Online Model Adaptation in Monte Carlo Tree Search Planning
Supplementary Material

1 Introduction

This document provides full details about the true environment used in the software attached.

2 True Environment Model

In this section we give full details about the simulator described in Section 'Realistic Simulator of the Environment" of the main paper. The simulator is used as the real-world environment model, i.e., the true one. Taking inspiration from [2, 1] we developed a realistic simulator of the environment to test our framework in silico. We notice that the simulator is inspired by real-world models but it has been extended with elements (e.g., room occupancy and VOC dynamics) useful for our application domain. For those elements we introduced complex transition functions that are interesting from a computational viewpoint but which are not validated with physical experiments in the real-world.

2.1 State

The *state* contains: *i*) number of persons present in the room, *ii)* the concentrations of CO_2 in the room, *iii*) the concentrations of VOC in the room, *iv*) the indoor temperature, and *v*) the outdoor temperature.

2.2 Actions

Actions are related to ventilation system, namely *stack ventilation* (open/close windows) or *mechanical ventilation* (turn on vents at low/high speed or turn off), *sanitization* system (turn on/off sanitizers) and combinations of them. Each action is characterized by a specific value called ACH represented with δ_a (see Table 1), which indicates the number of times the entire room air volume is replaced in one hour with that action. This characteristic allows actions to affect the environment air quality (which is part of the reward function as detailed later) since the air renewal process acts on the concentrations of CO_2 and VOC.

2.3 Transition Model

For each state-action pair the model returns the new state reached by performing the action from the state. The model consists of three *sub-models* used to describe the evolution of CO_2 concentration, VOC concentration and indoor temperature, respectively.

Table 1: Description of domain's actions and the corresponding ACH values

Action	Description	Ventilation	Sanitization	δ_a (h^{-1})
ALL_OFF	Windows closed, all devices off	No	No	0.1
WINDOW	Windows opened, all devices off	Stack	No	8.0
$VENT_L$	Windows closed, only vents on (low speed)	Mechanical	No	11.0
$VENT_H$	Windows closed, only vents on (high speed)	Mechanical	No	21.0
SANITIZER	Windows closed, only sanitizers on	No	Yes	0.1
WINDOW-SANITIZER	Windows opened and sanitizers on	Stack	Yes	8.0
$VENT_L$-SANITIZER	Vents (low speed) and sanitizers on	Mechanical	No	11.0
$VENT_H$-SANITIZER	Vents (high speed) and sanitizers on	Mechanical	No	21.0

2.3.1 Evolution of CO_2 Concentration

It considers both *stack ventilation system*, characterized by actions with $\delta_a < 9.97$ h^{-1}, and *mechanical ventilation systems*, related to actions with $\delta_a \geq 9.97$ h^{-1}.

- *Stack ventilation system in populated room.* We consider the model proposed in [2] that computes the next value of CO_2 concentration ($c^{i+1}_{CO_2}$ (ppm)), as function of the number of persons (p^i), the room volume (V) up to 420 m^3, the ACH (δ_a) and the time (t):

$$c^{i+1}_{CO_2} = A \cdot v \cdot k \cdot t + c^i_{CO_2}, \qquad v = p^i/V \tag{1}$$

$$k = B + C \cdot \delta_a + D \cdot \delta_a^{1.5} + E \cdot \delta_a^2 + F \cdot \delta_a^3 \tag{2}$$

  where $A = 180$ $m^3 \cdot ppm \cdot person^{-1} \cdot min^{-1}$, $t = 5$ min and $B = 1.350$, $C = -1.261$, $D = 0.945$, $E = -0.236$, $F = 0.005$ (every coefficient has its own unit of measure to make k dimensionless). Equation (1) can be used only with *stack ventilation compatible actions*, since Equation (2) has a local minimum at $\delta_a = 9.97$ h^{-1} and with mechanical ventilation compatible actions ($\delta_a \geq 9.97$ h^{-1}) it becomes an increasing function.

