



UNIVERSITY OF DEUSTO

# Contributions to Demand-Side Management by the application of Artificial Intelligence techniques in domestic, commercial, and industrial scenarios

Tesis doctoral presentada por

OIHANE KAMARA ESTEBAN

dentro del Programa de Doctorado en

INGENIERÍA INFORMÁTICA Y TELECOMUNICACIÓN

dirigida por

DR. CRUZ E. BORGES HERNÁNDEZ y DRA. ANA M. MACARULLA ARENAZA





UNIVERSITY OF DEUSTO

# Contributions to Demand-Side Management by the application of Artificial Intelligence techniques in domestic, commercial, and industrial scenarios

Tesis doctoral presentada por

OIHANE KAMARA ESTEBAN

dentro del Programa de Doctorado en

INGENIERÍA INFORMÁTICA Y TELECOMUNICACIÓN

dirigida por

DR. CRUZ E. BORGES HERNÁNDEZ y DRA. ANA M. MACARULLA ARENAZA

La doctoranda

El Co-director

La Co-directora



To my mother, my brother,  
and in loving memory of my father.



## Abstract

---

Electricity is, perhaps, the most identifiable and ubiquitous form of energy. Whether we are watching a movie, preparing dinner, working on the computer, manufacturing the engine of a car, using an elevator, or just reading the news feed on our mobile phone, electricity is always present either directly or indirectly. Trends in electricity consumption worldwide show that the global demand is expected to grow significantly in the forthcoming years driven, primarily, by the development of new technologies that help achieve a higher quality of life. In fact, the shift of worldwide economies from a subsistence perspective to industrial or service approaches, specially in developing countries, is what is leading the growth in electricity needs.

Electricity is considered a secondary energy source since it is obtained from the transformation of primary sources of energy, either renewable or non-renewable. Even though renewable technologies are slowly but steadily gaining ground as a clean and cost-effective generation alternative, the vast majority of the world's electricity is still being produced from non-renewable sources, such as coal, gas, or oil. In fact, if we analyse CO<sub>2</sub> emissions related to energy generation, the electricity sector is responsible for around 40 % of these emissions due to the use of fossil sources for electricity generation.

This growth scenario calls for the design and implementation of energy efficiency measures that ensure a reliable and adequate electricity supply that meets the global demand at all times, while reducing the greenhouse

gas emissions derived from its generation. Among these efficiency measures, the most favoured by electric utilities due to its cost-effectiveness and immediacy of results is Demand-Side management.

Demand-Side management strategies are actions designed to modify the behaviour of the customers in regards to the the amount and timing of electricity use for the collective benefit of the society and the utility. The emergence and settlement of the Smart Grid and associated smart devices have encouraged the implementation of these type of programs, thanks to the availability of real-time consumption data through intelligent monitoring and the possibility to manage the whole grid.

The thesis presented below comprehends the research done to push forward the State of the Art of Demand-Side management. We have identified the research opportunities from a vertical perspective: analysing the needs and best practice to standardise the communication of the devices that are part of the Smart Grid, highlighting the advantages of creating simulation scenarios that will help electric utilities to take decisions prior to physical or logical deployment of network elements, and finally, demonstrating cases of application of Artificial Intelligence techniques to implement Demand-Side management in commercial, domestic, and industrial scenarios.

## Resumen

---

La electricidad es la forma de energía más ubicua que se conoce a día de hoy. Ya sea que estemos viendo una película, preparando la cena, trabajando en el ordenador, fabricando el motor de un coche, usando un elevador, o simplemente leyendo las noticias en nuestro dispositivo móvil, la electricidad está, directa o indirectamente, siempre presente. A nivel mundial, la tendencia de consumo eléctrico demuestra que la demanda seguirá creciendo significativamente en los próximos años impulsada, principalmente, por nuevos desarrollos tecnológicos que buscan mejorar la calidad de vida. En concreto, la transición de economías de subsistencia a economías industriales o de servicios sobre todo en países en vías de desarrollo, es una de las causas que justifican el crecimiento global de la demanda de electricidad.

La electricidad es una fuente de energía secundaria ya que se obtiene a partir de la transformación de fuentes de energía primarias, ya sean renovables o no renovables. Aún y cuando el uso de tecnologías renovables está ganando terreno como alternativa limpia y rentable en términos económicos a los métodos clásicos de generación de energía, los combustibles fósiles, como el carbón, gas, o el petróleo, siguen monopolizando el sector de generación. De hecho, si analizamos las emisiones de CO<sub>2</sub> relacionadas con la generación de energía, el sector eléctrico es responsable de la producción del 40 % de esas emisiones debido al uso de estos combustibles.

Este escenario de crecimiento aboga por el diseño e implementación de medidas de eficiencia energética que aseguren un suministro eléctrico que, por un lado, satisfaga la demanda en cada momento y, por otro lado, sea capaz de minimizar las emisiones de gases de efecto invernadero derivadas de la generación de dicha electricidad. Entre estas medidas de eficiencia, las más favorecidas por las empresas eléctricas son aquellas relacionadas con la Gestión de la Demanda.

La Gestión de la Demanda engloba todo un conjunto de estrategias diseñadas para modificar el consumo eléctrico de los usuarios finales en lo que respecta a la cantidad y momentaneidad del consumo de electricidad, con el objetivo de generar un beneficio colectivo para sociedad y la empresa de distribución eléctrica. El surgimiento y afianzamiento de la red inteligente de energía eléctrica, también conocida como Smart Grid, ha alentado positivamente la implementación de este tipo de estrategias, gracias a que se dispone de una monitorización inteligente en tiempo real y una gestión global de todos los puntos de la red.

La tesis que se presenta a continuación engloba la investigación realizada en materia de Gestión de la Demanda. En este sentido, se han identificado las oportunidades de investigación desde un enfoque vertical: analizando las necesidades y mejores prácticas para estandarizar la comunicación de los dispositivos que forman parte de la Smart Grid, subrayando las ventajas de crear escenarios de simulación que ayuden a las empresas de distribución a la toma de decisiones antes del despliegue físico y lógico de elementos de la red eléctrica, y finalmente, demostrar casos de aplicación de técnicas de Inteligencia Artificial para implementar técnicas de Gestión de la Demanda en los ámbitos doméstico, comercial e industrial.

## Acknowledgements

---

*He'd been wrong, there was a light at the end of the tunnel, and it was a flamethrower. — Terry Pratchett, Mort*

Life is a like book. The adventure starts when we open the cover and write the first letter in sync with our first breath. Our handwriting evolves through the pages fuelled by the emotions, lessons, and hardships concealed beneath the white sheets. As the story progresses, the change of pace marks the beginning of a new chapter. Sometimes, we decide when the chapter ends. Other times, that decision may be forced upon us. In either case, a new page awaits.

The writing of this dissertation gives the finishing touches to a chapter that started four years ago. Aside from my own musings, the pages are vibrant with the colour and the essence of the people that have walked by my side on this journey, and to which I would like to show my gratitude.

First, thank you to my thesis advisors, Cruz E. Borges and Ana M<sup>a</sup> Macarulla for their patience, dedication, and support throughout these years. You definitely provided me with the tools and guidance necessary to successfully complete my dissertation. Rewriting the introduction a thousand times can surely be accounted for that.

Thank you to the University of Deusto and the Basque Government's Hezkuntza, Hizkuntza Politika eta Kultura Saila (Department for Education, Linguistics Policy and Culture) for providing me with the funding

necessary to carry out my research, initially through the FPI Deusto, and later through the PRE\_2013\_1\_516, PRE\_2014\_2\_139, PRE\_2015\_2\_0003 predoctoral grants.

Heartfelt thanks to Ainhoa Alonso for her encouragement on a personal and professional level, and for believing in my work ever since the beginning. Lovely thanks, as well, to my dearest former and current members of DeustoTech Energy. Our paths entwined at different pages of our lives, but I am sure that the bond and the memories are timeless, with many of you becoming lifelong friends. Thank you Aitor, Gorka, Iván, María, Lohitzune, Iraia, Tony, Ander, Cristina, Borja, Mikel, Christian, Virginia, Hector, Paula, and Nando. You have made this trip a lot more enjoyable.

To Gorka Azkune for his valuable collaboration in one of the chapters of this thesis, and for opening a new research path with his insights. Thank you also for your patience through the awfully long reviewing process.

To Maite, Oscar, and Alain for their support. We started this new chapter almost at the same moment, and I am happy to see that you have successfully made it to the final page. Special thanks, as well, to Unai Aguilera for his on-point encouragement and helpful advice during these times.

To Angela, thank you for the laughs and good moments. I think we have an awesome friendship in the making. Dancing session pending.

To all my friends, specially Ana, Unai López, Itsaso, Tania, Sergio and Mireya, for cheering on me from the sidelines even though I was absent for the last few months. Thank you for understanding and giving me the time and personal space necessary to clear my head and sprint towards the finish line.

To Josu, for holding my hand steady whenever the pen began to falter and the words refused to come out. Your presence, patience, and solace have, without a doubt, breathed life and meaning into this chapter. There are not enough words to express my appreciation and I look forward to

writing more chapters together. I will still fall asleep on boring movies, though.

My deepest gratitude goes to my family for their unwavering love throughout my life. To my wonderful *Amama*, the most selfless and loving person I have ever known. Thank you for your comfort even though you could not understand why I had to spend so many hours at the computer, but you knew I was writing in English because “all letters looked out of place”. To my parents, that spared no effort in providing us with the best possible environment to grow up. To *Amatzu*, for showing me what strength really means, for being unconditionally close even though we were miles apart, and for never letting go of our hands when life hit too hard. To, my brother, *Ander*, for being my personal support and mischievous accomplice ever since I can remember. Finally, to my dear *Aita*, a loving father, passionate reader, and advocate for critical thinking who would have loved to hold this dissertation in his hands. Your lessons and memories live in every word.

Thank you  
Oihane Kamara Esteban



# Table of Contents

<b>Table of Contents</b>	<b>i</b>
<b>List of Figures</b>	<b>v</b>
<b>List of Tables</b>	<b>ix</b>
<b>List of Listings</b>	<b>xi</b>
<b>Acronyms</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Context and motivation . . . . .	3
1.2 Dissertation objective . . . . .	23
1.3 Dissertation outline . . . . .	25
<b>2 Standardising Smart Grid communication protocols</b>	<b>27</b>
2.1 Introduction . . . . .	28
2.2 IEC 60870 . . . . .	30
2.3 Common Information Model (CIM) . . . . .	33
2.4 IEC60870 integration into CIM . . . . .	35
2.5 Conclusions . . . . .	44
<b>3 GeoWorldSim: Steps towards a Smart City Simulator</b>	<b>47</b>
3.1 Introduction . . . . .	48

3.2	Related Work . . . . .	51
3.3	Components of GeoWorldSim . . . . .	55
3.4	Materials and methods . . . . .	63
3.5	Results and discussion . . . . .	64
3.6	Conclusions . . . . .	67
<b>4</b>	<b>Agent-based Human Activity Simulation</b>	<b>69</b>
4.1	Introduction . . . . .	70
4.2	Related Work . . . . .	72
4.3	Design and development . . . . .	75
4.4	Materials and Methods . . . . .	84
4.5	Results . . . . .	89
4.6	Discussion . . . . .	99
4.7	Conclusions . . . . .	103
<b>5</b>	<b>Domestic DSM: Middleware for implementing strategies</b>	<b>109</b>
5.1	Introduction . . . . .	110
5.2	Materials and Methods . . . . .	113
5.3	Results and discussion . . . . .	118
5.4	Conclusions . . . . .	122
<b>6</b>	<b>Commercial DSM : Optimisation of the electric energy tariff</b>	<b>127</b>
6.1	Introduction . . . . .	128
6.2	Analytical optimisation of the 6.1 electric tariff . . . . .	133
6.3	Evolutionary algorithms . . . . .	134
6.4	Materials and methods . . . . .	138
6.5	Results and Discussion . . . . .	141
6.6	Conclusions . . . . .	145
<b>7</b>	<b>Industrial DSM : Energy-aware foundry production scheduling</b>	<b>147</b>
7.1	Introduction . . . . .	148
7.2	Related Work . . . . .	150
7.3	Materials and methods . . . . .	152

7.4	Results . . . . .	162
7.5	Discussion . . . . .	169
7.6	Conclusion . . . . .	173
<b>8</b>	<b>Utility tools for DSM: Short Term Load Forecasting</b>	<b>177</b>
8.1	Introduction . . . . .	178
8.2	Data imputation vs Forecasting . . . . .	179
8.3	Materials and methods . . . . .	181
8.4	Experimentation . . . . .	185
8.5	Results and discussion . . . . .	189
8.6	Conclusions . . . . .	193
<b>9</b>	<b>Utility tools for DSM: Spatial Load Forecasting</b>	<b>197</b>
9.1	Introduction . . . . .	198
9.2	Related Work . . . . .	199
9.3	Environment Modelling for Spatial Load Forecasting . . . . .	200
9.4	Results and discussion . . . . .	207
9.5	Conclusions . . . . .	214
<b>10</b>	<b>Conclusions</b>	<b>215</b>
10.1	Discussion . . . . .	216
10.2	Future work and open issues . . . . .	218
	<b>Bibliography</b>	<b>221</b>
<b>A</b>	<b>Publications</b>	<b>245</b>
A.1	Indexed journal articles . . . . .	245
A.2	Conference papers . . . . .	245
A.3	Book sections . . . . .	247



# List of Figures

1.1	World energy consumption in 2015 . . . . .	4
1.2	Comparative primary energy consumption over 2005-2015 . . . . .	5
1.3	World CO <sub>2</sub> emissions from fuel combustion by sector. 2014 . . . . .	5
1.4	Simplified diagram of electrical energy grid . . . . .	7
1.5	Overview of a Smart Grid. . . . .	11
1.6	World electricity production in 2015. . . . .	14
1.7	Hourly energy demand (MW) in peninsular Spain . . . . .	16
1.8	Classification of DSM strategies. . . . .	19
1.9	DSM techniques (Zhang and Grossmann, 2016). . . . .	21
2.1	Existing Smart Grid protocols. . . . .	28
2.2	Fixed and variable length frame content. . . . .	32
2.3	Data flow with operations Ox100 and Ox189. . . . .	34
2.4	Mapping of the Op.Code 0x189 frame into CIM classes. . . . .	37
2.5	Mapping of the Op.Code 0x139 frame into CIM classes. . . . .	38
2.6	Embedded SOA architecture. . . . .	39
2.7	Remote SOA architecture. . . . .	40
2.8	Remote real-time SOA architecture. . . . .	41
3.1	Ideal Smart Home Architecture. . . . .	56
3.2	Smart Home Simulator Architecture. . . . .	59
3.3	Diagram of the water deposit management system. . . . .	61

3.4	Diagram of the HVAC system. . . . .	62
3.5	Frequency of data reception . . . . .	66
3.6	HVAC system simulation . . . . .	66
3.7	Gas boiler power profile for a 1 hour simulated consumption profile. . . . .	67
4.1	Conceptual diagram of MASSHA's Human Simulation Model. . . . .	76
4.2	Distribution of the frequency of activation of two sensors in $DS_1$ (single-user scenario). . . . .	90
4.3	Distribution of the duration of activation of two sensors in $DS_1$ (single-user scenario). . . . .	90
4.4	Distribution for the frequency of activation of two sensors in $DS_3$ (multiple-user scenario). . . . .	92
4.5	Distribution for the duration of activation of two sensors in $DS_3$ (multiple-user scenario). . . . .	93
4.6	Distribution for the frequency of activation for two sensors in MASSHA and $DS_1$ (single-user scenario). . . . .	95
4.7	Distribution for the duration of activation for two sensors in MASSHA and $DS_1$ (single-user scenario). . . . .	95
4.8	Distribution for the frequency of activation for two sensors in MASSHA and $DS_3$ (multiple-user scenario) . . . . .	97
4.9	Distribution for the duration of activation for two sensors in MASSHA and $DS_3$ (multiple-user scenario). . . . .	97
4.10	Distribution of the amount of not similar sensors depending on the length of the <i>region of similarity</i> (epsilon) and <i>significance level</i> (alpha) for $DS_1$ (single-user). . . . .	100
4.11	Distribution of the amount of not similar sensors depending on the length of the <i>region of similarity</i> . (epsilon) and <i>significance level</i> (alpha) for $DS_3$ (multiple-user) . . . . .	102
5.1	Real and simulation architectures for the DSM strategy scenario. . . . .	110
5.2	FI-WARE architecture . . . . .	112
5.3	Centralised architecture schema . . . . .	115

5.4	Federation architecture schema . . . . .	115
5.5	Diagram of the Smart City scenario . . . . .	117
5.6	Diagram of the IoT Jungle scenario . . . . .	118
5.7	Diagram of the Ubiquitous World scenario . . . . .	118
5.8	Evolution of the number and type of messages per time in a simulation.	121
5.9	Bandwidth distribution per number of houses in the Smart City scenario.	122
5.10	Bandwidth used per number of houses in the IoT Jungle scenario. . . . .	123
5.11	Bandwidth used per number of houses in the Ubiquitous World scenario.	123
5.12	Comparison of the bandwidth needs of the architectures per scenario. . . . .	125
6.1	Outline of the cost function $\phi_m(p_i)$ . . . . .	135
6.2	Building A. . . . .	140
6.3	Building B. . . . .	140
6.4	Building C. . . . .	140
6.5	Building D. . . . .	140
6.6	Building E. . . . .	140
6.7	Mean load over a year for all buildings. Errors bar denote $\pm\sigma$ . . . . .	140
7.1	Phases of the casting process. . . . .	149
7.2	Optimisation methodology for the casting process. . . . .	153
7.3	Chromosome representation . . . . .	156
7.4	Initial population . . . . .	158
7.5	Implementation of the crossover operation . . . . .	159
7.6	Backtracking algorithm . . . . .	163
7.7	Box-plot per algorithm . . . . .	168
7.8	ECDF per algorithm . . . . .	168
7.9	Fitness landscapes for the genetic algorithm with fixed and variable melting time . . . . .	174
8.1	Average daily load for zone PDT of dataset Ciudad Real . . . . .	188
8.2	Typical load for CRAI building in the University of Deusto. . . . .	189
8.3	Summary of the results for the Very Short Term approach . . . . .	192

8.4	Summary of the results for Short Term problems . . . . .	193
9.1	Simplified class diagram of the environment. . . . .	206
9.2	Overview map of the SLF test area: Ciudad Real. . . . .	207
9.3	Probability density function of the random variable <i>hits + semi</i> . . . . .	211
9.4	Heatmap of the new settlement for Elderly people and Hospitals. . . . .	213
9.5	Heatmap of the new settlement for Families and Schools. . . . .	213
9.6	Heatmap of the new settlement for Young People and Sports Centres. . . . .	214

# List of Tables

3.1	Distribution of sensors in the Smart Home. . . . .	64
4.1	Example of activities and their corresponding actions. . . . .	78
4.2	Definition of activity attributes. . . . .	80
4.3	Definition of action attributes. . . . .	81
4.4	Format of the RAW datasets used to validate MASSHA. . . . .	86
4.5	Pre-process datasets used to validate MASSHA. . . . .	87
4.6	Internal Consistency – Example of the results for one of the executions on $DS_1$ (single-user scenario). . . . .	91
4.7	Internal Consistency – Example of the results for one of the executions on $DS_3$ (multiple-user scenario). . . . .	94
4.8	Coherence with real measurements – Extract of the results of the com- parison MASSHA vs $DS_1$ (single-user scenario). . . . .	96
4.9	Coherence with real measurements – Extract of the results of the com- parison MASSHA vs $DS_3$ (multiple-user scenario). . . . .	98
4.10	Extract of the results of the comparison MASSHA vs $DS_1$ . . . . .	106
4.11	Extract of the results of the comparison MASSHA vs $DS_3$ . . . . .	107
5.1	Results of the Kruskal-Wallis with a Conover post hoc statistical test. . .	124
6.1	Electric tariffs defined by the Spanish BOE . . . . .	130
6.2	Distribution of annual hours within the 6-period tariff. . . . .	132

6.3	Comparison of the results achieved by TM and GA optimisation methods.	143
7.1	Inputs and outputs to the casting optimisation process.	154
7.2	Results obtained with the GA (constant melting time).	163
7.3	Results obtained with the SA (constant melting time).	164
7.4	Comparison of the results obtained with the GA, SA, and backtracking algorithms.	165
7.5	Results obtained with GA (variable melting time).	166
7.6	Results obtained with SA (variable melting time).	167
7.7	Comparison of the three algorithms - Variable Melting Time	167
7.8	P-values for the comparison between algorithms	169
7.9	Best optimisation achieved per selection method (fixed melting time)	172
7.10	Best optimization achieved per selection method (variable melting time)	173
7.11	Best optimization achieved in the fastest time (fixed melting time)	175
7.12	Best optimization achieved in the fastest time (variable melting time)	175
8.1	Description of the components each of proposed model.	183
8.2	Weather Station Selection Results for GEFC12 (MAPE %)	190
8.3	Summary of the results. Mean MAPE $\pm \sigma$ (%)	191
8.4	P-values of the Friedman test for the comparison of the models.	191
8.5	Results for the data imputation problem	194
8.6	Results for the forecasting problem	195
9.1	Infrastructures considered per factor.	203
9.2	Agents types and their preferences	204
9.3	Experimental results for the different agent type composition	210

# List of Listings

2.1	CIM RDF message (beginning). . . . .	43
2.2	SPARQL query. . . . .	44
2.3	Results of the SPARQL query in XML (beginning). . . . .	45
5.1	Header example. . . . .	113
5.2	Payload example. . . . .	113
5.3	Control message example. . . . .	114



# Acronyms

**ABM** Agent-Based Modelling.

**AI** Artificial Intelligence.

**BA** Backtracking Algorithm.

**BOE** Boletín Oficial del Estado.

**CIM** Common Information Model.

**COSEM** Companion Specification for Energy Metering.

**DLMS** Device Language Message Specification.

**DR** Demand Response.

**DSM** Demand-Side Management.

**DSO** Distribution System Operator.

**EC** European Commission.

**EE** Energy Efficiency.

**ENTSO-E** European Network of Transmission System Operators.

**EPRI** Electric Power Research Institute.

**ERP** Enterprise Resource Planner.

**EU** European Union.

**GA** Genetic Algorithm.

**GDP** Gross Domestic Product.

**GHG** Greenhouse Gas.

**GIS** Geographic Information System.

**GSM** Global System for Mobile Communications.

**HVAC** Heating, Ventilating and Air Conditioning Systems.

**IoT** Internet of Things.

**IT** Information Technology.

**JSP** Job-Shop Scheduling.

**LTLF** Long Term Load Forecasting.

**MAPE** Mean Absolute Percentage Error.

**MAS** Multi-Agent System.

**MILP** Mixed Integer Linear Programming.

**MLR** Multiple Linear Regression.

**MO** Market operator.

**NIST** National Institute of Standards and Technology.

**NMLRB** Naïve Multiple Linear Regression Benchmark.

**NN** Neural Network.

**OECD** Organization for Economic Cooperation and Development.

**OSI** Open Systems Interconnection.

**OWL** Web Ontology Language.

**PV** Photovoltaic.

**RDF** Resource Description Framework.

**SA** Simulated Annealing.

**SCADA** Supervisory Control and Data Acquisition.

**SLF** Spatial Load Forecasting.

**SOA** Service-Oriented Architecture.

**SOAP** Simple Object Access Protocol.

**SPARQL** SPARQL Protocol and RDF Query Language.

**SQL** Structured Query Language.

**STLF** Short Term Load Forecasting.

**STLI** Short Term Load Imputation.

**SVM** Super Vector Machines.

**TSO** Transmission System Operator.

**UAV** Unmanned Aerial Vehicle.

**UML** Unified Modeling Language.

**USA** United States of America.

**VAT** Value Added Tax.

**VSTLF** Very Short Term Load Forecasting.

**VSTLI** Very Short Term Load Imputation.

**XML** eXtended Mark-up Language.

**XSD** XML Schema Definition.

*The growth that we want is one that brings real benefits to the people, raises quality and efficiency of development, and contributes to energy conservation and environmental protection.*

Li Keqiang

CHAPTER

# 1

## Introduction

**E**NERGY has become an essential and ever present companion of our daily activities. Indeed, energy needs have broadened in comparison to those in the past, mainly due to the globalisation and the new technological developments that have settled in all aspects of modern life. Take transportation, for example, where we may have to move long distances to get to our workplace or reach our vacation destination. The kilometres travelled per person and per year can be directly translated into the amount of energy needed to provide the fuel necessary to travel those journeys and the energy required to build the transport vehicles. Think of medicine and the advances achieved in life expectancy thanks to the performance of routine clinical studies. These studies rely on the use of sophisticated equipment that required energy to be manufactured and consumes energy to operate. Even more, regardless of our physical location, we dedicate a specific energy quota to stay informed of all the events taking place around us and also to be in constant communication with other people.

All in all, it is obvious that achieving a higher quality of life in regards of personal comfort, transport, and communication leads to a greater amount of energy consumption (Pasten and Santamarina, 2012). For this reason, it is reasonable to think that the greater the energy consumption per inhabitant of a particular town, city, society, or nation, the greater their quality of life and socio-economic development, specially for countries which economies are transitioning from subsistence-based to industrial or commercial-based. According to the United Nations, the energy consumed per inhabitant can be considered an indicator of sustainable development (Eurostat, 2017).

This continuous growth in energy consumption does not come without consequences. Even though renewable alternatives are slowly gaining ground and popularity as primary source for energy generation (International Energy Agency, 2016a), the majority of the energy produced is still based on fossil sources. Undeniably, the use of these sources has a direct, harmful impact on the environment, and contributes to climate change (Intergovernmental Panel on Climate Change, 2014). In the transition from fossil to renewable sources, it is therefore necessary to identify the sectors with a high energy consumption and define measures that promote energy efficiency and enable sustainable advancement on a socio-economic level.

Among the economic activities dependant on the use of energy, the electricity sector is the one with the highest energy consumption. Aside from the environmental effects, uncontrollable electricity consumption can lead to demand peaks and problems in the operation of the electrical grid. In order to mitigate these effects, electric utilities rely on the use of Demand-Side Management (DSM).

The goal of DSM strategies is to encourage the consumer to use less energy during peak demand hours or to move the time of energy use to off-peak times, such as nighttime and weekends. DSM does not necessarily involve a decrease in the total energy consumption, but does facilitate the reduction in the need for investments in networks and power plants to meet peak demands. Thus, DSM applied to electricity systems provides significant economic, reliability and environmental benefits, and facilitates the massive integration of renewable technologies into the electrical grid (Galvis and Costa, 2016).

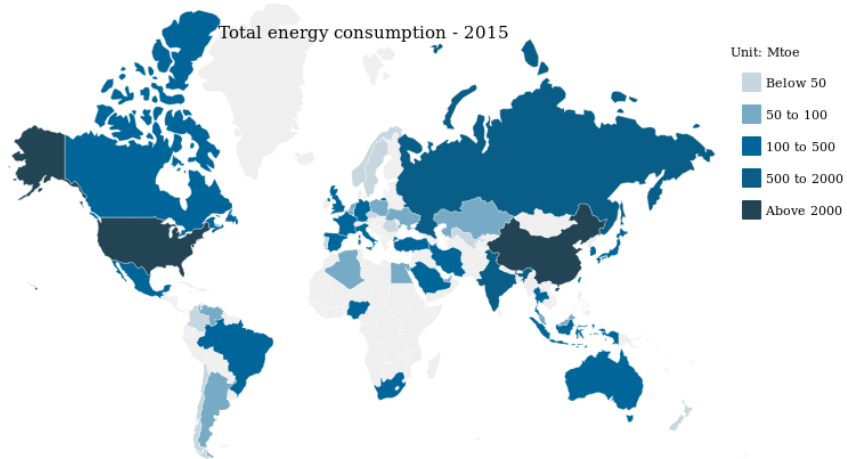
The suitability and success of the application of DSM rely heavily on the inter-connection of all the elements that shape each level of the electrical grid. The thesis presented below focuses on the study of DSM from a bottom-up approach, starting with the study of standardisation capabilities on existing monitoring standards, to the deployment of *in silico* simulations that help analyse the best communication architectures that can maximise the amount of relevant information sent to the utilities, to the analysis of which Artificial Intelligence techniques provide better suitability on domestic, commercial, and industrial DSM scenarios.

## 1.1 Context and motivation

### 1.1.1 World energy consumption and greenhouse gas emissions

During the twentieth century, the leading paradigm was focused on producing energy of good quality in great quantity. Without an abundant amount of energy, the pace of the world's economic growth and the improvement in quality of life would not have been possible to sustain. Undeniably, the global progress achieved in health care, medicine, communications, food production or transport, among others, is due to the extensive use of energy. Figure 1.1 shows the total amount of energy consumption in 2015. As seen, both the United States and China stand out as the greatest energy consumers, while other countries such as Canada, Mexico, Russia, Brazil, India, Australia, and great part of Europe stand as the second group of great energy consumers.

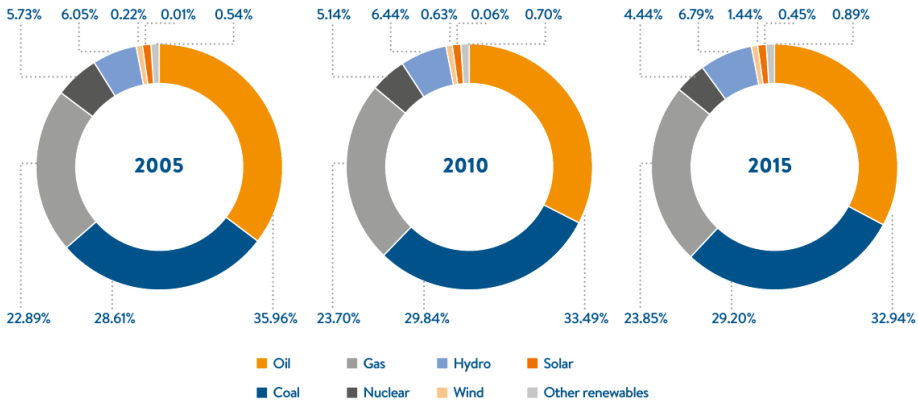
Far from diminishing and according to (U.S. Energy Information Administration, 2016), the world energy consumption is expected to grow by 48 % between 2012 and 2040. The majority of this growth will take place in countries that are not in the Organization for Economic Cooperation and Development (OECD), including countries where demand is driven by strong economic growth, particularly in Asia. Specifically, non-OECD countries, including China and India, account for more than half of the world's total increase in energy consumption.



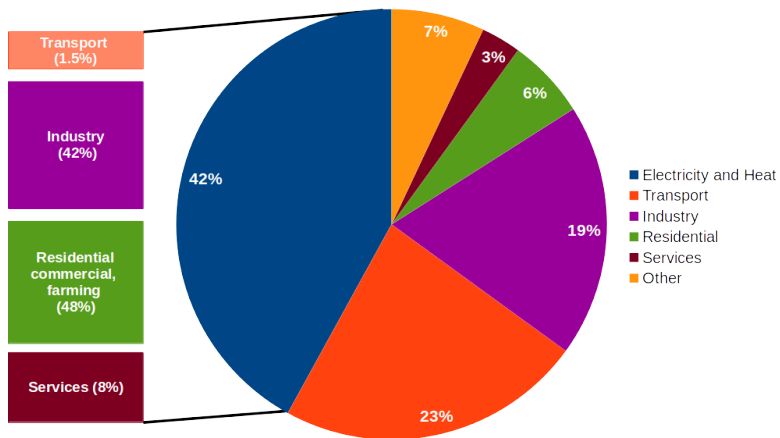
**Figure 1.1:** World energy consumption in 2015. The unit of measure is million tonnes of oil equivalent (Mtoe). 1 toe = 11.63 megawatt-hour (MWh) (Enerdata, 2016).

In this paradigm of ever-growing demand, the importance of an adequate energy management has increased significantly due to the burgeoning awareness of existing limits to the amount of fossil energy sources, as well as the negative environmental impacts associated with them. In fact, the use of fossil sources for energy generation is the reason behind the rise in emissions of CO<sub>2</sub>, a gas that contributes to the greenhouse effect and poses a great threat to the stability of the planets' climate. Indeed, the analysis of the energy consumption over the last 15 years, Figure 1.2, shows that there has been an important shift in the energy generation mix, led primarily by the replacement of coal-fired power with generation from natural gas and renewable sources (World Energy Council, 2016). Nevertheless, oil remains as the primary source for energy generation, and in turn, the reason behind the ever growing increment in Greenhouse Gas (GHG) emissions.

If we look at global GHG emissions from the perspective of the economic activities that lead to their production, we can identify the contribution of each sector to the global amount of CO<sub>2</sub> (International Energy Administration, 2016).



**Figure 1.2:** Comparative primary energy consumption over 2005-2015. (World Energy Council, 2016)



**Figure 1.3:** World CO<sub>2</sub> emissions from fuel combustion by sector. 2014 (International Energy Administration, 2016).

- **Electricity and Heat (42 %):** Burning of natural gas, coal, and oil for electricity and heat is the largest single source of global GHG emissions. Electricity is essential to households and businesses alike. The residential and commer-

cial sector, including farming and public services, consume 56 % of electricity output to power electronic devices, home appliances, televisions, heating and air conditioning, lighting, etc. Industry consumes 42 % to operate electric motors, Information Technology (IT) servers, and keep production lines running. Transportation, on its part, consumes 1.5 % of electricity.

- **Transportation (23 %):** Fossil fuels burned for road, rail, air, and marine transportation. Almost all (95 %) of the world's transportation energy comes from petroleum-based fuels, largely gasoline and diesel.
- **Industry (19 %):** Fossil fuels burned on site at facilities for energy. This sector also includes emissions from chemical, metallurgical, and mineral transformation processes not associated with energy consumption and emissions from waste management activities. Emissions from industrial electricity use are excluded from this category, and are instead covered in the Electricity and Heat sector.
- **Residential (6 %):** Onsite energy generation and burning fuels for heat in buildings or cooking in homes. Emissions from electricity use in buildings are excluded from this category and are instead covered in the Electricity and Heat sector.
- **Services (3 %):** Onsite energy generation and burning fuels for heat in public buildings or restaurants.
- **Other energy (7 %):** Includes agriculture, forestry, and other emissions from the energy sector which are not directly associated with electricity or heat production, such as fuel extraction, refining, processing, and transportation.

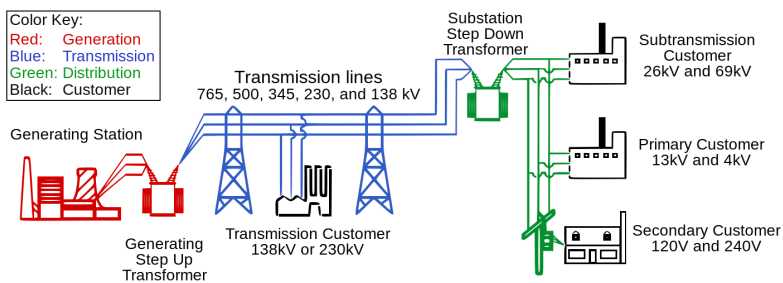
As seen in Figure 1.3, the electricity and heat generation sector stands out as the economic activity with the greatest contribution to CO<sub>2</sub>, accounting for 42 % of the total emissions. These emissions are primarily determined by the available capacity and relative operating costs of the different technologies used for generating electricity. As pointed out previously, most of energy generation is still highly dependent on the use of fossil fuels. Considering the harmful effects of GHG es and the pace of the world electricity demand, it is necessary to develop and deploy strategies that influence the way in which electricity is consumed to favour and adequate and

sustainable use of the resource. These strategies can only be effective with the precise knowledge of how the electrical grid operates, what factors influence the price of energy, and, specially, information about the consumption patterns of the main electricity consumers whether domestic, industrial, or commercial.

### 1.1.2 Electrical sector: the largest energy consumer

The electrical sector is a complex system responsible for distributing the power from the generator infrastructures to the end user, every minute of every day, year-round. The sum of the power requirements of thousands of different users connected to the electric grid, from individual households to large industrial and commercial infrastructures, is what is known as *electricity demand*. Due to the difficulty in storing energy at high scale, the electric grid must maintain a near real-time balance between the amount of energy being generated and the amount of energy being demanded. This balance is of great importance; a mismatch between electricity generation and demand would produce dangerous instability in the network itself; the customers would experience varying voltage, which could potentially damage their installations, or sudden blackouts, which could have serious consequences for applications needing continuous power.

The electrical grid, also known as electrical power system, consists of five major components, as depicted in Figure 1.4.



**Figure 1.4:** Simplified diagram of electrical energy grid. (Liscouski and Elliot, 2004).

- **Generating stations or power plants.** These are the facilities in charge of electricity generation: the process of using primary energy sources to generate electric power. Before reaching the transmission lines, the electrical energy from the generation stations passes through power transformers that increase the voltage so as to minimise transmission losses consequence of the Joule effect (von Meier, 2006). Available generation stations are described below:
  - **Combustion power plants**, which burn fossil fuels (coal, oil or natural gas) or biomass (household waste or plant matter), are the most widespread type of power station and the least expensive to build. Coal-fired plants are the most common, currently generating more than 40 % of the world's electricity, but also emitting the most CO<sub>2</sub> per kWh.
  - **Nuclear power plants** function on the heat produced by atoms of uranium-235 or plutonium-239 undergoing fission. These plants heat the water at high pressure via the heat released in the nuclear fission. This high-pressure steam, moves a turbine connected to an electricity generator.
  - **Hydraulic plants** use the energy of water as it tumbles from natural waterfalls or is moved by dams, tides or currents.
  - **Wind turbines** capture the wind's energy with their enormous blades and use it to spin turbines.
  - **Wave power** can be captured using a variety of systems, such as huge buoys that rise and fall with the swell and large tube sections installed on the water's surface. Underwater turbines work similarly to their onshore counterparts, only they are driven by ocean currents instead of wind. Other technologies are being developed to make use of temperature differences at various depths of the ocean.
  - **Solar Photovoltaic (PV) power plants**, or solar farms, consist of arrays of solar panels (covered in a semiconductor material, such as silicon) that convert sunlight directly into electricity.
  - **Geothermal plants** use the heat within the earth to heat water at a high temperature and high pressure, which is responsible for moving a series

of turbines connected to a generator. These plants are installed in areas in which the earth reaches high temperatures at shallow depths.

- **Transmission lines.** These are necessary to carry high-voltage electricity over long distances in order to connect generators with consumers. Transmission lines are generally interconnected with each other following a mesh structure. This layout allows to transport electricity between very distant points in any direction, thus maximising the reliability of the service. High voltages, between 220 kV and 1000 kV, are used for reducing transmission losses.
- **Substations.** The voltage needed at home is significantly lower than that carried by the transmission lines, therefore the purpose of substations is to convert high voltage to low voltage (or vice versa) in a prior step to entering the distribution network.
- **Distribution.** The distribution network is the part of the electrical energy grid that connects the substations to the places where electricity is being demanded through power lines. Electricity from smaller renewable sources, such as solar and wind, is generally fed into the distribution network.
- **Consumers.** These represent the end points of the electricity grid, the final stage where electricity is demanded.

Aside from the physical infrastructure, there also exist several profiles related to the logical operation, regulation, and coordination of the electrical grid that work together in order to provide an adequate overall management in terms of energy and costs.

- **Market operator (MO).** The MO runs the market settlement system. It matches the bids and offers which are submitted from buyers and sellers and determines the market price based on certain criteria according to the market structure. The MO monitors the delivery of energy and forwards payments from buyers to sellers.
- **Regulator.** The regulator is the governmental body entrusted with the responsibility to ensuring a fair and efficient operation of the electricity sector. It

sets the prices of the services and products that can be provided by the entities having monopolies. It also establish rules for the electricity market and investigate cases in which market power may be misused.

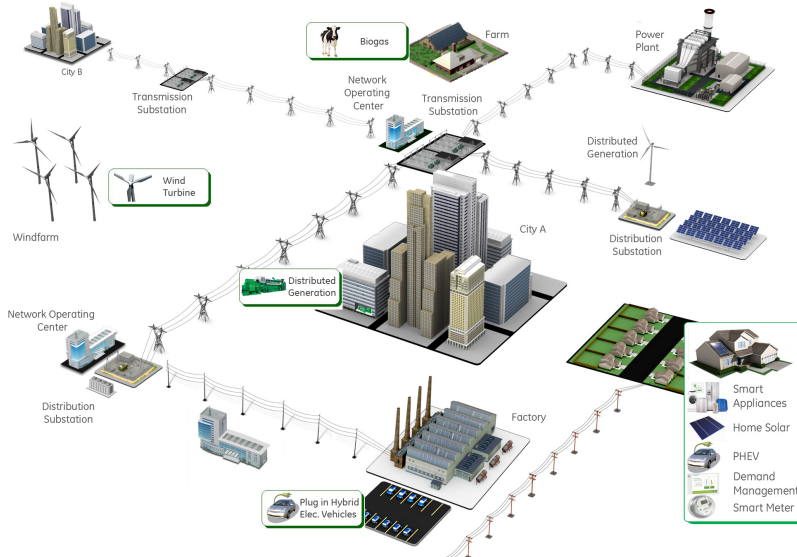
- **Transmission System Operator (TSO).** The TSO controls, coordinates and monitors the operation of the electric power system. It has the prime responsibility of ensuring the security and reliability of the power system. The TSO can provide supply of emergency reserves or reactive power from other entities in the system in order to preserve the system reliability and security. At European level, these are organised in the European Network of Transmission System Operators (ENTSO-E), which draws up 10-year network development plans and participates in the development of rules for operating the network.
- **Distribution System Operator (DSO).** The DSO is a natural or legal person responsible for operating, ensuring the maintenance of and, if necessary, developing the distribution system in a given area and, where applicable, its interconnections with other systems and for ensuring the long term ability of the system to meet reasonable demands for the distribution of electricity.
- **Customers.** A customer is an entity that consumes electricity. Small consumers are connected to the distribution system and they buy electricity from a retailer. Large consumers can, on the other hand, either buy electricity directly from the electricity market by bidding for purchase.

In recent years, and due to the increased availability and accessibility to technology, two new additional figures have arise to influence the operation of the electrical grid.

- **Prosumers.** A prosumer is a new entity concept that consumes but also can produce and store electricity. Prosumers will be able to own and operate small or large parts of the power grid, transport electricity and they will obtain economic optimization according to their energy utilization.
- **Aggregators.** Aggregators can offer services to aggregate energy production from different sources. Their services include the local aggregation of power demand and power supply from customers and prosumers.

## Smart Grid: the evolution of the electrical grid

Originally, electrical grids were designed to attend the needs of a unidirectional flow of energy, one that communicated large centralised sources of supply and distributed locations of consumption. In this paradigm, electricity demand fluctuated independently and the utility operators had to make all the necessary adjustments to meet such demand. However, the ageing of electrical grids, the increment in electricity demand, the emergence of electric vehicles, and the growing prevalence of distributed generation with the inclusion of renewable technologies, highlighted the need to redefine the concept and scope of the traditional grid into a much *smarter* one.



**Figure 1.5:** Overview of a Smart Grid.

A Smart Grid, depicted in Figure 1.5, can be defined as a next-generation electrical power system represented by an increment in the use of communications and IT in all the stages of electrical operation (Yeager, 2016). It typifies the concept of

a power grid in which the major operations (electricity generation, transmission, distribution, and consumption) are coordinated efficiently according to the following developments: (Borlase, 2012):

- Increased use of digital information and control technology to improve security, reliability, and efficiency of the electrical grid.
- Dynamic optimisation of grid resources and operations.
- Development and incorporation of DSM resources.
- Deployment of smart technologies, i.e. real-time, automated, interactive technologies that optimise the physical operation of appliances and consumer devices, for metering, communications concerning grid operations and status, and distribution automation.
- Integration of smart appliances and consumer devices.
- Deployment and integration of advanced electricity storage and peak-clipping technologies, including plug-in electric and hybrid electric vehicles, and thermal storage air conditioning.
- Provision to consumers of timely information and control options.
- Development of standards for communication and interoperability of appliances and equipment connected to the electric grid.
- Identification and lowering of unreasonable or unnecessary barriers to adoption of smart grid technologies, practices, and services.

Smart Grid applications improve the ability of electricity producers and consumers to communicate with one another and make decisions about how and when to produce and consume electrical power. In fact, the modernisation of the grid by using cutting-edge technologies, equipment, and controls that communicate and work together to deliver electricity more reliably and efficiently, can reduce the frequency and duration of power outages, reduce storm impacts, and restore service faster when outages occur. Consumers can better manage their own energy consumption and costs because they have easier access to their own data. Utilities also benefit from a modernised grid, including improved security, reduced peak loads, increased integration of renewable sources, and lower operational costs.

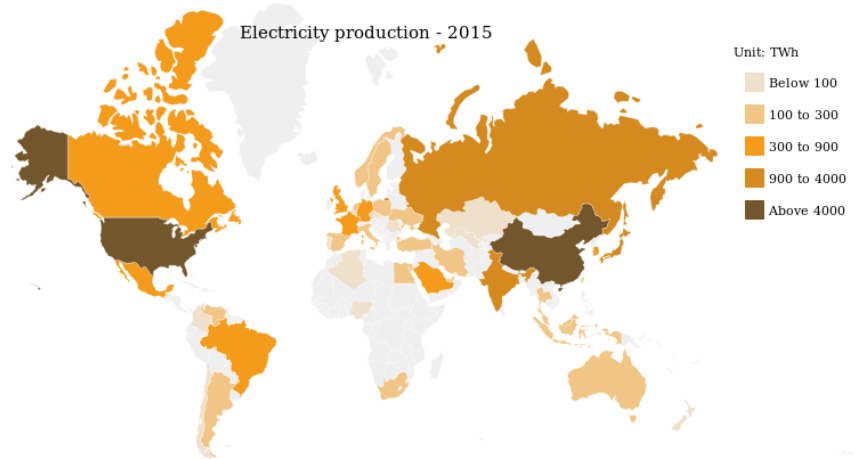
The emergence of the Smart Grid brought to light the need to coordinate large numbers of distributed resources through sensors, actuators, and monitoring systems from different manufacturers. However, each device *spoke* in its own particular language which hindered the direct communicating among them. Protocol harmonisation became one of the foremost and important tasks to undo this *Tower of Babel* and enable communication and interoperation within the Smart Grid. This harmonisation involved the creation of a common data model, the Common Information Model (CIM), that described not only each of the entities living in every stage of the Smart Grid, but also the relationships between them. Once defined, each device only had to map the particular attributes of its communication protocol to those of the common data model. This approach allows the seamless integration of all present and future resources in the Smart Grid.

As proof of concept, in **Chapter 2** (Penya et al., 2011), we focus on the harmonisation of the CIM and the IEC 60870, the standard adopted by the Spanish TSO for the management of high-voltage electricity meters, to show that in some cases a simple mapping suffices to achieve interoperability.

### **Trends in electricity consumption**

Electricity needs continue to rise guided by the socio-economic growth and the greater per capita consumption of the world population. Global power generation increases by 2.2 % a year and is expected to double by 2040, with countries such as China and India, expecting to increase their generation by 260 % and 180 %, respectively, between now and 2030 (International Energy Agency, 2016b). Currently, China is the world's largest generator of electricity (24 % of global output), followed by the United States (18 %), India (5.1 %), Russia (4.5 %) and Japan (4.5 %), then by Canada, Germany, Brazil, France and South Korea, as seen in Figure 1.6.

Of the total amount of CO<sub>2</sub> emissions derived from electricity generation, 73 % can be attributed to coal-fired power plants, which emit 950 grams of CO<sub>2</sub> for every kWh of electricity they generate, compared with 350 grams for gas-fired power plants. For power plants that run on renewable energies, such as hydro, wind, solar PV and solar thermal, the only CO<sub>2</sub> emissions are attributable to their construction.



**Figure 1.6:** World electricity production in 2015. The unit of measure is terawatt-hours (TWh). 1 TWh =  $10^6$  megawatt-hour (MWh) (Enerdata, 2016).

Accordingly, for every kWh of electricity generated, a solar PV system emits between 60 and 150 grams of CO<sub>2</sub> (depending on where the solar panels were manufactured), a wind turbine between 3 and 22 grams, and a hydropower plant 4 grams. As for nuclear power plants, even after the future need to dismantle aging facilities is factored in, CO<sub>2</sub> emissions still only represent 6 grams per kWh of electricity generated, a stark contrast with the 950 grams emitted by coal-fired power plants.

Europe's drive for sustainability has been further continued with the set objectives for 2030, aiming for 40 % emission reduction compared to 1990 levels and at least a 27 % share of renewable energy sources (European Commission, 2014). For the longer term, the European Commission (EC) targets a zero CO<sub>2</sub> emitting electricity sector in 2050. Those objectives for the electricity sector have a large impact on the expected development of electricity generation, but also on the evolution of demand. In an electric system that is largely based on renewable electricity sources, it is desired to have higher electricity consumption in moments when more renewable electricity is being produced, and a lower consumption in times of lower renewable production. Maintaining reliable and secure electricity services while seeking to

rapidly decarbonise power systems is a key challenge for countries throughout the world.

In this sense, a major innovation in Smart Grid is to also include the management of flexible loads, which is generally referred to as DSM since it involves the enhancement of energy systems on the electricity demand side. DSM is expected to play a crucial role in the improvement of grid efficiency and reliability as well as the creation of additional benefits for the consumers (Levy, 2006; Strbac, 2008); this has entailed tremendous research efforts across multiple disciplines, such as electrical engineering, civil and environmental engineering, economics, data science, behaviour science, and public policy.

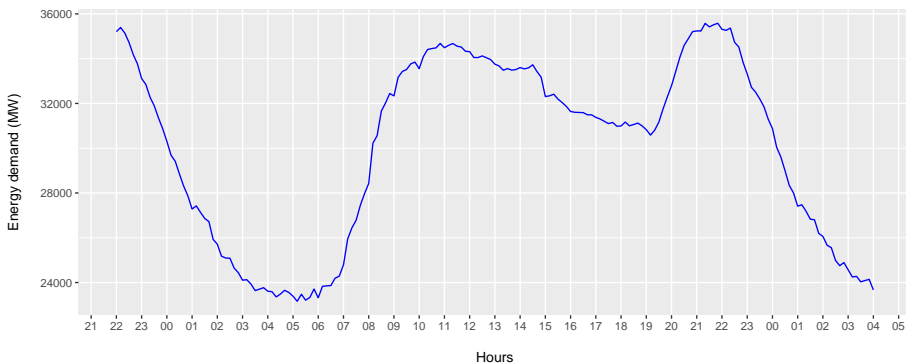
### 1.1.3 Demand-Side Management in the electrical sector

As noted in Section 1.1.2, the electrical grid is designed to reliably match supply and demand. This task has become increasingly challenging due to high fluctuations in the electricity demand and the increasing penetration of intermittent renewable energy into the electricity supply mix (International Energy Agency, 2016a). Also, the pressure to reduce GHG emissions have further amplified the need to improve the efficiency, reliability, and sustainability of the power grid. In order to achieve this, it is important to understand the basic concepts underlying its operation:

- Electricity consumption and production must balance at all times. Any significant imbalance could cause grid instability or severe voltage fluctuations, resulting in grid failures. Certain types of power plants, such as coal and nuclear power plants, have little short-term flexibility in adjusting their electricity output, so the amount of electricity supplied by them is already programmed and scheduled within the grid operation. Other plants, such as natural-gas are often used to meet peaks in demand. Finally, electricity generated from renewable sources, such as wind and solar PV, is generally used whenever these resources are available.
- Electricity consumption is variable over time in response to the habits of consumers, yearly seasons, climatology, work calendar, etc.

- Each energy generation facility has a different set of associated costs depending on the type of energy source used, maintenance costs, infrastructure amortisation, etc.

In this sense, forecasting of the electricity demand, as far in advance as possible, is an essential factor for the operation of the electrical power industry. The electric utility analyses the foreseen load to make decisions on making investments, planning for efficient management, schedule maintenance, and develop DSM strategies to influence the demand. The prediction of the electric load at a future time is a challenging problem because of the diverse characteristics of the electrical load and the high uncertainty associated with them.



**Figure 1.7:** Hourly energy demand (MW) in peninsular Spain. February 21<sup>th</sup>, 2017.

Figure 1.7 shows a typical daily variation of electric load for peninsular Spain. As seen, the maximum consumption takes place around 21:00-22:00, while the minimum consumption takes place around 5:00. The characteristics of the electric load depend on the nature of the users and the devices used. From this point of view, the electric load can be classified into six categories:

- **Domestic.** Electricity demand in the domestic sector is characterised by a great amount of consumers with low electricity consumption. There is no fixed load pattern that describes the general behaviour of this type of load. It varies from

home to home influenced by factors, such as location, weather, personal and professional characteristics of the inhabitants, etc.

- **Commercial.** Commercial demand consists on lighting for shops, fans, and electric appliances. This type of load takes place for more hours during the day in comparison to the domestic load, and has seasonal variations due to the extensive use of air conditioners and space heaters.
- **Industrial.** Electricity demand in the industrial sector is characterised by a small amount of consumers with high electricity consumption. The magnitude of the industrial load depends on the type of industry. Small scale industries demand up to 25 kW, medium scale industries demand between 25 kW, and 100 kW, and large-scale industries demand above 500 kW. Industrial loads are generally not weather dependent.
- **Municipal.** Municipal demand consists on street lighting and power required for water supply and drainage purposes. Street lighting requirements are practically constant through the night. For water supply, water is pumped to overhead tanks during off-peak periods to help improve the load factor of the power system.
- **Irrigation.** This type of load refers to the electric power needed by irrigation pumps to supply water to fields. Generally this type of load is supplied all night long.
- **Traction.** This type of load includes tram cars, trolley buses, railways etc. This type of load is characterised by a variable profile. The peak value takes place on the first hours of the morning, when people go to work. Then the load decreases and rises again on the evening, when people get back home.

In **Chapter 8** we present a comprehensive study of how the addition of information from consumers can improve electricity load forecasting for several of the previous categories.

The traditional control strategy of the electrical grid aims to balance the amount of energy being generated and the amount of energy being consumed by acting over the *generation*. In contrast, DSM strategies try to act over the *demand*. In order to achieve this goal, DSM strategies rely heavily on the constant monitoring of all the

entities living within the energy grid and the thorough analysis of the behavioural patterns that define the shape of the energy loads.

Existing definitions of DSM often represent electricity management from the grid's operator point of view. According to (Kreith and Goswami, 2016):

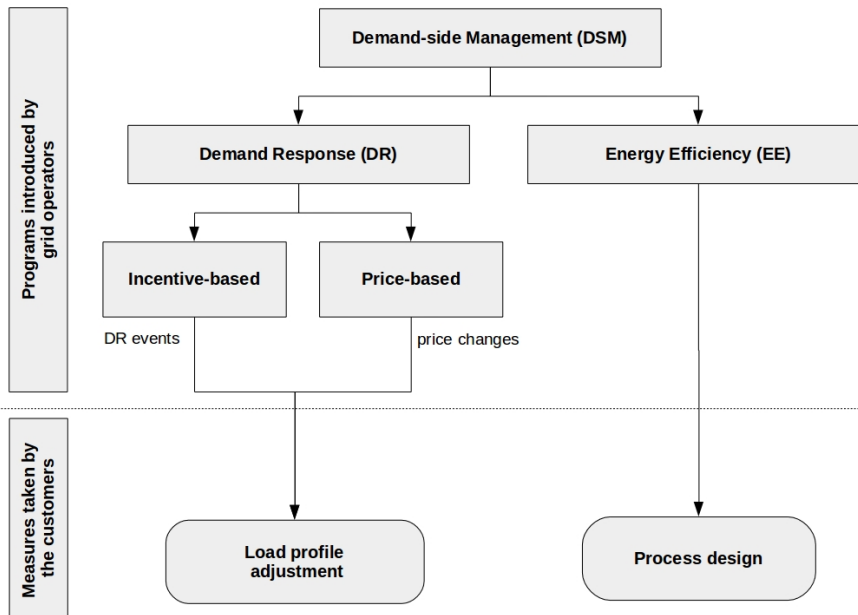
*“DSM involves all the activities planned and implemented by electric utilities that are designed to influence the customers' use of electricity to produce changes in the shape of the demand load.”*

In this definition, DSM only regards those activities carried out by the utilities and the government, while customers are considered simple passive entities that only react to DSM strategies. This definition is correct and highlights the origins of DSM as a concept proposed by the utilities (Gellings, 1985), however, it undervalues the decision-making capacity of the customers that can enable a real variation in their consumption patterns. Actually, final customers are the only ones that really perform that change. Utilities and governments can only provide incentives that encourage such activities. Therefore, the definition of DSM can be broadened to include both perspectives (Kreith and Goswami, 2016):

*“DSM refers to the set of activities or measures at the interplay between the grid operators and the electricity consumer designed to change the amount and/or timing of the consumer's use of electricity.”*

Specifically, DSM on the operator side involves the analysis of the demand, the identification of needs for load adaptation, and the creation of financial incentives for the consumer. On its part, the consumer reacts to these financial incentives and performs the actual physical load adjustment operations. For example, utilities may signal demand requests to their customers including simple off-peak metering, in which power is cheaper at certain times of the day, and smart metering, in which explicit requests or changes in price can be communicated to customers. The customer may adjust their power demand by postponing some tasks that require large amounts of electric power, or may decide to pay a higher price for their electricity.

