

Trabajo Fin de Grado  
Grado en Ingeniería de las Tecnologías Industriales

## **Desglose Automático de Consumos Energéticos en Edificios Terciarios y Obtención de Líneas Base**

Autor: Álvaro Harana Martín-Arroyo

Tutor: Servando Álvarez Domínguez

Cotutor: M<sup>a</sup> del Carmen Guerrero Delgado

Dpto. Ingeniería Energética  
Escuela Técnica Superior de Ingeniería  
Universidad de Sevilla

Sevilla, 2019





Trabajo Fin de Grado  
Grado en Ingeniería de las Tecnologías Industriales

# **Desglose Automático de Consumos Energéticos en Edificios Terciarios y Obtención de Líneas Base**

Autor:

Álvaro Harana Martín-Arroyo

Tutor:

Servando Álvarez Domínguez

Catedrático de universidad

Cotutor:

M<sup>a</sup> del Carmen Guerrero Delgado

Predoctoral FPI Mineco

Dpto. de Ingeniería Energética  
Escuela Técnica Superior de Ingeniería

Universidad de Sevilla

Sevilla, 2019



Trabajo Fin de Grado: Desglose Automático de Consumos Energéticos en Edificios Terciarios y Obtención de Líneas Base

Autor: Álvaro Harana Martín-Arroyo

Tutor: Servando Álvarez Domínguez

Cotutor M<sup>a</sup> del Carmen Guerrero Delgado

El tribunal nombrado para juzgar el Proyecto arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla, 2019

El Secretario del Tribunal



*A mis padres, hermana y  
pareja.*



# AGRADECIMIENTOS

---

Este trabajo fin de grado culmina una etapa que pronto empezó y en el tiempo perduró, pero como todo, un final tiene. Este momento no hubiera sido posible sin el apoyo incondicional de mi familia, que nunca dudaron de mí y no les faltó unas palabras de apoyo y aliento en los momentos más necesarios. Reconocer y agradecer infinitamente la predisposición para realizar el gran esfuerzo que supone ofrecer a un hijo la posibilidad de estudiar a libre elección aquello que deseaba ser.

Agradecer también los esfuerzos de M<sup>a</sup> del Carmen Guerrero que aún sabiendo de los recursos de tiempo limitados que tenía, aceptó sin dudar trabajar conmigo para llegar a buen fin.

Finalmente, no quiero olvidar a mi pareja quien ha sufrido los daños colaterales de convivir con alguien que muchas veces decía... “no puedo porque tengo que estudiar” y tener que aguantar ratos de encerrona acompañándome.

A todos ellos, gracias.

Álvaro Harana Martín-Arroyo

Sevilla, 2019

# RESUMEN

---

El cambio en materias energéticas es una realidad y una obligación para la sociedad, la concienciación del uso de la energía y del ahorro que todos deberíamos de buscar pone de manifiesto la necesidad de herramientas que permitan la visualización y el estudio de los consumos energéticos que realizamos, para poder actuar.

Las líneas que continúan tratarán de definir un método o procedimiento que permita al usuario desglosar el consumo energético total de un edificio terciario, de tal manera que a partir de consumos globales se pueda obtener consumos desglosados. A parte del desglose, se detallará una línea base de consumos que será la encargada de estimar los valores desglosados del consumo en momentos futuros como así se lleva a cabo en cualquier plan de Measurement and Verification (M&V).

Con todo esto, se pretende solucionar la necesidad de poder cuantificar y luego en un paso posterior calcular o estimar los consumos desglosados de una edificación terciaria, con el objetivo de aplicar medidas de ahorro específicas o detectar anomalías de consumo a tiempo para que los edificios sean más eficientes y reduzcan los gastos brutos de energía que consumen.



# ÍNDICE

---

<b>AGRADECIMIENTOS</b>	<b>VIII</b>
<b>RESUMEN</b>	<b>IX</b>
<b>ÍNDICE</b>	<b>XI</b>
<b>ÍNDICE DE TABLAS</b>	<b>XIII</b>
<b>ÍNDICE DE FIGURAS</b>	<b>XV</b>
<b>1 1 INTRODUCCIÓN</b>	<b>1-1</b>
1.1 <i>Contexto. Problemática y normativa asociada.</i>	1-2
1.1.1 ASHRAE Guideline.	1-3
1.1.2 IPMVP. EVO.	1-3
1.1.3 ISO 50006:2014	1-4
1.1.4 Conclusión general literatura y guías actuales.	1-6
1.2 <i>Modelos de línea base</i>	1-7
1.2.1 Fundamentos.	1-7
1.2.2 Tipos de modelos de línea base.	1-7
1.3 <i>Desglose de consumos</i>	1-10
1.4 <i>Clustering.</i>	1-10
1.4.1 Tipos de medidas de distancia.	1-12
1.4.2 Tipos de índices de evaluación.	1-13
1.4.3 Técnicas jerárquicas de clustering.	1-14
1.4.4 Técnicas no jerárquicas de clustering.	1-18
1.5 <i>Objetivos</i>	1-20
<b>2 2 METODOLOGÍA</b>	<b>2-21</b>
2.1 <i>Alcance.</i>	2-21
2.2 <i>Procedimiento de desglose de consumos.</i>	2-21
2.2.1 Antecedentes.	2-21
2.2.2 Situación de partida.	2-22
2.2.3 Procedimiento.	2-22
2.3 <i>Validación del procedimiento de desglose de consumos.</i>	2-29
2.4 <i>Procedimiento caracterización energética</i>	2-30
2.4.1 Línea base de consumos.	2-30
2.4.2 Utilidades. Gestión energética, detección de anomalías.	2-32
<b>3 3 CASO TEÓRICO. Edificio de oficinas</b>	<b>3-33</b>
3.1 <i>Alcance.</i>	3-33
3.2 <i>Ubicación y arquitectura del edificio problema.</i>	3-33
3.3 <i>Análisis de consumos del edificio a estudio.</i>	3-35
3.3.1 Uso y ocupación del edificio.	3-36
3.3.2 Climatología.	3-37
3.4 <i>CASOS A ESTUDIO.</i>	3-39
3.5 <i>CASO 1. Desglose de consumos.</i>	3-40
3.5.1 CASO 1. FASE 1. Preparación de los datos de entrada.	3-40
3.5.2 CASO 1. FASE 2. Clustering.	3-41
3.5.3 CASO 1. FASE 3. Desglose de consumos.	3-44
3.6 <i>CASO 2. Desglose de consumos.</i>	3-48

3.6.1	CASO 2. FASE 1. Preparación de los datos de entrada.	3-48
3.6.2	CASO 2. FASE 2. Clustering.	3-49
3.6.3	CASO 2. FASE 3. Desglose de consumos.	3-51
3.7	<i>CASO 3. Desglose de consumos.</i>	3-54
3.7.1	CASO 3. FASE 1. Preparación de los datos de entrada.	3-54
3.7.2	CASO 3. FASE 2. Clustering.	3-54
3.7.3	CASO 3. FASE 3. Desglose de consumos.	3-55
3.8	<i>Conclusiones.</i>	3-57
3.9	<i>Línea base de consumos.</i>	3-58
3.9.1	Edificio de oficinas. Línea base de consumos.	3-58
3.9.2	Estimación de Consumos totales. Resultados.	3-61
<b>4</b>	<b>4 CASO REAL. Colegio de arquitectos de cataluña.</b>	<b>4-62</b>
4.1	<i>Ubicación y arquitectura del edificio.</i>	4-62
4.2	<i>Análisis del consumo.</i>	4-63
4.3	<i>Climatología. Modos de operación.</i>	4-63
4.4	<i>Variables de entrada al procedimiento.</i>	4-64
4.5	<i>Clustering.</i>	4-65
4.6	<i>Desglose de consumos.</i>	4-66
4.6.1	Consumo fijo.	4-66
4.6.2	Consumo de ocupación.	4-66
4.6.3	Desglose. Estimación mensual.	4-67
4.7	<i>Línea base de consumos.</i>	4-67
4.7.1	Colegio de Arquitectos de Cataluña. Línea base de consumos.	4-67
4.7.2	Estimación de Consumos totales. Resultados.	4-69
<b>REFERENCIAS</b>		<b>I</b>

# ÍNDICE DE TABLAS

---

Tabla 1. Proceso, método de la distancia mínima 1.	1-15
Tabla 2. Proceso, método de la distancia mínima 2.	1-15
Tabla 3. Ejemplo clúster y componentes.	2-26
Tabla 4. Fragmento datos de partida, CASO 1, edificio teórico.	3-40
Tabla 5. Medidas estadísticas variables de entrada.	3-40
Tabla 6. Variables normalizadas, CASO 1, edificio teórico.	3-41
Tabla 7. Pruebas clustering, CASO 1, edificio teórico.	3-42
Tabla 8. Clústeres, CASO 1, edificio teórico.	3-43
Tabla 9. Clústeres variables normalizadas, CASO 1, edificio teórico.	3-43
Tabla 10. Cálculo consumo fijo, CASO 1, edificio teórico.	3-45
Tabla 11. Cálculo consumo de ocupación, CASO 1, edificio teórico.	3-45
Tabla 12. Resultados desglose de consumos, variante 1, CASO 1, edificio teórico.	3-46
Tabla 13. Errores, variante 1, CASO 1, edificio teórico.	3-47
Tabla 14. Errores, variante 2, CASO 1, edificio teórico.	3-47
Tabla 15. Datos de partida, CASO 2, edificio teórico.	3-48
Tabla 16. Variables normalizadas, CASO 2, edificio teórico.	3-48
Tabla 17. Pruebas clustering, CASO 2, edificio teórico.	3-49
Tabla 18. Resultados clústeres, CASO 2, edificio teórico.	3-50
Tabla 19. Cálculo consumo fijo, CASO 2, edificio teórico.	3-51
Tabla 20. Cálculo consumo ocupación, CASO 2, edificio teórico.	3-51
Tabla 21. Errores, variante 1, CASO 2, edificio teórico.	3-53
Tabla 22. Errores, variante 2, CASO 2, edificio teórico.	3-53
Tabla 23. Pruebas clustering, CASO 3, edificio teórico.	3-54
Tabla 24. Cálculo consumo fijo, CASO 3, edificio teórico.	3-55
Tabla 25. Cálculo consumo ocupación, CASO 3, edificio teórico.	3-55
Tabla 26. Errores, variante 1, CASO 3, edificio teórico.	3-56
Tabla 27. Errores, variante 2, CASO 3, edificio teórico.	3-57
Tabla 28. Datos de partida, línea base, caso teórico.	3-59
Tabla 29. Coeficientes, línea base de refrigeración, caso teórico.	3-59
Tabla 30. Coeficientes línea base de calefacción, caso teórico.	3-59
Tabla 31. Pruebas clustering, edificio real.	4-65
Tabla 32. Clústeres, caso real.	4-66
Tabla 33. Cálculo consumo de ocupación, caso real.	4-66
Tabla 34. Coeficientes línea base refrigeración, caso real.	4-68
Tabla 35. Coeficientes, línea base calefacción, edificio real.	4-68



# ÍNDICE DE FIGURAS

---

Figura 1. Consumo de energía primaria. Fuente BP Statistical Review	1-1
Figura 2. Consumo de energía primaria en España. Fuente BP Statistical Review	1-1
Figura 3. Desglose de consumos en edificio terciario.	1-2
Figura 4. Esquema de ruta M&V ASHRAE Guideline	1-3
Figura 5. Construcción y uso de la línea base.	1-4
Figura 6. Tipos de variables según ISO 50006:2014	1-5
Figura 7. Línea base mediante regresión	1-8
Figura 8. Fundamento SVM.	1-8
Figura 9. Fundamentos clustering.	1-9
Figura 10. Fundamentos modelo de caracterización inversa.	1-9
Figura 11. Centroides y clústeres.	1-11
Figura 12. Clasificación técnicas de clustering	1-11
Figura 13. Dendograma.	1-14
Figura 14. Metodología. Procedimiento de Desglose Automático.	2-23
Figura 15. Procedimiento de búsqueda de número óptimo de clústeres	2-24
Figura 16. Matlab, obtención de clústeres.	2-25
Figura 17. Ocupación, variable tipo día.	2-27
Figura 18. Procedimiento de validación del desglose de consumos.	2-29
Figura 19. Estimación, línea base de consumos.	2-31
Figura 20. Bandas de error de la estimación de consumos.	2-32
Figura 21. Ubicación edificio caso teórico.	3-33
Figura 22. Coordenadas caso teórico.	3-34
Figura 23. Modelo 3D caso teórico.	3-34
Figura 24. Planta edificio caso teórico.	3-34
Figura 25. Consumo real mensual desglosado, caso teórico.	3-36
Figura 26. Ocupación edificio caso teórico.	3-36
Figura 27. Evolución consumo temperatura media exterior.	3-37
Figura 28. Cálculo temperatura equivalente real.	3-38
Figura 29. Cálculo temperatura equivalente sin desglose previo.	3-38
Figura 30. Descripción variables de las pruebas a realizar para desglose en edificio teórico.	3-39
Figura 31. Número óptimo de clústeres edificio teórico.	3-42
Figura 32. Hetmap clústeres CASO 1, edificio teórico.	3-44
Figura 33. Desglose de Consumos Estimado, CASO 1, edificio teórico.	3-46
Figura 34. Hetmap clústeres CASO 2, edificio teórico.	3-50
Figura 35. Consumo clima, desviaciones, CASO 2, edificio teórico.	3-52
Figura 36. Desglose de consumos estimados, CASO 2, edificio teórico.	3-52

Figura 37. Hetmap clústeres CASO 3, edificio teórico.	3-55
Figura 38. Desglose de consumos estimados, CASO 3, edificio teórico.	3-56
Figura 39. Líneas bases, validaciones, edificio teórico.	3-60
Figura 40. Ajuste del Consumo estimado de climatización.	3-61
Figura 41. Ubicación construcción caso real.	4-62
Figura 42. Fachada caso real. Colegio de Arquitectos de Cataluña.	4-62
Figura 43. Cálculo temperatura equivalente, caso real.	4-64
Figura 44. Definición del número óptimo de clústeres.	4-65
Figura 45. Desglose estimado de consumos, caso real.	4-67
Figura 46. Validaciones líneas bases, caso real.	4-68
Figura 47. Ajuste de consumos línea base caso real.	4-69



# 1 INTRODUCCIÓN

El desarrollo económico de un país está ligado fuertemente tanto a la infraestructura energética que posee como a su consumo. Los informes del panorama energético actual muestran que en el año 2017 ha habido un incremento importante del 2,2% en el consumo de energía primaria [1]. El uso del carbón sigue aumentando en países en pleno desarrollo como es India, fuertemente dependiente de este recurso [2], y las emisiones de CO2 globales a la atmósfera se incrementan en un 1,6% en el último año [1].

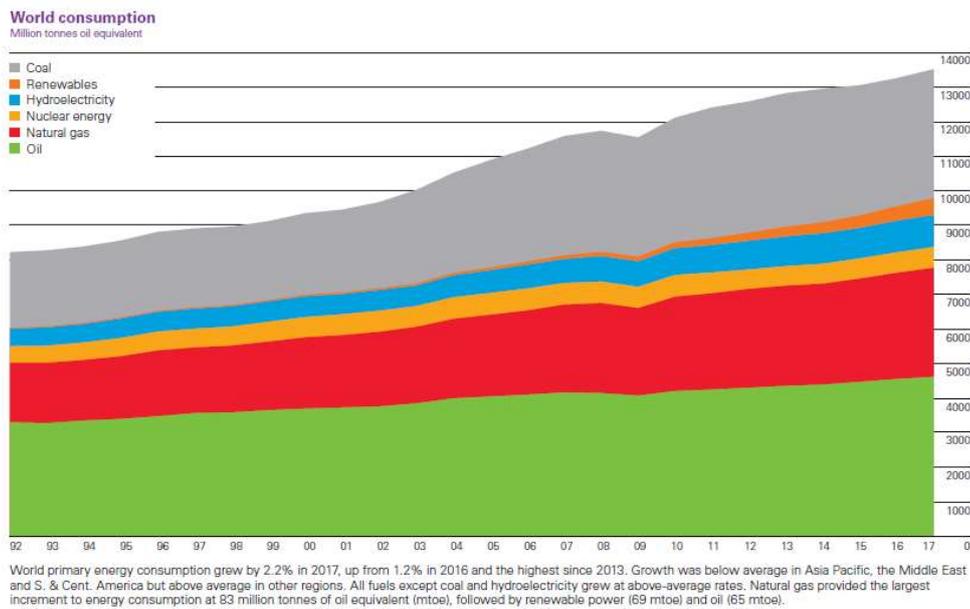


Figura 1. Consumo de energía primaria. Fuente BP Statistical Review

El escenario español no cambia mucho a lo largo de los años, mostrando una fuerte dependencia de las fuentes de energía no renovables, lo cual no ayuda a mejorar las predicciones a nivel global.

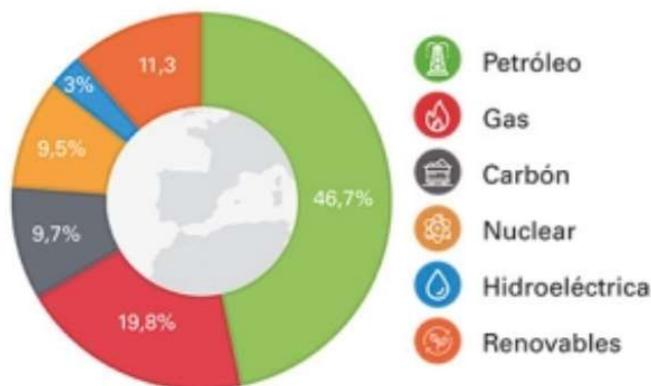


Figura 2. Consumo de energía primaria en España. Fuente BP Statistical Review

Todo esto lleva a pensar en la necesidad de cambio en materias energética y medioambiental como así muestran las políticas y acuerdos que se están llevando a cabo, que tienen como objetivo en el caso europeo, de conseguir un ahorro del 20% mejorando la eficiencia energética en 2020[3].

