Energy & Buildings 176 (2018) 95–108

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enbuild

Interactive visualization for NILM in large buildings using non-negative matrix factorization

Diego García[a,*], Ignacio Díaz[a], Daniel Pérez[a], Abel A. Cuadrado[a], Manuel Domínguez[b], Antonio Morán[b]

[a] Electrical Engineering Department, University of Oviedo, Edif. Departmental, Campus de Viesques s/n, Gijón 33204 Spain
[b] SUPPRESS Research Group, University of Leon, Escuela de Ingenierías, Campus de Vegazana, León 24007 Spain

ARTICLE INFO

Article history:
Received 5 February 2018
Revised 24 May 2018
Accepted 28 June 2018
Available online 20 July 2018

Keywords:
Energy efficiency
Visual analytics
Building energy consumption
NMF
Energy disaggregation
Interactive displays

ABSTRACT

Non-intrusive load monitoring (NILM) techniques have recently attracted much interest, since they allow to obtain latent patterns from power demand data in buildings, revealing useful information to the expert user. Unsupervised methods are specially attractive, since they do not require labeled datasets. Particularly, *non-negative matrix factorization* (NMF) methods decompose a single power demand measurement over a certain time period into a set of components or "parts" that are sparse, non-negative and sum up the original measured quantity. Such components reveal hidden temporal patterns which may be difficult to interpret in complex systems such as large buildings. We suggest to integrate the knowledge of the user into the analysis in order to recognize the real events inside the electric network behind the learnt patterns. In this paper, we integrate the available domain knowledge of the user by means of a visual analytics web application in which an expert user can interact in a fluid way with the NMF outcome through visual approaches such as *barcharts, heatmaps* or *calendars*. Our approach is tested with real electric power demand data from a hospital complex, showing how the interpretation of the decomposition is improved by means of interactive *data cube* visualizations, in which the user can insightfully relate the NMF components to characteristic demand patterns of the hospital such as those derived from human activity, as well as to inefficient behaviors of the largest systems in the hospital.

1. Introduction

One interesting approach in energy efficiency is the improvement of electric consumption strategies by means of power demand monitoring tools. Thanks to the growing amount of smart meters installed recently, a lot of measurements of power demand are being gathered today in buildings, household and industrial systems, containing useful hidden information that can help in making efficient decisions if it is extracted and presented intuitively. Thus, developing techniques that are able to extract characteristic patterns of how energy is being consumed from large volumes of demand data, as well as methods to visualize these patterns in an efficient and intuitive way have become promising research topics. A suitable approach with these features is *visual analytics* (VA) [1,2]. VA exploits the insightful synergies between *intelligent data analysis* (IDA), *data visualization* and *interaction* mechanisms, allowing the user to get knowledge from efficient visualizations of raw data and the IDA outcome. By means of interaction mechanisms, the user is brought into the analysis, being able to modify according to his expert knowledge of the system both the visualization and the IDA analysis.

Applying techniques based on IDA algorithms, it is possible to address issues that require a certain learning from the data such as forecasting energy consumption [3–5], getting temporal patterns in electric power demand [6] or the factorization of a total electric power demand into consumptions downstream that have not been measured individually. Decomposing a total consumption, in a sensorless way, is called in the literature *non-intrusive load monitoring* (NILM) [7–9]. This kind of techniques, that can be considered a parts-based representation of total energy, increases the energy awareness of the user, since the obtained components can provide insightful information of how the electric consumptions are distributed temporally and spatially in the network.

Despite recent efforts to develop novel NILM systems, few approaches have explored the NILM outcome within the VA paradigm. NILM techniques are suitable analysis tools for VA, since the obtained disaggregated consumptions reveal temporal patterns

* Corresponding author.
E-mail address: diegogarcia@isa.uniovi.es (D. García).

https://doi.org/10.1016/j.enbuild.2018.06.058

bearing insight when they are represented using suitable data visualizations such as week [10] or calendar [11] heatmaps.

NILM techniques can be classified into supervised and unsupervised methods. In supervised methods, optimization algorithms [12] are applied, and their principal disadvantage is that a labeled dataset is required. In many cases, obtaining labeled data can increase set-up costs of NILM systems, making unsupervised methods more suitable. Within unsupervised NILM techniques, *blind source separation* (BSS) and *factorial hidden Markov models* (FHMM) are the most typical approaches as it is explained in [13]. We choose to apply *non-negative matrix factorization* (NMF) [14] as BSS technique, due to the positive nature of electric power demand and because the imposed constraints on the resulting components may improve their interpretability. NMF techniques aim to separate total power demand into sources of consumption which can be associated with characteristic events or recognizable consumption patterns downstream. In large buildings, where there are many complex systems, the interpretation of the resulting decomposition is often difficult and may require an expert knowledge of the system.

In our work, we propose to apply state-of-the-art NMF techniques to energy-based data of a hospital complex within an interactive framework which allows to integrate the user knowledge in the analysis, showing that NILM methods combined with efficient visualization and interaction mechanisms result in an insightful power demand monitoring tool for large buildings. In Section 2, we formulate our approach and we describe the real electric consumption dataset from a hospital complex on which our approach is tested. Then, in Section 3, we define how to apply NMF to electric measurements and how to integrate NMF in an interactive visualization. Finally, in Section 4, the proposed power demand monitoring application is presented, showing different cases of study.

2. Problem formulation

In this section, we will introduce the two pillars of our work: NILM analysis and the specifications of the interactive data visualization. Then, we will explain the particular problem that will be addressed in Section 4.

2.1. NILM analysis

Several studies [7,15,16] suggest that an improvement in energy efficiency can be achieved through the increase of energetic awareness by means of feedback mechanisms. The extended deployment of smart meters is an example that getting feedback about demand has become a key factor in energy efficiency strategies. Based on this, we propose not only a simple visual supervision of the measured energy data, but also consider the integration of NILM techniques in a visual interface to reveal hidden patterns of how the energy is spent, providing a more detailed feedback information.

The idea of NILM was introduced three decades ago by Hart in [12]. Its application to time series of electrical demand results in the decomposition of the following parts

$$\mathbf{p}(t) = \mathbf{p}_1(t) + \mathbf{p}_2(t) + \ldots + \mathbf{p}_n(t) \quad (1)$$

where the electric power demand at time t is the sum of the n consumptions $\mathbf{p}_i$ corresponding to the appliances connected downstream. Note that if the network lacks generative systems, the components $\mathbf{p}_i$ are always positive. This non-negative nature of the components suggests that each $\mathbf{p}_i$ is a part of a whole. Getting insight into an object through learning its parts individually is an operation that occurs in the human perception process as it is shown in [17,18]. Thus, the parts-based representation expressed in Eq. (1) is potentially a highly interpretable feedback for the user.

A well-known decomposition method which attempts to learn the parts of a whole from the input data is *non-negative matrix factorization* introduced in [14]. NMF was conceived to address fundamentally image processing, being able to learn different parts from example images of an object. In recent years, more successful applications of NMF have been addressed in different fields, such as text analysis [19] or environmental data analysis [20], but few approaches have explored the possibilities of applying NMF to electric power demand data. In this paper, we suggest how to apply NMF techniques to energy consumption analysis in a large building, discussing their advantages and limitations as unsupervised NILM methods.

2.2. The need for interactive visualization

Energy demand monitoring tools in large buildings are based on the energy-based data gathered from smart meters. These measurements involve different kinds of attributes apart from the electric variables, such as time factors (hour, day, week, year), spatial factors (buildings, floor), environmental factors (temperature, humidity, occupancy) or even variables of characteristic processes or systems where a large demand is involved. Thus, energy demand analysis is a *multiway* problem. If all these factors are efficiently presented in an intuitive representation, an expert user can discover hidden insightful correlations between all these factors, thereby improving his knowledge about how the energy is spent. In this scenario, many applications suggest that the user can find these correlations visually by means of well designed visual encodings to transform information into appropriate visual representations. Most of the visual energy demand approaches are based on static 2D representations which can only plot a few factors. Through interaction, the user can explore a multiway dataset, setting a scenario by means of a number of conditions and using the visual feedback to confirm or reformulate hypotheses, by retuning the conditions, focusing on specific factors or modifying the visual encodings. This process has been modeled in [2] as an iterative process that improves the knowledge of the user through several cycles of interaction and visualization. It is known that the more fluid [21] the process of interaction is, the better immersion in the problem, increasing the chances to obtain useful knowledge as well as confidence in the results.

One example of interactive data exploration tool based on electric power demand is [22], where the multiway problem of energy demand analysis in several public buildings is addressed through a fluid interaction between coordinated views of the factors. In order to achieve the required fluidity in the interaction, the information from the buildings is structured according to the *data cube* approach [23,24]. The data cube provides mechanisms to create new hypotheses through filters on different attributes of the data which can be established quickly by means of Javascript implementations [25], receiving an immediate visual feedback (latencies < 16 ms).

The *visual analytics* paradigm suggests that the gained knowledge about the problem not only comes from the fluid interaction between data and the user, but also from the outcome of IDA algorithms. The integration of IDA outcomes in the interactive visualization reinforces them, since they can be visually confronted with the rest of the attributes by the user in an intuitive way. The newly discovered knowledge inspires the user to change the specifications both in data visualization and in IDA, closing the cycle.

In our approach, we propose to integrate the outcome of NILM techniques in an interactive visualization based on the data cube framework, considering design principles which improve both the fluid interaction and the efficient perception of the attributes.

2.3. Monitoring power demand in a hospital complex

The concepts explained above will be tested with real electric power demand data obtained from a hospital complex, which in-

Table 1
Variables used in our energy disaggregation approach.

Id	Label	Variable description
1	PotenciaPF	Total power demand (kW)
2	P1	Cooling machine 1 (kW)
3	P2	Cooling machine 2 (kW)
4	P3	Cooling machine 3 (kW)
5	P4	Cooling machine 4 (kW)
6	P5	Cooling machine 5 (kW)
7	T1	Cooling machine 6 (kW)
8	T2	Cooling machine 7 (kW)
9	Temperature	Temperature (°C)
10	DayOfYear	Day of year
11	Hour	Hour
12	Month	Month
13	WeekOfYear	Week of year
14	DayOfWeek	Day of week
15	Fecha	Timestamp

clude different measurements of current, voltage, power and quality parameters in several measuring points over the network. In our work, we will analyze the total power demand (which gathers the consumptions of all systems in the hospital), focusing our analysis on the heating, ventilation and air conditioning systems (HVAC) as well as on seven cooling machines, which feed with cold water the ring that keeps the diagnostic devices chilled. The thermal comfort system represents about 30% of overall consumption in large buildings [26]. Therefore the disaggregation of these consumptions is a relevant feedback in order to improve the efficiency through changes in the hospital such as configuration of HVAC or schedules. Moreover, the cooling machines have a large peak of consumption in the starting, fact that may produce problems in the electrical network of the hospital, such as exceeding the total contracted power. In the dataset used in this paper, the individual consumptions of each cooling machine are recorded. Therefore, it is not only possible to supervise the starting peaks, but it is also possible to prove the capacity of NMF methods to detect the consumptions of the HVAC system.

