PAPER • OPEN ACCESS

Arduino data-logger and artificial neural network to data analysis

To cite this article: G F Contreras Contreras *et al* 2019 *J. Phys.: Conf. Ser.* **1386** 012070

View the article online for updates and enhancements.

Arduino data-logger and artificial neural network to data analysis

G F Contreras Contreras[1], H J Dulcé-Moreno[2,3], and R Ardila Melo[2]

[1] Grupo de Invesitgación y Desarrollo en Electrónica y Telecomunicaciones, Universidad Francisco de Paula Santander, Cúcuta, Colombia

[2] Grupo de Investigación en Física y Materia Condensada, Universidad Francisco de Paula Santander, Cúcuta, Colombia

[3] Materials Science and Technology Research Group, Foundation of Researchers in Science and Technology of Materials, Colombia

E-mail: `ghiordyferneycc@ufps.edu.co`, `hectorjaimedm@ufps.edu.co`

Abstract. This work takes thermodynamic modelling through computer science for incubation process at domestic birds, that has presented energy consumption significantly high than energy used in processes. Thus, a data analysis was applied upon variables of temperature and relative humidity for heating zones, trying to know how much energy supplied by source was used, as well as, voltage and current variables are measured in the same moment that temperature and relative humidity are acquired. Then, data analysis was done using artificial neural networks models with samples obtained from sensors, where real process is highly time-variant, fixing environment conditions at the moment required. Therefore, with this system has been obtained an air flow of $3.4375 \ 10^{-2} \ \mathrm{m^3/J}$ using a anemometer respect to electrical energy supplied by fans, giving 9.4818 W of average power using ceramics resistances, and testing an adaptive controller where its variables are fitted using equations obtained from data analysis. In contrast, colombian farmers have decreased economic conditions to maintain them productions due to free trade agreements implemented lastly, indeed this system was developed using open-source software and hardware to avoid costs in acquisition by licensing politicians or periodic subscription to a specific product developed by companies.

1. Introduction

Thermodynamic systems have a process that include time-variant heating changes dependently from environment coordinate at the moment [1], theory's equations give a general description for physics phenomena, where the most have assumptions or only was designed for a specific number of variables, with static behavior over time [2]. Real process are highly time-variant, for it is known like dynamic process, and its behavior dependent of targeted variables.

Actually exists massive sensors in commercial market to all type of physics variable and counting with easy connection to digital systems like microcontrollers, microprocessors, digital signal processors (DSPs), and others [3,4]. Through these sensors can obtain data and manage itself using computers with math operations, where data is so important to know how a system changes over time for its featuring, managing, viewing or only understanding what happen with itself [5,6]. For it, in section 2 will be present a general review of procedures to understanding a thermodynamic system as the incubation is; section 3 shows 3 experiments used to analyze

data in thermodynamic system using data-logger, and finally section concluded the experimental results based in statistics metrics gotten for each experiment.

On other hand, Colombia's agriculture is a vulnerable economic sector for who was presented this researching, giving a approach of data analysis in birds' incubation systems, or similar production methods, aiming to improve techniques in industrial businesses with emerging tools like machine learning and computer science. Indeed, materials field involves engineering tools to solve problems related to measuring or featuring systems as much as materials and growing up the methods to study physics processes.

2. Methods and materials

The methodology was divided in forth parts: Design of the prototypes for testers and experiments, where the main issue is make simple as so as be possible; the second part aims to validate data-logger process, using experiments and birds' incubation system; the third part we introduce a controller to the system in order to manage automatically coordinates; and the last part take off data to a analysis into deep learning technique named artificial neural networks (ANN).

2.1. Design and requirements

Focusing to obtain a data analysis of values obtained from sensors installed in a thermodynamic system, thus the framework [7] gives arduino open-source hardware and software as an embedded system at several applications due to three main features:

- Doesn't require external manage to manipulate its behavior or main functions.
- Work in real-time where measurements are done with a sampling time (T_s).
- Includes computational intelligence, that is, the machine takes a decision when an event has occurred.

