




## Article

# Biclustering of Smart Building Electric Energy Consumption Data

Federico Divina , Francisco A. Gómez Vela  and Miguel García Torres 

Division of Computer Science, Universidad Pablo de Olavide, ES-41013 Seville, Spain;  
fgomez@upo.es (F.A.G.V.); mgarcia@upo.es (M.G.T.)

\* Correspondence: fdivina@upo.es; Tel.: +34-954-977-592

Received: 12 December 2018; Accepted: 3 January 2019; Published: 9 January 2019



**Abstract:** Nowadays, smart buildings can collect data regarding the electric energy consumption, which can then be analyzed to gain insights or to predict or identify abnormal energy consumption. Numerous models are applied to face this problem but they are based on a global point of view and cannot detect local patterns of abnormal consumption. This work lies in the former option, as we propose a way to analyze energy consumption data from smart buildings. In particular, we use energy consumption data collected by various buildings over a five-year period. These data were analyzed to gain insight into the functioning of the considered buildings, with the aim of detecting anomalous situations, which could indicate that some energy usage policy should be changed or that there is a fault in the sensor network. In particular, we propose an approach based on biclustering, which allows obtaining subgroups of buildings that show a similar behaviour over a specific period of time. To the best of our knowledge, this is the first application of biclustering to energy consumption data analysis. Results confirm that the proposed approach can help policy makers in detecting irregular situations, which can provide hints on how to improve the efficiency of buildings.

**Keywords:** smart buildings; biclustering; evolutionary computation

## 1. Introduction

Energy consumption of buildings is receiving more and more attention in today's economies. As buildings represent substantial consumers of energy worldwide, with this trend increasing over the past few decades due to rising living standards, this issue has drawn considerable attention from various stakeholders (e.g., inhabitants, policy makers, and industry). As pointed out in [1], the energy consumption in buildings has experienced an increment of 1.5% per year in Europe and of 1.9% per year in North America. This increment is even more alarming in China, where it has experienced an increment of 10% in the last twenty years [2]. Moreover, as pointed out in [3], it is estimated that the world energy consumption will increase from 549 quadrillion Btu in 2012 to 629 quadrillion Btu in 2020 and then to 815 quadrillion Btu in 2040, a 48% increase (1.4%/year). Non-OECD Asian countries (including China and India) account for more than half of the increase. It follows that actions are required to improve the efficiency of energy consumption in buildings, in order to decrease both the costs associated to it and the environmental impact this consumption has.

In fact, energy consumption is not only important on the economy level, but also for the repercussions that it has on the environment. For example, the European Union has recently issued a directive (European Directive 2009/28/EC [4]) that requires that all EU countries adopt a set of minimum energy efficiency requirements. More specifically, it stipulates a 20% reduction in energy consumption by 2020 relative to 1990 levels along with a 20% reduction in CO<sub>2</sub> emissions as well as 20% of all energy be produced by renewable technologies.

A relevant fraction of worldwide energy consumption is tightly related to indoor systems for residential, commercial, public, and industrial premises; it has been estimated that residential and commercial buildings account today for about 20% of the world's total energy consumption. In this domain, the energy bill due to environmental control, including heating, ventilation, and air conditioning (HVAC), is dominating, especially in the developed countries [3].

Thus, it is important to gain insights regarding the energy consumption of buildings, with the aim of improving their energy efficiency. In this sense, smart buildings [5], i.e., buildings that use sensors, actuators and microchips, to collect data and manage them according to a business functions and services, can help. In fact, such buildings can provide data regarding the electric energy consumption, which can then be used to both improve their efficiency and for automated fault detection and diagnostics.

In the field of analysis of energy consumption of smart buildings, different data analysis techniques are usually applied to help manage them, including classification [6], forecasting [7–9] and clustering techniques [10,11] of energy consumption patterns.

In this sense, it is useful to analyse the data regarding the consumption of energy over a period of time in order to extract useful information. This kind of data is usually organised as a time series. Time series are a common data type and they are widely used in diverse application areas, such as finance, economics, communication, automatic control, and online services, among others [12]. A common analysis technique applied to time series is clustering [12,13]. Clustering time series is to identify the homogeneous groups of time series data based on their similarity. It is an important and useful technique for exploratory study on the characteristics of various groups in a given time series dataset.

However, in some domains, clustering in only one dimension may not be enough. For example, we may consider the energy consumption data of a set of buildings over a specific period of time. In this case, clustering can be applied in order to extract general buildings profiles, i.e., groups of buildings that present the same behaviour over the whole time interval considered. In a similar way, we can extract day profiles, i.e., set of days with a comparable energy consumption common to all the considered buildings.

With biclustering, we can perform clustering in both dimensions. In this way, it is possible to detect groups of buildings that behave in an analogous way during a particular time interval, which does not have to be the whole interval considered. In fact, if data are organized as matrix, where, for instance, rows stand for buildings and columns for days, then a bicluster represents a submatrix, where only the data relative to a subset of buildings are considered over a specific subset of days.

Biclustering based methods have been widely proposed in the literature due to their capacity of discovering groups of instances with a common pattern under a subset of features [14,15]. These approaches have mainly been applied in bioinformatics, and in particular in the context of microarray data. Within this context, their main purpose is to find subsets of genes with a similar biological behaviour under a subset of experimental conditions [16]. To this aim, several approaches have been proposed to analyse gene expression datasets. For example, in [17], the authors proposed a scatter search approach based on linear correlations among genes to find biclusters. In [18], the authors proposed a multi-objective evolutionary algorithm (MOEA) for the detection of biclusters presenting interesting features.

However, biclustering techniques are not applied only within bioinformatics. They have also been used in other fields, such as social network datasets [19] or text mining [20]. For example, in [21], a biclustering method is used to identify the most suitable group of friends of the user in social network datasets.

Within biclustering of biological data, time series datasets have also been addressed (e.g., [22–25]). An interesting work was presented by Lee et al. [26], who proposed and tested a biclustering algorithm for time series, not only based on gene expression datasets, but also on artificial time series and apartment price datasets in metropolitan areas.

In this paper, we propose the use of biclustering methods to analyse energy consumption data. The main goal is to identify groups of buildings that present the same behaviour during a specific period of time. In particular, we are interested in detecting behaviours presenting peaks of consumption, as this may indicate some out of the norm behaviour that should be further investigated.

