A novel method for daylight harvesting optimization based on lighting simulation and data-driven optimal control

Tullio de Rubeis\textsuperscript{1}, Niko Gentile\textsuperscript{2}, Francesco Smarra\textsuperscript{3}, Alessandro D’Innocenzo\textsuperscript{3}, Dario Ambrosini\textsuperscript{1}, Domenica Paoletti\textsuperscript{1}

\textsuperscript{1} Dept. of Industrial and Information Engineering and Economics, Università degli Studi dell’Aquila, 67100 L’Aquila, Italy
\textsuperscript{2} Dept. of Architecture and the Built Environment, Lund University, 22100 Lund, Sweden
\textsuperscript{3} Dept. of Information Engineering, Computer Science and Mathematics, Università degli Studi dell’Aquila, 67100 L’Aquila, Italy

Abstract

To date, the best daylighting assessment technique is provided by climate-based simulation tools, which require remarkable efforts to create and calibrate realistic models. The data-driven approaches represent an interesting opportunity to support the physics-based modelling. This work proposes a novel method aimed at the optimization of energy use and luminous environment for a set of lighting control system solutions. The method processes experimental data of occupancy and lighting switch on/off events of an individual side-lit office in an academic building at high latitude via DIVA4Rhino; then, the climate-based simulation results provide the data necessary for the data-driven static optimal control that allow different control strategies of the lighting systems according to their lighting power density. The control allows optimal strategies giving priority to either energy saving or luminous environment improvement, depending on the energy efficiency of the lighting installation, while guaranteeing comfort base level. The results show that the method allows to achieve energy savings up to 18.6\% by maintaining high visual comfort levels.

Introduction

Lighting control systems in commercial buildings are ubiquitously promoted as energy saving measure, and building codes sustain such approach. With increasing market penetration of lighting-emitting diodes (LEDs) (U.S. Dept. of Energy, 2016), the average load for functional illumination decreases, while lighting design solutions increase. This leads to two major observations:

- on one hand, if LEDs decrease the load for functional illumination, the use of lighting control systems that actually save energy during ‘on’ lighting states, e.g. daylight harvesting, may lead to marginal savings;
- on the other hand, more design flexibility arises. If energy use is not the main target, lighting can be designed so that other amenities of the visual space are enhanced, e.g. by providing granular dimming or colour tuning.

In regards to the first observation, researches have showed that parasitic uses in the lighting system may account for 25-30\% of the total energy for lighting system with efficient light sources (Roisin et al., 2008; Aghemo et al., 2014). In extreme cases, with adequate daylight provision and low occupancy rates, lighting controls may even offset energy saving (Gentile et al., 2018; Gentile and Dubois, 2017). In such cases, the purpose of lighting control may shift from mere energy saving to improvement of the luminous environment, which is today more possible than even given the intrinsic design flexibility of LEDs.

Therefore, the lighting control system purpose becomes twofold:

- saving energy when loads for functional illumination are high (de Rubeis et al., 2017);
- improving some aspects of the luminous environment when loads for functional illumination are low.

In such context, this work proposes a novel method aimed at the optimization of energy use and the prediction of any indicator of visual comfort for a set of lighting control system solutions. In fact, visual comfort is extremely complex and it does not only depend on ‘hard’ measures, being closely linked to subjective perception (Boyce, 2014). For daylighting and electric lighting integration, a number of dimensions of lighting quality has been identified in a survey of recent literature (Gentile et al., 2016), where luminaire-based metrics seem to be better predictors of comfort (Kruijsselbrink et al., 2018). However, most of current standards and regulation relies on illuminance-based metrics and so are verified the vast majority of existing lighting designs. Therefore, in this paper, for illustrative purposes, we take the simplest case of horizontal illuminance on the work plane as indicator of visual comfort.