The energy-based dataset used to apply our approach is composed of the variables shown in Table 1. These variables were collected with a sampling period of one minute throughout the year 2016. Some samples were lost due to problems in the meter caused by a temporary network malfunction, so our dataset has a total of 526,981 samples. Such amount of samples difficults the agility of interaction if the data are not properly summed up using aggregation techniques. In the next section, we shall explain how the information is efficiently aggregated in the data cube paradigm, improving the fluidity in the interaction with the data.

3. Methods and techniques

In this section, methods and terminology about NMF will be explained in the context of its application to data from total electric demand of large buildings. Then, in order to reach a fluid interactive visualization, the data cube paradigm is introduced, summarizing briefly its elements and operations. Finally, in this section, we shall explain a set of suitable visualizations and interaction methods to visualize the data cube and the NMF outcomes.

3.1. Non-negative matrix factorization (NMF)

NMF is formulated [14,27] as an approximate matrix factorization of non-negative input expressed as $\mathbf{V} \approx \mathbf{WH}$. Let us define a M-dimensional vector $\mathbf{v}_j$ whose elements are non-negative and the matrix $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N] \in \mathbb{R}^{M \times N}$ as N observations of $\mathbf{v}_j$, $j = 1 \ldots N$. Considering $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_L] \in \mathbb{R}_+^{M \times L}$ and $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_L]^T \in \mathbb{R}_+^{L \times N}$, where $\mathbf{h}_\alpha^T$ is the αth row of matrix $\mathbf{H}$. NMF can be reformulated as:

$$\mathbf{v}_j = \sum_{\alpha=1}^{L} \mathbf{w}_\alpha h_{\alpha,j} \tag{2}$$

where $\mathbf{v}_j$ is the jth column of $\mathbf{V}$ and $h_{\alpha,j}$ is the αth element in jth column of $\mathbf{H}$. Therefore $\mathbf{v}_j$ is a linear combination of the columns in $\mathbf{W}$. The coefficients of this linear combination are the elements of the columns in $\mathbf{H}$. Thus, all the input observations are weighted sums of non-negative basis vectors. This non-negativity suggests that the columns $\mathbf{w}_i$ are "the parts of a whole" as it is shown in [28], and they will be referred to as *basis consumptions*. On the other hand, the elements of the rows in $\mathbf{H}$ contain information about the influence of the basis consumptions along the N observations. Note that NMF decomposition has infinite possible solutions since it is possible to transform the decomposition into a new one by means of any transformation matrix $\mathbf{K}$ such that $\mathbf{V} = \mathbf{WKK}^{-1}\mathbf{H}$.

Although there are infinite possible transformations, the product $\mathbf{w}_\alpha h_{\alpha,j}$ between one basis consumption α and its coefficient corresponding to a specific observation j is unique and it has units of measurement (kW in the case of total electric power demand). This product will be called a *component* throughout the rest of the article and it will be the object of analysis in later cases of study. Similarly, the pair $(\mathbf{w}_i, \mathbf{h}_i)$ will be referred to as the ith *factor*. In summary, the application of NMF to electrical consumptions aims to extract a certain number of basis consumptions that are a parts-based representation of the whole measurement, thus showing latent patterns, which may contain useful information in order to improve the knowledge about the system.

3.2. Complexity of NMF

Achieving an exact NMF decomposition $\mathbf{V} = \mathbf{WH}$ is considered a NP-hard issue [29]. Due to this complexity, we dismiss the exact NMF model and we use the basic NMF $\mathbf{V} \approx \mathbf{WH}$. One difficult issue in this basic approach is to determine the number of basis consumptions, since the accuracy of the reconstruction $\mathbf{WH}$ is directly related to the number L. The number of components can be chosen according to the reconstruction error, choosing the inflection point in the curve of the reconstruction error vs. the number of components. In certain complex problems, such as energy demand in large buildings, the number of components obtained with this automatic method is often high, making the efficient representation of the results more complicated. Therefore, in our approach we propose to integrate the choice of L within the interactive paradigm, being the user who sets the number of components according to the expert knowledge of the problem and the feedback received in the visual representation.

3.3. SVD-based initialization

Typically, $\mathbf{W}$ and $\mathbf{H}$ are estimated using a constrained optimization problem whose objective function is a similarity measurement between $\mathbf{V}$ and $\mathbf{WH}$ [28] and it is solved by means of gradient descent methods [30]. This optimization problem is not convex with respect to $\mathbf{W}$ and $\mathbf{H}$ at the same time; however, it is separately convex in either $\mathbf{W}$ or $\mathbf{H}$ [31]. As a nonconvex gradient-based problem, results of basic NMF rely on the initialization of $\mathbf{W}$ and $\mathbf{H}$ matrices.

In most of the NMF approaches, the matrices $\mathbf{W}$ and $\mathbf{H}$ are initialized applying nonnegative random techniques. Optimizing NMF with gradient descent methods, the random initializations converge to a local minima. Thus, each trained model will be different. Other methods suggest to apply an initial low rank decomposition of the input matrix $\mathbf{V}$ such as clustering or svd-based meth-

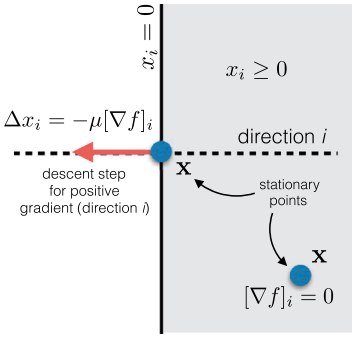

Fig. 1. Intuition about the possible NMF solution.

ods. In clustering methods typically the *k*-means algorithm is applied, which contains random elements, so basic clustering methods are not capable of ensuring the repeatability of the results. In [32], the authors propose *nonnegative double singular value decomposition* (*nndsvd*) as a good initialization of the matrix **W** and **H**. This svd-based initialization is the best positive low rank matrix decomposition based on the first L most relevant singular values of input matrix **V**. Without a random process in *nndsvd*, the obtained basis consumptions will be always the same. Moreover, in [32] it was shown that less iterations are needed in order to obtain the local minimum. In this paper, (**W, H**) will be initialized with *nndsvd* algorithm.

3.4. Objective function: sparseness

NMF, such as it has been formulated above, is a low rank decomposition that is purely additive since the reconstruction of the input matrix lacks cancellations by negative components. The interpretation of the components is more intuitive if these components are sparse. Thus, NMF decomposes the input matrix into positive blocks that have relevance only in a few elements. This idea has a tight connection with the human mechanism of learning an object by its parts. In the beginnings of NMF [27,28], it is shown that the basic NMF model applied to some real dataset reaches sparse basis consumptions and coefficients, being, therefore, a sparse parts-based representation. In [33] it is suggested that by analyzing NMF as an optimization problem, we can obtain some intuition about why the outcome is sparse. Let us define the following optimization problem:

$$J(\mathbf{W},\mathbf{H}) = \frac{1}{2}\|\mathbf{V}-\mathbf{WH}\|_F^2$$
$$(\mathbf{W}^*,\mathbf{H}^*) = \arg\min_{\mathbf{W},\mathbf{H}} J(\mathbf{W},\mathbf{H})$$

where the cost function J(**W, H**) is based on the Frobenius norm of the residual matrix $\mathbf{V}-\mathbf{WH}$. J(**W, H**) represents a measurement of similarity between the input matrix **V** and the estimated matrix **WH**. This problem is similar to the minimization:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x})$$

where the vector **x** gathers all the elements of **W** and **H**. If we use gradient-based methods to minimize the function $f(\mathbf{x})$, the set of stationary points has to satisfy the following requirements:

$$S = \{\mathbf{x} \in \mathbf{R}_+^n \,|\, \nabla f(\mathbf{x}) \geq 0 \text{ and } x_i[\nabla f(\mathbf{x})]_i = 0,\ i = 1\ldots n\}$$

The condition $x_i[\nabla f(\mathbf{x})]_i = 0,\ i = 1\ldots n$ expresses that, for every direction i, the solution must be either clipped to the positivity borders ($x_i = 0$) or lay on $x_i > 0$ being a fixed point ($[\nabla f(\mathbf{x})]_i = 0$). One representation of this condition is shown in Fig. 1 which underlines that the possible stationary points might be located on the boundary of the constrained space of possible values of **x**, explaining why many elements of **W** and **H** are zero, making the NMF a sparse positive decomposition. Despite this tendency to yield sparse results, some approaches [31] integrate new terms in the cost function in order to ensure the sparsity of the results. In this paper, we consider that the results obtained by the basic NMF are sparse enough, so that we do not integrate measurements of sparsity in the cost function.

3.5. Structure of matrix **V**

If we consider electrical consumption data as a time series $\{x(t)\}$, a restructuring is needed in order to construct **V** as an input matrix to NMF. Thus we can restructure the sequence by breaking it up into N contiguous windows of length M:

$$\{\underbrace{x(0), x(1), \ldots, x(M-1)}_{\mathbf{v}_1}, \ldots,$$
$$\underbrace{x((N-1)M), x((N-1)M+1), \ldots, x(NM-1)}_{\mathbf{v}_N}\}$$

The new windows conform the matrix **V** of observations as follows:

$$\mathbf{V} = \begin{pmatrix} x(0) & x(M) & \ldots & x((N-1)M) \\ x(1) & x(M+1) & \ldots & x((N-1)M+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(M-1) & x(2M-1) & \ldots & x(NM-1) \end{pmatrix} \quad (3)$$

Power demand time series in large buildings typically present a strong cyclic structure for periods of days, weeks, seasons or years. If the size of the window is chosen according to these periods of time, more interpretable results will be obtained, since the structure of the results allows a better association between basis consumptions and events in the network. The results of NMF are strongly dependent on the length of the windows. Thus, if we choose a weekly or annually instance of the load profile, NMF will decrease resolution in the obtained patterns, revealing weekly or annual patterns, losing resolution in daily events. Therefore, the size of the window should be chosen according to the desired analysis. In our work, we will focus on interpretable events in daily scale because they can be associated with daily behaviors that can be easily corrected, so each window will be a daily instance of the load profile.