Figure 1.8 shows a general classification of DSM strategies. The two main DSM categories are Energy Efficiency (EE) and Demand Response (DR) (River, 2005). The goal of EE is to reduce power consumption without variation in the type of tasks, while DR refers to load profile adjustments driven by market incentives.



**Figure 1.8:** Classification of DSM strategies.

EE can be achieved by improving the design of existing processes by optimal scheduling and control strategies that maximise the time in which the process runs at its most energy-efficient operating point.

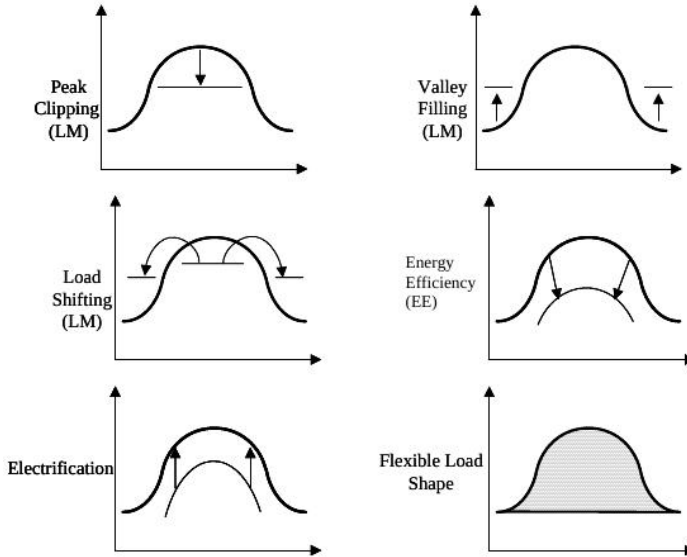
DR differentiates between dispatchable and non-dispatchable DR (Federal Energy Regulatory Commission, 2010), which are often also referred to as incentive-based and price-based DR (U.S. Department of Energy, 2006), respectively. Dispatchable DR refers to load adjustment capacities that consumers provide to the grid

operator in order to maintain the grid stability in times of emergency. The various types of dispatchable DR resources differ primarily in the amount of time the consumer has to respond to DR requests. Providing dispatchable DR requires a high degree of flexibility in the consumer's process since requests by the grid operator, which typically cannot be anticipated in advance, have to be met while maintaining the feasibility and safety of the process. The immediacy in responding to these requests determines the value of the DR service, and in turn the amount of money consumers get paid for providing such DR capacities. Non-dispatchable DR resources are not controlled by the grid operator; instead, consumers choose to adjust their power consumption profile based on price signals sent from the electricity market according to the needs of the power grid.

Any DSM technique implemented may result in one of the following forms of demand reduction (Zhang and Grossmann, 2016), as seen in Figure 1.9:

- Peak Clipping and valley filling refer to the reduction or increment of utility loads during peak or low demand periods. This can delay the need for additional generation capacity.
- Conservation refers to reduction in consumption by consumers. There is net reduction in both demand and total energy consumption. Strategic conservation can be implemented by motivating customers to use more energy-efficient appliances.
- Load Shifting involves shifting loads from on-peak to off-peak periods. The net effect is a decrease in peak demand, but there is no change in the total energy consumption.
- Flexible load shape involves making the load shape responsive to reliability conditions. In exchange for various incentives, the utilities may temporarily apply service constraints or other limitations to the quality of the energy supply.
- Electrification involves load building over all hours and is often associated with customer retention programs from the perspective of the utility.

In summary, DSM strategies offer electricity consumers the opportunity to shift their load either in response to price signals or in exchange for an incentive. Through



**Figure 1.9:** DSM techniques (Zhang and Grossmann, 2016).

this load shifting, DSM programmes offer the possibility to downsize the pressure on the transmission and the distribution network, i.e. increase the load on a voluntary basis either when the power supply is more available, for example, due to large output from variable renewable sources, or to reduce the load when the power supply is particularly tight or at peak levels, without significant impact on the daily life of consumers. In **Chapter 7** (Esteban and Peña, 2012) and **Chapter 6**, we introduce the use of Artificial Intelligence (AI) techniques in the foundry industry, and in a commercial building (a university), that automatically adjust their operation to these price signals, achieving theoretical savings of and 18 % and 30 %, respectively.

The electric power industry considers DSM as an increasingly valuable resource option whose capabilities and potential impacts are expanded by grid modernisation efforts. For example, sensors can perceive peak load problems and use automatic switching to divert or reduce power in strategic places, removing the chance of over-

load and the resulting power failure. Advanced metering infrastructure expands the range of time-based rate programs that can be offered to consumers. Smart customer systems such as in-home displays or home-area-networks can make it easier for consumers to changes their behaviour and reduce peak period consumption from information on their power consumption and costs. These programs also have the potential to help electricity providers save money through reductions in peak demand and the ability to defer construction of new power plants and power delivery systems specifically, those reserved for use during peak times.

Several deployments of Smart Homes and Smart Grids in real environments have been carried out so as to understand the particular aspects of DSM strategies. In any case, there are several important issues that must be taken into account, such as the lengthy process of configuration and maintenance of the electronic devices, as well as the amount of time needed to acquire a sufficient amount of data for analysis. Moreover, the high economic investment and time restrictions related to the deployment of Smart Homes and Smart Grids in real or scaled environments highlights the advantages of *simulation models*. A careful analysis in terms of real Smart Grid deployments and simulators shows that none of them were built with the purpose of analysing the human aspects of DR strategies: What are the potential impacts that may derive from the implementation of DSM strategies? What are the effects of deploying these solutions in case that human factors are not considered? What strategies will work better and how can they be implemented in order to avoid rejection from citizens?

In order to try to clarify all these questions, in **Chapters 3 and 4** (Kamara-Esteban et al., 2017), we present the first steps towards the implementation of a Smart City Agent Based Simulator, GeoWorldSim, for evaluating DSM strategies, following the GreenSoul methodology (Apostolou et al., 2016). One of the important pillars of the simulator is the capacity of reproducing human behaviour. This is an advantage, since we can reproduce exact patterns of different person characteristics and study how they are affected by load demand. Specifically, the Smart City Simulator have been used to perform the following studies:

- In **Chapter 5** (Kamara-Esteban et al., 2017), we evaluate several architectures for the appropriate communication middleware that retrieves the relevant information to the utilities for the domestic DSM strategies.
- In **Chapter 9** (Borges et al., 2014), we evaluate whether changes in future demographic settlements will affect the reliability of the electrical grid through Spatial Load Forecasting.

## 1.2 Dissertation objective

Energy demand is increasing globally and in consequence GHG emissions from this sector are on the rise as well. This trend is set to continue, driven primarily by the economic growth and the rising population. Solutions in this area go hand in hand with the worldwide deployment of policies that look forward a better management and usage of energy in domestic, commercial, and industrial scopes. In this line, load balancing through DSM strategies comes out as one of the most effective and immediate actions aimed at achieving efficiency in the use of energy resources. DSM strategies consist of the planning, implementing, and monitoring activities designed to encourage consumers to modify their levels of electricity demand.

Actually, the benefits of the application of DSM strategies can be analysed from two points of view. On the utility side, the ability of customers to shed loads during periods of peak demand through is beneficial to the entire electric system for two reasons. First, under tight electricity supply and demand conditions DSM can significantly reduce peak prices and overall price volatility for all users. Second, by reducing system peaks, DSM may reduce the need for expensive changes in the system infrastructure, such as new generation, transmission, and distribution facilities to meet these peaks in demand.

On the customer's side, DSM can offer real financial savings on the electricity bill. While overall energy reductions can be difficult to achieve, a substantial reduction over a short period of time during a price spike can be a much easier goal to accomplish. DSM allows to identify these spikes by scheduling power intensive processes outside of these times.

In this scenario, the main objective of the thesis is to understand and identify the DSM needs in all layers of the electrical power system. From the needs of standardisation capabilities on existing monitoring standards, to the deployment of in silico simulations that help analyse the best communication architectures that can maximise the amount of relevant information sent to the utilities, to the analysis of which AI techniques provide better suitability on domestic, commercial, and industrial DSM scenarios.

In order to achieve the main objective, the following specific and measurable objectives are also defined:

- Develop the first steps towards a Smart City simulation platform as a decision making tool prior to the deployment of new infrastructures and with the purpose of providing the social component of the application of DSM strategies.
- From the DSM customer's point of view:
  - Determine the appropriate communication characteristics needed to issue accurate DSM strategies to the domestic scenario characterised by the existence of a high amount of consumers with low electricity consumption.
  - Design and implement a solution based on the use of AI techniques to apply DSM to a real case industrial scenario, characterised by the existence of a small amount of consumers with high electricity consumption.
  - Design and implement a solution based on the use of AI techniques to apply DSM to a real case commercial scenario, where price signals can influence the fluctuations of the demanded load.
- From the DSM customer's point of view:
  - Design and implement a solution based on the use of AI techniques to issue spatial forecasts that let electric utilities decide the when, where and how to apply DSM strategies to lead with the expected future electricity demand.

## 1.3 Dissertation outline

DSM strategies rely on the coordinated communication of all the stages of the Smart Grid. Consistent with this paradigm, this dissertation has been structured following a bottom-up approach that describes the research done on each of stage.

**Chapter 1** describes the context and motivation of the research presented in this thesis. The chapter focuses primarily on the importance of the application of DSM strategies to influence the electricity demand of the end users, both for avoiding demand peaks that could lead to the destabilisation of the electrical grid, and for maximising the integration of renewable energy sources.

**Chapter 2** presents the harmonisation between the IEC 60870, the proprietary Spanish high-voltage meter protocol and the Common Information Model, to demonstrate the feasibility of achieving interoperability with the Smart Grid, enabling in this way the retrieval of on site information needed to characterise and implement DSM programs.

**Chapter 3** introduces an agent-based Smart City Simulator for the evaluation DSM strategies. The chapter focuses on the technical aspects of the implementation, from the characterisation and operation of appliances to the definition of an appropriate human behaviour model.

**Chapter 4** presents the agent-based human-behaviour engine behind the implementation of the Smart City Simulator. This chapter introduces the characterisation of human actions in line with the Belief-Desire-Intention paradigm, providing a sound validation methodology to evaluate the correspondence between synthetic and real human activity datasets.

**Chapter 5** focuses on the problem of domestic DSM from the point of view of analysing which is the best architecture solution to gather all the information relevant to the design and implementation of DSM strategies while maintaining appropriate response times in the real time monitoring.

**Chapter 6** demonstrates the feasibility of using AI for aiding commercial DSM strategies. Specifically, this chapter focuses on the design and development of a methodology that adjusts the terms of the electrical tariff of a university campus,

composed of 5 buildings, based on historical consumption data. The analysis estimates different alternatives of pricing, depending on the use of individual or joint consumptions.

**Chapter 7** demonstrates the feasibility of using AI for aiding industrial DSM strategies. Specifically, this chapter focuses on the design and development of a methodology that best schedules the execution of orders in a foundry plant in order to minimise the cost of the casting process while meeting production deadlines.

**Chapter 8** focuses on a methodology to improve Short Term Load Forecasting at substation level by gathering data from nearby substations to improve data imputation methods and predict future trends in electricity consumption.

**Chapter 9** focuses on the use of the agent-based systems to improve Spatial Load Forecasting. Specifically, this chapter presents a methodology that can help electric utilities in assessing whether the electrical infrastructure deployed at a certain location is able to support the demographic growth.

Finally, **Chapter 10** draws the conclusions we come up with after developing this research. We exhibit a critical analysis of the performed research, and define the future lines of work.

*Every language is a world. Without translation, we would inhabit parishes bordering on silence.*

George Steiner

CHAPTER

# 2

## Standardising Smart Grid communication protocols

**D** SM STRATEGIES rely on the continuous monitoring and coordination of large amount of resources through sensors, actuators, and communication protocols for efficient electricity generation and consumption management. Networking of state-of-the-art measuring and control technologies prevent overload situations and enable the integration of future network components such as energy accumulators and e-mobility.

However, the technological transition from the original distribution and transport networks into the Smart Grid, brought to light complications derived from the lack of interoperability among the large quantity of existing protocols and the need to devise a common language that enabled real-time operation and communication

among them, specially when integrating smart buildings into the Smart Grid. In fact, advances in building automation have grown independently to those in Smart Grids. As buildings acquire importance becoming a basic node of the Smart Grid, the need to create interoperability among the protocols from both sides becomes even more evident.

This chapter shows how integration can be performed by defining a harmonised knowledge/data model that allows to seamlessly connect smart buildings into the Smart Grid, facilitating in this way the retrieval of on site information needed to characterise and implement DSM programs.

## 2.1 Introduction

Standards play a key role in the development and deployment of technology in society. Common and interoperable standards encourage innovation and boost productivity by reducing or eliminating technical barriers. In this sense, one of the big challenges that Smart Grids face is the communication and interoperability of all the protocols available for the different stages of the Smart Grid operation, as shown in Figure 2.1.

<p><b>Smart Network Management</b></p>	<ul style="list-style-type: none"> <li>• Electromagnetic compatibility and power quality</li> <li>• Advanced network operation and control</li> <li>• Smart metering and power line communication</li> </ul>	<ul style="list-style-type: none"> <li>• IEC 61000 series</li> <li>• IEC 61968/61970/62325 (CIM)</li> <li>• IEC 61850 series, IEC 60870 series</li> <li>• IEC 62689 series</li> <li>• IEC 62351 series</li> <li>• IEC 60255 series</li> </ul>
<p><b>Smart Integration of Distributed Generation and e-mobility</b></p>	<ul style="list-style-type: none"> <li>• Integration of distributed generation</li> <li>• Integration of electric vehicles</li> <li>• Integration of new usages such as storage, heating and cooling.</li> </ul>	<ul style="list-style-type: none"> <li>• EN 50438</li> <li>• IEC 61850 series</li> <li>• TS 50549 - 1 &amp; 2</li> <li>• ISO/IEC 15118</li> <li>• IEC 62786</li> <li>• IEC 61851</li> </ul>
<p><b>Smart Markets and Active Customers</b></p>	<ul style="list-style-type: none"> <li>• Enable DSO to act as market facilitator and grid optimiser</li> <li>• Develop demand response and demand side management programmes</li> <li>• Aggregate distributed energy resources and e-mobility</li> <li>• Balance the power grid</li> </ul>	<ul style="list-style-type: none"> <li>• IEC 61968/61970/62325(CIM)</li> <li>• IEC 62056 (DLM/COSEM)</li> <li>• IEC 61850series</li> <li>• SEP 2.0, Open ADR</li> </ul>

**Figure 2.1:** Existing Smart Grid protocols.

As accounted in (Crapo et al., 2009b), the National Institute of Standards and Technology (NIST) originally identified 16 key Smart Grid standard protocols (National Institute of Standards and Technology, U.S. Department of Commerce, 2009). Later on, a report from the Electric Power Research Institute (EPRI) showed that the list of standard protocols could increase up to 77 items (EPRI, 2009c). This drawback aggravates when it comes to the management of high-voltage meters. For instance, ANSI C.12 (National Electrical Manufacturers Association, 2006) is the established standard for meter management in the United States of America (USA), while Companion Specification for Energy Metering (COSEM) (IEC, 2006) is so Europe-wide. Spain, though, remains still loyal to the IEC 60870 (Commission, 1996).

There have been some interesting new directions explored within this domain, especially those focusing on the bidirectional integration of two Smart Grid protocols. In general, all of these works intend to place one of the targeted protocols as the basis, and then, they map the properties of the other through a middle-ware layer, similar to the *protocol tunneling* achieved in other network domains (ANSI, 2010). This verge of research has recently turned to semantic technologies in their search for tools to accomplish interoperability. For instance, the appearance of the Unified Modeling Language (UML) model (Booch et al., 2005) of the CIM standard (IEC, 2010) has been highlighted as the reference *ontology* (Becker et al., 2010; Neumann and Nielsen, 2010; Britton and deVos, 2005) to be extended in order to include the rest of related protocols (Fairchild, 2002; Wang et al., 2003; Crapo et al., 2009b,a; Rohjans et al., 2010a,b; Peña and Peña, 2011) (approach envisioned and backed by the IEC (Smart Grid Strategic Group (SG3), 2010) and the EPRI (EPRI, 2008a, 2009a, 2008b, 2006, 2009b)), starting by CIM's sort of semantic ancestor in the USA, MultiSpeak. This methodology has also been promoted by the Technical Committees developing the involved protocols, as is the case of Device Language Message Specification (DLMS)/COSEM, where the DLMS Users Group itself has led a study to integrate this protocol into CIM (Taylor and Kmethy, 2010).

With the emergence of the *semantic web*, this collaboration has become easier and feasible. More accurately, one of the basic pillars of the semantic pyramid, namely *web services* or Service-Oriented Architecture (SOA), very much enlighten and facili-

tate this challenge since they have been specially conceived to allow communication between different applications and distinct sources of data. In a way, SOA sends and receives the information within standardised eXtended Mark-up Language (XML) envelopes by means of the Simple Object Access Protocol (SOAP), achieving the desired interoperability, providing as well a very easy service discovery and publication mechanism, and an efficient service description language. Hence, any ad-hoc, non-proprietary or standard communication or data exchange protocol is susceptible to be wrapped into a SOA mediator in order to fulfil the same functionality but in an open, interoperable way.

Aware of the opportunities that semantic models enable, the IEC has defined an OWL model of the CIM (IEC, 2010; Crapo et al., 2009a), setting the basis for many research projects that have extended the CIM to include related protocols (Taylor and Kmethy, 2010), combined the CIM ontology with other protocols' ontologies to achieve interoperability (Breedillet et al., 2010), etc. Still, the semantic paradigm has not been fully exploited, since SPARQL Protocol and RDF Query Language (SPARQL), a structured query language for semantic sources, has not been fully applied in this domain.

## 2.2 IEC 60870

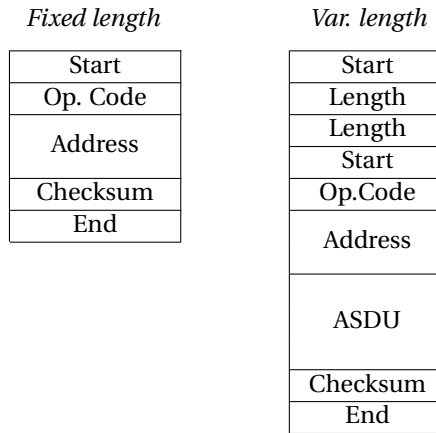
The IEC 60870 is the standard adopted by the Spanish TSO for the management of high-voltage electricity meters. This standard specifies the communication protocol between registers or accumulators and meters. It presents three layers, namely *physical*, *link*, and *application* (taking the Open Systems Interconnection (OSI) Model as reference).

- **Physical** layer: The IEC 60870 foresees three approaches to access the meters, one direct and the other two remote (via Global System for Mobile Communications (GSM) modem or TCP/IP, depending on the device model and its features). We focus here on the remote GSM alternative since all meters have a GSM modem attached. Therefore, the physical connection is performed through AT-HAYES commands (Telit Wireless Solutions, 2011).

- **Link** layer: This layer is dedicated to establish a communication session between the meter and the external querier by the assignment of an unique `session id` that lasts until the communication is closed. The vehicle to exchange data in this layer is the frame. The IEC 60870 specifies their type and order, being usually grouped in question-answer pairs, except for the `session-open` and `session-close` frames that receive a fixed length frame with `Op.Code 0x10 (ack)` in case the operation is successful. The connection in this layer is established as follows: the querier sends a fixed length frame with `Op.Code 0x40`, the meter responds with a `0x0B`, then the querier replies with an `Op.Code 0x40` and, if the connection is successful, the querier receives an `ack` frame.
- **Application** layer: The protocol enables non-password protected operations (i.e. non critical) such as inquiring the meter and vendor id, reading current date and time, official work-schedule dates (holidays and vacations), contracted power, consumed active and reactive power (accumulated and stored since the last reading), and so on. Password protected functions include all meter configuration operations that could lead to a deceitful or defective data retrieval such as changing date and time, modifying contracted power, holidays, password, or reading the meter's configuration information (since this information contains the password).

On a lower level, data is transmitted by using two types of frames. Fixed length frames are used to transmit acknowledge messages (`ack`, i.e. operation successful), error codes or information about the status of the communication. Variable length frames carry questions and answers along with the required data. Figure 2.2 shows the schema for fixed and variable length frames, with each cell representing a single byte. The meaning of each field is explained as follows:

- **Start**: This field indicates the head of the frame. In fixed length frames, this field is set to `0x10` whereas in variable length frames, it is assigned the value of `0x68`. In the latter case, this field appears twice in the frame due to security reasons.



**Figure 2.2:** Fixed and variable length frame content.

- **Operation Code:** This field determines the purpose of the frame.
- **Address:** It defines the MAC address of the target meter. This field comprises 2 bytes sent in *Little Endian* format.
- **Checksum:** Allows to check the integrity of the received data.
- **Length:** Specifies the amount of bytes of the frame. In variable length frames, this field appears twice due to security reasons.
- **ASDU** (Application Service Data Unit): The length of this field depends on the amount of data to be sent. It contains a byte specifying the type of ASDU (`cdt`), the address of the target (including `meter id` and an application layer address), and the data to be transferred.
- **End:** Similar to `Start`, this field signals the end of the frame. In both fixed and variable length frames its value is set to `0x16`.

IEC 60870 compliant meters store, every fifteen minutes, electricity consumption data in a cyclic buffer, i.e. a buffer that replaces older data with newer information), which capacity depends on the vendor and the model of the device (the average being about three months worth of data). As any simple meter control and management

protocol, the IEC 60870 offers basic meter management services and load data querying (absolute, relative, and accumulated values). All services that access private or sensitive data are password protected.

Figure 2.3 illustrates the frame flow issued when performing the following two operations: first, retrieving the meter vendor code (Op.Code 0x100) and, second, querying absolute integrated consumption readings (Op.Code 0x189). We assume that the connection is not yet established and that it will be closed after receiving the requested information.

Once the session is opened, we can already perform *open* functions such as the retrieval of the vendor id, which always includes the same number of frames (one for the query, a fixed length frame, and a variable length one for the answer). We have also portrayed an operation whose amount of needed frames depends on the length of the information to be transmitted, for instance reading the absolute integrated consumption, as in Figure 2.3. It starts with a fixed part (a request with Op.Code 0x189, an ack response), and then a loop that is repeated until all the information is transmitted. In this case, the querier issues an Op.Code 0x11 frame to continue getting data, which is a variable length frame with Op.Code 0x139 and whose ASDU shows ASDU type cdt 0x05 (meaning answering request). The last of the frames sent by the meter in this operation has also Op.Code 0x139 and cdt 0x10 (meaning transmission end).

## 2.3 Common Information Model (CIM)

The CIM is an abstract information model that provides data understanding through the identification of the relationships and associations of the data within a utility enterprise. This enhanced data understanding supports the exchange of data models and messages and increases the ability to integrate applications both within the enterprise and with trading partners. These trends go beyond exchange or updates of network models to the exchange of specific dynamic data within transactional messages in a real-time environment.

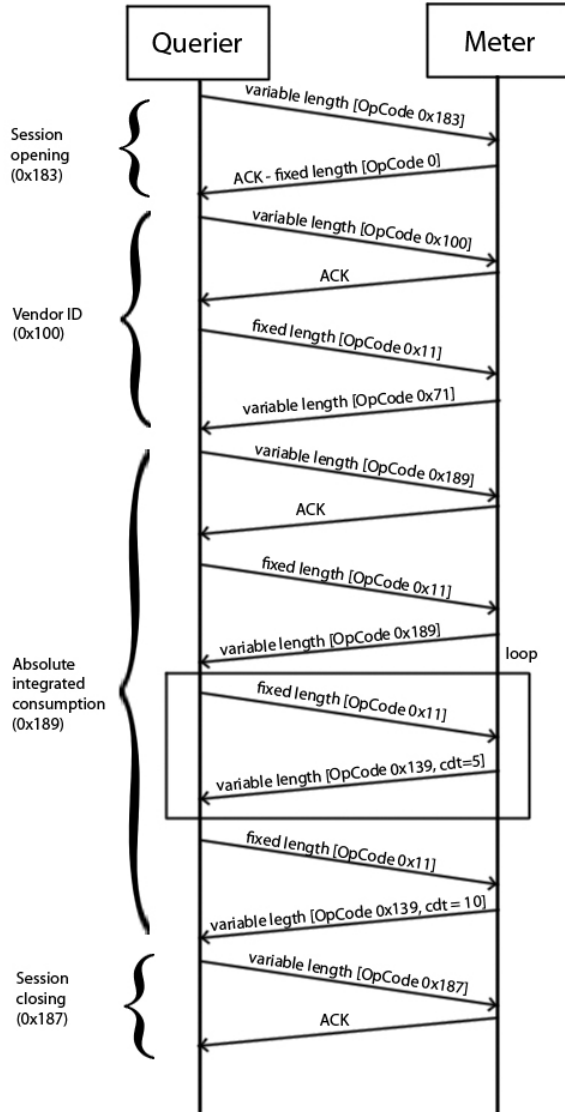


Figure 2.3: Data flow with operations 0x100 and 0x189.

Currently maintained as a UML model, the CIM defines a common vocabulary and basic ontology for aspects of the electric power industry. The CIM models the electric network itself by describing the basic components used to transport electricity. The standard that defines the core packages of the CIM is IEC 61970-301, with a focus on the needs of electricity transmission, where related applications include energy management system, Supervisory Control and Data Acquisition (SCADA), planning and optimisation. The IEC 61970-501 and 61970-452 standards define an XML format for network model exchanges using Resource Description Framework (RDF). The IEC 61968 series of standards extend the CIM to meet the needs of electrical distribution, where related applications include distribution management system, outage management system, planning, metering, work management, geographic information system, asset management, customer information systems and enterprise resource planning.

## 2.4 IEC60870 integration into CIM

The first step to achieve interoperability between CIM and, in this case, the IEC 60870 is to carefully analyse the services offered. Some protocols include an advanced data model (e.g. DLMS/COSEM (Taylor and Kmethy, 2010)) that require extending some of the CIM classes to reflect such structure, but this is not the case of the IEC 60870: a direct mapping between services suffices. The strategy is simple and consists on taking one by one each of the IEC 60870 services, looking for the equivalent in CIM and performing the conversion at frame level. In this way, every frame arriving from an IEC 60870 meter will be transformed into a CIM message by a wrapper (the different architectures and modalities we address will be discussed later in Section 2.4.1). For the sake of simplicity and due to space constraints, we will illustrate this process with the most popular service in a meter: querying and retrieving active power consumption values. Mapping the rest of IEC 60870 services into CIM messages follows the same strategy.

According to the CIM specification (IEC, 2009), a meter is handled as an end device and, therefore, it is represented using the CIM `MeterAsset` class, which is a

subclass of `EndDeviceAsset`. CIM also defines the message interface of a special *profile*, namely the `MeterReadings` profile, that contains all of the functions that can be performed on a meter. Using *CIMTool*, a tool that creates profiles from UML models (.xmi extension) and allows to export them into different languages: XML Schema Definition (XSD), Structured Query Language (SQL), Web Ontology Language (OWL) etc., we've completed the profile by adding all the necessary classes. This profile foresees several ways to retrieve load consumption values from a meter: periodic reads, manual, on-request, historical meter data access; we have selected the on-request modality as the equivalent for the IEC 60870 absolute integrated consumption retrieving service (`Op.Code 0x189`). Figure 2.4 shows the mapping of the `Op.Code 0x189` frame into CIM classes. Please note that there are some fields with internal information inherent to the IEC 60870 that do not need to be mapped.

In this way, the wrapper transforms the CIM message into an IEC 60870 frame. The meter, then, answers with the requested data (a series of `Op.Code 0x139` frames), mapped into CIM classes as depicted in Figure 2.5.

As mentioned previously, the information obtained from the `Op.Code 0x139` frames is recast into a `MeterReading` message (the actual format and schema can be found in (IEC, 2009)). This message specifies, among others, the identity of the meter or meters, the reading values (as `IntervalBlocks` for the periodic reads or single `Readings` for the rest), and their respective qualities and time-stamps (as well as other important information specified within the `ReadingType` class, such as interval length of the measure, type, unit, etc.).

### 2.4.1 Integration Architectures

There are different approaches to design the wrapper that translates IEC 60870 frames into CIM messages and vice versa. The first decision to be taken is to select the nature of that wrapper. We have selected a SOA based architecture against proprietary ad-hoc wrappers in order to achieve an open and scalable solution. In this way, taking into account the aforementioned connection variants of the IEC 60870 (direct,

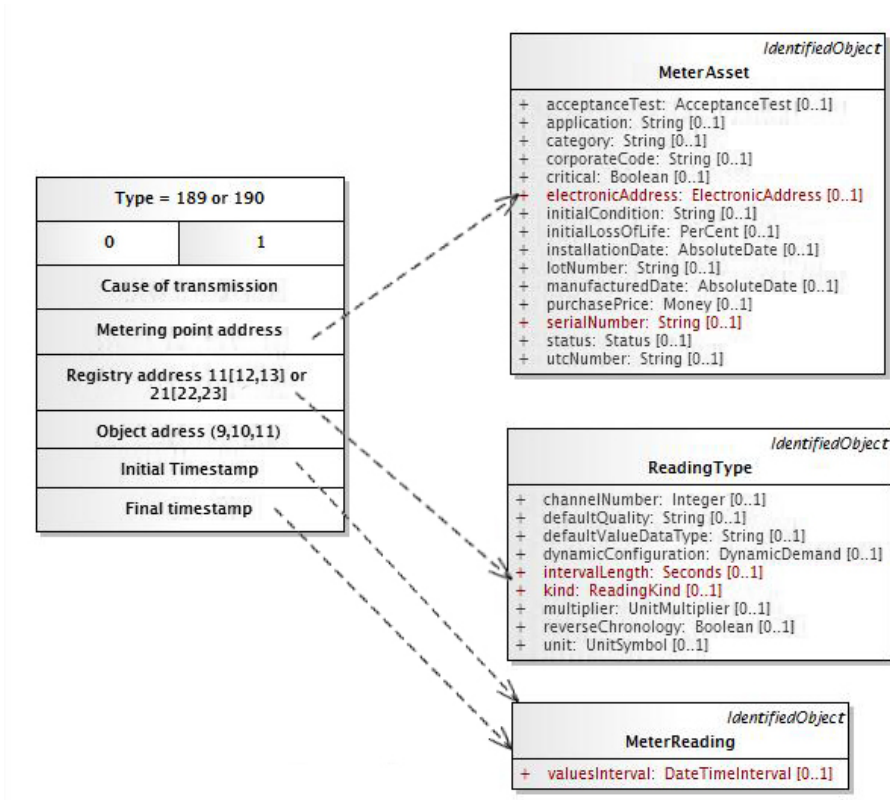


Figure 2.4: Mapping of the Op.Code 0x189 frame into CIM classes.

via GSM, or via TCP/IP), we envision three different approaches for enabling SOA management on IEC 60870 meters, as follows:

- **Embedded SOA wrapper:** It implies connecting directly to the meter and managing all of its inputs and outputs, as illustrated in Figure 2.6. The accomplishment of this goal may be difficult due to a possible hazardous environment and the need to invest on an embedded lightweight platform (which will have limited storage capacity as well) that may not be allowed by the electric Utility that owns the meter.

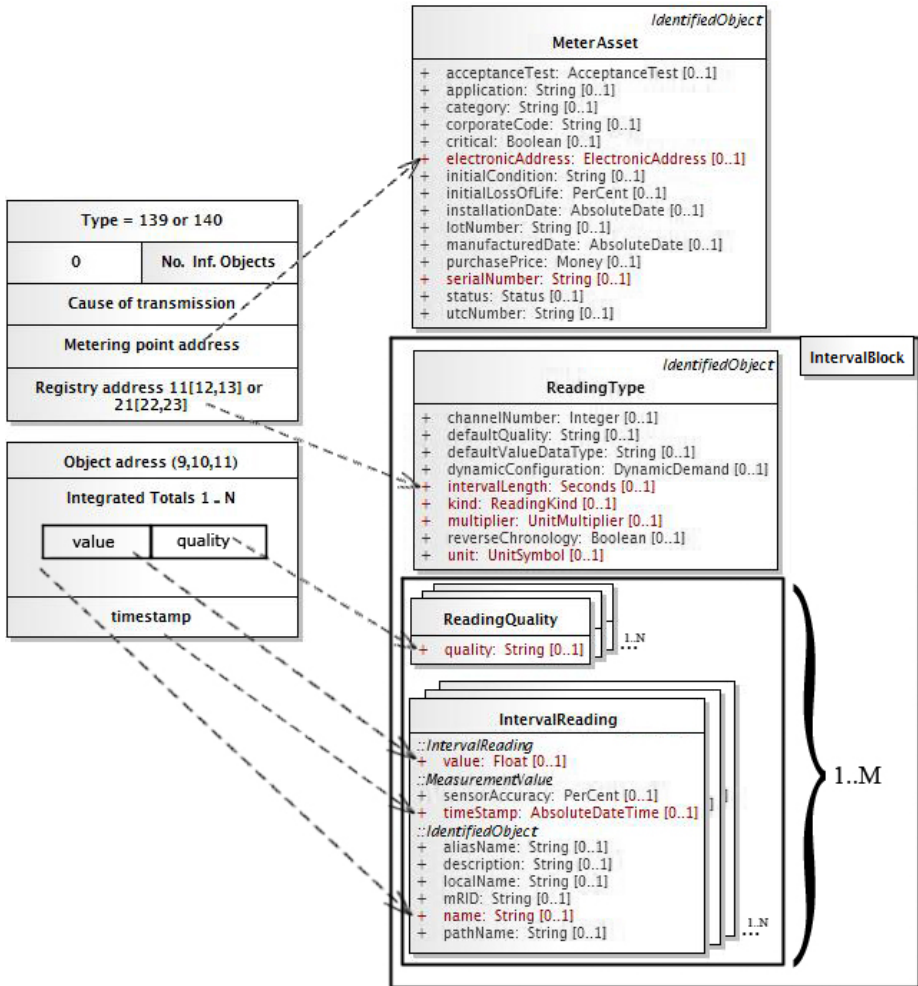
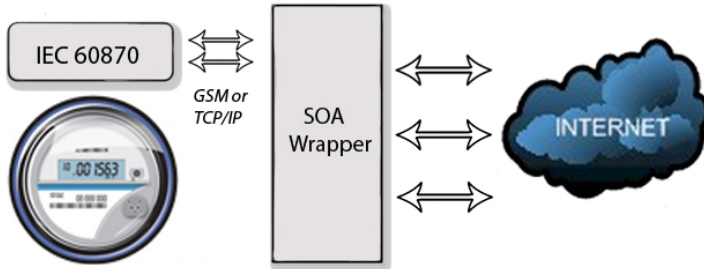


Figure 2.5: Mapping of the Op.Code 0x139 frame into CIM classes.

- **Remote SOA wrapper:** This alternative consists of a remote machine that, via GSM or TCP/IP, queries and stores the meter data in a database (avoiding in this way the problem of storage limit) while simultaneously offering decoupled



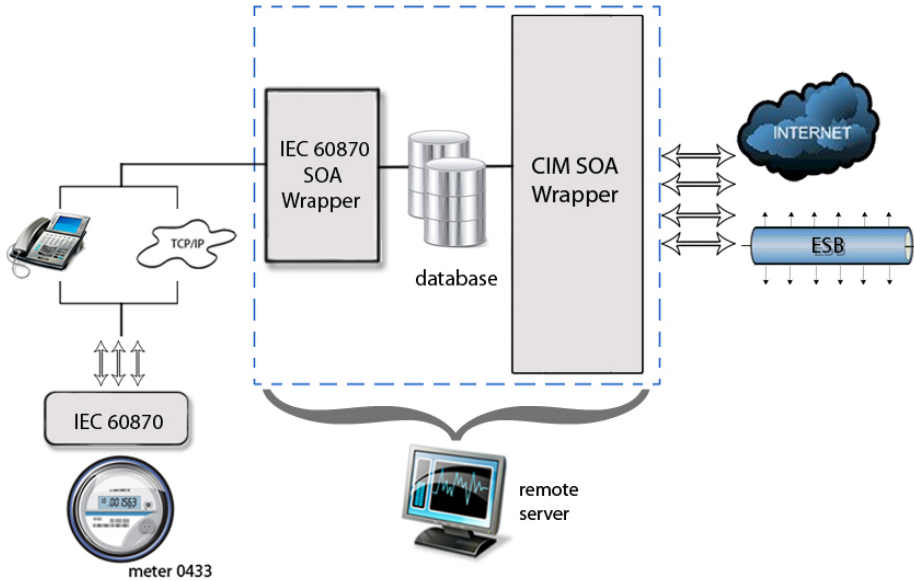
**Figure 2.6:** Embedded SOA architecture.

access to it (e.g. through SOA). The database could even be placed as a *cloud computing* service (Wei and Blake, 2010), if needed. This architecture is illustrated in Figure 2.7. As pointed out in (EPRI, 2008b), the database is not a CIM database, but a database implementing a CIM messages interface.

- **Remote real-time SOA wrapper:** In a similar vein to the previous one, the remote machine acts as a transmitter of the CIM requests, translating and delivering via GSM or TCP/IP to the meter and vice versa for the responses. This strategy still suffers from the limited storage problem (though it could be solved by querying the meter on a regular basis). This architecture is depicted in Figure 2.8.

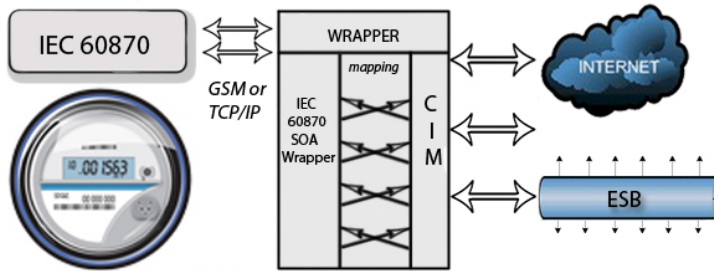
The SOA layer offers the following non-password protected services (please note that the `db` prefix means that this service is only offered in the database-enabled architecture while the `rt` prefix means that it is supported in the real-time one):

- ***db\_read-active*:** Allows the retrieval of the consumed active power of a given period (as long as that period exists in the database).
- ***rt\_read-active*:** Queries the accumulated information (stored every 15 minutes) on consumed active power (usually 3 months of data). Opposite to the previous service, this one does not allow to specify the period (it just retrieves all data of this type).



**Figure 2.7:** Remote SOA architecture.

- **rt\_read-integrated-active** : Queries the accumulated hourly information on consumed active power.
- **db\_read-reactive**: Returns the consumed reactive power of a given period (as long as that period exists in the database).
- **rt\_read-reactive**: Asks for the accumulated information (stored each 15 minutes) on consumed active power.
- **rt\_read-integrated-reactive**: Queries the accumulated hourly information on consumed reactive power.
- **read-vendorId**: Reads the vendor Id of the device.
- **read-deviceId**: Retrieves the Id of the meter.
- **read-dateTime**: Queries the current date and time of the device.
- **read-dateChange**: Allows retrieving the days in which the hour is changed (in order to save energy).
- **read-holidays**: Asks for the vacation days.



**Figure 2.8:** Remote real-time SOA architecture.

Similarly, sending the password allows the accomplishment of the following password-protected services (common to both architectures):

- **read-parameters:** Reads the configuration parameters of the meter, including the password.
- **change-parameters:** Writes new values on the configuration parameters of the meter, including the password.
- **change-dateTime:** Allows changing the current date and time of the device.
- **change-dateChange:** Updates the days whenever the hour is changed.
- **change-holidays:** Enables changing the vacation days.
- **read-contractedPower:** Retrieves the data on contracted power.
- **write-contractedPower:** Allows updating the contracted power data to another amount.
- **read-tariffInfo:** Reads the tariff information stored in the meter.
- **write-tariffInfo:** Writes new tariff information.

### 2.4.2 CIM Semantic Querying

Any of the architectures presented in Section 2.4.1 enables the interoperability of CIM and the IEC 60870 protocol. Still, this goal would have been accomplished in an easier way if we had performed the mapping merely using the CIM model expressed

in UML. Instead, we chose to use ontologies since semantic models do present a number of advantages (Peña and Peña, 2011) that allow to go further from mere syntactical interoperability. The Semantic Web paradigm comprises methods and technologies devised to make information readable and understandable to machines. To this end, RDF (Resource Description Framework) expresses knowledge in the form of triples (subject, object, predicate) in XML, RDF Schema defines the structure of the knowledge, OWL classifies the resources based on description logic and SPARQL allows querying RDF data stored persistently in knowledge bases.

Therefore, by performing the mapping with the CIM ontology (and not with the UML model), we can address CIM-compliant meters as RDF databases which information can be retrieved by SPARQL queries. The three architectures presented in Section 2.4.1 are compatible with this approach: all CIM messages must only be embedded or transformed into CIM RDF messages as presented in Code listing 2.1.

As seen, RDF and CIM information is mixed but identified with distinctive *namespaces*: `rdf` and `mr`. Acting this way shows another advantage since the same concept can be naturally used in semantic service-oriented architectures (Uslar et al., 2008): semantic web services differ from the *normal* ones in that they are annotated with semantic information (in our case identified by `mr`) regarding an ontology known beforehand (in our case, the CIM OWL).

SPARQL is the standard query language for RDF data. It is considered "data-oriented" since it only queries the information held in the models; meaning that there is no inference in the query language itself: SPARQL does not do anything other than take the description of what the application wants. The SPARQL query processor searches for sets of triples that match the given triple patterns, binding the variables in the query to the corresponding parts of each triple and returning the pertinent information in the form of a set of bindings or an RDF graph. Code listing 2.3 shows a SPARQL query that retrieves the load consumption from a CIM compliant meter and Code listing 2.1 presents an extract of the response that a CIM compliant meter may generate. In our case, the SPARQL query was received by the remote server (as proof-of-case we implemented the second database SOA architecture), whose database was

```

1 <?xml version="1.0"?>
2 <rdf:RDF
3   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4   xmlns:mr="http://iec.ch/TC57/2011/MeterReading#"
5   xmlns:xsd="http://www.w3.org/2001/XMLSchema#">
6 <mr:MeterReading>
7   <mr:MeterReading.IntervalBlocks>
8     <mr:IntervalBlock
9       rdf:about="http://iec.ch/TC57/2011/MeterReading#IntervalBlock16">
10    <mr:IntervalBlock.IntervalReadings>
11      <mr:IntervalReading
12        rdf:about="http://iec.ch/TC57/2011/MeterReading#IntevalReading16">
13        <mr:IntervalReading.ReadingQualities>
14          <rdf:Description
15            rdf:about="http://iec.ch/TC57/2011/MeterReading#ReadingQualityValid16">
16            <mr:ReadingQuality.quality
17              rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
18              >Valid</mr:ReadingQuality.quality>
19            </rdf:Description>
20          </mr:IntervalReading.ReadingQualities>
21          <mr:IntervalReading.value rdf:datatype="http://www.w3.org/2001/XMLSchema#float"
22            >14</mr:IntervalReading.value>
23        </mr:IntervalReading>
24      </mr:IntervalBlock.IntervalReadings>
25    </mr:IntervalBlock>
26  </mr:MeterReading.IntervalBlocks>
27 </rdf>

```

**Listing 2.1:** CIM RDF message (beginning).

```
1 PREFIX mr: <http://iec.ch/TC57/2011/MeterReading#>
2 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4
5 select ?beginDate ?endDate ?rdnKind ?val ?rdnUnit ?rdnMulti ?quality
6 where {
7   ?metRdns a mr:MeterReading;
8   mr:MeterReading.IntervalBlocks ?intBlock;
9   mr:MeterReading.MeterAsset ?metAsset;
10  mr:MeterReading.valuesInterval ?valInt.
11  ?valInt a mr:DateTimeInterval;
12
13  mr:DateTimeInteval.beginDate ?beginDate;
14  mr:DateTimeInterval.endDate ?endDate.
15
16  ?metAsset mr:Asset.serialNumber ?serialNumb.
17  ?intBlock a mr:IntervalBlock;
18  mr:IntervalBlock.IntervalReadings ?intRead;
19  mr:IntervalBlock.ReadingType ?rdnType.
20  ?rdntype mr:ReadingType.kind ?rdnKind;
21  mr:ReadingType.multiplier ?rdnMulti;
22  mr:ReadingType.unit ?rdnUnit.
23
24  ?intReads rdf:type mr:IntervalReading;
25  mr:IntervalReading.value ?val;
26  mr:IntervalReading.ReadingQualities ?rq.
27  ?rq mr:ReadingQuality.quality ?quality.
28
29  FILTER regex(?serialNumb, "M200050").
30  FILTER (?beginDate >= "2010-11-30T00:00:00"^^xsd:dateTime && ?endDate <=
31  "2010-11-30T04:00:00"^^xsd:dateTime ).
32  }
33 }
```

**Listing 2.2:** SPARQL query.

continuously receiving and recording IEC 60870 meter data. The answer was then mapped into a CIM SPARQL response and sent back to the querier.

## 2.5 Conclusions

This chapters illustrates the possibility of achieving semantic interoperability with the CIM when the source communication protocol does not define its own data model. In order to achieve this, we have proposed an specific wrapper relying on a direct protocol service mapping that discovers the protocol operations through a

```

1 <?xml version="1.0"?>
2 <?xml-stylesheet type="text/xsl" href="/xml-to-html.xsl"?>
3 <sparql xmlns="http://www.w3.org/2005/sparql-results#">
4   <head>
5     <variable name="beginDate"/>
6     <variable name="endDate"/>
7     <variable name="rdnKind"/>
8     <variable name="val"/>
9     <variable name="rdnUnit"/>
10    <variable name="rdnMulti"/>
11    <variable name="quality"/>
12  </head>
13  <results>
14    <result>
15      <binding name="beginDate">
16        <literal
17  datatype="http://www.w3.org/2001/XMLSchema#dateTime">2010-11-30T00:00:00</litera
18  l>
19        </binding>
20        <binding name="endDate">
21          <literal
22  datatype="http://www.w3.org/2001/XMLSchema#dateTime">2010-11-30T02:00:00</litera
23  l>
24          </binding>
25          <binding name="rdnKind">
26            <uri>http://iec.ch/TC57/2011/MeterReading#ReadingKind.power</uri>
27          </binding>
28          <binding name="val">
29            <literal datatype="http://www.w3.org/2001/XMLSchema#float">125</literal>
30          </binding>
31          <binding name="rdnUnit">
32            <uri>http://iec.ch/TC57/2011/MeterReading#UnitSymbol.W</uri>
33          </binding>
34          <binding name="rdnMulti">
35            <uri>http://iec.ch/TC57/2011/MeterReading#UnitMultiplier.k</uri>
36          </binding>
37          <binding name="quality">
38            <literal
39  datatype="http://www.w3.org/2001/XMLSchema#string">Valid</literal>
40            </binding>
41          </result>
42        </results>
43      </sparql>

```

**Listing 2.3:** Results of the SPARQL query in XML (beginning).

SOA. We have put forward two complementary architectures, one using a database for meter information storing and decoupled SOA access, and another with real-time direct meter management. Moreover, we have shown how to further exploit the opportunities that the semantic web paradigm offers by enabling SPARQL queries of CIM-compliant sources and, as a proof-of-case, we have implemented this approach with data from IEC 60870 meters.

Further works will concentrate on the problem of *query federation* problem (Peña and Peña, 2011) (how to answer to queries like: what is the current consumption of *all* the meters placed within neighbourhood). We will also investigate the way to apply C-Query ((Barbieri et al., 2010) (Continuous Query, a real time SPARQL variant to query real-time RDF stores) to our model in order to define a real-time management framework of CIM-compliant meters.

*A smart man makes a mistake, learns from it, and never makes that mistake again. But a wise man finds a smart man and learns from him how to avoid the mistake altogether.*

Roy H. Williams

CHAPTER

# 3

## GeoWorldSim: Steps towards a Smart City Simulator

**T**HE ADEQUACY OF DSM to a particular scenario is heavily dependent on its particular characteristics. In this line, deployment and maintenance of Smart Homes and Smart Grids in real environments is an expensive and lengthy process. Simulations play an important role by providing means of emulating the behaviour of the aforementioned systems and providing both what-if scenarios that may envision the resiliency of the power system to future changes, and estimations about the effect of applying certain DSM incentives.

These simulations, however, may suffer from lack of accuracy due to the inability to properly reproduce the operation of complex technologies such as solar panels, Heating, Ventilating and Air Conditioning Systems (HVAC), sewerage networks or

water provisioning. Within this context, this chapter presents the first steps towards the deployment of GeoWorldSim: a Smart City Simulator that is able to carry out more representational simulations by merging agent-based simulation of human behaviour with real-world modelling capabilities such as those provided by the Simulink software.

### 3.1 Introduction

The fast pace at which technology advances has endowed everyday objects with new characteristics that broaden their purpose. This purpose goes far beyond than that of simple inanimate entities which respond to stimuli. These objects are now able to “sense”, “think” and “act” together as a greater intelligent “being”. Smart Homes are a clear example of this paradigm. Being monitored by ambient intelligence, Smart Homes offer a better quality of life by introducing assistive services and a centralized control of automated appliances (Alam et al., 2012). User comfort, safety, security, and even energy conservation are optimized through the use of context-aware services which are able to sense the physical environment and adapt their behaviour accordingly (Robles and Kim, 2010).

Smart Home technologies have achieved significant improvements in health care efforts, such as assessing the wellness condition of the residents (Virone et al., 2002), generating alarms if any abnormal vital sign is detected, or even monitoring physiological data and recognizing sleep stages (Andoh et al., 2004). Of all the focus groups, elderly people are highly benefited from the inclusion of home automation in everyday living since this kind of deployment is able to recognize falls, immobility and reaction incapacity among patients with either physiological or psychological problems (Helal et al., 2003).

Another clear advantage of home automation is its unprecedented potential for saving energy, and consequently, saving costs. For example, the cost related to electric bills can be minimized when lights are automatically turned off in empty rooms, which in turn can be cooled or heated based on their level of occupancy (Casado-Mansilla et al., 2012)(Erickson and Cerpa, 2010). At the most basic level, home

automation allows to schedule the programmability of the devices so that the energy usage corresponds to the occupants' daily schedule. However, the introduction of more flexible automation systems may allow to further customize this scheduling to fit and predict the user's needs, extending its applicability to other areas such as water provisioning, waste management and heating control systems.

All in all, Smart Home technologies are based on what is known as the Internet of Things (IoT). IoT refers to a group of interconnected objects identifiable through a digital network that can sense and be controlled remotely, resulting in improved efficiency, accuracy, and economic benefits for the final user (Xia et al., 2012). The versatility of IoT ranges from Home Automation on a simple, isolated level, to its use in Smart Grids for the adequate management and balancing of resources for several buildings connected to a common grid.

The design of a system that can integrate such a large range of applications and technologies while meeting and balancing all the necessary requirements is a complex task. The technological, social and economic impacts of its introduction need to be carefully analysed as well as the added value and benefits it may provide. Within this context, simulations play an important role by providing a testing environment in which to model each subject separately, customizing both their individual operation and interactions with each other in order to evaluate the trade-off among the different components in a prior phase to that of a real scale implementation. Though the advantages and benefits of simulation environments are quite clear, there is still a high level of uncertainty. Due to the difficulty to model some real world processes, simulation engines tend to make artificial assumptions that may not closely resemble how a certain subject operates, whether it is a solar panel, HVAC, a water pump, a sewage network, etc.

### 3.1.1 IoT in Home Automation

IoT is able to bring together both the physical and the information world through the use of sensors, which monitor the environment, collect data and generate responses according to its dynamics. Within IoT, the concept of context awareness plays an

important role. Context awareness is the concept of leveraging information about the end user in order to improve the quality of the interaction (Perera et al., 2014). Having information about the situation under which they operate, devices can react accordingly and are even able to make assumptions and predictions about the needs of the user.

There is a wide range of Home Automation Systems available, which consist on a suite of products designed to work together through Wi-Fi or power line communications such as thermostats, security systems, blinds, lighting and door locks. Popular home automation protocols include X10, Z-Wave, Insteon, Wi-Fi, Bluetooth and Zigbee (Gomez and Paradells, 2010). These devices monitor, gather and send data about the environment to a cloud-based service that takes decisions according to manually programmed rules or inferences through the analysis of the occupants' behavioural patterns.

This whole collection of captured data gives rise to new Smart Digital Services that go beyond the services that an isolated embedded system can provide. In this new paradigm, Smart Homes can control basic environmental parameters such as light, temperature, and heating according to the choices and habits of the occupant (Das et al., 2002). Researchers can also add extended operations such as selection of TV programs, display of cooking recipes or even forgotten property-check services (Yamazaki, 2006). Other popular services include remote monitoring of domestic appliances (Khiyal et al., 2009) and load balancing for energy conservation (Zoha et al., 2012).

### **3.1.2 IoT in Smart Grids**

On a broader scale, IoT can be applied to transportation networks or Smart Grids, enabling a whole new myriad of services including, among others, energy load balancing, intelligent water provisioning, maintenance of sewerage networks, avoidance of traffic congestions, and waste management. Focusing on energy provisioning, Smart Grids are expected to spread the intelligence of the energy distribution system from a centralized core to many peripheral nodes in order to facilitate an accurate

monitoring of energy losses, precise control of the whole energy distribution network, load balancing and better blackout management. Both the abstract concept and the final purposes of these intelligent nodes are clearly comparable to those of the IoT. In this area, community energy solutions such as targeted energy efficiency, district energy, microgrids, local energy generation, and energy storage represent an important opportunity to fundamentally change the way the energy system operates. The quest for sustainable energy models is the main factor driving research on Smart Grid technology. Smart Grids represent the bridging paradigm to enable highly efficient energy production, transport, and consumption along the whole chain, from the source to the user.

However, the testing of IoT solutions within the scope of Smart Grids is quite a challenge due to the expensive nature concerning the construction of pilot sites or the developing of scale models. In addition, Smart Grid simulation platforms are scarce. Works like (Karnouskos and d. Holanda, 2009), (Oliveira et al., 2012) and (Pipattanasomporn et al., 2009) focus on the simulation of the usage of household appliances relying primarily on the load balancing of energy consumption for a certain grid layout under controlled conditions. As mentioned before, one of the issues with simulations is the inability or difficulty to replicate the exact operation of some of the components because the modelling language or software may have limitations in terms of capabilities to model the physical, communication or even control logic aspects.

## **3.2 Related Work**

Smart Homes and Smart Grids can be analysed from two different points of view depending on the nature of the context on which they are built:

### **3.2.1 Real environments**

The deployment of Smart Homes in real environments involves the physical and technological equipment of a building for precise monitoring. This process relates to

the analysis, acquisition, and installation of a set of devices for Home Automation, whether they are already on the market or they respond to a particular implementation for research purposes. In this case, there are several important issues that must be taken into account, such as the lengthy process of configuration and maintenance of the devices as well as the amount of time needed to acquire a sufficient amount of data for the analysis.

The MavHome project (Cook et al., 2003) is a clear example of Smart Home implementation in real environments, aimed at creating a sandbox that perceives the state of the home through sensors and acts upon it through device controllers. Its architecture is divided into four layers: a) physical (60 X10 hardware devices plugged into the home electric wiring system), b) communication (exchange of information among devices) c) information (generation of knowledge for decision making) d) decision (execution of actions based on the information supplied). Decision making is based on a finite-order Markov model which predicts inhabitant future actions in order to automate and improve repetitive activities to meet the house goals in terms of comfort and costs. Other projects like GatorTech (Helal et al., 2005) are heavily focused on elder care and proactive health. In this case, the authors propose a middleware architecture that comprises separate physical, sensor-platform, service, knowledge, context-management, and application layers. The central system collects the data provided by the sensors and gives indications to whether there is any anomaly in the resident's behaviour or vital signs.

In the case of Smart Grids, the deployment comprises the leverage of several Smart Homes connected to a common grid. Data from several sources is aggregated and analysed from a centralized point of view with the purpose of balancing resources in several areas, such as energy load, water usage and waste management.

Projects like Nemo&Coded (NEtworked MOonitoring & COntrol, Diagnostic for Electrical Distribution) (Penya et al., 2012) focus on the modelling, design, implementation and operation of networked hardware/software smart devices for the low voltage electrical distribution domain by building dynamic energy efficiency services (Penya et al., 2013). The infrastructure consists of an acquisition platform for collecting energy data in real time by guaranteeing the interoperability of their

hardware/software devices, independently of their equipment or communication technologies. The main novelty lies on the use of intelligent and autonomous distributed nodes to process semantic information known as PGDIN's (Espinoza et al., 2013), which integrate several of the large number of existing standards in the electricity sector, enabling the process of events and data in an autonomous way.

### 3.2.2 **Simulation environments**

The high economic investment and time restrictions related to the deployment of Smart Homes and Smart Grids in real or scaled environments highlights the advantages of simulation models. In order to support the implementation in the real Smart Home, it is necessary to demonstrate that things can be achieved in a simulator which deals with virtual appliances and devices that model a real Smart Home environment.

