



**HAL**  
open science

## Data-driven performance evaluation of ventilated photovoltaic double-skin facades in the built environment

Leon Gaillard, Guillaume Ruedin, Stéphanie Giroux-Julien, Marc Plantevit, Mehdi Kaytoue, Syamimi Saadon, Christophe Ménézo, Jean-François Boulicaut

### ► To cite this version:

Leon Gaillard, Guillaume Ruedin, Stéphanie Giroux-Julien, Marc Plantevit, Mehdi Kaytoue, et al.. Data-driven performance evaluation of ventilated photovoltaic double-skin facades in the built environment. 6th International Building Physics Conference, Jun 2015, Turin, Italy. pp.447-452, 10.1016/j.egypro.2015.11.694 . hal-01272611

**HAL Id: hal-01272611**

**<https://hal.science/hal-01272611v1>**

Submitted on 11 Feb 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



6th International Building Physics Conference, IBPC 2015

## Data-driven performance evaluation of ventilated photovoltaic double-skin facades in the built environment

Leon Gaillard<sup>a,\*</sup>, Guillaume Ruedin<sup>b</sup>, Stéphanie Giroux-Julien<sup>b</sup>, Marc Plantevit<sup>c</sup>, Mehdi Kaytoue<sup>d</sup>, Syamimi Saadon<sup>a</sup>, Christophe Ménézo<sup>a</sup>, Jean-François Boulicaut<sup>d</sup>

<sup>a</sup>INSA-Lyon, CETHIL UMR 5008, INSA-EDF Chair "Habitats and Energy Innovations", Villeurbanne 69621, France

<sup>b</sup>University of Lyon 1, CETHIL UMR 5008, 69621 Villeurbanne, France

<sup>c</sup>University of Lyon, CNRS, University of Lyon 1, LIRIS, UMR5205, 69621 Villeurbanne, France

<sup>d</sup>University of Lyon, CNRS, INSA-Lyon, LIRIS, UMR5205, 69621 Villeurbanne, France

---

### Abstract

In this paper we present a collaborative data science project bringing together experts from the fields of physical and information sciences to tackle to challenges and opportunities of analysing large and rich datasets obtained from the monitoring of building integrated PV systems operating in the built environment. We present data mining analysis techniques to classify data according to environmental conditions and system performance, to distinguish between nominal and anomalous behaviour, and to identify instrumentation faults. These methods were implemented using data from the RESSOURCES project to construct evaluate the performance of a PV envelope, and to validate a simplified physical model to predict thermal and aerodynamic behaviour.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

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

Peer-review under responsibility of the CENTRO CONGRESSI INTERNAZIONALE SRL

**Keywords:** Full-scale; PV double-skin façade; Monitoring; Datamining

---

### 1. Introduction

Europe strives for a decrease in green house gas emissions by a factor of 4 by 2050 relative to 1990 levels, and a reduction of 20% of energy consumption by 2020 whilst raising the contribution of renewable energy to 20%. The

---

\* Corresponding author. Tel.: +33-4-7243-8813; fax: +33-4-72438811.

E-mail address: [leon.gaillard@insa-lyon.fr](mailto:leon.gaillard@insa-lyon.fr)

building sector accounts for a large portion of emissions and energy consumption, and this is reflected in national regulations such as in France where by 2020 all new buildings should be net energy positive (BEPOS). Building integrated photovoltaic (BIPV) systems represent an interesting solution, which may include active or passive cooling techniques to control PV cell temperature whilst improving building thermal performance. For the large-scale deployment of PV envelopes, several technological hurdles must be overcome. Such components are often multifunction in nature, and must meet criteria including electrical and thermal production, but also daylighting, sound insulation and aesthetics. The complexity of the urban environment, heterogeneous and fluctuating in nature, also presents a major challenge to BIPV architectures [1]. Full scale experimentation under real conditions is essential in order to validate the concepts and predict the performance over the lifetime of the system [2].

