



Use of optimised MLP neural networks for spatiotemporal estimation of indoor environmental conditions of existing buildings

Miguel Martínez-Comesaña^{a,*}, Ana Ogando-Martínez^a, Francisco Troncoso-Pastoriza^b,
Javier López-Gómez^a, Lara Febrero-Garrido^b, Enrique Granada-Álvarez^a

^a Department of Mechanical Engineering, Heat Engines and Fluids Mechanics, Industrial Engineering School, Universidade de Vigo (University of Vigo), Maxwell s/n, 36310 Vigo, Spain

^b Defense University Centre, Spanish Naval Academy, Plaza de España, s/n, 36920 Marín, Spain

ARTICLE INFO

Keywords:

Air quality
IEQ
Interpolation
MLP
NSGA-II
Thermal comfort

ABSTRACT

Controlling the indoor environmental quality in real time is essential for the health, well-being and productivity of occupants of a building. In recent years, research has focused on improving monitoring devices and strategies and developing techniques for estimating indoor conditions. The use of machine learning algorithms in this context has increased considerably. However, monitoring data in real time from large multizone working areas is challenging. The aim of this work is to provide an interpolation methodology based on the use of optimised multilayered perceptron neural networks to estimate the indoor environmental conditions of a building in real time. These estimations are obtained without the need for neither monitoring in the occupied working area nor human intervention and considering low-cost sensors. The neural network is optimised by implementing the multiobjective genetic algorithm NSGA-II to find the best architecture in terms of error and complexity. This method was applied to the building of a research centre in north-western Spain, where interpolated values for indoor air temperature, relative humidity and CO₂ concentration were obtained. The results of this case study yielded relative errors close to 6% for temperature, 5% for relative humidity, and 12% for CO₂ concentration. These values validate the methodology developed for the estimation of indoor environmental conditions and the contribution of this research to the improvement of the monitoring and control of the indoor environmental quality of a building.

1. Introduction

Considering that people spend 80%–90% of their time indoors [1], indoor environmental quality (IEQ) has become a key study factor for many researchers. It is well known that both air quality and thermal, acoustic and visual comfort influence the well-being, health and productivity of a building's occupants [2,3]. Moreover, building consumption in Europe accounts for 40% of primary energy consumption and 36% of greenhouse gas emissions [4]. The practice of constructing green buildings, which are structures that make more efficient and sustainable use of resources, is currently widespread. These buildings are characterised by reduced energy consumption, but reducing consumption does not always ensure occupant satisfaction or improvements in IEQ [5,6]. Indeed, consumption is closely related to the IEQ since systems such as lighting, heating, cooling, and ventilation are responsible for maintaining suitable conditions inside the building. Therefore, IEQ monitoring is key to verify that the reduction in consumption is not detrimental to environmental quality.

This research focuses on thermal comfort and air quality, taking into account that air quality monitoring is essential at this time due to the COVID-19 pandemic. Among the four IEQ categories, the occupants of a building consider thermal comfort to be the most important parameter [2]. Numerous studies have been conducted in recent years to determine the effect of different indoor environmental conditions on thermal comfort and air quality and identify the most influential ones [3,7–9]. These studies revealed that indoor temperature and relative humidity are two of the most commonly used variables in the calculation of thermal comfort, while the indoor concentration of pollutants, especially carbon dioxide (CO₂), has been frequently used to assess air quality and the need for air changes. Identifying and monitoring these variables has become a challenge for many researchers. Among the most commonly used measurement methods for monitoring indoor environmental conditions, the use of measuring devices installed at fixed positions stands out [10–14]. This can be interesting in small spaces, where the use of a single point to represent the entire space

* Corresponding author.

E-mail address: migmartinez@uvigo.es (M. Martínez-Comesaña).

<https://doi.org/10.1016/j.buildenv.2021.108243>

Received 7 May 2021; Received in revised form 26 July 2021; Accepted 7 August 2021

Available online 13 August 2021

0360-1323/© 2021 The Authors.

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Abbreviations

NSGA	Nondominant Sorting Genetic Algorithm;
IEQ	Indoor Environmental Quality;
ANN	Artificial Neural Network;
MLP	Multilayered Perceptron;
CV(RMSE)	Coefficient of Variation of the Root Mean Squared Error;
NMBE	Normalised Mean Bias Error;
RMSE	Root Mean Squared Error;
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers;
ReLU	Rectified Linear Unit;
PBI	Penalty Boundary Intersection;
CINTECX	Centre for Research in Technology, Energy and Industrial Processes;
PLA	Polylactic Acid;
GFS	Global Forecast System;
NOAA	National Oceanic and Atmospheric Administration;

provides a simplification that does not result in an excessive loss of accuracy. However, in larger areas where the indoor environment is not homogeneous, it is difficult to represent the whole environment with a single measurement. For this reason, mobile measurement devices began to be used to take measurements at different locations in a room [15–18]. The present work combines both methods and uses fixed devices installed outside the occupied area, where they do not interfere with the performance of the building users, and a mobile measuring device that moves around the study area. The devices used in this research are notable for using low-cost sensors.

Accurately knowing the indoor environmental conditions of a building is a big step to be able to improve, if necessary, the environmental quality of the building. However, estimating the IEQ at any point of a building exactly is a complicated issue. To solve this problem, there have been different approaches in recent years. On the one hand, monitoring systems (sensors, positioning, devices, etc.) inside a building have been improved to better understand the characteristics of the environment throughout the building [19,20]. On the other hand, white box models based on simulations [21–23], spatial visualisation [24] or statistical methods [16] have also been used to attempt to estimate the indoor environmental conditions of a building. In addition, although not so widespread, grey box models focused on spatial interpolations such as *kriging* have been used to analyse the environmental quality inside buildings [25]. Furthermore, the application of advanced mathematical modelling techniques such as machine learning (black box) models has become more common in this field in recent years [17,26].

The application of artificial intelligence models in building energy and environmental analyses has increased in recent years [17,27–29]. These models stand out for their robustness, ease of implementation and strong performance [28–30]. Another reason for their growing use is that black box models have proven successful in learning complex patterns without the need for specific knowledge of the subject [30]. Among these models, which are defined as mathematical models capable of learning a pattern from data and extrapolating it to new data samples, the most used and well-known model is the artificial neural network (ANN). Specifically, ANNs have a notable ability to adjust and model nonlinear relationships between model inputs and outputs due to their massive interconnectivity [27,31]. Among ANN models, the most widely deployed is the multilayered perceptron (MLP) neural network. This type of ANN is characterised by being composed of

several layers: input layer, output layer and zero, one or more hidden layers between them [32]. MLP neural networks have already been applied in scientific fields such as biotechnology [31,33], sensors [34], aerospace science [35], environmental [32,36,37] and building energy analyses [27,28]. Furthermore, numerous studies have shown the usefulness of adjusting the model parameters to their optimal values [38–40]. Specifically, the use of genetic algorithms (GAs) for optimising the parameters and structure of MLP neural networks has been carried out in many studies [41–43]. Moreover, there are some genetic algorithms, such as the nondominant sorting genetic algorithm (NSGA-II), that in the optimisation process can focus on more than one objective [44–46]. Previous studies, in which the IEQ of a building is estimated using machine learning models, are focused on small spaces [15–17], require devices installed within the occupancy zone [15,17], use simple machine learning models [16] or without any type of optimisation [29]. Jin et al. [15,29] compared the performance of different machine learning models estimating the air change effectiveness in a small room. These studies are based on sensors installed in the occupancy zone considering temperature, illuminance and pollutant particles such as CO₂. Other study, conducted by Shen et al. [16], uses a classical multivariate regression to estimate particulate matter concentrations in a small space taking into account the emissions from the kitchen and the values of the outdoor infiltrations collected by the sensors placed throughout the analysed space.

The aim of this work is to introduce a methodology to estimate the indoor environmental conditions of an in-use building, both spatially and temporally, using optimised MLP neural networks [17]. In particular, this study focuses on estimating indoor temperature, relative humidity and CO₂ levels, but the methodology developed can be applied to any other variable of interest that may have an effect on the IEQ. The analysed building in this study was a centre for research in north-western Spain. The estimations were obtained by means of an interpolation technique, and the optimal architectures of the built MLPs were achieved with NSGA-II [47,48]. In this case, the two objective functions to minimise were the coefficient of variation of the root mean squared error (CV(RMSE)) and a complexity function considering the number of hidden neurons and layers [41]. The available data consisted of minute-frequency observations of the three variables of interest measured by sensors installed in eight fixed positions and an additional mobile device. Ten-minute-frequency meteorological variables of the area where the studied building is located were also available. In addition, three time variables (hour of the day, day of the week and day of the year) were considered as model inputs to improve the training process. A spatial analysis of the environmental conditions throughout the analysed multizone building, together with a temporal analysis over three specific days, was carried out after the overall good performance of the MLPs was demonstrated by a positional cross-validation experiment (run through the entire data sample separating each of the studied positions). The Normalised Mean Bias Error (NMBE), RMSE and, as recommended by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), CV(RMSE) were the error measures selected to evaluate the built MLPs [28,49].

The novelty of this work lies in the application of machine learning models, optimised with a genetic algorithm, to perform spatiotemporal estimations of the environmental conditions in a large open space inside a building without the need to monitor the occupation area. Furthermore, the methodology presented does not rely on physical simulations, which would require significant knowledge of the subject and model parameters. In this way, this paper contributes with a method that minutely estimates the values of certain variables related to the IEQ at any position in a multizone indoor space and, consequently, can show a real-time image of its distribution throughout the building. Thus, this method becomes essential to enable an effective evaluation of the IEQ conditions of the building, with the possibility of directly integrating this additional information into its digital twin representation.

2. Materials and methods

The aim of this work is to determine the environmental conditions inside a building without affecting the activity being carried out. Thus, this research attempts to develop a spatiotemporal interpolation methodology to obtain the environmental conditions within the occupation area from measurements performed at a set of fixed points outside that area. This methodology combines the use of neural networks with a genetic algorithm, as shown in Fig. 1.