The method processes experimental data of occupancy and switch on/off events of the lighting system by means of climate-based simulation and employs these data to implement a data-driven optimization algorithm. The proposed method is applied to an individual side-lit office in an academic building located in Lund, Sweden, with west orientation. The actual occupancy profile of the room and the real operation of the lighting system are logged by a SCADA system. The calibrated models are simulated via DIVA4Rhino. Based on the experimental data and the climate-based simulation results, a data-driven static optimal control, derived following the methodology proposed in Jain et al. (2018), is designed to optimize performance metrics for both energy consumption and luminous environment. The proposed approach introduces two main advantages:

- it particularly exploits his potential when the system has a lot of components interacting in complicated,
nonlinear ways, and assembling a single global model using standard technique can be prohibitive;
- once validated, based on the Lighting Power Density (LPD), it provides the possibility to optimize the energy saving of the lighting system (for high LPD) and/or visual comfort (for low LPD).

The significance of employing the proposed methodology lies on the fact that when dealing with complex systems (e.g. controlling \( m \) working plane illuminance levels with \( n \) luminaires drove by different dimming levels, as in the case of open office rooms), it allows to create models that can be used to setup optimization problems for energy savings and visual comfort.

The results obtained by applying the proposed method to an academic case study showed that the controller is able to provide energy savings while guaranteeing visual comfort. In our simulated testbed we achieved an energy saving up to 18.6%.

**Notation.** We denote by \([T]\) the number of leaves of the tree \( T \).

**Methods**

The method proposed in this work is based on the interaction between experimental data, climate-based lighting simulation, and data-driven modelling. As known, the climate-based simulation, which represents the best tool to assess building’s thermal performance (de Rubeis et al., 2018a), and daylight harvesting maximization (de Rubeis et al., 2018b) and the effects provided by lighting control systems, requires remarkable model creation and calibration efforts. Therefore, the availability of experimental and simulated data allows to create data-driven models that can be used to provide optimal lighting control strategies in a short time.

The proposed method, shown in Figure 1, is divided into four main steps. The first step is represented by the experimental campaign to know the actual occupancy profile of the room, the lighting switch on/off events, and the luminance values of the room’s structural components. Secondly, the 3D model is created via DIVA4Rhino simulation tool considering the weather file of the city where the building is located, the photometric properties of the room, and the actual occupancy profile. Then, the experimental data coupled with the climate-based simulation results are used to create data-driven models to setup the data-driven static optimal control through which different control strategies of the lighting system can be performed.

More precisely, the idea is to use historical data that can be collected into a building, together with the Machine Learning methodology of Regression Trees, to derive a mathematical model to:
- predict the illuminance levels at the working plane;
- employ such models to setup a static optimization problem in order to optimize the dimming level with the aim of either energy saving and/or luminous environment improving, depending on the lighting power density, within intervals of acceptability.

With the last step of the method, the results obtained by the control strategies have been collected in terms of energy savings and luminous performance.

![Figure 1: Block diagram of the proposed method.](image-url)

**Climate-based simulation**

In this work, a side-lit office room of an academic building located in Lund (Sweden) has been modelled in Rhinoceros 3D and simulated via DIVA4Rhino. The building that host the room and the surrounding buildings were considered in the model, although the climate-based simulation was performed only on the considered office room.

For this room, data on both actual occupancy and lighting use (on or off) have been logged by the SCADA system. The daylight modelling followed good practice (Dubois et al., 2019). In particular:
- in order to account for obstructions and reflections, the surrounding buildings have been modelled and their measured reflectances have been used (Figure 2);
- the ground reflectance of main ground materials has been measured and averaged;
• the window mullions, as well as the geometry of the opening, have been modelled in detail (Ibarra and Reinhart, 2013);
• although all surfaces were assumed perfectly diffusive, their reflectance was measured with a reference reflectance plate and luminance meter. Reflectance for lateral walls, which have bookshelves, varies with rooms; in such case, an average of measured reflectances was used.

In detail, the analyzed side-lit individual office room is shown in Figure 3. The room is located at the fourth of five floors of the building and it is used by teaching staff. The room surface is equal to 16.9 m² and it has a single window facing west, with a window-to-wall ratio equal to 30%. The total power installed for the lighting system is equal to 112 W provided by two fluorescent T5 fixtures with 2 x 28 W tubes each. As indicated in the European Standard EN 12464-1 (European standard, 2011) the maintained illuminance target on the work plane is equal to 500 lx. Given the proximity between Copenhagen and Lund (about 40 km) and their similar climatic conditions, the simulations were carried out considering the Copenhagen weather file.