3.6. Data cube paradigm and NMF

In some cases, the number of resulting components is very high, making it difficult to understand the representation based on parts. In order to achieve a better intuition of the results, in Section 1, we exposed the need of a visual interactive representation of the energy disaggregation, which permits to associate the components with characteristic behaviors of the system. One intuitive way to make these associations is to formulate hypotheses in form of questions such as: when does this component take the highest value?; where is this component more relevant?; is there any correlation between this component and any of the attributes?; which components show strong temporal patterns?

In this paper, we suggest that an interactive approach of a data cube allows us to formulate these questions in a visual and interactive way, through interaction mechanisms. Data cubes (OLAP, *on-line analytical processing*) [23,24] can be seen as data structures storing multidimensional *records*. Each *record* is composed of values of both measures and attributes. We call *measures* those fields[1]

[1] Some works [22–24] use the term *dimension*. We preferred the term *field* to avoid confusion with the geometrical dimension of the cube.

which are object of analysis. Each *attribute* takes values from a finite set, resulting from a group operation[2] on the field values. Attributes are the indices of the data cube, and their values define the coordinates of each cube cell. In this way, attributes can be seen as the sides of the data (hyper)cube. Specifically, each register is assigned to a cube cell according to its values of the attributes, it being possible for a cell to contain multiple records (or none). For each cell, representative values of the records it contains can be calculated by means of aggregation operations, such as summation, average, maximum, etc., applied on the measures. We will refer to these values as the outcome of each cell.

Data cubes admit operations that modify both the number of possible cells and the way in which their aggregated value is obtained, adjusting the hypercube outcome to the queries of the user. Our approach will enable the following three operations:

- *Selection operation*. We shall define the *selection* operation as the act of filtering the records according to a logical expression. Thanks to selection it is possible to set up the outcome of the data cube according to the constraints imposed by the user. *Dice selection* is a particular case of selection whereby the logical expression is only true for a contiguous subset of bins in one or more attributes, and therefore the analysis is focused on a region of the hypercube. In our approach, dice selections can be defined by filters of contiguous values in one or several attributes.
- *Projection operation*. A common practice in multiway problems is to focus the analysis on a subset of attributes in order to understand the whole problem incrementally. Within the data cube paradigm, this is addressed by the *projection operation* which reduces the sides of the cube by dimensionally squashing it into a subset of the initial attributes, thereby decreasing the number of cells.
- *Aggregation operation*. We shall define the *aggregation function* as the operation which returns a single value (more generally a set of values or even an object) that summarizes the set of records on each cell. In our approach, the aggregation function can be chosen among `count`, `sum` or `mean`. In the case of `sum` and `mean` operations, the user decides which measure will be aggregated.

In [22], an *interactive data cube* was proposed that allows the user to implement these operations on large datasets in a highly efficient manner (using the Javascript library `crossfilter.js`), being able to represent the results on "live" coordinated views such as conditional histograms, barcharts of average values, that are updated in real time (with latencies well below 0.1 seconds) upon changes in the filters. Our approach includes the functionalities of [22] and add new ones such as the incorporation of the NMF results as new dimensions and 2D projections of the data cube.

The highly interpretable parts-based representation that can be obtained algorithmically by means of NMF can be complemented by sensor measurements and features, thereby providing additional insight into the problem. We suggest turning each NMF component into both a new attribute of the data structure and a new measure object of analysis, so that the user can not only arrange the cube by the NMF components but also obtain aggregated values of the components as the outcome of the data cube cells. In order to achieve that, a transformation of the NMF outcome is needed. The resulting decomposition can be seen as

$$\mathbf{V} = \mathbf{w}_1\mathbf{h}_1^T + \mathbf{w}_2\mathbf{h}_2^T + \cdots + \mathbf{w}_L\mathbf{h}_L^T$$

where $\mathbf{w}_i$ is the ith column of $\mathbf{W}$ and $\mathbf{h}_i^T$ is the ith row of $\mathbf{H}$, so that the outer product $\mathbf{w}_i\mathbf{h}_i^T \in \mathbb{R}^{M\times N}$. Each of the summed matrices contains the components of the decomposition for a specific factor throughout the observations. These matrices should be rearranged, undoing the operation (3) in order to obtain a vector of dimensions MN which can be integrated into the data cube structure. Once the components are integrated, their interpretation is dramatically improved, since insightful correlations between the resulting latent patterns and initial measures of the data cube (external temperature, active power of cooling machines, etc.) can be discovered in a highly interactive way through dice and projection operations on the new NMF-based attributes.

In [22] only 1D projections are included, but we include new 2D projections onto certain combinations of attributes in order to provide a bidimensional representation of the cube like a pivot table. For instance, an efficient representation of temporal 2D projections on the pair of attributes DayOfWeek/Hour or DayOfWeek/WeekOfYear provides an adequate feedback in temporal context which can be easily connected to the desired analysis by selecting both the specific aggregation function and the measure which is aggregated.

Although the data cube approach achieves an efficient computation of the aggregated results, they are composed of numeric values which have to be visually encoded in order to provide an efficient feedback according to the principles of human visual system. Moreover, the interaction with the data cube through its operations has to be implemented by means of low latency interaction mechanisms which allow the user to explore the data in a fluid manner, being able to get a fast feedback on her hypotheses.

3.7. Visualization of the data cube and interaction mechanisms

Although the data cube approach is a useful tool for generating numeric reports of how the energy is spent, in this paper we aim to encode efficiently this information in an interactive web visualization. It can be noted that the data cube is based on the *binned aggregation principle* [34], since the records are arranged into bins along the attributes. This principle is a well-known data reduction technique which is often applied to manage large datasets (more than a million of records) in interactive visualizations [22,35,36].

Specific visual designs suitable for representing the aggregated values resulting from the data cube projections can be found in the literature [35] (such as barcharts, heatmaps, scatterplots, etc). In this paper, we do not only suggest representing the aggregated values with the mentioned static visual approach, but also integrating the data cube operations through interactive mechanisms in two types of interactive visualizations:

3.7.1. Barchart

The so-called *barchart*, which is shown in Fig. 2, is an efficient visual design for representing 1-D projections of the data cube, in which the aggregated values resulting from each cell are encoded by means of the height of the bars. The color is not relevant in the visual encoding, only the height is employed to encode the resulting aggregated value of each cell, since in order to compare magnitudes, length is a more effective channel than color, according to the principles of human perception [37,38]. At the top of each barchart in Fig. 2, there are three combo boxes by which the user defines both the projection and aggregation operations. The first combo box defines the attribute on which the data cube is projected, whereas the second combo box indicates which aggregation function is applied. If the selected function is either `mean` or `sum`, the measure to be aggregated is defined by the third combo box. The selection operation is implemented by means of *brush* and *drag* gestures as it is shown in Fig. 2. The brush event defines a selection of contiguous values (dice selection) in the attribute

[2] A partitioning process of the field values into a finite number of subsets (e.g. bins on a continuous variable, or specific classes such as day at week, etc.) assigning a unique value to each subset.

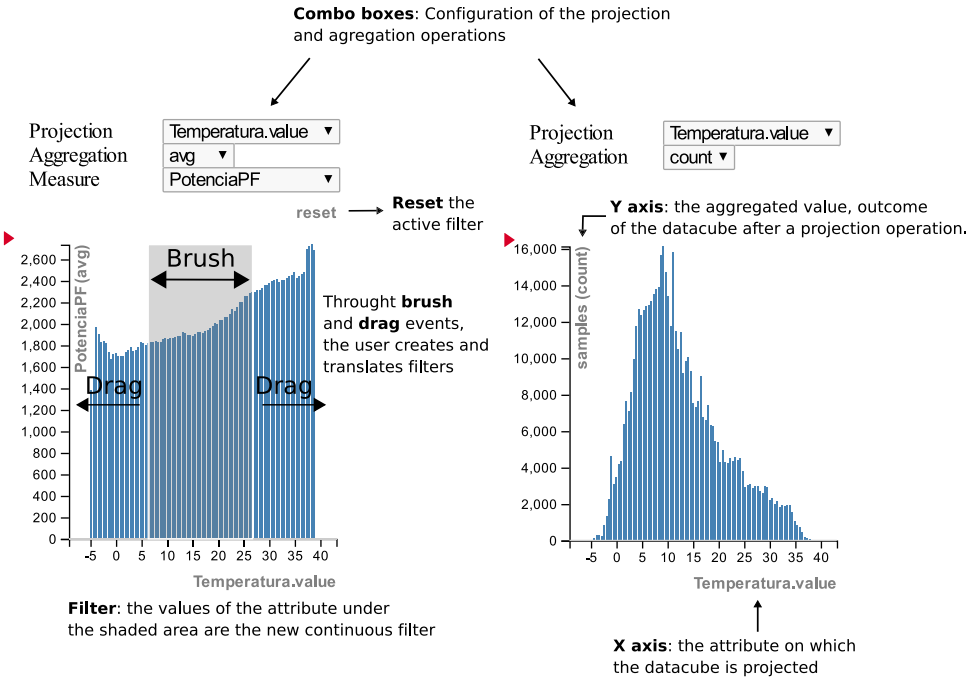

Fig. 2. Interactive barchart representing 1D projections of the data cube.

represented in the *x*-axis. This selection can be quickly redefined by means of drag events, so that the user can limit the analysis to different regions of interest, being able to spot insightful visual correlations in the rest of the represented projections.

3.7.2. Heatmap

An advisable visualization for representing 2D projections are the binned heatmaps [35], as shown in Fig. 3. In these kinds of visualizations, the bins of the 2-D projection are visually represented by a square whose color encodes the resulting aggregated value. Although the heatmap integrates combo boxes whereby the user defines the aggregation operation, the projection and selection operations are disabled. Our heatmaps are associated with `DayOfWeek/Hour`, shown in Fig. 3(a), and `DayOfWeek/WeekOfYear`, shown in Fig. 3(b). In the last one, the bins are arranged as a calendar, which is a widely used visualization. The calendar arrangement provides insight on many kinds of periodicities closely related to the human activity.