To do so, we use a biclustering approach based on a multi-objective evolutionary algorithm (MOEA) proposed in [18]. To the best of our knowledge, there are no other biclustering approaches to time series of electric energy consumption data. However, as stated above, biclustering methods have been successfully applied to other time series data, e.g., genetic expression data (e.g., [23,24]). Clustering techniques have been applied to analyse energy consumption dataset (see, for instance [27,28]). As mentioned above, in our proposal, we use an algorithm, called SMOB, that has been previously used on gene expression data [18].

To evaluate the usefulness of our approach, we used data produced by different buildings located in a university campus in Spain. The obtained results show that biclustering methods can help in gathering more information about the energy consumption behaviour of buildings and to detect anomalous situations, e.g., deficient management of the buildings or abnormal functioning of some sensors.

The paper is organised in the following way. In Section 2, we provide a description of the data used, some basic notions of biclustering and of the particular biclustering strategy applied to analyse the data. In Section 3, we describe the experimental setup and provide the results. In Section 4, we draw the main conclusions and identify possible future developments.

## 2. Our Proposal

In this section, we first provide a brief description of the data used in this paper, followed by some basic notions of biclustering and a brief description of the biclustering algorithm used in our proposal.

### 2.1. Data

The data used in this work are related to the electric energy consumption of various buildings located at the campus of the Pablo de Olavide University of Seville, Spain. The geographical location of the Campus is characterized by extremely high temperature, with maximum temperature higher than 40 °C, during the summer months, and by relatively mild winters, with minima of around 0–2 °C and maxima of around 18–20°. This climate is considered Hot-summer Mediterranean climate (Csa) according to the Köppen–Geiger classification. Table 1 shows the average temperatures, the average maximum and minimal temperatures, and rainfall registered between 2011 and 2015, the years in which the data used in this paper were collected.

**Table 1.** Average temperatures and rainfall in Seville.

Year	Avg. Temp.	Avg. Max. Temp.	Avg. Min. Temp.	Rainfall (mm)
2011	19.2	26.4	13.5	369.03
2012	18.7	25.8	12.4	324.13
2013	18.5	25.3	12.6	425.67
2014	18.9	25.6	13.3	636.50
2015	19.1	26.8	13.1	318.46

The campus consists of 33 buildings with different characteristics. However, not all buildings are monitored for energy consumption, and of the ones that are monitored, many are characterized by a low quality of the collected data. For this reason, we selected five buildings, namely those with more higher quality data. A brief description of the selected buildings is provided in Table 2 including year of construction (and refurbishment year if applicable), size and a short description of the purpose of the building.

**Table 2.** Description of the selected buildings. The first column presents the year of construction and the year of the refurbishment if applicable. The second column presents the size in m<sup>2</sup>. The third column presents a short description.

Building	Year (Refurbishment)	Size (m <sup>2</sup> )	Description
4	1956 (2000)	1948.16	A two-story building with four large classrooms
8	1956 (1999)	1948.16	A two-story building with four large classrooms
11	1956 (2005)	5231.01	A four-story building with several classrooms and offices
20	1997 (-)	6926.49	A devoted biology research centre, which contains many biological laboratories and offices.
45	2010 (-)	6992.18	A newly-built two-story building with computer labs, classrooms and offices. This is the largest building.

The five selected buildings automatically perform a read regarding the electric energy consumption every 15 min, giving in this way the possibility of monitoring the energy consumption in real time. It should be noted that the energy consumed includes lighting, air conditioning, office and lab equipment, but only the total energy consumption is registered, meaning that we cannot access finer grained energy consumption data. Moreover, the occupancy rate of the buildings and the external temperature are also monitored every 15 min. The external temperature data are the same for all buildings, while the occupancy rate is registered only for classrooms and computer laboratories, so it does not really reflect the real occupancy of a building, as it does not consider, for example, the occupancy of offices. For this reason, we did not consider these data in our study. Water consumption is also registered, but it is not relevant for this study, and thus it was not taken into account.

The data used in this paper were collected from 2011 to 2015. The resulting data had to be preprocessed, since they presented various issues. First, some reads were missing. In most cases, a series of  $n$  missing reads were followed by the cumulative consumption over that interval. In this case, we computed the average consumption over  $n + 1$  reads and used it for replacing the missing values. After this phase, we calculated the total electric energy consumption for each day, since we considered the daily consumption of the five selected buildings. However, if less than 4 h of reads were available for a day, it was discarded, as it was not considered representative.

After this preprocessing phase, we obtained 16,512 reads per year, i.e., 172 days for each building, and a total of 412,800 reads for the five buildings over the years taken into consideration, i.e., 860 days.

The data were then organized in a matrix, where rows stand for buildings and columns represent days. Notice that the month of August was not included in this study, since during that month the buildings of the campus remain closed, so there would be little meaning in considering it for this study. For the same reason, the Christmas period was also not considered.

Some of the data used in this work are available at [29]. The precise dataset used in this paper is available on demand.

## 2.2. Biclustering

In this section, we present the model of bicluster. In particular, we focus on how to assess the quality of a bicluster, i.e., how to estimate the coherence of the elements contained in the bicluster.

Let  $B = \{b_1, \dots, b_N\}$  be a set of buildings and  $D = \{d_1, \dots, d_M\}$  a set of days. The data can be viewed as an  $N \times M$  energy consumption matrix  $EM$ .  $EM$  is a matrix of real numbers, where each

entry  $e_{ij}$  corresponds to the electric energy consumption (expressed in kWh) of building  $i$  registered on day  $j$ .

Given such an  $N \times M$  matrix  $EM$  representing the data, a bicluster is defined as a  $I \times J$  sub-matrix  $(I, J)$  of  $EM$  that exhibits some coherent tendency, where  $I$  and  $J$  are sets of rows and columns, respectively, and  $|I| \leq |N|$  and  $|J| \leq |M|$ .

We define the volume of a bicluster  $(I, J)$  as the number of elements  $e_{ij}$  such that  $i \in I$  and  $j \in J$ . In our case, rows stand for buildings, while columns stand for days. It follows that a bicluster represents the energy consumption of  $I$  buildings registered over  $J$  days.

To determine whether a bicluster presents a coherent tendency, i.e., whether a group of buildings show a similar electric energy consumption behaviour during the same period of time, we used a measure called Transposed Virtual Error ( $VE^t$ ) [30]. The lower is the value of  $VE^t$ , the better the bicluster is considered.

