Journal of Ambient Intelligence and Humanized Computing (2023) 14:16749–16770
https://doi.org/10.1007/s12652-023-04685-8

ORIGINAL RESEARCH

A multi-objective reinforcement learning approach for furniture arrangement with optimal IEQ in multi-occupant offices

Patrizia Ribino[1] · **Marina Bonomolo**[2]

Received: 29 December 2021 / Accepted: 28 August 2023 / Published online: 9 September 2023

Abstract
Indoor Environmental Quality (IEQ) concerns several aspects of environmental comforts, such as thermal, visual and acoustics comfort. In particular, IEQ plays a relevant role in workers' satisfaction since it strongly influences health, well-being, and productivity. Specifically, it has been demonstrated that the furniture configuration in working spaces affects the occupant's comfort perception. Nevertheless, IEQ has been either neglected or partially addressed in the context of interior design. The contribution of this paper is to introduce a novel method for furniture layout optimisation in terms of IEQ requirements in multi-occupant offices. In particular, we explore the furniture arrangement task as a Multi-Objective Markov Decision Process (MOMDP), which is solved by a reinforcement learning (RL) agent. The goal is to determine optimal workstation positions that maximise workers' IEQ satisfaction and functional aspects of working spaces under analysis. Firstly, we formulated the furniture layout task as a MOMDP problem by defining reward functions in terms of thermal, acoustics and visual comfort. Then, we train the RL agent to produce optimal/suboptimal layout patterns through a Q-learning-based algorithm. We conducted experiments in two different offices. The experimental results demonstrated that the proposed multi-objective RL approach is able to determine optimal furniture arrangements that provide a balance among office occupants in terms of IEQ satisfaction. Moreover, numerical results show that the proposed approach can be a valuable tool for evaluating the conformity to the environmental comfort standard of working environments during the furniture layout design phase instead of applying corrections during the post-occupancy evaluation.

Keywords Furniture arrangement · Multi-objective reinforcement learning · Layout optimization · Indoor environmental quality

1 Introduction

More than 50% of workers in the world spend most of their time in offices (Vimalanathan and Babu 2014). It has been demonstrated that the quality of the working environment improves user satisfaction since it strongly influences occupants' health, well-being, and productivity (Frontczak and Wargocki 2011; Leaman and Bordass 1999; Colenberg et al. 2020). Improving user satisfaction is beneficial for the employee and the organization, which may increase its financial gains (Seppänen and Fisk 2003). In particular, recent studies show that desk location and office arrangement significantly impact occupant satisfaction since they are related to the environmental comfort perception (Kwon et al. 2019; Sant'Anna et al. 2018). However, in the specific problem of furniture layout arrangement, the indoor environmental quality (IEQ) has been either neglected or partially addressed. Indeed, as reported in Sect. 2, several furniture arrangement approaches focus mainly on functional and aesthetic aspects of an indoor layout. Conversely, indoor environmental quality concerns diverse sub-domains that affect human life, including visual, thermal and acoustics comfort. This study addresses the furniture layout problem, considering IEQ requirements for working environments with multi-occupant office end-use. In particular, in a furniture arrangement problem, a given space must be populated with

✉ Patrizia Ribino
patrizia.ribino@icar.cnr.it

Marina Bonomolo
marina.bonomolo@deim.unipa.it

[1] Istituto di Calcolo e Reti ad Alte prestazioni (ICAR), Consiglio Nazionale delle Ricerche, via Ugo La Malfa 153, 90100 Palermo, Italy

[2] Dipartimento di Ingegneria, Universita degli Studi di Palermo, Viale delle Scienze, 90100 Palermo, Italy

a set of furniture pieces, resulting in an optimal arrangement according to some design rules. This problem implies an approach that systematically enumerates alternative solutions while considering a broad spectrum of criteria. In this paper, we consider the problem of finding furniture arrangement patterns for spaces with office end-use that maximize the occupants' environmental comfort perception. Specifically, since multiple workers may occupy a working space, all occupants' satisfaction with IEQ is evaluated because different furniture arrangements may contribute to different comfort levels. Thus, an optimal office layout should maximize the comfort conditions of all occupants by providing an optimal trade-off.

The main contribution of this paper is a reinforcement learning method for finding optimal furniture arrangement patterns in terms of visual, thermal and acoustics comfort conditions and functional criteria for shared offices. In particular, we propose an approach based on a Multi-Objective Reinforcement Learning (MORL) technique. In so doing, we formulated the furniture layout problem for multi-occupant offices as a Multi-Objective Markov Decision Process (MOMDP) by defining states, actions, and reward functions in terms of IEQ indices (Piasecki et al. 2017) and functional criteria.

An evaluation of the proposed method has been conducted on two offices with two workers in each, whose physical characteristics differ. Experimental results show that the proposed approach is able to provide different office furniture arrangement patterns that maximize the level of satisfaction of all workers. The results of the case study also show that the proposed IEQ-based MORL approach is a helpful tool for evaluating the conformity to the environmental comfort standards during the furniture arrangement design phase instead of applying corrections during the post-occupancy evaluation phase. Indeed, a common method called *Post Occupancy Evaluation* (Choi and Lee 2018; Li et al. 2018) is used for evaluating the comfort conditions of an occupant after (s)he lived in that space and making some successive modifications to reach desired comfort conditions. The proposed approach, allowing the evaluation of the best comfortable conditions obtained in an environment beforehand, overcomes the POE approach by improving the decision process for office furniture arrangement design.

Finally, a comparison with the well-known multi-objective optimization algorithm NSGA-II (Deb et al. 2002) and with common single-objective RL techniques (i.e., DQN, A2C and PPO) (Arulkumaran et al. 2017; Grondman et al. 2012; Schulman et al. 2017) shows that the proposed approach provides better layout configurations and execution times.

The rest of the paper is organized as follows. Section 3 introduces the theoretical background. Section 2 shows relevant literature in the field of furniture arrangement. Sections 4 and 5 present the proposed approach and the experimental evaluation. In Sect. 6, a comparison with optimization approaches is conducted. Finally, in Sect. 7 conclusions and future works are drawn.

2 Related works

There is a considerable research effort in the furniture arrangement field. The majority of them focuses on the layout of furniture objects in an automated or semi-automated fashion. They can be classified into three main categories: procedural, data-driven, and optimization-based methods.

The methods from the first category are fast methods based on procedural layout generation (Akazawa et al. 2006; Germer and Schwarz 2009; Tutenel et al. 2009). They use room semantics and furniture layout rules for positioning furniture objects with respect to already arranged objects. The relationships between objects can be defined explicitly (e.g., defining that the sofa needs to face the TV and that it should be no more than five meters away from it) or implicitly, through the use of furniture features. Some approaches, for example, make scenes by using a semantic database that stores information about parent objects, e.g., a table can be a parent of TV, the backside of a bookshelf should contact a wall, and so on.

Data-driven methods address the problem by using information from existing layouts (Fisher et al. 2012; Zhao et al. 2016; Ma et al. 2016; Henderson et al. 2017; Yamakawa et al. 2017; Li et al. 2019; Wang et al. 2018). These approaches require a set of layout examples to generate new plausible arrangements of objects. Some of them requires a 3D scan of real environments to populate the virtual space with objects, others need the initial scene to augment the layout with additional objects, whilst some others use databases containing several layouts designed by humans.

Optimization-based methods generate realistic furniture arrangements by optimizing a cost function. Merrell et al. (2011) turn furniture layout guidelines into a probabilistic model and suggest sensible room layouts by sampling from the density function, as a user interactively moves furniture in a room. In contrast, Yu et al. (2011) learn the layout rules from 3D scene exemplars. These solutions optimize the layout of a given room with a given set of furniture. In Sanchez et al. (2003) and Kán and Kaufmann (2017), optimization of furniture arrangement based on evolutionary computing was proposed. Such methods utilize genetic algorithms to select and arrange furniture objects in a given room by optimizing the population of interior designs (individuals). They use a set of design guidelines to form a cost function that assesses each individual interior design in terms of design guidelines. The goal of the genetic algorithm is to find a set of furniture objects and their arrangement for a given room that minimizes

the cost function. In Kán and Kaufmann (2018), a method for automatic furniture arrangement, which combines optimization and procedural methods, is proposed by using greedy cost minimization to rapidly explore the space of possible solutions.

In addition to the previous categories, new methods recently exploit reinforcement learning for furniture arrangement. In particular, Wang et al. (2020) propose a reinforcement learning approach based on Monte Carlo tree search methods. The main idea is to construct a sequence of search trees to create a sequence of move actions iteratively. Each search tree corresponds to one move action decision. The final aim is to find a sequence of object movements from an initial layout to a target layout. Similarly, Di and Yu (2021) propose an RL based approach for producing furniture layouts in indoor scenes with optimal position and size. In order to define the reward functions, the authors consider constraints to be applied to object positions, such as moving towards the most right position, preventing the furniture from moving and coinciding with other elements and finally preventing the furniture from moving outside the room.

Although many approaches have been proposed to generate furniture arrangement patterns, aesthetic and functional rules are the key principles used for creating a new furniture layout for a specific room. None of the previous works considers indoor environmental quality as a requirement for the optimization process of furniture arrangement. To the best of our knowledge, only a recent work proposed by Vitsas et al. (2020) addresses the problem by taking into consideration only the illumination as a guideline for optimal furniture arrangement. Thus, the novelty introduced in this paper with respect to state of the art concerns the introduction of IEQ as further criteria for furniture arrangement optimization. Moreover, furniture arrangement optimization is also treated as a multi-objective problem for maximizing the comfort of all occupants.

Finally, the work proposed in this paper is based on a preliminary study presented in Ribino and Bonomolo (2021). A simple Q-learning approach based on IEQ optimization for a single-occupant office has been presented in such a study. The contribution submitted in this paper differs both for the method and the final purpose. The current work defines a multi-objective reinforcement learning approach to find optimal IEQ layout patterns in multi-occupant offices. Moreover, an experimental evaluation is here performed on real case studies.

3 Theoretical background

3.1 Indoor comfort conditions

IEQ is commonly described as the comfort conditions inside a building. It includes access to daylight and views, pleasant acoustic conditions, thermal comfort, and air quality, which is not considered for this paper. Fanger et al. (1970) established the first model of thermal comfort based on the Predicted Mean Vote (PMV). It predicts the mean response of a larger group of people according to the ASHRAE thermal sense scale (Fanger et al. 1970). PMV depends on a combination of environmental and individual variables such as the air temperature, the mean radiant temperature, the metabolic rate, etc. Thermal comfort standards use the PMV model to recommend acceptable thermal comfort conditions (Olesen and Parsons 2002). As concerns visual comfort, it is related to the quantity and quality of light within a given space at a given time. Both too little and too much light can cause visual discomfort. The European standard (UNI 2011) defines lighting requirements for indoor work areas in terms of quantity and quality of illumination, covering, for example, offices, restaurants and hotels. Finally, acoustic comfort is provided by minimizing noise. Buildings are designed to specific noise standards based on their use (e.g., the noise level in a library may differ from noise specifications in a public hall). Noise Ratings (NR) (Beranek et al. 1971) is one of the standards used to evaluate acoustic comfort. Each space has a recommended NR value, which is based upon the intended requirements of the indoor environment.

3.2 Reinforcement learning and multi-objective RL

In an RL approach, an autonomous agent learns to choose the most effective actions for achieving its goals by maximizing its overall rewards (Naeem et al. 2020; Van Otterlo and Wiering 2012). The agent learns the optimal policy by trails during the interactions with the environment, which can be described by a sequence of states, actions, and rewards. This sequential decision process is usually modelled as a Markov decision process (MDP). One of the most popular algorithms is Q-learning (Watkins and Dayan 1992) founded on the Bellman Equation:

$$Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha\left[R + \gamma \cdot \max_{a'} Q(s',a')\right] \quad (1)$$

where R is the reward received when agent moves from the state s to the state s'. The parameters α and γ are the learning rate and the discount factor. The first one determines to what extent newly acquired information overrides old information. The second one determines the relevance of future rewards. When the problem refers to the sequential decision-making problem with multiple objectives, the learning agent needs to simultaneously solve several tasks with different rewards. Such an RL problem is called a Multi-Objective RL (MORL) problem (Liu et al. 2014). In MORL, each objective has its associated reward signal, so the reward is not a scalar value but a vector $R = (r_1, \ldots, r_m)$ where m is the number of objectives. Moreover, when all the objectives

are directly correlated, a single objective can be derived by combining the multiple objectives. If all the objectives are uncorrelated, they can be optimized separately, and we can find a combined policy to optimize all of them. However, it is not always clear a priori which objectives might be correlated and how they influence each other. As the objectives are conflicting, there usually exists no single optimal solution. In those cases, we are interested in a set of trade-off solutions that balance the objectives.

