A Simplified Model Parameter Identification Methodology for Buildings Indoor Thermal Behavior Control: A Case Study in a Tropical Climate of Panama

Ana K. Rivera, Engineer[1], Josue Sánchez, Engineer[1], and Miguel Chen Austin, Ph.D.[1,2,3,*]
[1]Research Group Energética y Confort en Edificaciones Bioclimáticas (ECEB), Faculty of Mechanical Engineering, Universidad Tecnológica de Panamá, Ciudad de Panamá, [ana.rivera3, josue.sanchez5, miguel.chen]@utp.ac.pa
[2]Centro de Estudios Multidisciplinario en Ciencias, Ingeniería y Tecnología (CEMCIT-AIP), Ciudad de Panamá, Panamá
[3]Sistema Nacional de Investigación (SNI), Clayton Ciudad de Panamá, Panamá
*Corresponding author: miguel.chen@utp.ac.pa

Abstract– ***The design of building control strategies that maintain occupant comfort and improve energy efficiency relies on a thermal model to predict indoor temperature. The model should represent the building envelope and its systems. Other than controller design, the thermal model is useful for the evaluation and simulation of energy optimization strategies. In this paper a methodology to identify building thermal models, based on a thermal-electrical analogy, is proposed. The methodology can be applied to residential case studies, located in tropical climates. It consists of determining which candidate model best describes the building thermal dynamics. The models are trained multiple times to study parameter estimate dispersion, and if the estimates converge to a single value regardless of their initial value, the models are validated. The model with the lowest root-mean-square error (RMSE) is selected as the best model. When the methodology is tested in a residential case study located in Panama, the best network has a validation RMSE of 0.36°C, which is satisfactory for controller design purposes. The model is then used to tune a proportional integral derivative (PID) controller which is then successfully employed to maintain a desired indoor temperature. Tuning the controller with the identified model avoids the need for tedious trial and error controller tuning.***

Keywords—Building energy modeling, RC thermal network, System identification

I. INTRODUCTION

Buildings represent over one third of final energy consumption and nearly 40% of CO_2 emissions in the world [1]. Heating, ventilation and air conditioning (HVAC) systems consume around 40% of total energy in buildings [2]. In Panama, with its hot and humid tropical climate, air conditioning is among the systems with considerable consumption in commercial and residential sectors [3].

The Department of Energy of the United States (DOE) shows that between 4 and 20% of energy used in HVAC systems and lighting is wasted due to operative problems [4], revealing the energy efficiency issues present in buildings. Keeping in line with both international [5] and national [3] energy efficiency goals, it is imperative to take action to reduce energy consumption while maintaining occupant comfort and performance, and at the same time reduce operating costs. The deployment of control strategies is an effective way of improving building energy efficiency. This is proven by studies such as the DOE's, which estimates that it is possible to save up to 29% of actual consumption through the deployment of controllers, fault elimination and better sensor feedback in commercial buildings [6].

Control actions must maintain occupant comfort, which depends on the indoor temperature, among other factors. The design of such control strategies relies on a thermal model to predict indoor temperature, representing the building envelope and its systems. The thermal model is also useful to evaluate, simulate and implement energy optimization strategies [2], [7]. The model must represent the building accurately while remaining simple enough to maintain the computational efficiency necessary for good controller performance [8].

A. Modeling techniques

The thermal model can be obtained through three main modeling techniques: white-box, black-box and gray-box [4], [9]–[11]. White box models are built from physical equations, such as energy balances, that describe the heat transfer behavior of the building [12]. To formulate these models, it is necessary to know building thermal properties and time constants a priori, information that is difficult to obtain or unavailable. This, added to the fact that detailed heat transfer equations result in a high order model that is time consuming to solve, makes white box models unsuitable for the design and optimization of control strategies [4], [11]. The main use for white box models in building energy models is through energy simulation softwares, such as EnergyPlus (EP) and TRNSYS. In these softwares, a reference model of the building is built, and is simulated to generate the input-output data that will be used to train and validate black-box and grey box models, as is done in [8], [13], [14].

On the other hand, black box models make little to no use of the physical knowledge of the system to derive a model. The

Digital Object Identifier (DOI):
http://dx.doi.org/10.18687/LACCEI2022.1.1.187
ISBN: 978-628-95207-0-5 **ISSN:** 2414-6390

mathematical relations that describe the building are regressed from input and output data [15]. Autoregressive models are black models used in building energy modeling. An autoregressive integrated moving average with exogenous variables (ARIMAX) model is used in [16] to predict indoor temperature of a conference room using as inputs heat supplied or rejected by radiators and fancoils, outdoor temperature and radiative heat gains. In [8] the authors incorporate knowledge of the building's heat transfer processes, to make educated guesses on the structure of a fractional order autoregressive model with exogenous inputs (FARX) model to predict indoor temperature. Other popular black-box models are artificial neural networks, which derive nonlinear input-output relations through connections between neurons [17]. In [18], neural networks are used to design a predictive control scheme to optimize demand flexibility in a house in Holland. Similarly in [19] a house's HVAC subsystems are modeled with neural networks. The lack of physical insight in black box models facilitates the modelling process, as less expert knowledge is required. However, this makes the models highly dependent on the quality of the training data, often requiring large and informative datasets, which is the main shortcoming of black-box models [4], [20].