En la Unión Europea, el consumo energético en edificios supone un 40% del total y las emisiones de CO2 de estos un 36% [4]. Además, tal y como se estudia en [5], estos consumos seguirán incrementándose causados por el calentamiento global que hará sus efectos en los edificios si no se trata de implementar medidas de eficiencia energética en climatización.

La climatización en algunos edificios supone un 70% del consumo energético total del edificio en Europa [6,7], variando en función de la zona climática, meteorología, uso del edificio, ocupación, etc. Lo importante de este dato es ver el peso en temas de reducción de consumo que tendría la puesta en marcha de medidas eficientes en climatización, es por ello, que el Parlamento Europeo propone medidas a medio y largo plazo en renovación de edificios y en edificaciones nuevas. Como ejemplo a lo expuesto, el desglose del consumo energético de la terminal de un aeropuerto a continuación.

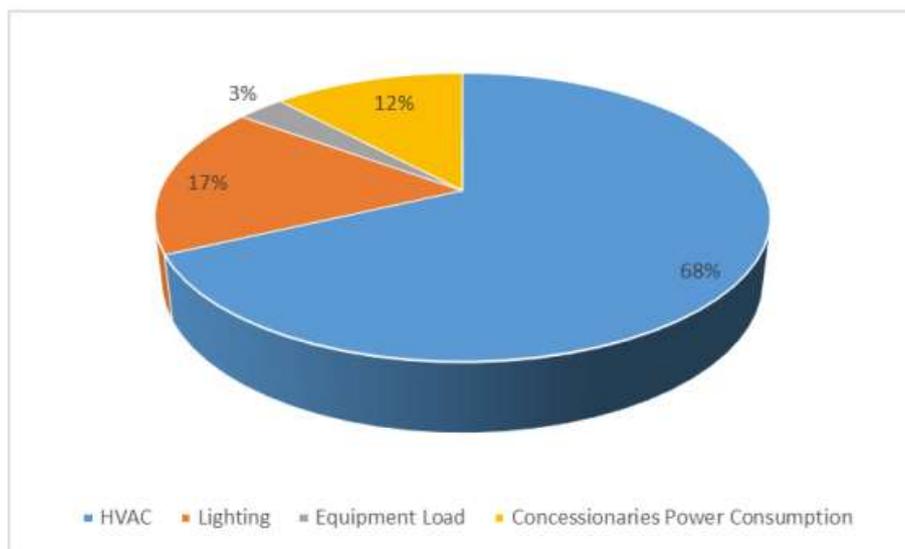


Figura 3. Desglose de consumos en edificio terciario.

Todo lo descrito en el anterior párrafo no solo contribuiría medioambientalmente hablando, sino que proporcionaría una reducción del gasto económico en energía del edificio y una rentabilidad y una tasa de recuperación de la inversión positiva.

El principal objetivo de esta introducción es mostrar la necesidad de cambio, concretamente en el apartado de confort térmico en edificios residenciales, que se caracteriza por tener mucho potencial de mejora. Y para todo esto como tarea exigida para conseguir los compromisos y como aval de los proyectos, la evaluación y estudio de las posibles mejoras y del ahorro o reducción de consumo para los edificios existentes, se convierte en requisito fundamental.

## 1.1 Contexto. Problemática y normativa asociada.

En este apartado se realizará un breve desarrollo de la normativa actual en torno a la estimación y cálculos de consumos en edificios que realizan las organizaciones para distintos fines como valorar proyectos de mejora, estudiar la eficiencia energética o detectar errores o problemas en el consumo.

El potencial de mejora en temas de consumos en climatización en un edificio es una realidad, mucho más alto que en otros aspectos, pero no hay una literatura extensa y/o procedimientos detallados que desarrollen el cálculo de estos consumos a partir de una situación de partida generalista.

La línea base es la herramienta principal que permitirá conocer el gasto en función de las variables de entrada para un supuesto futuro, para el caso que nos concierne, la climatización y tratamiento del aire, esta herramienta necesitará de los consumos de climatización del edificio a estudio. La situación común de inicio de cualquier programa de mejora no es unos datos particionados y claramente desglosados del consumo para poder alimentar dicha línea base, y es esta la razón por la que nace este proyecto, el cual es la fase inicial que permitirá conocer el reparto de consumos de una edificación y así poder valorar la posibilidad de comenzar un estudio de eficiencia y posteriormente usar los datos obtenidos para realizar los cálculos necesarios.

A continuación, se enumerará distintas instituciones, guías y normas que dedican su esfuerzo a los programas de M&V y a la eficiencia energética para disminuir el consumo energético de las construcciones.

### 1.1.1 ASHRAE Guideline.

Es una guía creada por ASHRAE, una asociación tecnológica que se centra en la eficiencia energética en calefacción, refrigeración y calidad de aire interior de los edificios. En su documento, la organización detalla que el principal fin de éste es valorar el ahorro energético tras una ECM (Energy Conservation Measure), es decir, la implementación de una medida de mejora de la eficiencia energética. Para ello habla que la comparación entre una situación pre-mejora y la situación post-mejora no es suficiente, ya que en estos dos escenarios hay factores externos influyentes como cambios en condiciones de meteorología o de ocupación, y no es posible valorar el impacto energético real que la ECM ha causado.

El texto da de forma generalista un enfoque global del proceso de M&V (Medición y Verificación) por el cual se crea una línea base y se obtiene los ahorros energéticos tras un periodo de medición y confección de la herramienta matemática y otro periodo final de comparación entre lo estimado y la medición de los consumos reales tras la implantación de la mejora.

Este documento a rasgos generales da idea al lector de todo el proceso de M&V, las técnicas de medición y calibración, la valoración del error obtenido, pero en su contra no detalla específicamente la obtención de una línea base de forma concisa, solo dando información acerca de la técnica de Modelos de Regresión Lineal, no muy recomendable debido a la poca aproximación y minimización del error que este método tiene.

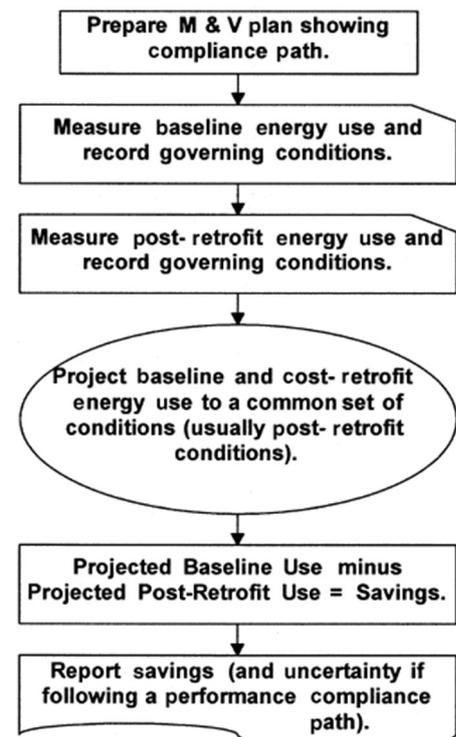


Figura 4. Esquema de ruta M&V ASHRAE Guideline

### 1.1.2 IPMVP. EVO.

International Performance Measurement and Verification Protocol, creado por EVO (Efficiency Valuation Organization). Al igual que en la ASHRAE Guideline, en su texto se desarrolla el proceso de M&V tradicional y se dan las bases para aplicar el método en proyectos de mejora de la eficiencia energética. Esta organización desarrolla a parte del IPMVP otras guías y protocolos, aparte de certificaciones, para ayudar a la sociedad a desarrollarse en temas energéticos. Entre estas, la guía IEEFP provee al lector las herramientas necesarias para la evaluación económica y de riesgos para proyectos de mejora de eficiencia energética.

Las distintas etapas del plan de M&V que se detallan en el IPMVP [8] son las siguientes:

- a) Asignación de las responsabilidades y evaluación del riesgo del proyecto.
- b) Definición del alcance y medición de la situación inicial.
- c) Creación del modelo de línea base.
- d) Instalación de las medidas de ahorro energético.
- e) Medición del consumo post-implementación y evaluación de los ahorros.

f) Realización de reportes periódicos y control.

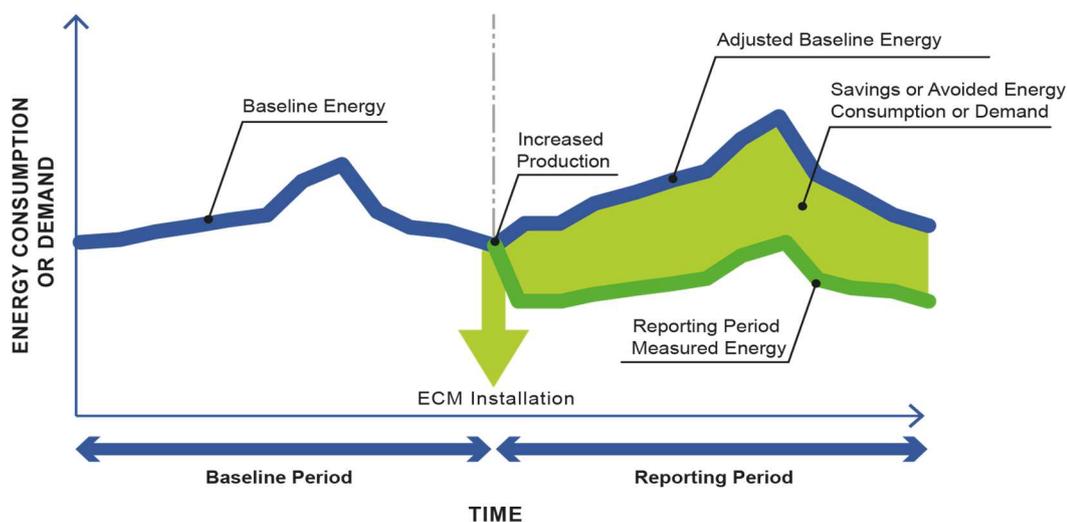


Figura 5. Construcción y uso de la línea base.

En el IPMVP podemos encontrar varios métodos de medidas y consiguientemente de análisis en el plan de M&V. Estos métodos se clasifican en dos tipos dependiendo del enfoque del plan, la diferencia entre ellos está en considerar el equipo a instalar / instalación donde se realiza la mejora aislado del resto de influencias que le rodea. En la guía se clasifican los métodos aislado A y B y los no aislados C y D.

A: método más simple. La predicción del consumo se basará en un único parámetro clave (key parameter), el resto de parámetros necesarios para la confección del modelo de línea base se obtendrán por aproximación o por datos históricos. El periodo de medición será corto y se realizarán mediciones periódicas o de forma continua. El ahorro se obtendrá por cálculos y estimaciones.

B: se medirán varios parámetros de forma aislada, en un breve periodo de tiempo y de forma continua o periódica, se obtendrá la línea base. Las mediciones se prolongarán tras la implantación de la ECM y se calculará el ahorro como la diferencia entre la estimación (tras ajustes si es necesario) menos la medición post-mejora.

C: se realizará una medición continua del consumo en un periodo de medición y post-mejora, se construirá una línea base con herramientas de regresión que dependerán de variables independientes. Es necesario conocimiento de las distintas influencias de las instalaciones, así como los equipos existentes y todo lo que afectará al consumo.

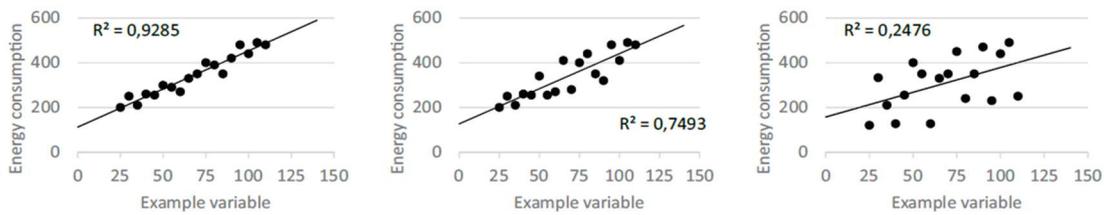
D: necesitará de la calibración de los modelos con los consumos reales de la instalación. Se requiere métodos de computación para el modelo y que éste contemple todas las variables influyentes en temas de equipos, perfiles de uso, rendimientos...

### 1.1.3 ISO 50006:2014

La norma ISO50006 del 2014 pretende dar una guía al lector de como cumplir los requerimientos exigidos en la norma ISO50001 en materias de eficiencia energética. Para este propósito define en su texto el uso EnPI's y EnBI's, indicadores de rendimiento energético y líneas bases de consumo y rendimiento.

EnPI's: parámetros clave de un equipo, instalación o edificio completo que proveerá al analista una imagen relevante de la situación y permitirá tomar decisiones de mejora en función de los datos. En la imagen que sigue, extraída de la propia norma, se puede ver el proceso de selección de aquellas variables que podrían formar parte del conjunto óptimo de las mismas:

EnB's: líneas base que son referencias de consumo o rendimiento para un periodo determinado de tiempo, que permitirán evaluar cambios en eficiencia entre dos puntos en el tiempo y evaluar ahorros antes y después de llevar a cabo una mejora energética en las instalaciones.



a) Significant variable

b) Less significant variable

c) Not significant variable

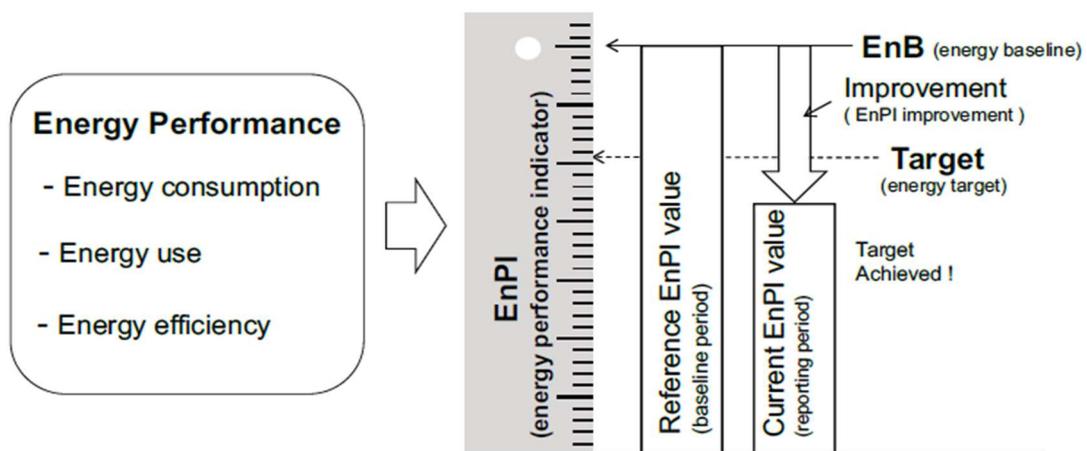


Figura 6. Tipos de variables según ISO 50006:2014

La norma define las siguientes etapas para la calificación de la eficiencia energética y la posterior toma de decisiones:

1) Recaudación de información inicial.

Antes de comentar el proyecto de evaluación, será necesario analizar evaluaciones anteriores en el caso en el que las hubiere, marcar el alcance del proyecto actual y definir el límite de las EnPI's a utilizar.

Al igual que se detalla en el plan de M&V de EVO, es posible analizar la eficiencia energética desde varios tamaños de vista o enfoques, para entenderlo el siguiente ejemplo muestra los tres principales enfoques que se podría llevar a cabo en unas instalaciones para la medición y verificación: habitación, planta 1ª del edificio o edificio completo. Esto anterior marcaría los límites de los EnPI's que se utilizarían en el análisis

A parte de todo esto, en esta etapa es necesario identificar los flujos energéticos existentes en nuestro estudio y las variables de salida del mismo, ya que es posible que para un proyecto sea más interesante analizar el consumo total, pero en otro sea más interesante medir el consumo por tonelada producida de acero, por ejemplo.

2) Definición de indicadores energéticos.

Es necesario realizar un estudio del consumo para definir las variables clave de las que depende principalmente. En la definición de EnPI's hemos visto el análisis de la correlación entre una variable y el objetivo de estudio o salida (consumo generalmente), esto y el análisis de lo viable que es realizar mediciones de esta y la independencia de una en particular del resto que se utilizarán harán que se conforme el conjunto de variables clave que definirán nuestro sistema y harán posible la evaluación del mismo.

Entre las variables o EnPI's más comunes están la radiación solar, la temperatura interior o exterior, el día del año, la hora, etc.

3) Líneas base.

Las líneas bases se construirán a partir de los valores de los EnPI's en el periodo de medición, resultando un modelo del sistema a estudiar dependiente de los valores de las mediciones que ofrecerá la posibilidad de obtener las variables representativas que se quiera estimar posteriormente.

Una vez definidos los objetivos, que se quiere obtener de estas líneas bases, comenzará el proceso de medición para la construcción de esta. Se acordará el periodo de medición, se desarrollará el modelo y posteriormente se validará comprobando los resultados con mediciones reales.

4) Periodo de reporte, uso de los indicadores y líneas base.

Cuando el fin de las líneas bases no sean la mera comparación entre dos situaciones parecidas, se procederá a la normalización de los datos y resultados para poder valorar de mejor manera cualquier cambio o desarrollo aplicado a la instalación.

En este periodo se procederá a reportar los resultados obtenidos entre la comparación de la situación real y lo calculado mediante las líneas bases. Esto permitirá conocer la diferencia en términos absolutos, ratios o porcentuales de una situación estimada por la experiencia anterior y la real o de una situación estimada en condiciones pasadas y la situación post-mejora después de implementar medidas de eficiencia.

5) Mantenimiento de los EnPI's y EnB's.

Tras el paso del tiempo y el desarrollo de cambios en las instalaciones medidas, es posible que las mediciones de los indicadores y los resultados de las líneas base ya no sean de calidad como en situaciones pasadas. Después de la observación de lo descrito, será necesario desarrollar cambios o nuevos indicadores y ajustar las líneas base.

