



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



FACULTAD DE
INGENIERÍA
UDELAR

Caracterización del consumo eléctrico residencial en Uruguay

Análisis de datos de consumo eléctrico y aplicación de
técnicas de inteligencia computacional para su
desagregación

Juan P. Chavat

Tesis de Maestría presentada al Programa de Posgrado en Informática-PEDECIBA, Facultad de Ingeniería de la Universidad de la República, en cumplimiento parcial de los requerimientos para la obtención del título de Magíster en Informática-PEDECIBA.

Tutores

Sergio Nesmachnow

Tribunal

Pablo Rodríguez-Bocca
Cecilia Galarza
Mario Vignolo

Montevideo, Uruguay
Febrero de 2023



UNIVERSIDAD
DE LA REPUBLICA
URUGUAY



Caracterización del consumo eléctrico residencial en Uruguay

Análisis de datos de consumo eléctrico y aplicación de
técnicas de inteligencia computacional para su
desagregación

Juan P. Chavat

Programa de Maestría en Informática–PEDECIBA
Facultad de Ingeniería
Universidad de la República

Montevideo – Uruguay
Febrero de 2023



UNIVERSIDAD
DE LA REPUBLICA
URUGUAY



Caracterización del consumo eléctrico residencial en Uruguay

Análisis de datos de consumo eléctrico y aplicación de
técnicas de inteligencia computacional para su
desagregación

Juan P. Chavat

Tesis de Maestría presentada al Programa de
Posgrado en Informática-PEDECIBA, Facultad de
Ingeniería de la Universidad de la República, como
parte de los requisitos necesarios para la obtención
del título de Magíster en Informática-PEDECIBA.

Director:

Prof. Dr. Sergio Nesmachnow

Director académico:

Prof. Dr. Sergio Nesmachnow

Montevideo – Uruguay

Febrero de 2023

Chavat, Juan P.

Caracterización del consumo eléctrico residencial en Uruguay / Juan P. Chavat. - Montevideo: Universidad de la República, Facultad de Ingeniería, 2023.

XXIV, 135 p.: il.; 29, 7cm.

Director:

Sergio Nesmachnow

Director académico:

Sergio Nesmachnow

Tesis de Maestría – Universidad de la República, Programa en Informática–PEDECIBA, 2023.

Referencias bibliográficas: p. 119 – 126.

1. análisis de datos, 2. consumo eléctrico, 3. NILM, 4. desagregación, 5. series temporales. I. Nesmachnow, Sergio, . II. Universidad de la República, Programa de Posgrado en Informática–PEDECIBA. III. Título.

INTEGRANTES DEL TRIBUNAL DE DEFENSA DE TESIS

Pablo Rodríguez-Bocca

Cecilia Galarza

Mario Vignolo

Montevideo – Uruguay
Febrero de 2023

RESUMEN

El consumo eléctrico a nivel global ha aumentado y lo seguirá haciendo de forma ininterrumpida en corto y mediano plazo. La tendencia despierta en investigadores, proveedores y consumidores, el interés de entender cómo se consume el recurso. Un insumo para el entendimiento son los datos de consumo, obtenidos en procesos de monitoreo no intrusivo de cargas o mediante la instalación de medidores dentro de los establecimientos. Un buen entendimiento del uso permite planificar la demanda basándose en datos reales (e.g., elaborar tarifas, controlar picos de consumo, etc.) y promover el uso consciente y óptimo del recurso (e.g., identificar dispositivos de mayor consumo, detectar comportamientos anómalos, etc). Esta tesis presenta una caracterización del consumo eléctrico residencial en Uruguay. Se presenta un algoritmo de desagregación de consumo eléctrico residencial, utilizando técnicas de inteligencia computacional y datos sintéticos de consumo. Se crea un conjunto de datos de consumo eléctrico de hogares en Uruguay, Electricity Consumption Dataset of Uruguay, ECD-UY, que cuenta con tres subconjuntos: consumo total, consumo del termotanque, y consumo desagregado de electrodomésticos. Finalmente, se presenta un índice de consumo potencial construido a partir de datos socio-económicos y datos de consumo de electrodomésticos presentes en el mercado local. El rendimiento del algoritmo de desagregación fue comparado con el de dos métodos de la literatura: Combinatorial Optimization y Factorial Hidden Markov Models, obteniendo rendimientos satisfactorios en todas las instancias evaluadas. El conjunto de datos ECD-UY es el primero de su tipo referente a consumo residencial de un país de América Latina. Por último, el índice de consumo potencial fue validado utilizando regresiones lineales y datos de consumo total presentes en ECD-UY, presentando valores de R^2 altos para los niveles geográficos de departamentos y segmentos censales.

Palabras clave: análisis de datos, consumo eléctrico, NILM, desagregación, series temporales.

ABSTRACT

Global electricity consumption has been experiencing a steady increase and is projected to continue this upward trajectory in the short to medium term. This trend has sparked interest among researchers, providers, and consumers to gain a deeper understanding of consumption patterns. Data pertaining to consumption, obtained through non-intrusive load monitoring or by employing metering devices within establishments, provides crucial insights. A comprehensive understanding of usage enables demand planning based on real data (e.g., tariff formulation, peak consumption management, etc.) and fosters conscious and optimal resource utilization (e.g., identifying high-consumption appliances, detecting anomalies, etc). This thesis offers a characterization of residential electricity consumption in Uruguay. It introduces a residential electricity disaggregation algorithm that leverages computational intelligence techniques and synthetic consumption data. The thesis presents the Electricity Consumption Dataset of Uruguay (ECD-UY), which encompasses three subsets: total consumption, water heater consumption, and disaggregated appliance consumption. Furthermore, we introduce a potential consumption index built from socioeconomic data and local appliance consumption data. The performance of the disaggregation algorithm has been compared with two methodologies from the existing literature: Combinatorial Optimization and Factorial Hidden Markov Models. The results indicate satisfactory performance across all evaluated instances. The ECD-UY dataset represents the first of its kind related to residential consumption in a Latin American country. Lastly, the potential consumption index was validated using linear regressions and total consumption data from the ECD-UY, demonstrating high R^2 values for geographical levels of departments and census segments.

Keywords:

data analysis, Electricity consumption, NILM, disaggregation, time series.

Lista de figuras

2.1	Etapas del análisis de datos, basada en Runkler (2020) y Tsai et al. (2015)	8
3.1	Taxonomía de firmas de electrodomésticos propuesta por Hart (1992) . Los rectángulos en celeste corresponden a las utilizadas por Hart.	17
4.1	Resumen de la etapa de entrenamiento del algoritmo DS	41
4.2	Resumen de la etapa de evaluación del algoritmo DS	41
4.3	Porcentaje de tiempo encendido/apagado de los electrodomésticos en los conjuntos de datos sintéticos.	49
4.4	Valor-F de CO, FHMM, y DS para la heladera (izquierda) y la jarra eléctrica (derecha) en las instancias #4-5, #4-10, y #4-15.	58
4.5	Valor-F de CO, FHMM y DS para los electrodomésticos heladera (izquierda) y jarra eléctrica (derecha) en las instancias #5-5, #5-10 y #5-15.	62
4.6	Valor-F de los algoritmos CO, FHMM, y DS para las instancias con período de muestreo de 5 minutos, para el electrodoméstico lavarropa.	64
5.1	Vista general del sistema de recolección, comunicación, y procesamiento de los datos de consumo eléctrico residencial.	67
5.2	Dispositivos instalados por la empresa UTE, para la recolección de datos de consumo eléctrico en los hogares. Imágenes brindadas por UTE.	71
5.3	Arquitectura y procesos del sistema de comunicación de los datos de consumo total de los hogares.	72

5.4	Arquitectura y procesos del sistema de comunicación empleado en la recolección de los datos de consumo eléctrico de termotanques.	73
5.5	Histograma de cantidad de hogares por intervalo de cantidad de días monitoreados, teniendo en cuenta todos los días y solo los que cumplen el criterio de completitud (i.e., cuentan con al menos 95 % de los registros).	82
5.6	Consumo promedio durante un día, calculado cada 15 minutos.	83
5.7	Ejemplo de un mes de consumo total de un hogar (cliente #69806), con presencia de <i>gaps</i> y valores <i>outliers</i> y su corrección (línea roja y verde).	84
5.8	Ejemplo de 24 horas de consumo de un termotanque (cliente #115609) y dos posibles correcciones a los <i>gaps</i> y valores <i>outliers</i> : utilizando consumo cero (círculos rojos) o consumo igual a la potencia nominal del dispositivo (círculos verdes).	85
5.9	Potencia promedio total y del termotanque durante un día, utilizando datos de 135 hogares con rango de fechas desde el 15/07/2019 al 30/11/2019.	86
5.10	Ejemplo de patrones de consumo de termotanques.	86
5.11	Distribución de los hogares con datos de consumo de termotanque e información de clientes, en el departamento de Montevideo, Uruguay.	87
5.12	Potencia media demandada por hora (histograma) y acumulada en un día (barras apiladas), para cada electrodoméstico del subconjunto de consumo desagregados.	89
6.1	Potencia media consumida por electrodoméstico utilizado en el índice.	101
6.2	Porcentaje de hogares por departamento presentes en ECD-UY con consumo eléctrico total, luego de filtrar por año y percentil.	102
6.3	Porcentaje de hogares por sección censal de Montevideo presentes en ECD-UY con consumo eléctrico total, luego de filtrar por año y percentil.	103
6.4	Porcentaje de hogares por segmento censal de Montevideo presentes en ECD-UY con consumo eléctrico total, luego de filtrar por año y percentil.	103

6.5	Valores calculados del índice para los departamentos de Uruguay.	105
6.6	Valores obtenidos para el índice calculado para las secciones censales de Montevideo	106
6.7	Valores obtenidos para el índice calculado para los segmentos censales de Montevideo	108
6.8	Consumo potencial y consumo real promedio para cada departamento de Uruguay.	110
6.9	Valores obtenidos para el índice y valor de consumo promedio calculado a partir de consumo real, para los departamentos de Uruguay.	110

Lista de tablas

3.1	Conjuntos de datos de consumo energético.	28
4.1	Descripción de los electrodomésticos extra incluidos en instancias del conjunto de datos sintético. I representa el intervalo de muestreo expresado en segundos.	46
4.2	Detalle de las instancias de los conjuntos de datos sintéticos. Los usos se codifican como EN para entrenamiento y EV para evaluación.	49
4.3	Resultados de CO, FHMM y DS para la instancia #1.	53
4.4	Resultados de CO, FHMM y DS para la instancia #2.	54
4.5	Resultados de CO, FHMM y DS para la instancia #3.	55
4.6	Resultados de CO, FHMM y DS para la instancia #4-5.	56
4.7	Resultados de CO, FHMM y DS para la instancia #4-10.	57
4.8	Resultados de CO, FHMM y DS para la instancia #4-15.	58
4.9	Resultados de CO, FHMM y DS para la instancia #5-5.	59
4.10	Resultados de CO, FHMM y DS para la instancia #5-10.	60
4.11	Resultados de CO, FHMM y DS para la instancia #5-15.	61
5.1	Resumen de los tres subconjuntos de consumo eléctrico y el conjunto de información de clientes, contenido en ECD-UY. *Los períodos pueden variar, dependiendo del cliente.	67
5.2	Características de los hogares del subconjunto de datos de consumo eléctrico desagregado por electrodomésticos.	70
5.3	Descripción de los registros correspondientes al subconjunto de datos de consumo eléctrico total de los hogares.	76
5.4	Descripción de los registros correspondientes al subconjunto de datos de consumo eléctrico de termotanques.	77

5.5	Descripción de los registros del subconjunto de consumo eléctrico desagregado por electrodomésticos.	79
5.6	Descripción de los registros correspondientes al conjunto de información de clientes, contenido en el archivo <code>customers.csv</code>	80
5.7	Cantidad de días por intervalo de cantidad de registros.	81
5.8	Cantidad de días por cantidad de hogares, contando todos los días y solo los que cumplen el criterio de completitud (i.e., cuentan con al menos 95 % de los registros).	82
5.9	Información detallada sobre el monitoreo de los electrodomésticos y los <i>gaps</i> detectados.	88
6.1	Descripción de los datos de ECH (2019) utilizados para la construcción del índice.	99
6.2	Información de electrodomésticos utilizada para la creación del índice.	101
6.3	Lista de departamentos de Uruguay, junto al valor resultante del índice.	105
6.4	Lista de las secciones de Montevideo, junto al valor resultante del índice.	107
6.5	Lista truncada de los segmentos de Montevideo, junto al valor resultante del índice.	109
1.1	Descripción de los archivos de información de clientes, relación cliente-medidor y consumo eléctrico del subconjunto de consumo total de los hogares.	130
1.2	Descripción de los archivos de relaciones entre cliente-medidor, cliente- <i>timer</i> , y cliente-historico de contratos, del subconjunto de consumo eléctrico de termotanques.	132
1.3	Descripción del archivo con información de clientes del subconjunto de consumo eléctrico de termotanques.	133
1.4	Descripción de los archivos de consumo eléctrico total y de termotanques del subconjunto de consumo eléctrico de termotanques.	134
1.5	Descripción de los archivos de información de electrodomésticos y señales medidas (e.g., consumo eléctrico) del subconjunto de consumo eléctrico de electrodomésticos.	135
1.6	Detalle de las señales medidas en el subconjunto de consumo eléctrico de electrodomésticos.	135

Lista de siglas

CO Combinatorial Optimization [36](#), [50](#), [52](#), [53](#), [54](#), [55](#)

CSV Comma-Separated Values [42](#)

DS Desagregación Simple [39](#), [52](#), [53](#), [54](#), [55](#)

ECH Encuesta Continua de Hogares [4](#)

FHMM Factorial Hidden Markov Model [36](#), [52](#), [53](#), [54](#), [55](#)

HDF Hierarchical Data Format [42](#)

INE Instituto Nacional de Estadísticas [4](#)

YAML Yet Another Markup Language [42](#)

Tabla de contenidos

Lista de figuras	XIII
Lista de tablas	XVII
Lista de siglas	XIX
1 Introducción	1
2 Análisis del consumo eléctrico residencial y monitoreo no intrusivo de las cargas	7
2.1 Análisis de datos de consumo eléctrico	7
2.2 Monitoreo no intrusivo de cargas	10
2.2.1 Monitoreo no intrusivo	10
2.2.2 Desagregación de consumo energético	11
2.2.3 Formulación matemática del problema de desagregación de cargas eléctricas	11
2.2.4 Descripción de un escenario de desagregación con características de uno real	12
3 Trabajos relacionados	15
3.1 Monitoreo no intrusivo de cargas	15
3.2 Conjuntos de datos de consumo de energía	24
3.3 Análisis de datos de consumo eléctrico residencial	28
3.4 Resumen	32
4 Algoritmo de monitoreo no intrusivo de consumo eléctrico basado en similitud de patrones	35
4.1 Introducción	35
4.2 Descripción general	36

4.3	Etapas del algoritmo	38
4.3.1	Etapa de entrenamiento	38
4.3.2	Etapa de evaluación	39
4.4	Implementación	42
4.5	Análisis experimental	43
4.5.1	Conjuntos de datos	43
4.5.2	Instancias de evaluación	46
4.5.3	Métodos base para la comparación de resultados	50
4.5.4	Métricas de evaluación de resultados	51
4.6	Resultados	52
4.6.1	Instancias sin presencia de ruido ni variación de período de muestreo	52
4.6.2	Instancias sin presencia de ruido, con variación de período de muestreo	56
4.6.3	Instancias con presencia de ruido y variación de período de muestreo	59
4.7	Conclusiones	64
5	Conjunto de datos de consumo eléctrico residencial en Uruguay	65
5.1	Introducción	65
5.2	Metodología	67
5.2.1	Recolección de datos	67
5.2.2	Comunicación de datos	71
5.2.3	Pre-procesamiento de datos	74
5.3	Estructura de datos	76
5.4	Validación	80
5.4.1	Consumo eléctrico total de los hogares	80
5.4.2	Consumo eléctrico de termotanques	83
5.4.3	Consumo eléctrico desagregado por electrodomésticos	87
5.5	Repositorio de datos	89
5.6	Utilización de los datos	90
5.7	Aplicabilidad	91
5.8	Conclusiones	93

6	Análisis del consumo eléctrico residencial en Uruguay	95
6.1	Introducción	95
6.2	Motivación e hipótesis	96
6.3	Abordaje propuesto para el análisis de consumo	97
6.4	Descripción de los datos	98
6.4.1	Datos censales	98
6.4.2	Datos de características de electrodomésticos	100
6.4.3	Datos de consumo eléctrico	100
6.5	Implementación	104
6.6	Análisis experimental	104
6.6.1	Resultados obtenidos	104
6.6.2	Validación del enfoque propuesto	108
6.7	Conclusiones	110
7	Conclusiones y trabajo futuro	113
7.1	Conclusiones	113
7.2	Trabajo futuro	117
	Referencias bibliográficas	119
	Anexos	127
	Anexo 1 Descripción de los datos entregados por UTE	129
1.1	Consumo eléctrico total de los hogares	129
1.2	Consumo eléctrico de termotanques	129
1.3	Consumo eléctrico desagregado de electrodomésticos	131

Capítulo 1

Introducción

De acuerdo al informe de la [IEA \(2015\)](#), el consumo de energía eléctrica residencial en el mundo ha aumentado de forma ininterrumpida. Se espera que la demanda de energía eléctrica en el año 2050 duplique la registrada en el año 2010 ([Larcher y Tarascon, 2015](#)). En los últimos años se han llevado a cabo investigaciones con el objetivo de entender mejor las características del consumo eléctrico en pos de lograr un uso más eficiente en fábricas, residencias y edificios en general ([Ford, 2009](#); [Luján et al., 2019](#); [Orsi y Nesmachnow, 2017](#)). En este sentido, el monitoreo no intrusivo de cargas (*Non-Intrusive Load Monitoring*, NILM) es una técnica para obtener datos detallados del consumo energético de un lugar (e.g., hogar, fábrica, oficina) sin necesidad de instalar medidores en cada dispositivo. Posteriormente, se procesan los datos obtenidos para mejorar la planificación del consumo energético u ofrecer incentivos al consumidor para que mediante un cambio de hábitos genere ahorros en su consumo ([Konstantakopoulos et al., 2016](#)). También, para las compañías proveedoras del servicio, la información recolectada es un insumo importante para la planificación de la demanda y la generación de políticas comerciales más ajustadas a la realidad.

NILM permite desagregar el consumo eléctrico total para obtener el detalle de consumo de un subconjunto de dispositivos conectados a la red. Generalmente, esta técnica es aplicada en redes de energía eléctrica, aunque también puede ser aplicada a redes de gas, agua, u otra fuente de energía, incluso combinarlas ([Keramati et al., 2021](#); [Ellert et al., 2016](#)). Para el caso de las redes eléctricas, los algoritmos de desagregación se basan en el análisis de mediciones de corriente, tensión, potencia, consumo acumulado total agregado o desagregado por electrodomésticos, medidos por pinzas amperimétricas, me-

didores *plug-in*, o medidores inteligentes. NILM se diferencia de otras técnicas de recolección de datos por no requerir más *hardware* que un medidor, evitando la instalación de sensores u otro tipo de dispositivos en el interior de los hogares. Por este motivo, la técnica es clasificada como no-intrusiva. Particularmente, la técnica es aplicada a nivel residencial. Una vez instalado el medidor inteligente en un hogar, éste transmite los datos de consumo total para que posteriormente sean procesados y analizados. Los dispositivos conectados a la red eléctrica generan firmas distinguibles en las curvas de consumo total que pueden ser clasificadas por diferentes algoritmos. La clasificación de las firmas genera información desagregada del consumo energético. Esta tesis presenta un algoritmo de desagregación simple, que sirve como acercamiento al abordaje del problema de desagregación a través de técnicas de inteligencia computacional.

Una forma de abordar el problema de la desagregación de consumo eléctrico es utilizando algoritmos de inteligencia computacional. El insumo principal de éstos algoritmos son los datos históricos de consumo total agregado y los consumos etiquetados por electrodoméstico. Existen numerosos conjuntos de datos de consumo eléctrico general y también particularmente de consumo residencial. Es importante que los datos utilizados en los algoritmos de inteligencia computacional pertenezcan al contexto geográfico en el que se van a desagregar los datos. En este sentido, esta tesis presenta el primer conjunto de datos de consumo eléctrico residencial de Uruguay, ECD-UY.

La electricidad en Uruguay, a los suscriptores de distribución, es provista en su totalidad por la empresa pública estatal Administración Nacional de Usinas y Trasmisiones Eléctricas (UTE). Uruguay cuenta con una población de 3,4 millones de habitantes, de los cuales 1,3 residen en su capital, Montevideo. El país ha sido mundialmente reconocido por estar a la cabeza de los países más desarrollados en energías renovables (Watts, 2015). La red eléctrica cubre el 99,8 % de las áreas totales del país, por lo cual la electrificación es considerada universal (Ministerio de Industria, Energía y Minería, 2018). Según datos de julio de 2019, UTE provee de energía eléctrica a 1.498.164 clientes, de los cuales el 90,5 % corresponde a clientes residenciales (Ministerio de Industria, Energía y Minería, 2019). En promedio, la empresa provee 228 kWh de energía a los clientes residenciales, 246 kWh para los de Montevideo y 216 kWh para el resto del país. El 86,2 % de los hogares en Uruguay cuenta con termotanque eléctrico, dispositivo que representa en promedio un 27 % del consumo total de

los hogares, con picos de hasta 35 %. Este electrodoméstico puede ser manejado por los clientes y representa un potencial almacenamiento de energía en forma de calor. En este documento se presenta una caracterización del consumo eléctrico residencial en Uruguay, tomando como base el conjunto de datos locales ECD-UY.

Esta tesis corresponde al trabajo realizado como parte de la Maestría en Informática-PEDECIBA. En el transcurso de esta maestría se realizaron numerosas publicaciones que abordan las temáticas tratadas en este documento. A continuación se enumeran las publicaciones junto a una breve descripción.

- En los artículos ‘Household energy disaggregation based on pattern consumption similarities’ (Chavat et al., 2019) y ‘Nonintrusive energy disaggregation by detecting similarities in consumption patterns’ (Chavat et al., 2020) se presentó un enfoque para la desagregación de consumo de energía eléctrica residencial, basado en la detección de patrones similares en registros de consumo etiquetados. El algoritmo fue evaluado con datos sintéticamente generados a partir de consumo real registrado en el conjunto de datos UK-DALE, utilizando el framework `nilmtk`. Los resultados reportados mostraron que a pesar de las complejidades introducidas en los escenarios validados, el algoritmo logró resultados precisos. El artículo inicial fue presentado en la conferencia Ibero-Americana de Ciudades Inteligentes (ICSC-CITIES) del año 2019 y luego extendido para su publicación en la Revista Facultad de Ingeniería de la Universidad de Antioquía (REDIN) en el año 2020. En estos trabajos, mis principales aportes estuvieron en la generación de datos sintéticos, la creación de escenarios de validación, la conceptualización de variantes del algoritmo propuesto, la implementación del algoritmo en el framework `nilmtk`, el relevamiento de trabajos relacionados y la redacción y revisión científica de los artículos.
- En los artículos ‘Analysis of residential electricity consumption by areas in Uruguay’ (Chavat y Nesmachnow, 2020a) y ‘Data analysis approach for characterizing residential energy consumption based on statistics of household appliances ownership’ (Chavat y Nesmachnow, 2020b) se presentó un procedimiento para medir el consumo eléctrico potencial en hogares de Uruguay, considerando agrupaciones por área censal (sección, segmento y departamento). El procedimiento utiliza como datos

de entrada una combinación de variables socioeconómicas de la Encuesta Continua de Hogares (ECH) del Instituto Nacional de Estadísticas (INE), información de consumo de los electrodomésticos relevados por la encuesta y datos de consumo real residencial en Uruguay. Los resultados finales validaron la correlación entre el índice y el consumo real de las áreas censales. Los artículos fueron presentados en la conferencia Ibero-Americana de Ciudades Inteligentes (ICSC-CITIES) del año 2020 y en el Primer Taller Internacional sobre Tecnologías y Sistemas Avanzados de Información y Computación (AICTS) del año 2020. En estos trabajos, mis principales aportes radicarón en la presentación del abordaje planteado, la creación de los escenarios de validación, recolección y preprocesamiento de datos, la implementación del índice propuesto, el relevamiento de trabajos relacionados, la redacción y revisión científica de los artículos y la presentación en las conferencias.

- En el artículo ‘ECD-UY: Detailed household electricity consumption dataset of Uruguay’ ([Chavat et al., 2022a](#)) se introdujo el conjunto de datos ECD-UY, describiendo el proceso de recolección y comunicación de los datos, el pre-procesamiento y la validación técnica realizada, entre otros. ECD-UY se compone por registros de consumo eléctrico residencial en hogares de Uruguay y se divide en tres subconjuntos de consumo eléctrico y uno de información de clientes. Los subconjuntos de consumo abarcan el consumo eléctrico total de más de 110.000 hogares, el consumo eléctrico de 268 termotanques y el consumo eléctrico desagregado de 9 hogares. En esta instancia se trabajó conjuntamente con representantes de UTE. Los datos fueron agregados a un repositorio público para su libre descarga. El artículo fue presentado en la revista *Scientific Data - Nature* en el año 2021 y publicado en 2022. En este trabajo, mis principales aportes fueron el estudio de la modalidad de artículo de datos, la propuesta de presentación formal de los datos, el relevamiento de conjuntos y de artículos de datos existentes, el pre y pos procesamiento de datos, la validación técnica, la redacción y revisión científica del artículo y su presentación ante la empresa UTE.
- En los artículos ‘Demand Response Control in Electric Water Heaters: Evaluation of Impact on Thermal Comfort’ ([Porteiro et al., 2020](#)) y ‘A Thermal Discomfort Index for Demand Response Control in Residential Water Heaters’ [Porteiro et al. \(2021\)](#) se presentó la construcción de

un índice que mide la disconformidad de los clientes al aplicar manejo activo de la demanda, consistente en la interrupción del termotanque. Basado en datos de consumo eléctrico real (de ECD-UY) y utilizando métodos de aprendizaje automático y modelos lineales, el índice integra los diferentes patrones de uso del termotanque y estima la temperatura del agua en cada momento. El índice es evaluado en un escenario real y dos simulados. Los resultados obtenidos demostraron la viabilidad de diseñar una estrategia de gestión que decida qué dispositivos pueden ser interrumpidos, manteniendo el índice de disconformidad en valores bajos. En estos trabajos, mis principales aportes radicaron en la recolección y el posprocesamiento de los datos y la redacción y revisión científica del artículo.

Durante el desarrollo de este trabajo, colaboré activamente en el proyecto UTE-UdelaR *Inteligencia computacional para caracterizar los usos de energía eléctrica en clientes residenciales* financiado por UTE y la Comisión Sectorial de Investigación Científica (CSIC).

Lo que resta del documento se estructura de la siguiente manera. El Capítulo 2 introduce las temáticas abordadas en este trabajo. El Capítulo 3 presenta los trabajos relacionados más relevantes, seleccionados de la literatura académica disponible. En el Capítulo 4 se brindan detalles de un algoritmo simple para abordar el problema de desagregación de consumo eléctrico y su validación utilizando instancias de datos sintéticos. En el Capítulo 5 se presenta el conjunto de datos ECD-UY, basado en datos de consumo eléctrico residencial proporcionados por la empresa UTE. Se describe el proceso de recolección, comunicación y pre-procesamiento de los datos, el formato de sus subconjuntos y la validación técnica realizada para los mismos. En el Capítulo 6 se presenta un índice de consumo potencial basado en datos socioeconómicos y de características de consumo de los electrodomésticos del mercado Uruguayo. El índice fue procesado para departamentos de Uruguay, secciones y segmentos censales de Montevideo. Finalmente, el Capítulo 7 presenta las conclusiones generales de esta tesis y un conjunto de puntos relevantes para continuar trabajando a futuro.

Capítulo 2

Análisis del consumo eléctrico residencial y monitoreo no intrusivo de las cargas

Este capítulo presenta los principales conceptos abordados en esta tesis. La Sección 2.1 describe el análisis de datos en general y particularmente el de consumo eléctrico residencial. La Sección 2.2 describe el proceso de monitoreo no intrusivo de cargas y presenta una formulación matemática del problema de desagregación de consumo eléctrico.

2.1. Análisis de datos de consumo eléctrico

El análisis de datos es el proceso de inspeccionar, limpiar, transformar y modelar los datos con el fin de tomar decisiones, descubrir información útil o entender diferentes realidades a partir de los datos.

Tukey (1962) definió el análisis de datos como el conjunto de procedimientos para el análisis, las técnicas para interpretar los resultados, la planificación de su recolección, y todas las técnicas de estadística matemática que aplican al análisis de los datos. El autor del trabajo es considerado uno de los precursores del análisis de datos (Hoaglin, 2003). Cuando los datos utilizados se presentan en volúmenes tales que no pueden ser cargados en una única máquina o implican que los métodos tradicionales de análisis de datos (desarrollados para procesos centralizados) no puedan ser aplicados directamente sobre los datos, el área de estudio es *big data* (Tsai et al., 2015). Para explicar lo gran-

de (*big*) de los datos, se utilizan las 3Vs: volumen (gran tamaño), velocidad (rápida generación) y variabilidad (múltiples tipos o fuentes). Este concepto es de gran importancia en áreas como el marketing, el control y prevención de enfermedades y ciudades inteligentes, entre otros.

Los procesos pueden ser enmarcados bajo el concepto de descubrimiento del conocimiento en las bases de datos (KDD), que se compone de seis etapas: selección, preprocesamiento, transformación, análisis de datos (también conocida como minería de datos), interpretación y evaluación. También, el proceso estándar de la industria para la minería de datos (CRISP-DM) comprende seis etapas, identificadas de la siguiente forma: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y desarrollo. [Runkler \(2020\)](#) planteó que en proyectos de análisis de datos tradicionales, para simplicidad, se pueden agrupar las diferentes etapas de KDD y CRISP-DM en cuatro grandes grupos: preparación, preprocesamiento, análisis y postprocesamiento. Por otro lado, [Tsai et al. \(2015\)](#) clasificó las etapas en tres grandes grupos: entrada de datos, análisis de datos y salida. La Figura 2.1 presenta un esquema de los grandes grupos principales, que reúnen las diferentes etapas identificadas por los investigadores.

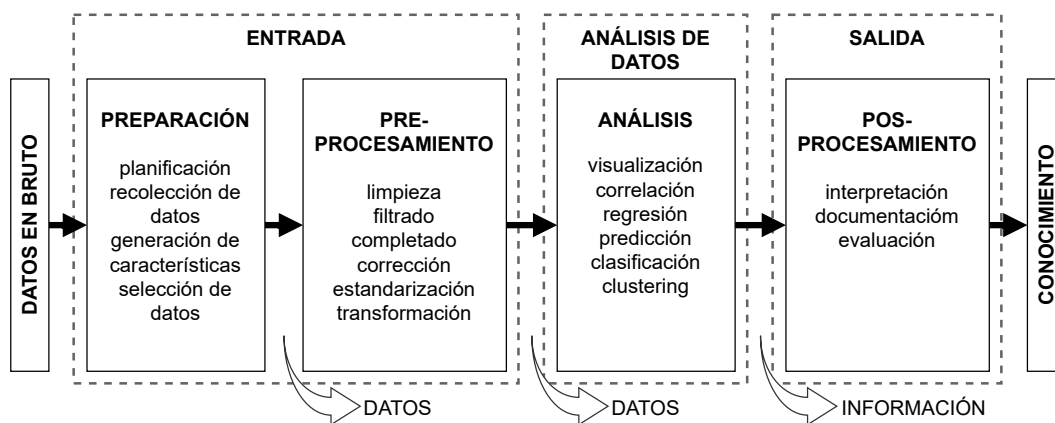


Figura 2.1: Etapas del análisis de datos, basada en [Runkler \(2020\)](#) y [Tsai et al. \(2015\)](#).

En la etapa de preparación se recolectan y seleccionan los datos que son relevantes para el análisis, descartándose el resto. Luego, en la etapa de preprocesamiento se detectan, limpian y filtran los datos innecesarios, inconsistentes e incompletos, para luego pasar a la etapa de transformación, que se encarga de llevar el formato de los datos a uno apto para las siguientes etapas. En la etapa de preprocesamiento también se pueden ejecutar procesos o reducir la

complejidad o el tamaño de los datos. La preparación y el preprocesamiento son etapas previas a la de análisis de datos. La etapa de análisis de datos se encarga de descubrir los patrones, reglas e información oculta en los datos. Los métodos aplicados en el análisis no se limitan a los específicos del tratamiento de datos, sino que abarcan también la estadística y la inteligencia computacional, entre otros. Una vez analizados los datos, la etapa de pos-procesamiento evalúa, interpreta y mide (e.g., costo computacional, tiempo de procesamiento) los resultados. Típicamente, el proceso concluye en reportes a nivel de usuario que pueden complementarse con resúmenes de indicadores o componentes visuales para facilitar la comprensión. El análisis de datos tradicional no está pensado para el procesamiento a gran escala, por lo que la adaptación de las etapas tradicionales es objeto de estudio bajo el concepto de big data.

El análisis de datos de consumo eléctrico es un área principal de investigación dentro de la planificación y gestión de sistemas de energía, motivado en parte por la extensa instalación de medidores inteligentes, que permiten contar con grandes volúmenes de datos. La instalación de los medidores responde al interés de los gobiernos y compañías proveedoras de energía en promocionar el consumo eficiente de consumo energético ([Benzi et al., 2011](#)).

Específicamente, el consumo eléctrico residencial refiere a la electricidad consumida en edificaciones destinadas a viviendas residenciales. Estos datos tienen, en general, la forma de series temporales en las que cada registro cuenta con una serie de indicadores de consumo de electricidad. En algunos casos, los indicadores refieren al consumo de un área de edificaciones (e.g., un país, un departamento, una sección censal, etc), mientras que en otros casos refieren a la edificación en concreto (e.g., edificio de apartamentos, complejo de viviendas u hogares individuales). De igual forma, los registros pueden indicar el consumo acumulado de un período de tiempo (i.e., consumo diario, semanal, etc) o directamente el consumo del instante en el que fue registrado (e.g., valor de corriente, intensidad o potencia).

Las principales limitaciones para un análisis preciso del consumo eléctrico residencial radican en el volumen, la fugacidad y el orden aleatorio de los datos generados por los medidores, y su posterior transmisión desde los múltiples dispositivos a una ubicación central (e.g., centro de datos). Por la naturaleza de los datos, es probable que existan patrones de consumo que se repitan. Es conveniente que las técnicas de análisis sean incrementales y capaces de acumular el conocimiento de patrones previamente detectados ([De Silva et al., 2011](#)).

En las redes eléctricas modernas, el análisis de consumo de energía basado en datos reales permite una planificación precisa del consumo. Además, el análisis posibilita la implementación de tarifas personalizadas o políticas energéticas que disminuyan los picos de consumo en determinados días u horarios e incremente la eficiencia de las cadenas de suministro de electricidad. Los análisis también pueden seccionar el consumo, por ejemplo, según tipo de edificios, área geográfica, o características socioeconómicas del cliente (Ali et al., 2016).

2.2. Monitoreo no intrusivo de cargas

Esta sección describe el monitoreo no intrusivo de cargas, particularmente en hogares residenciales, y brinda una formulación matemática del problema de desagregación.

2.2.1. Monitoreo no intrusivo

El monitoreo de cargas trata de supervisar, mediante la obtención de datos, el consumo de un determinado lugar o conjunto de dispositivos. Existen dos formas de realizar el monitoreo, una utilizando un fuerte componente de *hardware*, instalando medidores en cada lugar o dispositivo; y otra utilizando un *hardware* básico y un complemento de *software* complejo. Un posible escenario de la segunda forma de monitoreo (*hardware* básico y *software* complejo) es contar con un medidor general de cargas y un *software* que clasifique el estado de los dispositivos, usando como insumo únicamente la señal de consumo total. El único *hardware* necesario para el escenario planteado es el medidor general del lugar, por ejemplo el medidor de consumo general de un hogar. El hecho de utilizar los datos de consumo total evita la instalación de medidores adentro de los establecimientos, por este motivo, esta forma de monitoreo se define como un proceso no intrusivo. Además, al utilizar un único dispositivo para la medición total, los costos de *hardware* son comparativamente menores a los de la forma intrusiva, que requiere la instalación de múltiples dispositivos. El proceso no intrusivo fue presentado por primera vez, y de forma independiente, en Hart (1992) y Sultanem (1991), a comienzos de la década de 1990.

2.2.2. Desagregación de consumo energético

La señal de consumo total agregado contiene la firma de consumo de los dispositivos encendidos en cada momento. Las firmas se definen como la medición de ciertos parámetros sobre el consumo total, que entregan información sobre el consumo energético de dispositivos individuales conectados a la red, en un período de tiempo dado. Por ejemplo, para un método de desagregación de consumo eléctrico, la firma de consumo puede estar integrada por los valores observables de voltaje en un período de tiempo dado.

Dada una señal de consumo eléctrico total agregado, la desagregación consiste en obtener las múltiples señales de consumo de los dispositivos encendidos conectados, responsables del consumo eléctrico en cada momento. El problema de desagregación puede ser catalogado como un problema de clasificación en los casos en los que los estados de consumo de los dispositivos a ser desagregados son clasificados. Las técnicas más populares tienen en cuenta las diferentes firmas que dejan los dispositivos a desagregar en la señal de consumo total agregado, y a partir de éstas se clasifican sus estado. Los hogares presentan múltiples configuraciones de dispositivos (i.e., diferentes electrodomésticos conectados a la red eléctrica del hogar) con variedad de firmas de consumo según el modelo o tipo de uso que se les de. Por lo tanto, en la desagregación es complejo encontrar el conjunto de características que ayudan a resolver el problema con un alto nivel de precisión (Figueiredo et al., 2012).

2.2.3. Formulación matemática del problema de desagregación de cargas eléctricas

El problema de desagregación de cargas eléctricas puede ser expresado de forma matemática. La definición matemática del problema considera los siguientes elementos:

- Un rango de tiempo T , discretizado en intervalos t .
- Un conjunto de electrodomésticos $E = \{e_1, \dots, e_m\}$.
- Una función $C : E \times T \rightarrow \mathbb{R}$ que dado un electrodoméstico e_i y un intervalo t retorna un valor real x_t^i correspondiente al consumo del electrodoméstico en el intervalo, i.e., $x_t^i = C(e_i, t)$.