Grey box models aim to overcome the shortcomings of white box and black box techniques. As with white box models, the model structure is derived from physical equations, which can be simplified to obtain a lower order, more computationally efficient model. However, model parameters and coefficients do not need to be known beforehand and are estimated from input-output data, much like in black box models. Therefore, it is possible to incorporate the available physical knowledge of the building into the structure and optimization constraints and estimate the missing information through input and output data [20]. RC thermal networks are the most popular gray box modelling techniques used to represent buildings. These are based in a thermal-electrical analogy, where materials properties, climate conditions and building subsystems are modeled as electrical components, in the form of capacitances, resistances, current and voltage sources [21], [22]. Several network topologies have been proposed in the literature to characterize building thermal dynamics. In [23], a 3R2C network is selected as the best model to represent a house with large glazing area and considerable envelope thermal inertia. In [24] a 6R5C network is simplified to a 2R1C network and able to successfully model a residence equipped with air conditioner, with an absolute error of 0.72°C when the model is trained with 20 days of data. In [25] a methodology was developed to identify thermal models based on RC networks for a passive residence located in Panama. As a hybrid modelling technique, grey-box models are also able to couple different model structures to improve or complement model estimation. In [13], a gaussian process (GP) model (black-box model) is used to correct the error in temperature estimation when obtained from a 4R4C model. The model performance was superior when incorporating GP than with the standalone RC network, especially when the occupant schedule was varied. In [26], a machine learning algorithm and a 6R4C network are coupled to predict the temperature difference, and the mean temperature, respectively, of two stories of a house. In [27] heat gains through the envelope are modeled by several RC networks, which are coupled with humidity and predicted mean vote (PMV) models. The integrated model is then used to develop a model predictive control (MPC) scheme which achieves up to 19.4% energy savings.

B. *Parameter estimation*

While grey-box and parametric black-box modelling techniques define the model structure, it remains necessary to estimate model parameters and coefficients through some parameter estimation method. The prediction error method defines an estimation problem as presented by (1), where the estimated parameters $\hat{\theta}_N$ are the parameter values that minimize a cost function $V(\theta, Z^N)$, within a search space Z^N, which corresponds to the training dataset, according to some criteria. Cases of PEM arise depending on how the cost function is defined, such as the least squares estimator where $V(\theta, Z^N)$ is defined as the quadratic norm of the prediction error and the maximum likelihood estimator where it is a likelihood function [28]. PEM is used by several authors [13], [24], [26], [29] to estimate model parameters.

$$\hat{\theta}_N = \text{minimize } V(\theta, Z^N) \tag{1}$$

Once the cost function is defined, the optimization algorithm finds the parameters that minimizes it. In [18], [29] the Levenberg-Marquardt algorithm is used to minimize the cost function, while in [26] it is done through particle swarm optimization.

C. *Contribution*

In this paper a methodology is proposed to identify RC thermal networks and determine the model that best represents building thermal dynamics while maintaining the computational efficiency needed for control purposes. The methodology is developed for application in residential buildings situated in tropical climates. To formulate the methodology, a candidate model selection procedure is carried out which consists of an evaluation of the networks' parameter dispersion. Additionally, several datasets are evaluated to determine the characteristics of the training dataset that generalize best. The methodology is designed to by applied to residential case studies situated in tropical climates, where important non-linear phenomena such as solar radiation and humidity are prominent factors.

The paper is outlines as follows: section II outlines the case studies considered as well as the procedures employed to develop the methodology. Section III details the results and some important insights from them. Conclusions and further works are discussed in Section IV.

II. METHODOLOGY

A. *Case Studies*

The case studies, one-story houses situated in Panama, are built in DesignBuilder (DB), a graphical interface for EP. Two case studies are used to develop and test the methodology, Case 1, and Case 2, respectively. Geometrically they are exactly alike (see Fig. 1) but differ in the materials used for construction. The characteristics of each case study are summarized in TABLE I. In any case, the house is occupied by one person all day. There are no elements surrounding the house that might mitigate solar radiation. The house is equipped with a fancoil air conditioning (AC) system, where the supply air mass flow rate is kept constant to maintain linearity in (2). The air conditioner is turned on and off according to an operating schedule shown in TABLE II. The meteorological data used in DB consist of a typical year in Panama City.

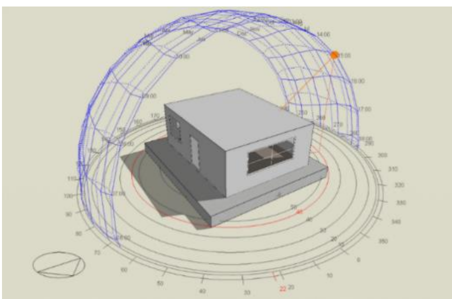

Fig. 1 3D model used for the Case studies in Designbuilder.

TABLE I
THERMAL PROPERTIES FOR CASE STUDIES
U: Heat transfer coefficient
SHGC: Solar Heat Gain Coefficient

	Case 1	Case 2
U wall (W/m^2-K)	3.767	2.174
U roof (W/m^2-K)	2.941	2.326
U floor (W/m^2-K)	0.25	
SHGC windows	0.72	
U windows (W/m^2-K)	3.772	
Window-wall ratio	30%	
Floor area (m^2)	47.858	

TABLE II
AIR CONDITIONER OPERATING SCHEDULE AND SETPOINTS

12:00 A.M - 8:00 A.M.	22°C
8:00 A.M - 12:00 P.M.	Off
12:00 P.M. - 3:00 P.M.	24°C
3:00 P.M. - 6:00 P.M.	Off
6:00 P.M. - 12:00 A.M.	18°C

B. *Training datasets*

A dynamic simulation is run in DB to obtain the input and output data used for model training and validation. Four months are simulated: February (typically the driest month of the year), November (typically the rainiest), April and December (two months with intermediate climate conditions). These months are divided in twelve possible training datasets, varying in length, detailed in TABLE III. All data is sampled at a 1 min interval.

TABLE III
TRAINING DATASETS

	February	April	November	December
10 days	D1	D4	D7	D10
20 days	D2	D5	D8	D11
Full month	D3	D6	D9	D12

C. *Workflow*

To develop the methodology, the workflow detailed in Fig. 2 was followed. The procedures “evaluation and selection of candidate models” and “evaluation and selection of training dataset”, detailed in the homonymous sections, where applied to case study 1. The main results from these procedures, the candidate models, and the best training dataset, as shown in the gray squares of Fig. 2, are used to formulate the final methodology, which will then be tested on Case 2.