Most RL approaches on multi-objective tasks rely on single-policy algorithms to learn Pareto optimal solutions. In a single-policy approach, the purpose is to obtain the best policy that simultaneously satisfies the preferences among the multiple objectives assigned by the designer or defined by the application domain. Single-policy MORL algorithms employ scalarization functions to define a utility over a vector-valued policy and thereby reducing the dimensionality of the underlying multi-objective environment to a single, scalar dimension. In such an approach, scalar $\hat{Q}$-values are extended to $\hat{Q}$-vectors that store a $\hat{Q}$-value for each objective, i.e.,

$$\hat{Q}(s,a) = (\hat{Q}_1(s,a), \dots, \hat{Q}_m(s,a)) \tag{2}$$

where m is the number of objectives. When an action is selected in a given state of the environment, a scalarization function f is applied to the $\hat{Q}$-vectors of each action in order to obtain a single, scalar $SQ(s,a)$ estimate. This work adopts the weighted sum approach (Marler and Arora 2010), which computes a linearly weighted sum of Q-values for all the objectives, as follows:

$$SQ(s,a) = \sum_{i=1}^{m} w_i Q_i(s,a) \tag{3}$$

where $\sum_{i=1}^{m} w_i = 1$. The weights allow putting more or less emphasis on each objective. The Q-value update rule for each objective can be expressed by the following equation:

$$Q_i(s,a) \leftarrow (1-\alpha) Q_i(s,a) + \alpha \left[r_i + \gamma \cdot \max_{a'} Q_i(s',a') \right] \tag{4}$$

4 Materials and methods

4.1 Problem formulation

The problem addressed in this paper concerns the optimization of the furniture arrangement in multi-occupant offices. The primary purpose is to find office layout patterns that maximize the satisfaction of all occupants in terms of thermal, acoustics and visual comfort while satisfying some basic office functional requirements (e.g., the layout provides easy access to tools and people when needed and whether there is sufficient space between occupants). For achieving our aim, we modeled the multi-objective optimization problem by using Multi-objective Markov Decision Process as follows:

$$MOMDP = \langle F, S, s_0, s_f, A, P, R \rangle \tag{5}$$

where

- F is the set of furniture items that must be placed in a given office.
- S is the space of possible states. In particular, s_0 is an initial state represented by an initial arrangement of the furniture set, and $s_f \in S$ is a final state representing an office layout pattern providing an optimal comfortable situation for all office occupants.
- A is the space of available actions. It is represented by a finite set of possible movements that can be performed for each object $f \in F$. To explore all possible furniture positions and orientations, the following set of actions are considered $act(f) = \{MoveUp, MoveDown, MoveRight, MoveLeft, Rotate, NoMove\}$.
- $P : S \times A \times S \rightarrow [0,1]$ is a probability distribution function. $P(s' \mid s,a)$ is the probability of transition from a state s to a state s' due to an object movement.
- $R : S \times A \times S \rightarrow \mathbb{R}^d$ is a vector-valued function specifying the immediate reward obtained for each objective by moving an object from a position to a new position defined as follows:

$$R(s' \mid s,a) = (R_1(s' \mid s,a), R_2(s' \mid s,a), \dots, R_d(s' \mid s,a)) \tag{6}$$

where $d \geq 2$ is the number of objectives and each scalar value $R_i(s' \mid s,a)$ is the reward value for the ith objective.

Let assume n different items to be placed in an office $(f_1, f_2, \dots, f_n)$ (e.g., desks, printers, chairs, bookshelves). An office layout pattern is characterized by an arrangement of such items in some office locations $(l_1, l_2, \dots l_n)$. We use a MORL approach to find optimal office layout patterns, in other word we find mappings $(f_1, f_2, \dots, f_n) \rightarrow (l_1, l_2, \dots l_n)$ which maximize a vector R of reward. In our problem, there is a unique location for placing each item f_i. Conversely, each location l_j can be assigned to any furniture. A reference system is associated with the office plant, whose origin corresponds to the top left corner. Each furniture has a location defined by a couple of coordinates (x, y), corresponding to the centre of the furniture, and size s represented by a couple $(width, depth)$. Each learning step, the MORL agent places an item f_j at a new location $l_j = (x_j, y_j)$. Then, a new layout pattern is determined, and a new reward R is evaluated considering the impact of the current office layout pattern on comfort conditions.

4.2 Agent knowledge

During the learning process, the MORL agent needs to know the geometric features of the office under analysis, the features of each item to be placed, and the values of environmental parameters, namely illuminance (lx), the air temperature and the mean radiant temperature (°C) and the sound pressure level (dB). Whereas the features of the room and the furniture are known by design, the evaluation of the environmental state depends on the knowledge of scalar fields defining the spatial distribution of the environmental parameters. Since the knowledge of the whole environment state would require a significant number of sensors placed in the room resulting in an expensive solution, in this work, we evaluate the spatial distribution of the environmental parameters using appropriate tools. In particular, we used the lighting simulation software Dialux Evo (DIAL 2018) to reproduce the illumination field inside an environment. It represents an inexpensive trade-off since it requires only a model of the environment and an initial validation to assure a proper position of the measurement point.

As concerns the acquisition of the sound pressure level, we use Eqs. (7)–(10) considering the Sabine model (Sabine and Egan 1994). For a continuing sound source in a room, the sound level is the sum of direct and reverberant sound (Lewis and Bell 1994). The sound pressure level for a receiver is given by the following equation:

$$L_{Receiver} = L_{Source} + 10log(Q/(4\pi r^2) + 4/Ac_R) \tag{7}$$

where $L_{Receiver}$ is the sound pressure level (dB) received in a given point of the room at r distance from the source. L_{Source} is the sound power level from the source of the noise. Q is the directivity factor that measures the directional characteristic of a sound source. It assumes predefined values according to the locations of the sound source. Finally, the room constant Ac_R[1] expresses the acoustic property of a room according to:

$$Ac_R = S\bar{\alpha}_{abs}/(1 - \bar{\alpha}_{abs}) \tag{8}$$

where S is the room total surface and $\bar{\alpha}_{abs}$ is the mean absorption coefficient of the room computed as follows:

$$\bar{\alpha}_{abs} = \frac{\sum_{i=1}^{n} S_i \alpha_i}{\sum_{i=1}^{n} S_i} \tag{9}$$

where S_i is an individual surface in the room (m^2) made with a specific material that is characterized by an absorption coefficient α_i.[2] Moreover, we assume that several sources of noise can affect the environment. Thus, we need to consider the total contribution of each source to the receiver that can be obtained as follows:

$$L_{TotReceiver} = 10 * log \sum_{j=1}^{k} 10^{0.1 * L_j} \tag{10}$$

where k is the number of sources and L_j is the sound pressure level of the jth source.

Concerning the temperature, since the amount of radiant heat lost or received by the human body is the algebraic sum of all radiant fluxes exchanged by its exposed parts with the surrounding sources, the mean radiant temperature can be calculated from the measured temperature of surrounding surfaces and their positions with respect to the person. We assume that slight temperature differences exist between the surfaces of the enclosure; thus, we use the linear form Guo et al. (2020):

$$T_{mr} = T_1 F_{p-1} + T_2 F_{p-2} + \cdots + T_n F_{p-n} \tag{11}$$

where T_{mr} is the mean radiant temperature, T_i is the temperature of the ith surface of the office, F_{p-i} is the angle factor between a person and the ith surface.

4.3 Reward function

As previously said, several factors influence user satisfaction with their workplace. The most common ones concern their desks position and orientation that are strictly correlated to environmental comfort as a whole. In multi-occupant offices where each employee performs independent tasks, more considerable distances between occupants correspond to a more comfortable space for their working activities. Moreover, access to daylight and outside view, the distance from noise sources and surfaces with low thermal insulation corresponds to a higher degree of IEQ satisfaction. Hence, it is necessary to find an optimal trade-off between office occupants in shared environments to guarantee the same degree of satisfaction. Hence, in order to find office layout patterns that optimize users satisfaction in a multi-occupant office and some functional office requirements, we defined the reward function as follows:

$$R = (R_{user_1}, R_{user_2}, \ldots, R_{user_k}, R_{dist-users}) \tag{12}$$

where R_{user_i} is the reward associated with the ith occupant, higher rewards correspond to better occupant working position. $R_{dist-users}$ is the reward associated with the distance

[1] In this paper, we named the room acoustic propriety with the symbol Ac_R instead of the classical R to avoid misunderstandings with the reward function.

[2] Common absorption coefficients can be found at https://www.acoustic.ua/st/web_absorption_data_eng.pdf.

between occupants; larger distances correspond to a more comfortable space. $R_{dist-users}$ is defined as follows:

$$R_{dist-users} = \{D_{ij}, \forall i,j \; with \; i \neq j\} \tag{13}$$

where

$$D_{ij} = \sqrt{(x_{desk_i} - x_{desk_j}) + (y_{desk_i} - y_{desk_j})} \tag{14}$$

Concerning R_{user_i}, it is computed as follows:

$$R_{user_i} = R_{IEQ} + R_{door} \tag{15}$$

where R_{door} associates a quantitative measure to the compliance with a layout design rule that recommends placing any item far from the door for not impeding its regular use and R_{IEQ} is the reward value that varies according to the comfort perception of a single user.

In this work, we use the indoor environment quality index as a measure of R_{IEQ} reward. In particular, the IEQ index refers to the building's indoor environment quality considering the occupants' satisfaction level presented on a 0–100% scale. In this study, according to the IEQ model proposed in Piasecki et al. (2017), we evaluated the IEQ index as the weighted sum of three indoor comfort sub-indexes (also presented on a 0–100% scale): thermal comfort TC_{index}, acoustic comfort AC_{index}, visual comfort VC_{index}. Hence, we formulated the reward function R_{IEQ} as follows:

$$R_{IEQ} = p_{TC} * TC_{index} + p_{AC} * AC_{index} + p_{VC} * VC_{index} \tag{16}$$

TC_{index} refers to the percentage of people accepting the thermal environment:

$$TC_{index} = 100 - PD_{TC} \tag{17}$$

where PD is the Predicted Percentage Dissatisfied which indicates the number of people dissatisfied with the thermal environment as follows:

$$PD_{TC} = 100 - 95 \times e^{(-0.03353PMV^4 - 0.2179PMV^2)} \tag{18}$$

AC_{index} refers to the ability of buildings to provide an environment with minimal noise:

$$AC_{index} = 100 - PD_{Acc} \tag{19}$$

where PD_{Acc} is the predicted percentage dissatisfied indicating the number of occupants dissatisfied of sound pressure level (SPL) with the change in noise level from *Recommended* to *Actual* value:

$$PD_{Acc} = 2 * (Actual_{SPL} - Reccomended_{SPL}) \tag{20}$$

We assume that the source of this noise may be external or internal to the office. Internal noise, for example, can be generated by HVAC (Heating, Ventilation and Air Conditioning) systems. Conversely, external noise can be propagated by external devices or people.

VC_{index} is based on the amount of light falling on the working plane defined as follows:

$$VC_{index} = 100 - PD_L \tag{21}$$

where PD_L is the percentage of persons dissatisfied with minimum daylighting or the probable percentage of people switching on artificial lighting. The function of daylight illuminance $E_{min}[lx]$ and the predicted percentage of dissatisfied occupants was calculated with the following equation:

$$PD_L = \frac{(-0.0175 + 1.0361)}{1 + exp[4.0835 \times (log(E_{min}) - 1.8223)} \times 100 \tag{22}$$

Moreover, p_{TC}, p_{AC} and p_{VC} in Eq. (16) are weights related to the user preferences with respect to each kind of comfort.

Finally, it is worth noting that, to eliminate the effects of the variation in the scale about the different metrics in the multi-objective problem, R_{door} and $R_{dist-users}$ are normalized distance values according to the following equation:

$$\overline{dist} = \frac{dist - min_{dist}}{max_{dist} - min_{dist}} \times 100 \tag{23}$$

where $dist$ is the distance value to be normalized, max_{dist} and min_{dist} are the maximum and minimum distances in the office under study, with respect to the door and between users for the normalization of R_{door} and $R_{dist-users}$ respectively.

4.4 IEQ-based multi-objective Q-learning algorithm for optimal furniture layouts

Algorithm 1 IEQ-based Multi-Objective Q-learning for optimal furniture layouts

Require: MOMDP model, Room Plant
Ensure: Opt Layout Pattern $(l_{1_f}, l_{2_f}, ..., l_{n_f})$

$< S, F, s_0, s_f, A, P, R > \leftarrow MOMDP$;
$F \leftarrow (f_1, f_2, ..., f_n)$;
for $f_i \in F$ **do**
  $act(f_i) \leftarrow [MoveUp(f_i), MoveDown(f_i),$
    $MoveLeft(f_i), MoveRight(f_i),$
    $Rotate(f_i), NoMove(f_i)]$;
  $AvailableAction.append(act(f_i))$;
end for
$MOQ \leftarrow [Q_1, Q_2, ..., Q_d]$;
$s_0 \leftarrow (l_{1_0}, l_{2_0}, ..., l_{n_0})$;
$s \leftarrow s_0$;
for each episode e **do** ▷ Learning Process
  $act \leftarrow \epsilon$ *-greedy action selection* ▷ (see Alg.2)
  MakeAction(act);
  Observe the new state s';
  for i=1,...,d **do**
    evaluate R_i; ▷ (see eq.12-22)
    $Q_i(s,a) \leftarrow (1-\alpha) \cdot Q_i(s,a) + \alpha \cdot \left[r_i + \gamma \cdot \max_{a'} Q_i(s', a')\right]$;
  end for
  $s \leftarrow s'$
end for
$s_f \leftarrow$ *optimal state selection* ▷ (see Alg.3)
$(l_{1_f}, l_{2_f}, ..., l_{n_f}) \leftarrow s_f$

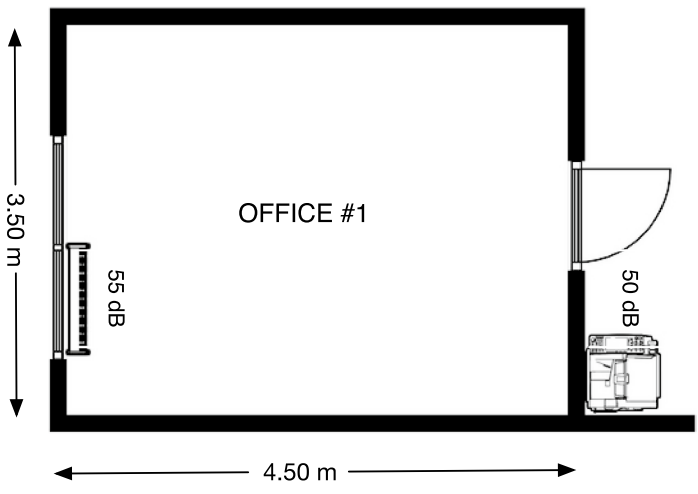
The pseudo-code of the IEQ-based multi-objective Q-learning is shown in Alg. 1. Starting from the MOMDP model and the characteristics of the space to be configured, Alg. 1 provides a mapping of the furniture (f_1,f_2, … ,f_n) to room locations (l_1, l_2, … l_n) that maximize the environmental comfort of each occupant. Firstly, Alg. 1 initializes the set of elements to be placed in the room plane. For each furniture f_i, a set of actions *Act*(f_i) is available. The multi-objective Q-learning (MOQ) matrix is initialized according to the number of objectives to be optimized as well as the the initial positions of the furniture. The *i*th element of MOQ matrix is the Q-learning matrix related to the *i*th objective. For each learning step, starting from the initial state s_0, the agent chooses an action from the list of the available actions according to the selection strategy defined in Alg. 2.