A multiobjective genetic algorithm is used to find the optimal neural network architecture that minimises a set of target functions in an iterative process. For each iteration, using the independent and dependent variables (measured outside and inside the occupation area, respectively) and the architecture proposed by the genetic algorithm, the neural network is trained and tested to generate a model. The error metric evaluated in that testing process and the complexity of the neural network architecture are the objective functions that the genetic algorithm attempts to minimise. Once the optimal model is identified and trained, it is no longer necessary to conduct measurements within the occupation zone.

2.1. Machine learning model

Machine learning models are data-driven models based on mathematical algorithms that are highly successful in learning complex patterns and producing accurate estimates without the need for extensive knowledge in the field of their application [26,28]. The algorithm adopted to estimate the environmental conditions inside an in-use building is an ANN. Such algorithm try to replicate the functioning of the human brain network and therefore has an extraordinary ability to model non-linear relationships [27]. In this work, the ANN implemented is the multilayered perceptron (MLP) neural network. This machine learning model and the process to select its optimal architecture are explained in the following sections.

2.1.1. Multilayered perceptron neural network

MLP is one of the most widely used ANNs. Its structure is divided into three types of layers that are fully connected between them: the input layer, where the model inputs are introduced; the output layer, where the results of the trained model are obtained; and the hidden layers, which are each of the intermediate layers between the previous ones and can be zero, one or more. The connections between them are based on a weight structure (see Fig. 2) whose values are altered through model training, building a stable framework with decision capacity [32,50,51].

Assuming that the number of hidden layers in the MLP neural network is G (see Fig. 2) and the neurons in each of the layers are respectively set as $n_0, n_1, \dots, n_{G+1}$, the inputs and outputs of the first hidden layer have the shape presented in Eqs. (1) and (2) [32,50,52]:

$$i_n = \sum_{j=1}^{n_0} w_{jn} f_0(x_j) + b_n \quad (1)$$

$$o_n = f_1(i_n) = (1 + e^{-i_n})^{-1} \quad (2)$$

where i_n , o_n and b_n are the input, output and threshold of the n th neuron in the first hidden layer, respectively. In addition, while w_{jn} represents the weight value connecting the j th input neuron with the n th neuron in the first hidden layer, $f_k(\cdot)$ represents the activation function of the layer k [32,50]. Therefore, the input and output of the neurons in the output layer are obtained as shown Eqs. (3) and (4):

$$i_1 = \sum_{j=1}^{n_G} w_{j1} o_j + b_1 \quad (3)$$

$$o_1 = i_1 \quad (4)$$

where i_1 , o_1 and b_1 correspond to the input, output and threshold of the neuron in the output layer. Moreover, w_{j1} represents the weight value

that connects the j th neuron in the last hidden layer to the neuron in the output layer.

The MLP training process is based on the error backpropagation method. In this way, this algorithm needs real values to fit the weights estimating the errors between the MLP predicted values and the real ones [53]. Furthermore, during model training, the updates of the weights are based on an average update (*batch-learning*) obtained by entering all the patterns as inputs (an *epoch*) and accumulating all updates of the weights [54]. The training process is usually based on a gradient descent algorithm. Nevertheless, a conventional gradient descent presents significant disadvantages such as problems in searching for local optima and a slow convergence rate [50,55]. Minibatch gradient descent, an algorithm based on batch gradient descent and stochastic gradient descent, avoids these potential problems. This algorithm, which is used for training the MLPs presented in this study, offers advantages regarding the search for global optima and a higher convergence rate [32,56]. Additionally, to avoid overfitting problems, the following stop criterion is selected [57]: training is stopped when the performance of the MLP, represented by the mean square error (MSE: $\frac{1}{n} \sum_{i=1}^N (\hat{Y}_i - Y_i)^2$) and tested with an independent validation sample (*test sample*), begins to decrease or stagnates. The selected threshold, in form of the number of epochs in which the model is allowed to not improve, is set at 50. Last, all MLPs built in this analysis are trained using a rectified linear unit (ReLU) activation function: $\max(0, x)$ [58], the adaptive moment estimation (Adam) optimiser [59] and a Gaussian kernel as initialiser. Further information about MLP training can be found in [27,32].

2.1.2. Model selection

The optimal architectures of the built MLPs (number of hidden layers and hidden neurons) are obtained by solving a multiobjective problem with a genetic algorithm. This type of algorithm attempts to replicate biological evolution to solve optimisation problems [48,60]. They start with a random population based on individuals, which are represented by chromosomes that are structured into genes. In this context, genes are values for the explanatory variables (also known as *covariates*) of the model. Moreover, the optimisation is carried out using three individual operators [47,61]:

- *Crossover*: Swapping a segment of a specific chromosome with the segment of another chromosome at random. This provides a combination of solutions and convergence in a subspace.
- *Mutation*: Random modification of parts of one solution. This increases the diversity in the population and allows escape from local optima.
- *Selection*: The best chromosomes are selected (*elitism*) to pass on to the next generations.

In this paper, the selected genetic algorithm is the nondominated sorting genetic algorithm (NSGA-II), a robust multiobjective algorithm widely used in numerous real-world applications [47,61]. It was originally developed as an improved version of the NSGA due to the incorporation of elitism, a fast nondominated ranking approach and diversity along the Pareto optimal front using a crowding distance operator. Concretely, it is one of the most powerful algorithms in multiobjective optimisation applied in building performance analysis [47].

The multiobjective problem is based on two metrics that are to be minimised: the CV(RMSE) (Eq. (7)) on the one hand and a complexity function (Eq. (5)) on the other. The complexity function, which is a slight modification of the function used in [41], is based on the number of neurons and layers of the neural network architecture and has the following form:

$$\text{Complexity} = 0.25 \frac{h}{H} + 0.75 \frac{\sum_{i=0}^H n_i}{N} \quad (5)$$

where h is the number of layers used in the evaluated architecture, and H represents the maximum value allowed for the h parameter (in this

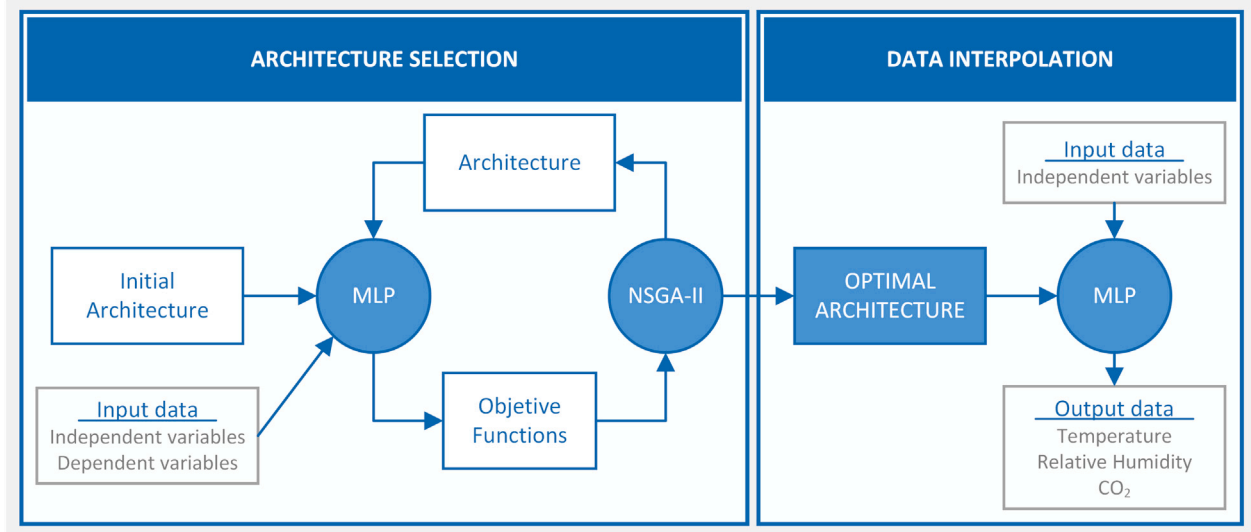


Fig. 1. Interpolation methodology implemented to determine the temperature, relative humidity and CO₂ levels in the area of interest from measurements conducted outside that area.

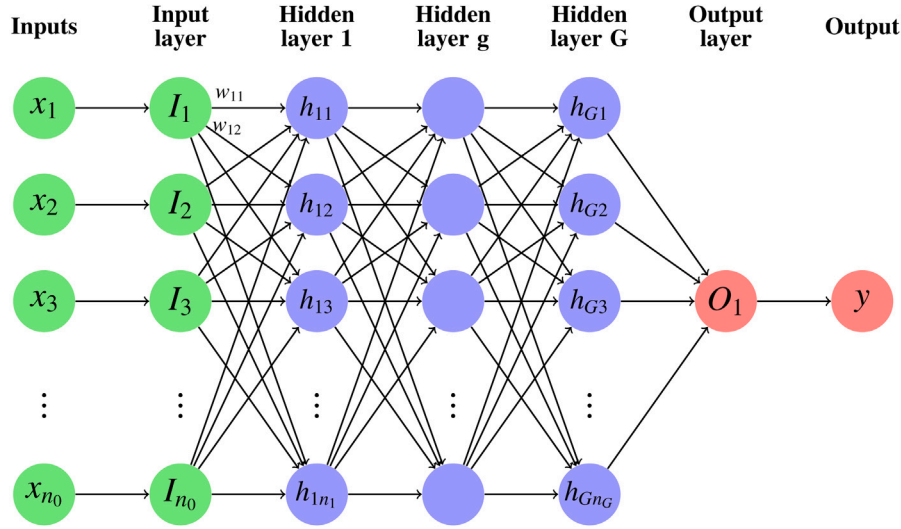


Fig. 2. MLP neural network architecture with N inputs, G hidden layers, an univariate output and weights w_{jn} .

case, 3). In addition, n_1, n_2 and n_3 are the neurons in each layer, and N is the maximum number of neurons that can be in the model (in this case, 600). It can be seen that the complexity that comes from the number of layers has less impact in our modification. This mitigates the rejection of too many multilayered structures. On the other hand, the decomposition function used to find the optimum point in the Pareto front obtained by the genetic algorithm was the penalty boundary intersection (PBI) [62]. Further information about the NSGA-II and its implementation can be found in [44,45,48,60].

