

# Identifying services for short-term load forecasting using data driven models in a Smart City platform.

Joaquim Massana, Carles Pous, Llorenç Burgas, Joaquim Melendez, Joan Colomer

*University of Girona, Campus Montilivi, P4 Building,  
Girona, E17071, Spain,*

*{joaquim.massana, carles.pous, llorenc.burgas, joaquim.melendez, joan.colomer}@udg.edu*

---

## Abstract

The paper describes an ongoing work to embed several services in a Smart City architecture with the aim of achieving a sustainable city. In particular, the main goal is to identify services required in such framework to define the requirements and features of a reference architecture to support the data-driven methods for energy efficiency monitoring or load prediction. With this object in mind, a use case of short-term load forecasting in non-residential buildings in the University of Girona is provided, in order to practically explain the services embedded in the described general layers architecture. In the work, classic data-driven models for load forecasting in buildings are used as an example.

*Keywords:* Short-term load forecasting, Data mining, Services, Building, Smart city architecture

---

## 1. Introduction

The concept of Smart City appears due to the mobilization of people to the cities. This increase of people has an impact on city services such as transportation, utilities, communications, waste management, health services and much other. In order to avoid services degradation, and have an idea of the effect of such increase of people for a particular service, it is necessary to manage each service by constantly monitoring it. Therefore, it is needed to provide the system with mechanisms for collecting data. This is the first step towards getting to a Smart City. But what it really makes the city smart is

to process and analyse the data and returns as response some kind of action to ensure the provision of services at satisfactory levels of quality. Hence, it is necessary to integrate these monitoring devices with the applications that perform the analysis of this data and are able to provide an action [12].

The synergy of computational and physical components, specifically the use of cyber-physical systems (CPSs), led to the advancement of such integration. At different scale, neighbourhoods, communities or buildings can also be considered large CPS continuously operated accordingly to demand affected by the activities of users. As important is to know physical system constraints as consumers behaviour, and interactions between both. Major Information and Communication Technology (ICT) vendors have made efforts for developing Smart City transversal platforms oriented to integrate city information and making it available to end-users. On the other hand, the utilities (water, electricity, gas, etc.) have their proprietary solutions specifically designed to operate and supervise these infrastructures and providing managing and billing services. This work falls in between these two scopes and shares the IoT (internet of Things) vision, focusing not only in making data available but also providing the required services to facilitate advanced data analysis, monitoring and assessment procedures in the domain of urban energy distribution and consumption. This paper aims to analyse a specific use case in order to identify services that are required in a platform that supports the development of energy monitoring and assessment applications for urban infrastructures.

Several general architectures for Smart Cities are proposed in the literature, but few examples of their implementation and how to embed services on them are given. According to the existing Smart City architectures, the present work proposes an implementation of a practical case, a complete short-term load forecasting system, explaining the singularities layer by layer trying to cover this gap.

The utilities are the main users of the load prediction systems, who, thanks to the load prediction, manage the maintenance and the control of the distribution systems, buying fuel at the best price or shaping the consumption curve in order to have a flat consumption curve following several strategies. So, this is a tool for the utilities who manage the distribution systems, and in particular to help them to forecast the electrical consumption.

## 2. Context and related work

In the bibliography, taking into account the existence of different visions, several definitions of Smart City are found. In [14] the Smart City is defined as “a city well performing in a forward-looking way in economy, people, governance, mobility, environment, and living, built on the smart combination of endowments and activities of self-decisive, independent and aware citizens”. Otherwise in [6] it is said that Smart City is “a city that monitors and integrates conditions of all of its critical infrastructures, including roads, bridges, tunnels, rail/subways, airports, seaports, communications, water, power, even major buildings, can better optimize its resources, plan its preventive maintenance activities, and monitor security aspects while maximizing services to its citizens”. The paper [31] says that “the use of Smart Computing technologies to make the critical infrastructure components and services of a city which include city administration, education, healthcare, public safety, real estate, transportation, and utilities more intelligent, interconnected, and efficient”.

Some papers, like [26], coincide that Smart Cities are composed by three main dimensions. The first one is the technology dimension, where several technologies are used to monitor, control and share in the city processes. The second one is the human dimension, where creativity, relationships, education and knowledge are the base of the human infrastructure to provide social benefits to the Smart City. The third one is the institutional dimension, where the administration promotes regulations, policies and community participation to grow properly and sustainably.

On the basis of the reviewed works, the common Smart City challenges are:

- Establish a base Smart City architecture to provide a common framework for the sector.
- Dispose and extend standardized Smart City policies that lead to the growth and the proliferation of Smart City services and initiatives.
- Design a list of the essential Smart City services such as Smart water, Smart Governance, Smart buildings, etc.
- Define the basic guidelines in order to perform operations, maintenance, improvements and the scalability in the Smart Cities.

Therefore, it has sense to contribute in the field with a suitable Smart City architecture, selected for developing the services oriented to consumption prediction. It provides the basis where the smart services are going to operate. The following paragraphs summarize different works done in the field of Smart Cities covering proposed architectures and services implied and some of them particularized for short-term load forecasting (STLF). From the point of view of services, there are some papers cited.

A complete guide for design the Smart City architectures and all the functionalities from the data point of view is proposed in [32]. A summary of the main issues of the application systems and the difficulties and challenges in the construction of the Smart City is presented in [30]. A broad view of energy services and their usage, functionality and development challenges are explained in [17]. In order to improve operations and maintenance, reduce the cost of operation, provide enhanced energy management capabilities and provide scalability in the Smart City architecture a guidelines are highlighted in [3]. Several Smart City architectures and their requirements are exposed and commented in [12]. The work [25] comes up with a model for analysis of interactions with a Smart City, providing a larger scale simulation among several Smart City systems. A wide survey of technologies, protocols, and architecture for an urban internet of things in Smart Cities is shown in [35].

So, there is no defined criteria about the number and the function of layers of the Smart City architecture. The work [18] presents a three layers architecture: information storage layer, application layer and user interface layer. The paper [3] suggests a five layers architecture: smart infrastructure, smart database, smart building manager, smart interface and integration layer. The publication [4] proposes a five layers architecture: stakeholder layer, service layer, business layer, infrastructure layer and information layer. In [13] the Smart City architecture is divided in two layers: knowledge processors and semantic information brokers. The paper [21] proposes a Smart City architecture with three parts: the physical network, the communications infrastructure and the flow of information. The study [2] divides the Smart City in two layers: monitoring layer and development layer. The work [32] proposes a five layers architecture: data acquisition, data transmitting, data storage, support service, domain service and event application.