Before defining  $VE^t$ , one important observation that can be extracted from an analysis of biclusters is that the range of values contained in a bicluster can vary substantially. Therefore, to make an appropriate comparison between each building and the general pattern, a previous standardisation process of the bicluster would enable the values of the bicluster to be scaled to a common range. This mechanism would also be responsible for softening the behaviour of a building, since the most important aspect is to characterise its tendency rather than its numerical values.

Thus, we performed a standardisation of the biclusters  $\mathcal{B}$ , whose elements  $\hat{b}_{ij}$  are obtained as follows:

$$\hat{b}_{ij} = \frac{\hat{b}_{ij} - v_{r_i}}{\sigma_{r_i}}, 1 \leq i \leq |I|, 1 \leq j \leq |J| \quad (1)$$

After having standardised the bicluster, we could proceed to compute  $VE^t$ , which is a measure that allows simultaneously detecting different kinds of interesting patterns in a bicluster, which is an improvement over other measures, such as the Mean Square Residue [31]. The basic idea behind  $VE^t$  is to measure how rows follow the general tendency within the bicluster.

The first step to compute  $VE^t$  of a bicluster  $\mathcal{B}$  is to obtain a new row from  $\mathcal{B}$ , which represents a *virtual pattern* ( $\rho = \{\rho_1, \rho_2, \dots, \rho_J\}$ ), where  $\rho_j$ ,  $1 \leq j \leq J$ , is defined as the mean of the  $j$ th column:

$$\rho_j = \frac{1}{I} \sum_{i=1}^I e_{ij} \quad (2)$$

The virtual pattern symbolizes the general behaviour of the set of buildings for the given bicluster. With  $\rho$ , we can compute how a specific building follows the general tendency of the bicluster by computing the differences between the energy consumption of building  $i$  and the values of  $\rho$  for each column of the bicluster. The  $VE^t$  of a bicluster  $\mathcal{B}$  is then computed as follows:

$$VE^t(\mathcal{B}) = \frac{1}{I \cdot J} \sum_{i=1}^I \sum_{j=1}^J |\hat{b}_{ij} - \hat{\rho}_j| \quad (3)$$

where  $\hat{\rho}_j$  is the standardised virtual condition and  $\hat{b}_{ij}$  is the standardised element.

$VE^t$  has been proven to be zero for biclusters containing perfectly coherent patterns.

A part from a coherent tendency, we were also interested in finding biclusters with fluctuating patterns yet coherent trends, since such bicluster may shows behaviour out of the normal and thus indicating a situation that has to be further studied, or a possible malfunctioning of some reading sensors. For this reason, we used the row variance as an accompanying score. The *row variance* of a bicluster  $\mathcal{B}$  is defined as:

$$var(\mathcal{B}) = \frac{\sum_{i \in I, j \in J} (e_{ij} - e_{iJ})^2}{|I| \cdot |J|} \quad (4)$$

where  $e_{ij}$  corresponds to the mean of the  $i$ th row of  $\mathcal{B}$ . Notice that a bicluster with row variance equal to zero is a constant bicluster.

Our goal was to find biclusters of maximum size, with low  $VE^t$  and with a relatively high row variance.

In fact, the problem cannot be addressed by only optimizing the  $VE^t$  of biclusters. In fact, this approach may lead to the discovery of uninteresting biclusters. For instance, flat biclusters will have a low value of  $VE^t$ , or, again, biclusters containing few days will typically have lower values of  $VE^t$ , if compared to biclusters characterized by higher volume. This is because the more days and buildings are contained in a bicluster, the less likely the behaviour of the buildings is to be similar. Such biclusters are not very interesting for detecting anomalies, and, to solve this issue, other properties of the biclusters are usually optimized, e.g., the volume.

In particular, we were interested in finding biclusters with high volume, good quality (being quality measured by an appropriate metric such as  $VE^t$ ) and relatively high variance. Thus, we could individuate at least three objectives to be optimized and these objectives were usually in conflict with each other. For example, a bicluster consisting of just one element has  $VE^t$  equal to zero, or, again, a constant bicluster has gene variance equal to zero, but also  $VE^t$  equal to zero.

For this reason, the problem of finding biclusters can be straightforwardly seen as a multi-objective problem. Moreover, by addressing this problem as a multi-objective problem, it is not necessary to combine all the objectives into single cost function, which might become complicated, especially when both maximization and minimization are involved. Finding a way to combine the objectives in a single function can be problematic, and may require more parameters to set [32].

Moreover, we wanted to minimize the overlapping among the biclusters discovered. In this way, we could find groups of buildings that present the same variations regarding the energy consumption in the same period of time. This could discover particular patterns or anomalies that would help in taking action aimed at improving the efficiency of buildings.

### 2.3. Sequential Multi-Objective Biclustering

In this section, we provide a short description of the algorithm used in this paper. The algorithm is called SMOB (Sequential MultiObjective Biclustering), and for a more detailed description of the algorithm, we refer the reader to [18,33].

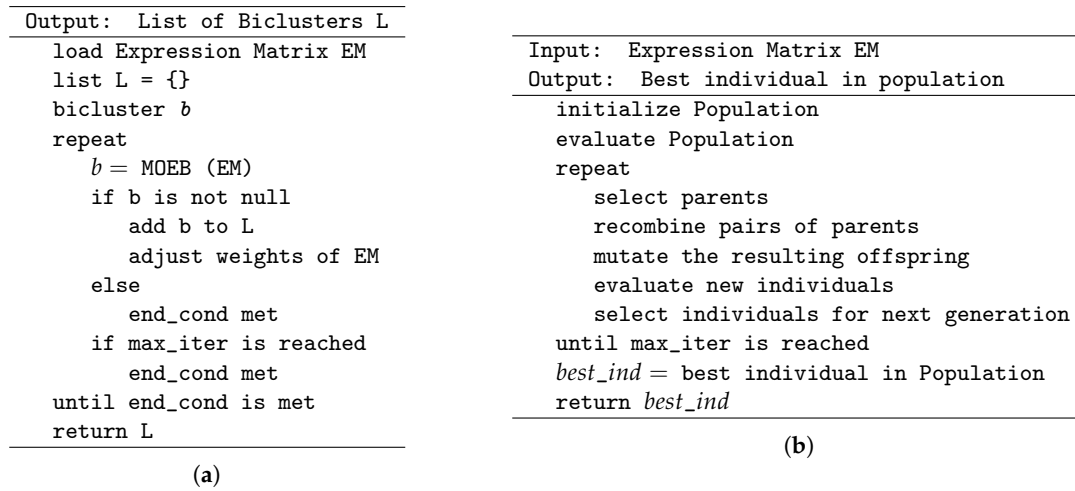
SMOB is a multi-objective EA and, in this study, we considered three objectives to be optimized:  $VE^t$ , the row variance and the volume of the biclusters. Thus, we aimed at finding biclusters with a coherent tendency, with great volume and presenting fluctuating patterns. Notice that, in this work, we used  $VE^t$ , but it would be trivial to use any other quality measure.