2.2. Error measurements

The error metrics considered in this work to measure the accuracy of the built MLPs are the root mean square error (RMSE), coefficient of variation of the RMSE (CV(RMSE)) and normalised mean bias error (NMBE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (6)$$

$$CV(RMSE) = 100 \times \frac{\sqrt{\sum_{i=1}^N (y_i - \hat{y}_i)^2 / N}}{\bar{y}} \quad (7)$$

$$NMBE = 100 \times \frac{\sum_{i=1}^N (y_i - \hat{y}_i)}{\sum_{i=1}^N (y_i)} \quad (8)$$

These metrics are recommended by ASHRAE [49] and have been used in similar studies [24,47,63]. Additionally, Martínez et al. [47] demonstrated that CV(RMSE) is the best metric in building simulation analysis. CV(RMSE) is also one of the objective functions taken into account in the multiobjective problem that seeks the optimal MLP structures.

3. Experimental system

The building under study is the Centre for Research in Technology, Energy and Industrial Processes (CINTECX) located on the campus of the University of Vigo in north-western Spain and facing northwest. The general climate of the region where the building lies is a mixture between Cfb (atlantic influence) and Csb (mediterranean-atlantic transition) according to the Köppen-Geiger classification [64]. Ten research

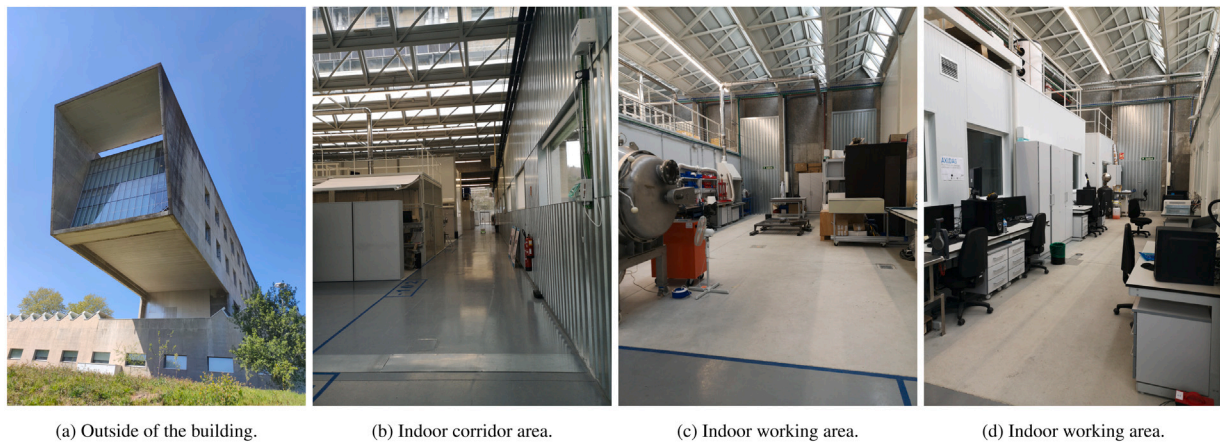


Fig. 3. Photographs of the studied building.

groups work in this building in a variety of areas including energy, electronics and automation, transport, sustainability and biomedical engineering. The diversity of activities that take place inside makes it attractive for this study. The measuring devices are installed on the ground floor of the building, where the laboratories are located. This space is where the research groups carry out the experimental part of their research, and it is a particularly interesting area for this reason. Research involves a wide variety of equipment, such as boilers, heaters, welding equipment, and engines that can affect the comfort of occupants and disrupt air quality. The current heating and cooling system consists of 4 fan coils distributed throughout the room. Fig. 3 shows pictures of the building and some of the working areas.

The data required for the study are gathered by two different data sources: first, a set of measuring devices are used to take measurements inside the building; second, the weather data from outside the building are gathered from a nearby meteorological station, complemented with the outputs of a numerical weather prediction model when direct values are not available.

3.1. Monitored data

The indoor data source used in this research consists of 9 tailor-made devices, shown in Fig. 4, containing sensors for the measurement of indoor temperature, relative humidity and CO₂ levels. One of the devices is mobile (D0) and is used to collect data from multiple positions inside the working area. It also has a platform of variable height that facilitates the collection of data at 0.3 m intervals between heights of 0.55 m and 1.75 m. The remaining devices (D1 to D8) are wall-mounted in fixed locations (where they do not disturb the users of the building) and at heights between 1.9 and 3.2 m. These devices are characterised by their reduced cost in terms of hardware components and in terms of manufacturing and assembly. The structural components are mainly manufactured by 3D printing in polylactic acid (PLA).

All measuring devices are equipped with the same sensors to measure the variables of interest at a frequency of one minute. The temperature and relative humidity are measured by SHT31-D sensors, while MHZ-14 sensors are used to acquire CO₂ values. The technical specifications of the sensors used in this work, which are low-cost (less than 35€ for the most expensive ones), are detailed in Table 1.

Some of the wall-mounted devices (D1, D2, D3, D4 and D7) are placed at the borders of the area of interest in which measurements are collected and interpolations are performed. This is a multizone open space of approximately 825 m² and a height of 6 m. Fig. 5 shows the position of the fixed devices and the different locations monitored with the mobile device. Only spaces where building users work and spend most of their time are studied. Closed laboratories, illustrated in Fig. 5 with a shaded background, may have very particular environmental

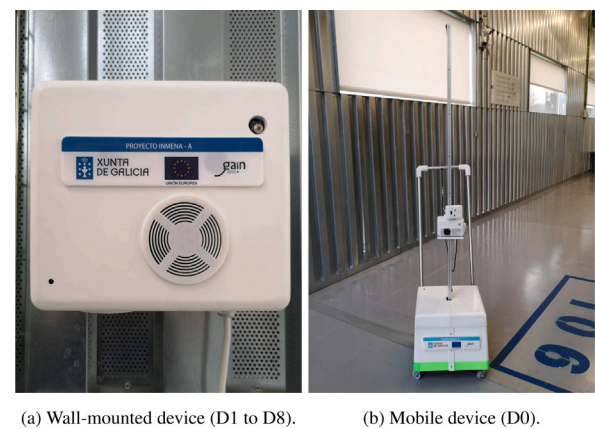


Fig. 4. Measuring devices installed in building.

Table 1

Technical specifications of used sensors.

Variable	Sensor	Accuracy tolerance	Measuring range	Resolution	Response time
Temperature	SHT31-D	±0.3 °C	−40 to 125 °C	0.015 °C	8 s
Relative humidity	SHT31-D	±2%	0 to 100%	0.01%	2 s
CO ₂	MHZ-14	±50 ppm ±5%	0 to 5000 ppm	10 ppm	<30 s

Table 2

Device coordinates given as length (x), width (y) and height (z).

Device	x [m]	y [m]	z [m]	Device	x [m]	y [m]	z [m]
D1	0	16	2.4	D5	11.5	7.8	1.9
D2	4	0	3	D6	20.6	9.4	2.2
D3	50.1	3	3.2	D7	30.6	15.7	2.5
D4	47.1	14.4	3.2	D8	44	6.3	2.1

conditions that are not representative of the rest of the spaces and that have no effect on the workers since they are not usually occupied during the experimental tests. The envelope of this part of the building is constructed of concrete (35.5 cm wide) and is characterised by a large number of windows (approximately a 40% of the exterior envelope). The two shorter sides of the building are fully glazed, and part of the sawtoothed roof is also glazed. In addition, one of the long sides of the building has windows (see Fig. 3). The specific positions where wall-mounted devices are placed are presented in Table 2, measured in metres and disaggregated into length, width and height.

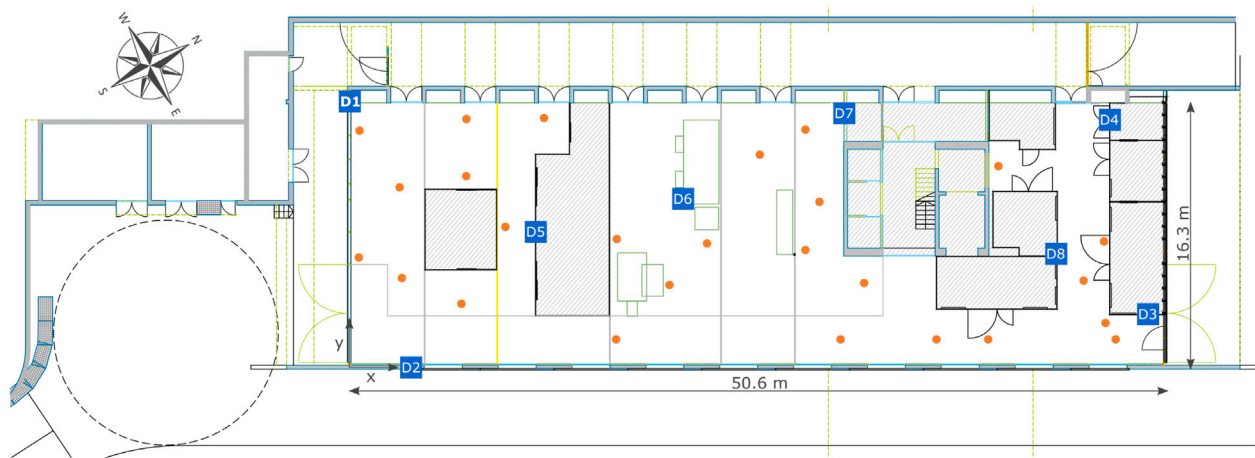


Fig. 5. Blueprint of the study area. The blue boxes represent the positions of the wall-mounted devices, the orange dots indicate the positions in which the mobile device was temporarily located, and the shaded areas depict closed laboratories that were omitted from this study. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Weather data

The main source of environmental meteorological data is an automatic weather station belonging to the MeteoGalicia weather agency. This station is located 300 m north-east of the centre of the studied building with an uphill height difference of 43 m. The weather variables retrieved from this station are the mean air temperature, relative humidity and global solar irradiance. Measurements were performed at a height of 1.5 m from the local ground surface.