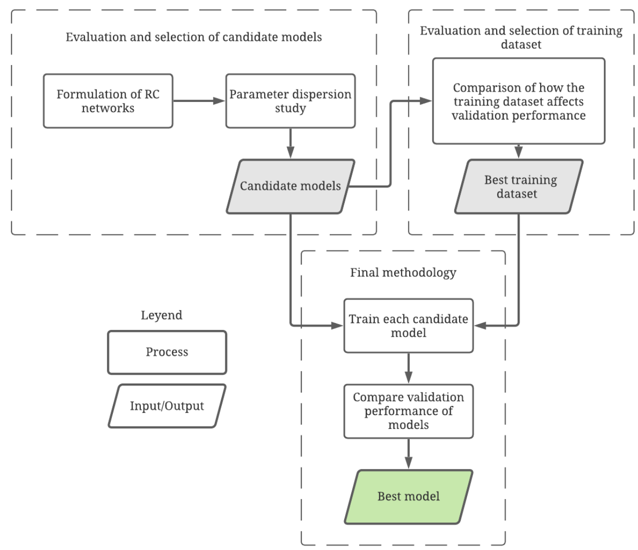

Fig. 2 Workflow followed in this work.

D. *Model formulation*

The candidate models are proposed in order of increasing complexity, starting from the simplest network (1R1C) to the most complex one (4R3C). In total, 10 networks are initially proposed, shown in Fig. 3.

The nodes in the network include indoor temperature (Ti) for all networks, and roof-wall (Tw) and floor (Tf) temperatures for some networks. The inputs considered are outdoor temperature (Tout), internal gains (Qint), global horizontal irradiation (ϕ), AC supply temperature (Ts) and mean radiant temperature (Tmr). Ts and ϕ are related to the heat flows Qac and Qsol, respectively. These heat flows are given by (1) and (2), where Fj and G are estimated parameters and ṁ and Cp are known constants.

$$\mathrm{Qsol_j} = \mathrm{F_j} \cdot \phi \tag{1}$$

$$\mathrm{Q_{ac}} = \mathrm{G} \cdot \dot{\mathrm{m}} \cdot \mathrm{Cp}(\mathrm{T_s} - \mathrm{T_i}) \tag{2}$$

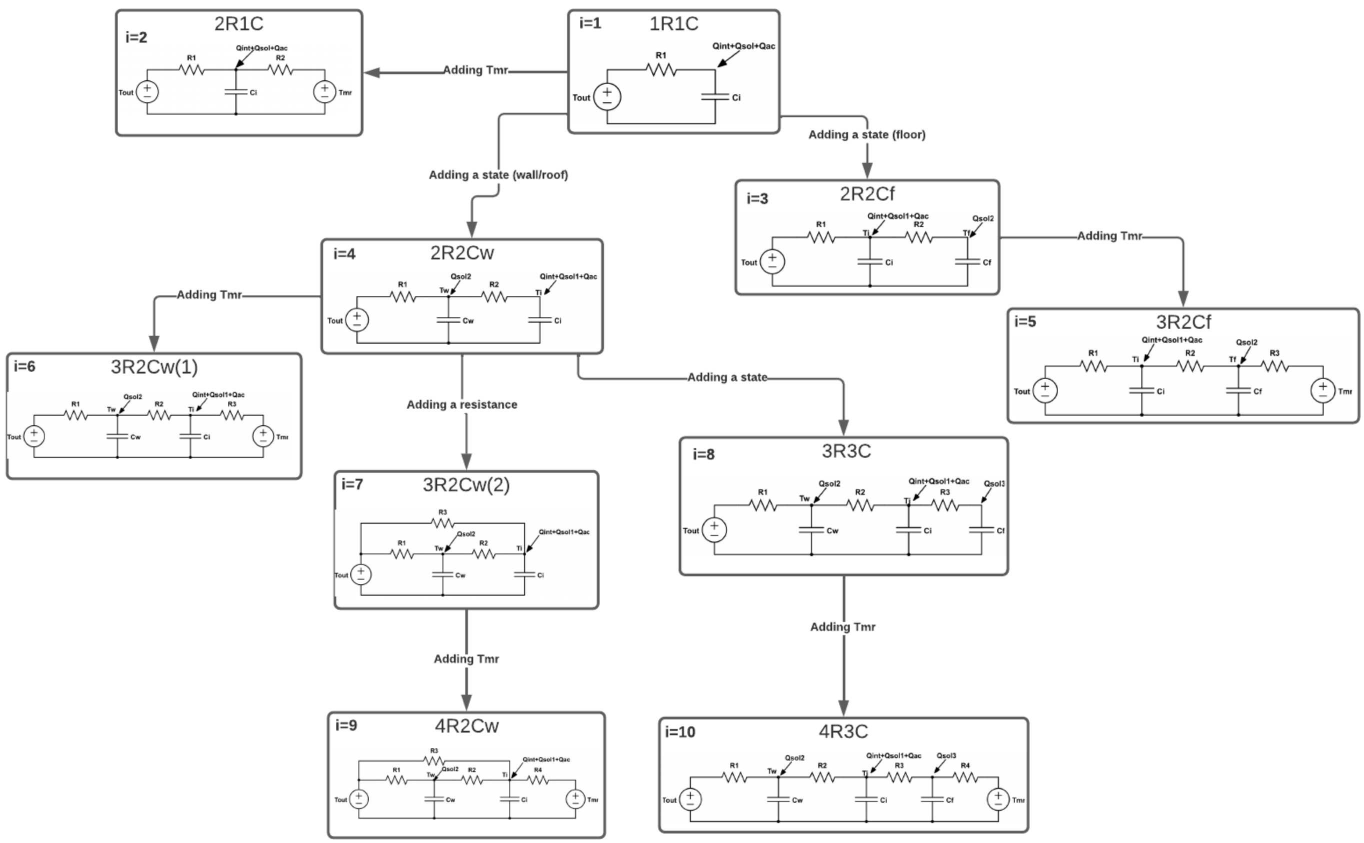

Fig. 3 Networks proposed

For each model, the thermal dynamics are described by the ordinary differential equations obtained when applying Kirchhoff's Laws to the thermal networks. These are rearranged in state-space representation, which clearly defines model inputs, outputs, states, and parameters.