In relation with Smart City services, a short-term load forecasting model for non-residential building on the basis of occupancy and temperature is presented in [23]. A principal component analysis is used for monitoring the electric consumption of buildings in [7]. In order to organize the power

production of distributed generation sources in relation with energy storage system and reduce the operational costs of microgrids a smart energy manager system is provided in [10]. In the work [22], the need to include the cogeneration power generation in electricity balancing and grid stabilization is pointed out. The benefits of a home energy control box for optimizing energy consumption from electrical vehicle charging in residential buildings is seen in [24]. In [19] an energy system planning which incorporates renewable energy services, energy storage technologies and system regulation strategies is provided. A smart energy distribution and management system for monitoring power consumption and users situation and controlling appliances is presented in [8]. An energy information system (real data acquisition, visualization, analysis and switching) which admits the integration of several sensors is provided in [20]. The paper [9] describes a smart lighting solution which allows the integration of the communications and logic on the current street lighting infrastructure. A design and implementation of occupancy sensor platform for individual offices is presented in [1].

Taking into account the energy signatures, in [5] the importance of energy signatures which can help to improve the energy efficiency and monitor the consumption, is pointed out. The use of the energy signatures in order to evaluate the energy performance of chillers using several design options and operating strategies is seen in [34]. In [28] the addition of occupancy as a variable in energy signature model PRISM is analysed.

With regard to baseline models and measuring and verification methods, the work [15] proposes a calibration methodology of the building energy models which can deal with energy retrofit options. In [33] a calibration procedure of the energy performance model on the basis of monthly data through a base load analysis approach is proposed. A statistical evaluation of the performance of various commercial building baseline models analysing the importance of the weather and the morning adjustment factors is seen in [11]. Measuring and verification guiding principles for the assessment of energy efficiency insisting in the need of unambiguous contractual models are highlighted in [27].

In the following sections, a suitable Smart City architecture, selected for developing the services oriented to consumption prediction is detailed. It provides the basis where the smart services are going to operate. After that, a use case to explore the smart-x service in line with the proposed architecture layers is provided.

### 3. Vision

As it can be seen in the bibliography, there is an extensive proposal of architectures to face common challenges that arise in the Smart Cities concept. But, a reference architecture that allows the entire operation of a Smart City has not been designed yet. The subject has been treated cautiously due to the number of technologies that involves, and mainly because it has not been established an standard for integrating these technologies in order to generate a coherent, flexible, scalable, repeatable and effective system. Furthermore, some of the approaches deal with Smart Cities from a theoretical viewpoint which distances itself from the real world. The proposed architectures focus on different aspects from the point of view of technology, human-system interaction or logic [32]. Most of the proposals from the technological aspect, divide the architecture in layers. There could be some slight differences, but as seen in the previous section they have some features in common.

The proposed 5-layers architecture is composed by: data acquisition layer, data transmission layer, data storage layer, preprocessing layer, services layer and application layer, as shown in Figure 1. This architecture delivers better definition of the function of each layer and it is oriented to the services implementation.

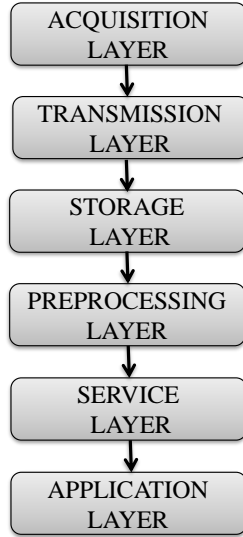


Figure 1: Proposed layers architecture.

- **Data acquisition layer** is responsible for collecting and storing external data. It can capture any kind of information including images, video, sound, and others. In particular circumstances, some preprocessing can be done here, in order to store the data filtered or more elaborated.
- **The data transmission layer** is in charge of end-to-end communications. Network technologies and protocols are taken into account at this level.
- **The data storage layer** has to be able to support large-scale complex data. Also, it has to guarantee that the data is reliable and must provide for the introduction of new data from new sensors or new available information. That is, it has to be scalable. At the same time, the layer has to provide access methods to the data.
- **Preprocessing layer** Once the data is stored, since they come from different types of sensors or information sources, the architecture has to prevent from duplications, outliers, errors, missing values and inconsistency. These kind of actions are carried out by the preprocessing layer.
- **Services layer.** Following with the most common layers that constitute the majority of architectures proposed for a Smart City from the technological point of view, there is the services layer. This layer makes possible the usability of the data, usually by means of modules of software that provide the data requested by the user in a transparent manner.
- **Applications layer** The last layer is the applications layer. It is responsible for interacting directly with the user. It shows the data to the user in a comprehensible manner such as graphical form, table or other type of display, and facilitates the interaction with the platform.

A use case focusing on forecasting electrical energy to improve its management is proposed in the next section. This example is intended to help to identify services involved, required functionalities and possible interactions. The architecture has to consider not only the infrastructure itself but also the interaction with consumers, providing performance indicators and using forecasting models. The use case shows

that acquisition, preprocessing, analysis and modelling of data are required processes to provide a set of goal oriented services to systematically exploit data for energy management purposes. In this particular case, classic data-driven methods for forecasting are proposed, but the idea behind can easily be extrapolated to other methods.

#### **4. A use case: short-term load forecasting**

The aim of this use case is to present the implementation of a smart-x service following the architecture reference of the Smart Cities. The prediction of the load consumption is a need in the Smart Cities and a well-known research domain. The consumption of the non-residential buildings is determined by several factors such as previous consumption, occupancy, temperature and temperature set point. There are several ways to deal with the load forecasting depending on the horizon, available data or used model. In general the main objective is to forecast with high accuracy and few data. The process is usually composed by data selection, data pre-processing, model selection, training process, model evaluation and results exploitation.

So, in the next sections, with the aim to generate an auto-regressive (AR) model to predict the consumption of the buildings a complete process of short-term load forecasting using only real consumption data is explained, respecting the same architecture layers explained in the Section 3.