Data values from the aforementioned weather station are retrieved using a RSS service provided by MeteoGalicia [65]. Data are provided with a 10-min temporal resolution and a time delay of 15–25 min from real time.

The complementary meteorological data source, used as a backup for replacing missing or invalid data from the main weather source, is the forecast data obtained from the Global Forecast System surface flux (GFS sflux) model. GFS is a global-scale numerical weather prediction model operated by the National Oceanic and Atmospheric Administration (NOAA) of the United States. All GFS products are executed four times each day starting at 00 h UTC and every 6 h, and their output files are stored at the NOAA Model Archive and Distribution System (NOMADS) repository for a 10-day interval [66]. The GFS sflux version provides hourly forecasts over a ~ 13 km resolution horizontal grid of points. A detailed explanation of the GFS model and products is available at [67].

For each analysed time, the GFS sflux output file corresponding to the shortest forecast term available is used to extract weather data at the nearest grid point. This point is located at 42.11519°N 8.67187°W , 5.94 km away from the studied building. From the forecast variables available in the GFS output files, the ones most similar to the three aforementioned weather variables are temperature at 2 m height, relative humidity at 2 m height, and shortwave downward solar irradiance at ground level. These are the same meteorological variables that were used in a previous study by some of the authors [68], where the data estimations of GFS sflux were compared with data observations from the MeteoGalicia network, achieving a similar estimation performance.

3.3. Data preprocessing

This analysis is focused on interpolating specific variables related to the indoor environment of a building. As already mentioned, the variables selected for interpolation are temperature, relative humidity and CO_2 concentration (see Fig. 6). The data available in this study are minute observations of the three variables of interest measured from sensors located in eight fixed devices (see Table 2) and from

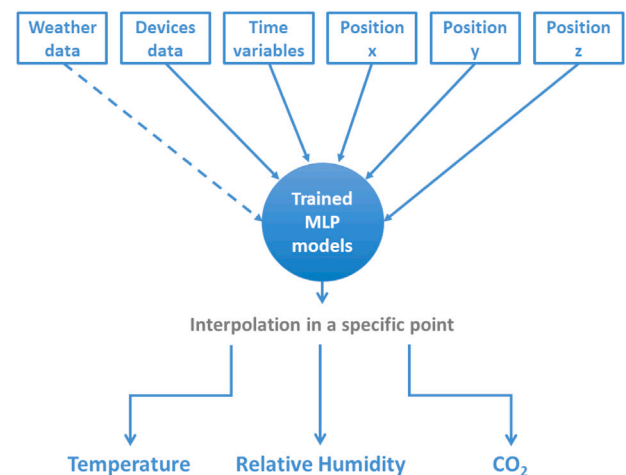


Fig. 6. Summary of the interpolation process after training the machine learning model. Boxes represent the inputs of model, and the variables at the bottom represent outputs of the model. The dashed line in weather data indicates that these data are not taken into account in the model built to interpolate CO_2 .

a mobile device (see Fig. 5). The availability of these data extends from November 18, 2020, through March 10, 2021. The entire sample contains data from 26 different positions of the mobile device (see Fig. 5) and slightly fewer than 140,000 observations (although at many times not all devices were sending data). In this case, the algorithm is trained to withstand up to two nonoperating wall-mounted devices at each analysed instant. The structure of the input and output variables of the study presented in this paper is summarised in Fig. 6.

The black box models are trained by combining data from the wall-mounted devices, weather observations, three ancillary temporal variables, the positions of the mobile device and the measurements it takes (which are the desired values). Specifically, the positions of the mobile device are transformed into distances to the fixed devices, considering the x-, y- and z-axes, to make the trained model more accurate. The data monitored by the mobile device are time-interpolated because the sensors do not take measurements during its vertical movement. In addition, a smoothing process based on the *moving average technique* [69] is applied to the device data to balance out the combined effects of this time interpolation and the sensor errors. For the external weather data, each 10-min frequency observation value is replicated to generate data inputs of the required frequency. Finally, three temporal

variables (hour of the day, day of the week and day of the year) are computed from the timestamp of the sensor data.

Each of the models built in this study considers a different set of variables together with the time variables and the distances to the wall-mounted devices that are common in all models. Thus, to interpolate indoor temperatures and relative humidity values, only the values of the variable itself measured by the fixed devices together with the outdoor temperature and solar irradiance are considered. For CO₂ levels, only the values from wall-mounted devices are taken into account, due to the low correlation between indoor CO₂ levels and the analysed meteorological variables. Other variables such as occupancy rate, air exchange rate or spatial parameters are significantly related to the CO₂ levels inside a building, but they are not considered in this study. The reasons for this are the existence of numerous previous studies that support this approach [15,24,70,71], the difficulty of monitoring this type of variables (occupancy rate is usually estimated using different techniques [72–74]) and, as shown below, the accuracy obtained by the presented method is comparable to other studies that consider the aforementioned variables [75]. Finally, a time lag of one minute is introduced in the environmental variables inside the building to model the inertia of these parameters [27].

4. Results and discussion

A methodology based on optimised MLP neural networks to interpolate temperature, relative humidity and CO₂ values in any location in a large open space inside a building is designed in this paper. To validate this methodology, a case study is presented in the CIN-TECX building located at the University of Vigo. The available data comprise four months of observations measured every minute (with approximately a 13% of missing values due to sensor errors and the movement of the mobile device). On the one hand, in Sections 4.1 and 4.2, the MLP neural network efficiency is analysed by taking into account the optimisation results and error metrics presented (RMSE, CV(RMSE) and NMBE). First, Section 4.1 introduces the selection process of the optimal MLP architecture; second, Section 4.2 presents the average accuracy of the selected models interpolating data by positions throughout the entire sample and on specific days. In this section, positional analysis is based on accuracy measurements of the model in positions where it has no information since these values were previously removed from the training dataset (positional cross-validation). Furthermore, three validation samples (days) are studied to show the temporal performance of the MLP compared to the monitored values. On the other hand, in Section 4.3, the usefulness of the optimised and validated MLP neural networks is presented by spatially studying the indoor environmental conditions in the building and by observing their changes over time. Finally, the full implementation of the proposed methodology and the results presented hereinafter are based on the Python programming language [76].

4.1. MLP optimisation

The optimised MLP neural network structures are obtained with the genetic multiobjective algorithm NSGA-II. The CV(RMSE) (Eq. (7)) and the complexity function (Eq. (5)) are the objective functions that the genetic algorithm minimise to find optimal structures. In addition, Table 3 shows the parameters of the NSGA-II used in the evaluation process of the possible architectures and in the optimisation stopping criterion.

The most relevant parameters are the neuron choices (10–200 in each of the layers), representing 7220 possible permutations, and the stopping criterion. This criterion does not take into account the specific number of evaluations but the tolerance in the feasible space. In addition, among the options for each Pareto front, the MLP structure is selected through the decomposition function *PBI*, giving more weight to the error function: 0.75/0.25. The process of selecting the optimal

Table 3

Parameters and functions used in the optimisation through NSGA-II, comprising the general parameters related to the multiobjective algorithm and the specific parameters related to the optimal selection.

General parameter	Value	Termination parameter	Value
Neurons options	10–200 (10 by 10)	Tolerance (<i>tol</i>)	0.0025
Layer options	1–3	N° max evals (<i>n_{max}</i>)	4000
Population	50	Last genes considered (<i>n_{last}</i>)	10
Mutation	0.9	Decomposition function	<i>PBI</i>
Crossover	0.1		

Table 4

Results of the MLP architecture optimisation through the NSGA-II and a brute-force research.

Method	Variables	Hidden layers	Hidden neurons	CV(RMSE) [%]	Complexity	Evaluations
NSGA-II	Temperature	2	30–20	6.83	0.23	1999
	Relative humidity	2	110–10	8.31	0.32	1099
	CO ₂	1	20	11.28	0.11	999
Brute-force search	Temperature	1	10	7.34	0.10	7220
	Relative humidity	1	110	8.25	0.22	7220
	CO ₂	1	20	12.78	0.11	7220

point on the Pareto front depends on the objective of the study: in this case, the priority is to maximise the accuracy of the model rather than minimise the computational time needed to train it. For simplification and considering that the aim of the study is to effectively interpolate data at different positions and times, the evaluation is focused on three samples separated in time, where the mobile device was in three different positions. Therefore, the algorithm decides according to the average error obtained on these samples.

Finally, Fig. 7 presents the Pareto fronts obtained for each of the variables analysed. Due to their different behaviours, the results vary: while the optimisation of the network structure for temperatures needed 1999 evaluations to stop, for relative humidity and CO₂ levels, it only needed 1099 and 999, respectively (see Table 4). The selected optimal points on the Pareto fronts are shown in Fig. 7, and the MLP hidden layers architecture they represent is presented in Table 4. In addition, the performance of NSGA-II optimisation is compared with a traditional brute-force search to find the most suitable MLP architecture. Considering a selection process analogous to that used with NSGA-II (but using a weighted sum function to select the optimal point since, in this case, there is no Pareto front), Table 4 shows that NSGA-II is more efficient for this type of search. The results in terms of the objective functions are similar, but the evaluations needed to achieve them and therefore the computation time required are much higher. The reason is that the time taken by NSGA-II to choose or discard options is negligible. Specifically, in the case of temperatures, the brute-force search, with 5221 more evaluations, finds a simpler optimal architecture but with a higher average error (from 6.825 to 7.337). In the case of relative humidity, the brute-force search reached a simpler solution with an average error in the same range (from 8.307 to 8.246). With respect to CO₂ concentration, the brute-force search selects an optimal architecture with the same complexity and with a slightly higher average error (from 11.278 to 12.780) due to the randomness present in the process.

4.2. MLP performance

Once the optimal structures are known (see Table 4), two different analyses are carried out: positional cross-validation and accuracy analysis focused on specific days. First, the average results obtained by interpolating data at each of the studied positions (after having removed those data from the training sample) are shown in Table 5. These results demonstrate the effectiveness of the optimised MLPs for both spatial and temporal interpolation.

