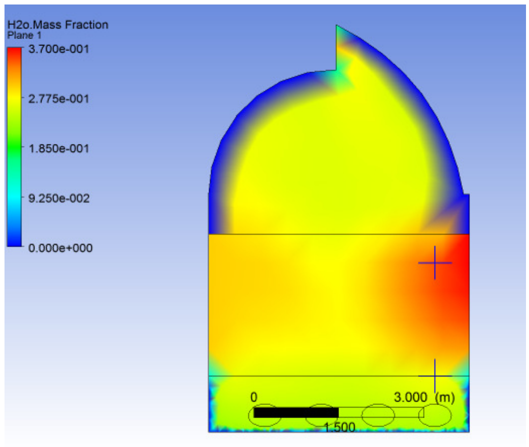
A dependency analysis was done to obtain the topological structure of the network, which represents the variables and their causal dependencies. After obtaining the parametric learning network, the conditional probabilities were calculated showing the relationships of influences between variables.

IV. Results and Discussion

A. CFD Model

To provide validity to the CFD model, the results obtained from the solution of the equations is calculated using the ANSYS software were compared with the values of the variables of air flow and measured temperature inside the greenhouse. An important part of this step is the graphical representation of the variables that govern the flow, to have a quick and pleasant view of the results obtained, as shown in Figure 2.





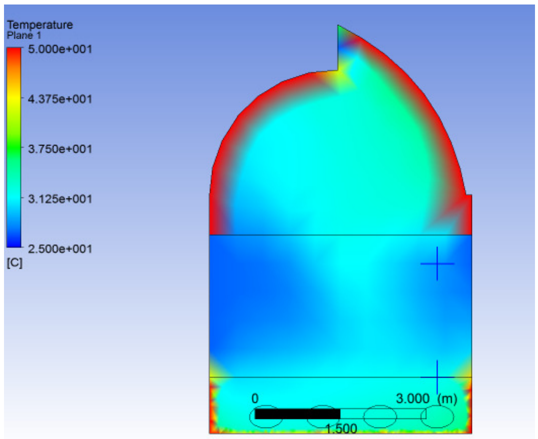


Fig.2. CFD model result inside a greenhouse A) CO_2 Concentration (ppm), B) Air Humidity (%), C) and Temperature (°C).

B. Climate conditions inside greenhouse

Humidity and CO_2 concentration shows an inverse relationship to Temperature. The temperature shows a heterogeneous behavior inside the greenhouse, forming gradients ranging from the outside temperature near windows, rising in the crop space. The air flow speed decreases by the action of the crop, forming a barrier, stagnating in the lower space, adjacent to the floor. Moreover, the elevated temperature causes that the CO_2 concentration and humidity, decreases in the same proportion.

By simulating the crop space beyond the properties of a porous medium, it consumes CO_2 as result of photosynthesis process, and increasing its humidity as result of the transpiration process, their concentrations are different in the space above the crop.

Figure 3 shows how is the relationship between CO_2 concentration and humidity, with temperature in CFD model compared with the data set obtained to measures inside the greenhouse. Lineal regression shows that both CO_2 concentration and humidity in CFD model are similar with the measures data set; with 5% of difference in the intercept of humidity, and 39 ppm of CO_2 concentration. The slops of the equations in both cases are very similar too. This indicates that the approximation by the CFD model is correct.

International Journal of Agriculture Innovations and Research
Volume 4, Issue 1, ISSN (Online) 2319-1473