Algorithm 2 $\epsilon - greedy$ action selection for MORL

Require: Current State s, Available Actions $AvAct$, MOQ matrix, weights vector W
Ensure: $\epsilon - greedy$ act

$(w_1, ..., w_d) \leftarrow W$;
$QScalarisedVector \leftarrow \varnothing$;
$MOQ \leftarrow [Q_1, Q_2, ..., Q_d]$;
for $act \in AvAct(s)$ **do**
  $Qvector \leftarrow (Q_1(s, act), ..., Q_d(s, act))$;
  $Qvalue(s, act) \leftarrow \sum_{i=1}^{d} w_i Q_i(s, act)$;
  $QScalarisedVector.append(Qvalue(s, act))$;
end for
$num \leftarrow random.uniform(0, 1)$;
if $num \leq \epsilon$ **then**
  $act \leftarrow random(AvAct(s))$;
else
  $act \leftarrow max(QScalarisedVector(s, :))$;
end if

According to Alg. 3, when an action is selected in a particular state, a scalarization function is applied to the Q-vector of each action to obtain a scalar estimate ScalarisedQ(s, a). In particular, for each available action in a given state s, the algorithm computes a new Q-value attributed at state s through a weighted sum function. The weight associated with each element of the *Qvector* expresses the relevance attributed to every single objective of the problem. Then, an ϵ-greedy selection is applied to the scalarized vector that returns the optimal action (the one with the greatest value) with probability $1 - \epsilon$ and random action with probability ϵ. After action selection in Alg. 1, the learning agent acquires the new state of the environment obtained after the execution of the chosen action. Hence, for each objective, a reward is evaluated for the new observed state s' according to Eqs. 12–22. As the Q-matrix has been extended to incorporate a separate value for each objective, these values are updated individually. The single-objective Q-learning update rule is extended for a multi-objective environment. More precisely, for each triplet of state, action and objective, the Q-values are updated, by using the corresponding reward (i.e., $R_{user}, R_{dist-users}$), into the direction of the



									x
80	50	130	130	130	110	100	90	90	
80	50	130	130	130	110	100	90	90	
440	350	350	300	300	300	100	90	90	
540	350	350	300	300	200	200	150	90	
800	350	300	300	300	200	200	150	125	
700	350	370	300	300	200	200	150	110	
300	350	370	330	200	190	190	180	200	

y

Fig. 1 Office #1 and illuminance values on its surface

best-scalarized action of the next state. Finally, an optimal layout pattern is obtained through the Alg. 3. In this case, a scalarization process on the whole MOQ-matrix has been performed, thus obtaining a scalarized Q matrix from which we find the state having the maximum scalarizedQ value. Hence, the optimal layout pattern that maximizes the users satisfaction with respect to their working environment is given by $(f_1, f_2, \ldots, f_n) \leftarrow (l_{1_f}, l_{2_f}, \ldots, l_{n_f})$. It is worth noting that as the proposed approach considers preferences during the optimization process, the obtained solutions are equivalent.

Algorithm 3 Optimal State Selection

Require: MOMDP, MOQ Matrix, weights vector W
Ensure: Opt State S_{opt}

$< S, F, s_0, s_f, A, P, R > \leftarrow MOMDP$;
$MOQ \leftarrow [Q_1, Q_2, ..., Q_d]$
$(w_1, ..., w_d) \leftarrow W$
$ScalarisedQ \leftarrow \varnothing$
for $s \in S$ **do**
  for $a \in A$ **do**
    $Qvector \leftarrow (Q_1(s,a), ..., Q_d(s,a))$;
    $Qvalue(s,a) \leftarrow \sum_{i=1}^{d} w_i Q_i(s,a)$;
    $ScalarisedQ(s,a) \leftarrow Qvalue(s,a)$;
  end for
end for
Find s where $ScalarisedQ(s,:)$ *is* Max;
$s_{opt} \leftarrow s$

5 Experimental evaluation

The experimental evaluation has been conducted on two real offices that show different physical characteristics as described in the following.

OFFICE #1: The size of the first office is $4.5 \times 3.5 \times 2.8$ m. The room has a great north-west facing window and a door located as it is shown in Fig. 1. The room has 15.75 m^2 of gray stoneware floor with absorption coefficient $\alpha_{abs} = 0.3$, 15.75 m^2 of roof with $\alpha_{abs} = 0.1$, 4 m^2 of 4 mm glass windows with $\alpha_{abs} = 0.3$ and 40.8 m^2 of walls with $\alpha_{abs} = 0.1$. The acoustic property of the room calculated according to Eq. (8) is $AC_R = 13.65$. The air conditioning system installed under the window produces a noise of $55dB$. For this study, a setpoint temperature of 23 °C has been assumed. Outdoor the room, there is an aisle crossed by several workers and a printer near the office door. The worker chattering and the printer produce an average sound pressure of 50 dB. Acquisition of daylight illuminance values shows that illuminance values around 300 lx characterize the central space of the room; the area close to the windows shows values between 500 and 800 lx. The top left corner of the room has the lowest illuminance level (i.e., 50–130 lx). In order to validate the data acquired by the simulation software, we collected direct measures in some points through a luxmeter. Figure 1 illustrates the plant of office #1 along with the illuminance on its surface. Each cell is represented with its lux value. Moreover, the

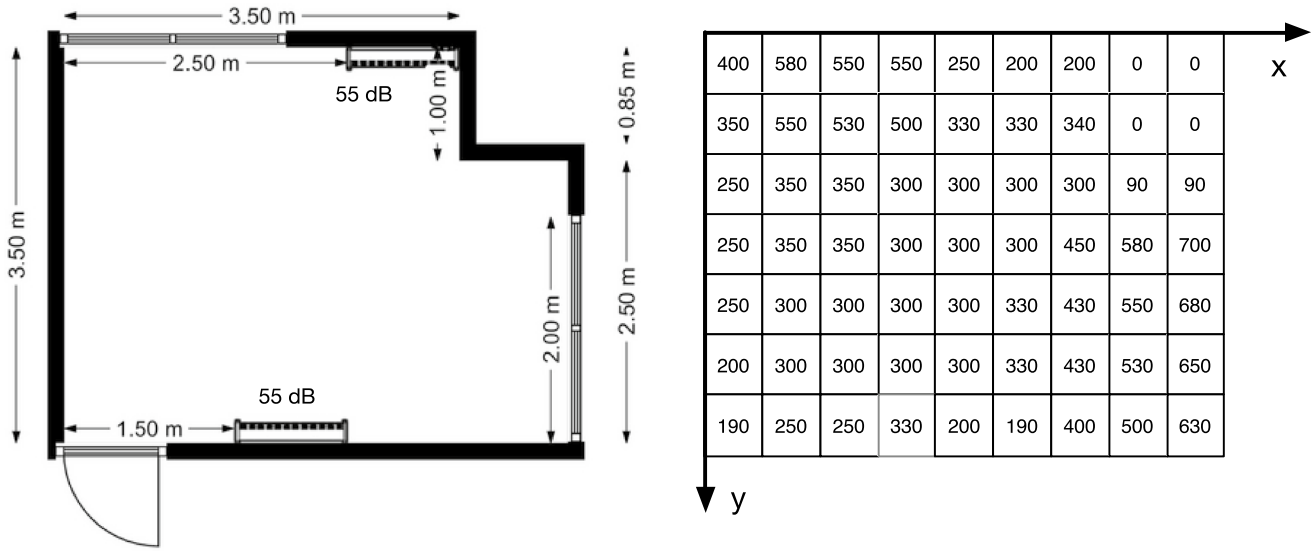


Fig. 2 Office #2 and illuminance values on its surface

coordinates of each cell room are taken according to the reference system, as shown in Fig. 1.

OFFICE #2: The size of the second office is slightly different from the previous one. Shape and sizes are shown in Fig. 2. The room has two windows; the first is a north-west facing window, and the second is a north-facing window. The office door is located at an angle of the bottom wall, as is shown in Fig. 2. The room has 14.75 m^2 of gray stoneware floor with absorption coefficient $\alpha_{abs} = 0.3$, 14.75 m^2 of roof with $\alpha_{abs} = 0.1$, 8 m^2 of 4 mm glass windows with $\alpha_{abs} = 0.3$ and 36.8 m^2 of walls with $\alpha_{abs} = 0.1$. The acoustic property of the room calculated according to Eq. (8) is $AC_R = 14.28$. Two air conditioning systems are installed on opposite walls, producing a sound pressure of 55 dB individually. Also, in this case, a setpoint temperature of 23 °C has been assumed. Acquisition of daylight illuminance values shows that illuminance values around 300 lx characterize the central space of the room; the area close to the north-west window shows values between 500 and 700 lx. The area near the north window shows lower values between 400 and 550 lx.

Each office can host two occupants. Hence, each room has to be configured with two desks and two desk chairs. The size of desks is 150 cm × 50 cm. The desk position constrains the position of the desk chairs. Thus, four possible orientations are considered (see Fig. 3). In our prototype, the room plants are represented as a grid where the dimension of each cell is 0.5 × 0.5 m. The MORL agent performs a cell movement or a 90° clockwise rotation at each step. We assumed that users have the same degree of preference concerning visual, acoustic, and thermal comfort in these case studies. Thus, in Eq. (16), $p_{TC} = p_{AC} = p_{VC} = 0.33$. Conversely, we assumed that the objectives have different weights. Thus, the weights vector of Alg. 3 $W = (0.4, 0.4, 0.2)$ indicates that the IEQ satisfaction of both users is more relevant than their reciprocal distance. Moreover, acoustic and thermal reward and users' distance are evaluated with respect to the desk chair position. Conversely, the visual reward is evaluated according to the working plane, namely the desk position.

5.1 Results and discussions

We executed 20 tests for each office under study. Each test performs 40,000 episodes. To measure the effectiveness of the solutions, we consider the sum of the reward values defined in Eq. (12) for each optimal/sub-optimal layout obtained at the end of every single test. Table 1 reports a synthesis of such data for each office. In particular, Office #2 provides better performance with respect to Office #1. Moreover, results show that optimal/sub-optimal layout patterns have a slight difference in terms of environmental comfort, as it is shown by the small difference between the max and min of the total reward. The low values of standard deviation indicate that all layouts found by the RL agent are very similar in terms of user satisfaction.

For each office, we also show a comparison between two layouts having a total reward higher and lower than the mean value, respectively. Two optimal layout patterns are depicted in Figs. 4 and 5 for Office #1. Their features are reported in Tables 2 and 3, respectively. As we can see, the Layout Pattern #1 balances the occupants' satisfaction in terms of environmental comfort and functional constraints. Indeed, both occupants are positioned far from the door (thus not impeding its regular use). The distance between users is enough to guarantee occupant movement inside the room, and the comfort indexes related to each aspect of the environmental

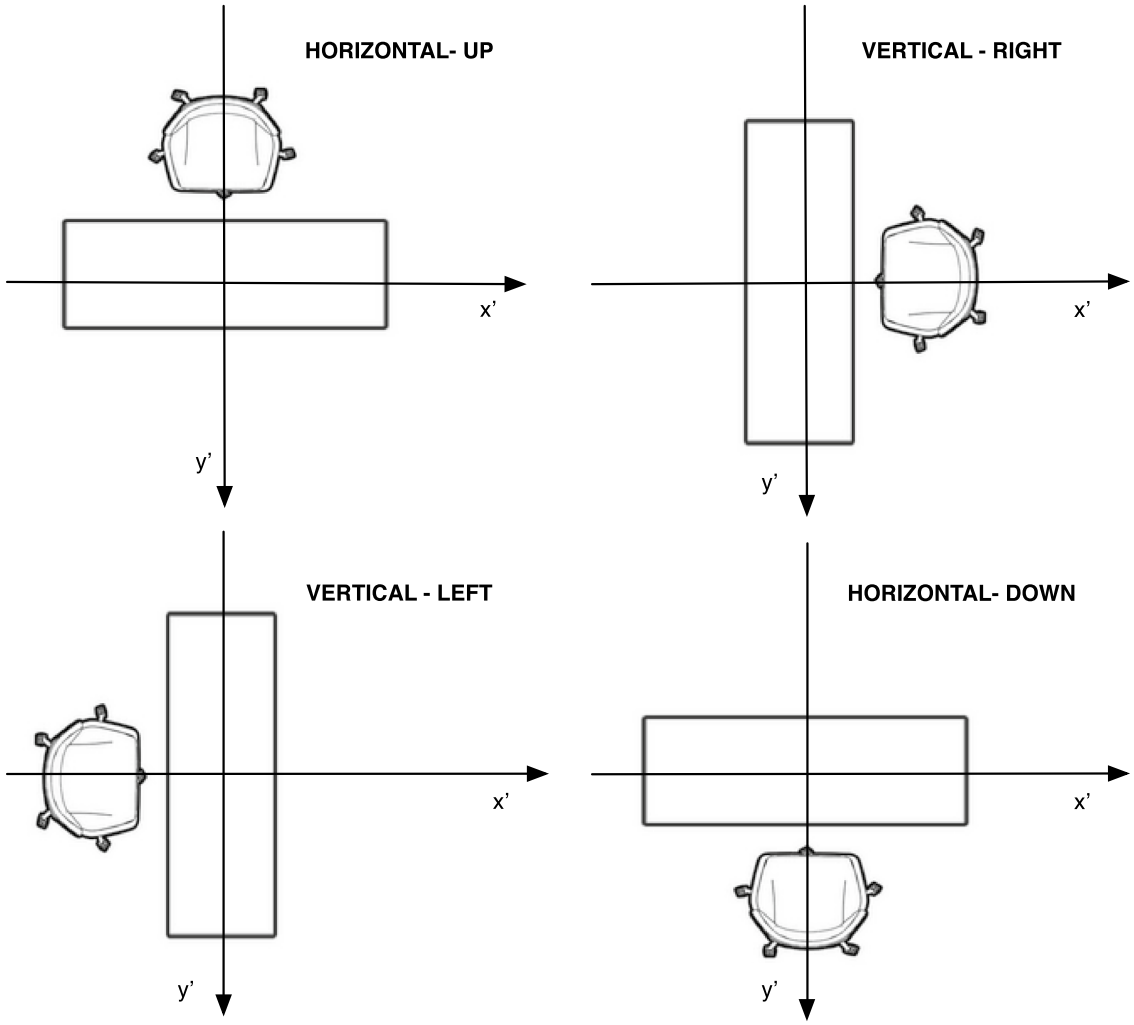


Fig. 3 Possible desk orientations

Table 1 Data synthesis of test results

	Office #1	Office #2
Mean	558.80	561.27
Max	561.46	563.21
Min	553.66	558.92
DevStand	1.63	1.20

quality are comparable. Layout Pattern #2 provides very similar comfort conditions with respect to Layout Pattern #1. Conversely, the different furniture arrangements provide a lower distance between users.