Para actualizar los resultados de las líneas base se usarán los datos del periodo de medición para desarrollar un modelo estadístico que permita recalcular los valores corregidos en la situación actual, o se usarán datos en condiciones estándar en el caso de no haber mediciones, para el mismo fin.

#### 1.1.4 Conclusión general literatura y guías actuales.

EVO con el IPMV junto con ASHRAE con su ASHRAE's Guideline [9] son las dos organizaciones más representativas en la actualidad al hablar de los planes de M&V de eficiencia energética al igual que la norma ISO 50006:2014. En sus textos se desarrolla el proceso de M&V tradicional y se dan las bases para aplicar el método en proyectos de mejora de la eficiencia energética.

La medición y verificación es un proceso costoso y largo, pero es clave para demostrar si se logra un ahorro de energía al desplegar medidas de eficiencia energética, reducir el riesgo del proyecto y detectar anomalías posteriores [10]. Las guías y normativa anteriormente expuestas detallan el método tradicional de M&V, el cual queda obsoleto por no capturar con precisión las tendencias complejas del comportamiento energético de las edificaciones [11]. Es por ello que en la actualidad se está investigando y desarrollando técnicas novedosas de medición y verificación más conocidas como M&V 2.0.

Los planes de M&V 2.0 se basan en algoritmos de aprendizaje automático "machine learning" que tratan grandes cantidades de datos medidos para modelar el consumo futuro de un edificio de forma más precisa [12]. En los más usados encontramos las técnicas MR [11,24,25], GP, ANN [13,28], DNN [13] y SVM[26,27]. Y entre estos destaca el ANN que imita el trabajo del cerebro humano, realizando un aprendizaje y modificando su estructura mediante la retroalimentación [13] consiguiendo un error del 6%. Estos métodos usan como inputs variables generalistas: temperatura, humedad, hora, día y ocupación (muy influyente) [14,15] y adaptan las frecuencias y duraciones del proceso a los requerimientos del proyecto [16], de tal manera que se optimizan costes (llegando a un ahorro del 11% para un nivel de incertidumbre válido [17,18]).

En conclusión, se investiga nuevas técnicas óptimas y de menor error, pero las guías no se actualizan. Lo expuesto en apartados anteriores sirve a modo de guía, pero no se detallan los métodos de obtención de línea base y estimación de forma extendida. Es por ello que finalizamos esta conclusión comentando la necesidad de información en los métodos utilizados, así como métodos generalistas y adaptables para cualquier tipo de edificación y emplazamiento de la misma.

## 1.2 Modelos de línea base

### 1.2.1 Fundamentos.

Una de las fases en el programa de M&V es la modelización de la línea base. Esta curva permitirá predecir el consumo energético del edificio que se estudia, para verificar el ahorro energético obtenido tras realizar un proyecto de mejora de eficiencia. Dado el impacto económico que un proyecto supone, esta fase se convierte en una de las más importantes del M&V [19,20]. Se distinguen dos situaciones: la situación de modelado, en la que se recaba información para conseguir un ajuste óptimo del modelo de línea base, y la situación post mejora, en la cual, una vez realizada la mejora de eficiencia, es posible obtener el consumo real y el consumo que tendría el edificio si no se hubiese implementado dicha mejora, gracias al uso de la línea base.

### 1.2.2 Tipos de modelos de línea base.

Son numerosos los distintos modelos de línea base que se desarrollan en la actualidad, entre ellos encontramos tanto los modelos específicos de benchmarking para construcciones y edificaciones determinadas [21-23] así como modelos matemáticos. A continuación, se da una breve visión de estos:

- Modelos de Benchmarking.  
Estos modelos generalistas, usan la regresión para estimar el consumo energético del edificio en estudio a partir de las mediciones reales de otros edificios de la misma tipología. Este método es independiente de la infraestructura del edificio, eficiencias de los equipos y de los materiales constructivos. Su validez se compromete a la hora de elegir los indicadores de los que dependerá el resultado. Comparando los indicadores medidos en el edificio de estudio y los de los edificios dato, tras la normalización de los mismos, se realiza una regresión para obtener los valores estimados del consumo.
- Modelos de regresión.  
Son los más sencillos y utilizados para realizar los estudios de línea base. En las guías más famosas se incluyen anexos donde describen el proceso de construcción de los mismos. Estos modelos se caracterizan por realizar una aproximación de una variable salida (dependiente) en función de los valores de una o varias variables entrada (independientes). Con ello se obtiene una ecuación lineal (caso regresión simple) que estima el valor de una forma sencilla, las principales desventajas de este modelado son la poca correlación de los resultados vs los valores reales, además de no tener en cuenta momentos pasados. Es razonable pensar que las variables a estudiar mantienen una relación sencilla respecto a las entradas, pero en la realidad raramente ocurre. Los modelos de regresión lineal y polinómica tienen las siguientes etapas:
  - 1.- Definición de la/s líneas bases a utilizar: a veces es interesante modelar el modelo a partir de varias líneas bases dependientes entre ellas.
  - 2.-Elección de las variables independientes: dada la poca complejidad del modelo y por consiguiente su coste, es importante realizar un buen estudio sobre que variables son las ideales y óptimas para medir e incluir en nuestras ecuaciones.
  - 3.-Construcción del modelo: a partir de los valores medidos es fácil llevar a cabo un análisis de los datos y construir el modelo lineal (ecuación de regresión lineal). Este paso se simplifica bastante dadas las herramientas informáticas que pueden ayudarnos para este fin, como puede ser Microsoft Excel.
  - 4.-Validación: tras el periodo de medición y construcción es obligado validar las líneas bases y estudiar el error obtenido. En el caso en el que no se obtengan valores aceptables sería necesario pensar en incluir alguna variable más en el/los modelos, aumentar el número de mediciones o utilizar otro método de modelado de la línea.

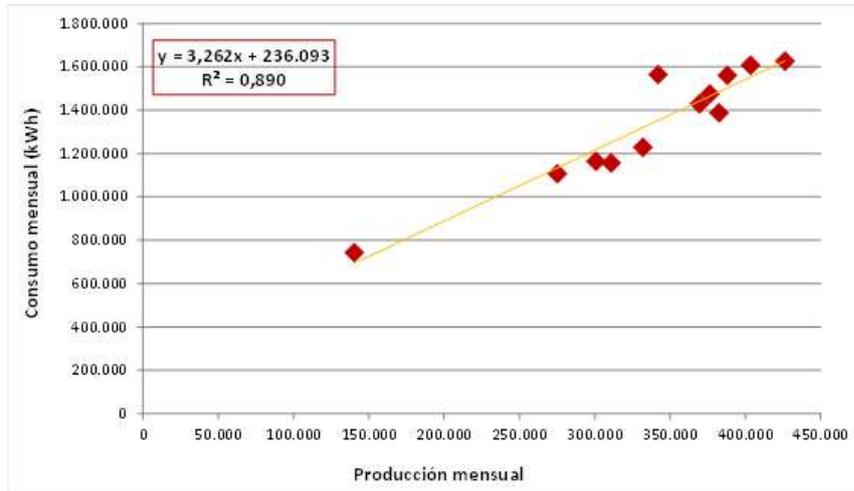


Figura 7. Línea base mediante regresión

- Modelos computacionales. Machine Learning.

Estos algoritmos son la base de los modelos de M&V 2.0, se caracterizan por usar algoritmos matemáticos de alto nivel para agrupar y clasificar datos mediante métodos complejos de agrupamiento y así realizar luego estimaciones. Dependiendo del método de clasificación que realicen se distinguen tres grupos: geométricos basados en distancias, probabilísticos y lógicos basados en comportamiento (árboles de decisión). Entre sus ventajas destaca los valores de error que se obtienen y entre sus desventajas encontramos la necesidad de especialización y conocimientos de aquel que lo modela y el alto nivel computacional necesario para llevarlos a cabo en comparación con otros métodos. Se detallan a continuación tres de ellos:

SVM: máquina de vectores soporte. Es una técnica de reconocimiento de patrones que basa su funcionamiento en la transformación y simplificación del espacio dimensión de los datos utilizados haciendo uso de funciones kernel. Cada medición es un vector que consta de varios atributos (por ejemplo: día, tipo de día, radiación, temperatura, etc.) y la salida (consumo eléctrico). Haciendo uso de los multiplicadores de Lagrange se simplifica el sistema obteniendo una función que determina la variable salida para una entrada de las variables atributos. Este método tiene ciertas desventajas, en el proceso de construcción (entrenamiento) requiere que los datos estén bien separados para poder hacer una segmentación precisa. Otra cuestión importante es que este método es utilizado mayormente para obtener variables de salida discretas, es decir este método no es el más acertado a la hora de determinar un consumo (variable continua), aunque si sería útil si este consumo tomara por ejemplo tres valores: consumo alto (1), consumo medio (2), consumo bajo (3).

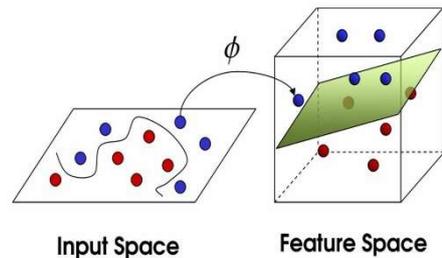


Figura 8. Fundamento SVM.

GP: regresión gaussiana. Este método probabilístico permite obtener el valor de una observación desconocida, basándose en el principio en que cualquier punto en un espacio probabilístico es gobernado por una variable aleatoria que sigue una distribución normal de media 0. De esta manera el algoritmo realiza una normalización de cada una de las variables de entrada y un análisis de las dependencias entre las mismas. Usando funciones kernel, realiza una medición de la proximidad de una medida y estima el valor de la salida.

Algoritmos de agrupamiento (clustering) + regression baseline: no es un método concreto, y existen infinidad de ellos en función de varios factores. El principio de estos métodos consiste en agrupar los datos de las mediciones midiendo la semejanza en conjuntos (clústeres). Esto se realiza para obtener un espacio reducido (reducción de la dimensionalidad) que permita la construcción de una línea base con datos más representativos, ya que un clúster es un representante de muchas medidas. Estos métodos se diferencian entre ellos dependiendo de la función de similitud que se utilice y los índices utilizados de evaluación. Posteriormente se desarrollará este método de agrupamiento junto con el modelo de línea base basado en funciones de transferencia para el desarrollo, validación y resolución de casos teóricos y prácticos.

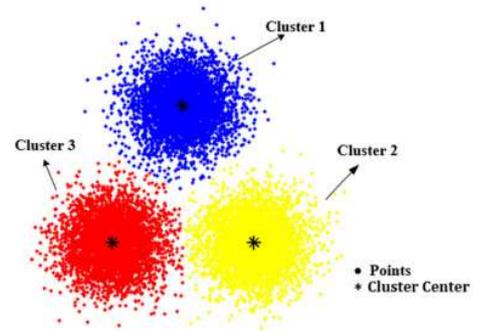


Figura 9. Fundamentos clustering.

- Modelo de caracterización inversa. Consumo HVAC edificio.

El modelo de caracterización inversa se basa en el uso de las funciones de transferencia para la predicción del consumo de un edificio. La función de transferencia proporciona un resultado a partir de las distintas excitaciones del sistema (variables independientes) en el instante de cálculo y en los momentos precedentes a éste.

Estas herramientas son usadas para la determinación de una respuesta dinámica, y en el caso de los edificios el efecto de la inercia térmica que se traduce en un desfase temporal entre entradas y salida del sistema hacen que esta técnica sea muy propicia para este fin, debido a la respuesta dinámica de la instalación de HVAC.

Sea representado el modelo mediante el esquema que sigue, en el se puede observar las siguientes variables:

$U(t)$  excitaciones medibles con incertidumbre

$V(t)$  perturbaciones: excitaciones no medibles, sucesos no explicables...

$Y(t)$  variable objetivo medible



Figura 10. Fundamentos modelo de caracterización inversa.

Para el caso de los edificios, y más en el caso de grandes edificios terciarios, la componente de perturbación  $V(t)$  es ínfima respecto a la componente  $U(t)$ . Esto anterior se debe a que el consumo energético en términos de climatización no está ligado al comportamiento humano, sino a la temperatura, radiación y ocupación, así es que es posible hacer la componente no medible despreciable para el objetivo del modelo.

De esta manera, este modelo de línea base modela la respuesta  $y(t)$  (Consumo) en función de las entradas teniendo en cuenta factores transitorios y desfases, de una manera cómoda matemáticamente hablando.

$$f(t) = \sum_{i=1}^m \sum_{j=0}^n a_{ij} Y_i(t-j) - \sum_{k=1}^d d_k f(t-j)$$

Donde:

$f(t)$  variable objetivo.

$Y_i$  m excitaciones, variables independientes.

$a_{ij}$  coeficientes de ajuste de las variables independientes.

$d_k$  denominadores de la función, dependencias de la función objetivo con momentos anteriores.

### 1.3 Desglose de consumos

Visto las necesidades de concienciación energética, el crecimiento del consumo energético y la problemática medioambiental es necesario un cambio en la concepción del uso de la energía. Actualmente la información disponible sobre el consumo del edificio es muy genérica, lo que dificulta el estudio de dichos consumos, así como la caracterización y la concienciación al usuario.

Con objeto de solventar la problemática de la monitorización existente en el trabajo desarrollado se presenta una propuesta de procedimiento de desglose de consumos en los distintos usos/servicios del edificio en cuestión.

El procedimiento de desglose propuesto se base en un algoritmo de clustering. El clustering es un método matemático que se utiliza para agrupar valores en unos representantes llamados clústeres que en el apartado 1.4 se extenderá su explicación. Aunque no es común encontrar ejemplos en los que se utilice esta herramienta para el desglose de consumos, dado la cantidad de datos que se maneja a la hora de estudiar el consumo de un edificio, es clara la opción de utilizar el clustering para conseguir un método más simplificado, fácil y adaptable a casos diferentes entre ellos.

Es decir, el uso del clustering podría llegar a generalizar/automatizar el desglose de consumos en la medida de lo posible de tal forma que permitiera desglosar de forma cómoda el consumo total de una edificación.

### 1.4 Clustering.

El clustering es uno de los métodos dentro de los llamados métodos de reducción de la dimensionalidad de la minería de datos. La minería de datos engloba a todos aquellos procesos que tratan grandes volúmenes de información para procesarlos, detectar patrones y/o relaciones ocultas con el fin de reducir la complejidad de estos.

El clustering o clustering de vectores se encarga de reducir la dimensionalidad del espacio definido por un conjunto de datos originales y transformarlo en un espacio de menor dimensión que facilite las tareas posteriores que se deban emprender. Esta reducción de la dimensionalidad se consigue con el agrupamiento de las entradas en distintas particiones, de tal manera que los datos englobados dentro de una partición pasarán a ser representados todos por un mismo conjunto de atributos que son los atributos de la partición.

Imaginemos un conjunto de vectores dato Q donde cada componente del vector es un atributo del dato. Estos vectores originales pueden separarse en q grupos ( $q < Q$ ) dado la similitud entre datos y disimilitud entre los grupos. El método de clustering consiste en esta separación, de tal manera que se forman clústeres o grupos de datos que se definen por su centroide. El centroide de un clúster es un vector de atributos calculado mediante la media de los atributos de los datos de dentro de cada clúster, que pasará a representar todos los datos que engloba. Con esto conseguiremos pasar de un espacio de dimensión Q datos a un espacio de dimensión q clústeres.

En el esquema se muestra el principio de este método, los puntos representan vectores datos que se han agrupado por similitud en tres clústeres. Cada clúster tiene su centroide que es el punto medio de los atributos de los integrantes. Pasamos de un espacio con 16 mediciones a un espacio con 3 clústeres.

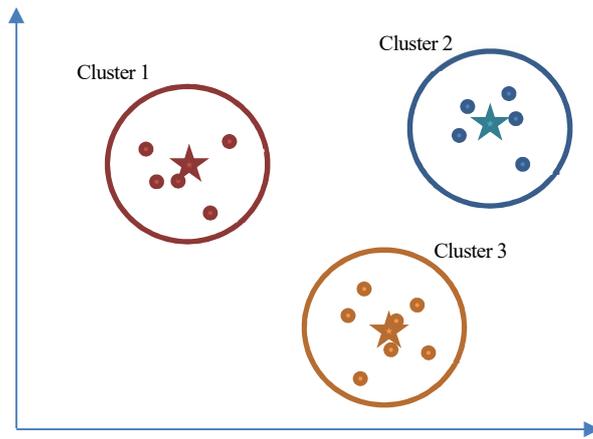


Figura 11. Centroides y clústeres.