As part of the RESSOURCES project, realistic prototype PV envelopes were designed and constructed on real buildings in France to study their behavior under natural ventilation operating conditions: two for individual houses and one for an office building. Each facade comprised tinted double glass PV modules of variable cell configurations plus fully transparent tinted panels. This paper makes use of results from the HBS Technal prototype, presented in figure 1. The prototype, of vertical pleated geometry, was installed at Toulouse on the W.N.W wall of a three-storey, open-plan office building. Measuring 7.4 m in height, 4.0 m in width, and with an airgap of 60-80cm, the prototype covers the first two floors of the building. The PV components are divided into a stack of three separate arrays, hereafter referred to as ‘blocs’, connected to a constant load, with a total rated power of 1.2 kW. The data used in this paper span a three month period, 21/06/2012 – 20/09/2012. Previous results [3] have shown that despite the level of geometrical complexity of the HBS prototype and the variability of the environment, the behaviour of the system follows regular and periodic trends over timescales of one day and one year.

The data retrieved from this and other PV installations contain a wealth of information that could be further exploited to improve the understanding of the technology, but novel methods are needed to analyze sizeable datasets comprising a large number of parameters, a multiplicity of spatial and temporal scales, and a level of instrumentation that is typically too coarse to reveal the detailed interaction of phenomena such as wind and shading effects. To this end, the CETHIL and LIRIS laboratories initiated with the INSA-Lyon/EDF research chair a collaborative data science project in 2012, initially within the framework of the AMADOUER project (CNRS MASTODONS programme), and subsequently by the INSA-Lyon BQR project SOLSTICE. To date this work has focused upon well-established data mining techniques and the creation of data exploration tools.

### Nomenclature

$G_{dir}$	direct solar radiation (pyrheliometer), ( $Wm^{-2}$ )
$G_i$	in-plane total solar radiation (pyranometer), ( $Wm^{-2}$ )
$G_{iT}$	total horizontal radiation (rooftop), ( $Wm^{-2}$ )
$G_{ref}$	radiation intensity at reference conditions, 1000 ( $Wm^{-2}$ )
$k$	number of clusters, (-)
$P_c$	rated electrical power, (W)
$P_{dc}$	instantaneous produced electrical power, (W)
$\eta$	instantaneous performance ratio (-)

## 2. Data mining methods

The analysis was developed with the aid of the open source software Knime [4]. A workflow was first assembled to manage the interface with the MySQL database repository, and handle the tasks of spatial and temporal aggregation, and filtering, that were previously undertaken using Scilab. Several data mining techniques were then appended to this analysis chain, and in particular the k-means clustering algorithm. Rarely used in scientific research of solar systems in urban environments, these methods have been demonstrated effective in other domains such as life sciences research (gene expression studies) for many years [5]. The k-means method is one of the oldest and most widely used clustering routines [6, 7]. The algorithm regroups data into  $k$  partitions or clusters in  $n$  dimensions, which for the current application correspond to indicators of system behavior or environmental factors. The only

criterion used to define the clusters is the distance of each constituent point to its centre. Points are associated to the cluster with the nearest centre in terms of the chosen dimensions; data belonging to the same cluster are therefore localized in this space. The partitioning is obtained by iteration, beginning with a random set of  $k$  seeds. Each data point is then associated to the nearest seed, resulting in an initial clustering. Cluster centroids are determined (in our case using a Euclidean metric), and adopted as a new set of cluster centres. The clustering is then repeated and this process continues until cluster centres have stabilized. The obtained clusters provide additional information for a data analysis by regrouping similar data together, which may therefore help to distinguish periods of nominal behavior and anomalies. A major advantage of the  $k$ -means algorithm is its speed, which scales well with the size of dataset and number of dimensions. Among its shortcomings are the need to choose the number of clusters and set of dimensions by hand, and to a lesser degree, a certain sensitivity to the starting point of the calculation. To overcome these weaknesses in the present analysis, the  $k$ -means tool was used multiple times for different configurations before choosing the best set of results. Beyond the centroids of the clusters and their distributions, the  $k$ -means algorithm alone does not provide any information regarding the physical explanation of cluster association. It is therefore necessary to follow up the analysis with other methods, conventional or otherwise.