Analogously, optimal furniture layout patterns for Office #2 and their features are shown in Figs. 6 and 7 and tables 4 and 5 respectively. Also in the case of Office #2, the optimal layout patterns balance the occupants' satisfaction in terms of environmental comfort and functional constraints. The Layout Pattern #1 balances the occupants'

Table 2 Office #1—layout pattern #1 features

	User #1	User #2
Desk location	(2, 2), Horizontal-up	(5, 2), Horizontal-down
Acoustic comfort index	88.12	88.04
Thermal comfort index	95.0	95.0
Visual comfort index	94.92	95,37
Door distance	304 cm	360 cm
Users distance	304 cm	

satisfaction in terms of environmental comfort and functional constraints. Conversely, Layout Pattern #2 provides a lower level of thermal comfort due to the proximity of the worker chair to the window. Indeed, near the window, the mean radiant temperature of the windows is lower than

Fig. 4 Office #1—layout pattern #1

Table 3 Office #1—layout pattern #2 features

	User #1	User #2
Desk location	(2, 1), Vertical-right	(5, 2), Horizontal-DOWN
Acoustic comfort index	88.03	88.04
Thermal comfort index	95.0	95.0
Visual comfort index	94.92	95,37
Door distance	300 cm	380 cm
Users distance	200 cm	

the other surfaces, thus producing dissatisfaction in terms of thermal comfort.

Finally, in Figs. 8, 9, 10, 11, 12 and 13 are reported the thermal, visual and acoustics indexes of both offices for each user that are individually obtained for each test. It is worth noting that the proposed algorithm provides optimal layout patterns that maximize the environmental comfort for each occupant for a given office with given physical features. Such features influence the environmental comfort, and it may occur that the optimal environmental quality that can be obtained for an office does not satisfy the recommended standard thresholds. Indeed, according to the IEQ standards, the comfort indexes thresholds that define high-quality environments for office end-use are the following ones: $TCindex > 90$, $VCindex > 97$ and $ACindex > 95$.

Analyzing the results makes it possible to note that optimal layout patterns are configurations with the maximum reachable value for that office. However, in some cases, it may occur that the optimal layout patterns do not reach values of comfort recommended by the standards due to the physical configuration of the office. Indeed, we can see that all optimal layouts for Office #1 satisfy the level of thermal comfort with respect to the standard (see Fig. 8). Conversely, the acoustic comfort is out of the range of the acceptable levels (see Fig. 9). Since the standard for acoustics recommends 45 dB, the room size is not enough to reduce the combined effect of the two noise sources at the recommended value. At each position, the acoustic pressure level is around 50 dB. Thus, for office #1, we cannot have a notable improvement in terms of acoustic comfort, meaning that the office needs some design correction, such as replacing the old

Fig. 5 Office #1—layout pattern #2

Table 4 Office #2—layout pattern #1 features

	User #1	User #2
Desk location	Desk: (5, 8), Vertical, left	(0, 2), Horizontal, down
Acoustic comfort index	84,86	84.91
Thermal comfort index	95.0	95.0
Visual comfort index	98.24	97.66
Door distance	353 cm	255 cm
Users distance	370 cm	

Table 5 Office #2—layout pattern #2 features

	User #1	User #2
Desk location	(1, 2), Horizontal-up	(4, 7), Horizontal-down
Acoustic comfort index	84.92	84.86
Thermal comfort index	93.0	95.0
Visual comfort index	97.5	97.66
Door distance	316 cm	353 cm
Users distance	353 cm	

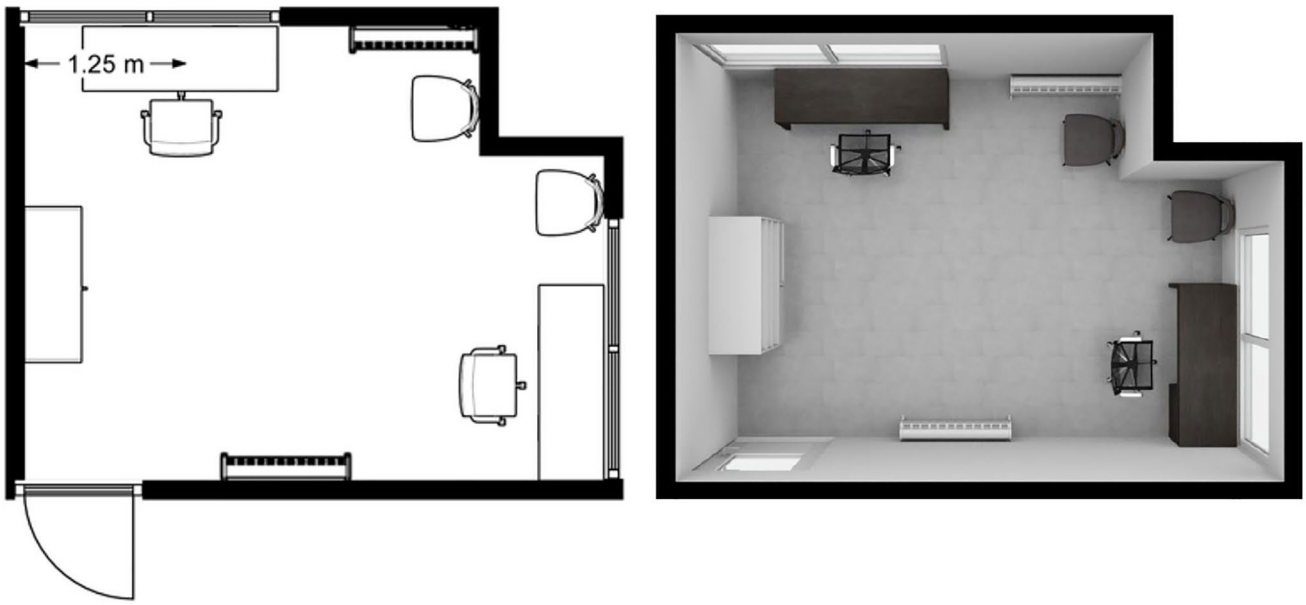


Fig. 6 Office #2—layout pattern #1

conditioning system. The experiments also show that in the office #1 is not possible to reach acceptable levels of visual comfort standard during daylight (see Fig. 10). This situation is due to the window's orientation that does not allow high light distribution. Hence, artificial lights also have to be used in the hour of daylight to reach the recommended lux level on the work-plane.

The results obtained on Office#2 show some differences in terms of IEQ. As we can see, there is no variability in terms of acoustics comfort index both for office #1 and office #2. Analogously to Office #1, also in Office #2 the level of acoustics comfort is far from the standard (see Fig. 12). The acoustics pressure level to the worker position is around 52 dB. As concerns thermal comfort, there is no difference between offices, both satisfying the level of standard (see Fig. 11). Differently from layouts of Office#1, an improvement is shown about the visual comfort due to the presence of two windows with different orientations (see Fig. 13).

6 Algorithm comparison

To demonstrate the effectiveness of the proposed approach for a multi-objective layout optimization in terms of indoor environmental quality, we first compare the proposed method with three mainstream single-objective RL methods, including Deep Q-Learning (DQN), Actor-Critic (A2C) and Proximal Policy Optimization (PPO). Then, we compare the proposed method with the Non-Dominated Sorting Genetic Algorithm II (NSGA-II), one of the most popular multi-objective optimization algorithms. In particular, DQNs (Arulkumaran et al. 2017) are Q-learning methods where Q-tables are replaced with Neural Networks for approximating Q-values for each action-state pair. A2Cs (Grondman et al. 2012) belong to a hybrid class of RL approaches that combine value estimation and policy gradient methods. An actor-critic method generally consists of an actor

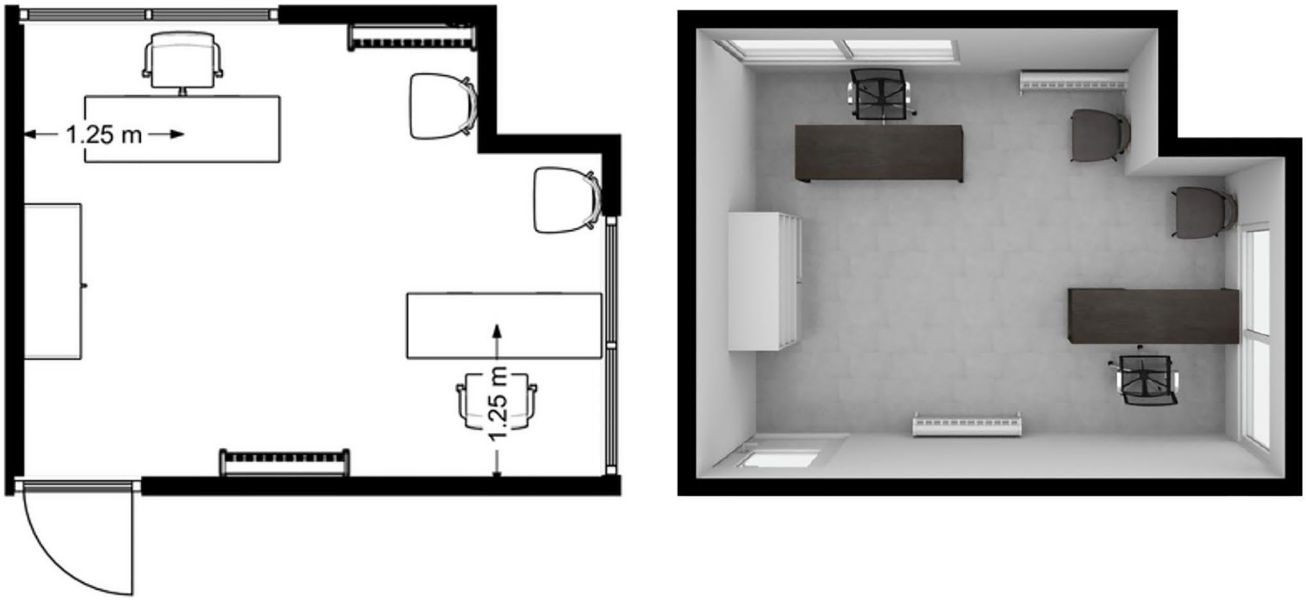


Fig. 7 Office #2—layout pattern #2

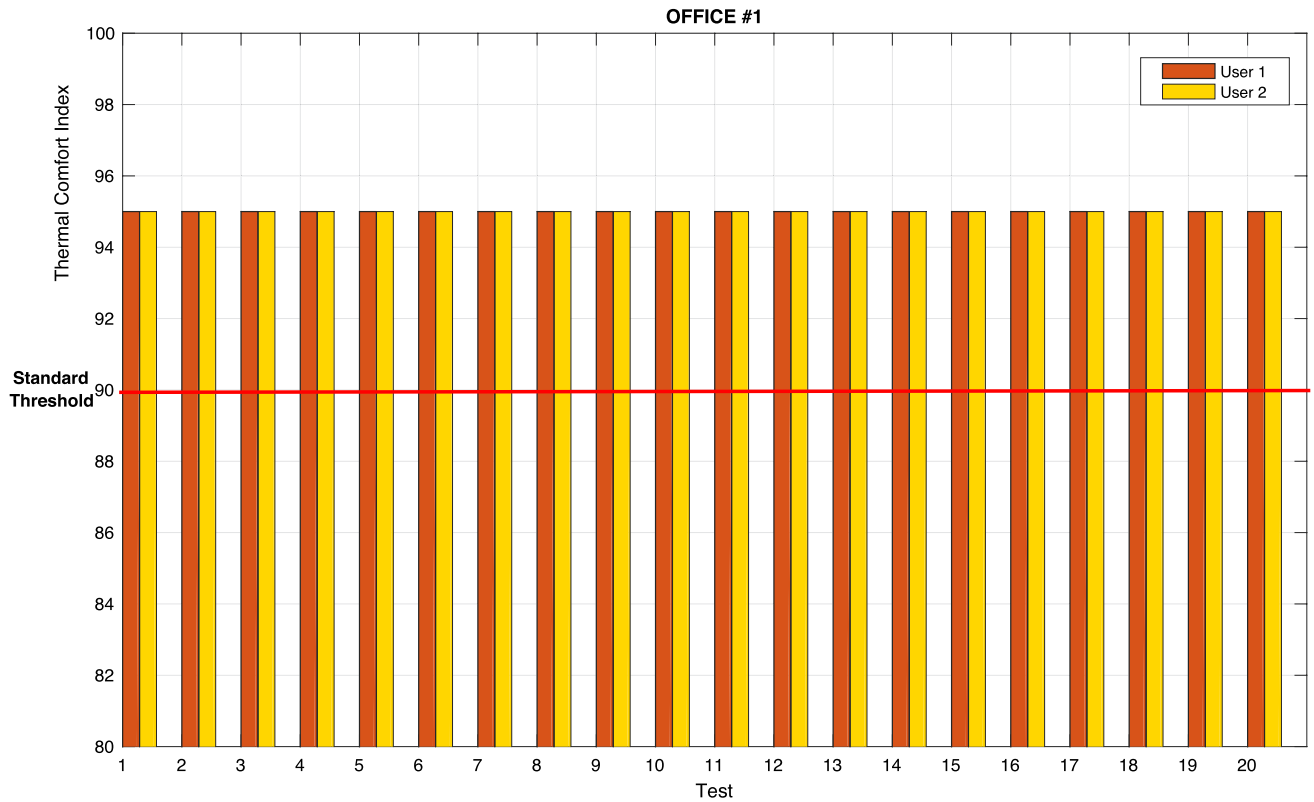


Fig. 8 Thermal comfort index office #1

that changes the policy to maximize its value, as estimated by the critic. The critic uses an approximation architecture to learn a value function, which is then used to update the actor's policy parameters in the direction of performance improvement. PPO (Schulman et al. 2017) is a family of policy optimization methods that use multiple epochs of