- El consumo agregado del hogar en un instante dado t , x_t . Se cumple la igualdad $\sum_{i=1}^{i=m} x_t^i = x_t$, donde m es la cantidad de electrodomésticos.
- Un vector $\vec{y}^t \in \{0, 1\}^m$ cuyos elementos y_i^t indican el estado del electrodoméstico i en el intervalo de tiempo t . El valor 1 indica que el electrodoméstico se encuentra encendido mientras que el valor 0 indica que se encuentra apagado.

El escenario más sencillo del problema se da cuando el consumo de los electrodomésticos es constante. De esta forma, la función C para un cierto electrodoméstico i pasa a retornar únicamente dos valores: cero (cuando está apagado) o el valor constante de consumo (cuando está encendido). Sea c_i la constante de consumo para el electrodoméstico i , entonces:

$$x_t = f(\vec{y}^t) = \sum_{i=1}^{i=m} C(e_i, t) \times y_i = \sum_{i=1}^{i=m} c_i \times y_i^t = \vec{c} \times \vec{y}^t \quad (2.1)$$

Para los casos en que la función f es inyectiva (i.e., para cada x_t existe un único \vec{y}^t posible), el problema es trivial. Para los demás casos, más ajustados a los escenarios reales, la serie temporal $\{x_t\}_{t \in T}$ debe ser analizada para aprender y deducir el estado de los electrodomésticos de forma correcta.

2.2.4. Descripción de un escenario de desagregación con características de uno real

Supongamos que existe un escenario con los siguientes electrodomésticos y sus correspondientes consumos: heladera (250 W), lavarropa (1500 W), lavavajilla (2250 W), jarra eléctrica (2000 W) y *home theatre* (80 W). Con esta configuración de electrodomésticos–consumos, el consumo total agregado no es una función inyectiva. Por ejemplo, podemos encontrar dos estados de electrodomésticos que generan el mismo consumo total agregado: si la heladera y la jarra eléctrica están encendidas (2250 W en total), y si solo está encendido el lavavajilla (2250 W en total). Suponiendo que el vector \vec{y}^t tiene como elementos a cada uno de los electrodomésticos del escenario planteado, y respeta el orden en el que fueron enunciados, entonces la Ecuación 2.2 muestra la ambigüedad mencionada.

$$f((1, 0, 0, 1, 0)) = f((0, 0, 1, 0, 0)) = 2250 W \quad (2.2)$$

Para resolver estas ambigüedades, se estudian las formas de las firmas de consumo que dejan los electrodomésticos en el consumo total y los patrones de consumo, entre otras características. Además, otros patrones más complejos pueden ser detectados al tener en cuenta el vecindario temporal del instante de consumo a clasificar. Por ejemplo, se puede buscar la función que genera la series de vectores \vec{y}^t para un rango de tiempo, tratando de maximizar el total de los estados de electrodomésticos bien clasificados en todo el rango, como lo representa la Ecuación 2.3.

$$\max \sum_{t \in T} \prod_{i=1}^m 1_{\{\hat{y}_t^i = y_t^i\}} \quad (2.3)$$

Capítulo 3

Trabajos relacionados

Este capítulo expone la revisión literaria de los trabajos relacionados a la temática de la investigación presentada y de los conjuntos de datos utilizados en la literatura científica. El capítulo se subdivide en trabajos relacionados al monitoreo no intrusivo de cargas, conjuntos de datos de consumo de energía y análisis de datos de consumo eléctrico.

3.1. Monitoreo no intrusivo de cargas

El proceso de monitoreo no intrusivo de cargas fue propuesto de forma independiente en [Hart \(1992\)](#) y [Sultanem \(1991\)](#) a principios de la década de 1990. Ambos trabajos representan el inicio de la línea de investigación en esta temática, y por lo tanto se presenta una revisión extensa de estos.

[Hart \(1992\)](#) planteó por primera vez el monitoreo no intrusivo de consumo eléctrico de electrodomésticos del hogar, asignándole el nombre de *Nonintrusive Appliance Load Monitoring* (NALM), que tiempo después pasó a llamarse directamente *Non-Intrusive Load Monitoring* (NILM). Se observó que los enfoques existentes se basaban fuertemente en el *hardware* instalado de forma intrusiva en los conectores de los dispositivos a monitorear, y relegaban la tarea del *software* únicamente a la recolección de los datos. El autor presentó un enfoque basado en *hardware* simple y *software* de análisis complejo, permitiendo eliminar la intrusión permanente en los hogares (incorporación del término *no intrusivo* al proceso). El trabajo presenta dos formas de resolver el problema: i) *Manual-Setup NALM* (MS-NALM), con una única etapa intrusiva para observar las firmas de cada electrodoméstico y ii) *Automatic-Setup*

NALM (AS-*NALM*), basada en el conocimiento previo de las firmas de los electrodomésticos objetivo. Entre las posibles aplicaciones planteadas se destacan: auditorías de consumo energético (instalación temporal del sistema y análisis de características de consumo para elaborar sugerencias que reduzcan el costo), detección de fallas de los electrodomésticos (monitoreo de consumo o ciclos anómalos), detección de presencia humana en los lugares, con fines de seguridad (registro de consumo cuando no debería existir), gestión de la demanda (control de electrodomésticos con cargas aplazables moviendo su uso fuera de horarios pico), y monitoreo de dispositivos inaccesibles físicamente (e.g., ubicaciones submarinas o extraterrestres), entre otras.

Hart también propuso una clasificación de los electrodomésticos a partir del tipo de activación: encendido/apagado, que presentan consumo constante desde su encendido a su apagado (único estado), de máquinas de estado finito (*Finite State Machine*, FSM), que presentan estados y transiciones identificados por la potencia que requieren, y de estados continuamente variables, que varían su potencia de forma continua (e.g., luces con intensidad regulable, herramientas con velocidades de operación variable, etc). El trabajo apuntó únicamente a los electrodomésticos de encendido/apagado y FSM, afirmando que los continuamente variables presentan consumos insignificantes en los hogares (no así en las industrias).

Las firmas de consumo son el insumo principal para la identificación del estado de los electrodomésticos y se reflejan en el consumo total agregado. Hart propuso una clasificación de firmas por niveles. En el primer nivel se encuentran las firmas intrusivas y las no intrusivas. Las firmas intrusivas se miden mediante la observación pasiva de la carga, mientras que las no intrusivas se subdividen en firmas de estado estacionario (información de cambio de estado presente de forma continua) y transitorias (información de cambio de estado durante la transición de estados). A nivel experimental, Hart utilizó firmas estacionarias por tres razones: i) los indicadores continuos son más fácil de detectar que los momentáneos, ii) la sumatoria de los cambios de potencia entre ciclos de cambios de estado son cero (*Zero Loop-Sum Constraint*, ZLSC) y iii) cumplen aditividad cuando ocurren a la vez. Dentro de las firmas de estado estacionario, se encuentran las de frecuencia fundamental (miden potencia, corriente y admitancia de la carga total), las de frecuencias armónicas (medidas por sensores en el rango de frecuencia de interés) y las de consumo de corriente continua. Los prototipos implementados por Hart se basan en los cambios sus-

tanciales observados en el valor de potencia normalizada (admitancia ajustada por un factor de escala constante). La Figura 3.1 presenta la taxonomía de los tipos de firmas estudiados en el artículo. Las marcas en celeste indican las firmas utilizadas en los prototipos implementados.

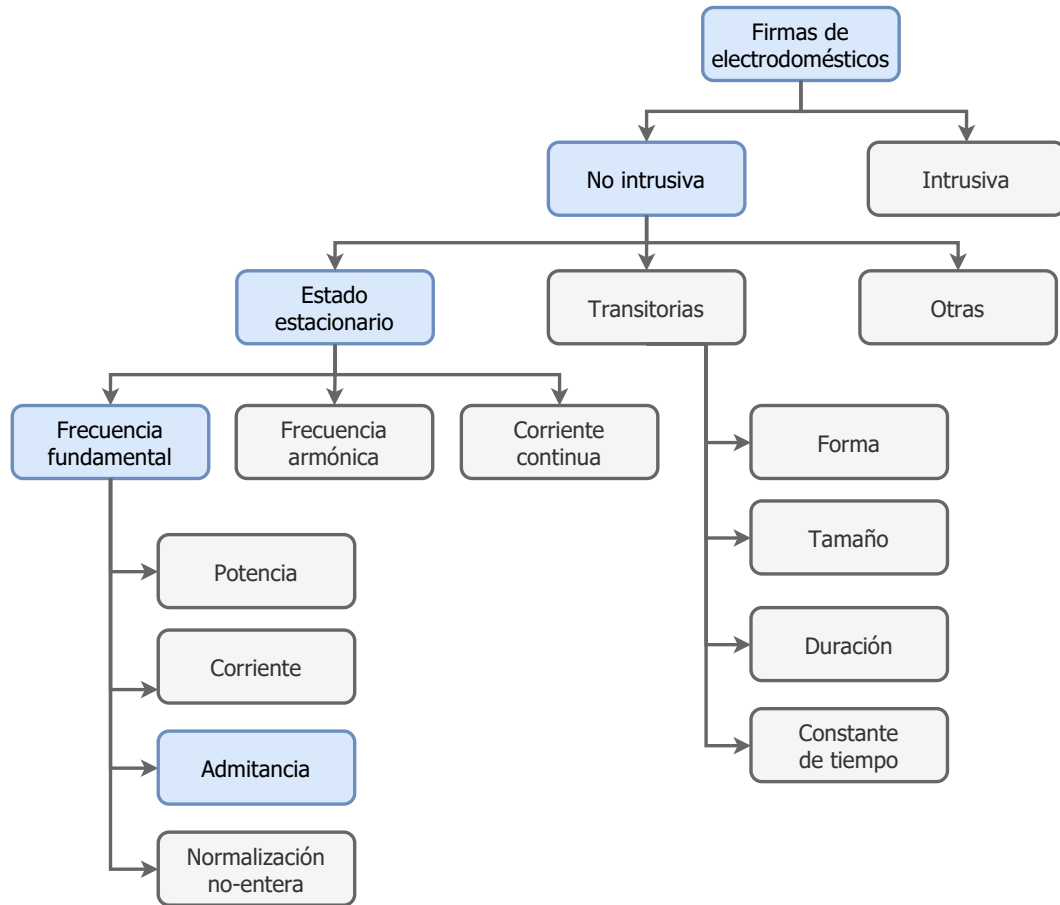


Figura 3.1: Taxonomía de firmas de electrodomésticos propuesta por Hart (1992). Los rectángulos en celeste corresponden a las utilizadas por Hart.

Teniendo en cuenta que los electrodomésticos se conectan en paralelo a la red eléctrica y que la potencia consumida es aditiva, el autor definió un modelo para el análisis de la señal total del circuito. El modelo se expresa en la Ecuación 3.1, donde $a_i(t)$ representa el estado encendido/apagado de un electrodoméstico en el tiempo t .

$$a_i(t) = \begin{cases} 1, & \text{si el dispositivo } i \text{ está encendido en el tiempo } t \\ 0, & \text{si el dispositivo } i \text{ está apagado en el tiempo } t \end{cases} \quad (3.1)$$

Luego, las cargas con p fases para un electrodoméstico i se modelan como vectores $P_i \in \mathbb{R}^p$, donde la suma de sus componentes representa la carga total consumida cuando está encendido. La Ecuación 3.2 define el consumo total como $P(t) \in \mathbb{R}^p$, donde t representa el instante de consumo y $e(t)$ representa el ruido/error registrado.

$$P(t) = \sum_{i=1}^n P_i + e(t) \quad (3.2)$$

El modelo involucra la resolución de una optimización combinatoria en el que se busca el vector $a(t)$ que minimice el error, teniendo el conocimiento de P_i y $P(t)$, como muestra la Ecuación 3.3.

$$\hat{a}(t) = \arg_a \min |P(t) - \sum_{i=1}^n a_i P_i| \quad (3.3)$$

El problema de optimización combinatoria resultante es NP-completo, y por lo tanto computacionalmente intratable para gran cantidad de electrodomésticos (valores de n grandes). Una opción es utilizar algoritmos heurísticos que permitan calcular soluciones de calidad aceptable para la mayoría de los casos, pero la aplicabilidad de estos algoritmos es limitada en la práctica ya que no se conoce la totalidad de los vectores P_i , la cantidad de electrodomésticos no es fija, y ante dispositivos desconocidos se buscarían una combinación usando los ya conocidos. Incluso, aún conociendo todos los P_i , una pequeña variación en la medición de $P(t)$ podría provocar grandes cambios en $a(t)$, prediciendo erróneamente eventos de encendido y apagado simultáneos. Para evitar la predicción de encendidos y apagados simultáneos, se utilizó el principio *switch continuity*, que establece que en pequeños intervalos de tiempo pocos electrodomésticos presentan cambios en su estado. Además, se restringe a que ningún electrodoméstico tenga consumo negativo para eliminar la ambigüedad producida entre el encendido de uno y el apagado de un generador de energía (se asume que no existen generadores eléctricos en la red).

La desagregación puede ser abordada como un problema de comunicación donde los electrodomésticos son tomados como los transmisores, los cables como el canal y las firmas como los códigos transmitidos. La solución es diseñar un receptor capaz de decodificar los mensajes, o sea, las emisiones en forma de cambios de estados de los electrodomésticos. La ventaja de este enfoque

radica en que los mensajes son transmitidos a una tasa baja a través de un cable de cobre relativamente corto y grueso, mientras que las desventajas son que no se cuenta con una tabla de códigos transmitidos, existen transmisiones simultáneas y la naturaleza de los errores introducidos al canal dificulta su entendimiento. Hart propuso aplicar una generalización del algoritmo Viterbi (VA), que permite trabajar con diferentes tipos de corrupciones en los datos transmitidos por el canal. Dada una FSM que representa la fuente de mensajes (los electrodomésticos), la generalización de VA corrige de forma óptima las inserciones, los borrados, las fusiones y otros tipos de errores que podrían ocurrir en el canal. La solución planteada consiste en la ejecución de los siguientes pasos:

1. Medición de características de consumo.
2. Agrupamiento de los cambios detectados desde el paso previo.
3. Construcción de los modelos de electrodomésticos.
4. Seguimiento del comportamiento de los electrodomésticos utilizando el enfoque de decodificador.
5. Cálculo de los indicadores de interés.
6. Etiquetado de los electrodomésticos basado en los datos recolectados.

En su trabajo, Hart implementó tres prototipos basados en modelos encendido/apagado: MS-NALM (un hogar), AS-NALM (tres hogares) y AS-NALM (diez hogares). Se logró detectar satisfactoriamente los electrodomésticos del tipo encendido/apagado, pero no así los del tipo FSM. Para el tercer prototipo se contó con los datos reales para la comparación de resultados, obteniendo una diferencia de predicción menor al 10%. Finalmente, el autor concluyó que NALM es una solución que no necesita de *hardware* complejo y su instalación es económica y no intrusiva. AS-NALM logró adaptarse a los nuevos electrodomésticos, sin limitantes en la cantidad de electrodomésticos. Por otro lado, la solución planteada no permitió identificar electrodomésticos de consumo bajo, continuamente variable ni constantemente encendidos, y es propensa a errores ante electrodomésticos desconocidos.

Motivado por el creciente consumo eléctrico del área residencial en la década de 1990 y la necesidad de contar con información precisa sobre el reparto del consumo por electrodomésticos, [Sultanem \(1991\)](#) desarrolló una herramienta capaz de monitorear la carga de los hogares e identificar los equipos encendidos. El autor aseguró poder reconocer la procedencia de las señales eléctricas a partir del análisis de los parámetros de consumo que caracterizan al dispositivo. La dificultad surge cuando diferentes señales se entremezclan en una única señal compartida de consumo total agregado, en la que los dispositivos se encienden/apagan mientras que otros permanecen encendidos. La herramienta propuesta aislaba los parámetros característicos de consumo de los dispositivos de la señal de consumo total agregado, en un proceso de monitoreo en tres etapas: i) aprendizaje de características de la carga, ii) adquisición de eventos, e iii) identificación de cargas. Se midieron tres parámetros de la carga, la corriente, el voltaje y la potencia activa, y se dedujeron dos, la potencia aparente y la reactiva. Además, se tuvieron en cuenta la duración del consumo, la forma de la señal de corriente en la transición a encendido y los armónicos de corriente producidos por los dispositivos. Las características medidas fueron utilizadas para clasificar los electrodomésticos en seis grupos:

1. *Resistivos*. Dispositivos con cero uso de potencia reactiva, gran uso de potencia activa, tiempo casi nulo de transición a encendido y ausencia de armónicos en la señal de corriente. Ejemplos de estos electrodomésticos son los calefactores o las bombillas de luz.
2. *Compresor/bomba*. Dispositivos con motor eléctrico que accionan un compresor o una bomba. Son identificados por un uso sustantivo de potencia reactiva, transición a encendido larga y de forma característica, y señal de corriente armónica impar. Ejemplos de estos electrodomésticos son las heladeras o los lavavajillas.
3. *Motor eléctrico*. Complementa el grupo anterior, ya que son dispositivos con motor eléctrico pero no destinados a un compresor/bomba. Se caracterizan por presentar menores transiciones de encendido. Ejemplos de estos electrodomésticos son los ventiladores.
4. *Bajo consumo*. Dispositivos que presentan consumo bajo de electricidad. Se caracterizan por transiciones de encendido con duración muy corta, y un espectro de armónicos variado. Ejemplos de estos electrodomésticos son la televisión y los equipos de audio.

5. *Control de potencia.* Este grupo reúne a los dispositivos que operan a potencia variable. Su identificación es particularmente difícil, por lo que el autor no los utilizó para su trabajo. Ejemplos de estos electrodomésticos son las lámparas halógenas y las aspiradoras.
6. *Iluminación fluorescente.* Dispositivos de iluminación caracterizada por transiciones de encendido largas y en dos pasos, generación del tercer armónico de corriente alto y cambios significantes en la fase corriente-voltaje. Ejemplos de estos dispositivos son los tubos de luz.

Los eventos de encendido/apagado ocurren a la vez que otros dispositivos están en uso. Por esta razón, la variación de corriente fue utilizada para la detección de eventos en la señal. Para detectar las variaciones, se compararon las mediciones de un instante con las mediciones de instantes inmediatamente anteriores. Si se detectaban cambios, se calculaban los parámetros de identificación. Una vez detectado un evento y obtenidos los valores, se utilizaba un algoritmo de identificación para asignar el evento ocurrido a uno de los electrodomésticos. El algoritmo de identificación utilizado buscaba la serie de eventos más coincidentes, dentro de un margen de error, registradas para electrodomésticos procesados en la etapa de aprendizaje de características. El prototipo desarrollado demostró la posibilidad de monitorear cargas de un hogar de forma no intrusiva y contar con información de reparto de consumo entre los tipos de electrodomésticos. Se concluye que la adquisición de los valores de parámetros fue satisfactorio, permitiendo detectar los eventos de encendido/apagado de los dispositivos, mientras que el algoritmo de identificación fue capaz de identificar el electrodoméstico de forma satisfactoria en más del 95 % de los casos.

En trabajos más recientes, la resolución de la desagregación en procesos de monitoreo no intrusivo de cargas ha sido tratada como un problema de aprendizaje automático, al que pueden aplicarse las estrategias de aprendizaje supervisado y no supervisado. Los métodos de aprendizaje supervisado están basados en conjuntos de datos de consumo etiquetado para cada dispositivo del circuito y de la señal agregada. El objetivo del entrenamiento en éstos métodos es generar modelos que aprendan a desagregar la señal de los dispositivos a partir de la señal agregada. Las técnicas frecuentemente aplicadas para este enfoque son los métodos de aprendizaje bayesiano y las redes neuronales. Por otro lado, en el enfoque no supervisado se busca aprender las firmas de posibles dispositivos a partir de la señal agregada, sin conocer a priori qué dispositivos hay dentro del circuito.

Como en todos los problemas de aprendizaje automático, contar con datos de mediciones reales resulta clave para poder aplicar las diferentes técnicas algorítmicas. [Bonfigli et al. \(2015\)](#) presentaron un relevamiento de los conjuntos de datos disponibles y las principales técnicas utilizadas para su desagregación. Un relevamiento detallado de los conjuntos de datos presentes en la literatura científica es presentado en detalle en la Sección 3.2. Entre las técnicas de aprendizaje no supervisado, los autores destacaron las basadas en Modelos Ocultos de Markov (*Hidden Markov Models*, HMM). Las HMM definen un número de estados ocultos en el que se mueve el modelo y las transiciones que representan condiciones de funcionamiento de los dispositivo.

[Kelly y Knottenbelt \(2015a\)](#) desarrollaron una investigación partiendo de la misma premisa que [Sultanem \(1991\)](#), donde se observó que las personas aprenden a detectar la presencia de electrodomésticos en datos agregados de consumo, simplemente reconociendo un conjunto de características presentes en sus firmas de consumo. Bajo esta idea sería posible definir manualmente un conjunto de características, pero demandaría mucho tiempo y se obtendrían características poco robustas a los ruidos y las variaciones de modelos de electrodomésticos. Por esta razón, los autores plantearon aplicar técnicas de aprendizaje automático para la clasificación de imágenes y aprender de forma automática las características de las firmas de los electrodomésticos. Se evaluaron tres redes neuronales profundas, y se utilizaron dos métricas diferentes para su comparación. Para evaluar la robustez de los algoritmos, se midieron los resultados al procesar modelos de electrodomésticos ausentes en la etapa de entrenamiento.

Las redes neuronales propuestas por los autores contaban con entre uno y 150 millones de parámetros entrenables, por lo que fue necesario disponer de grandes cantidades de datos de entrenamiento. Los autores utilizaron el conjunto de datos UK-DALE, que contiene registros de potencia total agregada y desagregada por electrodomésticos (refrigerador, lavarropas, lavavajillas, jarra eléctrica y microondas) de cinco casas, con intervalos de muestreo de seis segundos. El enfoque consistió en entrenar una red neuronal por cada electrodoméstico, tomando como entrada una secuencia de consumo total agregado y retornando como resultado la predicción de la potencia demandada por el electrodoméstico asociado. Las redes fueron entrenadas utilizando 50 % de datos reales y 50 % de datos sintéticos, generados con las firmas de los electrodomésticos del conjunto de datos, y utilizando la herramienta `nilmtk`.

El trabajo evaluó tres arquitecturas de redes neuronales. La primer arquitectura fue una Red Neuronal Recurrente (*Recurrent Neural Network*, RNN) del tipo Gran Memoria de Corto Plazo (*Long Short-Term Memory*, LSTM). La segunda arquitectura se enfocó en “limpiar” la señal agregada de consumo para obtener la correspondiente al electrodoméstico objetivo, utilizando una red Auto-Codificadora (*AutoEncoder*, AE), particularmente una auto-codificadora para la eliminación de ruido (*denoising AutoEncoder*, dAE). Las redes AE, compuestas por una parte codificadora y otra decodificadora, tienen como objetivo reconstruir una entrada dada. La tercer arquitectura se enfocó en detectar el inicio y fin del uso de los electrodomésticos y la potencia promedio demandada en ese tiempo. Gráficamente, el proceso implicó detectar rectángulos que envuelven a la señal del electrodoméstico, donde los lados del rectángulo representan el comienzo y fin respectivamente, y la altura la potencia promedio demandada. La evaluación experimental comparó los resultados de las redes implementadas con dos algoritmos base en la literatura, implementados en `nilmtk`: Optimización Combinatoria (CO) y Modelo Factorial Oculto de Markov (FHMM). Se concluyó que las redes dAE y de detección de rectángulos presentaron un rendimiento satisfactorios, especialmente en los datos de evaluación, indicando que el enfoque es acertado para la resolución del problema de desagregación. Las redes LSTM presentaron un mal rendimiento en la identificación de electrodomésticos multiestado, pudiéndose deber a que los eventos relevantes en la señal de potencia demandada quedan fuera del largo de las secuencias de entrada a la red. La comparaciones de resultados entre las redes no fue del todo justa, ya que tienen un número de parámetros entrenables muy variado (un millón en la LSTM y 150 millones en la dAE). En las conclusiones se remarcó que para lograr un buen desempeño de las redes es fundamental contar con más datos de consumo de diferentes hogares y modelos de electrodomésticos para lograr que los algoritmos generalicen adecuadamente.

En el relevamiento literario se identificó que varias investigaciones utilizaron la herramienta `nilmtk` (Batra et al., 2014) que ofrece un *framework* para el análisis de datos NILM. Una de las funcionalidades permite utilizar múltiples conjuntos de datos al convertirlos a un modelo de datos estándar. Además, `nilmtk` proporciona algoritmos para el preprocesamiento de los datos (e.g., *downsample* y normalización de voltage), estadísticas para la descripción de los datos, algoritmos de desagregación (e.g., CO y FHMM) y métricas para aplicar en la evaluación de rendimiento de los modelos.

Finalmente, algunos trabajos han abordado el uso de NILM y las técnicas de desagregación sobre otros recursos del hogar aparte de la electricidad. [Ellert et al. \(2016\)](#) presentaron un método para la desagregación del consumo de agua de electrodomésticos, basándose en la posible correlación entre los consumos de agua y electricidad. El método utilizaba el resultado de NILM aplicado al consumo eléctrico y la medición del contador principal de agua para aplicar luego una modificación del algoritmo Viterbi (*Capped Viterbi Algorithm*). La evaluación se realizó utilizando el conjunto de datos AMPds ([Makonin et al., 2016](#)), detallado en la siguiente sección. Los resultados mostraron una alta precisión y un error medio cuadrático por debajo de $0.02L^2/min^2$.

3.2. Conjuntos de datos de consumo de energía

Numerosos conjuntos de datos de consumo energético, en específico consumo eléctrico, han sido publicados en los últimos 10 años por la comunidad científica y la industria. Estos conjuntos se conforman principalmente por series temporales de mediciones de consumo eléctrico en diferentes edificios o para diferentes dispositivos, dentro de una zona geográfica determinada (e.g., país, departamento, barrio, etc). Los conjuntos de datos presentan múltiples características como el período de muestreo, el largo de tiempo monitoreado, la cantidad de edificios que se monitorean, el país de origen, el tipo de desagregación (cuando ésta existe) y el acceso a los datos, entre otros.

En cuanto al período de muestreo, los conjuntos se pueden clasificar en conjuntos de períodos bajos a muy bajos, medios o largos a muy largos. Los conjuntos con períodos de muestreo bajos a muy bajos cuentan con una frecuencia de muestreo de 50 kHz a 5MHz. Ejemplos de estos conjuntos son BLOND ([Kriechebaumer y Jacobsen, 2018](#)), COOLL ([Picon et al., 2016](#)) o HFED ([Gulati et al., 2014](#)). Los conjuntos con períodos de muestreo medios cuentan con frecuencias de muestreo de 0.5 Hz a 16 kHz. Ejemplos de estos conjuntos son UK-DALE ([Kelly y Knottenbelt, 2015b](#)), REDD ([Kolter y Johnson, 2011](#)) o GREEND ([Kleiminger et al., 2015](#)). Por último, los conjuntos con períodos de muestreo largos a muy largos cuentan con períodos de 20 segundos a 30 minutos. Ejemplos de estos conjuntos son AMPds2 ([Makonin et al., 2016](#)), BERDS ([Gupta et al., 2010](#)) o SMART* ([Barker et al., 2012](#)).

Con respecto a la duración de monitoreo y registro de los valores, se pueden encontrar conjuntos con una acumulación de datos correspondientes a 36 días, como es el caso de REDD, o de hasta 4 años, como el conjunto DRED (Uttama et al., 2015). En algunos tipos de conjuntos de datos, como los que registran señales de electrodomésticos, no tiene sentido esta característica. Además, en muchos casos las duraciones varían dependiendo del edificio que se monitorea.

En relación a la cantidad de edificios monitoreados, el rango comienza en uno, con ejemplos como iAWE (Batra et al., 2013) o IHEPCDS (Hebrail, 2012), y se extiende hasta los 722, como es el caso de Dataport (Parson et al., 2016). Al igual que con la duración, para algunos tipos de conjunto de datos esta característica carece de sentido.

Sobre el origen de los conjuntos de datos, se puede afirmar que proceden de múltiples países. Sin embargo, a nivel de continentes, la mayoría de los conjuntos tienen origen en países Europeos, Norteamericanos o Asiáticos, quedando con mínima o nula participación los países de América del Sur, África u Oceanía.

En lo que refiere al tipo de valores registrados, los conjuntos se clasifican en: consumos desagregados por electrodomésticos, consumos desagregados por clase/grupos, señales de electrodomésticos, eventos de electrodomésticos o consumo agregado. En el caso de la desagregación por electrodoméstico, se registran los consumos totales agregados del edificio y el consumo de un conjunto de electrodomésticos seleccionados. Ejemplo de este tipo de conjuntos son DEDDIAG (Wenninger et al., 2021) y ECO (Kleiminger et al., 2015). La desagregación también puede darse por clase de electrodomésticos o por agrupación de estos (e.g., enchufes/electrodomésticos de una habitación), como por ejemplo en el conjunto iAWE. Por último, en lo que refiere al consumo de dispositivos, los conjuntos se clasifican por aquellos que registran únicamente los eventos de encendido/apagado, como es el caso de BERDS, o los que registran únicamente las señales de consumo (i.e., la firma que demarca el consumo de un dispositivo). Generalmente, la señal es registrada con un período de muestreo bajo (o muy bajo), lo que permite trabajar con una señal de excelente resolución. Ejemplos de este tipo de conjuntos son COOLL, HFED o Tracebase (Reinhardt et al., 2012). Otro tipo de conjunto de datos, quizás el más básico, es el que registra únicamente el consumo total agregado de los edificios, como el conjunto HUE (Makonin, 2019).

Por último, se relevó la forma de acceso a los datos. Se constató que se puede acceder de forma pública a todos los conjuntos relevados, aunque en algunos casos es necesario realizar una solicitud formal (Dataport, ECO, EDP, GREEND y REDD). En todos los casos, el uso que se le da a los datos es libre a excepción del conjunto Dataport, que restringe el uso con fines no académicos.

Comúnmente, los conjuntos reúnen únicamente datos de consumo eléctrico, aunque existen excepciones. Por ejemplo, el conjunto SMART* incluye datos climáticos para el rango de fechas monitoreadas, AMPds2 incluye consumo de gas natural y agua de los hogares, y el conjunto IDEAL incluye consumo de gas, datos de sensores contextuales (como temperatura, humedad y luz) para cada dormitorio, la temperatura de las cañerías de calefacción y del agua caliente y el resultado de encuestas realizadas a los hogares.

La Tabla 3.1 resume las características de los conjuntos de datos relevados, identificados por la abreviación del nombre asignada en el artículo que lo presenta. En caso de no encontrarse ninguna abreviación para el conjunto, se conformó una utilizando la primer letra de las palabras que conforman su nombre (e.g., IHEPCDS). Las características se presentan de la siguiente manera:

- Período de muestro. Para los casos iguales o menores a 1 segundo, el período se expresó como una frecuencia.
- Duración. El valor presentado corresponde a la duración máxima monitoreada, cuando aplica. Es necesario aclarar que según el edificio monitoreado que se tome en cuenta, esta duración puede variar.
- Cantidad de edificios. Este valor indica la cantidad de edificios monitoreados, cuando corresponde. Los edificios pueden ser hogares u oficinas.
- País. El valor indica el código de país de origen de los datos. Los códigos de país están definidos por la norma ISO-3166-1 (<https://laendercode.net/es/2-letter-list.html>).
- Tipo de valores. DE (n) indica que el conjunto contiene datos de consumo desagregados de hasta n electrodomésticos, DC (n) indica que existe desagregación de hasta n clases, SE refiere al registro de las señales de los electrodomésticos (i.e., su firma de consumo) y EE refiere al registro de eventos de los electrodomésticos (prendido/apagado).
- Acceso a los datos. AP indica que el conjunto de datos es de acceso público, SA indica que es de acceso público y de uso libre para fines

académicos, y S indica que se requiere una solicitud formal para su utilización.

<i>nombre</i>	<i>período de muestreo</i>	<i>duración</i>	<i>edificios</i>	<i>país</i>	<i>tipo de valores</i>	<i>acceso</i>
AMPds2 (Makonin et al., 2016)	1 min	2 años	1	CA	DE (11)	AP
BERDS (Gupta et al., 2010)	20 s	6 meses	7	US	EE (4)	AP
BLOND (Kriechbaumer y Jacobsen, 2018)	50–250 kHz	50–213 días	1	DE	DE (16)	AP
COOLL (Picon et al., 2016)	100 kHz	–	–	FR	SE (12)	AP
DP (Parson et al., 2016)	1 min	4 años	722	US	DE (10)	SA/S
DEDDIAG Wenninger et al. (2021)	1 Hz	3.5 años	15	DE	DE (10)	AP
DRED (Uttama et al., 2015)	1 Hz 1 min	2-4 años	5	NL	DE (5)	AP
ECO (Kleiminger et al., 2015)	1 Hz	8 meses	6	CH	DE (12)	AP/S
EDP (Saldanha y Beausoleil, 2012)	1 min	1 año	12	CA	DE (8)	AP/S
ENERTALK (Shin et al., 2019)	15 Hz	122 días	22	KR	DE (5)	AP
GREEND Kleiminger et al. (2015)	1 Hz	1 año	9	IT, AT	DE (9)	AP/S
HFED (Gulati et al., 2014)	10kHz 5MHz	–	–	IN	SE (15)	AP
HUE (Makonin, 2019)	1 hora	3 años	22	CA	AG	AP
iAWE (Batra et al., 2013)	1 Hz	73 días	1	IN	DE (11) DC (11)	AP
IDEAL (Pullinger et al., 2021)	1 Hz	23 meses	255	GB	DE (12)	AP
IHEPCDS (Hebrail, 2012)	1 min	47 meses	1	FR	DC (3)	AP
REDD (Kolter y Johnson, 2011)	1 Hz 15 kHz	36 días	6	US	DE (10-24)	AP/S
REFIT (Murray et al., 2017)	8 s	2 años	20	GB	DE (9)	AP
SMART* (Barker et al., 2012)	30 min	3 años	7	US	DE (5-31)	AP
TB (Reinhardt et al., 2012)	1 Hz	–	10	DE	SE (31)	AP

UK-DALE (Kelly y Knottenbelt, 2015b)	6 s 16 kHz	2–4 años	5	GB	DE (5)	AP
VT (Pipattanasomporn et al., 2014)	1 Hz 1 min	4 meses	2	US	DE (7)	AP

Tabla 3.1: Conjuntos de datos de consumo energético.

3.3. Análisis de datos de consumo eléctrico residencial

La revisión de literatura permitió identificar múltiples enfoques en la caracterización de consumo eléctrico aplicado a diferentes áreas geográficas (e.g., países, ciudades, áreas censales). En muchos casos se identifica el uso de herramientas estadísticas para llevar a cabo el análisis de los datos disponibles, como por ejemplo las regresiones logísticas o multinivel que también fueron utilizadas en esta tesis. A continuación se presenta una revisión de algunos de los trabajos que resultaron de relevancia.

[Chávez et al. \(2017\)](#) realizaron un estudio del consumo eléctrico en la ciudad de La Plata, Argentina. Al momento de presentar el trabajo, el sector generador del país afrontaba dos grandes problemas: los picos de consumo no podían ser cubiertos (incrementaban 5% por año), y tenían una matriz de generación poco diversa. En 2014, el sector residencial de Argentina se conformaba por 13.382.765 clientes, con un consumo promedio anual de 3017,80 kWh. La fuente principal de energía eran el gas natural (64,92%), la electricidad (23,70%), el gas licuado de petróleo (9,65%) y otros (1,73%). Ya que el 90% de los edificios conectados a la red se encontraba en áreas urbanas, los autores decidieron enfocarse en mejorar el consumo en éstas áreas ya que tendría grandes impactos en el consumo total. El trabajo identificó 1010 áreas censales (determinadas por el censo local realizado en 2010) que presentaron consumo eléctrico similar al procesarlas con técnicas de *clustering* en ocho grupos. El algoritmo de *clustering* utilizado fue el *k-means*, que procesó seis consumos bimensuales promedios en cada área. Los resultados fueron luego relacionados a seis variables socioeconómicas (cantidad de dormitorios, personas por hogar, porcentaje de hogares que no cubren las necesidades básicas, cobertura de red

de gas, tipología del hogar y calidad de la edificación) para estudiar su impacto en el consumo. El trabajo concluyó que a medida que la cantidad de personas por hogar o la cantidad de personas por dormitorio aumenta, se incrementa rápidamente el consumo eléctrico. Lo inverso ocurre con la cobertura de la red de gas, que a menor cobertura mayor es el consumo eléctrico. En cuanto a la tipología del hogar y la calidad de la edificación, cuanto mayor es la cantidad de edificios de apartamentos en el área, menor es el consumo eléctrico, y cuanto más precarias son las edificaciones, mayor es el consumo eléctrico.

En un trabajo similar, se analizaron los datos de consumo eléctrico de 3941 medidores inteligentes en Irlanda, con períodos de muestreo de 30 minutos y combinados con variables socioeconómicas, demográficas y de características de cada edificación (McLoughlin et al., 2012). El trabajo propuso encontrar las características que influyen en mayor medida al consumo eléctrico. En general, las características más analizadas, consecuentemente con la facilidad de recolección, son el tipo de edificación, el ingreso del hogar, la posesión de electrodomésticos y la cantidad de ocupantes. Se comenzó utilizando los primeros seis meses de datos recolectados para la elaboración del análisis, luego se les ofreció a los clientes cuatro planes comerciales. Se estudió el efecto del tipo de edificación, características de los ocupantes y electrodomésticos en el hogar, a través de cuatro parámetros: consumo eléctrico total, demanda máxima, factor de carga (a menor valor, mayores picos de consumo) y el tiempo en valores de consumo máximo. Se implementaron dos modelos de regresión lineal, uno para estudiar cómo afecta el tipo de edificación y las características de los ocupantes en el consumo total, y el otro para estudiar cómo la posesión de electrodomésticos afecta en el consumo. Los autores concluyen que el consumo eléctrico es afectado negativamente a mayor cantidad de dormitorios y ante la presencia de ocupantes de entre 36 y 55 años o profesionales. Por otro lado, el consumo es afectado positivamente en presencia de edificios de apartamentos y ocupantes de clase social baja o media. En cuanto a la posesión de electrodomésticos, los hogares que utilizan electricidad para calentar agua o cocinar, presentan mayores niveles de demanda que el resto. La demanda máxima es afectada directamente por la composición de los hogares, el número de dormitorios, y los electrodomésticos para calentar agua o cocinar. Entre los electrodomésticos presentes en las encuestas realizadas para la recolección de datos, el secaropa, el lavavajila y el horno eléctrico son los que contribuyen en mayor medida a los niveles máximos de demanda. En cambio, el factor de

carga es sensible al tipo de edificación y la cantidad de dormitorios, pero presenta valores bajos en presencia de ocupantes jóvenes, aún con calentadores de agua y hornos eléctricos. Los electrodomésticos que reducen más el factor de carga son el calentador instantáneo de ducha y el horno. En momentos de máxima demanda, el número de ocupantes afecta más que el tipo de edificación. Sobre el momento del día con mayor demanda, cuanto más adultos son los ocupantes, la demanda máxima se registra durante la mañana, y con ocupantes de mediana edad se registra tarde en el día.