The Interactive Smart Home Simulator (ISS) (Van Nguyen et al., 2009) proposes a context aware simulation system comprised of several electronic devices distributed on an apartment. The simulator is comprised of a context retriever which requests and receives sensor information from the home appliances. A centralized server is the central control in charge of taking decisions according to changes in the environment and the behaviour of the house occupants. This simulator demonstrates the exchange and update of information as perceived by the Smart Home on changes in the environment

The Intelligent Project Home (IHome) (Lesser et al., 1999) uses multi-agent system technology to the problem of managing an intelligent Smart Home environment. The simulation is populated with distributed intelligent home-control agents that control appliances and negotiate over shared resources with the objective to automate the tasks made by humans. Each household appliance is represented by an agent and the model coordinates over electricity, hot water, noise or sound levels, and room temperature in each of the modelled rooms. The agents make decisions about the activity that the occupant does depending on the availability of the resources, whether there is not hot water, the temperature outside, etc.

Within Smart Grids, projects like MASGrip (Oliveira et al., 2012) propose a multi-agent system that models the internal operation of smart grids. This system considers all the typically involved players, each player being represented by one agent with the capability of representing the actual corresponding player and to simulate its actions. The agents involve not only the devices within the Smart Homes, but also an agent that simulates negotiations within the Electricity Market (MASCEM). In order to efficiently analyse the quality of the simulated smart grid management, it is essential to provide the means for the MASGrip system to be able to perform the required negotiations in the electricity market, information not always available.

Karnouskos et al. (Karnouskos and De Holanda, 2009) propose an agent simulation system that simulates discrete heterogeneous devices that consume and/or produce energy, that are able to act autonomously and collaborate. The bottom layer is the simulator layer and contains all the agents which represent energy generating and/or consuming devices. The agents utilize the full communication capabilities offered by the agent platform and execute in order to offer the desired functionality. The simulator focuses heavily on maximizing user comfort.

All in all, Smart Homes and Smart Grids constitute an interdisciplinary domain considering the variety of technologies that come together to reach a common objective. Within this context, the main objective of this chapter is to design and test a simulation platform that is able to replicate the behaviour of the residents of a small neighbourhood on their own houses. Each house is equipped with several smart appliances simulating energy, water, and gas consumption, and resource providers such as a solar panel, a hybrid heating system, and a water deposit. The platform will evaluate the feasibility of integrating these two modelling approaches in order to bridge the gap between the configuration of a simulated and a real Smart Home. This simulation platform will improve the state of the art in two aspects: on the one hand, it will serve as a Computer Aided Design solution for Smart Homes or Smart Grids, while on the other hand, it can be used as a benchmark system for the testing of new control algorithms for the Smart Home or Smart Grids services (from energy boxes to traffic light control and virtual power plants).

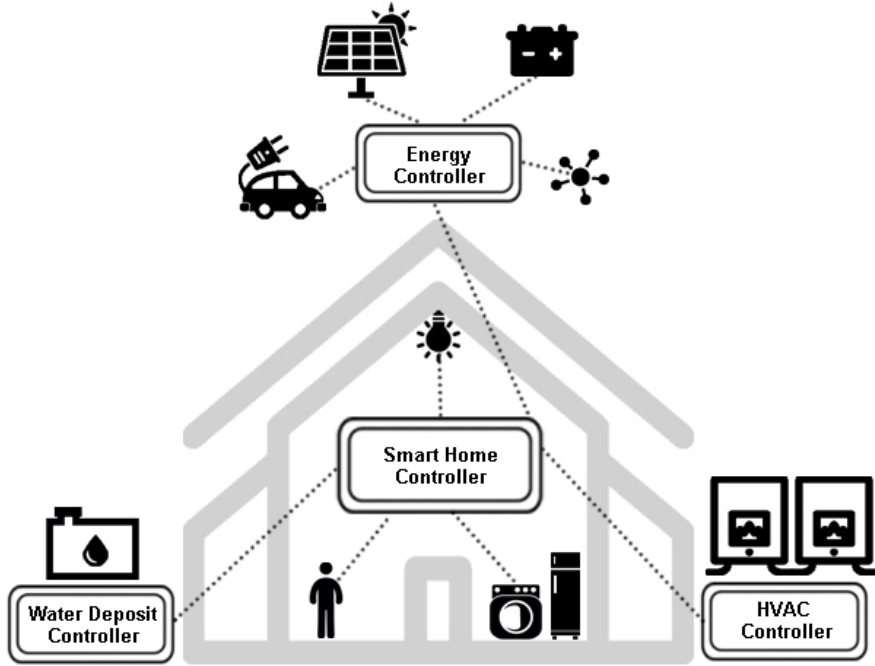
### 3.3 Components of GeoWorldSim

GeoWorldSim is a Smart City Agent Based Simulator developed for analysing how demand-response strategies can influence human behaviour in order to achieve an optimal operation of the electrical grid. To this end, the main interest lies in the proper characterisation of the energy consumption of appliances within homes, specifically those which usage can be timetabled along the day, such as washing machines or dishwashers. However, the great majority of the publicly available energy consumption datasets provide hourly data about the aggregated load per energy meter. Information detailing the specifics of the interaction between the house occupants and the house appliances is not always available. Even so, the datasets that detail appliances usage only comply with the monitoring of a certain amount of buildings over a limited number of days. In order to replicate these datasets for a whole city, the *Block Bootstrap* methodology (Kunsch, 1989) could be used by randomly resampling data from the original datasets. However, this data would be heavily dependant on the season in which the appliances where monitored, failing to characterise seasonal loads, i.e. loads that take place in winter against loads that take place in summer.

#### 3.3.1 Ideal Smart Home Architecture

Every Smart Home project focuses on the creation of an environment that behaves as an intelligent agent which perceives the state of the home through sensors and acts upon the environment through actuators with the objective of maximising the comfort and productivity of its occupants. In order to achieve these goals, the house must predict, reason and adapt to its inhabitants. We propose to go a step further and build a system that integrates all complex systems within a home, such as HVAC, energy and water provisioning, etc., and actuates upon them so as to reach an optimum level of efficiency and comfort. This proof of concept can then be later extrapolated to the balancing and leverage of resources within Smart Grids.

The architecture envisioned according to the components described in Section 3.4 is detailed in Figure 3.1.



**Figure 3.1:** Ideal Smart Home Architecture.

- **Energy Controller:** This controller is in charge of attending the energy demand within a Smart Home considering all the energy sources connected to it. In our case, we consider four energy sources: the actual electric distribution network, the existence of solar panels, solar batteries and the introduction of the electric car. When energy is demanded the controller balances the petition, and determines the source of the electricity based on issues such as the load, the hourly energy price, and the saturation of the distribution network. If needed, the controller may also feed the distribution network directly.
- **HVAC Controller:** This controller governs the behaviour of an hybrid boiler system so as to meet the demands in terms of heating and hot water while main-

taining a constant temperature. The purpose of the controller is to optimise the system performance while meeting the objectives of energy efficiency,

- **Water Controller:** This controller act upon a deposit that feeds water to the house. In case the water reaches a certain threshold, the controller may activate a pump to maintain the level at a certain setpoint.
- **Smart Home Controller:** This controller is the central axis of the system. It collects data from the sensors attached to the appliances and distributes the information to the other controllers in the house. The system also comprises a module for activity recognition and forecasting as well as an optimisation algorithm that analyses the occupants' behaviour so as to leverage the resources and meet the comfort and efficiency objectives.

### 3.3.2 Human Behaviour model

Humans in GeoWorldSim follow the human activity definition and execution model, developed in Chapter 4, to emulate daily duties and the consequent interaction with appliances. This model is based on environmental multi-agent theories to model and simulate intelligent environments, inhabitants and their interactions.

### 3.3.3 Agent-based system

The need to go further in the modelling of appliances and human behaviour called for the use of a Multi-Agent System (MAS). MAS are computer systems composed by intelligent pieces of software called agents that emulate the behaviour of an entity (Wooldridge and Jennings, 1995). Agents perceive their environment, process perceptions, respond, and act rationally like the entity they represent. All of the agents are executed in parallel, either cooperating or competing against each other to achieve a certain goal. In this way, these systems can create networks of agents working together to address complex issues. The use of MAS involves readjusting the traditional centralised governing paradigm to a bottom-up dynamical model where every entity that lives in the simulation is able to make its own decisions. In this sense, we differentiate between agents and passive entities. In the first case, agents

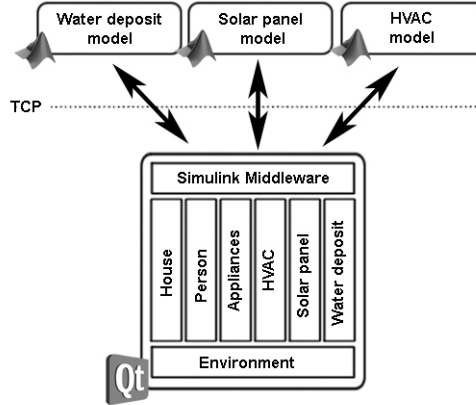
are characterised by its set of skills, methods, and a state defined by the values of their attributes. By contrast, passive entities are those which although having a *logic*, they just react to stimuli initiated by other entities.

Another important concept within MAS technologies is the Agent Environment. The Agent Environment can be considered as an independent and living first-class entity called to play a relevant role in modelling dynamic real world problems (Ricci et al., 2015). From a functionality point of view, the Agent Environment consists of both a physical and a social environment which together form the Intelligent Environment where both objects and people are situated. On the one hand, the physical environment ensures none of the spatial constraints are violated and, on the other hand, the social environment is responsible for managing communications, roles and knowledge by giving agents a social and communicative context.

Following this roadmap, GeoWorldSim defines two types of agents: humans and house appliances. Humans or house occupants are modelled as active agents capable of taking decisions and interacting with house appliances inside a Smart Home. House appliances, on their side, are modelled as passive agents which react to the interaction initiated by the occupants. The Agent Environment in particular, consists of mainly of the skills inherent to each agent and that define how each one of them operates, as well, as the communication context on which they interact. With these capabilities, the home can control many aspects of the environment such as climate, water, lighting, maintenance, and entertainment. Intelligent automation of these activities can reduce the amount of interaction required by inhabitants, reduce energy consumption and other potential wastages, and provide a mechanism for ensuring the health and safety of the environment occupants. Figure 3.2 shows the communication between the simulator and the Simulink models.

### 3.3.4 **Qt implementation**

In search of a highly abstract and scalable solution that suited this new approach, the energy box has been developed from scratch using the C++ programming language. It is built upon Qt Framework, a widely used cross-platform framework for developing



**Figure 3.2:** Smart Home Simulator Architecture.

native C++ applications along with interesting extensions, such as signals and slots for enabling remote event invocation, a meta-object compiler, and the ability to detach pieces of code and move them to separate threads.

The use of Qt has allowed us to create a complete independent and robust simulation engine based on our understanding of what an Agent Environment should be. All the Qt features make it possible to model more accurate agents and passive entities by using an event-driven paradigm. The execution flow of an agent will be heavily influenced by events such as messages from other agents/entities, sensor outputs, or even its own actions, just like in real life, where humans are constantly bombed by stimuli. Parallelisation and queueing management is controlled by splitting the workload into individual threads/event queues and by moving agents from one to the other by working out the most efficient execution flow. The QMutex class of Qt Framework complements parallelisation by providing access serialisation to critical resources and attributes between threads. Sensor messaging is built upon the system of Qt's signals and slots: one of its central features and, probably, the characteristic that differs the most from the functionalities provided by other frameworks. Signals and slots follow a publish/subscribe mechanism where objects emit notifications

that are 'listened' by other objects in the environment. A class that emits a signal neither knows nor cares which slots receive it and may send additional arguments of any type. It is up to the other agents to create a slot for the signals they want to listen to if they want to know when they have been triggered.

Each household appliance is modelled as a passive agent, i.e. an agent that responds to stimuli but does not interact with others on its own. When an agent executes an action that uses any of these appliances through their get and set methods, the passive agents react by activating/deactivating themselves and emitting signals of consumption to Simulink through the TCP/IP protocol.

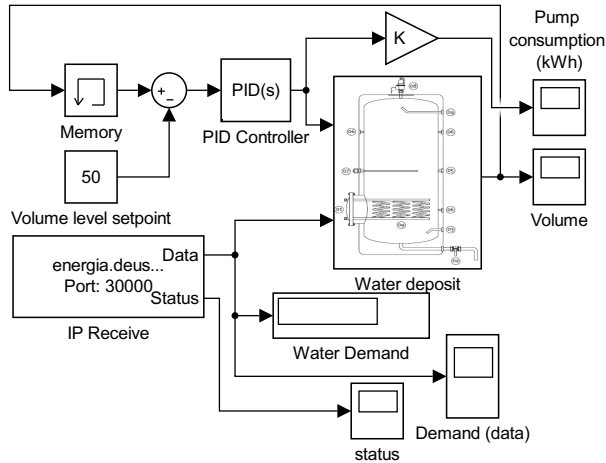
### 3.3.5 Simulink implementation

Simulink is a block diagram environment specifically designed for the implementation of model based simulations. This visual programming environment is integrated in the programming environment of MATLAB. Simulink provides libraries with pre-designed and configurable models of mechanical, electronic and physical elements. Therefore, it provides an ideal environment to develop real simulation models of complex systems such as HVAC systems and solar panels, as well as more common systems like water deposits. The following models simulate the behaviour of those services within the Smart Home simulator architecture.

The water deposits are simple models that simulate how water is consumed according to the Smart Home demand and how the water deposit is filled through a water pump. Figure 3.3 shows the block diagram implemented in Simulink to simulate the behaviour of the common water deposit for different water consumption profiles.

Figure 3.4 shows the simulated HVAC system, which consists on two gas boilers connected in parallel. Each boiler is defined by its efficiency curve and maximum power (350 kW). The incoming demand is distributed between both boilers, being one of them the main boiler. The following commercial heating systems have been studied in order to build a more realistic simulation model:

- **Biomass boilers:** Pellematic (10-56 kW), Viessman pyrotic (390-1250 kW)



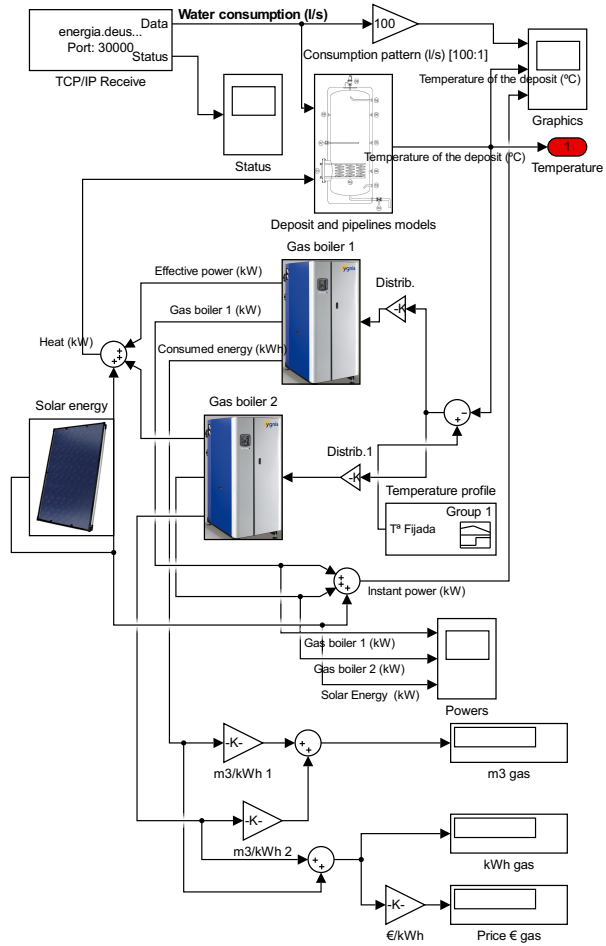
**Figure 3.3:** Diagram of the water deposit management system.

- **Gas boilers:** Ygnis Varmax (127-477 kW), Beretta Power Plus (50-100 kW), Viessman Vitodens (1.9-35 kW)
- **Gasoil:** Junkers Suprapur-O (22-35 kW)
- **Heat pumps:** Carrier 30RQS (39-160 kW)

The last modelled installations were been the solar panels. The objective is to study their effect on the performance of the global system, specifically with the introduction of the electric car. The generated energy is simulated with solar irradiance data and is stored in a sizeable battery module depending on the consumption profiles.

Each simulated installation is controlled by PID controllers in order to maintain the established setpoints. These controllers have been configured to correct the possible deviations in the minimum response times and without excessive overshoots. The objective is to maintain the established deposit level and hot water temperature level.

The Simulink models are executed independently. They receive the real time simulated demand data regarding electricity, and hot or cold water from the Smart



**Figure 3.4:** Diagram of the HVAC system.

Home simulator. Then, the Simulink models execute the simulation for the modelled water pump, boiler system, and solar panel, for the specific demand at each point in time. Once the simulation finishes, Simulink sends the result of the simulation back to the Smart Home simulator. The data transmission between both modelling

platforms is achieved through the use of the TCP/IP communication protocol. The main problem relies in the synchronization of all the simulations in order to maintain the same simulation time over all the systems. Each simulation can be run in a different device and each model has a different computational cost, which causes de-synchronization between the models.

### 3.4 Materials and methods

For the purpose of this case study, we selected a small urbanisation composed of 12 single residential homes with a similar inner layout consisting of a kitchen, a bedroom, a bathroom and a living room. Each house is configured in the same way and is equipped with the following technologies:

- **Sensorised household objects:** Every single appliance in the house is monitored through consumption sensors depending on the types of resources it needs to operate: energy, water or gas. For example, the sensor associated to a single light switch monitors the energy consumption, while the sensor associated with the washing machine monitors energy, gas, and water consumption. The distribution and type of resource needed by each sensor is shown in Table 3.1.
- **Solar panel:** The house is equipped with a solar panel that converts the energy of light directly into electricity through the photovoltaic effect. The electricity produced can then be fed into the house main electricity supply or stored in a battery for further use.
- **HVAC system:** Heating and air conditioning is provided by a system composed of two gas boilers (350 kW). The HVAC system is equipped with an independent PID control system that continuously maintains the temperature of the water deposit.
- **Water deposit:** A single deposit feeds water to the house. The deposit is equipped with a pump that keeps the amount of water at a certain setpoint.

**Table 3.1:** Distribution of sensors in the Smart Home.

Place	Sensors	Energy	Gas	Water
Bathroom	1 toilet			X
	1 shower faucet		X	X
	1 sink faucet		X	X
	1 lightswitch			
	1 door			
Bedroom	1 closet			
	1 cabinet			
	1 door			
	1 lightswitch	X		
Hall	1 door			
	1 lightswitch	X		
Kitchen	3 cabinets			
	3 drawers			
	1 washing machine	X	X	X
	1 burner		X	
	1 dishwasher	X	X	X
	1 sink faucet		X	X
	1 oven	X		
	1 microwave	X		
	1 freezer	X		
	1 door			
1 lightswitch	X			
Living room	1 TV	X		
	1 radio	X		
	1 lamp	X		

The objective of the simulator is to be able to not only replicate the behaviour of the Smart Homes in terms of activity recognition and automation, but to go further and provide an intelligent system that based on the resources available, whether it is an HVAC, a water deposit or a solar panel, is able to manage itself according to the occupants behaviour. We intend to show that the use of real simulation models concerning these elements will help improve the accuracy of the simulation.

### 3.5 Results and discussion

The results obtained by using the proposed simulation architecture are quite promising. The agent-based Smart Home Simulator is capable of reproducing a person's behaviour for different simulation times and during different activity peaks of the day.

This module generates the consumption profiles according to the human activities simulation and sends the information to the parallel executed Simulink modules which are able to simulated common services of the building.

Figure 3.5 shows the frequency at which the consumption data is received by Simulink models. As can be seen, the data reception is very stable and regular. The transmission has been set to be received every 60 seconds of the simulation time. However, each model runs at a different simulation speed due to their computational cost. A simulation of the consumption generated for 1 h, takes 184.5 s for the water deposit model and approximately 2070 s to simulate the HVAC system. Therefore, the data transmission must be synchronised between the different simulation modules to maintain the same time line between all of them.

Currently, the data transmission is synchronised based on the estimated computational cost for each Simulink model. However, in order to maintain the same simulation time across the models, in future works, the complete simulation will be carried out by time segments. Each model will simulate the response to the generated demand for a limited time segment and waits until all the simulations have finished. During these simulation pauses, the Simulink models send back their results to the Smart Home simulator. Finally, once all the simulation of services have finished, the Smart Home simulator sends a new batch of consumption data for the next time segment. Thereby, the simulation time will be maintained across all the models with acceptable deviations, despite being executed in different platforms. An specific Simulink block will be developed to manage the TCP/IP data transmission and to synchronise the simulation of each model.

All the architecture has been tested for different simulation times. Figures 3.6 and 3.7, show the simulation of the HVAC system for 1 hour of consumption data. In Figure 3.6, the yellow line shows the consumption profile received from the Smart Home Agent, the red one is the total instant power needed by the installed heating system and the blue line is the temperature of the water deposit which is set to maintain a constant temperature of 75 °C during the day.

Figure 3.7, shows the power consumption of the HVAC system disaggregated for the two installed boilers. As previously stated, the heat power demand is divided

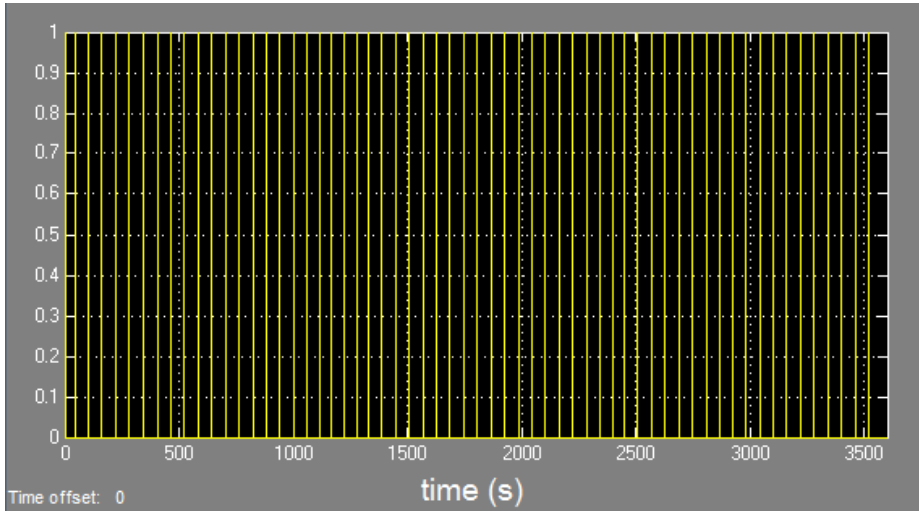


Figure 3.5: Frequency of data reception

between the boilers. Currently the distribution between both is fixed, but in a future version of the model the demand will be intelligently distributed based on the efficiency, used fuel and responsiveness of each boiler.

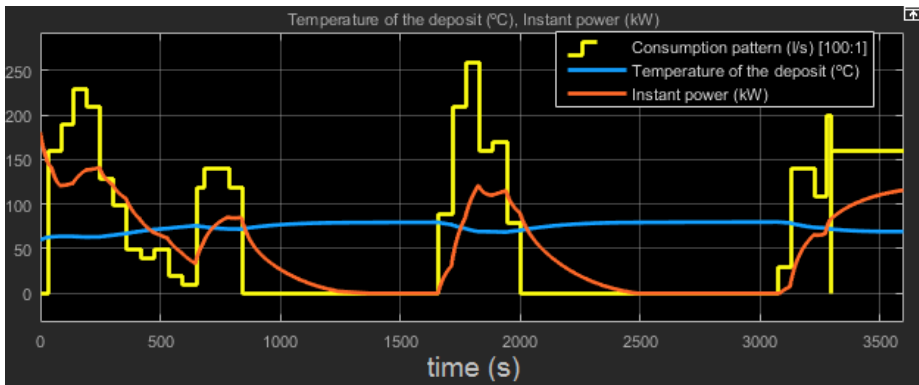
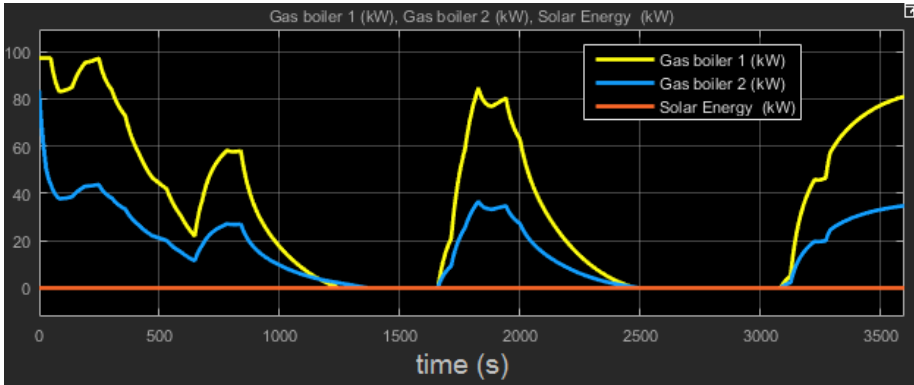


Figure 3.6: HVAC system simulation: Temperature of the deposit and the instant accumulated power.



**Figure 3.7:** Gas boiler power profile for a 1 hour simulated consumption profile.

### 3.6 Conclusions

The main contribution of the research presented in this chapter is the feasibility to integrate an agent-based Smart City simulator with the modelling capabilities offered by platforms like Matlab/Simulink, which are specifically designed to solve engineering and scientific problems. The advantages of this kind of approach rely in the ability to accurately model dynamic complex systems such as HVAC, solar panels or microgrids, bridging the gap between simulation and real implementation along with the benefits in terms of cost, maintenance and deployment. In this sense, the future developments considered as a follow up of the work presented in this chapter are detailed next:

- **Improve the HVAC model and solar panel models.** The objective is to enlarge the energy agent controller so as to accurately design the way the system works, even introducing the electric vehicle. Once available the extended models, the simulations of the global system will support the decision making when dimensioning new buildings or to improve the efficiency of the existing installations.

- **Improve the synchronisation between the simulated services.** An specific Simulink block will be developed to manage the TCP/IP data transmission and to synchronize the simulation of each model in order to maintain the same simulation time across all the models and platforms.
- **Build a microgrid with several sources of generation.** The next step in the simulation will focus on integrating the Smart City within a Smart Grid comprised of several sources of generation and services such as waste collection and management, sewerage networks, etc

*Every human being is intended to have a character of his own; to be what no others are, and to do what no other can do.*

William Ellery Channing

CHAPTER

# 4

## Agent-based Human Activity Simulation

**H**UMAN BEHAVIOUR plays a pivotal role in the analysis of viable DSM strategies. Therefore, a proper characterisation of how humans act is important so that the scenario studied resembles the real one as close as possible. DSM strategies aim at reducing the amount of energy supply that a utility must have available in order to meet the peak energy demand, thereby improving the operational efficiency of the energy grid. Research on this area demands the execution of complex experiments involving humans interacting with intelligent environments in order to generate meaningful datasets, both for development and validation. Running such experiments is generally expensive and troublesome, slowing down the research process.

This chapter presents MASSHA: the engine behind the simulation of human behaviour for the Smart City Simumatot presented in Chapter 3. Specifically, MASSHA models the behaviour of the occupants of a sensorised environment from a single-

user and multiple-user point of view. The accuracy of MASSHA is tested through a sound validation methodology, providing examples of application with three real human activity datasets and comparing these to the activity datasets produced by the simulator.

## **4.1 Introduction**

Intelligent environments are thought to be a key resource in assisting older adults and people with disabilities through their daily life by following the ambient assisted living paradigm. The use of sensors and computing devices allows intelligent environments to monitor the activity of inhabitants and plan suitable interventions depending on the activities and behaviour of the monitored person.

There are many challenges to face in the area of intelligent environments and human activity recognition, such as the type and layout of the sensors to be used, or the definition of accurate algorithms for activity recognition and behaviour modelling (Kafali et al., 2014), (Cvetković et al., 2016). However, many of the research directed to address these issues finds a common operational problem which, in the best case, slows down the research, hindering the capability of running meaningful experiments in intelligent environments.

This operational problem has already been identified by many researchers (Helal et al., 2012), (Rashidi and Cook, 2011), (Azkune et al., 2015), (Buchmayr et al., 2011). Intelligent environments are expensive to build and maintain. Whenever there is a need to test new sensor layouts, it is generally difficult to reconfigure the environment, thus researchers have to be very careful with the initial design. In addition, the existence of a proper intelligent environment does not guarantee the generation of needed datasets, since recruiting humans for experiments is a troublesome process. In consequence, many of the datasets generated in intelligent environments do not allow to test and verify complex theories concerning human behaviour (Helal et al., 2012).

The solution to this problem is not trivial. In this line, many researchers have proposed the development of several simulation approaches for the definition of

intelligent environments. Simulated intelligent environments offer many positive features, e.g. the generation of large datasets, total control of the environment and sensor layouts, cost-effective experiments and ability to define very specific experiments. Synnott et al. (Synnott et al., 2015) divide simulators into two main groups: model-based simulators and interactive simulators. We present a model-based simulator to mitigate the problems of experimenting in intelligent environments. Specifically:

- Section 4.3 presents the MASSHA simulator. MASSHA is an agent-based simulator for human activities in sensorized spaces that, in contrast with previous approaches, is based on environmental multi-agent theories to model and simulate intelligent environments, inhabitants and their interactions. Multi-agent environments allow a bottom-up modelling, where all agents collaborate and compete against each other while interacting with the environment. In this way, a model is defined and run through a simulation, whose results vary depending on both the characteristics of the agents and their decision processes. The advantage of this approach is the feasibility to parameterize different scenarios and evaluate how the agents behaviour and interactions are affected each time. MASSHA's scope goes far beyond the replication of human activities and its functionality can be extended to other scenarios, allowing a myriad of researchers to benefit from it, for cases such as the evaluation of electrical distribution systems, adequate sizing of water and sewerage networks, accessibility studies or traffic simulations.
- Section 4.6 introduces a rigorous methodology to validate the activity datasets generated by the simulator which also aims to be used by other researchers to assess the validity of their approaches.

MASSHA has been designed mainly for researchers who want to work on human populated intelligent environments. The objective of our simulator is to provide a useful tool to generate all the required realistic data for specific research questions. We test MASSHA for activity recognition data generation. More concretely, we explore two scenarios:

- Single-user scenarios, where only one person is being monitored. Sequential and overlapping activities can be performed. The majority of the research in intelligent environments has traditionally been focused on such scenarios, thus the importance of proving the validity of the tool in these conditions.
- Multi-user scenarios, where multiple persons are monitored in the same intelligent environment. Those scenarios are more challenging. However, the full potential of MASSHA is shown in multi-user scenarios, due to the agent-based approach adopted. To the best of our knowledge, this is the first simulator to address the multi-user scenario for activity datasets.

## 4.2 Related Work

Following the categorisation presented by Synnott et al. (Synnott et al., 2015), simulators for intelligent environments can be classified into two main groups:

- **Model-based approaches:** these approaches use activity models, which can be obtained by diverse ways, to create synthetic datasets that store information about the activation of sensors during the execution of activities.
- **Interactive approaches:** a human user can interact with virtual environments and sensors set up by the simulation; depending on the actions executed by the user, the state of the virtual environment and sensors changes accordingly, storing that information in a synthetic dataset.

For a complete review of both approaches, it is recommended to read the work of Synnott et al. (Synnott et al., 2015). Model-based approaches for synthetic data generation using simulators demand the specification of activity models explaining the generation of events over time, the probability of events occurring, the time taken for each event during the performance of specific activities and the definition of the behaviour of monitored virtual people as a generator of activities over time.

An early example of a simulator called DiaSim was developed by Bruneau et al. (Bruneau et al., 2009). The DiaSim simulator executes pervasive computing applications by creating an emulation layer and developing simulation logic using

a programming framework, where the models are defined. It is more focused on simulating applications such as fire situations, intrusions and so on to identify potential conflicts, hence it cannot be used for human activity recognition. In contrast, MASSHA's architecture makes it possible to model human activity/behaviour and simulate these type of situations for analysis.

Helal et al. (Helal et al., 2011) developed a simulator called Persim, which has been enhanced in the new version Persim-3D (Helal et al., 2012). Persim is an event driven simulator of human activities in intelligent environments. Persim is capable of capturing elements of space, sensors, behaviours (activities), and their inter-relationships. The simulator allowed users to define activities by specifying the sensors involved in each activity, the order of sensor activations, the maximum and minimum typical sensor values and activity duration. Based on these parameters a list of sensor data could be generated in the Sensory Dataset Description Language. In contrast with MASSHA, Persim only offers single-user single-activity scenarios, activities are defined as fixed sequences of actions and dependencies between activities cannot be modelled.

An earlier example of a model-based simulator was provided by Bouchard et al. (Bouchard et al., 2010). They developed a simulator called SIMACT, where the activity models are captured using a form-based interface for users. Those forms allow to specify scripts that detail the series of steps involved in the performance of activities within an environment. SIMACT users could define behaviour related parameters such as the order of events, the time taken for each event and the objects involved in the event. Besides, users could also define actions associated with the completion of each step which modify the state of the environment. Activity scripts could be replayed in real-time or fast-forwarded and could be replayed with adjustments to event timings. While SIMACT only applies an activity-based approach, MASSHA follows the idea of Persim of a sensor-based approach for simulation, which is more realistic and provides richer information.

A different approach was presented by Mendez-Vazquez et al. (Mendez-Vazquez et al., 2009), who demonstrated the use of Markov chains describing the order of events, combined with Poisson distribution to calculate a range of realistic activity

times and probability distributions to calculate a range of sensor values to generate a simulated activity dataset. Their tool considered activities such as reading, sleeping, walking and sitting together, adding interesting metrics including time and energy expenditure. A similar approach was shown by Okeyo et al. (Okeyo et al., 2012), who use a synthetic data generator tool to simulate time intervals between sensor activations. Their research is focused on sensor data stream segmentation, so the tool generates varying patterns of sensor activations in order to verify their approach.

Kormányos and Pataki (Kormanyos and Pataki, 2013) developed a simulator to model the activity of a single inhabitant within an intelligent environment. Individual behaviour profiles allow modelling concepts like typical sleep amount, and the change in current state such as thirst and tiredness, explicitly introducing the state of the monitored person as an actor in the activity generation process. Their simulator was capable of outputting data from simulated motion sensors, RFIDs and water consumption.

General purpose large scale agent-based modelling software such as Repast (North et al., 2005) and MASON (Luke et al., 2004) provide interactive tools for the several stages of agent modelling, from the model specification to the visualization of the results. The communication between agents is achieved through peer-to-peer mechanisms that allow the direct invocation of methods among them. In MASSHA, the communication between agents and the environment follows the publish-subscribe mechanism: an agent in MASSHA can subscribe to certain types of events, whether from the environment or from other agents. When an event takes place, all subscribers in MASSHA are automatically notified. This approach allows the communication to be totally independent, providing a robust and scalable context for communication and interaction regardless of the complexity of the human activity model.

Finally, (Azkune et al., 2015) presented another model-based simulator, where varying activity executions, time-lapses and sensor error models are included. They provide an original way to capture realistic activity models using surveys to target users.

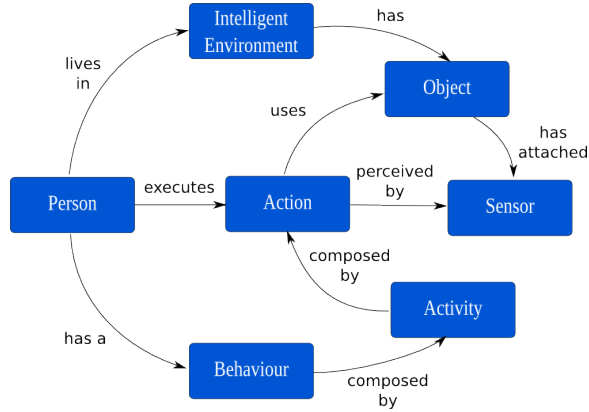
The major problem of model-based simulators is their dependence on activity models. This implies that the quality and accuracy of the resulting datasets, heavily relies on the quality of the activity description model and the associated parameters. However, obtaining accurate activity models is complicated, since human behaviour is very complex and full of details which are hard to capture. The work presented in MASSHA builds on the experience of previous research in the field of human behaviour by simplifying the process of the creation of activity datasets. The purpose is to provide a sensor-based simulation platform that generates activity datasets based on previously defined expert definitions of human behaviour. By simulating the interaction between occupants and sensors when carrying out daily activities, MASSHA produces accurate datasets that can then be used not only for human activity recognition -the main target of this work-, but also in diverse research domains, such as simulating the energy demand within a building, proper sizing of water and energy networks or load balancing.

## 4.3 Design and development

### 4.3.1 MASSHA's Human Simulation Model

The ability to simulate human behaviour within intelligent environments, such as smart homes or heavily monitored habitats, is closely related to the definition of a descriptive model that outlines the interactions between human actions and the entities that are receivers of those actions, all the while preserving the nature of decision making, inherent to human needs and desires. In this context, the human simulation model conceived for MASSHA is defined by the concepts of *Person*, *Intelligent Environment*, *Behaviour*, *Activity*, *Action*, *Object* and *Sensor*. Figure 4.1 shows an overview of these concepts along with the relationships among them.

The *Person* is the central axis of the behaviour model, an entity that lives in an *Intelligent Environment* (smart home, smart office, or even a smart city) whose definition can be as complex as desired, and is governed by spatial and temporal laws and constraints, such as the existence of physical barriers, sensors' detection ranges, avail-



**Figure 4.1:** Conceptual diagram of MASSHA's Human Simulation Model.

ability of open/closed roads, business calendars, saving light periods, etc. Specifically, for this research, our *Intelligent Environment* represents a smart home that houses a set of inanimate objects with specific *Sensors* attached to them. MASSHA currently follows the dense sensing monitoring approach which relies on miniaturised simple sensors installed on *Objects* of interest, such as cabinets, drawers, or taps, that trace a *Person's* interaction through a boolean state (i.e. activated/deactivated). For this research, we also implemented motion sensors, i.e. sensors that get activated when a moving entity is in their range of detection. It is important to point out that MASSHA's model is not solely restricted to this monitoring approach, but allows the definition and implementation of any other type of sensors (i.e. occupancy, temperature, noise, etc.).

*Objects*, on their part, may also have their own set of constraints. For example, when a *Person* opens a door (executes action *open* over an object of type *door*), the state of the door changes to *open* and thus, can not be opened again by another person until the state of the door changes again to *closed*. The main idea behind this is the desire to maintain simplicity in the objects' behaviour, really focusing on the interaction initiated by a person or another object.

Each *Person* has a specific *Behaviour* that defines the characteristics of its interaction with the *Intelligent Environment*. At this point, the behaviour model follows the well established hierarchy of *Behaviour*, *Activity* and *Action* (Chen et al., 2012). *Actions* are short-timed events *executed* by a *Person* that can be *perceived* by *Sensors*, such as grabbing a cup of tea or opening a door. In the dense sensing monitoring approach, actions are directly mapped to the activation and deactivation of sensors. *Activities* are seen as *sequences of actions* in a time frame ranging from minutes to hours, such as preparing breakfast, making coffee or brushing teeth. The difference between an *Action* and an *Activity* is not only the time lapse, but also the inherent higher semantic level due to the existence of more complex interactions between objects and people. The *Behaviour* helps define the way in which the person performs the *Activities*, in regards to order, relationships, and priorities according to both the person's personal preferences and the constraints of the *Intelligent Environment* in which the person lives.

#### 4.3.2 Parameterising the Human Behaviour Model

The selection of which activities to carry out during the day and the order in which to perform them is driven not only by the person's needs and preferences, but also by any event that might take place in the intelligent environment and directly influence the person's behaviour. From the moment a person wakes up, they have a list of daily activities that must be fulfilled by the time the day comes to an end. The person will then try to complete as many activities as possible, performing, at least, the most essential in the impossibility of performing all of them. Therefore, the key point in emulating a person's behaviour is the capability to complete as many daily activities as possible, taking into consideration different aspects and fitting them to the person's needs. To this end, the behaviour of our simulated person is based on three main concepts:

- **Activities:** These model every single possible activity in the use case. Activities consist of a set of basic attributes that define their priorities, their approximate duration or the conditions for their completion. Some examples of activities

include *wake up*, *prepare dinner*, *do laundry* and *take shower*. In addition, each activity has a set of associated *actions* that establish the sensors that can be activated during the activity execution. For example, performing the *preparing dinner* activity may involve the execution of several actions such as *open cupboard*, *open fridge*, and *use microwave* which are then perceived by the corresponding sensors, as shown in Table 4.1.

**Table 4.1:** Example of activities and their corresponding actions.

Activity	Actions
Preparing breakfast	Open drawer   Open fridge   Open cabinet   Use microwave
Washing dishes	Open drawer   Open cabinet   Use dishwasher
Shower	Use toilet   Use shower faucet HOT   Use shower faucet COLD
Preparing dinner	Use oven   Use drawer   Use cabinet   Use microwave   Use fridge
Do laundry	Use door   Use hamper   Use washing machine   Use laundry dryer

- **TODO list:** Every person has a list of all the activities they need to perform during the day. This list can be extended automatically since there are activities that will add prerequisites and others that will generate new needs, e.g. the *preparing dinner* activity generates a new *washing dishes* activity, even though it was not originally defined in the person's TODO list. The TODO list is reset every single day.
- **DONE list:** Every person keeps track of the activity that is being done at each moment, as well as the activities that are already finished. It is necessary to point out that the execution of certain activities may remove others from this list. For example, after cooking food, a person will add the *cooking activity* to the *DONE* list. However, when the person executes the *eat* activity, the *cooking* activity will be removed from the *DONE* list, since the execution of another *cooking* activity will be necessary if the agent wants to eat again.

The definition of these concepts is heavily influenced by the Belief-Desire-Intention paradigm (BDI) which provides mechanisms for separating the process of selecting a

plan from the execution of other currently active plans (Velleman and Bratman, 1991). In general, people tend to perform similar activities throughout the day, though they do not execute them in the same order, invest the same amount of time, or give them equal importance. When selecting an activity over others, there are several factors that come into play, such as the priority of the activity, the amount of time it is expected to take, any possible time restrictions (e.g. the person must perform a *go to work activity* at a certain hour, so all the morning activities, or at least the most important ones, should be finished by then), or even the existence of mandatory preconditions must be fulfilled before the activity is carried out (e.g. the *take shower* activity must be performed before the *dressing* activity). In order to allow greater flexibility, we have selected those attributes that, to our consideration, influence most the selection process. The attributes that parameterise each type of activity within MASSHA are shown in Table 4.2.

As stated in Section 4.3.1, each activity defines its own set of actions that are directly related to the interaction between the person and the objects living in the intelligent environment. When an activity is chosen for execution, the person performs the actions within the activity. The execution of several actions is what defines the degree of completion of an activity. Note that for an activity to be completed, performing each and every single one of its actions is not always necessary. For example, *preparing lunch* is related to actions such as *use oven*, *use microwave* or *use burner*, but they are not always used every time the activity is executed. In order to differentiate and characterise how they help fulfil their associated activities, MASSHA specifies a set of particular attributes for each action, as shown in Table 4.3.

Aside from this parameters, the intelligence and behaviour of the persons simulated in MASSHA can be modeled as complex as desired. In terms of the including conditions such as the Alzheimer's disease, depression, or strokes the researcher would need to know and understand how the existence of these conditions affect human behaviour in order to build a behaviour model that takes this issues into consideration. For example, if a person stops the execution of an activity abruptly, but the sensors layout throughout the house do not detect any further interaction, could be an indication of the person suffering a stroke.

**Table 4.2:** Definition of activity attributes.

Name	Description
Type	Alphanumerical code that identifies the activity type.
Attendance	Attribute for describing whether a person has to be physically present during the execution of the activity. An attended activity will monopolise the attention of the agent until it finishes (i.e. <i>prepare dinner</i> ), whereas an unattended activity will allow the agent to just initiate it and carry on with other activities (i.e. <i>watch TV</i> ). This attribute is key in enabling a concurrent behaviour among agents and a more realistic representation of an intelligent environment where the simultaneous activation of several sensors may take place.
Estimated duration	Amount of time the activity will take, according to what the person estimates, even though there may be some deviations in reality
Priority	Number that determines the importance an agent gives to a certain activity. This priority is modified according to the preferences of the person and the compliance with the daily schedule.
Desirable start time	Indicative start time at which the person would like to start performing the activity. If possible, the agent will try to schedule the start of the activity at this time.
Desirable end time	Indicative end time at which the agent would like to have the activity already finished. If possible, the agent will try to schedule the end of the activity before this time.
Mandatory start time	Compulsory time at which the activity must start. Necessary to model time critical activities such as going to work or watching a certain TV show.
Mandatory end time	Compulsory time at which the activity must finish.
Preconditions	Activity codes for modelling prerequisites, i.e. other related activities that must be already finished by the time this activity is chosen. These codes will be checked against the person's <i>DONE</i> list to ensure their fulfilment. If there are pending preconditions, these missing activities will be added to the <i>TODO</i> list and the person will choose once again which activity they will perform.
Outcomes	Performing some activities may lead to the appearance of other new activities, like needing to <i>take a shower</i> after <i>doing exercise</i> , or <i>toileting</i> after <i>having lunch</i> . For this purpose, there is a list of activity codes so that when an activity finishes, these new outcomes are appended to the person's <i>TODO</i> list.

### 4.3.3 Multi-Agent-System Implementation

The use of Multi-Agent Systems (MAS) allows to face the problem of human behavioural modelling from a bottom-up perspective, by focusing on the characteristics

**Table 4.3:** Definition of action attributes.

Name	Description
Duration mean	Average duration of the action
Duration deviation	Deviation of the duration of the action.
Finish increment	Amount of completion fulfilled with the execution of this action.
Frequency	Defines how much an specific action within an activity is prone to be executed.
Associated sensor	Sensor attached to the household appliances and that gets activated every time a person uses them.

of each individual separately. The mapping of each concept in our conceptual human simulation model into MAS is as follows:

- **Person:** Each person living in the smart home is implemented as an intelligent *active* agent, i.e. an entity aware of the characteristics and constraints of the intelligent environment, with a specific set of preferences, capable of taking their own decisions and with the ability to interact and communicate with both the agents and the environment.
- **Intelligent Environment:** Implemented as an Agent Environment, the Intelligent Environment emulates the characteristics of the smart home and represents the habitat in which the agents live. The Agent Environment has been recognised as an independent and living first-class entity (Weyns and Michel, 2015), that enables agents to interact and communicate while providing a centralised synchronisation protocol that supports the simultaneity of actions for system consistency, an overlay network that represents relationships between agents, and notification of contextual events for the production, and delivery of event notifications to create dynamic agent contexts.
- **Object:** Each object in the smart home is modelled as an intelligent *passive* agent, i.e. an entity that has its own specific behaviour and only reacts to stimuli initiated by other entities in the Intelligent Environment, never interacting with others on its own. When an agent of type Person interacts with any of the agents

of type Object (microwave, tap, fridge, etc.), the latter reacts by executing its own behaviour and notifies it to its attached sensors.

- **Sensor:** Each sensor in the smart home is modelled as a simple entity that emits notifications about a change in its state in accordance to the behaviour of its associated object.
- **Activity:** Each activity is modelled as a simple entity with several attributes that define its own characteristics (see Table 4.2). Each activity can store a list of related actions.
- **Action:** Actions inherit all the properties of activities, and are directly associated to the use of an *Object*. See Table 4.3.
- **Behaviour:** The behaviour of each person is modelled as a skill. Skills are the degree of competence a person has when carrying out a certain task. In particular, for MASSHA, the behaviour skill stores a list of all the activities the person has to carry out and provides a roulette-wheel based mechanism to select which one of them will be executed at each moment.

Prior to the simulation, both active and passive agents are loaded into the environment. The smart home is built in terms of spaces (kitchen, bathroom, bedrooms, etc.) and paths that allow movement within the house. The house is assigned a specific number of occupants, which in turn, are configured with a particular *TODO* list with the set of activities that each need to perform.

When the simulation starts, each person living in the smart home “wakes up” and initiates their behaviour algorithm. Specifically, the algorithm is invoked every time the person needs to choose a new activity for execution as follows:

- The algorithm searches the *TODO* list for activities which have a mandatory execution time.
  - If there exists an activity whose scheduled mandatory time falls close to the current simulation time, that activity is *selected* for execution.
  - If there are no activities that fulfil this condition, the algorithm inserts the pending activities into its roulette-wheel. This roulette-wheel creates a weighted distribution based on activities’ priority and according to the

distribution suggests the most suitable pending activity to be *selected* for execution. Each activities' priority is increased when the current simulation is close to the *desirable start* and *desirable end* times defined for that activity.

- The algorithm analyses whether the selected activity has any pending preconditions.
  - If the selected activity does not have any pending preconditions, the activity is executed right away.
  - If the selected activity has preconditions, the algorithm updates the *TODO* list with all pending preconditions, setting their priority to be higher than that of the initially selected activity.
- When the person has finished executing the activity, the behaviour algorithm starts once again.

```

procedure GET-NEXT-ACTIVITY( TODO-LIST , DONE-LIST )
inputs: TODO and DONE activities lists
while TODO-LIST is not empty do
  for activity in TODO-LIST do
    if activity.mandatory_start_time is near NOW then
      selected := activity;
      return selected;
    end
  end
  new ROULETTE-WHEEL( TODO-LIST );
  highest_priority_activity := ROULETTE-WHEEL.getHighestPriority();
  if highest_priority_activity.preconditions are all in DONE-LIST then
    selected := highest_priority_activity;
    return selected;
  end
  else
    TODO-LIST.insert( highest_priority_activity.preconditions );
    // Go back to while start
  end
end

```

**Algorithm 1:** Pseudocode for Activity selection model.

### 4.3.4 Inputs and Outputs

The configuration of the activities in MASSHA is very simple. MASSHA loads a JavaScript Object Notation file (JSON) annotated with the activities the person likes to do along with information like time and precedence restrictions to be considered.

The JSON file is composed of two parts:

- **Activity configuration:** Configures all the possible activities that the agent can execute. This is used to set the default parameters when an activity is created within the simulation. Each activity is composed of several actions, i.e. activation of sensors throughout the environment. Each action defines parameters such as the frequency of activation within the activity, estimated duration when invoked, and how much their activation helps to reach the objective of the activity execution.
- **TODO List:** Defines the specific activities the agent will have to carry out, along with the basic priorities, priority increment, mandatory start times, etc. All of these parameters are used by the activity selection algorithm to decide which activity to execute at each moment.

With these preferences, MASSHA creates a personalised version of the person's activity schedule and behaviour, emulating his/her interaction with the objects around. In addition, MASSHA is able to generate output in any format needed. For the activity simulation scenario, the output files follow the format shown in Table 4.4, however, MASSHA can be configured to register any kind of data, such as energy consumption monitoring derived from the usage of objects, or even the equivalent CO<sub>2</sub> generation.

## 4.4 Materials and Methods

### 4.4.1 Datasets

In order to validate MASSHA in both single-user and multiple-user paradigms, we considered several reference datasets. After a careful analysis, the datasets that were

finally selected were those collected by (Tapia et al., 2004) and (Cook and Schmitter-Edgecombe, 2009), since they present two different implementations of dense sensing monitoring approaches and the distribution of the sensors throughout the household is accurately specified.

- **Single-user:** The dataset chosen to evaluate the single-user case provides information about the activation of sensors located inside two different single-person apartments, with about 77 and 84 reed switch sensors installed for monitoring around 16 activities of daily living (see (Tapia et al., 2004) for detailed information). The first subject ( $DS_1$ ) was a 30-year-old professional woman, whereas the second subject ( $DS_2$ ) was an 80-year-old woman. Both subjects spent most of their time at home and lived alone in their one-bedroom apartments. Once the sensors were installed, they started collecting data for 14 days. During the study, the subjects used a custom annotation tool to build a detailed record of their activities. The datasets are formatted in comma separated files (CSV), where for each activity, the day, start time and end time as annotated by the subject appear. Additionally, for each annotated activity the activation and deactivation times for the involved sensors are listed.
- **Multiple-user:** The dataset chosen to evaluate the multiple-user case provides information about the activation of sensors located in the WSU smart apartment testbed (Cook and Schmitter-Edgecombe, 2009) during the 2009-2010 academic year. Specifically, the dataset provides information from 24th August 2009 to 1st May 2010. At that time, the apartment housed two residents, R1 and R2. Sensor events are annotated with the corresponding activity that was being performed, distinguishing among 13 activities for R1 and 12 for R2. WSU apartments rely on a slightly different monitoring approach, stressing the usage of motion sensors all along the apartments. However, they also use object-attached sensors.

Since we address both single-user and multiple-user scenarios, we believe the combination of these datasets provides a solid reference to evaluate the performance of MASSHA in terms of human behaviour modelling. Furthermore, we also show

how MASSHA can model different intelligent environments, ranging from the heavily sensorised apartments from (Tapia et al., 2004) to the lighter sensorised and motion-based monitoring approach from (Cook and Schmitter-Edgecombe, 2009).

**Table 4.4:** Format of the RAW datasets used to validate MASSHA ((Tapia et al., 2004)).

timestamp	id sensor	sensor	action	activity
2003-03-27 06:43:40	67	Cabinet_67	ON	Toileting
2003-03-27 06:43:43	67	Cabinet_67	OFF	Toileting
2003-03-27 06:44:06	100	Toilet Flush_100	ON	Toileting
2003-03-27 06:44:20	101	Light switch_101	ON	Toileting
2003-03-27 06:44:35	57	Medicine cabinet_57	ON	Toileting
2003-03-27 06:44:36	58	Medicine cabinet_58	ON	Toileting
⋮	⋮	⋮	⋮	⋮
2003-04-11 22:24:17	71	Drawer_71	ON	Grooming
2003-04-11 22:24:17	71	Drawer_71	OFF	Grooming

In a pre-processing stage, we analysed the original datasets to identify the most relevant information that helped build a human behaviour model. Table 4.4 shows the format for the original datasets. The objective is that, when the behaviour model is executed in MASSHA, the resulting dataset is similar to the original in terms of frequency, duration and hourly activation of the sensors. First of all, we computed the sequence of activities, time-slots and mean duration for each dataset and monitored day, so as to identify any behavioural patterns in terms of the order in which the activities were executed and the sensors that were activated each time. This data was then used to build the configuration file specified in Section 4.3.4. Table 4.5 contains an example of the activity model extracted from the reference datasets, used as input to MASSHA.

**Table 4.5:** Pre-process datasets used to validate MASSHA.

activity	sensor	freq	mean (s)	deviation (s)
$activity_1$	$sensor_1$	$f_1^1$	$\mu_1^1$	$\sigma_1^1$
$activity_1$	$sensor_2$	$f_1^2$	$\mu_1^2$	$\sigma_1^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$activity_1$	$sensor_n$	$f_1^n$	$\mu_1^n$	$\sigma_1^n$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$activity_m$	$sensor_1$	$f_m^1$	$\mu_m^1$	$\sigma_m^1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$activity_m$	$sensor_n$	$f_m^n$	$\mu_m^n$	$\sigma_m^n$

#### 4.4.2 Validation Methodology

From now on, we will name the dataset generated by MASSHA as *simulated dataset*, whereas we will name *reference dataset* to the dataset we want to replicate. The validation methodology is based on the following definition:

**Definition 1** (Similarity). *Two persons are considered to behave similarly if and only if:*

- *The frequency and duration of every action within a certain activity are analogous.*
- *The temporal distribution of the actions throughout the day is analogous.*

Within the sense of *similarity*, we differentiate between two concepts: assessment of the *internal consistency* and *coherence with real measurements*.

- **Assessment of the internal consistency.** The objective is to verify whether MASSHA is able to build datasets that emulate the behaviour of a certain activity model. For this analysis, we first built an empirical distribution of the frequency and duration of the activation for each sensor available in the reference datasets. After defining a *TODO* list (see Section 4.3.3) based on the activity patterns

of each subject, we ran several simulations to generate simulation datasets by recording the person's interaction with sensors through the execution of actions within activities. The frequency and duration values obtained from the simulations were then analysed to verify whether they lay inside the confidence intervals for the expected values given by the reference datasets. Since the latter did not follow any particular distribution, the confidence intervals were built using bootstrapping methods (Davison and Hinkley, 1997). In addition, the existence of about 100 sensors made it compulsory to take into consideration the *look-elsewhere effect* (Randall Munroe, 2011), overcome with the use of the Bonferroni correction (Derrac et al., 2011).

- **Coherence with real measurements.** The objective is to verify whether the simulator can replicate the reference dataset by following the evaluation methodology proposed by Helal *et al.* (Helal et al., 2011). In this case, we built a *TODO* list that detailed the activities to be performed for a period of 14 days. In this context, the simulated and reference datasets are *similar* if the distributions of frequencies, duration and distributions of sensors' activations are equivalent (Robinson et al., 2005) with respect to a *region of similarity*, i.e. the percentage of differences that can be considered *experimental errors*. This region should be defined following expert advice or other (not statistical) procedures (Walker and Nowacki, 2011). Since there is no clear procedure to fix this value within this context, we have plotted the relation between the length of the region of similarity and the number of not similar sensors. The use of this method provides more trustworthy results over other tests like those of Kolmogorov-Smirnov or Anderson-Darling, since these methods can only assess the dissimilarity, but never the similarity. After the region of similarity is defined, two one side confidence intervals for a statistic are tested whether they lie within the region of similarity. In this case, the null hypothesis is rejected and can be concluded that both samples come from the same population, allowing to conclude the existence of equivalence. Finally, as in the previous case, we used the Bonferroni correction to overcome the *look-elsewhere effect*.