##### *4.1. Data acquisition*

###### *4.1.1. Sensors*

There are several sources of data: electrical load data, weather data and indoor data. Different sensors, placed in distinct places are collecting data with an hourly sampling rate.

- Electric load data: electrical load data (kW) is collected using the Campus Infrastructure Monitoring System (CIMS). The CIMS is composed of several Schneider power meters installed at the university buildings. The consumption data is collected in PI, PII, PIII, PIV, Faculty of Science, Faculty of Law and Faculty of Economics buildings. There are three different types of power meters: PowerLogic ION 7350 (power, current, voltage, frequency, power factor, current and harmonic distortion), PowerLogic ION 6200 (power, current, voltage and frequency) and PowerLogic PM-810 (power, power factor and frequency). With



the devices properly configured, the data is transferred by the log inserter from each device to the database every 15 minutes. The communication between meters and data storage is performed with the PowerLogic ION Enterprise 5.6 software.

- Weather data: data of temperature ( $^{\circ}C$ ) using a HMP-35AC sensor of Vaisala, relative humidity (%) using a Humicap sensor of Vaisala and solar radiation ( $W/m^2$ ) using a CM11 sensor of Kipp & Zonen are collected outside the buildings by the Department of Physics.
- Indoor data: Only for the case of PIV, indoor ambient and occupancy data are collected inside the building. A wireless sensor network (WSN) is collecting data of temperature ( $^{\circ}C$ ) using a MCP9700A sensor of Microchip, relative humidity (%) using a 808H5V5 sensor of Sencera, light level (lux) using a PDV-P9203 sensor of Optoelectronics and presence using a passive infra-red sensor of Parallax. In summary, there are 6 sensors badges capturing ambient data and 2 capturing people activity.

#### *4.1.2. Dataset*

Taking into account that in this work an AR model is implemented, only consumption data is used. In the paragraphs that follow, the used data is explained with a brief introduction of its location.

The experiments are conducted using data from PI, PII, PIII and PIV, Faculty of Science, Faculty of Law, Faculty of Economics buildings located at the University of Girona, as seen in Figure 2. The buildings have classrooms, offices and laboratories.

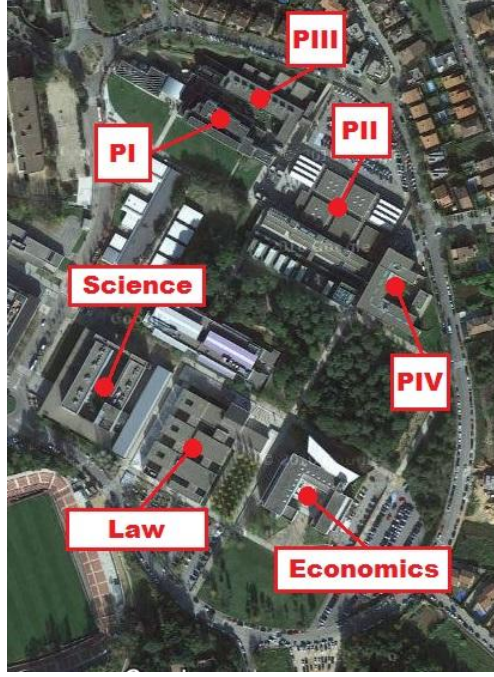


Figure 2: Location of the buildings in the campus of the University of Girona.

In Table 1, the architectural characteristics for each university building are shown.

Building	Floors	Year	Volume ( $m^3$ )	Frontage area ( $m^2$ )	Glass area ( $m^2$ )
PI	6	1983	26150	3791	610
PII	6	1992	25560	2326	1351
PIII	3	2003	11346	1785	310
PIV	3	2003	12000	1836	630
Faculty of Science	3	1997	34810	4903	1233
Faculty of Law	6	1999	32290	6420	1675
Faculty of Economics	5	1997	32287	4770	1375

Table 1: Architectural features of the buildings.

In Table 2, the specifications of the heating system for each university building are seen.

Building	Heating system	
	System	Boiler brand (power)
PI	Gas boiler + Fancoil	Fer (442 kW)
PII		Robur (327 kW)
PIII		Dietrich (310 kW)
PIV		Ygnis (824 kW)
Faculty of Science		Dietrich (560 kW)
Faculty of Law		Wiessman (575 kW)
Faculty of Economics		Dietrich (310 kW)

Table 2: Heating system features of the buildings.

In Table 3, the specifications of the cooling system for each university building are explained.

Building	Cooling system	
	System	Refrigeration brand (power)
PI	Compression refrigeration system + Fancoil	Mitsubishi (160 kW)
PII		Ygnis (269 kW)
PIII		Carrier (255 kW)
PIV		Climaveneta (618 kW)
Faculty of Science		Daikin (430 kW)
Faculty of Law		Teva (1113 kW)
Faculty of Economics		Carrier (255 kW)

Table 3: Cooling system features of the buildings.

The number of data instances of PI is 27375, covering a total of 38 months, from 1st September, 2011 to 15th October, 2014. The total of instances of PII is 16589, covering a total of 24 months, from 21st November, 2012 to 15th October, 2014. The number of instances of PIII and PIV is 16590, covering a total of 24 months, from 23rd November, 2012 to 15th October, 2014. The number of instances of Faculty of Science is 27366, covering a total of 38 months, from 1st September, 2011 to 14th October, 2014. The number of instances of Faculty of Law is 27379, covering a total of 38 months, from 1st September, 2011 to 15th October, 2014. The number of instances of Faculty of Economics is 27379, covering a total of 38 months, from 1st September, 2011 to 15th October, 2014.

In the Figure 3 the consumption of a week in spring and summer is observed:

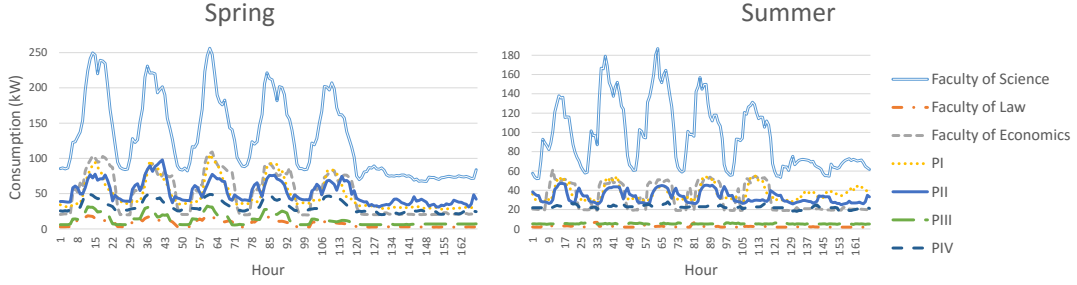


Figure 3: Consumption vs. hour.