Utilizando el mismo conjunto de datos (consumo eléctrico y encuestas de Irlanda), se estudió la posibilidad de inferir características de los hogares utilizando como entrada el consumo eléctrico y el número de residentes total y residentes menores, llamándolo *Smart Census* (Anderson et al., 2017). La idea consistió en caracterizar las áreas censales en lugar de los hogares de forma individual. Los indicadores utilizados se basaron en los relevados en la revisión bibliográfica realizada por los autores, considerando la magnitud de carga, las estadísticas, las propiedades temporales y algunos indicadores detallados en McLoughlin et al. (2012) (e.g., factor de carga, potencia consumida). Para el análisis, primero se identificaron las características de los hogares que permitían una mejor inferencia de los indicadores de perfil de consumo, utilizando un modelo de regresión multinivel y variables explicativas como los ingresos y el estado laboral del cabeza de hogar, la presencia de niños y el número de residentes en el hogar. Luego, mediante una regresión logística, se seleccionaron los indicadores que mejor invierten la dirección del modelo. En total, se construyeron cuatro modelos diferentes, variando la combinación de variables explicativas utilizadas. Los resultados fueron procesados únicamente para la predicción del estado laboral del cabeza de hogar, obteniendo una precisión cercana al 60%. Se concluyó que, a pesar de que la precisión alcanzada fue para una única característica, es factible inferir características de los hogares a partir de su perfil de consumo eléctrico.

En un estudio realizado en Brasil, se analizó el consumo eléctrico residencial desde 1985 a 2013 (Villareal y Moreira, 2016). El consumo residencial en Brasil representa el 26% del total, con un crecimiento promedio anual de 4,9%, conformado principalmente por la ducha eléctrica (19%), la heladera (18%), la iluminación (15%), la televisión (11%), el aire acondicionado (5%), el congelador (5%) y el lavarropa (1,8%), de acuerdo a datos de los años 2012 y 2013. El trabajo analizó el impacto de un conjunto de variables socioeconómi-

cas en el consumo eléctrico del país, utilizando métodos de regresión lineal. Las variables utilizadas fueron el número de hogares en el país, los ingresos familiares, los tipos de tarifa eléctrica, la posesión de electrodomésticos y las políticas sociales/económicas que se aplicaron en el período con influencia en el consumo eléctrico. Se seleccionaron tres políticas para el estudiar: contención del consumo eléctrico, facilitación de acceso a la electricidad para familias de bajos recursos y programas de eficiencia energética. Los valores de elasticidad obtenidos permitieron concluir que al aumentar 1 % el número de residentes, la electricidad consumida aumenta 1,53 %, al aumentar 1 % los ingresos familiares, el consumo eléctrico aumenta 0,19 % y al aumentar 1 % la tarifa eléctrica, el consumo eléctrico se reduce en un 0,23 %. Los modelos presentaron coeficientes de determinación altos (mayores a 0,9), lo que muestra una relación fuerte entre las variables analizadas y el consumo eléctrico.

En Uruguay, [Laureiro \(2018\)](#) presentó una investigación sobre el consumo eléctrico residencial basado en características socioeconómicas (ingreso per capita y composición del hogar), características del hogar (tamaño, tipo, material de construcción y equipamiento), destino de uso de la electricidad (cocina, calefacción, ventilación, calentamiento de agua sanitaria y conservación de alimentos) y la temperatura ambiente. Mediante técnicas de regresión de mínimos cuadrados ordinarios (Ordinary Least Squares, OLS) y regresión cuantílica (Quantile Regression, QR) se procesaron datos de 2994 hogares y de una encuesta sobre consumo eléctrico realizada por UTE e INE en el año 2015. Las casas seleccionadas contaban con un consumo mensual promedio mayor a 50 kWh y se ubicaban en ciudades con al menos 10000 habitantes. Un análisis superficial primario arrojó que el ingreso per capita es un factor relevante aunque no único, que los calentadores de agua y los aires acondicionados impactan directamente sobre el consumo eléctrico y que los electrodomésticos de confort térmico son más frecuentes en edificios con mayor consumo. El análisis con OLS concluyó que: i) el ingreso per capita tiene un alto valor de elasticidad, ii) el incremento de 1 % en los metros cuadrados de un hogar resulta en el incremento de 0,06 % en la electricidad consumida, iii) en las casas se consume en promedio 10,8 % más que en los apartamentos, iv) el consumo eléctrico se incrementa en 8,2 % por cada aire acondicionado y en 17,2 % por cada termotanque, y v) las variables regionales no impactan significativamente en el consumo. Al dividir los datos por nivel socioeconómico bajo, medio o alto, la regresión OLS muestra algunas diferencias en el consumo eléctrico entre los

subconjuntos: i) a mayor ingreso per capita, menor es el impacto de esta variable, ii) cada miembro extra mayor de 10 años en hogares socioeconómicamente bajos impactan menos que los de hogares socioeconómicamente medios/altos, iii) el tipo y tamaño del hogar impacta más en contextos socioeconómicamente bajos, iv) cada aire acondicionado o termotanque extra, impacta más en contextos socioeconómicamente medios/altos, y v) electrodomésticos de calefacción impactan más en contextos socioeconómicos altos que en medio/bajo. Para descubrir los comportamientos diferenciales que pueden quedar ocultos con OLS, el trabajo aplicó QR y los resultados muestran que: i) el impacto del ingreso per capita decrete de forma progresiva hacia los cuantiles altos, ii) los miembros extra mayores a 10 años impactan apenas más solo en cuantiles altos mientras que los menores de 10 años impactan igual en todos los cuantiles, iii) el tamaño de la edificación impacta más en cuantiles altos que en los bajos, iv) el tipo de edificación impacta significativamente solo en cuantiles medios y altos, mientras que los materiales directamente no reflejan ningún impacto, v) los aires acondicionados impactan más en cuantiles bajos que en altos y los termotanques tienen el mismo impacto en todos los cuantiles, vi) electrodomésticos de cocina, lavarropa, secarropa y calefacción de agua sanitaria muestran un impacto en forma de U invertida (bajo en los cuantiles bajos y altos y alto en los cuantiles medios). El trabajo concluye que a pesar de que el ingreso per capita es una variable determinante en el consumo eléctrico, no es la única que tiene un impacto significativo. Por lo tanto, es importante incluir en el análisis variables como la composición del hogar, las características de la edificación y los diferentes usos energéticos, entre otras.

Esta tesis presenta en el Capítulo 6 un estudio del consumo eléctrico potencial a partir de la creación de un índice que evalúa zonas censales, datos de posesión de electrodomésticos y características de los dispositivos en plaza. Se plantea una validación de los resultados del índice utilizando técnicas de regresión y datos de consumo eléctrico real en Uruguay.

3.4. Resumen

El proceso de monitoreo no intrusivo de cargas comenzó a estudiarse en la década de 1990 y trata de analizar las características de la carga total agregada de un hogar con el fin de desagregarla en cargas individuales por electrodomésticos. La principal ventaja de este proceso radica en que no es ne-

cesario instalar grandes cantidades de dispositivos en el interior de los hogares, sino que basta con monitorear la carga principal correspondiente al consumo total agregado. En sus inicios, para la desagregación se utilizaron técnicas de programación dinámica (algoritmo Viterbi, modelos escondidos de Markov) y actualmente se implementan técnicas de inteligencia computacional (RNN, LSTM, etc). Además, se ha experimentado la aplicación de NILM sobre otros recursos, como el consumo de agua del hogar. Este trabajo presenta una técnica de desagregación basada en patrones de consumo similares, que sirve de acercamiento a la problemática de desagregación de cargas eléctricas.

Los métodos de inteligencia computacional generalmente hacen un uso exhaustivo de datos. Particularmente, los métodos del tipo supervisado utilizan los datos para su etapa de entrenamiento. En los últimos años se han presentado múltiples conjuntos de datos de consumo eléctrico, particularmente de consumo residencial, utilizados para el entrenamiento de algoritmos de desagregación. Los conjuntos varían el período de muestreo, el país de origen, el tipo de desagregación y su accesibilidad, entre otras características. Algunos de los conjuntos más citados en la literatura científica son AMPds2, REDD y UK-DALE. Esta tesis utiliza el conjunto UK-DALE en la creación de datos sintéticos utilizados para el entrenamiento del algoritmo de desagregación presentado en el Capítulo 4, y en el Capítulo 5 presenta un conjunto de datos de consumo eléctrico residencial en Uruguay, nombrado ECD-UY.

Los datos de consumo eléctrico se utilizan en combinación con información socioeconómica de los países para caracterizar los diferentes usos. Por ejemplo, utilizando técnicas de agregación, es posible relacionar la variación de consumo eléctrico a partir de las características de los hogares y la composición de las personas que la habitan. Otro propósito posible es entender cómo se consume la electricidad por zonas geográficas y así planificar su demanda o elaborar mejores planes comerciales. Además, algunos trabajos infieren las características socioeconómicas de un hogar a partir del consumo registrado, o predicen el consumo a partir de las características de un área geográfica. Esta tesis presenta en el Capítulo 6 un índice de consumo eléctrico potencial a partir de la combinación de información de características de electrodomésticos e información socioeconómica por áreas censales, validado por regresiones lineales OLS y datos de consumo eléctrico residencial del conjunto ECD-UY.

Capítulo 4

Algoritmo de monitoreo no intrusivo de consumo eléctrico basado en similitud de patrones

Este capítulo presenta un algoritmo de inteligencia computacional para la desagregación de consumo eléctrico residencial, mediante procesos de monitoreo no intrusivo de cargas. El algoritmo es entrenado con datos sintéticos, generados a partir del consumo real registrado en el conjunto de datos UK-DALE. Se brindan detalles de las técnicas utilizadas para la creación de los datos y el posterior entrenamiento y evaluación del algoritmo. El algoritmo presentado brinda un acercamiento al abordaje del problema de desagregación de consumo eléctrico utilizando técnicas de inteligencia computacional aplicadas a NILM en escenarios sintéticos menos complejos que los reales.

4.1. Introducción

Se plantea un algoritmo de desagregación de consumo eléctrico residencial basado en el reconocimiento de los estados de encendido y apagado de los electrodomésticos. El reconocimiento de los estados consiste en la búsqueda de patrones de consumo similar, registrados en un histórico de consumos etiquetados.

El algoritmo recibe el nombre de Desagregación Simple (DS) y se compone de la etapa de entrenamiento y la etapa de predicción. Para la etapa de entrenamiento se utilizan datos sintéticos creados a partir del consumo eléctrico

real registrado en el conjunto de datos UK-DALE. La metodología para la generación de los datos es presentada en detalle en la Sección 4.5.1. Los datos sintéticos generados se utilizan también durante la evaluación del algoritmo, recreando escenarios complejos.

Para estudiar la robustez del algoritmo se plantearon escenarios de evaluación donde se recrea la superposición de consumo de electrodomésticos con demandas similares, se agrega ruido a la señal total de consumo, y se varía el período de muestreo de datos (i.e., intervalos de tiempo entre dos registros consecutivos de consumo). Los resultados de los experimentos son procesados con la herramienta NILM Toolkit (nilmtool) y comparados con los obtenidos por dos algoritmos base, presentes en la literatura revisada: Combinatorial Optimization (CO) y Factorial Hidden Markov Model (FHMM). La evaluación muestra que DS obtiene resultados precisos bajo escenarios complejos, superando el rendimiento de los algoritmos base.

El trabajo realizado cumple el objetivo de brindar un primer acercamiento a la resolución del problema de desagregación, evitando dificultades que presentan los escenarios reales (e.g., electrodomésticos con consumos variables). El algoritmo fue presentado en forma de artículo en el Congreso Iberoamericano de Ciudades Inteligentes (ICSC-CITIES 2019) (Chavat et al., 2019), y seleccionado para extensión y publicación en la revista arbitrada Revista Facultad de Ingeniería Universidad de Antioquía (Chavat et al., 2020).

4.2. Descripción general

Sea $f : \{0, 1\}^m \rightarrow R$ una función que dado un conjunto de estados de electrodomésticos (i.e., encendido/apagado), retorna el consumo total agregado. Se considera una función $g_{E,Z} : R^{2d+1} \rightarrow R^m$ en la que el número positivo d determina un vecindario temporal para las predicciones. Por lo tanto, como se observa en la Ecuación 4.1, la función $g_{E,Z}$ recibe un vector $X^i \in R^{2d+1}$ con los valores de consumo total agregado de un hogar i , en el intervalo de un instante t , y retorna un vector \hat{Y}_t^i correspondiente al estado predicho para cada uno de los m electrodomésticos del hogar i , en el instante t (i.e., la configuración estimada de los electrodomésticos del hogar).

$$(\hat{y}_t^{i_1}, \hat{y}_t^{i_2}, \dots, \hat{y}_t^{i_m}) := g_{E,Z}(x_{t-d}^i, \dots, x_t^i, \dots, x_{t+d}^i) \quad (4.1)$$

La función $g_{E,Z}$ contiene elementos aleatorios y se define utilizando información del conjunto de datos de entrenamiento. El conjunto E se conforma por vectores con los estados de los electrodomésticos y se define como $E = \{e_t \mid t = 1, \dots, n \wedge e_t \in \{0, 1\}^m\}$, mientras que Z se conforma por los valores de consumos total agregado y se define como $Z = \{z_t \mid t = 1, \dots, n \wedge z_t \in R\}$. Los vectores del conjunto de estados E son dominio de la función f , mientras que los valores de consumo total agregado del conjunto Z , son el codominio, como se muestra en la Ecuación 4.2.

$$z_t = f((e_t^1, e_t^2, \dots, e_t^m)) \quad (4.2)$$

Los parámetros de la función $g_{E,Z}$ fueron seleccionados de forma empírica buscando maximizar la sumatoria definida por la Ecuación 4.3, donde A es el conjunto de configuraciones ambiguas definido como $A = \{y \in \{0, 1\}^m \mid \exists y' \in \{0, 1\}^m \wedge y' \neq y \wedge f(y') = f(y)\}$, y la sumatoria definida por la Ecuación 4.4, donde T es el conjunto de instantes procesados y la sumatoria define la cantidad de predicciones correctas del estado del dispositivo.

$$\sum_{y_t \in A} \prod_{i=1}^m 1_{\{\hat{y}_t^i = y_t^i\}} \quad (4.3)$$

$$\sum_{t \in T} \prod_{i=1}^m 1_{\{\hat{y}_t^i = y_t^i\}} \quad (4.4)$$

El algoritmo toma como insumo de entrada los valores detallados a continuación:

- Vector X con instancias de consumo total agregado x_t .
- Conjunto Z de valores de consumo real total agregado z_t^i para el hogar i , usado en la etapa de entrenamiento del algoritmo. El período de muestreo de estos valores (i.e., tiempo entre t y $t + 1$) debe ser igual al utilizado por X .
- Conjunto E de estados de electrodomésticos e_t^i , cuyos instantes de registro corresponden a los del conjunto Z , usado en la etapa de entrenamiento.
- Parámetro d , que define el margen (temporal) de valores de consumo a tener en cuenta en la desagregación.

- Parámetro δ , que define el margen de valores de consumo a tener en cuenta en la desagregación.
- Parámetro H , que define un umbral para clasificar el consumo total entre alto y bajo.

El resultado final del algoritmo es un vector $\vec{y} \in R^m$ con la desagregación del consumo de un hogar.

4.3. Etapas del algoritmo

A continuación se presentan las dos etapas del algoritmo DS: etapa de entrenamiento y etapa de evaluación. La etapa de entrenamiento utiliza un conjunto de datos de consumo total agregado y desagregado mientras que la etapa de evaluación utiliza un conjunto de datos únicamente con valores de consumo total agregado.

4.3.1. Etapa de entrenamiento

La primer etapa del algoritmo DS corresponde a la etapa de entrenamiento, que se encarga de procesar los datos de entrenamiento para construir un vector de descriptores de consumo, que relaciona cada registro de consumo con los registros de consumo inmediatamente anteriores y posteriores. Este proceso se detalla de forma algorítmica en el Algoritmo 1, donde M_Z corresponde al vector de descriptores de consumo y Z al conjunto de consumo total agregado utilizado en el entrenamiento. Los valores del vector ($z_j \in M_Z$) son descriptores del consumo de los dispositivos encendidos en el instante correspondiente al registro. A estos elementos se les llama descriptores de la firma de consumo. La iteración principal del algoritmo ocurre en las líneas 3–11, donde se itera sobre cada elemento del conjunto de entrenamiento. En cada iteración, el algoritmo cuenta cuántos registros cercanos (i.e., registros vecinos temporalmente) presentan un consumo similar (líneas 5–9). Al finalizar cada iteración se actualiza el vector de descriptores en la posición del registro procesado (línea 10). El vector resultante es utilizado en la etapa de evaluación para buscar instantes pasados que presenten un consumo similar al que se desea desagregar.

Algoritmo 1: Etapa de entrenamiento del algoritmo Desagregación Simple (DS)

Datos: Conjunto de datos de entrenamiento Z

Resultado: Vector M_Z de descriptores de firmas de consumo

```
1 inicio
2    $M_Z \leftarrow$  vector de largo igual a  $Z$ 
3   para  $z_i \in Z$  hacer
4     similares  $\leftarrow$  0
5     para  $\{z_j \in Z : |j - i| < d\}$  hacer
6       si  $z_j > z_i - \varphi$  entonces
7         similares  $\leftarrow$  similares + 1
8       fin si
9     fin para
10     $M_Z[i] \leftarrow$  similares
11  fin para
12 fin
```

4.3.2. Etapa de evaluación

La segunda etapa corresponde a la etapa de evaluación, que se encarga de desagregar los datos de consumo real. El Algoritmo 2 describe de forma algorítmica esta etapa. La primer iteración, correspondiente a las líneas 3–8, presenta la misma lógica que la iteración principal del entrenamiento, pero aplicada al vector de consumo a ser desagregado. El resultado de la iteración es un vector M_X , de descriptores de consumo de X (al igual que lo es M_Y de Y). La segunda iteración (líneas 9–19) itera sobre el conjunto de datos de predicción en busca de similitudes con los registros del conjunto de datos de entrenamiento. Anidada a la segunda iteración se encuentra la tercera (líneas 11–13), que evalúa la diferencia entre los registros del conjunto de entrenamiento y los registros evaluados (X). Si la diferencia entre los registros comparados es menor a un umbral (δ) y el registro de entrenamiento representa un consumo alto (i.e., mayor al valor H), el registro de entrenamiento es apartado para ser procesado posteriormente (i.e., agregado al conjunto I). Los parámetros δ y H son detallados en la Sección 4.2. Una vez terminada la iteración pueden ocurrir dos situaciones dependiendo de la cantidad de registros seleccionados (cardinal de I). Si existen registros apartados (i.e., el conjunto I no es vacío), se comparan los conjuntos de descriptores de consumo de esos registro y se filtran aquellos que minimizan la distancia a algún descriptor (i.e., si son cer-

canos a un descriptor, se pasan al conjunto J), como se puede observar en la línea 15. Si no existen registros apartados (i.e., I es vacío), se comparan los registros de consumo del conjunto de entrenamiento con los del conjunto evaluado, manteniendo los que minimizan la diferencia (i.e., si son cercanos a un consumo histórico, se pasan al conjunto J), como se observa en la línea 17. De los registros resultantes (i.e., contenidos en J), se selecciona uno de forma aleatoriamente (línea 18), y se obtiene el correspondiente vector de estados de electrodomésticos del conjunto de entrenamiento (i.e., para el z_k , se obtiene el correspondiente $e_k \in \{0, 1\}^m$).

Algoritmo 2: Etapa de predicción del Algoritmo DS

Datos: Conjunto de datos de evaluación X

Resultado: Vector y con resultado de estados de dispositivos

```

1 inicio
2    $M_X \leftarrow$  vector de largo igual a  $X$ ;
3   para  $x_i \in X$  hacer
4     similares  $\leftarrow$  0;
5     para  $\{x_j \in X : |j - i| < d\}$  hacer
6       si  $x_j > x_i - \varphi$  entonces
7         similares  $\leftarrow$  similares + 1;
8      $M_X[i] \leftarrow$  similares;
9   para  $x_i \in X$  hacer
10     $I \leftarrow \emptyset$ ;
11    para  $z_j \in Z$  hacer
12      si  $|z_j - x_i| \leq \delta$  AND  $x_i > H$  entonces
13         $I \leftarrow I \cup \{j\}$ ;
14    si  $I$  no es vacío entonces
15       $J \leftarrow \operatorname{argmin}\{|M_Z(I(\cdot)) - M_X(i)|\}$ ;
16    en otro caso
17       $J \leftarrow \operatorname{argmin}\{|z(\cdot) - x(i)|\}$ ;
18     $k \leftarrow \operatorname{aleatorio}\{1, \dots, \operatorname{length}(J)\}$ ;
19     $y(i, \cdot) \leftarrow E(J(k), \cdot)$ ;

```

Conceptualmente, el algoritmo DS en su etapa de predicción relaciona el valor de consumo a desagregar con las variaciones de consumo en un contexto de tiempo cercano. Ambas características son clave para el problema de desagregación de consumo eléctrico. Para una mejor comprensión del algoritmo DS, las Figuras 4.1 y 4.2 presentan una síntesis de las etapas de entrenamiento, de evaluación y de los procesos que conforman al algoritmo.

Figura 4.1: Resumen de la etapa de entrenamiento del algoritmo DS

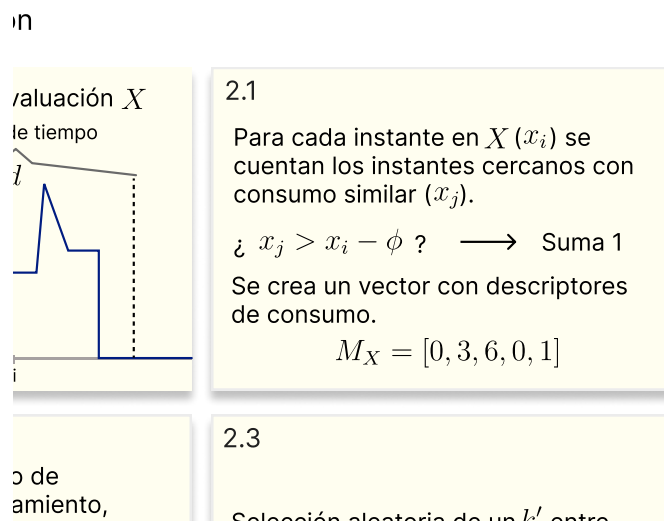


Figura 4.2: Resumen de la etapa de evaluación del algoritmo DS

4.4. Implementación

El algoritmo DS fue implementado utilizando el lenguaje Python versión 3 y las bibliotecas Pandas y numpy. La implementación del algoritmo sigue la estructura de *pipeline* de ejecución de los algoritmos presentes en nilmtk.

Para la generación de los datos sintéticos, se implementaron dos *scripts* auxiliares, también en lenguaje Python. El primer *script* normaliza los valores de consumo total agregado y los de consumo de los electrodomésticos presentes en el conjunto UK-DALE. El cálculo de normalización consiste en reemplazar todos los valores mayores a un umbral por un valor de consumo constante, y los que no superan el umbral se establecen en cero. Para los escenarios en los que se incluye ruido como parte del consumo total, el *script* calcula una distribución de probabilidad exponencial y la suma al valor final. Para generar consumo de electrodomésticos extra, el *script* ejecuta modelos probabilísticos exponenciales que simulan el comportamiento de diferentes dispositivos. Los electrodomésticos extra son aquellos que no son objetivo en la desagregación sino que simulan consumo eléctrico para generar ruido en la señal de consumo total agregado. Una vez normalizados los valores, se crea una estructura de directorios para almacenar los **metadatos** del conjunto de datos, siguiendo los lineamientos establecidos en NILM metadata (Kelly y Knottenbelt, 2014). Además, se escriben los datos de consumo para cada hogar, tanto los consumos desagregados como el total agregado. Los metadatos son escritos en formato Yet Another Markup Language (YAML) y los datos de consumo en formato Comma-Separated Values (CSV).

El segundo *script* interpreta la estructura y los datos creada por el primero, y lo traduce desde el formato CSV al formato Hierarchical Data Format (HDF) versión cinco. El resultado de este *script* es un conjunto de datos de consumo eléctrico sintético, total y desagregado por subconjunto de dispositivos objetivo, con la misma frecuencia de muestreo que el conjunto original.

La implementación del algoritmo DS, la de los *scripts* de generación datos sintéticos, y las modificaciones realizadas a nilmtk, se encuentra disponibles en el repositorio público <http://gitlab.com/jpchavat/nilm-scripts>.

4.5. Análisis experimental

Esta sección presenta el análisis experimental del algoritmo DS. Las instancias experimentales fueron ejecutadas como parte de un proceso nilmtk, utilizando los conjuntos de datos sintéticos creados a partir de UK-DALE.

4.5.1. Conjuntos de datos

Los conjuntos de datos utilizados en los experimentos fueron creados especialmente para el análisis experimental del algoritmo DS, a partir de datos reales de un hogar del conjunto UK-DALE. Teniendo en cuenta el impacto en el consumo que tiene cada electrodoméstico monitoreado en el conjunto de datos reales, se seleccionó un subconjunto con el cual trabajar. Los electrodomésticos considerados son: heladera, lavarropa, jarra eléctrica, lavavajilla y *home theater*. A continuación se describe el procedimiento de generación de datos sintéticos y se brindan detalles sobre la simulación de electrodomésticos que no son objetivo de la desagregación.

Procedimiento de generación

Se generaron diferentes instancias de datos sintéticos a partir de los datos reales. El procedimiento utilizado se describe a continuación:

- Primero se lee la estructura de datos UK-DALE y se genera una igual con metadatos propios, siguiendo la estructura *NILM Metadata* (Kelly y Knottenbelt, 2014).
- Luego se establecen dos vectores para normalizar el consumo eléctrico de los electrodomésticos. El primer vector se conforma de un valor por electrodoméstico que se calcula en base a la mediana de consumos máximos, el segundo está compuesto por valores de consumo constante, previamente definidos para cada electrodoméstico.
- Por último, se generan series temporales de consumo para cada electrodoméstico y para la señal de consumo total agregado. Cada registro de consumo en UK-DALE que supera cierto umbral (definido en 5,0 W) se reemplaza por el valor establecido en el punto anterior (mediana de los máximos o constante de consumo). Los registros cuyo consumo no superan el umbral establecido son considerados como instantes con el electrodoméstico apagado, y por ello se les asigna valor de consumo cero.

Los conjuntos de datos sintéticos resultantes tienen el mismo período de muestreo que el conjunto de datos original, con la particularidad de no presentar *gaps* (i.e., no presentan valores faltantes consecutivos). Los *gaps* del conjunto original fueron rellenados con valores correspondientes al promedio entre el consumo del extremo anterior y el posterior a cada *gap*. De esta forma, si el período de muestreo del conjunto original es de seis segundos, los datos sintéticos aseguran un registro de consumo exactamente cada seis segundos.

La metodología aplicada para la generación de datos sintéticos es genérica para todos los conjuntos de datos origen que cumplan con el formato *NILM metadata* y permite seleccionar qué hogares y qué electrodomésticos se tienen en cuenta en el proceso de construcción. La selección de hogares y electrodomésticos sobre un mismo conjunto origen permite generar diferentes variantes de datos. Por ejemplo, tomando como origen el conjunto UK-DALE, podríamos generar un conjunto de datos a partir de la casa uno y cierto subconjunto de electrodomésticos y otra variante a partir de la casa dos y otro subconjunto de electrodomésticos. La implementación de los procedimientos de generación de datos se encuentra disponible de forma pública en <http://gitlab.com/jpchavat/nilm-scripts>, en los *scripts* `normalize-nilm-tk-dataset.py` y `convert-csv-to-h5.py`.

Simulación de electrodomésticos no monitoreados

Los electrodomésticos extra no son considerados como objetivo en la desagregación, sino que se simulan para agregar ruido al consumo total agregado con el objetivo de evaluar el algoritmo DS ante la presencia de señales de consumo desconocidas. A continuación se describe el procedimiento de generación del consumo de estos dispositivos.

Tomando en cuenta que los eventos de encendido/apagado de los electrodomésticos se producen de forma continua e independiente y con una frecuencia promedio casi constante, se asume que siguen un proceso de Poisson. El tiempo que un electrodoméstico permanece apagado (T_{OFF}) es modelado como la discretización de una distribución exponencial de parámetro λ , mientras que el tiempo que permanece encendido (T_{ON}) se modela como la discretización de una distribución exponencial de parámetro μ . Las Ecuaciones 4.5 y 4.6 describen de forma matemática los valores T_{OFF} y T_{ON} , donde los parámetros U_1 y U_2 son números aleatorios con distribución uniforme en el rango $[0, 1]$ y $\lfloor x \rfloor$ representa la parte entera del valor x .

$$T_{OFF} = \left\lfloor \frac{1}{\lambda} \log(1 - U_1) \right\rfloor + 1, \quad (4.5)$$

$$T_{ON} = \left\lfloor \frac{1}{\mu} \log(1 - U_2) \right\rfloor + 1 \quad (4.6)$$

El Algoritmo 3 describe el procedimiento aplicado para la generación de consumo de electrodomésticos extra. El vector de salida \vec{y} es inicializado como un vector de valores cero y de largo N , en la línea 2. La iteración principal, línea 4, se repite hasta completar la generación de N intervalos. Los valores T_{OFF} y T_{ON} , correspondientes a los valores de las Ecuaciones 4.5 y 4.6, son calculadas en cada iteración, en las líneas 5 y 6 respectivamente. En la línea 7 se avanza el puntero m tantos intervalos como T_{OFF} lo indique, marcando el período que el dispositivo permanece apagado. La iteración de la línea 8 asigna el valor de encendido por una cantidad de intervalos definida por T_{ON} . Finalmente, en la línea 10 se actualiza el valor del puntero m , avanzando T_{ON} posiciones.

Algoritmo 3: Generación de consumo de dispositivos extra

Entrada:

Cantidad de intervalos a generar N
 Parámetro Poisson para apagado λ
 Parámetro Poisson para encendido μ

Resultado:

Vector de estados \vec{y} del dispositivo simulado

```

1 inicio
2    $\vec{y} \leftarrow \vec{0}$ ;
3    $m \leftarrow 0$ ;
4   mientras  $m < N$  hacer
5      $T_{OFF} \leftarrow \lfloor (-1/\lambda) \log(1 - rand[0, 1]) \rfloor + 1$ ;
6      $T_{ON} \leftarrow \lfloor (-1/\mu) \log(1 - rand[0, 1]) \rfloor + 1$ ;
7      $m \leftarrow m + T_{OFF}$ ;
8     para  $i = m, \dots, \min(m + T_{ON}, N)$  hacer
9        $y[i] \leftarrow 1$ ;
10     $m \leftarrow m + T_{ON}$ ;

```

Una vez finalizado el algoritmo DS, se obtiene un vector $\vec{y} \in \{0, 1\}^N$ cuyos valores representan los estados de encendido/apagado del electrodoméstico.

La actividad de los electrodomésticos extra es generada a partir de la discretización de dos distribuciones exponenciales de parámetros λ y μ . En el conjunto de datos sintéticos se generan ocho dispositivos extra: tres lámparas de 8 W, tres lámparas de 10 W, un microondas de 2000 W y un televisor de 40 W. Considerando que el algoritmo DS no tiene como objetivo la desagregación de los dispositivos extra, el consumo aportado en la señal de consumo total puede ser interpretado como ruido. En escenarios reales, no se puede asegurar que el consumo a desagregar esté conformado por el consumo de electrodomésticos idénticos (i.e., igual modelo, uso, nivel de consumo, etc.) a los utilizados en el entrenamiento del algoritmo DS. La presencia del consumo de electrodomésticos extra (ruido en el consumo total agregado) aporta a los escenarios sintéticos características presentes en los escenarios reales. La Tabla 4.1 detalla los electrodomésticos extra generados, los valores de λ y μ utilizados, la cantidad de instancias agregadas y el consumo eléctrico que representan.

<i>electrodoméstico</i>	λ	μ	<i>cantidad</i>	<i>consumo</i>
Lámpara	I/3600	I/300	3	8 W
Lámpara	I/15000	I/300	3	10 W
Microondas	I/7200	I/100	1	2000 W
Televisión	I/30000	I/7500	1	40 W

Tabla 4.1: Descripción de los electrodomésticos extra incluidos en instancias del conjunto de datos sintético. I representa el intervalo de muestreo expresado en segundos.

4.5.2. Instancias de evaluación

Las instancias de evaluación se definen a partir de los conjuntos de datos sintéticos que varían: i) la cantidad de electrodomésticos presentes en la señal de consumo agregado, ii) el conjunto de electrodomésticos objetivo a ser desagregados, y iii) el período de muestreo de los datos.

Instancias de conjuntos de datos sintéticos

Las instancias de conjuntos de datos sintéticos fueron generadas en dos etapas. En la primer etapa se generaron 5 instancias (#1, #2 y #3, #4 y #5) variando la cantidad y las características de consumo de los electrodomésticos, y el conjunto de electrodomésticos objetivo a ser desagregados. En todos estos casos, el período de muestreo original de UK-DALE fue submuestreado

a períodos de 5 minutos. En la segunda etapa se trabajó con las instancias #4 y #5 como base y se generaron 2 instancias más por cada una, en la que se submuestra la original a 10 y 15 minutos. Las instancias que tienen como base la instancia #4 son nombradas como sub-instancias #4-5, #4-10 y #4-15. De igual forma, las instancias que tienen como base la instancia #5 son nombradas como sub-instancias #5-5, #5-10 y #5-15.

El conjunto de datos de entrenamiento fue acotado a un rango de tiempo definido desde el día 2013-01-01 a las 00:00:00 horas al 2013-07-01 a las 00:00:00 horas, mientras que el conjunto de datos de evaluación fue acotado al rango de tiempo definido desde el día 2013-07-01 a las 00:00:00 al 2013-12-31 a las 23:59:59. De esta forma, se define el conjunto de entrenamiento y el de evaluación con un 50% de los datos disponibles. Las primeras cuatro instancias tienen como objetivo evaluar la eficacia del algoritmo DS ante escenarios con ambigüedad en el consumo de energía. La quinta instancia, aparte de la ambigüedad en el consumo, suma a la evaluación la presencia de consumo de electrodomésticos que no son objetivo de la desagregación (i.e., ruido en la señal de consumo total).

A continuación se describen las instancias/sub-instancias de datos generadas, que determinan diez escenarios de evaluación.

Instancia #1. Se normalizan los consumos de los electrodomésticos según la mediana de consumos máximos de cada período de activación (i.e., períodos en los que el electrodoméstico permanece encendido) y se filtran los *outliers* (i.e., valores anómalos, detectados utilizando el criterio de aquellos por fuera de la franja de desviación estándar). Luego, el conjunto se divide en un subconjunto de entrenamiento y un subconjunto de evaluación, de acuerdo al rango de fecha establecido para estos subconjuntos. Esta instancia tiene como objetivo evaluar el algoritmo DS trabajando con valores cercanos a los reales, pero manteniendo consumos constantes a lo largo del tiempo.

Instancia #2. Se normalizan los consumos de los electrodomésticos, de forma que dos electrodomésticos presentan un consumo constante idéntico. Luego, el conjunto se divide en un subconjunto de entrenamiento y un subconjunto de evaluación, de acuerdo al rango de fecha establecido para estos subconjuntos. Esta instancia tiene como objetivo evaluar el algoritmo DS ante un escenario básico de ambigüedad de consumo de dos electrodomésticos.

Instancia #3. Se normalizan los consumos de los electrodomésticos de forma similar a la instancia #2, con la particularidad de que la ambigüedad se presenta entre la suma del consumo del *home theatre* y el lavavajilla, frente a la suma de consumo de la heladera y la jarra eléctrica. Luego, el conjunto se divide en un subconjunto de entrenamiento y un subconjunto de evaluación, de acuerdo al rango de fecha establecido para estos subconjuntos. Esta instancia tiene como objetivo evaluar el algoritmo DS ante un escenario complejo de consumos ambiguos.

Instancia #4. Se normalizan los consumos de los electrodomésticos de dos formas diferentes, dependiendo la fecha del registro. Si el registro pertenece al primer rango de fechas, destinado a la etapa de entrenamiento, los valores se normalizan de igual forma que en la instancia #2. Si el registro pertenece al segundo rango de fechas, destinado a la etapa de evaluación, los valores de normalización son los de la instancia #2 con pequeñas variaciones en el consumo de los electrodomésticos (a excepción del lavarropa). Por ejemplo, el valor constante de consumo de la heladera fue normalizado a 260 W en lugar de 250 W. El objetivo de esta instancia es evaluar el algoritmo DS en un escenario con valores de consumo diferentes a los utilizados durante el entrenamiento.

Instancia #5. Se toma como base el conjunto de evaluación de la instancia #4 y se le agregan consumos de electrodomésticos extra, modelados por la discretización de una variable aleatoria exponencial (según el proceso explicado en la Sección 4.5.1). Esta instancia evalúa el algoritmo DS en presencia de electrodomésticos que no son objetivo de la desagregación.

Sub-instancia de #4 y #5. A partir de las instancias #4 y #5, se crean tres sub-instancias en las que se varían los períodos de muestreo en 5 minutos (#4-5 y #5-5), 10 minutos (#4-10 y #5-10) y 15 minutos (#4-15 y #5-15). Las sub-instancias con diferentes períodos de muestreo fueron definidas con el fin de evaluar la robustez del algoritmos DS en escenarios donde los registros se encuentran más dispersos, como es el caso de los datos generados por los medidores inteligentes de UTE, que cuentan con un período de muestreo de 15 minutos.