Son diversos los métodos de clustering que se usan, todos ellos quedan expuestos en el esquema que sigue:

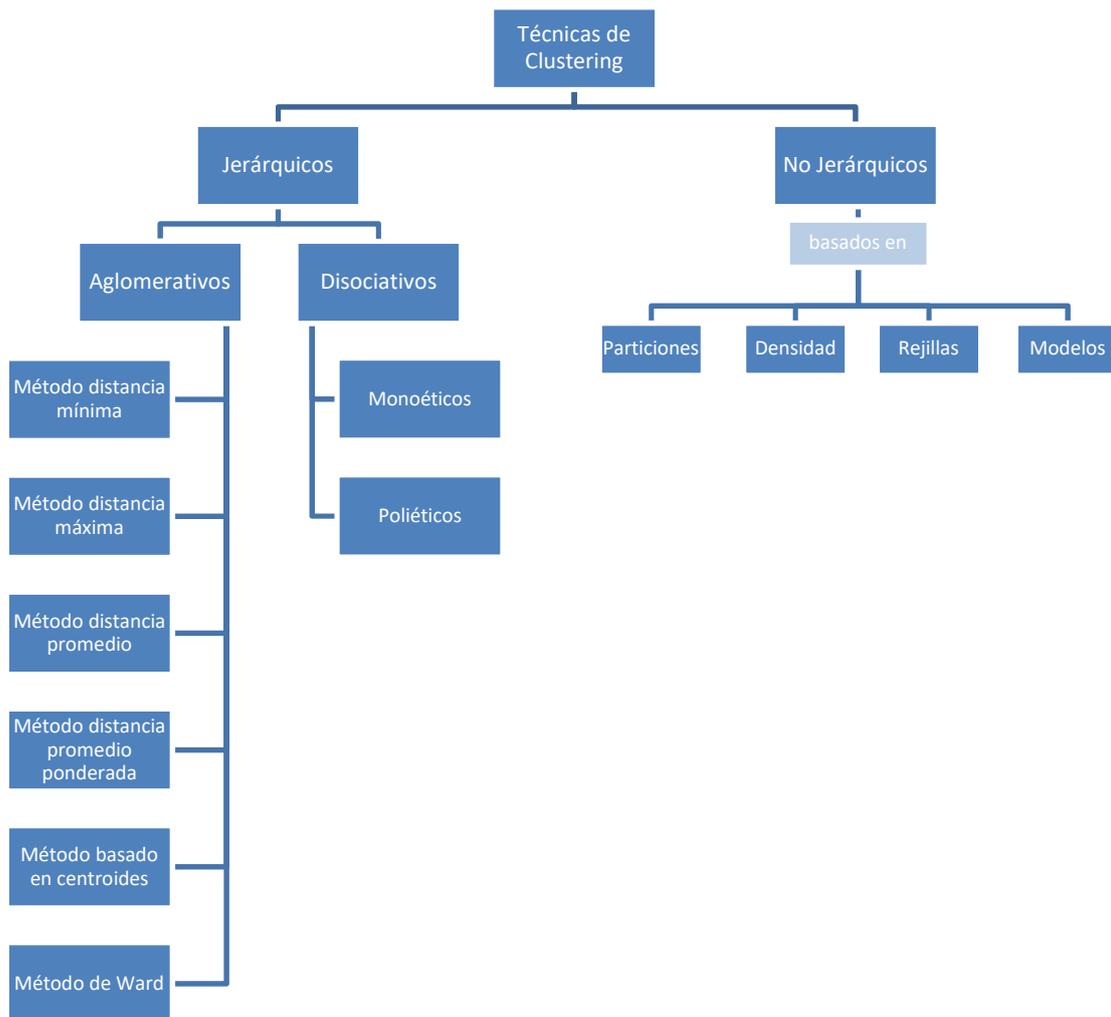


Figura 12. Clasificación técnicas de clustering

Dos puntos fundamentales a la hora de aplicar un método de clustering son la medida de la distancia y el índice de evaluación. Se entiende por medida de la distancia como la herramienta matemática que usa las distintas técnicas para segregar los datos. Para un mejor entendimiento, la distancia entre dos puntos dato es equivalente a la medida de la similitud. El índice de evaluación no es más que la valoración del método, es un índice que permite hacer visible el error causado en la aproximación realizada al pasar de Q datos a q clústeres. Estos dos temas serán los tratados en los dos puntos que siguen y serán los precedentes a la explicación y detalle de las distintas técnicas de clustering que más se usan en la actualidad.

### 1.4.1 Tipos de medidas de distancia.

Todo método de clustering basa su segregación de datos en la similitud de estos mismos, para el cálculo de esta similitud/distancia entre ellos hay diversas utilidades matemáticas.

Supongamos un conjunto vectorial de datos expresados como  $x_1, x_2, x_3, x_4 \dots x_m$  donde  $x_i$  representa un vector dato formado por  $n$  atributos, de tal manera que  $x_i = (x_{i1}, x_{i2} \dots x_{in})$

La distancia entre un dato  $x_s$  y  $x_t$  queda en función del método de la siguiente manera:

#### 1.4.1.1 Distancia euclidean.

$$d_{st}^2 = (x_s - x_t)(x_s - x_t)$$

#### 1.4.1.2 Distancia seclidean.

$$d_{st}^2 = (x_s - x_t) \cdot V^{-1} \cdot (x_s - x_t)'$$

Ves la matriz diagonal construida a partir  $S^2$ , donde  $S$  es la matriz de desviaciones estándar.

#### 1.4.1.3 Distancia mahalnobis.

$$d_{st}^2 = (x_s - x_t) \cdot C^{-1} \cdot (x_s - x_t)'$$

$C$  es la matriz de covarianza.

#### 1.4.1.4 Distancia cityblock.

$$d_{st} = \sum |x_{sj} - x_{tj}|$$

#### 1.4.1.5 Distancia minkowski.

$$d_{st} = \sqrt[p]{\sum |x_{sj} - x_{tj}|^p}$$

#### 1.4.1.6 Distancia chebychev.

$$d_{st} = \max_j \{|x_{sj} - x_{tj}|\}$$

#### 1.4.1.7 Distancia cosine.

$$d_{st} = 1 - \frac{x_s \cdot x_t'}{(x_s \cdot x_s') \cdot (x_t \cdot x_t')}$$

#### 1.4.1.8 Distancia correlation.

$$d_{st} = 1 - \frac{(x_s \cdot \bar{x}_s) \cdot (x_t \cdot \bar{x}_t)'}{\sqrt{(x_s \cdot \bar{x}_s) \cdot (x_s \cdot \bar{x}_s)' \cdot (x_t \cdot \bar{x}_t) \cdot (x_t \cdot \bar{x}_t)'}}$$

Siendo  $\bar{x}_s, \bar{x}_t$  las medias de los valores  $x_s, x_t$  respectivamente.

#### 1.4.1.9 Distancia hamming.

$$d_{st} = \frac{\#(x_{sj} \neq x_{tj})}{n}$$

#### 1.4.1.10 Distancia hamming.

$$d_{st} = \frac{\#[(x_{sj} \neq x_{tj}) \cap (x_{sj} \neq 0) \cup (x_{tj} \neq 0)]}{\#[(x_{sj} \neq 0) \cup (x_{tj} \neq 0)]}$$

#### 1.4.1.11 Distancia spearman.

$$d_{st} = 1 - \frac{(r_s - \bar{r}_s)(r_t - \bar{r}_t)'}{\sqrt{(r_s - \bar{r}_s)(r_s - \bar{r}_s)'} \sqrt{(r_t - \bar{r}_t)(r_t - \bar{r}_t)'}}$$

$r_s$  y  $r_t$  son los vectores de rango de  $x_s$  y  $x_t$

### 1.4.2 Tipos de índices de evaluación.

Como se describía anteriormente, el índice de evaluación permite reconocer la eficiencia del clustering en función del número de clústeres en los que se segrega un conjunto espacial de datos.

Este valor permitirá decidir un número de clústeres óptimo de tal manera que la reducción de la dimensionalidad no se vea comprometida por el error causado. Los índices más conocidos son los siguientes:

#### 1.4.2.1 Índice de Davies-Bouldin.

$$DB = \frac{1}{n} \cdot \sum_{i=1}^n \max_{i \neq j} \left( \frac{\sigma_i - \sigma_j}{d(c_i, c_j)} \right)$$

Donde n es el número de clústeres, c representa el centroide y  $\sigma$  la distancia media de los datos al clúster que pertenecen.

Los índices i y j representan dos clústeres distintos, de tal manera que el índice de Davies-Bouldin toma mejores resultados cuando los elementos dentro de un clúster están próximos a su centroide, pero los centroides de los distintos clústeres están separados entre ellos.

#### 1.4.2.2 Índice de Calinski-Harabasz.

$$VRC_k = \frac{SS_B}{SS_W} \cdot \frac{(N - k)}{(k - 1)}$$

$SS_B$  es la varianza entre clústeres y  $SS_W$  es la varianza entre elementos de un mismo clúster y N el número de observaciones. La varianza entre clústeres debe ser alta mientras que la varianza entre elementos de un clúster pequeña, de tal manera que una buena segmentación es aquella que maximiza  $VRC_k$ .

#### 1.4.2.3 Índice de Silhouette.

$$Sil(i) = \frac{a(i) - b(i)}{\max(a(i), b(i))}$$

a(i) es la distancia media del punto i a los puntos del mismo clúster, b(i) es el mínimo de las distancias medias entre el punto i de los puntos de otros clústeres.

### 1.4.3 Técnicas jerárquicas de clustering.

Las técnicas jerárquicas parten de una situación inicial sin un número de clústeres definido. Esta estrategia crea o agrupa clústeres de tal manera que lo realizado provoque el error mínimo. En el esquema anterior se recogían dos tipos fundamentales para estos métodos: los aglomerativos y los disociativos. Los primeros comienzan con tantos grupos como mediciones y va agrupando los datos hasta conseguir una segregación de  $q$  clústeres, los segundos comienzan con un único clúster que engloba a todas las mediciones y las va separando en conjuntos más disgregados.

Las técnicas jerárquicas de clustering ofrecen la posibilidad de hacer gráfico el proceso, obteniéndose el llamado dendograma. Este diagrama muestra las distintas agrupaciones y cuándo se crean, es decir el valor de similitud o distancia que provoca la disociación de los grupos. La principal desventaja de estos métodos es la decisión de cuántos clústeres se han de formar. La solución de mínimo error es sabido que sería tantos clústeres como datos tenemos, pero no tendría sentido el clustering en ese caso, por eso es ahí donde la experiencia y el conocimiento son la parte fundamental, para haciendo uso del dendograma determinar cuántos clústeres como máximo deberían de formarse en el proceso.

En el dendograma mostrado se representa la distancia o nivel de inconsistencia en el eje vertical y los clústeres que se forman en el eje horizontal. De tal manera que para un nivel de inconsistencia de 3.5 tenemos 4 clústeres como partición (cortes con líneas verticales al trazar una horizontal en el nivel de inconsistencia). Cuanto más cerca del eje horizontal nos situemos la disociación es mayor y por tanto la distancia entre clústeres menor. Como se explicaba antes es elección del usuario determinar el nivel de error máximo deseado en el problema de clustering para así determinar el número de clústeres que se formarán.

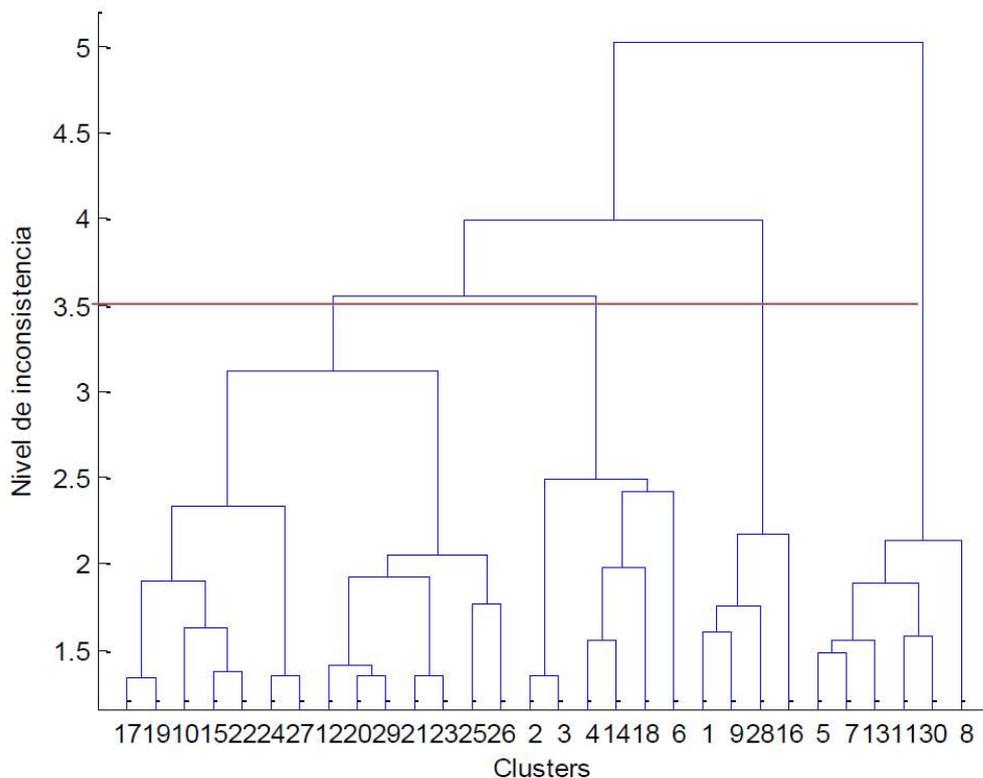


Figura 13. Dendograma.

#### 1.4.3.1 Métodos jerárquicos aglomerativos de clustering.

- Método de la distancia mínima o similitud máxima.

La estrategia realiza la decisión de si agrupar o no dos clústeres comparando la distancia o similitud entre ellos.

Si inicialmente existían  $n$  datos,  $n_i$  y  $n_j$  elementos dentro de los clústeres  $C_i$  y  $C_j$  y en la agrupación número  $k$  tenemos  $n-k$  clústers, la decisión se basa en la comparación de los elementos más parecidos:

$$d(C_i, C_j) = \text{Min}_{\substack{x_l \in C_i \\ x_j \in C_j}} \{d(x_l, x_m)\} \quad l = 1, \dots, n_i \quad ; \quad m = 1, \dots, n_j$$

$$s(C_i, C_j) = \text{Max}_{\substack{x_l \in C_i \\ x_j \in C_j}} \{s(x_l, x_m)\} \quad l = 1, \dots, n_i \quad ; \quad m = 1, \dots, n_j$$

Los clústeres  $C_i$  y  $C_j$  se unirán si la distancia o la similitud cumplen las condiciones siguientes:

$$d(C_i, C_j) = \text{Min}_{\substack{i_1, j_1=1, \dots, n-k \\ i_1 \neq j_1}} \{d(C_{i_1}, C_{j_1})\} =$$

$$= \text{Min}_{\substack{i_1, j_1=1, \dots, n-k \\ i_1 \neq j_1}} \left\{ \text{Min}_{\substack{x_l \in C_{i_1} \\ x_j \in C_{j_1}}} \{d(x_l, x_m)\} \right\} \quad l = 1, \dots, n_{i_1} \quad ; \quad m = 1, \dots, n_{j_1}$$

$$s(C_i, C_j) = \text{Max}_{\substack{i_1, j_1=1, \dots, n-k \\ i_1 \neq j_1}} \{s(C_{i_1}, C_{j_1})\} =$$

$$= \text{Max}_{\substack{i_1, j_1=1, \dots, n-k \\ i_1 \neq j_1}} \left\{ \text{Max}_{\substack{x_l \in C_{i_1} \\ x_j \in C_{j_1}}} \{s(x_l, x_m)\} \right\} \quad l = 1, \dots, n_{i_1} \quad ; \quad m = 1, \dots, n_{j_1}$$

	A	B	C	D	E	F	G
A	0						
B	2,15	0					
C	0,7	1,53	0				
D	1,07	1,14	0,43	0			
E	0,85	1,38	0,21	0,29	0		
F	1,16	1,01	0,55	0,22	0,41	0	
G	1,56	2,83	1,86	2,04	2,02	2,05	0

Tabla 1. Proceso, método de la distancia mínima 1.

De tal manera que si una partición inicial tiene la tabla de distancias o matriz de proximidades:

	A	B	(C,E)	D	F	G
A	0					
B	2,15	0				
(C,E)	0,7	1,38	0			
D	1,07	1,14	0,29	0		
F	1,16	1,01	0,41	0,22	0	
G	1,56	2,83	1,86	2,04	2,05	0

Tabla 2. Proceso, método de la distancia mínima 2.

En la primera etapa se unirán los individuos C y E dado que su distancia es la mínima con un valor de 0.21 y se formará el primer clúster del proceso. El proceso acaba hasta llegar a un único clúster que agrupe los elementos y permita construir en función de la distancia el dendograma del método.

- Método de la distancia máxima o similitud mínima.

$$d(C_i, C_j) = \text{Max}_{\substack{x_l \in C_i \\ x_j \in C_j}} \{d(x_l, x_m)\} \quad l = 1, \dots, n_i ; \quad m = 1, \dots, n_j$$

$$s(C_i, C_j) = \text{Min}_{\substack{x_l \in C_i \\ x_j \in C_j}} \{s(x_l, x_m)\} \quad l = 1, \dots, n_i ; \quad m = 1, \dots, n_j$$

Midiendo la distancia entre los elementos más dispares, a diferencia de la técnica anterior, se construye el árbol de decisión de este método.

$$\begin{aligned} d(C_i, C_j) &= \text{Min}_{\substack{i1, j1=1, \dots, n-k \\ i1 \neq j1}} \{d(C_{i1}, C_{j1})\} = \\ &= \text{Min}_{\substack{i1, j1=1, \dots, n-k \\ i1 \neq j1}} \left\{ \text{Max}_{\substack{x_l \in C_{i1} \\ x_j \in C_{j1}}} \{d(x_l, x_m)\} \right\} \quad l = 1, \dots, n_{i1} ; \quad m = 1, \dots, n_{j1} \end{aligned}$$

Así los clústeres  $C_i$  y  $C_j$  se unirán si cumplen:

A partir de la tabla de distancias y de forma análoga al método anteriormente ejemplificado, este método proporciona el dendograma final.

- Método de la distancia promedio no ponderada.

Los clústeres  $i$  y  $j$  son comparados mediante la medida de la distancia promedio de los elementos que los componen, sin tener en cuenta la cantidad de elementos que componen cada uno de los clústeres.

$$d(C_i, C_j) = \frac{d(C_{i1}, C_j) + d(C_{i2}, C_j)}{2}$$

- Método de la distancia promedio ponderada.

En este caso los clústeres se comparan teniendo en cuenta el tamaño, si un clúster  $c_i$  se subdivide en dos clústeres más. La distancia entre los clústeres  $i$  y  $j$  se define como sigue:

$$d(C_i, C_j) = \frac{n_{i1}d(C_{i1}, C_j) + n_{i2}d(C_{i2}, C_j)}{n_{i1} + n_{i2}}$$

De esta manera la distancia respeta el peso que tendrá un clúster formado por más elementos, más grande, que un clúster formado por menos elementos.

- Método basado en centroides.