- *Stack ventilation system in unpopulated room.* It could be either *active* (actions with $\delta_a > 0.10$) or *inactive* (actions with $\delta_a = 0.10$). In the *active* case we also consider the closeness between the CO_2 concentration in the room and the outdoor one. Then, at time i, we compute

$$\Delta c^i_{CO_2} = c^i_{CO_2} - c^{i,out}_{CO_2}. \tag{3}$$

  In this case the estimated CO_2 concentration is given by Equation (4). If the value of $\Delta c^i_{CO_2}$ is higher than the threshold $\varepsilon_1 = 100$ ppm we compute CO_2 concentration at time $i+1$ as a function of ACH (δ_a) and CO_2 concentration at time i ($c^i_{CO_2}$), otherwise it is expressed in terms of CO_2 concentration at time i ($c^i_{CO_2}$), the difference between the internal and external CO_2 concentration at time i ($\Delta c^i_{CO_2}$) and the threshold (ε_1):

$$c^{i+1}_{CO_2} = \begin{cases} (1 - \frac{\delta_a}{200}) \cdot c^i_{CO_2} & \Delta c^i_{CO_2} > \varepsilon_1 \\ c^i_{CO_2} - \frac{(\Delta c^i_{CO_2})^2}{20 \cdot \varepsilon_1} & otherwise. \end{cases} \tag{4}$$

  In the *inactive case*, instead, the concentration of CO_2 must not change over time, i.e., $c^{i+1}_{CO_2} = c^i_{CO_2}$ (following the observations provided in [2]). Then Equation (1) holds, since $\alpha = 0$ when $p^i = 0$.

- *Mechanical ventilation system.* We introduce a new equation to compute the CO_2 concentration in the case of *mechanical ventilation compatible actions*. Equation (5) computes the value of $c_{CO_2}^{i+1}$ as exponential function of k'. Notice that when $c_{CO_2}^{i+1} < c_{CO_2}^{i,out}$ we always set the value of $c_{CO_2}^{i+1} = c_{CO_2}^{i,out}$.

$$c_{CO_2}^{i+1} = \exp\left(\frac{15.20 - k'}{35.00}\right) \cdot c_{CO_2}^{i} \tag{5}$$

with

$$k' = \begin{cases} m_{k'} \cdot p^i + q_{k'} & \delta_a \in (9.97, 15.00) \\ \frac{p_{max} \cdot \delta_a}{\frac{10.00}{49.00} \cdot p^i + 39.80} & \delta_a \in [15.00, +\infty) \end{cases} \tag{6}$$

The computation of k' (Equation (6)) changes with respect to the ACH (δ_a), since we want to imitate a different evolution for the CO_2 concentration on the basis of the air renewal effect that each action has on the environment.

The values of $m_{k'}$ and $q_{k'}$ are the result of a computation to find the linear equation (i.e., first branch of Equation (6)) given p_{max}, g_{min} and g_{max}. In particular,

$$m_{k'} = \frac{g_{min} - g_{max}}{p_{max}}, \qquad q_{k'} = g_{max} \tag{7}$$

since the two points of the line are (0, g_{max}) and (p_{max}, g_{min}). Then,

$$\begin{aligned} g_{min} &= 14.90000 \\ g_{max} &= 0.36660 * \delta_a + 13.34499 \end{aligned} \tag{8}$$

Moreover, p_{max} in Equations (6) and (7) represents the maximum occupancy of the room, that we assume to be a constant value set to 50 persons. Notice that changing the value of p_{max}, the equations for the mechanical ventilation system do not longer hold. In particular, with $\delta_a = 15.00$, we can observe a relevant jump discontinuity for k' (using the same value of p^i for both cases of Equation (6)), and thus for $c_{CO_2}^{i+1}$ as well. For instance, consider two actions a_1 and a_2 with $\delta_{a_1} = 14.90$ h^{-1} and $\delta_{a_2} = 15.00$ h^{-1}. If we set $p_{max} = 100$ and compute the first case of Equation (6) with δ_{a_1} and the second one with δ_{a_2} (using $p^i = 50$ for both cases), we get $k'_{a_1} = 16.854$ (instead of the original 14.900) and $k'_{a_2} = 29.998$ (instead of the original 14.999) respectively. Thus, moving only 0.1 h^{-1} from δ_{a_1} to δ_{a_2}, we observe the jump discontinuity between k'_{a_1} and k'_{a_2}. This also causes an inconsistency in the computation of $c_{CO_2}^{i+1}$: with k'_{a_1} the term that multiplies $c_{CO_2}^{i}$ in Equation (5) becomes 0.954, while with k'_{a_2} becomes 0.655 .