SMOB adopts a sequential covering strategy, as depicted in Figure 1a. A MOEA (whose scheme is reported in Figure 1b) is called  $n$  times, and each time a bicluster is returned. The returned bicluster is stored in a list  $L$  that contains all the biclusters found so far. To avoid overlapping among biclusters, we associated a weight  $w(e_{ij})$  with each element  $e_{ij}$  of the expression matrix. These weights are adjusted right after a bicluster is returned.  $w(e_{ij})$  is equal to the number of biclusters stored in  $L$  that contain  $e_{ij}$ . When evaluating a bicluster, the weights of its elements are used in order to penalize biclusters overlapping with elements of  $L$ , as it explained in the following. Notice that the sequential coverage strategy adopted is such that the order in which biclusters are discovered does not reflect their quality nor their relevance.

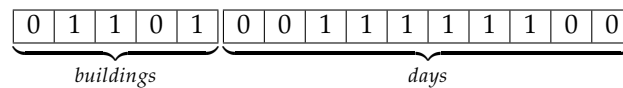
In the algorithm, each individual encodes a single bicluster. The encoding of biclusters is the one proposed in [18], where bit strings are evolved. However, the mutation and crossover operators differ from that approach, as it explained below. A bit is associated to each building (row) and each day (column) considered. If a bit is set to one, it means that the relative building/day belongs to the bicluster, otherwise it does not. Figure 2 shows an example of an individual representing a bicluster with three buildings (rows) and six days (columns). Notice that the days are consecutive. This is a necessary condition we imposed on individuals, since we were considering a time series, and having a



biclusters with no consecutive days would have little meaning for our purpose. This is a difference with respect to the algorithm proposed in [33].



**Figure 1.** (a) The scheme of Sequential Multi-Objective Biclustering (SMOB) is shown. A scheme of the multi-objective evolutionary algorithm used for finding biclusters is shown (b).



**Figure 2.** Example of an individual encoding a biclusters containing three buildings and six consecutive days.

As mentioned above, since we considered time series data, we adapted the mutation and crossover operators used in [18], as we were interested in biclusters containing consecutive days (or almost consecutive, since after the preprocessing phase of the data, some days may be missing, and thus some gaps can be created). However, an individual is considered valid if the columns contained in the encoded biclusters are consecutive. For this reason, the crossover will act only on genes representing the rows, i.e., the buildings, since rows do not have to be consecutive. Mutation may add or delete a row or a column, with the restriction that columns must be consecutive. The columns contained in a bicluster will always be consecutive, meaning that there will be a set of consecutive 1 s in the individual. Mutation will only act at the beginning or the end of such interval. Mutation on the columns will have an higher probability that on the rows, since the number of columns is much higher than the number of rows.

Tournament selection was used, and selected individuals undergo crossover and mutation. Elitism was applied by letting the non-dominated individuals survive to the next generation. Individuals were evaluated using a strategy similar to NSGA-II [34]. Individuals were divided into different non-dominated fronts, and individuals belonging to the same front had the same starting fitness  $rank(x)$ . For instance, a non-dominated individual  $x_1$  will have  $rank(x_1) = 0$ .

The fitness of an individual  $x$  is defined as:

$$f(x) = rank(x) + sh(x) + P(x) \quad (5)$$

where  $sh(x)$  is the phenotypic sharing [35] and  $P(x) = 1 - \frac{V(x) - \sum_{i,j \in x} w(e_{ij})}{V(x)}$ , where  $V(x)$  is the volume of  $x$ . In our implementation,  $sh(x)$  was the minimal Euclidean distance, computed on the objectives to the other individuals. Thus,  $\forall y \neq x : sh(x) = \min \left( \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \right)$ , where  $n$  is the number of objectives,  $y$  is an individual, and  $x_i, y_i$  are the objectives used.

As mentioned above in this section,  $P(x)$  was used to avoid overlapping among biclusters. From the definition of  $P(x)$ , it follows that, if a bicluster has low volume and it covers elements of the expression matrix that are already contained in many biclusters already found,  $P(x)$  will be high. On the other hand, if the bicluster has a high volume and it overlaps with few biclusters, the penalty will be lower. If the bicluster  $x$  does not overlap with any bicluster, then  $P(x)$  is zero.

### 3. Experimental Results

In this section, we present a summary of the most representative results obtained on the data presented in Section 2.1.

Firstly, we present the general behaviour of the five selected buildings in order to identify some of their interesting features. Table 3 shows the average electric energy consumption of the considered buildings for each year. This information can provide us an idea of the kind of energy consumption profile each building presents. From the presented table, we can conclude that Building 20 exhibits the highest energy consumption. This was expected, since this building hosts many biological laboratories, which require a high amount of energy to operate, and also for its size (one of the largest buildings in the experiment, see Table 2 for more details). It can also be noticed that Building 8 exhibits the lower energy consumption. It is worth mentioning the difference, in terms of energy consumption, between this building and Building 4 that is a similar building (see Table 2). On the other hand, Buildings 11, 45 and 4 present a similar general consumption rate, even if their characteristics are very different, being Buildings 45 and 11 much larger.

**Table 3.** Mean electric energy consumption information per building, in kWh. In the last column, the five-year average consumption is presented in terms of Kwh/m<sup>2</sup>.