Este método basa su comparación en la medida de la distancia de los centroides de los clústeres  $i$  y  $j$  que se comparan. Existe dos variantes para este método: el primero en el que la distancia no es ponderada por el tamaño del clúster y un segundo en el que si se pondera la medida con el tamaño (cantidad de mediciones) que tiene el clúster.

Haciendo uso de la distancia euclídea al cuadrado, y los clústeres  $C_i = \{C_{i1}, C_{i2}\}$  y  $C_j$

La variante ponderada queda:

$$d^2(C_i, C_j) = \sum_{l=1}^n (\sigma_l^j - \sigma_l^i)^2 = \frac{n_{i1}}{n_{i1} + n_{i2}} d^2(C_{i1}, C_j) + \frac{n_{i2}}{n_{i1} + n_{i2}} d^2(C_{i2}, C_j) - \frac{n_{i1}n_{i2}}{(n_{i1} + n_{i2})^2} d^2(C_{i1}, C_{i2})$$

Donde  $n_{i1}$  es el número de mediciones dentro del subclúster  $C_{i1}$  y  $\sigma^j$  es el centroide para  $C_j$ .

Para el caso no ponderado:

$$\begin{aligned} d(C_i, C_j) &= \frac{1}{2} d(C_{i1}, C_j) + \frac{1}{2} d(C_{i2}, C_j) - \frac{1}{4} d(C_{i1}, C_{i2}) \\ s(C_i, C_j) &= \frac{1}{2} s(C_{i1}, C_j) + \frac{1}{2} s(C_{i2}, C_j) + \frac{1}{4} [1 - s(C_{i1}, C_{i2})] \end{aligned}$$

Tenga en cuenta que la obtención del centroide del clúster no es más que la media ponderada de cada una de las componentes de las mediciones que componen el clúster. El caso no ponderado causa efectos no deseados en el caso de muy dispares tamaños de clústeres que se comparan.

Este tipo de métodos tiene cierta complicación en la observación del dendograma obtenidos, la razón de esto es que, de una fase a otra del método de agrupamiento, los clústeres se agrupan y van cambiando los valores de sus centroides y por ello las medidas de las distancias varían cada vez que avanzamos en las etapas de agrupamiento.

- Método de Ward.

Esta técnica realiza un agrupamiento de los dos clústeres que proporcionen un menor incremento de la suma de las diferencias cuadráticas de cada elemento al centroide de cada clúster.

$$E_k = \sum_{i=1}^{n_k} \sum_{j=1}^n (x_{ij}^k - m_j^k)^2$$

La expresión anterior representa la suma de los cuadrados de los errores del clúster k. i es el índice que denota cada uno de los elementos dentro del clúster k y j es el cada una de las componentes que representan los vectores de los puntos, m es el vector de componentes del centroide del clúster k. Así la suma de todos los errores de todos los clústeres en los que se divide el espacio es:

$$E = \sum_{k=1}^h E_k$$

K es el índice de cada uno de los clústeres y h es el número total de clústeres en los que se segmenta nuestros datos en la etapa del proceso que nos situemos.

Este método se inicializa con un dato en cada clúster, lo cual dará tantos  $E_k$  como datos tengamos iguales a 0 y por lo tanto un E global nulo. Tras esto, el método evalúa el incremento del valor E entre el momento inicial y el hipotético caso en el que dos clústeres determinados se unan. Aquella combinación de clústeres que haga el menor incremento será la elegida para unirse y formar un nuevo clúster y así reducir la dimensión del espacio.

Esto de forma iterativa en varias etapas proporciona un agrupamiento progresivo de aquellos clústeres que optimicen la función  $\Delta E$ .

#### 1.4.3.2 Métodos jerárquicos disociativos de clustering.

Estos métodos al contrario que los anteriores realizan una disociación de clústeres a partir del estado inicial en el que un clúster agrupa todos los datos. Al igual que los anteriores realizan una comparación usando la distancia como herramienta de cálculo para la separación, de tal manera que el objetivo sea maximizar la distancia entre clúster y disminuir la distancia entre elementos y centroide del clúster. Se distinguen principalmente dos tipos de técnicas.

- Métodos moetéuticos.

Estas técnicas realizan una separación en clústeres de los datos atendiendo únicamente a una componente de los elementos datos. Por lo tanto, solo es usada en caso en el que los elementos o mediciones no sean más que un valor y no sea un vector de valores.

- Métodos polítéticos.

Es el caso contrario al método moetéutico y realiza la segregación atendiendo a los valores de los vectores de variables de los elementos dato.

- Evaluación del método jerárquico. Coeficiente cofenético.

Los métodos jerárquicos de agrupamiento o disociativos de clustering realizan una segregación del conjunto de datos que a veces rompen las relaciones originales entre los mismos, resultando un espacio disgregado y poco conciso. Es por ello, que cuando se procede a realizar clustering a un problema y no está fijado que método jerárquico a realizar, se hace uso del coeficiente cofenético. Este coeficiente proporciona el valor de correlación entre los  $\frac{n(n-1)}{2}$  elementos de la parte superior de la matriz de proximidades observada y los correspondientes en la llamada matriz cofenética, C, cuyos elementos,  $c_{ij}$ , se definen como aquellos que determinan la proximidad entre los elementos  $i$  y  $j$  cuando éstos se unen en el mismo clúster.

Aquel método que para un mismo número de clústeres tenga un coeficiente cofenético mayor será el más apropiado para aplicar.

#### 1.4.4 Técnicas no jerárquicas de clustering.

Las técnicas no jerárquicas de clustering a diferencia de las jerárquicas parten de una situación inicial ya agrupada. Esta organización tiene fijado desde el inicio el número de clústeres en los que se va a agrupar los datos y una organización aleatoria de los mismos. En conclusión, los métodos no jerárquicos se encargan de reorganizar los datos para obtener una partición mejorada de la que se le da al algoritmo de inicio.

Estos métodos realizan un problema de optimización, estudiando los distintos movimientos de un elemento a los distintos clústeres existentes y realizando el movimiento el cual proporcione el mejor resultado. Es de remarcar que el número de clústeres se proporciona desde el inicio o se incluye su cálculo como parte intrínseca del algoritmo.

##### 1.4.4.1 Métodos basados en particiones.

A partir de un conjunto de objetos, se realizan particiones, y los datos se asocian a cada una de las particiones, de tal manera que iterativamente se reubican estos objetos para que cada objeto solo puede pertenecer a una partición.

##### 1.4.4.2 Métodos basados en densidad.

La agrupación de los objetos en clústeres se desarrolla mediante la observación del número de objetos que hay dentro de cada clúster, densidad. Por lo tanto, la creación o no creación de un clúster se decide observando la densidad que este tendría si esta entre los valores deseados o no.

##### 1.4.4.3 Métodos basados en rejillas.

El conjunto de datos representados en el espacio  $n$  dimensional es superpuesto por una rejilla con tamaño definido (número de clústeres en los que particionar) y los elementos dentro de cada celda pasarán a formar parte del clúster en cuestión.

##### 1.4.4.4 Métodos basados en modelos.

Haciendo uso de modelos estadísticos se realiza la segmentación de los datos en función de los valores arrojados para la herramienta estadística utilizada.

##### 1.4.4.5 K-means.

K-means es uno de los métodos más utilizados a la hora de hablar de clustering. Su funcionamiento se basa en la obtención de un espacio con un número determinado de clústeres que minimice la función de coste  $E$ , realizando una segmentación de mínima varianza.

El método consta de las siguientes etapas:

1. Segmentación inicial de todos los datos, de forma arbitraria o previamente estudiada por el usuario.  
Cálculo de los centroides de los clústeres iniciales.
2. Mover cada dato al clúster más cercano si cumple la condición:

$$\|x_i - q_l\| < \|x_i - q_k\|$$

Siendo  $x_i$  el elemento que se mueve,  $q_l$  el centroide del clúster al que se mueve y  $q_k$  el centroide del clúster del que proviene.

3. Recalcular los centroides de la nueva segmentación.
4. Realizar los pasos 2 y 3 hasta que ningún elemento se mueva de un clúster a otro.

Como conclusión a este método decir que es simple pero efectivo, necesita gran cantidad de computación cuando el número de datos es extenso. Hay que proporcionar el número de clústeres en los que se va a dividir el conjunto de datos. Esto último es una de las principales decisiones que el usuario deberá de tomar a la hora de emplear este método ya que no es conveniente un número grande de clústeres, pero tampoco es conveniente un número pequeño ya que el error de aproximación tomaría valores muy elevados.

#### 1.4.4.6 DBSCAN.

A partir de la densidad de puntos, esta técnica permite segregar un espacio a partir de dos valores previamente dados: radio del clúster y número de puntos mínimo.

Como anteriormente se ha comentado el algoritmo requiere dos datos de inicio:

- A. Radio,  $\epsilon$ : este valor es el radio o distancia que define una circunferencia límite situando un punto en su centro.
- B. MinPts: la variable es el número mínimo de puntos (vectores dato del espacio sin agrupar) que debe de haber dentro de la circunferencia de radio  $\epsilon$  alrededor de un punto, para que éste se configure como centroide o núcleo de un clúster.

El método centrándonos en un punto determinado del espacio inicial distingue tres tipos de puntos:

- A. Punto alcanzable: centrándonos en un punto determinado y trazando la circunferencia de radio  $\epsilon$ , un punto es alcanzable mirando desde el punto de estudio si se encuentra dentro de su circunferencia.
- B. Ruido: es un punto no alcanzable desde el punto observado.
- C. Núcleo: es un punto que tiene como mínimo un número de puntos alcanzables, este número mínimo es el MinPts.

El algoritmo estudia cada punto del espacio dato, mirando qué puntos son alcanzables desde cada punto que tenemos. Si un punto cumple el mínimo de puntos alcanzables se configura como núcleo y pasa a ser un clúster de nuestro espacio agrupado. Iterativamente se va realizando la operación de recolocación de puntos en clústeres hasta que no sea posible continuar y no haya ningún punto que cumpla el MinPts.

Este método es bastante utilizado, es ágil y como característica fundamental, no necesita proporcionar el número de clústeres de inicio, pero por desventaja los datos  $\epsilon$  y MinPts serán variables altamente influyentes que deberán ser estudiadas para obtener resultados eficientes.

#### 1.4.4.7 NEURAL GAS.

Esta técnica utiliza el método de agrupamiento por particiones, sin utilizar un conjunto de puntos vecinos al clúster restringido.

Manteniendo la función de densidad de los datos del algoritmo, el algoritmo segmenta el espacio obteniendo una serie de vectores prototipo (clústeres). Seguidamente reubica los datos mediante el cálculo de la distancia del dato a los clústeres vecinos, así el método ubica el dato en el clúster donde más separado esté de los puntos vecinos.

A este método es necesario proporcionarle el número de iteraciones que debe de realizar como valor de entrada.

#### 1.4.4.8 Gaussian mixture model.

El método utiliza la función de densidad de la suma de las componentes gaussianas como método de ajuste:

$$p(\vec{x}) = \sum_{i=1}^M w_i (N(\vec{x} | \mu_i, \Sigma_i))$$

El vector  $x$  es un vector con  $d$  componentes de datos,  $w_i$  el peso de cada dato y  $N(\vec{x})\mu_i, \sum i$  la función de densidad gaussiana.

Este método crea un vector formado por la media y la matriz de covarianzas que representa la distribución gaussiana del conjunto de datos. Seguidamente calcula la probabilidad de cada dato pertenecer a un clúster determinado y tras actualizar la distribución, recalcula sus valores característicos. Tras realizar esto, realiza una prueba de convergencia y valora el cambio producido al actualizarse hasta llegar al valor objetivo de convergencia definido por el usuario.

## 1.5 Objetivos

Visto las necesidades de concienciación energética, el crecimiento del consumo energético y la problemática medioambiental es necesario un cambio en la concepción del uso de la energía. Actualmente la información disponible sobre el consumo del edificio es muy genérica, lo que dificulta el estudio de dichos consumos, así como la caracterización y la concienciación al usuario.

Con objeto de solventar la problemática de la monitorización existente en el trabajo desarrollado se presenta una propuesta de procedimiento de desglose de consumos en los distintos usos/servicios del edificio en cuestión.

Posteriormente, dado que el procedimiento de desglose de consumos nos permite conocer el consumo diario asociado a la climatización del edificio se desarrolla un modelo de caracterización diario de dicho consumo.

El procedimiento de desglose y el modelo de caracterización (línea base) será desarrollado y validado en un caso teórico comentado en el apartado 3. Posteriormente dicho procedimiento y modelo de línea base es aplicado a datos de un edificio real (apartado 4).

# 2 METODOLOGÍA

---

## 2.1 Alcance.

En lo que sigue se procede a la redacción del fundamento de este trabajo fin de grado, el cual trata de definir una metodología de actuación inicial para poder poner en práctica un programa de mejora de la eficiencia y hacer visual lo que una certificación o medición eléctrica no puede llegar a hacer. Así mismo pretende servir como procedimiento generalista aplicable a cualquier necesidad de desglosar el consumo energético de forma rápida y fiable.

El fin principal del texto es concienciar del uso de la energía en el caso de los edificios terciarios y grandes terciarios dada la clara necesidad de cambio en políticas de gestión energética que se demanda en la actualidad, para ello se listan los alcances fundamentales que se cumplirán:

- Desglosar de forma automática el consumo total de un edificio de uso terciario en los niveles característicos de consumo a partir de las mediciones del consumo eléctrico total de un año.
- Aplicar un modelo de predicción para consumos futuros del edificio en situaciones distintas a las de referencia.

A parte de los alcances, este proyecto defenderá posteriormente la validez de sus resultados comparando los resultados obtenidos con mediciones reales del edificio para el denominado año de validación.

## 2.2 Procedimiento de desglose de consumos.

### 2.2.1 Antecedentes.

El consumo de un edificio terciario es posible dividirse en sus tres partes fundamentales:

$$C = C_{fijo} + C_{ocupación} + C_{HVAC}$$

Donde:

- A.  $C_{fijo}$  es el consumo constante que posee un edificio. Este consumo no varía durante el año, y si varía es de una manera ínfima ya que es posible que muchos sistemas dependan algo de la ocupación. Los equipos asociados son aquellos que funcionan permanentemente, tales como cámaras de seguridad, sistemas de comunicaciones, medición, ascensores, etc.
- B.  $C_{ocupación}$  se define como el consumo que tiene el edificio debido a su uso. Este valor es variable y depende de la ocupación del edificio y del porcentaje de uso en un momento determinado. Es importante comentar la linealidad existente entre este consumo y su valor. Gastos que repercuten en este consumo son aquellos como la luz, la conexión de equipos de trabajo a la red, equipos informáticos, maquinaria etc.
- C.  $C_{HVAC}$  es el gasto debido a los equipos de calefacción y refrigeración, humidificación, etc. del edificio. Su valor es fuertemente variable en función del calendario, climatología y ubicación del edificio. Equipos que imputan su gasto en esta variable son los equipos de refrigeración, calefacción, renovación y humidificación del aire.

## 2.2.2 Situación de partida.

El primer objetivo del TFG llevado a cabo es desglosar de forma automática el consumo total de un edificio de uso terciario en los niveles característicos de consumo a partir de las mediciones del consumo eléctrico total de un año.

El procedimiento de desglose de consumos llevado a cabo en el trabajo se divide en tres fases, de las cuales las fases 1 y 2 se fundamentan en el trabajo llevado a cabo en el trabajo fin de máster “Sistema de Gestión Energética para Edificios EMS siguiendo la Hoja de Ruta BS-ISO 50006:2014” y la fase 3 se ha desarrollado en el presente trabajo fin de grado.

Dicho procedimiento de desglose comienza con la recepción de los datos necesarios para su desarrollo. Los datos son los siguientes:

- A. Ubicación y fecha de datos: la ubicación del edificio, junto con la fecha de las mediciones permitirá la obtención de los datos meteorológicos a partir de las bases de datos del MAGRAMA (Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente) en caso de no disponerlos a partir de otra fuente de medición más cercana.
- B. Datos de consumo: para un año de estudio como mínimo, los datos de consumo eléctrico del edificio serán recogidos en base horaria. En el caso de disponer de las mediciones en una base menor a la horaria, éstas se pasarán a la base requerida para realizar el método con un número menor de datos, ya que no afectará al resultado.
- C. Calendario, uso del edificio: para un año genérico, será necesario tener una visión esquemática del porcentaje de uso y ocupación para una semana normal de trabajo, los días de ocupación reducida o nula y los periodos vacacionales.

Una vez los datos están en disposición para el estudio, comienza el procedimiento de desglose.

## 2.2.3 Procedimiento.

### 2.2.3.1 FASE 1. Preparación de los datos de entrada.

En esta fase se constituirán los datos de entrada del algoritmo de clustering. Una vez recibidos los datos de entrada y recogidos los valores climatológicos es necesario pasar los datos a base diaria para luego normalizar estas variables. La razón fundamental de la normalización reside en disminuir el orden de magnitud de las variables, ya que el algoritmo de clustering se desarrolla haciendo uso de medidas de distancia. Con esto, la computación y el entendimiento de las distancias entre puntos se optimiza y se obtienen valores más fiables del procedimiento que en el caso de usar los datos en su unidad original de medida.

Esta normalización se consigue tipificando las mediciones de la siguiente forma:

$x_{normalizada} = \frac{x_{real} - \mu_x}{\sigma_x}$  Siendo  $x$  el valor de una variable de entrada,  $\mu$  es la media de la variable  $x$  del conjunto anual de valores y  $\sigma$  su desviación típica.