E. *Parameter estimation*

Model parameter estimation is implemented in MATLAB's System Identification Toolbox. For the estimation problem, the initial values of the parameters are chosen randomly within 50% and 150% of an approximated physical value. All parameters are restricted to be positive, since negative resistances and capacitances have no physical meaning, and resistances have an upper constraint of 1 since these are physically smaller than 1.

F. *Evaluation and selection of candidate models*

The networks proposed in Fig. 3 are selected or discarded according to a parameter dispersion criterion. We are only interested in identifiable models, meaning that there exists a unique parameter vector that produces a global minimum in the cost function. Therefore, the model's parameter estimates must converge to a unique value regardless of its initial value.

To study parameter estimates convergence, the following procedure is followed:

1. Train model ten times with dataset D1, using a different initial value for parameters in each identification.
2. Calculate de coefficient of variation (CV) for the 10 parameter estimates.
3. Repeat steps 1 and 2 for datasets D2-D12.
4. Average the CV obtained for all datasets.
5. If the average CV is smaller or equal than 10%, the model is selected. Otherwise, it is discarded for its non-identifiability.
6. Repeat steps 1-6 for all models in Fig. 3.

G. *Evaluation and selection of training dataset*

To determine the best training dataset for the model identification methodology, it is of interest to study the length of the training dataset and the month most appropriate to recollect data.

The models that were selected from the previous procedure each have twelve different identified versions, corresponding to estimates obtained from each of the twelve training datasets.

Each of these versions is validated using the three months not used for training. The validation RMSE is averaged and the training dataset that produced the lowest value is selected as the best training dataset for the methodology.

III. RESULTS AND DISCUSSION

A. *Evaluation and selection of candidate models (applied to Case 1)*

The candidate models for the methodology are obtained from applying the "evaluation and selection of candidate models" procedure to Case 1. The results to applying this procedure to the network 1R1C when training with dataset D1 is graphically shown in the scatter plot of Fig. 4 for parameter Ci. Here, the 10 random initial values for each parameter (which correspond to the data points that produce a higher RMSE, since the parameter is not chose to be optimal) are matched to their corresponding estimated value, which was reached when the optimization was completed. The parameter's coefficient of variation is higher than 10% for this dataset, with the estimated parameters seemingly not converging to a single value. However, when repeating this procedure through the remaining eleven datasets, the CV is nearly zero. For example, in Fig. 5 the dispersion of parameter Ci is shown for training with D2. Here, the parameter clearly reaches the same value regardless of its initial value. Calculating the average CV over all datasets, the dispersion criterion is met. These results are summarized in TABLE IV. Therefore, network 1R1C is selected as a candidate model for the final methodology.

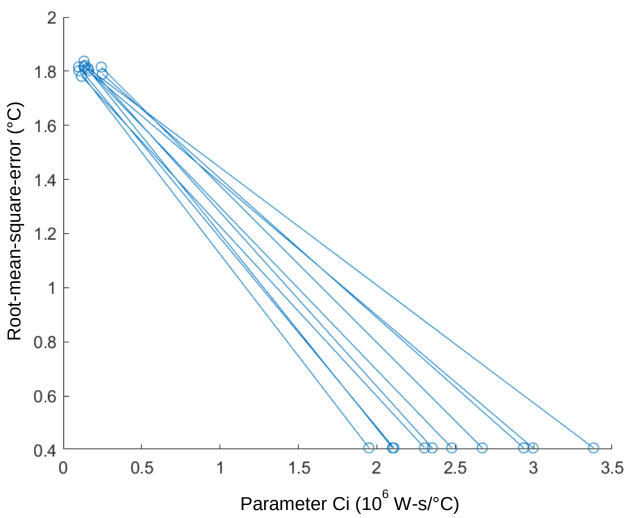

Fig. 4 Initial vs estimated values of parameter Ci when training with D1 in network 1R1C

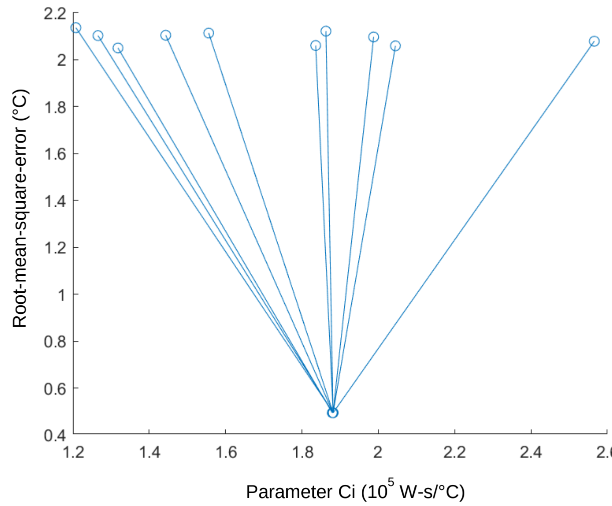

Fig. 5 Initial vs estimated value of parameter Ci when training with D2 in network 1R1C

TABLE IV
PARAMETER DISPERSION IN NETWORK 1R1C

Training dataset	CV (%)			
	Ci	R1	F1	G
D1	18.24	17.33	18.23	18.20
D2	0.04	0.00	0.00	0.00
D3	9.08	8.47	9.10	9.05
D4	6.27	5.90	6.28	6.27
D5	0.52	0.52	0.52	0.52
D6	0.14	0.04	0.05	0.04
D7	0.04	0.00	0.00	0.00
D8	0.02	0.00	0.00	0.00
D9	0.02	0.00	0.00	0.00
D10	9.16	9.02	9.16	9.15
D11	10.19	10.64	10.20	10.18
D12	0.03	0.01	0.01	0.01
Avg.	4.48	4.33	4.46	4.45

Fig. 6 shows the dispersion of parameter Ci for network 2R1C when training with D1. The CV obtained for each dataset are summarized in TABLE V. Since the dispersion criterion is not met, this network is discarded, and therefore not considered for the final methodology. For this network, the estimated value of parameter F1 in every training dataset is zero, which is why its CV cannot be calculated. This can be due to the nature of the mean radiant temperature as it models all radiative exchanges between surfaces. Therefore, the model no longer requires any additional radiative heat gain sources which translates into a zero F1 coefficient.