La Tabla 4.2 lista las instancias previamente descritas junto al período de muestreo y los valores de consumo de cada electrodoméstico agregado. La Figura 4.3 grafica el porcentaje de tiempo encendido/apagado de cada electrodoméstico, igual en todas las instancias de evaluación.

<i>inst.</i>	<i>uso</i>	<i>consumo de electrodoméstico (W)</i>					<i>período de muestreo extra (min.)</i>	
		<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>		
#1	EN,EV	117	3325	2390	2741	93	5	No
#2	EN,EV	250	2000	2500	2500	80	5	No
#3	EN,EV	300	1800	2200	2300	200	5	No
#4	EN	250	2000	2500	2500	80	5	No
#4-5	EV	260	2000	2400	2600	70	5	No
#4-10	EV	260	2000	2400	2600	70	10	No
#4-15	EV	260	2000	2400	2600	70	15	No
#5-5	EN,EV	250	2000	2500	2500	80	5	Si
#5-10	EN,EV	250	2000	2500	2500	80	10	Si
#5-15	EN,EV	250	2000	2500	2500	80	15	Si

Tabla 4.2: Detalle de las instancias de los conjuntos de datos sintéticos. Los usos se codifican como EN para entrenamiento y EV para evaluación.

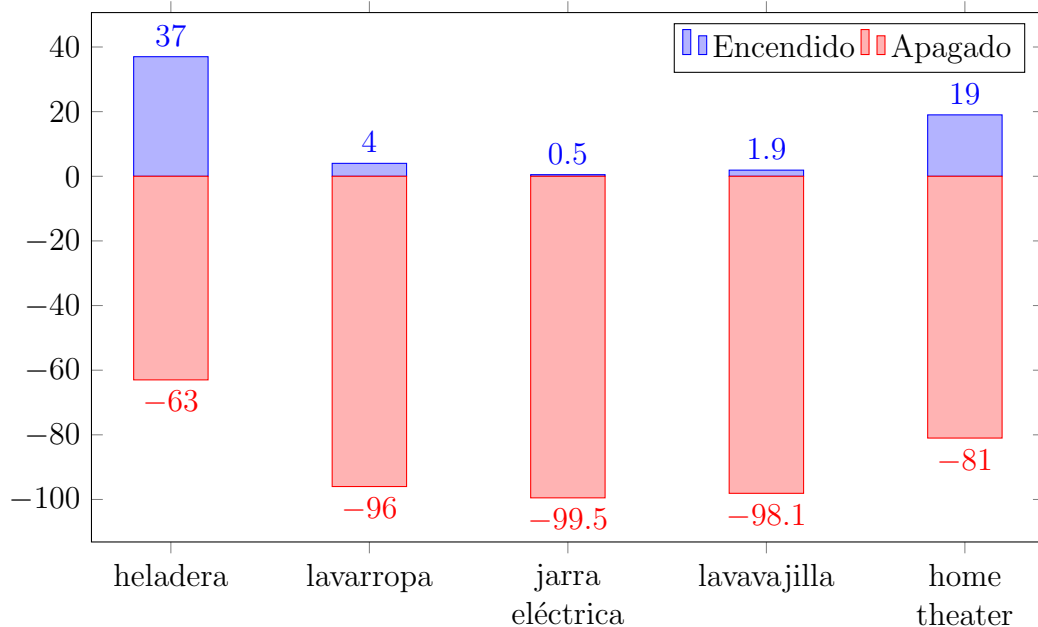


Figura 4.3: Porcentaje de tiempo encendido/apagado de los electrodomésticos en los conjuntos de datos sintéticos.

4.5.3. Métodos base para la comparación de resultados

Para poder comparar los resultados del algoritmo DS, las mismas instancias de prueba fueron procesadas por dos métodos de base seleccionados de la implementación de nilmtk: Optimización Combinatoria (*Combinatorial Optimization*, CO) y Modelos Ocultos de Markov Factorial (*Factorial Hidden Markov Model*, FHMM). A continuación se brindan detalles de ambos métodos.

Optimización Combinatoria

El método CO fue presentado por [Hart \(1992\)](#) e implementado en la herramienta nilmtk ([Batra et al., 2014](#)). CO busca la combinación óptima de estados de electrodomésticos que minimice la diferencia entre la suma del consumo del electrodoméstico predicho y el consumo agregado observado, teniendo en cuenta un conjunto de modelos de electrodomésticos. Siendo $P(t)$ el consumo total agregado para el tiempo t , P_i el consumo del electrodoméstico i y $a(t)$ un vector de n valores booleanos que indica si el electrodoméstico i se encuentra encendido en el tiempo t , se puede expresar CO con la ecuación 4.7. El método busca el vector \hat{a} que minimiza la diferencia entre el consumo total agregado y los estados predichos de los electrodomésticos encendidos.

$$\hat{a}(t) = \arg \min_a \left| P(t) - \sum_{i=1}^n a_i(t) \cdot P_i \right| \quad (4.7)$$

Dada la complejidad del algoritmo CO (exponencial en el número de electrodomésticos), es impracticable aplicar el método en escenarios con un gran número de electrodomésticos.

Modelos Ocultos de Markov Factorial

El método FHMM fue presentado por [Ghahramani y Jordan \(1996\)](#) y adaptado por [Kim et al. \(2011\)](#) al problema de desagregación de consumo eléctrico. Los Modelos Ocultos de Markov (*Hidden Markov Model*, HMM) son utilizados para modelar datos secuenciales utilizando probabilidades de transición. FHMM extiende los HMM, lo que permite modelar múltiples secuencias independientes de estados ocultos. FHMM maneja tres parámetros: i) probabilidades a priori, ii) probabilidades de las transiciones para cada estado y hacia cada estado posible, y iii) matriz de emisiones. La complejidad de este algoritmo es aún mayor que la del algoritmo CO, y por lo tanto, escala de peor forma.

4.5.4. Métricas de evaluación de resultados

Para la evaluación del algoritmo DS y su comparación con los métodos base, se utilizó un conjunto de métricas que consideran el vector $x_i^{(n)}$, los estados reales del electrodoméstico n para el hogar i y el vector $\hat{x}_i^{(n)}$, con los estados predichos del electrodoméstico n y el hogar i . Se consideraron las definiciones de *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), y *False Negative* (FN), como se indica en las Ecuaciones 4.8–4.11.

$$TP = \sum_i \text{AND}(x_i^{(n)} = 1, \hat{x}_i^{(n)} = 1) \quad (4.8)$$

$$FP = \sum_i \text{AND}(x_i^{(n)} = 0, \hat{x}_i^{(n)} = 1) \quad (4.9)$$

$$TN = \sum_i \text{AND}(x_i^{(n)} = 0, \hat{x}_i^{(n)} = 0) \quad (4.10)$$

$$FN = \sum_i \text{AND}(x_i^{(n)} = 1, \hat{x}_i^{(n)} = 0) \quad (4.11)$$

Para medir el rendimiento, se consideraron cinco métricas basadas en las relaciones previas, detalladas a continuación:

- *Precisión*. La precisión (*Precision*, en inglés) de una predicción se define como un estimador de la probabilidad condicional de las predicciones de encendido, como se observa en la Ecuación 4.12.
- *Recall*. El *recall* se define como la probabilidad condicional de que un electrodoméstico esté encendido, dado que la predicción es de encendido, como se observa en la Ecuación 4.13.
- *Valor-F*. El Valor-F se define como la media armónica entre la precisión y el *recall*, como se detalla en la Ecuación 4.14.
- *Error de Asignación de Energía Total*. El error en la asignación de energía total (*Total Energy Error assigned*, TEE), se define como el error en la asignación del consumo total asignado, como se puede observar en la Ecuación 4.15.
- *Error Normalizado de Asignación de Potencia*. El error normalizado en la asignación de potencia (*Normalized Error in Assigned Power*, NEAP) se define como la media del error normalizado en la asignación de potencias, como se detalle en la Ecuación 4.16.

$$precisión = \frac{TP}{TP + FN} \quad (4.12)$$

$$recall = \frac{TP}{TP + FP} \quad (4.13)$$

$$Valor-F = \frac{2 \times precisión \times recall}{precisión + recall} \quad (4.14)$$

$$TEE^{(n)} = \left| \sum_t y_t^{(n)} - \sum_t \hat{y}_t^{(n)} \right| \quad (4.15)$$

$$NEAP^{(n)} = \frac{\sum_t |y_t^{(n)} - \hat{y}_t^{(n)}|}{\sum_t y_t^{(n)}} \quad (4.16)$$

Se puede observar que la métrica de exactitud (*accuracy*), si bien es muy popular, no fue tomada en cuenta. Esto ocurre porque en el problema de desagregación de consumos eléctricos, en general, el tiempo de encendido de los electrodomésticos es muy menor comparado con el de apagado (ver Figura 4.3), resultando en un desbalance entre los estados apagados y encendidos. Por lo tanto, un método que prediga que todos los electrodomésticos están apagados, tendría un valor de exactitud alto pero a lo que compete a la desagregación sería muy malo.

4.6. Resultados

Esta sección presenta los resultados obtenidos en el análisis experimental del algoritmo presentado y los métodos base. A partir de una evaluación empírica, se tomó como *rule-of-thumb* los valores de parámetros $\delta = 100$, $d = 10$, $H = 500$ y $\varphi = 250$.

4.6.1. Instancias sin presencia de ruido ni variación de período de muestreo

Las Tablas 4.3–4.6 reportan los resultados numéricos del algoritmo DS, y los métodos base CO y FHMM, para las instancias #1 a la #4, sin presencia de ruido en el consumo y con período de muestreo constantes de cinco minutos.

<i>CO</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	292.18	2318.84	2767.39	6651.69	1118.64
NEAP	0.8663	0.7644	5.9284	2.6279	2.1975
precisión	0.8324	0.9863	0.7153	0.9758	0.8413
<i>recall</i>	0.5584	0.4827	0.0228	0.2301	0.2814
Valor-F	0.6684	0.6481	0.0442	0.3724	0.4218
<i>FHMM</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	306.46	3209.08	3399.42	5371.37	948.72
NEAP	0.8843	0.8367	6.8117	2.7134	2.3119
precision	0.7576	0.9817	0.7810	0.9768	0.5799
recall	0.5408	0.5078	0.0258	0.2377	0.2199
Valor-F	0.6311	0.6694	0.0500	0.3823	0.3188
<i>DS</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	23.87	0.00	0.00	0.00	29.67
NEAP	0.0218	0.0000	0.0000	0.0000	0.1497
precisión	0.9839	1.0000	1.0000	1.0000	0.9409
recall	0.9942	1.0000	1.0000	1.0000	0.9121
Valor-F	0.9891	1.0000	1.0000	1.0000	0.9263

Tabla 4.3: Resultados de CO, FHMM y DS para la instancia #1.

Los resultados en la Tabla 4.3 muestran cómo el algoritmo DS fue capaz de resolver el problema de desagregación en un escenario sin ambigüedades en el consumo de electrodomésticos, ni ruido en el consumo total. Los valores obtenidos para el Valor-F variaron entre 0,92 y 1,0 para DS, mientras que para CO y FHMM los valores no superaron 0,67 en cualquiera de los electrodomésticos (cercano a 0,6 para la heladera y el lavarropa, a 0,3 para el lavavajilla y *home theater*, y prácticamente nulo para la jarra eléctrica).

En la Tabla 4.4 se muestran los resultados de la instancia #2. Se puede observar cómo el Valor-F obtenido por el algoritmo DS, para electrodomésticos con ambigüedades de consumo, disminuye hasta en un 9% con respecto a la instancia #1. El Valor-F para el resto de los electrodomésticos permanece similar a la instancia #1. Por otro lado, CO muestra una disminución del 50% en la predicción de los electrodomésticos con consumo ambiguo, mientras que FHMM presenta valores similares a la instancia #1, exceptuando el caso de la jarra eléctrica cuyo Valor-F disminuye un 66%.

<i>CO</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	2228.32	1701.36	5595.52	7206.13	685.29
NEAP	1.0053	1.5412	9.8478	3.0491	1.6285
precisión	0.6973	0.8271	0.6715	0.9807	0.7781
recall	0.5123	0.2457	0.0111	0.1184	0.2907
Valor-F	0.5907	0.3789	0.0219	0.2113	0.4233
<i>FHMM</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	1401.84	962.88	7904.24	5016.79	431.27
NEAP	0.9007	1.1175	13.2448	2.1841	1.7049
precision	0.7687	0.9149	0.7007	0.9787	0.6649
recall	0.5573	0.4790	0.0084	0.2379	0.2850
Valor-F	0.6461	0.6288	0.0166	0.3828	0.3990
<i>DS</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	0.00	0.00	42.50	42.50	14.88
NEAP	0.0000	0.0000	0.1788	0.0473	0.1264
precisión	1.0000	1.0000	0.9416	0.9681	0.9460
recall	1.0000	1.0000	0.8866	0.9843	0.9289
Valor-F	1.0000	1.0000	0.9133	0.9761	0.9374

Tabla 4.4: Resultados de CO, FHMM y DS para la instancia #2.

Los resultados detallados en la Tabla 4.5 corresponden a la instancia #3 e indican que los registros de Valor-F para el algoritmo DS disminuyeron con respecto a los obtenidos en la instancia #1, 3 % para el lavarropa, 6 % para el lavavajilla y 25 % para la jarra eléctrica. Por el contrario, en el caso del *home theater*, el Valor-F registró un aumento del 6 % mientras que para la heladera permaneció igual. En cuanto al algoritmo CO, se registró una disminución del Valor-F de entre 42 % y 62 % para la desagregación del lavarropa, jarra eléctrica y lavavajilla, comparado con la instancia #1. Por otro lado, el Valor-F para el algoritmo FHMM disminuyó en todos los casos, con un máximo de 66 % para la jarra eléctrica, a excepción del *home theatre* que aumentó 33 %.

<i>CO</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	1690.42	2194.13	6298.06	6720.05	949.90
NEAP	0.9386	1.6483	12.1919	3.0818	1.7343
precisión	0.8217	0.8678	0.5876	0.9826	0.8432
recall	0.5754	0.2400	0.0073	0.1212	0.3250
Valor-F	0.6768	0.3760	0.0145	0.2157	0.4692
<i>FHMM</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	2069.24	1655.52	6273.13	6895.43	1561.12
NEAP	1.1036	1.2927	12.1388	3.1483	2.0024
precision	0.4318	0.9067	0.6387	0.9797	0.7645
recall	0.4512	0.3677	0.0087	0.1380	0.2942
Valor-F	0.4413	0.5232	0.0171	0.2419	0.4249
<i>DS</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	4.50	82.80	15.40	89.70	13.60
NEAP	0.0221	0.0668	0.5000	0.1092	0.0377
precisión	0.9893	0.9771	0.7372	0.9266	0.9845
recall	0.9886	0.9570	0.7566	0.9629	0.9780
Valor-F	0.9889	0.9670	0.7468	0.9444	0.9812

Tabla 4.5: Resultados de CO, FHMM y DS para la instancia #3.

Por último, la Tabla 4.6 muestra los resultados obtenidos para la instancia #4-5. Los valores resultantes demuestra la robustez de DS frente a los cambios menores del consumo de los electrodomésticos utilizados para entrenamiento y el utilizado para evaluación. El Valor-F registrado fue alto para todos los electrodomésticos desagregados, ubicándose entre 0,85 y 0,99. El electrodoméstico con menor Valor-F (0,85), al igual que en las instancias #2 y #3, fue la jarra eléctrica. Si se compara con el Valor-F obtenido en la instancia #1, la jarra eléctrica experimentó una disminución del 15 %, mientras que el resto de los electrodomésticos aumentaron/disminuyeron su valor en menos de un 2 %. Para el caso de CO, y comparando el Valor-F con respecto a los obtenidos en la instancia #1, disminuyó entre 13 % (heladera) y 55 % (lavarropa) en el caso de todos los electrodomésticos menos el *home theater*, que aumentó 8 %. Haciendo la misma comparación para FHMM, el Valor-F para la heladera y el lavavajilla prácticamente no varió (menor a 1,6 %), pero disminuyó considerablemente para el lavarropa (11 %), el *home theatre* (22 %) y la jarra eléctrica (67 %).

<i>CO</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	2543.42	2239.86	5208.68	7414.75	637.84
NEAP	0.9819	1.7824	9.6185	3.0921	1.8408
precisión	0.6597	0.7653	0.6533	0.9826	0.7895
recall	0.5202	0.1823	0.0121	0.1193	0.3205
Valor-F	0.5817	0.2944	0.0238	0.2128	0.4559
<i>FHMM</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	1829.65	699.35	8567.56	5080.58	560.53
NEAP	0.9218	1.1591	14.6967	2.2034	1.9148
precision	0.7209	0.8403	0.6971	0.9797	0.6961
recall	0.5453	0.4634	0.0083	0.2383	0.2931
Valor-F	0.6210	0.5974	0.0163	0.3834	0.4125
<i>DS</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	182.69	62.00	42.60	111.00	145.00
NEAP	0.0440	0.0142	0.3221	0.0821	0.2720
precisión	0.9985	1.0000	0.8066	0.9758	0.9666
recall	0.9957	0.9860	0.8984	0.9787	0.9166
Valor-F	0.9971	0.9930	0.8500	0.9773	0.9409

Tabla 4.6: Resultados de CO, FHMM y DS para la instancia #4-5.

4.6.2. Instancias sin presencia de ruido, con variación de período de muestreo

Las Tablas 4.6–4.8 presentan los resultados numéricos de DS, CO y FHMM, para las instancias #4-5, #4-10 y #4-15, en las que se varían los períodos de muestreo en 5, 10 y 15 minutos.

Al compararse la instancia con período de muestreo de 5 minutos con la de 15 minutos, el Valor-F para el algoritmo DS disminuye un máximo un 2,2% en todos los electrodomésticos, menos para la jarra eléctrico que disminuye su rendimiento 10,5%. En cuanto a la métrica TEE, se mantuvo en todos los casos por debajo de 183kW, y la NEAP por debajo de 0,52, con el peor valor registrado en la desagregación de la jarra eléctrica.

Si se tienen en cuenta los métodos CO y FHMM, en general los valores del Valor-F son menores que los obtenidos por DS. Si la comparación se hace para cada método y sobre la variación de los períodos de muestreo, se encuentran

<i>CO</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	1280.93	746.31	1607.04	4981.13	395.14
NEAP	1.0765	1.5259	6.8454	4.1716	2.1725
precision	0.4443	0.9471	0.6842	0.9764	0.792
recall	0.4132	0.2742	0.0121	0.0623	0.2694
Valor-F	0.4282	0.4252	0.0238	0.1171	0.402
<i>FHMM</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	581.89	280.14	3315.45	3020.55	460.4
NEAP	0.9491	1.2129	12.1956	2.6429	2.2588
precision	0.807	0.9188	0.594	0.9685	0.8431
recall	0.5167	0.4506	0.0068	0.2137	0.277
Valor-F	0.6301	0.6046	0.0135	0.3502	0.417
<i>DS</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	92.62	22.0	9.2	55.8	73.27
NEAP	0.0428	0.01	0.3321	0.0915	0.2675
precision	0.999	1.0	0.8195	0.9705	0.9703
recall	0.9965	0.9901	0.879	0.9743	0.9179
Valor-F	0.9977	0.995	0.8482	0.9724	0.9434

Tabla 4.7: Resultados de CO, FHMM y DS para la instancia #4-10.

resultados variados. Para CO, el rendimiento del Valor-F mejora un 70 % para el lavarropa, pero disminuye más de 50 % para la jarra eléctrica. Para el caso de FHMM, el rendimiento del Valor-F permanece similar en las diferentes instancias (desde variaciones despreciables a 10 %). En cuando al TEE, tanto para CO como para FHMM, en general se observan reducciones de hasta una décima parte. En el caso de la métrica NEAP, varía tanto incrementandose como decrementando, al rededor de 20 %.

Los gráficos en la Figura 4.4 sintetizan los resultados del Valor-F para los tres algoritmos evaluados, en las tres variaciones de períodos de muestreo; para la heladera y jarra eléctrica. Se graficaron estos dos electrodomésticos ya que presentan el mayor y menor tiempo de activación entre los electrodomésticos monitoreados. En general, DS obtiene mejores resultados que los obtenidos por CO o FHMM, destacandose la robustez al aumentar los períodos de muestreo en las instancias de prueba.

<i>CO</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	670.46	551.67	1477.47	2706.1	191.76
NEAP	1.0734	1.4782	9.2879	3.3986	2.0149
precision	0.5722	0.9498	0.6867	0.9712	0.5858
recall	0.4674	0.2714	0.0067	0.0993	0.2092
Valor-F	0.5145	0.4222	0.0133	0.1802	0.3084

<i>FHMM</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	295.81	255.21	2085.94	2036.45	209.06
NEAP	0.9604	1.2312	12.3144	2.6414	2.0098
precision	0.7612	0.9281	0.6867	0.951	0.6825
recall	0.4861	0.4459	0.0067	0.2021	0.2471
Valor-F	0.5933	0.6024	0.0132	0.3333	0.3629

<i>DS</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	61.38	20.0	8.3	59.7	48.73
NEAP	0.0441	0.0136	0.5236	0.1216	0.2657
precision	0.9982	1.0	0.7590	0.9424	0.9705
recall	0.9960	0.9866	0.7590	0.9703	0.9191
Valor-F	0.9971	0.9933	0.7590	0.9561	0.9441

Tabla 4.8: Resultados de CO, FHMM y DS para la instancia #4-15.

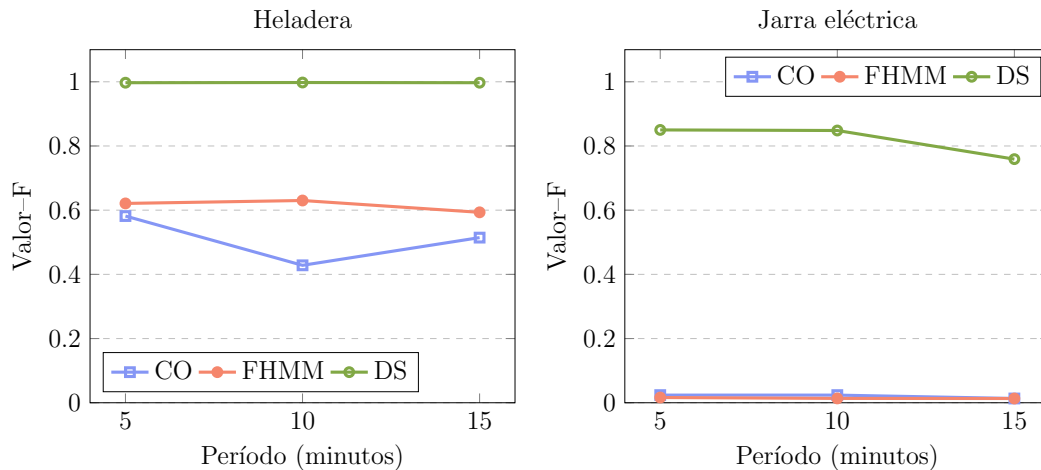


Figura 4.4: Valor-F de CO, FHMM, y DS para la heladera (izquierda) y la jarra eléctrica (derecha) en las instancias #4-5, #4-10, y #4-15.

4.6.3. Instancias con presencia de ruido y variación de período de muestreo

Las Tablas 4.9–4.11 reportan los resultados de DS, CO y FHMM evaluados en las instancias #5-5, #5-10 y #5-15. Estas instancias, aparte de variar los períodos de muestreo en 5, 10 y 15 minutos, agrega el consumo de electrodomésticos extra al consumo total agregado para generar ruido en la señal.

<i>CO</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	2217.76	1639.73	5802.36	8796.79	840.29
NEAP	1.0307	1.6097	10.1835	3.6616	1.9241
precision	0.6714	0.7493	0.6241	0.9865	0.7221
recall	0.4906	0.2591	0.0101	0.093	0.2525
Valor-F	0.567	0.3851	0.0199	0.17	0.3742
<i>FHMM</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	1080.94	1476.69	8210.2	5593.23	599.52
NEAP	0.8928	1.2786	13.6761	2.408	1.7635
precision	0.8592	0.8769	0.7263	0.9787	0.7177
recall	0.56	0.3932	0.0084	0.2127	0.262
Valor-F	0.6781	0.543	0.0166	0.3495	0.3839
<i>DS</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	0.5	36.0	20.0	20.0	10.16
NEAP	0.0001	0.0686	0.219	0.058	0.1269
precision	0.9999	0.9698	0.9051	0.9671	0.9428
recall	1.0	0.9619	0.8794	0.9747	0.9311
Valor-F	0.9999	0.9658	0.8921	0.9709	0.9369

Tabla 4.9: Resultados de CO, FHMM y DS para la instancia #5-5.

Con respecto a los resultados del Valor-F en el algoritmo DS, la comparación entre las instancias de 5 y 15 minutos de período de muestreo (ambos extremos), indica que el valor disminuye hasta un 13% para el lavarropa, jarra eléctrica y lavavajilla, mientras que para la heladera y el *home theatre* permanece con valores similares. Los valores de TEE se encuentran por debajo de 65 kW, mientras que los de NEAP por debajo de 0,53, para todos los electrodomésticos.

<i>CO</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	782.62	1164.42	2063.2	4600.55	325.91
NEAP	0.9966	1.6854	8.0059	3.9545	1.8397
precision	0.6627	0.9334	0.782	0.9606	0.7179
recall	0.4833	0.2249	0.0108	0.0729	0.2315
Valor-F	0.5589	0.3625	0.0213	0.1355	0.3501
<i>FHMM</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	373.1	635.38	3384.24	3238.49	483.63
NEAP	0.9853	1.3113	12.0121	2.8532	2.0886
precision	0.7949	0.9434	0.5865	0.9587	0.7987
recall	0.4918	0.3924	0.0064	0.1908	0.2492
Valor-F	0.6077	0.5543	0.0127	0.3183	0.3799
<i>DS</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	0.5	58.0	17.5	17.5	6.32
NEAP	0.0002	0.1232	0.3534	0.0925	0.1234
precision	0.9998	0.9252	0.8496	0.9469	0.946
recall	1.0	0.9503	0.8071	0.9601	0.9316
Valor-F	0.9999	0.9376	0.8278	0.9534	0.9388

Tabla 4.10: Resultados de CO, FHMM y DS para la instancia #5-10.

En las tres instancias evaluadas, el Valor-F registrado para DS fue mayor que los registrados para CO o FHMM. Además, DS presentó menores valores de TEE y NEAP que CO y FHMM, en todos los casos.

Para el caso de CO, el Valor-F disminuyó en 9 % para la heladera y 13,5 % para el *home theatre*, mientras que para el resto de los electrodomésticos permaneció con valores similares. Los valores de TEE experimenta, en general, una disminución de hasta un tercio, mientras que los valores de NEAP permanecen similares. En cuanto a FHMM, el Valor-F disminuye hasta un 17 % en todos los casos menos el del lavarropa (incrementa 3,5 %). Los resultados de TEE muestran una disminución de hasta un quinto en todos los casos, mientras que NEAP permanece igual.

Al compararse los resultados de todas las instancias de #4 y #5 (i.e., ausencia vs. presencia de ruido en las variaciones de período de muestreo), los resultados de DS varían según el caso:

<i>CO</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	439.5	764.0	1571.18	2728.08	261.97
NEAP	1.0791	1.5812	9.43	3.492	2.0609
precision	0.6438	0.9552	0.6506	0.9597	0.6819
recall	0.4326	0.2385	0.0067	0.0989	0.2149
Valor-F	0.5175	0.3817	0.0134	0.1794	0.3268
<i>FHMM</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	202.69	403.58	2172.2	2279.77	251.7
NEAP	1.0065	1.308	12.3287	2.9379	2.042
precision	0.7144	0.9349	0.6506	0.9625	0.7069
recall	0.4617	0.4043	0.0061	0.1723	0.2221
Valor-F	0.5609	0.5645	0.0121	0.2922	0.3381
<i>DS</i>					
<i>métrica</i>	<i>heladera</i>	<i>lavarropa</i>	<i>jarra eléctrica</i>	<i>lavavajilla</i>	<i>home theater</i>
TEE (kW)	0.5	30.0	65.0	65.0	4.0
NEAP	0.0002	0.1452	0.5301	0.1268	0.1241
precision	1.0	0.9376	0.8916	0.8991	0.9453
recall	0.9998	0.9189	0.6789	0.972	0.9317
Valor-F	0.9999	0.9281	0.7708	0.9341	0.9384

Tabla 4.11: Resultados de CO, FHMM y DS para la instancia #5-15.

- #4-5 vs. #5-5: el Valor-F disminuye hasta 3% para el lavarropa y el *home theatre*, incrementa 4,5% para la jarra eléctrica, y presenta valores similares para los demás dispositivos. Con respecto a los valores de TEE y NEAP, ambos resultados mostraron una disminución.
- #4-10 vs. #5-10: el Valor-F disminuye entre 0,5% y 9,1% para todos los electrodomésticos, a excepción de la heladera, que no varió. Con respecto a TEE, los valores disminuyeron para todos los electrodomésticos menos la jarra eléctrica. En cuanto a NEAP, todos los valores obtenidos resultaron menores a 0,35 para todos los casos, en ambas instancias.
- #4-15 vs #5-15: el Valor-F disminuye un 6,5% para el caso del lavarropa y el lavavajilla, aumenta hasta un 1,4% para la jarra eléctrica y el *home theatre*, y se mantiene en valores similares para la heladera. Los valores de TEE resultaron menores para la heladera y mayores para el resto de los electrodomésticos, pero siempre por debajo de 62kW. En cuanto a NEAP, los resultados muestran valores similares en ambas instancias.

El Valor-F para DS disminuyó en las tres instancias para los casos del lavarropa, jarra eléctrica y lavavajilla, mientras que para el resto de los casos permaneció en valores similares. Los resultados entre las instancias con presencia de ruido y variaciones en el período de muestreo (cada vez mayores) muestran una clara tendencia a disminuir el rendimiento en los electrodomésticos con menor tiempo de activación. Los gráficos de la Figura 4.5 sintetizan los resultados del Valor-F para CO, FHMM y DS, para los casos de la heladera y jarra eléctrica, en las instancias #5-5, #5-10, y #5-15. Se puede observar cómo el Valor-F de la heladera representa prácticamente una línea horizontal, o sea que su valor no varía demasiado entre instancias, mientras que para la jarra eléctrica la línea muestra un descenso a medida que aumenta el período de muestreo, o sea, que el Valor-F disminuye. El mismo comportamiento se esperaría para CO y FHMM, que se puede comprobar en el caso de la heladera pero no es comprobable para la jarra eléctrica, ya que los valores obtenidos son muy bajos.

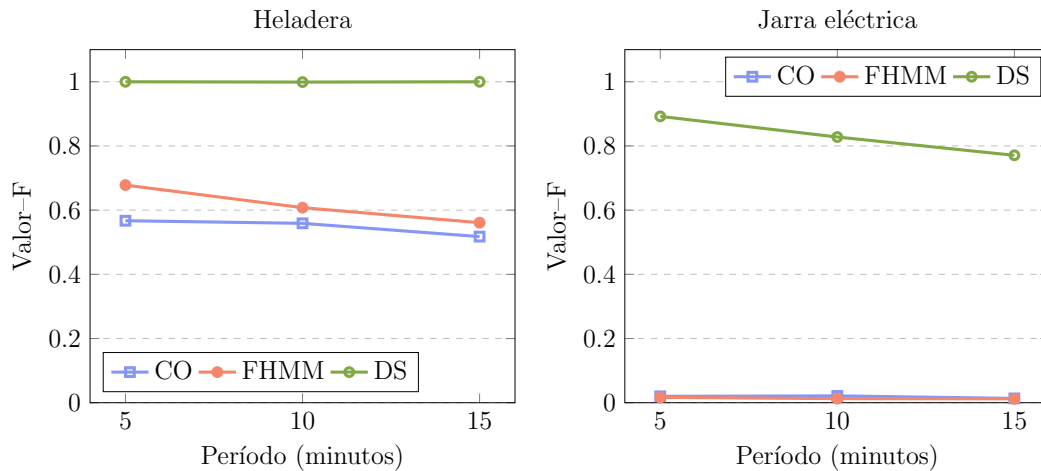


Figura 4.5: Valor-F de CO, FHMM y DS para los electrodomésticos heladera (izquierda) y jarra eléctrica (derecha) en las instancias #5-5, #5-10 y #5-15.

Los resultados obtenidos sugieren que la presencia de consumo de electrodomésticos extra, i.e., ruido en la señal de consumo total, afectan negativamente en el Valor-F del algoritmo DS, hasta en un 9,1%. El efecto negativo en este valor es más evidente en electrodomésticos con bajos niveles de activación. En los casos de CO y FHMM, los resultados sugieren que el Valor-F también es afectado negativamente, pero independientemente al tiempo de activación de los electrodomésticos. En general, en los dos algoritmos, a medida que el período de muestreo aumenta, el Valor-F disminuye hasta en un 13%.

Resumen

En todas las instancias, el algoritmo DS alcanza niveles satisfactorios de desagregación de consumo eléctrico de los electrodomésticos. Además, comparando con CO y FHMM, los resultados muestran una clara ventaja de DS.

El Valor-F para las instancias sin ruido y períodos de muestreo de 5 minutos (#1, #2, #3 y #4-5), en el caso de DS resulta en promedio hasta un 60% mejor que CO y hasta 57% mejor que FHMM. Al considerar las instancias con variaciones de período de muestreo y sin ruido (instancias #4-5, #4-10 y #4-15), las mejoras de DS frente a CO llegan en promedio a un 69%, mientras que frente a FHMM a 59%. Particularmente, en la instancia de mayor período de intervalo y para electrodomésticos con bajo tiempo de activación, los resultados de DS muestran mejoras de hasta 98% (e.g., la desagregación de la jarra eléctrica en instancia #4-15, comparado con CO).

En las instancias de evaluación con ruido y variación del período de muestreo (#5-5, #5-10 y #5-15), DS mostró mejoras entre 98% (mejor caso en instancia #5-15) y 43% (peor caso en instancia #5-5) sobre los resultados de CO, promediando 69%. Comparado con FHMM, DS mostró mejoras entre 98% (mejor caso en las tres instancias) y 32% (peor caso en instancia #5-5), promediando 61%.

De forma sistemática, DS obtuvo los valores más bajos tanto de TEE como de NEAP para todas las instancias. Los bajos resultados obtenidos en el caso de la jarra eléctrica en instancias con presencia de ambigüedad de consumo, periodos largos de muestreo o ruido, sugieren que los dispositivos con bajo tiempo de activación (0,5% del tiempo encendido, para la jarra eléctrica) afecta negativamente en los resultados. Cuanto más complejo sea el conjunto de datos, más datos de consumo se necesitan en el conjunto de datos de evaluación, especialmente para capturar el comportamiento de encendido/apagado de los dispositivos con menor tiempo de activación.

Los gráficos de la Figura 4.6 resumen los resultados del Valor-F obtenido por los algoritmos CO, FHMM y DS en todas las instancias con período de muestreo de 5 minutos, para la heladera y el lavarropa. Los electrodomésticos graficados presentan el mayor tiempo de activación (la heladera) y el tiempo promedio de activación (el lavarropa). En las instancias graficadas, se puede observar cómo DS obtiene los mayores valores de Valor-F en todas las instancias evaluadas.

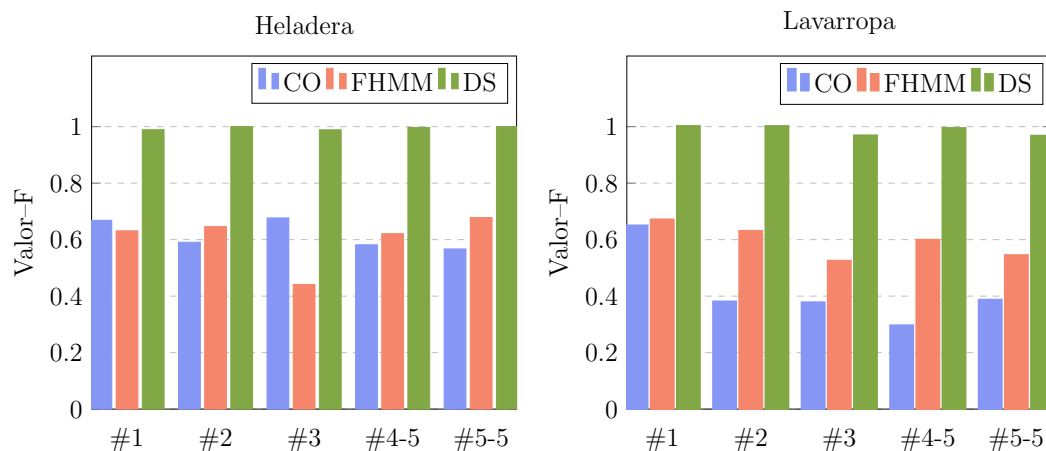


Figura 4.6: Valor-F de los algoritmos CO, FHMM, y DS para las instancias con período de muestreo de 5 minutos, para el electrodoméstico lavarropa.

4.7. Conclusiones

Este capítulo presenta un algoritmo para abordar el problema de desagregación de consumos eléctricos basado en patrones similares de consumo. El algoritmo recibe el nombre de Desagregación Simple (DS) y consta de dos etapas: entrenamiento y evaluación. Para el análisis experimental se crearon diferentes instancias de datos sintéticos basados en los datos del conjunto UK-DALE. El procedimiento para la creación de los datos fue presentado como parte de este trabajo. Las instancias de análisis propuestas cubren escenarios realistas como ambigüedad de consumos entre electrodomésticos, diferentes períodos de muestreo y presencia de consumos extra a los electrodomésticos desagregados (actuando como ruido en la señal total de consumo). Los resultados fueron comparados con los obtenidos por dos algoritmos base de la literatura revisada: CO y FHMM. DS obtuvo rendimientos satisfactorios en todas las instancias del problema evaluadas, superando en general el rendimiento de los dos algoritmos base. En particular, comparando con CO y FHMM, DS obtuvo valores de Valor-F hasta 64% mejores en las instancias de la #1 a la #4-5, hasta 69% mejores en instancias #4-5 a #4-15 y hasta 98% mejores en las instancias #5-5 a la #5-15. Los resultados del análisis experimental demuestran que DS es un algoritmo efectivo para abordar el problema de desagregación de consumos eléctricos en los escenarios planteados.