Building	2011	2012	2013	2014	2015	Average (Kwh/m <sup>2</sup> )
4	7.87	7.31	6.36	8.13	9.35	$4.00 \times 10^{-3}$
8	1.29	1.50	1.47	1.33	1.35	$7.12 \times 10^{-4}$
11	8.65	10.62	7.52	5.85	5.56	$1.46 \times 10^{-3}$
20	26.98	24.65	23.39	22.11	29.87	$3.66 \times 10^{-3}$
45	6.85	8.95	8.11	8.49	9.53	$1.19 \times 10^{-3}$

With the aim to find some general consumption profiles, a preliminary study on these buildings was also conducted using a clustering technique. In particular, we applied the well-known k-means clustering algorithm with the aim of finding electric energy consumption profiles for the buildings of the campus. We selected three clusters for k-means clusters since there are three different types of buildings, so we expect that they will be grouped together. The conclusions regarding the considered buildings were that only Buildings 4 and 11 were grouped together, while the others were found to have a overall different profile. Again, it is surprising that Building 8 is not grouped together with Buildings 4 and 11, as the characteristics of the three buildings are similar (see Table 2).

To obtain biclusters, we applied SMOB to the five selected buildings for considering the five years, and we extracted 45 biclusters. We also ran experiments considering each year separated, but the results are similar to the ones considering the whole year range, and, for this reason, as well as for reasons of space, they are not reported in this paper.

The parameter of the algorithm used in the experiments are summarized in Table 4, which were tuned after performing various preliminary results.

The average features of the 45 biclusters found on the data are presented in Table 5.

We can notice that, on average, four buildings were grouped together for an average period of 51 days (see Table 5). This result is surprising, considering that at least two buildings (Buildings 20 and 8) show average consumptions that are significantly different from the other three buildings (see Table 3), and that also the characteristics and the use of the buildings are different. This results show that, even if the general energy consumption profile of the buildings is different, there are period of



time where buildings present the same behaviour. Such results suggests that the managers of the buildings should check the energy consumption pattern of the buildings more closely, to notice out of the normal behaviours that could yield a unnecessary use of energy or management issues.

**Table 4.** Parameters used in the experiment.

Parameter	Value
Generations	100
Population size	200
Tournament size	4
Crossover probability	0.85
Mutation probability	0.2

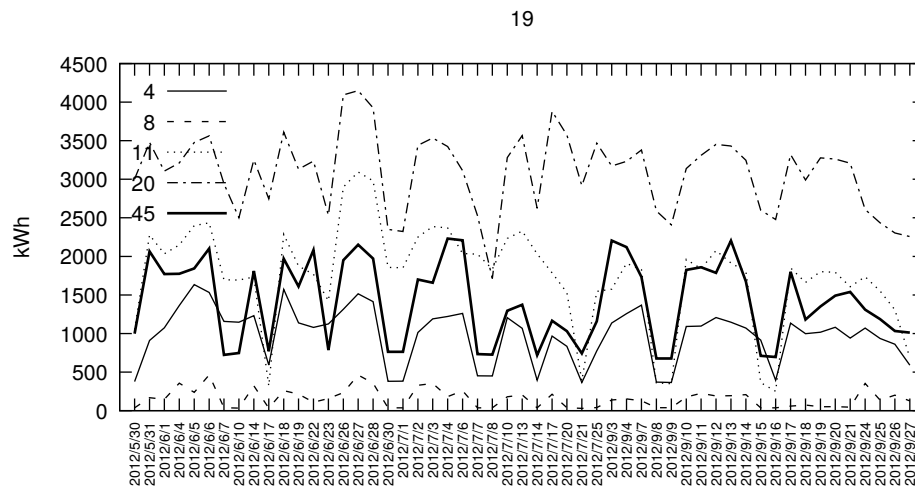
**Table 5.** Average information for the 45 biclusters found on the data. Standard deviation between brackets.

Measure	Average
$VE^t$	0.17 (0.11)
Var	$7.37 \times 10^5$ ( $1.14 \times 10^6$ )
Volume	210.38 (56.43)
Buildings	4.07 (0.69)
Days	51.38 (11.13)

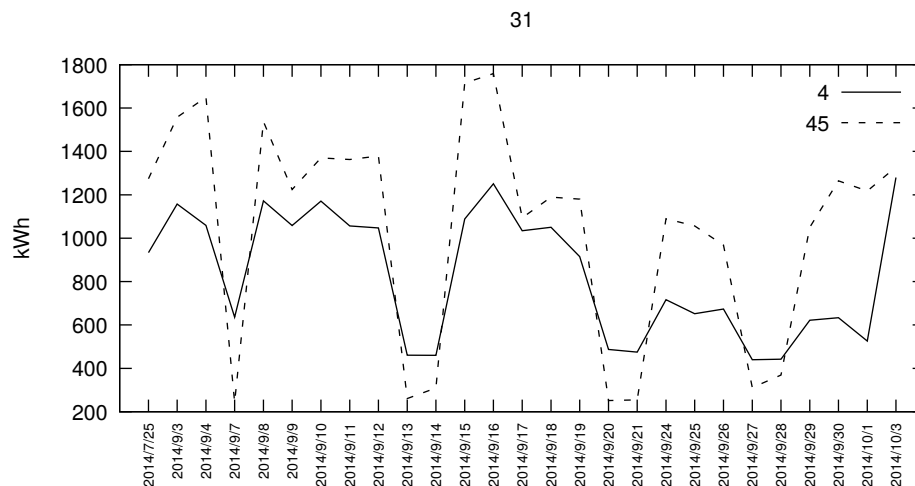
We can test if this hypothesis is valid by checking more closely some of the biclusters found. In this sense, we have selected four representative biclusters to show how this method is able to identify abnormal behaviours in a period of time.

For example, considering Bicluster 19 depicted in Figure 3, we can notice that on 7–8 July, which was a weekend, the consumption of Building 20 decreased to the levels of Building 11 (17.13 kWh). This is odd, since in Buildings 20 there are many biological laboratories, which need constant energy to, for example, maintain the temperature. In contrast, in Building 11 there are only offices and classrooms, which on a Saturday should not consume much energy. Moreover, on the day after, a Sunday, the consumption of Building 11 was higher than that of Building 20. This fact represent an abnormal consumption for Building 20 since the average consumption during weekends in July through the years is 25.61 kWh, which is significantly higher than the consumption depicted in the bicluster. Moreover, on 14 July, a Sunday, we can see that the consumption of Building 11 was lower than that of building 20, which is the expected behaviour and can be found again on 21 July, which was a Saturday.