## 4.5 Results

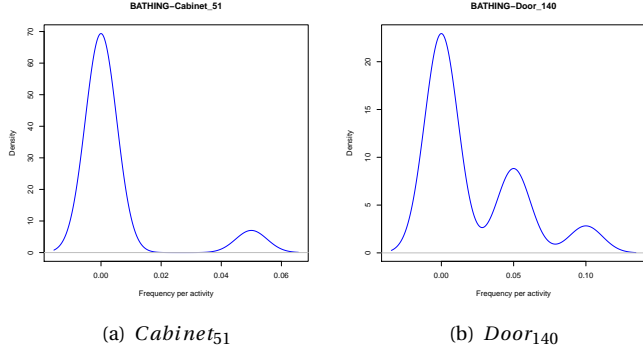
### 4.5.1 Assessment of internal consistency

For the *assessment of internal consistency*, we defined a person's activity model and ran 100 simulations in MASSHA to build the distribution of frequencies and duration of activation for all the sensors in each of the single-user and multiple-user scenarios. In both cases, every simulation configured a set of daily activities for each of the occupants living in the smart home. Even though we could have selected any random activity models as reference, we used those obtained from the datasets described in Section 4.4.1.

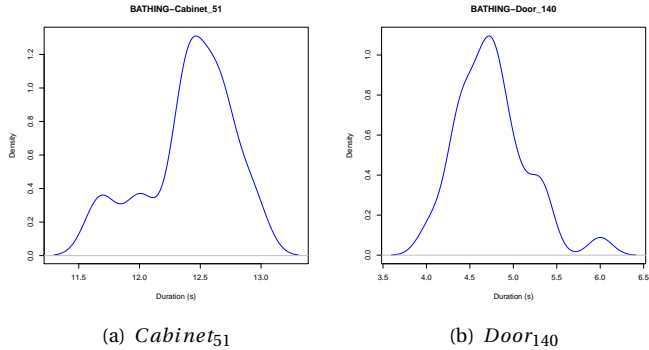
#### Single-user scenario

Figure 4.2 and Figure 4.3 show, respectively, a distribution of the *frequency* and *duration* of activation of two sensors within the *Bathing* activity for the 100 simulations performed in the single-user scenario. Clearly, the plots are quite irregular and do not follow any standard distribution. Being a common issue among all the sensors identified in the reference datasets for the single-user scenario, we needed to use bootstrapping techniques in order to build the confidence intervals needed to evaluate the accuracy of the datasets generated by MASSHA. In particular, for  $DS_1$  and  $DS_2$ , the attendance attribute gains importance, since several activities, like *WatchingTV* or *ToiletFlush*, do not need for the person to be physically present, allowing the simulation of parallel activities.

Table 4.6 contains an extract of the results of one of the simulations carried out for  $DS_1$ . In this example, the duration of activation (column *duration*) from the reference dataset for all the sensors of the *Bathing* activity lie within the 0.05-confidence interval (columns *low C.I. (%)* and *upper C.I. (%)*). On the other hand, for 8 sensors, the frequency of activation (column *frequency*) from the reference dataset for the same sensors do not lay within the 0.05-confidence interval (columns *low C.I. (s)* and *upper C.I. (s)*).



**Figure 4.2:** Distribution of the frequency of activation of two sensors in  $DS_1$  (single-user scenario).



**Figure 4.3:** Distribution of the duration of activation of two sensors in  $DS_1$  (single-user scenario).

Overall, for  $DS_1$ , 94 % of the reference values for the frequencies of activation lie within the 0.05-confidence interval, whereas, 88 % do so for  $DS_2$ . Further, for  $DS_1$ , 98 % of the reference values for the durations of activation lie within the 0.05-confidence interval, while 99 % do so for  $DS_2$ .

**Table 4.6:** Internal Consistency – Example of the results for one of the executions on  $DS_1$  (single-user scenario).

activity	sensor name	freq. (%)	low C.I. (%)	upper C.I. (%)	similar freq.	dur. (s)	low C.I. (s)	upper C.I. (s)	similar dur.
BATHING	Burner94	1.6	1.3	2.2	✓	2.2	2.1	2.3	✓
BATHING	Cabinet138	0.4	0.3	0.8	✓	20.5	19.2	21.1	✓
BATHING	Cabinet51	0.4	0.3	0.7	✓	13.1	12.3	13.7	✓
BATHING	Cabinet59	0.4	0.2	0.7	✓	41.6	39	44	✓
BATHING	Cabinet61	0.4	0.3	0.9	✓	12.7	12.5	13.8	✓
BATHING	Cabinet66	0.4	0.3	0.8	✓	41.9	39.1	43.3	✓
BATHING	Cabinet67	8.6	7.6	9.3	✓	79.7	79.1	81.8	✓
BATHING	Cabinet72	0.8	0.7	1.3	✓	18.1	17.4	18.7	✓
BATHING	Cabinet79	2.3	1.9	3.3	✓	3 611	3 559.7	3 728.6	✓
BATHING	Cabinet83	0.4	0.3	0.9	✓	23.2	22.3	24.9	✓
BATHING	Closet81	1.6	1.8	2.8	✗	10.6	10.4	10.9	✓
BATHING	Dishwasher70	0.8	0.6	1.3	✓	2 060	1 984.4	2 164.4	✓
BATHING	Door130	3.4	3.1	4.4	✓	141.4	138.4	145.5	✓
BATHING	Door140	1.6	1.4	2.3	✓	4.7	4.6	4.9	✓
BATHING	Door54	2.6	2.4	3.5	✓	2.5	2.5	2.6	✓
BATHING	Drawer62	1.6	1.7	2.6	✗	2 515.8	2 433.9	2 575.9	✓
BATHING	Drawer71	0.4	0.4	0.9	✓	7.4	6.9	7.7	✓
BATHING	Drawer75	0.4	0.3	0.9	✓	16.4	16	17.5	✓
BATHING	Drawer78	0.4	0.2	0.7	✓	2.1	2	2.2	✓
BATHING	Drawer82	2.4	2.3	3.3	✓	1 866.8	1 804.6	1 888	✓
BATHING	ExhaustFan96	4	4	5.2	✗	1 098	1 074.7	1 120.7	✓
BATHING	Freezer137	1.6	1.2	2.1	✓	726.9	700.5	740.7	✓
BATHING	Garbagedisposal98	0.4	0.2	0.6	✓	1	1	1	✓
BATHING	Jewelrybox139	1.1	0.8	1.6	✓	12.1	11.9	12.8	✓
BATHING	Lamp76	0.4	0.3	0.8	✓	7 164.2	6 732.4	7 374.7	✓
BATHING	Lightswitch101	4.4	3.9	5	✓	2 122.1	2 073.6	2 167.8	✓
BATHING	Lightswitch104	0.4	0.3	0.8	✓	2.1	2	2.3	✓
BATHING	Lightswitch108	0.4	0.3	0.8	✓	5 706.3	5 445.3	5 998.5	✓
BATHING	Lightswitch92	2	2	3.1	✗	3 675.1	3 589.2	3 807	✓
BATHING	Lightswitch95	0.4	0.4	0.9	✗	1.1	1	1.1	✓
BATHING	Medicinecabinet57	7.1	6.3	7.8	✓	2 538.6	2 482.1	2 581.8	✓
BATHING	Medicinecabinet58	5.3	4.8	6.5	✓	36.5	35.9	37.3	✓
BATHING	Refrigerator91	0.4	0.2	0.8	✓	18.2	17.3	19.1	✓
BATHING	Showerfaucet93	28.6	20.3	23.2	✗	1 083.6	1 059.9	1 090.9	✓
BATHING	Sinkfaucet-cold88	3.4	3.5	4.5	✗	13.4	13.1	13.6	✓
BATHING	Sinkfaucet-hot68	3.5	3.3	4.3	✓	12.8	12.5	13.1	✓
BATHING	Toaster131	0.4	0.3	0.8	✓	1	0.9	1	✓
BATHING	ToiletFlush100	4.6	3.4	4.4	✗	3 369.4	3 334.9	3 503.3	✓
BATHING	WashingMachine142	0.4	0.3	0.8	✓	34.3	32.6	36	✓

† Lower C.I.: Lower limit of the confidence interval

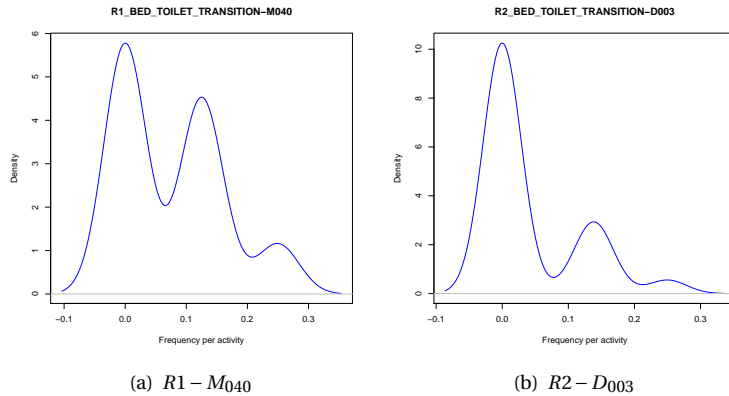
† Upper C.I.: Upper limit of the confidence interval

† freq.: frequency

† dur.: duration

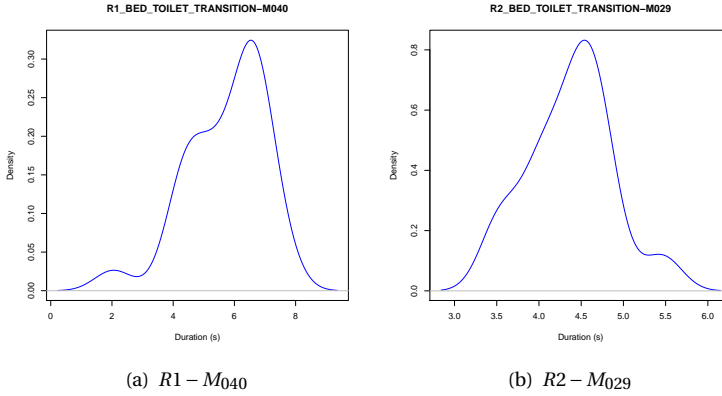
### Multiple-user scenario

Figure 4.4 and Figure 4.5 show, respectively, the *frequency* and *duration* of activation of two sensors for the 100 simulations performed within the *Bed toilet transition* activity for the two occupants of the smart home, *R1* and *R2*. As in the single-user scenario, the plots do not follow any standard distribution. Being a common issue among all the sensors in the multi-user activity dataset, we used bootstrapping techniques in order to build the confidence intervals needed to evaluate the accuracy of the generated dataset.



**Figure 4.4:** Distribution for the frequency of activation of two sensors in  $DS_3$  (multiple-user scenario).

Table 4.7 contains an extract of the results of one of the simulations carried out for  $DS_3$ . In this example, all frequencies of activation from the reference dataset for the sensors within the activities *R1-Housekeeping* and *R2-Housekeeping* (column *frequency*) lie within the 0.05-confidence interval (columns *low C.I. (%)* and *upper C.I. (%)*). On the other hand, for 6 sensors, the duration of activation (column *duration*) from the reference dataset does not lay within the 0.05-confidence interval (columns *low C.I. (s)* and *upper C.I. (s)*).



**Figure 4.5:** Distribution for the duration of activation of two sensors in  $DS_3$  (multiple-user scenario).

Overall, for  $DS_3$ , 99% of the reference values for frequencies of activations lie within the 0.05-confidence interval, while for 88% of the reference values for the duration of activations lie within the 0.05-confidence interval.

#### 4.5.2 Coherence with real measurements

Regarding the *coherence with real measurements*, the objective is to assess whether MASSHA is able to create datasets that are similar to those recorded in real conditions. To this end, we estimated the duration, frequency, and hourly execution of daily activities in order to build a human behaviour model (see Section 4.3.1) for the single-user and multiple-user scenarios. MASSHA was then configured to simulate the execution of activities for a 14-day period in both cases.

##### Single-user scenario

Table 4.8 shows a comparison between the frequency and duration of activation for sensors within the activities *Grooming*, *PreparingBeverage*, *PreparingBreakfast*, and *PreparingLunch* for MASSHA and  $DS_1$ . Figure 4.6 compares the distributions

**Table 4.7:** Internal Consistency – Example of the results for one of the executions on  $DS_3$  (multiple-user scenario).

activity	sensor name	freq. (%)	lower C.I. (%)	upper C.I. (%)	similar freq.	dur. (s)	lower C.I. (s)	upper C.I. (s)	similar dur.
R1-BED-TOILET-TRANSITION	D004	3.3	0.9	5.9	✓	100.1	98.2	100.7	✓
R1-BED-TOILET-TRANSITION	M029	6.6	2.2	8.9	✓	3.2	2.9	4.8	✓
R1-BED-TOILET-TRANSITION	M037	7.1	5	12.2	✓	4.3	4.4	5	✗
R1-BED-TOILET-TRANSITION	M038	6.7	2.4	9.4	✓	4.2	4.5	5.1	✗
R1-BED-TOILET-TRANSITION	M039	6.9	3	9.6	✓	4.1	3.8	4.7	✓
R1-BED-TOILET-TRANSITION	M040	6.9	6	14.4	✓	5.9	4.8	6.9	✓
R1-BED-TOILET-TRANSITION	M041	9.4	9.3	22	✓	8.2	6.5	10.5	✓
R1-BED-TOILET-TRANSITION	M043	6.8	3.5	12.2	✓	4.3	3.2	4.4	✓
R1-BED-TOILET-TRANSITION	M044	6.7	3.5	10	✓	6.9	5.4	7.7	✗
R1-BED-TOILET-TRANSITION	M045	6.8	3.7	11.5	✓	4.4	3.8	4.6	✓
R1-BED-TOILET-TRANSITION	M046	10.2	5.2	14.5	✓	6.4	3.4	9.5	✓
R1-BED-TOILET-TRANSITION	M047	6.9	2.8	11.1	✓	12.7	12.2	14	✓
R1-BED-TOILET-TRANSITION	M050	10.7	4.9	13.5	✓	3.6	2.7	4.8	✓
R2-BED-TOILET-TRANSITION	D003	38.7	33.2	54	✓	226.5	224.7	227.5	✓
R2-BED-TOILET-TRANSITION	D005	2.4	1.8	4.5	✓	18.1	17.8	20	✓
R2-BED-TOILET-TRANSITION	M029	2.5	0.2	3.5	✓	4.5	4	4.8	✓
R2-BED-TOILET-TRANSITION	M030	5	1.9	8.6	✓	6.3	6.3	6.7	✗
R2-BED-TOILET-TRANSITION	M032	4.3	4.5	13.4	✓	5.5	4.6	7.7	✓
R2-BED-TOILET-TRANSITION	M033	2.6	0.3	4.5	✓	3.2	0.4	4.7	✓
R2-BED-TOILET-TRANSITION	M034	2.6	0.3	3.5	✓	1.1	1.8	4	✗
R2-BED-TOILET-TRANSITION	M035	2.6	0.6	5.1	✓	2	0.3	4.3	✓
R2-BED-TOILET-TRANSITION	M036	2.5	1	6.4	✓	8.6	7.4	13.2	✓
R2-BED-TOILET-TRANSITION	M037	4.8	1.6	7	✓	11.3	10.2	12.2	✓
R2-BED-TOILET-TRANSITION	M038	5.1	3.4	13.3	✓	8.4	9	15.5	✗
R2-BED-TOILET-TRANSITION	M039	10.2	6.8	14.9	✓	5.1	3.5	6.8	✓
R2-BED-TOILET-TRANSITION	M040	7.3	3.8	12.3	✓	4.4	3.1	5.4	✓
R2-BED-TOILET-TRANSITION	M041	10.2	6.7	14.8	✓	6.2	5.8	9.7	✓
R2-BED-TOILET-TRANSITION	M046	25	14.8	28.9	✓	5.8	3.9	7.7	✓
R2-BED-TOILET-TRANSITION	M047	9.7	5.5	15.4	✓	5.4	4.1	7.1	✓

† Lower C.I.: Lower limit of the confidence interval

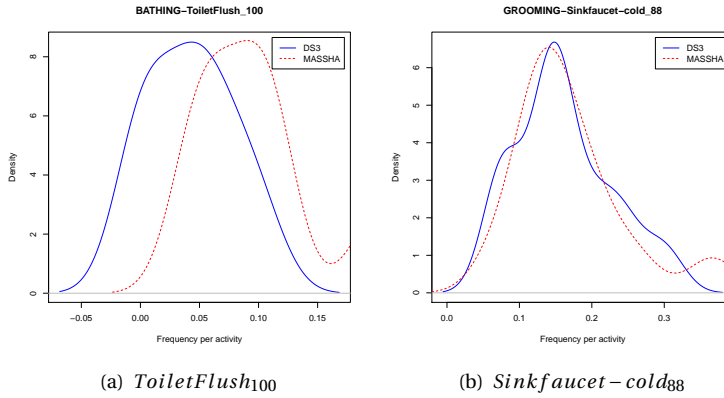
† Upper C.I.: Upper limit of the confidence interval

† freq.: frequency

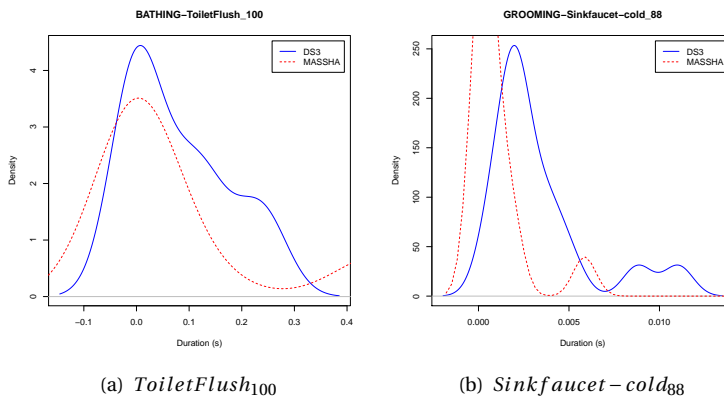
† dur.: duration

for the *frequency* of activation of two sensors between MASSHA and  $DS_1$ . Figure 4.7 compares the distributions for the *duration* of activation of two sensors for the same datasets. As can be seen, the distributions for the real and simulated datasets are similar, meaning that MASSHA is able to replicate the original single-user dataset in terms of frequencies and durations of activation.

Finally, Table 4.8 shows the results of the similarity test for the frequency and duration for a extract of the sensors and activities taking as *region of similarity* a 15 % of distance (see Section 4.4.2 for details).



**Figure 4.6:** Distribution for the frequency of activation for two sensors in MASSHA and DS<sub>1</sub> (single-user scenario).



**Figure 4.7:** Distribution for the duration of activation for two sensors in MASSHA and DS<sub>1</sub> (single-user scenario).

**Table 4.8:** Coherence with real measurements – Extract of the results of the comparison MASSHA vs  $DS_1$  (single-user scenario).

activity	sensor name	mean freq.	mean freq.	similar frequency	mean dur.	mean dur.	similar duration
		W (%)	M (%)		W (s)	M (s)	
GROOMING	Cabinet67	8.68	9.13	✓	6.29	0.08	✗
GROOMING	Door130	10.32	3.06	✗	0.19	0.12	✓
GROOMING	ExhaustFan96	6.84	2.25	✓	0.12	0.45	✓
GROOMING	Jewelrybox139	5.81	3.21	✓	0.03	0.45	✓
GROOMING	Lightswitch101	7.6	5.89	✓	6.46	2.93	✗
GROOMING	Lightswitch120	5.73	1.25	✓	0.03	1.22	✓
GROOMING	Medicinecabinet57	14.54	11.28	✓	63	56.59	✗
GROOMING	Medicinecabinet58	12.45	13.28	✓	6.63	0.8	✗
GROOMING	Sinkfaucet-cold88	16.97	16.12	✓	1.91	0.23	✓
GROOMING	Sinkfaucet-hot68	17.97	19.58	✓	2.09	0.33	✓
GROOMING	ToiletFlush100	8.06	1.73	✓	7.04	3.16	✗
PREPARINGBEVERAGE	Cabinet80	21.5	16.5	✗	4.15	0.14	✓
PREPARINGBEVERAGE	Freezer137	22.14	6.07	✗	18.6	15.29	✗
PREPARINGBEVERAGE	Refrigerator91	24.27	21.25	✗	12.93	0.42	✗
PREPARINGBREAKFAST	Cabinet73	12.81	2.5	✗	9.18	0.02	✗
PREPARINGBREAKFAST	Drawer84	11.61	7.5	✓	1.36	0.02	✓
PREPARINGBREAKFAST	Freezer137	9.7	2.92	✗	21.7	4.96	✗
PREPARINGBREAKFAST	Refrigerator91	11.71	6.56	✗	3.53	0.02	✓
PREPARINGBREAKFAST	Toaster131	27.06	10.63	✗	4.88	5.65	✓
PREPARINGDINNER	Cabinet55	7.02	5	✓	1.35	0.06	✓
PREPARINGDINNER	Cabinet80	6.62	7.5	✓	1.34	0.02	✓
PREPARINGDINNER	Drawer84	7.51	3.33	✓	0.35	0.02	✓
PREPARINGDINNER	Freezer137	18.3	16.43	✗	18.75	4.24	✗
PREPARINGDINNER	Refrigerator91	18.98	15.63	✗	3.26	0.21	✓
PREPARINGLUNCH	Cabinet53	6.17	4.77	✓	2.38	0.42	✓
PREPARINGLUNCH	Cabinet55	6.82	6.14	✓	1.55	0.75	✓
PREPARINGLUNCH	Cabinet59	4.76	3.33	✓	0.36	0.06	✓
PREPARINGLUNCH	Cabinet72	7.52	3.33	✗	0.15	0.02	✓
PREPARINGLUNCH	Cabinet80	5.75	0.71	✓	1.26	0.02	✓
PREPARINGLUNCH	Dishwasher70	5.37	2.86	✓	12.57	11.36	✗

† W: Reference dataset

† M: MASSHA dataset

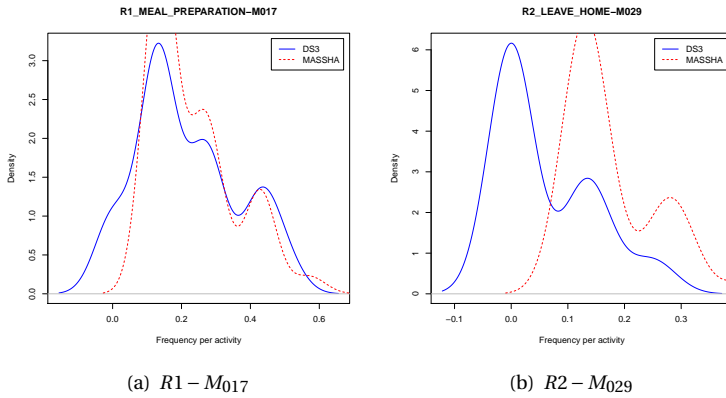
† mean freq.: mean frequency

† mean dur.: mean duration

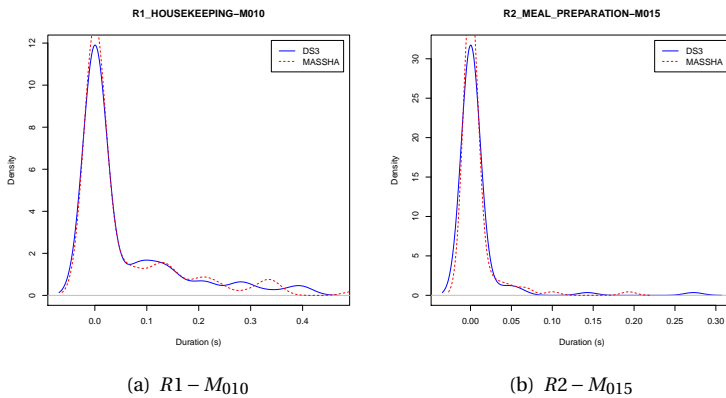
### Multiple-user scenario

Table 4.9 shows a comparison between the frequency and duration of activation for sensors within the activities  $R1\_Eating$ ,  $R2\_LeaveHome$  for MASSHA and  $DS_3$ . Figure 4.8 compares the distributions for the *frequency* of activation of two sensors between MASSHA and  $DS_3$ . Figure 4.7 compares the distributions for the *duration* of activation of two sensors for the same datasets. As can be seen, the distributions for

the real and simulated datasets are similar, meaning that MASSHA is able to replicate the original multiple-user dataset in terms of frequencies and durations of activation.



**Figure 4.8:** Distribution for the frequency of activation for two sensors in MASSHA and  $DS_3$  (multiple-user scenario)



**Figure 4.9:** Distribution for the duration of activation for two sensors in MASSHA and  $DS_3$  (multiple-user scenario).

**Table 4.9:** Coherence with real measurements – Extract of the results of the comparison MASSHA vs  $DS_3$  (multiple-user scenario).

activity	sensor name	mean freq. W (%)	mean freq. M (%)	similar frequency	mean dur. W (s)	mean dur. M (s)	similar duration
R1-EATING	M013	13.84	1.56	✓	4.24	1.91	✓
R1-EATING	M014	17.26	5.57	✓	3.68	3.39	✓
R1-EATING	M034	17.93	12.41	✓	13.14	10.6	✓
R1-EATING	M035	29.49	26.46	✗	6.26	3.83	✓
R1-LEAVE-HOME	D001	21.51	11.43	✓	10.27	5.5	✓
R1-LEAVE-HOME	M024	24.82	14.94	✓	13.89	8.88	✓
R1-LEAVE-HOME	M025	21.8	14.96	✓	9.44	6.32	✓
R1-LEAVE-HOME	M026	22.54	13.33	✓	5.12	2.93	✓
R1-LEAVE-HOME	M027	20.66	18.03	✗	3.76	3.78	✓
R1-LEAVE-HOME	M028	22.93	14.22	✓	5.42	3.66	✓
R1-LEAVE-HOME	M043	22.31	12.59	✓	2.09	1.35	✓
R2-EATING	M006	26.31	22.76	✗	7.38	7.45	✓
R2-EATING	M007	21.61	12.15	✓	6.65	3.64	✓
R2-EATING	M008	18.87	14.38	✗	3.02	2.31	✓
R2-EATING	M009	21.22	13.87	✓	4.86	3.68	✓
R2-EATING	M015	16.64	6.43	✓	4.23	1.46	✓
R2-EATING	M016	19.51	8.02	✓	5.27	2.18	✓
R2-EATING	M017	21.34	13.9	✓	14.1	8.6	✓
R2-EATING	M018	14.73	1.19	✓	3.25	0.16	✓
R2-EATING	M046	14.96	4.91	✓	0.39	0.12	✗
R2-EATING	M051	13.78	1.79	✓	0.85	0.18	✓
R2-LEAVE-HOME	D001	17.01	10.24	✓	17.01	10.24	✓
R2-LEAVE-HOME	M024	18.11	9.67	✓	18.11	9.67	✓
R2-LEAVE-HOME	M025	16.69	7.01	✓	16.69	7.01	✓
R2-LEAVE-HOME	M026	16.83	9.09	✓	16.83	9.09	✓
R2-LEAVE-HOME	M027	16.33	7.79	✓	16.33	7.79	✓
R2-LEAVE-HOME	M028	16.55	9.1	✓	16.55	9.1	✓
R2-LEAVE-HOME	M029	18.1	6.03	✓	18.1	6.03	✓
R2-LEAVE-HOME	M030	16.96	10.65	✗	16.96	10.65	✗
R2-LEAVE-HOME	M032	16.57	9.55	✗	16.57	9.55	✗
R2-LEAVE-HOME	M035	23.69	17.87	✗	23.69	17.87	✗
R2-LEAVE-HOME	M036	16.63	4.99	✓	16.63	4.99	✓

† W: Reference dataset

† M: MASSHA dataset

† mean freq.: mean frequency

† mean dur.: mean duration

## 4.6 Discussion

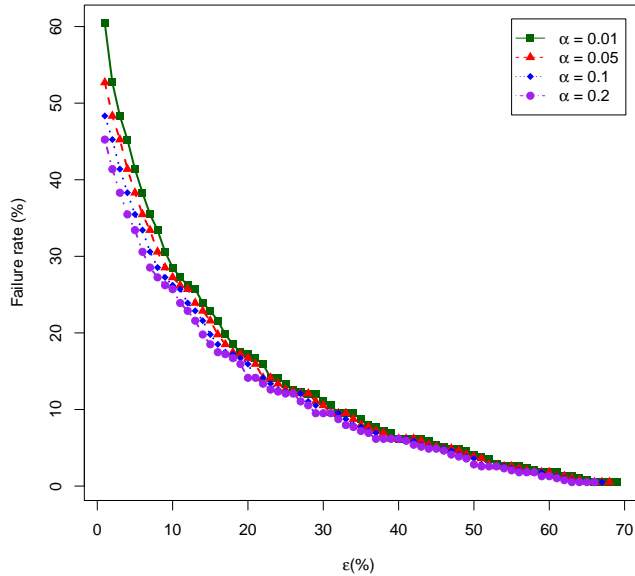
### 4.6.1 Overall assessment

Results validate that MASSHA is able to generate activity datasets that resemble those recorded and annotated in a smart home, regardless of the monitoring approach and the number of occupants in the house. In this sense, the similarity between the reference and the simulated datasets is given by the resemblance of the *frequency* and *duration* of activation of the sensors depending on the activity being performed at each moment. The particular characteristics and significance of the results obtained for the single-user and multiple-user scenarios are discussed next.

#### Single-user scenario

Figure 4.10 synthesises the performance of MASSHA regarding the consistency with real measurements. In the general case, identifying the acceptable threshold for the region of similarity is not feasible. The specific value may vary depending on the research focus, the variability of human behaviour and the monitoring time. Thus, every practitioner should check whether the results shown in Figure 4.11 satisfy their requirements, before deciding whether they use MASSHA for their experiments. However, taking into account the behaviour variability we observed in  $DS_1$  and  $DS_2$ , as well as the low number of monitored days (14 days), we believe MASSHA's performance is good. For a region of similarity of 20 %, the failure rate of produced sensor activations and their duration is around 15 %. So the similarity between the simulated dataset and the real one is around 85 % for a reasonable region of similarity, under the conditions described. In scenarios where activity models can be acquired more accurately, MASSHA's performance is expected to be even better.

Since the annotation of  $DS_1$  and  $DS_2$  included detailed information regarding sensor activation per hour, we decided to perform an additional test to verify whether MASSHA was able to reproduce this behaviour as well. Table 4.10 shows, for every sensor in  $DS_1$  and for every hour of the day, whether the real hourly activations of sensors matched those simulated by MASSHA. For example, for sensor *Burner*<sub>106</sub>,



**Figure 4.10:** Distribution of the amount of not similar sensors depending on the length of the *region of similarity* (epsilon) and *significance level* (alpha) for  $DS_1$  (single-user).

MASSHA coincided with its activations for hours  $h_1$ ,  $h_6$ ,  $h_{11}$ ,  $h_{14}$ ,  $h_{17}$ ,  $h_{20}$ ,  $h_{22}$  and  $h_{23}$  while did not coincide with its activations for hours  $h_2$ ,  $h_7$ ,  $h_{12}$  and  $h_{15}$ . Overall, the hourly activations of the real dataset and that produced by MASSHA seem to be quite different. The main cause is that the original dataset contains annotations and monitoring for a short period of time (only 14 days). This causes some of the activities to be wrongly annotated and the definition of actions within activities to be somewhat sloppy. Moreover, the subjects did not follow any consistent patterns neither throughout the day nor the week. Once again, this issue can be overcome by increasing the monitoring period in order to achieve a better definition of behaviour patterns and a more consistent outline of activities and related activation of sensors.

Please note that, in the case of the duration of activation, the amount of variance recorded from real environment (due to errors in the annotation methods). Moreover, MASSHA is able to closely recreate datasets recorded by humans since the amount of sensors without similarity is low even with a small region of similarity. In fact, to get a 15 % of disagreement, the *region of similarity* should only be of 21 %. Figure 4.10 shows the distribution of the amount of sensors that are not similar depending on the length of the *region of similarity* (epsilon) and the significance level (alpha). As expected, an increment in both values produces a reduction of the failure rate. Moreover, it can be observed that the failure rate is more sensitive to changes on the region of similarity than to the significance level.

#### 4.6.2 Multiple-user scenario

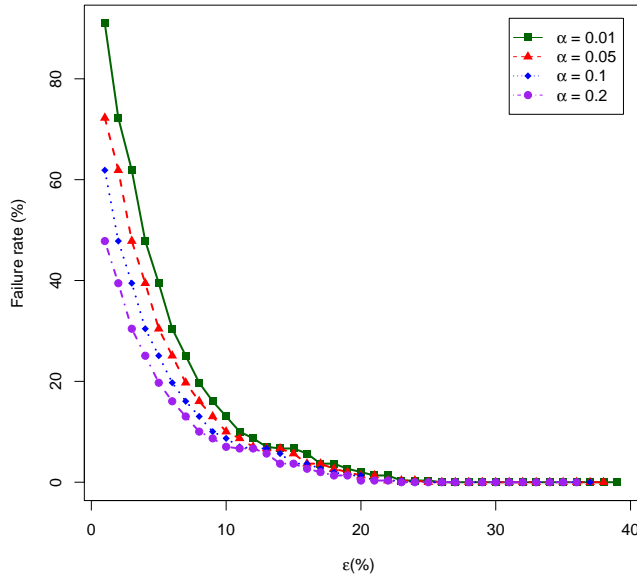
Figure 4.11 synthesises the performance of MASSHA regarding the consistency with real measurements in a multiple-user scenario. We can observe how the failure rate for sensor activations is generally lower than in Figure 4.10, which describes the single-user scenario. For example, given a region of similarity of 10 % MASSHA achieves a failure rate of around 15 %, taking into account the different significance levels depicted in the graphic. Please, notice that to achieve a similar failure rate in the single-user scenario, the region of similarity was around 20 %, which shows that the performance of the simulator is better in the multiple-user scenario.

In general, it can be argued that it is harder to simulate two persons instead of one. However, we obtained better results in the multiple-user scenario. This can be explained by three main factors:

- The monitoring time: in the case of  $DS_3$  we have around 9 months of sensor activations and annotated activities, whereas for  $DS_1$  and  $DS_2$  we only have 14 days. A longer monitoring period allows us obtaining more accurate behaviour models, which produces better simulations.
- Regularity of behaviours: although  $DS_3$  presents two occupants in the same house, their behaviour is more regular than  $DS_1$  and  $DS_2$ . Of course, this is also related to the monitoring time, since longer periods usually show better

the behaviour patterns. As MASSHA relies on behaviour models, more regular behaviours tend to make the simulation easier.

- The monitoring approach:  $DS_3$  uses motion sensors extensively to monitor the occupants' activities. As MASSHA models spatial and temporal restrictions realistically, we can specify typical trajectories when performing activities. We found that location- and motion-based monitoring approaches can be simulated accurately.



**Figure 4.11:** Distribution of the amount of not similar sensors depending on the length of the *region of similarity*. (epsilon) and *significance level* (alpha) for  $DS_3$  (multiple-user)

Following the approach to validate the single-user scenario, we decided to perform an additional test to verify whether MASSHA was able to reproduce the hourly activations of sensors. Table 4.11 shows, for every sensor in  $DS_3$  and for every hour

of the day, whether the real hourly activations of sensors matched those simulated by MASSHA. For example, for sensor *M001*, MASSHA coincided with its activations for hours  $h_7$ ,  $h_{15}$ ,  $h_{16}$ ,  $h_{18}$ ,  $h_{20}$ ,  $h_{21}$ , and  $h_{22}$ , while did not coincide with its activations for hours  $h_9$ ,  $h_{13}$ , and  $h_{14}$ . The overall activations of the real dataset and that produced by MASSHA, seem to be quite different. However, in contrast with the results for the single-user scenario shown in Table 4.10, it seems that MASSHA was able to perform better with the multiple-user dataset.

## 4.7 Conclusions

Since MASSHA is based on MAS theories and technologies, the ground for improvements is huge. One of the main limitations has to do with the behaviour being very conditioned by the definition of the *TODO* list. The inputs for MASSHA hereby described allow a great degree of flexibility in configuring user profiles, but requires thorough parameterisation. While we have demonstrated that MASSHA is able to reproduce multiple-user scenarios, it is still necessary to implement collaborative and competitive scenarios where human agents share or monopolise activities.

The next two subsections show the envisaged short-term (Section 4.7.1) and mid-term developments with the intend to overcome these limitations (Section 4.7.2).

### 4.7.1 Short-Term Developments

Among the short-term developments three main blocks have been distinguished, namely the improvement of the activity management system, developments to improve the usability of MASSHA and the developments to run experiments in complex scenarios:

- **Improve the management of activities:** at the moment, MASSHA heavily relies on the *TODO* list described in Section 4.3.3 to plan the temporal distribution of activities. The next iteration of MASSHA will test the use of Hidden Markov Models, Probabilistic Finite State Machines, Neural Networks or Support Vector Machines to model the temporal distribution of activities.

- **Developments to improve usability:** in order to easily design and deploy scenarios (intelligent environments, sensors, objects and persons), the graphical interface still needs to be improved. The objective is to provide an easy-to-use scenario editor for researchers to create, edit and remove sensors and agents, since currently, MASSHA only supports script-based definitions for those entities.
- **Developments to run experiments in complex scenarios:** MAS technologies are very suitable to run multi-users concurrent-activities scenarios, as shown in this thesis. Currently, MASSHA's person behaviour modelling scripts allow modelling overlapping activities in a natural way by using the unattended tag. Adding more active agents to implement more persons with different behaviours is also straightforward. In this sense, the performance of MASSHA in multi-user scenarios has already been shown in Section 4.5. However, we would like to model and simulate even more complex scenarios, given that the decentralised approach of MASSHA will enable the interaction among active agents, setting inter-related dependencies among the activities performed by different agents. For example, a person *A* may be performing a cooking activity in the home, while another person *B* calls him/her, because *B* has an independent behaviour. Such a case may change the activities of person *A*, who will have to decide between attending the phone call or continuing cooking. MASSHA's infrastructure is already prepared for such scenarios, so experiments have to be run in order to verify its functionality.
- **Evaluate simulated datasets with a human activity recognition model:** go a step further in validating MASSHA by assessing the similarity between the results obtained from an external activity recognition model on both the simulated and reference datasets.

One of the main challenges of model-based simulators, as identified in (Helal et al., 2012) and (Synnott et al., 2015), is to acquire accurate and meaningful activity models. Having detailed information would allow MASSHA to reach domains beyond smart homes such as road traffic (Pijoan et al., 2016) or smart grids (Kamara-Esteban et al., 2016) simulation. Some researchers propose using programming by demonstration

approaches (Helal et al., 2012), whereas some others show how surveys to target users can be used (Azkune et al., 2015). Nowadays, there is not a single perfect way for that task, but research continues in the area. As MASSHA's full potential appears specially in multi-agent environments, a system to acquire easily different activity models and behaviours has to be found. In that sense, the survey-based approach presented by Azkune et al. (Azkune et al., 2015) seems more suitable. However, surveys may be a good way to acquire the high-level behaviour of a person, but they tend to get quite unmanageable when capturing the details of specific activities.

#### 4.7.2 Mid-Term Developments

It is noticeable that term *Environment* has appeared regarding two different contexts. On the one hand, Intelligent Environments are those compound of sensors, actuators and computer systems for Human Activity Recognition that monitor some activities and try to have some beneficial influence on them. On the other hand, Agent Environments are the virtual worlds where agents live and interact together in a Multi Agent System.

Research contributions about agent environments so far have mainly focused on cases where the first-class citizens acting, perceiving, and interacting have typically been artificial agents, either software or robots. But what about agent environments where *both human and agents* interact as first-class citizens? As stated in (Ricci et al., 2015) an appropriate environment definition can bring together real world scenarios and applications by addressing key issues as time management, sensor processing, augmented reality and actuators. It is here where the two environments need to come together to form the glue that links these two areas of knowledge converging humans and agents. MASSHA aims to join these two forces creating a mixed environment driven by humans and agents where the real world is mapped to a virtual world and it is possible to spread the changes done from one to the other. Mixed environments make possible for citizens to live together with agents that assist humans providing a qualitative and quantitative leap in several sectors such as demand response, healthcare, logistics, domotics or entertainment.

**Table 4.10:** Extract of the results of the comparison MASSHA vs  $DS_1$ .

sensor name	$h_0$	$h_1$	$h_2$	$h_3$	$h_4$	$h_5$	$h_6$	$h_7$	$h_8$	$h_9$	$h_{10}$	$h_{11}$	$h_{12}$	$h_{13}$	$h_{14}$	$h_{15}$	$h_{16}$	$h_{17}$	$h_{18}$	$h_{19}$	$h_{20}$	$h_{21}$	$h_{22}$	$h_{23}$	✓ (%)	✗ (%)
Burner106	✓	✓	✗				✓	✗			✓	✗		✓	✗		✓				✓	✓	✓	✓	67	33
Burner118	✓	✗			✗	✗	✗			✗		✓	✗		✓	✗				✗	✗	✗	✗	✗	20	80
Burner94	✗	✓			✗	✗	✗	✓	✓	✗	✗	✗	✗	✓	✗						✓	✗	✗	✗	33	67
Burner99					✗	✗				✓					✗							✗			20	80
Cabinet132	✓			✗	✗	✗				✗	✗				✗	✗		✓							30	70
Cabinet138	✓		✓		✓	✓	✓				✗	✓		✓	✗						✓		✓		83	17
Cabinet51							✗	✗	✗	✗	✗	✗	✓	✗	✗			✗	✓						27	73
Cabinet53	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	9	91
Cabinet55	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	13	88
Cabinet61	✗	✗	✗				✗	✗	✓	✓			✓	✗	✗	✓	✓	✓	✓	✗	✓	✓	✗	✗	41	59
Cabinet63		✗	✗	✗	✗	✓	✓	✓		✓	✗	✗	✗	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	40	60
Cabinet66	✗	✓	✗	✗	✗	✗	✓	✗	✗		✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	13	87
Cabinet72	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	4	96
Cabinet73	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	8	92
Cabinet79	✓	✗	✗	✓	✗	✓	✗	✓		✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	38	62
Cabinet80	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	4	96
Cabinet83							✓	✓	✓	✓	✗	✗	✗	✗		✗					✓	✓			60	40
Cereal145		✓				✓	✓	✓	✓		✗	✗									✓	✓			78	22
Closet81	✓	✗	✗	✗		✗	✗	✗	✗	✗	✗	✗	✓	✗	✓	✗	✗	✗	✗	✓	✓	✗	✗	✗	26	74
Coffeemachine119	✗					✓	✓	✓	✓	✓	✓	✓									✓	✓	✓	✓	78	22
Containers60	✗		✓		✗	✗	✗	✗	✗	✗	✗	✗			✓	✗	✓	✓			✓	✓	✓	✓	44	56
Dishwasher70	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	14	86
Door140	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	4	96
Door141	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	4	96
Door54	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	4	96
Drawer125		✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	71	29
Drawer135		✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	✗	✗			✗	✗	✗	✗	35	65
Drawer146	✓	✗	✓		✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✗		✗	✓	✓			✓		✓	67	33
Drawer52	✗	✓	✗		✗	✓	✓	✓	✓	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	56	44
Drawer62	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✗	✗			✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	38	62
Drawer69		✓	✓	✗	✗		✓	✗				✓	✓		✗	✗	✓	✓				✓	✓	✓	60	40
Drawer71		✓	✓	✗	✗		✓	✗	✗		✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	22	78
Drawer75	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	4	96
Drawer78	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	40	60
Drawer82	✓	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	22	78
Drawer86		✓			✗	✗	✓	✓		✓	✗	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	50	50
Drawer87							✓		✓						✗										67	33
DVD56	✓						✗		✓						✓							✓			75	25
GarbageDisposal98		✓			✗	✗	✗	✓	✓	✓		✓			✓	✗		✗			✓	✓	✗	✗	53	47
Lamp76	✓	✓	✓				✗	✓	✓	✓	✓	✓			✓	✗	✗	✓	✓		✓	✓	✓	✗	67	33
Lightswitch104	✓	✓				✓				✓	✗				✓				✗	✓	✓		✓	✓	80	20
Lightswitch105		✓		✓				✓		✓		✓	✓			✓		✗	✓	✓		✓		✓	82	18
Lightswitch107			✓				✓	✓	✓			✓	✓			✓		✓	✓		✓	✓	✓	✗	89	11
Lightswitch108			✗		✗	✓	✓	✓	✗	✗		✓	✗			✓	✓	✓	✗	✗	✓	✓	✓	✓	57	43
Lightswitch120	✗	✗				✓	✓	✓	✗	✗		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	44	56
Lightswitch92	✓	✗	✗	✓	✗	✗	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	25	75
Lightswitch95	✓			✗		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓	81	19
Lightswitch97									✓						✓								✓	✓	80	20
Microwave143	✓	✗	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	17	83
Oven129						✗	✗	✗	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓			33	67
Refrigerator126	✓		✗	✓			✗				✓	✓	✓		✗	✗	✓	✓	✗				✓		55	45
Refrigerator144			✓			✓					✓			✗	✓			✓							86	14
Toaster131	✗	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	8	92
WashingMachine142	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	5	95
✓ (%)	32	35	27	24	13	15	26	24	28	24	25	17	29	16	32	6	27	23	18	27	23	22	24	32		
✗ (%)	67	65	72	75	86	84	73	75	72	75	74	82	70	83	67	93	72	76	81	72	76	77	75	67		

**Table 4.11:** Extract of the results of the comparison MASSHA vs  $DS_3$ .

sensor name	$h_0$	$h_1$	$h_2$	$h_3$	$h_4$	$h_5$	$h_6$	$h_7$	$h_8$	$h_9$	$h_{10}$	$h_{11}$	$h_{12}$	$h_{13}$	$h_{14}$	$h_{15}$	$h_{16}$	$h_{17}$	$h_{18}$	$h_{19}$	$h_{20}$	$h_{21}$	$h_{22}$	$h_{23}$	✓ (%)	✗ (%)
M001									✓	✗				✗	✗	✓	✓		✓	✓	✓	✓	✓	✓	70	30
M005																✗			✗	✓	✓	✓	✓	✓	50	50
M006																				✗	✗	✓	✓	✓	88	13
M007								✓	✓	✓				✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	85	15
M008								✓	✓	✓				✓	✗	✗	✓	✓	✓	✓	✓	✗	✓	✓	77	23
M009								✗	✗	✗				✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	69	31
M010																				✓	✓	✗	✓	✓	86	14
M013								✓	✓	✗						✓	✓			✓				✓	67	33
M014													✓	✓	✗	✗	✗	✗	✗	✓				✓	64	36
M015								✗	✗	✗				✓	✓	✓	✓	✓	✗	✗	✓			✓	75	25
M016								✗	✗	✗				✓	✗	✗	✓	✓	✓	✓	✓	✗		✓	54	46
M017								✓	✗	✗				✓	✓	✗	✓	✓	✓	✗	✗			✗	67	33
M018								✗	✓					✓	✗	✗	✓	✓	✓	✓	✗	✗		✗	55	45
M019								✓	✓	✗				✗	✗	✗	✓	✓	✓	✓	✗	✓		✗	45	55
M021								✓	✓	✓				✓	✗	✗	✓	✓	✓	✗	✗	✓	✗	✓	69	31
M022								✓	✓	✓				✓	✗	✗	✗			✓				✓	60	40
M024				✗				✓	✓	✓				✓	✗	✗	✗	✗	✗	✓	✓	✗	✗	✓	47	53
M025								✓	✓	✓				✗	✗	✗	✓	✓	✗	✓	✓	✓	✓	✗	57	43
M026								✓	✓	✓				✗	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	79	21
M027								✓	✓	✓				✗	✓	✓	✓	✓	✓	✗	✗	✓	✗	✓	71	29
M028	✗							✓	✗	✓				✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✗	69	31
M030								✓	✓	✓				✓	✓	✓	✓	✓	✓		✗	✗		✓	91	9
M032								✗	✓	✓				✓	✓	✓	✓	✓		✗	✓		✓	✓	82	18
M033								✗	✓	✓				✓	✗	✗	✗			✓	✓		✗	✓	54	46
M034	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓				✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	65	35
M035		✓	✗	✓	✓	✓	✓	✓	✓	✓				✓	✓	✓	✗			✓	✗		✗	✓	75	25
M036								✓	✓	✓				✓	✓	✓	✓	✓	✓	✗	✓	✓	✗	✗	73	27
M037	✓							✓	✗	✓				✓	✓	✓	✓	✓	✓	✗	✗		✓	✓	73	27
M038	✓							✓	✗	✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	91	9
M039				✗				✓	✓	✓				✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	67	33
M040								✓	✓	✓	✗			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	56	44
M041								✗	✗	✗				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	67	33
M043	✓							✓	✓	✓				✓	✓	✗	✗		✗					✓	60	40
M044	✗							✓	✓	✓	✗			✓	✓	✓	✓	✓	✓	✓				✓	78	22
M045	✗	✓		✓	✗	✗		✓	✓	✓				✓	✓	✗	✓	✓	✓			✗	✗	✓	50	50
M046	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓				✓	✓	✓	✓	✓	✓	✗				✓	77	23
M047	✗	✗		✓	✗	✓	✓	✓	✓	✓				✓	✓	✓	✓	✓	✓			✗		✓	64	36
M048	✓	✓						✓	✓	✓				✓	✗	✗	✓	✓	✓					✗	67	33
M049	✓	✓				✗	✗	✗	✗	✓				✓	✓	✗	✓	✓	✓			✗		✓	45	55
M050	✓							✓	✓	✓				✓	✓	✗	✓	✓	✓					✓	80	20
M051								✗	✗					✓	✓	✓	✓	✓	✓				✓	✓	75	25
✓ (%)	58	71	50	40	80	50	81	63	96	70	0	0	0	84	63	54	71	87	67	55	68	67	67	68		
✗ (%)	42	29	50	60	20	50	19	37	4	30	100	100	100	16	37	46	29	13	33	45	32	33	33	32		



*The first rule of any technology used in a business is that automation applied to an efficient operation will magnify the efficiency. The second is that automation applied to an inefficient operation will magnify the inefficiency.*

Bill Gates

CHAPTER

# 5

## **Domestic DSM: Middleware for implementing strategies**

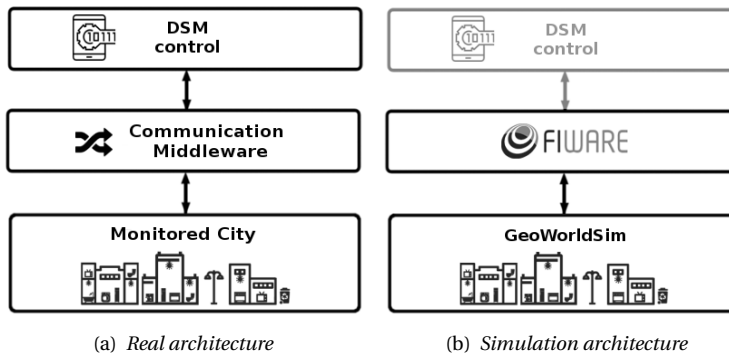
**D**OMESTIC ELECTRIC-BASED DSM requires sufficient controllable energy use, i.e. typical loads that are more than 400W and DSM programmes that involve over 100,000 homes. The accurate definition and application of these type of strategies rely heavily on the capacity of getting the most relevant information regarding the electricity profile of the users, specially the mobile loads, which events can be deferrable.

This chapter presents how we have used the Smart City Simulator, developed in Chapter 3 to integrate the development of a human activity model developed in Chapter 4 with the communication middleware known as FI-WARE in order to test the best architectures available for the implementation of DSM strategies.

## 5.1 Introduction

Figure 5.1 shows the ideal architecture of a DSM scenario and how we have mapped each element to our simulation approach. The bottom layer of the architecture represents a city which energy demand is constantly monitored. Information regarding the energy consumption of devices within the city is sent to a common communication middleware which centralises all the information, registering all the events and notifying them to the DSM Control. The DSM Control analyses all the data received and provides a set of signals back to the monitored city so as to influence citizens behaviour with the objective of leveraging energy demand and energy production.

Following these schema, our approach consists on developing an agent-based system (GeoWorldSim) that emulates the behaviour of a city in terms of energy consumption, and evaluates the feasibility of deploying several communication architectures depending on the amount and type of devices connected to the communication middleware chosen, in this case FI-WARE. The development and implementation of the DSM Control will be carried out in future research.



**Figure 5.1:** Real and simulation architectures for the DSM strategy scenario.

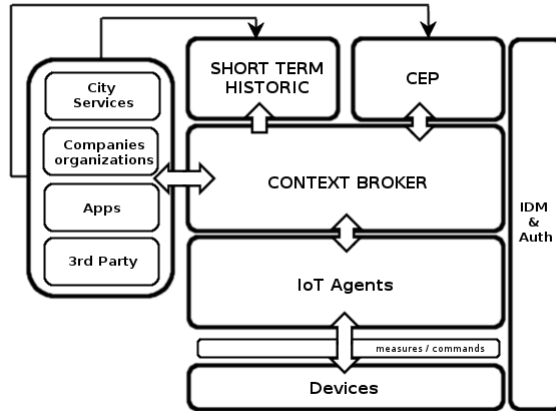
### 5.1.1 FI-WARE Middleware

GeoWorldSim relies on FI-WARE as the central middleware for creating a link between the monitored devices and the core of the DSM Control. FI-WARE is a new cloud infrastructure created by the European Commission and several major European ICT companies for the development and global deployment of services and applications of the Future Internet (Consortium, 2016). FI-WARE is based on a set of modules, known as General Enablers (GE), which provide services such as shared cloud computing resources, big data analysis, data management, modules for integrating web applications to the FI-WARE infrastructure, and security control.

Figure 5.2 depicts a general overview of FI-WARE's architecture, with the central core of the solution being the Orion Context Broker GE. This GE is a C++ implementation of the Publish/Subscribe mechanism, providing interfaces to the NGSI9 and NGSI10 API's with services such as registering context producer applications (e.g. sensor humidity within a room), update control information (e.g. send updates about changes in humidity), subscription and notification on changes on context information (e.g. when the humidity changes) or with a given frequency (e.g. retrieve humidity values every hour), and query context information about registered applications.

The Context Broker has the ability of managing Context Information at large scale by keeping virtual representations of physical devices. The interaction with the devices takes place on the event of modifying or updating their virtual representations in the Context Broker. When a petition of registration or update is performed, the APIs map each device to a specific Context Provider in order to retrieve all the necessary information through a Uniform Resource Identification protocol. Device types are grouped in domains according to their hierarchical schema and queries can be done in less or more deepness of the hierarchy. Section 5.1.1 describes in detail this hierarchical representation and the messages needed for communication.

According to FI-WARE documentation, the proper way to connect any device to the platform is through the use of an entity called IoTAgent which bridges system specific communication with FI-WARE middleware, regardless of the protocol and



**Figure 5.2:** FI-WARE architecture (Consortium, 2016).

data representation. Specifically, in our system, we have implemented a particular IoTAgent in charge of binding the REST API synchronous calls of the Context Broker to the asynchronous and non-blocking Qt Signal and Slots communication architecture.

### **Messaging format**

Each FI-WARE message must contain a minimum set of properties for the Context Broker to be able to identify the service, domain and device the message belongs to. Therefore, the smallest piece of information to be sent will on average be of 180 bytes.

Additionally, messages can describe other entity attributes, values and control commands. Each attribute must contain a name and type of value. An example of such control command can be seen in Source Codes 5.1–5.3.

Entities are registered in FI-WARE Context Broker using HTTP POST operations and sending the entire headers and payload description. Updates, however, are reported using a HTTP PATCH operation and just sending the headers together with the attributes that need to be updated, reducing the size of the bandwidth needed.

```
1 {
2   "content-type": "application/json",
3   "X-Auth-Token": "[TOKEN]",
4   "Fiware-Service": "HouseManager",
5   "Fiware - ServicePath ": " / "
6 }
```

**Code 5.1:** Header example.

```
1 {
2   "devices": [{
3     "device_id": "[DEV_ID]",
4     "entity_name": "[ENTITY_ID]",
5     "entity_type": "thing",
6     "protocol": "IoTA-UL",
7     "timezone": "Europe/Madrid",
8   }]
9 }
```

**Code 5.2:** Payload example.

## 5.2 Materials and Methods

In order to assess the advantages and drawbacks of several architectures in regards of implementing a DSM solution in a Ubiquitous World, at city level, we propose the analysis of two middleware architectures (Centralised and Federated) and two publication strategies (Periodical and Event-based):

- **Centralised Architecture.** As seen in Figure 5.3, this architecture consists of a single Context Broker. This configuration provides an easy management and a clear monetising of the services provided. However, since all the users share the same Context Broker instance, the bandwidth requirements are very high,

```
1  {
2    {
3      "device_id": "[DEV_ID]",
4      "entity_name": "[ENTITY_ID]",
5      "entity_type": "thing",
6      "timezone": "Europe/Paris",
7      "endpoint": "http://{{DEV_IP}}:{{PORT}}",
8      "attributes": [
9        {
10         "object_id": "t",
11         "name": "temperature",
12         "type": "int"
13       }
14     ],
15     "commands": [
16       {
17         "name": "ping",
18         "type": "command",
19         "value": "[Dev_ID]@ping|%s"
20       },
21     ]
22   }
23 }
```

**Code 5.3:** Control message example.

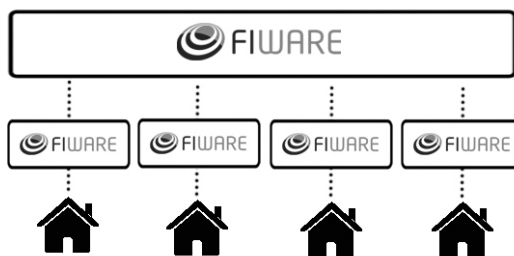
as seen in Section 5.3, the resilience to failures decreases significantly, and the process to guarantee privacy to citizen data grows in complexity.