Capítulo 5

Conjunto de datos de consumo eléctrico residencial en Uruguay

Este capítulo presenta el conjunto de datos de consumo eléctrico residencial de hogares Uruguayos, compuesto por tres subconjuntos con diferentes características de consumo (consumo total del hogar, consumo de termotanque y consumo desagregado de un conjunto de electrodomésticos). El conjunto de datos está destinado al estudio de diferentes comportamientos de los clientes y a la detección de patrones de consumo en los hogares. Estos estudios permiten implementar mejoras en el servicio de distribución de la energía, la elaboración de mejores planes comerciales y de generación eléctrica, metodologías para facturación personalizada, entre otras. El trabajo presentado en este capítulo resultó en la generación del conjunto de datos ECD-UY, su presentación pública y la elaboración de un artículo científico de datos con su descripción y validación técnica ([Chavat et al., 2022a](#)).

5.1. Introducción

En Uruguay, el servicio de energía eléctrica es provisto por la empresa estatal Administración Nacional de Usinas y Transmisiones Eléctricas (UTE). El país ha sido reconocido por el desarrollo en las energías renovables y el gran porcentaje de producción eléctrica proveniente de éstas fuentes ([Watts, 2015](#); [Barbosa et al., 2017](#)). Uruguay cuenta con 3,4 millones de habitantes, de los cuales 1,3 viven en Montevideo, su capital. Según los últimos datos disponibles (julio de 2019) la empresa UTE brinda energía eléctrica a 1.498.164

clientes, de los cuales el 90,5 % son del tipo residencial. Además, según datos reportados por el [Ministerio de Industria, Energía y Minería \(2018\)](#), la tasa de electrificación de viviendas ocupadas es de 99,8 % (99,9 % en zonas urbanas y 98,9 % en zonas rurales), por lo cual la electrificación es considerada universal.

Según datos reportados en el año 2020, en promedio UTE provee de 235 kWh por mes a sus clientes residenciales, 251 kWh a los clientes de Montevideo y 226 kWh a los clientes del resto del país ([Usinas y Trasmisiones Eléctricas, 2019](#)). El electrodoméstico con mayor tasa de consumo residencial es el termostanque con 27 % en promedio y presente en el 87,3 % de los hogares, según datos del [Instituto Nacional de Estadística \(2019\)](#). El funcionamiento de este electrodoméstico, que puede ser visto como un elemento de almacenamiento térmico, podría ser manejado en su totalidad por la empresa proveedora de electricidad, aprovechando los momentos de abundancia de generación eléctrica con fuentes renovables.

El conjunto de datos presentado en este capítulo, nombrado ECD-UY a partir de *Electricity Consumption Dataset of Uruguay*, consta de tres subconjuntos de consumo eléctrico y un conjunto de información de clientes. El primer subconjunto, *Total Household Consumption* (THC), corresponde al consumo total de 110.953 hogares distribuidos por todo el país, con una frecuencia de muestreo de 15 minutos y una extensión de 23 meses. El segundo subconjunto, *Electric Water Heater consumption* (EWH), corresponde al consumo eléctrico de 268 termostanques ubicados en diferentes ciudades del país, con una frecuencia de muestreo de un minuto y una duración de casi 3 meses. El tercer subconjunto, *Disaggregated Appliance Consumption* (DAC), corresponde al consumo eléctrico desagregado de un conjunto de electrodomésticos y el consumo total de 9 hogares, con una frecuencia de muestreo de un minuto y una duración de 21 días. Por último, el conjunto de información de clientes presenta información geográfica, de tarifas y de potencia contratada, de 110.994 clientes. La Tabla 5.1 resume las características de los datos presentados.

En las secciones subsiguientes se presenta el diseño del sistema de recolección de datos, los métodos de comunicación utilizados, el pre-procesamiento y la limpieza de los registros, la estructura resultante, la validación técnica y los posibles usos de ECD-UY. La Figura 5.1 presenta una vista general de las etapas transcurridas para la creación del conjunto de datos.

<i>subconjunto</i>	<i>hogares</i>	<i>consumo total</i>	<i>consumo desag.</i>	<i>período</i>	<i>fecha inicio*</i>	<i>fecha fin*</i>
THC	110.953	si	no	15 min.	1/01/2019	3/11/2020
EWH	268	si	-	1 min.	2/07/2019	26/10/2020
DAC	9	si	si	1 min.	27/08/2019	16/09/2019
Info. de clientes	110.994	-	-	-	-	-

Tabla 5.1: Resumen de los tres subconjuntos de consumo eléctrico y el conjunto de información de clientes, contenido en ECD-UY. *Los períodos pueden variar, dependiendo del cliente.

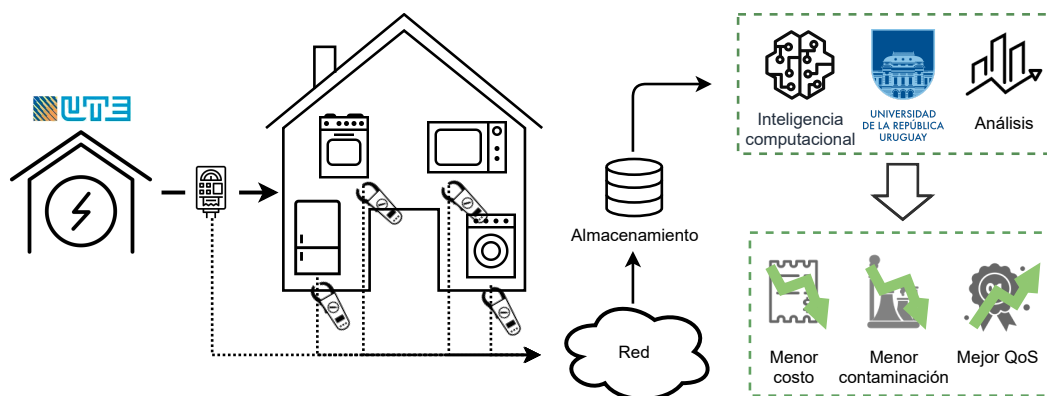


Figura 5.1: Vista general del sistema de recolección, comunicación, y procesamiento de los datos de consumo eléctrico residencial.

5.2. Metodología

En esta sección se brindan detalles sobre la metodología aplicada en la recolección, comunicación, pre-procesamiento y limpieza de los datos. Una parte de las metodologías descritas fueron aplicadas por la empresa UTE, mientras que la otra formó parte de este trabajo de tesis. Las subsecciones siguientes se subdividen en una parte por cada subconjunto de datos.

5.2.1. Recolección de datos

Los procesos de recolección de datos fueron independientes para cada subconjunto de ECD-UY y se llevaron a cabo enteramente por la empresa UTE. A continuación se brindan detalles sobre los procesos, dispositivos y metodologías aplicadas en esta etapa, para cada subconjunto de datos.

Consumo eléctrico total de los hogares

Los datos de consumo eléctrico total de los hogares fueron recolectados por los sistemas de telemetría de UTE. Al momento de escribir este trabajo, el sistema de telemetría es integrado por aproximadamente 600.000 medidores inteligentes (*smart meters*), que cubren el consumo de 40% de los clientes residenciales. La empresa tiene como objetivo, para el año 2023, llegar a cubrir el 100% de los clientes residenciales.

Los dispositivos utilizados en la recolección son marca Kaifa, modelos MA110P (mayormente), MA309P o MA309D. Todos los modelos siguen los estándares IEC 62052-11, 62052-11/21/23 y 62056-21/46/53/61/62 (*i.e.*, estándares internacionales que aseguran la calidad de dispositivos de medición de electricidad). Los dispositivos permiten la medición de la energía activa y reactiva, la tensión, la corriente eléctrica y la frecuencia. Además, cuentan con interconexión a través de un modem PLC/RF/FPRS/3G/RS-485 con velocidad de hasta 9600 bps. En la siguiente subsección se brindan mayores detalles sobre la comunicación de los datos. El período de muestreo de los dispositivos es configurable en 5/10/15/30/60 minutos, siendo 15 minutos el valor por defecto. En la página web oficial de la marca, <https://kaifametering.com/>, se encuentra mayor cantidad de información sobre estos dispositivos. La Figura 5.2a muestra un medidor Kaifa, modelo MA110P, previo a ser instalado por la empresa.

El intervalo de muestreo utilizado fue de 15 minutos, que corresponde a la configuración por defecto que traen de origen los medidores. A pesar de que este subconjunto no cuenta con períodos de muestreo tan cortos como el de los demás subconjuntos (un minuto), reúne el consumo eléctrico de 400 veces más hogares. El manejo de los datos generados en tal cantidad de hogares, con un período de muestreo menor (*e.g.*, un minuto), implica lidiar con grandes cantidades de datos, lo cuál requiere mayor infraestructura por parte de la empresa, lo cuál no ha sido considerado al momento.

Consumo eléctrico de termotanques

En el marco de un plan comercial que tuvo por objetivo estudiar los patrones de consumo eléctrico de los hogares, la empresa UTE recolectó datos de consumo de termotanques en 268 hogares. En todos los casos, los termotanques monitoreados fueron del tipo acumulación con resistencia eléctrica. El termo-

tanque fue definido como el dispositivo a estudiar ya que está presente en la mayoría de los hogares Uruguayos y es un electrodoméstico considerado de alto impacto en el total del consumo (27 % en promedio, llegando a picos de 37 %). Para motivar a los clientes a participar de este plan, se les ofreció un incentivo económico correspondiente a 6 USD por mes, lo que equivalía a 50 kWh de consumo eléctrico al momento, y la funcionalidad de poder controlar de forma remota (vía aplicación web/móvil) el encendido del dispositivo. Para evitar sesgos, los clientes a los que se les ofreció participar del plan cumplían con los siguientes criterios: estaban ubicados en subestaciones representativas con consumo histórico, ingresos y nivel socio-económico promedio, o correspondía a otras estaciones con muestras de consumos relevantes y clases de tarifas diferentes. Además, los hogares debían estar ubicados dentro de ocho departamentos de Uruguay (Canelones, Montevideo, Salto, Paysandú, Maldonado, Río Negro, Colonia o San José), seleccionados especialmente para desplegar el plan comercial.

Los dispositivos para el monitoreo instalados por la compañía fueron del tipo interruptores inteligentes (*smart switches*) de la marca Sonoff, modelo IM160810001. Estos dispositivos tienen la forma de un sobre-enchufe y se instalan en los enchufes que utilizan los electrodomésticos a monitorear. El modelo utilizado incluye un chip (modelo HLW8012) de monitoreo de energía eléctrica en una sola fase, que permite medir un máximo de 3500 W, un rango de tensión de 90-250 V AC, un máximo de corriente de 15 A, con un rango de frecuencia de 80-160 MHz, bajo el estándar 802.11 b/g/n. En la Figura 5.2b se observa uno de los interruptores inteligentes, marca Sonoff, utilizados por UTE para el monitoreo del consumo eléctrico de los termotanques.

Consumo eléctrico desagregado por electrodomésticos

La recolección de los datos de consumo eléctrico desagregado por electrodoméstico se llevó a cabo en el marco de un plan piloto para estudiar los patrones de uso de electrodomésticos en hogares. Se monitorearon nueve hogares ubicados en Montevideo y Canelones, algunos en áreas urbanas y otros sobre la costa sur Uruguaya. El criterio de selección de los electrodomésticos a monitorear y su posterior etiquetado, estuvo a cargo de los clientes participantes e incluyó el microondas, el lavarropa, la heladera, el termotanque, el horno, el aire acondicionado, el deshumidificador y el secarropa. Tres de los

nueve hogares monitoreados correspondieron a apartamentos y los restantes seis a casas. Los hogares presentaban tres ocupantes en promedio, dos adultos y un niño. La Tabla 5.2 resume las características de los hogares monitoreados en el plan piloto.

<i>id. del hogar</i>	<i>departamento</i>	<i>tipo de hogar</i>	<i>ubicación</i>	<i>adultos/niños</i>
1	Montevideo	Casa	Urbana	2 / 0
2	Canelones	Casa	Costa	2 / 0
3	Montevideo	Apartamento	Urbana	2 / 1
4	Montevideo	Apartamento	Urbana	2 / 1
5	Montevideo	Casa	Urbana	2 / 2
6	Canelones	Casa	Costa	2 / 2
7	Montevideo	Casa	Urbana	2 / 2
8	Montevideo	Casa	Urbana	2 / 1
9	Montevideo	Apartamento	Urbana	2 / 0

Tabla 5.2: Características de los hogares del subconjunto de datos de consumo eléctrico desagregado por electrodomésticos.

Para la recolección de los datos se utilizaron pinzas amperimétricas de la marca Jiangyin Spark, modelo XH-SCT-T10A que registraron el consumo de los diferentes dispositivos en períodos de muestreo de un minuto. La Figura 5.2c muestra una de las pinzas amperimétricas instaladas por UTE para la medición de los consumos eléctricos de electrodomésticos.

Información de clientes

La información de los clientes de los subconjuntos de consumo total y consumo de termotanques forma parte de la base de datos de clientes de UTE. Durante la recolección de los datos por parte de la empresa, se anonimizaron los registros de forma que se pudieran asociar los registros de consumo a los clientes sin permitir identificar su identidad. Para los clientes de los nueve hogares del subconjunto de consumos desagregados de electrodomésticos se encuestó a los clientes para obtener los datos de los hogares y sus ocupantes. En total, se cuenta con la información de 110.994 clientes.



(a) Medidor inteligente marca Kaifa, modelo MA110P. (b) Dispositivo Sonoff, modelo IM160810001 (c) Pinza amperimétrica, marca Jiangyin Spark, modelo XH-SCT-T10A

Figura 5.2: Dispositivos instalados por la empresa UTE, para la recolección de datos de consumo eléctrico en los hogares. Imágenes brindadas por UTE.

5.2.2. Comunicación de datos

Los datos recolectados fueron transmitidos a servidores centrales de la empresa UTE, a través de diferentes mecanismos. Esta subsección presenta los mecanismos y procesos de comunicación utilizados para cada subconjunto de datos.

Consumo eléctrico total de los hogares

Los valores de consumo eléctrico total recolectados por los medidores inteligentes se envían y almacenan en la infraestructura avanzada de medición (*Advanced Metering Infrastructure*, AMI) montada por UTE. La AMI es un componente fundamental en las redes eléctricas modernas y a parte de recibir los datos y almacenarlos, establece una comunicación bidireccional con el medidor inteligente. La conexión con el medidor permite, entre otras cosas, ejecutar tareas de control de dispositivos con el fin de optimizar el uso de la electricidad. Además, la AMI actúa de nexo entre los sistemas de facturación y el sistema de manejo de la demanda.

En la mayoría de los hogares (86%), la comunicación entre los medidores y la AMI se establece por protocolo 3G. En los casos en que el medidor se encuentra fuera de la cobertura 3G (14% de los medidores), se utilizan conexiones cableadas mediante puerto RS-485 o PLC, a un *hub* conectado a la red de Internet (*e.g.*, fibra óptica) de la empresa nacional de telecomunicaciones, ANTEL. La Figura 5.3 presenta un diagrama que sintetiza la arquitectura y los procesos del sistema de comunicación para los datos de consumo eléctrico total de los hogares.

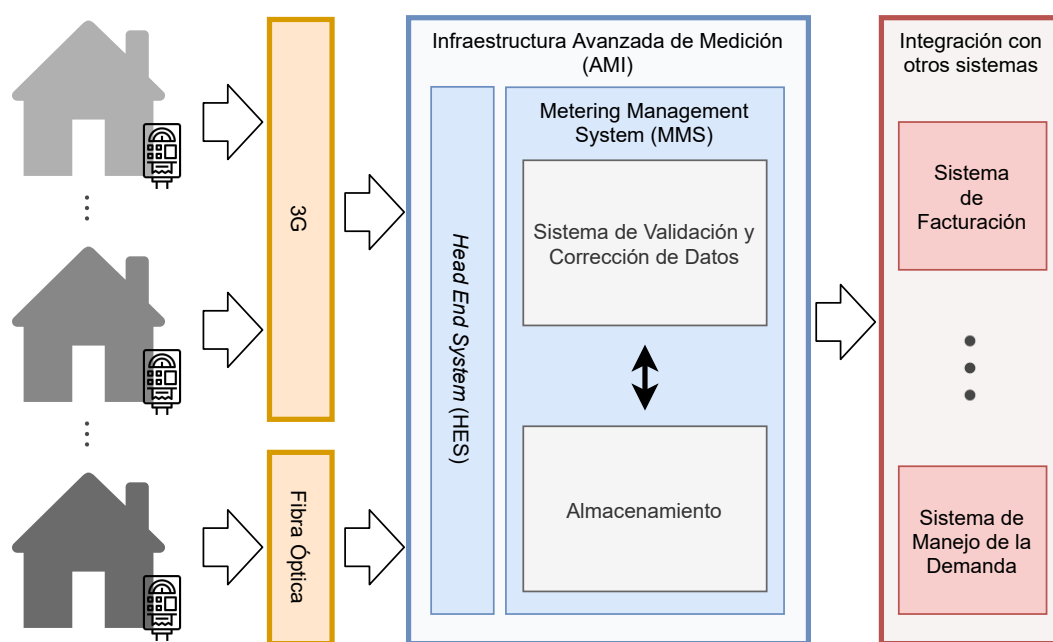


Figura 5.3: Arquitectura y procesos del sistema de comunicación de los datos de consumo total de los hogares.

Consumo eléctrico de termotanques

Los registros de consumo de termotanque fueron comunicados primero a través de la conexión a Internet del hogar y luego a través de una conexión a Internet 2G del propio medidor. Al comienzo del plan comercial, los medidores no contaban con módem y utilizaban la conexión inalámbrica de cada hogar para establecer un canal bidireccional por el cual enviar (*e.g.*, mediciones) y recibir (*e.g.*, comandos de encendido/apagado) datos. Luego, los medidores fueron sustituidos por modelos más modernos capaces de establecer conexiones 2G por sus propios medios. De esta forma, la comunicación pasó a ser independiente del cliente y permitió mejorar la robustez del sistema, evitando

gran parte de la pérdida de paquetes causadas por interrupciones en la conexión. Para asegurar los datos transmitidos se utilizaron conexiones a través de túneles VPN. El responsable del manejo de la conexión fue el chipset STM32 integrado al medidor, que ejecutaba un código implementado por UTE en lenguaje C. Los datos recolectados se transmitieron con una frecuencia de un minuto, utilizando el protocolo de red MQTT. En la infraestructura de UTE, una plataforma de manejo de demanda implementada utilizando el *framework* Spring Boot y el lenguaje Java, recibió los datos.

La arquitectura y los procesos empleados en el sistema de comunicación son presentados en el diagrama de la Figura 5.4.

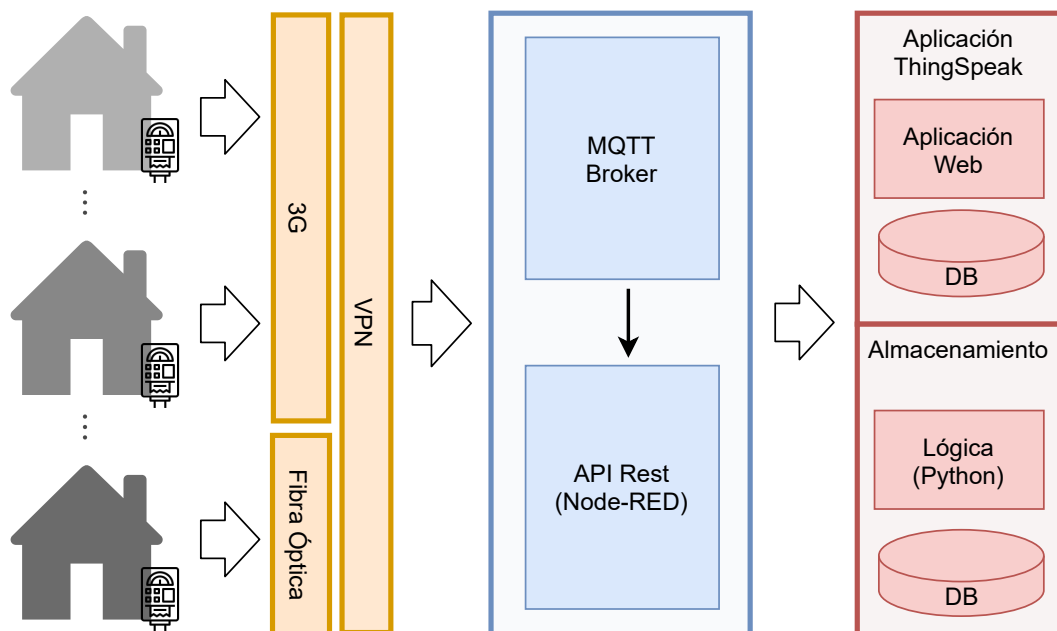


Figura 5.4: Arquitectura y procesos del sistema de comunicación empleado en la recolección de los datos de consumo eléctrico de termotanques.

Consumo eléctrico desagregado por electrodomésticos

El consumo registrado para cada electrodoméstico se transmitió, utilizando el protocolo Zigbee 3.0, desde las pinzas amperimétricas a un *gateway/hub* instalado en cada hogar. Luego, utilizando la conexión inalámbrica del hogar, se enviaron los registros a un servicio externo de terceros que centralizó los datos de consumo de los 9 hogares y les asoció metadatos del hogar del que procedían. Los datos fueron descargados regularmente por UTE desde el sistema tercero y almacenados en sistemas de la empresa.

Las pinzas amperimétricas midieron y enviaron los registros de consumo con un período de muestreo de un minuto. Para los casos en los que la conexión se encontraba interrumpida, las pinzas contaban con una memoria *buffer* que permitió retener los paquetes hasta que la conexión fuera restablecida.

5.2.3. Pre-procesamiento de datos

Los datos fueron pre-procesados con la intención de dejar una base común de partida para los futuros usuarios del conjunto de datos. De esta manera, es posible comparar los resultados de métodos que abordan un mismo problema y utilizan los datos de ECD-UY para su procesamiento. Para esta etapa se implementaron varios *scripts* Unix y Jupyter *notebooks*.

En términos generales, las transformaciones del pre-procesamiento fueron aplicadas con los siguientes fines: i) estandarización de las unidades, formato de fechas, nombre de columnas y archivos, ii) eliminación de columnas innecesarias y iii) unificación de dos o más columnas. Las transformaciones aplicadas a los datos de consumo y a la información de clientes se detalla en las siguientes subsecciones y el formato original de los datos es descripto en el Anexo 1.

Consumo eléctrico total de los hogares

En la AMI de UTE, los datos fueron corregidos y validados antes de ser derivados al sistema de manejo de datos. Las correcciones se aplicaron al detectar anomalías o pérdida de los registros. Una vez que los datos eran procesados por la AMI, se dejaban disponibles para ser usados por los módulos de análisis de la empresa. Los archivos entregados por UTE no cumplían un criterio claro, por lo que todos los registros fueron ordenados de forma cronológica y separados en un archivo por mes, logrando archivos de un tamaño más manejable (entre 780 MB y 7.1 GB). Además, entre los archivos originales se encontraba uno que reflejaba la relación entre el medidores y los clientes. Luego de validar que para cada cliente existe un único medidor, se reemplazó el identificador del medidor por el identificador del cliente, eliminando la necesidad del archivo, simplificando la estructura y reduciendo el tamaño total del subconjunto. Finalmente, se eliminaron los registros que no tenían un cliente correspondiente en el conjunto de clientes y se cambió el formato de la marca de tiempo de ISO-8601 a *epoch time*, lo que redundó en una reducción del tamaño del archivo de 20%.

Consumo eléctrico de termotanques

En una primera instancia, UTE calculó los valores medios de potencia, voltaje y potencia instantánea a partir de los valores medidos por el interruptor inteligente Sonoff (específicamente, por el chip HLW8012 de una fase), utilizando un *firmware* desarrollado en lenguaje C y ejecutado por el chipset STM32. Luego, a partir de los datos entregados y para aquellos consumos de termotanques con los que se cuenta con información del cliente, se reemplazó el identificador del medidor por el identificador del cliente y la columna de marca de tiempo se formateó a *epoch time*. Ambos cambios redujeron el largo de los registros y por lo tanto el tamaño total del subconjunto. Por último, se detectó la presencia de termotanques monitoreados con una cantidad de registros menor a la de un día (1.440 registros, uno por minuto) y se procedió a eliminarlos. Como resultado final de esta etapa, se logró una drástica reducción en el tamaño del subconjunto, pasando de 6,6 GB a 2,4 GB.

Consumo eléctrico desagregado por electrodomésticos

UTE recolectó la información de los clientes y la relacionó a los identificadores de los electrodomésticos. Luego, la empresa relacionó el consumo de los electrodomésticos a identificadores de medidores y estos a identificadores de clientes. Con el fin de simplificar las relaciones entre los datos, se modificó de forma conveniente su estructura. A los datos que describen los electrodomésticos, se les agregó el identificador del cliente, eliminando el identificador del medidor. Originalmente, las diferentes medidas de consumo (*e.g.*, energía activa/reactiva, potencia activa/reactiva y cada una de sus facetas) de cada electrodoméstico y del consumo total eran registradas como columnas de una única entrada e identificadas por una marca de tiempo. Este formato resultaba en una gran cantidad de características por entrada, por lo que cada característica fue transformada en un registro independiente, lo que facilitó la manipulación de los datos y el análisis del consumo de los electrodomésticos. Además, los datos de consumo fueron separados en dos colecciones, una para el consumo total y otra para el consumo por electrodoméstico. En la subsección 5.3 se brinda mayor información sobre la estructura resultante.

En la colección de información de electrodomésticos, se eliminaron los registros que no contaban con registros de consumo y también se eliminaron aquellos registros de consumos sin su correspondiente información de electrodoméstico. En total, se eliminaron 34 electrodomésticos y 1.163.714 registros de consumo.

Los datos originales contaban también con mediciones de termostatos internos a algunos de los hogares. Visto que la calidad de los datos de los termostatos era mala (presentaban una gran cantidad de *gaps* y *outliers*) y que no se relacionaban directamente con la temática del conjunto de datos, se decidió eliminar esos registros.

5.3. Estructura de datos

Esta sección describe la estructura final de cada subconjunto, resultado de aplicar el pre-procesamiento detallado en la subsección 5.2.3.

Consumo eléctrico total de los hogares

El subconjunto de consumo total de los hogares está integrado por un archivo por mes con el consumo eléctrico total, rotulado `consumption_data_`**AAAAMM**`.csv`. En el nombre del archivo, el texto **AAAAMM** es sustituido por el año (AAAA) y el número del mes (MM) al que corresponden los registros. La Tabla 5.3 lista las características registradas, el nombre de la columna a la que corresponden y una breve descripción. Los registros de este subconjunto se relacionan con la información de clientes mediante el valor de de la columna *id*, correspondiente al identificador de cliente.

<code>consumption_data_</code> AAAAMM <code>.csv</code>		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
<code>datetime</code>	texto	Marca de tiempo del registro, en formato <i>epoch time</i>
<code>id</code>	número	Valor único para identificar al cliente
<code>value</code>	número	Valor de la energía activa, en kWh

Tabla 5.3: Descripción de los registros correspondientes al subconjunto de datos de consumo eléctrico total de los hogares.

Consumo eléctrico de termotanques

El subconjunto de consumo eléctrico de termotanques está conformado por dos archivos, ambos con datos de consumo eléctrico. Por un lado, el archivo `consumption_data_customers.csv` almacena los datos de consumo de hogares que cuentan con un cliente identificado y por lo tanto cuentan con un identificador de cliente. Las entradas de este archivo se relacionan con la información de cliente mediante el valor de la columna `customer_id`. Por otro lado, el archivo `consumption_data_timers.csv` almacena los datos de consumo de hogares identificados por un número de medidor y que no cuentan con información de cliente. La separación planteada permite procesar el consumo de termotanques de clientes de forma fácil, sin necesidad de aplicar filtros previamente al procesamiento. La Tabla 5.4 presenta las características contenidas en los archivos del subconjunto de consumo eléctrico de termotanques, el tipo de dato y una breve descripción.

<code>consumption_data_customers.csv</code>		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
<code>datetime</code>	texto	Marca de tiempo de los registros, en formato <i>epoch time</i>
<code>customer_id</code>	número	Valor único que identifica al cliente
<code>power</code>	número	Potencia instantánea, en watts (W)
<code>voltage</code>	número	Tensión instantánea, en Volts (V)

<code>consumption_data_timers.csv</code>		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
<code>datetime</code>	texto	Marca de tiempo de los registros, en formato <i>epoch time</i>
<code>meter_id</code>	número	Valor único que identifica al medidor
<code>power</code>	número	Potencia instantánea, en watts (W)
<code>voltage</code>	número	Tensión instantánea, en Volts (V)

Tabla 5.4: Descripción de los registros correspondientes al subconjunto de datos de consumo eléctrico de termotanques.

Consumo eléctrico desagregado por electrodomésticos

El subconjunto de consumo eléctrico desagregado por electrodoméstico está integrado por tres archivos. Por un lado, `appliances.csv` presenta información sobre los electrodomésticos monitoreados y por otro lado `appliance_consumption_data.csv` y `total_consumption_data.csv` presentan los registros de consumo eléctrico por electrodoméstico y total, respectivamente. La Tabla 5.5 brinda mayores detalles sobre las características registradas en cada archivo. Los dos archivos de consumo eléctrico se relacionan con el de la información de los electrodomésticos mediante el valor de identificador de medidor (columna `meter_id`), mientras que la información de los electrodomésticos se relaciona con el conjunto de información de clientes mediante el valor del identificador de cliente (columna `id`).

Información de clientes

La información disponible de los clientes de los tres subconjuntos se encuentra en un único archivo rotulado `customers.csv` y consiste en el identificador del cliente, las características del servicio eléctrico contratado (e.g., la tensión, el tipo de tarifa y la potencia contratada) y las características georeferenciales (e.g., el departamento y la sección/segmento/zona censal) del hogar. Cada registro del conjunto se relaciona con los demás subconjuntos a partir del identificador de cliente (columna `id`). La Tabla 5.6 enumera los datos contenidos en el archivo, detallando el nombre de columna asignado, el tipo de valor y una breve descripción del mismo.

<code>appliances.csv</code>		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
<code>id</code>	número	Valor único que identifica al cliente
<code>appl_meter_id</code>	número	Valor único que identifica al electrodoméstico
<code>appl_desc</code>	texto	Descripción del electrodoméstico (tal cuál fue ingresada por el cliente)
<code>appl_type</code>	texto	Tipo de electrodoméstico (basado en nilm-metadata (Kelly y Knottenbelt, 2014))
<code>total_consumption_data.csv</code>		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
<code>datetime</code>	texto	Marca de tiempo de los registros, en formato <i>epoch time</i>
<code>meter_id</code>	número	Valor único que identifica al electrodoméstico
<code>aenergy</code>	número	Energía activa acumulada en el último minuto, en Wh
<code>aenergy_ph{1,2,3}</code>	número	Energía activa acumulada en el último minuto, fases 1, 2 y 3, en Wh
<code>reenergy</code>	número	Energía reactiva, en VARh
<code>reenergy_ph{1,2,3}</code>	número	Energía reactiva acumulada en el último minuto, fases 1, 2 y 3, en VARh
<code>apower</code>	número	Potencia activa, en W
<code>apower_ph{1,2,3}</code>	número	Potencia activa, fases 1, 2 y 3, en W
<code>rpower_ph{1,2,3}</code>	número	Potencia reactiva, fases 1, 2 y 3, en VARh
<code>current_ph{1,2,3}</code>	número	Corriente eléctrica, fases 1, 2 y 3, en A
<code>pfactor</code>	número	Factor de potencia (eficiencia energética)
<code>pfactor_ph{1,2,3}</code>	número	Factor de potencia, fases 1, 2 y 3
<code>voltage_ph{1,2,3}</code>	número	Tensión, fases 1, 2 y 3, medido en V
<code>appliance_consumption_data.csv</code>		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
<code>datetime</code>	texto	Marca de tiempo de los registros, en formato <i>epoch time</i>
<code>meter_id</code>	número	Valor único que identifica al electrodoméstico
<code>aenergy</code>	número	Energía activa acumulada, en Wh
<code>apower</code>	número	Potencia activa, en W
<code>apower_ph{1,2,3}</code>	número	Potencia activa, fase 1, 2 y 3, en W

Tabla 5.5: Descripción de los registros del subconjunto de consumo eléctrico desagregado por electrodomésticos.

<code>customers.csv</code>		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
<code>id</code>	número	Valor único que identifica al cliente/hogar
<code>tension</code>	texto	Tensión a la cual el cliente se conecta a la red eléctrica (230V ó 400V)
<code>tariff</code>	texto	Tipo de tarifa contratada. Valores posibles: TCB, consumo básico; TRS, consumo residencial simple; TRD, consumo residencial doble banda; TRT, consumo residencial triple banda; TCBT, descuento al TCB por plan social para hogares de bajos ingresos
<code>power</code>	número	Potencia contratada, en W
<code>department</code>	número	Departamento en el cual se ubica el hogar
<code>section</code>	número	Sección censal en la cual se ubica el hogar
<code>segment</code>	número	Segmento censal en el cual se ubica el hogar
<code>zone</code>	número	Zona censal en la cual se ubica el hogar

Tabla 5.6: Descripción de los registros correspondientes al conjunto de información de clientes, contenido en el archivo `customers.csv`.

5.4. Validación

Esta sección presenta una serie de validaciones realizadas a los subconjuntos de consumo eléctrico, que dan soporte a la calidad técnica de los datos.

5.4.1. Consumo eléctrico total de los hogares

Los 110.953 hogares que componen el subconjunto de consumo eléctrico total se distribuyeron entre los 19 departamentos de Uruguay y fueron monitoreados en promedio durante 539,2 días. Cada día cuentan con un promedio de 95,2 registros, de los 96 que debería tener si no hubiesen existido pérdida de registros (*i.e.*, uno cada 15 minutos). El subconjunto fue pre-procesado de tal forma que todos los hogares monitoreados cuentan con información de clientes.

Para evaluar los registros por día, se ejecutó un experimento. Primero, se definió un criterio de completitud que establece que un día con al menos 95 % de los registros que debería tener (*i.e.*, al menos 91 de los 96) se considera como completo. Luego, se establecieron intervalos de cantidad de registros por día y se contaron cuántos días corresponden a cada intervalo. Los resultados obtenidos muestran que el 97,16 % de los días están completos. La Tabla 5.7 muestra para cada intervalo la cantidad de días y su correspondiente porcentaje.

<i>intervalo de registros</i>	<i>cantidad de días</i>	<i>%</i>
(91, 96]	58.122.666	97,16 %
(86, 91]	355.269	0,59 %
(81, 86]	92.690	0,15 %
(76, 81]	289.154	0,48 %
(72, 76]	76.332	0,13 %
(0, 72]	887.232	1,48 %
<i>total</i>	59.823.343	100 %

Tabla 5.7: Cantidad de días por intervalo de cantidad de registros.

En cuanto a la cantidad de hogares y los días monitoreados, se ejecutó un experimento que los relaciona. Primero se establecieron intervalos cada 60 días, a excepción del menor intervalo que abarca de 0 a 360 días. Luego, se contó la cantidad de hogares cuyos días monitoreados se encuentran en cada uno de los intervalos, primero se tuvieron en cuenta todos los días y luego solo aquellos que cumplen el criterio de completitud previamente establecido. Los resultados arrojaron que en ambas instancias del experimento, el intervalo de cantidad de días con mayor cantidad de hogares (23,11 % y 27,08 %) abarca desde 480 a 540 días. Además, se constató que la mayoría de los hogares (98,83 % y 97,17 %) cuentan con entre 360 y 690 días monitoreados. En la Tabla 5.8 se observan los resultados de este experimento. Por último, se calculó el promedio de días monitoreados en los hogares para ambas instancias (sin filtrar y filtrando los días por el criterio de completitud) y se constató que el número decrece de 539,2 a 525,2 días.

Al aplicar el criterio de completitud, la cantidad de hogares desciende de 110.953 a 96.565. Esto se explica por la existencia de hogares que no cuentan con al menos un día con más de 91 registros. Esta observación queda reflejada en la Figura 5.5, que presenta un histograma de la cantidad de clientes por intervalo de cantidad de días, diferenciado por cada instancia del experimento.

Como validación general de los consumos de electricidad registrados, se calculó el promedio anual por hogar y se comparó con los reportados por la empresa en su reporte anual ([Usinas y Trasmisiones Eléctricas, 2019](#)). El resultado muestra que ambos promedios son cercanos, 3.132 kWh, según los datos de ECD-UY, frente a 3.012 kWh de acuerdo al reporte de UTE. Por lo tanto, la diferencia entre ambos es de 3,8 %, lo cual valida la buena calidad de los datos registrados y de la representatividad de los hogares considerados

<i>intervalo de días</i>	<i>todos los días</i>		<i>días completos</i>	
	<i>hogares</i>	<i>%</i>	<i>hogares</i>	<i>%</i>
(660, 690]	18.400	16,58 %	0	0,00 %
(600, 660]	14.105	12,71 %	14.820	15,35 %
(540, 600]	21.010	18,94 %	19.032	19,71 %
(480, 540]	25.643	23,11 %	26.148	27,08 %
(420, 480]	18.630	16,79 %	20.069	20,78 %
(360, 420]	11.870	10,70 %	13.768	14,26 %
(0, 360]	1.294	1,17 %	2.728	2,83 %
<i>total</i>	110.952	100 %	96.565	100 %

Tabla 5.8: Cantidad de días por cantidad de hogares, contando todos los días y solo los que cumplen el criterio de completitud (i.e., cuentan con al menos 95 % de los registros).