An embedded system must have the following main facts: reliability, maintainability, and availability such as are implemented for data-logger into code structure, using the paradigm of object-oriented programming and the code is released on electronic repository under open-source Massachusetts Institute of Technology (MIT) license, allowing public use and implementation of the system or modification for specific application or variable, under conditions that there isn't legacy warranty of the system or liability.

Figure 1 shows flow chart for data-logger, since system with temperature and relative humidity (RH) coordinates measured with sensors at the same time that voltage and current are recorded, this information go to board connected with Atmega328 through serial digital communications protocol, where also are connected a panel and micro secure digital (microSD) card memory, finally a serial to USB protocol is used as external monitor for data-acquisition in real-time.

The atmega328 microcontroller process information obtained to carry out data to microSD card memory, this saves the information obtained using comma-separated values (CSV) format with notepad file. On other hand, the panel mentioned previously has the following options:

(i) New file: Close the current file and create a new file using Electrically Erasable Programmable Read-Only Memory (EEPROM) data where will be information about what number is counted until this moment (maximum 255 files).

(ii) Reboot system: Allows overwrite all files saved since log 00 until 254. This action is not recommended when information wasn't saved previously, only is added in case that system has been collapsed by wrong data flow.
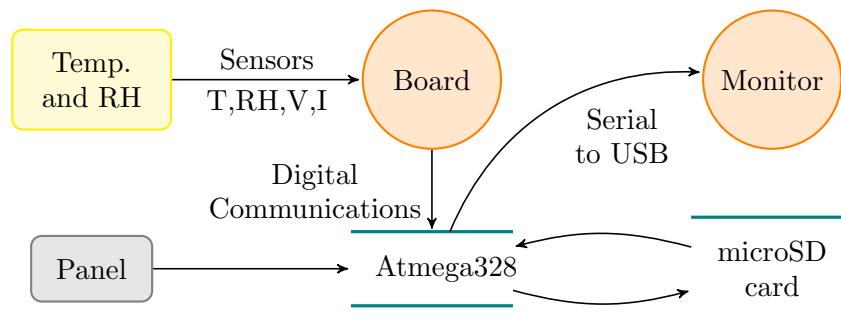
Figure 1. Data flow from thermodynamic system to micro SD storage using atmega328 incorporated on Arduino Board.

2.2. Laboratory proofs

In laboratory proofs has been necessary define which thermodynamic system will test the data-logger behavior [2], for it we've been proposed an incubator system as part of our research, using a system with ceramics resistors, these resistors are known for heating process, because temperature value doesn't affect resistance value significantly [8]. On other hand, was required to define ventilation conditions to the incubator system using the most optimized fan [2], to reduce loss of energy unnecessary at secondary, to determine that device, we evaluate volumetric flow rate describe in Equation (1) using 5 fans with 6 cm of ratio which are implemented in the incubation system to measure air flow velocity with and anemometer, and their energy consumption with measurements of currents supplied when are 12 V in their input.

$$Q = \frac{dV}{dt} = v * A \tag{1}$$

Table 1 shows the 5 fans used with their respective model in industry, flow velocity rate, current consumption, volumetric flow rate computed with Equation (1), and dQ/dP relationship in m^3/J that represent the level of energy transferred to move a volume.

Being the fan number 1 was selected to the incubation system due to its grades obtained, nevertheless errors' values defined as 1% for anemometer and 4% for ammeter didn't impact to the selection due to computing process were reducing values.

Table 1. Results of test for ventilation conditions.