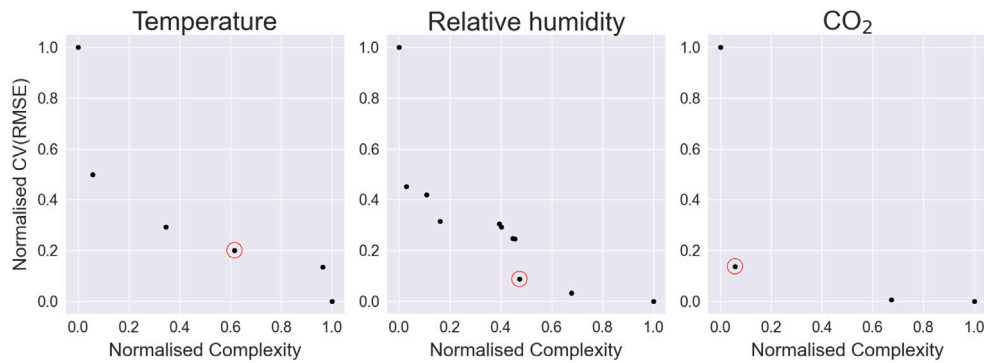


Fig. 7. Pareto fronts with the optimum point selected in the multiobjective problem related to the MLP structure optimisation. The two normalised functions taken into account are the CV(RMSE) and a complexity function.

Table 5

Numerical results of the interpolations of the three variables selected to study the environmental conditions in the analysed building. The average RMSE, CV(RMSE) and NMBE obtained through a positional cross-validation process are presented with their standard deviations (SDs).

Variables	MLP architecture	RMSE	SD	CV(RMSE) [%]	SD	NMBE [%]	SD
Temperature [°C]	45-30-20-1	0.92	0.39	5.90	2.50	−1.38	3.70
Relative humidity [%]	45-110-10-1	2.89	1.37	4.91	2.33	−0.01	3.56
CO ₂ [ppm]	43-20-1	59.80	20.89	12.16	4.25	−0.02	8.46

In the case of temperature, whose model has 45 inputs, the average CV(RMSE) obtained throughout the entire sample is 5.90% with a standard deviation of ± 2.50 (see Table 5). Considering its absolute value (RMSE), on average, the error produced by the model is 0.92 °C, varying by ± 0.39 . Moreover, considering the NMBE results, the average error is -1.38% with a standard deviation of ± 3.70 . This means that, on the one hand, the model presents good accuracy, as the RMSE it yielded is not far for the sensor error (± 0.3 °C). On the other hand, as presented in Table 5, this model makes interpolations, on average, slightly above the real values due to the negative NMBE. Regarding the relative humidity results, where the built MLP needs 45 inputs, the average CV(RMSE) obtained by analysing all the positions considered in this analysis is 4.91% with a standard deviation of ± 2.33 (see Table 5). Considering absolute units, the average error is 2.89% with a variability of ± 1.37 . In addition, the average NMBE obtained is -0.01% , and its standard deviation is ± 3.56 . In this case, the model built to interpolate relative humidity, as for indoor temperatures, is accurate, showing an average RMSE close to the error sensor ($\pm 2\%$). Furthermore, the small and negative NMBE means that this model produces interpolations that, on average, are very close to the monitored values but a minimum above. Last, Table 5 also presents the results of the MLP neural network focused on CO₂ levels, which requires 43 inputs. Thus, the average CV(RMSE) and the RMSE yielded in this case are 12.16% and 59.80 ppm with standard deviations of ± 4.25 and ± 22.02 , respectively. With respect to the NMBE results, this MLP exhibits, on average, an NMBE of -0.02% and a standard deviation of ± 8.46 (see Table 5). As with the other variables, the average RMSE obtained by analysing CO₂ levels is very close to the sensor error (± 50 ppm), and the NMBE results show that the interpolations of this model, on average, are also very close to the monitored data but slightly above due to its negative value.

By analysing all of the models according to their relative results (see Table 5), it can be observed that the MLP focused on the interpolation of relative humidity is the most accurate. It presents the lowest average CV(RMSE) and NMBE as well as the least variable results. By contrast, the least accurate model is the MLP built to interpolate CO₂ levels. It shows a higher average CV(RMSE) with a higher variability in its results, and although the average NMBE obtained is not the highest, its variability is much higher than that presented by the other models. Nevertheless, all of the built MLP neural networks show a high

accuracy, with the obtained values being slightly higher than the real values. This is especially positive considering the sensor errors and the fact that the interpolation method allows missing data in two of the fixed devices. Consequently, considering the distance of the average absolute errors to the sensor error, the CO₂ model becomes the most accurate model due to its RMSE of 59.80 ppm compared with the sensor error of ± 50 ppm. However, both relative humidity and indoor temperature MLPs, despite having lower relative errors, perform worse in this case. They present RMSE values of 2.89% and 0.92 °C with sensor errors of $\pm 2\%$ and ± 0.3 °C, respectively. These facts indicate that the highest relative errors in the CO₂ interpolations are mainly caused by a higher variability in the monitored data (see Table 5) but also by a higher sensor error.

On the other hand, with the aim of showing the MLP performance interpolating data over an entire day, Fig. 8 presents the results of the three built MLPs interpolating data on three specific days. Two of them represent typical conditions, while the third, which is different for each variable, is specifically selected to show a particular behaviour. This analysis showcases the accuracy of the optimised MLP neural networks by interpolating data temporally at known positions for the model (where the mobile device was located at some point in time). Additionally, the CV(RMSE) obtained by the models on each of the days and for each of the variables is presented in Fig. 8. It can be seen that for every day analysed, the built models correctly replicate the daily behaviour of the three monitored variables. Specifically, as presented in Table 5, temperature and relative humidity interpolations yield better results (CV(RMSE) below 5%) than in the case of CO₂ levels (CV(RMSE) below 9%). Fig. 8 also shows that the evolution of the CO₂ levels during the day is more variable, a fact that significantly influences the higher errors obtained by the model. In addition, further evidence that the built MLP neural networks work efficiently is that on atypical days, the errors, although increasing, remain at acceptable levels (see Fig. 8). Furthermore, it can be observed that in comparison with the results in Table 5, the errors obtained in this case are significantly lower. This is because in these interpolations, the model does have information about the specific positions analysed (the mobile device is often in the same position for more than one day).

In short, the results presented in this section demonstrate the effectiveness of the built MLPs interpolating data temporally and spatially in a large multi-zone room. They obtained average NMBE values close to zero and small average RMSE values considering the error shown by the sensors used in this study. Compared to previous studies analysing a simpler space, the errors obtained with the presented methodology are smaller or remain in the same range [24,70,71].

4.3. Building performance analysis

The main objective of the developed spatiotemporal interpolation methodology is the identification of possible uncomfortable or even

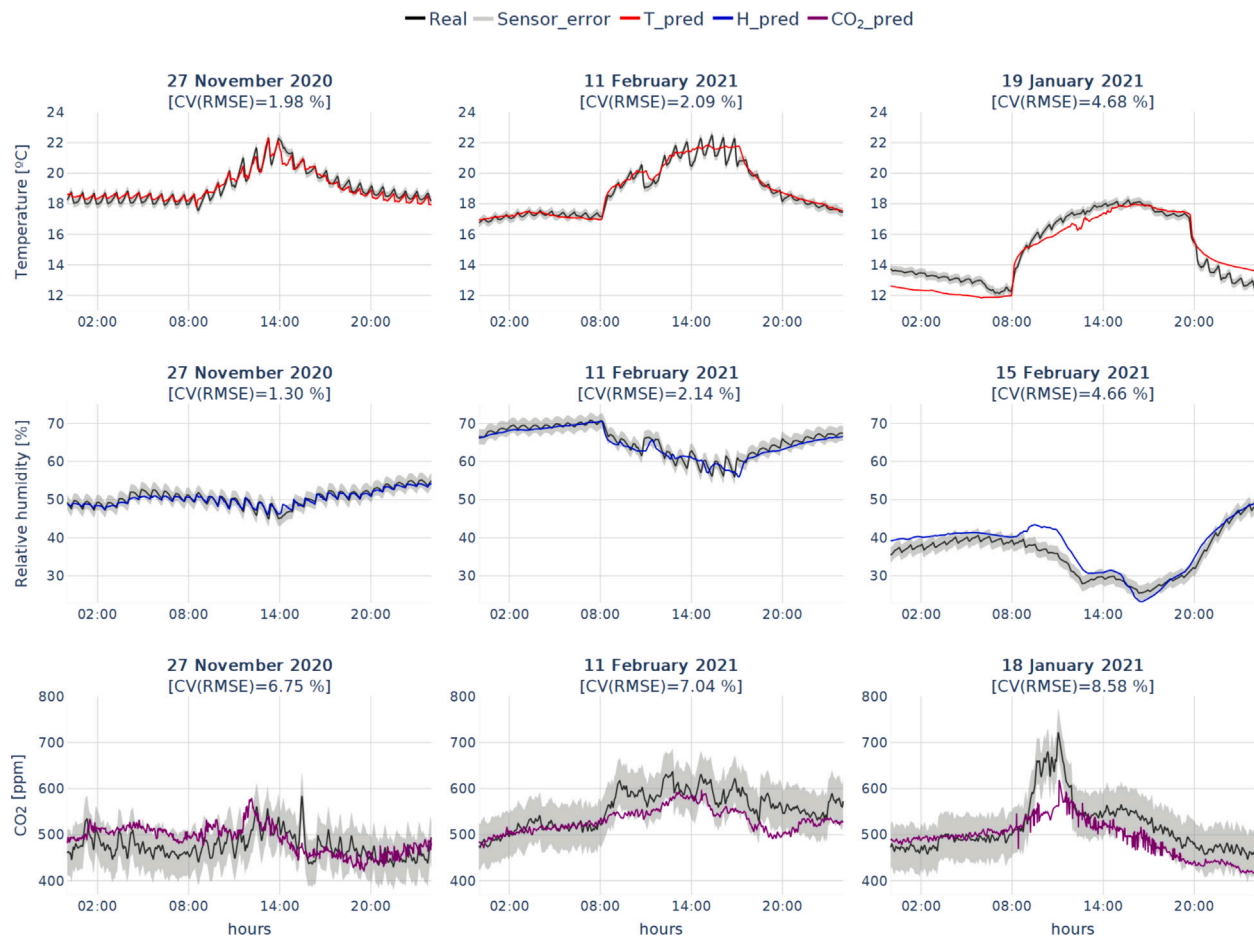


Fig. 8. Results of the interpolations on specific days and thus in positions where the mobile device was at some stage, together with the errors yielded. Indoor temperatures are in the first row, relative humidity is in the second row, and CO₂ levels are in the third row.

unhealthy situations within a building. This is possible because interpolation provides real-time monitoring data for the entire building and not just for the individual positions where the measuring equipment is installed. This section presents the interpolated data obtained for temperature, relative humidity and CO₂ concentration in the study area and at a height of 1.5 m above the floor. Fig. 9 shows, as an example, the spatial distribution of these variables at three points in time corresponding to February 26, 2021, with a time difference of 30 min. In this case, a door near device D3 was opened in the time interval presented in the figure, which caused the conditions inside the building to change slightly. This behaviour is correctly captured by the model, influencing all points in the entire space to different degrees.