Bicluster 31 (Figure 4) also presents some facts that are worth noticing. In this bicluster, only two buildings, namely Buildings 4 and 45, are included. These are very different buildings (see Table 2), being Building 4 a smaller building, while Building 45 is much bigger and has many computer laboratories. From our preliminary clustering study, they were considered to have an overall different energy consumption profile. This result shows how a biclustering method is able to find similar consumption patterns over certain periods of time, even when the overall consumption characteristics of each building are different.



**Figure 3.** Representation of Bicluster 19. In this figure, the abnormal consumption of Building 20 on 7–8 July period is depicted.

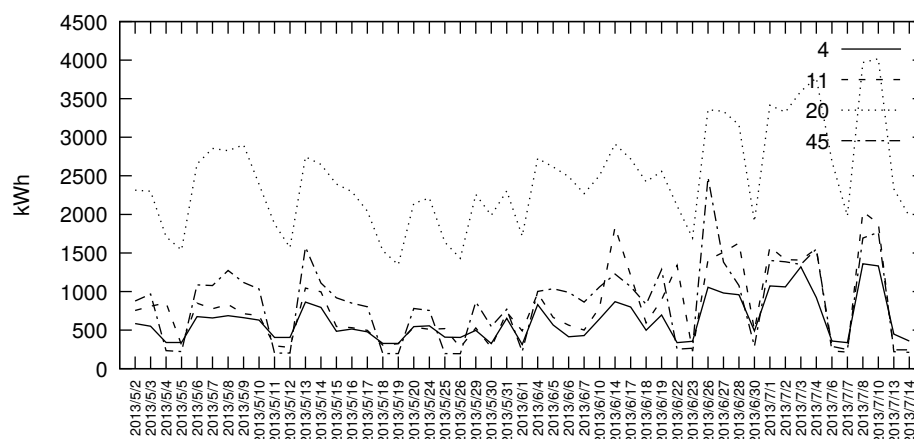


**Figure 4.** Representation of Bicluster 31. In this biclustering, is worth mentioning that two different buildings present a similar consumption pattern during a period of time.

Moreover, it can be observed that the behaviours of the two buildings captured in this bicluster are similar; the energy consumption of the two buildings drops during weekends (for example on 7, 13 and 20 September). However, it is strange that exactly during these periods, the consumption of Building 45 is lower than that of Building 4. This may reflect a bad management of Building 4, or may be due to the fact that Building 45 is newer and better equipped as far as energy efficiency is concerned.

The same kind of behaviour can be found in Bicluster 35 (see Figure 5), where, in addition, we can noticed that Building 4 presents peaks of consumption that are higher than those of Buildings 45 and 11, which are bigger buildings. This is remarkable, as it represents an anomalous energy consumption given the size of the experimental buildings.

35

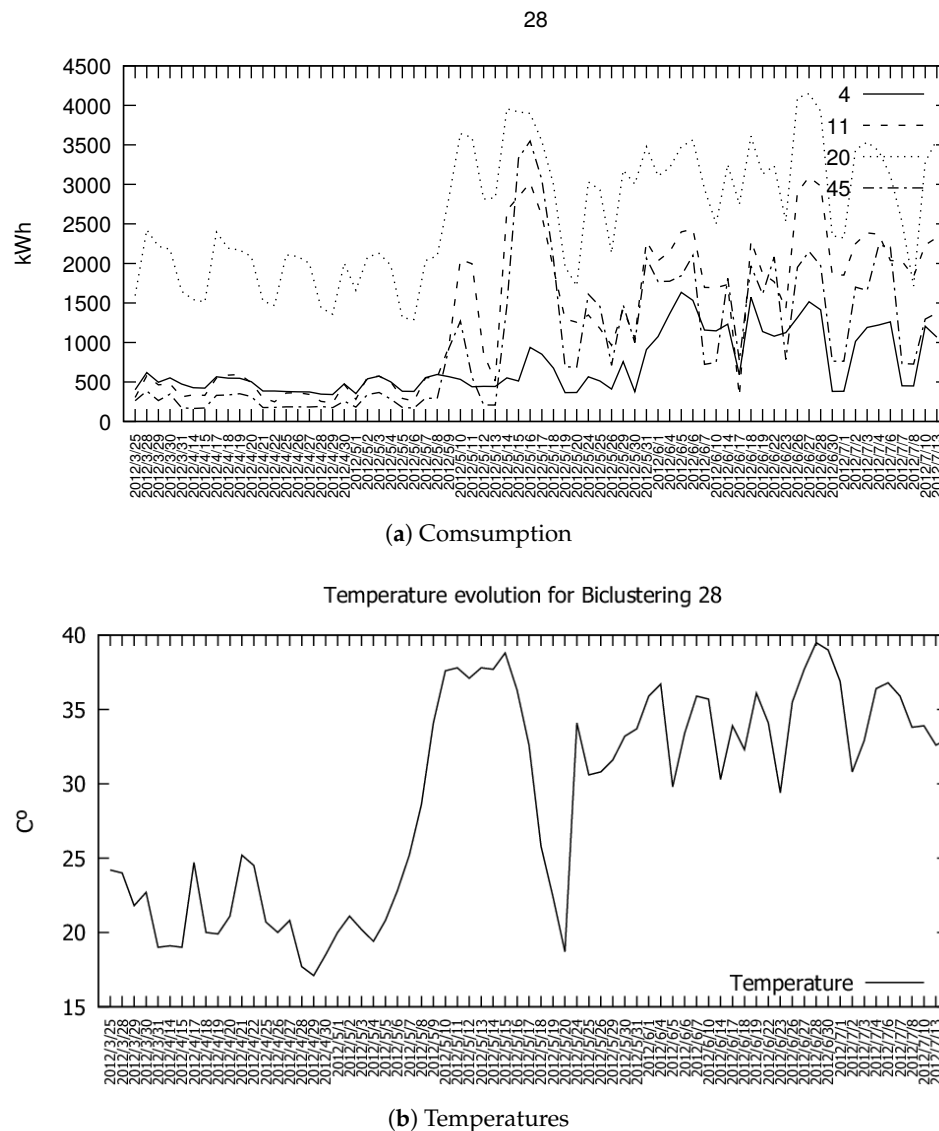


**Figure 5.** Representation of Bicluster 35. In this figure, the efficiency of Building 45 is shown during the weekends.