In the Figure 4 the consumption of a week in autumn and winter is seen:

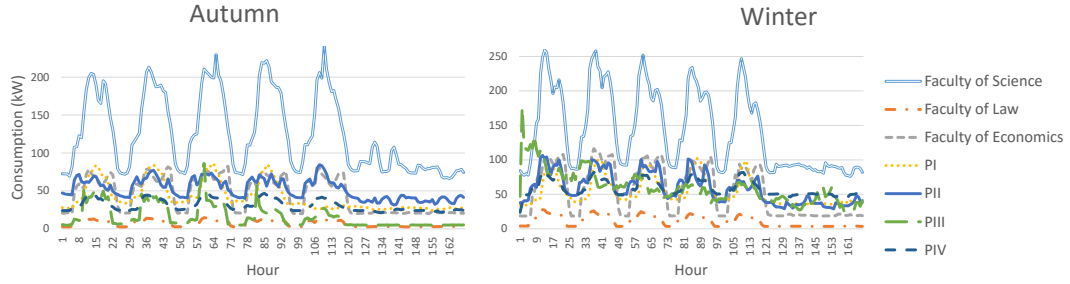


Figure 4: Consumption vs. hour.

#### 4.2. Data transmission

There are three data sources. The Department of Physics, serving weather data, the CIMS, disposing of consumption data and the WSN of PIV, collecting ambient and occupancy data.

The Department of Physics uses a wired sensor network to collect the data from the several instruments of the weather station. At the same time, the CIMS captures the consumption of different buildings using a wired network too. In the case of the indoor data a WSN is employed.

The WSN is composed by 8 motes of the Libelium-brand which send the measured data to a central hub, called Meshilum, trough XBee radio modules that communicate by means of the ZigBee protocol. The Meshilum data is

accessed using an Ethernet connection. The Libelium technology is based on Arduino and the topology of the network is star.

In the Figure 5 the WSN of PIV building is seen:

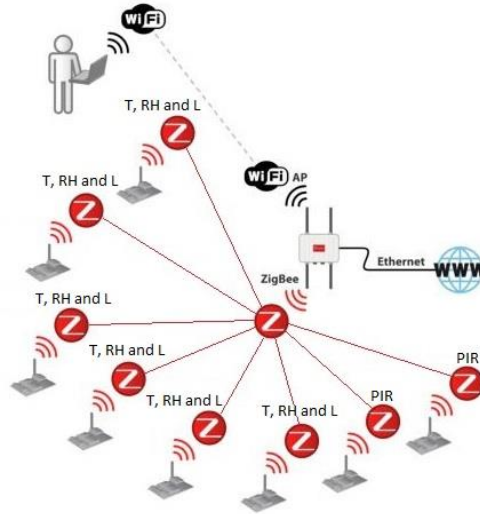


Figure 5: Wireless sensor network of PIV building.

#### 4.3. Data storage

The data come from 3 distinct sources: meteorological data provided by the Department of Physics, consumptions data for each buildings provided by CIMS and indoor data collected by the WSN. Each source presents distinct configurations, and even owners, that made impossible a direct actuation over the distinct databases storing the information further than data access. Data presents distinct formats for each of the sources and a homogenization step is mandatory.

The CIMS data is stored in a MSSQL database for the Schneider software ION. Department of Physics data is accessed via an SFTP server and the WSN data is stored in a MySQL database inside the Meshlium. The solution implemented is an homogenization server whose tasks consist in periodically connect to the distinct data sources, check for updates and update a local MySQL database with an homogeneous format for all the data and provide a simple interface for user to select and download the desired data.

#### 4.4. Data preprocessing

In the following sections, the steps to clean and uniform the data are explained as seen in the Figure 6.



Figure 6: Block diagram of the preprocessing.

##### 4.4.1. Missing values

Given the mistakes in sensor readings, there is always a small amount of lost values. The percentage of missing values needs to be minimized. There are several methods used to filter the missing values such as removing or averaging them. In our case, the instances with missing values are deleted.

##### 4.4.2. Normalization

If data has different scales and units normalization is needed. The use of the same data scale improves the forecasting. The normalization range used is from 0 to 1, as seen in the Equation 1.

$$x_{in} = \frac{x_i}{\sqrt{\sum_i (x_i^2)}} \quad (1)$$

Where:

$x_{in}$  is the normalized instance.

$x_i$  is the instance.

##### 4.4.3. Outliers

The performance of the model is increased if the outliers are filtered. The more restrictive the process, the greater amount of data lost. The outliers filtering process [29] consists in detection and substitution. In the present case, the process of detection consists in identifying  $n$  outliers based on the euclidean distance to their  $k$  nearest neighbours. Then, according to an outlier detection process, the outliers are removed.

#### *4.4.4. Feature selection*

With the aim of removing irrelevant features, redundant and non-correlated attributes are removed. Reducing the size of the database, the computational cost is reduced. The feature selection process is composed by two blocks that perform linear correlations. The first block, in order to eliminate the useless attributes, removes the features with low correlation with the class attribute. The second block, with the aim of deleting the duplicate attributes, removes the attributes with high correlation among them.

#### *4.4.5. Instance selection*

The number of instances is reduced in order to minimize the computational cost. The selected training data is a 30% random sub-sample. Samples about this percentage reduce the computational time while maintain the forecasting performance levels.

### *4.5. Data service*

There is the intention to explore the performance limits of the AR model. So, the experiments are realized using only consumption data taking into account that these data models are simpler and useful in 1-hour ahead forecasts.

#### *4.5.1. Methodology*

A methodology for predicting the load consumption 1-hour ahead is proposed. This methodology consists of several blocks as shown in Figure 7. The preprocessed data is split, 1/3 to test and 2/3 to train. Then, with the training data a grid search of the suitable training parameters is performed over the selected model (AR). The final step is the validation of the model using test data and the performance indicator calculation.