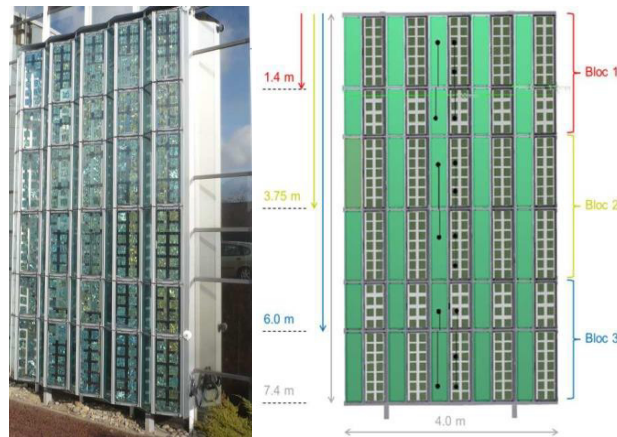


Fig. 1. Prototype vertically-pleated, PV double-skin facade installed at HBS Technal, Toulouse France as part of the RESSOURCES project. Right: schematic illustrating bloc arrangement.

### 3. Clustering according to electrical performance and solar radiation

The present analysis addresses the relationship between electrical power and incident radiation. Previous studies of correlation plots revealed multiple structures, indicating a superposition of different behavior in the dataset [3]. The  $k$ -means algorithm was used to isolate these features. The algorithm was found to provide a good partitioning of data for  $k=6$  clusters defined by the measured radiation intensities (total façade, total horizontal, and direct normal) and the performance ratio of each bloc, which is a measure of yield relative to incident radiation, both normalized to reference values, as defined by equation 1.

$$\eta = (P_{dc} / P_c) / (G_i / G_{ref}) \quad (1)$$

In figure 2, the resulting clusters are shown for the three blocs, with marker styles indicating the cluster association common between the three graphs. Cluster names are arbitrary. The graph legends are given in order of decreasing population. As can be seen in figure 2, some clusters are localised to certain regions of the graph. For example, cluster 6 is primarily located at low incident radiation and low performance for all three graphs. Similarly, cluster 2 appears associated to periods of high performance for all three blocs. In contrast, cluster 4 is characterized by a raised performance for bloc 1, a moderate performance for bloc 2 and a low performance for bloc 3.