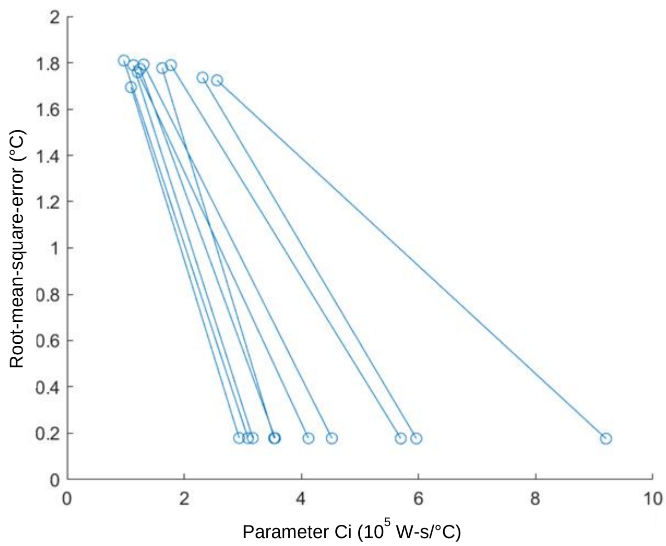

Fig. 6 Initial vs estimated values of parameter Ci when training D1 in network 2R1C

TABLE V
PARAMETER DISPERSION IN NETWORK 2R1C

Training dataset	CV (%)				
	Ci	R1	R2	F1	G
D1	42.3	29.8	31.0	N/A	41.9
D2	32.8	33.5	35.1	N/A	32.5
D3	20.3	21.1	22.7	N/A	20.1
D4	37.3	26.5	27.4	N/A	37.0
D5	22.1	25.2	26.3	N/A	21.9
D6	35.3	30.8	31.8	N/A	35.0
D7	44.9	42.2	59.1	N/A	44.9
D8	40.6	33.3	47.4	N/A	40.6
D9	25.8	22.9	32.5	N/A	26.1
D10	31.7	34.6	36.8	N/A	31.4
D11	34.3	34.1	35.7	N/A	33.9
D12	23.4	24.6	25.5	N/A	23.2
Avg.	32.6	29.9	34.3	N/A	32.4

A network that required further analysis was network 2R2Cf. During preliminary analysis, the network did not meet the dispersion criterion. However, with closer study, it was found that when training with D5 the CV was over 80%, and it was the only dataset that produced estimates with a CV higher than 10%, which caused the average to be higher than the criterion. In Fig. 7, the scatter plot of estimates is shown for parameter Ci. It is shown that some estimations converge to a value with a higher RMSE than others, indicating that the algorithm may have gotten stuck on a local minimum for those estimations. Since there is a global minimum (around 0.3°C), only the estimations that converged to this value where considered. When the CV is recalculated for this dataset, the average along all twelve datasets does meet the criterion, and therefore this network is selected as a candidate model.

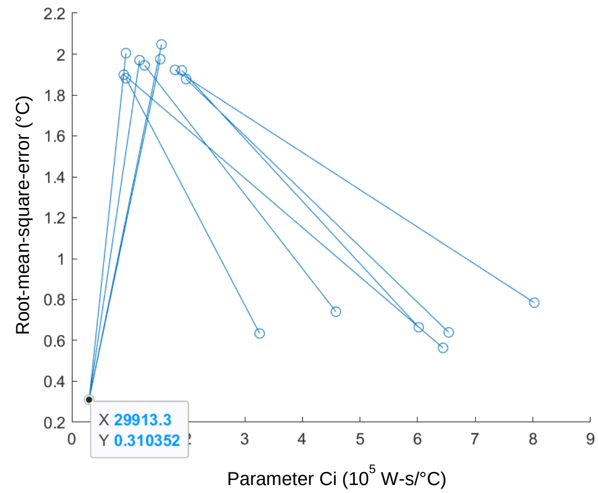

Fig. 7 Initial vs estimated value of Ci when training with D5 for network 2R2Cf

Among the initially proposed 10 models, only the 1R1C and 2R2Cf met the dispersion criterion and where therefore selected as final candidates. The resulting candidate models have few parameters and can produce consistent estimates. This is consistent with the results of [30], by reducing the number of degrees of freedom the cost function becomes notably convex, and therefore a numerical solver finds a global minima regardless of the initial parameter values.

B. *Evaluation and selection of training dataset (applied to Case 1)*

The result from applying this procedure to Case 1 is shown in TABLE VI and TABLE VII for networks 1R1C and 2R2Cf, respectively. While the best validation performance is obtained with D6, it is not significantly better than when training with any of the other eleven datasets. The standard deviation of the average RMSE is less than 0.04 °C. It is concluded that in the final methodology the training dataset can consist of data from any month of the year. Since it is preferable to train with the less amount of data to reduce identification time, the training dataset is recommended to be 10 days long.