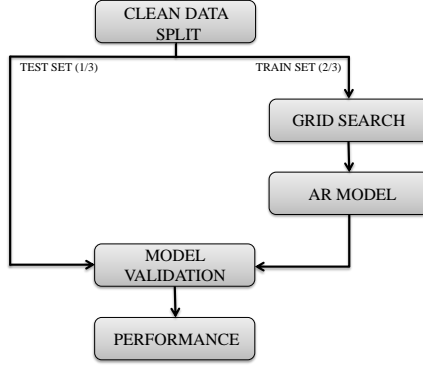


Figure 7: Block diagram of the forecasting model.

#### 4.5.2. Grid search

Grid search method performs a search through the ranks of pairs of training parameters and chooses the best ones. The main tested parameters of the regression model are the following ones: ridge parameter and feature selection.

#### 4.5.3. AR model

The AR model [16] specifies that the output variable depends linearly on its own previous values. Taking into account that the occupancy data is only available for PIV building and the temperature variable does not increase the accuracy of this model due to the partial disaggregation of the heating ventilating and air conditioning system, AR model is a proper model to apply in short-term load forecasting (1-hour ahead). So, the consumption depends on the past values of consumption, as can be seen in Equation 2.

$$X_t = C + \sum_{i=1}^p \varphi_i X_{t-i} + \epsilon_t \quad (2)$$

Where:

$X_t$  is the output variable.

$X_{t-1}$  are the previous values of the output.

$\varphi_i, \dots, \varphi_p$  are the parameters of the model.

$C$  is a constant.

$\epsilon_t$  is white noise.



#### 4.5.4. Validation

The validation process contrasts the model generated with training data (65%) against the test data (35%). The mean absolute percentage error (MAPE) indicator is used to validate the model due to its popularity in the forecasting field. The first period of time is used to predict the last period of time.

The MAPE performance indicator, showed in Equation 3, does not depend on the magnitude of the unit of measurement, and is used to compare models. If the MAPE is small, the model is accurate. In the topic, a range between 1% and 20% is considered acceptable.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_{m(i)} - y_{p(i)}}{y_{m(i)}} \right| \times 100 \quad (3)$$

Where:

$N$  is the number of observations.

$y_m$  is the measured output.

$y_p$  is the predicted output.

#### 4.6. Smart application

On the basis of the outputs of the model, the smart application can offer several services such as prediction charts, energy saving information or corrective actions effectiveness. The users can access to the application interfaces to increase the expert knowledge in order to reduce the consumption or to monitor the forecasting accuracy. In the present paper analytic and graphic results are explained.

##### 4.6.1. Analytical results

Table 4 shows the MAPE indicator for all the buildings, where CC is the correlation coefficient and the computation time is the time to perform the experiment.

Building	Computing time (ms)	C.C	MAPE (%)
Faculty of Science	3672	0.986	5.21
Faculty of Law	4031	0.968	16.75
Faculty of Economics	3734	0.972	15.58
PI	4109	0.981	6.06
PII	2797	0.943	7.84
PIII	3235	0.876	30.11
PIV	2250	0.967	6.83

Table 4: AR model results for all buildings.

The Faculty of Law and The Faculty of Economics present intermediate level of accuracy due to the variability in the consumption profile. PIII has low level of accuracy as a result of inconsistent data.

#### 4.6.2. Graphic results

In this section several charts are presented. First three MAPE charts are presented. Then, seven prediction consumption charts are showed.

- MAPE vs. Hour:

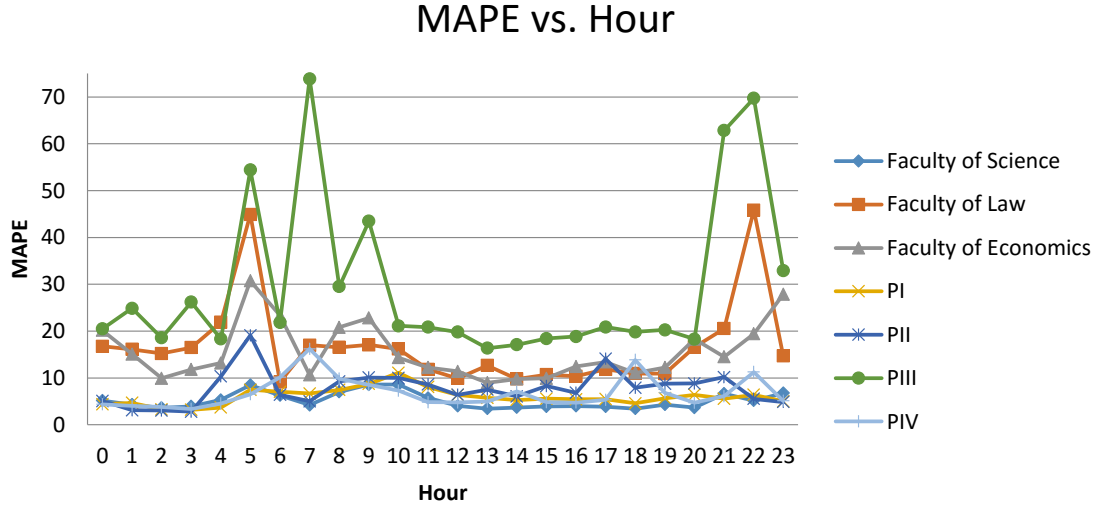


Figure 8: MAPE vs. Hour.

As can be seen in Figure 8, there are three time slots which present poor quality prediction, some night hours (4:00 to 6:00) due to the cleaning and security services and at the beginning (7:00 to 10:00) and the end (20:00 to 22:00) of the school day given the variability in the human behaviour.