The model has been checked via quick assessment of daylight factors throughout selected rooms. Further information on the model are available on a previous paper (Gentile and Dubois, 2017).

The yearly occupancy profile was logged by a SCADA system with a sampling time of 10 min, on 24/7 basis. The period between 19th January and 16th February turned out to be missing due to problems in the logging system; therefore, it has been hypothesized that during this period the office room would always be unoccupied.

Table 1: Surfaces properties.

<table>
<thead>
<tr>
<th>Surface</th>
<th>Reflectance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office walls</td>
<td>0.80</td>
</tr>
<tr>
<td>Office lateral walls</td>
<td>0.60</td>
</tr>
<tr>
<td>Office floor (grey linoleum)</td>
<td>0.19</td>
</tr>
<tr>
<td>Office ceiling</td>
<td>0.90</td>
</tr>
<tr>
<td>Exterior brick wall</td>
<td>0.28</td>
</tr>
<tr>
<td>Window and door frame</td>
<td>0.88</td>
</tr>
<tr>
<td>Office desk</td>
<td>0.35</td>
</tr>
<tr>
<td>Surrounding building</td>
<td>0.52</td>
</tr>
<tr>
<td>Surrounding roofs</td>
<td>0.20</td>
</tr>
<tr>
<td>Surrounding walls (brick walls)</td>
<td>0.28</td>
</tr>
<tr>
<td>Ground</td>
<td>0.25</td>
</tr>
<tr>
<td>Door glazing (transmittance)</td>
<td>0.90</td>
</tr>
<tr>
<td>Window glazing (transmittance)</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Based on the measured data, the office room resulted occupied for 561.8 h/yr (i.e. 6.41% of the total 8760 h/yr); the switch on/off events revealed that the lighting system was turned on 1.45% of the year (i.e. 127.2 h/yr), namely the 22.63% of the occupied time. Moreover, for 16.3 h/yr (0.19% of the year) the lighting system was switched on, although the room was unoccupied. For a better understanding of the actual occupancy profile of the office room, see Figure 4. Since the simulation software uses an hourly occupancy schedule, some simplifications have been made: if the room was occupied less than 30 min, then it was considered unoccupied; instead, if it was occupied more than 30 min, an hourly occupancy was appraised.
The modelled lighting system was controlled by an occupancy detector and a daylight dimming controller, with an overall standby power equal to 1 W.

The simulations were performed by hypothesizing two measuring grids (at work plane and ceiling) whose nodes’ distance has been defined following the European Standard EN 12464-1.

The climate-based simulation at the work plane eight provided the results summarized in Table 2.

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
<th>U.M.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daylight Autonomy (DA900)</td>
<td>43%</td>
<td>-</td>
</tr>
<tr>
<td>Continuous DA (cDA900)</td>
<td>63%</td>
<td>-</td>
</tr>
<tr>
<td>Mean Daylight Factor (DF)</td>
<td>1.8%</td>
<td>-</td>
</tr>
<tr>
<td>Energy Consumption (lighting system)</td>
<td>10.6 kWh/yr</td>
<td></td>
</tr>
<tr>
<td>Energy Consumption (standby power)</td>
<td>0.2 kWh/yr</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Climate-based simulation results.

It is worth noting that the total energy consumption is very low due to the low lighting power density (6.6 W/m²), the limited time of room occupancy (6.41% of the entire year), and the good daylight provision. In fact, the limited use of the lighting system is highlighted also by the simulated dimming schedule, shown in Figure 5.