TABLE VI
VALIDATION RESULTS FOR TRAINED 1R1C NETWORKS

Training dataset	Avg. validation Fit, %	Avg. Validation RMSE, °C
D1	92.38	0.6160
D2	92.06	0.6416
D3	91.95	0.6509
D4	92.43	0.5985
D5	92.59	0.5851
D6	92.66	0.5794
D7	91.94	0.6539
D8	92.09	0.6422
D9	92.10	0.6415

D10	92.33	0.6113
D11	92.38	0.6073
D12	92.47	0.6004
Standard deviation	0.24	0.0261

TABLE VII
VALIDATION RESULTS FOR TRAINED 2R2Cf NETWORKS

Training dataset	Avg. validation Fit, %	Avg. Validation RMSE, °C
D1	95.93	0.3293
D2	95.20	0.3884
D3	95.18	0.3896
D4	95.60	0.3476
D5	95.83	0.3294
D6	95.87	0.3265
D7	94.86	0.4162
D8	95.19	0.3893
D9	95.23	0.3863
D10	95.04	0.3959
D11	95.23	0.3811
D12	95.55	0.3548
Standard deviation	0.3526	0.0305

This seemingly low sensitivity to the month in which the data is recollected can be due to the use of a principle-based model structure. Since the equations already describe the heat transfer processes on their own, the model is less reliant on data to appropriately describe the system. This is a clear advantage of RC networks to black-box techniques, the ability to generalize well with limited training data.

C. *Final methodology*

The final methodology, now complete with results from the previous sections is summarized in Fig. 8. The candidate models are trained 10 times with the dataset determined as best (which resulted in any dataset 10 days long). The parameters estimate dispersion is once again studied, as it is important to verify identifiability when applying the methodology to a new case study. If the parameters' CV is greater than 10% the model is discarded. Otherwise, it is validated. The models' performance is compared, and the RC network with the lowest validation RMSE is selected as the best model, that which best represent the case study's heat transfer dynamics.

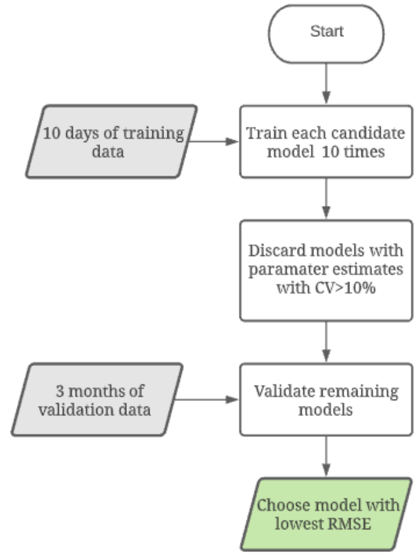

Fig. 8 Final methodology

D. *Methodology applied to Case 2*

The methodology is now implemented to determine the model that best represents Case study 2.

The dataset used for training is D10 and validation datasets are D3, D6 and D9. TABLE VIII and TABLE IX present the dispersion criteria that resulted from identifying 10 times models 1R1C and 2R2Cf with D10, respectively. For both networks, a CV lower than 10% is achieved. Therefore, both networks are validated.

TABLE VIII
PARAMETER DISPERSION IN NETWORK 1R1C (Case 2)

	Average	Standard deviation	CV, %
Ci	1.19E+04	1.08E+01	0.0908
R1	2.69E-02	6.58E-06	0.0244
F1	0.00E+00	N/A	N/A
G	6.46E-02	1.38E-05	0.0213

TABLE IX
PARAMETER DISPERSION IN NETWORK 2R2Cf (Case 2)

	Average	Standard deviation	CV, %
Ci	1.57E+04	1.02E+01	0.0650
Cf	1.62E+06	1.04E+03	0.0641
R1	9.57E-02	8.70E-05	0.0909
R2	1.88E-02	1.00E-05	0.0532
F1	1.47E-02	1.05E-05	0.0713
F2	1.29E+00	9.14E-04	0.0709
G	1.24E-01	7.54E-05	0.0609

The average validation performance is summarized in TABLE X . The indoor temperature simulated by models 1R1C and 2R2Cf is shown in Fig. 9 for two randomly chosen days of April. These temperatures are compared to that of the reference model built in DB.

TABLE X
VALIDATION PERFORMANCE FOR CASE 2

Network	Avg. Validation RMSE, °C	Avg. validation Fit, %
1R1C	0.5944	88.34
2R2Cf	0.3573	93.02

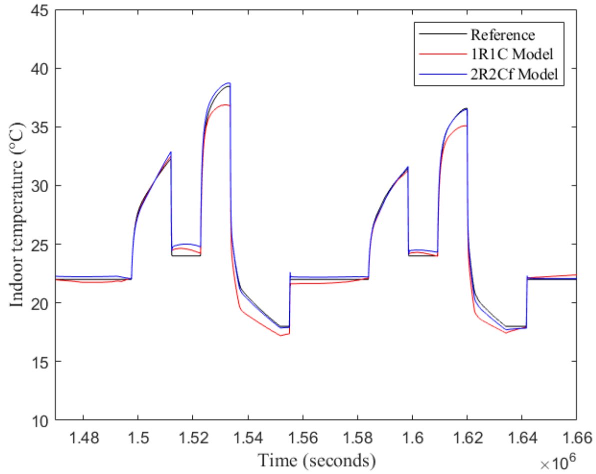

Fig. 9 Simulated indoor temperature by models

According to the validation performance results, the best model is network 2R2Cf, as it has the lowest RMSE among all candidate models. The model performance is satisfactory for control purposes, since the RMSE is less than 0.5°C [30].

E. Controller design

The thermal models identified have as input the supply temperature of the air cooled by an air conditioning system that operates at a constant flow rate (e.g., mini-splits). This by itself is not a controllable input in real air conditioning systems, however, if there is a model that relates the supply temperature to some controllable variable (e.g., refrigerant flow) it could be feasible to design a controller with the identified model. However, this is outside the scope of this research.

The purpose of this section is to design a controller that drives the supply temperature (manipulated variable) in such a way that the interior temperature remains close to the setpoint value, following the control loop in Fig. 10.

The controller will be designed taking the 2R2Cf model as the control loop plant. The gains of said controller will be adjusted with the PID Tuner of the Design and Analysis of Control Systems toolbox in MATLAB.

Once the controller gains have been adjusted, it will be inserted into another control loop that uses a state space black box model identified with N4SID (reference) as a plant. This model is of fifth order and has an average validation RMSE of less than 0.316°C, when trained with a full month of data.

