



International Workshop on Enabling ICT for Smart Buildings (ICT-SB 2014)

## A multi-sensor office-building database for experimental validation and advanced control algorithm development

N. Zarkadis<sup>a,\*</sup>, A. Ridi<sup>b</sup>, N. Morel<sup>a</sup>

<sup>a</sup>Laboratoire d'Énergie Solaire et de Physique du Bâtiment (LESO-PB), Ecole Polytechnique  
Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

<sup>b</sup>Department iCoSys, University of Applied Sciences of Western Switzerland, Boulevard de Pérolles 80, 1700 Fribourg, Switzerland

---

### Abstract

Data from multiple sensors and actuators has been recorded from December 2001 to March 2008 at the LESO-PB building of the EPFL campus, using the EIB/KNX bus standard. Sensors and actuators provided data on room temperature, presence, lighting level, windows opening, blinds position, electric lights and heating power. Weather data has also been collected and includes ambient temperature, solar radiation on a horizontal surface (direct and diffuse components), wind speed and direction and rain alarm. The data was written continuously to a MySQL database, each EIB/KNX telegram being recorded into the database. The primary goal of that setup was the experimental validation and checking of advanced control algorithms for the actuation of solar shadings (outside fabric blinds), electric lighting and heating equipment. With the aid of the database it was shown that such control algorithms were able to reduce significantly the energy consumption while keeping the same comfort or even improving it. Since 2002, the database has been used in various research projects carried out in LESO-PB. As of 2014 the database has been made freely accessible to the scientific community as a tool to work on.

© 2014 Published by Elsevier B.V. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).  
Selection and Peer-review under responsibility of the Program Chairs.

*Keywords:* smart building database; office building; sensors; actuators; machine learning algorithms development & testing; public domain

---

\* Corresponding author. Tel.: +41 21 693 57 14; fax: +41 21 693 27 22.  
E-mail address: [nikos.zarkadis@epfl.ch](mailto:nikos.zarkadis@epfl.ch)

## 1. Introduction

Nowadays, in the energy consumption domain, buildings have a great impact from the economic and environmental points of view. Energy consumption of the building sector is estimated to be between 20% and 40% of the total energy use. Buildings incorporate numerous sub-systems that each has a specific function and contributes differently to the total energy consumption. In office buildings about the 50% of energy consumption is due to the heating, ventilation, air conditioning (HVAC) system while other sources of consumption like lighting, appliances and domestic hot water play a major role.<sup>1</sup>

Smart systems inside the building environment can achieve energy saving while preserving the human comfort. This “intelligence” can be mainly provided through mathematical or data-driven models. In first case mathematical equations are derived from the physics of the building. Information on the building, its structure, orientation and materials are generally needed as parameters. In the second case, models are generated purely from data without necessarily a relation with the building physics. Sensors and actuators distributed in the building provide data to the models, which generally use machine learning techniques. The energy saving is fulfilled by “intelligent” decisions applied through actuators. As a backdrop, sensors, actuators and a control unit have to be installed. Depending on the communication protocol chosen, these nodes can be wire connected, increasing the installation time and cost.

In many cases scientific researchers are interested in testing machine learning algorithms without recording data. As a consequence, already-existing building databases are their solution, but only few of them are publicly available and mostly oriented to the activity recognition.<sup>2, 3, 4</sup> Moreover, databases are usually recorded during a small period of time (in the order of months), which is a short time compared to building’s life cycle. When this is the case, seasonal effects that are important for some sub-systems (e.g. the HVAC system) cannot be detected or verified.

In this paper we present a publicly available database. The database is filled with telegrams coming from sensors and actuators recorded continuously for more than 6 years at the LESO-PB building of the EPFL campus. The EIB/KNX bus standard has been used for the data transfer. The LESO-PB database contains information about *air temperature, presence, electric lighting, illuminance, window opening, blind position, electric heating and weather data*. Models and machine learning algorithms on the visual comfort, the occupant presence and behavior and on the season identification and prediction have already been realized and are presented herein. Researchers will be able to compare, analyse and test existing models as well as develop new models and algorithms.

In Section 2 the LESO-PB building, the sensors and actuators are presented. More information on the structure of LESO-PB database is detailed in Section 3. A selection of research work performed using LESO-PB database is provided in Section 4 while the Section 5 concludes the paper.