**Figure 5: Dimming schedule of the lighting system.**

**Data-driven optimal control with Regression Trees**

In this section we introduce a data-driven framework to optimally control dimming level in order to optimize both the energy consumption and the visual comfort inside a building. For the sake of simplicity, we state the problem formulation for one room only, but it can be easily extended to multiple rooms. The idea is to modify the construction algorithm of a well-known machine learning methodology, the Regression Trees, to obtain a data-driven model that can be used in a static optimal control problem formulation. Such model is built upon historical datasets that are created collecting data from the sensors inside the room (or data provided by climate-based simulations). In particular, this section is divided in two subsections:

1. identification: where we show how to modify the Regression Trees learning methodology in order to obtain control-oriented models. To this aim we follow the procedure already introduced in Behl et al. (2016), and Jain et al. (2018);
2. control: where we setup the static optimal control problem.

**Model identification using Regression Trees**

Let an historical dataset $D = \{X, Y\}$ be given, composed by data collected from the sensors during normal operation of the building. The dataset is composed by $n$ samples. In particular:

- $Y = \{y_k\}_{k=1,...,n}$ is the set of data associated to the variables (outputs) we want to predict, i.e. the illuminance target values on the working plane $y_k \in \mathbb{R}^p$;
- $X = \{d_k, u_k\}_{k=1,...,n}$ is the set of data (features) associated to the variables we can directly measure, i.e. solar radiation and illuminance values at the ceiling $d_k \in \mathbb{R}^q$, and to the variables we can control, i.e. the dimming levels $u_k \in \mathbb{R}^m$.

The dataset $D$ is used to train a regression tree, so that it can be used as a model in run time to predict the response $y$ given the current measurements of $d$ and $u$.

In this paper we limit our considerations to the case of a single room, as described above. As output variable we limit our prediction to a single point of the measuring plane, so $p = 1$. Furthermore, as features of $d$ we consider a single illuminance node on the ceiling, and the solar radiation measurement, so $q = 2$. Since in the room there are two luminaires, but with only one control signal, we have that $m = 1$.

Our goal is to learn, using Regression Trees, a model of the following form

$$\hat{y}_k = f(d_k, u_k)$$

that relates the value of the response variables with the value of the features, so that we can use it in an optimal control problem. To this aim, the main issue is that the classical Regression Trees algorithm does not allow to obtain a closed-form expression of the function $f$, so it cannot be used in an optimal control problem. This is because the regression tree $T$ is obtained partitioning the dataset into smaller regions $\mathcal{R}_i, i = 1, ..., |T|$ (the leaves of the tree), and assigning to each region the average value of the output part of the samples contained in it. Consequently, the prediction provided by the tree is a constant (see Breiman et al., 1984, or more briefly Jain et al., 2018 for further details on how to train a regression tree).

A solution to this issue follows the procedure described in Jain et al. (2018), and consists in assigning to each region an equation as (1). More precisely, the process is divided in 3 steps:

1. we split the dataset $X$ into to 2 subsets: $X_d = \{d_k\}_{k=1,...,n}$, the dataset containing non-control variables, and $X_c = \{u_k\}_{k=1,...,n}$, the dataset containing control variables;
2. we train a regression tree $T$ using the dataset $D_d = \{X_d, Y\}$ (it will be clear in the next section why this choice is necessary to make our model suitable for control);
3. we associate to each region $R_i$ of $T$ an affine model of the form

$$\hat{y}_k = \beta_0^i + \beta_1^i u_k$$

(2)
that relates the lighting level $y_k$ to the dimming level $u_k$. Parameters $\beta_0$ and $\beta_1$ are obtained solving a classical least square problem using data from $X_k$ associated to the samples of each region. A graphical representation of this procedure is depicted in Figure 6.

![Figure 6: Regression tree structure and leaf model derived using the dataset split procedure defined above.](image)

**Optimal control using Regression Trees**

Once the Regression Trees-based model is created, model (2) can be used to setup an optimal control problem to control dimming level while saving energy and guaranteeing visual comfort. In particular, the following problem can be setup:

$$\min_{u_k} E_{\text{max}} u_k + \rho \| \hat{y}_k - y_r \|^2$$

subject to:

$$\hat{y}_k = \beta_0 + \beta_1 u_k$$

$$0 \leq u_k \leq 1$$

$$\hat{y}_k \geq y_{\text{min}}.$$  

The variable $y_r$ represents the reference illuminance value that we want to have in the room. The weight $\rho$ is used to decide if we want to give more importance to the visual comfort or to the energy saving. So the objective is to minimize the weighted sum of the energy consumption and the visual comfort, subject to the fact that the dimming level has to be bounded by 0 (switched off) and 1 (maximum power), and that the (estimated) illuminance level $\hat{y}_k$ cannot be lower than a minimum comfort threshold $y_{\text{min}}$. $E_{\text{max}}$ is the energy consumption of the light when the dimming level is 1. Clearly, different optimization functions can be chosen leveraging this data-driven modelling framework.