Fig. 11 shows the closed loop response when the desired temperature is 18°C, starting from a similar initial condition. The response of the 2R2Cf loop stays closest to the desired temperature, with a -1.06% overshoot, compared to a 2% overshoot for the reference model. The identified model can successfully be used to tune a controller that will control another plant, avoiding the need for a tedious trial and error procedure.

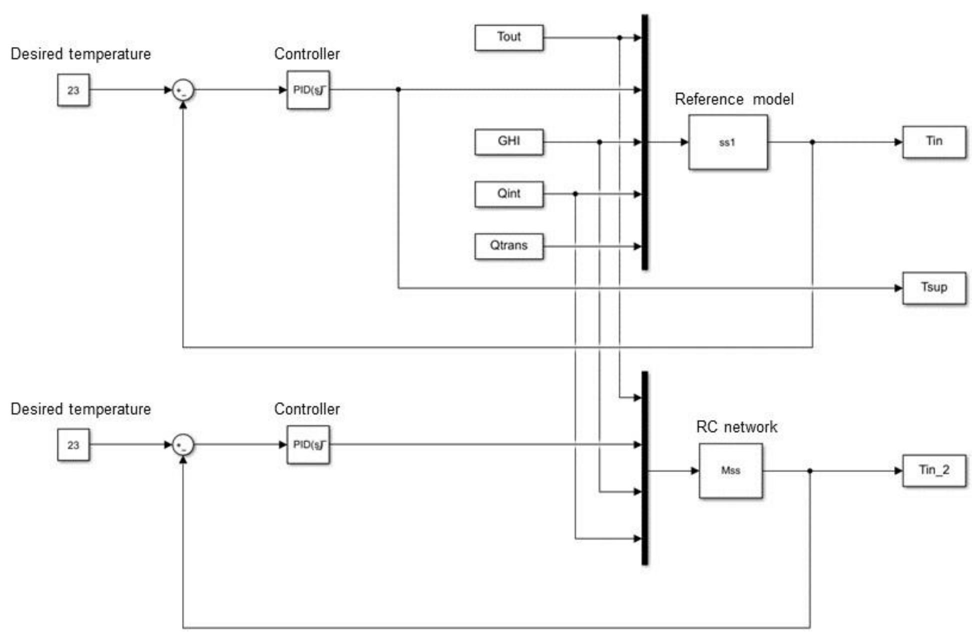

Fig. 10 Control loop implemented in Simulink

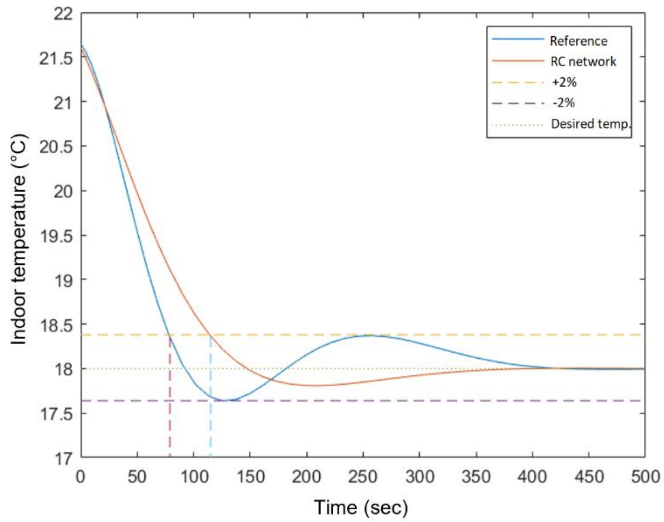

Fig. 11 Closed loop response

IV. CONCLUSIONS AND FURTHER WORK

In this study, a methodology to identify linear grey box models to represent home thermal dynamics was developed. The procedures through which the methodology was obtained were detailed and the resulting methodology may be applied to residential, one-zone case studies situated in a tropical climate.

We show that models with few parameters are generally more identifiable models, and the models can successfully represent the thermal dynamics of a house. Moreover, a linear model can accurately describe a small case study affected by prominent nonlinear phenomena, such as solar radiation and a high humidity.

There is also an apparent advantage to RC networks regarding the amount of identification data needed. The principle-based structure makes the models less reliant on data to achieve satisfactory performance, as it was obtained that the models performed similarly regardless of the dataset length, or the month used for training.

When the methodology was implemented in a different case study, a RMSE of 0.36°C was obtained for the best model, a 2R2C network. This model was then used to tune a PID controller which was then used to maintain a desired temperature in a plant modeled by another technique.

Further work will investigate the coupling of the thermal model with a hygric model since humidity is very important to maintain comfort in tropical climates. Also, it is relevant to include inputs or parameters that consider variable occupant behaviour to predict indoor temperature. It is of interest to design more complex control strategies, such as MPC, using the developed models and to study their effectiveness.

ACKNOWLEDGEMENTS

The authors would like to thank the Faculty of Mechanical Engineering at the Universidad Tecnológica de Panamá for their collaboration. Part of this publication has received funding from the National Secretariat of Science, Technology, and Innovation (SENACYT) under the project code FID18-056, FIED19-R2-005, and the Sistema Nacional de Investigación (SNI).

REFERENCES

[1] "Buildings – Topics," *IEA*. https://www.iea.org/topics/buildings (accessed Jun. 09, 2021).

[2] A. Afram and F. Janabi-Sharifi, "Black-box modeling of residential HVAC system and comparison of gray-box and black-box modeling methods," *Energy Build.*, vol. 94, pp. 121–149, May 2015, doi: 10.1016/j.enbuild.2015.02.045.

[3] Panamá. Secretaría Nacional de Energía, *Plan energético nacional, 2015-2050*. 2017.

[4] X. Li and J. Wen, "Review of building energy modeling for control and operation," *Renew. Sustain. Energy Rev.*, vol. 37, pp. 517–537, Sep. 2014, doi: 10.1016/j.rser.2014.05.056.

[5] "Sustainable Development Goals," *unfoundation.org*. https://unfoundation.org/what-we-do/issues/sustainable-development-goals/ (accessed Feb. 08, 2022).