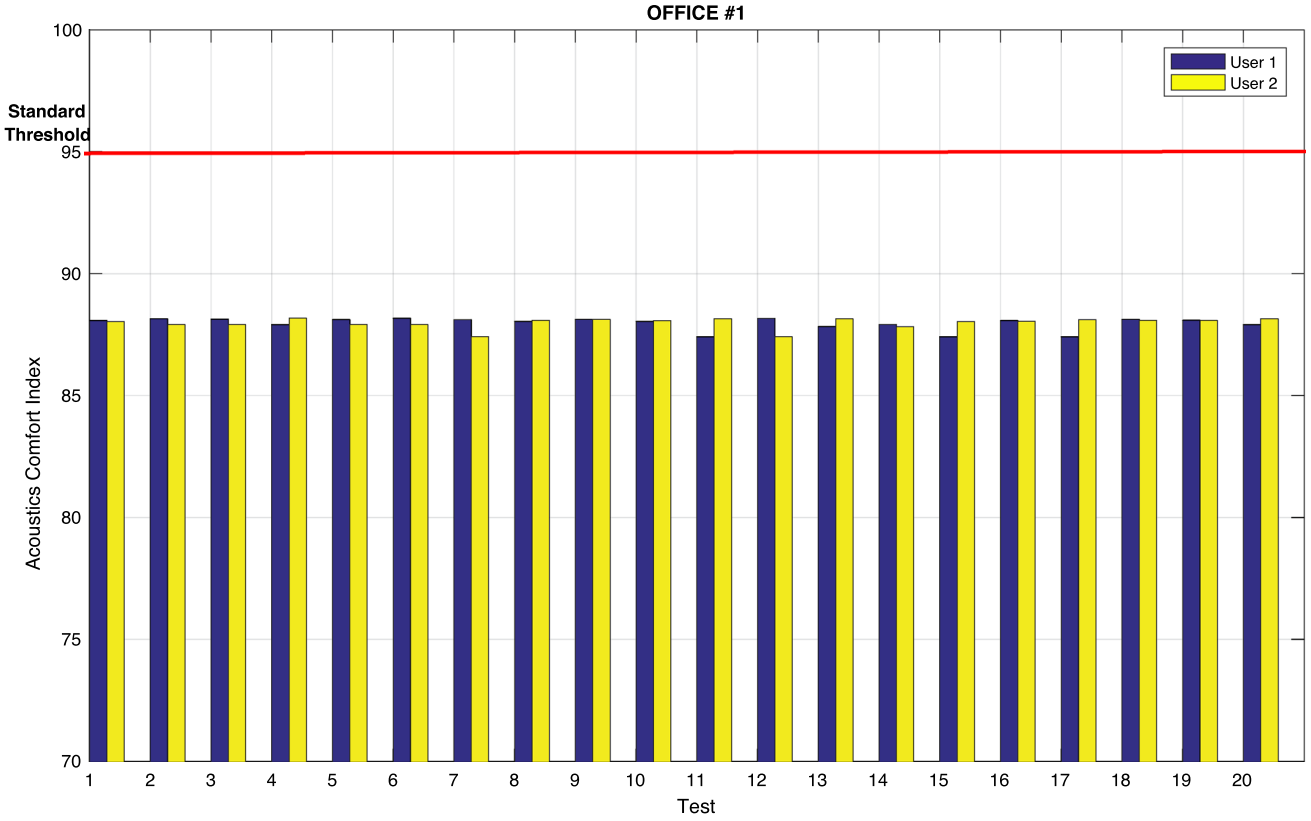


Fig. 9 Acoustics comfort index office #1

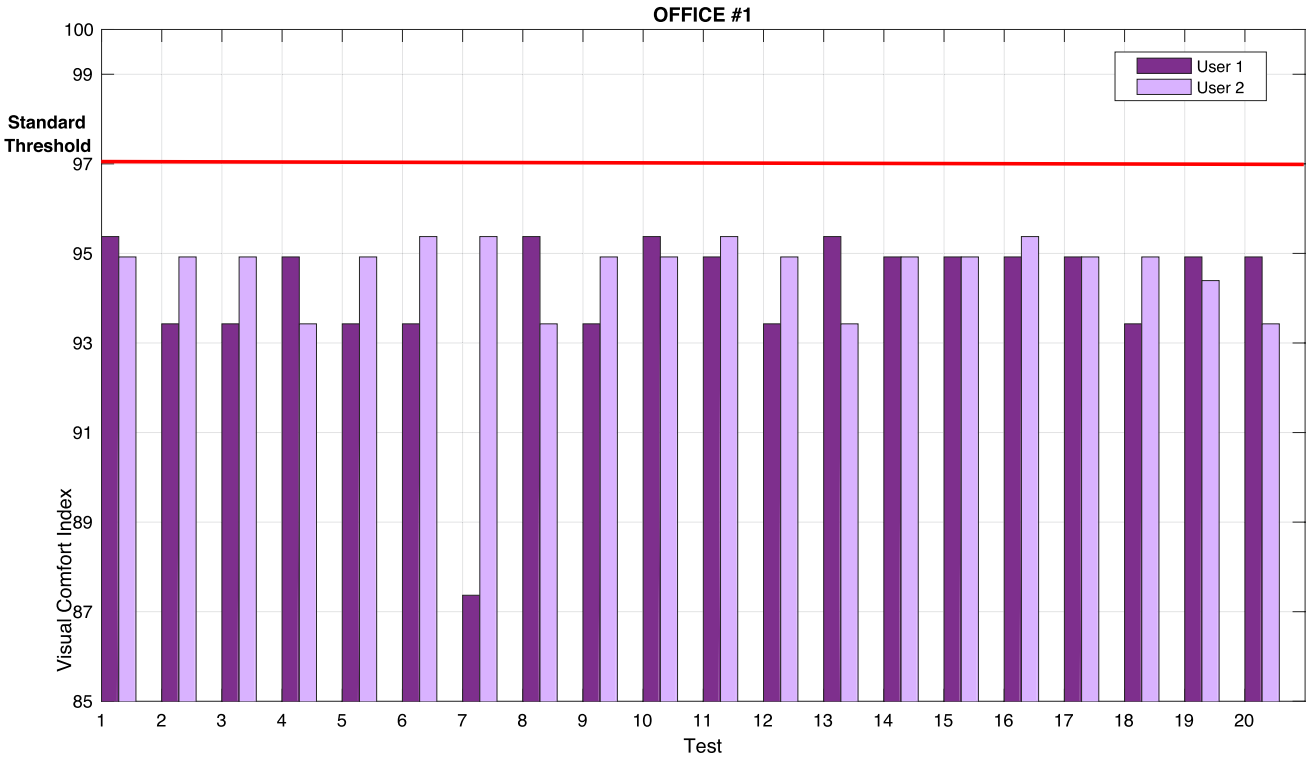


Fig. 10 Visual comfort index office #1

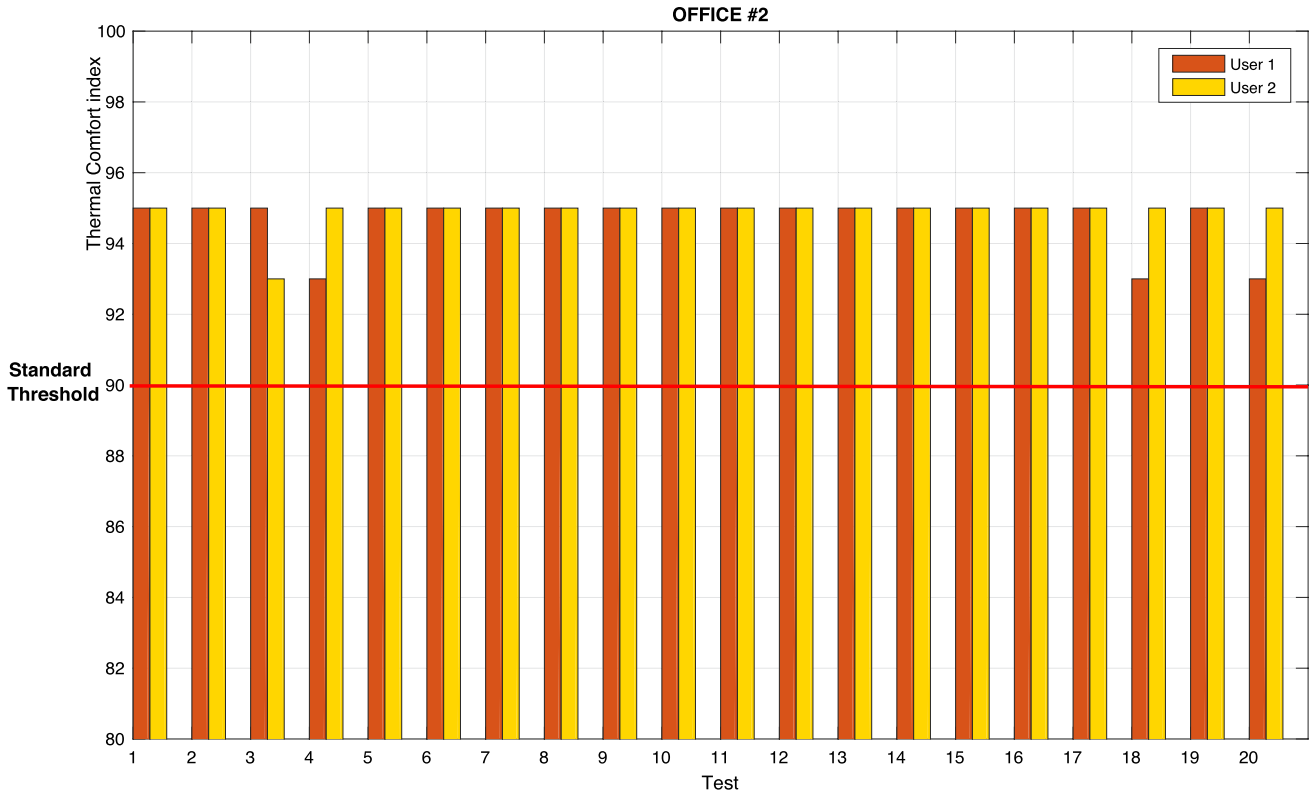


Fig. 11 Thermal comfort index office #2

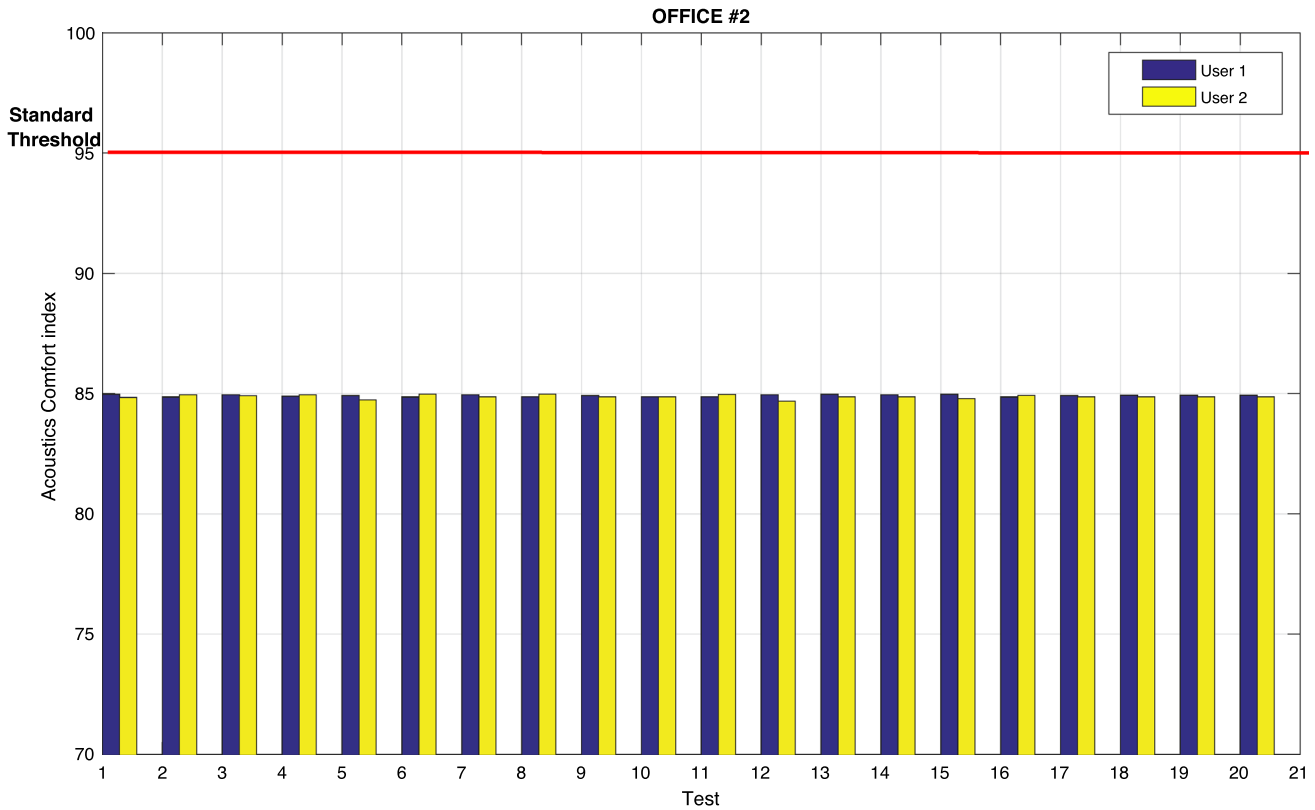


Fig. 12 Acoustics comfort index office #2

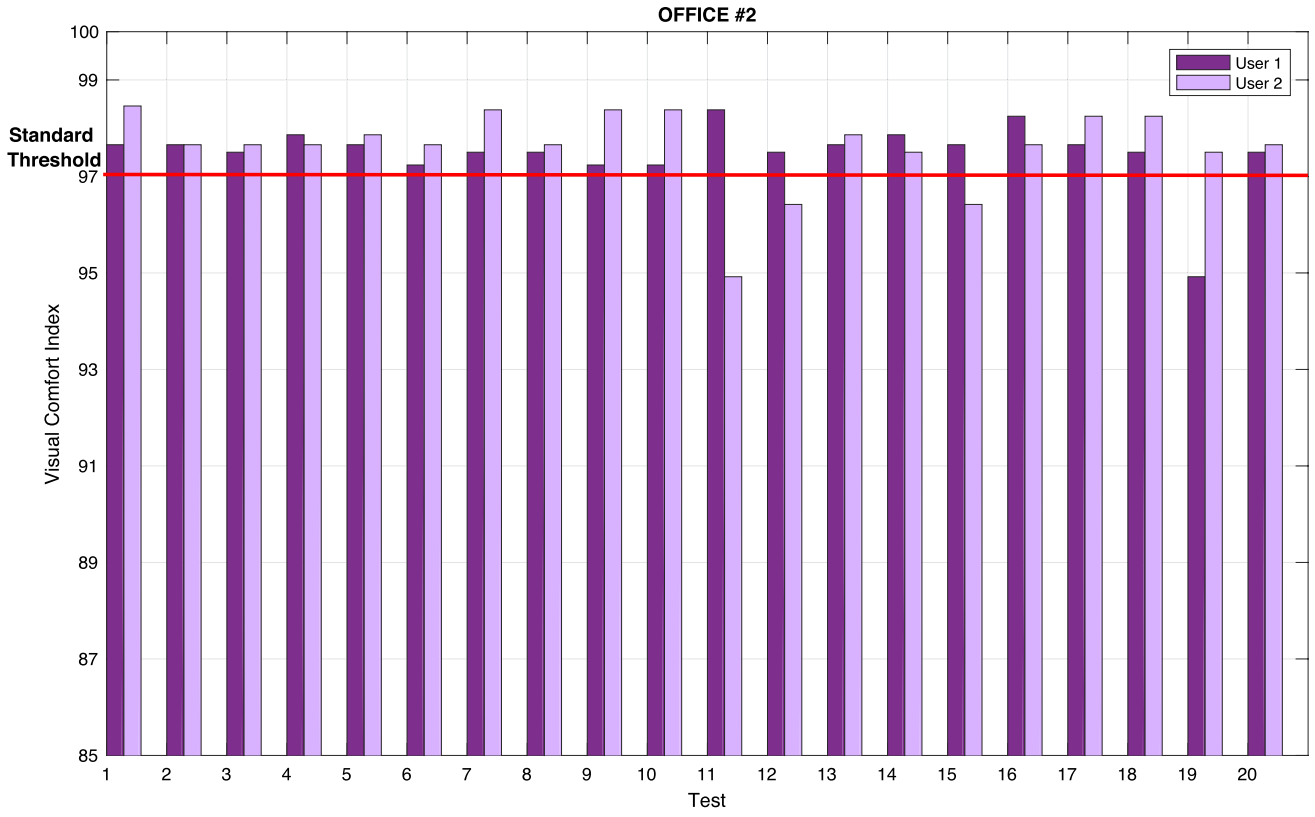


Fig. 13 Visual comfort index office #2

stochastic gradient ascent to perform each policy update by also using the actor-critic method. Conversely, NSGA-II (Deb et al. 2002) is a space exploration engine based on genetic algorithms. It is a multi-objective algorithm based on non-dominated sorting, fast crowded distance estimation procedure and simple crowded comparison operator. An NSGA-II solves a multi-objective optimization problem by generating a population of candidate solutions that evolve toward a near-optimal solution through a number of generations. As a result, it is able to find a set of optimal solutions called non-dominated solutions, which provide a suitable compromise between all objectives without degrading them.

To compare such methods with the proposed MORL, we conducted a set of experiments on the model of the OFFICE #1. Due to the randomness of the algorithms, each experiment was repeated ten times. The average values have been considered as the final results for the comparison. The evaluation criteria for the comparison are:

- The value of the optimal solution, given by the sum of the indexes of the environmental comfort of the occupants:

$$IEQ_{Tot} = R_{user_1} + R_{user_2} + R_{dist-users} \tag{24}$$

where R_{user_1}, R_{user_2}, and $R_{dist-users}$ are computed according to Eqs. (15) and (13). Higher IEQ_{Tot} denotes better layout configurations.
- The standard deviation of the solutions as measure of the variability of the algorithm performance. A low standard deviation implies a more stable algorithm.
- The execution time as measure of the time taken by each algorithm to provide an optimal solution.

It is worth noting that solving a multi-objective optimization problem is commonly addressed through two steps: finding a set of representative Pareto-optimal solutions and choosing a single preferred solution from the obtained set. Standard multi-objective evolutionary algorithms are applied to simultaneously provide a set of non-dominated solutions. Such algorithms treat each objective equally important and search in the solutions space without applying any preference strategy in their search process (Tang et al. 2020), although incorporating preferences into the search process of multi-objective evolutionary algorithms has gained attention recently (Tang et al. 2020; Wang et al. 2017; Bechikh et al. 2015). NSGA-II, as a standard multi-objective evolutionary algorithm, does not assign preferences to the objectives, thus giving them the same importance. Since our approach, unlike NSGA-II, incorporates weights for the objectives, the