Another observation is that, in general, Building 45 is more efficient than even building 11, which does not contain any laboratories. From this bicluster, it seems that Building 45 is more efficient, especially during weekends, than the other buildings observing the amount of energy consumed over the weekend (the consumption plates in the figure on weekends). This fact highlights the relevance of good buildings management and the need to upgrade energy efficiency.

Finally, we would like to highlight another interesting fact. By visually inspecting Bicluster 28 (see Figure 6a), we can also notice that there is an unusual activity after 8 May 2012. We can immediately notice a peak of energy consumption during 13–19 May for Buildings 45 and 11. Buildings 4 and 20 also present a peak, but less pronounced. Checking the evolution of the temperatures (depicted in Figure 6b), a sudden rise in temperatures was observed during this period. As shown in Figure 6a,b, the consumption of energy by the four buildings presented has a direct correlation with the evolution of the temperatures in this period. Thus, thanks to this result, we can argue that temperatures have a direct impact on electricity consumption on campus, especially during sudden changes in temperatures (caused because of an increment of the use of air conditioning).

Other similar observations can be extracted from the other biclusters. This observations, could help the managers of the buildings to take actions aimed at improving the efficiency of the buildings, which can substantially reduce the maintaining costs relative to these buildings and their impact on the environment. With biclustering, we can analyse the behaviour of buildings during a specific period of time. As the results presented have shown, we can detect peaks of consumption that may not be considered normal during a specific time interval. This is relevant since these peaks could be due to a non-optimal management of the energy consumption or some kind of problem in the management. Thus, we can conclude that the biclustering methods for the analysis of energy consumption data are a new tool in the energy management.



**Figure 6.** (a) Biclustering 28 representation, where it is possible to check that from 13 to 19 May there was a large increase in consumption. (b) The temperature evolution in the same period. It is possible to see how there is a direct correlation between the consumption and the evolution of the temperatures.

#### 4. Conclusions and Future Works

In this paper, we have presented an analysis based on biclustering of data, generated over five years, regarding the energy consumption relative to different buildings located on a university campus. To the best of our knowledge, no other biclustering approach has been applied before to this kind of data. The data were generated by sensors that provide a read of electric energy consumption every 15 min.

The aim of this study was to be able to detect functioning patterns in a more refined way than what simple clustering could allow. In fact, clustering would allow us to detect global behaviours, i.e., if some buildings present the same energy consumption profile over the whole time period considered. Instead, biclustering techniques allow discovering buildings showing a similar behaviour during a specific period of time, which does not necessarily correspond to the whole period considered. In this way, we can detect peaks of consumption that may not be considered normal during a specific time interval. This is relevant since such peaks could be due to non-optimal management of the energy consumption or to some kind of problem in the buildings. In both cases, the analysis of the resulting biclusters can help the building managers to improve the efficiency of the buildings, reducing, in this

way, both the costs associated to the energy consumption and CO<sub>2</sub> emissions. Results obtained on the dataset used in this paper show that biclustering can allow drawing such conclusions.

In this paper, we have shown how biclustering methods could fit to study concrete periods of time on energy consumption time series. In particular, through the analysis of the results, it is possible to check how biclusters offer different and more precise conclusions than clustering methods, which can provide global results that may be insufficient in some cases, as discussed in this paper. In this sense, it is possible to argue that the use of biclustering for this kind of analyses is more appropriate than traditional clustering methods on some occasions, such as temporary malfunction or bad energy saving policy.

As future works, we intend to apply the same kind of analysis to more information (regarding the consumption of 2016, which has been recently published), since it should be more precise and contain less noise.

In the study presented in this paper, we have not included data relative to holiday periods, since our main aim was to gain insights regarding the normal use of the buildings. However, in future works, we are planning to include such periods to verify the energy consumption behaviour of the buildings without students. With such data, we may be able, for instance, to detect malfunctions in the building management system with clock-based controls of ventilation and air conditioning. Such situations may be difficult to detect during a normal use of the buildings, since controls based on indoor climate indicators such as CO<sub>2</sub> and temperature could be triggered if there were people present in the buildings. Moreover, we are planning to extend the study to other sources of available data. We also have to include a detailed post-processing analysis considering the occupation rate of the buildings.

We believe that the results obtained can help managers in improving the efficiency of buildings, reducing in this way the associated costs.

**Author Contributions:** conceptualization, F.D.; methodology, F.D.; software, F.D.; validation, F.D., F.A.G.V. and M.G.T.; formal analysis, F.D. and F.A.G.V.; investigation, F.D. and M.G.T.; resources, F.D.; data curation, F.D.; writing—original draft preparation, F.D., F.A.G.V. and M.G.T.; writing—review and editing, F.D., F.A.G.V. and M.G.T.; visualization, F.D. and F.A.G.V.; supervision, F.D.; project administration, F.D.; funding acquisition, F.D., F.A.G.V. and M.G.T.

**Funding:** This research was funded by Spanish Ministry of Economic and Competitiveness and the European Regional Development Fund, grant number TIN2015-64776-C3-2-R (MINECO/FEDER).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Yu, Z.; Haghighat, F.; Fung, B.C.; Yoshino, H. A decision tree method for building energy demand modeling. *Energy Build.* **2010**, *42*, 1637–1646. [CrossRef]
2. Cai, W.; Wu, Y.; Zhong, Y.; Ren, H. China building energy consumption: Situation, challenges and corresponding measures. *Energy Policy* **2009**, *37*, 2054–2059. [CrossRef]
3. U.S. Energy Information Administration—International Energy Outlook. 2016. Available online: <https://www.eia.gov/outlooks/ieo/index.php> (accessed on 12 November 2018).
4. Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009 on the Promotion of the Use of Energy From Renewable Sources And Amending and Subsequently Repealing Directives 2001/77/EC and 2003/30/EC; Official Journal of the European Union: 2009; L140. Available online: <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:140:0016:0062:en:PDF> (accessed on 16 November 2018).
5. Snoonian, D. Smart buildings. *IEEE Spectr.* **2003**, *40*, 18–23. [CrossRef]
6. Li, X.; Bowers, C.P.; Schnier, T. Classification of energy consumption in buildings with outlier detection. *IEEE Trans. Ind. Electron.* **2010**, *57*, 3639–3644. [CrossRef]
7. Hernandez, L.; Baladron, C.; Aguiar, J.M.; Carro, B.; Sanchez-Esguevillas, A.J.; Lloret, J.; Massana, J. A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 1460–1495. [CrossRef]