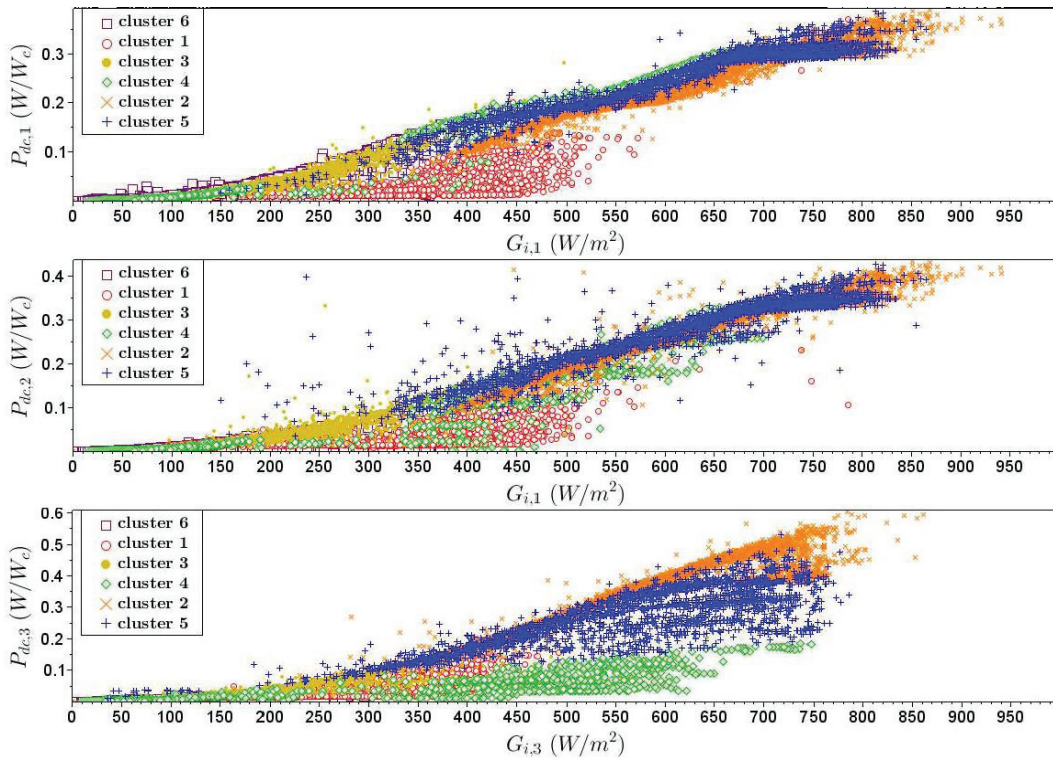


Fig. 2. Electrical power versus in-plane radiation for blocs 1-3 (from top to bottom respectively). Markers indicate cluster association.

It is also interesting to study possible temporal correlations in the clusters, which can be achieved effectively with the aid of carpet plots visualizations [3]. In figure 3, carpet plots are shown for global horizontal radiation, bloc 3 electrical performance and cluster association. For each plot, each pixel corresponds to an instantaneous value, at a given date (horizontal axis) and time (vertical axis). For solar radiation and electrical performance, the colour scales indicate the radiation intensity normalized to 1000 W/m<sup>2</sup> and power output normalized to the rated power of the array respectively. For the clusters, the colour scale indicates the cluster associated of each data point. To aid visibility, missing data points were suppressed by copying previous values (up to three consecutive measurements).

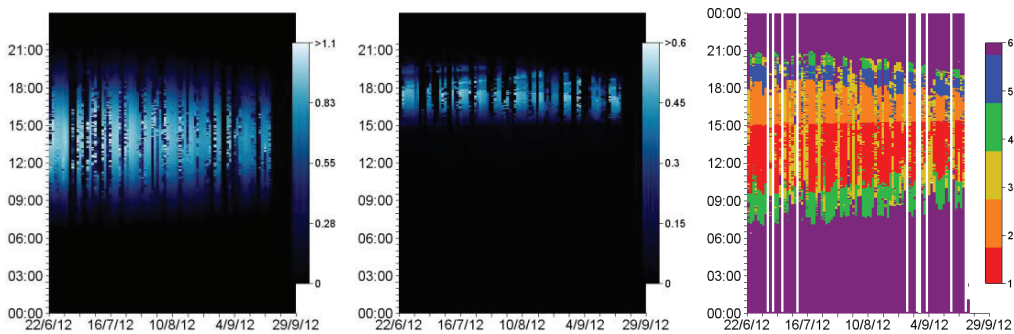


Fig. 3. Carpet plots of total horizontal radiation (left), bloc 3 power (centre), and k-means cluster (right) upon axes of date and time.



The solar radiation plot is characterised by a seasonal envelop (modified from a simple sinusoid by horizon features), and fleeting variations dues to cloud cover. These long and short term variations are matched electrical performance distribution, which is also strongly affected by the façade orientation. In addition, a horizontal banding is clearly visible, indicating power losses correlated with time of day, which are caused by fixed obstacles in the vicinity of the façade. The carpet plot visualisation of clusters shows a clear relationship with time of day for most of the clusters, which therefore further adds to the description of these clusters. Night time is completely accounted for by cluster 6. Similarly, cluster 4 regroups dawn and dusk, early morning and early evening for most of the days in the sample. Comparing the cluster associations to the other carpet plots, it is clear that cluster 3 is attributed cloudy periods. Moreover, days with partial cloud cover are divided into several clusters, which may indicate the presence of various competing phenomena during such days. The rest of the dataset is primarily attributed to one of three clusters, which appear correlated to the horizontal banding observed in the electrical performance carpet plot.