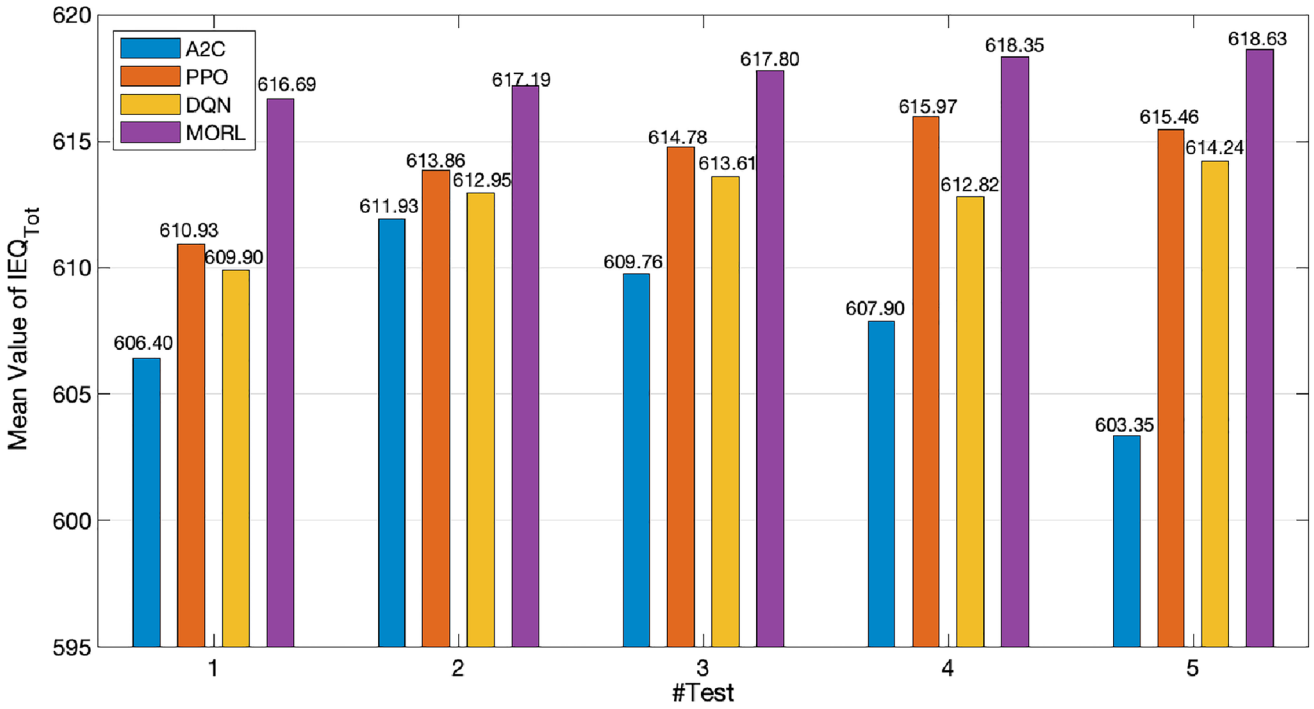


Fig. 14 Mean values of IEQ_{Tot} related to the optimal layout configurations

solutions provided by the proposed algorithm will be located on a preferred area of the Pareto front, i.e., our region of interest. Hence, to make an unbiased quantitative comparison between our approach and the NSGA-II, the experiments have been carried out by treating all the objectives equally without preferences, thus giving all algorithms the same working conditions.

Finally, the experiments have been conducted on a PC with the following configuration: 3 GHz 8-Core Intel Xeon E5 with 64 GB RAM, macOS Mojave operating system (Version 10.14.6). All the algorithms have been implemented in Python. In particular, the RL-based algorithms have been developed using Tensorforce (Kuhnle et al. 2017), an open-source deep reinforcement learning framework based on Tensorflow (Abadi et al. 2016). The hyperparameters are tuned according to the Bayesian Optimization and Hyperband (Falkner et al. 2018) method provided by Tensorforce. Vice versa, for the implementation of the NSGA-II, the python library provided in Pham Ngo Gia et al. (2018) has been used.

6.1 Comparison with RL methods

In this section, a comparison with the three single-objective RL methods is conducted. To make this comparison, we translate the multi-objective problem into a single-objective one by considering a unique reward through the weighted sum technique (Marler and Arora 2010) to be adopted by the single-objective RL methods. In particular, the experiment consists of twenty tests, five tests for each algorithm by varying the number of learning steps (i.e., $1 \times 10^3, 2 \times 10^3, \ldots, 5 \times 10^3$). Each test has been repeated ten times.

Figures 14, 15 and 16 show the mean values of IEQ_{Tot} related to the optimal layout configurations found from each algorithm, the standard deviation and the execution time of each RL algorithm. As we can see from the graphs shown in these figures, the proposed MORL finds, on average, better solutions than single-objective RL methods for each test, with a standard deviation that decreases by increasing the learning steps, meaning that MORL consolidates its learning process with 5×10^3 learning steps. Moreover, the execution times of the proposed algorithm are better than A2C and DQN for each test, while they are better than PPO during the first three tests (see Test #1,#2 and #3 in Fig 16) and similar to the consecutive ones (see Test #4 and #5 in Fig 16). The PPO and DQN algorithms propose quite similar solutions in terms of IEQ_{Tot} even if they need more learning steps to reach more consolidated solutions. They significantly differ in terms of execution times. DQN spends, on average, ten