## 2. LESO-PB office building

In this Section a short description of the LESO-PB building is given, along with some details regarding the installed sensors and actuators whose signals are recorded in the LESO-PB database.

### 2.1. Building description

LESO-PB is a 3-floor, 20-room building. Offices are mostly south-oriented and about half of them have a single user and the other half two or three occupants. The south facade of the building integrates advanced anidolic daylighting systems that efficiently redistribute daylight in the offices.<sup>5</sup> LESO’s construction is a heavy one with thick walls and substantial thermal mass. All windows have a wooden frame and are double glazed with IR coating, U-value  $1.4\text{W/m}^2\cdot\text{K}$  and g-value 0.54. The windows of all south-looking offices are protected by two external textile blinds, one for the normal window (lower) and one for the anidolic one (upper). The building features no active

cooling or ventilation system and it is naturally ventilated by a stack effect. A detailed description of the building, including an exhaustive analysis of the building's energy flows, is given by Altherr and Gay<sup>6</sup>. Floor drawings of the three building floors are shown in Figure 1.



Fig. 1. LESO-PB building: (a) ground floor; (b) first floor; (c) second floor; (d) southern façade.

## 2.2. Sensors and actuators

For each room, the following sensor and actuator data are available:

- **Air temperature** (integrated in the control box for the heating and the electric lighting); it measures a weighted average between air temperature and wall temperature (due to the fact that the sensor itself is contained in the box, located on the room wall).
- **Presence** (infra-red sensor mounted on the ceiling).
- **Lighting level** (actually, the sensor is a conventional luminance sensor mounted on the ceiling and looking downwards, measuring the luminance in a given cone; it is used as a illuminance sensor, but this is only valid if the objects in the sensing cone are not too dark or too bright).
- **Window opening** (on/off sensors, one for each openable window).
- **Blind position**: the blinds position is not directly recorded; instead, the basic commands to actuators, i.e. “blind up”, “blind down”, “stop”, are recorded as EIB/KNX telegrams. From these commands and the respective timestamps the actual position is calculated and inserted into the database.
- **Heating**: electric radiators are controlled by an on/off controller, using a pulse width modulation for implementing a proportional controller, with a cycle time of several minutes; like for the blinds, the significant variable, i.e. the average heating power, is not directly available, but only the elementary commands on and off with the respective timestamps, allowing to reconstruct the heating power from the on/off commands.
- **Electric lighting data** (on/off and dimming status).

Additionally, **weather data** (for the whole building) is also recorded: ambient temperature, solar radiation on a horizontal surface (direct and diffuse components), global horizontal illuminance, wind speed and direction and rain alarm.

### 3. The LESO-PB database

#### 3.1. Data acquisition

The LESO building is equipped with a commercial building management system that can read out the status of the building's sensors and actuators, and send commands to the latter. It was installed in 1999. As of August 2004, 240 sensors and actuators were on this system. *Eibserver* is a Java program that allows a PC connected to the EIB system via its serial port to listen to, and send, EIB telegrams. It keeps in memory at all times the complete known state of the LESO building ; when any variable is modified, that change is committed to memory and logged to disk.

The Eibserver program logs all bus events to disk and a Perl script that starts every day at 5 a.m., reads in the logfiles that have been written during the last 24 hours and inserts the corresponding values in a MySQL database. The MySQL database includes all the data recorded by Eibserver since it began recording data in 2001.

Table 1. LESO-DB database tables structure.

Field name	Description	MySQL field format
index	primary index	int(11)
date	date stamp of EIB telegram	date
time	time stamp of EIB telegram	time
room	room number of the device generating the telegram	char(3)
type	source of the event (user, controller, sensor)	varchar(5)
device	name of the device considered	varchar(20)
action	kind of event (sensor, actuator command, etc.)	varchar(5)
<i>data</i>	numeric data depending on the device considered	double

#### 3.2. Data format and structure

The database is made up of ten tables (sharing an identical structure), corresponding to the ten categories of data that Eibserver records. These tables and their schema are described in detail on the website: <http://www.wattict.com/>, where also public access to the database can be obtained for members of the scientific community. It should be noted that although the data spans the date interval 18 December 2001 - 5 March 2008, not all the tables are covering the whole interval. Moreover, some crashes of the data acquisition computers caused some intervals to be empty. The effective data interval varies from table to table.