No.	Model	Speed (m/s)	Current (A)	Q (m^3/s)	$\frac{dQ}{dP}(10^{-2}m^3/J)$
1	Artic PWM PST	$3.50 \pm 1\%$	$0.096 \pm 4\%$	$0.0396 \pm 1\%$	3.4375 ± 0.1719
2	Artic Silent	$2.00 \pm 1\%$	$0.060 \pm 4\%$	$0.0226 \pm 1\%$	3.1389 ± 0.1569
3	BOK BDH1202SS	$4.20 \pm 1\%$	$0.220 \pm 4\%$	$0.0475 \pm 1\%$	1.7992 ± 0.0900
4	Car Fan 12V	$3.00 \pm 1\%$	$0.110 \pm 4\%$	$0.0339 \pm 1\%$	2.5682 ± 0.1284
5	SXD8025S12M	$3.00 \pm 1\%$	$0.230 \pm 4\%$	$0.0339 \pm 1\%$	1.2283 ± 0.0614

2.3. On-off adaptive-predictive controller

The on-off adaptive-predictive controller (OOAPC) is a device that allows manage automatically the input source using conditions desired of temperature, only require two fixed temperatures to connect or disconnect source for heating system, is submitted to characterize system behavior with a certain action of control is applied over thermodynamic system. OOAPC predicts the

temperature that system is going to reach in a determined time of inertia defined as $a_j = 60$ seconds for this case, this value is multiplied with temperature gradient dT/dt to get temperature step and previous temperature is added giving a result for prediction described in Equation (2).

$$P(t + a_j) = a_j \frac{dT}{dt} + T_{(t-1)} \tag{2}$$

To compute gradient is enough take n number of samples of the sensors measurements and obtain mean value over time spent to the sampling procedure as is showed in Equation (3).

$$\frac{dT}{dt} = \frac{1}{n * T_s} \sum_{i=1}^{n} T_i \tag{3}$$

With this prediction value, the controller turn off an electromagnetic relay (ER) to supply energy towards the heating resistor connected at normally closed port of the ER when temperature is 37.0 °C or less; controller turn on the ER when temperature of prediction is 37.5 °C or greater, these two actions are made with assumption of a inertial time fixed by a_j.

2.4. Artificial neural networks modeling

ANNs is a technique of deep learning into a machine learning area included in artificial intelligence science, that allows us to learn about of one system from its data when input and output is known [9]. In literature has been found a relationship between temperature and humidity, defined experimentally as inverse [5, 10–12], that is a tricky trouble into a heating interchange systems and model of this phenomenon is uncertain. Thus, ANN is applied over data recorded using data-logger to emulate system in order to obtain more information about this case and how it's modelled [2], using 10% of samples to train and 90% of samples to test two models of ANNs which are presented in next sections.

2.4.1. Single neuron-layer model. Is a model with a single dense neuron unit into a one layer that is similar to linear regression model where the neuron weight (w) is the slope and neuron bias (b) is the intersection in dependent axis like is described in Equation (4).

$$y = w * x + b \tag{4}$$

The data input (x) is T as the temperature and the output (y) is RH as the relative humidity, obtaining a model $RH(T)$, also a $T(RH)$ model is considered to analyze.

2.4.2. Multiple neurons model. The another model consists in six neurons distributed uniformly at two hidden layers, that is a approach that system isn't linearly fitted as is presented in Figure 2, where hidden layers contains neurons a_0, a_1, a_2 and b_0, b_1, b_2 for hidden layer 1 and hidden layer 2, respectively.
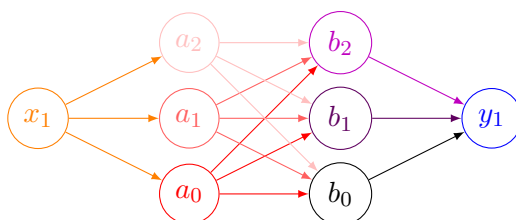
Figure 2. ANN dense with six neurons in two hidden-layers for one input and one output neurons.

That neurons units are dense type of neurons, as well as input layer and output layer have singles neurons defined as x_1 and y_1, respectively. The model for multiple is very complex in

comparison with single neuron-layer model (SNLM), due to how its weights and bias are related. Literature has the assumption that this type of models require much data due to great quantity of variables or change the training mode in order to improve as fast as be possible [13].