In equation (6), the numerical terms are not dimensionless because of the need to make k' dimensionless. This also happens in equations (4), (8), (13), (14), (15) and in the expert's model equations which are relative to CO_2 and VOC variation. Moreover, in equations (4) and (5) the term t, representing the time difference between two subsequent time-steps, is not present. The rationale is that we assume to acquire samples every 5 minutes (i.e., $t = 5$ min) and it is implicit in these equations.

2.3.2 Evolution of VOC Concentration

The estimated concentration of VOC (c_{VOC}^{i+1} ($\mu g \cdot m^{-3}$)) depends on the concentration of VOC that each person emits in the room (VOC_p), and the VOC removed by the sanitizer (VOC_r^i), the number of persons (p^i) present in the room at time i, the room volume (V), the VOC concentration (c_{VOC}^i) at time i and the relative variation of CO_2 concentration ($\phi_{c_{CO_2}}$):

$$c_{VOC}^{i+1} = \left(\frac{VOC_p \cdot p^i - VOC_r^i}{V} + c_{VOC}^i \right) \cdot (1 + \phi_{c_{CO_2}}) \tag{9}$$

where

$$VOC_p = \frac{VOC_h}{60 \text{ min} \cdot h^{-1}} \cdot t = 520.83 \ \mu g \cdot person^{-1} \tag{10}$$

with $VOC_h = 6250 \ \mu g \cdot h^{-1} \cdot person^{-1}$ [1], $t = 5$ min and $VOC_r = 10000$ µg that depends on the effectiveness of the sanitizer.

We set $\phi_{c_{CO_2}}$ to 0 when its value is greater than 0, otherwise it corresponds to the ratio between the difference of CO_2 concentration between two subsequent time-steps and the current CO_2 concentration:

$$\phi_{c_{CO_2}} = \begin{cases} 0 & \phi_{c_{CO_2}} > 0 \\ \frac{c_{CO_2}^{i+t} - c_{CO_2}^i}{c_{CO_2}^i} & otherwise. \end{cases} \tag{11}$$

Notice that when $c_{VOC}^{i+1} < c_{VOC}^{i,out}$ we always set the value of $c_{VOC}^{i+1} = c_{VOC}^{i,out}$.

2.3.3 Evolution of Internal Temperature

Depending on ΔT^i, the temperature course can follow a quadratic, linear or constant course. When the action performed by the agent opens the windows, the next value for internal temperature (T_{in}^{i+1} ($^\circ C$)) is computed as a function of the difference between the indoor and the outdoor temperature ($\Delta T^i = T_{out}^i - T_{in}^i$), the heat generated by people in the room (h_p) and by sanitizers in operation (h_s) and the current indoor temperature (T_{in}^i):

$$T_{in}^{i+t} = \begin{cases} j + \frac{h_p + h_s}{4} + T_{in}^i & \Delta T^i > 0 \\ -j + \frac{h_p + h_s}{4} + T_{in}^i & \Delta T^i \leq 0 \end{cases} \tag{12}$$

To compute j we distinguish the following cases on the basis of ΔT^i and the thresholds $\varepsilon_2 = 1.6$ and $\varepsilon_3 = 2.5$. The value of j depends on the difference between the indoor and the outdoor temperature (ΔT^i) and the time difference between two subsequent time-steps (t):

$$j = \begin{cases} 0.080 \cdot (\Delta T^i)^2 \cdot t & |\Delta T^i| < \varepsilon_2 \\ \frac{1.084 \cdot |\Delta T^i| - 0.711}{5} \cdot t & \varepsilon_2 \leq |\Delta T^i| \leq \varepsilon_3 \\ 0.4 \cdot t & |\Delta T^i| > \varepsilon_3 \end{cases} \tag{13}$$

The value of h_p is computed as a function of the number of people currently in the room (p^i), the time difference between two subsequent time-steps (t) and the room volume (V):

$$h_p = \frac{p^i \cdot t}{5 \cdot V} \tag{14}$$

while the value of the heat produced by sanitizers in operation is computed as follows:

$$h_s = \frac{VOC_r \cdot t}{1000 \cdot V} \tag{15}$$

where VOC_r is the VOC removed by sanitizers, t is the time difference between two subsequent time-steps and V is the room volume. Notice that we set h_s to zero when the sanitizers are deactivated.