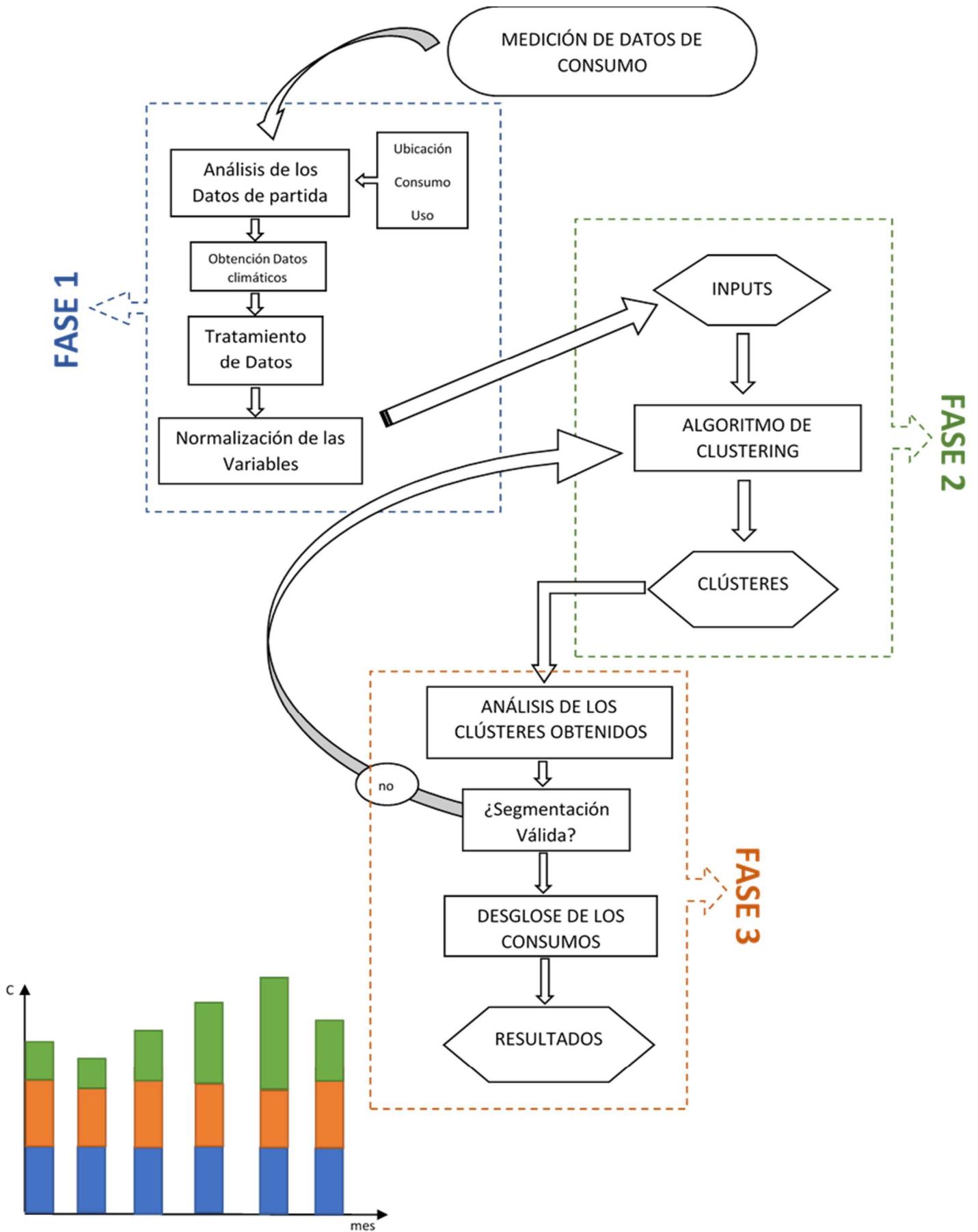


Figura 14. Metodología. Procedimiento de Desglose Automático.

### 2.2.3.2 FASE 2. Clustering.

Una vez las variables de entrada están definidas y normalizadas, es el momento de iniciar el algoritmo de clustering.

El algoritmo de clustering que se empleará será el método K-means, aunque no es el único tipo de algoritmo que se puede utilizar para este cometido. La técnica de tipo no jerárquico requiere previamente del número de clústeres en los que se desea segmentar el espacio de datos, la fase 2 será la encargada de determinar primeramente el número de clústeres óptimo en el que seccionar el conjunto de datos para luego realizar el agrupamiento propiamente dicho.

- Determinación del número de clústeres.

El número de clústeres a seccionar el espacio influirá en la calidad del resultado del algoritmo de clustering, por ello previamente hay que realizar una buena selección de este número. Esto se obtiene mediante el índice de Davies-Bouldin, el cual disminuye en función de la calidad del agrupamiento realizado.

Este índice se encarga de comparar las distintas posibilidades de agrupamiento y reflejar con un valor menor aquel en la que se obtenga distancias pequeñas entre los datos dentro de cada clúster (clúster compacto) y mayor para grandes distancias entre los distintos clústeres.

El intervalo de búsqueda del número óptimo será un valor que irá cambiando entre las distintas pruebas que se realizarán. 2 clústeres como mínimo, en el caso en el que el comportamiento del edificio fuese uniforme en todo el año y 21 como máximo, aunque podrá variar según las necesidades, resultado de distinguir entre los 7 días de la semana y tres estaciones fundamentales: invierno (calefacción), verano (refrigeración) y estaciones intermedias (no HVAC), aunque este intervalo será flexible en función de la situación de partida.

El cálculo del número óptimo se realizará con la llamada a una función programada en Matlab. Esta función realizará 150 iteraciones y devolverá para cada iteración, el número óptimo de clústeres encontrados dentro de los límites impuestos y el valor del índice de Davies-Bouldin.

Tras llamar a la función, ésta comprueba los valores óptimos del índice para cada una de las iteraciones, de tal manera que si son menores a 0.9 el procedimiento lo considera válido y será el caso del mínimo absoluto del índice, el número óptimo de clústeres que daremos a nuestro algoritmo de clustering. Este valor de 0.9 proviene de la experiencia, como así se recoge en el TFM en el que nos basamos. En caso de no encontrar una solución que cumpla este margen superior, entonces tendremos que ampliar el intervalo de búsqueda que inicialmente fijamos en [2,21].

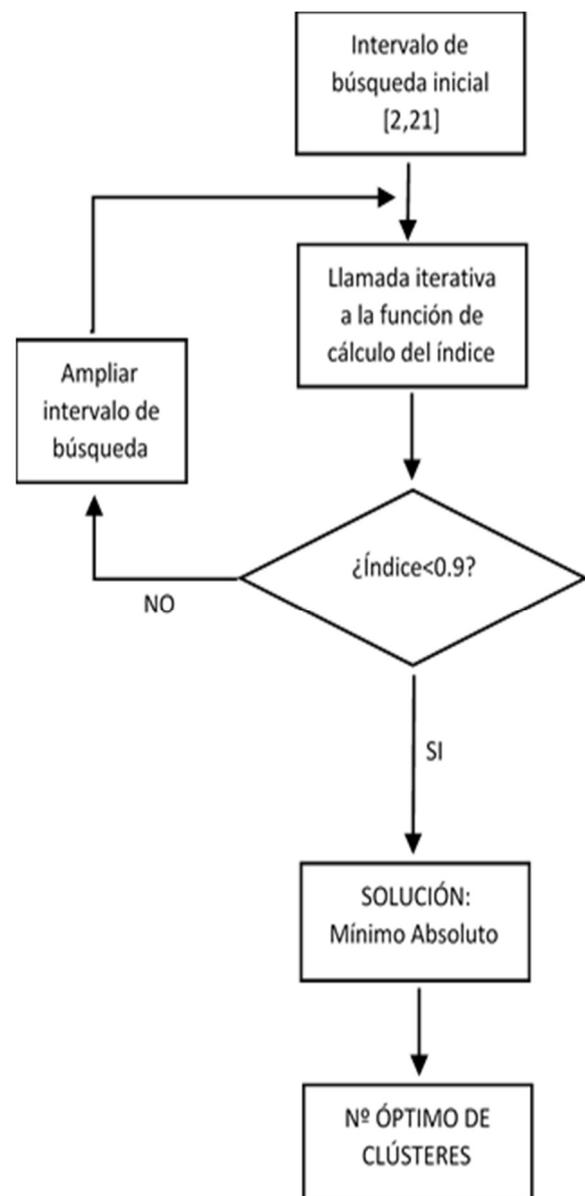


Figura 15. Procedimiento de búsqueda de número óptimo de clústeres

La razón del uso de iteraciones en el algoritmo no es más que el agrupamiento inicial que necesita el programa. Es decir, K-Means parte de una situación ya agrupada, en este caso aleatoriamente que luego modifica y analiza. Dado que la situación de partida es importante, se emplea las iteraciones para que en cada una de ellas la situación de partida en la que comienza a analizar sea la solución de la anterior iteración. Muchos algoritmos de clustering no contemplan el uso de la iteración para tener un agrupamiento inicial bueno y conseguir mejores resultados. Así si comparamos los resultados para unos mismos valores de inicio, el número de clústeres óptimo e índice pueden variar en función de la situación agrupada inicial que usa el algoritmo. Se ha fijado 150 iteraciones ya que no supone mucho coste computacional y los resultados a partir de estas 150 pruebas no sufren apenas cambios.

- Clustering.

Una vez tenemos el número de clústeres en los que agrupar los datos, y los datos normalizados, es el momento de iniciar el algoritmo de clustering propiamente dicho.

El procedimiento de clustering se desarrolla en Matlab debido a que en este software ya encontramos dentro de su biblioteca la función K-Means.

La función utiliza como entradas una matriz X con los vectores de datos dispuestos en cada fila, donde cada componente del vector se aloja en una columna de la matriz, y k es el número de clústeres en los que vamos a agrupar los datos, previamente obtenido en el punto anterior.

La llamada a la función tiene la siguiente estructura:

$$[idx, C, sumd, D] = kmeans(X, k)$$

Las salidas de la función son las siguientes:

- *idx* es un vector columna, que representa el índice del clúster al que pertenece cada punto de la matriz de datos de entrada, X.
- *C* es una matriz con los centroides de cada uno de los clústeres que se han formado en el proceso.
- *sumd* es un vector columna que devuelve la suma de las distancias de los puntos dentro de un clúster a su centroide.
- *D* es una matriz con la distancia de cada punto dato a cada centroide obtenido.

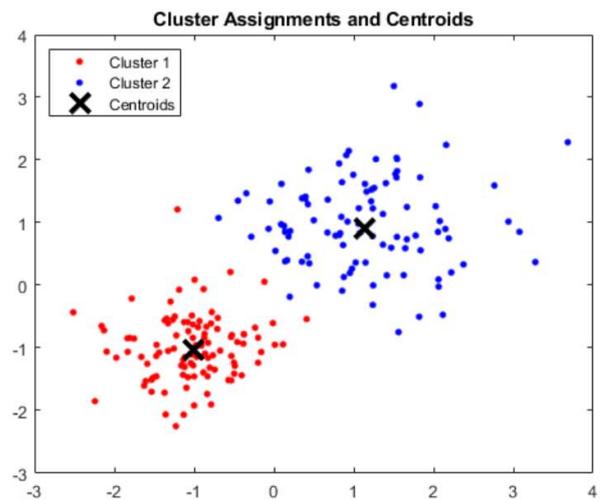


Figura 16. Matlab, obtención de clústeres.

### 2.2.3.3 FASE 3. Análisis de Clústeres y Desglose.

La fase 3 como etapa final tiene el objetivo de caracterizar y definir cada uno de los consumos básicos del edificio que anteriormente se exponen, para luego calcularlos en base mensual y poder ser estudiados.

El análisis de los resultados agrupados en los distintos clústeres comienza mediante la visualización de los elementos de un mismo clúster. Para este fin se hará uso de los índices del clúster que arroja Matlab a cada punto dato.

Sirva como ejemplo la siguiente tabla en la que se muestra el clúster 39 con los distintos vectores dato que lo componen y el valor de las variables su centroide del mismo.

Cluster 39	Consumo [kWh]	Temp Media [°C]	Radiación [MJ/m2]	Tipo de día
02/12/2013	1542	9.43	3.24	1
16/12/2013	1405	11.01	2.71	1
23/12/2013	1410	7.27	2.94	1
30/12/2013	1456	6.17	2.78	1
13/01/2014	1276	11.59	1.79	1
20/01/2014	1526	8.07	3.07	1
27/01/2014	1236	12.77	2.75	1
03/02/2014	1527	7.63	2.12	1
10/02/2014	1523	10.11	4.07	1
17/02/2014	1442	10.46	2.96	1
10/11/2014	1482	11.32	3.22	1
17/11/2014	1382	12.45	3.19	1
Centroide	1434	9.86	2.90	1

Tabla 3. Ejemplo clúster y componentes.

Una vez obtenido el desglose de los clústeres, así como el valor de los centroides, es el momento de analizar si la segmentación es válida. En caso negativo, la fase 2 se determinaría como inválida y habría que volver a realizarla modificándolo y buscando un número mayor de clústeres para agrupar los datos que cumpla con los requerimientos. Los requerimientos a los que se tiene que obligar el agrupamiento de los datos son los siguientes:

- Índice de Davies Bouldin menor a 0,9.
- Conservación de un 80% de la característica de variable entera en los valores de los centroides de los clústeres, para las variables definidas como no decimales de inicio. Generalmente las de ocupación.
- Variedad de clústeres en los periodos de no funcionamiento de los sistemas de climatización.

En el apartado 3 se podrá observar el tratamiento de estos requerimientos y la importancia de cada uno de ellos.

- Modos de operación. Temperatura equivalente.

Se define la temperatura equivalente como una temperatura que tiene en cuenta los efectos convectivos y radiantes con los que está excitado el edificio. Es por tanto una vía equivalente a la formulación de las excitaciones exteriores. Dicha temperatura nos permitirá determinar los regímenes de operación del edificio así como los periodos de no climatización.

La temperatura equivalente será un parámetro esencial en el proceso de desglose de consumos que se presenta. A partir de ella será posible obtener el consumo fijo y el consumo de ocupación de la edificación, por lo que será fundamental conocer el valor antes de proceder con el clustering. Antes de desglosar permitirá saber si el agrupamiento nos ha permitido obtener la variedad de clústeres que nos vendrá bien para realizar el desglose. Es decir, podremos saber si tenemos suficientes clústeres dentro de la zona de no trabajo del sistema de HVAC como para poder realizar el desglose de forma óptima.

Una vez obtenido el valor, será posible saber en qué caso el sistema trabajará en modo de calefacción, refrigeración o simplemente no trabajará. Para saber esto último es importante hacer un análisis previo de la zona en la que se trabaja y definir un intervalo de temperaturas centrado en la temperatura equivalente que nos marcará el rango de temperaturas en el que el equipo de climatización no funciona.

Para la obtención de la temperatura equivalente real es necesario el consumo de climatización del edificio. Dado que partimos de consumos globales del edificio y dicha temperatura es necesaria para el procedimiento de desgloses desarrollado el cálculo de este parámetro se ha estimado a partir del consumo total del edificio, buscando el corte entre las rectas de consumo total en meses de refrigeración y calefacción en función de la temperatura (ver 3.3.2.1).

Para los casos que analizamos, se ha optado por elegir un intervalo de 2°C de amplitud, el cual marcará si un día, clúster, hora... determinada funciona en un modo u otro.

- Ocupación. Variable tipo de día.

La ocupación del edificio es una variable esencial para la definición del consumo por uso. Este problema se resuelve con la definición de los tipos de días clave que existen en el calendario laboral.

La variable de ocupación es una de las entradas al algoritmo de clustering que no puede faltar en el caso que la ocupación sea irregular ya que se encargará de dar la información de ocupación de cada uno de los componentes y clústeres del problema.

En caso de no claridad de los tipos de día para un edificio, es posible definirse los valores de la variable haciendo uso de intervalos para el porcentaje de uso diario que se de en el edificio.

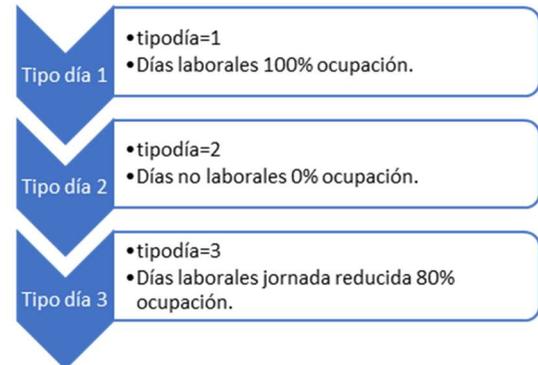


Figura 17. Ocupación, variable tipo día.

$$\%uso = \frac{n^{\circ} usuarios \times jornada\ de\ ocupación(horas)}{n^{\circ} usuarios \times 24\ horas}$$

Así se asociará un tipo de día en función del porcentaje de uso que tenga el edificio, por ejemplo:

- [0%,10%] de uso -> Tipo día=1
- [20%,50%] de uso-> Tipo día=2
- [50%,70%] de uso -> Tipo día=3
- [70%,80%] de uso -> Tipo día=4
- [90%,100%] de uso -> Tipo día=5

- Obtención del consumo fijo del edificio.

El consumo fijo del edificio  $C_{fijo}$  es el que tiene el edificio por simplemente ser un edificio, aunque muchos de los equipos consumidores que lo repercuten son dependientes levemente de la ocupación. Este consumo es casi constante en la curva de consumo del edificio.

Para la obtención del consumo fijo será necesario tener el calendario de funcionamiento del edificio, o en el caso que nos concierne la variable tipo de día.

Este consumo se hará visible en aquel/os clústeres que engloben los días de no ocupación, donde el consumo por uso no existirá. De tal manera que el consumo medio de todos estos días o en su defecto del clúster/es será el valor que tomará el consumo fijo siempre y cuando el sistema de climatización no trabaje.

Remarcar que, debido a la posible inexactitud de los datos de entrada, no es aconsejable suponer que el consumo en climatización no existe cuando no hay ocupación.

De esta manera sabiendo qué días no hay consumo por ocupación y buscando clústeres en la zona intermedia o de no consumo del sistema HVAC, podremos conocer el valor del consumo fijo que supondremos constante para la estimación del consumo diario.