3. Experimental results
Three experiments was made to validate our system, where the first consist in the varying of the voltage value for input supply to the heating system, second experiment implements OOAPC and two fixed values in thresh starting in a low temperature environment (transient state), and in the third experiment OOAPC remains under steady state conditions with a closest temperature to the desired value.

3.1. Experiment number 1
Starting experiment sessions, the system was undergoing to conditioner air flow of 18.0 °C until reach 25.0 °C in the system, these air flow remains during all time that experiment spent. In Figure 3 is showed the temperature and RH curves when input voltage is manually varied from 2 V to 12 V in steps of 2 V as is presented in curves of Figure 4, this task is repeated 3 times over 30 minutes of execution. Red curve is the temperature on heating source and cyan curve correspond to its RH, orange curve is the temperature at load zone and blue curve its RH. In another plot green curve is voltage in electrical energy source to run all devices included in the system, while purple curve is the current consumed, this curves have critical behaviors for source zone but load trends to linear behavior and soft manipulation, when energy supply seems has problems with maximum voltage.
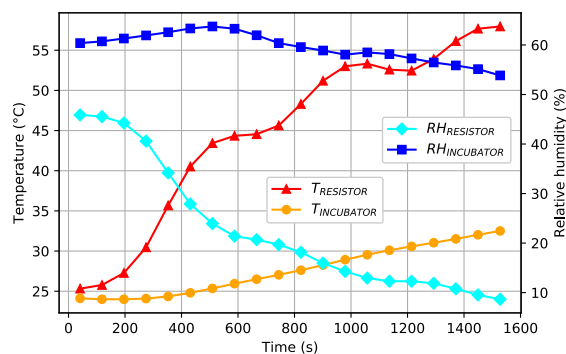
Figure 3. Temperature and RH curves at supply zone and load zone for experiment 1.
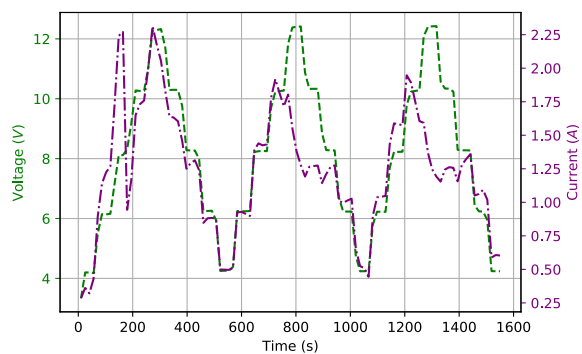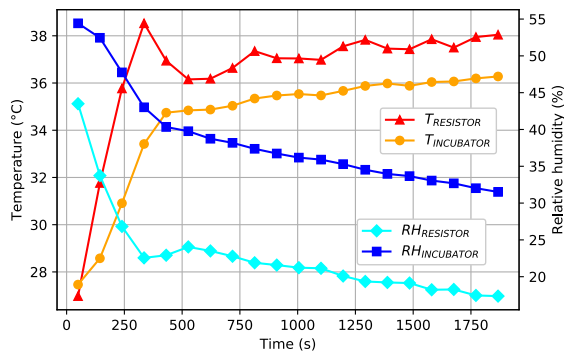

Figure 4. Voltage and current curves at energy supply for experiment 1.

The next has been apply the neural networks models SNLM and multiple neurons model (MNM) using 695 samples to train and 6257 samples to test both models, the training is done using mean squared error (MSE) with Adam Optimizer fixed in 0.1 through 500 epochs, giving the linear Equation $RH(T) = -0.9066 * T + 83.9731$, where temperature is in Celsius degrees and RH is % units, this linear Equation is obtained from SNLM model that reaches 1.9163 of MSE and 1.1389 of mean absolute error (MAE). The MNM model improved grades respect to SNLM obtaining 1.8397 of MSE and 1.0961 of MAE.