Concerning the origin of the data passed on to the database, it is always possible to distinguish controller commands from user commands. In particular, by looking at the prefix of the device's name ("type" field) one knows whether it was a user-initiated action or a command sent through Eibserver. The database columns (fields) have always the same meaning, except for the table "useraction" and are described in Table 1.

### 4. Selected research performed with the LESO-PB database

In this section a selection of research work examples based on the LESO-PB database is presented. The diversity of the developed models and machine learning algorithms on visual comfort, occupant presence and behavior and on the season identification and prediction clearly demonstrate the huge potential and the versatility of the LESO-PB database.

#### 4.1. Bayesian optimization of visual comfort

An intelligent controller for blinds and electric lighting for small office rooms that optimizes the visual comfort by maintaining the workplane illuminance levels inside acceptable levels was proposed by Lindeloef<sup>7</sup>. This study was based on the application of the Bayesian theorem<sup>8</sup>: the light conditions in a room are modified by the occupant's interaction with the electric lights and/or with the blinds, when the occupant experiences visual discomfort. It is then assumed that by the end of the occupant's actions the visual comfort has been restored. Following this principle, Lindeloef studied the interactions of the user with the electric lighting system and the daylighting system (blinds and solar shadings) recorded in the database. Similar actions recorded within the same minute are considered as a single user action. Actions performed less than two minutes before the user's departure from the room are ignored.

Additional recorded information included outdoor irradiance, user presence, window openings, indoor and outdoor temperatures. A daylight illuminance model for office rooms was built and combined with the user's visual discomfort probability, inferred from the user interactions with the building's systems described above. Based on these models, a user-adaptive and self-commissioned control algorithm has been developed which controls automatically the solar shadings and/or the electric lighting providing the best possible lighting conditions. Results have shown that the visual comfort is improved when compared to a state of the art non-adaptive controller while energy savings can be achieved in office rooms when adequate direct solar gains are available.

#### 4.2. Simulating Occupant Presence and Behaviour in Buildings

Two models were developed by Page<sup>9</sup> in the framework of his research using the LESO-PB database. First, a model of occupant presence and behaviour modelling with the use of Markov chains was elaborated. Based on data recorded from the presence sensors, personalized patterns of occupancy were produced which can reproduce quite accurately one's presence in a building. This model can be used as a useful input to supplementary models (such as an appliance use model) and predict energy consumption patterns. Regarding the data recorded in LESO-DB database, the development of the occupancy model required the following treatment steps (that could serve as a proposed protocol for future research):

- Removal of days with gaps in data acquisition (related to problems either in the sensor, the bus or the server).
- Removal of absence periods of less than 2 minutes ("noise" that usually corresponds to a user staying almost still for more than 30s).
- The resulting time series were then presented with time-steps of presence or absence of 15 minutes (by allocating to that time-step the state with the longest total duration during the 15-minute period).
- Classification of absence periods into "short absence" (less than a day) and "long absence" (more than a day). Both are used as inputs to the model but differently: long absences correspond usually to weekends or vacation time (the timely occurrence of which has to be nonetheless predicted) and are not considered for the day-to-day occupancy model elaboration.

A second model proposed by Page was a stochastic model of window opening. User's interactions with the windows as well as room and outside air temperatures are considered and correlated to indoor pollution-related dissatisfaction and/or occupants' thermal comfort.

#### 4.3. Season identification and prediction model

A model of season prediction based on Hidden Markov Models was developed to identify the heating, cooling and intermediate (no heating or cooling) seasons<sup>10</sup>. In building control strategies it is important to adapt the system's behaviour to the weather conditions and thus, the "season" variable plays an important role. However, simplistic definitions of the season based solely on calendar or on outdoor temperatures that ignore building

characteristics and user behaviour are normally used<sup>11</sup>. In this model, three possible states of the "season" variable (heating, cooling and intermediate season) were considered as well as the probability of transition from one state to the next.

Database records from a single-occupant office room and 8 actuators and sensors including window opening, blinds position, external and internal temperature and solar irradiation were used for the elaboration of the model. Data regarding the heating power were used to verify the developed model as the use of heating is considered to be a characteristic of the heating season. Results showed that season was correctly predicted 91% of times (for heating or cooling seasons and it had 69% accuracy when identifying the intermediate season.