Otherwise, if the action performed by the agent closes the windows, the value of T_{in}^{i+1} depends on he heat generated by people in the room (h_p) and by sanitizers in operation (h_s) and the current value of the internal temperature (T_{in}^{i})

$$T_{in}^{i+t} = h_p + h_s + T_{in}^{i} \tag{16}$$

where h_p and h_s are computed in the same way as in Equations (14) and (15).

2.3.4 Evolution of Room Occupancy and Outdoor Temperature

We assume their values over time are given by occupation schedule of the room and weather forecast.

2.4 Reward Model

The reward model considers four components: air quality r_q, comfort r_w, energy consumption r_E and the energy factor EF (set to 0.1):

$$r = \frac{r_q + r_w + EF \cdot r_E}{2 + EF}. \tag{17}$$

The air quality component (r_q) is computed on the bases of the presence of persons in the room, and it corresponds to the mean between the reward due to CO_2 concentration ($r_{c_{CO_2}}$) and the reward due to VOC concentration ($r_{c_{VOC}}$) when there are occupants, otherwise it is set to 1.

$$r_q = \begin{cases} \frac{r_{c_{CO_2}} + r_{c_{VOC}}}{2} & p^i > 0 \\ 1 & otherwise. \end{cases} \tag{18}$$

The values of $r_{c_{CO_2}}$ and $r_{c_{VOC}}$ linearly decrease when the values of c_{CO_2} and c_{VOC} are below their maximum acceptable concentration values (i.e., $\varepsilon_4 = 1000$ ppm and $\varepsilon_6 = 600$ $\mu g \cdot m^{-3}$ respectively). Conversely, these rewards quadratically decrease when their values are below the maximum concentration values we suppose the environment could reach (i.e., $\varepsilon_5 = 2500$ ppm and $\varepsilon_7 = 1500$ $\mu g \cdot m^{-3}$ respectively), otherwise they are set to 0. More precisely, we compute $r_{c_{CO_2}}$ and $r_{c_{VOC}}$ as follows:

$$r_{c_{CO_2}} = \begin{cases} m_{CO_2} \cdot c_{CO_2}^i + q_{CO_2} & c_{CO_2}^i < \varepsilon_4 \\ x_{CO_2} \cdot (c_{CO_2}^i)^2 + y_{CO_2} \cdot c_{CO_2}^i + z_{CO_2} & \varepsilon_4 \leq c_{CO_2}^i < \varepsilon_5 \\ 0 & c_{CO_2}^i \geq \varepsilon_5 \end{cases} \tag{19}$$

$$r_{c_{VOC}} = \begin{cases} m_{VOC} \cdot c_{VOC}^i + q_{VOC} & c_{VOC}^i < \varepsilon_6 \\ x_{VOC} \cdot (c_{VOC}^i)^2 + y_{VOC} \cdot c_{VOC}^i + z_{VOC} & \varepsilon_6 \leq c_{VOC}^i < \varepsilon_7 \\ 0 & c_{VOC}^i \geq \varepsilon_7 \end{cases} \tag{20}$$

with

$$m_{CO_2} = \frac{r_{max_{CO_2}} - r_{min_{CO_2}}}{c^{i,out}_{CO_2} - \varepsilon_4}, \qquad q_{CO_2} = r_{max_{CO_2}} - m_{CO_2} \cdot c^{i,out}_{CO_2}. \tag{21}$$

We set $r_{max_{CO_2}} = 1.0$ and $r_{min_{CO_2}} = 0.7$ since we want to find a linear equation that starts at ($c^{i,out}_{CO_2}$, $r_{max_{CO_2}}$) and finishes at (ε_4, $r_{min_{CO_2}}$).