Applying the previous approaches to the NMF outcome, the user can visually explore the NMF decomposition together with all the fields gathered from the building, being able to get useful correlations between them. Thus, visualizing interactively a data cube which integrates the discovered patterns, results in significant cross-fertilization between both approaches.

4. Results

The described approach was tested with real electric power demand from a hospital complex, developing the visual energy monitoring tool shown in Fig. 4. Only the total power demand (`PotenciaPF`) from Table 1 was the input to the NMF model, while the rest of the variables were only incorporated into the data cube structure. The measurements were gathered with a sample period of one minute, so the dataset obtained throughout one year contains a large amount of records which might be computationally hard to analyze. For this reason, we suggested to reduce the number of records by means of a 1:3 downsample operation (average as aggregation function), obtaining a resolution of three minutes in the records.

Applying NMF decomposition requires to split the time series `PotenciaPF` into windows in order to generate the columns of the input matrix $\mathbf{V}$ described in (3). We set the size of the windows so that the obtained columns were daily observations of the total demand of the hospital. With this arrangement, the resulting coefficient vector $\mathbf{h}_i$ represents the daily contributions of basis consumption $\mathbf{w}_i$ throughout the year.

In this process, we discarded those days which presented missing records, because NMF would learn the gaps in the data as components, instead of other relevant patterns in the network. Both the preprocessing of the data and the NMF decomposition were developed in a *Python* script which makes use of several well-known data analysis libraries such as *Pandas* or *Scikit-learn*. This script was called from a Python web server based on *Flask*, which hosts the *Javascript* visualizations, and thereby both approaches were merged into one website allowing the user to configure the data analysis and the visualization within the same interface.

The hyperparameters of the NMF algorithm, such as the number of components L or the maximum number of iterations, are chosen by the user in a dialog window as it is shown in Fig. 5. By default, we suggest the configuration shown in Fig. 5, which ensures a representative set of sparse components.

In order to improve the interpretation of the NMF outcome, we represent the basis consumptions $\mathbf{w}_i$, integrating a new visualization outside of the data cube visualization shown in Fig. 6. Representing the basis consumptions through *sparklines* allows the final user to spot and compare the learned daily patterns, using a small area of the dashboard.

In the remainder of the section, we shall discuss the NMF analysis according to different number of components. Then, we shall present the most relevant patterns discovered through NMF and we shall show how powerful the interactive visualization of the data cube is in order to interpret the NMF decomposition.

4.1. Analysis and interpretation of NMF components

In order to analyze the NMF decomposition we trained several NMF models with different numbers of components L, collecting

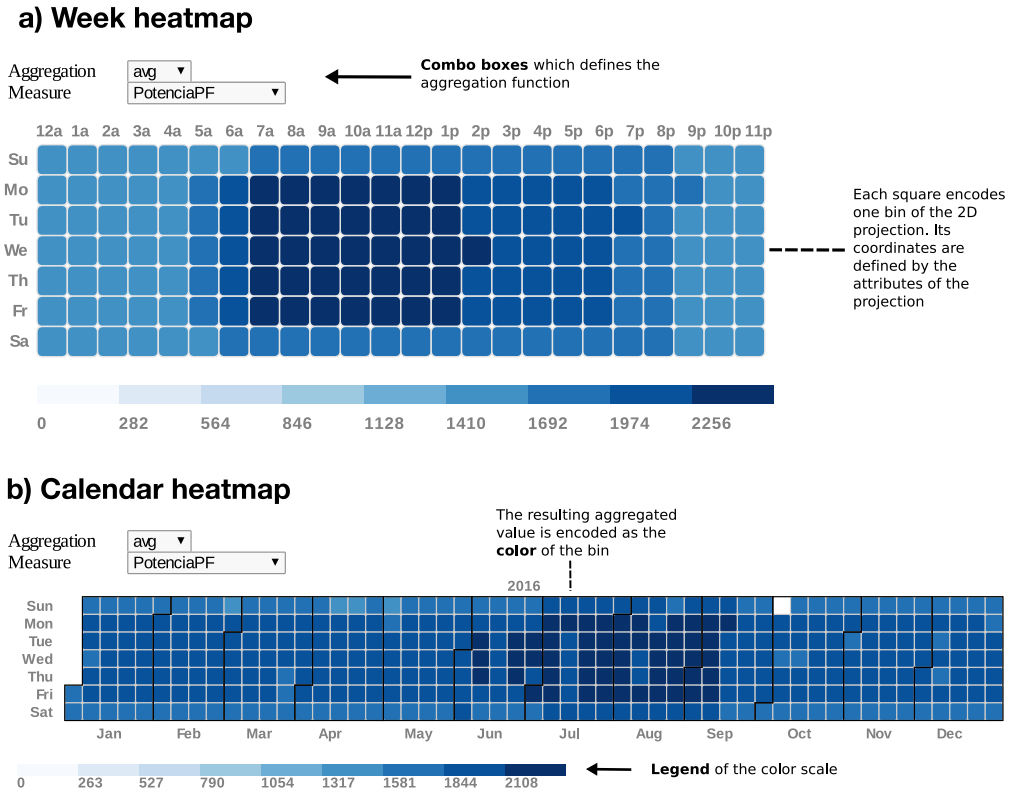

Fig. 3. Heatmaps which represent the 2D projections of the data cube. (a) Represents the average of the total power demand of the hospital (`PotenciaPF`) grouped by hour and day of week. (b) Represents the average of the total power demand grouped by days of year arranged in a calendar distribution.

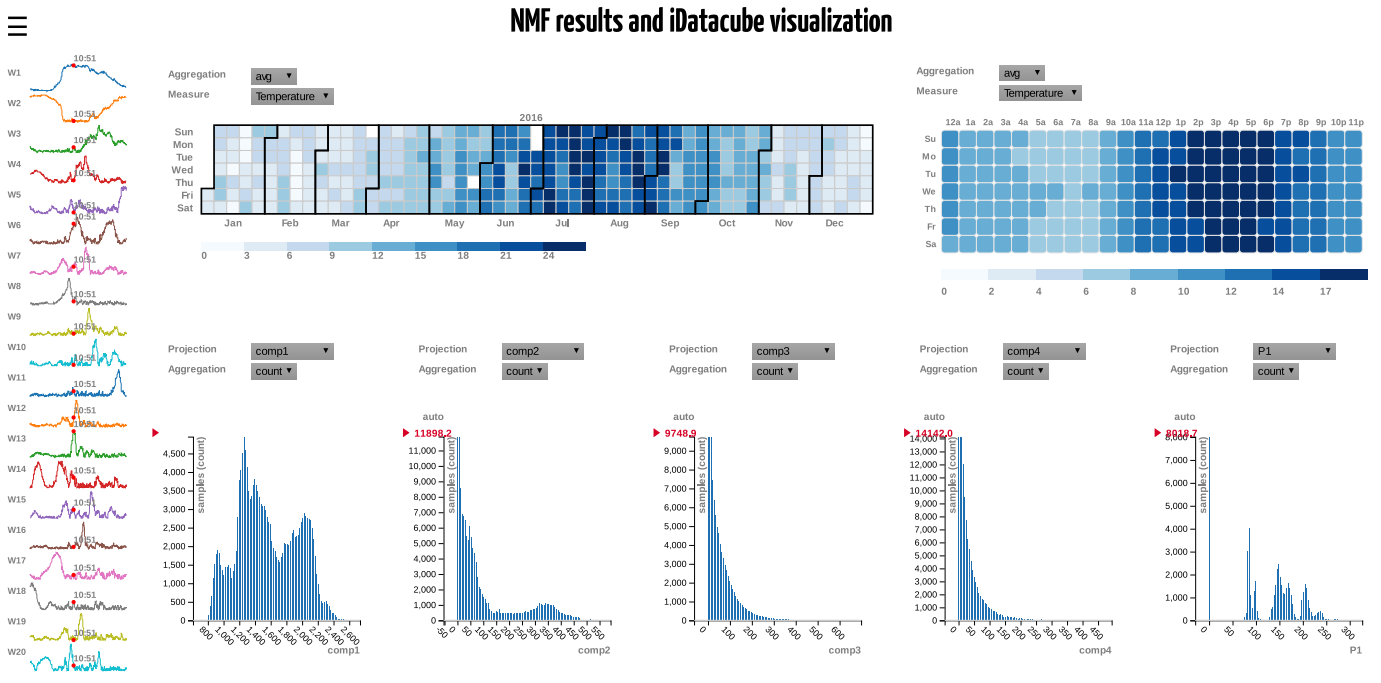

Fig. 4. Application snapshot.

the Frobenius norm of $\mathbf{V} - \mathbf{WH}$ as the reconstruction error of the trained models.

$$\text{reconstruction error} = \|\mathbf{V} - \mathbf{WH}\|_F^2$$

Thus, 29 models were trained with $L = 2, 3, \ldots, 30$ numbers of components. In Fig. 7, it is shown how the error falls when the number of components is increased. We consider that a decomposition of 20 factors is sufficient. Therefore, in our application this will be the number of components by default.

Fig. 8 shows some of the 20 trained basis consumptions and their coefficients, where, due to the svd-based initialization method, the first basis consumptions, such as $\mathbf{w}_1$, $\mathbf{w}_2$ and $\mathbf{w}_3$,

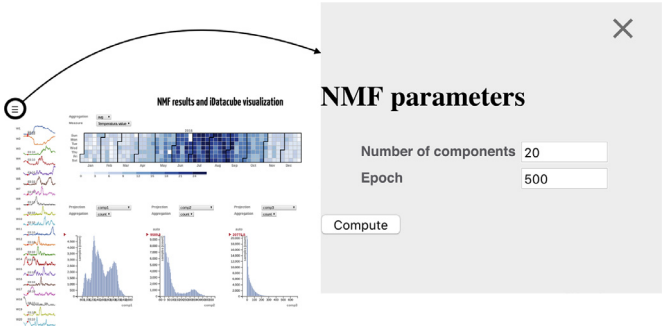

Fig. 5. Dialog window in which the user can choose the parameters of NMF model.