At each time step $k = 0, 1, 2, \ldots$, we:

1. get the measurement of $d_k$ from the sensors;
2. narrow down the leaf of the tree to get model (2);
3. solve the static optimization problem (3);
4. apply the optimal dimming level $u^*_k$ provided by (3);
5. repeat the process at $k + 1$.

**Simulation results**

In this section we first provide results concerning data-driven identification and describe the procedure used to derive the model, and then we discuss the setup of the optimal control strategy showing the closed-loop simulation results.

**Identification**

In order to derive the illuminance model at the working plane we first build a dataset $D = \{X_d, X_c, Y\}$, and then we use it and the procedure introduced above to derive a data-driven model as the one depicted in Figure 6 and described by Equation (2). The features in $X_d$ are the solar radiation and the illuminance at the ceiling, the feature in $X_c$ is the dimming level, and the output to predict features in $Y$ is the illuminance at the working plane. The historical data composing the dataset $D$ have been generated using the DIVA4Rhino simulation model above described, with a sampling time of 1 hour. In particular, the results provided by the node placed at the central point of the measuring grid have been employed (Figure 7).

![Figure 7: Measuring grid (the node selected to provide the data useful for the data-driven optimization is indicated with a red star).](image)

We split the dataset $D$ into a training dataset containing all the samples from January to November, and a validation dataset containing all the samples of December. The training dataset is used to derive the data-driven model, while the testing dataset is used to validate it.

**Remark:** although one year of data has been considered to train the data-driven model, the amount of data necessary can be much less. In particular, the size of the dataset is related to the amount of information that it contains, e.g. one week of data that covers a wide variety of situations is a good dataset. Furthermore, once the model is created, also using few data, it can be updated as more data are generated.

Validation results, obtained by comparing the prediction of the data-driven model with the illuminance values at the working plane, are plotted in Figure 8 showing the good quality of the prediction. In particular, let $n_s$ be the number of samples in the testing dataset, the Root Mean Square Error (RMSE), computed as

$$\text{RMSE} = \sqrt{\frac{1}{n_s} \sum_{i=1}^{n_s} (y_{\text{obs}}(i) - y_{\text{pred}}(i))^2}.$$
RMSE = \sqrt{\frac{1}{n_2} \sum_{i=1}^{n_2} (y_i - \hat{y}_i)^2}, \tag{4}

between the real \( y \) and the predicted \( \hat{y} \) trajectories is equal to 43.32 lux, while the Normalized RMSE (NRMSE) in percentage, computed as

\[
NRMSE = \frac{\text{RMSE}}{\max(y) - \min(y)} \times 100, \tag{5}
\]

is equal to 7.99%.

**Figure 8: Validation results on testing data from December.**

**Closed-loop control**

Provided the models identified in the previous section, here we run closed-loop optimal control simulations that implement the control problem (3). We defined an illuminance target value equal to \( y_r = 550 \text{ lux} \) and a minimum illuminance value \( y_{\text{min}} = 400 \text{ lux} \). The illuminance target value differs from the one previously used in the paper (500 lux) to emphasize the effect the control can have in managing the trade-off between energy saving and visual comfort. We ran simulations for 3 different levels of \( \rho \), i.e. \( \rho = 10 \), \( \rho = 10^{-3} \), and \( \rho = 10^{-6} \), to highlight that this problem formulation allows different levels of visual comfort at the cost of energy consumption depending on the lighting power density. In particular, the weight \( \rho \) is defined in three different conditions:

- \( \rho = 10^{-6} \) for high lighting power density (e.g. 12 W/m²) to give more importance to energy consumption minimization;
- \( \rho = 10 \) for low lighting power density (e.g. 4 W/m²) to privilege visual comfort;
- \( \rho = 10^{-3} \) a trade-off between the two boundary cases. These categories were chosen after consultation of a large worldwide survey of existing installations (Dubois & Gentile 2016) and expected efficiency of retrofitted lighting systems (Dubois et al., 2015).