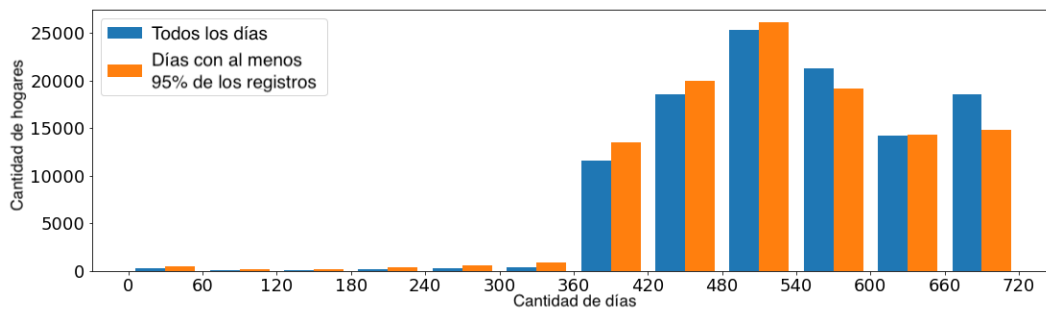


Figura 5.5: Histograma de cantidad de hogares por intervalo de cantidad de días monitoreados, teniendo en cuenta todos los días y solo los que cumplen el criterio de completitud (i.e., cuentan con al menos 95 % de los registros).

para el monitoreo de consumo. Además, utilizando la información de clientes se calculó la cantidad de registros cuyo consumo sobrepasa la potencia contratada, que resulta en un número despreciable que representa al 0,20 % de los registros.

Por último, utilizando todos los clientes del subconjunto se calculó el consumo promedio en bloques de 15 minutos. El resultado, que se observa en la Figura 5.6, demuestra que a pesar de que existen diferentes perfiles de consumo, se sigue un patrón de consumo global que incluye un valle durante las horas de la noche (mínimo alrededor de las cuatro de la madrugada) y dos picos de consumo durante el día, el primero cercano al mediodía y el segundo sobre las nueve horas de la noche. Estos resultados validan la consistencia de los registros al compararlos con los perfiles de consumo reportados en el reporte anual de la empresa.

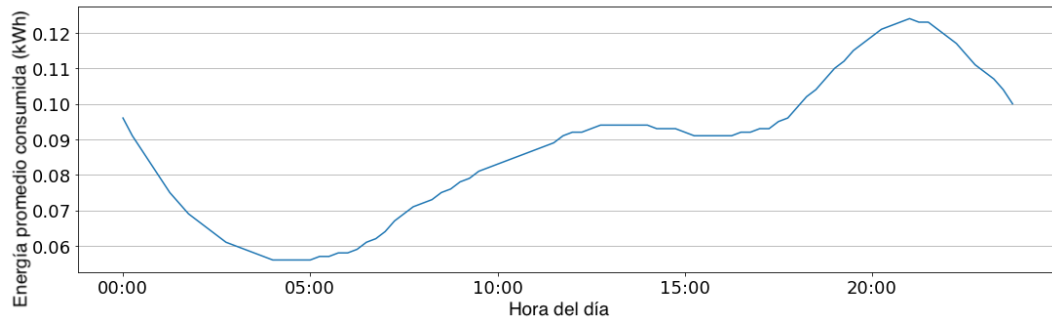


Figura 5.6: Consumo promedio durante un día, calculado cada 15 minutos.

5.4.2. Consumo eléctrico de termostatos

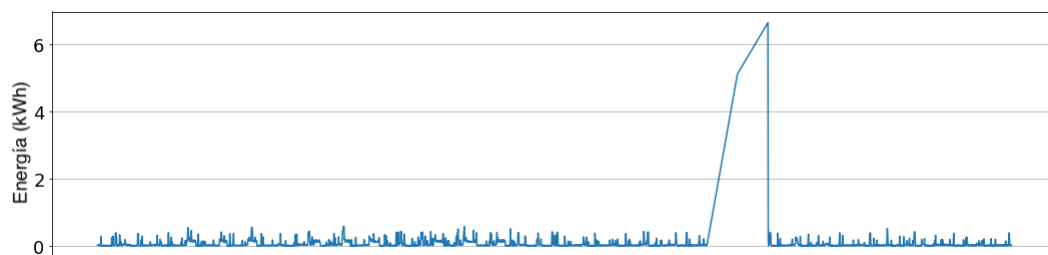
De los 268 hogares en los cuales se monitoreó el consumo eléctrico de termostatos, 166 (62 %) cuentan con información de clientes. Al cruzar los datos de los 166 hogares que cuentan con información de clientes con los datos de consumo eléctrico total de hogares (filtrado por el rango de fechas en el que se monitorean los dispositivos), el resultado indica que 135 hogares (50,4 %) cuentan con registros en el subconjunto de consumo eléctrico total. Por lo tanto, 50,4 % de los termostatos monitoreados podrían ser relacionados con el cliente y el consumo total del hogar.

Con el propósito de evitar procesar valores anómalos (*e.g.*, *outliers*, valores negativos) y evitar los *gaps* como parte de la validación, los registros fueron filtrados por un criterio de percentil y luego se rellenaron los registros faltantes con un criterio de relleno.

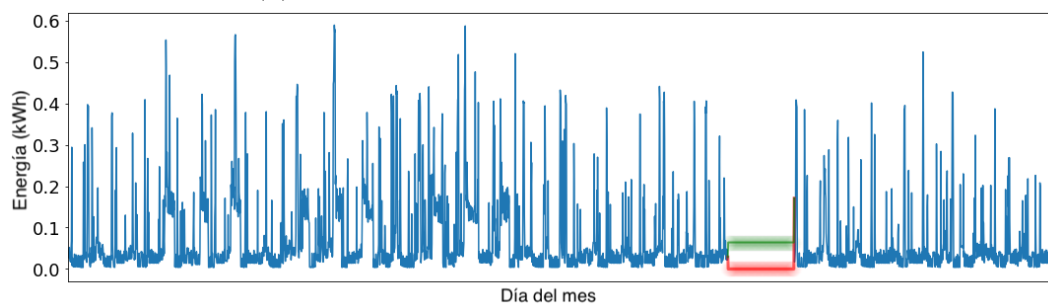
Para filtrar los valores anómalos en el consumo total, se calculó el percentil 99 de forma transversal a todos los hogares. El valor del percentil calculado, 2,186 kWh, se aplicó como límite superior y todos los registros con valores mayores fueron eliminados. Para el consumo de los termostatos se calculó el percentil 97 individualmente para cada uno de los hogares. El cálculo por hogar permitió preservar las características de los dispositivos y las incidencias del contexto en el que operan (*e.g.*, clima, modelo del termostato, aislamiento térmico, etc). Luego, para cada hogar se fijó el percentil como límite superior y se eliminaron los registros con valores mayores.

En cuanto al relleno de valores faltantes y *gaps*, se aplicó el criterio único de asignar valores cero en los lugares que faltaba el registro. Es necesario aclarar que existe un sinnúmero de criterios aplicables, algunos más sencillos (*e.g.*, asignación de ceros, valor medio o máximo) y otros más complejos (*e.g.*,

aplicación de una función de interpolación de los valores inmediatos anteriores/posteriores a los faltantes, recreación de valores utilizando redes neuronales antagónicas) (Weerakody et al., 2021). La Figura 5.7 presenta dos gráficas de consumo, antes y después de aplicar las correcciones (rellenado de *gaps* y eliminación de *outliers*) en el consumo total de un mes (Setiembre de 2019) para un determinado cliente (#69806). Se aplicaron dos criterios de relleno básico, asignación de consumo cero (en rojo) y asignación de consumo medio (en verde). De la misma forma, la Figura 5.8 muestra el consumo eléctrico de un termostato durante un día (1^{ro} de Agosto de 2019), para un cliente (#115609), antes y después de aplicar las correcciones. Los criterios aplicados fueron el de asignación de consumo cero (en rojo) y el de asignación de la potencia nominal del dispositivo (color verde).



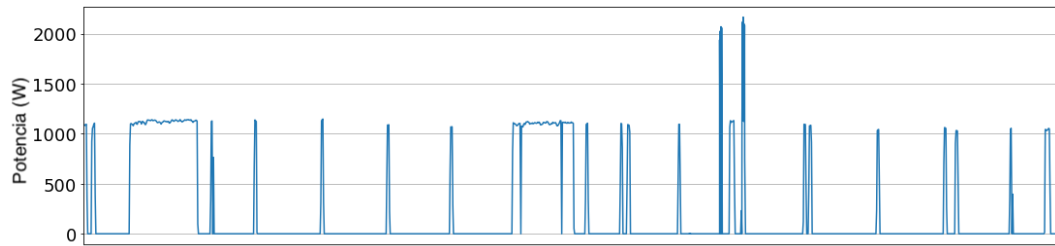
(a) Consumo total antes de aplicar correcciones



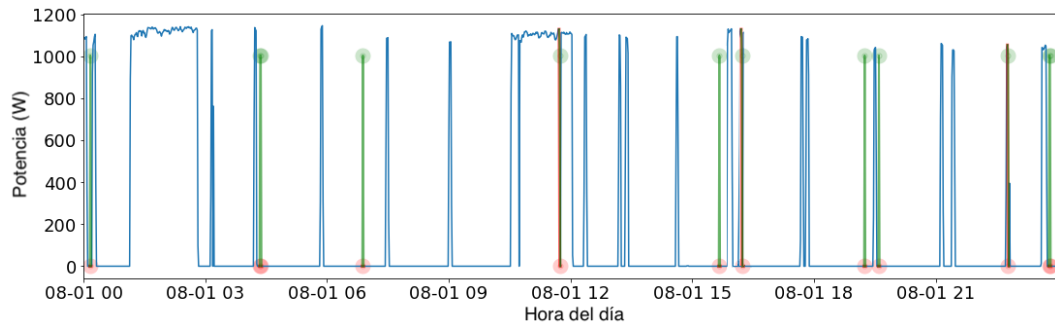
(b) Consumo total luego de aplicar correcciones, usando consumo cero (rojo) o consumo medio (verde)

Figura 5.7: Ejemplo de un mes de consumo total de un hogar (cliente #69806), con presencia de *gaps* y valores *outliers* y su corrección (línea roja y verde).

Como se explicó previamente, la forma de tratar las anomalías puede variar y no necesariamente debe ser única. En general, cómo tratarlas es una decisión que depende del propósito de uso de los datos y compete al usuario final. Por esta razón, los resultados del pre-procesamiento realizado para la validación, no son incluidos como resultado final del conjunto de datos.



(a) Consumo de termostato, antes de aplicar correcciones



(b) Consumo de termostato, luego de aplicar correcciones utilizando valor cero (rojo) o potencia nominal (verde)

Figura 5.8: Ejemplo de 24 horas de consumo de un termostato (cliente #115609) y dos posibles correcciones a los *gaps* y valores *outliers*: utilizando consumo cero (círculos rojos) o consumo igual a la potencia nominal del dispositivo (círculos verdes).

Con el fin de conocer las características de los *gaps* presentes en el subconjunto, se midió la longitud de cada uno y se calculó el tiempo promedio, resultando en una extensión de 4 minutos y 19 segundos. El tiempo promedio equivale a 4,3 registros consecutivos, para el período de muestreo del subconjunto.

En cuanto al consumo del termostato respecto al consumo total, se calculó la potencia promedio requerida en períodos de 15 minutos y se comparó con la potencia promedio total. Los resultados demuestran la relevancia del consumo del termostato en los hogares, donde en promedio significa un 27% del consumo total, llegando a picos de 35%. La Figura 5.9 grafica los resultados obtenidos en forma de potencia promedio total y potencia promedio del termostato a lo largo de un día. Los datos utilizados para la gráfica son los registros de 135 hogares con marcas de tiempo desde el 7 de Julio de 2019 al 11 de Noviembre de 2019.

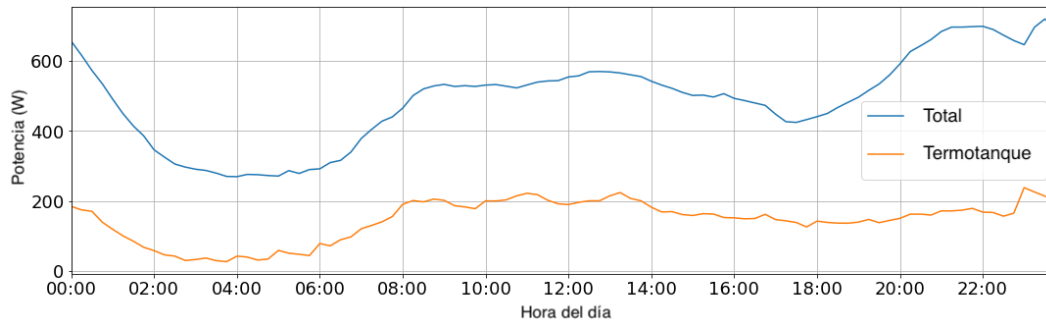
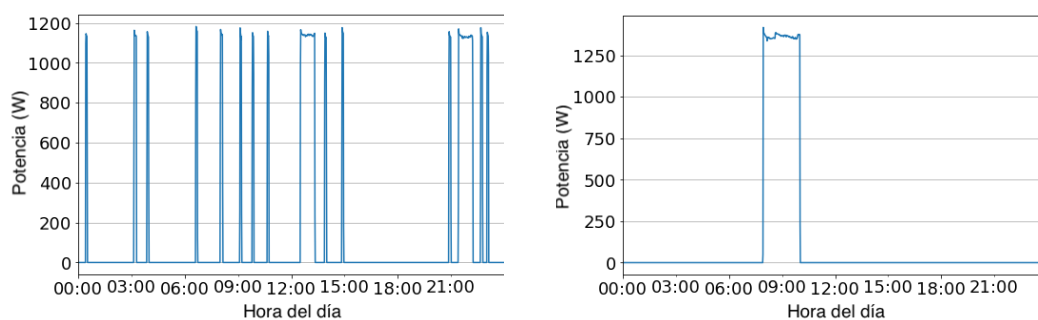


Figura 5.9: Potencia promedio total y del termotanque durante un día, utilizando datos de 135 hogares con rango de fechas desde el 15/07/2019 al 30/11/2019.

En lo que respecta únicamente al consumo del termotanque, durante la validación se identificaron dos patrones de consumo predominantes. Uno corresponde a los termotanques que consumen energía eléctrica a lo largo del día, con encendidos intermitentes para mantener la temperatura del agua (cortos) y otros (más largos) para calentar el agua que ingresó al tanque luego de algún uso. Este patrón de consumo se observa en la Figura 5.10a, que grafica el consumo de un día para uno de los termotanques que lo cumplen. El otro patrón de consumo corresponde a aquellos dispositivos que consumen electricidad solo por algunos períodos a lo largo del día. Se asume que estos dispositivos se encienden manualmente o son manejados por algún interruptor automático que los enciende en una franja horaria preestablecida. La Figura 5.10b grafica el consumo de un día para uno de los dispositivos que cumple este patrón.



(a) termotanque con encendidos intermitentes (#4729, 30 de Octubre de 2019). Consumo diario promedio 3.4 kWh.

(b) termotanque encendido solo para usos específicos (#65374, 2 de Noviembre de 2019). Consumo diario promedio 2.8 kWh.

Figura 5.10: Ejemplo de patrones de consumo de termotanques.

En cuanto a la distribución geográfica de los hogares con consumo de termotanque, 118 de los 135 se ubican en el departamento de Montevideo. Más de la mitad se concentran en una única zona censal, como se observa en la Figura 5.11.

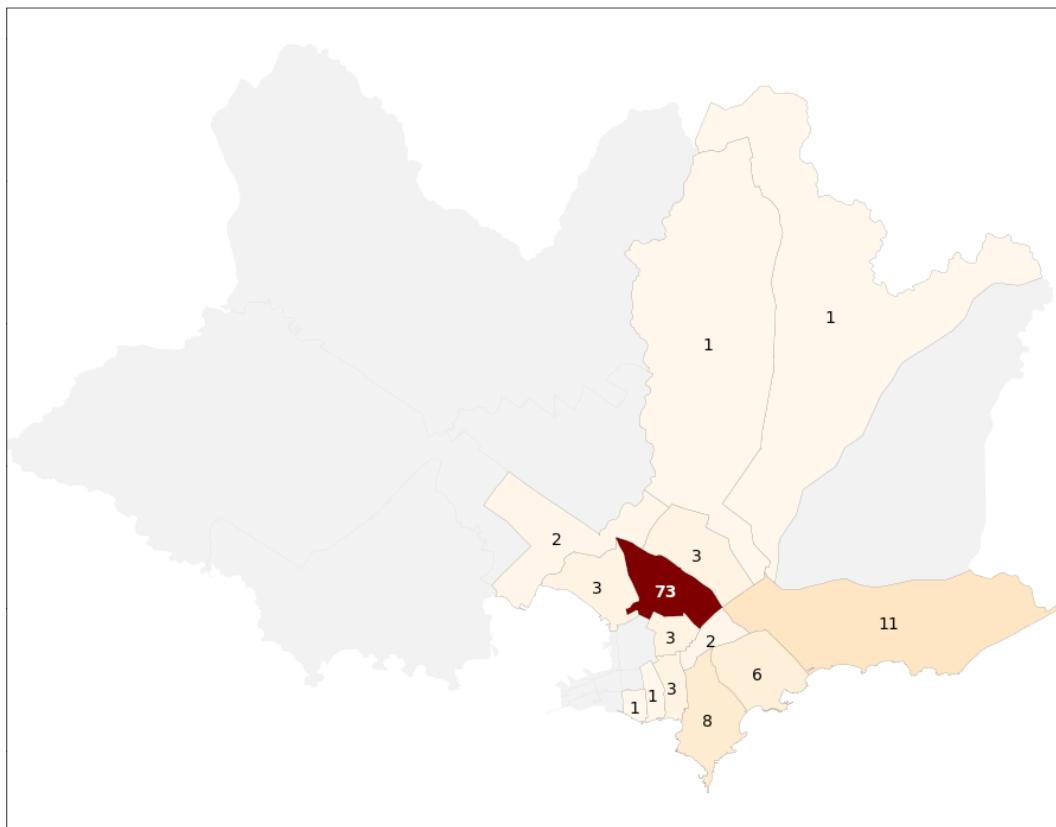


Figura 5.11: Distribución de los hogares con datos de consumo de termotanque e información de clientes, en el departamento de Montevideo, Uruguay.

5.4.3. Consumo eléctrico desagregado por electrodomésticos

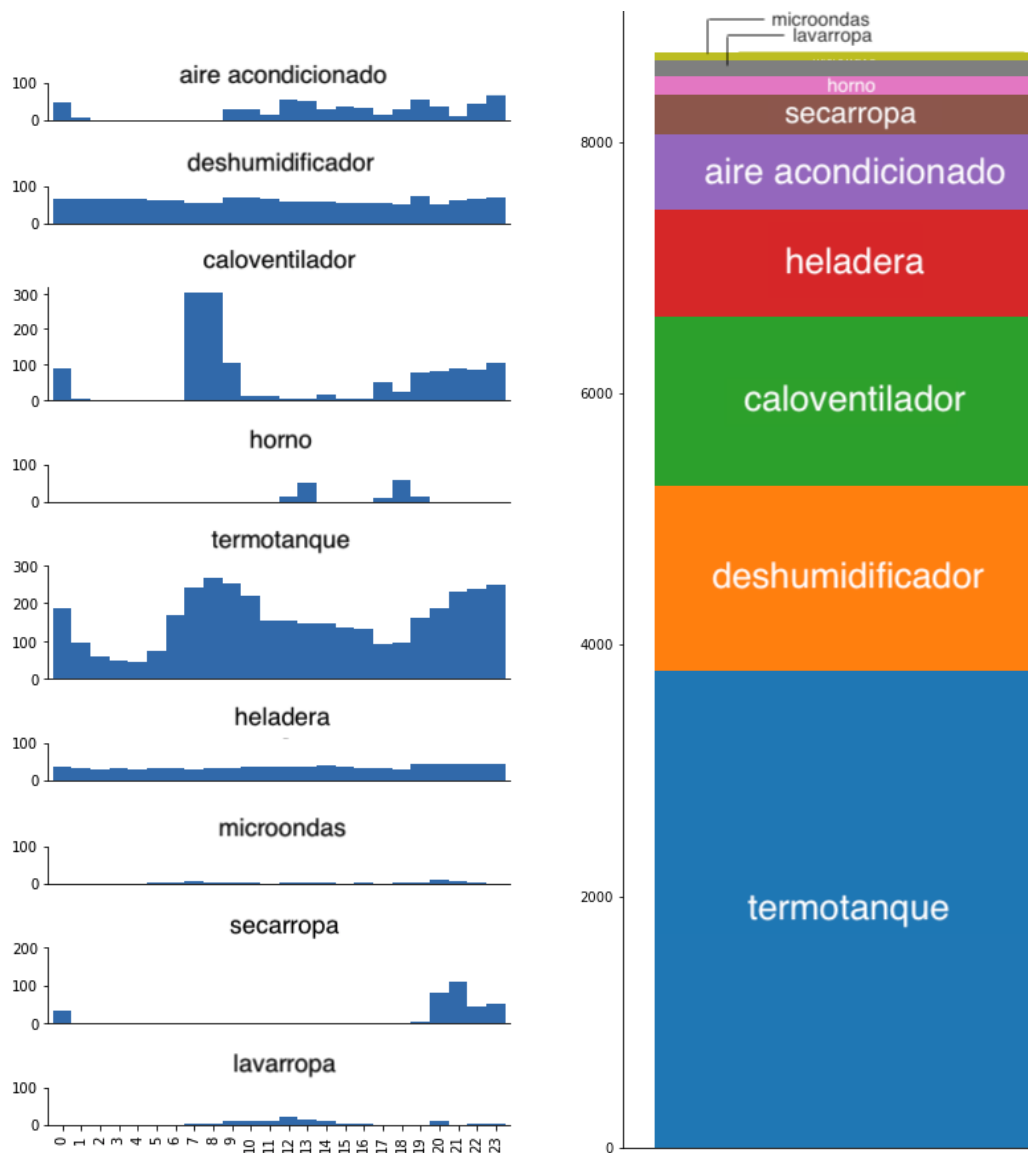
El total de los hogares del subconjunto de consumos desagregados por electrodomésticos cuenta con información de clientes y los períodos monitoreados duran en promedio 19 días. Diferentes experimentos detectaron la presencia de *gaps* cuyo largo está directamente relacionado al electrodoméstico y se extienden de una a casi siete horas. La pérdida de estos datos corresponde, principalmente, a fallas en la comunicación y recepción de los registros. La Tabla 5.9 brinda información sobre la cantidad de electrodomésticos monitoreada, la cantidad y duración media de los *gaps* y la duración media que fueron monitoreados.

<i>electrodoméstico</i>	<i>cantidad</i>	<i>cantidad media de gaps</i>	<i>duración media de gaps</i>	<i>duración media de monitoreo</i>
caloventilador	2	177.0	01:50:59	19 días 19:05:00
horno	1	120.0	01:07:33	19 días 19:05:00
secarropa	1	116.0	01:40:46	19 días 19:05:02
lavarropa	5	114.0	02:17:09	19 días 16:07:59
termotanque	5	107.6	01:30:54	19 días 19:04:59
microondas	4	104.5	01:37:33	19 días 19:04:59
heladera	5	81.2	06:38:48	19 días 13:05:59
deshumidificador	1	74.0	03:25:36	18 días 04:41:58
aire acondicionado	3	73.3	01:50:51	19 días 19:04:59

Tabla 5.9: Información detallada sobre el monitoreo de los electrodomésticos y los *gaps* detectados.

Con el fin de evitar el procesamiento de valores anómalos en los experimentos de validación, se calculó el percentil 99 de consumo y todos los valores menores a cero o mayores al percentil fueron identificados como *outliers*. El percentil fue calculado para cada electrodoméstico, permitiendo preservar las características del contexto del hogar y de cada dispositivo (*e.g.*, modelo, usos). Luego, todos los *outliers* identificados y los *gaps* detectados fueron rellenados con valores de consumo cero.

Para entender el funcionamiento de los diferentes electrodomésticos y su participación relativa al consumo total, se procesaron dos gráficos. El primer gráfico en la Figura 5.12a se conforma de histogramas de potencia media demandada por cada electrodoméstico, para cada hora del día. Este gráfico evidencia los tiempos de uso específicos para algunos electrodomésticos (*e.g.*, horno, lavarropa, termotanque, caloventilador), mientras que en otros se identifica un uso casi constante durante el día (*e.g.*, deshumidificador, heladera). El segundo gráfico en la Figura 5.12b muestra una gráfica de barras apiladas que acumula el consumo eléctrico medio (Wh) de cada electrodoméstico a lo largo de un día. Se observa cómo el termotanque, deshumidificador y el caloventilador tienen la mayor cuota de participación en el consumo total.



(a) Histograma de potencia media demandada por hora y electrodoméstico (b) Barras acumuladas de la demanda media por electrodoméstico

Figura 5.12: Potencia media demandada por hora (histograma) y acumulada en un día (barras apiladas), para cada electrodoméstico del subconjunto de consumo desagregados.

5.5. Repositorio de datos

El conjunto resultante ECD-UY fue almacenado en un repositorio público de Figshare y está disponible para su descarga en <https://doi.org/10.6084/m9.figshare.c.5428608.v1> (Chavat et al., 2022b). Los archivos en el repositorio cuentan con un prefijo que identifica el subconjunto al que corresponden:

THC (del inglés *Total Household Consumption*) corresponde al subconjunto de consumo total de los hogares, EWH (del inglés *Electric Water Heater consumption*) corresponde al subconjunto de consumo de termotanques y DEC (del inglés *Disaggregated Electricity Consumption by appliances*) corresponde al subconjunto de consumo desagregado por electrodomésticos. Los archivos se encuentran en formato CSV ([Shafranovich, 2005](#)) y sus columnas corresponden a las detalladas en la sección [5.3](#).

Para reducir el tamaño de almacenamiento necesario y facilitar la descarga de los datos, los archivos de gran tamaño fueron comprimidos. Una versión reducida de estos archivos fue incluida sin comprimir, con el fin de permitir previsualizar su contenido en la web de los mismos. Los archivos reducidos tienen formato CSV, su contenido corresponde a las primeras 10000 líneas del archivo original y se nombran igual al original pero anteponiendo el prefijo ‘REDUCED’.

5.6. Utilización de los datos

Para utilizar los datos de ECD-UY se puede emplear cualquier software/librería que cuente con manejo de archivos CSV. Durante el transcurso de esta tesis, tanto para el pre-procesamiento como para los posteriores análisis, se utilizaron herramientas Unix tales como `awk`, `split`, `sort` o `uniq`, y `Jupyter notebooks` con lenguaje Python y librerías tales como `Pandas` y `Numpy`. Además, para ejecuciones en el *cluster* se utilizó la librería `Dask` ([McKinney, 2012](#); [Rocklin, 2015](#)), que permite trabajar con archivos de tamaño mayor a la cantidad de memoria RAM disponible y paralelizar los procesamientos. El procesamiento de los datos fue ejecutado, en gran medida, utilizando la plataforma del centro de supercomputación nacional, `Cluster-UY` ([Nesmachnow y Iturriaga, 2019](#)).

Dependiendo del tipo de procesamiento a aplicar, puede resultar útil transformar el formato de CSV a Apache Parquet ([Vohra, 2016](#)). El formato Apache Parquet tiene una estructura orientada a columnas y sus archivos son binarios comprimidos que pueden ser utilizados para el procesamiento eficiente en plataformas como Apache Hadoop, o similares

Previo a la utilización de los datos, se debería tomar una decisión de cómo tratar los *gaps* presentes. Se sugiere aplicar la siguiente regla: si el largo del *gap* es menor a 7,5 minutos (i.e., la mitad del mayor período de muestreo entre los

subconjuntos), se considera como corto, sino, se considera como largo. A los *gaps* cortos se les rellena aplicando algún método que tenga en cuenta valores previos y posteriores, como por ejemplo una interpolación o un promedio de los valores previos y siguientes. En cuanto a los *gaps* largos, se podría interpretar que el dispositivo monitoreado permaneció sin consumo por un período largo de tiempo (*i.e.*, apagado, en el caso de un electrodoméstico) y por lo tanto los valores se podrían rellenar con cero.

En cuanto a los períodos de muestreo irregulares, librerías como Pandas cuentan con funciones específicas de remuestreo de series temporales (método *resample* en Pandas) que utilizan una función de recreado/rellenado de los valores. Para el remuestreo es recomendable utilizar ventanas de tiempo regulares definidas por los períodos de muestreo teóricos de los subconjuntos de datos y asignar al intervalo regularizado el valor máximo o promedio de los valores comprendidos en cada intervalo regularizado. El Jupyter Notebook implementado para la validación del subconjunto de consumo de termotanques presenta un ejemplo de regulación de períodos de muestreo. El código está disponible para ser descargado desde <https://github.com/jpchavat/ecd-uy/blob/master/EWH-subset-Technical-validation.ipynb>.

5.7. Aplicabilidad

Los datos expuestos en el conjunto de datos ECD-UY son un insumo importante para las investigaciones con fines académicos o industriales, que abordan problemas relacionados al consumo eléctrico en contextos similares al Uruguayo. A continuación se enumeran algunos ejemplos de aplicabilidad del conjunto de datos ECD-UY.

Monitoreo no intrusivo de cargas (NILM). Las técnicas NILM permiten obtener datos de consumo eléctrico desagregados por electrodomésticos de un hogar, sin necesidad de realizar mediciones intrusivas o instalar dispositivos extra al medidor inteligente. Para el entrenamiento de los métodos NILM que utilizan inteligencia computacional (Chavat et al., 2020; Esteban et al., 2021), es imprescindible contar con datos de consumo eléctrico desagregado, como los que incluye ECD-UY.

Diseño e implementación de estrategias de manejo de la demanda.

El manejo de la demanda en redes eléctricas inteligentes (*smart grids*) permite modificar, reducir o desplazar la demanda eléctrica de un hogar, para utilizar el recurso de forma eficiente. Por ejemplo, se puede lograr una reducción de picos de consumo desplazando el uso de ciertos electrodomésticos a horarios valle. De esta forma, la compañía de suministro eléctrico obtiene un uso más homogéneo de la energía a lo largo del día y el cliente puede verse beneficiado con precios más favorables. Para diseñar y aplicar estas estrategias, es fundamental contar con una caracterización del consumo y de los clientes lo más exacta posible, lo cual se logra a partir del análisis de los datos contenidos en ECD-UY.

Evaluación de técnicas de respuesta a la demanda. La creación de índices para evaluar las técnicas de respuesta a la demanda son de gran importancia para medir el descontento causado a los clientes en su aplicación, como por ejemplo al interrumpir de forma remota un termotanque (Porteiro et al., 2021). Para estas evaluaciones es necesario contar con datos de consumo como los contenidos en ECD-UY, que permiten simular y validar la aplicación de las técnicas.

Diseño de sistemas inteligentes para la planificación de consumo eléctrico. Los sistemas inteligentes para la planificación de consumo eléctrico le generan a sus usuarios recomendaciones de uso del recurso que minimizan la pérdida de confort. El seguimiento de las recomendaciones permite un uso eficiente de la electricidad, lo cual redundará en una reducción de los costos. Los datos de consumo real disponibles en ECD-UY son un insumo necesario para el entrenamiento y la validación de estos sistemas (Nesmachnow et al., 2021b,a).

Además, los datos de ECD-UY son de gran valor para las empresas de suministro eléctrico, particularmente para UTE, en el análisis de patrones de consumo de sus clientes. Los estudios de los patrones relacionan la forma de consumo con variables exógenas como el clima o las características socio-económicas de los clientes (Chavat y Nesmachnow, 2020b,a). Contar con conocimiento preciso de los patrones de consumo ayuda en la planificación de la demanda, la planificación de extensiones de las redes eléctricas o el diseño de políticas que incentiven un uso más eficiente de los recursos (*e.g.*, facturas desagregadas, franjas tarifarias ajustadas al tipo de cliente, etc).

5.8. Conclusiones

Este capítulo presentó el conjunto de datos ECD-UY, primero de su tipo en el cono sur del continente Americano. El conjunto está compuesto por tres subconjuntos de datos de consumo eléctrico y un subconjunto de información de clientes. Se presentan las metodologías de recolección, comunicación y pre-procesamiento de los datos, la estructura final de los diferentes subconjuntos y una validación técnica de la información que lo conforma. Finalmente, se presenta el repositorio de datos públicos del cual se puede obtener el conjunto, se brindan instrucciones de su utilización y se enumeran ejemplos de aplicabilidad de los datos.

Capítulo 6

Análisis del consumo eléctrico residencial en Uruguay

Este capítulo presenta un análisis del consumo eléctrico residencial en Uruguay a partir de la creación de un índice de consumo basado en datos de posesión y características de electrodomésticos. Para la creación del índice, se tomaron en cuenta variables socioeconómicas disponibles en la Encuesta Continua de Hogares del Instituto Nacional de Estadísticas e información de electrodomésticos del mercado local. El índice se validó en tres zonas censales utilizando datos reales de consumo eléctrico presentes en el conjunto ECD-UY.

6.1. Introducción

El análisis y la caracterización del consumo eléctrico es un insumo importante para los métodos de manejo de la demanda orientados a mejorar el uso de los recursos energéticos. Uno de los posibles enfoques del manejo de la demanda consiste en motivar a los clientes a cambiar su comportamiento de consumo y así generar ahorros en la energía que consumen. El análisis de datos provee información precisa sobre la forma en la que los clientes consumen la energía y puede ser utilizada en la elaboración de políticas que impacten sobre la conducta de consumo, planes comerciales o la delineación de nuevas tarifas.

Este capítulo presenta un índice que modela el nivel de consumo eléctrico residencial a partir de información de posesión de electrodomésticos por áreas censales y las características de los electrodomésticos en el mercado local. Para la construcción del índice se utilizaron datos de posesión de electrodomésticos

obtenidos de la Encuesta Continua de Hogares (ECH) que realiza el Instituto Nacional de Estadísticas (INE) todos los años e información de características de electrodomésticos, recolectada de tiendas locales con presencia en Internet. Finalmente, la validación se procesó utilizando información de consumo real presente en el conjunto ECD-UY. El índice se calculó para tres niveles de áreas censales: departamentos, secciones y segmentos.

El índice y el análisis de consumo eléctrico presentados en este capítulo fueron publicados en un artículo elaborado para la Conferencia Ibero-Americana de Ciudades Inteligentes (ICSC-CITIES 2020) (Chavat y Nesmachnow, 2020a) y en otro elaborado para el Primer Taller Internacional sobre Tecnologías y Sistemas Avanzados de Información y Computación (AICTS 2020) (Chavat y Nesmachnow, 2020b).

6.2. Motivación e hipótesis

Esta sección describe el problema abordado, las principales motivaciones, las preguntas y las hipótesis planteadas.

Se propone la construcción de un índice que modele los niveles de consumo eléctrico residencial por áreas censales, basado en características socioeconómicas e información sobre la posesión de electrodomésticos locales. El índice puede ser utilizado para procesar diferentes análisis sobre el consumo eléctrico de Uruguay y extenderse a otros países con insumos de datos similares.

La pregunta que se formula es: ¿Es posible construir un índice que modele los niveles de consumo eléctrico a partir de datos estadísticos de posesión de electrodomésticos?

En el mercado local existen electrodomésticos considerados de consumo eléctrico intensivo, como los aires acondicionados o los termotanques. Este tipo de electrodoméstico determinan en gran medida el nivel de consumo eléctrico de los hogares. Algunos de estos electrodomésticos no son de primera necesidad, y por lo tanto no todos los hogares los poseen. Además, se puede suponer que cada área censal comparte entre sus casas niveles similares de posesión de electrodomésticos. Por lo tanto, al cuantificar el nivel de posesión de los electrodomésticos en un área censal, es posible calcular el consumo eléctrico esperado de los hogares en cada área.

De la pregunta previa y las consideraciones anteriores, se formula la siguiente hipótesis. Cuanto más electrodomésticos de consumo intensivo se poseen en un hogar, mayor es el consumo eléctrico esperado. De igual forma, cuanto más electrodomésticos de consumo intensivo se poseen en promedio en un área censal, mayor es el consumo eléctrico esperado para el área.

6.3. Abordaje propuesto para el análisis de consumo

Se propone abordar el análisis de consumo eléctrico a través de la creación de un índice que cuantifica el consumo eléctrico potencial por áreas censales. Para la construcción del índice, se utilizan datos estadísticos georreferenciados de la ECH del año 2019 e información de las características de los electrodomésticos en el mercado local, recabada de tiendas con presencia en Internet.

Los datos de las encuestas fueron procesados para determinar la tenencia promedio de electrodomésticos por área censal. Simultáneamente, se calculó la potencia promedio consumida por los electrodomésticos relevados, y se les asignó una frecuencia de uso para ponderar su uso (e.g., el uso de una heladera es constante, mientras que el de una televisión es intermitente). Las frecuencias se asignaron por regla del pulgar, basado en la experiencia de uso como usuario.

Tras calcular el promedio de tenencia, consumo y frecuencia de uso, los valores del índice se calcularon como lo expresa la Ecuación 6.1. Dado un tipo de área censal r , con m áreas diferentes, una matriz $A^{(r)} \in \mathbb{R}^{m \times n}$ con una fila por área censal y una columna por electrodoméstico; un vector $\vec{c}^{(r)} \in \mathbb{R}^n$ con la potencia promedio (en vatios) de n electrodomésticos; y un vector $\vec{f}^{(r)} \in \mathbb{R}^n$ que cuantifica la frecuencia de uso de cada n electrodoméstico, se obtuvo el valor del índice para cada área censal como resultado de la multiplicación de los componentes. El vector resultante $index^{(r)} \in \mathbb{R}^m$ representa el valor del índice para cada área en unidad de vatios (W). El índice cuantifica la potencia del consumo potencial por hogar en cada área, lo que facilita la comparación y análisis de eficiencia energética entre diferentes áreas censales.

$$index^{(r)} = A_{m,n}^{(r)} \cdot \vec{c}^{(r)} \cdot \vec{f}^{(r)} \cdot \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad (6.1)$$

6.4. Descripción de los datos

En esta sección se brindan detalles de la recolección y pre-procesamiento de los datos utilizados en el cálculo del índice para diferentes áreas censales de Uruguay. .

6.4.1. Datos censales

Los datos utilizados que provienen de censos corresponden a los de la ECH del año 2019, relevados por el INE. La institución recolecta datos de diferentes características con una frecuencia mensual, cuatrimestral y anual, presentada en forma de encuesta continua de hogares cada año. Las ECH recolectan información del mercado laboral, ingresos de los hogares y los individuos, entre otros datos socioeconómicos, en base a una muestra representativa de hogares georreferenciados por área censal y distribuidos en Uruguay.