- **Federated Architecture.** In this architecture, each house has its own Federated Context Broker instance (FCB), that in turn is connected to another Context Broker instance on a higher hierarchical level. The low level Context Broker stores and aggregates data from the house's appliances and publishes them upstream. It also transmits other events relevant to the DSM strategy, such as the need to schedule loads that can be *delayed*, like the operation of washing machines, dishwashers or laundry driers. Urban equipment like streetlights



**Figure 5.3:** Centralised architecture schema

and waste containers publish to their own private low level Context Broker which also sends the data to the high level Context Broker. Figure 5.4 shows an example of the Federation Architecture where several Context Brokers form a hierarchy. Our results suggest that the use of a single high level Context Broker gives solutions to several of the drawbacks of the Centralised Architecture. For example, a) the privacy of citizen data is better preserved since there is a less amount of data travelling outside the home network (Symantec Security Response, 2016); b) the architecture is more resilient: if one Context Broker malfunctions, the rest of the network can continue operating normally; c) the bandwidth need is several orders of magnitude lower. From a service provider company, however, this system is less attractive in terms of exploitation, due to the complexity of its management.



**Figure 5.4:** Federation architecture schema

- **Periodical Publication Strategy.** In this strategy, all sensors publish their state or consumption over a fixed period of time. This strategy makes the development of DSM controls simpler and similar to real time monitoring. However, the energy consumption of the devices and the bandwidth requirements increase significantly and become unmanageable for large sets of houses or appliances.
- **Event based Publication Strategy.** In this strategy, the sensors only publish data when their state changes (e.g. switched on/off, opened/closed, etc). When a sensor is switched off, in addition to notifying this change, it also provides information about the power consumption occurred during the time the associated appliance was operative. This approach solves the drawbacks of the periodic publication strategy but increases the complexity of the DSM control since these sensor events need to be mapped to particular consumption patterns.

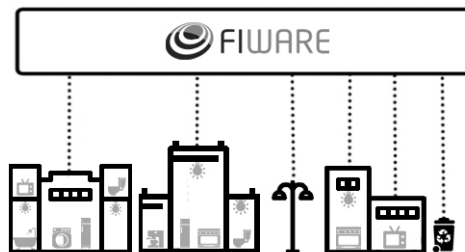
We also identified three future scenarios that pave the way to a fully Ubiquitous World on which to test the previous architectures and publication strategies.

- **Smart City (SC).** In this scenario we will simulate the traditional vision envisaged by the managers of the myriad of distribution grids that feed a city. In this sense, houses periodically provide information about the aggregated load (whether it is water, energy or amount of waste) for billing purposes. The publication frequency is fairly low, which makes it feasible to connect the publication feed to a single centralised Context Broker. The federation architecture has not been tested in this scenario, since it would be an over-engineering example. Figure 5.5 shows a graphical representation of this scenario.
- **IoT Jungle (IOTJ).** Within this scenario, all the IoT devices within a house publish information to a centralised Context Broker. This publication can be performed both periodically or event based triggered by a change in the state of the device. Depending on the publishing method, the bandwidth needs and size of published message may vary since, for example, event based mes-

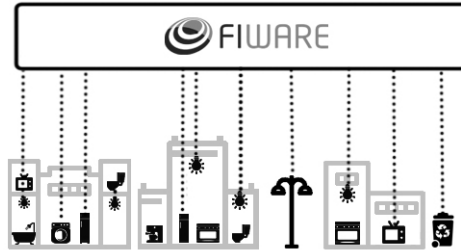
sages need to send bigger pieces of information. Figure 5.6 shows a graphical representation of this scenario.

- **Ubiquitous World (UW).** This scenario combines the advantages of the SC and IOTJ scenarios, a vision envisaged by IoT researchers. In this sense, all the devices within a house are connected to a private FCB. This Context Broker is in charge of storing and aggregating all the data received from the devices. This aggregation is then published to a centralised Context Broker. Urban equipment also publish to their own FCB. Each FCB, in addition, publishes other events relevant to the DSM strategy, such as the need to schedule loads that can be *delayed*. Figure 5.7 shows a graphical representation of this scenario. This architecture is fully aligned with the *Edge Computing* paradigm (Skala et al., 2015) in which services are moved from centralised nodes to the extremes of the network where the information is generated.

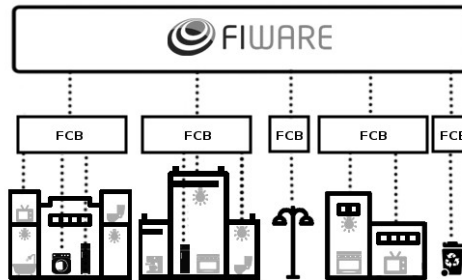
The aforementioned scenarios only cover the Periodical Publication Strategy for the first case. The reason is that the amount of messages increases with the number of houses and quickly flow the network driver. Simple scenarios could require more than 250 MB/s if the publication period is short. If the publication period is too large, this strategy loses its purpose.



**Figure 5.5:** Diagram of the Smart City scenario



**Figure 5.6:** Diagram of the IoT Jungle scenario



**Figure 5.7:** Diagram of the Ubiquitous World scenario

### 5.3 Results and discussion

The simulation setup is composed of an Acer TravelMatep238 laptop Core i5 2.3 GHz with 16GB of RAM using a Gigabit Ethernet class connection to reach the cloud server (Education, 2016) (where GeoWorldSim is being run) and the FI-WARE-LAB cloud environment (FI-WARE, 2016) (where the middleware is deployed). The FI-WARE-LAB cloud environment provided a working production instance of the FI-WARE stack without any Service Level Agreement. As in any other cloud environment, the virtual instance is under our control (namely, we can guarantee that no one else is using that instance) but we share virtual machines with several other users (we cannot guarantee the amount of bandwidth at our disposal nor the overall load of the

physical server). In order to assess the different middleware architectures, we defined two Key Performance Indicators (KPI) of interest:

- **Bandwidth.** Number of bytes transmitted per unit of time. In this particular example, we have recorded the number of bytes per message without including the overheads related to the specific protocols used to transport the information. Our objective is to estimate the real bandwidth needs in a real deployment without assuming a simultaneity factor. Note that this value is specially relevant in the Smart City scenario, as the deployment is traditionally performed over SIGFOX or LoRa networks that have limited bandwidth (Adelantado et al., 2016).
- **Number of Messages Lost:** number of messages emitted by GeoWorldSim that are not processed by the Context Broker because they may get lost during the transmission. This value gives information about the quality of the service provided by the architecture. In our experiments, none of the transmitted messages were lost due to the fact that the Context Broker uses an asynchronous database storage, and therefore, during high load peaks, message insertion gets asynchronously scheduled. This is a good indicator of the reliability of the solution.

In order to accelerate the procurement of results and balance both the length of the configuration file and RAM consumption, the simulation time was accelerated. We can define that 1 second of the simulation corresponds to  $n$  second in terms of real time.  $n$  have been taken accordingly to ease the deployment of the simulation. Modification in  $n$  has a direct impact in the relation between the measured KPIs and the real KPIs (i.e. if  $n = 1$  have been used). In this sense, an speedup factor of  $n$  would be equivalent to:

- multiply by  $n$  the number of houses
- divide by  $n$  the bandwidth

Several city sizes were tested for each of the scenarios defined in 5.2, specifically, Village (250 houses), Small City (2500 houses), Medium City (25 000 houses), Large

City (250 000 houses) and Megalopolis (2 500 000 houses). Each house accommodates two persons that interact with 18 smart devices while performing their daily duties. In addition, the scenarios were provided with ( $5 \times \text{\#number of houses}$ ) streetlights and ( $1 \times \text{\#number of houses}$ ) recycling containers.

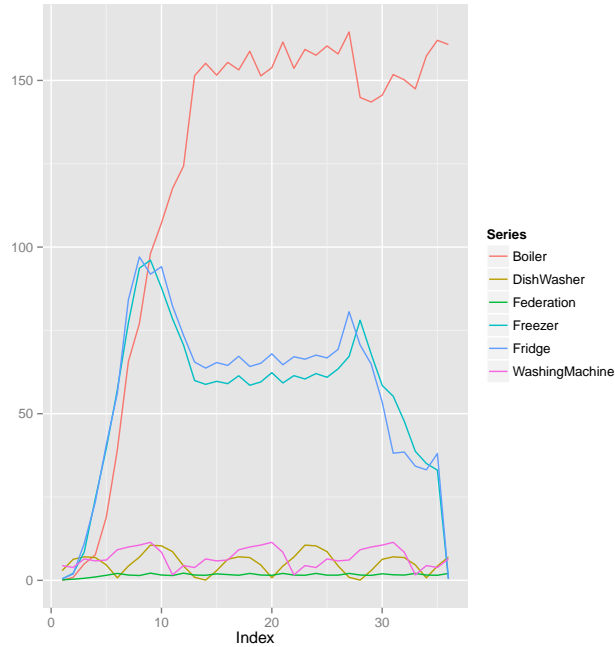
The simulation has been carried out for 6 hours, the busiest of the day, that is 3 in the morning and 3 in the evening:

- **Morning.** This portion of the simulation covers the time ranging from the moment the citizens wake up to the moment they go to work. The appliances activated are those related to the activities that involve getting ready for going outside (showering, dressing, grooming), and preparing breakfast. The number of appliances used is rather low but all of them are used almost at the same time.
- **Evening.** This portion of the simulation covers the period of time between having dinner and going to bed. This period includes the prime time. The appliances activated range from the ones typically used in the kitchen to the ones used for entertainment. In this time period, the number of appliances used is high and diverse but their activation is more spread over time.

Since each simulation run is done over a long period of time, we consider that there is no need to repeat the experimentation due to the fact that all the possible variability that may appear will probably be already captured. A video of a complete simulation can be found in <http://bit.ly/2fbX54p>.

For each simulation we register two logs: the timestamp and types of message sent by GeoWorldSim and the equivalent log from the FI-WARE Context Broker so as to easily extract the KPIs values. For example, Figure 5.8 shows a graphical representation of the number of messages per device transmitted during the simulation time. The bandwidth is measured by multiplying the number of messages by their length and dividing the result by the unit of time.

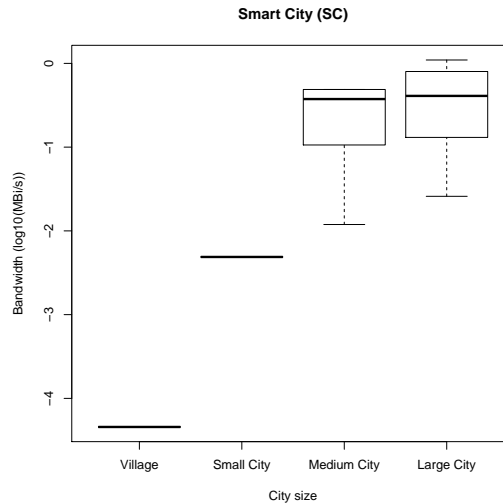
Figure 5.9 shows several boxplots (Benjamini, 1988) depicting the bandwidth distribution used per number of houses in the Smart City scenario. Please note that this is a log-log plot. As can be seen, the bandwidth grows and its saturation takes



**Figure 5.8:** Evolution of the number and type of messages per time in a simulation.

place at the *Medium City* level, with a saturation bandwidth around 1 MB/s. Take into consideration that this scenario is the only one where the amount of messages allows to perform a simulation of the Megalopolis. Clearly, even though we were able to perform this simulation, the computational needs superseded the bandwidth restrictions of this case.

The IoT Jungle represents the most problematic scenario. Figure 5.10 shows, as in the previous case, several boxplots describing the distribution of the bandwidth used per size of the city. As can be seen, the saturation is reached under the same case *Medium City*, but it uses 10 times more bandwidth than the Smart City scenario. However, in this case, it seems that the bandwidth demands of the *Large City* have produces a change of phase and the bandwidth collapses (Al-Bahadili, 2012).

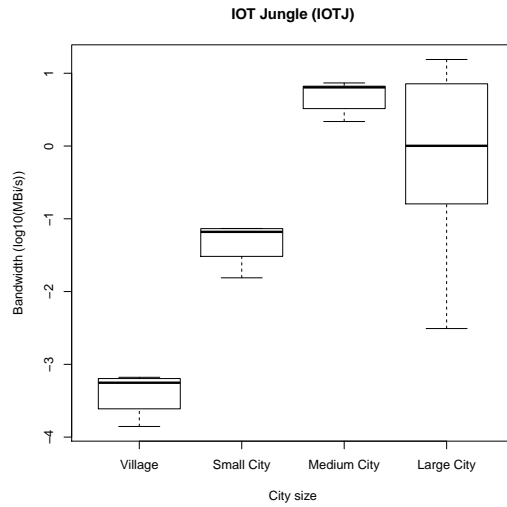


**Figure 5.9:** Bandwidth distribution per number of houses in the Smart City scenario.

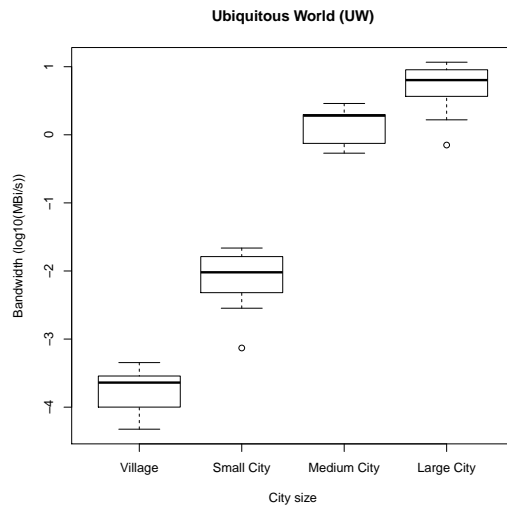
Federation architectures need to have several Context Brokers operating simultaneously. We only have access to a single instance of a Context Broker, therefore we cannot directly assess this architecture. However, we can perform the simulation in two steps. First, we simulate the behaviour on a node on the federation (a house) and register all the messages that should have been sent to the federation. Then we repeat this process for every node. Finally, we replay the message record from the nodes to the centralized Context Broker. Figure 5.11 shows the bandwidth used per number of houses in the Ubiquitous World scenario. In this case the saturation takes place in the *Large City* case, in contrast with the previous two scenarios.

## 5.4 Conclusions

Results suggest that the most suitable architecture to carry on DSM controls at city level is the event based Federation (Ubiquitous World Scenario). Even though the control strategies to be built will be more complex, the bandwidth demand is several



**Figure 5.10:** Bandwidth used per number of houses in the IoT Jungle scenario.



**Figure 5.11:** Bandwidth used per number of houses in the Ubiquitous World scenario.

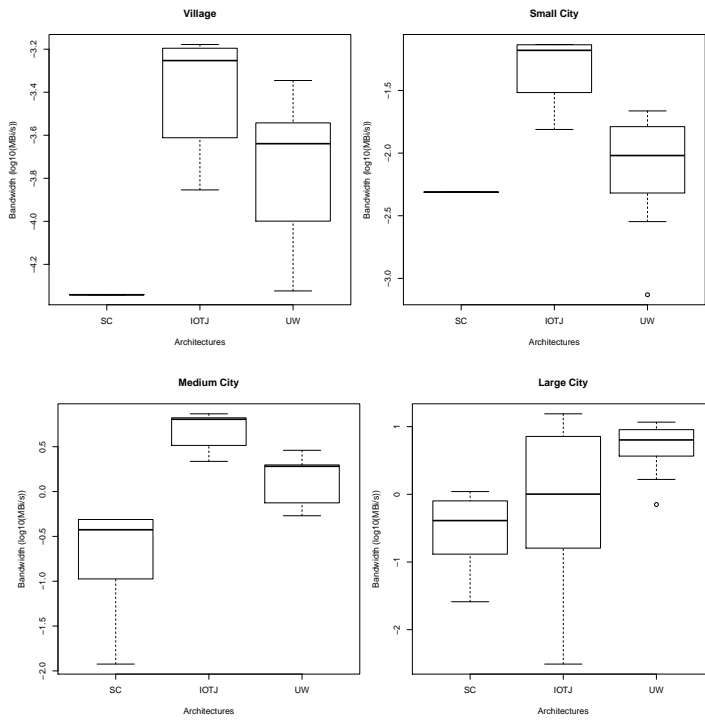
**Table 5.1:** Results of the Kruskal-Wallis with a Conover post hoc statistical test.

Test result	Village	Small City	Medium City	Large City
Kruskal-Wallis	$1.4 \times 10^{-6}$	0.0037	$9.4 \times 10^{-8}$	0.00093
SC-IOTJ	$2.0 \times 10^{-11}$	0.00087	$2 \times 10^{-16}$	<b>0.064</b>
SC-UW	$1.1 \times 10^{-7}$	<b>0.11</b>	$1.5 \times 10^{-9}$	0.00017
IOTJ-UW	0.0019	0.011	$1.1 \times 10^{-8}$	0.028

orders of magnitude lower and the overall architecture is both more secure and resilient.

Figure 5.12 shows the boxplots comparing the bandwidth needs for the three architectures in the four scenarios. Table 5.1 shows the p-values of a Kruskal-Wallis test with a Conover post hoc (Pohlert, 2014). Several conclusions can be drawn out from this information:

- The bandwidth needs for the Smart City scenario are statistically significant lower than those of the other architectures in almost all the scenarios but in turn, this architecture provides less information.
- The bandwidth needs for the IoT Jungle are statistically significant higher than those of the other architectures in almost all scenarios but the amount of information that this architecture provide is similar to that provided by the Ubiquitous World architecture.



**Figure 5.12:** Comparison of the bandwidth needs of the architectures per scenario.



*In 80% of the world, energy will be bought where it is economic. You have to help the rest of the world get energy at a reasonable price.*

Bill Gates

CHAPTER

# 6

## Commercial DSM : Optimisation of the electric energy tariff

**T**HE ECONOMIC CRISIS has motivated the need to audit all expenses in order to identify the key areas where saving techniques can be applied to reduce global costs. Nowadays, most of these techniques focus on minimising costs within the electric scope, not only in terms of reducing the overall consumption, but also in configuring an specific electric contract that meets the energy needs of the infrastructure while avoiding overruns.

The knowledge on how the energy is consumed within an infrastructure provides a great advantage when negotiating contracts with the electricity supplier. For this, most companies turn to *energy managers* whose labour is to regulate and monitor energy consumption with the aim of improving its efficiency by implementing new

policies and changes where necessary. They are responsible for knowing in detail all the variables that come into play within the electric market such as the methodology for the calculation of electricity prices and the advantages and disadvantages of the current electric tariffs.

Electricity contracts can be challenging to understand. With so many new electric companies entering the market and unclear terms and conditions, residential consumers and businesses alike need to carefully understand what type of service they need and how to get the best rates. The billing concepts are not always understandable for an average user, which together with the inclusion of several taxes of different type hinders its comprehension. The result of this misinformation is that most companies choose contracts based on what a sales representative from the electric company advised them rather than taking into account their own energy needs. Therefore, the optimisation of the energy bill requires a technical and thorough analysis of the parameters that influence it, the facility's consumption patterns, and the possibilities of negotiation within the electric market.

The accurate definition and implementation of electricity tariffs is one of the most used strategies in the commercial DSM. In this sense, the optimisation of the electricity tariff is the first step towards the influence of the end user electricity consumption. This chapter presents the the design and development of a methodology that adjusts the terms of the electrical tariff of a university campus based on historical consumption data. The analysis estimates different alternatives of pricing, depending on the use of individual or common tariffs.

## 6.1 Introduction

The electricity bill in Spain defines a series of terms that can be classified into regulated and non-regulated.

- **Regulated terms.** These include a series of concepts which are common and applicable to *every* electricity supply contract in Spain, regardless of the supplier. The scope and methodology of calculus are defined in the Boletín Oficial del Estado (BOE), the official gazette of the Government.

- **Power term.** When negotiating an electric contract, companies choose the type of tariff that better suits their needs, depending on the voltage of their electric infrastructure and the amount of power they expect to consume. The available tariffs are summarised in Table 6.1. Some of these tariffs split the days in *periods* which are assigned an annual power term price by the BOE. In order to calculate the monthly bill, the amount of contracted power in every period is multiplied by the proportional price for the days invoiced.
  - **Power excess.** Penalty applied each time the company demands more power than the amount originally contracted.
  - **Rental of equipment:** Amount paid to the electricity supplier for the use of the measurement and control equipment. This is a fixed rate.
  - **Electricity tax and Value Added Tax (VAT).** Tax that applies to the total of the power term, the energy term and the supplement for reactive energy.
- **Non-Regulated terms.** These include a series of concepts, common to all the electricity bills, which price can be negotiated between customer and supplier and, therefore, can vary among contracts.
    - **Energy term.** The price of energy varies depending on the hour of the day and month. The energy term of the monthly bill is calculated by multiplying the price of energy negotiated with the electricity supplier by the amount of energy measured by the meter.
    - **Complement for consumption of reactive energy.** The billing of high electricity consumers in the industrial and commercial sectors includes a charge for the generation of reactive energy. Its purpose is to promote EE by issuing an economic penalty for inefficient consumption. This inefficiency is billed according to the power factor, an indicator that quantifies an entire electric infrastructure regarding the total active energy consumed in comparison to the amount of electric energy supplied. The complement for reactive energy is calculated by multiplying the measured values of reactive energy by the penalisation negotiated with the

electricity supplier. The easiest way to avoid this penalisation consists on the installation of capacitor banks.

**Table 6.1:** Electric tariffs defined by the Spanish BOE (Boletín Oficial del Estado, 2014).

Tariff	Voltage ( $v$ )	Power ( $p$ )	Periods
2.0A	$v < 1 \text{ kV}$	$p < 10 \text{ kW}$	1
2.0DHA	$v < 1 \text{ kV}$	$p < 10 \text{ kW}$	2 <sup>†</sup>
2.0DHS	$v < 1 \text{ kV}$	$p < 10 \text{ kW}$	3 <sup>†</sup>
2.1A	$v < 1 \text{ kV}$	$p < 10 \text{ kW}$	1
2.1DHA	$v < 1 \text{ kV}$	$10 \text{ kW} \leq p < 15 \text{ kW}$	2 <sup>†</sup>
2.1DHS	$v < 1 \text{ kV}$	$10 \text{ kW} \leq p < 15 \text{ kW}$	3 <sup>†</sup>
3.0	$v < 1 \text{ kV}$	$15 \text{ kW} \leq p < 450 \text{ kW}$	3 <sup>‡</sup>
3.1	$1 \text{ kV} \leq v < 36 \text{ kV}$	$15 \text{ kW} \leq p < 450 \text{ kW}$	3 <sup>‡</sup>
6.1	$1 \text{ kV} \leq v < 36 \text{ kV}$	$450 \text{ kW} \leq p$	6 <sup>§</sup>
6.2	$36 \text{ kV} \leq v < 73 \text{ kV}$	$450 \text{ kW} \leq p$	6 <sup>§</sup>
6.3	$73 \text{ kV} \leq v < 150 \text{ kV}$	$450 \text{ kW} \leq p$	6 <sup>§</sup>
6.4	$150 \text{ kV} \leq v$	$450 \text{ kW} \leq p$	6 <sup>§</sup>

<sup>†</sup> Each period is assigned an equal number of hours per day in summer and in winter. The power limit contracted is the same for all periods.

<sup>‡</sup> Each period is assigned an equal number of hours per day in summer and in winter. The power limit contracted for every period can vary, the only condition is that these amounts have to be configured in ascending order.

<sup>§</sup> Each period is assigned different hours and days throughout the year. The power limit contracted for every period can be different, though at least one of those values should be greater than 450 kW and they have to be configured in ascending order.

Taking into account the previous terms of the electricity bill, the optimisation of the contracted electricity tariff can be achieved in three different ways.

- **Optimisation of the power term.** This methodology focuses on adjusting the amount of contracted power to the real amount of power expected to be demanded by the infrastructure. The main objective of this approach is the minimisation of penalties for excess in the amount of power demanded.

- **Optimisation of the reactive term.** This methodology focuses on compensating the generation of reactive energy, derived from the use of some electrical and mechanical devices, with the installation of the appropriate number of capacitor batteries.
- **Reduction of power and energy consumed:** This methodology focuses on the implementation of policies and best practices to promote EE, such as replacing lighting fixtures and electric equipment with much more efficient alternatives.

The configuration of an appropriate electricity contract is as important as knowing the specific needs and energy capabilities of the infrastructure. The BOE establishes the voltage steps, power ranges, and billing conditions that characterise each of the available electricity tariffs. As shown in Table 6.1, these tariffs are classified into two groups: low voltage rates ( $v \leq 1$  kV), for small and medium energy consumers, and high voltage rates ( $v > 1$  kV) for large industrial consumers with high energy needs. The selection of the proper electricity tariff is determined by the amount of voltage on which the infrastructure is connected to the electric grid and the minimum and maximum amount of power expected to be demanded.

We will focus on the optimisation of the regulated term for the 6.1 high voltage electricity tariff, though the optimisation methodology is generic enough to be applicable to any of the other tariffs. The 6.1 tariff splits the hours of the days in periods, as shown in Table 6.2. The configuration of the period calendar is heavily determined by the global and seasonal annual energy demand within the country. For example, the hours belonging to  $P_1$  are the ones in which the global electricity demand reaches its maximum peak, so the cost of demanding power or consuming energy on this period has a higher cost than doing so on  $P_6$ . The seasonality is heavily determined by the economic activity of the region, that is why the period calendar is different for peninsular Spain than for any of the islands.

When contracting energy supply under this type of tariff, the energy manager of the infrastructure has to estimate the global power  $p_i$  expected to be demanded on each period  $P_i$  in ascending order, as seen in the Inequality (6.1).

$$p_1 \leq p_2 \leq p_3 \leq p_4 \leq p_5 \leq p_6. \quad (6.1)$$

**Table 6.2:** Distribution of annual hours within the 6-period tariff.

From	To	Jan	Feb	Mar	Apr	May	Jun <sub>1</sub>	Jun <sub>2</sub>	Jul	Aug	Sep	Oct	Nov	Dec
00:00	01:00	6	6	6	6	6	6	6	6	6	6	6	6	6
01:00	02:00	6	6	6	6	6	6	6	6	6	6	6	6	6
02:00	03:00	6	6	6	6	6	6	6	6	6	6	6	6	6
03:00	04:00	6	6	6	6	6	6	6	6	6	6	6	6	6
04:00	05:00	6	6	6	6	6	6	6	6	6	6	6	6	6
05:00	06:00	6	6	6	6	6	6	6	6	6	6	6	6	6
06:00	07:00	6	6	6	6	6	6	6	6	6	6	6	6	6
07:00	08:00	6	6	6	6	6	6	6	6	6	6	6	6	6
08:00	09:00	2	2	4	5	5	4	2	2	6	4	5	4	2
09:00	10:00	2	2	4	5	5	3	2	2	6	3	5	4	2
10:00	11:00	1	1	4	5	5	3	2	2	6	3	5	4	1
11:00	12:00	1	1	4	5	5	3	1	1	6	3	5	4	1
12:00	13:00	1	1	4	5	5	3	1	1	6	3	5	4	1
13:00	14:00	2	2	4	5	5	3	1	1	6	3	5	4	2
14:00	15:00	2	2	4	5	5	3	1	1	6	3	5	4	2
15:00	16:00	2	2	4	5	5	4	1	1	6	4	5	4	2
16:00	17:00	2	2	3	5	5	4	1	1	6	4	5	3	2
17:00	18:00	2	2	3	5	5	4	1	1	6	4	5	3	2
18:00	19:00	1	1	3	5	5	4	1	1	6	4	5	3	1
19:00	20:00	1	1	3	5	5	4	2	2	6	4	5	3	1
20:00	21:00	1	1	3	5	5	4	2	2	6	4	5	3	1
21:00	22:00	2	2	3	5	5	4	2	2	6	4	5	3	2
22:00	23:00	2	2	4	5	5	4	2	2	6	4	5	4	2
23:00	00:00	2	2	4	5	5	4	2	2	6	4	5	4	2

The traditional method (TM) for power optimisation most commonly used by energy managers nowadays consists on analysing the power consumption of the installation and identifying the maximum amount of power demanded on each one of the six periods. These values are then altered to fulfil the Inequality (6.1) and comply with the power term configuration of the electricity tariff. This approach significantly reduces the cost derived from the power term. However, there is great margin of improvement since the existence of power peaks can overstate the power requirements of the facility, thus incurring in high penalties for power excesses.

In this scenario, we propose the use of evolutionary algorithms as a technique to optimise the regulated term of the electricity bill by finding the best combination of power values and power excesses that can minimise the annual electricity cost of a commercial infrastructure. Specifically, we will focus on the optimisation of the

power term, since the rest of regulated concepts in the electricity bill are fixed and do not have any range for variation. Aside from the evident economic benefits, the accurate adjustment of the contracted power to the real needs of the infrastructure helps the overall management of the power networks by reducing the amount of power needed to be supplied at each moment. Therefore the importance of electricity pricing as one of the pillars of DSM.

## 6.2 Analytical optimisation of the 6.1 electric tariff

The mathematical properties of the power term and excesses defined in the regulated term of de 6.1 electricity tariff are analysed below.

Let  $p := (p_1, \dots, p_6)$  denote the power limit assigned to every period  $P_i$ . Please note that  $p$  is a vector of integers that fulfil the Inequality 6.1 with at least one of its values greater than 450 kW, see Table 6.1. On the other hand, let  $\pi_j^{i,m}$  denote the maximum power demanded in the  $j$  quarter (which corresponds to a specific hour in period  $P_i$  on month  $m$ ). Now, for every month  $m$ , the cost function  $\phi_m$  summarising the relationship between the power term and excesses is:

$$\phi_m(p) := \sum_{i=1}^6 \left( c_i p_i + 1.4064 k_i \sqrt{\sum_{j \in J} (\pi_j^{i,m} - p_i)^2} \right),$$

where  $c_i$  denotes the power cost of the period  $P_i$ ,  $k := (1, 0.5, 0.37, 0.37, 0.37, 0.17)$  is a vector of coefficients used to assign importance to power excesses on the most critical periods (both two latter quantities already defined in the BOE), and  $J$  denotes the set of quarters where the power measured  $\pi_j^{i,m}$  is higher than the power limit  $p_i$  contracted for this period. Namely,  $J$  is the set of quarters  $j$  such that:

$$p_i < \pi_j^{i,m}. \quad (6.2)$$

This function can have as much as 35 000 jump discontinuities depending on the cardinality of  $J$ . This results in a highly non-continuous, non-linear integer optimisation problem in which the use of classical integer programming algorithms

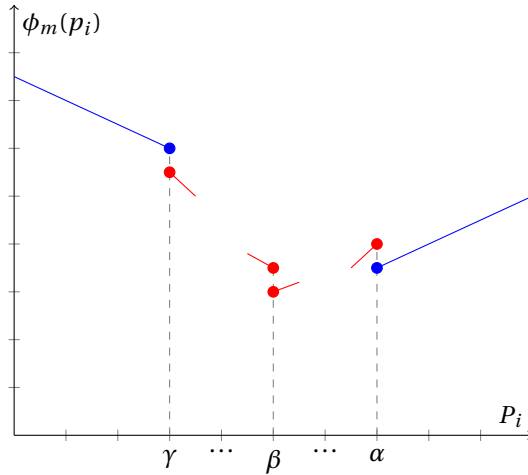
is not feasible. A graphical approach to the problem gives us additional insights about its complexity. As seen in Figure 6.1,  $\phi_m(p_i)$  is a monotonic linear function that increases steadily until one  $p_i$  fulfils Inequality (6.2) ( $\alpha$  label in Figure 6.1). In this situation we have the first jump discontinuity. Now,  $\phi_m(p_i)$  could be monotonic decreasing or increasing depending on whether  $c_i$  is bigger or not than  $1.4064k_i$  (€/kW of power excess, also defined in the BOE). The latter case means that the minima has been reached and the process can stop. In the former case, however, the process should continue until a quarter (or a discontinuity jump)  $k \in J$  is found so that:

$$c_i \leq 1.4064k_i \frac{\sum_{k \geq j} (\pi_k^{i,m} - p_i)}{\sqrt{\sum_{k \geq j} (\pi_k^{i,m} - p_i)^2}}.$$

Note that this process only finds the best set of  $p_i$  for a particular month. Since these coefficients are the same for all the months of the year, the problem becomes combinatorial. A brute-force approach would help in testing all the possibilities, however, this would be better accomplished by using more efficiency techniques such as evolutionary algorithms.

### 6.3 Evolutionary algorithms

Evolutionary algorithms are AI techniques that have gained great importance over the last decade since they are capable of coping with difficult to characterise (Koza, 1992), large search spaces which purpose is to optimise complex functions (non-linear, non-differential, etc.). This type of algorithms provide a solution where other methods are not feasible, too complex or a solution can not be found within a reasonable amount of time. Currently, the application of these algorithms covers different fields, from the design of the optimal placement of windmills in a wind farm in order to produce the maximum amount of energy at the lowest infrastructure cost (Chen et al., 2013), to the provision of lighting on motorways (Gómez-Lorente et al., 2013), optimisation of food packaging (Enitan and Adeyemo, 2013), structural optimisation



**Figure 6.1:** Outline of the cost function  $\phi_m(p_i)$ .

design of wind turbine blades (Zhu et al., 2014), configuration of power distribution systems (Tomoiağâ et al., 2013), optimal operation of a multi-reservoir system (Xu et al., 2012a), short-term power forecasting model for photovoltaic plants, operation and power flow Control of multi-terminal DC Networks for grid integration of offshore wind farms (Pinto et al., 2012), optimal charging scheduling of electric vehicles in Smart Grids linearisation and input-output decoupling for non-linear control of proton exchange membrane fuel cells (Chang and Chen, 2014).

This type of algorithms work with populations of individuals where each individual represents a candidate solution. In this paper, an individual is equivalent to the definition of 6 power values that meet the problem constraints, i.e., the values are between the minimum and maximum set by the chosen tariff and are sorted in ascending order. Individuals undergo an iterative process that applies a series of operations, combining them together and creating new individuals. Each transformation and selection cycle constitutes a generation, so that after a certain number of generations the best individual is close to the optimum solution. The main components of the evolutionary algorithms are:

- **Individual codification.** Representation of a candidate solution.
- **Population of individuals.** Set of individuals on which to apply the genetic operators in order to build new and better candidate solutions.
- **Fitness function.** Involves the definition of an objective against which each individual is tested for suitability for the environment under consideration. In this study, the fitness function is equivalent to the replication of the electric bill, which evaluates whether a particular combination of power values maximises or minimises the annual electricity cost.
- **Selection procedure.** Mechanisms that determine how to choose the individuals, which will later act as operands for the transformation procedures, in such a way that the best candidate solutions are kept within each iteration and the genetic diversity of the population is maintained.
- **Transformation procedures.** Operations that allow to build new individuals from old ones.

Genetic Algorithm (GA) are a particular type of evolutionary algorithms. These algorithms try to simulate the process of natural selection as a process optimiser by defining a set of operations that can be applied to the population of candidate solutions. The operations available are as follows:

- **Selection procedure:** Depending on the characteristics of the problem, there are several types of selection techniques:
  - **Elitist.** Ensures the selection of the fittest individuals of every generation.
  - **Roulette.** The probability of selecting an individual is proportional to the difference between its ability and its competitors’.
  - **Tournament.** constructs population subgroups which members compete against each other. Only the best individual of each tournament is selected.
  - **By rank.** Each individual in the population is assigned a numerical rank according to its fitness defining its selection priority among the rest.
- **Transformation procedures.** Genetic algorithms use two procedures to evolve the population:

- **Mutation.** Involves the alteration, at random, of a portion of the structure of the individual.
- **Crossover.** combines the structure of two individuals to produce an artificial offspring, simulating the recombination process that takes place during sexual reproduction. Common forms of crossover techniques include:
  - \* **Single point.** One crossover point is selected, code from the beginning of the individual to the crossover point is copied from one parent, the rest is copied from the second parent.
  - \* **Two point.** Two crossover points are selected, code from the beginning of the individual to the first crossover point is copied from one parent, the part from the first to the second crossover point is copied from the second parent and the rest is copied from the first parent.
  - \* **Uniform.** Each equivalent element of both individuals is compared and exchanged based on a fixed probability, typically 0.5.

There are several ways in which GA can be implemented. The most common, which covers the use of the most important operators is described in Algorithm 2.

The selection of the operators and methodology used to construct a genetic algorithm strongly affects the process and its ability to find a solution close to the global optimum. In fact, the following are the most important aspects that should be taken into account when designing a genetic algorithm:

- **Range of variables.** It is important to note that the algorithm will not be able to find a good solution if the range of variables is not correctly defined and/or the initial population is not properly distributed.
- **Variability of candidates.** The heterogeneity of the individuals within the population is a key for the "goodness" of the final solution. A higher genetic diversity increases the possibility of combining and generating new solutions, allowing a better exploration of the search space and coming up with solutions that are closer to the global optimum.

- **Final local search.** Once the algorithm has come up with an optimal solution, a local search can be carried out to improve the final result.

**Data:** size  $a$  of population, rate  $b$  of crossover, rate  $c$  of mutation, number  $d$  of iterations

**Result:** A population of possible solutions  $X$

**for**  $i = 1$  to  $a$  **do**

    Create a random candidate solution  $i$ ;

    Assign a fitness value to candidate solution  $i$ ;

    Introduce candidate solution  $i$  in the population  $X$

**end**

$j = 0$ ;

**while**  $j < d$  **do**

    [Selection] Select two candidate solutions from population  $X$ ;

    [Crossover] With crossover probability  $b$ , combine the parents to form a new offspring;

    [Mutation] With mutation probability  $c$ , mutate new offspring;

    [Accepting] Place new offspring in new population;

    [Replace] Use new generated population for a further run of algorithm;

$j = j + 1$ ;

**end**

**Algorithm 2:** Pseudocode of a genetic algorithm.

## 6.4 Materials and methods

The commercial infrastructure which electricity consumption was evaluated is comprised of six buildings: five located in Bilbao and one in Donosti; both cities situated on the north of Spain. Currently, each building feeds electricity from its own supply point and therefore is managed under a particular electric contract. Five of the buildings (four in Bilbao and the one in Donosti) have a 6.1 tariff (six-period tariff) while the remaining one has a 3.1 tariff (three-period tariff).

As seen in Figure 6.7, the behaviour of the demand curves of four of the buildings with the 6.1 tariff is quite similar:

- Days can be classified into three types: weekdays, Saturdays and holidays.
- The highest energy consumption takes place on weekdays, consistent with the hours of use of the buildings.
- The peak hours take place between 10:00 and 18:00.
- On Saturday, the demand curve follows a similar pattern, but with a much lower consumption and the peak hours finishing at 14:00.
- On holidays, the curve is almost flat.

The only difference of the building D with the previous ones is the existence of a fourth day type, bank holidays, where the shape of the load profile is similar to a weekday but the peak load is similar to a Saturday. Based on this configuration, we propose the analysis of two case studies:

- **Individual case:** Analyse the electric consumption per building and apply a the electricity tariff optimisation algorithm to each dataset separately.
- **Joined case:** Simulate that the buildings with the 6.1 tariff located in Bilbao are all connected to a single supply point and are all managed under the same electric contract. This means that the consumption values of the four buildings will be added together to generate a global dataset on which to apply the genetic algorithm.

In both case studies, the purpose of applying the genetic algorithm to the consumption datasets is to find the best combination of power values that meets the demand of the infrastructure while minimising the economic cost derived from a contract configured with higher power values than those that the facility needs. The analysis of the best solution will depend not only on the degree of optimisation achieved in each case, but also on the possibility of modifying the current electric configuration if the best optimisation is given by a different layout regarding the supply points.

The energy consumed by the whole infrastructure is monitored by an automated system, which periodically collects the data recorded by the buildings' meters and

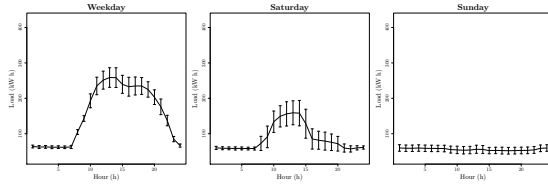


Figure 6.2: Building A.

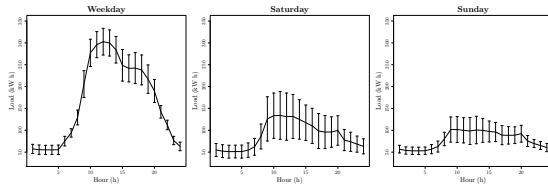


Figure 6.3: Building B.

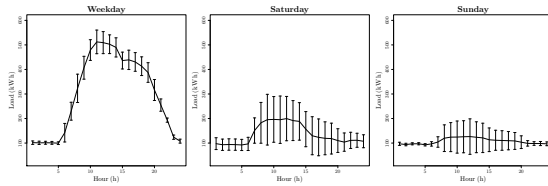


Figure 6.4: Building C.

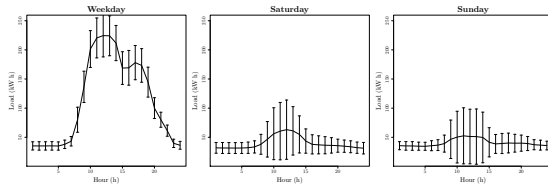


Figure 6.5: Building D.

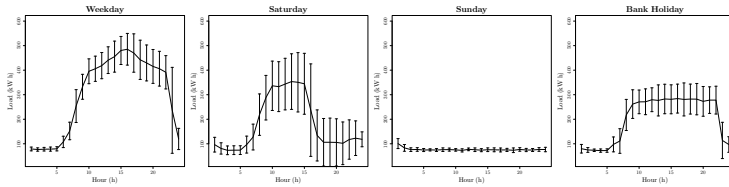


Figure 6.6: Building E.

Figure 6.7: Mean load over a year for all buildings. Errors bar denote  $\pm\sigma$ .

stores it in a centralised database. The monitoring module manages all aspects of telemetering, from the identification of missing or corrupted data up to the management of events in case of meter malfunction or communication failure. The centralised database also stores additional data regarding the electricity contracts associated with each building, the definition of the tariff's periods, the cost of the power term established by the BOE, and the distribution of bank holidays within the year. All this information feeds the system's optimisation module, which replicates the annual bills in order to simulate scenarios with different sets of contracted power values to find the combination that meets the infrastructure needs while minimising the global cost.

In addition, the historical evolution of electric consumption in the six buildings is accessible via the visualisation module which provides a web interface to interact with the functionalities provided by the automated system. Users can monitor the status of meters and demand curves, query information regarding the electricity contracts of each building, analyse monthly bills and run simulations for power optimisation.

## 6.5 Results and Discussion

The implementation of the genetic algorithm and the bill replication module altogether called for a deep study of the parameters and characteristics influencing the electricity contract of each building in order to identify common and particular aspects. The process of optimisation performed by the algorithm in both case studies is as follows:

- The system collects data from the database regarding the annual consumption, contracts, and tariff parameters per building.
- A partial calculus of the annual billing is done for those concepts that do not influence the power term such as cost of active energy, reactive energy and equipment rental. This partial invoice is added as input to the core of the optimisation algorithm. It is important to note that in the joined study use

case, in which all buildings are evaluated as a single facility, the values of energy consumption are added together to form a single dataset.

- The genetic algorithm then generates a set of candidate solutions, i.e. combinations of six power values that satisfy the constraints set by the type of tariff (upper and lower limit values of power that can be demanded; in the case of the 6.1 tariff, one of the values must be higher than 450 kW).
- The candidate solutions undergo a series of transformation through the use of genetic operators.

Using the previous approximation, three different experiments were carried out in order to assess the quality of the GA: tariff optimisation, savings forecasting and convergence.

### 6.5.1 Tariff Optimisation

The first experiment assesses the percentage of reduction achieved in the electricity bill using the TM, i.e. the optimisation technique explained in Section 6.1, in comparison to the proposed GA methodology. Namely, the objective is to look for the power limits  $p_1, \dots, p_6$  of every period that minimise the electric bill of the previous year. The results can be seen in Table 6.3. Columns  $\Delta T_{PM}$  and  $\Delta T_{GA}$  show the *percentage of reduction* on the period of time comprised by the dates in columns *From* and *To* for the TM and the GA respectively. Note that negative values denote an increment in the electricity bill while positive values denote reductions in the electricity bill.

In the *individual case*, the GA almost always finds a better combination of parameters that minimises the annual cost than the TM. As expected, the improvement is scarce. For the *joined case*, however, the savings are significantly higher, representing the 30% in reduction. However this change may not be always possible to contract due to physical or legal restrictions, but should always be considered in the design phase of a complex of buildings such as a university campus.

**Table 6.3:** Comparison of the results achieved by TM and GA optimisation methods.

Building	From	To	$\Delta T$ TM (%)	$\Delta T$ GA (%)	Days	$\Delta V$ PM (%)	$\Delta V$ GA (%)	Conv (%)
A	2012-09-01	2013-08-31	-1.68	0.03	211	-1.69	0.59	0.03
A	2012-10-01	2013-09-30	-1.66	0.09	181	-1.95	0.88	0.03
A	2012-11-01	2013-10-31	-1.51	0.18	150	-1.78	1.01	0.04
A	2012-12-01	2013-11-30	-1.24	0.22	120	-1.50	1.61	0.03
A	2013-01-01	2013-12-31	-0.63	0.91	89	-0.47	2.44	0.04
A	2013-02-01	2014-01-31	0.06	1.47	58	-1.23	1.23	0.04
A	2013-03-01	2014-02-28	0.10	2.09	30	6.49	6.95	0.03
B	2012-09-05	2013-08-31	1.56	3.02	211	5.92	9.54	0.04
B	2012-10-01	2013-09-30	1.97	3.45	181	6.52	10.18	0.04
B	2012-11-01	2013-10-31	2.25	3.88	150	6.39	10.49	0.05
B	2012-12-01	2013-11-30	2.54	4.43	120	6.34	10.31	0.04
B	2013-01-01	2013-12-31	3.82	5.69	89	7.39	11.86	0.04
B	2013-02-01	2014-01-31	4.37	6.61	58	8.01	13.62	0.03
B	2013-03-01	2014-02-28	5.76	8.18	30	15.79	26.04	0.04
C	2012-09-01	2013-08-31	-1.18	0.31	211	-0.31	2.48	0.05
C	2012-10-01	2013-09-30	-1.06	0.45	181	0.29	3.66	0.05
C	2012-11-01	2013-10-31	-0.69	0.76	150	0.62	3.90	0.03
C	2012-12-01	2013-11-30	-0.87	0.95	120	-0.35	4.07	0.03
C	2013-01-01	2013-12-31	-0.41	1.62	89	0.19	4.59	0.03
C	2013-02-01	2014-01-31	0.03	2.18	58	-0.57	3.78	0.03
C	2013-03-01	2014-02-28	0.33	2.88	30	3.34	11.22	0.03
D	2012-09-10	2013-08-31	10.40	12.68	211	9.36	10.39	0.09
D	2012-10-01	2013-09-30	10.80	13.03	181	7.79	9.50	0.07
D	2012-11-01	2013-10-31	11.25	13.35	150	5.16	8.22	0.07
D	2012-12-01	2013-11-30	11.46	13.54	120	1.45	6.69	0.09
D	2013-01-01	2013-12-31	9.77	12.83	89	2.72	10.52	0.08
D	2013-02-01	2014-01-31	9.59	13.03	58	5.77	12.70	0.09
D	2013-03-01	2014-02-28	9.09	12.94	30	34.00	29.11	0.09
E	2012-09-10	2013-08-31	-2.26	0.48	211	-6.06	3.38	0.03
E	2012-10-01	2013-09-30	-1.91	0.74	181	-4.64	4.73	0.03
E	2012-11-01	2013-10-31	0.33	1.22	150	0.90	5.18	0.05
E	2012-12-01	2013-11-30	0.37	1.50	120	0.92	4.99	0.03
E	2013-01-01	2013-12-31	0.41	1.75	89	0.95	6.44	0.03
E	2013-02-01	2014-01-31	0.52	2.19	58	1.27	8.10	0.05
E	2013-03-01	2014-02-28	0.57	2.53	30	1.95	12.50	0.03
JOIN	2012-09-10	2013-08-31	38.91	39.53	211	35.30	37.87	0.03
JOIN	2012-10-01	2013-09-30	38.38	39.05	181	37.05	39.42	0.05
JOIN	2012-11-01	2013-10-31	36.80	28.61	150	40.04	33.65	0.03
JOIN	2012-12-01	2013-11-30	36.22	28.55	120	42.47	35.69	0.03
JOIN	2013-01-01	2013-12-31	36.08	28.72	89	43.45	36.87	0.05
JOIN	2013-02-01	2014-01-31	35.76	28.88	58	44.26	37.45	0.03
JOIN	2013-03-01	2014-02-28	35.48	29.09	30	47.37	42.50	0.03

### 6.5.2 Savings Forecasting

The second experiment assesses the ability to forecast future savings for both the *individual case* and the *joined case*. In this experiment we change the power limits using the data comprised between *From* and *To* dates for the period comprised between the *To* date until 2014-03-31. Columns  $\Delta V$  PM and  $\Delta V$  GA of Table 6.3 contain the percentage of reduction over this period for the TM and the GA respectively. As in the previous experiment, negative values denote an increment in the electric bill, while positive values denote a reduction.

As seen, the GA achieves a significant reduction in the annual electricity cost for the *individual case* and the *joined case*. Therefore, these results suggested by the GA proposed can provide long-term savings.

### 6.5.3 Convergence

Since the GA is stochastic by nature, its convergence properties have to be tested empirically. In this context, convergence refers to the ability of the proposed GA methodology to find a local minima. In order to measure this ability, the *tariffs optimisation* experiment was repeated 100 times for each period of time defined by the *From* and *To* columns. The variability of the results was examined by using the Coefficient of Variation (Reed et al., 2002) which normalises the standard deviation with respect to the mean value of the sample. In this sense, it represents the percentage of central variation of the sample. Namely, a low value means that the method always finds the same local minimum. On the other hand, high variability may mean that the function has several different local minima or that the method fails to converge. Column *Conv* of the results in Table 6.3 suggest that, in all cases, the method converges to the same local minimum, giving an indication that it may be the absolute one.

## 6.6 Conclusions

The optimisation of the electricity bill is a secure alternative in terms of cost savings given the high energy consumption in which business and particulars alike incur today. Of the possible alternatives for optimisation, the one that balances contracted power values is the most affordable, since its application does neither involve a change in the electric system nor a change of supplier, and is of immediate application when renewing electric contracts. This chapter proposes an artificial intelligence based methodology, namely evolutionary algorithms, to find the best combination of power values that minimises the annual electricity cost while meeting the infrastructure needs, in this case a public infrastructure comprised of several buildings.

In this sense, two case studies were proposed to evaluate the degree of optimisation achieved within each building separately in comparison to the results obtained by simulating that all the buildings are connected to the same single supply point. The results show that the optimisation achieved with the GA is greater than that obtained with the TM.



*Conserving energy and thus saving money, reducing consumption of unnecessary products and packaging and shifting to a clean-energy economy would likely hurt the bottom line of polluting industries, but would undoubtedly have positive effects for most of us.*

David Suzuki

CHAPTER

# 7

## Industrial DSM : Energy-aware foundry production scheduling

**I**NDUSTRIAL DSM plays a pivotal role in the management of the electric grid due to the considerable amount of electricity needed by power intensive processes. In contrast with domestic DSM, which relies heavily on the constraints defined by human preferences on the degree of personal comfort, the deployment of industrial DSM strategies is eased because of the singular characteristics of this economic sector:

- The individual power consumption in the industrial sector is very high, which motivates and facilitates the participation in DSM programs.

- In most cases, the initial capital investment needed to implement DSM strategies is very low because the plant may already have an advanced metering infrastructure deployed.
- Since industrial processes operate in isolated environments, human comfort related constraints are generally not an issue. Therefore, the DSM strategies can focus primarily on the sole characteristics, scheduling, and planning of the processes.

Nevertheless, industrial processes are complex and conditioned to strict safety requirements. Prior to the application of DSM techniques to any industrial activity, it is therefore necessary that operators assure the process is flexible enough to avoid adverse disturbances caused by sudden changes in the plant operation.

Within the industrial sector, foundries stand out as one of the most energy intensive industries. Consequently, the application of DSM strategies in this field can be beneficial in terms of controlling the electricity demand load shape so as to avoid demand peaks, and saving costs by scheduling the castings on the hours of the day in which the price of electricity is lower.

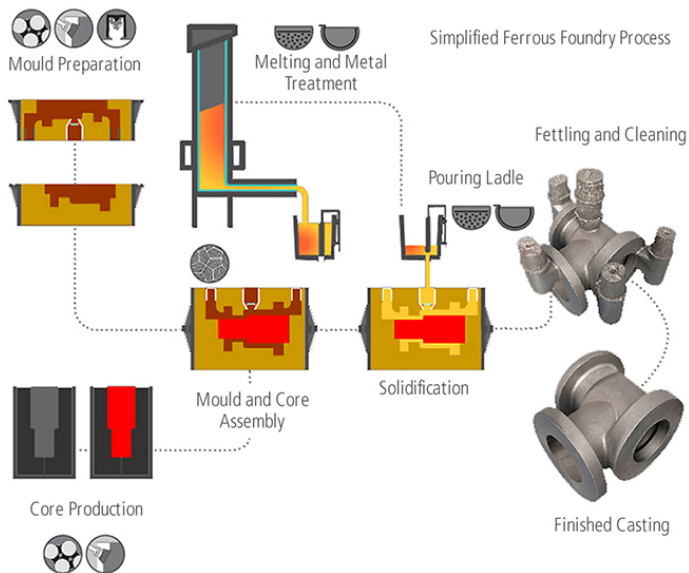
This chapter focuses on the application of DSM strategies to the operation of a foundry plant. Specifically, this approach relies on the use of AI techniques to schedule the optimal operation of the casting process that minimises the electricity cost of the process while meeting the production deadlines.

## **7.1 Introduction**

Casting by foundry plants is one of the main branches of the metallurgical industry. The casting process involves the production of pieces by pouring molten metal into prefabricated moulds. This molten metal is produced from several sources, the most common being scrap, pig iron, ingots, and alloys. Due to the nature of their processes, foundries are a very energy intensive industry, which leads to consider energy accounting (i.e. determine where and how energy is being consumed and how efficient the energy management is) one of the important pillars involved in the process of decision making. Energy accounting has to take notice of the factors that

directly affect the energy consumption of this type of industry, as noted in (Nieves et al., 2010): the continued rise in energy prices reduces the profit margin, while new European Union (EU) directives on electricity market deregulation (e.g., ED 2003/54) penalise excessive consumption.

The extreme conditions under which casting is carried out deem complexity to the process. In order to build the final pieces, the casting process involves several sequential phases, as seen in Figure 7.1: source metals are placed into electric arc furnaces while moulding machines build and prepare the sand moulds where the casting will be poured. The heat needed to melt the metal is produced with an electric arc from three carbon or granite electrodes. Once the casting is poured into the moulds, the electric arc furnace is lined up with refractories, which slowly decompose and remove the slag.



**Figure 7.1:** Phases of the casting process.

Nowadays, the scheduling and dispatch of production orders in the foundry industry is carried out in such a way that the actual cost of energy consumed is not always taken into account. The cost of energy fluctuates during the day, being more expensive in the mornings, and more affordable in the evenings or at night. There exists a general notion about the low-priced hours that would benefit the casting process in terms of electricity costs, but there is no clear proof whatsoever whether that is the best solution to approach. The distribution of the quantity of production orders and the configuration of an scheduled plan within the furnaces available is absolutely based on expert's knowledge.

## **7.2 Related Work**

The optimisation of the energy consumption in industrial processes is at a point with a new revolution of intelligent based systems that allow better control of the variables involved in order to make the processes more efficient and to make the most of the production. Nowadays, many of these decisions are made only with the help of experience and intuition.

GA and Simulated Annealing (SA) are widely used on job shop scheduling problems. Being NP-hard problems, these algorithms provide a way of reducing the search space of the problem and selecting only those individuals whose combination may lead to the finding of an accurate solution. There is a lot of literature concerning artificial intelligence applied to job shop scheduling (Wei and Blake, 1992), (Mati et al., 2011), (Kachitvichyanukul and Sitthitham, 2011), (Zhang and Wu, 2011). These algorithms, however, refer to different types of jobs which processes involve more than one machine (i.e. the making of an specific piece is dependent on several jobs that are carried out along several machines). The object of those algorithms is to find the most suitable order in which to produce the pieces, configuring them across different machines to minimise the duration of the overall production process and avoid or shorten time leaks. The huge energy consumption in the melting process along with the much lower consumption of the subsequent steps advise to focus only on the first part (i.e. the melting phase). The situation faced in the problem proposed

is quite different as we focus only on one step of the job shop scheduling, in this case, the optimisation of energy consumption on foundry processes. The jobs involved are specifically carried out in a line of production and the main objective is to find an accurate ordering of the orders along the furnaces that minimises the amount of energy consumed.