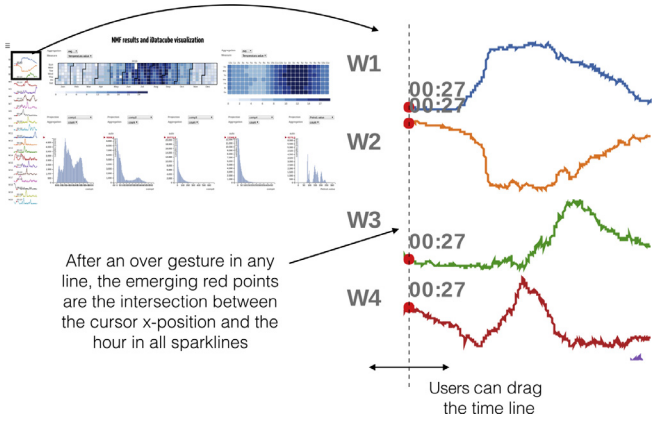

Fig. 6. Snapshot of sparklines of NMF basis consumptions.

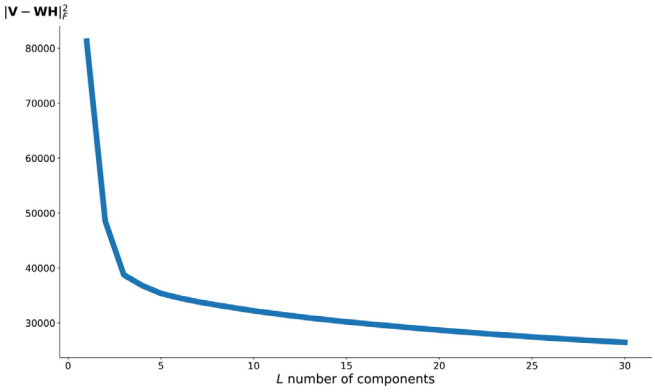

Fig. 7. Reconstruction error according to the number of components.

are the most relevant, revealing patterns which are associated to weekly or seasonal consumptions throughout the year as it can be seen in their coefficients. The last basis consumptions (from the fourth onwards) learnt the unlikely events associated with deviations in normal behavior or occasional faults, being relevant only in a few hours of specific days.

The first factor ($\mathbf{w}_1$, $\mathbf{h}_1$) is the most relevant, having influence throughout the year, specially in weekdays.[3] It has the form of a normal daily profile in the hospital and it can be interpreted as a "canvas" on which the rest of the components will be aggregated. The second factor ($\mathbf{w}_2$, $\mathbf{h}_2$) shows an important influence in weekends and holidays, where the first factor takes less significance. Thus, the NMF has recognized two kinds of days in its two most relevant factors which can be associated with the human activity in the hospital. The third factor ($\mathbf{w}_3$, $\mathbf{h}_3$) shows a significant influence during summer, revealing a rise of the power demand from 2:00 p.m. to 8:30 p.m. Moreover, its coefficients ($\mathbf{h}_3$) show a large peak at the end of May which coincides with a heavy decrease of $\mathbf{h}_1$. The decomposition underlines this day as a deviation from the normal profile which had a large consumption in the afternoon and evening, demonstrating the capacity of NMF to detect anomalies. From the tenth factor onwards, the magnitude of the coefficients decreases, obtaining basis consumptions that are only relevant in specific days, as can be observed in $\mathbf{h}_{11}$, $\mathbf{h}_{16}$ and $\mathbf{h}_{20}$. Such factors are often composed of basis consumptions that have a large peak of consumption for a couple of hours, being patterns without periodicity. Hence, NMF decomposition not only finds periodic and relevant patterns in electric power demand but also can learn non-periodic and less relevant patterns which would be more difficult to obtain by other methods.

So far, hidden latent patterns have been found by means of NMF and they have been temporally located, but the cause of the consumption is yet to be found. This is a difficult issue in NILM decompositions which may require to integrate extra expert knowledge into the analysis by means of the interactive data cube visualization.

4.2. Analysis using the visual analytics application

The factors presented above were integrated into the interactive data cube by developing the application shown in Fig. 4, which includes the methods explained in Section 3.7. By using different configurations of the available interactive tools, the NMF factors can be contextualized with the rest of the attributes, thereby reinforcing their analysis. Such configurations are given by:

- NMF parameters. Although in this section the *number of factors* obtained is set to 20, the user could redefine it. The *maximum number of iterations* of the decomposition can be also indicated.
- The selected attributes in the *projection operation* in each 1D visualization of the data cube (barcharts).
- The aggregation operation in each visualization of the data cube that in turn is defined by the *aggregation function* and the measure on which it is applied.
- The *filters* that the user has defined in any of the data cube views.

Thanks to this approach, the user is part of an incremental analysis in which, after acquiring new knowledge from the representation of the NILM analysis, both the number of factors and the data cube representations can be redefined. This process is shown in detail in Fig. 9. It is not possible to recompute a new decomposition in a fluid way, so changes in NMF parameters involve long latencies in the application. On the other hand, changes in visualizations and data cube operations involve low latencies. In the rest of the section we suggest useful configurations of the application in order to get new knowledge of the system from the NMF decomposition.

4.2.1. Large coefficients/weights in the components

One of the most powerful configurations in order to understand the NMF outcome is to filter those records that have the highest values in each of the obtained components, spotting the emergent patterns in the rest of the data cube visualizations. Under this configuration, the analysis is focused on the records where the filtered component has the largest influence. After this kind of filtering, several components reveal periodic patterns which are easily recognizable by the user as it is shown in Fig. 10. The components that present periodic patterns are often associated with human activity in the hospital. For instance, Fig. 10a shows how after filtering by the largest values of the component 1, an emerging week-

[3] Fig. 8 is not big enough to see clearly the weekly patterns due to space problems but it is possible to see the periodicity of the peaks and the seasonal patterns.

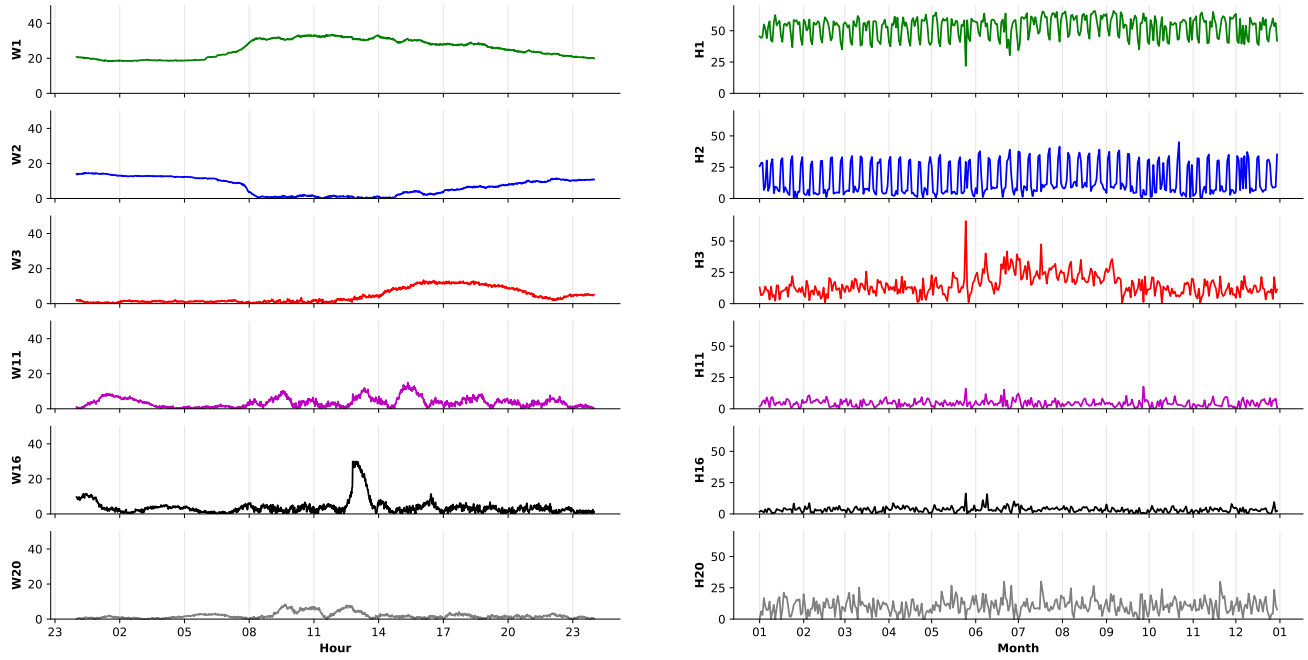

Fig. 8. Some components of a 20 factors NMF model. The first three factors are more relevant both in daily profile and throughout the year than the last factors.

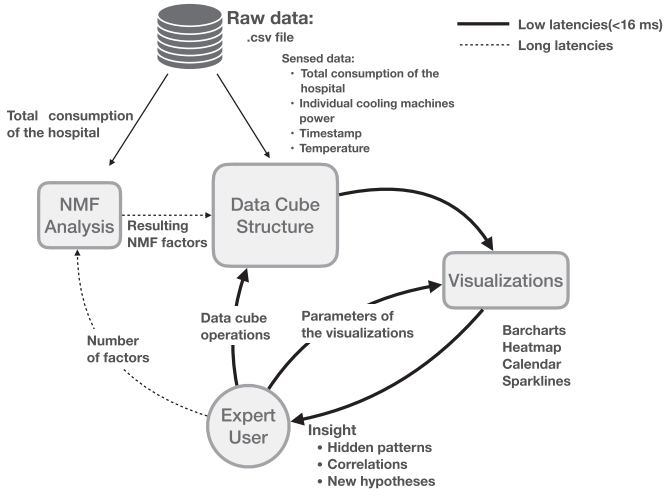

Fig. 9. Conceptual scheme of the application. Our method involves two levels of interaction: long latencies interactions, denoted in the scheme by dashed lines, and low latencies (fluid), denoted by bold lines.

day pattern can be spotted in the heatmaps, which have *count* as the aggregation function. On the other hand, in Fig. 10b, when the largest values of the second component are filtered, weekends and holidays are highlighted in the heatmaps. Other components show seasonal patterns, such as the third component which is filtered by its largest values in Fig. 10c. In the calendar the largest values of the third component are related to the seasonal behavior of the hospital and its week heatmap shows how this component has more influence between afternoons and evenings.

These findings may lead to new questions or hypotheses like "why is the third component higher in the afternoons?" or "does this have some connection with cooling machines?" which can be answered with other configurations of the application.

4.2.2. Visual correlation between factors and the rest of the attributes

Establishing correlations between attributes, in a visual way, allows the user to find hidden relations between NMF components and the rest of the attributes. Although the possible correlations established visually through the different views of the data cube do not necessarily imply causation, such relations may lead the expert user to the right connection between the variables of the system. The configuration whereby the user can achieve visual correlations is set when an NMF attribute is projected on a barchart whose aggregation function is the averaged value of another measure. Two correlated variables return a barchart where the y-axis values increase or decrease monotonically along the x-axis as it is exemplified with two examples in Fig. 11. The components 3 and 8 were chosen to get visual correlations through the *average* aggregation operation.