#### 4. Numerical model validation using clustered data

While the clustering of experimental data provides a means to classify data according to apparent features, the results alone do not lead to a detailed physical explanation of the behaviour. In order to identify the dominant physical phenomena and to create new knowledge, it is necessary to evoke numerical or empirical models. By way of an example, here we present the assessment of a numerical model for naturally ventilated PV double-skin facades, developed by [8]. This stationary, coupled thermal/aerodynamic model can be used to predict surface and air temperatures, as well as the mass flow rate under natural convection conditions. It comprises an iterative solution to energy balance equations for the envelop discretised into a stack of zones, and a single loop air pressure balance equation considering pressure losses, stack effect due to the heating of air and wind effect at the inlets [9]. Partial transparency of the facades is also taken into account.

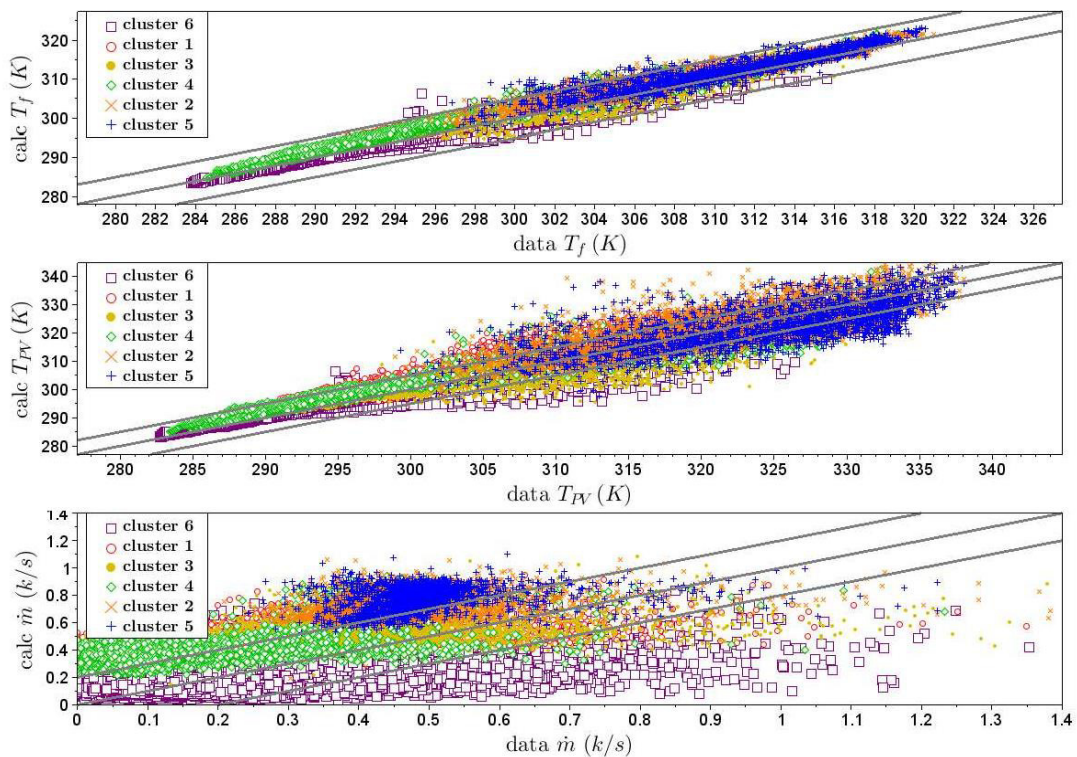


Fig. 4. Numerical model validation using clustered data. From top to bottom: cavity air and PV surface temperatures (bloc 1), and mass flow rate.