- MAPE vs. weekday:

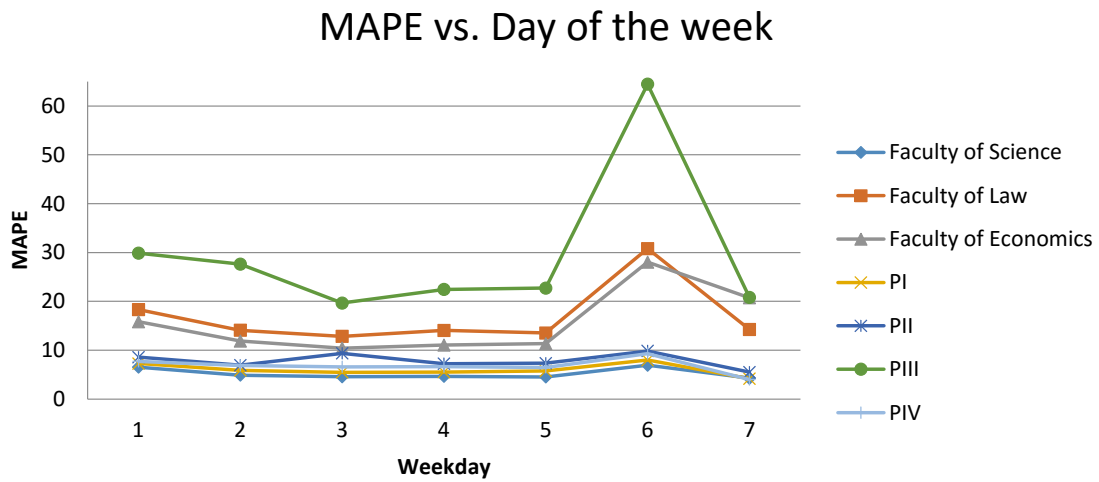


Figure 9: MAPE vs. weekday.

As is shown in Figure 9, Saturday presents low accuracy in the forecasting due to random activities realized in the buildings. In addition, the beginning and the end of the week, that is Mondays and Fridays, taking into account the high dispersion of human behaviours, present lower forecasting precision.

- MAPE vs. Month:

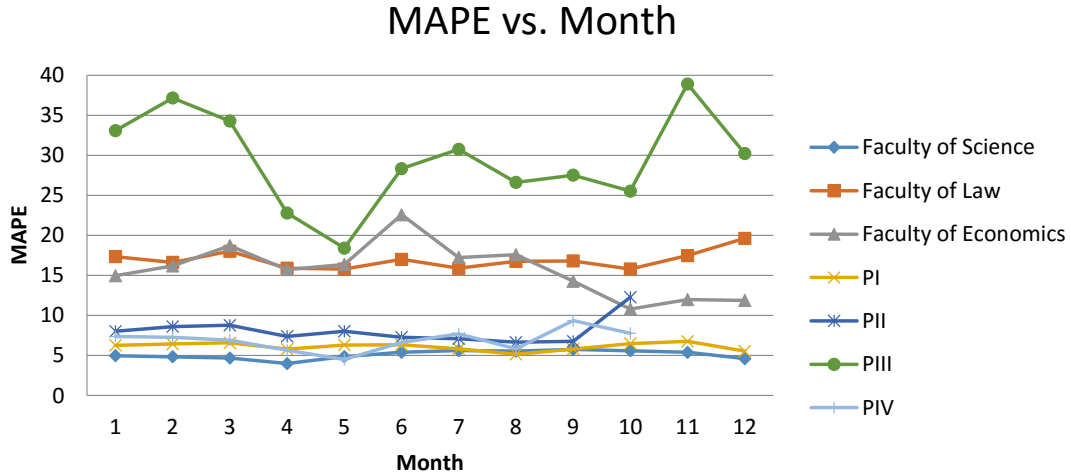


Figure 10: MAPE vs. Month.

As presented in Figure 10, there is no clear conclusions about which months are better predicted but in general, months with high variability and unclear profiles such as December, June or March present worse prediction.

The following charts show the consumption prediction for each building.

- Faculty of Science, Faculty of Law and Faculty of Economics prediction:

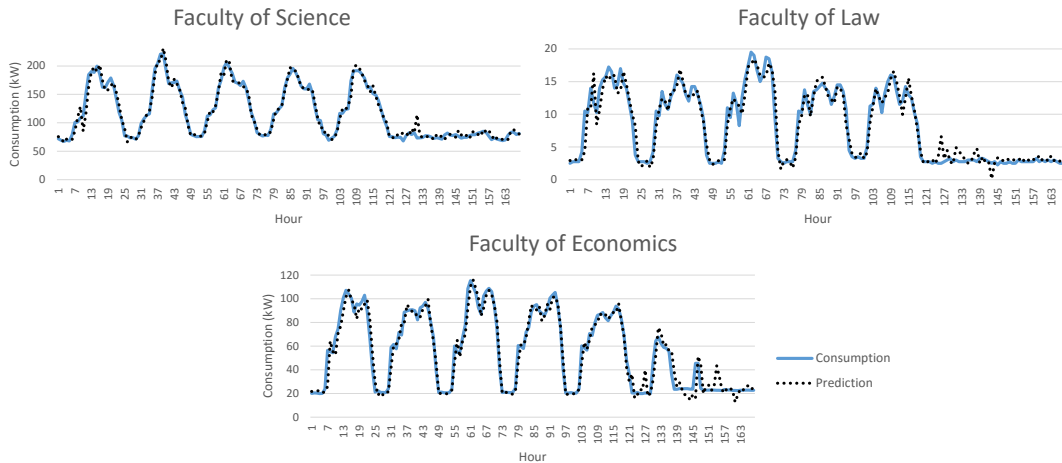


Figure 11: Consumption vs. hour.

As is seen in Figure 11, the worst prediction level is found in Saturdays due to the random after-school activities in some buildings. Besides, in the early hours of Mondays, due to the irregularity of some services, the prediction accuracy decays.