En el caso de un edificio con ocupación permanente, no se distinguirá entre  $C_{fijo}$  y  $C_{uso}$  sino que el desglose será realizado en dos partes: fija+uso que llamaremos  $C_{usuario}$  y HVAC.

En caso de haber heterogeneidad en los clústeres en cuanto al tipo de día, es decir, que los centroides tomen valores muy decimales en la variable tipo día, será necesario repetir la fase 2 del procedimiento y hacer un agrupamiento más extenso de los datos.

- Consumo ocupación.

El consumo por uso u ocupación es aquel que depende en gran medida de la ocupación y del porcentaje de uso del edificio.

A parte de la ocupación del edificio, es necesario la distinción de los modos de operación de un edificio cualquiera para la obtención de este. Para ello se hace uso de la temperatura equivalente y la variable tipo día que marcará las distintas situaciones de ocupación a las que se expondrá la edificación.

Para la obtención de los consumos por uso se procederá tal y como sigue:

- A. Búsqueda del clúster o clústeres que trabajen en modo de operación de estaciones intermedias (Consumo HVAC=0). Para ello se utilizará el intervalo de temperaturas de no consumo HVAC centrado en la temperatura equivalente.
- B. Tabular los consumos de uso en función del tipo de día. Se obtendrá un consumo de uso característico para cada tipo de día, para ello se realizará la diferencia del consumo total y el consumo fijo previamente obtenido.
- C. En caso de existir varios clústeres que se sitúen en la zona de entretiempo del mismo tipo de día, se utilizará aquel cuya temperatura sea más cercana a la temperatura equivalente.

- Consumo HVAC.

El consumo HVAC se obtendrá trivialmente para cada uno de los días del año. Cada día estará asociado unos consumos fijos y de ocupación característicos y el consumo de climatización se obtendrá por diferencia del consumo total real menos los fijo y ocupación estimados.

Es posible que algunos resultados no concuerden, obteniéndose consumos de HVAC negativos, etc. Para evitar esto hay que recordar las siguientes reglas:

- A. Si el consumo total diario es menor al consumo fijo estimado, el consumo fijo estimado será igual al consumo total de ese día.
- B. Si el consumo de ocupación diario es mayor a la diferencia del consumo total diario y el consumo fijo estimado, entonces el consumo ocupación diario modificará su valor al resultado de dicha diferencia.
- C. Un día situado en la zona de no trabajo del sistema de climatización, mantendrá siempre un valor nulo en su consumo de climatización estimado. En caso de tener un consumo total mayor a la suma de los consumos fijo y ocupación estimados, el consumo de ocupación modificará su valor y absorberá la falta.
- D. El consumo total estimado y el consumo total real diario tienen que ser iguales.

- Presentación de resultados.

Una vez calculados los consumos diarios de todo el periodo y corregido los posibles errores de cálculo es el momento de calcular los resultados en base mensual y su presentación en forma gráfica o tabulada para el análisis de los resultados.

## 2.3 Validación del procedimiento de desglose de consumos.

Partimos de la situación en la que se dispone del desglose real de los consumos del edificio para las tres secciones del consumo que se han hecho mención en el año de mediciones.

Con ello podremos realizar el cálculo del error mensual, así como un error promedio y total del desglose. En función de los resultados obtenidos y el nivel de exigencia del proyecto, el procedimiento se dará por válido o será sometido a modificación para obtener resultados mejores.

Esta sección será la encargada de englobar todos los errores del procedimiento. El resultado de ésta será resultado de la suma del error de clustering de cada partida, así como el propio error de desglose. Este error contemplará tales como:

- A. Error en el cálculo de la temperatura equivalente.
- B. Error en la elección de las variables utilizadas.
- C. Error al determinar el intervalo de no operación del sistema HVAC.
- D. Error en el cálculo del consumo fijo del edificio.
- E. Error en el cálculo del consumo de uso del edificio.

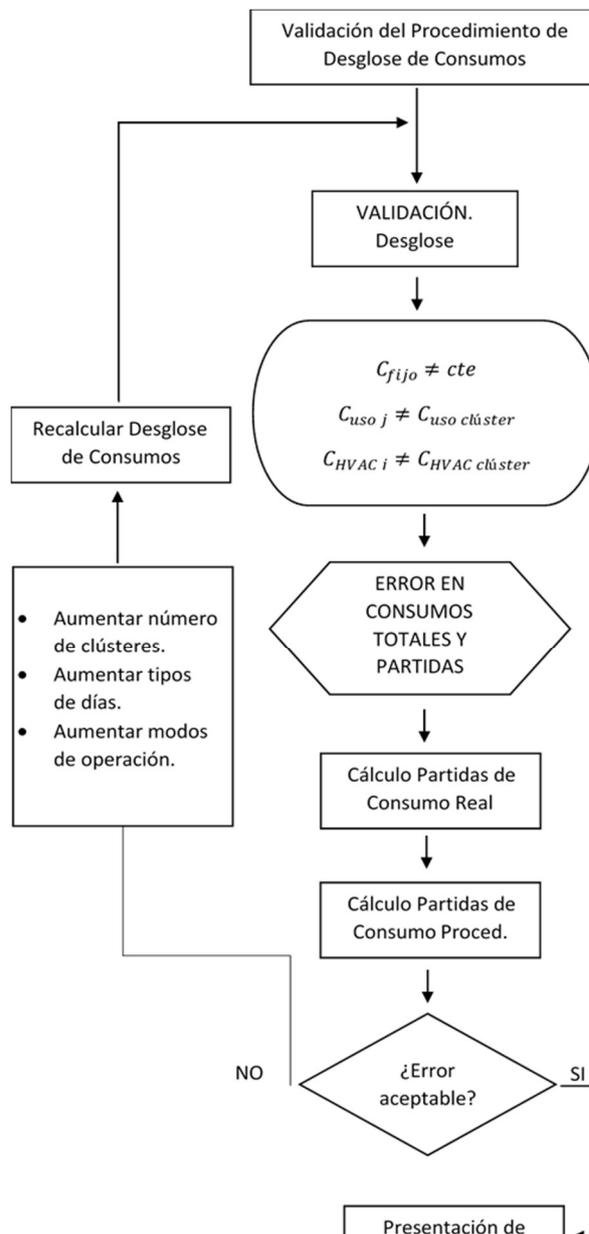


Figura 18. Procedimiento de validación del desglose de consumos.

Para la comprobación el usuario deberá de tener los datos desglosados en los consumos fijo, uso y HVAC del año de mediciones, o en su defecto del consumo usuario y HVAC. Una vez tengamos estos resultados, el error se calculará en base mensual a partir de los resultados del procedimiento y el cálculo de los consumos reales de los distintos meses del edificio. El método de validación se resume mediante el diagrama que sigue, se propone el cálculo de los errores realizados en el desglose de cada uno de los consumos obtenidos. Previamente será necesario calcular el consumo mensual de cada uno de las partidas y tras fijar un máximo acordado, en función de la precisión necesaria del procedimiento, este se dará por válido o no. En caso de no ser válido el desglose, será necesario aplicar una o varias de las acciones detalladas para volver a realizar el desglose.

## 2.4 Procedimiento caracterización energética

En esta sección se expone el procedimiento de caracterización de consumos desarrollado. La línea base de consumos permitirá estimar el consumo global del edificio a partir de los datos meteorológicos y de uso del mismo. Dicho procedimiento, como se ha comentado anteriormente, tiene como objetivo la predicción del consumo del edificio que resulta imprescindible para realizar la verificación de los ahorros obtenidos tras una mejora implementada en el edificio o también puede ser usado como indicador efectivo a la hora de detectar cualquier tipo de anomalía en el consumo tras compararlo con la medición real.

La caracterización del consumo de climatización se construirá usando como referencia el modelo de caracterización inversa que desarrolló D<sup>o</sup> José Sánchez Ramos en su tesis “Metodología de Caracterización Inversa para Edificios” adaptado y haciendo uso de los resultados obtenidos mediante el clustering [29].

### 2.4.1 Línea base de consumos.

La línea base de consumos se apoya en el uso de funciones de transferencia para realizar la predicción y obtención de la salida del modelo. De esta manera con el agrupamiento de los datos anteriormente desarrollado, obtendremos la ecuación que gobierne el consumo del edificio.

La línea base que a continuación se presenta será el paso posterior al desglose de consumos de la edificación, esta calculará el consumo desglosado del edificio gracias a la previa construcción de una línea base de ajuste del consumo de climatización multiparámetro. En una situación real los datos que se disponen generalmente será el consumo total del edificio, es por ello que para poder construir la línea base de climatización será necesario tener los consumos desglosados inicialmente, así como disponer de los consumos característicos de ocupación y fijos de la edificación para su estimación.

$$C_{total,estimado} = C_{clima,estimado} + C_{ocupación,estimado} + C_{fijo,estimado}$$

Comenzando por los consumos más triviales de este cálculo, el consumo fijo estimado tomará el valor obtenido en el desglose de consumos del edificio que previamente se ha realizado.

En cuanto al consumo de ocupación, será condición necesaria para poder calcular su valor que los niveles de uso del edificio, es decir, el porcentaje de ocupación no hayan variado para poder asociar un consumo de ocupación al día que se estima en función del valor de la variable de ocupación que se tenga, que se obtuvo en el desglose de consumos.

Ya para el cálculo del consumo de climatización, será necesario realizar una línea base propiamente dicha. Se llevará a cabo la estimación mediante la construcción de una línea base basada en funciones de transferencia, que permitirá calcular la demanda de energía de climatización, en función de las características meteorológicas del día a estimar y días previos, dado que todos los edificios tienen lo conocido como efecto memoria, es decir, las condiciones de días anteriores influyen notablemente en el comportamiento del día en cuestión.

#### 2.4.1.1 Línea base de climatización.

La caracterización del consumo de climatización se lleva a cabo según el esquema descrito en la figura 19.

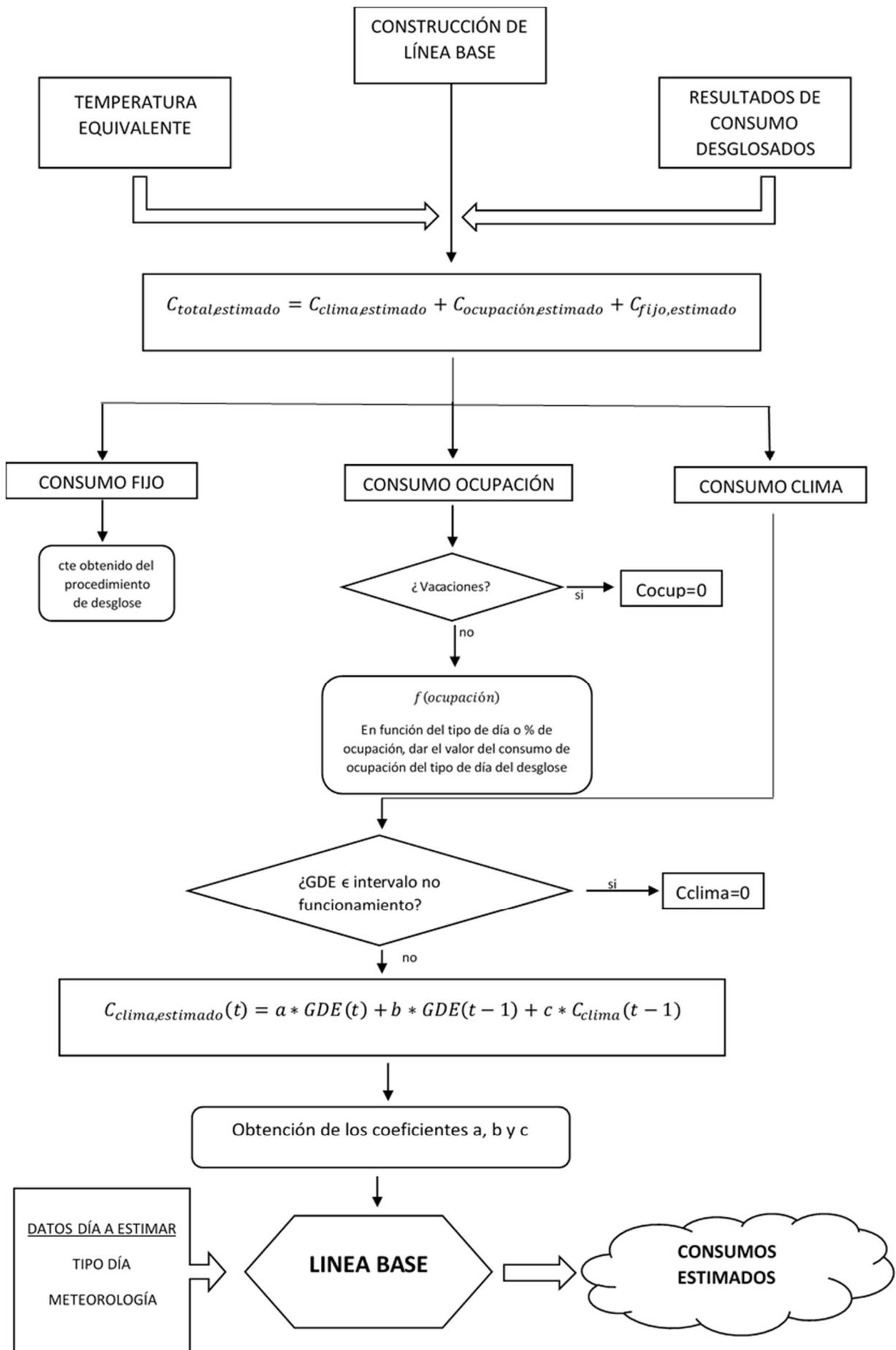


Figura 19. Estimación, línea base de consumos.

### 2.4.1.2 Evaluación del proceso de estimación.

La evaluación del proceso de estimación se realizará para un periodo de tiempo o año distinto del que se utilizó para el cálculo de la línea base.

Para valorar la calidad de las estimaciones, calcularemos los consumos de los distintos días que representen los puntos del periodo de validación y posteriormente se realizará una comparativa entre los resultados obtenidos y los valores reales medidos.

En el diagrama tipo que continua se ha representado un valor  $x$  que no es más el error permitido que dará por válido el proceso y que debe ser asumido por las partes implicadas en el procedimiento.

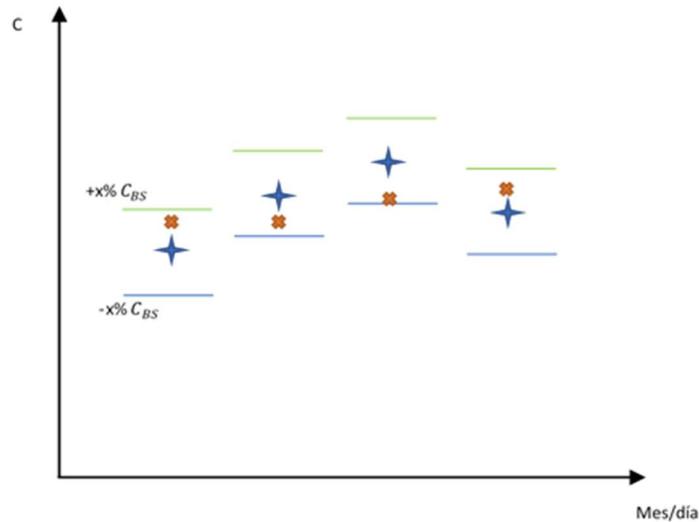


Figura 20. Bandas de error de la estimación de consumos.

### 2.4.2 Utilidades. Gestión energética, detección de anomalías.

Una de las funciones que tiene el método de M&V que se detallaba en la introducción del proyecto, era la detección de anomalías de consumo, uno de los objetivos de cualquier proceso de gestión energética que se realice a un edificio.

Este proceso parte de la predicción del consumo del edificio usando la metodología ya desarrollada de línea base. Una vez obtengamos el modelo y las predicciones de consumo del edificio para las condiciones reales de operación, es hora de aplicar los límites superiores e inferiores entre los cuales deberá de estar el consumo real de la edificación. De tal manera que el usuario encargado de la gestión energética o de forma automática se podrá desarrollar una serie de avisos cuando los consumos reales se encuentren fuera de los límites esperados de consumo.

# 3 CASO TEÓRICO. EDIFICIO DE OFICINAS

## 3.1 Alcance.

En el capítulo que se expone a continuación se hará uso de un caso conocido de consumos en un edificio terciario, en concreto un edificio de oficinas, para la prueba del procedimiento de desglose automático.

El edificio de oficinas es fruto de la hibridación de los datos reales de consumo por uso de un edificio conocido de las mismas características y de los consumos de climatización obtenidos mediante simulación empleando la herramienta unificada Lider-Calener.

El motivo de emplear un edificio fruto de la invención es debido a la inexistencia de mediciones desglosadas de consumo para un edificio gran terciario de calidad en un largo periodo de tiempo como el procedimiento requiere para su validación. Por ello, se ha desarrollado este caso inicial para contrastar los resultados obtenidos del procedimiento y poder tener una visión general del alcance y la credulidad de los números que aporta.

Entre los objetivos del apartado que sigue encontramos:

1. Validación de un procedimiento estandarizado de desglose útil para edificios diferentes.
2. Visualización del consumo diario total de un edificio mediante varios clústeres que representen todos los días del año. Es decir, reducir la dimensionalidad del problema inicial.
3. Desglosar y separar el consumo en climatización diario para posteriormente obtener una línea base que nos permita a partir de los clústeres que representan los días del año, obtener el consumo en climatización para un día/s en diferentes condiciones.

## 3.2 Ubicación y arquitectura del edificio problema.

El edificio objeto de estudio de naturaleza simulada se ubica en la zona norte de la ciudad de Sevilla, al sur de España en la zona climática B4. La instalación se encuentra dentro de la isla de la Cartuja, área de la ciudad donde se alojan la mayor parte de los grandes edificios de oficinas de empresas privadas y organismos públicos, así como varias facultades y el estadio olímpico.