A detailed analysis of the temperature distribution shows a high degree of consistency with the building envelope. The areas with higher temperatures are areas where the envelope has a large amount of glazing that magnifies the influence of solar radiation, leading to an increase in temperature compared to more shaded areas. As previously mentioned, the envelope is glazed on the shorter sides of the study area as well as on a section of the roof. The comparison of the temporal results identifies how the temperature in the area around device D3 decreases as the time elapses after the door is opened. However, there are no major variations on the opposite side of the room.

Regarding relative humidity, although it is modelled independently from temperature using its own MLP, it is closely related to the temperature results. In the warmer areas, it can be observed that the relative humidity is slightly lower than that in the cooler areas. This is a consistent behaviour considering that without additional water, the air becomes drier as the temperature increases and more humid

as the temperature decreases. As with temperature, relative humidity increases with the intake of air from outside when the door is opened.

The CO₂ concentration results show a significant relationship with the building occupancy and building activity. Experimental boiler tests are carried out around the D3 device. Although the tests are carried out under controlled conditions and the installation is properly constructed, it is possible that at certain times the CO₂ levels may increase slightly compared to the rest of the building, highlighting one of the main reasons why it is interesting to monitor the entire building. It is precisely in this area that the effect of opening the door can be noticed, as the CO₂ concentration is reduced. Close to device D5 is where the highest levels are registered. This is an area with higher occupancy than the rest of the building.

In addition to detecting occasional problems, distribution maps are also useful for analysing the overall performance of buildings. Fig. 9 shows that the indoor temperature is not homogeneous at the instants analysed, but there is a variation of up to 4.5 °C between some areas. This may be partly due to the building envelope and the large size of the study area but also to the operation of the heating system. A more in-depth analysis of temperature values over a longer time span could provide information about deficiencies in heating and cooling systems. In contrast, the relative humidity does not show such a large variation in this period. It remains at approximately 10% between the most and less humid areas of the space. The difference in CO₂ concentration inside the building is up to 200 ppm, but at all monitored times, the levels remain within the recommended ranges. Another fact that can be deduced from the behaviour of the three variables in the analysed interval is that opening a small door is not sufficient to ensure the ventilation of such a large space.

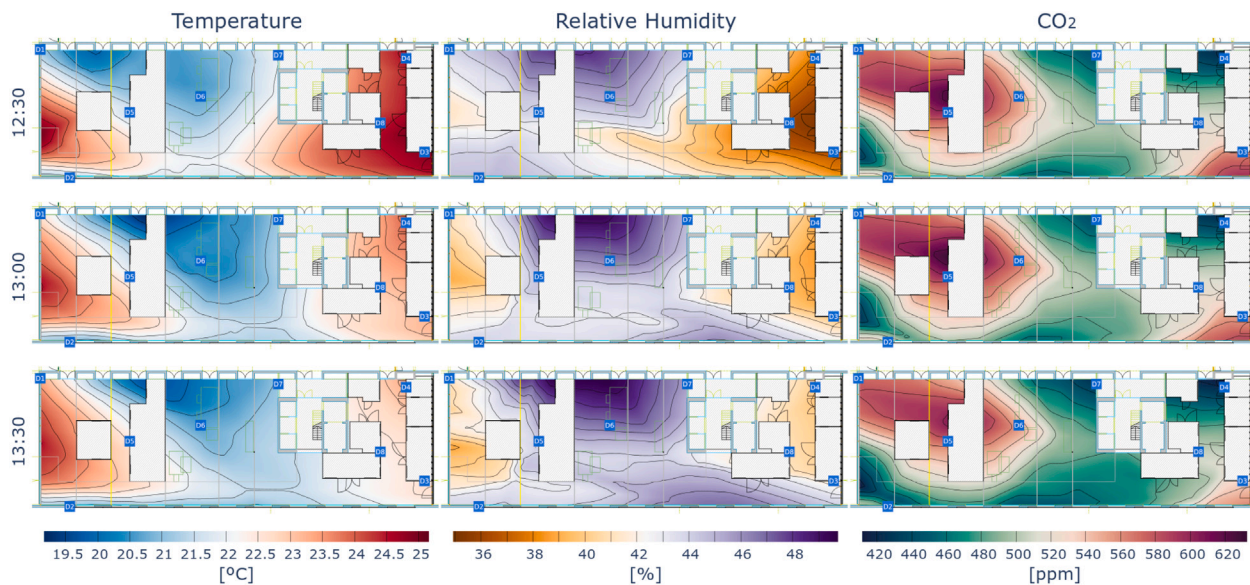


Fig. 9. Distribution map of temperature, relative humidity and CO₂ concentration in study area at three points in time on February 26, 2021, at height of 1.5 m above floor.

5. Conclusions

A new methodology to estimate the indoor environmental conditions in a large open space of an in-use building through artificial intelligence is presented in this paper. In this way, a picture of the distribution of these variables throughout the analysed space can be visualised every minute even for instants when not all wall-mounted devices are sending data (the model was trained supporting a maximum of two malfunctioning devices). The specific models used in this study were three MLP neural networks, one for each of the variables considered, which were responsible for estimating the IEQ values at any location in the building.

The results achieved show that based on a low-cost monitoring system, it is possible to efficiently estimate the indoor environmental conditions of a building by interpolating indoor temperatures, relative humidity and CO₂ levels with machine learning. The MLP neural network architectures were optimised using the NSGA-II algorithm to select the best structure considering two objective functions (in this case, prediction error and model complexity). In this way, positional cross validations were carried out to demonstrate the efficiency of the built models. The results for indoor temperature interpolations show that the built model is accurate since it obtained an average CV(RMSE) of 5.90% and ASHRAE, with hourly data, recommends an average CV(RMSE) of less than $\pm 30\%$ to consider a model calibrated. With regard to relative humidity, the average RMSE obtained of 2.89% in comparison with the sensor error of $\pm 2\%$, demonstrates that the built MLP is also accurate. Last, in the case of CO₂ levels, considering a sensor error of 50 ppm, the average RMSE of 59.80 ppm shows the good performance of the model. From a thermal comfort and air quality point of view, the conclusion is that the methodology presented in this paper is effective in estimating the IEQ in a large open building space.

The main limitation of this research was the amount of data. A more extensive data sample would not only provide data from the different seasons of the year but also from more positions of the mobile device. Future lines of research would be to evaluate if the number of devices can be further optimised without affecting the quality of the results, the study of other algorithms or additional inputs, such as building design parameters, that improve the performance of the presented model or the possible contribution of more accurate sensors.

The outcomes of this research are evidence that the developed methodology can contribute to improving the monitoring and control of the environmental conditions inside a building. On the one hand, the

possibility of analysing, by means of distribution maps, the variables of interest in the whole space makes it possible to identify locations where conditions are not comfortable or even where workers' health is at risk. Moreover, this contributes to the assessment of building performance, identifying deficiencies in the heating and cooling systems and even in the building envelope. This is beneficial not only for the health and well-being of workers but also for their productivity. Furthermore, this methodology also contributes with advantages over the existing models of estimating the IEQ of a building. Simple continuous monitoring of data is very costly, and simulation models require a great deal of expertise and handle many different parameters. Other methods based on artificial intelligence only use fixed devices, need very expensive equipment, or are focused on small spaces. With the system presented in this paper, which is built based on inexpensive assemblies and sensors and is faster than traditional simulations, only the movement of the mobile device to the positions where the data are most significant is needed. Thus, after the model is trained, no further human intervention is required. In addition, it does not require the monitoring or previous estimation of variables such as the occupancy rate, air exchange rate or others similar variables which are more costly and complex to monitor. On the other hand, the fact that the case of study is a large multizone open space enables this methodology to be extrapolated to similar buildings and, in principle, to smaller rooms, since the problem difficulty would be lower. Likewise, this system allows other variables related to the IEQ of a building to be easily considered, increasing the possible extrapolation of this method. Lastly, the usefulness of the NSGA-II to adjust the hyperparameters of a machine learning model efficiently and quickly has been demonstrated.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements and funding

This research was partially supported by the Ministry of Science, Innovation and Universities of the Spanish government under the SMARTHerm project (RTI2018-096296-B-C2). The authors also want to thank the Ministry of Science, Innovation and Universities, Spain (grants FPU17/01834 and FPU19/01187) and the University of Vigo, Spain (grant 00VI 131H 641.02). Funding for open access charge: Universidade de Vigo/CISUG, Spain.