The first two barcharts in Fig. 11(a) show how the third component has a strong influence on summer afternoons, so that one might think that this component is correlated with the ambient temperature. After that, the user could configure a new barchart to project the averaged values of the ambient temperature on the attribute corresponding to the third component and thus, confirming that the temperature and the component are directly related. With this information, the user might reformulate the hypotheses, looking for relations between the third factor and the cooling machines which would explain the increment of consumption in the hospital from 1:00 p.m. These hypotheses may be argued by setting a data cube view in which the averaged values of consumption of a cooling machine of type T are grouped by the third component as it is shown in the last barchart in Fig. 11(a). After that configuration, the barchart shows a strong positive correlation between the consumption of the cooling machine of type T and the third component.

In the same way, Fig. 11(b) shows how in the winter months, the 8th component has learnt a peak of consumption at 9:00 a.m. Analyzing this component by means of the configuration of the interactive data cube described above, it is found that the ambient temperature is inversely related to the 8th component for medium

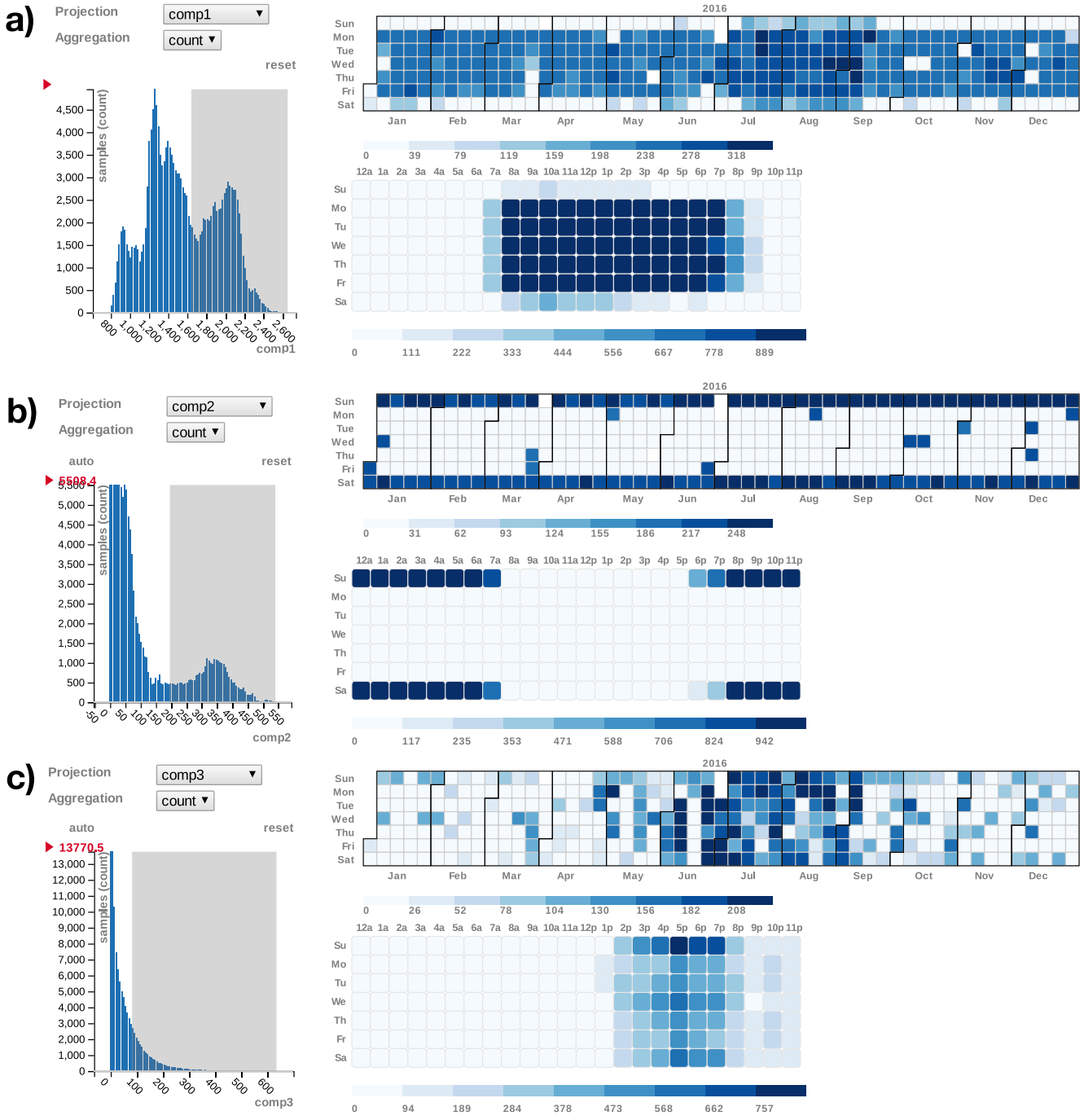

Fig. 10. Filtering the largest values of the first three components by means of drag gestures in their associated barcharts. (a) Weekday patterns are spotted in both the calendar and weekly heatmaps of the largest values of the first component. In (b), the highest values of the second component highlight the weekends and holidays. The calendar in (c) shows that the largest values of the third component are related to the seasonal behavior.

and low values of this component. On the other hand, the last barchart in Fig. 11(b) shows that the consumption of the P1 is slightly related to the 8th component.

This visual-based correlation analysis allows us to establish relationships by sections between two attributes. Thus, for example, in the barchart in Fig. 11(a) that shows the correlation between component 3 and the T2 machine, two well-differentiated sections can be observed: a section of strong correlation for low and medium values of the component; and another section with weaker correlation for high values of the component. Such kind of analysis can not be achieved in a typical approach based on global statistical descriptors, such as linear correlation. In these cases, the outcome is composed of a unique global measure in which the local relations are not appreciated.

The Tables 2 and 3 help us to validate this visual-based correlation analysis. Table 2 shows the Pearson correlations of the components 8 and 3 against the consumption of machines P1 and T2 as well as the ambient temperature, conditioned to several ranges of the components. These ranges are: low values (L) between [0–100], medium values (M) between [100–200], high values (H) between [200-350]. We also include a whole-range correlation (WR) spanning the three previous ranges [0–350]. Table 2 shows that the

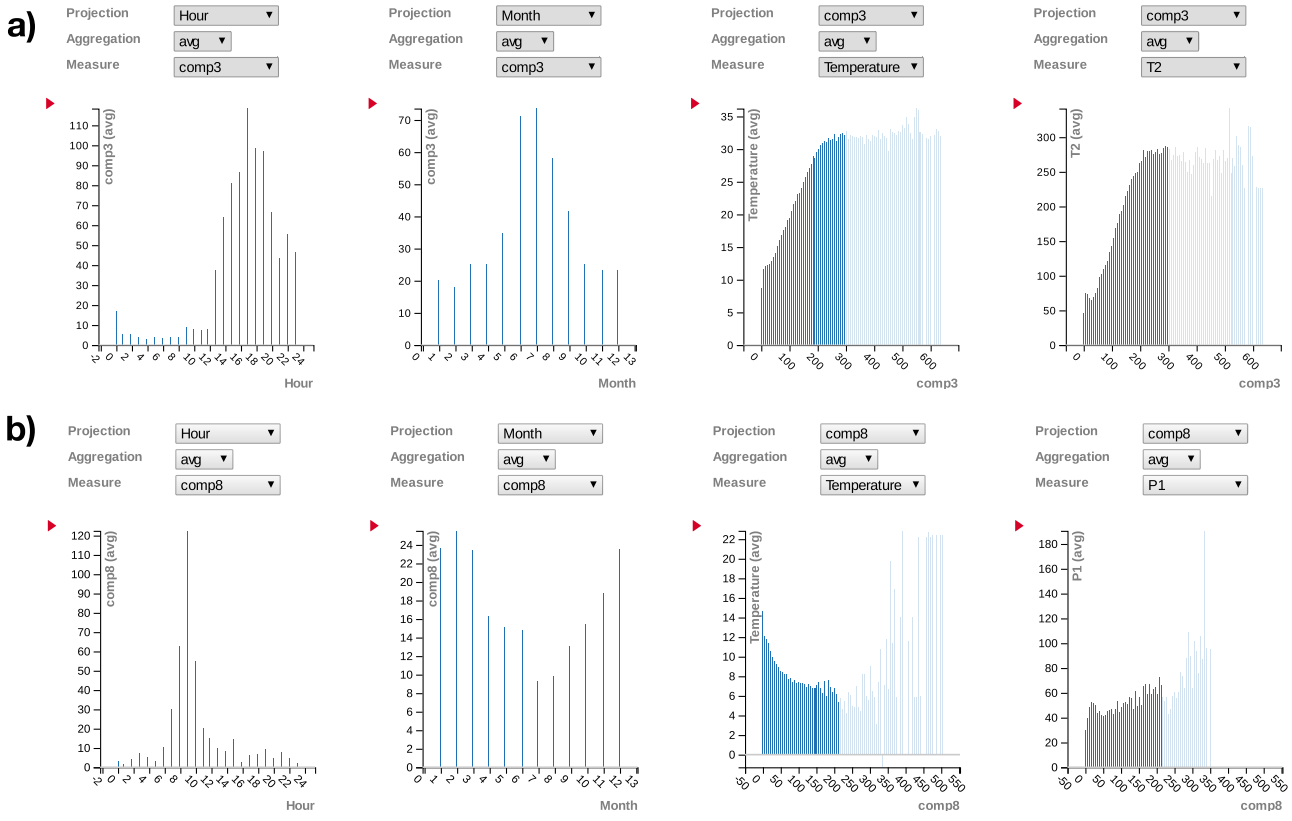

Fig. 11. Configuration of the interactive data cube whereby an insightful visual correlation was found in the third and eighth components. In (a), relations can be established between the third component, the ambient temperature and the consumptions of the cooling machines of type T. In the same way, (b) shows how the low and medium values of the eighth component can be strongly related to low temperatures and slightly related to the consumption of the P1 machine.