Methods based on evolutionary algorithms have proven to achieve better results than other classical approaches in solving NP-hard problems, such as Job-Shop Scheduling (JSP), (Mati et al., 2011; Kachitvichyanukul and Sittitham, 2011; Zhang and Wu, 2011; Watanabe et al., 2001; Wan et al., 2011). These algorithms, however, refer to processes which manufacture is carried out on several steps. The objective of those algorithms is to find the most suitable order in which to produce the pieces, configuring them across different machines in order to minimise the duration of the overall production process and avoid or shorten time leaks.

Other works like the one presented in (Santos and Dourado, 1999) propose a system that looks for the best assignment for all the production sections in a particular area of a paper industry, optimising energy costs through the use of GA. The investigation is focused on programming maintenance shut-downs in adequate time so as to prevent disturbances that may lead to mass energy losses.

Particularly, regarding the foundry process, the high energy consumption that takes place during the melting phase in contrast with the much lower consumption of the subsequent steps, advise to focus optimisation on the first part. There are several studies concerning job-scheduling within foundries. For example, (Deb et al., 2003) proposes the use of GA to schedule the planing of orders along several days according to their casting weights and due dates. They construct a scenario based on the possibility of having two furnaces working intermittently through the whole week. Their results show that the basic operators of GA do not prove to be enough to solve the problem in a specific amount of time. Therefore a problem-specific set of operators had to be defined. As mentioned before, the investigation is focused on finding the best plan based on the due date of the orders, not taking into account the overall cost of the production, a parameter which would be ideal to consider.

Works like (Gravel et al., 2000) and (Xu et al., 2010) also use GA to seek the best processing sequence in foundries for  $n$  orders on  $m$  parallel machines. In the same sense, (Zhang et al., 2008) seeks an optimisation in the times of production to prevent energy losses between tasks. In the venue of energy optimisation, (Kümmel et al., 1988) proposes the concept of energy conservation in industrial systems. The objective function is the amount of primary energy saved by maximising the use of secondary energy within the limits established by the laws of thermodynamics.

## 7.3 Materials and methods

### 7.3.1 Description of the case study

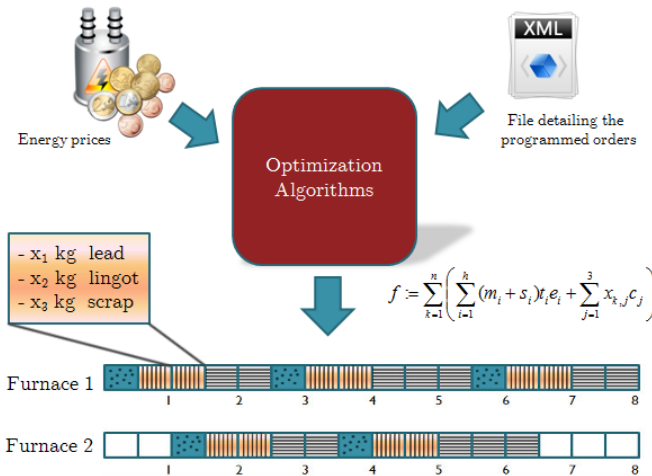
The main line of production of the foundry plant used as object of study consists of two ABB electric furnaces, each one with a maximum capacity of 6 tons. According to the delivery date and nature of the orders, an Enterprise Resource Planner (ERP) issues a production plan: a list of jobs detailing the amount of pieces to be manufactured along with their reference numbers and the weight of each individual piece. The production plan is then handed to the casting operator, who decides how to schedule the operation of the furnaces to meet the production deadlines.

Specifically for this production plant, the whole casting process is composed of three phases:

- **Filling phase:** Involves setting up the furnaces by slowly heating them while filling it with the metallic resources needed to create the casting. Depending on the availability of the resources, the casting operator will load and specific combinations of materials. These combinations are built in such a way that the total amount of casting is always 6 tons. This phase lasts always 1 hour: 51 minutes for loading the furnace and 9 minutes for casting correction.
- **Melting phase:** The furnace reaches the desired temperature and starts melting the metallic resources. This phase lasts always 1 hour.
- **Emptying phase:** The casting is ready and the furnace is emptied into the sand moulds. The duration of this phase is always 1 hour.

In addition to the particular characteristics of the casting process, the foundry plant itself includes a set of operative restrictions inherent to the working schedule:

- The amount of time in which the furnace is in melting phase must not exceed an hour in any case.
- Two furnaces can not be in melting phase at the same time.
- The amount of casting produced must be equal to the maximum capacity of the furnace.
- The amount of pieces produced must be such that the whole casting is consumed on each emptying.
- The total material used in the melting phase must not exceed the level of existing stock on that day.



**Figure 7.2:** Optimisation methodology for the casting process.

Figure 7.2 optimisation methodology is summarised in Figure 7.2. The optimisation algorithm receives data detailing the configuration of the furnaces and the production plan, as well as information concerning the price of energy for an specific day and the particular characteristics of the casting process. With these data,

the optimisation algorithm produces an elaborated scheduling of the production phase which explains the amount of stock to be used and the pieces that should be produced to meet the expected deadlines. Table 7.1 details the inputs and outputs to the optimisation process.

**Table 7.1:** Inputs and outputs to the casting optimisation process.

Concept	Description
Inputs	<p>Maximum capacity of each furnace.</p> <p>Stock limit of the materials to be melted.</p> <p>Temperature of each furnace when it is being loaded.</p> <p>Temperature of each furnace when it is melting.</p> <p>Temperature of each furnace when it is being unloaded.</p> <p>Amount of time it takes to load each furnace.</p> <p>Amount of time it takes to melt an entire cast.</p> <p>Amount of time it takes to unload or empty each furnace specifying whether it depends on the size of the pieces or any other issue (i.e. time per piece, average, etc.).</p> <p>Cost of energy for each hour of the day.</p> <p>Daily list of orders planned to be carried out: date of delivery, number of pieces, reference numbers, etc.</p> <p>The list of orders that were carried out on each day, the energy consumption for the day and its equivalence in economic cost.</p>
Outputs	<p>Schedule indicating the ideal placement of the orders on each furnace with details such as the starting and end hour for each phase will also be provided.</p> <p>Brief summary of the completion of orders intended to be carried out on a certain date as well as the economic cost derived.</p>

### 7.3.2 Optimisation algorithms

#### Genetic algorithm

GA (Koza, 1992) are iterative and stochastic processes that operate on a set of individuals (population) and simulate the process of evolution. The individuals are formed by a set of rules or variables that describe their composition and each rule represents a particular feature inherent to the characteristics and behaviour of said individual. The overall evaluation of the rules against the problem defines the fitness of the solution (i.e. how fit the candidate solution is in comparison with the others).

The algorithm proposes a set of operations to be performed on the individuals. These operations include the combination and mutation of rules to create new offspring which forms a new pool of candidate solutions, and are subjected to a second round of fitness evaluation. The expectation is that the average fitness of the population will increase each round, and so by repeating this process for hundreds or thousands of rounds, very good solutions to the problem can be discovered.

GA have been used on a wide range of areas such as optimising the design of antenna arrays (Lee et al., 1999), methods for fingerprint core point extraction (Gravel et al., 2000), studies of optimal economic power dispatch (Yoshimi et al., 1993), find the optimum design of a concert hall (Sato et al., 2002) or multi-FPGA partitioning (Hidalgo et al., 2001).

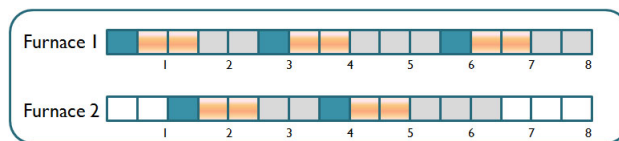
- **Chromosome encoding:** For the case presented, the large number of possible combinations between the different variables made the search space impossibly large. To solve this problem, we decided to encode the solution in form of subgroups of individuals within individuals. The furnaces were represented as two arrays of length equal to that of the shifts defined in the foundry schedule. Each of the arrays accommodates a series of orders while verifying that all the restrictions previously mentioned are observed.

As mentioned earlier, the casting process involves three phases. At first, the furnace is filled with the materials to be melted. Then, as the temperature on the furnace becomes constant, the melting process starts. Once the casting is produced, the final process, which involves pouring said casting into the

moulds in order to make the pieces, takes place. Bearing in mind the innards of each of the three phases and how each one of them contemplates a particular set of variables, we opted for the creation of an specific gene for each phase.

- *Filling gene*: This gene contains information about the filling state and the initial configuration of the furnace.
- *Melting gene*: This gene contains an array with equal length to the number of different types of stock considered. Each slot in the array represents the amount of material to be used in every case. The total amount of material used must not exceed the maximum capacity of the furnace nor the total amount of stock available.
- *Emptying gene*: This type of gene contains an array of orders that specifies the quantity, type of pieces to be produced with the casting and the reference number. The final layout of orders in this gene will serve as an indication of which and how many moulds of what type need to be prepared each time.

All of the above genes come together to form the *casting gene*. This gene will act as a single entity to schedule the orders on the two furnaces through the hours of the shifts. This means, that the genes mentioned above will not act independently but that a triplet of them will form a gene that is going to be the operational unit in the genetic algorithm. This triplet will follow a defined sequence as the nature of the process obliges a particular order of functioning. Given the restrictions of the problem and the particular codification of the genes mentioned before, the resulting chromosome is depicted in Figure 7.3.



**Figure 7.3:** Chromosome representation

The green boxes indicate the times when the oven is in the filling phase. The orange boxes refer to the times when it is melting and the gray boxes to the emptying phase of the casting. Each set of boxes filling and emptying cast is considered as a single entity and the sequence of tasks can not be broken or constructed in a different order. We will call *gene* to this group of activities since they will be part of the solutions that are combined with others.

As shown in Figure 7.3, the codification of the chromosome contains no coincident melting genes, that is, the furnaces are never in melting status simultaneously. Furthermore, the order and execution of tasks (filling phase, melting phase, emptying phase) is consequent, so this solution may be considered valid.

- **Initial population:** Since the chromosome is divided in sets of genes and sub-genes, a special initialization is required for each type before the chromosome is built. First, the filling gene is constructed by assigning groups of predefined quantities of stock. These quantities come as triplets of aggregates that define how much material of each type should be used, along with special additives that help produce a casting of quality. These additives are established in relation to the amount and type of stock used, so there is no possibility to vary them beforehand.

After the initialization of the *filling gene* comes the initialization of the *emptying gene*. For that, we take into account the planing produced by the ERP. Combining in a random way the different orders and number of pieces scheduled, we form groups of pieces which moulds will empty the casting produced during the melting process. The total weight of the overall gene is the same as the maximum capacity of the furnace and the individual weight will determine the amount of time the furnace is in emptying state. Once the *casting gene* is formed said gene is randomly placed on a slot in the schedule of one of the two furnaces. This process goes on as long as there is space on the schedules in which the orders can be placed. Once the chromosome is built, restrictions are checked to determine whether the whole chromosome is valid or not, i.e.

checking whether there are times when the furnaces are in melting status at the same time or if the overall planing complies with the predefined shifts.

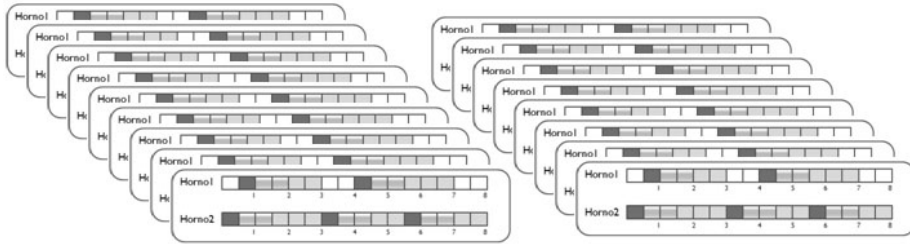


Figure 7.4: Initial population

All the chromosomes are initialised using the method previously mentioned and a population of candidate solutions is then created. Each one of these chromosomes represent a possible solution to the scheduling problem.

- **Fitness calculation:** A genetic algorithm requires a criteria to compare solutions in order to find which one is better. To achieve that a fitness function is defined to measure the capability of a candidate solution. For the presented problem, we have defined the fitness function as follows:

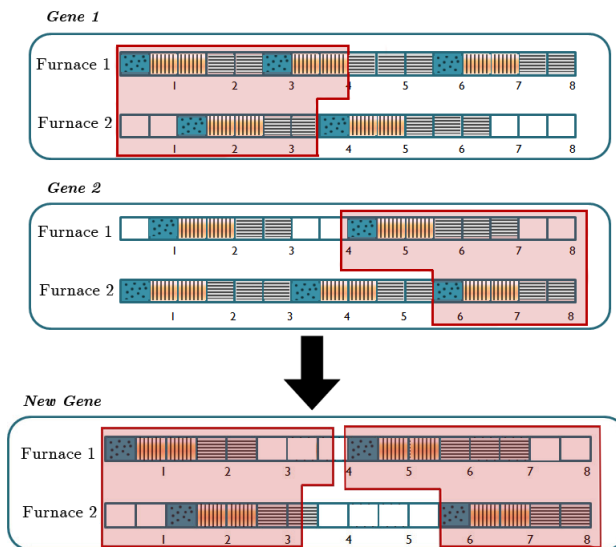
$$k_i = w_i t_i e_i + \sum_{j=0} x_j c_j + w_p t_i e_i$$

where  $w_i$  is the consumption of the furnace in melting state,  $w_p$  refers to the consumption in standby state (whether the furnace is not operative or it is being emptied),  $t_i$  is the duration of a certain state and  $x_j$  and  $c_j$  are parameters that show the amount, type and cost of stock used to produce the casting.

- **Operation criterion:** In each generation, a part of the existing population is selected to provide a new series of chromosomes. The solutions are selected through a fitness-based process, where those with better fitnesses are more prone to be selected. For the selection of the individuals to be processed by the algorithm, we implemented the *roulette wheel selection* algorithm. The roulette wheel selection is a method for choosing members of the population

of chromosomes in proportion to their fitness. Although it does not guarantee that the fittest member goes on to the next generation, it provides a mechanism for making it more eligible among other solutions. Each solution is given a percentage that defines the possibility of being chosen (those with better fitness will have a wider range of percentage). Every time there is a need for choosing a chromosome from the pool of solutions, this method is executed and the randomly generated number will provide with the new candidates.

The operations involved in GA: selection, crossover, and mutation are executed within the scope of the chromosomes considering each gene as a single minimal entity. Given the nature of the problem and the final codification of the solutions, we designed an intra level set of operations, meaning each type of gene (filling gene, melting gene, emptying gene and casting gene) implements said operations individually. Moreover, we implemented these operations for the chromosome as a whole in order to interact with other solutions. Fig. 7.5 presents an example of this design.



**Figure 7.5:** Implementation of the crossover operation

The overall process involves the following steps: first we select two chromosomes or solutions from the chromosome pool through the roulette wheel algorithm. Given the two candidates, the algorithm checks to see whether the crossover operation should be applied at chromosome level. If so, we chose a point of cutting for the two arrays. It is important to take into account that the point of cut may not be identical in both furnaces due to the complexity of the encoding of the genes, i.e. we may choose a cutting point in the first furnace that cuts in half the casting gene in the second furnace. The point of cutting must take into account the distribution of the casting genes along the overall chromosome.

The new offspring can then be mutated (i, e. that the elements of the chromosomes are changed independently). This variations are mainly caused by *errors* in copying genes from parents, but is also a way to assume genetic welfare in the nature. In the problem at hand, it was important for the new chromosomes to be coherent with the order planing defined before. The problem could be found specially in the emptying gene. The emptying gene defines the amount of pieces to be made with the casting and referencing the order to which they belong. Thus, we used the mutation operator to repair possible inconsistencies among emptying genes.

New genes are added to the population and the process starts again. New parents are selected for each new child and the process continues until a new population of solutions of appropriated size is generated. These processes ultimately result in a population which is the next generation of chromosomes, different from the initial one. The average fitness will have increased by this procedure since only the best organisms from the first generation are selected for breeding. The algorithm finishes when maximum number of generations has been produced.

### **Simulated Annealing**

GA tend to converge towards local optima instead of the global optimum of the problem since this type of strategies does not sacrifice short-term fitness in order to gain

longer-term fitness. While both GA and SA (Kondrak and Beek, 1997) may converge prematurely to suboptimal points, the latter has a formal proof of convergence.

Each step of the SA algorithm attempts to replace the current solution by a random solution. The new solution may then be accepted with a probability that depends both on the difference between the corresponding fitness function values and on a global parameter  $T$  (known as the temperature), that is gradually decreased during the process. The dependency is such that the choice between the previous and current solution is almost random when  $T$  is large, but increasingly selects so-called *downhill* solution as  $T$  approaches zero. The allowance for *uphill* moves potentially saves the method from becoming stuck at local optima which are the drawback of greedier methods.

SA has been used in studies such as joint configuration of the optical and electrical layer in intelligent optical networks (Qin et al., 2002), optimisation of texture synthesis (Huiqiu and Junhua, 2010), Unmanned Aerial Vehicle (UAV) route planning (Meng and Xin, 2010) or logistic network design (Jin and Li-xin, 2009).

The initial feasible solution from which the algorithm must be started is taken at random from the set of possible solutions. Generally, when the codification of the atoms of the genetic algorithm is simple, such search is accomplished by using random numbers, but for the present case, since the smaller units of operation are complex, that is not possible. In order to solve this issue, we have implemented a greedy local search algorithm which provides candidate solution to start the evaluation.

In every iteration, as the cooling temperature drops, the configuration of the problem scheduling is modified. A probability rate indicates which part of the current candidate solution is going to be changed, meaning sometimes the probability will lead to the modification of the *emptying atoms*, randomly modifying the number of pieces of the orders that must be carried out, the sets of stock that must be used or the layout and ordering of the orders along the two furnaces.

At the end of the loop comes the decision whether to accept or not the current solution based on an operation of acceptance probability and the fitness function. The higher the temperature, the more prone the algorithm is to accept worse solutions.

As the temperature drops, the probability decreases and the algorithm will choose only the best solutions.

Finally, the termination criteria for the simulated algorithm depends on the amount of temperature defined at the beginning. Higher temperature results in a wider range of search the whole problem space.

### **Backtracking algorithm**

In contrast with the genetic algorithm and simulated annealing, the backtracking algorithm is a general algorithm that checks almost every solution given a computational problem. The algorithm builds in an incremental way the candidates to solutions and abandons the developing of a partial candidate as soon as it determines that the path to develop said solution does not lead to a valid solution.

We start the backtracking algorithm with the definition of the possible states involved in the problem and a set of restrictions that must be taken into account when developing the search tree. Each node or state in the tree represents a status of the furnace. The depth of the tree will be equal to that of the predefined shifts scheduled in the foundry plant. The set of restrictions will dictate the valid order for the node development in each iteration, (i.e. if a previous node dictates that the furnace was in melting mode, the next iteration will only consider the node which contemplate the furnace being emptied). That way, the number of options to be evaluated decreases importantly. In the same venue, whenever the current fitness of the incomplete branch being evaluated (i.e., a branch whose total depth has not been totally developed) is higher than that of a complete branch previously evaluated the algorithm stops developing said branch and processes the next one.

## **7.4 Results**

### **7.4.1 Constant melting time**

The testing was carried out over two datasets detailing the production that took place on the foundry plant for two weeks. These datasets provided information about



will not provide a better solution to the problem since the algorithm has reached maximum fitness.

The results obtained from the execution of the SA are detailed in Table 7.3. In the same sense as the GA, the SA produces worse result with a small temperature, deeming the search space with few candidates to evaluate and decrementing the possibility of selecting worse solutions in order to explore it in a deeper manner. The increment of temperature leads to finding better solutions as well as a growth in the computational time to resolve the problem.

**Table 7.3:** Results obtained with the SA (constant melting time).

Temperature	Execution time (min)	Cost variation (%)
20000	12	+17
25000	17	+11
30000	24	+9
35000	29	+5
40000	32	-3
60000	56	-3

SA performs better in comparison to the GA, meaning the use of worse solutions leads to the developing of better candidates to evaluate. As seen in Table 7.3, SA reached a level in which the increasing of the temperature does not improve the overall fitness function of the final candidate.

One of the problems faced was that the definition of the shift limited the amount of options from where to choose from. That means, that the amount of codifications of individuals that could not considered valid, incremented, as opposed to using the twenty-four hour schedule. SA and GA performed worse when there was a limitation in the scheduling scope. To solve that issue, we opted for the use of a backtracking algorithm that proved to be better at finding solutions in the search space at the cost of a much higher execution since the restrictions of the problems provided a search space accurate for the search of the problem.

Bearing in mind the goal of minimising energy cost, Table 7.4 shows the execution time and the optimisation achieved by each algorithm. The column of cost variation

**Table 7.4:** Comparison of the results obtained with the GA, SA, and backtracking algorithms.

Algorithm	Execution time (min)	Cost variation (%)
GA	31 minutes	-1%
SA	56 minutes	-3%
<b>BA</b>	<b>165</b>	<b>-7</b>

shows how much the solution given by each algorithm differs from the actual cost of energy consumption. A positive value indicates that the result given by the algorithm exceeds that of the scheduling issued by the expert. On the contrary, a negative value means that the algorithm was able to minimise the cost.

#### 7.4.2 Variable melting time

Though reached, the optimisation of the casting process was, however, hindered by the fact that the amount of time the furnace is in melting status is always fixed (an hour). In order to produce a casting which properties comply with this condition, the foundry expert selects a specific aggregation of stocks from a list of material combinations. This list details the amount of material of each type (scrap, ingot, pig iron, and additives) to be used during the casting process, being that specific synthesis the one that allows to produce one-hour castings. This limitation, along with the one concerning the predefined shifts, reduced the search space significantly. The set of candidate solutions constructed shared very similar characteristics, thus, affecting the offspring's genetic diversity.

When constructing the casting genes for the GA, a new procedure determined the amount of casting time (all within a minimum and a maximum range). Then, for every iteration of the mutation operator, the procedure was invoked again to adjust new times and construct a valid candidate solution. The results for this model are shown in Table 7.5.

In the case of the SA, the same procedure was used for each iteration of the algorithm, in addition with a probability that determined which part of the candidate

**Table 7.5:** Results obtained with GA (variable melting time).

Population size	Execution time (min)	Cost variation (%)
60	9	+30
80	9	+31
100	13	+27
120	13	+23
180	18	+9
200	18	+5
250	20	-2
300	23	-6
<b>400</b>	<b>40</b>	<b>-8</b>

solution was prone to be changed. Table 7.6 summarises the results obtained in the evaluation of the SA.

Finally, for the Backtracking Algorithm (BA), the introduction of a variable melting time model allowed diversification in the development of the branches. The algorithm constructed a search tree based on the different possibilities for node representation. Each node of the tree defined both a status of the furnace and the duration of said status. For the several possibilities that could be developed regarding the duration, two were chosen at a random, bearing in mind the upper and lower limits for the melting phase. The set of restrictions dictates the valid order for the node development in each iteration, (i.e. if a previous node dictates that the furnace was in melting mode, the next iteration will only consider the node which contemplate the furnace being emptied). In the same venue, whenever the current fitness of the incomplete branch being elaborated was higher than that of a complete branch previously evaluated, the algorithm stopped developing said branch and began processing the next one. The total execution time of the BA increased significantly while the deviation decreased due to the ability of the algorithm to explore and find better solutions.

Table 7.7 shows a comparison of the results obtained by the three algorithms. Once again, there is a significant improvement in the minimisation of the total energy

**Table 7.6:** Results obtained with SA (variable melting time).

Temperature	Execution time (min)	Cost variation (%)
20 000	11	+19
25 000	19	+12
30 000	21	+9
35 000	25	+5
40 000	25	-2
60 000	46	-5
<b>100 000</b>	<b>128</b>	<b>-10</b>

cost, while the BA gives the best solution of them all. Moreover, it doubles the benefit achieved in the case of the constant melting time.

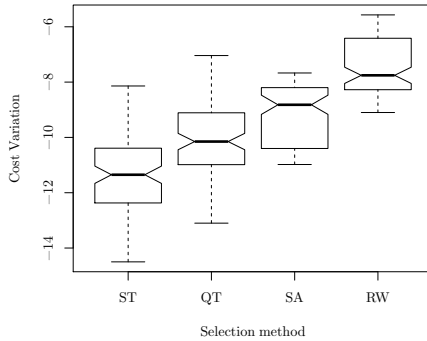
The comparison of the two groups of tests, GA and SA with and without variable melting time, shows that the set of algorithms incorporating the variability time perform better than those run in the first place. In addition, the execution time decreases since the algorithms spend more time in the operations themselves than in the construction of valid candidates and the verification of restrictions.

**Table 7.7:** Comparison of the three algorithms - Variable Melting Time

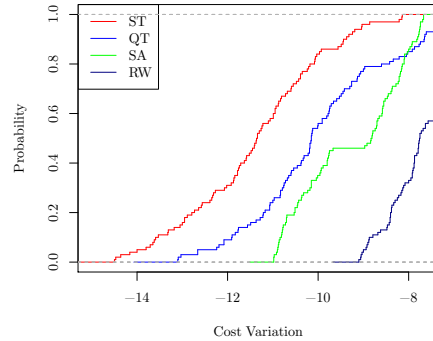
Algorithm	Execution time (min)	Cost variation (%)
GA	40	-8
SA	128	-10
BA	<b>291</b>	<b>-15</b>

### 7.4.3 Statistical tests

Since the algorithms presented are stochastic by nature, a statistical analysis is needed in order to compare their performances. Namely, the objective is to assess if the different cost variations seen previously are significant from a statistical point of view. In order to test this, every experiment was repeated 100 times.



**Figure 7.7:** Box-plot per algorithm



**Figure 7.8:** ECDF per algorithm

Figure 7.7 presents a *box-plot* of the results. Please note that in this figure, the bold line represents the median cost variation of the repetitions, the top and bottom part of the *box* represent the first and third quartile (*i.e.* the interquartile distance), while the *whisker* marks indicate the minimum and maximum value. Finally, since the *notches* of the different boxes do not overlap there is *strong evidence* that the medians differ (Chambers, 1983, p. 62).

As seen, all methods consistently improve the schedule used while presenting a low variance (the size of the boxes is small), which can be traduced in terms of good robustness. Moreover, the figure hints that the roulette wheel method is worse than SA and, in turn, the latter is worse than  $q$ -tournament, while the simple tournament comes up as the best of all. This hypothesis is tested trough a comparison of empiric cumulative distribution functions (ECDF), namely, for every pair of distributions there is a comparison whether one of them is stochastically smaller (*i.e.* has bigger cost reductions with more probability) than the other. Note that this means that the ECDF of the distribution is in the left part of the figure. Figure 7.8 displays the outcomes of this experiment.

Finally, a *Kolmogorov–Smirnov* test with Bonferroni correction (García and Herrera, 2008) between every pair of algorithms was used to confirm the main hypothesis (note that more powerful tests, like the *Friedman Test* with *post hoc* analysis (Derrac et al., 2011), are not needed due to the clearness of the results): 7.8 shows that the

results given by the GA with ST are stochastically smaller than those obtained with any other algorithm (with a very high significance level ( $4.28 \cdot 10^{-4}$  according to the Bonferroni correction); in turn, the  $q$ -tournament results are stochastically smaller than those of both the roulette-wheel, the SA, etc.

**Table 7.8:** P-values for the comparison between algorithms

	ST	QT	RW	SA
ST	–	$2.48 \cdot 10^{-7}$	$2.32 \cdot 10^{-16}$	$4.2 \cdot 10^{-39}$
QT		–	$3.57 \cdot 10^{-5}$	$3.72 \cdot 10^{-25}$
RW			–	$2.17 \cdot 10^{-13}$
SA				–

## 7.5 Discussion

This section focuses on the computational experiments carried out with the three different GA selection techniques, in order to give an overview of the approach followed.

### 7.5.1 GA: The importance of the selection method

Even though BA obtains the best results when minimizing energy cost, GA offers better possibilities of improvement in an acceptable amount of time when looking for an approximate solution. Moreover, the execution time of GA would hinder the search for an ideal solution when working with dynamic scheduling (i.e. the arrival of new orders with higher priority which modifies the whole planing). Other methods like Mixed Integer Linear Programming (MILP), can help solve this problem. However, given the amount of variables introduced by the problem, it is expected that the execution times of a MILP algorithm will follow closely behind those of the proposed BA, and even then, it is with high probability that the algorithm will not scale appropriately as pointed out in (Cardoen et al., 2010), (Roland et al., 2006). If the algorithm was to be extended to schedule the orders for several days instead of

a single day, the number of branches to be developed, and in turn, the execution time, would both increase exponentially, making it impossible to deliver a valid solution. Therefore, in order to test possible improvements on the GA, the following experiments focus on the variation that can be obtained by altering the selection procedure.

In particular, GA define a selection mechanism to choose individuals from the population (also referred to as mating pool), on which to perform the genetic operators that will create new offspring. Since the individuals in the mating pool are the ones whose genes are inherited by the next generation, it is desirable for this pool to contain *good genetic material*.

The purpose of defining a selection operator is to give preference to better individuals (those solutions which are closer to the global optimum) by allowing them to be chosen as candidates to pass their genes on to the next generation. Since the goodness of each individual depends on its fitness, it is expected that, as the generations evolve, the fitness of the population increases. A critical parameter to be improved in GA is what is known as the *selection pressure*, which is the process of selecting the best individuals for the next generation. If this parameter is low, then the rate of convergence towards the optimum solution is small. If the selection pressure is high, the system is likely to be stuck in a local optimum due to the loss of diversity in the population. Therefore, the selection methods control the selection pressure, which in turn determines how fast the algorithms converge (R.Sivaraj, 2011; Blicke and Thiele, 1995).

In order to evaluate the impact of the selection procedure on the performance of the GA, two additional selection processes were tested against the results obtained with the *roulette wheel* selection methodology (RW), contemplating both the fixed and variable melting time alternatives. These methods, based on *tournament selection*, present several benefits: they are efficient to code, work on parallel architectures and the selection pressure can be adjusted easily.

The methods tested are described as follows:

- *Q-tournament (QT)*: This type of selection involves running several *tournaments* among a few individuals chosen at random from the population. The

winner of each tournament (the one with the best fitness) is selected and used as one of the operands for the genetic operators. The number of individuals chosen to be evaluated in every operation is determined by  $q$ . In the experiments presented,  $q \geq 3$ . A ranked placement operator is used to choose the best two different individuals among parents and children.

- *Simple tournament (ST)* (Alonso et al., 2009): Similar to the  $q$ -tournament, this type of selection involves choosing a random individual at the beginning of the generation and retrieving, in sequential order, the rest of the individuals so as to ensure population diversity. The ranked placement operator is the same as the one used in the Q-tournament approach.

The idea behind the different selection methodologies is that the most successful individuals in each competition will produce better offspring than those which perform poorly. Genes from *good* individuals propagate throughout the population: two good parents will most likely produce offspring that is better than either of them. Thus, each successive generation will become more suited to their environment (Goldberg et al., 1989).

The algorithm, was modified to change its behaviour in accordance to three input parameters: selection mechanism ( $q$ -tournament, or simple tournament), size of the population and number of iterations. Both the selection mechanism and the size of the population influence the diversity of the mating pool and the possibility of creating better candidate solutions based on the chosen parents. The number of iterations will, however, influence the execution time of the algorithm, i.e, time needed in order to give a valid solution. The combination of these three characteristics will determine the amount of optimization achieved within each phase.

Figure 7.9 portrays the fitness landscapes obtained from the execution of the fixed-melting and variable-melting GA with the two new selection procedures and compares them with of the roulette-wheel.

### 7.5.2 Best solution

The optimal solution is given by the combination of population size, number of iterations and methodology selected that gives the best optimization to the foundry problem. Just like in the previous experiments, this optimization is defined in terms of percentage about how much the energy cost reduces (negative value) or not (positive value) the original energy cost for a day of production.

#### Fixed melting time

In case of the GA with fixed melting time, the results point out that the methodology which best minimizes the energy cost is the one given by the combination of the *simple tournament* selection method, a population size of 600 *individuals*, and the algorithm being run over 500 iterations, as shown in 7.9.

**Table 7.9:** Best optimisation achieved per selection method (fixed melting time)

Method	Pop.size	Generations	Time (min)	Cost variation (%)
<i>RW</i>	650	550	98.02	3.12
<i>QT</i>	650	600	76.93	-2.45
<b><i>ST</i></b>	<b>600</b>	<b>500</b>	<b>78.97</b>	<b>-4.68</b>
<i>SA</i>	-	-	103	-3
<i>BA</i>	-	-	165	-7

#### Variable melting time

In case of the GA with variable melting time, the results show that the best solution is given by the *q-tournament* selection method when run with a population size of 650 *individuals* and 500 iterations, as shown in 7.10.

In both cases, the tests of the new selection methods enhance significantly the results obtained by the SA algorithm. Beyond this point, the subsequent iterations of the algorithm with values for parameters over the threshold given for the best optimization, do not provide any improvement.

**Table 7.10:** Best optimization achieved per selection method (variable melting time)

Method	Pop.size	Generations	Time (min)	Cost variation (%)
<i>RW</i>	650	500	78.11	-9.29
<i>QT</i>	650	450	76.75	-10.11
<i>ST</i>	<b>650</b>	<b>500</b>	<b>94.54</b>	<b>-13.60</b>
<i>SA</i>	-	-	128	-10
<i>BA</i>	-	-	291	-15

### 7.5.3 Best solution per time

An important factor that must be taken into account when evaluating the validity of the optimization achieved is the amount of time it takes for the algorithm to deliver the final result. No matter how optimal the solution may be, if the total execution time exceeds that of the deadline, the solution can not be considered valid. Time is crucial in some processes so it is imperative to find a balance between both aspects.

#### Fixed melting time

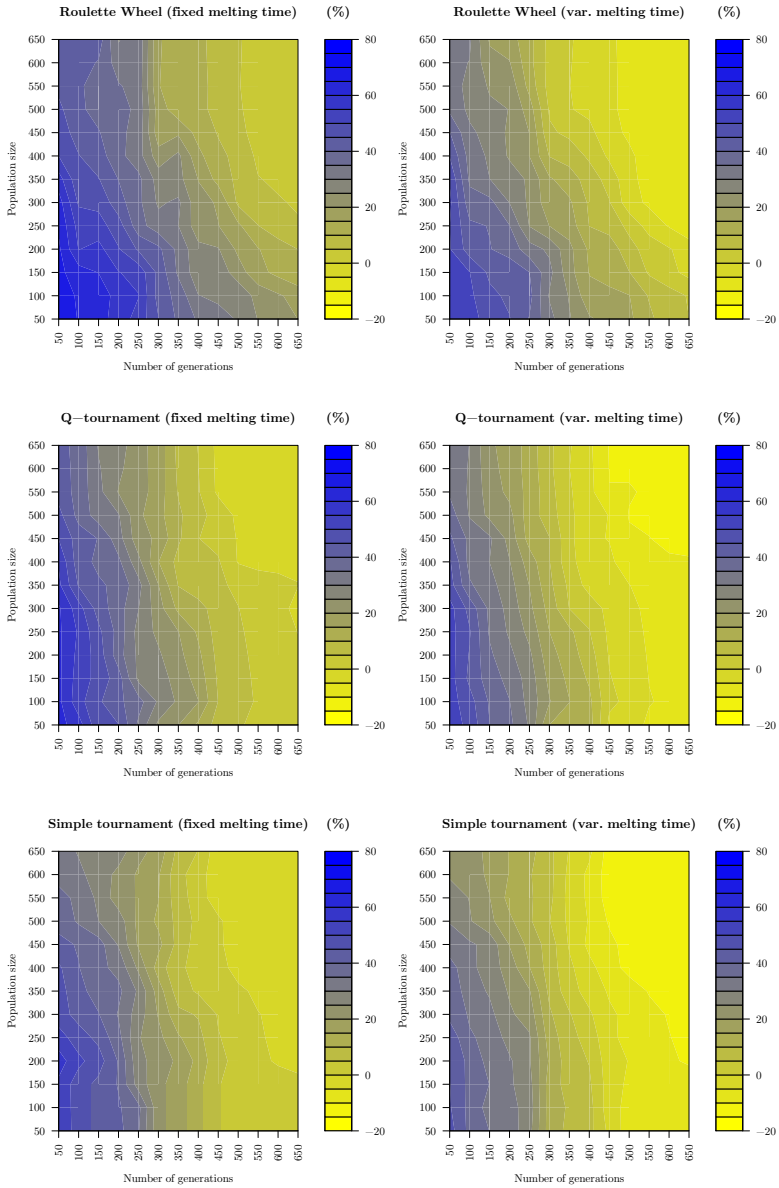
In case of the GA with fixed melting time, the results show that the methodology which best minimizes the energy cost is the one given by the combination of the *simple tournament* selection method, a population size of *550 individuals*, and the algorithm being run over *500 iterations*, as shown in 7.11.

#### Variable melting time

In case of the GA with variable melting time, the results show that the best solution is given by the *simple tournament* selection method when run with a population size of *500 individuals* and *50 iterations*, as shown in 7.12.

## 7.6 Conclusion

Foundry processes consume a significant amount of energy. New regulation laws, which penalize excessive consumption along with the rise in energy prices, call for



**Figure 7.9:** Fitness landscapes for the genetic algorithm with fixed and variable melting time

**Table 7.11:** Best optimization achieved in the fastest time (fixed melting time)

Method	Pop.size	Generations	Time (min)	Cost variation (%/min)
<i>RW</i>	550	650	98.02	3.12
<i>QT</i>	650	300	30.02	-0.84
<b><i>ST</i></b>	<b>550</b>	<b>500</b>	<b>56.71</b>	<b>-4.56</b>
<i>SA</i>	-	-	77.21	-2.78
<i>BA</i>	-	-	113	-0.00032

**Table 7.12:** Best optimization achieved in the fastest time (variable melting time)

Method	Pop.size	Generations	Time (min)	Cost variation (%/min)
<i>RW</i>	650	350	61.41	-8.92
<i>QT</i>	550	50	11.27	-5.77
<b><i>ST</i></b>	<b>500</b>	<b>50</b>	<b>10.45</b>	<b>-7.69</b>
<i>SA</i>	-	-	27.30	-6.32
<i>BA</i>	-	-	315	-0.008

the implementation of energy-efficient approaches concerning their operation. This chapter presents two search space methods used to build a model capable of detailing the appropriate schedule of processes and orders within a foundry plant, taking into account the price of energy for each hour of the day. The performance given by the two optimization approaches (GA and SA) is compared to that given by the BA, which provides the best solution within the search space. It is demonstrated that, though the results given by both optimization algorithms fall behind those given by the BA, an optimization is still achieved in a reasonable amount of execution time. Moreover, an analysis of the influence of the selection method for the GAs analysed and tested statistically.

One of the main drawbacks of the initial analysis was that there was no possibility to modify the aggregations of stock. Being as predefined as they were, the stocks were configured to melt at a certain temperature for a constant period of time. Therefore, there was no possibility to test what consequence the combination of different types

of material would particularly have on the duration of the melting process. In order to overcome this obstacle, a contribution to the state of the art is presented, where these algorithms were modified to accept variable slots of times within the melting phase. The results show a better optimization in terms of energy cost since the algorithms are able to benefit from the increment in genetic welfare. In addition, the introduction of variable slots of times confirms that a modelling of the type and amount of stocks coming together to produce the casting, would result beneficial in terms of making the most of the fluctuation of energy prices when scheduling the processes.

Although the backtracking algorithm stands out as the best approach in terms of optimization, the long execution time and its inability to scale properly, suggests the use of GA or SA in order to achieve an approximate solution in an adequate amount of time. Moreover, the modification of parameters for the genetic algorithm such as the population size and the number of generations along with the selection method chosen, proves that with the adequate combination a better optimization can be achieved in a desirable time.

*Nobody wants a prediction that the future will be more or less like the present, even if that is, statistically speaking, an excellent prediction.*

Nathan Myhrvold

CHAPTER

# 8

## Utility tools for DSM: Short Term Load Forecasting

**L**OAD ANALYSIS represents one of the most important activities within DSM and the overall management of electrical networks. This analysis includes short and long-term forecasting, pattern recognition, variability examination, and even data completion. Specifically, load forecasting is necessary to quantify the amount of electricity that is going to be needed to meet the expected demand. According to the foreseen load the electrical distribution company can elaborate DSM strategies aimed at lowering peak demands or provide incentives and mechanisms to flatten the load curve.

An adequate and accurate load forecasting relies heavily on the quality of the consumption data available and the characterisation of the forecasting scope. However,

some of these data may be missing or corrupt due to glitches or malfunctioning of the existing energy meters. Several works show that load prediction within a specific substation can be improved by adding information from neighbouring primary substations due to the fact that geographically close substations may share common weather-related (e.g. temperature, humidity, wind direction, etc.) and human-related (e.g. work-calendar, holidays, cultural consumption patterns, etc.) features.

This chapter presents an analysis on the effects of retrieving information from neighbouring substation to improve load forecasting by enhancing missing data imputation and by testing this approach on an extensive collection of datasets and experiments. Results show that the consideration of data from surrounding substations significantly improves energy load forecasting while lowering data imputation errors.

## 8.1 Introduction

The electrical grid has been designed to attend the electricity demand of its users, regardless of changes and fluctuations (Labeeuw and Deconinck, 2013). An accurate prediction has crucial consequences on day-ahead network operation plans, for the tiniest bias or error may increase operation costs, yielding losses ranging from hundreds of thousands or even millions of euros to the operator (Alfares and Nazeeruddin, 2002a).

Thus, forecasting or the art of predicting the behaviour of different-scale loads (e.g. country-wide (Khosravi and Nahavandi, 2014; Goude et al., 2014; Hinojosa and Hoese, 2010; Terciyani et al., 2014), zone-wide (Borges et al., 2013a; Xu et al., 2012b), building (Borges et al., 2013b; Pournaras et al., 2014; Sechilariu et al., 2013), etc.) plays an critical role in the real-time control and operation of the grids and it remains an essential task for the electric utilities. There are many factors that affect demand loads, such as climate changes, usage patterns, season, day of the week, time of the day, etc.

There have been two types of solutions to cope with this challenge, namely Statistical methods as ARIMA (Fusco and Ringwood, 2010), Multiple Linear Regression

(MLR) (Hong et al., 2011), Semiparametric (Liu et al., 2006) or Non-Parametric Regression (Vilar et al., 2012) and AI methods from Neural Network (NN) (Hassan et al., 2012; Agarwal et al., 2006), to Support Vector Machines Super Vector Machines (SVM) (Jain and Satish, 2009). Each one present its pros and cons; whereas statistical methods show poorer records than their counterparts (Alfares and Nazeeruddin, 2002a; Hinojosa and Hoese, 2010), they do not have drawbacks such as difficult parametrization, non-obvious selection of variables and over-fitting. Furthermore, in the praxis they normally require much more historical data to *discover* the patterns inherent on it (Hinojosa and Hoese, 2010; De Silva et al., 2011).

The work presented in (Borges et al., 2013a) explored an intuition that had already been explored an intuition that had already been validated in other disciplines (Osborne et al., 2011; Chaouch, 2014; Goncalves Da Silva et al., 2014; Xie et al., 2014): adjacent zones, sharing commonalities such as similar weather, may help improve the forecasting of a single substation. The results of this work were promising but lacked generality. One of the main prerequisites of almost all forecasting methods is to have a complete database of historical data. However, as all practitioner know, this is hardly ever the case. The art of completing the missing information in a database is known as *data imputation*. While there is a large knowledge about the data imputation problem as a technique (Schafer and Olsen, 1998; Myrtveit et al., 2001; Ibrahim et al., 2005), there are not many previous references on the context of electric load information. To our knowledge, the only attempt to tackle this problem in this particular context is the seminar work presented in (Mateos and Giannakis, 2013, 2012; Kim and Giannakis, 2013).

## 8.2 Data imputation vs Forecasting

Data imputation and forecasting are two approaches that share a similar structure, but yield rather different results. *Forecasting* deals with predicting the future evolution of a single variable by analysing information from the past, whether it relates to the previous behaviour or states of said variable or the behaviour of other related variables that may be of direct influence. Specifically, in this research, forecasting is

focused on predicting the amount of energy that is going to be consumed, or needs to be delivered, at a primary substation during the next hour. This restriction hinders the search for appropriate solutions since it precludes the use of information stemming from time-points beyond the one being analysed, i.e., weather data from 12:00h can not be used to predict the energy load at that hour; instead, the methodology must either use past data or use *predicted* weather data for 12:00h.

On its part, *data imputation* deals with figuring out the value for a specific variable in the near past. Data imputation is useful when, for instance, the reading of a meter is either missing or labelled as incorrect. To clarify the example, suppose that a primary substation is missing data from 12:00h due to a malfunction in the metering process, and that said data may be needed to predict the amount of energy load at 13:00h. In order to fill the gap at 12:00h, we may use data from previous time-points (e.g. 10:00h, 11:00h), as well as data from following time-points (e.g. 13:00, 14:00). This research, in particular, will focus on the *online* data imputation problem, in which the missing values are the last values of a time-series.

Another basic difference between *forecasting* and *data imputation* is the information that is used as input for each methodology. On the one hand, *forecasting* predicts the value of load  $L$  at time  $t + h$  with  $h \geq 1$  by using data previous to time  $t$ . As such, the following equation is a valid forecasting model for  $L$ .

$$L_{t+h} = f(L_t, \dots, L_{t-i}, X_t, \dots, X_{t-j}), \quad (8.1)$$

where  $i, j$  are two integer constants,  $f$  represents the forecasting function and  $X$  refers to a set of exogenous variables.

On the other hand, *data imputation* estimates the value of load  $L$  at time  $t + h$  with  $h \geq 1$ , but adding information given by exogenous variables prior to time  $t + h$ . As such, the following equation is a valid imputation model for  $L$ :

$$L_{t+h} = f(L_t, \dots, L_{t-i}, X_{t+h}, \dots, X_{t-j}), \quad (8.2)$$

where  $f, X, i$  and  $j$  are two integer constants, just as in Equation (8.1). Please note that, for both cases,  $t$  is measured in hours.

### 8.2.1 Short Term vs Very Short Term

Practitioners include another differentiation for both *forecasting* and *data imputation* depending on the prediction time frame. If the prediction time frame spans between 1 and 24 hours, the methodology is known as Very Short Term Load Forecasting (VSTLF) or Very Short Term Load Imputation (VSTLI). On the contrary, if the prediction time frame spans between 24 and 168 hours, the methodology is known as Short Term Load Forecasting (STLF) or Short Term Load Imputation (STLI). In either case, the objective is to predict load  $L$  from  $t + 1$  to  $t + h$ .

The importance of this classification lies in the fact that both approaches are needed by the energy utility on different management levels. While STLF/STLI are directly related to demand-response or energy purchase scenarios, VSTLF/VSTLI are usually related to the immediate or real-time control of the energy grid.

## 8.3 Materials and methods

### 8.3.1 Baseline model

The definition of a baseline model is a basic step in the process of establishing a reference against which new forecasting and data imputation techniques are to be tested. For this case, the model chosen as baseline is the Naïve Multiple Linear Regression Benchmark (NMLRB) provided in (Hong et al., 2011), in which the authors introduce a methodological approach based on MLR to build STLF benchmarking models for comparative assessments. The complete model was successfully tested in the 2012 Global Energy Forecasting Competition (GEFC12) (Hong et al., 2014), winning the top prize with an impressive 3 % Mean Absolute Percentage Error (MAPE) error (see 8.4 for more information), 2 points ahead of the second best model.

The NMLRB model, which presents an important interaction effect among variables, requires values for temperature, month, hour, and day-type as input data. Specifically, this model considers the following day-types  $\{mon, tue-thu, fri, sat, sun\}$  as well as special dates such as Christmas or Easter. Moreover, square and cubic powers terms of the temperature are also included in order to approximate a non-linear

behaviour as follows:

$$L_{t+h} = \beta_0 + \beta_1 Tr + \beta_2 D_t + \quad (8.3)$$

$$+ \beta_3 H_t + \beta_4 M_t + \beta_5 D_t H_t + \quad (8.4)$$

$$+ \beta_6 M_t T_t + \beta_7 M_t T_t^2 + \beta_8 M_t T_t^3 + \quad (8.5)$$

$$+ \beta_9 H_t T_t + \beta_{10} H_t T_t^2 + \beta_{11} H_t T_t^3 + \quad (8.6)$$

$$+ \beta_{12} L_t + \beta_{13} L_{t-1} + \beta_{14} L_{t-2} + \quad (8.7)$$

$$+ \beta_{15} L_{t-24} + \beta_{16} L_{t-25} + \beta_{17} L_{t-26}, \quad (8.8)$$

where  $L(t)$  denotes the load at hour  $t$ ,  $Tr$  the trend,  $D_t$  represents the day-type at time  $t$ ,  $H_t$  the hour of the day at time  $t$ ,  $M_t$  the month at time  $t$ ,  $T_t$  the temperature at time  $t$ , and  $\beta_i$  are real coefficients.

The NMLRB model can be decomposed in two major components: the *auto-regressive* component and the *meteorological* component. The auto-regressive component (lines (8.7) and (8.8)) captures the dynamic of the load profile by analysing and estimating its *shape* based on previous patterns. The meteorological component (lines (8.5) and (8.6)) is responsible for capturing the *height* of the load profile, in other terms, the amount of energy to be consumed at each moment. Furthermore, the NMLRB model has a set of *constant terms* (lines (8.3) and (8.4)) that are defined separately due to the difficulty of assigning ownership to either of the two aforementioned components.

Among this paradigm, the contribution of the current research lies in the addition of a third component, *neighbouring zones' load profiles*, to the original components already defined in the NMLRB model. The objective is to analyse whether this approach improves energy load forecasting and data imputation accuracy for both STLF/STLI and VSTLF/VSTLI. As such, the additional terms to be included to the

**Table 8.1:** Description of the components each of proposed model.

Model name	Auto-regressive	Meteorological	Neighbourhood
<i>RW</i>	✗	✗	✗
<i>AUTO</i>	✓	✗	✗
<i>MET</i>	✗	✓	✗
<i>NBHD</i>	✗	✗	✓
<i>LF</i>	✓	✓	✗
<i>NAUTO</i>	✗	✓	✓
<i>NMET</i>	✓	✗	✓
<i>GLF</i>	✓	✓	✓

NMLRB model are as follows:

$$\sum_{k \in K} \gamma_1^k L_t^k + \gamma_2^k L_{t-1}^k + \gamma_3^k L_{t-2}^k + \quad (8.9)$$

$$+ \gamma_4^k L_{t-24}^k + \gamma_5^k L_{t-25}^k + \gamma_6^k L_{t-26}^k, \quad (8.10)$$

where  $K$  is the set of neighbouring zones of the energy substation being modelled,  $L_t^k$  represents the energy load of the neighbouring zone  $k$  at time  $t$ , and  $\gamma_i^k$  refers to the model parameters that need to be estimated.

### 8.3.2 Proposed models

In order to assess the contribution of the three components of the new model, *auto-regressive*, *meteorological*, and *neighbouring*, in terms of *forecasting* and *data imputation* accuracy, there is a need for a comprehensive study that shows the advantages and disadvantages of the model combinations. 8.1 summarises the main features of the models proposed for this research.

- **AUTO.** Composed of lines (8.7) and (8.8) of the NMLRB model, this refers to the traditional auto-regressive approach. This model is expected to perform reasonably well for *forecasting* approaches since it is able to capture the shape of the energy load based on the analysis of previous patterns. In the case of

*data imputation*, however, the model will perform worse due to its inability to capture the existing correlation between the substation load being analysed and its neighbouring substations' loads.

- **MET.** Composed of lines (8.5) and (8.6) of the NMLRB model, this model accounts only for the meteorological component. It is expected to perform badly in both *forecasting* and *data imputation* problems since it barely captures the dynamic behaviour of the energy load profile, and can not exploit the correlation with other substations' loads.
- **NBHD.** Composed of lines (8.9) and (8.10), this model focuses only on the correlation between the current substation and its neighbouring substations. The model will perform similar or better than the *AUTO* model, for both *forecasting* and *data imputation* problems due to its auto-regressive characteristics. In fact, this model does not only extract the energy load profile of the current substation, but also the correlation with neighbouring substations.
- **LF.** Being one of the most used models to issue a STLF, this refers to the complete NMLRB model proposed in (Hong et al., 2011). It is expected to perform well on both *forecasting* and *data imputation* problems as it is able to completely capture the dynamic behaviour of the energy load.
- **NAUTO.** Composed of lines (8.5), (8.6) of the NMLRB model, and lines (8.9) and (8.10) of the proposed improvement, this model is similar to the *LF* model. The difference is that the *auto-regressive* component is swapped for the *neighbouring* component. It is expected to perform reasonably well since it can retrieve the energy load dynamics by taking advantage of the correlation with neighbouring substations, and estimate the height of the energy load through the analysis of the meteorological variables.
- **NMET.** Composed of lines (8.7), (8.8), (8.9) and (8.10) of the baseline model, this model is composed of the *meteorological* and *neighbouring* components. This model learns quite well the dynamics of the energy load curve, however the effect of the meteorological variables will be inferred through the *neighbouring* component. Therefore, the results should be quite similar as those of the *AUTO* model for both *forecasting* and *data imputation*.

- **GLF.** This is the full model that comprises the two original components from the NMLRB model, *auto-regressive*, and *meteorological*, as well as the *neighbouring* addition proposed in this research. It is composed of lines (8.7), (8.8), (8.5), (8.6), (8.9) and (8.10). The results of this model are expected to outperform any of the results given by the other models.  
Please note that all of the aforementioned models also contemplate the *constant terms* defined in lines (8.3) and (8.4).
- **RW.** Random-walk model *RW* defined as  $L_{t+h} := L_t$ . This model does not include any of the three components previously explained and is used as contrast model.

### 8.3.3 Modification of the models to fit VSTLF, STLI and VSTLI

The NMLRB model was originally designed to satisfy the restriction of the STLF approach. The authors also adapted it to fit VSTLF and Long Term Load Forecasting (LTLF) methodologies. The former was just a matter to define  $h = 1$ , whereas the latter consisted on adding socio-economics variables such as Gross Domestic Product (GDP), labour force, population growth, etc.

In the same sense, these modifications were also applied to our *GLF* model so as to adapt the model for the STLF/STLI and VSTLF/VSTLI. Specifically, for the *data imputation problems*, all components of the model were adapted except for the *auto-regressive* part (i.e. lines (8.3), (8.4), (8.5), (8.6), (8.9) and (8.10)) to include terms until  $t + h$ . This brings similar characteristics to some models when applied to different problems. For example, the *MET* model will be the same for both STLF and VSTLF, while the *AUTO* model will be the same for STLF and STLI.

## 8.4 Experimentation

### 8.4.1 Experimental Settings

The experiments were carried out on two x64 workstations using Weka 3.8 (Hall et al., 2009) with the *LinearRegression* model, trained using Weighted Least-Mean

Squares. For each of the datasets provided, each of the proposed model was trained with a whole year of data. Once built, the performance of the model was tested by measuring the error between predicted or imputed data and the real data from the original datasets. The forecasting or data imputation horizon was of 100 days for STLf/STLI and of 240 hours for VSTLf/VSTLI, selected at random and not included in the training set.

Error quantification for each of the different models was addressed through the use of MAPE: a simple unit-free measure that can straightly compare the results obtained among several datasets, and is one of the most widely-used error measurements in STLf. MAPE is calculated as follows:

$$\text{MAPE} := \frac{100}{|V|} \sum_{t \in V} \frac{|R_t - L_t|}{|R_t|},$$

where  $V$  the validation set,  $|V|$  its cardinality,  $R_t$  the real load at time  $t$  and  $L_t$  the predicted or imputed one.

## 8.4.2 Datasets

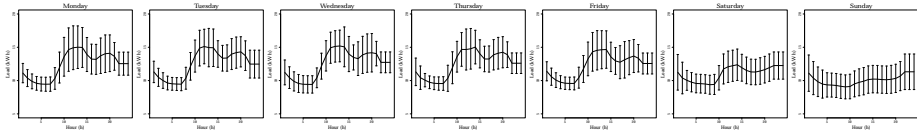
The experiments were carried out on six different datasets, chosen as such to cover the entire energy scope from one of the biggest world electricity markets to an isolated group of four monitored buildings. Each dataset consists of several time-series containing energy load measurements from different system points, typically geographically *close* substations. The information was either obtained directly from the energy meter, or from the websites of the TSO, hence the private status of two of the studied datasets. The main characteristics of each dataset are detailed as follows:

- **Global Energy Forecasting Competition (GEFC12) (Hong et al., 2014)** Used in the GEFC12 competition, this dataset belongs to an unknown TSO from the USA. It contains energy and weather measurements from January 2004 to June 2008 for twenty zones, labelled as such with numbers from 1 to 20. Most of the zones share a similar energy load profile, but some of them contain erroneous data. Specifically, zone 3 and 7 are almost identical while zone 2 is a

scaled down mirror of zone 6. Zone 9, on its part, does not seem to share any similarities with the rest of the datasets, particularly in relation to the weather variables provided.

For this research, the WEKA model was trained with data from 2004/01/01 to 2004/12/31. In fact, the datasets contain, for each of the twenty zones, a first year of complete measurements with data missing from the following eight weeks. The aim of the competition was to fill these gaps as well as to predict the following week after the end of the dataset for both the twenty zones and the aggregated load.