References

- ## References
- [1] R. American Society of Heating, A.-C. Engineers, ASHRAE Guideline 10-2011, Interactions Affecting the Achievement of Acceptable Indoor Environments, ASHRAE, 2011.
 - [2] M. Frontczak, P. Wargocki, Literature survey on how different factors influence human comfort in indoor environments, *Build. Environ.* 46 (4) (2011) 922–937, <http://dx.doi.org/10.1016/j.buildenv.2010.10.021>, URL: <https://www.sciencedirect.com/science/article/pii/S0360132310003136>.
 - [3] M.S. Andargie, E. Azar, An applied framework to evaluate the impact of indoor office environmental factors on occupants' comfort and working conditions, *Sustainable Cities Soc.* 46 (2019) 101447, <http://dx.doi.org/10.1016/j.scs.2019.101447>, URL: <https://www.sciencedirect.com/science/article/pii/S2210670718307674>.
 - [4] M. Economidou, B. Atanasiu, C. Despret, J. Maio, I. Nolte, O. Rapf, J. Laustsen, P. Ruysssevelt, D. Staniaszek, D. Strong, S. Zinetti, Europe's buildings under the microscope. A country-by-country review of the energy performance of buildings, 2011.
 - [5] A. Steinemann, P. Wargocki, B. Rismanchi, Ten questions concerning green buildings and indoor air quality, *Build. Environ.* 112 (2017) 351–358, <http://dx.doi.org/10.1016/j.buildenv.2016.11.010>, URL: <https://www.sciencedirect.com/science/article/pii/S0360132316304346>.
 - [6] R. Elnaklah, I. Walker, S. Natarajan, Moving to a green building: Indoor environment quality, thermal comfort and health, *Build. Environ.* 191 (2021) 107592, <http://dx.doi.org/10.1016/j.buildenv.2021.107592>, URL: <https://www.sciencedirect.com/science/article/pii/S0360132321000081>.
 - [7] J. Kim, T. Hong, J. Jeong, C. Koo, M. Kong, An integrated psychological response score of the occupants based on their activities and the indoor environmental quality condition changes, *Build. Environ.* 123 (2017) 66–77, <http://dx.doi.org/10.1016/j.buildenv.2017.06.046>, URL: <https://www.sciencedirect.com/science/article/pii/S0360132317302810>.
 - [8] J. Park, V. Loftness, A. Aziz, T.-H. Wang, Critical factors and thresholds for user satisfaction on air quality in office environments, *Build. Environ.* 164 (2019) 106310, <http://dx.doi.org/10.1016/j.buildenv.2019.106310>, URL: <https://www.sciencedirect.com/science/article/pii/S0360132319305207>.
 - [9] N. Ma, D. Aviv, H. Guo, W.W. Braham, Measuring the right factors: A review of variables and models for thermal comfort and indoor air quality, *Renew. Sustain. Energy Rev.* 135 (2021) 110436, <http://dx.doi.org/10.1016/j.rser.2020.110436>, URL: <https://www.sciencedirect.com/science/article/pii/S1364032120307231>.
 - [10] U. Haverinen-Shaughnessy, R.J. Shaughnessy, E.C. Cole, O. Toyinbo, D.J. Moschandreass, An assessment of indoor environmental quality in schools and its association with health and performance, *Build. Environ.* 93 (2015) 35–40, <http://dx.doi.org/10.1016/j.buildenv.2015.03.006>, URL: <https://www.sciencedirect.com/science/article/pii/S0360132315001092>, Special Issue: Indoor pollutants, chemistry and health- Selected papers presented at Indoor Air 2014 conference in Hong Kong.
 - [11] S. Vilcekova, L. Meciarova, E.K. Burdova, J. Katunska, D. Kosicanova, S. Doroudiani, Indoor environmental quality of classrooms and occupants' comfort in a special education school in Slovak Republic, *Build. Environ.* 120 (2017) 29–40, <http://dx.doi.org/10.1016/j.buildenv.2017.05.001>, URL: <https://www.sciencedirect.com/science/article/pii/S0360132317301749>.
 - [12] S. Zuhair, R. Mantou, C. Griffin, M. Hajdukiewicz, M.M. Keane, J. Goggins, An Indoor Environmental Quality (IEQ) assessment of a partially-retrofitted university building, *Build. Environ.* 139 (2018) 69–85, <http://dx.doi.org/10.1016/j.buildenv.2018.05.001>, URL: <https://www.sciencedirect.com/science/article/pii/S0360132318302610>.
 - [13] M. Karami, G.V. McMorro, L. Wang, Continuous monitoring of indoor environmental quality using an Arduino-based data acquisition system, *J. Build. Eng.* 19 (2018) 412–419, <http://dx.doi.org/10.1016/j.jobe.2018.05.014>, URL: <https://www.sciencedirect.com/science/article/pii/S2352710218301025>.
 - [14] C. Giraldo-Soto, A. Ekoreka, A. Barragan, L. Mora, Dataset of an in-use tertiary building collected from a detailed 3D mobile monitoring system and building automation system for indoor and outdoor air temperature analysis, *Data Brief* 31 (2020) <http://dx.doi.org/10.1016/j.dib.2020.105907>, URL: <https://www.sciencedirect.com/science/article/pii/S2352710218301025>.
 - [15] M. Jin, S. Liu, S. Schiavon, C. Spanos, Automated mobile sensing: Towards high-granularity agile indoor environmental quality monitoring, *Build. Environ.* 127 (2018) 268–276, <http://dx.doi.org/10.1016/j.buildenv.2017.11.003>, URL: <https://www.sciencedirect.com/science/article/pii/S0360132317305012>.
 - [16] H. Shen, W. Hou, Y. Zhu, S. Zheng, S. Ainiwaer, G. Shen, Y. Chen, H. Cheng, J. Hu, Y. Wan, S. Tao, Temporal and spatial variation of PM2.5 in indoor air monitored by low-cost sensors, *Sci. Total Environ.* 770 (2021) 145304, <http://dx.doi.org/10.1016/j.scitotenv.2021.145304>, URL: <https://www.sciencedirect.com/science/article/pii/S0048969721003703>.
 - [17] J. Li, H. Li, Y. Ma, Y. Wang, A. Abokifa, C. Lu, P. Biswas, Spatiotemporal distribution of indoor particulate matter concentration with a low-cost sensor network, *Build. Environ.* 127 (2017) <http://dx.doi.org/10.1016/j.buildenv.2017.11.001>.
 - [18] Y. Yang, J. Liu, W. Wang, Y. Cao, H. Li, Incorporating SLAM and mobile sensing for indoor CO2 monitoring and source position estimation, *J. Cleaner Prod.* 291 (2021) 125780, <http://dx.doi.org/10.1016/j.jclepro.2020.125780>, URL: <https://www.sciencedirect.com/science/article/pii/S0959652620358261>.
 - [19] Z. Han, R.X. Gao, Z. Fan, Occupancy and indoor environment quality sensing for smart buildings, in: 2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings, 2012, pp. 882–887, <http://dx.doi.org/10.1109/I2MTC.2012.6229557>.
 - [20] B. Dong, V. Prakash, F. Feng, Z. O'Neill, A review of smart building sensing system for better indoor environment control, *Energy Build.* 199 (2019) 29–46, <http://dx.doi.org/10.1016/j.enbuild.2019.06.025>, URL: <https://www.sciencedirect.com/science/article/pii/S0378778819309302>.
 - [21] K. Li, H. Su, J. Chu, C. Xu, A fast-POD model for simulation and control of indoor thermal environment of buildings, *Build. Environ.* 60 (2013) 150–157, <http://dx.doi.org/10.1016/j.buildenv.2012.11.020>, URL: <https://www.sciencedirect.com/science/article/pii/S0360132312003186>.
 - [22] G.A. Ganesh, S.L. Sinha, T.N. Verma, Numerical simulation for optimization of the indoor environment of an occupied office building using double-panel and ventilation radiator, *J. Build. Eng.* 29 (2020) 101139,