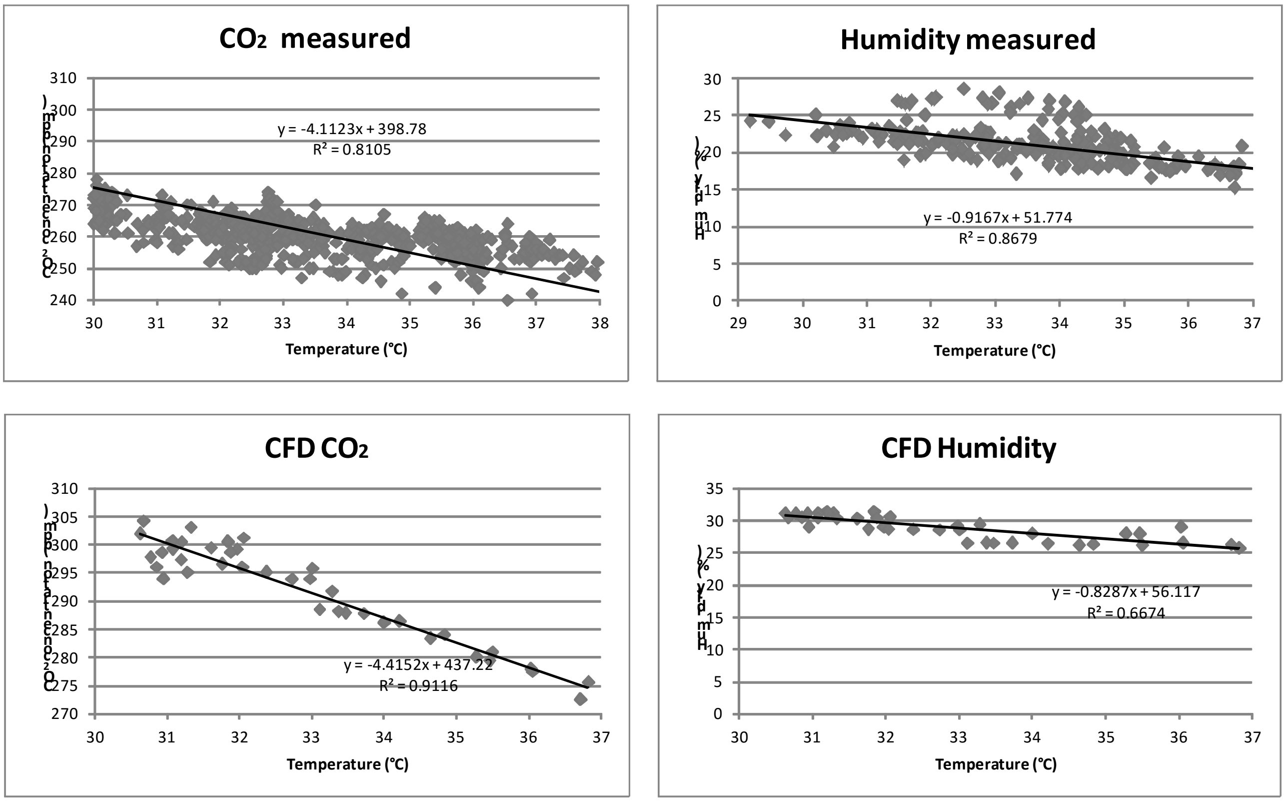


Fig.3. Relationships between Humidity and CO_2 concentration with Temperature.

C. Validation to CFD model using BN model

Validation consisted of verifying the accuracy and validity of the results given by the CFD model. Commonly, validation was performed on a different data set of the variables used in the developed model. At this work, the validation process consists in compared two BN models; the first develop from the data set of the measured sensors, and the second BN model from data set generated by the CFD model.

A BN model was obtained from 4200 data records of CFD model and a second BN model was obtained of 2000 data records of the measured sensors. Both models were compared and its inferences are shown in Table 2.

Table 2: Correlation coefficient between CO_2 and Humidity with Temperature

	Temperature		CO_2	$P(CO_2 \mid T°)$	Humidity	$P(H \mid T°)$
CFD model	34 °C	1.00	233 ppm	0.77	24. %	0.74
Data measures	33.3°C	1.00	240 ppm	0.88	17.4 %	0.88

In both BN models, variables were discretized into three intervals. It took into account the higher temperature to obtain the conditional probabilities of the CO_2 concentration and humidity. The values of the variables with respect to temperature for both BN models are very similar, indicating that the values obtained from the CFD model are correctly approximate to values obtained by measurements inside the greenhouse. The inferences calculated by the BN models for both: sensors measurements and the data obtained from the CFD model are similar in temperature and CO_2 concentration.

Correlation coefficients indicate that both CO_2 concentration and humidity are inversely related to temperature, which present both in the measurements of the sensors and when the data is obtained from the CFD model, as showed in Table 3. This indicates that the CFD model behaves properly and attached to reality.