3.2. Experiment number 2
Second experiment implements OOAPC device with its measuring point located in load zone, source zone doesn't represent a reliable place to set temperature desired. How was looked at experiment 1, colors in curves represent the same variables, but is evident a trending to a defined value for Figure 5 in temperature values, remaining softest to load zone and some variant for

source zone while RH curves have similar inverse behavior. In the Figure 6 current a voltage have a mid-point until desired temperature is near, when these event occurs, curves trends to oscillate creating a quasi-periodic signal to establish the desired temperature.

Figure 5. Temperature and RH curves at supply zone and load zone for experiment 2.
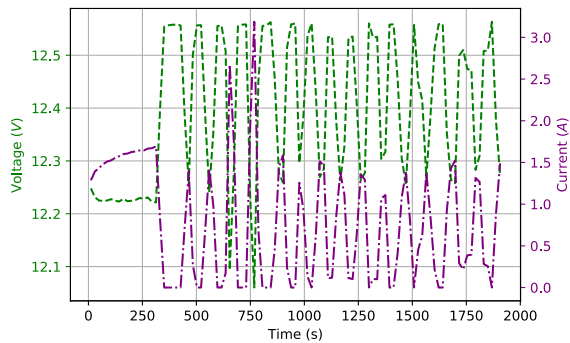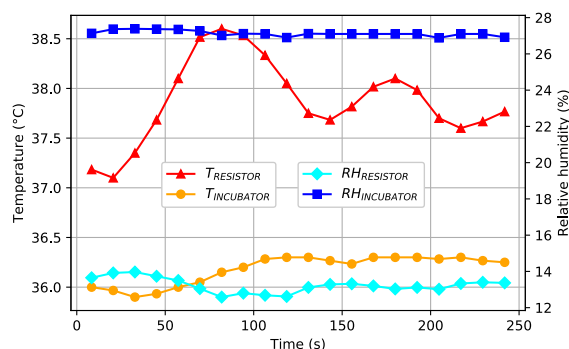

Figure 6. Voltage and current curves at energy supply for experiment 2.

For this experiment, data analysis using ANN models gives the SNLM model better than MNM model due to the metrics applied, MSE for SNLM was 2.2857 when MNM gave 5.1834 and MAE was 1.2820 for SNLM and 1.9001 for MNM, that is, SNLM model is almost 44.10% more accuracy than MNM model for transient state. Nevertheless this experiment used only 838 samples to train and 7551 to test giving a linear regression model described in Equation (5) and using the same parameters for training mentioned in the experiment 1.

$$RH(T) = -2.5105 * T + 125.2680 \tag{5}$$

3.3. Experiment number 3

This experiment aims to steady state behavior, when desired temperature has been reached by OOAPC and the importance is find how much varies its values like Figure 7 where has been seen a quasi-static RH and temperature trending to mid-point as much as the time pass, getting a quasi-periodic signal in Figure 8 to maintain the desired conditions.

Figure 7. Temperature and RH curves at supply zone and load zone for experiment 3.
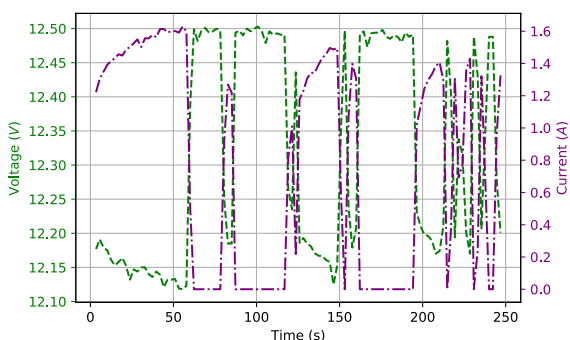
Figure 8. Voltage and current curves at energy supply for experiment 3.