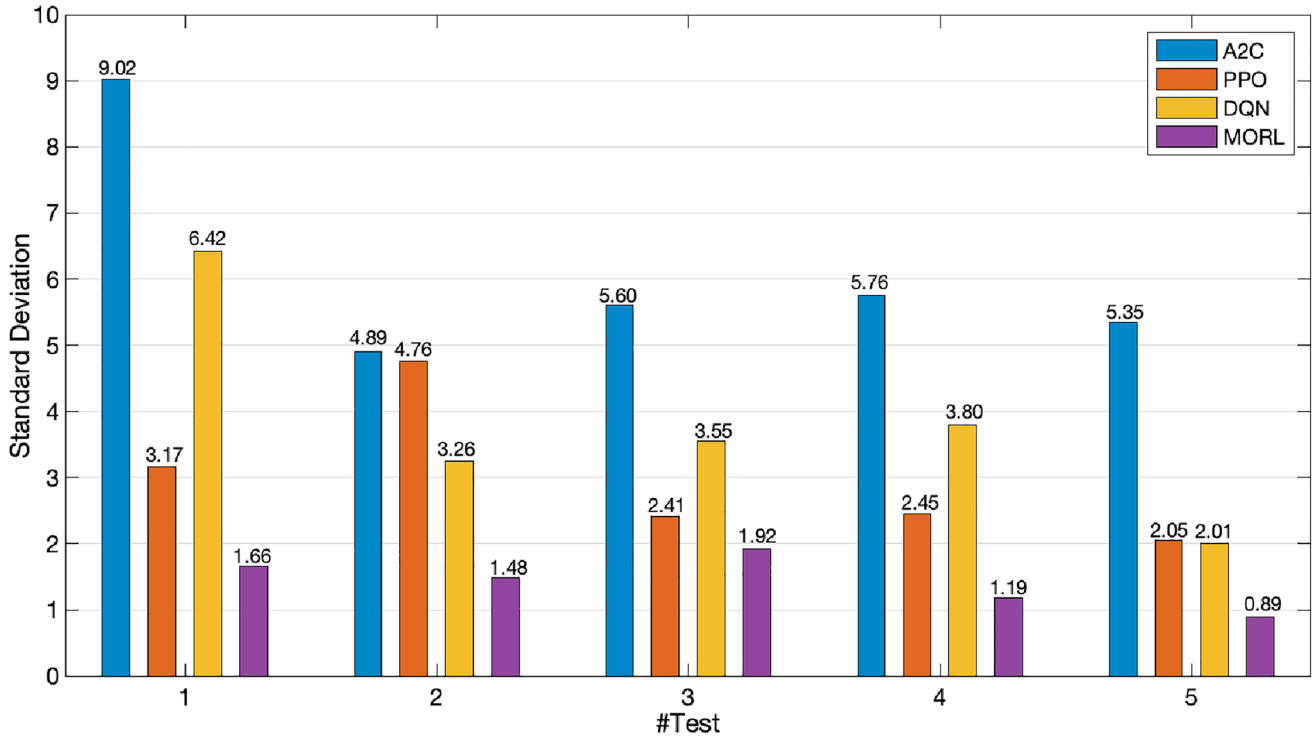


Fig. 15 Standard deviation of IEQ_{Tot} related to the optimal layout configurations

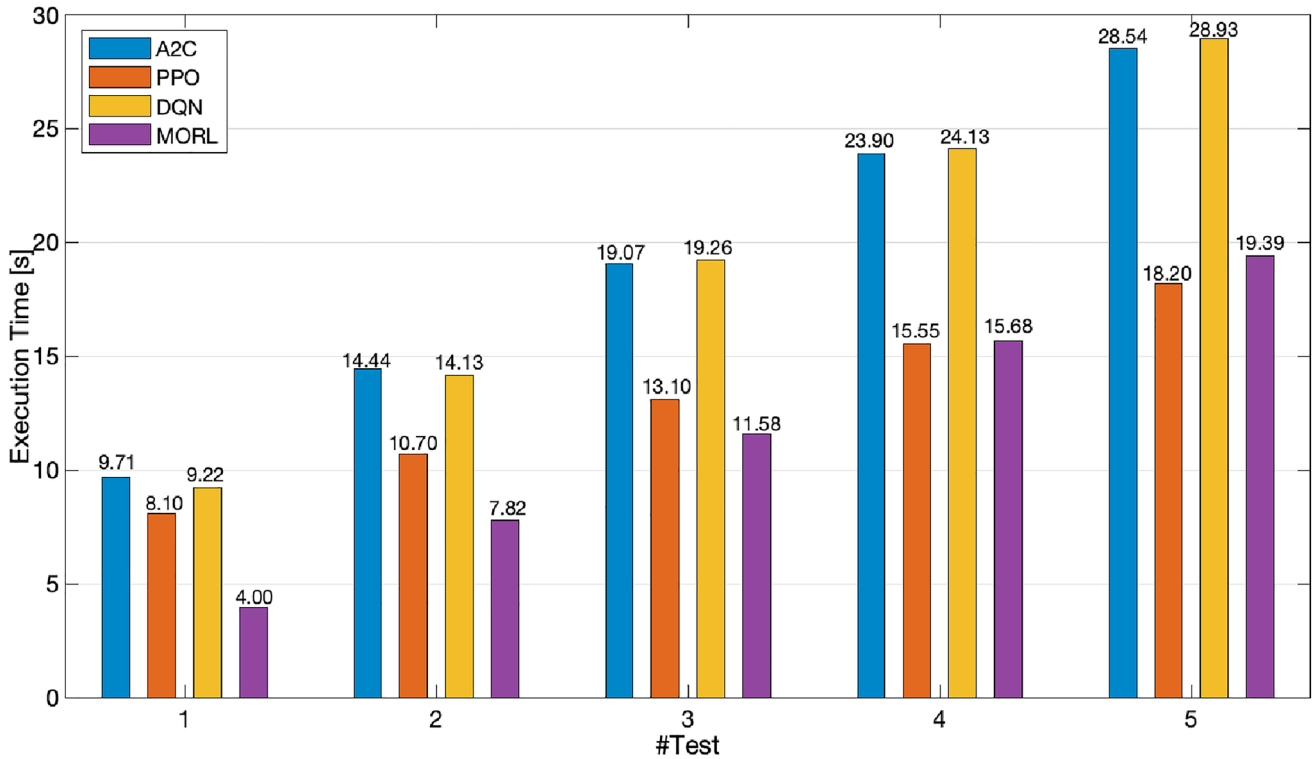


Fig. 16 Execution times

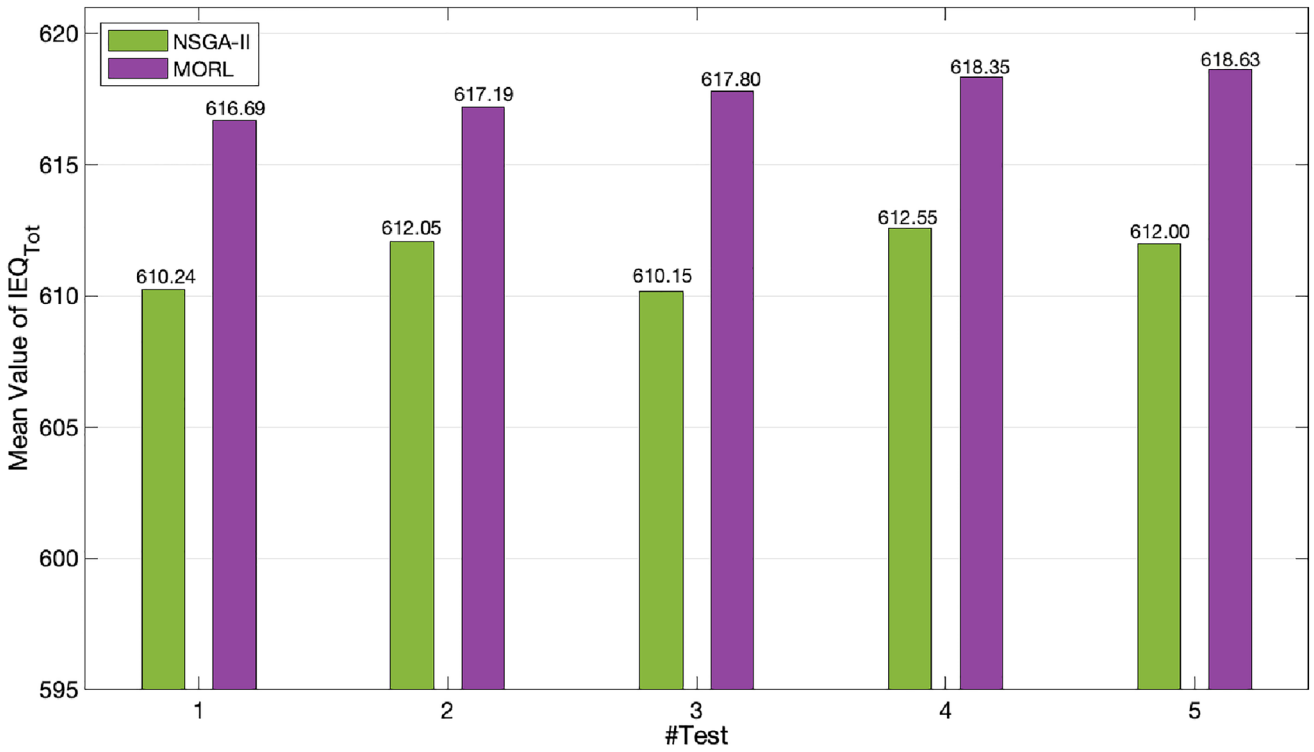


Fig. 17 Mean values of IEQ_{Tot} related to the optimal layout configurations

seconds more with respect to PPO for finding better solutions (see Test #5 in Figs. 14 and 16). The worst-performing algorithm for this problem is the A2C that finds the worst solutions with respect to the other algorithms with a higher variability and execution times in each test. Finally, the results of the comparison indicate that MORL is effective for the problem under study, but an extension of PPO to support multi-objective problems may be promising.

6.2 Comparison with NSGA-II

In this section, the proposed MORL is compared with NSGA-II. Particularly, we conducted five more tests using the NSGA-II algorithm with a population of 20 individuals, varying the number of generations per test (i.e., $1 \times 10^3, 2 \times 10^3, \ldots, 5 \times 10^3$). Also, in this case, the tests have been repeated ten times. Equation (24) was computed for every solution obtained from the NSGA-II Pareto front, and the solutions displaying the highest IEQ_Tot value were selected.

Figures 17, 18 and 19 show the mean values of IEQ_{Tot} related to the optimal layout configurations, the standard deviation and the execution times for both algorithms. Also, in this case, MORL outperforms the NSGA-II in terms of optimal solutions found (Fig. 17) in comparable times (Fig. 19), by also showing a lower variability (Fig. 18). MORL discovers better layout configurations with higher frequency. Additionally, further tests on NSGA-II showed that it could find the same optimal solutions as MORL (i.e., those with $IEQ_{Tot} > 618$) using a population of 50 individuals that evolved over 5000 generations. NSGA-II spent around 105 s to find such solutions, five times more than MORL.

Besides a quantitative comparison of these two approaches, we also want to highlight a difference between our approach and the NSGA-II concerning the search in the solutions space. Indeed, although both MORL and NSGA-II deal with multi-objective optimization, NSGA-II, as a standard multi-objective evolutionary algorithm, does not assign preferences to the objectives and gives them the same importance (Wang et al. 2017). As opposed to the NSGA-II, our algorithm incorporates weights for the objectives, which will lead to solutions located on a preferred area of the Pareto front.

For a layout configuration problem based on environmental comfort, it is more important to give different priorities for each objective since the comfort perception is widely influenced by individuals' preferences (Roskams and Haynes 2021; Castaldo et al. 2018). Adopting the proposed MORL approach, the optimal solutions are directed toward the best

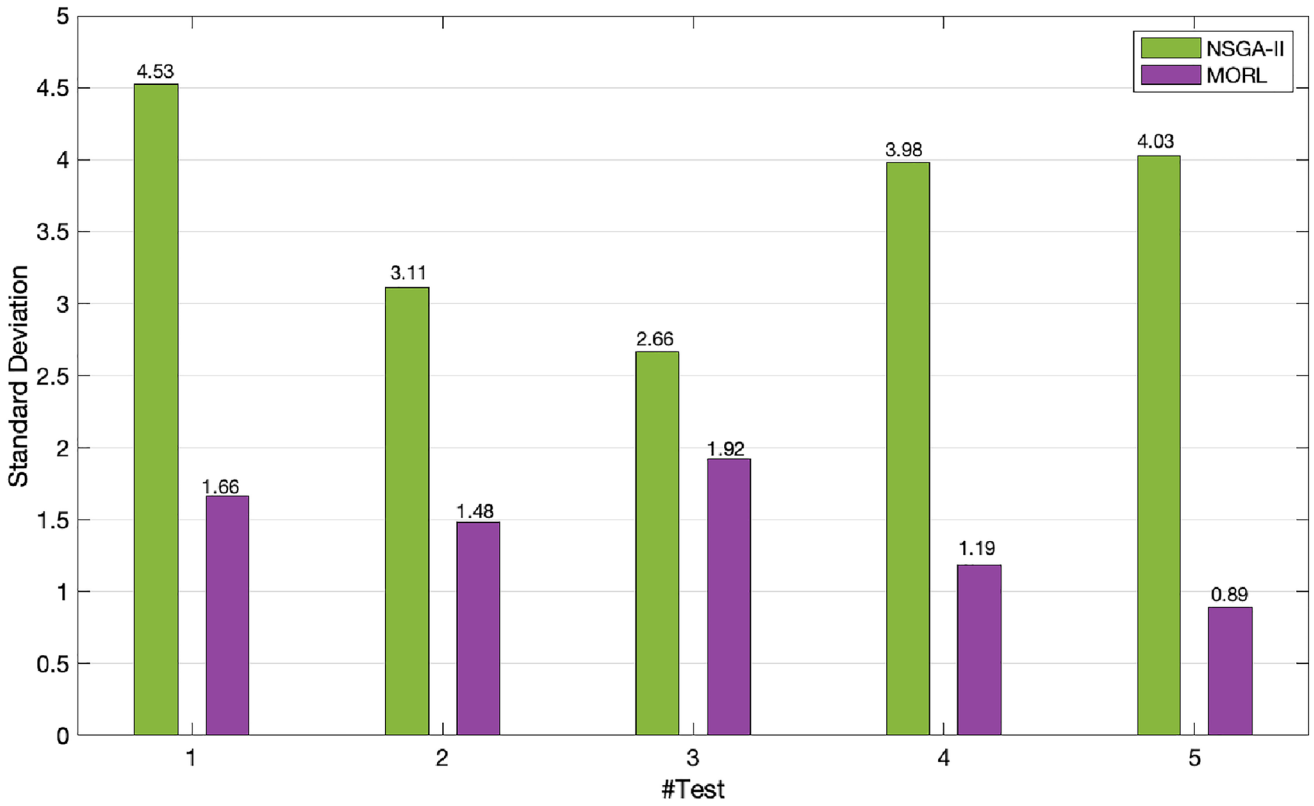


Fig. 18 Standard deviation of IEQ_{Tot} related to the optimal layout configurations

ones that satisfy the required preferences. Indeed, as modeled in Eq. (16), p_{TC}, p_{AC} and p_{VC} are weights related to the user preferences with respect to each kind of comfort, whilst the weights associated with each element of the *Qvector* in Algorithm 3 express the relevance attributed to every single objective of the problem.