Table 3: Comparison of inferences between measures data set and CFD model

Correlation coefficient	Temperature/CO_2	Temperature/Humidity
CFD model	-0.98057339	-0.96686119
Data set measures	-0.94082092	-0.8998763

Although the correlation analysis of variables is a probabilistic method, that indicates the degree to which two variables are related or matched, the BN model showing relationships among variables can be used for many variables at the same time. Similarly, The BN may show degrees of independence in the variables that linear correlation analysis not shows.

Comparing the results of Table 2, we observe that the conditional probabilities of CO_2 and humidity with respect to temperature are less than the correlation values of Table 3. This indicates that although the model is properly validated CFD based on linear correlation analysis, the BN models show that CFD model can be more refined to fit it to data measured inside the greenhouse.

The advantages of using BN models to validate a CFD model based on this work are: comparing with others validation methods is an economic technique and consumes little time and few computer resources increased accuracy, include uncertainty by calculating inferences and quantify the degree of dependence or independence between the variables studied. Besides the possibility of used incomplete data sets for both the CFD model as measurements in the interior of the greenhouse. The main disadvantage is that it takes at least two measured variables to establish the relationship between the two.

V. Conclusion

In this work has been possible to validate a CFD model using BN models with an acceptable degree of accuracy. BN have shown the relationships between variables CO_2 concentration and humidity with respect to temperature. Lineal regression shows that both CO_2 concentration and humidity in CFD model are similar with the measures data set. It has also been possible to quantify these relationships by calculating inferences in dependent probability distributions. The BN are helpful in CFD models with multiple variables. Still are required data sets of measurements by sensors inside the greenhouse for validation.

Acknowledgements

This research was partially supported by the Mexican National Council of Science (CONACYT), FIFI 2010 de la Facultad de Ingeniería (UAQ).

References

[1] Bournet PE, Boulard T (2010) Effect of ventilator configuration on the distributed climate of greenhouses: A review of experimental and CFD studies. Comput. Electron Agric 74:195–217.

[2] Correa M, Bielza C, Paimes-Teixeira J, Alique JR (2009) Comparison of Bayesian networks and artificial neural networks for quality detection in a machining process. Expert Syst Appl 36(3): 7270-7279.

[3] Chen Q, Lee K, Mazumdar S, Poussou S, Wang L, Wang M, Zhang Z (2010) Ventilation performance prediction for buildings: model assessment. Build Environ 45: 295-303.

[4] De la torre-Gea G, Soto-Zarazúa GM, Guevara-González R, Rico-García E (2011) Bayesian Networks for defining relationships among climate factors. IJPS 6(18): 4412-4418.

[5] De la torre-Gea G, Soto-Zarazúa GM, López-Crúz I, Torres-Pacheco I, Rico-García E (2011) Computational fluid dynamics in greenhouses: A review. AJB 10(77): 17651-17662.

[6] De la torre-Gea G, Rico-García E (2012) Bayesian Networks applied to a Computational Fluid Dynamics model for climate conditions inside a greenhouse. RCS 60: 85-92.

[7] Gámez JA, Mateo JL, Puerta JM (2011) Learning Bayesian networks by hill climbing: efficient methods based on progressive restriction of the neighborhood. Data Min. Knowl Discov 22: 106-148.

[8] Garrote L, Molina M, Mediero L (2007) Probabilistic Forecasts Using Bayesian Networks Calibrated with Deterministic Rainfall-Runoff Models. In: Vasiliev et al (ed) Extreme Hydrological Events: New Concepts for Security, Springer, pp 173-183.

[9] Guoliang L (2009) Knowledge Discovery with Bayesian Networks. Dissertation, National University of Singapore.

[10] Hruschka E, Hruschka E, Ebecken NFF (2007) Bayesian networks for imputation in classification Problems. J Intell Inform Syst 29: 231-252.

[11] Mediero OL (2007) Probabilistic forecast flood flows Through Bayesian Networks Applied to a Distributed Hydrological Model. Dissertation, Polytechnic University of Madrid.

[12] Pearl J (1988) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann, San Mateo CA.

[13] Reyes P (2010) Bayesian networks for setting genetic algorithm parameters used in problems of geometric constraint satisfaction. Intell Artificial 45: 5-8.

[14] Wang S, Li X, Tang H (2006) Learning Bayesian Networks Structure with Continous Variables. In: Li et al (ed) Lecture Notes in Computer Science, Heidelberg: Springer-Verlang pp 448-456.