- [37] H. Moayedi, M. Gör, Z. Lyu, D.T. Bui, Herding behaviors of grasshopper and harris hawk for hybridizing the neural network in predicting the soil compression coefficient, *Measurement* 152 (2020) 107389, <http://dx.doi.org/10.1016/j.measurement.2019.107389>, URL: <http://www.sciencedirect.com/science/article/pii/S0263224119312564>.
- [38] G. Zhou, H. Moayedi, M. Bahiraei, Z. Lyu, Employing artificial bee colony and particle swarm techniques for optimizing a neural network in prediction of heating and cooling loads of residential buildings, *J. Cleaner Prod.* 254 (2020) 120082, <http://dx.doi.org/10.1016/j.jclepro.2020.120082>, URL: <http://www.sciencedirect.com/science/article/pii/S0959652620301293>.
- [39] E.T. Mohamad, R.S. Faradonbeh, D.J. Armaghani, M. Monjezi, M.Z.A. Majid, An optimized ANN model based on genetic algorithm for predicting ripping production, *Neural Comput. Appl.* 28 (1) (2017) 393–406, <http://dx.doi.org/10.1007/s00521-016-2359-8>, URL: <https://doi.org/10.1007/s00521-016-2359-8>.
- [40] S. Nazari, M. Bahiraei, H. Moayedi, H. Safarzadeh, A proper model to predict energy efficiency, exergy efficiency, and water productivity of a solar still via optimized neural network, *J. Cleaner Prod.* 277 (2020) 123232, <http://dx.doi.org/10.1016/j.jclepro.2020.123232>, URL: <http://www.sciencedirect.com/science/article/pii/S0959652620332777>.
- [41] M.A. Janati Idrissi, H. Ramchoun, Y. Ghanou, M. Ettaouil, Genetic algorithm for neural network architecture optimization, 2016, pp. 1–4, <http://dx.doi.org/10.1109/GOL.2016.7731699>.
- [42] M. Khandelwal, A. Marto, S.A. Fatemi, M. Ghorogi, D.J. Armaghani, T.N. Singh, O. Tabrizi, Implementing an ANN model optimized by genetic algorithm for estimating cohesion of limestone samples, *Eng. Comput.* 34 (2) (2018) 307–317, <http://dx.doi.org/10.1007/s00366-017-0541-y>.
- [43] M. Aboutalebi, O. Bozorg-Haddad, H. Loaiciga, Optimal monthly reservoir operation rules for hydropower generation derived with SVR-NSGAI, *J. Water Resour. Plann. Manage.* 141 (2015) [http://dx.doi.org/10.1061/\(ASCE\)WR.1943-5452.0000553](http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000553).
- [44] D.S. Yeung, J. Li, W.W.Y. Ng, P.P.K. Chan, Mlpnn training via a multiobjective optimization of training error and stochastic sensitivity, *IEEE Trans. Neural Netw. Learn. Syst.* 27 (5) (2016) 978–992, <http://dx.doi.org/10.1109/TNNLS.2015.2431251>.
- [45] J. Blank, K. Deb, Pymoo: Multi-objective optimization in python, *IEEE Access* 8 (2020) 89497–89509, <http://dx.doi.org/10.1109/access.2020.2990567>, URL: <http://dx.doi.org/10.1109/ACCESS.2020.2990567>.
- [46] M. Hemmat Esfe, P. Razi, M.H. Hajmohammad, S.H. Rostamian, W.S. Sarsam, A.A. Abbasian Arani, M. Dahari, Optimization, modeling and accurate prediction of thermal conductivity and dynamic viscosity of stabilized ethylene glycol and water mixture al₂O₃ nanofluids by NSGA-II using ANN, *Int. Commun. Heat Mass Transfer* 82 (2017) 154–160, <http://dx.doi.org/10.1016/j.icheatmasstransfer.2016.08.015>, URL: <https://www.sciencedirect.com/science/article/pii/S0735193316302421>.
- [47] S. Martínez, P. Eguía, E. Granada, A. Moazami, M. Hamdy, A performance comparison of multi-objective optimization-based approaches for calibrating white-box building energy models, *Energy Build.* 216 (2020) 109942, <http://dx.doi.org/10.1016/j.enbuild.2020.109942>, URL: <http://www.sciencedirect.com/science/article/pii/S0378778819336850>.
- [48] A. Jahani, L.M. Khanli, M.T. Hagh, M.A. Badamchizadeh, Ee-CTA: Energy efficient, concurrent and topology-aware virtual network embedding as a multi-objective optimization problem, *Comput. Stand. Interfaces* 66 (2019) 103351, <http://dx.doi.org/10.1016/j.csi.2019.04.010>, URL: <http://www.sciencedirect.com/science/article/pii/S0920548918302708>.
- [49] G.R. Ruiz, C.F. Bandera, Validation of calibrated energy models: Common errors, *Energies* 10 (10) (2017) <http://dx.doi.org/10.3390/en10101587>, URL: <https://www.mdpi.com/1996-1073/10/10/1587>.
- [50] M. Bahiraei, S. Nazari, H. Moayedi, H. Safarzadeh, Using neural network optimized by imperialist competition method and genetic algorithm to predict water productivity of a nanofluid-based solar still equipped with thermoelectric modules, *Powder Technol.* 366 (2020) 571–586, <http://dx.doi.org/10.1016/j.powtec.2020.02.055>, URL: <http://www.sciencedirect.com/science/article/pii/S0032591020301595>.
- [51] N.I.M. Yusoff, D. Ibrahim Alhamali, A.N.H. Ibrahim, S.A.P. Rosyidi, N. Abdul Hassan, Engineering characteristics of nanosilica/polymer-modified bitumen and predicting their rheological properties using multilayer perceptron neural network model, *Constr. Build. Mater.* 204 (2019) 781–799, <http://dx.doi.org/10.1016/j.conbuildmat.2019.01.203>, URL: <http://www.sciencedirect.com/science/article/pii/S0950061819302399>.
- [52] D. Tien Bui, V.-H. Nhu, N.-D. Hoang, Prediction of soil compression coefficient for urban housing project using novel integration machine learning approach of swarm intelligence and multi-layer perceptron neural network, *Adv. Eng. Inform.* 38 (2018) 593–604, <http://dx.doi.org/10.1016/j.aei.2018.09.005>, URL: <http://www.sciencedirect.com/science/article/pii/S1474034618302246>.
- [53] Y. Liu, S. Liu, Y. Wang, F. Lombardi, J. Han, A stochastic computational multi-layer perceptron with backward propagation, *IEEE Trans. Comput.* 67 (9) (2018) 1273–1286, <http://dx.doi.org/10.1109/TC.2018.2817237>.
- [54] L.N. Smith, A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay, 2018, *ArXiv abs/1803.09820*.
- [55] C.O. Sakar, S.O. Polat, M. Katircioglu, Y. Kastro, Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks, *Neural Comput. Appl.* 31 (10) (2019) 6893–6908, <http://dx.doi.org/10.1007/s00521-018-3523-0>.
- [56] M.W. Khan, M. Zeeshan, M. Usman, Traffic scheduling optimization in cognitive radio based smart grid network using mini-batch gradient descent method, in: 2019 14th Iberian Conference on Information Systems and Technologies (CISTI), 2019, pp. 1–5, <http://dx.doi.org/10.23919/CISTI.2019.8760693>.
- [57] M. Li, M. Soltanolkotabi, S. Oymak, Gradient descent with early stopping is provably robust to label noise for overparameterized neural networks, in: S. Chiappa, R. Calandra (Eds.), in: *Proceedings of Machine Learning Research*, vol. 108, PMLR, Online, 2020, pp. 4313–4324, URL: <http://proceedings.mlr.press/v108/li20j.html>.
- [58] K. Eckle, J. Schmidt-Hieber, A comparison of deep networks with ReLU activation function and linear spline-type methods, *Neural Netw.* 110 (2019) 232–242, <http://dx.doi.org/10.1016/j.neunet.2018.11.005>, URL: <http://www.sciencedirect.com/science/article/pii/S0893608018303277>.
- [59] S. Bock, M. Weiß, A Proof of Local Convergence for the Adam Optimizer, 2019, pp. 1–8, <http://dx.doi.org/10.1109/IJCNN.2019.8852239>.
- [60] T. Vo-Duy, D. Duong-Gia, V. Ho-Huu, H. Vu-Do, T. Nguyen-Thoi, Multi-objective optimization of laminated composite beam structures using NSGA-II algorithm, *Compos. Struct.* 168 (2017) 498–509, <http://dx.doi.org/10.1016/j.compstruct.2017.02.038>, URL: <http://www.sciencedirect.com/science/article/pii/S0263822316323728>.
- [61] X.-S. Yang, Chapter 5 - genetic algorithms, in: X.-S. Yang (Ed.), *Nature-Inspired Optimization Algorithms*, Elsevier, Oxford, 2014, pp. 77–87, <http://dx.doi.org/10.1016/B978-0-12-416743-8.00005-1>, URL: <http://www.sciencedirect.com/science/article/pii/B9780124167438000051>.
- [62] N. Namura, K. Shimoyama, S. Obayashi, Expected improvement of penalty-based boundary intersection for expensive multiobjective optimization, *IEEE Trans. Evol. Comput.* 21 (6) (2017) 898–913, <http://dx.doi.org/10.1109/TEVC.2017.2693320>.
- [63] J. New, Suitability of ASHRAE guideline 14 metrics for calibration, ASHRAE Trans. (2016) URL: <https://www.osti.gov/servlets/purl/1266021>.
- [64] M. Kottek, J. Grieser, C. Beck, B. Rudolf, F. Rubel, World Map of the Köppen-Geiger climate classification updated, *Meteorol. Z.* 15 (3) (2006) 259–263, <http://dx.doi.org/10.1127/0941-2948/2006/0130>.
- [65] X. de Galicia, Meteogalicia, 2021, URL: <https://www.meteogalicia.gal/web/RSS/rssIndex.action>, (accessed on 10 March 2021).
- [66] NOAA, NOAA Operational model archive and distribution system (NOMADS), 2021, URL: <https://nomads.ncep.noaa.gov/>, (accessed on 10 March 2021).
- [67] NOAA, The global forecast system (GFS) - global spectral model (GSM), 2021, URL: https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/documentation.php, (accessed on 10 March 2021).
- [68] J. López Gómez, F. Troncoso Pastoriza, E.A. Fariña, P. Eguía Oller, E. Granada Álvarez, Use of a numerical weather prediction model as a meteorological source for the estimation of heating demand in building thermal simulations, *Sustainable Cities Soc.* 62 (2020) 102403, <http://dx.doi.org/10.1016/j.scs.2020.102403>, URL: <https://www.sciencedirect.com/science/article/pii/S2210670720306247>.
- [69] C.N. Babu, B.E. Reddy, A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data, *Appl. Soft Comput.* 23 (2014) 27–38, <http://dx.doi.org/10.1016/j.asoc.2014.05.028>, URL: <https://www.sciencedirect.com/science/article/pii/S1568494614002555>.
- [70] A. Costantino, E. Fabrizio, A. Ghigini, M. Bariani, Climate control in broiler houses: A thermal model for the calculation of the energy use and indoor environmental conditions, *Energy Build.* 169 (2018) 110–126, <http://dx.doi.org/10.1016/j.enbuild.2018.03.056>, URL: <https://www.sciencedirect.com/science/article/pii/S0378778817338598>.
- [71] A. Pantazaras, S.E. Lee, M. Santamouris, J. Yang, Predicting the CO₂ levels in buildings using deterministic and identified models, *Energy Build.* 127 (2016) 774–785, <http://dx.doi.org/10.1016/j.enbuild.2016.06.029>, URL: <https://www.sciencedirect.com/science/article/pii/S0378778816305187>.
- [72] A. Franco, F. Leccese, Measurement of CO₂ concentration for occupancy estimation in educational buildings with energy efficiency purposes, *J. Build. Eng.* 32 (2020) 101714, <http://dx.doi.org/10.1016/j.jobe.2020.101714>.
- [73] P. Tien, S. Wei, J.K. Calautit, J. Darkwa, C. Wood, A vision-based deep learning approach for the detection and prediction of occupancy heat emissions for demand-driven control solutions, *Energy Build.* 226 (2020) 110386, <http://dx.doi.org/10.1016/j.enbuild.2020.110386>.
- [74] X. Feng, D. Yan, T. Hong, Simulation of occupancy in buildings, *Energy Build.* 87 (2015) 348–359.
- [75] P.K. Sharma, A. Mondal, S. Jaiswal, M. Saha, S. Nandi, T. De, S. Saha, IndoAirSense: A framework for indoor air quality estimation and forecasting, *Atmos. Pollut. Res.* 12 (1) (2021) 10–22.
- [76] G. Van Rossum, F.L. Drake, Python 3 Reference Manual, CreateSpace, Scotts Valley, CA, 2009.