In figure 4, a model-data comparison is presented in terms of the air temperature at bloc 1 (upper section), the mean PV facade temperature (bloc 1), and the mass flow rate in the cavity. In each scatter plot, the cluster associations are included using the markers as in figure 2. Overall the model performs reasonably well, but is more accurate for temperatures than air flow rate. Several distinct structures are visible indicating that the performance of the model varies during the study period. Moreover, the clusters appear to differentiate some of this structure.

Calculated and measured values of temperature and mass flow rate agree fairly well for the lower end of the ranges (clusters 6 and 4). The predicted mass flow rate is significantly narrower than the measured data for these clusters and other cloudy periods (cluster 3), and for sunny periods where little direct radiation is received by the façade (cluster 1). For these daytime clusters the predicted mass flow rate is greater than during the night, but again the model does not reproduce the full range of observed mass flow rates. During periods of significant direct solar radiation incident on the façade (clusters 2 and 5), the  $T_{PV}$  and  $T_f$  are predicted to good accuracy by the model, albeit with a large degree of dispersion. The model tends to overestimate the mass flow rate for this subset.

## 5. Discussion and conclusion

Configured for incident solar radiation and electrical response, the k-means algorithm clustered data in terms of cloud cover and local shadowing effects. These environmental factors appear to account for some of the variation in model performance, and notably for mass flow rate. Overall the model tends to slightly overestimate the stack effect during periods when it dominates air flow in the cavity, and to underestimate peaks in mass flow rate during periods of low radiation, which are in part due to wind phenomena. The similarity in model behaviour for clusters 2 and 5 is expected, because the pyranometers are not generally affected by shadowing at the same time as the adjacent PV cells. In this work we have introduced data mining methods into the domain of Building Integrated PV Components for ventilated operating conditions in urban environments. We have demonstrated their utility for experimental evaluation and numerical model validation. Data mining classification methods including decision tree algorithms are currently under consideration to further contribute to the monitoring activity which will be essential for producing energy buildings. We are also working to systematise the configuration of the clustering algorithm regarding dimension selection and the quantitative comparison of results for different numbers of clusters.

## Acknowledgements

This work was financially supported by the French national research programme PREBAT managed by ADEME through the RESSOURCES project (convention 075C0076), the INSA-Lyon/EDF research chair and INSA-Lyon BQR SOLSTICE and CNRS (MASTODONS) AMADOUER projects.

## References

- [1] Eltawil MA, Zhao Z. Grid-connected photovoltaic power systems: Technical and potential problems – a review. *Renewable and Sustainable Energy Reviews* 14 (2010) 112-129
- [2] Bazilian MD, Leenders F, Van Der Ree BG, Prasad D, Photovoltaic cogeneration in the built environment. *Solar Energy* 71 1 (2001) 57-69
- [3] Gaillard L, Giroux-Julien S, Ménézo C, Pabiou H, Experimental evaluation of a naturally ventilated PV double-skin envelope in real operating conditions, *Solar Energy* 103 (2014) 223-241
- [4] Berthold MR et al. KNIME: The Konstanz Information Miner, *Data Analysis, Machine Learning and Applications*. Springer Berlin Heidelberg; 2008, p. 319-326
- [5] Hackl H, Burkard TR, Sturm A, Rubio R, Schleiffer A, Tian S, Quackenbush J, Eisenhaber F, Trajanoski Z. Molecular processes during fat cell development revealed by gene expression profiling and functional annotation, *Genome Biology* 6 (2005) 13, R108
- [6] Steinhaus H, Sur la division des corp materiels en parties. *Bull. Acad. Polon. Sci. IV (C1.III)* (1956) 801–804
- [7] Anil KJ, Data clustering: 50 years beyond K-means, *Pattern Recognition Letters*, 31 (2010) 8 651-666
- [8] Saadon S, Gaillard L, Menezo C, Simulation study of a naturally ventilated building integrated photovoltaic (BIPV) envelope, *IBPC 2015*, Turin, Italy; 2015
- [9] Brinkworth BJ, Marshall RH, Ibarahim Z, A validated model of naturally ventilated PV cladding, *Solar Energy*, 69 (2000) 67-81