La georreferenciación de los datos se establece como mínimo a los departamentos, secciones censales y segmentos censales. La definición de los niveles de georreferenciación se detalla a continuación:

- *Departamento.* Áreas que corresponden al límite político-administrativo de Uruguay, en total son 19.
- *Sección censal.* Áreas que corresponden al primer nivel de división de los departamentos, coincidiendo con los utilizados en el censo del año 1963.
- *Segmento censal.* Áreas que corresponden a la subdivisión de las secciones censales y son una porción de territorio que agrupa unidades menores con límites físicos reconocibles en el terreno. Pueden incluir pequeños centros poblados.

Se seleccionó y se trabajó con un subconjunto de datos de los relevados en la ECH. Una parte de los datos seleccionados permiten la georreferenciación de los hogares, mientras que la otra parte brindan información sobre la posesión de electrodomésticos del hogar. La Tabla 6.1 enumera el nombre clave, la descripción y el tipo de dato de los datos seleccionados.

Preparación de los datos

En análisis primarios se detectó que los registros de ECH que no corresponden al departamento de Montevideo no cuentan con datos de sección ni

<i>nombre</i>	<i>descripción</i>	<i>tipo</i>
dpto	Código del departamento (1-19)	Número
nomdpto	Nombre del departamento	Texto
secc	Sección censal	Número
segm	Segmento censal	Número
nombarrio	Nombre del barrio	Texto
d18	Fuente de energía para la iluminación (1: eléctrica; 2-4: otra)	Número
d260	Fuente de energía para la calefacción (1: eléctrica; 2-6: otra)	Número
d20	Fuente de energía para cocinar (1: eléctrica; 2-6: otra)	Número
d21.1	Termotanque (1: si; 2: no)	Número
d21.2	Calentador inst. de ducha (1: si; 2: no)	Número
d21.3	Heladera (1: si; 2: no)	Número
d21.4	TV de tubo (1: si; 2: no)	Número
d21.4_1	Cantidad de TVs de tubo	Número
d21.5	TV LCD/Plasma (1: si; 2: no)	Número
d21.5_1	Cantidad de TVs LCD/Plasma	Número
d21.6	Radio (1: si; 2: no)	Número
d21.9	Reproductor DVD (1: si; 2: no)	Número
d21.10	Lavarropa (1: si; 2: no)	Número
d21.11	Secarropa (1: si; 2: no)	Número
d21.12	Lavavajilla (1: si; 2: no)	Número
d21.13	Microondas (1: si; 2: no)	Número
d21.14	Aire Acondicionado (1: si; 2: no)	Número
d21.14_1	Número de Aires Acondicionados	Número
d21.15	Computadoras portátiles (1: si; 2: no)	Número
d21.15_2	Cantidad de computadoras del Plan Ceibal	Número
d21.15_4	Cantidad de otras computadoras portátiles	Número

Tabla 6.1: Descripción de los datos de ECH (2019) utilizados para la construcción del índice.

segmento censal. Por este motivo, la evaluación del índice para los niveles de área censal de sección y segmento se limitó al departamento de Montevideo. Una primera transformación fue aplicada sobre las columnas con valores Si/No, transformaron los valores 1 y 2 a valores 0 para el caso de No y 1 para el caso de Si. El cambio facilitó los cálculos posteriores ya que al multiplicar por cero se anulan los valores de cantidades o consumo de un hogar. De igual forma, las columnas que indican la fuente de energía de cierta actividad o dispositivo fueron transformadas al valor 1 para el caso de electricidad y cero para el resto de los casos. Además, se formaron dos datos únicos, uno con la sumatoria

de la cantidad de computadoras portátiles (sean o no del Plan Ceibal) y otro con el valor resultante de multiplicar la fuente de energía para la iluminación y el número de habitaciones del hogar. Para distinguir entre los hogares con calefacción por aire acondicionado u otro medio eléctrico, el valor de la fuente de energía para la calefacción fue definido en cero para los hogares con aire acondicionado, y en 1 para los demás casos. De esta forma, es posible tratar la presencia de dispositivos de aire acondicionado diferente a la presencia de otros dispositivos eléctricos utilizados para la calefacción del hogar. Por último, se multiplicaron los datos de presencia de determinado tipo de electrodoméstico, por el dato correspondiente a su cantidad (e.g., presencia y cantidad de computadoras portátiles), transformándose en un único dato.

Para asegurar la integridad de los datos transformados se procesaron diferentes validaciones. Por ejemplo, se validó que los datos que indican la cantidad de cierto electrodoméstico con valores mayor a cero, presenten el valor correspondiente a ‘Si’ en el dato de presencia del electrodoméstico, o su opuesto, cantidades cero del electrodoméstico deben tener valor ‘No’ en el dato de presencia del mismo.

6.4.2. Datos de características de electrodomésticos

Basado en la lista de electrodomésticos que releva la ECH, se recolectó la potencia consumida de hasta cinco modelos diferentes de cada electrodoméstico, de tiendas locales con presencia en Internet, para luego procesar la potencia media por electrodoméstico. Además, se asignó a cada dispositivo una frecuencia de uso para ponderar la potencia media consumida en los cálculos del índice. La Tabla 6.2 presenta la información resultante del proceso de recolección y pre-procesamiento de los datos junto a la frecuencia de uso asignada, mientras que en la Figura 6.1 se observa una gráfica con la potencia media calculada y la desviación estándar de cada electrodoméstico.

6.4.3. Datos de consumo eléctrico

Los datos de consumo eléctrico fueron recolectados por la empresa UTE, utilizando los medidores inteligentes instalados en los hogares en el contexto de un plan de modernización de la infraestructura.

electrodoméstico	potencia media (W)	frecuencia de uso	potencia ponderada por frecuencia de uso
iluminación	11.8	media	8.85
calefacción	1200.0	alta	1200.00
horno	1380.0	alta	1380.00
termotanque	1600.0	alta	1600.00
calentador de ducha	1810.0	media	1357.50
heladera	199.4	alta	199.40
TV de tubo	124.8	media	93.60
TV plana	85.6	media	64.20
radio	20.2	baja	10.10
reproductor DVD	10.5	baja	5.25
lavarrpa	740.0	media	555.00
secarropa	3154.0	media	2365.50
lavavajilla	1409.6	media	1057.20
microondas	1068.0	media	801.00
aire acondicionado	1290.0	alta	1290.00
notebook	57.0	media	42.75

Tabla 6.2: Información de electrodomésticos utilizada para la creación del índice.

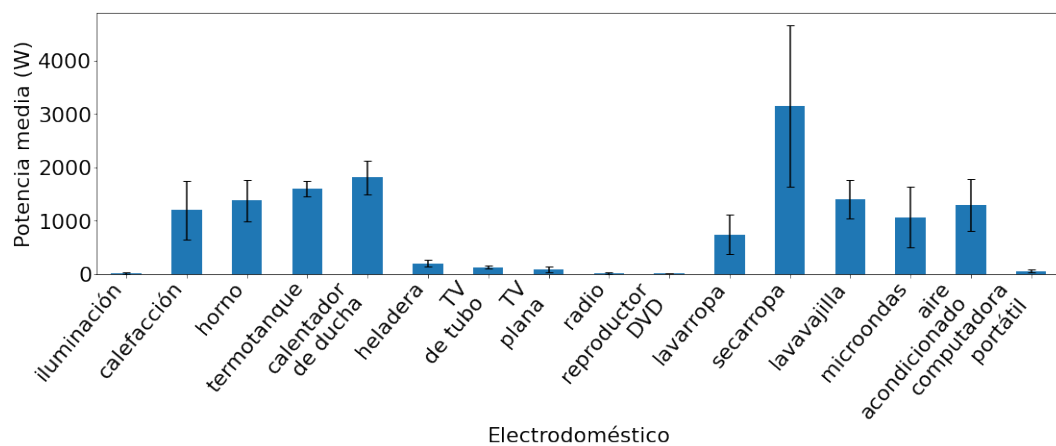


Figura 6.1: Potencia media consumida por electrodoméstico utilizado en el índice.

El subconjunto de consumo total está integrado por registros de consumo eléctrico total de 110.953 hogares, con período de muestreo de 15 minutos y datos de geolocalización. Del total de los hogares, se seleccionan los que registran consumo en el año 2019 (para coincidir con el año de la ECH), y se filtran los consumos promedios que se detectan como *outliers* (utilizando el rango intercuartílico). Los hogares resultantes son 83.597 y se distribuyen en 19 departamentos. En cuanto a los hogares en Montevideo, se distribuyen en 26 secciones censales y 211 segmentos censales.

Un dato importante para dimensionar el nivel de cobertura de los datos de consumo real es el porcentaje de clientes monitoreados, con respecto al total. Para contar con los datos de representatividad por área se procesaron los datos demográficos de población por departamento y por cantidad de hogares ocupados en secciones y segmentos censales en Montevideo. Los datos fueron obtenidos del censo nacional del año 2011 realizado por INE. Se distribuyó la cantidad de clientes residenciales de UTE en cada departamento, ponderando por la población de cada uno con respecto a la población total de Uruguay. Para las áreas censales, se asumió que todos los hogares de Montevideo cuentan con conexión eléctrica, por lo tanto el valor de hogares ocupados coincide con la cantidad de clientes residenciales por área censal. Esta asunción es posible ya que el acceso de los hogares al tendido eléctrico es considerado universal en Uruguay (99,8%). La Figura 6.2 muestra para cada departamento el porcentaje de hogares de los que se dispone de datos de consumo real y geolocalización.

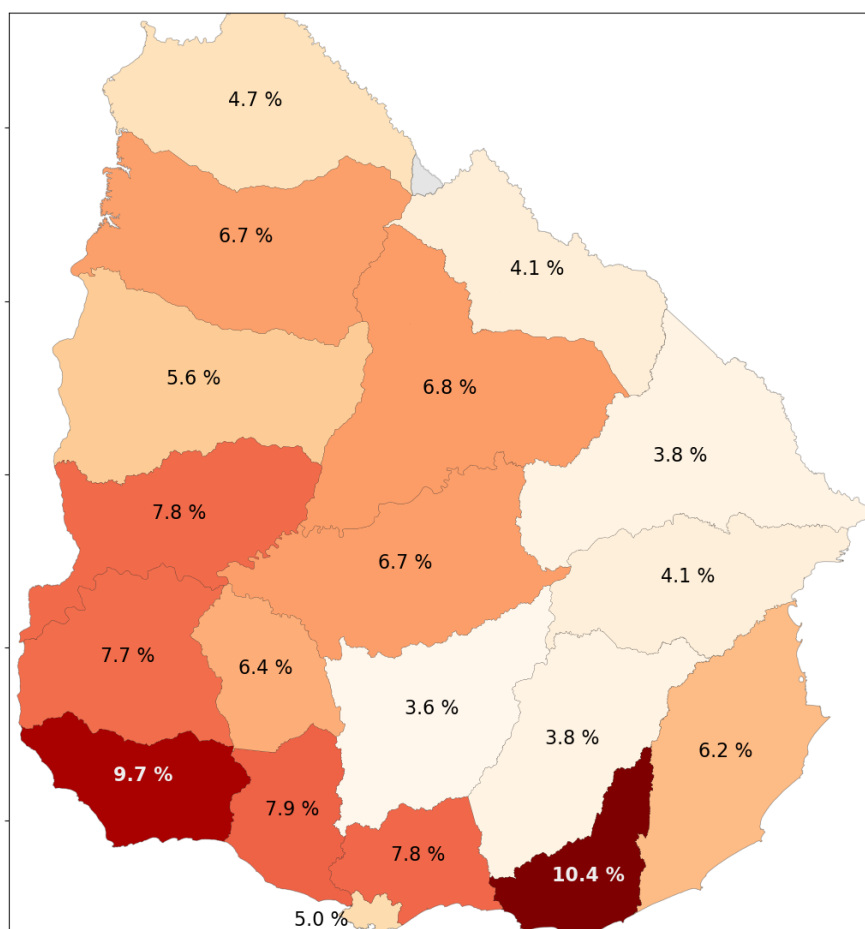


Figura 6.2: Porcentaje de hogares por departamento presentes en ECD-UY con consumo eléctrico total, luego de filtrar por año y percentil.

Las Figuras 6.3 y 6.4 muestran las secciones y segmentos de Montevideo coloreados según una escala que indica el porcentaje de hogares de los que se dispone de datos de consumo real y geolocalización.

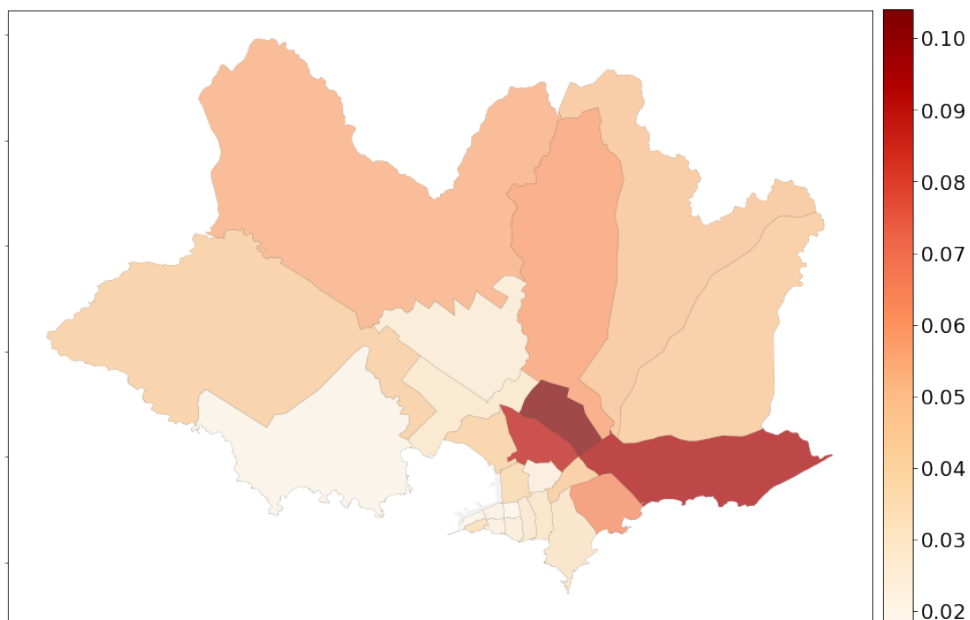


Figura 6.3: Porcentaje de hogares por sección censal de Montevideo presentes en ECD-UY con consumo eléctrico total, luego de filtrar por año y percentil.

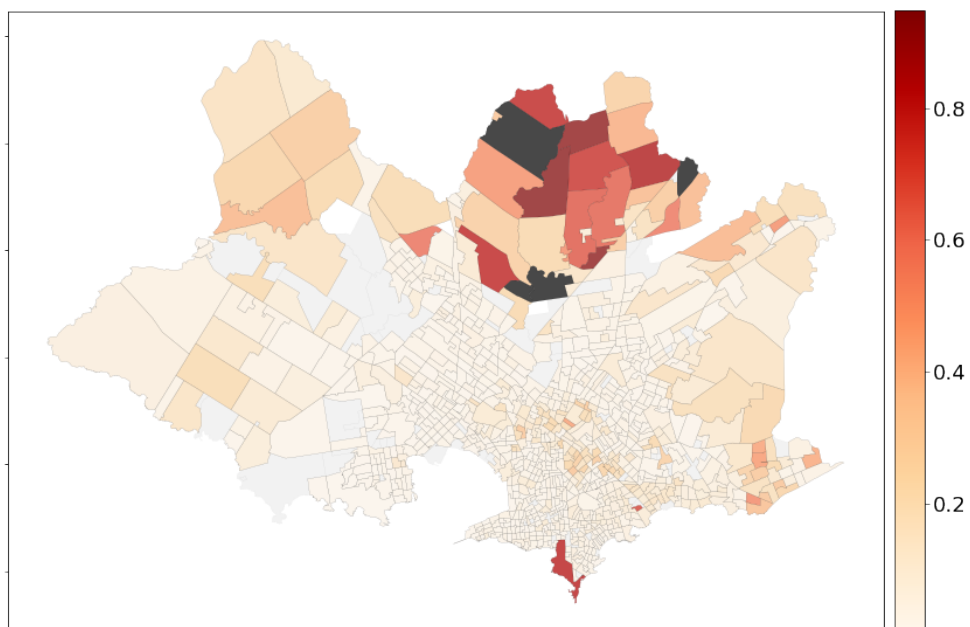


Figura 6.4: Porcentaje de hogares por segmento censal de Montevideo presentes en ECD-UY con consumo eléctrico total, luego de filtrar por año y percentil.

6.5. Implementación

La implementación del índice consistió en la carga de los datos, la construcción de la matriz de demanda de potencia y posesión de electrodomésticos, el procesamiento del índice para cada área censal y la visualización y validación de los resultados.

El procesamiento general se llevó a cabo en una computadora personal y el cálculo de promedios de consumo por área censal se llevó a cabo utilizando la plataforma del centro de supercomputación nacional, Cluster-UY (Nesmachnow y Iturriaga, 2019). El código fue implementado utilizando Jupyter Notebook y el lenguaje Python versión 3. Además, se utilizaron las librerías Pandas, su extensión para procesamiento de mapas GeoPandas, Dask y Numpy. El *notebook* resultante está disponibles para su descarga en <https://gist.github.com/jpchavat/f1799c73ebadf001cbebf0f804657ba>.

6.6. Análisis experimental

Esta sección presenta los resultados obtenidos al procesar el índice en tres niveles censales de Uruguay: por departamento, por sección censal de Montevideo y por segmento censal de Montevideo. El índice fue validado utilizando datos de consumo real disponibles en el conjunto ECD-UY.

6.6.1. Resultados obtenidos

Los resultados del índice para todos los experimentos presentaron diferencias de hasta 65 % entre el menor y el mayor valor del área evaluada.

Resultados por departamento. Montevideo es el departamento que obtuvo los mayores valores del índice, mientras que Cerro Largo obtuvo los menores. En general, el grupo de departamentos que obtuvieron mayores valores del índice se encuentran localizados sobre la costa oeste y suroeste del país. La Figura 6.5 grafica los diferentes departamentos con una escala de colores que representa el valor del índice obtenido en cada caso. Se observa cómo, comenzando desde el departamento de Colonia (ubicado en el extremo suroeste) en dirección hacia el noreste, los valores del índice van de más a menos. La Tabla 6.3 muestra la lista completa de departamentos y valor del índice obtenido, ordenado de mayor a menor.

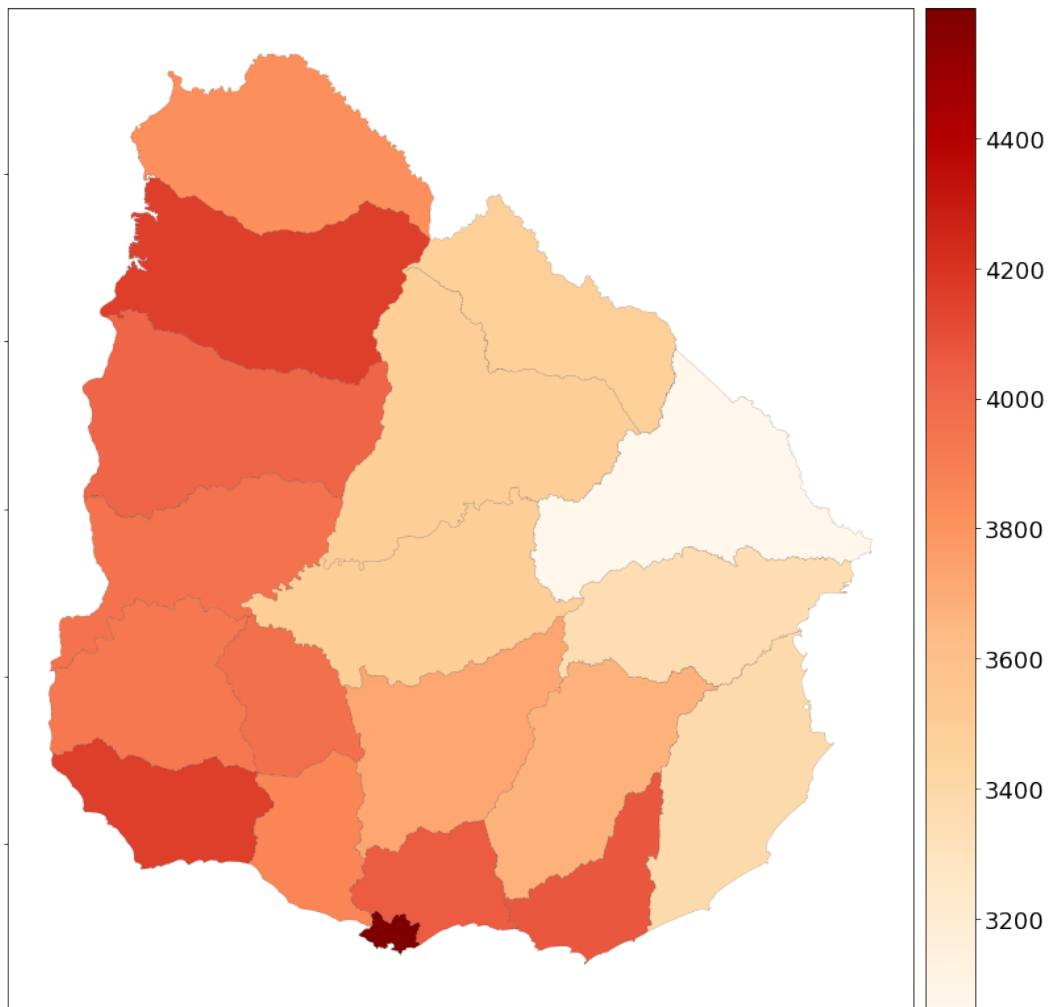


Figura 6.5: Valores calculados del índice para los departamentos de Uruguay.

valor del índice	departamento	valor del índice	departamento
4599.3	Montevideo	3819.4	Artigas
4155.9	Salto	3727.3	Florida
4155.0	Colonia	3678.0	Lavalleja
4069.5	Maldonado	3497.5	Durazno
4055.2	Canelones	3489.5	Tacuarembó
4021.5	Paysandú	3475.6	Rivera
3973.0	Flores	3387.1	Rocha
3955.1	Río Negro	3358.1	Treinta Y Tres
3928.7	Soriano	3059.9	Cerro Largo
3862.8	San José		

Tabla 6.3: Lista de departamentos de Uruguay, junto al valor resultante del índice.

Resultados por sección censal. Las secciones censales de Montevideo que presentaron los mayores valores del índice se localizan sobre la costa sureste del departamento. La sección con mayor valor del índice es la número 10 y cubre los barrios Carrasco Norte, Buceo, Malvín, Malvín Norte, Punta Gorda, Unión, Las Canteras y Carrasco, mientras que la de menor valor es la número 11, que cubre los barrios Ituzaingó, Jardines del Hipodromo, Flor de Maroñas, Punta de Rieles, Bella Italia, Manga, Toledo Chico, Piedras Blancas, Villa García, Villa Española e Unión. Si se tiene en cuenta la sección con mayor y menor valor, la diferencia registrada es de 66 %. En la Figura 6.6 se aprecia cómo partiendo desde la costa sureste, y en dirección hacia el noroeste, el valor del índice va descendiendo progresivamente, sección a sección. La Tabla 6.4 lista todos los valores resultantes para las secciones procesadas, junto a un listado de los barrios que la comprenden, ordenadas de mayor a menor.

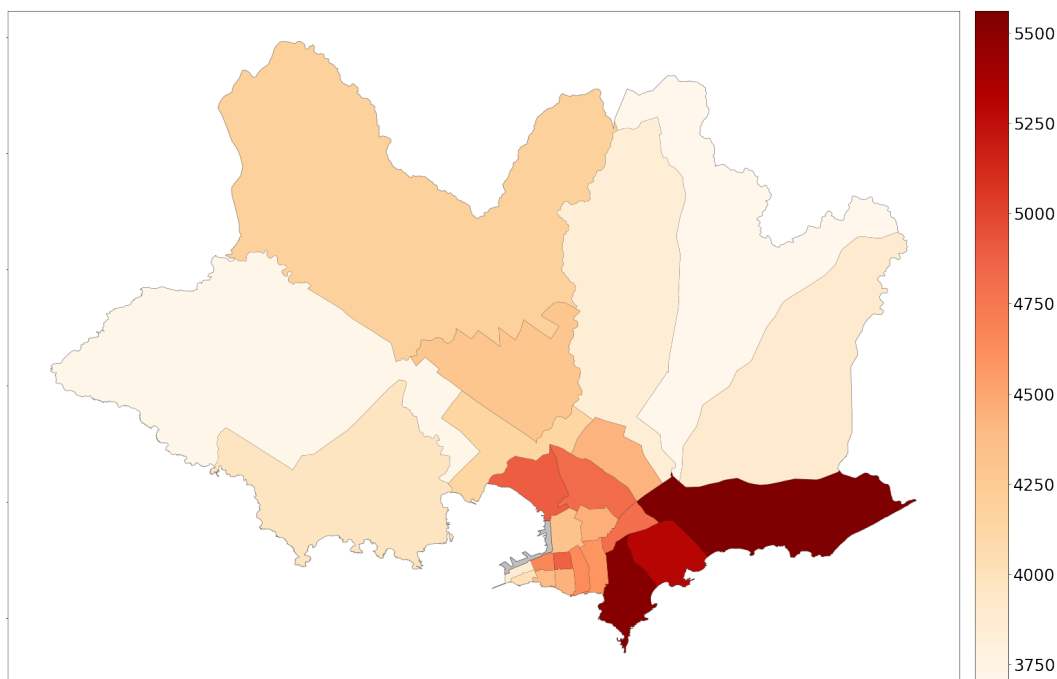


Figura 6.6: Valores obtenidos para el índice calculado para las secciones censales de Montevideo

valor del índice	sección	barrio
5561.1	10	Carrasco Norte, Buceo, Malvín, Malvín Norte, Punta Gorda, Unión, Las Canteras, Carrasco
5523.1	18	Punta Carretas, Pocitos, Cordón, Tres Cruces, Parque Batlle, V. Dolores, Parque Rodó
5303.1	24	Pocitos, Pque. Batlle, Villa Dolores, Buceo
4896.5	14	Prado, Nueva Savona, Reducto, Capurro, Bella Vista
4883.5	6	Centro (Norte)
4820.0	12	Reducto, Atahualpa, La Figurita, Jacinto Vera, Larrañaga, Brazo Oriental, Mercado Modelo, Bolivar
4805.1	23	Tres Cruces, La Blanqueada, Larrañaga
4664.5	1	Ciudad Vieja (Noreste), Centro
4635.6	7	Cordón, Palermo
4592.2	15	Cordón, Palermo, Parque Rodó
4460.0	19	La Comercial, Villa Muñoz, Retiro
4442.4	5	Centro (Sur), Barrio Sur
4438.3	22	Cerrito, Brazo Oriental, Villa Española, Bolivar, Mercado Modelo, Castro, P. Castellanos
4383.4	4	Centro (Suroeste), Ciudad Vieja (Sureste), Barrio Sur
4325.3	8	Aguada
4295.0	21	Peñarol, Lavalleja, Conciliacion, Sayago, Nuevo París Paso de las Duranas, Belvedere
4184.7	9	Colón Centro y Noroeste, Colón Sureste, Abayubá, Lezica, Melilla
4141.4	20	Aires Puros, La Teja, Prado, Nueva Savona, Belvedere, Nuevo París
4016.1	3	Ciudad Vieja (Sur)
3972.0	13	Casabó, Pajas Blancas, Paso de la Arena, La Paloma, Tomkinson, Cerro
3889.5	99	Flor de Maroñas, Maroñas, Parque Guaraní, Unión Bañados de Carrasco, Villa García, Manga Rural, Punta Rieles, Bella Italia, Las Canteras
3864.3	2	Ciudad Vieja (Norte)
3848.6	17	Casavalle, Manga, Las Acacias, Villa Española, Piedras Blancas, Castro, P. Castellanos, Manga, Toledo Chico
3715.0	16	Tres Ombúes, Victoria, Nuevo París, Paso de la Arena Casabó, Pajas Blancas
3696.6	11	Ituzaingó, Jardines del Hipodromo, Flor de Maroñas, P. Rieles, Bella Italia, Manga, Toledo Chico, Piedras Blancas, Villa García, Villa Española, Unión

Tabla 6.4: Lista de las secciones de Montevideo, junto al valor resultante del índice.

Resultados por segmento censal. Los resultados del índice calculados para los segmentos censales en Montevideo mostraron cómo segmentos ubicados sobre la costa sureste acumulan los mayores valores, mientras que los segmentos ubicados a las afueras de la ciudad concentran los menores valores del índice. La Figura 6.7 muestra el mapa de los segmentos censales de Montevideo, donde la escala de colores representa el valor del índice resultante para cada segmento. La Tabla 6.5 muestra una lista truncada de los segmentos censales ordenada por los valores resultantes del índice. En la tabla queda en evidencia la gran diferencia entre los menores y los mayores valores obtenidos para los segmentos, llegando a ser de hasta 6 veces.

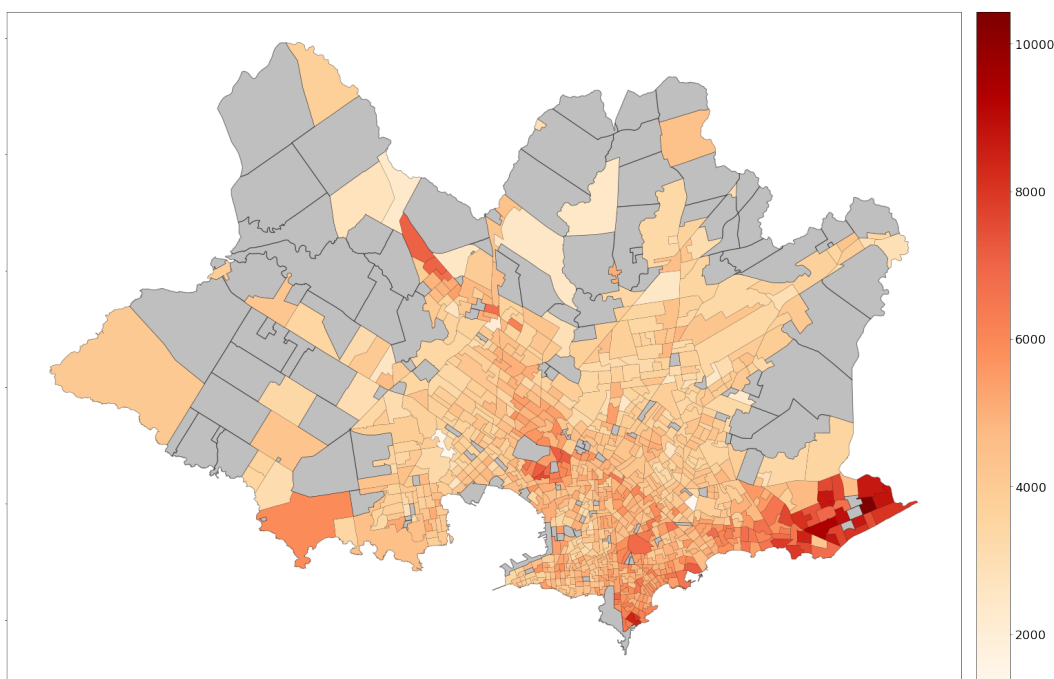


Figura 6.7: Valores obtenidos para el índice calculado para los segmentos censales de Montevideo

6.6.2. Validación del enfoque propuesto

Para validar el índice propuesto se calculó el consumo eléctrico mensual promedio de las diferentes áreas censales, a partir de datos de consumo eléctrico real. Los valores de consumo promedio obtenidos se contrastaron con los valores obtenidos por el índice y ambos valores fueron procesados por regresiones lineal OLS para medir su correlación, considerando la métrica de R^2 . El cálculo de consumo promedio mensual está basado en el consumo real de 108.701 hogares

valor del índice	sección	segmento	barrio
10434.8	10	75	Carrasco
9380.6	10	64	Carrasco
8920.8	10	66	Carrasco
8800.0	10	77	Carrasco
8753.3	10	73	Carrasco
...
2405.5	9	51	Leizica, Melilla
2381.7	99	208	Bañados de Carrasco
2081.2	9	25	Colón Centro y Noroeste
1344.2	10	218	Unión
1304.7	13	113	Tres Ombúes, Victoria

Tabla 6.5: Lista truncada de los segmentos de Montevideo, junto al valor resultante del índice.

monitoreados en el subconjunto de datos ECD-UY correspondiente al consumo total de los hogares.

Para los resultados de los departamentos, se realizó una inspección visual sobre el consumo real promedio y el valor procesado por el índice. Al ordenar los departamentos por el consumo potencial fue posible contrastarlos con el consumo real y observar que varios de los departamentos se encuentran en el orden correcto. La Figura 6.8 muestra una gráfica de barras donde para cada departamento se visualiza el consumo potencial *vs.* el consumo real promedio. Esta comparación es posible para los departamentos ya que la cantidad de estos es una cifra manejable, no así para las secciones o segmentos censales. Para todos los casos se utilizaron los resultados de la regresión OLS. En el caso de los departamentos, utilizando los datos de todos los departamentos sin importar el nivel de representatividad, obtenemos un valor alto de R^2 igual a 0,79. La Figura 6.9 muestra el consumo mensual promedio calculado para los departamentos de Uruguay, junto al valor resultante del índice. Se observa cómo los puntos, en su gran mayoría, permanecen cercanos a la recta diagonal, indicando la correlación entre ambos valores.

Para el caso de las secciones censales se procesaron diferentes regresiones OLS en las que se varían las secciones tomadas en cuenta según el nivel de representatividad de los datos de consumo real promedio. Comenzando sin restricción de mínimo de representatividad de los datos (25 secciones) el resultado obtenido de R^2 es 0,01, restringiendo a un mínimo de 6% (4 secciones) el valor de R^2 es 0,03. Si bien se observa un aumento del R^2 al utilizar datos

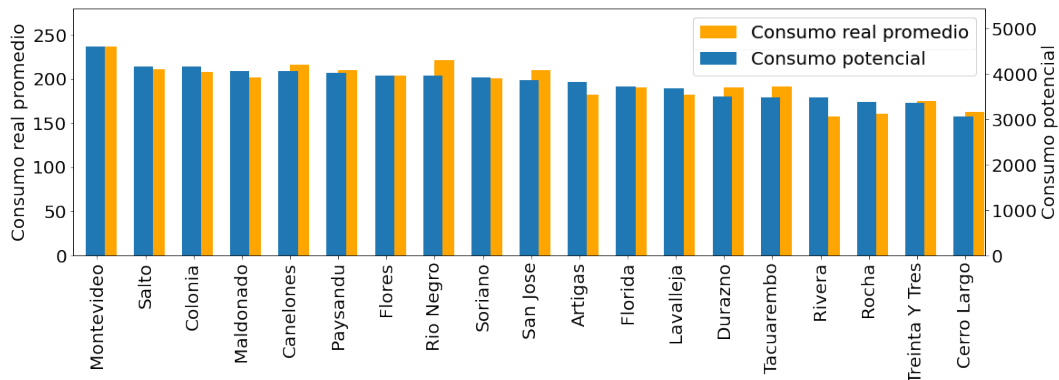


Figura 6.8: Consumo potencial y consumo real promedio para cada departamento de Uruguay.

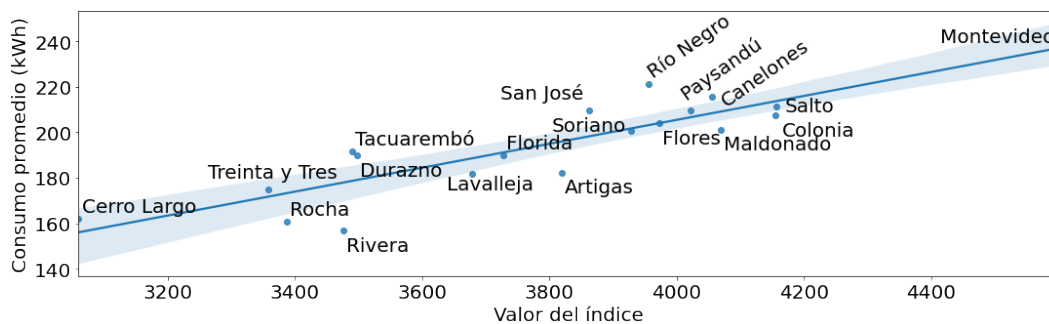


Figura 6.9: Valores obtenidos para el índice y valor de consumo promedio calculado a partir de consumo real, para los departamentos de Uruguay.

con mejor representación, los valores permanecen en niveles muy bajos.

Por último, se procesaron instancias de OLS para el caso de los segmentos censales. Los resultados obtenidos del R^2 van desde 0,05 al no aplicar filtro de datos por representatividad (877 segmentos), aumentando de forma progresiva, por ejemplo hasta 0,33 al filtrar los datos por aquellos que tienen al menos un 30 % de representatividad (21 segmentos).

6.7. Conclusiones

Este capítulo presenta un índice que cuantifica el consumo eléctrico potencial de los hogares, basado en datos socioeconómicos de posesión de electrodomésticos y características de los electrodomésticos en el mercado local. El índice se procesó para tres niveles de áreas censales: departamentos, secciones y segmentos. Los datos socioeconómicos fueron provistos por la ECH del año 2019, la información de las características de los electrodomésticos se

recolectó de tiendas locales con presencia en internet, y los datos de consumo real, utilizados durante la validación, fueron obtenidos del conjunto de datos ECD-UY.

Una inspección visual de los resultados obtenidos muestra a los departamentos con mayor valor de índice sobre la costa sur y oeste del país, mientras que para las secciones y segmentos en Montevideo, los mayores valores se encuentran sobre la costa sureste, disminuyendo a medida que se avanza hacia las afueras de la ciudad. Estos datos visuales coinciden en gran medida con los procesados al utilizar datos de consumo real. Por otro lado, se procesaron múltiples instancias de regresiones lineales OLS para evaluar la correlación entre los valores del índice y el consumo real de las áreas censales, a través del R^2 obtenido. Para los tres escenarios de pruebas se procesaron diferentes instancias variando los datos de consumo real utilizados a partir del porcentaje de representatividad que significan para cada área censal. Para los departamentos se obtuvo un R^2 entre 0,65 y 0,79 con variaciones de representatividad mínimas del 0 % al 6 %. En cuanto a las secciones censales, el R^2 registrado varió desde 0,01 sin restricción de representatividad hasta 0,03 con restricción de mínima representatividad de 6 %. Finalmente, para segmentos censales se registró un R^2 de 0,05 sin aplicar restricciones de representatividad, y fue creciendo progresivamente al aumentar el mínimo porcentaje de representatividad, por ejemplo hasta 35 % resultando en un valor de R^2 igual a 0,37. La poca correlación entre los valores del índice y el consumo real de las secciones censales podría deberse a que su definición para Montevideo abarca segmentos censales muy disímiles entre ellos, dificultando la posibilidad de caracterizar apropiadamente el perfil de consumo de los hogares en cada una de estas áreas.