- **Pennsylvania, Jersey, and Maryland Interconnection (PJM) (PJM Interconnection System Operator, 2014)** This dataset belongs to the PJM Interconnection that comprises the states of Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia, and the District of Columbia. PJM is one of the biggest electricity market systems in the world. Energy data for five primary substations (AEP, COMED, DAY, DOM, DUQ) was recorded from January 2006 until December 2010. Unlike the GFCE12 dataset, all the PJM zones present in the a regular energy profile. For this dataset, the WEKA model was trained with data from 2006/01/01 to 2006/12/31.
- **New York Independent System Operator (NYISO) (New York Independent System Operator, 2001)** This dataset contains energy measurements of eleven primary substations located in New York City (CAPITL, CENTRL, DUNWOD, GENESE, HUDVL, LONGIL, MHKVL, MILLWD, NYC, NORTH and WEST), ranging from June 2007 to October 2010. Particularly, the NORTH zone presents a *flat* load curve, quite different from any of the load profiles of the remaining zones. This could be explained by the fact that NORTH may be a substation supplying an industrial area, hence the particularity of its load profile. For this dataset, the WEKA model was trained with data from 2007/01/01 to 2007/12/31.
- **Canary Islands (CAN) (Red Eléctrica de España, 2014)** This dataset contains energy load profiles of the Canary Islands, a Spanish archipelago located just



**Figure 8.1:** Average daily load for zone PDT of dataset Ciudad Real. Error bars denote  $\pm\sigma$ .

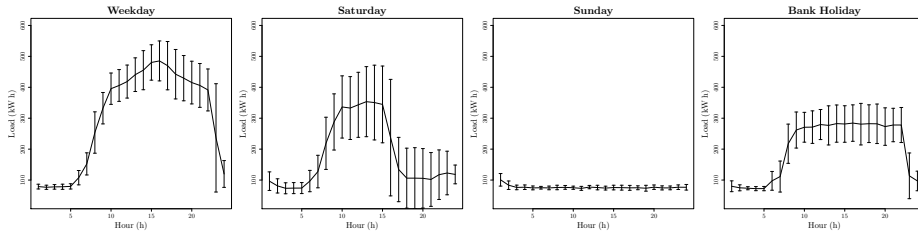
off the Northwest coast of mainland Africa. The energy measurements, ranging from March 2012 until May 2013 and given by the Spanish TSO, were collected from the primary substations located on each of the seven islands (HIERRO, FUERTEVE, GCANARIA, GOMERA, PALMA, LANZAROT and TENERIFE). In this case, all zones show a similar and homogeneous energy load profile.

For this dataset, the WEKA model was trained with data from 2012/03/12 to 2013/03/11.

- **Ciudad Real (CR)** This dataset belongs to the province of Ciudad Real, located in the south-western area of the autonomous community of Castilla-La Mancha, Spain. The dataset contains records from 10 substations (CIU, CTV, DA2, DAI, MAL, PDP PDT, PEM, PLM, SOL and VPE), collected from January 2009 until December 2012 by the DSO). All zones present a uniform behaviour, differentiating among weekdays, Saturdays and Sundays (as portrayed in 8.1). Each substation, supplies to a different mix of consumer types.

For this dataset, the WEKA model was trained with data from 2009/01/01 to 2009/12/31.

- **Bilbao Campus of the University of Deusto** This dataset contains energy measurements of four different buildings in the Bilbao Campus of the University of Deusto (CEN, CRAI, GEO and LIT), from September 2012 to December 2013. Three of these buildings are standard university buildings (*i.e.* classrooms and teachers' offices), while the remaining one refers to the data of the campus main library. The load profile presents a different behaviour depending on the day and scholar calendar. For example. 8.2 shows the typical behaviour of the main library, where the peak activity takes place on weekdays and Saturdays until late at night.



**Figure 8.2:** Typical load for CRAI building in the University of Deusto. Error bars denote  $\pm\sigma$ .

### 8.4.3 Weather Data

Weather data for datasets with energy measurements of substations in the USA was obtained from Wunderground<sup>1</sup>. Similarly, weather data for Spanish zones was procured by AEMET<sup>2</sup>. In either case, for the PJM, NYISO, CR, CAN and UD datasets, the data was retrieved from the weather substation closest to each energy substation. The GEF12 dataset, in particular, already provided anonymised weather data (i.e. there is no information about the relationship between the meteorological information and the zone it corresponds to). For this case, a brute-force test training with only one year of training data and six months of forecasting was performed, pairing every weather station with every zone in order to highlight pairs with the best results. The outcome of this approach is summarised in 8.2.

## 8.5 Results and discussion

Experimental results derived from the execution of each proposed model are summarised in 8.3. Each column shows the mean and standard deviation of the MAPE achieved by each model in all the datasets for the STLF/STLI and VSTLF/VSTLI approaches previously described. 8.5 8.6 display the results in full detail.

As expected, the *RW* model presents the worst results. On the contrary, the proposed model (*GLF*) achieves better results for each the four problems than those

<sup>1</sup><http://www.wunderground.com/>

<sup>2</sup><http://www.aemet.es/>

**Table 8.2:** Weather Station Selection Results for GEFC12 (MAPE %). In bold the selected weather station for every substation.

Zone	WS <sub>1</sub>	WS <sub>2</sub>	WS <sub>3</sub>	WS <sub>4</sub>	WS <sub>5</sub>	WS <sub>6</sub>	WS <sub>7</sub>	WS <sub>8</sub>	WS <sub>9</sub>	WS <sub>10</sub>	WS <sub>11</sub>
1	7.94	8.13	7.75	7.95	<b>7.69</b>	7.86	8.13	8.14	7.77	8.18	8.10
2	8.17	8.06	7.90	8.12	8.09	<b>7.58</b>	7.60	7.72	7.99	8.06	7.96
3	7.79	<b>7.39</b>	8.17	7.89	8.14	7.97	7.89	7.94	8.16	8.16	8.13
4	7.67	7.85	7.82	7.91	7.94	8.12	<b>6.69</b>	8.11	7.28	8.11	7.70
5	7.90	7.95	8.10	7.99	8.10	7.98	8.02	8.13	8.07	<b>7.53</b>	8.11
6	7.88	7.78	7.92	7.86	7.86	<b>6.80</b>	7.93	7.92	8.12	7.89	8.10
7	7.64	<b>6.21</b>	8.15	8.06	7.99	6.32	8.08	8.09	7.91	7.89	8
8	7.78	8.14	7.87	8.05	7.90	<b>6.54</b>	7.95	7.83	7.95	8.08	8.14
9	7.97	8.11	8.14	8.17	8.10	<b>7.87</b>	8.03	8	7.98	7.99	8.02
10	8.19	7.93	8.13	8.02	8.15	8.12	<b>7.70</b>	8.02	8.17	8.19	8.09
11	7.92	8	7.91	7.90	7.54	8.01	7.98	7.96	8.15	7.96	<b>7.35</b>
12	7.86	7.92	7.96	<b>6.87</b>	8.11	8.18	7.96	7.55	7.93	8.11	7.96
13	7.49	6.96	8.11	8.16	7.56	7.99	7.88	8.16	7.95	<b>6.59</b>	7.92
14	7.88	7.73	7.39	8.17	8.11	7.93	8.12	<b>6.96</b>	7.95	7.94	8.10
15	7.93	7.56	7.82	8.12	<b>6.54</b>	7.83	8.17	8.08	8	8.04	8.15
16	7.76	<b>7.13</b>	8.04	8.17	7.78	7.67	8.12	7.97	8.10	8.16	7.94
17	<b>7.84</b>	8.16	7.95	8.11	8.12	7.91	8.05	7.99	8.08	8.17	8.16
18	7.98	7.95	8.05	7.87	8.10	7.94	8.08	7.88	<b>7.86</b>	8.11	7.90
19	8.01	7.90	7.89	<b>7.85</b>	7.95	7.97	8.12	8.13	8.05	8.14	8.06
20	8.15	7.84	8.04	7.86	8.08	<b>7.63</b>	8.04	7.90	7.86	7.95	7.96

**Table 8.3:** Summary of the results. Mean MAPE  $\pm \sigma$  (%)

Problem	<i>RW</i>	<i>AUTO</i>	<i>MET</i>	<i>NBHD</i>	<i>NMET</i>	<i>NAUTO</i>	<i>LF</i>	<i>GLF</i>
<i>VSTLF</i>	6.98 $\pm$ 2.78	5.07 $\pm$ 3.11	8.48 $\pm$ 6.85	6.31 $\pm$ 6.91	2.76 $\pm$ 2.02	4.85 $\pm$ 5.83	2.78 $\pm$ 2.11	2.63 $\pm$ 2.11
<i>VSTLI</i>	6.98 $\pm$ 2.78	5.07 $\pm$ 3.11	8.48 $\pm$ 6.85	5.69 $\pm$ 7.16	2.74 $\pm$ 2.42	5.19 $\pm$ 6.58	2.78 $\pm$ 2.11	2.67 $\pm$ 2.49
<i>STLF</i>	13.41 $\pm$ 7.57	12.23 $\pm$ 10.36	8.38 $\pm$ 7.16	9.69 $\pm$ 7.42	8.91 $\pm$ 6.31	7.85 $\pm$ 7.37	7.37 $\pm$ 6.45	7.25 $\pm$ 6.57
<i>STLI</i>	13.41 $\pm$ 7.57	12.23 $\pm$ 10.36	8.38 $\pm$ 7.16	6.1 $\pm$ 8.05	5.41 $\pm$ 6.74	5.68 $\pm$ 7.72	7.37 $\pm$ 6.45	5.25 $\pm$ 6.88

of the base model (*LF*). In between, the picture is blurry and seems to be problem-dependent, which makes necessary to discuss the behaviour of each model individually.

The MAPE of the models for *VSTLF* and *VSTLI* are smaller than their corresponding *STLF* and *STLI*, as seen in Tables 8.5 and 8.6. The reason behind this phenomenon is that the prediction or data imputation being issued is of an immediate and short temporal horizon, where energy processes do not fluctuate much over time. The separate comparison of the output of the best models for *forecasting* and *data imputation* drew out more interesting conclusions. When considering the Very Short Term time horizon, the MAPE difference among models is small and there is little margin for improvement. However, the assessment of the results for the Short Term version showed the existence of a big difference in favour of the data imputation version of the models. This results are coherent considering that, for data imputation, the inclusion of information of neighbouring substations is similar to issuing the Very Short Term version of the problem.

Following this results, a Friedman rank test with Bergmann *post hoc* (see (Derrac et al., 2011)) was carried out in order to ascertain whether the differences between models were statistically significant. With a significance level  $\alpha$  was set at 0.05, there exist statistical differences between the models in all cases since the p-values are way below  $\alpha$ , as seen in Table 8.4.

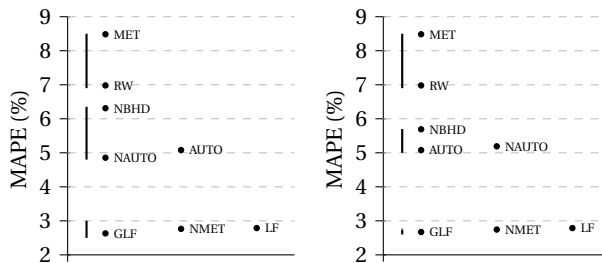
**Table 8.4:** P-values of the Friedman test for the comparison of the models.

Problem	<i>VSTLF</i>	<i>STLF</i>	<i>VSTLI</i>	<i>STLI</i>
p-value	$1.09 \cdot 10^{-10}$	$1.62 \cdot 10^{-10}$	$1.06 \cdot 10^{-10}$	$1.21 \cdot 10^{-10}$

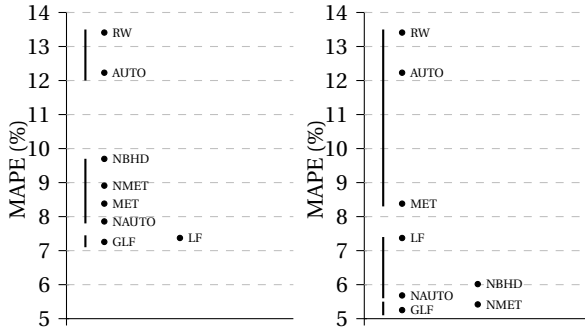
The results of the Bergmann procedure are summarised in 8.38.4, where the models have been ranked according to the results in terms of MAPE on the vertical axis. Models marked or classified together with a line are models among which no statistical differences were found using the Bergmann procedure. The placement of the models on the horizontal axis is exclusively an aesthetic issue. The figures show that, in all cases, the models can be classified into three groups: the group containing the *RW* model (worst group), the group that contains the *GLF* model (best group), and a group with the rest of the models (middle group).

Specifically, for VSTLF and VSTLI, the groups are quite homogeneous: the worst group is composed by the *MET* and *RW* models. The behaviour of the *MET* model is remarkable, having performed worse than the *RW* model. The middle group is composed by *NBHD*, *AUTO* and *NAUTO*. Finally, the best group is composed by *GLF* and *NMET*. Please note that, for VSTLI, *LF* forms its own group, while for VSTLF it is included in the best group.

On the contrary, for STLF and STLI, the groups are quite different, the only similarity being that the worst group is also composed of the *RW* and *AUTO* models. While for STLF, the best group is composed of the *LF* and *GLF*, for STLI, the best group is composed by *GLF* and *NMET* (with *NAUTO* and *NBHD* being close in terms of MAPE). The common characteristic among this models is that they all consider information given by neighbouring substations, demonstrating that the addition of the *neighbouring* part to the original model accomplishes an important improvement.



**Figure 8.3:** Summary of the results for the Very Short Term approach. Left figure displays the results for VSTLF. Right figure displays the results for VSTLI.



**Figure 8.4:** Summary of the results for Short Term problems. Left figure displays the results for STLF. Right figure displays the results for STLI.

## 8.6 Conclusions

This chapter focuses on how load forecasting can be improved by retrieving information from nearing substations. This improvement directly influence the accuracy of predictions and in turn the DSM strategies to be implemented by the utility. In this way, a cutting-edge STLF algorithm on four different situations, ranging from classical STLF and VSTLF to STLI and VSTLI. The objective of this integral validation was to verify whether adding surrounding data enhances the results of data imputation, as it does with STLF. Results show that adding adjacent substation data ameliorates the quality of both forecasting and data imputation.

**Table 8.5:** Results for the data imputation problem. Left table contains the results for STLI while right table contains VSTLI. MAPE (%).

Dataset	AUTO	RW	NMET	GLF	NAUTO	LF	NBHD	MET	Dataset	AUTO	RW	NMET	GLF	NAUTO	LF	NBHD	MET
HIERRO	6.51	8.30	5.40	5.48	5.55	5.65	5.50	5.88	HIERRO	3.79	5.59	3.41	3.42	5.42	3.62	5.34	5.81
FUERTEVE	2.86	4.73	1.91	1.89	2.22	2.17	2.23	3.30	FUERTEVE	1.64	5.70	1.24	1.28	2.31	1.36	2.28	3.57
GCANARIA	3.36	4.13	1.61	1.68	1.78	1.95	1.67	2.41	GCANARIA	1.46	4.78	0.93	0.96	1.92	1.07	1.84	3.05
GOMERA	3.19	4.84	2.53	2.49	2.93	2.65	3.09	3.80	GOMERA	1.75	4.38	1.52	1.59	2.92	1.67	2.98	4.43
PALMA	4.59	6.32	3.73	3.85	4.80	3.96	4.79	5.84	PALMA	2.39	6.14	1.97	2.00	4.41	1.95	4.43	6.29
LANZAROT	2.43	3.70	1.35	1.34	1.41	1.79	1.41	2.28	LANZAROT	1.37	4.69	0.92	0.96	1.52	1.11	1.53	2.59
TENERIFE	2.75	3.93	1.87	2.02	2.31	2.29	2.28	2.91	TENERIFE	1.40	4.93	0.76	0.83	2.07	0.93	2.03	3.36
CIU	10.82	10.03	9.22	9.43	13.24	9.77	12.87	12.95	CIU	6.14	9.24	3.98	4.03	10.82	4.82	10.51	18.51
CTV	7.90	9.90	5.14	4.52	4.67	4.97	5.32	5.37	CTV	5.24	7.26	4.02	4.03	5.50	4.54	5.66	9.83
DA2	12.34	13.24	11.06	11.90	12.77	11.43	12.45	13.34	DA2	6.99	8.03	6.22	6.10	12.68	5.94	14.02	15.03
DAI	12.17	14.71	12.02	11.32	12.41	10.99	14.35	12.41	DAI	7.33	6.89	6.87	6.42	16.62	7.22	16.86	21.13
MAL	7.56	9.17	4.68	4.54	4.85	5.90	5.11	6.74	MAL	4.57	6.44	3.58	3.51	4.82	3.95	5.00	6.95
PDP	7.48	10.18	6.44	6.51	7.90	7.01	8.16	8.91	PDP	5.02	6.26	4.54	4.47	7.67	4.53	7.99	8.47
PDT	7.68	7.46	4.96	4.93	5.19	4.95	5.41	5.66	PDT	4.65	5.24	3.68	3.56	5.34	4.02	5.81	8.74
PEM	13.97	16.18	11.11	11.26	11.55	12.98	11.41	14.19	PEM	5.37	4.35	5.20	5.45	9.05	5.62	8.44	12.03
PLM	12.48	14.51	9.55	10.07	10.18	12.08	10.18	14.11	PLM	7.07	8.52	6.48	6.34	10.47	6.47	10.56	13.71
SOL	10.80	14.06	7.13	7.22	7.64	8.58	7.56	9.94	SOL	6.23	8.62	5.82	5.28	7.16	5.41	8.04	9.93
VPE	11.57	10.60	7.72	7.70	8.13	7.89	8.22	9.30	VPE	4.64	5.44	4.40	4.08	8.23	4.19	8.23	10.26
1	12.03	18.08	5.34	5.04	4.85	8.90	5.40	10.16	1	6.54	8.93	1.85	1.77	4.88	3.80	5.43	10.25
10	10.13	14.23	4.94	4.76	5.65	6.36	5.25	7.03	10	2.08	6.23	1.68	1.63	4.71	1.64	4.73	6.64
11	10.26	16.17	1.85	1.92	1.94	6.60	1.88	7.78	11	5.03	7.26	1.06	0.95	1.51	1.93	1.56	7.30
12	11.71	18.02	2.50	2.43	2.51	8.35	2.58	9.80	12	6.07	8.22	1.28	1.23	2.18	3.08	2.29	8.78
13	9.35	14.75	3.18	2.71	2.80	6.67	3.29	8.03	13	3.64	8.23	2.45	2.35	2.83	3.30	3.08	8.31
14	16.65	23.87	4.00	4.00	4.10	9.10	4.16	10.55	14	4.35	9.99	2.98	2.94	3.85	3.66	4.05	10.09
15	10.93	16.79	3.45	3.40	3.53	7.89	3.62	9.53	15	3.65	7.51	2.78	2.68	3.38	3.37	3.42	9.08
16	14.14	22.42	3.44	3.13	3.22	8.94	3.35	10.99	16	3.29	8.62	2.34	2.16	2.83	2.86	3.03	10.85
17	10.04	14.91	2.21	2.22	2.24	5.67	2.24	6.39	17	2.80	6.25	1.68	1.58	2.18	1.90	2.31	6.28
18	11.59	18.99	2.11	2.08	2.16	6.70	2.14	7.95	18	3.60	8.49	1.70	1.51	1.87	2.35	2.07	7.59
19	13.48	20.83	2.60	2.61	2.63	8.64	2.59	10.01	19	3.87	9.23	2.16	2.12	2.43	3.26	2.40	10.01
2	6.56	11.68	0.00	0.00	0.00	4.82	0.00	5.73	2	1.94	5.28	0.00	0.00	0.00	1.04	0.00	5.15
20	7.43	12.24	1.99	1.79	1.91	5.38	2.14	6.20	20	2.37	6.47	1.55	1.41	1.71	2.01	1.82	5.73
3	6.56	11.68	0.00	0.00	0.00	4.22	0.00	4.93	3	1.93	5.28	0.00	0.00	0.00	1.04	0.00	4.66
4	7.94	12.04	4.90	4.90	5.17	6.72	5.14	7.63	4	3.88	8.66	3.46	3.40	5.58	3.38	5.63	7.06
5	12.30	20.01	3.57	3.43	3.53	8.42	3.65	10.05	5	3.95	10.65	2.08	1.95	2.93	2.12	3.19	8.45
6	6.71	11.95	0.00	0.00	0.00	4.91	0.00	5.75	6	2.21	5.46	0.00	0.00	0.01	1.04	0.00	5.15
7	6.56	11.68	0.00	0.00	0.00	4.22	0.00	4.93	7	1.93	5.28	0.00	0.00	0.00	1.04	0.00	4.66
8	9.02	15.38	4.23	3.88	3.89	7.25	4.25	8.10	8	2.34	7.56	1.87	1.72	3.64	1.69	4.02	7.12
9	51.13	51.05	42.34	42.99	50.57	42.51	53.19	49.54	9	14.08	17.57	12.69	13.69	41.65	13.47	43.32	43.39
CAPITL	11.52	10.80	2.21	2.07	2.01	3.37	2.13	3.52	CAPITL	6.90	5.72	1.63	1.59	1.85	2.17	1.97	3.36
CENTRL	9.22	8.54	1.51	1.50	1.51	2.90	1.55	3.00	CENTRL	4.47	4.57	1.25	1.15	1.33	1.77	1.42	2.73
DUNWOD	12.84	12.22	3.03	2.85	2.84	4.26	2.99	4.78	DUNWOD	6.32	6.24	1.90	1.92	2.52	2.77	2.72	4.95
GENESE	10.35	9.34	1.85	1.57	1.56	2.95	1.83	3.04	GENESE	5.07	5.23	1.63	1.48	1.79	2.07	2.10	3.28
HUDVL	11.70	11.30	2.35	1.97	1.99	3.51	2.35	3.72	HUDVL	7.18	5.96	1.83	1.65	1.80	2.40	1.94	3.51
LONGIL	13.87	13.05	3.16	2.19	2.21	4.58	3.04	4.78	LONGIL	8.77	6.44	2.04	1.76	2.05	2.75	2.71	4.15
MHKVL	11.63	10.59	3.20	3.12	3.19	4.53	3.28	4.81	MHKVL	5.25	5.83	2.08	1.93	2.71	2.47	2.78	4.43
MILLWD	16.08	16.68	5.65	5.36	5.47	7.25	5.72	7.35	MILLWD	5.92	7.42	2.77	2.59	4.36	3.80	4.64	7.00
NORTH	4.40	4.33	2.75	2.43	4.85	2.51	5.31	5.19	NORTH	1.67	2.01	1.04	1.01	4.14	1.09	4.38	5.28
NYC	11.95	10.13	3.18	2.36	2.32	4.04	3.06	4.11	NYC	7.23	5.01	1.55	1.39	1.89	1.95	2.42	3.73
WEST	8.08	7.54	1.87	1.80	1.84	2.74	1.90	2.87	WEST	4.06	4.25	1.49	1.50	1.94	1.83	1.94	2.93
AEP	7.06	9.44	2.30	1.46	1.78	3.07	2.36	3.47	AEP	3.90	4.07	1.11	0.98	1.53	1.24	1.70	3.09
COMED	7.11	8.89	4.04	2.29	2.58	2.79	4.14	3.13	COMED	5.17	4.63	1.42	1.06	2.30	1.32	3.59	3.26
DAY	8.34	11.12	0.00	0.00	0.00	3.71	0.00	4.06	DAY	5.09	5.00	0.00	0.00	0.00	1.41	0.00	3.51
DOM	10.15	13.03	5.61	2.82	3.61	3.87	6.24	4.37	DOM	6.77	5.68	1.74	1.37	2.91	1.61	5.23	3.92
DUQ	8.34	11.12	0.00	0.00	0.00	4.55	0.00	4.76	DUQ	5.09	5.00	0.00	0.00	0.00	1.48	0.00	4.01
CEN	32.82	18.45	13.64	12.52	11.93	16.61	16.36	17.05	CEN	6.45	10.21	5.29	4.95	10.21	6.65	16.36	17.05
CRAI	46.72	36.12	21.59	22.14	22.75	22.51	22.56	22.92	CRAI	14.71	16.95	8.09	7.98	18.96	1.00	22.56	22.92
GEO	52.89	20.84	18.28	18.50	18.56	21.49	20.21	21.85	GEO	14.15	13.39	9.00	9.10	16.90	1.02	20.21	21.85
LIT	29.66	11.65	17.99	18.17	18.41	17.03	17.04	16.93	LIT	13.88	7.66	4.18	4.21	14.76	0.50	17.04	16.93

**Table 8.6:** Results for the forecasting problem. Left table contains the results for STLFL while right table contains VSTLFL. MAPE (%).

Dataset	AUTO	RW	NMET	GLF	NAUTO	LF	NBHD	MET	Dataset	AUTO	RW	NMET	GLF	NAUTO	LF	NBHD	MET
HIERRO	6.51	8.30	5.50	5.58	5.63	5.65	5.52	5.88	HIERRO	3.79	5.59	3.42	3.53	5.49	3.62	5.44	5.81
FUERTEVE	2.86	4.73	2.28	2.18	2.68	2.17	2.84	3.30	FUERTEVE	1.64	5.70	1.23	1.27	2.63	1.36	2.62	3.57
GCANARIA	3.36	4.13	1.94	1.92	2.07	1.95	2.06	2.41	GCANARIA	1.46	4.78	0.95	1.03	2.07	1.07	1.97	3.05
GOMERA	3.19	4.84	2.61	2.64	3.28	2.65	3.39	3.80	GOMERA	1.75	4.38	1.44	1.48	2.97	1.67	3.04	4.43
PALMA	4.59	6.32	4.06	4.16	5.21	3.96	5.17	5.84	PALMA	2.39	6.14	1.66	1.68	4.57	1.95	4.68	6.29
LANZAROT	2.43	3.70	1.80	1.73	1.79	1.79	1.88	2.28	LANZAROT	1.37	4.69	0.98	1.03	1.67	1.11	1.69	2.59
TENERIFE	2.75	3.93	2.07	2.28	2.56	2.29	2.45	2.91	TENERIFE	1.40	4.93	0.81	0.90	2.21	0.93	2.23	3.36
CIU	10.82	10.03	9.49	8.78	14.44	9.77	14.95	12.95	CIU	6.14	9.24	4.19	4.53	11.77	4.82	11.49	18.51
CTV	7.90	9.90	6.52	5.29	5.28	4.97	6.80	5.37	CTV	5.24	7.26	4.20	4.24	5.89	4.54	6.28	9.83
DA2	12.34	13.24	11.11	11.35	12.54	11.43	12.46	13.34	DA2	6.99	8.03	6.15	6.03	12.95	5.94	14.22	15.03
DAI	12.17	14.71	12.02	11.20	12.42	10.99	14.34	12.41	DAI	7.33	6.89	7.34	7.10	16.80	7.22	17.04	21.13
MAL	7.56	9.17	5.99	5.85	6.18	5.90	6.49	6.74	MAL	4.57	6.44	3.84	3.79	5.23	3.95	5.68	6.95
PDP	7.48	10.18	7.19	7.20	8.81	7.01	9.32	8.91	PDP	5.02	6.26	4.59	4.48	8.18	4.53	8.23	8.47
PDT	7.68	7.46	5.81	5.21	6.64	4.95	6.30	5.66	PDT	4.65	5.24	3.80	3.76	5.67	4.02	6.20	8.74
PEM	13.97	16.18	12.68	12.84	13.25	12.98	13.16	14.19	PEM	5.37	4.35	5.44	5.60	9.32	5.62	8.46	12.03
PLM	12.48	14.51	11.34	11.51	12.55	12.08	12.57	14.11	PLM	7.07	8.52	6.33	6.24	10.91	6.47	11.27	13.71
SOL	10.80	14.06	8.87	8.79	9.40	8.58	9.82	9.94	SOL	6.23	8.62	5.84	5.54	8.05	5.41	9.23	9.93
VPE	11.57	10.60	7.85	7.65	9.14	7.89	9.51	9.30	VPE	4.64	5.44	4.24	4.21	8.22	4.19	9.27	10.26
1	12.03	18.08	11.11	8.81	8.79	8.90	11.56	10.16	1	6.54	8.93	1.90	1.66	5.08	3.80	5.57	10.25
10	10.13	14.23	8.75	6.55	6.72	6.36	9.00	7.03	10	2.08	6.23	1.58	1.53	4.80	1.64	4.86	6.64
11	10.26	16.17	10.05	6.79	6.85	6.60	10.08	7.78	11	5.03	7.26	1.18	1.09	1.84	1.93	1.95	7.30
12	11.71	18.02	11.57	7.95	8.21	8.35	11.86	9.80	12	6.07	8.22	1.42	1.40	2.66	3.08	2.99	8.78
13	9.35	14.75	8.82	6.51	6.68	6.67	8.98	8.03	13	3.64	8.23	3.29	3.25	4.34	3.30	4.43	8.31
14	16.65	23.87	14.93	8.77	8.85	9.10	14.90	10.55	14	4.35	9.99	3.68	3.61	5.27	3.66	5.68	10.09
15	10.93	16.79	9.97	7.60	8.03	7.89	10.10	9.53	15	3.65	7.51	3.21	3.24	4.75	3.37	4.77	9.08
16	14.14	22.42	13.45	8.80	9.00	8.94	13.58	10.99	16	3.29	8.62	2.74	2.60	3.95	2.86	4.23	10.85
17	10.04	14.91	8.92	5.34	5.29	5.67	8.90	6.39	17	2.80	6.25	1.83	1.75	2.73	1.90	2.87	6.28
18	11.59	18.99	11.01	6.50	6.55	6.70	10.89	7.95	18	3.60	8.49	2.26	2.20	2.75	2.35	2.96	7.59
19	13.48	20.83	12.22	8.60	8.63	8.64	12.11	10.01	19	3.87	9.23	3.28	3.20	4.05	3.26	4.07	10.01
2	6.56	11.68	6.42	4.89	4.89	4.82	6.59	5.73	2	1.94	5.28	0.91	0.90	0.90	1.04	0.91	5.15
20	7.43	12.24	6.95	5.36	5.39	5.38	7.13	6.20	20	2.37	6.47	1.93	1.95	2.53	2.01	2.62	5.73
3	6.56	11.68	6.42	4.26	4.26	4.22	6.59	4.93	3	1.93	5.28	0.91	0.91	0.91	1.04	0.91	4.66
4	7.94	12.04	7.63	6.79	7.01	6.72	8.15	7.63	4	3.88	8.66	3.54	3.50	5.56	3.38	5.65	7.06
5	12.30	20.01	11.63	8.49	8.48	8.42	11.58	10.05	5	3.95	10.65	2.16	2.13	3.18	2.12	3.43	8.45
6	6.71	11.95	6.58	4.98	4.98	4.91	6.74	5.75	6	2.21	5.46	0.92	0.91	0.90	1.04	0.92	5.15
7	6.56	11.68	6.42	4.26	4.26	4.22	6.59	4.93	7	1.93	5.28	0.91	0.91	0.91	1.04	0.91	4.66
8	9.02	15.38	9.02	7.24	7.50	7.25	10.16	8.10	8	2.34	7.56	1.77	1.74	3.96	1.69	4.26	7.12
9	51.13	51.05	45.66	45.03	51.46	42.51	55.17	49.54	9	14.08	17.57	12.24	13.38	41.51	13.47	42.88	43.39
CAPITL	11.52	10.80	6.14	3.27	3.36	3.37	6.18	3.52	CAPITL	6.90	5.72	2.78	2.13	2.70	2.17	3.10	3.36
CENTRL	9.22	8.54	5.04	2.95	2.97	2.90	5.04	3.00	CENTRL	4.47	4.57	2.26	1.76	2.14	1.77	2.43	2.73
DUNWOD	12.84	12.22	7.17	3.99	4.41	4.26	7.38	4.78	DUNWOD	6.32	6.24	3.26	2.59	3.77	2.77	3.60	4.95
GENESE	10.35	9.34	5.43	3.07	2.99	2.95	5.37	3.04	GENESE	5.07	5.23	2.52	2.07	2.46	2.07	2.96	3.28
HUDVL	11.70	11.30	6.74	3.45	3.43	3.51	6.79	3.72	HUDVL	7.18	5.96	3.21	2.37	3.06	2.40	3.53	3.51
LONGIL	13.87	13.05	7.90	4.73	4.77	4.58	8.20	4.78	LONGIL	8.77	6.44	3.67	2.80	3.44	2.75	4.17	4.15
MHKVL	11.63	10.59	6.12	4.31	4.62	4.53	6.32	4.81	MHKVL	5.25	5.83	2.83	2.40	3.48	2.47	3.64	4.43
MILLWD	16.08	16.68	10.34	6.87	7.57	7.25	11.05	7.35	MILLWD	5.92	7.42	4.22	3.41	5.49	3.80	5.48	7.00
NORTH	4.40	4.33	2.93	2.00	4.69	2.51	5.53	5.19	NORTH	1.67	2.01	1.20	1.03	4.78	1.09	4.60	5.28
NYC	11.95	10.13	6.30	3.92	3.97	4.04	6.40	4.11	NYC	7.23	5.01	2.23	1.91	2.80	1.95	3.02	3.73
WEST	8.08	7.54	4.41	2.80	2.93	2.74	4.55	2.87	WEST	4.06	4.25	2.12	1.78	2.52	1.83	2.75	2.93
AEP	7.06	9.44	5.58	2.87	3.10	3.07	5.84	3.47	AEP	3.90	4.07	1.43	1.19	1.93	1.24	2.14	3.09
COMED	7.11	8.89	5.46	3.08	2.85	2.79	5.88	3.13	COMED	5.17	4.63	1.60	1.28	2.36	1.32	3.90	3.26
DAY	8.34	11.12	6.79	3.37	3.67	3.71	6.79	4.06	DAY	5.09	5.00	1.71	1.42	1.42	1.41	1.45	3.51
DOM	10.15	13.03	8.14	3.54	3.93	3.87	8.49	4.37	DOM	6.77	5.68	2.50	1.63	3.07	1.61	5.70	3.92
DUQ	8.34	11.12	6.79	4.30	4.54	4.55	6.79	4.76	DUQ	5.09	5.00	1.71	1.51	1.51	1.48	1.45	4.01
CEN	32.82	18.45	13.43	14.00	15.67	16.61	16.36	17.05	CEN	6.45	10.21	0.56	0.58	1.25	0.65	16.36	17.05
CRAI	46.72	36.12	21.77	23.13	23.32	22.51	22.56	22.92	CRAI	14.71	16.95	1.01	1.02	2.23	1.00	22.56	22.92
GEO	52.89	20.84	18.36	19.72	21.23	21.49	20.21	21.85	GEO	14.15	13.39	0.96	0.96	2.03	1.02	20.21	21.85
LIT	29.66	11.65	15.30	15.79	16.97	17.03	17.04	16.93	LIT	13.88	7.66	0.48	0.48	1.57	0.50	17.04	16.93



*But I don't think I have any particular talent for prediction, because when you have three or four elements in hand, you don't have to be a genius to reach certain conclusions.*

Antonio Tabucchi

CHAPTER

# 9

## Utility tools for DSM: Spatial Load Forecasting

**U**RBAN AREAS account for 67 % of global final energy consumption and 76 % of fossil fuel-related CO<sub>2</sub> emissions (Intergovernmental Panel on Climate Change, 2014). With the global urban population expected to increase by an additional 2.5 billion people between 2010 and 2050 and associated expansion of urban areas, the urban shares in total energy use and GHG emissions are also expected to increase. It is not, however, just the rate or scale of urbanisation that matters for urban energy use. An important, and often overlooked, factor in DSM strategies is the future spatial patterns of urban development. The high level of cities expansion, ageing power grids and even electrical market practices, present major challenges to the distribution utilities when planning the network expansion to offer a reliable and economical service inside the service area.

This chapter focuses on the coupling of Geographic Information System (GIS) and MAS provided by the Smart City Simulator developed in Chapter 5 to prove how

this integration can be beneficial to provide accurate Spatial Load Forecasting (SLF) to help electric utilities assess and determine the resiliency of the electrical grid to future demographic changes. This type of analysis are the first and foremost step to design and implement a DSM policy in an specific area.

## 9.1 Introduction

One of the biggest challenges that electrical distribution companies face is the growth in demand for electrical power. Under the current economic conditions, this problematic boosts exponentially: distribution companies aim at getting the most out of the existing infrastructures, especially when their renovation can be expensive. Thus, long term load forecasts are extremely useful for energy suppliers, DSO, financial institutions, and other participants in the electric energy generation, distribution, and retail sectors. With buildings consuming approximately 34 % of energy demand, responsible for 6 % of GHG emissions, and urbanisation increasing in most areas of the world, it is clear that innovative technologies and design in urban areas can be instrumental in achieving long-term sustainability of the global energy system.

The SLF methodology connects consumer consumption patterns and expected demographic growth over a certain location. The strength of this technique lies in the granularity of the geographic projections. SLF is extremely useful for DSO and TSO who are concerned about whether the current electrical grid will be able to serve load growth in terms of physical infrastructure and overall capacity. Current load forecasting techniques used by electric utilities consider the geographical area where the grid is located as a static entity, failing to appraise the dynamics associated with the development of new infrastructures. For example, the construction of a school may facilitate the growth of nearby residential load, while the construction of a highway may favour the growth of industrial load. In this context the inherent dynamic of MAS systems can provide additional benefits to SLF by allowing to consider the dynamic growth of urban centers and the spatial nature of cities.

## 9.2 Related Work

The concept of LTLF involves social, economic, policy and technical issues to which we must add the problems of having limited amounts of information and the difficulty to operate with the scarce data available (Willis, 1996). In this sense, LTLF is closely linked to urban evolution since its main models derive from the fields of geography and sociology. Nowadays, these models are being combined with others stemming from AI. This integration brings together the best qualities of both areas: AI provides the learning ability and the capacity to interpret the available data, while social and spatial models define the natural behaviour and characteristics of the problem at hand. The technical literature shows a wide range of methodologies and models for LTLF which can generally be classified into two broad categories: statistical methods and artificial-intelligence-based methods. Statistical load forecasting methods comprise approaches mainly based on time series and regression models (Alfares and Nazeeruddin, 2002b). In turn, artificial intelligence methods comprise methodologies like networks (Senjyu et al., 2002), genetic algorithms (Lai, 1998), support vector machines (Pai and Hong, 2005), and fuzzy logic (Farahat, 2004).

The growth in demand for electric power can be decomposed into two axes. On the one hand, there is the so-called *vertical growth*, which represents a boost in power demand due to an increment in the electrification of the existing households. On the other hand, the *horizontal growth* involves the appearance of new settlements on a specific area, motivated by the natural evolution of the population. The analysis of these phenomena is crucial since an improper estimation may lead to the saturation of the electrical infrastructure and the loss of power supply, along with the consequent economic losses and social distress.

In this sense, there is a special type of LTLF that deserves a closer look due to its economic importance: SLF. SLF uses a model built over a GIS to get together data related to electric distribution, land use, and development indicators. In this way, urban infrastructure engineers will be able to predict, years in advance, any new load and how it affects the electric system, helping them to determine whether the current infrastructure should be upgraded or not. Failing to do so leads to the

inability to cope with load peaks, appearance of brownouts, blackouts and, in general, low-quality supply. Furthermore, there is a lengthy amount of previous research on agent-based simulations using GIS, whether it is a geographic phenomena or a phenomena with an important geographic component, covering several aspects of Urban Modelling (Li and Muller, 2007; Crooks et al., 2008) and Housing (Jordan et al., 2010).

Agent-Based Modelling (ABM) is experiencing a notable boost in new fields lately due to its versatility and ability to model and simulate human behaviour in very diverse disciplines, as seen in (Axtell et al., 2002; Malleson et al., 2010; Crooks and Heppenstall, 2012). Paradoxically, though ABM is a well-known and intensively used tool in related areas SLF remains *terra incognita* for this paradigm. SLF is a crucial task for the majority of the stakeholders in the electric sector due to its capacity to calculate the future evolution of energy demand on a certain zone. So far, the only attempt to bring together ABM and SLF is, to our notice, (Borges et al., 2012). Still, ABM and GIS should be coupled tighter in order to improve the quality of forecasts.

### 9.3 Environment Modelling for Spatial Load Forecasting

The MAS system presented hereby is able to simulate the variation in the load of the transformers and electrical substations located on a certain city. To this end, we have modelled the behaviour, evaluations and decisions a human takes when choosing a new place to live. Figure 9.1 describes a simplified class diagram of the code implementation. Our configuration is based on the environmental approach proposed by *Russel and Norvig* (Russell et al., 2003; Weyns et al., 2004) being:

- **Accessible:** The agents have access to the whole environment.
- **Non-deterministic:** A change in the state of the environment depends on the management of threads by the Operating System on which the simulation is running.
- **Dynamic:** The environment can change while the agent deliberates.
- **Discrete:** The number of percepts is limited and centralised.

Our Environment is populated by both Passive Entities, used for representing the city infrastructure like substations, transformers and existing buildings, and Autonomous Agents in the form of greenfields and buyers. The pseudo-code for the model is described in Algorithm 3.

```

Data: Initial state of the environment
while !termination(state) do
  for greenfield in greenfields do
    PERCEPT1[greenfield] = Get-Around-Facilities(greenfield, state)
    PERCEPT2[greenfield] = Get-Around-Constructed-Buildings(greenfield,
      state)
    ACTION1[greenfield] =
      Determine-Dwelling-Characteristics[PERCEPT1,PERCEPT2(agent)]
    ACTION2[greenfield] = Set-Dwelling-Price[ACTION1(agent)]
  end
  for agent in agents do
    PERCEPT[agent] = Get-List-Greenfields(agent, state)
    ACTION1[agent] = Evaluate-Greenfields[PERCEPT(agent)]
    while !assigned && greenfields-available do
      ACTION2[agent] = Get-Greenfield[ACTION1(agent)]
    end
    ACTION3[CT] = Add-Load[ACTION2(agent)]
  end
  state = UPDATE-FN(actions, agent, state)
end

```

**Algorithm 3:** Buyer's model for the selection of greenfields.

### Substations and transformers

Substations and transformers are modelled as Passive Entities as their only function is to receive the power demands of the new households (see Section 9.3 below for more details).

Each transformer  $t$  and substation  $s$  has a nominal power, a demanded power ( $E_t$  and  $E_s$ ) and a simultaneity factor ( $S_t$  and  $S_s$ ). Since the electric grid and its elements need to size their nominal power in order to manage demand peaks, the

simultaneity factor is critical. It is calculated comparing the maximum power demand with the contracted power of the clients. In that way, it adjusts the theoretical total consumption of the clients to real conditions. Unfortunately, the dataset we have does not include power measurements for the transformers (for more details see Section 9.3.1). In fact, the lowest level where we had real measures was at substation outputs. Therefore, we had to calculate the simultaneity factor  $S_t$  at this level, inheriting the simultaneity factor to all the transformers connected to it. The formula used is:

$$S_t := \frac{r_s}{\sum_{t \in T_s} \sum_{c \in C_t} p_c},$$

where  $r_s$  denotes the maximum measure registered at the substations output  $s$ ,  $T_s$  denotes the set of transformers connected to that substation output,  $C_t$  denotes the set of clients connected to transformer  $t$  and finally  $p_c$  denotes the contracted power of client  $c$ .

### **Buildings**

Buildings are those residential dwellings situated in the environment. Every building has its physical representation, concerning location, area, building levels, and the electrical information regarding the amount of clients, the transformer they are connected to and how much power they demand. Buildings are modelled as Passive Entities, so that greenfields (Autonomous Agents) can ask them in order to forecast which type of construction they will harbour.

### **Greenfields**

Greenfields are modelled as agents that emulate the land plots where new buildings can be constructed. Local authorities define a location as *accessible to citizens* if it is within 5 minutes walking. That is, considering that the speed of a pedestrian is 4 km/h, this distance corresponds to 300 meters. Therefore, every greenfield has the skill to perceive its surroundings up to 300 meters. As a first approximation, these agents will search and calculate the straight-line distance  $d_i$  to several public

facilities (e.g. green zones, public transports, parking spaces, and the like). Table 9.1 shows a comprehensive list of these public facilities. One of the problems faced when working with developable land use, is that, though not yet urbanised, some of these areas have already been split into smaller parcels while others comprise a whole rural area. Although clipped greenfields give some clues about the type of buildings they may contain, without such information, it is still hard to determine the type of construction and how many citizens will host each greenfield. So, as next step, in order to overcome this problem the greenfield estimates, using a spatial moving window smoothing (Kiesel and Wenkel, 2007), the type of building, number of dwellings and building levels expected according to all adjacent buildings within 300 meters.

Finally, based on the loaded information, each plot establishes its price per square meter and searches for its nearest transformer substation, which the new settlements will connect to. The process is as follows: first a Voronoi diagram is calculated from the set of transformers that serve more than one client (single-client transformers are owned individually, which means that they are not accessible by the DSO). Then, the plot makes the connection with the transformer whose area of influence presents the largest intersection with its own surface.

**Table 9.1:** Infrastructures considered per factor.

Factor	Infrastructures considered
Health	Hospitals, clinics
Education	Schools, colleges, kindergardens, universities
Sports	Public swimming pools, pitch, stadiums
Cultural	Art centres, theatres, community centres, conference centres, museums, libraries, cinemas
Food Shops	Food and convenience shops, department stores, supermarkets

### Buyers

The buyers are agents that emulate the people looking for a new house. Since every person has different preferences about the presence of (or distance to) a particular

public facility, we have encoded them in a vector  $a_i$  that describes how important each infrastructure is to a particular agent  $i$ . Moreover, agents have an individual budget limit and a degree of greediness depending on which, they will query a different number of greenfield. Further, we have identified three primary target groups sharing a common preference pattern: Elderly, Families and Singles. The accurate values of the preference vector have been issued using a uniform random variable with the mean described in Table 9.2 and a 10 % of standard deviation.

**Table 9.2:** Agents types and their preferences. Average values.

Type	HEALTH	EDUCATION	SPORTS	CULTURAL	FOOD	AFFORD
ELDERLIES	1	0.2	0	0.7	0.8	1800 €
FAMILIES	0.9	1	0.7	0.3	0.5	1750 €
SINGLES	0.2	0.2	1	1	0.8	1700 €

When the agents have been loaded into the Environment, they select a number of greenfields that will be asked for information. The agent will then select the plot that maximises the following function  $f$ :

$$f(a, d) := \begin{cases} -1 & \text{if plot price} > \text{agent budget} \\ \sum_{j \in J} \frac{a_{ij}}{d_j} & \text{in other case,} \end{cases}$$

where  $a_i$  is the preference vector of the agent  $i$ ,  $d_j$  is the distance from each building to the infrastructure  $j$  (see Table 9.2) and  $J$  are the categories in Table 9.1.

Next, the agent will try to buy this plot. In case some other agent has already bought it, the current agent will try to acquire its next preferred option until the plots reach the minimum desired quality set. Please note that it may be possible for an agent not to get a greenfield. Finally, the electrical load generated by each agent is added to the corresponding electrical infrastructure following the function  $l$ :

$$l(a, d) := E_t + I_a \cdot S_t \cdot A \cdot P_c,$$

where  $E$  is the previous load in that particular electrical infrastructure,  $I_a$  is the electrical intensity of agent  $a$  (i.e. how much power will the new settlement need),  $S_t$  is the simultaneity factor (see Section 9.3) of the loads in that particular infrastructure,  $A$  is the area covered by this plot, and  $P_c$  is the power intensity of the area, measured as:

$$P_c := \frac{|B_{300}| \sum_{c \in C_{300}} p_c}{|C_{300}| \sum_{b \in B_{300}} s_b},$$

where  $B_{300}$  is the set of buildings within 300 meter radius,  $C_{300}$  is the set of clients within a 300 meters radius,  $|\cdot|$  denotes the set cardinality,  $p_c$  is the contracted power by client  $c$ , and  $s_b$  is the total surface of the building (measured as the constructed area times the floor count).

By combining these data, when an agent is assigned to an available greenfield, the system analyses the consumption of the neighbouring parcels and size of the buildings in order to predict how much power this new settlement will need. This amount is then added to the total load of the transformer from which the plot feeds.

### 9.3.1 Data

The experiments have been carried out with real data from Ciudad Real, a Spanish middle-sized city with about 32 thousand power consuming clients and a surface of 400 km<sup>2</sup>, including the surrounding municipalities of Miguelturra, Carrión de Calatrava and Poblete. A map of the zone can be seen in Figure 9.2. Electric infrastructure and clients' measurements were provided by the corresponding utility (Gas Natural Fenosa, a Spanish DSO) while buildings and landuse data was obtained by conflating the Spanish cadastre records with the VGI source OpenStreetMap (Pijoan and Borges, 2012; Borges et al., 2013c). Both datasets are stored in a PostgreSQL relational database bolstered with the PostGIS geographical extension for manipulating spatial data. Although the Spanish cadastre registers the price for each plot, said information is private and is not provided in their open-data initiative. The most detailed prices found were extracted from recent appraisals performed by the

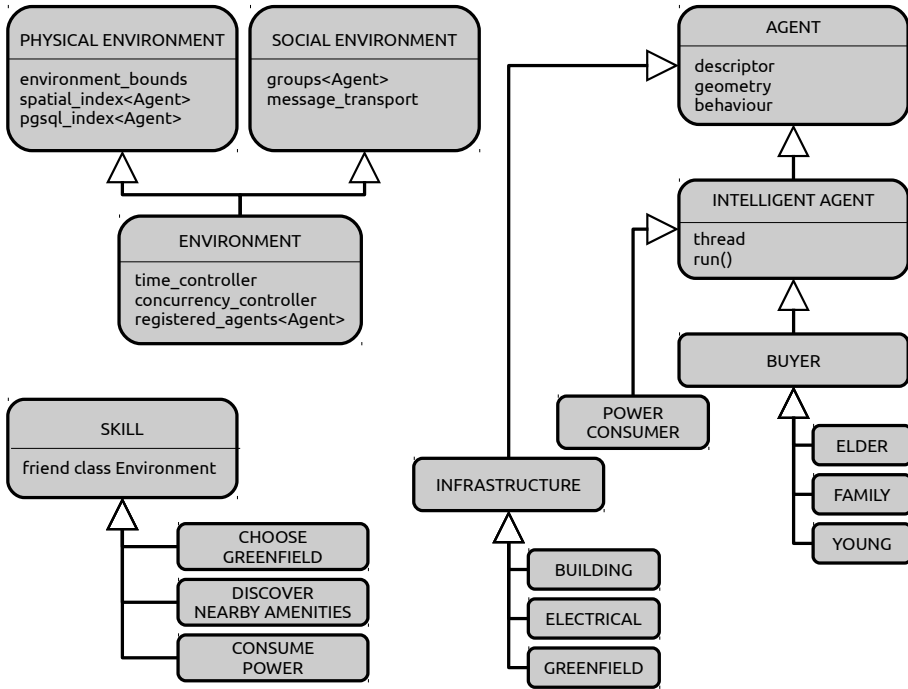
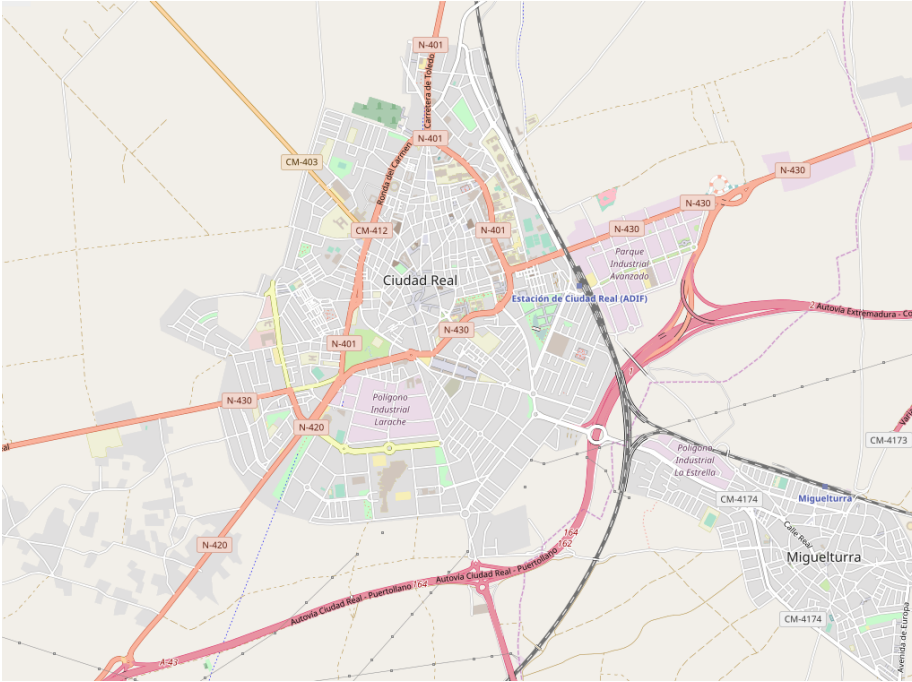


Figure 9.1: Simplified class diagram of the environment.

appraiser Tasamadrid (Tinsa, 2013), which provides average prices per square meter grouped by postcode.

Despite the spatial data being only a snapshot of the city in 2013, all electric power clients are geolocated and have a registration date. So that, we can assume that the date of the building’s creation is the same as the first client assigned to that particular building. The information covers the years from 2005 to 2010. Therefore, we get a complete historical map of the evolution of the city and can determine which areas were inhabited between those years. Nevertheless, we cannot ensure the moment in which the rural areas were zoned into greenfields as the status of the plots change by local laws and is not recorded in public databases.



**Figure 9.2:** Overview map of the SLF test area: Ciudad Real.

## 9.4 Results and discussion

For each independent year which electrical growth we want to forecast, the validation process creates a historical map that describes the status of the buildings of the city. In addition, it creates several distributions of agents in order to contrast the process outcome with the real settlements:

- **Environment.** The validation process identifies the buildings that were registered as such from a given year onwards and marks them as available green-fields, considering the date in which the building was created is the same as the oldest settlement registered.

- **Agent System.** All the scenarios will create the same amount of agents as new possible settlements that may appeared on that year. The main difference will be the amount of agents of each type created at each case.
- **Greediness.** We have validated the system considering agents with and without greediness. Obviously, the experiments that do not consider agent's greediness obtain better results as the agents are aware of the whole environment. We only provide results experiments where the greediness is considered as it is a more realistic scenario.
- **Plots' price.** The unitary price of the plots is estimated using recent appraisals performed by the appraiser Tasamadrid.

We have defined different metrics in order to test the results obtained with this model. The error can be split into two categories: spatial errors and effective errors.

- **Spatial Errors.** Errors related to the spatial component of the forecast. Namely, we measure how many agents correctly select the year  $y$  in which a greenfield has been built ( $hits_y$ ), how many agents incorrectly select the year  $y$  in which a greenfield has been built ( $semi_y$ ) and how many agents have completely failed by selecting a greenfield that even today has not been built ( $fails_y$ ). Thus, we have measured:

$$hits_y := \frac{a_y}{b_y}, \quad semi_y := \frac{f_y}{\sum_{j=y+1}^{2010} b_j}, \quad fails_y := 1 - hits,$$

where  $a_y$  denotes the number of agents that correctly select a plot that is built in the year  $y$ ,  $f_y$  denotes the number of agents that incorrectly set on a plot in the year  $y$  which in reality was built in the following years,  $b_y$  denotes the number of plots that have indeed been built in the year  $y$ . The variables  $hits$ ,  $semi$  and  $fails$  are just mean values over of  $hits_y$ ,  $semi_y$  and  $fail_y$  respectively.

- **Effective Errors.** In this category we measure the load forecasting error. Traditionally, this error is calculated using the MAPE error (Hyndman and Koehler, 2006):

$$mape := \frac{1}{6|S|} \sum_{s \in S} \sum_{y=2005}^{2010} \frac{|r_y^s - p_y^s|}{r_y^s},$$

where  $S$  denotes the set of transformers,  $|\cdot|$  denotes the set cardinality operator,  $p_y^s$  denotes the model's forecast for the maximum load of substation  $s$  at year  $y$  and  $r_y^s$  denotes the real one.

Please note that, in this model, MAPE is not a reliable source of error as there is a large variability among the experiments. This situation is due to technical or electrical infrastructure exploitation issues that force a building to not be always connected to its nearest transformer. In order to mitigate this problem, a possibility solution is to measure the error one level higher in the electrical infrastructure, that is, at substation level. However, this measurement would blur the results and hinder the possibility to identify zones where the model is not working correctly. Moreover, in terms of performance, the calculation of this value is quite complex due to the number of database queries that would need to be performed which, in turn, increases the execution time of each simulation.

We have tested the forecasting ability of the model in two situations: for one year ahead and five years ahead forecasts. Moreover, as a contrast method we used random assignation of agents to greenfields. In order to calculate an approximated p-value we have used a Monte Carlo approach. Namely, we have repeated 100 times the random assignation for every year to approximate the density function of the random variable  $hits + semi$  under the null hypothesis that our method and the random Monte Carlo approach would derive the same results. From the approximate density function we calculate the associated  $p - value$  (see  $pval$  in Table 9.3)

In the first case, each year is evaluated independently, and the outcome is not considered for the following evaluation. Figure 9.3 shows the empirical probability density function constructed with 100 repetitions of the contrast method. Table 9.3 shows the numerical results for 1 year ahead forecast (mean value of the 5 years). The first three columns represent the percentage of agent of every type that we have created in that particular experiment. Please note that the 0/0/0 row represent the Monte Carlo approach.