- PI, PII, PIII and PIV prediction:

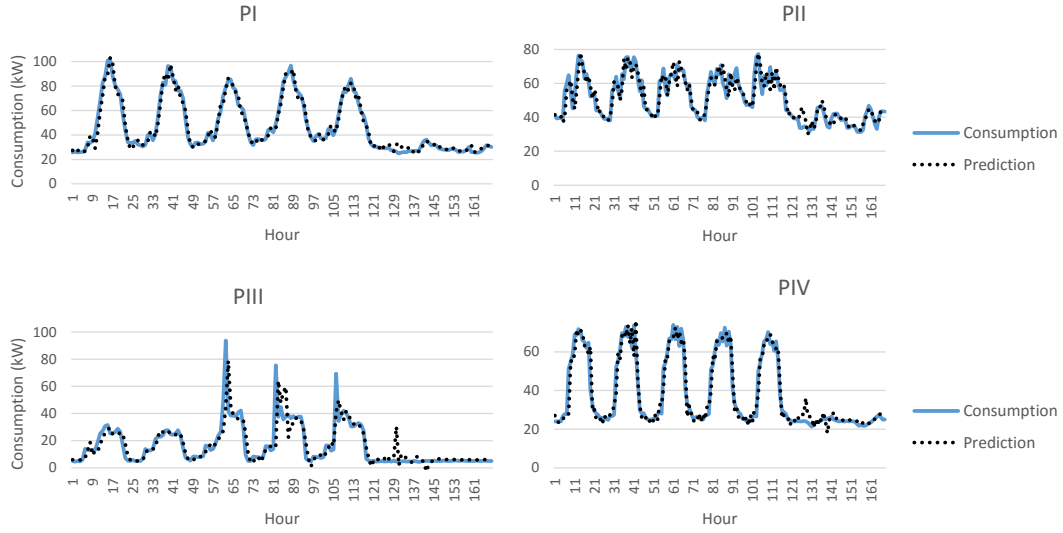


Figure 12: Consumption vs. hour.

As shows in Figure 12, PIII presents the poorest forecasting quality due to the inconsistency of the data. As in the previous figure, Saturdays are the worst predicted days.

## 5. Discussion

Following the smart-x architecture layers a use case is performed:

- In the acquisition layer the robustness is key, in the present case some misbehaviours of the sensors comported data loss and outliers.
- In relation with the transmission layer, some ZigBee reception problems entailed data loss, so a previous study of the sensor distribution is needed.

- The storage layer must be safe, standard and scalable, the load consumption database presented integration difficulties.
- With reference to the preprocessing layer, outliers or missing values in the sensor measures lead to a low accuracy in the forecasting. Using a software with suitable preprocessing tools is completely necessary to obtain fine results.
- In reference to the service layer, although occupancy and weather data are variables that partially explain consumption in buildings, the AR models are simple and quick. In summary, the presented model depends only on the consumption, so weather data is not needed, making it more economic and compact. In this case, the prediction is performed 1-hour ahead, that means better performance results than 24-hours ahead models, where exogenous data is usually needed. From the results, it's obvious that some buildings present better forecasting accuracy than others. In order to provide fine predictions using AR models, the building must have clean and consistent data. Cyclic and well-defined consumption patterns deliver proper AR model predictions. If there are some random or undefined activities in the buildings, the autocorrelation is low. The buildings with big amount of classrooms or offices with concrete schedule are easy to predict, in the other hand, buildings with rooms with non-defined activities are hard to predict. Similarly, there are some time-slots with high variability in the human behaviour in the nights or at the beginning and the end of the school day that present difficulties to be predicted.
- Taking into account the appliance layer, there is the need to make it accessible and upgradeable. In the present case some efforts have been employed to present the results (charts, tables, etc.) through web services.

In relation with the case study, some actions to take for possible energy saving improvements can be derived from the system analysis:

- Compress work schedules, reducing the hours flexibility. Specify the entering and leaving work, the mealtime and the lunchtime periods.
- Suppression of the HVAC system during weekends and holidays. Adjust the HVAC operation time downwards.

- Move the cleaning service to day hours.
- Control the HVAC in order to have temperature, relative humidity and light level inside the proper range, proposed by the authorities, taking into account the homogeneity along the building.

## 6. Conclusions

Urban development involves the use of intelligent services taking advantage of monitored data and providing an action to improve or maintain the quality of these city services. This paper focuses on the particular case of the electricity, identifying services that can help in increasing the energy efficiency in urban infrastructures.

The work proposes a use case of short-term load forecasting in non-residential buildings with real data in order to practically explain the services embedded in the described Smart City general layers architecture. These layers are responsible of collecting the data from the sensors, transmitting the data to the central hub, storing, cleaning and standardising this data, applying the forecasting methodology and finally provide an application to show the results. The use case provided as a demonstration, consists of predicting the consumption in 7 university buildings. The load forecasting is performed using AR models showing that the results differs according to the profile of the building and the quality of the data. When the data is complete and the consumption pattern is cyclic and clear, the results are fine. Also, the service allows to test the prediction accuracy from different points of view, such as analysing which is the best month predicted or the same for the days of the week.

As a future work, more services have to be defined to help providing more information to the users in order to improve the energy efficiency of the buildings. For example, defining an index related to the efficiency of the building can be a good contribution to the subject.

## Acknowledgments

This research project has been partially funded through BR-UdG Scholarship of the University of Girona granted to Joaquim Massana Raurich. Work developed with the support of the research group SITES awarded with distinction by the Generalitat de Catalunya (SGR 2014-2016), the MESC

project funded by the Spanish MINECO (Ref. DPI2013-47450-C2-1-R) and the European Unions Horizon 2020 research and innovation programme under grant agreement No 680708.

## References

- [1] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, and T. Weng. Occupancy-driven energy management for smart building automation. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, pages 1–6. ACM, 2010.
- [2] M. Al-Hader and A. Rodzi. The smart city infrastructure development & monitoring. *Theoretical and Empirical Researches in Urban Management*, 2(11):87, 2009.
- [3] M. Al-Hader, A. Rodzi, A. R. Sharif, and N. Ahmad. Smart city components architecture. In *Computational Intelligence, Modelling and Simulation, 2009. CSSim'09. International Conference on*, pages 93–97. IEEE, 2009.
- [4] L. Anthopoulos and P. Fitsilis. From digital to ubiquitous cities: Defining a common architecture for urban development. In *Intelligent Environments (IE), 2010 Sixth International Conference on*, pages 301–306. IEEE, 2010.
- [5] C. Aoun. The smart city cornerstone: Urban efficiency. *Published by Schneider electric*, 2013.
- [6] B. Bowerman, J. Braverman, J. Taylor, H. Todosow, and U. Von Wimmersperg. The vision of a smart city. In *2nd International Life Extension Technology Workshop, Paris*, volume 28, 2000.
- [7] L. Burgas, J. Melendez, J. Colomer, J. Massana, and P. Carles. Principal component analysis for monitoring electrical consumption of academic buildings. *Energy Procedia*, 62:555–564, 2014.
- [8] J. Byun, I. Hong, B. Kang, and S. Park. A smart energy distribution and management system for renewable energy distribution and context-aware services based on user patterns and load forecasting. *Consumer Electronics, IEEE Transactions on*, 57(2):436–444, 2011.