Table 2
Sample-based Pearson coefficients between the components 3 and 8 and both the individual consumptions of the cooling machines P1 and T2 and the ambient temperature conditioned to low (L), medium (M), high (H) values of the component as well as the whole-range correlation (WR).

	comp3				comp8			
	L	M	H	WR	L	M	H	WR
P1	0.03	-0.10	-0.04	-0.10	0.09	0.05	0.08	0.09
T2	0.22	0.27	0.03	0.46	-0.15	-0.08	0.02	-0.15
Temp	0.41	0.32	0.13	0.63	-0.21	-0.04	0.04	-0.22

Table 3
Group-based correlations between the components 3 and 8 and both the individual consumptions of the cooling machines P1 and T2 and the ambient temperature conditioned to low (L), medium (M), high (H) values of the component as well as the whole-range correlation (WR).

	comp3				comp8			
	L	M	H	WR	L	M	H	WR
P1	-0.38	-0.96	-0.60	-0.91	0.25	0.70	0.39	0.61
T2	0.95	0.99	0.03	0.94	-0.84	-0.78	0.48	-0.44
Temp	0.98	0.99	0.75	0.96	-0.93	-0.48	0.48	-0.48

results obtained analytically have the same sign but are weaker than the visually inferred relationships. For example, the correlation between component 3 and the ambient temperature spotted in Fig. 11(a) is strong and positive for medium values, while the corresponding linear correlation in Table 2 is 0.32, weaker than we expected. This is because each bar of the chart represents the average value of several samples, attenuating the effect of noise and variations at the local level, so that the observed relationships between atributes represent the global trend more clearly. The same correlations of the previous table, but this time on the average values within each group defined on the components (100 groups per component), are shown in Table 3. Comparing these two tables we can validate that the group-based correlations are much stronger than the sample-based correlations. A limitation of the group-based correlation approach appears when the number

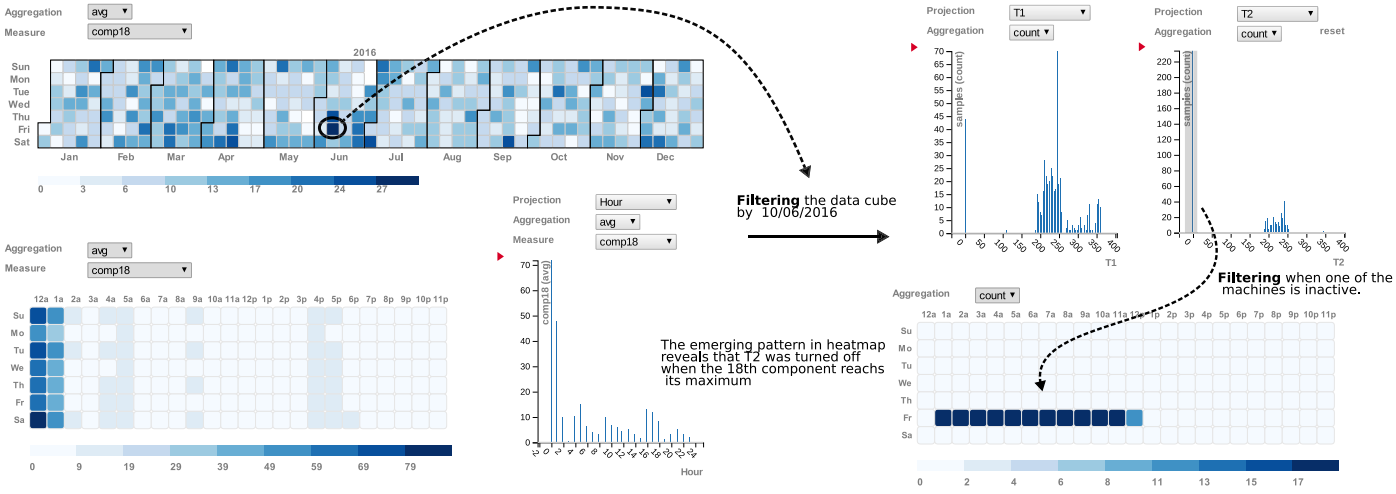

Fig. 12. The exploration of the 18th component as an example of analysis of a non periodic component. The user focus the analysis on one of the highlighted days, looking for evidences in the rest of the measures that can explain the causation behind the component.

of records arranged in the groups are not sufficient to produce a representative mean and thereby introduce noise in the barcharts. This kind of noise can be spotted in the groups which include the highest values of the components in Fig. 11(a) and (b). For this reason, in the aplication, those bars which contain less than 100 samples have been represented with a high transparency, losing visual relevance. Otherwise, they are totally opaque.

By focusing the analysis on average values we are losing local information, which in certain scenarios may be useful. Following the approach of the interactive data cube it is possible to focus the analysis on local behaviors, by means of filters in other attributes such as the time or month of the year, reducing the samples arranged into each group to those specific to certain condition.

The former examples show how to extract insight of the system through the synergies between the NMF decomposition and the interactive data cube approach since the recognized patterns such as the former components are related to important subsystems (cooling machines), revealing when they have more demand. For instance, in the previous analysis, it was found that in cold days the hospital presents a peak of consumption at 9:00 a.m. learned by the eighth component which is weakly related to the consumption of P machines, while the consumption of the machines of type T can be related to the increase of total consumption on summer afternoons learned by the third component.

4.2.3. Study of the non-periodic components

Resulting patterns do not necessarily appear in temporal 2-D projections of the data cube but they may emerge in attributes outside the NMF, such as the consumptions of specific large subsystems. For instance, the calendar representation of the 18th component, which represents the averaged values of that component grouped by day of year, shows no recognizable pattern as it is shown in Fig. 12. Thus, it is difficult to associate the peak of consumption at midnight learned by this component with an obvious behavior of the hospital.

One possible way to interpret this kind of components is to focus the analysis on those days where the component of analysis is more relevant and then trying to look for the causes of this relevance through filters on the rest of the attributes. In the example shown in Fig. 12, the 10th of June is selected, since the 18th component presents a large influence in this day. Then, thanks to previous explorations we know that machines of type T are activated in summer. Therefore, the attributes chosen to filter were the electric power demands of the two machines of type T, finding that the peak at midnight of the component is because one of the T machines is turned off. This fact caused the active T machine had to handle all the cooling demand, increasing its consumption, and therefore the total electric demand increased too. In a similar way, it can be proved that in winter days, where the 18th component is relevant, a machine of type P is turned off instead of a machine of type T. Thus, the user can identify the peak of the 18th component as the increment of the total consumption due to the shutdown of a cooling machine.

The former analysis not only confirms that NMF factors learn patterns associated with real events in the network but also that the synergies between NMF and interactive data cube allow to find these relations insightfully even when the components are not periodic.

4.2.4. Going deeper into the anomalies

Throughout the exploration of the components, we noted that some NMF factors highlight days in which their coefficients are much greater than in their neighboring days, revealing an abnormal behavior. These components learnt a fault which can be analyzed in detail by means of the data cube functionalities explained above. Within the hospital electric power demand, the component 3 shown in Fig. 13 learnt an anomaly on the second of May where its influence is greater than in the rest of May. Filtering by that day, we found that only P1 was active until 4:00 p.m. when a second cooling machine was required likely due to a rise in cooling demand. The control program which regulates the activation of the cooling machines interpreted that a machine of type T was necessary so T1 was started. T machines have a higher cooling capacity than P machines, being more efficient to use one machine of type T instead of two machines of type P. The situation illustrated in Fig. 13 is an example of how the control sent the “start” command to T1 and then sent the “stop” command to P1, but something wrong happened, since P4 was chilling only a few minutes about 5:00 p.m. after P1 was turned off. This inefficient start-up of P4 is learned by the third component as a peak at 5 p.m. The starting of P4 was probably due to the fact that the machines of type T take a long time to start up, so T1 could not handle the cooling demand, making it unavoidable that a machine of type P had to be turned on again.

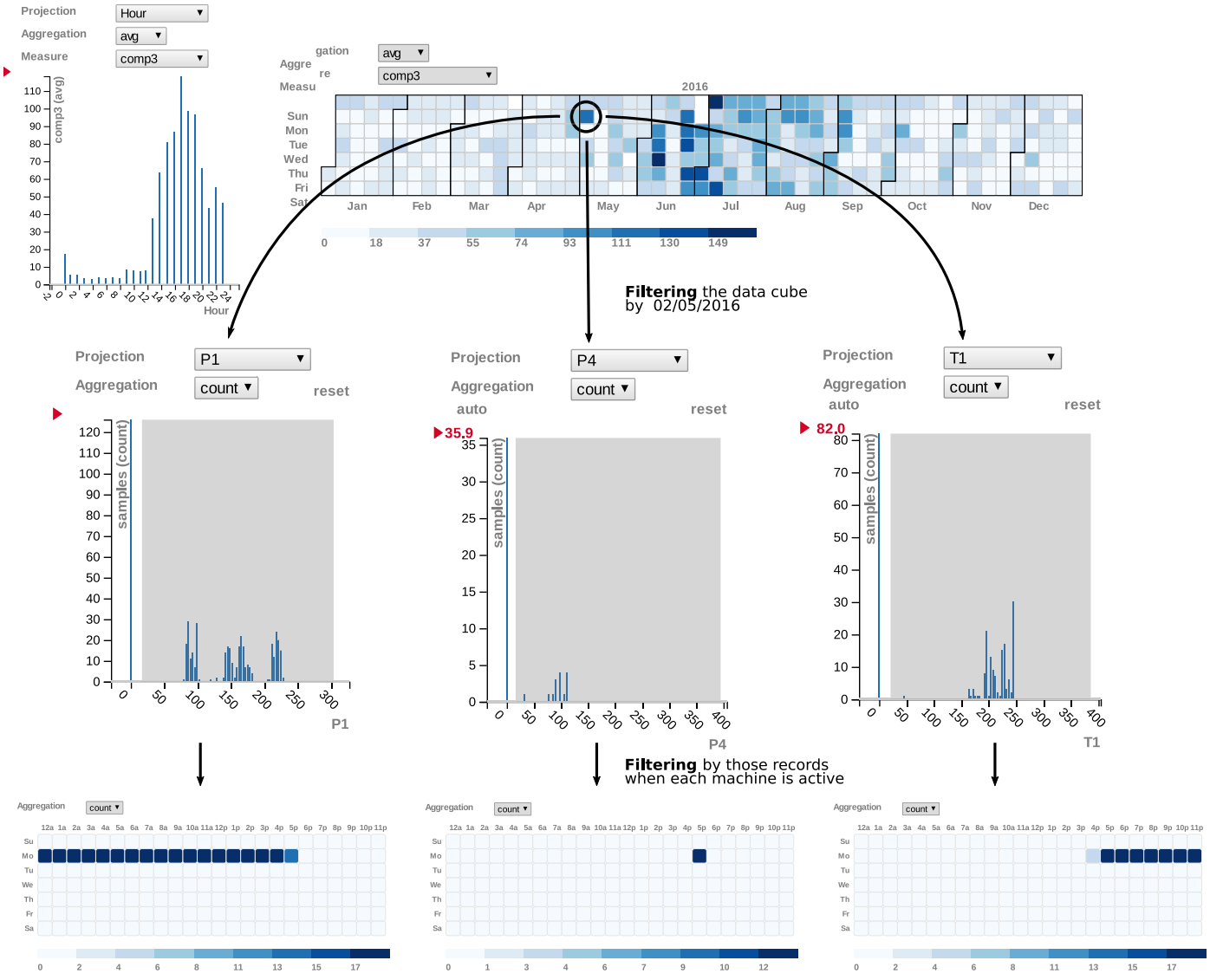

Fig. 13. Searching for the causes of the anomalies. After selecting the second of May, an anomaly in the starting of the cooling machines is discovered by means of filters in the individual consumptions of the cooling machines.