8. Jain, R.K.; Smith, K.M.; Culligan, P.J.; Taylor, J.E. 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* **2014**, *123*, 168–178. [[CrossRef](#)]
9. Divina, F.; Gilson, A.; Gómez-Vela, F.; García Torres, M.; Torres, J.F. Stacking Ensemble Learning for Short-Term Electricity Consumption Forecasting. *Energies* **2018**, *11*, 949. [[CrossRef](#)]
10. Capozzoli, A.; Lauro, F.; Khan, I. Fault detection analysis using data mining techniques for a cluster of smart office buildings. *Expert Syst. Appl.* **2015**, *42*, 4324–4338. [[CrossRef](#)]
11. Fan, C.; Xiao, F.; Wang, S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Appl. Energy* **2014**, *127*, 1–10. [[CrossRef](#)]
12. Liao, T.W. Clustering of time series data—A survey. *Pattern Recognit.* **2005**, *38*, 1857–1874. [[CrossRef](#)]
13. Aghabozorgi, S.; Shirkhorshidi, A.S.; Wah, T.Y. Time-series clustering—A decade review. *Inf. Syst.* **2015**, *53*, 16–38. [[CrossRef](#)]
14. Tanay, A.; Sharan, R.; Shamir, R. Biclustering algorithms: A survey. In *Handbook of Computational Molecular Biology*; CRC Press: Boca Raton, FL, USA, 2005; Volume 9, pp. 122–124.
15. Pontes, B.; Giráldez, R.; Aguilar-Ruiz, J.S. Biclustering on expression data: A review. *J. Biomed. Inform.* **2015**, *57*, 163–180. [[CrossRef](#)] [[PubMed](#)]
16. Anitha, S.; Chandran, C. Review on Analysis of Gene Expression Data Using Biclustering Approaches. *Bonfring Int. J. Data Min.* **2016**, *6*, 16. [[CrossRef](#)]
17. Nepomuceno, J.A.; Troncoso, A.; Aguilar-Ruiz, J.S. Scatter search-based identification of local patterns with positive and negative correlations in gene expression data. *Appl. Soft Comput.* **2015**, *35*, 637–651. [[CrossRef](#)]
18. Divina, F.; Pontes, B.; Giráldez, R.; Aguilar-Ruiz, J.S. An effective measure for assessing the quality of biclusters. *Comput. Biol. Med.* **2012**, *42*, 245–256. [[CrossRef](#)]
19. Gnatyshak, D.V.; Ignatov, D.I.; Semenov, A.; Poelmans, J. Analysing online social network data with biclustering and triclustering. Concept Discovery in Unstructured Data. In Proceedings of the 2nd International Workshop, CDUD 2012, Leuven, Belgium, 6–10 May 2012; pp. 30–39.
20. Li, F.; Li, M.; Guan, P.; Ma, S.; Cui, L. Mapping publication trends and identifying hot spots of research on internet health information seeking behavior: A quantitative and co-word biclustering analysis. *J. Med. Internet Res.* **2015**, *17*. [[CrossRef](#)]
21. Sun, Z.; Han, L.; Huang, W.; Wang, X.; Zeng, X.; Wang, M.; Yan, H. Recommender systems based on social networks. *J. Syst. Softw.* **2015**, *99*, 109–119. [[CrossRef](#)]
22. Eren, K.; Deveci, M.; Küçüktunç, O.; Çatalyürek, Ü.V. A comparative analysis of biclustering algorithms for gene expression data. *Briefings Bioinform.* **2012**, *14*, 279–292. [[CrossRef](#)]
23. Xue, Y.; Liao, Z.; Li, M.; Luo, J.; Hu, X.; Luo, G.; Chen, W.S. A new biclustering algorithm for time-series gene expression data analysis. In Proceedings of the 2014 Tenth International Conference on Computational Intelligence and Security (CIS), Kunming, China, 15–16 November 2014; pp. 268–272.
24. Denitto, M.; Farinelli, A.; Bicego, M. Biclustering of Time Series Data Using Factor Graphs. In Proceedings of the SAC '17 Symposium on Applied Computing, Marrakech, Morocco, 3–7 April 2017; pp. 28–30. [[CrossRef](#)]
25. Carreiro, A.V.; Anunciação, O.; Carriço, J.A.; Madeira, S.C. Prognostic prediction through biclustering-based classification of clinical gene expression time series. *J. Integr. Bioinform.* **2011**, *8*, 73–89. [[CrossRef](#)]
26. Lee, J.H.; Lee, Y.R.; Jun, C.H. A biclustering method for time series analysis. *Ind. Eng. Manag. Syst.* **2010**, *9*, 131–140. [[CrossRef](#)]
27. Benítez, I.; Quijano, A.; Díez, J.L.; Delgado, I. Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers. *Int. J. Electr. Power Energy Syst.* **2014**, *55*, 437–448. [[CrossRef](#)]
28. Ahmad, A.; Hassan, M.; Abdullah, M.; Rahman, H.; Hussin, F.; Abdullah, H.; Saidur, R. A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renew. Sustain. Energy Rev.* **2014**, *33*, 102–109. [[CrossRef](#)]
29. Electrical Energy Consumption Data of the Pablo de Olavide University. Available online: <https://datos.upo.gob.es/dataset/?id=consumo-de-electricidad> (accessed on 5 September 2018)).
30. Pontes, B.; Giráldez, R.; Aguilar-Ruiz, J.S. Quality Measures for Gene Expression Biclusters. *PLoS ONE* **2015**, *10*, e0115497. [[CrossRef](#)] [[PubMed](#)]
31. Cheng, Y.; Church, G.M. Biclustering of Expression Data. In Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology, San Diego, CA, USA, 19–23 August 2000; pp. 93–103.



32. Corne, D.; Deb, K.; Fleming, P.J. The good of the many outweighs the good of the one: Evolutionary multi-objective optimization. *IEEE Connect. Newsl.* **2003**, *1*, 9–13.
33. Divina, F.; Aguilar-Ruiz, J.S. A multi-objective approach to discover biclusters in microarray data. In Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation, London, UK, 7–11 July 2007; pp. 385–392.
34. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evolut. Comput.* **2002**, *6*, 182–197. [[CrossRef](#)]
35. Zitzler, E.; Thiele, L. Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Evolutionary Algorithm. *IEEE Trans. Evolut. Comput.* **1999**, *3*, 257–271. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).