The same method applies to m_{VOC} and q_{VOC} with the two points of the line that are ($c^{i,out}_{VOC}$, $r_{max_{VOC}}$) and (ε_6, $r_{min_{VOC}}$) with $r_{max_{VOC}} = 1$ and $r_{min_{VOC}} = 0.7$.

To model the quadratic course of $r_{c_{CO_2}}$ and $r_{c_{VOC}}$, we assume their curve start respectively at $r_{min_{CO_2}}$ and $r_{min_{VOC}}$ (both set to 0.7) and reach the 0 when $c^{i+1}_{CO_2}$ is equal to ε_5 for CO_2 and when c^{i+1}_{VOC} is equal to ε_6 for VOC. In particular,

$$x_{CO_2} = \frac{r_{min_{CO_2}}}{(\varepsilon_5 - \varepsilon_4)^2}, \qquad y_{CO_2} = \frac{-2 \cdot \varepsilon_5 \cdot r_{min_{CO_2}}}{(\varepsilon_5 - \varepsilon_4)^2}, \qquad z_{CO_2} = \frac{(\varepsilon_5)^2 \cdot r_{min_{CO_2}}}{(\varepsilon_5 - \varepsilon_4)^2} \tag{22}$$

and

$$x_{VOC} = \frac{r_{min_{VOC}}}{(\varepsilon_7 - \varepsilon_6)^2}, \qquad y_{VOC} = \frac{-2 \cdot \varepsilon_7 \cdot r_{min_{VOC}}}{(\varepsilon_7 - \varepsilon_6)^2}, \qquad z_{VOC} = \frac{(\varepsilon_7)^2 \cdot r_{min_{VOC}}}{(\varepsilon_7 - \varepsilon_6)^2} \tag{23}$$

Equations (19) and (20) change depending on the outdoor concentration of CO_2 ($c^{i,out}_{CO_2}$) and the outdoor concentration of VOC ($c^{i,out}_{VOC}$). In our experiments, we set the value of $c^{i,out}_{CO_2}$ to 400 ppm and of $c^{i,out}_{VOC}$ to 30 $\mu g \cdot m^{-3}$ respectively.

The value of the comfort component (r_w) is computed as a weighted mean between temperature and noise rewards when there are persons in the room, while it is set to 1 where there are no occupants.

$$r_w = \begin{cases} \frac{3 \cdot r_T + r_n}{4} & p^i > 0 \\ 1 & p^i = 0 \end{cases} \tag{24}$$

The temperature reward (r_T) is expressed as exponential function of the indoor temperature:

$$r_T = \exp\left(-\left(\frac{T^i_{in} - 20}{5} \right)^2 \right) \tag{25}$$

Finally, the noise reward (r_n) and the value of the energy consumption reward (r_E) are constant values associated to the action performed by the agent as we show in Table 2. The higher the value of these two components (i.e., close to 1) the lower the related energy consumption and the noise pollution.

Table 2: Energy and noise rewards for each action of the environment

Action	r_E	r_n
ALL_OFF	1.0	1.0
WINDOW	1.0	1.0
$VENT_L$	0.7	0.8
$VENT_H$	0.1	0.2
SANITIZER	0.9	0.8
WINDOW-SANITIZER	0.9	0.8
$VENT_L$-SANITIZER	0.6	0.6
$VENT_H$-SANITIZER	0.0	0.0

Notice that the reward, and each of its components (i.e., r_q, r_w, and r_E), has a value in the range [0, 1].

References

[1] Xiaochen Tang, Pawel K. Misztal, William W Nazaroff, and Allen H. Goldstein. Volatile organic compound emissions from humans indoors. *Environmental Science & Technology*, 50(23):12686–12694, 2016. PMID: 27934268.

[2] T Teleszewski and Katarzyna Gładyszewska-Fiedoruk. The concentration of carbon dioxide in conference rooms: a simplified model and experimental verification. *International Journal of Environmental Science and Technology*, 16(12):8031–8040, 2019.