5. Conclusions

This paper presents a novel power demand monitoring approach where intelligent data analysis, data visualization and interaction have been successfully integrated into a web application in which the results of a state-of-the-art NILM technique can be insightfully interpreted, finding relevant hidden patterns in the power demand of a hospital complex. On the one hand, we have shown that NMF is a suitable NILM technique, since the non negativity constraints result in compatible solutions with the underlying physical nature of the problem (positive aggregated demand is a sum of positive demands). Under such constraints, solutions that are not feasible are naturally dismissed, thereby increasing the chances of achieving interpretable parts-of-a-whole description of daily electric power demand. In NMF analysis, it is necessary to indicate the number of factors into which the aggregated consumption will be decomposed. Thus, we suggest that the application provides to users the possibility of defining interactively this parameter. Once the number of factors have been set, the application recalculates the decomposition, showing visually the new results so that the user can evaluate how interpretable is the decomposition according to both the number of chosen components and his own expert knowledge.

On the other hand, it is shown that a large number of factors are needed in order to achieve an accurate decomposition, making the interpretation of the decomposition less insightful. This paper addresses this common issue in NILM decompositions through the interactive data cube approach which allows the user to set and/or modify a broad set of conditions (filters) over both sensed data and NMF components, getting immediate visual feedback on the representation of the results. The fluidity in the filtering mechanisms establishes a continuous loop of analysis whereby the pre-attentive abilities of the visual perception system allow to spot any significant or recognizable relations between sensed data and NMF components. Moreover, the discovered relations are contextualized on a temporal framework thanks to specific visualizations such as calendar or week heatmaps, where the association between them and real consumptions or events in the electric network is improved.

Acknowledgments

The authors would like to thank financial support from the Spanish Ministry of Economy (MINECO), FEDER funds from the

EU under grant DPI2015-69891-C2-1/2-R and from the Principado de Asturias government through the pre doctoral grant “Severo Ochoa”.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.enbuild.2018.06.058.

References

[1] D.A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, H. Ziegler, Visual analytics: scope and challenges, in: Visual Data Mining, Springer, 2008, pp. 76–90.
[2] J.J. Van Wijk, The value of visualization, in: Visualization, 2005. VIS 05. IEEE, IEEE, 2005, pp. 79–86.
[3] H. xiang Zhao, F. Magoulés, A review on the prediction of building energy consumption, Renew. Sustain. Energy Rev. 16 (6) (2012) 3586–3592, doi:10.1016/j.rser.2012.02.049. URL http://www.sciencedirect.com/science/article/pii/S1364032112001438
[4] R.K. Jain, K.M. Smith, P.J. Culligan, J.E. Taylor, Forecasting energy consumption of multi-family residential buildings using support vector regression: investigating the impact of temporal and spatial monitoring granularity on performance accuracy, Appl. Energy 123 (Supplement C) (2014) 168–178, doi:10.1016/j.apenergy.2014.02.057. URL http://www.sciencedirect.com/science/article/pii/S0306261914002013
[5] A.H. Neto, F.A.S. Fiorelli, Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption, Energy Build. 40 (12) (2008) 2169–2176, doi:10.1016/j.enbuild.2008.06.013. URL http://www.sciencedirect.com/science/article/pii/S0378778808001448
[6] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu, C. Toader, Load pattern-based classification of electricity customers, IEEE Trans. Power Syst. 19 (2) (2004) 1232–1239, doi:10.1109/TPWRS.2004.826810.
[7] K. Ehrhardt-Martinez, K.A. Donnelly, S. Laitner, et al., Advanced Metering Initiatives and Residential Feedback Programs: A Meta-review for Household Electricity-saving Opportunities, 2010.
[8] M. Zeifman, K. Roth, Nonintrusive appliance load monitoring: review and outlook, IEEE Trans. Consum. Electron. 57 (1) (2011).
[9] L.K. Norford, S.B. Leeb, Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms, Energy Build. 24 (1) (1996) 51–64, doi:10.1016/0378-7788(95)00958-2.
[10] I. Yarbrough, Q. Sun, D. Reeves, K. Hackman, R. Bennett, D. Henshel, Visualizing building energy demand for building peak energy analysis, Energy Build. 91 (2015) 10–15.
[11] J.J. Van Wijk, E.R. Van Selow, Cluster and calendar based visualization of time series data, in: Information Visualization, 1999.(Info Vis’ 99) Proceedings. 1999 IEEE Symposium on, IEEE, 1999, pp. 4–9.
[12] G.W. Hart, Nonintrusive appliance load monitoring, Proc. IEEE 80 (12) (1992) 1870–1891.
[13] A. Zoha, A. Gluhak, M. Imran, S. Rajasegarar, Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey, Sensors (Switz.) 12 (12) (2012) 16838–16866, doi:10.3390/s121216838. Cited By 127
[14] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791.
[15] W. Gans, A. Alberini, A. Longo, Smart meter devices and the effect of feedback on residential electricity consumption: evidence from a natural experiment in Northern Ireland, Energy Econ. 36 (2013) 729–743.
[16] J. Carroll, S. Lyons, E. Denny, Reducing household electricity demand through smart metering: the role of improved information about energy saving, Energy Econ. 45 (2014) 234–243.
[17] E. Wachsmuth, M. Oram, D. Perrett, Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque, Cereb. Cortex 4 (5) (1994) 509–522.
[18] S.E. Palmer, Hierarchical structure in perceptual representation, Cogn. Psychol. 9 (4) (1977) 441–474.
[19] V.P. Pauca, F. Shahnaz, M.W. Berry, R.J. Plemmons, Text mining using non-negative matrix factorizations, in: Proceedings of the 2004 SIAM International Conference on Data Mining, SIAM, 2004, pp. 452–456.
[20] C. Rösch, T. Kohajda, S. Röder, M. von Bergen, U. Schlink, Relationship between sources and patterns of VOCs in indoor air, Atmos. Pollut. Res. 5 (1) (2014) 129–137, doi:10.5094/APR.2014.016. URL http://www.sciencedirect.com/science/article/pii/S1309104215303500
[21] N. Elmqvist, A.V. Moere, H.-C. Jetter, D. Cernea, H. Reiterer, T. Jankun-Kelly, Fluid interaction for information visualization, Inf. Vis. 10 (4) (2011) 327–340.
[22] I. Díaz Blanco, A.A. Cuadrado Vega, D. Pérez López, M. Domínguez González, S. Alonso Castro, M.A. Prada Medrano, Energy analytics in public buildings using interactive histograms, Energy Build. 134 (2017) 94–104, doi:10.1016/j.enbuild.2016.10.026. URL http://www.sciencedirect.com/science/article/pii/S0378778816312439
[23] J. Gray, S. Chaudhuri, A. Bosworth, A. Layman, D. Reichart, M. Venkatrao, F. Pellow, H. Pirahesh, Data cube: a relational aggregation operator generalizing group-by, cross-tab, and sub-totals, Data Min. Knowl. Discov. 1 (1) (1997) 29–53.
[24] A. Datta, H. Thomas, The cube data model: a conceptual model and algebra for on-line analytical processing in data warehouses, Decis. Support Syst. 27 (3) (1999) 289–301.
[25] GitHub, Crossfilter.js project github, 2017. URL https://github.com/crossfilter
[26] L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, Energy Build. 40 (3) (2008) 394–398.
[27] P. Paatero, U. Tapper, Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values, Environmetrics 5 (2) (1994) 111–126, doi:10.1002/env.3170050203. Cited By 1714
[28] D.D. Lee, H.S. Seung, Algorithms for non-negative matrix factorization, in: Advances in Neural Information Processing Systems, 2001, pp. 556–562.
[29] S.A. Vavasis, On the complexity of nonnegative matrix factorization, SIAM J. Optim. 20 (3) (2009) 1364–1377.
[30] C.-J. Lin, Projected gradient methods for nonnegative matrix factorization, Neural Comput. 19 (10) (2007) 2756–2779.
[31] P.O. Hoyer, Non-negative matrix factorization with sparseness constraints, J.Mach.Learn.Res. 5 (November) (2004) 1457–1469.
[32] C. Boutsidis, E. Gallopoulos, Svd based initialization: a head start for nonnegative matrix factorization, Pattern Recognit. 41 (4) (2008) 1350–1362.
[33] N. Gillis, F. Glineur, Using under approximations for sparse nonnegative matrix factorization, Pattern Recognit. 43 (4) (2010) 1676–1687.
[34] N. Elmqvist, J.-D. Fekete, Hierarchical aggregation for information visualization: overview, techniques, and design guidelines, IEEE Trans. Vis. Comput. Graph. 16 (3) (2010) 439–454.
[35] Z. Liu, B. Jiang, J. Heer, imMens: real-time visual querying of big data, in: Computer Graphics Forum, 32, Wiley Online Library, 2013, pp. 421–430.
[36] L. Lins, J.T. Klosowski, C. Scheidegger, Nanocubes for real-time exploration of spatiotemporal datasets, IEEE Trans. Vis. Comput. Graph. 19 (12) (2013) 2456–2465.
[37] T. Munzner, Visualization Analysis and Design, CRC Press, 2014.
[38] J. Mackinlay, Automating the design of graphical presentations of relational information, ACM Trans. Graph. (Tog) 5 (2) (1986) 110–141.