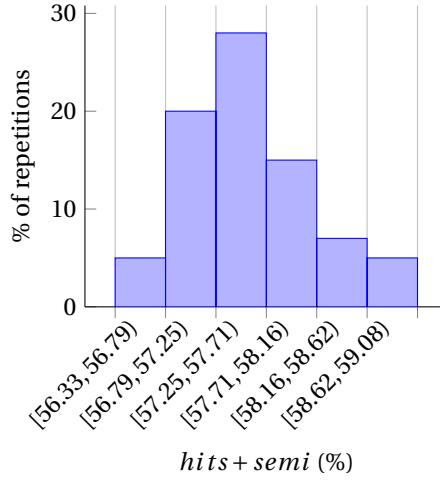
On the other hand, for the 5 year forecasting, we use a *rolling forecast* approach (Hyn-dman and Athanasopoulos, 2014). Namely, we will use the foretasted results of the previous years to predict the following year. As for the 1 year ahead case, Figure 9.3(b)

shows the empirical probability density function constructed with 100 repetition of the contrast method and Table 9.3 shows the numerical results of the five years ahead forecast.

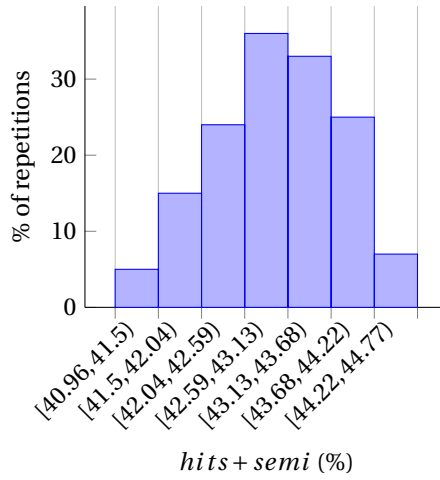
**Table 9.3:** Experimental results for the different agent type composition. All measures are in %.

Elderly	Families	Young	1 year forecast				5 years forecast			
			<i>hits</i>	<i>semi</i>	<i>fails</i>	<i>pval</i>	<i>hits</i>	<i>semi</i>	<i>fails</i>	<i>pval</i>
0	0	0	0.14	0.44	0.42	0.41	0.18	0.25	0.57	0.45
100	0	0	0.17	0.49	0.34	0.00	0.34	0.31	0.35	0.00
0	100	0	0.22	0.50	0.28	0.00	0.34	0.35	0.32	0.00
0	0	100	0.19	0.55	0.26	0.00	0.29	0.37	0.34	0.00
50	25	25	0.17	0.51	0.31	0.00	0.30	0.35	0.35	0.00
25	50	25	0.21	0.49	0.30	0.00	0.33	0.34	0.34	0.00
25	25	50	0.17	0.55	0.28	0.00	0.27	0.37	0.36	0.00
80	10	10	0.18	0.49	0.33	0.00	0.34	0.31	0.35	0.00
10	80	10	0.22	0.49	0.29	0.00	0.34	0.35	0.32	0.00
10	10	80	0.19	0.55	0.26	0.00	0.29	0.36	0.34	0.00
60	20	20	0.17	0.50	0.33	0.00	0.34	0.32	0.35	0.00
20	60	20	0.21	0.49	0.30	0.00	0.33	0.34	0.33	0.00
20	20	60	0.18	0.54	0.28	0.00	0.28	0.37	0.35	0.00

Figure 9.3 shows, as expected, a large difference between the errors of the experiment in the 1 year ahead forecasting and the 5 years ahead forecasting. Not only do the results of the random assignment are better in the 1 year ahead forecasting, but they also have considerably less variance. Table 9.3 confirms this expectation. In the case of the 1 year ahead forecasting the mean value of *hits* + *semi* is around 70 %, while in the case of 5 years forecasting it is around 66 %. Please note that this value is quite high considering that the random assignment has achieved 58 and 43 % (respectively) in those experiments. The results of using the proposed MAS model are in all cases significantly better than those corresponding to the random Monte



(a) 1 year ahead forecast



(b) 5 years ahead forecast

**Figure 9.3:** Probability density function of the random variable *hits + semi*.

Carlo approach. Take into account that the first row corresponds to the Monte Carlo approach, so in this particular case, we are comparing the same results.

The distribution of *hits* and *semi* in both experiments is also interesting. While in the 1 year ahead forecasting experiment *semi* is quite high (around 50 %) in the 5 years ahead forecasting experiment is near 30 %. The explanation of these results is straightforward: in the 1 year ahead forecasting experiment, the number of greenfields that (if occupied) score a *semi* decrease with every year, while in the case of 5 years ahead forecasting they are constant. Please note that it is impossible to score a *semi* in the last year of the 1 year ahead forecasting experiment since we do not have information of the next year and all the houses of previous years have already been occupied.

The results show how different agent type distributions affect the accuracy of the prediction delivered by our model. As the typical house buyers in Spain are families, it is expected that models with a high percentage of families agents will perform better. The results from the 5 years ahead forecasting confirm that hypothesis. The best agent mix consists of considering only agents of Family type while in the mixed cases, those with high percentage of Family agents perform at least as well as the other. On the contrary, it seems that in the short term (1 year ahead forecasting), the Young are the main variable. This can be explained as Young frequently change their homes, so in the short term, buildings near their favourite places are occupied in a short time.

Qualitatively, we can see that the solutions follow the logic imposed by the vector of preferences in the settlement of the agents. Figures 9.4 to 9.6 show the distribution of Elderly, Families and Young agents through three different heatmaps. The first one shows how the final distribution of the Elderly and the localisation of Clinics and Hospitals. In the second one, we present the final distribution of Families and the localisation of Schools. Finally, in the third one, we present the final distribution of Young agents and the localisation of Sport Centres. In the three cases the final distribution loosely approximate to the localisation of the corresponding infrastructures.



**Figure 9.4:** Heatmap of the new settlement for Elderly people and Hospitals.



**Figure 9.5:** Heatmap of the new settlement for Families and Schools.



*The journey of a thousand miles begins with one step.*

Lao Tzu

# CHAPTER 10

## Conclusions

**D**SM is an important tool for enabling an efficient use of the energy resources available. DSM can mitigate electrical system emergencies, minimise blackouts and increase system reliability, reduce dependency on expensive imports (in some countries), reduce energy prices, provide relief to the power grid and generation plants, defer investments in generation, transmission and distribution networks and contribute to lower environmental emissions.

New smart power grids demand a huge effort in redesigning and enhancing current power networks, as well as integrating distributed generation, renewable energies and the electric vehicle. Contrary to the previous centralised and unidirectional model (i.e., from centralised generation via the transmission and distribution grids to the customers), smart grids will change to a bidirectional power flow model that will include distributed generation (mainly from renewable sources) and the electric vehicle.

The power network operation needs to be safe, reliable and, as long as possible, cost effective. The new scenario enabled by the massive adoption of distributed generation and storage and the intensive application of information technologies render the old centralised and static architecture not feasible. The main reason is that gathering the complete information to achieve optimal management, if possible at all, would entail a prohibitive cost in infrastructure and time. Therefore, a trade-off must be found between distributing the intelligence and the costs associated to this decision.

During this final chapter, we will revise the main conclusions extracted from our research, summarise its limitations and draw out the lines for future developments.

## 10.1 Discussion

Through the performed research, we have made some contributions in the DSM field, relying on previous works and proposing some concepts in order to automatise the processes involved. Below we list the tasks and challenges we have addressed, together with the designed approaches:

**Chapter 1** describes the context and motivation of the research presented in this thesis. The chapter focuses primarily on the importance of the application of DSM strategies to influence the electricity demand of the end users, both for avoiding demand peaks that could lead to the destabilisation of the electrical grid, and for maximising the integration of renewable energy sources.

Chapter 3 introduces an agent-based Smart City Simulator for the evaluation DSM strategies. The chapter focuses on the technical aspects of the implementation, from the characterisation and operation of appliances to the definition of an appropriate human behaviour model.

**Chapter 4** presents the agent-based human-behaviour engine behind the implementation of the Smart City Simulator. This chapter introduces the characterisation of human actions in line with the Belief-Desire-Intention paradigm, providing a sound validation methodology to evaluate the correspondence between synthetic and real human activity datasets.

**Chapter 5** focuses on the problem of domestic DSM from the point of view of analysing which is the best architecture solution to gather all the information relevant to the design and implementation of DSM strategies while maintaining appropriate response times in the real time monitoring.

**Chapter 6** demonstrates the feasibility of using AI for aiding commercial DSM strategies. Specifically, this chapter focuses on the design and development of a methodology that adjusts the terms of the electrical tariff of a university campus, composed of 5 buildings, based on historical consumption data. The analysis estimates different alternatives of pricing, depending on the use of individual or joint consumptions.

**Chapter 7** demonstrates the feasibility of using AI for aiding industrial DSM strategies. Specifically, this chapter focuses on the design and development of a methodology that best schedules the execution of orders in a foundry plant in order to minimise the cost of the casting process while meeting production deadlines.

**Chapter 8** focuses on a methodology to improve Short Term Load Forecasting at substation level by gathering data from nearby substations to improve data imputation methods and predict future trends in electricity consumption.

**Chapter 9** focuses on the use of the agent-based systems to improve Spatial Load Forecasting. Specifically, this chapter presents a methodology that can help electric utilities in assessing whether the electrical infrastructure deployed at a certain location is able to support the demographic growth.

The major conclusions from this thesis can be summarised as follows:

The major problem of model-based simulators is their dependence on activity models. This implies that the quality and accuracy of the resulting datasets, heavily relies on the quality of the activity description model and the associated parameters. However, obtaining accurate activity models is complicated, since human behaviour is very complex and full of details which are hard to capture.

The grid computing paradigm presents inspiring characteristics for modelling agent environments that will help the smart grid vision come true. Indeed, grid computing applications are based on a group of distributed nodes carrying out a task coordinated by a central entity: power networks do have many nodes (e.g.,

meters and substations) and tasks that require participation and coordination (e.g., invoicing: demands retrieving the consumption data from all clients, applying the corresponding prices, and notifying the clients) among all of the nodes. MAS and agent environments can tackle this challenge by distributing the intelligence all over the grid by means of individual intelligent agents controlling a number of assets.

The management and configuration of power grids is closely linked with urban evolution. Social dynamics dictate the appearance of new settlements as well as the amount of energy demanded depending on the economic availability, the age and preferences of new costumers and the own evolution of the city. The edification of new infrastructures such as hospitals, schools, parks, malls; even the construction of new roads heavily influence the citizens' decisions and mobility.

Proper modelling of the dynamics of human behaviour will help in achieving more accurate simulations. Though simulation of the exact process of human thought is a complex task, the introduction of AI techniques in our system, such as Fuzzy Logic, will allow to mimic human decision making by building a wide range of behavioural rules in response to the stimulus and events of the environment, instead of the fixed decision approach used in our experiments. This modelling would then provide a parameterised sandbox to analyse and classify what kind of infrastructural and social changes in the city influence the most the variation of the electricity demand.

## 10.2 Future work and open issues

The next step towards the fulfilment of GeoWorldSim is in test the different middlewares currently available (Zdravković et al., 2016). In this sense, we will develop several of the services needed to perform demand-response in different middlewares and perform a qualitative and quantitative analysis to assess its suitability.

Then, an adequate simulation of the Smart Grid goes hand in hand with the precise definition of how the electrical system works and the implementation of several new agent types: producers of energy, storage and control nodes. Up to this point, we have used simplistic load models, since the objective of the study is the resiliency of the architecture over a great number of devices publishing power

consumption. The future steps in this area involve introducing a realistic model of the different grids to be able to simulate different smart controllers. This smart energy controllers will listen to what the energy needs of devices within a house, leveraging demand with energy generation coming from photovoltaic panels, batteries or even the distribution network. Moreover, simple models of both the new agents and the electrical distribution system could be implemented directly in GeoWorldSim but more complex ones would require being coupled via specialised tools like Matlab.

Finally, another line of future work has to do with the development a full human behaviour model that dictates the execution of activities of every agent in the simulation platform. Now, it heavily relies on a previous definition of a list of possible activities in order to schedule the temporal distribution. The next iteration in this area will test the use of Hidden Markov Models, Probabilistic Finite State Machines, Neural Networks or Support Vector Machines not only to model an adequate temporal distribution of activities but also to be able to predict the behaviour change when a persuasive methodology is used (Apostolou et al., 2016). Moreover, given that the communication systems is actually working in a hardware-in-the-loop configuration, this simulator could be used by a real demand-response controller to evaluate the impact produced by and specific control decision.



# Bibliography

- Adelantado, F., Vilajosana, X., Tuset-Peiró, P., Martínez, B., and Melià, J. (2016). Understanding the limits of lorawan. *CoRR*, abs/1607.08011.
- Agarwal, V., Bougaev, A., and Tsoukalas, L. (2006). Kernel regression based short-term load forecasting. In Kollias, S., Stafylopatis, A., Duch, W., and Oja, E., editors, *Artificial Neural Networks - ICANN 2006*, volume 4132 of *Lecture Notes in Computer Science*, pages 701–708. Springer Berlin Heidelberg.
- Al-Bahadili, H. (2012). *Simulation in Computer Network Design and Modeling: Use and Analysis*. IGI Global.
- Alam, M. R., Reaz, M. B. I., and Ali, M. A. M. (2012). A review of smart homes-past, present, and future. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 42(6):1190–1203.
- Alfares, H. and Nazeeruddin, M. (2002a). Electric load forecasting: literature survey and classification of methods. *International Journal of Systems Science*, 33(1):23–34.
- Alfares, H. K. and Nazeeruddin, M. (2002b). Electric load forecasting: literature survey and classification of methods. *International Journal of Systems Science*, 33(1):23–34.

- Alonso, C., Montaña, J., Puente, J., and Borges, C. (2009). A New Linear Genetic Programming Approach Based on Straight Line Programs: Some Theoretical and Experimental Aspects. *International Journal on Artificial Intelligence Tools*, 18(5):757–781.
- Andoh, H., Watanabe, K., Nakamura, T., and Takasu, I. (2004). Network health monitoring system in the sleep. In *SICE annual conference*, pages 1421–1424.
- ANSI (2010). *ANSI/CEA-852.1: 2010 - Enhanced Protocol for Tunneling Component Network Protocols Over Internet Protocol Channels*.
- Apostolou, G., Krinidis, S., Ioannidis, D., Tzovaras, D., Borges, C. E., Casado-Mansilla, D., and de Ipiña, D. L. (2016). Greensoul; a novel platform for the reduction of energy consumption in communal and shared spaces. In *2016 4th International Symposium on Environmental Friendly Energies and Applications (EFEA)*, pages 1–6.
- Axtell, R. L., Epstein, J. M., Dean, J. S., Gumerman, G. J., Swedlund, A. C., Harburger, J., Chakravarty, S., Hammond, R., Parker, J., and Parker, M. (2002). Population growth and collapse in a multiagent model of the kayenta anasazi in long house valley. *Proceedings of the National Academy of Sciences*, 99(suppl 3):7275–7279.
- Azkune, G., Almeida, A., López-de Ipiña, D., and Chen, L. (2015). Combining Users' Activity Survey and Simulators to Evaluate Human Activity Recognition Systems. *Sensors*, 15(4):8192–8213.
- Barbieri, D., Braga, D., Ceri, S., and Grossniklaus, M. (2010). An execution environment for c-sparql queries. In *In: Proc. Intl. Conf. on Extending Database Technology, EDBT*, pages 441–452.
- Becker, D., Saxton, T., and Goodrich, M. (2010). Cim standard for dynamic model exchange. In *IEEE Power and Energy Society General Meeting*, pages 1–3.
- Benjamini, Y. (1988). Opening the box of a boxplot. *The American Statistician*, 42(4):257–262.

- Blickle, T. and Thiele, L. (1995). A comparison of selection schemes used in genetic algorithms. Technical report, Gloriastrasse 35, CH-8092 Zurich: Swiss Federal Institute of Technology (ETH) Zurich, Computer Engineering and Communications Networks Lab (TIK).
- Boletín Oficial del Estado (2014). Orden IET/107/2014, de 31 de enero, por la que se revisan los peajes de acceso de energía eléctrica para 2014. *Boletín Oficial del Estado*, (28):7147–7169.
- Booch, G., Rumbaugh, J., and Jacobsen, I. (2005). *The Unified Modeling Language User Guide, 2nd ed.* Addison-Wesley.
- Borges, C., Peña, A., and Peña, Y. K. (2013a). On the influence of surrounding load demand to improve primary substation stlf. In *IECON 2013 - 39th Annual Conference on IEEE Industrial Electronics Society*. IEEE.
- Borges, C., Peña, Y., Fernandez, I., Prieto, J., and Bretos, O. (2013b). Assessing tolerance-based robust short-term load forecasting in buildings. *Energies*, 6(4):2110–2129.
- Borges, C., Peña, Y., and Pijoan, A. (5<sup>th</sup> June 2012). Agent based spatial load forecasting. In *Proceedings of 3<sup>rd</sup> International Workshop on Agent Technologies for Energy Systems (ATES 2012) held at the 11<sup>th</sup> International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, pages 107–108, Valencia (Spain). ACM press.
- Borges, C., Pijoan, A., Sorrosal, G., Oribe-García, I., González, M., and Esteban, O. K. (2013c). Uso de fuentes de información geográfica voluntarias en proyectos de ingeniería. In *Actas de las VII Jornadas de SIG Libre, Girona 6–8 marzo, España*. Servicio de SIG y Teledetección - SIGTE - de la Universitat de Girona.
- Borges, C. E., Kamara-Esteban, O., Pijoan, A., and Peña, Y. K. (2014). Multi-Agent GIS System for Improved Spatial Load Forecasting (Demonstration). In Lomuscio, A., Scerri, P., Bazzan, A., and Huhns, M., editors, *Proceedings of the 13th International*

- Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014)*, Paris (France). ACM.
- Borlase, S. (2012). *Smart Grids: Infrastructure, Technology, and Solutions*. CRC press.
- Bouchard, K., Ajroud, A., Bouchard, B., and Bouzouane, A. (2010). *Advances in Computer Science and Information Technology*, volume 6059 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Breedillet, P., Lambert, E., and Schultz, E. (2010). Cim, 61850, cosem standards used in a model driven integration approach to build the smart grid service oriented architecture. In *Proceedings of the IEEE 1<sup>st</sup> International Conference on Smart Grid Communications (SmartGridComm)*, pages 467–471.
- Britton, J. and deVos, A. (2005). Cim-based standards and cim evolution. In *IEEE Transactions on Power Systems*, pages 758–764. Digital Object Identifier 10.1109/PESS.1999.787426.
- Bruneau, J., Jouve, W., and Consel, C. (2009). DiaSim: A parameterized simulator for pervasive computing applications. In *Mobile and Ubiquitous Systems: Networking & Services, MobiQuitous, 2009. MobiQuitous' 09. 6th Annual International*, pages 1–10. IEEE.
- Buchmayr, M., Kurschl, W., and Küng, J. (2011). A Simulator for Generating and Visualizing Sensor Data for Ambient Intelligence Environments. *Procedia Computer Science*, 5:90–97.
- Cardoen, B., Demeulemeester, E., and Beliën, J. (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, 201(3):921–932.
- Casado-Mansilla, D., López-de Armentia, J., and López-de Ipina, D. (2012). Will eco-aware objects help to save the world? In *Ubiquitous Computing and Ambient Intelligence*, pages 17–24. Springer.

- Chambers, J. (1983). *Graphical methods for data analysis*. Chapman & Hall statistics series. Wadsworth International Group.
- Chang, L.-Y. and Chen, H.-C. (2014). Linearization and input-output decoupling for nonlinear control of proton exchange membrane fuel cells. *Energies*, 7(2):591–606.
- Chaouch, M. (2014). Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves. *Smart Grid, IEEE Transactions on*, 5(1):411–419.
- Chen, L., Hoey, J., Nugent, C., Cook, D., and Yu, Z. (2012). Sensor-based activity recognition. *IEEE Transactions on Systems, Man, and Cybernetics-Part C*, 42(6):790–808.
- Chen, Y., Li, H., Jin, K., and Song, Q. (2013). Wind farm layout optimization using genetic algorithm with different hub height wind turbines. *Energy Conversion and Management*, 70:56–65.
- Commission, I. E. (1996). *IEC60870: Telecontrol equipment and systems - Part 5: Transmission protocols - Section 102: Companion standard for the transmission of integrated totals in electric power systems*. IEC.
- Consortium, F-W. (2012–2016). FI-WARE architecture. <http://cordis.europa.eu/fp7/ict/netinnovation/deliverables/fi-ware/fi-ware-d231b.pdf>.
- Cook, D. J. and Schmitter-Edgecombe, M. (2009). Assessing the quality of activities in a smart environment. *Methods of information in medicine*, 48(5):480–5.
- Cook, D. J., Youngblood, M., Heierman, E. O., Gopalratnam, K., Rao, S., Litvin, A., and Khawaja, F. (2003). Mavhome: An agent-based smart home. In *Pervasive Computing and Communications, 2003.(PerCom 2003). Proceedings of the First IEEE International Conference on*, pages 521–524. IEEE.

- Crapo, A., Griffith, K., Khandelwal, A., Lizzi, J., Moitra, A., and Wang, X. (2009a). Overcoming challenges using the CIM as a semantic model for energy applications. In *Proceedings of the Grid-Interop Forum 2010*.
- Crapo, A., Wang, X., Lizzi, J., and Larson, R. (2009b). The semantically enabled smart grid. In *Proceedings of the Grid-Interop Forum 2009*, pages 177–185.
- Crooks, A., Castle, C., and Batty, M. (2008). Key challenges in agent-based modelling for geo-spatial simulation. *Computers, Environment and Urban Systems*, 32:417–30.
- Crooks, A. and Heppenstall, A. (2012). Introduction to agent-based modelling. In Heppenstall, A. J., Crooks, A. T., See, L. M., and Batty, M., editors, *Agent-Based Models of Geographical Systems*, pages 85–105. Springer Netherlands.
- Cvetković, B., Janko, V., Romero, A. E., Kafalı, Ö., Stathis, K., and Luštrek, M. (2016). Activity recognition for diabetic patients using a smartphone. *Journal of Medical Systems*, 40(12):256.
- Das, S. K., Cook, D. J., Battacharya, A., Heierman, E. O., and Lin, T.-Y. (2002). The role of prediction algorithms in the MavHome smart home architecture. *Wireless Communications, IEEE*, 9(6):77–84.
- Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge. ISBN 0-521-57391-2.
- De Silva, D., Yu, X., Alahakoon, D., and Holmes, G. (2011). Incremental pattern characterization learning and forecasting for electricity consumption using smart meters. In *2011 IEEE International Symposium on Industrial Electronics (ISIE)*, pages 807–812.
- Deb, K., Reddy, A. R., and Singh, G. (2003). Optimal scheduling of casting sequence using genetic algorithms. In *Materials and Manufacturing Processes*, pages 409–432.

- Derrac, J., García, S., Molina, D., and Herrera, F. (2011). A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and Evolutionary Computation*, 1(1):3–18.
- Education, B. G. (2015–2016). I2Basque Network. <http://www.i2basque.es/>.
- Enerdata (2016). Total primary production-global energy statistical yearbook 2016. <https://yearbook.enerdata.net/CO2-emissions-data-from-fuel-combustion.html>.
- Enitan, A. M. and Adeyemo, J. (2013). Food processing optimization using evolutionary algorithms. *African Journal of Biotechnology*, 10(72):16120–16127.
- EPRI (2006). *Harmonization of CIM with IEC Standards: Draft Report for CIM and other IEC Working Groups*, 1.0 edition.
- EPRI (2008a). *CIM-XML Interoperability Including CIM-Based Tools Test: The Power of the Common Information Model (CIM) to Exchange Power System Data*, 1.0 edition.
- EPRI (2008b). *The Common Information Model for Distribution: An Introduction to the CIM for Integrating Distribution Applications and Systems*, 1.0 edition.
- EPRI (2009a). *Distribution CIM Modeling of Two North American Feeders*, 1.0 edition.
- EPRI (2009b). *Enterprise Service Bus Implementation Profile: Integration Using IEC 61968*, 1.0 edition.
- EPRI (2009c). *Report to NIST on the Smart Grid Interoperability Standards Roadmap*.
- Erickson, V. L. and Cerpa, A. E. (2010). Occupancy based demand response HVAC control strategy. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, pages 7–12. ACM.
- Espinoza, A., Peña, Y., Nieves, J. C., Ortega, M., Peña, A., and Rodríguez, D. (2013). Supporting business workflows in smart grids: An intelligent nodes-based approach. *IEEE Transactions on Industrial Informatics*, 9(3):1384–1397.

- Esteban, O. and Peña, Y. (2012). Energy-aware foundry production scheduling. In *IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society*, pages 2827–2832.
- European Commission (2014). A policy framework for climate and energy in the period from 2020 to 2030. *COM (2014)*, 15.
- Eurostat (2017). *Sustainable development in the European Union. 2015 monitoring report of the EU Sustainable Development Strategy*. European Union.
- Fairchild, R. (2002). Automatic one-line creation tool for use with cim xml model exchange. In *Proceedings of the Power Engineering Society Summer Meeting*, volume 1, page 375.
- Farahat, M. (2004). Long-term industrial load forecasting and planning using neural networks technique and fuzzy inference method. In *Universities Power Engineering Conference, 2004. UPEC 2004. 39th International*, volume 1, pages 368–372. IEEE.
- Federal Energy Regulatory Commission (2010). National action plan on demand response.
- FI-WARE (2015–2016). Cloud FI-WARE. <https://account.lab.fiware.org/>.
- Fusco, F. and Ringwood, J. (2010). Short-term wave forecasting with ar models in real-time optimal control of wave energy converters. In *Industrial Electronics (ISIE), 2010 IEEE International Symposium on*, pages 2475–2480.
- Galvis, J. C. and Costa, A. (2016). Demand side management using time of use and elasticity price. *IEEE Latin America Transactions*, 14(10):4267–4274.
- García, S. and Herrera, F. (2008). An extension on “statistical comparisons of classifiers over multiple data sets” for all pairwise comparisons. *Journal of Machine Learning Research*, 9:2677–2694.

- Gellings, C. W. (1985). The concept of demand-side management for electric utilities. *Proceedings of the IEEE*, 73(10):1468–1470.
- Goldberg, D. E. et al. (1989). *Genetic algorithms in search, optimization, and machine learning*, volume 412. Addison-wesley Reading Menlo Park.
- Gomez, C. and Paradells, J. (2010). Wireless home automation networks: A survey of architectures and technologies. *IEEE Communications Magazine*, 48(6):92–101.
- Gómez-Lorente, D., Rabaza, O., Espín Estrella, A., and Peña-García, A. (2013). A new methodology for calculating roadway lighting design based on a multi-objective evolutionary algorithm. *Expert Systems with Applications*, 40(6):2156–2164.
- Goncalves Da Silva, P., Ilic, D., and Karnouskos, S. (2014). The impact of smart grid prosumer grouping on forecasting accuracy and its benefits for local electricity market trading. *Smart Grid, IEEE Transactions on*, 5(1):402–410.
- Goude, Y., Nedellec, R., and Kong, N. (2014). Local short and middle term electricity load forecasting with semi-parametric additive models. *Smart Grid, IEEE Transactions on*, 5(1):440–446.
- Gravel, M., Price, W. L., and Gagné, C. (2000). Scheduling jobs in an alcan aluminium foundry using a genetic algorithm. *International Journal of Production Research*, 38(13):3031–3041.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1):10–18.
- Hassan, S., Khosravi, A., Jaafar, J., and Belhaouari, S. (2012). Load forecasting accuracy through combination of trimmed forecasts. In Huang, T., Zeng, Z., Li, C., and Leung, C., editors, *Neural Information Processing*, volume 7663 of *Lecture Notes in Computer Science*, pages 152–159. Springer Berlin Heidelberg.
- Helal, A., Cho, K., and Lee, W. (2012). 3D modeling and simulation of human activities in smart spaces. In *Ubiquitous Intelligence & Computing and 9th International*

- Conference on Autonomic & Trusted Computing (UIC/ATC), 2012 9th International Conference on*, pages 112–119. IEEE.
- Helal, S., Lee, J., and Hossain, S. (2011). Persim-Simulator for human activities in pervasive spaces. In *Intelligent Environments (IE), 2011 7th International Conference on*, pages 192–199. IEEE.
- Helal, S., Mann, W., El-Zabadani, H., King, J., Kaddoura, Y., and Jansen, E. (2005). The Gator Tech Smart House: a programmable pervasive space. *Computer*, 38(3):50–60.
- Helal, S., Winkler, B., Lee, C., Kaddoura, Y., Ran, L., Giraldo, C., Kuchibhotla, S., and Mann, W. (2003). Enabling location-aware pervasive computing applications for the elderly. In *Pervasive Computing and Communications, 2003.(PerCom 2003). Proceedings of the First IEEE International Conference on*, pages 531–536. IEEE.
- Hidalgo, J., Baraglia, R., Perego, R., Lanchares, J., and Tirado, F. (2001). A parallel compact genetic algorithm for multi-fpga partitioning. In *Parallel and Distributed Processing, 2001. Proceedings. Ninth Euromicro Workshop on*, pages 113–120.
- Hinojosa, V. H. and Hoese, A. (2010). Short-term load forecasting using fuzzy inductive reasoning and evolutionary algorithms. *Power Systems, IEEE Transactions on*, 25(1):565–574.
- Hong, T., Pinson, P., and Fan, S. (2014). Global energy forecasting competition 2012. *International Journal of Forecasting*, 30(2):357 – 363.
- Hong, T., Wang, P., and Willis, H. (2011). A naïve multiple linear regression benchmark for short term load forecasting. In *Power and Energy Society General Meeting, 2011 IEEE*, pages 1–6.
- Huiqiu, C. and Junhua, G. (2010). Based on simulated annealing algorithm wang-tile texture synthesis optimization technology. In *Computer Design and Applications (ICCD), 2010 International Conference on*, volume 5, pages V5–142–V5–145.
- Hyndman, R. J. and Athanasopoulos, G. (2014). *Forecasting: principles and practice*. OTexts.

- Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679 – 688.
- Ibrahim, J., Chen, M., Lipsitz, S., and Herring, A. (2005). Missing-data methods for generalized linear models: A comparative review. *Journal of the American Statistical Association*, 100(469):332–346.
- IEC (2006). *International Standard IEC 62056-62 Electricity metering - Data exchange for meter reading, tariff and load control. Part 62: Interface classes*, 2 edition.
- IEC (2009). *IEC 61968-9 Application integration at electric utilities - System interfaces for distribution management. Part 9: Interfaces for meter reading and control*, 1.0 edition.
- IEC (2010). *IEC 61968 Application integration at electric utilities - System interfaces for distribution management. Part 11: Common Information Model (CIM)*, 1.0 edition.
- Intergovernmental Panel on Climate Change (2014). *Climate Change 2014: Mitigation of Climate Change*, volume 3. Cambridge University Press.
- International Energy Administration (2016). CO2 emissions from fuel combustion 2016.
- International Energy Agency (2016a). Medium-term renewable energy market report 2016.
- International Energy Agency (2016b). *World Energy Outlook 2016*. International Energy Agency.
- Jain, A. and Satish, B. (2009). Clustering based short term load forecasting using support vector machines. In *PowerTech, 2009 IEEE Bucharest*, pages 1–8.
- Jin, Q. and Li-xin, M. (2009). Combined simulated annealing algorithm for logistics network design problem. In *Intelligent Systems and Applications, 2009. ISA 2009. International Workshop on*, pages 1–4.

- Jordan, R., Birkin, M., and Evans, A. (2010). Agent-based simulation modelling of housing choice and urban regeneration policy. In 11<sup>th</sup> *International Conference on Multi-Agent-Based Simulation*, pages 152–166.
- Kachitvichyanukul, V. and Sitthitham, S. (2011). A two-stage genetic algorithm for multi-objective job shop scheduling problems. *Journal of Intelligent Manufacturing*, 22(3, SI):355–365.
- Kafali, Ö., Romero, A. E., and Stathis, K. (2014). *Activity Recognition for an Agent-Oriented Personal Health System*. Springer International Publishing, Cham.
- Kamara-Esteban, O., Pijoan, A., Alonso-Vicario, A., and Borges, C. E. (2016). On-demand energy monitoring and response architecture in a ubiquitous world - Video . Accessed: 2017-04-06.
- Kamara-Esteban, O., Pijoan, A., Alonso-Vicario, A., and Borges, C. E. (2017). On-demand energy monitoring and response architecture in a ubiquitous world. *Personal and Ubiquitous Computing*, pages 1–15.
- Karnouskos, S. and d. Holanda, T. N. (2009). Simulation of a smart grid city with software agents. In *Computer Modeling and Simulation, 2009. EMS '09. Third UKSim European Symposium on*, pages 424–429.
- Karnouskos, S. and De Holanda, T. N. (2009). Simulation of a smart grid city with software agents. In *Computer Modeling and Simulation, 2009. EMS'09. Third UKSim European Symposium on*, pages 424–429. IEEE.
- Khiyal, M. S. H., Khan, A., and Shehzadi, E. (2009). SMS based wireless home appliance control system (HACS) for automating appliances and security. *Issues in Informing Science and Information Technology*, 6:887–894.
- Khosravi, A. and Nahavandi, S. (2014). Load forecasting using interval type-2 fuzzy logic systems: Optimal type reduction. *Industrial Informatics, IEEE Transactions on*, 10(2):1055–1063.

- Kiesel, J. and Wenkel, K.-O. (2007). Spatial generalization methods based on the moving window approach and their applications on landscape analysis. In Olgierd Hryniewicz, Jan Studzinski, M. R. E., editor, *Shaker Verlag*, pages 619–626. Shaker Verlag. ISBN 978-3-8322-6397-3.
- Kim, S.-J. and Giannakis, G. B. (2013). Load forecasting via low rank plus sparse matrix factorization. In *Signals, Systems and Computers, 2013 Asilomar Conference on*, pages 1682–1686.
- Kondrak, G. and Beek, P. V. (1997). A theoretical evaluation of selected backtracking algorithms. *Artificial Intelligence*, 89:541–547.
- Kormanyos, B. and Pataki, B. (2013). Multilevel simulation of daily activities: Why and how? In *2013 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, pages 1–6. IEEE.
- Koza, J. R. (1992). *Genetic programming: on the programming of computers by means of natural selection*, volume 1. MIT press.
- Kreith, F. and Goswami, D. Y. (2016). *Energy management and conservation handbook*. CRC Press.
- Kümmel, R., Groscurth, H., and van Gool, W. (1988). Energy optimization in industrial models. In *System Modelling and Optimization*, volume 113 of *Lecture Notes in Control and Information Sciences*, pages 518–529.
- Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, pages 1217–1241.
- Labeeuw, W. and Deconinck, G. (2013). Residential electrical load model based on mixture model clustering and markov models. *Industrial Informatics, IEEE Transactions on*, 9(3):1561–1569.
- Lai, L. L. (1998). *Intelligent system applications in power engineering: evolutionary programming and neural networks*. John Wiley & Sons, Inc.

- Lee, Y., Marvin, A., and Porter, S. (1999). Genetic algorithm using real parameters for array antenna design optimisation. In *High Frequency Postgraduate Student Colloquium, 1999*, pages 8–13.
- Lesser, V., Atighetchi, M., Benyo, B., Horling, B., Raja, A., Vincent, R., Wagner, T., Xuan, P., and Zhang, S. X. (1999). The UMASS intelligent home project. In *Proceedings of the third annual conference on Autonomous Agents*, pages 291–298. ACM.
- Levy, R. (2006). A vision of demand response–2016. *The Electricity Journal*, 19(8):12–23.
- Li, Y. and Muller, B. (2007). Residential location and the biophysical environment: exurban development agents in a heterogeneous landscape. *Environment and Planning B*, 34:279–295.
- Liscouski, B. and Elliot, W. (2004). Final report on the august 14, 2003 blackout in the united states and canada: Causes and recommendations. *A report to US Department of Energy*, 40(4).
- Liu, J. M., Chen, R., Liu, L.-M., and Harris, J. L. (2006). A semi-parametric time series approach in modeling hourly electricity loads. *Journal of Forecasting*, 25(8):537–559.
- Luke, S., Cioffi-Revilla, C., Panait, L., and Sullivan, K. (2004). Mason: A new multi-agent simulation toolkit. In *Proceedings of the 2004 swarmfest workshop*, volume 8, page 44.
- Malleson, N., Heppenstall, A. J., and See, L. M. (2010). Crime reduction through simulation: An agent-based model of burglary. *Computers, Environment and Urban Systems*, 34(3):236–250.
- Mateos, G. and Giannakis, G. (2012). Spatiotemporal load curve data cleansing and imputation via sparsity and low rank. In *Smart Grid Communications (SmartGrid-Comm), 2012 IEEE Third International Conference on*, pages 653–658.

- Mateos, G. and Giannakis, G. (2013). Load curve data cleansing and imputation via sparsity and low rank. *Smart Grid, IEEE Transactions on*, 4(4):2347–2355.
- Mati, Y., Dauzere-Peres, S., and Lahlou, C. (2011). A general approach for optimizing regular criteria in the job-shop scheduling problem. *European Journal of Operational Research*, 212(1):33–42.
- Mendez-Vazquez, A., Helal, A., and Cook, D. (2009). Simulating events to generate synthetic data for pervasive spaces. In *Proceedings of the Workshop on Developing Shared Home Behaviour Datasets to Advance HCI and Ubiquitous Computing Research in Conjunction with CHI 2009*, pages 4–9, Boston, MA.
- Meng, H. and Xin, G. (2010). UAV route planning based on the genetic simulated annealing algorithm. In *Mechatronics and Automation (ICMA), 2010 International Conference on*, pages 788–793.
- Myrtveit, I., Stensrud, E., and Olsson, U. (2001). Analyzing data sets with missing data: An empirical evaluation of imputation methods and likelihood-based methods. *IEEE Transactions on Software Engineering*, 27(11):999–1013.
- National Electrical Manufacturers Association (2006). *ANSI C.12*, 6 edition.
- National Institute of Standards and Technology, U.S. Department of Commerce (2009). *NIST Framework and Roadmap for Smart Grid Interoperability Standards*, 1.0 edition.
- Neumann, S. and Nielsen, T. (2010). Cim interoperability challenges. In *IEEE Power and Energy Society General Meeting*, pages 3–5.
- New York Independent System Operator (2001). *Day-Ahead Scheduling Manual*.
- Nieves, J., Santos, I., Penya, Y. K., Brezo, E., and Bringas, P. G. (2010). Enhanced foundry production control. In *Proceedings of the 21st international conference on Database and expert systems applications: Part I*, pages 213–220.

- North, M. J., Howe, T. R., Collier, N. T., and Vos, J. (2005). The repast symphony runtime system. In *Agent 2005 Conference on Generative Social Processes, Models, and Mechanisms*. Argonne, Illinois, USA: Argonne National Laboratory. Citeseer.
- Okeyo, G., Chen, L., Wang, H., and Sterritt, R. (2012). Dynamic sensor data segmentation for real-time knowledge-driven activity recognition. *Pervasive and Mobile Computing*, 10:155–172.
- Oliveira, P., Pinto, T., Morais, H., and Vale, Z. (2012). Masgrip: A multi-agent smart grid simulation platform. In *Power and Energy Society General Meeting, 2012 IEEE*, pages 1–8.
- Osborne, M., Rogers, A., Roberts, S., Ramchurn, S., and Jennings, N. (2011). *Bayesian Time Series Models*, chapter 16.- Bayesian Gaussian process models for multi-sensor time-series prediction. Cambridge University Press.
- Pai, P.-F. and Hong, W.-C. (2005). Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. *Electric Power Systems Research*, 74(3):417–425.
- Pasten, C. and Santamarina, J. C. (2012). Energy and quality of life. *Energy Policy*, 49:468–476.
- Peña, A. and Peña, Y. K. (2011). Distributed semantic repositories in smart grids. In *Proceedings of the 9<sup>th</sup> International Conference on Industrial Informatics (INDIN)*. in press.
- Penya, Y., Peña, A., and Esteban, O. (2011). Semantic integration of IEC 60870 into CIM. In *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*, pages 428–433.
- Penya, Y. K., Borges, C. E., Peña, A., and Esteban, O. (2013). Service-orientation vs. real-time: Integrating smart-homes into the smart-grid. In *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, pages 5755–5760.

- Penya, Y. K., Nieves, J. C., Espinoza, A., Borges, C. E., Peña, A., and Ortega, M. (2012). Distributed semantic architecture for smart grids. *Energies*, 5(11):4824–4843.
- Perera, C., Zaslavsky, A., Christen, P., and Georgakopoulos, D. (2014). Context aware computing for the internet of things: A survey. *Communications Surveys & Tutorials, IEEE*, 16(1):414–454.
- Pijoan, A. and Borges, C. (2012). Tutorial de importación de datos de Catastro a OSM (cat2osm). In *Actas de las VI Jornadas de SIG Libre, Girona 21–23 marzo, España*. Servicio de SIG y Teledetección - SIGTE - de la Universitat de Girona.
- Pijoan, A., Kamara-Esteban, O., and Borges, C. E. (2016). GTPlat - Green Travelling Scenario simulation tool. Accessed: 2017-04-06.
- Pinto, R. T., Rodrigues, S. F., Wiggelinkhuizen, E., Scherrer, R., Bauer, P., and Pierik, J. (2012). Operation and power flow control of multi-terminal dc networks for grid integration of offshore wind farms using genetic algorithms. *Energies*, 6(1):1–26.
- Pipattanasomporn, M., Feroze, H., and Rahman, S. (2009). Multi-agent systems in a distributed smart grid: Design and implementation. In *Power Systems Conference and Exposition, 2009. PSCE '09. IEEE/PES*, pages 1–8.
- PJM Interconnection System Operator (2014). *Transmission Reliability Margin TRM Implementation Document*.
- Pohlert, T. (2014). The pairwise multiple comparison of mean ranks package (pncmr). *R package*.
- Pournaras, E., Vasirani, M., Kooij, R., and Aberer, K. (2014). Decentralized planning of energy demand for the management of robustness and discomfort. *Industrial Informatics, IEEE Transactions on*, PP(99):1–1.
- Qin, T., Zhang, H., and Guo, Y. (2002). A modified simulated annealing algorithm for joint configuration of the optical and electrical layer in intelligent optical networks. In *Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE*, volume 3, pages 2847–2850 vol.3.

- Randall Munroe (2011). xkcd - A Webcomic - Significant. Accessed: 2015-12-13.
- Rashidi, P. and Cook, D. (2011). Discovering activities to recognize and track in a smart environment. *Knowledge and Data Engineering, IEEE Transactions on*, 23(4):527–539.
- Red Eléctrica de España (2014). *Canary electricity demand in real time*. Red Eléctrica de España.
- Reed, G. F., Lynn, F., and Meade, B. D. (2002). Use of coefficient of variation in assessing variability of quantitative assays. *Clinical and Diagnostic Laboratory Immunology*, 9(6):1235–1239.
- Ricci, A., Rodriguez-Aguilar, J., Pijoan, A., and Zambonelli, F. (2015). Mixed environments for MAS: Bringing humans in the loop. In Weyns, D. and Michel, F., editors, *Agent Environments for Multi-Agent Systems IV*, volume 9068 of *Lecture Notes in Computer Science*, pages 52–60. Springer International Publishing.
- River, C. (2005). Primer on demand-side management with an emphasis on price-responsive programs. *prepared for The World Bank by Charles River Associates, Tech. Rep.*
- Robinson, A. P., Duursma, R. A., and Marshall, J. D. (2005). A regression-based equivalence test for model validation: shifting the burden of proof. *Tree Physiology*, 25(7):903–913.
- Robles, R. J. and Kim, T.-h. (2010). Review: context aware tools for smart home development. *International Journal of Smart Home*, 4(1).
- Rohjans, S., Uslar, M., and Appelrath, H.-J. (2010a). OPC UA and CIM: Semantics for the smart grid. In *Proceedings of the IEEE Transmission and Distribution Conference and Exposition*, pages 1–8.
- Rohjans, S., Uslar, M., and Niesse, A. (2010b). Towards a semantic communication interface between information and automation layer in smart grids. In *Proceedings*

- 
- of the 4<sup>th</sup> International Conference on Integration of Renewable and Distributed Energy Resources. Poster presentation.
- Roland, B., di Martinelly, C., and Riane, F. (2006). Operating theatre optimization: A resource-constrained based solving approach. In *Service Systems and Service Management, 2006 International Conference on*, volume 1, pages 443–448.
- R.Sivaraj, D. (2011). A review of selection methods in genetic algorithm. *International Journal of Engineering Science and Technology*, 3(5):3792–3797.
- Russell, S. J., Norvig, P., Canny, J. F., Malik, J. M., and Edwards, D. D. (2003). *Artificial Intelligence: A Modern Approach*. Prentice Hall.
- Santos, A. and Dourado, A. (1999). Constrained GA applied to production and energy management of a pulp and paper mill. In *Proceedings of the 1999 ACM symposium on Applied computing*, SAC '99, pages 324–332.
- Sato, S., Otori, K., Takizawa, A., Sakai, H., Ando, Y., and Kawamura, H. (2002). Applying genetic algorithms to the optimum design of a concert hall. *Journal of Sound and Vibration*, 258(3):517–526.
- Schafer, J. and Olsen, M. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, 33(4):545–571.
- Sechilariu, M., Wang, B., and Locment, F. (2013). Building integrated photovoltaic system with energy storage and smart grid communication. *Industrial Electronics, IEEE Transactions on*, 60(4):1607–1618.
- Senjyu, T., Takara, H., Uezato, K., and Funabashi, T. (2002). One-hour-ahead load forecasting using neural network. *Power Systems, IEEE Transactions on*, 17(1):113–118.
- Skala, K., Davidovic, D., Afgan, E., Sovic, I., and Sojat, Z. (2015). Scalable distributed computing hierarchy: Cloud, fog and dew computing. *Open Journal of Cloud Computing (OJCC)*, 2(1):16–24.

- Smart Grid Strategic Group (SG3), I. (2010). *IEC Smart Grid Standardization Roadmap*, 1.0 edition.
- Strbac, G. (2008). Demand side management: Benefits and challenges. *Energy policy*, 36(12):4419–4426.
- Symantec Security Response (2016). IoT devices being increasingly used for DDoS attacks. <https://www.symantec.com/connect/blogs/iot-devices-being-increasingly-used-ddos-attacks>.
- Synnott, J., Nugent, C., and Jeffers, P. (2015). Simulation of Smart Home Activity Datasets. *Sensors (Basel, Switzerland)*, 15(6):14162–79.
- Tapia, E. M., Intille, S. S., and Larson, K. (2004). *Pervasive Computing*, volume 3001 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Taylor, D. and Kmethy, G. (2010). DLMS/COSEM/CIM study report: 61968-9 message profiles for DLMS/COSEM. Technical report, DLMS User Association, IEC TC 13 WG15, and IEC TC 57 WG14.
- Telit Wireless Solutions (2011). AT commands reference guide.
- Terciyani, E., Demirci, T., Kucuk, D., Sarac, M., Cadirci, I., and Ermis, M. (2014). Enhanced nationwide wind-electric power monitoring and forecast system. *Industrial Informatics, IEEE Transactions on*, 10(2):1171–1184.
- Tinsa (2013). Precio medio del  $m^2$  por provincia. <http://www.elmundo.es/suivienda/sv/tasaciones/>.
- Tomoiağâ, B., ChindriÅŞ, M., Sumper, A., Sudria-Andreu, A., and Villafafila-Robles, R. (2013). Pareto optimal reconfiguration of power distribution systems using a genetic algorithm based on nsga-ii. *Energies*, 6(3):1439–1455.
- U.S. Department of Energy (2006). Benefits of demand response in electricity markets and recommendations for achieving them. *U.S. Dept. Energy, Washington, DC, USA, Tech. Rep.*

- U.S. Energy Information Administration (2016). *International Energy Outlook 2016 (IEO2016)*. United States Energy Information Administration.
- Uslar, M., Rohjans, S., Schulte, S., and Steinmetz, R. (2008). Building the semantic utility with standards and semantic web services. In *Proceedings of the OTM Confederated International Workshops and Posters On the Move to Meaningful Internet Systems*, pages 1026–1035. Springer-Verlag.
- Van Nguyen, T., Kim, J. G., and Choi, D. (2009). ISS: the interactive smart home simulator. In *Advanced Communication Technology, 2009. ICACT 2009. 11th International Conference on*, volume 3, pages 1828–1833. IEEE.
- Velleman, J. D. and Bratman, M. E. (1991). Intention, plans, and practical reason. *The Philosophical Review*, 100(2):277.
- Vilar, J. M., Cao, R., and Aneiros, G. (2012). Forecasting next-day electricity demand and price using nonparametric functional methods. *International Journal of Electrical Power & Energy Systems*, 39(1):48 – 55.
- Virone, G., Noury, N., and Demongeot, J. (2002). A system for automatic measurement of circadian activity deviations in telemedicine. *Biomedical Engineering, IEEE Transactions on*, 49(12):1463–1469.
- von Meier, A. (2006). *Electric Power Systems: A Conceptual Introduction*. Wiley survival guides in engineering and science. Wiley.
- Walker, E. and Nowacki, A. (2011). Understanding equivalence and noninferiority testing. *Journal of General Internal Medicine*, 26(2):192–196.
- Wan, M., Xu, X., and Nan, J. (2011). Flexible job-shop scheduling with integrated genetic algorithm. In *Advanced Computational Intelligence (IWACI), 2011 Fourth International Workshop on*, pages 13–16.
- Wang, X., Schulz, N., and Neumann, S. (2003). Cim extensions to electrical distribution and cim xml for the iee radial test feeders. In *IEEE Transactions on Power Systems*, volume 18, pages 1021–1028.

- Watanabe, M., Furukawa, M., Mizoe, A., and Watanabe, T. (2001). GA applications to physical distribution scheduling problem. *Industrial Electronics, IEEE Transactions on*, 48(4):724–730.
- Wei, Y. and Blake, M. B. (1992). Job shopscheduling by simulated annealing. *Operations Research*, 40(1):72–75.
- Wei, Y. and Blake, M. B. (2010). Service-Oriented Computing and Cloud Computing: Challenges and Opportunities. *IEEE Internet Computing*, 14(6):72–75.
- Weyns, D. and Michel, F. (2015). *Agent Environments for Multi-Agent Systems IV*. Springer International Publishing.
- Weyns, D., Parunak, H. V. D., Michel, F., Holvoet, T., and Ferber, J. (2004). Environments for multi-agent systems. In *First International Workshop (E4MAS 2004)*, volume 3374.
- Willis, H. (1996). *Spatial Load Forecasting*. Marcel Dekker Inc., New York.
- Wooldridge, M. and Jennings, N. R. (1995). Intelligent agents: theory and practice. *The Knowledge Engineering Review*, 10:115–152.
- World Energy Council (2016). *World Energy Resources 2016*. World Energy Council.
- Xia, F., Yang, L. T., Wang, L., and Vinel, A. (2012). Internet of things. *International Journal of Communication Systems*, 25(9):1101.
- Xie, L., Gu, Y., Zhu, X., and Genton, M. (2014). Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch. *Smart Grid, IEEE Transactions on*, 5(1):511–520.
- Xu, B., Zhong, P.-A., Wan, X., Zhang, W., and Chen, X. (2012a). Dynamic feasible region genetic algorithm for optimal operation of a multi-reservoir system. *Energies*, 5(8):2894–2910.

- Xu, X., Yin, E., and Zou, F. (2010). Researches on optimal scheduling for aluminum industry continuous casting and rolling production. In *Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference on*, volume 1, pages 348–352.
- Xu, Y., Dong, Z.-Y., Xu, Z., Meng, K., and Wong, K. P. (2012b). An intelligent dynamic security assessment framework for power systems with wind power. *Industrial Informatics, IEEE Transactions on*, 8(4):995–1003.
- Yamazaki, T. (2006). Beyond the smart home. In *Hybrid Information Technology, 2006. ICHIT'06. International Conference on*, volume 2, pages 350–355. IEEE.
- Yeager, K. (2016). *Evolution of the Smart Distribution Grid*. John Wiley Sons, Ltd.
- Yoshimi, M., Swarup, K., and Izui, Y. (1993). Optimal economic power dispatch using genetic algorithms. In *Neural Networks to Power Systems, 1993. ANNPS '93., Proceedings of the Second International Forum on Applications of*, pages 157–162.
- Zdravković, M., Trajanović, M., Sarraipa, J., Jardim-Gonçalves, R., Lezoche, M., Aubry, A., and Panetto, H. (2016). Survey of Internet-of-Things platforms. In *6th International Conference on Information Society and Technology, ICIST 2016*, volume 1, pages 216–220, Kopaonik, Serbia. ISBN: 978-86-85525-18-6.
- Zhang, P., Xu, Z., and Du, F. (2008). Optimizing casting parameters of ingot based on neural network and genetic algorithm. In *Natural Computation, 2008. ICNC '08. Fourth International Conference on*, volume 4, pages 545–548.
- Zhang, Q. and Grossmann, I. E. (2016). Planning and scheduling for industrial demand side management: advances and challenges. In *Alternative Energy Sources and Technologies*, pages 383–414. Springer.
- Zhang, R. and Wu, C. (2011). A simulated annealing algorithm based on block properties for the job shop scheduling problem with total weighted tardiness objective. *Computers & Operations Research*, 38(5):854–867.

- Zhu, J., Cai, X., Pan, P., and Gu, R. (2014). Multi-objective structural optimization design of horizontal-axis wind turbine blades using the non-dominated sorting genetic algorithm ii and finite element method. *Energies*, 7(2):988–1002.
- Zoha, A., Gluhak, A., Imran, M. A., and Rajasegarar, S. (2012). Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors*, 12(12):16838–16866.



# Publications

## A.1 Indexed journal articles

- **Kamara-Esteban, O.**, Pijoan, A., Alonso-Vicario, A., and Borges, C. E. (2017). *On-demand energy monitoring and response architecture in a ubiquitous world*. Personal and Ubiquitous Computing, 1-15. ISSN: 16174909. DOI: 10.1007/s00779-017-1014-4.
- **Kamara-Esteban, O.**, Azkune G., Pijoan, A., and Borges, C. E. *MASSHA: An Agent-Based Approach for Human Activity Simulation in Intelligent Environments*. Pervasive and Mobile Computing. In review process.

## A.2 Conference papers

- **Esteban, O. K.**, and Penya, Y. K. (2012, October). *Energy-aware foundry production scheduling*. In IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society (pp. 2827-2832). IEEE. DOI: 10.1109/IECON.2012.6389447.

- **Penya, Y. K., Kamara, O.,** and **Pena, A.** (2011, November). *IEC60870 meter SOA management*. In Innovative Smart Grid Technologies Asia (ISGT), 2011 IEEE PES (pp. 1-7). IEEE. DOI: 10.1109/ISGT-Asia.2011.6167092.
- **Penya, Y. K., Pena, A.,** and **Esteban, O. K.** (2011, October). *Semantic integration of IEC 60870 into CIM*. In Smart grid communications (SmartGridComm), 2011 IEEE international conference on (pp. 428-433). IEEE. DOI: 10.1109/Smart-GridComm.2011.6102360.
- **Penya, Y. K., Borges, C. E., Pena, A.,** and **Kamara-Esteban, O.** *Service-orientation vs. real-time: Integrating smart-homes into the smart-grid*. In Industrial Electronics Society, IECON 2013-39th Annual Conference of the IEEE (pp. 5755-5760). IEEE. DOI: 10.1109/IECON.2013.6700077
- **Kamara-Esteban, O., Sorrosal, G., Pijoan, A., Castillo-Calzadilla, T., Iriarte-Lopez, X., Macarulla-Arenaza, A. M.,** and **Borges, C. E.** (2016, July). *Bridging the Gap between Real and Simulated Environments: A Hybrid Agent-Based Smart Home Simulator Architecture for Complex Systems*. In Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), 2016 Intl IEEE Conferences (pp. 220-227). IEEE. DOI: 10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0052.
- **Pijoan, A., Kamara-Esteban, O.,** **Borges, C. E.** and **Penya, Y. K.**. *GIS and MAS tight coupling for Spatial Load Forecasting*. In 13th International Conference on Autonomous Agents and Multiagent Systems 2014. ISBN: 978-1-4503-2738-1.
- **Borges, C. E., Kamara-Esteban, O.,** **Pijoan, A.,** and **Penya, Y. K.**. *Multi-Agent GIS System for Improved Spatial Load Forecasting (Demonstration)*. In 13th International Conference on Autonomous Agents and Multiagent Systems. ISBN: 978-163439131-3.

### A.3 Book sections

- Pijoan, A., **Kamara-Esteban, O.**, and Borges, C. E. *Environment Modelling for Spatial Load Forecasting*. Environments for Multi-Agent Systems IV ISSN: 03029743. ISBN: 978-331923849-4. DOI: 10.1007/978-3-319-23850-0\_12.



## Declaration

---

I, Oihane Kamara Esteban, herewith declare that this dissertation is my own original work, carried out as a doctoral student at the University of Deusto. All assistance received and notions from other sources have been identified as such, acknowledging their correspondent contributions and citing them properly.

This work contains no material which has been presented in identical or similar form to any examination board, except where due acknowledgment is made in the dissertation.



This dissertation was finished writing on June 4<sup>th</sup>, 2017.