Capítulo 7

Conclusiones y trabajo futuro

Este capítulo final incluye los principales resultados y conclusiones obtenidos en esta tesis de maestría y formula algunas líneas de investigación como trabajo futuro.

7.1. Conclusiones

Esta tesis estudió y caracterizó el consumo eléctrico residencial en Uruguay a través del análisis de datos de consumo en tres actividades de investigación: el estudio del monitoreo no intrusivo de cargas y la implementación de un algoritmo de desagregación simple, la creación y presentación de un conjunto de datos de consumo eléctrico residencial de hogares Uruguayos y la creación de un índice basado en datos socioeconómicos para predecir el consumo potencial de energía eléctrica residencial en diferentes áreas censales.

El creciente consumo de energía eléctrica en el mundo a nivel residencial a motivado múltiples investigaciones para comprender las características de consumo y lograr usos más eficientes del recurso. Una de las áreas de investigación trata sobre el monitoreo no intrusivo de cargas, que permite obtener datos detallados de consumo a partir del procesamiento y análisis del consumo total de un hogar, empresa o edificio. El monitoreo no intrusivo de cargas se diferencia de otras técnicas por no requerir la instalación de una gran cantidad de dispositivos de monitoreo, sino que basta contar con un medidor de consumo general del lugar. La información recolectada a partir de esta técnica puede ser pos-procesada por diferentes algoritmos, por ejemplo de inteligencia computacional, para obtener el consumo desagregado por electrodoméstico. En

general, para el entrenamiento de algoritmos de inteligencia computacional se requiere contar con datos de consumo etiquetados y correspondientes al contexto geográfico al cual se desean aplicar. Uruguay es un país de 3,4 millones de habitantes, de los cuales 1,3 viven en su capital, Montevideo. La provisión de energía eléctrica para uso residencial está limitada a la empresa estatal UTE que atiende 1.498.164 clientes y cubre el 99,8% de las áreas totales del país con su red eléctrica. El 90,5% de los clientes son residenciales y consumen en promedio 228 kWh por mes. La empresa se encuentra en proceso de instalación de medidores inteligentes en todos sus clientes, lo cual le permitirá contar con datos de consumo total de los clientes en tiempo real. Una característica de importancia en el consumo residencial de Uruguay es la amplia presencia del termotanque en los hogares (86,2% del total) que representa en promedio el 27% del consumo total. Dadas las características de uso de este dispositivo (i.e., aseo personal y/o usos en la cocina), se puede pensar en el termotanque como un dispositivo de almacenaje de energía con una carga desplazable a franjas horarias más convenientes.

El estudio del monitoreo no intrusivo de cargas concluyó en la presentación del algoritmo de Desagregación Simple para desagregar cargas eléctricas totales. El algoritmo de desagregación se basa en la similitud de patrones similares de consumo y consta de una etapa de entrenamiento y otra de evaluación. Para validar el algoritmo de desagregación se creó un procedimiento que genera datos sintéticos basados en el conjunto UK-DALE. Los escenarios de validación tuvieron en cuenta escenarios reales que incluyeron ambigüedad de consumos entre electrodomésticos, diferentes períodos de muestreo y presencia de cargas extra a las cargas objetivo de desagregación (e.g., en forma de electrodomésticos o de ruido en la señal de consumo total). Para medir los resultados se establecieron 5 métricas: TEE, NEAP, *precision*, *recall* y Valor-F. En cuanto a la comparación de los resultados, se utilizaron dos algoritmos base presentes en la literatura revisada, CO y FHMM. Los resultados de DS reflejaron un rendimiento satisfactorio en todas las instancias del problema evaluadas, superando en general los resultados obtenidos para los algoritmos base. El análisis experimental demostró que DS es un algoritmo apropiado para abordar la desagregación de cargas, además de servir como acercamiento al problema de desagregación de consumo residencial.

El estudio de NILM y la construcción del algoritmo DS evidenciaron la necesidad de contar con datos reales de consumo eléctrico residencial corres-

pondientes al contexto geográfico en el cual se aplicó el estudio. Por esta razón, en este trabajo se colaboró junto a la empresa UTE para crear y presentar el conjunto de datos ECD-UY que consta de datos reales de consumo eléctrico residencial en Uruguay. ECD-UY se conformó por tres subconjuntos de datos de consumo y uno de información de clientes. Los subconjuntos de datos de consumo se conforman por el consumo total de 110.953 hogares, el consumo eléctrico desagregado del dispositivo termotanque de 268 hogares y el consumo desagregado de electrodomésticos de 9 hogares. Además, el subconjunto de información de clientes brinda datos que permiten georreferenciar los clientes a niveles de áreas censales e identificar los tipos de contratos que mantienen con la empresa. Este trabajo reunió y presentó las metodologías de recolección, comunicación y pre-procesamientos de los datos, dio estructura final a los subconjuntos y validó técnicamente la información que los compone. Finalmente, el conjunto fue liberado en un repositorio público junto a instrucciones de uso y ejemplos de utilización, quedando disponible para futuras investigaciones relacionadas al consumo eléctrico residencial.

Por último, este trabajo vinculó los datos de consumo eléctrico residencial con variables socioeconómicas por áreas censales. Particularmente, se presentó un índice de consumo potencial por área censal basado en indicadores de posesión de electrodomésticos por área censal y las características de los electrodomésticos en el mercado local. El índice fue procesado para tres niveles de áreas censales: departamentos de Uruguay, secciones censales y segmentos censales de Montevideo. Para la construcción del índice en cada uno de los tres niveles se utilizaron datos de la ECH del año 2019 realizada por el INE e información de electrodomésticos recolectada de tiendas locales con presencia en Internet. Para la validación se utilizaron datos reales de consumo eléctrico residencial disponibles en el conjunto de datos ECD-UY. Se validó primero mediante una inspección visual comparando el consumo real y el reportado por el índice, y luego se procesaron diferentes regresiones OLS para analizar la correlación entre consumo real y valores del índice. Para los departamentos, tanto el índice como el consumo real mostraron valores altos de consumo en los departamentos de la costa sur y el litoral oeste del país, decreciendo progresivamente a medida que se avanza hacia el noroeste. Para las secciones censales de Montevideo, el índice presentó valores altos para las áreas de la costa este de Montevideo mientras que los valores de consumo real lo hicieron para la costa este y las áreas periféricas. Por último, para los segmentos cen-

sales de Montevideo, tanto el índice como el consumo real evidenciaron que los segmentos de alto consumo corresponden a la costa este y algunas áreas puntuales en el centro de la ciudad, mientras que los segmentos de bajo consumo se encuentran en la periferia de la ciudad. Para las regresiones OLS se ejecutaron múltiples instancias, variando las áreas censales utilizadas según la representación de los datos con la que se contaba. Con respecto a los departamentos, las OLS mostraron que el índice tiene una alta correlación con el consumo real, alcanzando a valores de R^2 de hasta 0,79. En cuanto a las secciones censales, el valor de R^2 obtenido fue muy bajo (entre 0,01 y 0,03), lo cual coincidió con la inspección visual. Finalmente, las regresiones OLS procesadas para los segmentos censales presentaron valores de R^2 de hasta 0,37. Los valores bajos en el procesamiento de las secciones censales se debió a que la definición de sección censal en Montevideo abarca segmentos censales muy disímiles entre ellos, lo cual dificulta la correcta caracterización del consumo eléctrico a ese nivel de área censal.

En forma de resumen, a continuación se listan los principales aportes que esta investigación incluye:

- Una revisión de los trabajos relacionados a la caracterización del consumo eléctrico, la técnica de monitoreo no intrusivo, los conjuntos de datos de consumo eléctrico y el análisis del consumo eléctrico en conjunto con variables socioeconómicas.
- Un algoritmo de desagregación simple basado en similitud de patrones de consumo eléctrico residencial utilizando la técnica NILM.
- Un conjunto de datos reales de consumo eléctrico residencial en Uruguay, conformado por los subconjuntos de consumo total, de termotanque y desagregado por diferentes electrodomésticos.
- Un índice para predecir el consumo eléctrico potencial a partir de diferentes áreas censales, basado en indicadores socioeconómicos y validado con datos de consumo eléctrico residencial real.

Esta tesis resultó en siete publicaciones, cinco como autor principal y dos como co-autor, que abordan temas incluidos en este informe y otras líneas de trabajos relacionados. A continuación se listan los trabajos resultantes:

- ‘Household energy disaggregation based on pattern consumption similarities’ presentado en la conferencia Ibero-Americana de Ciudades Inteligentes (ICSC-CITIES) del año 2019 ([Chavat et al., 2019](#)).

- ‘Nonintrusive energy disaggregation by detecting similarities in consumption patterns’ publicado en la Revista Facultad de Ingeniería de la Universidad de Antioquía (REDIN) en el año 2020 ([Chavat et al., 2020](#)).
- ‘Analysis of residential electricity consumption by areas in Uruguay’ presentado en la conferencia Ibero-Americana de Ciudades Inteligentes (ICSC-CITIES) del año 2020 ([Chavat y Nesmachnow, 2020a](#)).
- ‘Data analysis approach for characterizing residential energy consumption based on statistics of household appliances ownership’ presentado en el Primer Taller Internacional sobre Tecnologías y Sistemas Avanzados de Información y Computación (AICTS) del año 2020 ([Chavat y Nesmachnow, 2020b](#)).
- ‘Demand Response Control in Electric Water Heaters: Evaluation of Impact on Thermal Comfort’ presentado en la conferencia Ibero-Americana de Ciudades Inteligentes (ICSC-CITIES) del año 2020 ([Porteiro et al., 2020](#)).
- ‘A Thermal Discomfort Index for Demand Response Control in Residential Water Heaters’ publicado en la revista Applied Sciences de MDPI en el año 2021 ([Porteiro et al., 2021](#)).
- ‘ECD-UY: Detailed household electricity consumption dataset of Uruguay’ publicado en la revista Scientific Data de Springer Nature en el año 2022 ([Chavat et al., 2022a](#)).

Además, durante el transcurso de esta tesis se realizaron múltiples colaboraciones en el proyecto UTE-UdelaR *Inteligencia computacional para caracterizar los usos de energía eléctrica en clientes residenciales* financiado por UTE y la Comisión Sectorial de Investigación Científica de la Universidad de la República.

7.2. Trabajo futuro

Esta tesis establece una base para el estudio del consumo eléctrico residencial en Uruguay y por lo tanto, es necesario continuar explorando diferentes líneas de investigación que profundicen en el uso y aplicabilidad de lo estudiado en escenarios reales.

En primer lugar, el algoritmo DS fue entrenado utilizando cargas sintéticas que simulaban dispositivos con consumo constante, pero los escenarios reales

exigen que el algoritmo reconozca patrones de consumo más complejos. Para mejorar el rendimiento de DS sería necesario agregar a la etapa de entrenamiento datos de consumo correspondientes a dispositivos de consumo multi-estado o continuamente variables, como los presentes en los subconjuntos DAC y EWH del conjunto de datos ECD-UY. En segundo lugar, se debería llevar a cabo un estudio en profundidad sobre los valores de los parámetros de DS. El ajuste de estos parámetros proporcionaría al algoritmo la capacidad de detectar patrones que omitió en el análisis experimental llevado a cabo.

En cuanto al conjunto de datos ECD-UY, es necesario establecer procedimientos que permitan mantener el conjunto de datos actualizados a los consumos de cada año y sumarle nuevos clientes a medida que la empresa avance en la instalación de los medidores inteligentes. Contar con varios años de consumo permitiría realizar estudios comparativos año a año y predecir el consumo futuro, mientras que el aumento de la cantidad de datos mejoraría la precisión de la caracterización del consumo. Como complemento de los datos reales de consumo, se plantea el estudio de técnicas de aumentación de datos basadas en redes generativas antagónicas aplicadas a las series temporales (Yoon et al., 2019) con el fin de crear consumo sintético desagregado basado en los subconjuntos DAC y EWH de ECD-UY.

Con respecto al índice de consumo potencial presentado, las principales líneas de trabajo futuro incluyen el estudio de la incidencia de otras variables socioeconómicas en el consumo eléctrico residencial, como los niveles de ingresos y la cantidad de integrantes del hogar, las características edilicias, el nivel educativo, entre otros. Además, la actualización del índice para nuevas versiones de la ECH (suspendida en los años 2020 y 2021 por restricciones sanitarias relacionadas a la COVID-19) y la integración de las ECH anteriores para estudiar la evolución del índice en el tiempo. Por último, se debería trabajar para contar con datos más certeros sobre la frecuencia de uso de los electrodomésticos, algunos de ellos podrían ser obtenidos de los subconjunto DAC y EWH de ECD-UY. Por último, para mejorar la validación sería adecuado recalcular los valores de representatividad de cada área censal utilizando información de cantidad de viviendas correspondientes al mismo año que la ECH y los datos de consumo real.

Referencias bibliográficas

- Ali, U., Buccella, C., y Cecati, C. (2016). Households electricity consumption analysis with data mining techniques. In *42nd Annual Conference of the IEEE Industrial Electronics Society*, pages 3966–3971.
- Anderson, B., Lin, S., Newing, A., Bahaj, A., y James, P. (2017). Electricity consumption and household characteristics: Implications for census-taking in a smart metered future. *Computers, Environment and Urban Systems*, 63:58–67.
- Barbosa, L., Bogdanov, D., Vainikka, P., y Breyer, C. (2017). Hydro, wind and solar power as a base for a 100 % renewable energy supply for South and Central America. *PloS one*, 12(3):e0173820.
- Barker, S., Mishra, A., Irwin, D., Cecchet, E., Shenoy, P., y Albrecht, J. (2012). Smart*: An Open Data Set and Tools for Enabling Research in Sustainable Homes. In *Workshop on Data Mining Applications in Sustainability*, number 112 in SustKDD’12, page 108.
- Batra, N., Gulati, M., Singh, A., y Srivastava, M. (2013). It’s different: Insights into home energy consumption in India. In *5th Workshop on Embedded Systems For Energy-Efficient Buildings, BuildSys’13*, pages 1 – 8.
- Batra, N., Kelly, J., Parson, O., Dutta, H., Knottenbelt, W., Rogers, A., Singh, A., y Srivastava, M. (2014). NILMTK : An Open Source Toolkit for Non-intrusive Load Monitoring Categories and Subject Descriptors. In *5th international conference on Future energy systems, e-Energy ’14*, pages 265–2761.
- Benzi, F., Anglani, N., Bassi, E., y Frosini, L. (2011). Electricity Smart Meters Interfacing the Households. *Transactions on Industrial Electronics*, 58(10):4487 – 4494.

- Bonfigli, R., Squartini, S., Fagiani, M., y Piazza, F. (2015). Unsupervised algorithms for non-intrusive load monitoring: An up-to-date overview. In *15th International Conference on Environment and Electrical Engineering*, pages 1175 – 1180.
- Chavat, J., Graneri, J., y Nesmachnow, S. (2019). Household energy disaggregation based on pattern consumption similarities. In *Ibero-American Congress of Smart Cities*, pages 54–69.
- Chavat, J., Graneri, J., y Nesmachnow, S. (2020). Nonintrusive energy disaggregation by detecting similarities in consumption patterns. *Revista Facultad de Ingeniería Universidad de Antioquia*, (98):27–46.
- Chavat, J. y Nesmachnow, S. (2020a). Analysis of residential electricity consumption by areas in Uruguay. In *Ibero-American Congress of Smart Cities*, pages 42–57.
- Chavat, J. y Nesmachnow, S. (2020b). Data analysis approach for characterizing residential energy consumption based on statistics of household appliances ownership. In *1st International Workshop on Advanced Information and Computation Technologies and Systems*, CEUR Workshop Proceedings, pages 17–23.
- Chavat, J., Nesmachnow, S., Graneri, J., y Alvez, G. (2022a). ECD-UY: Detailed household electricity consumption dataset of Uruguay. *Scientific Data*, 9(1):1–16.
- Chavat, J., Nesmachnow, S., Graneri, J., y Alvez, G. (2022b). ECD-UY: Detailed household electricity consumption dataset of Uruguay. <https://doi.org/10.6084/m9.figshare.c.5428608.v1>. (Último acceso: Octubre 2022).
- Chávez, P., Barbero, D., Martini, I., y Discoli, C. (2017). Application of the k-means clustering method for the detection and analysis of areas of homogeneous residential electricity consumption at the Great La Plata region, Buenos Aires, Argentina. *Sustainable Cities and Society*, 32:115–129.
- De Silva, D., Yu, X., Alahakoon, D., y Holmes, G. (2011). A Data Mining Framework for Electricity Consumption Analysis From Meter Data. *Transactions on Industrial Informatics*, 7(3):399–407.

- Ellert, B., Makonin, S., y Popowich, F. (2016). Appliance water disaggregation via non-intrusive load monitoring (NILM). *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, 166:455–467.
- Esteban, M., Fiori, I., Mujica, M., y Nesmachnow, S. (2021). Computational Intelligence for Characterization and Disaggregation of Residential Electricity Consumption. In *Ibero-American Congress of Smart Cities*, pages 58–73. Springer.
- Figueiredo, M., de Almeida, A., y Ribeiro, B. (2012). Home electrical signal disaggregation for non-intrusive load monitoring (NILM) systems. *Neurocomputing*, 96:66–73.
- Ford, R. (2009). *Reducing domestic energy consumption through behaviour modification*. PhD thesis, Oxford University.
- Ghahramani, Z. y Jordan, M. (1996). Factorial hidden markov models. In *Advances in Neural Information Processing Systems*, pages 472–478.
- Gulati, M., Ram, S., y Singh, A. (2014). An in depth study into using EMI signatures for appliance identification. In *1st Conference on Embedded Systems for Energy-Efficient Buildings*, pages 70–79.
- Gupta, S., Reynolds, M., y Patel, S. (2010). ElectriSense: Single-point sensing using EMI for electrical event detection and classification in the home. In *12th International Conference on Ubiquitous Computing, UbiComp '10*, pages 139–148.
- Hart, G. W. (1992). Nonintrusive Appliance Load Monitoring. *IEEE*, 80(12):1870–1891.
- Hebrail, G. (2012). UCI Machine Learning Repository: Individual Household Electric Power Consumption Data Set. <http://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption> (Último acceso: Octubre 2021).
- Hoaglin, D. (2003). John W. Tukey and Data Analysis. *Statistical Science*, 18(3):311–318.

- IEA (2015). World Energy Outlook 2015. Technical report, International Energy Agency, París, Francia.
- Instituto Nacional de Estadística (2019). Microdatos de la encuesta continua de hogares. Technical report, INE, Uruguay. <https://www.ine.gub.uy/encuesta-continua-de-hogares1> (Último acceso: Octubre 2021).
- Kelly, J. y Knottenbelt, W. (2014). Metadata for Energy Disaggregation. In *2nd IEEE International Workshop on Consumer Devices and Systems*, pages 578–583.
- Kelly, J. y Knottenbelt, W. (2015a). Neural NILM: Deep Neural Networks Applied to Energy Disaggregation. In *2nd International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 55–64, New York, New York, USA.
- Kelly, J. y Knottenbelt, W. (2015b). The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Scientific data*, 2:150007.
- Keramati, M., Azizi, E., Momeni, H., y Bolouki, S. (2021). Incorporating coincidental water data into non-intrusive load monitoring. *Sustainable Energy, Grids and Networks*, 32.
- Kim, H., Marwah, M., Arlitt, M., Lyon, G., y Han, J. (2011). Unsupervised disaggregation of low frequency power measurements. In *International conference on data mining*, pages 747–758.
- Kleiminger, W., Beckel, C., y Santini, S. (2015). Household occupancy monitoring using electricity meters. In *International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '15*, pages 975–986.
- Kolter, J. y Johnson, M. (2011). REDD: A public data set for energy disaggregation research. In *Workshop on data mining applications in sustainability*, number 1, pages 59–62.
- Konstantakopoulos, I., Ratliff, L., Jin, M., Spanos, C., y Sastry, S. (2016). Smart building energy efficiency via social game: A robust utility learning framework for closing-the-loop. In *1st International Workshop on Science of Smart City Operations and Platforms Engineering in Partnership with Global City Teams Challenge*, pages 1–6.

- Kriechbaumer, T. y Jacobsen, H. (2018). BLOND, a building-level office environment dataset of typical electrical appliances. *Scientific Data*, 5:1–14.
- Larcher, D. y Tarascon, J. (2015). Towards greener and more sustainable batteries for electrical energy storage. *Nature chemistry*, 7(1):19.
- Laureiro, P. (2018). Determinantes del consumo de energía eléctrica del sector residencial en Uruguay. *Serie Documentos de investigación estudiantil*, FCS, Udelar.
- Luján, E., Otero, A., Valenzuela, S., Mocskos, E., Steffemel, L., y Nesmachnow, S. (2019). Cloud Computing for Smart Energy Management (CC-SEM Project). In *Smart Cities*, volume 978 of *Communications in Computer and Information Science*. Springer.
- Makonin, S. (2019). HUE: The hourly usage of energy dataset for buildings in British Columbia. *Data in Brief*, 23:103744.
- Makonin, S., Ellert, B., Bajić, I., y Popowich, F. (2016). Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014. *Scientific Data*, 3:1–12.
- McKinney, W. (2012). *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. O’Reilly Media, Inc.
- McLoughlin, F., Duffy, A., y Conlon, M. (2012). Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study. *Energy and Buildings*, 48:240–248.
- Ministerio de Industria, Energía y Minería (2018). Tasa de electrificación urbana y rural. Technical report, MIEM, Uruguay. <https://www.gub.uy/ministerio-industria-energia-mineria/datos-y-estadisticas/datos/series-estadisticas-energia-electrica> (Último acceso: Octubre 2021).
- Ministerio de Industria, Energía y Minería (2019). Número de clientes de energía eléctrica por sector. Technical report, MIEM, Uruguay. <https://www.gub.uy/ministerio-industria-energia-mineria/datos-y-estadisticas/datos/series-estadisticas-energia-electrica> (Último acceso: Septiembre 2022).

- Murray, D., Stankovic, L., y Stankovic, V. (2017). An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. *Scientific Data*, 4:1–12.
- Nesmachnow, S., Colacurcio, G., Rossit, D., Toutouh, J., y Luna, F. (2021a). Optimizing household energy planning in smart cities: a multiobjective approach. *Revista Facultad de Ingenieria Universidad de Antioquia*, 101:8–19.
- Nesmachnow, S. y Iturriaga, S. (2019). Cluster-UY: Collaborative Scientific High Performance Computing in Uruguay. In *Communications in Computer and Information Science*, pages 188–202.
- Nesmachnow, S., Rossit, D., Toutouh, J., y Luna, F. (2021b). An explicit evolutionary approach for multiobjective energy consumption planning considering user preferences in smart homes. *International Journal of Industrial Engineering Computations*, 12(4):365–380.
- Orsi, E. y Nesmachnow, S. (2017). Smart home energy planning using IoT and the cloud. In *IEEE URUCON*, pages 1–4.
- Parson, O., Fisher, G., Hersey, A., Batra, N., Kelly, J., Singh, A., Knottenbelt, W., y Rogers, A. (2016). Dataport and NILMTK: A building data set designed for non-intrusive load monitoring. *IEEE Global Conference on Signal and Information Processing*, pages 210–214.
- Picon, T., Meziane, M., Ravier, P., Lamarque, G., Novello, C., Le Bunetel, J., y Raingeaud, Y. (2016). COOLL: Controlled On/Off loads library, a public dataset of high-sampled electrical signals for appliance identification. *arXiv preprint arXiv:1611.05803*.
- Pipattanasomporn, M., Kuzlu, M., Rahman, S., y Teklu, Y. (2014). Load profiles of selected major household appliances and their demand response opportunities. *IEEE Transactions on Smart Grid*, 5(2):742–750.
- Porteiro, R., Chavat, J., y Nesmachnow, S. (2021). A thermal discomfort index for demand response control in residential water heaters. *Applied Sciences*, 11(21).
- Porteiro, R., Chavat, J., Nesmachnow, S., y Hernández-Callejo, L. (2020). Demand response control in electric water heaters: Evaluation of impact on thermal comfort. In *Ibero-American Congress of Smart Cities*, pages 74–89.

- Pullinger, M., Kilgour, J., Goddard, N., Berliner, N., Webb, L., Dzikovska, M., Lovell, H., Mann, J., Sutton, C., Webb, J., y Zhong, M. (2021). The IDEAL household energy dataset, electricity, gas, contextual sensor data and survey data for 255 UK homes. *Scientific Data*, 8(1):1–18.
- Reinhardt, A., Baumann, P., Burgstahler, D., Hollick, M., Chonov, H., Werner, M., y Steinmetz, R. (2012). On the accuracy of appliance identification based on distributed load metering data. In *Sustainable Internet and ICT for Sustainability*, pages 1–9.
- Rocklin, M. (2015). Dask: Parallel computation with blocked algorithms and task scheduling. In *14th Python in Science Conference*, pages 130–136.
- Runkler, T. A. (2020). *Data analytics*. Springer.
- Saldanha, N. y Beausoleil, I. (2012). Measured end-use electric load profiles for 12 Canadian houses at high temporal resolution. *Energy and Buildings*, 49:519–530.
- Shafranovich, Y. (2005). RFC 4180: Common format and mime type for comma-separated values (csv) files. *The International Society*, 54:258.
- Shin, C., Lee, E., Han, J., Yim, J., Rhee, W., y Lee, H. (2019). The ener-talk dataset, 15 hz electricity consumption data from 22 houses in Korea. *Scientific Data*, 6(11):1–13.
- Sultanem, F. (1991). Using appliance signatures for monitoring residential loads at meter panel level. *IEEE Transactions on Power Delivery*, 6(4):1380–1385.
- Tsai, C., Lai, C., Chao, H., y Vasilakos, A. (2015). Big data analytics: a survey. *Journal of Big data*, 2(1):1–32.
- Tukey, J. (1962). The future of data analysis. *The annals of mathematical statistics*, 33(1):1–67.
- Usinas y Trasmisiones Eléctricas (2019). Memoria anual 2020. Technical report, UTE. https://portal.ute.com.uy/sites/default/files/generico/Memoria_UTE_2020.pdf (Último acceso: Octubre 2021).

- Uttama, A., Reyes, A., y Prasad, V. (2015). LocED: Location-aware Energy Disaggregation Framework. In *2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, number 10 in BuildSys '15, pages 45–54.
- Villareal, M. y Moreira, J. (2016). Household consumption of electricity in Brazil between 1985 and 2013. *Energy Policy*, 96:251–259.
- Vohra, D. (2016). Apache parquet. In *Practical Hadoop Ecosystem*, pages 325–335. Springer.
- Watts, J. (2015). Uruguay makes dramatic shift to nearly 95 % electricity from clean energy. *The Guardian*.
- Weerakody, P., Wong, K., Wang, G., y Ela, W. (2021). A review of irregular time series data handling with gated recurrent neural networks. *Neurocomputing*, 441:161–178.
- Wenninger, M., Maier, A., y Schmidt, J. (2021). DEDDIAG, a domestic electricity demand dataset of individual appliances in germany. *Scientific Data*, 8(1).
- Yoon, J., Jarrett, D., y van der Schaar, M. (2019). Time-series Generative Adversarial Networks. In *Advances in Neural Information Processing Systems*, number NeurIPS, pages 1–11.

ANEXOS

Anexo 1

Descripción de los datos entregados por UTE

Este anexo describe la estructura original de los datos entregados por UTE y que sirvieron de insumo para la creación del conjunto de datos ECD-UY.

1.1. Consumo eléctrico total de los hogares

El subconjunto de datos de consumo total de hogares contó con dos entregas de archivos: la primera en Noviembre de 2019 y la segunda en Noviembre de 2020. Los datos de la primer entrega (Noviembre de 2019) corresponden al consumo total agregado de 160.513 hogares entre las fechas 1/1/2015 y 1/10/2019. Los datos de la segunda entrega (Noviembre de 2020) corresponden al consumo total de 168.000 hogares entre las fechas 1/1/2015 y el 3/11/2020.

Los archivos correspondieron a información de clientes (`Datos-Clientes.csv`), relación entre medidores y clientes (`Relacion-Medidor-Cliente.csv`) y múltiples archivos de datos de consumo eléctrico (`Datos-Medidores-<N>.csv`). La Tabla 1.1 lista los tipos de valores de cada archivo, junto a una descripción de los mismos.

1.2. Consumo eléctrico de termotanques

Para el subconjunto de consumo eléctrico de termotanques se contó con dos entregas de archivos: la primera en Noviembre de 2019 y la segunda en Noviembre de 2020. Los datos entregados en Noviembre de 2019 corresponden

Datos-Clientes.csv		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
ID Generado	texto	Valor único que identifica al cliente
Tension	texto	Tensión a la cual el cliente se conecta a la red eléctrica (baja: BT, media MT, y un valor en V o kV, e.g., 230V, 15KV)
Tarifa	texto	Código de tipo de tarifa contratada (e.g., TCB, TRS, etc)
DscTarifa	texto	Descripción del tipo de tarifa
PotenciaContratada	número	Potencia contratada, en W
PotenciaPunta	número	Potencia contratada para horario punta, en W
PotenciaValle	número	Potencia contratada para valle punta, en W
PotenciaLlano	número	Potencia contratada para horario llano, en W

Relacion-Medidor-Cliente.csv		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
timer	número	Valor único que identifica al medidor
ID Generado	número	Valor único que identifica al cliente

Datos-Medidores-<N>.csv		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
medidor	número	Valor único que identifica al medidor
dia	número	Registro de fecha, expresado en milisegundos desde el 01-01-1970
periodo	número	Cantidad de períodos de 15 minutos desde el comienzo del día, comenzando en 1
carga	número	Energía consumida en el último período de 15 minutos, en kWh

Tabla 1.1: Descripción de los archivos de información de clientes, relación cliente-medidor y consumo eléctrico del subconjunto de consumo total de los hogares.

al consumo de 521 termotanques entre el 12 de Julio de 2017 y el 30 de Setiembre de 2018. Por otra parte, los datos entregados en Noviembre de 2020 corresponden al consumo de 819 termotanques entre el 2 de Julio de 2019 y el 20 de Noviembre de 2020, destacándose el incremento de la cantidad de hogares que cuentan con información de cliente y consumo total del hogar. De esta forma, se puede contar con información geolocalizada de los hogares y cruzar datos con el subconjunto de consumo eléctrico total. En cuanto al consumo total, la primer entrega contaba con estos datos ya que no existía forma de identificar los clientes de termotanque dentro del subconjunto de consumos totales. Esto fue solucionado en la segunda entrega y por lo tanto ya no existía

la necesidad de archivos de consumo total dentro del subconjunto de consumo de termotanques.

En ambas entregas, los archivos provistos correspondieron a información de clientes (`Cientes.csv`), relaciones clientes-medidores y cliente-*timer* (`Relacion-Cliente-Medidor.csv` y `Relacion-Cliente-Timer.csv`), datos históricos sobre planes contratados por los clientes (`Relacion-Hist-Punto-Consumo-Potencia.csv`, solo en la primer entrega), datos de consumo total (`OM-<MM>-<AAAA>[-<MM>-<AAAA>].csv`, solo en la primera entrega), y datos de consumo de termotanque (`Termotanque-<MM>-<AAAA>.csv`). La Tabla 1.2 describe los tres archivos que contienen relaciones entre clientes, medidores, *timers* y tipo de contratos, por otro lado, la Tabla 1.3 describe el archivo de información de clientes y la Tabla 1.4 describe los dos archivos de registros de consumo eléctrico.

1.3. Consumo eléctrico desagregado de electrodomésticos

El subconjunto de consumo eléctrico desagregado de electrodomésticos contó con una única entrega de dos tipos de archivos. Un archivo contiene la información de los dispositivos monitoreados (`datos-inicio-dispositivos.csv`), mientras que los restantes archivos (`FEED-<AAAA>-<MM>-<DD>-a-<MM>-<DD>-FIng.csv`) contienen registros de tipos múltiples (e.g., consumo eléctrico). Ambos tipos de archivos se describen en detalle en la Tabla 1.5.

El subconjunto de datos registra 13 diferentes señales en los archivos `FEED-<AAAA>-<MM>-<DD>-a-<MM>-<DD>-FIng.csv`. Los códigos de tipo de señales, su descripción y unidad de medida, se listan en la Tabla 1.6.

Relacion-Cliente-Medidor.csv.		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
Meas Point	número	Valor único que identifica al medidor
Identificador	número	Valor único que identifica al cliente
Relacion-Cliente-Timer.csv		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
Timer	número	Valor único que identifica al <i>timer</i> (rotulado <i>id_nodo</i> en segunda entrega)
Identificador	número	Valor único que identifica al cliente (rotulado <i>Id</i> en segunda entrega)
fecha_desde	texto	Fecha de inicio del periodo registrado
fecha_hasta	texto	Fecha de finalización del periodo registrado
Relacion-Hist-Punto-Consumo-Potencia.csv (solo en la primer entrega)		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
Identificador	número	Valor único que identifica al cliente
tension	texto	Tensión a la cual el cliente se conectó a la red eléctrica durante las fechas indicadas
fechaDesde	texto	Fecha de inicio del periodo registrado
fechaHasta	texto	Fecha de finalización del periodo registrado
potenciaContratada	número	Potencia (W) contratada en el período registrado
consumo	número	Consumo eléctrico (kWh) en el período registrado
fmpi	número	Facturación promedio mensual, teniendo en cuenta los últimos seis meses

Tabla 1.2: Descripción de los archivos de relaciones entre cliente-medidor, cliente-*timer*, y cliente-historico de contratos, del subconjunto de consumo eléctrico de termotanques.

<i>Clientes.csv</i>		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
Identificador	número	Valor único que identifica al cliente
TENSION	texto	Type of voltage contracted red[IDEM columna tension de Datos-Cliente.csv ?]
ID_TIPO_TARIFA	texto	Type of tariff contracted
DESC_TARIFA	texto	Description of the type of tariff contracted
POTENCIA_CONTRATADA	número	Contracted power
POTENCIA_LLANO	número	Contracted power on flat time
POTENCIA_PUNTA	número	Contracted power on peak time
POTENCIA_VALLE	número	Contracted power on valley time
depto	número	Department in Uruguay where the household is located, númeroed from 1 to 19 by INE
seccion	número	Censal section in Uruguay where the household is located, assigned by INE
segmento	número	Censal segment in Uruguay where the household is located, assigned by INE
zona	número	Censal zone in Uruguay where the household is located, assigned by INE

Tabla 1.3: Descripción del archivo con información de clientes del subconjunto de consumo eléctrico de termotanques.

OM-<MM>-<AAAA>[-<MM>-<AAAA>].csv y extraccion-OM-<N>.csv		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
Day	número	Fecha, expresada en milisegundos desde el 01-01-1970
Last Update Ms	número	Fecha de la última modificación de marca de consumo, en formato <i>epoch time</i>
Magnitude	número	Código de tipo de señal al que corresponde el registro
Meas Int	número	Intervalos de 15 minutos desde el último registro (constante en 1)
Meas Point	número	Valor único que identifica al medidor
Period	número	Cantidad de períodos de 15 minutos desde el comienzo del día, comenzando en 1
Rowkey	número	Valor que identifica el registro en los sistemas internos de UTE
Source	texto	Origen de la carga
Value	número	Energía consumida en el período, en kWh
Número de registros	N/A	Sin información
Quality	N/A	Sin información
Val Res	N/A	Sin información
Termotanque-<MM>-<AAAA>.csv.		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
Carga	número	Potencia instantánea al momento de la medición
Fecha	texto	Marca de tiempo del instante de la medición, en formato <AAAA>-<MM>-<DD><hh>:<mm>:<ss>.<m>
Id	número	Valor único que identifica al medidor
Rowkey	número	Identificador interno del registro
Rssi	número	Intensidad de la señal del medidor
Voltaje	número	Tensión, en V

Tabla 1.4: Descripción de los archivos de consumo eléctrico total y de termotanques del subconjunto de consumo eléctrico de termotanques.

datos-inicio-dispositivos.csv		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
Usuario	número	Valor único que identifica al cliente
MAC	número	Valor único que identifica al electrodoméstico
Tipo	texto	Descripción del electrodoméstico, tal cual fue dada por el cliente
FEED-<AAAA>-<MM>-<DD>-a-<MM>-<DD>-FIng.csv		
<i>columna</i>	<i>tipo</i>	<i>descripción</i>
user	número	Valor único que identifica al cliente
timesource	texto	Marca de tiempo del instante de la medición, en formato <DD>/<MM>/<AAAA><hh>:<mm>:<ss>
assdid	texto	Valor único que identifica al electrodoméstico
unit	texto	Unidad de medida, dependiendo del tipo de señal (signal)
measure	número	Valor medido al momento del registro, para el tipo de señal correspondiente
signal	texto	Uno de los 13 tipos de señales medida, detalladas en la Tabla 1.6

Tabla 1.5: Descripción de los archivos de información de electrodomésticos y señales medidas (e.g., consumo eléctrico) del subconjunto de consumo eléctrico de electrodomésticos.

<i>código</i>	<i>detalle</i>	<i>unidad</i>
ACTIVEENERGY	Energía activa	Wh
REACTIVEENERGY	Energía reactiva	VArh
ACTIVEPOWER	Potencia activa	W
CALCULATEDACTIVEPOWER	Potencia activa calculada	W
AMBIENTTEMP	Temperatura ambiente	°C
SETPOINT	Temperatura de referencia	°C ×100
PHASE1VOLTAGE	Tensión fase 1	V
PHASE2VOLTAGE	Tensión fase 2	V
PHASE3VOLTAGE	Tensión fase 3	V
PHASE1CURRENT	Corriente fase 1	A
PHASE2CURRENT	Corriente fase 2	A
PHASE3CURRENT	Corriente fase 3	A
POWERFACTOR	Factor de potencia	N/A

Tabla 1.6: Detalle de las señales medidas en el subconjunto de consumo eléctrico de electrodomésticos.