**Remark:** by visual comfort we mean a certain predefined visual illuminance level that can be comfortable for the occupants, and that can be setup as set point. It is out of the scope of this paper to consider more complex model for the comfort, although it can be taken into account since the optimal control problem formulation (3) can be generalized in that sense.

The optimal control was active only when the room was occupied, while it was automatically turned off when the room was empty. Occupancy and disturbance data have been taken directly from the testing dataset.

The results of the simulations are shown in Figures 9 and 10. A simulation period of 4 days, from December 13th to December 17th, has been chosen, since that was the busiest period in December, i.e. when the room resulted to be more occupied (see Figure 5). In particular, the top graph of Figure 9 shows the predicted illuminance values on the working plane. Although on December 14th and 15th at noon the room was empty and so the lighting system was switched off, the high illuminance levels are due to the external solar radiation. Moreover, Figure 9 reveals how the illuminance values go farther from the illuminance target value (550 lux) as the value of \( \rho \) increases.

**Figure 9: Controlled illuminance levels (top), and dimming levels applied for different values of \( \rho \) (bottom), with respect to occupancy profile from December 13th to December 17th.**

As a consequence, as depicted in Figure 10, it is possible to appreciate the energy saving due to the lower dimming levels applied into the room when a higher lighting power density is hypothesized (Figure 9, bottom graph).

**Figure 10: Energy consumption for different values of \( \rho \), from December 13th to December 17th.**

Over the 19 occupied hours of the lighting system (i.e. the occupied hours during the considered period of 4 days) it is worth noting to observe that:
• when the dimming control is on the choice of the lower $\rho$ ($\rho = 10^{-6}$), i.e. the energy saving represents the main goal, the energy consumption resulted equal to 1492 Wh;
• when the dimming control has the highest value of $\rho$ ($\rho = 10$), i.e. the visual comfort is privileged, the energy consumption resulted equal to 1832 Wh and the maintained illuminance is closer to the illumination target;
• the control strategies provided by the data-driven optimization, changeable with respect to the lighting power density, allow energy savings up to 18.6% (i.e. 340 Wh).

Conclusion

In this paper, a novel method aimed at the optimization of energy usage and luminous environment for a set of lighting control system solutions is proposed. The method, based on experimental data, calibrated climate-based lighting simulation, and data-driven optimization, has been tested on an actual side-lit office room in an academic building at high latitude.

The results obtained have shown that the data-driven optimization allows different control strategies as a function of the lighting power density. In fact, with high lighting power density (e.g. 12 W/m$^2$) the controller favors the energy saving, while with low lighting power density (e.g. 4 W/m$^2$) visual comfort is optimized maintaining illuminance values on the working plane very close to the assigned illumination target.

In particular, in the testbed simulated in this work, the impact of the trade-off between energy savings and visual comfort results in an energy saving up to 18.6% at the price of going far from the visual comfort reference, and was obtained by comparing different lighting power densities and the resulting control strategies.

Although the proposed method has been used for the study of a simplified case study, its application to a more complex case would exploit its potential.

Therefore, in future work the following aspects will be analyzed:
• considering multiple nodes on the work plane, also with furniture, and multiple luminaires controlled by different control signals;
• extension to Random Forests as in Jain et al., (2017) and Smarra et al. (2018);
• extend the case study setup to take into account sub-hourly profiles for occupancy.

Acknowledgement

Eng. T. de Rubeis thanks the Fondazione Ferdinando Faluaro that partially funded his work in this research activity. This work was supported partially by the Italian Government under Cipe resolution n. 135 (Dec. 21, 2012), project INNovating City Planning through Information and Communication Technologies (INCIPICT).

References