[6] N. Fernandez, S. Katipamula, W. Wang, Y. Xie, and M. Zhao, "Energy savings potential from improved building controls for the US commercial building sector," *Energy Effic.*, vol. 11, no. 2, pp. 393–413, Feb. 2018, doi: 10.1007/s12053-017-9569-5.

[7] P. Hietaharju, M. Ruusunen, and K. Leiviskä, "A Dynamic Model for Indoor Temperature Prediction in Buildings," *Energies*, vol. 11, no. 6, p. 1477, Jun. 2018, doi: 10.3390/en11061477.

[8] L. Chen, B. Basu, and D. McCabe, "Fractional order models for system identification of thermal dynamics of buildings," *Energy Build.*, vol. 133, pp. 381–388, Dec. 2016, doi: 10.1016/j.enbuild.2016.09.006.

[9] V. S. K. V. Harish and A. Kumar, "A review on modeling and simulation of building energy systems," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 1272–1292, Apr. 2016, doi: 10.1016/j.rser.2015.12.040.

[10] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo, and P. Chatellier, "Modeling and forecasting building energy consumption: A review of data-driven techniques," *Sustain. Cities Soc.*, vol. 48, p. 101533, Jul. 2019, doi: 10.1016/j.scs.2019.101533.

[11] A. Kathirgamanathan, M. De Rosa, E. Mangina, and D. P. Finn, "Data-driven predictive control for unlocking building energy flexibility: A review," *Renew. Sustain. Energy Rev.*, vol. 135, p. 110120, Jan. 2021, doi: 10.1016/j.rser.2020.110120.

[12] T. Söderström and P. Stoica, *System Identification*. Prentice Hall, 1989.

[13] F. Massa Gray and M. Schmidt, "A hybrid approach to thermal building modelling using a combination of Gaussian processes and grey-box models," *Energy Build.*, vol. 165, pp. 56–63, Apr. 2018, doi: 10.1016/j.enbuild.2018.01.039.

[14] M. Dahl Knudsen and S. Petersen, "Demand response potential of model predictive control of space heating based on price and carbon dioxide intensity signals," *Energy Build.*, vol. 125, pp. 196–204, Aug. 2016, doi: 10.1016/j.enbuild.2016.04.053.

[15] J. Drgoňa *et al.*, "All you need to know about model predictive control for buildings," *Annu. Rev. Control*, vol. 50, pp. 190–232, 2020, doi: 10.1016/j.arcontrol.2020.09.001.

[16] F. Paschke and T. Zaiczek, "Identification of dynamic models for the short-term temperature prediction in a single room," *IFAC-Pap.*, vol. 51, no. 2, pp. 79–84, 2018, doi: 10.1016/j.ifacol.2018.03.014.

[17] S. Haykin, S. S. Haykin, and S. A. HAYKIN, *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1999.

[18] C. Finck, R. Li, and W. Zeiler, "Economic model predictive control for demand flexibility of a residential building," *Energy*, vol. 176, pp. 365–379, Jun. 2019, doi: 10.1016/j.energy.2019.03.171.

[19] A. Afram, F. Janabi-Sharifi, A. S. Fung, and K. Raahemifar, "Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system," *Energy Build.*, vol. 141, pp. 96–113, Apr. 2017, doi: 10.1016/j.enbuild.2017.02.012.

[20] A. Foucquier, S. Robert, F. Suard, L. Stéphan, and A. Jay, “State of the art in building modelling and energy performances prediction: A review,” *Renew. Sustain. Energy Rev.*, vol. 23, pp. 272–288, Jul. 2013, doi: 10.1016/j.rser.2013.03.004.

[21] G. Parnis, “Building Thermal Modelling Using Electric Circuit Simulation,” Master of Engineering by Research, University of New South Wales, Sydney, Australia, 2012.

[22] M. G. Davies, *Building Heat Transfer*, 1st ed. Wiley, 2004. doi: 10.1002/0470020555.

[23] Z. Wang and Y. Chen, “Data-driven modeling of building thermal dynamics: Methodology and state of the art,” *Energy Build.*, vol. 203, p. 109405, Nov. 2019, doi: 10.1016/j.enbuild.2019.109405.

[24] J. Wang, C. Y. Tang, M. R. Brambley, and L. Song, “Predicting home thermal dynamics using a reduced-order model and automated real-time parameter estimation,” *Energy Build.*, vol. 198, pp. 305–317, Sep. 2019, doi: 10.1016/j.enbuild.2019.06.002.

[25] A. K. Rivera, J. Sánchez, D. Mora, and M. C. Austin, “Inspection of the Identification of Thermal Modeling Process in Buildings: A Proposed Methodology for a Case Study in Tropical Climate,” in *Advances in Automation and Robotics Research*, Cham, 2022, pp. 172–180.

[26] B. Cui *et al.*, “A hybrid building thermal modeling approach for predicting temperatures in typical, detached, two-story houses,” *Appl. Energy*, vol. 236, pp. 101–116, Feb. 2019, doi: 10.1016/j.apenergy.2018.11.077.

[27] S. Yang *et al.*, “A state-space thermal model incorporating humidity and thermal comfort for model predictive control in buildings,” *Energy Build.*, vol. 170, pp. 25–39, Jul. 2018, doi: 10.1016/j.enbuild.2018.03.082.

[28] A. K. Tangirala, *Principles of System Identification: Theory and Practice*. CRC Press, 2018.

[29] Z. Wang, Y. Chen, and Y. Li, “Development of RC model for thermal dynamic analysis of buildings through model structure simplification,” *Energy Build.*, vol. 195, pp. 51–67, Jul. 2019, doi: 10.1016/j.enbuild.2019.04.042.

[30] O. M. Brastein, D. W. U. Perera, C. Pfeifer, and N.-O. Skeie, “Parameter estimation for grey-box models of building thermal behaviour,” *Energy Build.*, vol. 169, pp. 58–68, Jun. 2018, doi: 10.1016/j.enbuild.2018.03.057.