- [9] M. Castro, A. J. Jara, and A. F. G. Skarmeta. Smart lighting solutions for smart cities. In *Advanced information networking and applications workshops (WAINA), 2013 27th international conference on*, pages 1374–1379. IEEE, 2013.
- [10] C. Chen, S. Duan, T. Cai, B. Liu, and G. Hu. Smart energy management system for optimal microgrid economic operation. *Renewable Power Generation, IET*, 5(3):258–267, 2011.
- [11] K. Coughlin, M. A. Piette, C. Goldman, and S. Kiliccote. Statistical analysis of baseline load models for non-residential buildings. *Energy and Buildings*, 41(4):374–381, 2009.
- [12] W. M. da Silva, A. Alvaro, G. Tomas, R. A. Afonso, K. L. Dias, and V. C. Garcia. Smart cities software architectures: a survey. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pages 1722–1727. ACM, 2013.
- [13] L. Filipponi, A. Vitaletti, G. Landi, V. Memeo, G. Laura, and P. Pucci. Smart city: An event driven architecture for monitoring public spaces with heterogeneous sensors. In *Sensor Technologies and Applications (SENSORCOMM), 2010 Fourth International Conference on*, pages 281–286. IEEE, 2010.
- [14] R. Giffinger and N. Pichler-Milanović. *Smart cities: Ranking of European medium-sized cities*. Centre of Regional Science, Vienna University of Technology, 2007.
- [15] Y. Heo, R. Choudhary, and G. A. Augenbroe. Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings*, 47:550–560, 2012.
- [16] S. R. Huang. Short-term load forecasting using threshold autoregressive models. In *Generation, Transmission and Distribution, IEE Proceedings-*, volume 144, pages 477–481. IET, 1997.
- [17] S. Karnouskos, P. G. D. Silva, and D. Ilic. Energy services for the smart grid city. In *Digital Ecosystems Technologies (DEST), 2012 6th IEEE International Conference on*, pages 1–6. IEEE, 2012.

- [18] N. Komninos. The architecture of intelligent cities: Integrating human, collective and artificial intelligence to enhance knowledge and innovation. In *Intelligent Environments, 2006. IE 06. 2nd IET International Conference on*, volume 1, pages 13–20. IET, 2006.
- [19] G. Krajačić, N. Duić, Z. Zmijarević, B. V. Mathiesen, A. A. Vučinić, and M. da Graça Carvalho. Planning for a 100% independent energy system based on smart energy storage for integration of renewables and CO<sub>2</sub> emissions reduction. *Applied thermal engineering*, 31(13):2073–2083, 2011.
- [20] I. Kunold, M. Kuller, J. Bauer, and N. Karaoglan. A system concept of an energy information system in flats using wireless technologies and smart metering devices. In *Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), 2011 IEEE 6th International Conference on*, volume 2, pages 812–816. IEEE, 2011.
- [21] L. Lugaric, S. Krajcar, and Z. Simic. Smart cityPlatform for emergent phenomena power system testbed simulator. In *2010 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*, pages 1–7. IEEE, 2010.
- [22] H. Lund, A. N. Andersen, P. A. Østergaard, B. V. Mathiesen, and D. Connolly. From electricity smart grids to smart energy systems—a market operation based approach and understanding. *Energy*, 42(1):96–102, 2012.
- [23] J. Massana, C. Pous, L. Burgas, J. Melendez, and J. Colomer. Short-term load forecasting in a non-residential building contrasting models and attributes. *Energy and Buildings*, 92:322–330, apr 2015.
- [24] K. Mets, T. Verschueren, W. Haerick, C. Develder, and F. De Turck. Optimizing smart energy control strategies for plug-in hybrid electric vehicle charging. In *Network Operations and Management Symposium Workshops (NOMS Wksp), 2010 IEEE/IFIP*, pages 293–299. Ieee, 2010.
- [25] B. Morvaj, L. Lugaric, and S. Krajcar. Demonstrating smart buildings and smart grid features in a smart energy city. In *Energetics (IYCE), Proceedings of the 2011 3rd International Youth Conference on*, pages 1–8. IEEE, 2011.

- [26] T. Nam and T. Pardo. Conceptualizing smart city with dimensions of technology, people, and institutions. In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, pages 282–291. ACM, 2011.
- [27] K. Park, Y. Kim, S. Kim, K. Kim, W. Lee, and H. Park. Building energy management system based on smart grid. In *Telecommunications Energy Conference (INTELEC), 2011 IEEE 33rd International*, pages 1–4. IEEE, 2011.
- [28] A. Rabl and A. Rialhe. Energy signature models for commercial buildings: test with measured data and interpretation. *Energy and buildings*, 19(2):143–154, 1992.
- [29] S. Ramaswamy, R. Rastogi, and K. Shim. Efficient algorithms for mining outliers from large data sets. *ACM SIGMOD Record*, 29(2):427–438, 2000.
- [30] K. Su, J. Li, and H. Fu. Smart city and the applications. In *Electronics, Communications and Control (ICECC), 2011 International Conference on*, pages 1028–1031. IEEE, 2011.
- [31] D. Washburn, U. Sindhu, S. Balaouras, R. A. Dines, N. Hayes, and L. E. Nelson. Helping CIOs understand smart city initiatives. *Growth*, 17(2), 2009.
- [32] R. Wenge, X. Zhang, C. Dave, L. Chao, and S. Hao. Smart city architecture: a technology guide for implementation and design challenges. *Communications, China*, 11(3):56–69, 2014.
- [33] J. Yoon, E. Lee, and D. E. Claridge. Calibration procedure for energy performance simulation of a commercial building. *Journal of solar energy engineering*, 125(3):251–257, 2003.
- [34] F. W. Yu and K. T. Chan. Energy signatures for assessing the energy performance of chillers. *Energy and buildings*, 37(7):739–746, 2005.
- [35] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi. Internet of things for smart cities. *Internet of Things Journal, IEEE*, 1(1):22–32, 2014.