## 5. Conclusions & Outlook

In this article we present the LESO-PB database along with selected examples of research work that were carried out using this data. A vast amount of data has been collected over a period of about 6 years from a multitude of sensors, building devices and actuators on an occupied office building, without interrupting the normal workflow of its occupants. The research work performed with the help of the LESO-PB database draws from a wide domain of research and thus demonstrates the database's versatility for different applications. In specific, models of visual comfort optimization, occupant presence and behaviour, thermal comfort, season identification and more were developed using this database.

Since quality data availability is often scarce, we decided to make the LESO-PB database available to all the scientific community under the website: <http://www.wattict.com/>. For this, an effort has been put in this article to provide a concise description of the database, its structure and the potential for research work to be performed with. In the future, we intend to expand it with new measurements from newly installed sensors as well as to update it regularly with most recent data. We will also consider a "versioning" system where different versions of the database will be made available as subsets making data treatment more flexible and less resource-intensive.

Researchers are strongly encouraged to either follow on the work presented in the previous Section for comparing the results of their machine learning algorithms or to build and test new algorithms of their own such as activity recognition, appliances use, behavioural models and more. In either case, the LESO-PB database, with its vast temporal spectrum and its wide array of recorded variables offers scientific researchers a huge potential for future work to be performed upon.

## Acknowledgements

The presented work has been supported by the Hasler Foundation in Switzerland through the research project Green-Mod.

## References

- <sup>1</sup> Perez-Lombard, L., Ortiz, J., & Pout, C. A review on buildings energy consumption information. *Energy and buildings*; 2008. 40(3), 394-398.
- <sup>2</sup> T.L.M. van Kasteren, G. Englebienne, Ben Kroese. *Human Activity Recognition from Wireless Sensor Network Data: Benchmark and Software*. Activity Recognition in Pervasive Intelligent Environments of the Atlantis Ambient and Pervasive Intelligence series, Atlantis Press, Volume Editors: Liming Chen, Chris Nugent, Jit Biswas, Jesse Hoey, 2010.
- <sup>3</sup> S. S. Intille, K. Larson, E. Munguia Tapia, J. Beaudin, P. Kaushik, J. Nawyn, R. Rockinson. Using a live-in laboratory for ubiquitous computing research. *Proceedings of PERVASIVE 2006*, vol. LNCS 3968, K. P. Fishkin, B. Schiele, P. Nixon, and A. Quigley, Eds. Berlin Heidelberg: Springer-Verlag, 2006, pp. 349-365.
- <sup>4</sup> D. Cook. Learning setting-generalized activity models for smart spaces. *IEEE Intell Syst.* 2010 September 9; 2010(99).
- <sup>5</sup> Scartezini, J.-L. and Courret, G. Anidolic daylighting systems. *Solar Energy*; 2002. Vol. 73, 2, pp. 123-135.

---

<sup>6</sup> René Altherr and Jean-Bernard Gay. A low environmental impact anidolic facade. *Building and Environment*; 2002. 37(12):1409–1419.

<sup>7</sup> Lindeloef, D. Bayesian optimization of visual comfort. PhD thesis no. 3918, EPFL; 2007.

<sup>8</sup> Kevin B. Korb, Ann E. Nicholson. *Bayesian Artificial Intelligence*. cRc Press, 2003.

<sup>9</sup> Page, J. Simulating occupant presence and behaviour in buildings. PhD thesis no. 3900, EPFL; 2007.

<sup>10</sup> A Ridi, N Zarkadis, G Bovet, N Morel, J Hennebert. Towards Reliable Stochastic Data-Driven Models Applied to the Energy Saving in Buildings. CISBAT 2013, Lausanne, Switzerland, September 4-6, 2013. *Proceedings of CISBAT 2013 Cleantech for Smart Cities and Buildings*, vol. I, p. 501-505.

<sup>11</sup> De Dear, Richard, Brager, G. S. Developing an adaptive model of thermal comfort and preference. UC Berkeley: Center for the Built Environment, ASHRAE Transactions; 1998. Vol 104 (1), pp. 145-167.