Hence, as it is also widely recognized in the literature, for general multi-objective optimization problems, NSGA-II is an effective algorithm. However, for optimizing furniture arrangement in terms of environmental comfort, MORL performs more effectively and supports users preferences.

7 Conclusions

The paper presents an approach for finding furniture arrangements patterns in multi-occupant offices that maximize user satisfaction in terms of indoor environmental quality and functional requirements. The approach is based on multi-objective reinforcement learning that allows an agent to learn optimal solutions that reach a trade-off between office occupants. There are several advantages to applying reinforcement learning. Mainly, since RL algorithms rely on a mathematical MDP framework, they can theoretically guarantee convergence toward an optimal solution. Moreover, an RL agent learns optimal policies directly from interactions with the unknown environment without model definition. It may execute an intensified search by exploitation and a diversified search by exploration, making it an efficient method for various NP-hard problems. An RL agent gradually learns the best (or near-best) strategy based on trial and error through interactions with the environment to improve overall performance. Moreover, the analysis of the results gives evidence that the proposed approach provides advantages not only for finding the best arrangement for multi-occupant offices but also for highlighting possible corrections to improve indoor environmental quality (e.g., reducing noise from devices or improving lighting design). The current prototype is developed to find optimal room configurations with office end-use. However, it can be extended as a generalized tool for dealing with different functional environments and furniture. Finally, we are working on extending the proposed work to address a more complex optimization problem which incorporates the air

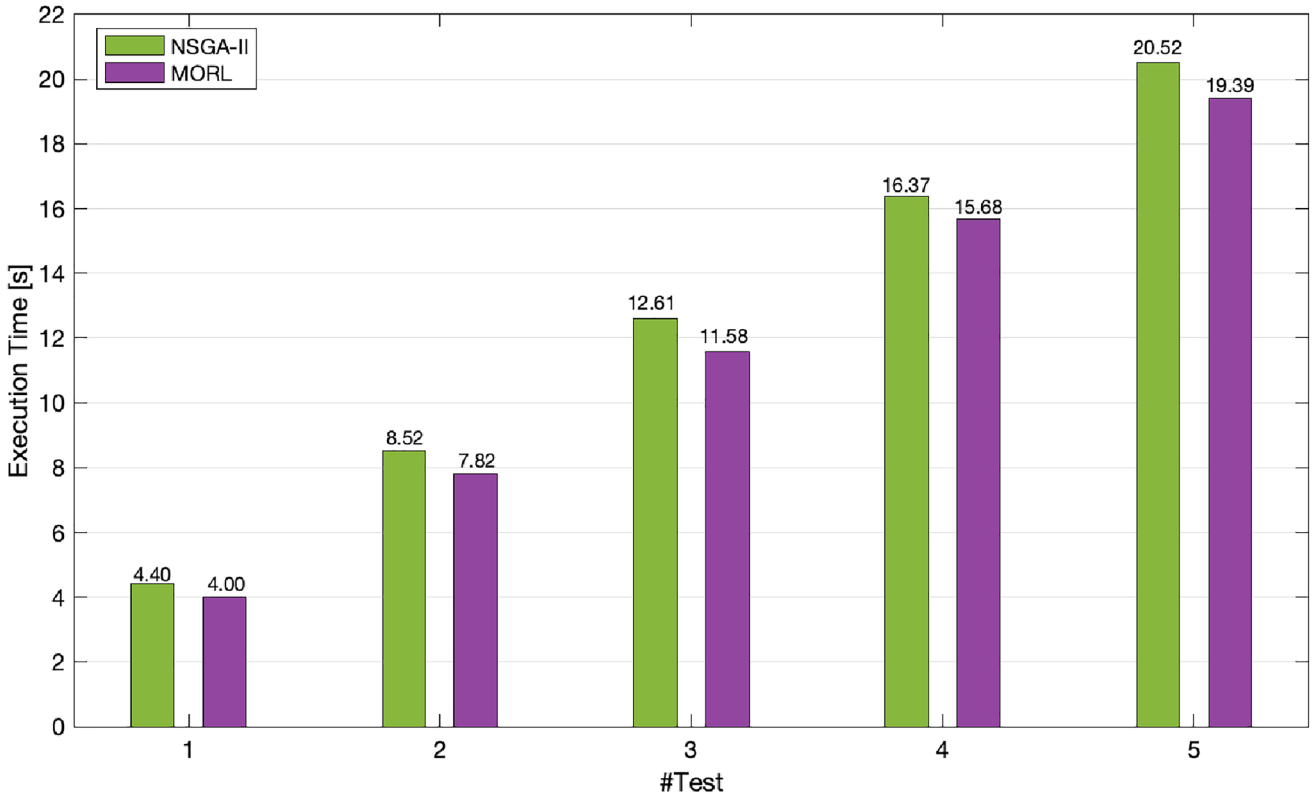


Fig. 19 Execution times

quality factor and other elements that influence visual comfort for spaces with different end-use.

Funding Open access funding provided by ICAR - PALERMO within the CRUI-CARE Agreement.

Data availability All data generated or analyzed during this study are included in this published article.

Declarations

Conflict of 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.

References

Abadi M, Barham P, Chen J, et al (2016) {TensorFlow}: a system for {Large-Scale} machine learning. In: 12th USENIX symposium on operating systems design and implementation (OSDI 16), pp 265–283

Akazawa Y, Okada Y, Niijima K (2006) Interactive learning interface for automatic 3d scene generation. In: 7th International Conference on Intelligent Games and Simulation, GAME-ON 2006, pp 30–35

Arulkumaran K, Deisenroth MP, Brundage M et al (2017) Deep reinforcement learning: a brief survey. IEEE Signal Process Mag 34(6):26–38

Bechikh S, Kessentini M, Said LB et al (2015) Preference incorporation in evolutionary multiobjective optimization: a survey of the state-of-the-art. In: Advances in computers, vol 98. Elsevier, pp 141–207

Beranek LL, Blazier WE, Figwer JJ (1971) Preferred noise criterion (pnc) curves and their application to rooms. J Acoust Soc Am 50(5A):1223–1228

Castaldo VL, Pigliautile I, Rosso F et al (2018) How subjective and non-physical parameters affect occupants' environmental comfort perception. Energy Build 178:107–129

Choi JH, Lee K (2018) Investigation of the feasibility of poe methodology for a modern commercial office building. Build Environ 143:591–604

Colenberg S, Jylhä T, Arkesteijn M (2020) The relationship between interior office space and employee health and well-being–a literature

Deb K, Pratap A, Agarwal S et al (2002) A fast and elitist multiobjective genetic algorithm: Nsga-ii. IEEE Trans Evol Comput 6(2):182–197
Di X, Yu P (2021) Deep reinforcement learning for producing furniture layout in indoor scenes. arXiv preprint arXiv:2101.07462
DIAL (2018) Dialux evo 8.2, professional lighting design software. Available at https://www.dial.de/en/dialux-desktop/
Falkner S, Klein A, Hutter F (2018) Bohb: Robust and efficient hyperparameter optimization at scale. In: International Conference on Machine Learning, PMLR, pp 1437–1446
Fanger PO, et al (1970) Thermal comfort. analysis and applications in environmental engineering. Thermal comfort Analysis and applications in environmental engineering
Fisher M, Ritchie D, Savva M et al (2012) Example-based synthesis of 3d object arrangements. ACM Trans Graph (TOG) 31(6):1–11
Frontczak M, Wargocki P (2011) Literature survey on how different factors influence human comfort in indoor environments. Build Environ 46(4):922–937
Germer T, Schwarz M (2009) Procedural arrangement of furniture for real-time walkthroughs. In: Computer Graphics Forum, Wiley Online Library, pp 2068–2078
Grondman I, Busoniu L, Lopes GA, et al (2012) A survey of actor-critic reinforcement learning: Standard and natural policy gradients. IEEE Trans Syst Man Cybern Part C (Applications and Reviews) 42(6):1291–1307
Guo H, Aviv D, Loyola M et al (2020) On the understanding of the mean radiant temperature within both the indoor and outdoor environment, a critical review. Renew Sustain Energy Rev 117(109):207
Henderson P, Subr K, Ferrari V (2017) Automatic generation of constrained furniture layouts. arXiv e-prints pp arXiv–1711
Kán P, Kaufmann H (2017) Automated interior design using a genetic algorithm. In: Proceedings of the 23rd ACM symposium on virtual reality software and technology, pp 1–10
Kán P, Kaufmann H (2018) Automatic furniture arrangement using greedy cost minimization. In: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), IEEE, pp 491–498
Kuhnle A, Schaarschmidt M, Fricke K (2017) Tensorforce: a tensorflow library for applied reinforcement learning. Web page, https://github.com/tensorforce/tensorforce
Kwon M, Remøy H, Van den Bogaard M (2019) Influential design factors on occupant satisfaction with indoor environment in workplaces. Build Environ 157:356–365
Leaman A, Bordass B (1999) Productivity in buildings: the'killer' variables. Build Res Inf 27(1):4–19
Lewis HB, Bell L (1994) Industrial noise control, fundamentals and applications, New York: M
Li M, Patil AG, Xu K et al (2019) Grains: Generative recursive autoencoders for indoor scenes. ACM Trans Graph (TOG) 38(2):1–16
Li P, Froese TM, Brager G (2018) Post-occupancy evaluation: state-of-the-art analysis and state-of-the-practice review. Build Environ
Liu C, Xu X, Hu D (2014) Multiobjective reinforcement learning: a comprehensive overview. IEEE Trans Syst Man Cybern Syst 45(3):385–398
Ma R, Li H, Zou C et al (2016) Action-driven 3d indoor scene evolution. ACM Trans Graph 35(6):173–1
Marler RT, Arora JS (2010) The weighted sum method for multi-objective optimization: new insights. Struct Multidiscip Optim 41(6):853–862
Merrell P, Schkufza E, Li Z et al (2011) Interactive furniture layout using interior design guidelines. ACM Trans Graph (TOG) 30(4):1–10
Naeem M, Rizvi STH, Coronato A (2020) A gentle introduction to reinforcement learning and its application in different fields. IEEE Access (2020)
Olesen BW, Parsons K (2002) Introduction to thermal comfort standards and to the proposed new version of en iso 7730. Energy Build 34(6):537–548
Pham Ngo Gia B, Tram Loi Q, Quan Thanh T, et al (2018) NSGA-II Python Library. https://github.com/baopng/NSGA-II
Piasecki M, Kostyrko K, Pykacz S (2017) Indoor environmental quality assessment: Part 1: Choice of the indoor environmental quality sub-component models. J Build Phys 41(3):264–289
Ribino P, Bonomolo M (2021) An rl-based approach for ieq optimization in reorganizing interior spaces for home-working. In: intelligent environments 2021: workshop proceedings of the 17th international conference on intelligent environments, IOS Press, p 179
Roskams MJ, Haynes BP (2021) Testing the relationship between objective indoor environment quality and subjective experiences of comfort. Build Res Inf 49(4):387–398
Sabine WC, Egan MD (1994) Collected papers on acoustics
Sanchez S, Roux O, Luga H, et al (2003) Constraint-based 3d-object layout using a genetic algorithm
Sant'Anna D, Dos Santos P, Vianna N et al (2018) Indoor environmental quality perception and users' satisfaction of conventional and green buildings in brazil. Sustain Cities Soc 43:95–110
Schulman J, Wolski F, Dhariwal P, et al (2017) Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347
Seppänen O, Fisk WJ (2003) A conseptual model to estimate cost effectiveness of the indoor environment improvements. In: Healthy Buildings 2003. Singapore 7.-11.12. 2003. p 368–374
Tang R, Li K, Ding W et al (2020) Reference point based multi-objective optimization of reservoir operation: a comparison of three algorithms. Water Resour Manage 34:1005–1020
Tutenel T, Bidarra R, Smelik RM, et al (2009) Rule-based layout solving and its application to procedural interior generation. In: CASA workshop on 3D advanced media in gaming and simulation
UNI E (2011) 12464-1: 2011. Light and lighting Lighting of work places Part 1
Van Otterlo M, Wiering M (2012) Reinforcement learning and markov decision processes. In: Reinforcement learning. Springer, p 3–42
Vimalanathan K, Babu TR (2014) The effect of indoor office environment on the work performance, health and well-being of office workers. J Environ Health Sci Eng 12(1):1–8
Vitsas N, Papaioannou G, Gkaravelis A, et al (2020) Illumination-guided furniture layout optimization. In: Computer Graphics Forum, Wiley Online Library, pp 291–301
Wang H, Liang W, Yu LF (2020) Scene mover: Automatic move planning for scene arrangement by deep reinforcement learning. ACM Trans Graph (TOG) 39(6):1–15
Wang K, Savva M, Chang AX et al (2018) Deep convolutional priors for indoor scene synthesis. ACM Trans Graph (TOG) 37(4):1–14
Wang S, Ali S, Yue T et al (2017) Integrating weight assignment strategies with nsga-ii for supporting user preference multiobjective optimization. IEEE Trans Evol Comput 22(3):378–393
Watkins CJ, Dayan P (1992) Q-learning. Machine learning 8(3–4):279–292
Yamakawa T, Dobashi Y, Okabe M, et al (2017) Computer simulation of furniture layout when moving from one house to another. In: Proceedings of the 33rd Spring Conference on Computer Graphics, pp 1–8
Yu LF, Yeung SK, Tang CK, et al (2011) Make it home: automatic optimization of furniture arrangement. In: ACM Transactions on Graphics (TOG)-Proceedings of ACM SIGGRAPH 2011, v 30,(4), July 2011, article no 86 30(4)
Zhao X, Hu R, Guerrero P et al (2016) Relationship templates for creating scene variations. ACM Trans Graph (TOG) 35(6):1–13