Finally, for steady state was used 108 samples to train and 980 samples to test remaining the training parameters from the previous experiments, but the metrics was significantly reduced for both models, where MSE gave 0.2661 for SNLM model and 0.6111 for MNM model, MAE

gave 0.4661 for SNLM model and 0.0196 for MNM model. Indeed, SNLM model remains as the better model for two or three experiments, denoting what OOAPC controller may be the difference for these results, trending the system to a linear regression such as for this case is $RH(T) = 0.7192 * T + 0.7029$, this Equation is very different respect to Equations (**??**) and (5), in this experiment RH didn't vary notably that can generate this great difference and giving a uncertain hypothesis about of temperature and RH are related.

4. Discussion

The data-logger system is robust and embedded device to acquire data from real environment in order to process that data using open-source computational tools like jupyter notebook, RStudio or Spyder [5]. Real systems must be delimited before laboratory proofs, this gives initial conditions to improve results and make it more comparable with another works in literature likes [3, 6, 14]. Sensors cannot read in real-time, they ever delay a defined time to acquire and process the values obtained, for it is necessary assume predictions over them, that can make more adaptive for time-variant systems with certain parameter of setting for its adaptability supported on data sheets' parameters like sampling frequency or sensibility. Data analysis using ANN tools got interesting results, where in three different environments SNLM and MNM models responded with high accuracy using a relation 1:9 for cross validation, that is, SNLM and MNM are models optimal for time-variant systems as was purposed by [2].

5. Conclusion

With the results of experiment 1, was obtained a average power value in 10.9681 W for 1570 seconds spent such as the energy consumption was 17.210 KJ when energy supply is manually varied, implicating high variant behaviour for the system and requiring MNM to improve accuracy in its modelling. With the results of experiment 2 was obtained a average power value in 9.7136 W for 1918 seconds spent such as the energy consumption was 18.630 KJ when energy supply is in transient state using OOAPC controller. With the results of experiment 3 was obtained a average power value in 9.4818 W for 250 seconds spent such as the energy consumption was 2.370 KJ when energy supply is in steady state using OOAPC controller. Both experiment 2 and 3 demonstrate high performance for OOAPC when it optimizes energy consumption while systems trend to linear model or SNLM was a near representation for itself.

References

[1] Wang J 2009 *Progress in Natural Science* **19(1)** 125
[2] Reyes-Rosas A *et al.* 2012 *Revista Chapingo. Serie horticultura* **18(1)** 125
[3] Fernandes M, Canito A, Bolón-Canedo V, Conceição L, Praça I and Marreiros G 2019 *International Journal of Information Management* **46** 252
[4] Sousa R, Pereira M, Pereira F M Q and Araujo G 2019 *Journal of Parallel and Distributed Computing* **130** 126
[5] Schito E, Pereira L D, Testi D and da Silva M G 2019 *Data in Brief* **24** 103788
[6] Nayak P, Mukherjee A K, Pandit E and Pradhan S K 2018 *Rice Science* **25(1)** 1
[7] González Morales J *et al.* 2017 *Diseño e implementación de un control de temperatura y humedad para un prototipo de incubadora artificial de pollos* (Cali: Pontificia Universidad Javeriana)
[8] Contreras Contreras G F and Dulcé-Moreno H J 2018 Diseño y construcción de una incubadora de aves de bajo consumo energético *5th International Week of Science, Technology and Innovation* (San José de Cúcuta: Universidad Francisco de Paula Santander) p 235
[9] Matich D J 2001 *Redes neuronales: Conceptos básicos y aplicaciones* (Rosario: Universidad Tecnológica Nacional)
[10] Laukkarinen A and Vinha J 2017 *Energy Procedia* **132** 711
[11] Sallam G A and Elsayed E 2018 *Ain Shams Engineering Journal* **9(1)** 1
[12] Xu X, Zhong Z, Deng S and Zhang X 2018 *Energy and Buildings* **162(1)** 163
[13] Xu X, Zhong Z, Deng S and Zhang X 2019 *Materials & Design* **162** 300
[14] Bou-Llusar J C and Satorra A 2019 *BRQ Business Research Quarterly* 1