La razón de ubicar el edificio en esta zona se debe a la disposición de una estación meteorológica en la Escuela Superior de Ingenieros de Sevilla de la que podremos disponer fácilmente las mediciones de los datos climáticos necesarios para nuestro problema.

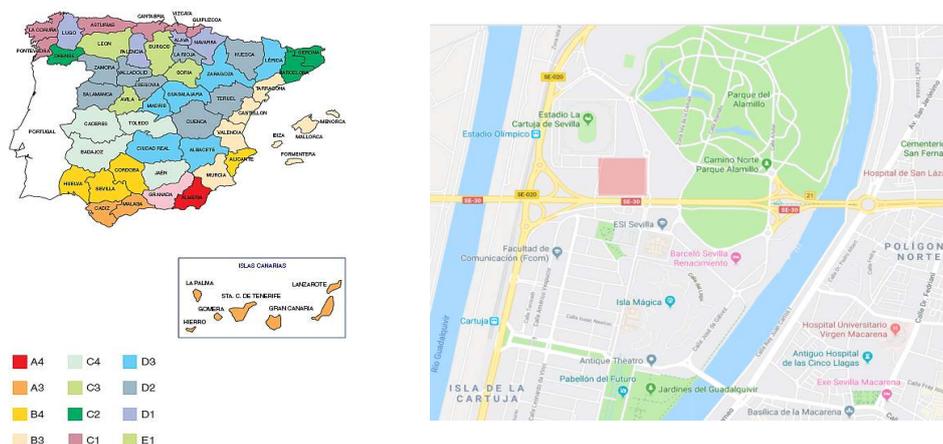


Figura 21. Ubicación edificio caso teórico.

Esta construcción se encuentra en un solar separado, no comparte ningún muro con alguna otra edificación y no recibe sombra de ningún edificio o estructura colindante. El total de su superficie habitable se distribuye en dos plantas divididas en diferentes espacios. Las coordenadas de su ubicación son las siguientes:

Latitud	37,41º
Longitud	-6,00º

Figura 22. Coordenadas caso teórico.

La construcción es un edificio de oficinas y despachos de varias empresas privadas. Está distribuido en dos plantas, éste cuenta a parte de las oficinas y despachos con cafetería, aseos, vestuarios y un gimnasio para los trabajadores de los distintos organismos que allí se alojan.

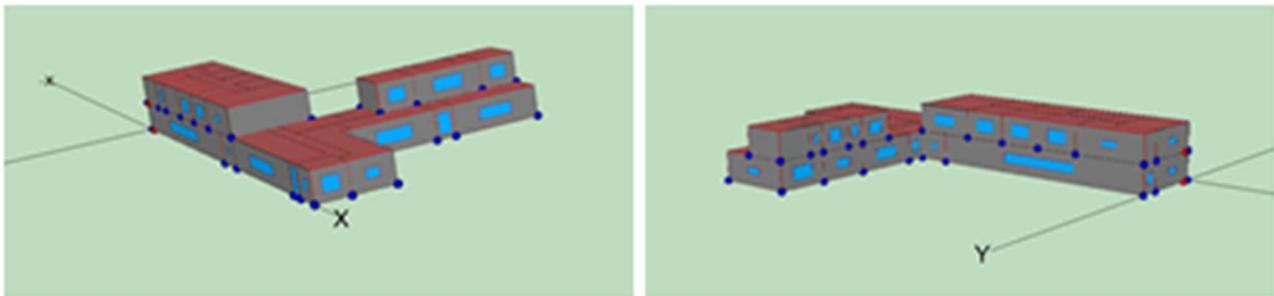


Figura 23. Modelo 3D caso teórico.

Como así se puede distinguir en los planos de zonificación, las oficinas se encuentran en la planta 1, y el resto de las habitaciones en la planta baja.

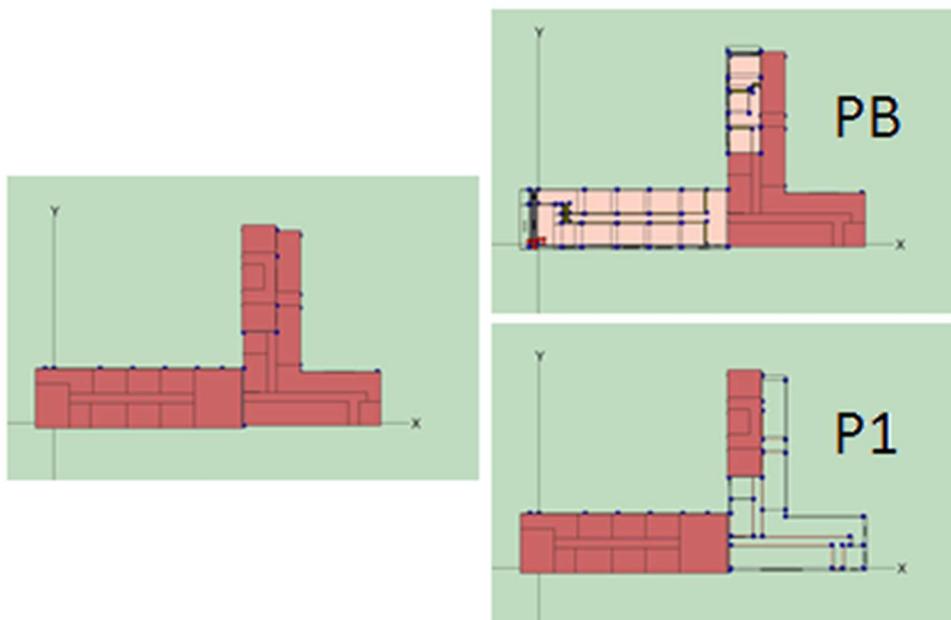


Figura 24. Planta edificio caso teórico.

El horario de uso de este edificio es de lunes a viernes de 9:00 a 21:00 y sábados de 9:00 a 15:00. Su ocupación es generalmente homogénea a lo largo de todo el año, y dado que es un edificio simulado, no se ha distinguido días festivos.

Aunque anteriormente se comenta el horario de uso y que la ocupación es generalmente homogénea, en la realidad rara vez es así, y en la componente real de los consumos por ocupación de nuestro edificio podemos ver que no difiere de los puramente reales, dado que encontramos días en los que los consumos no se corresponden totalmente con el horario descrito por el “cliente”. Ya que con este problema se quiere probar la validez del procedimiento de atacar a una situación real es por ello que se decidió de incluir consumos reales para poder ver su efectividad en el caso en el que los datos proporcionados no sean todo lo concisos que le gustaría al usuario.

### 3.3 Análisis de consumos del edificio a estudio.

El edificio a estudio utiliza como única fuente de energía la electricidad para cubrir todas las necesidades de sus ocupantes y del propio edificio, entre ellas:

- Sistemas de seguridad y antiincendios.
- Ascensores.
- Sistemas de domótica.
- Iluminación y equipos informáticos.
- ACS.
- Climatización.
- Otros.

Todas estas necesidades reportan su consumo a cada uno de los grupos que se explicaban en el procedimiento. El primero de ellos es el consumo fijo, en él, se engloban los consumos de los sistemas de seguridad y antiincendios, domótica, otros y ascensores. El consumo por uso, en este caso reflejará los consumos de ACS e iluminación y el último consumo clima es el formado por los equipos de acondicionamiento de aire.

Aunque los consumos se clasifiquen como fijo, uso y clima, es cierto que en realidad en menor o mayor medida todos estos consumos son dependientes de la ocupación del edificio, y es posible que consumos como el fijo, en la realidad no sean del todo constantes, sino ligeramente variables.

Para el desglose, en una primera parte, se analizará la capacidad del procedimiento de separar el consumo en estas tres partes y en una posterior de separarlo en solo dos: consumo clima y consumo usuario. Éste último representará el consumo asociado a los ocupantes y será la suma del consumo fijo y ocupación.

Dado que los datos de inicio lo permiten ya que el calendario contiene días de no ocupación, se podrá obtener las dos variantes de desglose y valorar el uso del desglose bicomponente o tricomponente.

$$C_{total} = C_{clima} + C_{usuario} = C_{fijo} + C_{ocupación} + C_{clima}$$

De partida disponemos de los datos de consumo desglosados en sus tres componentes medidos en base horaria de todos días de un año de medición, año 2007. En el gráfico que sigue se muestra la tendencia de consumo pasados a base mensual.

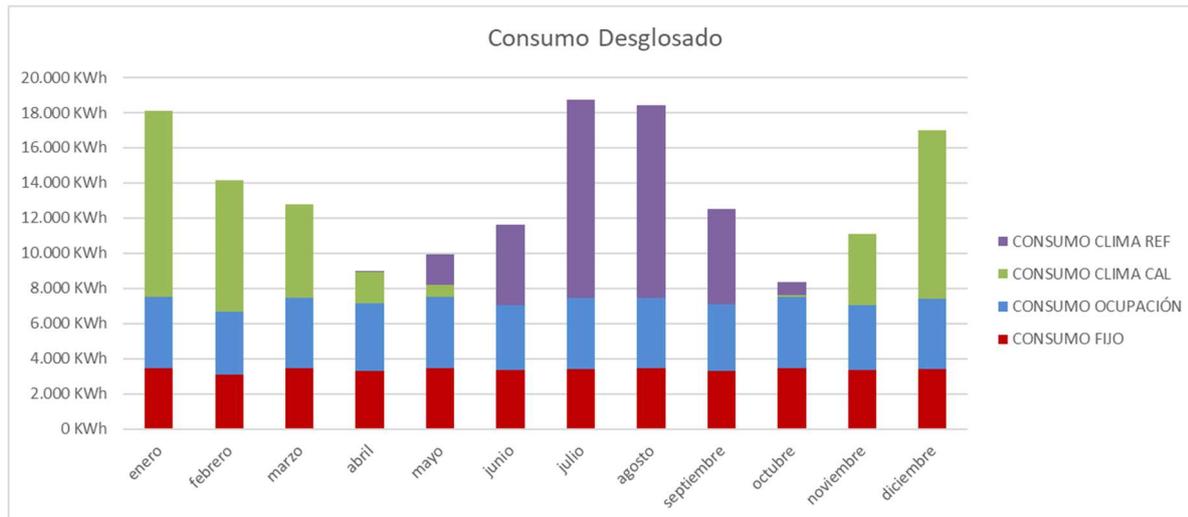


Figura 25. Consumo real mensual desglosado, caso teórico.

El consumo de climatización se ha subdivido en refrigeración y calefacción, para distinguir las estaciones de invierno y verano, pero dado que no es un requerimiento del procedimiento, en el desglose solo se tratará el consumo independientemente de si es en modo calefacción o refrigeración. El modo de uso del sistema será posible obtenerlo en el caso en el que fuera necesario comparando la temperatura exterior media del día con la temperatura equivalente del edificio.

El reparto de consumos graficado muestra la tendencia normal de consumos de edificios en nuestra zona climática, en el que se puede observar una estación intermedia o de entretiempo, donde el consumo en clima es mínimo en comparación con el resto de los meses. Esta estación se ubica entre los meses de abril-mayo y octubre-noviembre.

### 3.3.1 Uso y ocupación del edificio.

El edificio como se explicaba anteriormente mantiene un uso casi homogéneo a efectos de cálculo, aunque no real del todo, siguiendo el siguiente patrón:

	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO
7:00							
9:00	👤	👤	👤	👤	👤	👤	
12:00	👤	👤	👤	👤	👤	👤	
15:00	👤	👤	👤	👤	👤	👤	
18:00	👤	👤	👤	👤	👤		
21:00	👤	👤	👤	👤	👤		

Tipo día =1

Tipo día =2

Tipo día =3

Figura 26. Ocupación edificio caso teórico.

Desde el inicio es trivial ver la distinción entre tres tipos de día característicos en los que el edificio trabajará.

### 3.3.2 Climatología.

La climatología a la que se enfrenta el edificio es la variable esencial de la que dependerá el consumo eléctrico más fluctuante del edificio, el consumo de climatización. De la gama de variables climatológicas que se dispone, la de más fácil obtención e influencia en el consumo de climatización, es la temperatura media exterior del edificio, como así se observa en el gráfico.

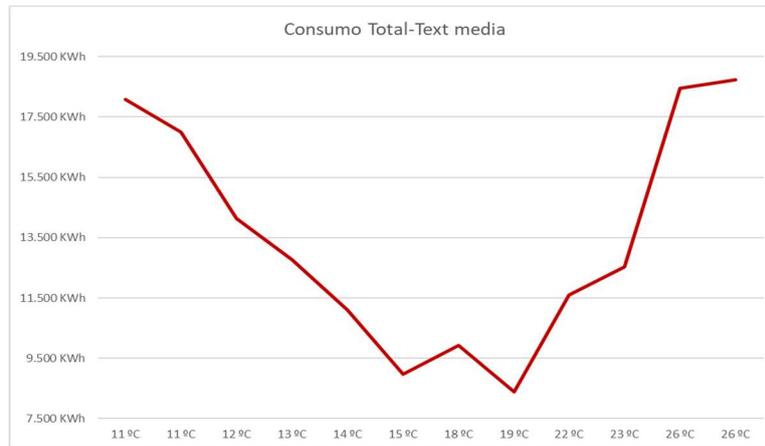


Figura 27. Evolución consumo temperatura media exterior.

Aunque el gráfico se ve influenciado por puntos donde el recinto no está ocupado y por lo tanto el consumo de ocupación es nulo, se puede observar como el consumo total de la edificación aumenta al disminuir la temperatura a partir de 15°C o al aumentar a partir de 19°C.

Otra variable que se estudiará incluir en el desglose será la radiación global por metro cuadrado. La estación meteorológica nos ofrece esta posibilidad y se realizará el estudio de si contemplarla favorece los resultados o no.

#### 3.3.2.1 Climatización. Modos de operación.

La temperatura equivalente marcará el modo de operación del sistema en función de la temperatura media exterior diaria que tengamos. Será un valor fundamental para el desglose de consumos y nos permitirá obtener los valores de consumo fijo y consumo de ocupación de la instalación.

Dado que disponemos del desglose de consumos es fácil calcular el valor real de este parámetro, pero en el caso real de estudio de un edificio, esto no es trivial. En este apartado calcularemos su valor de forma exacta y lo compararemos con el valor obtenido de la forma aproximada y que se usaría en el caso real.

Para el cálculo exacto de la temperatura equivalente se procede utilizando los valores de consumo en climatización en función de la temperatura media exterior. Recordar que para la representación que se presenta a continuación, se ha excluido los valores de consumo menores a 150Wh día para evitar ruido en su cálculo.

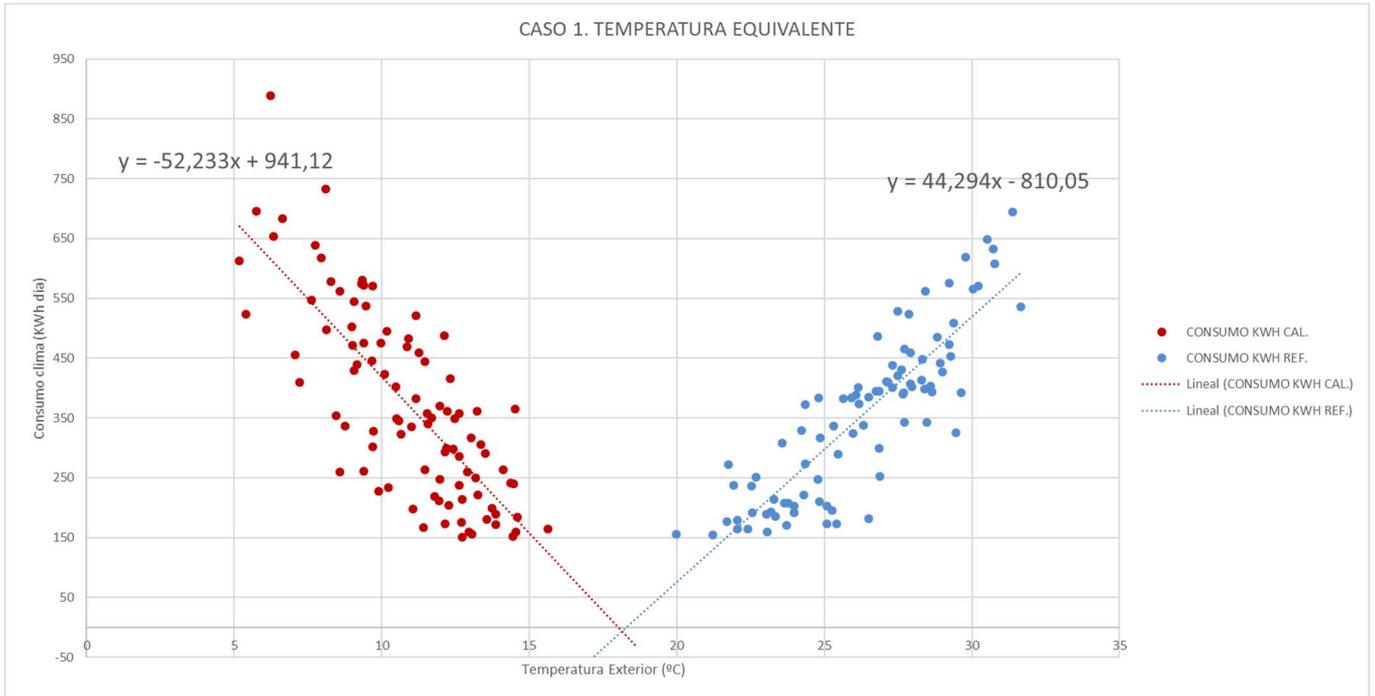


Figura 28. Cálculo temperatura equivalente real.

El corte de las rectas de regresión del consumo de refrigeración y de calefacción nos da un resultado de 18,12°C.

Para el cálculo de la temperatura equivalente en el caso real, se procede utilizando directamente el consumo total del edificio y la temperatura media diaria exterior. Dado que en esta situación no tendríamos certeza en qué mes se refrigera o calefacta el edificio tendremos que usar los datos de julio y agosto y de enero y diciembre únicamente para el cálculo.

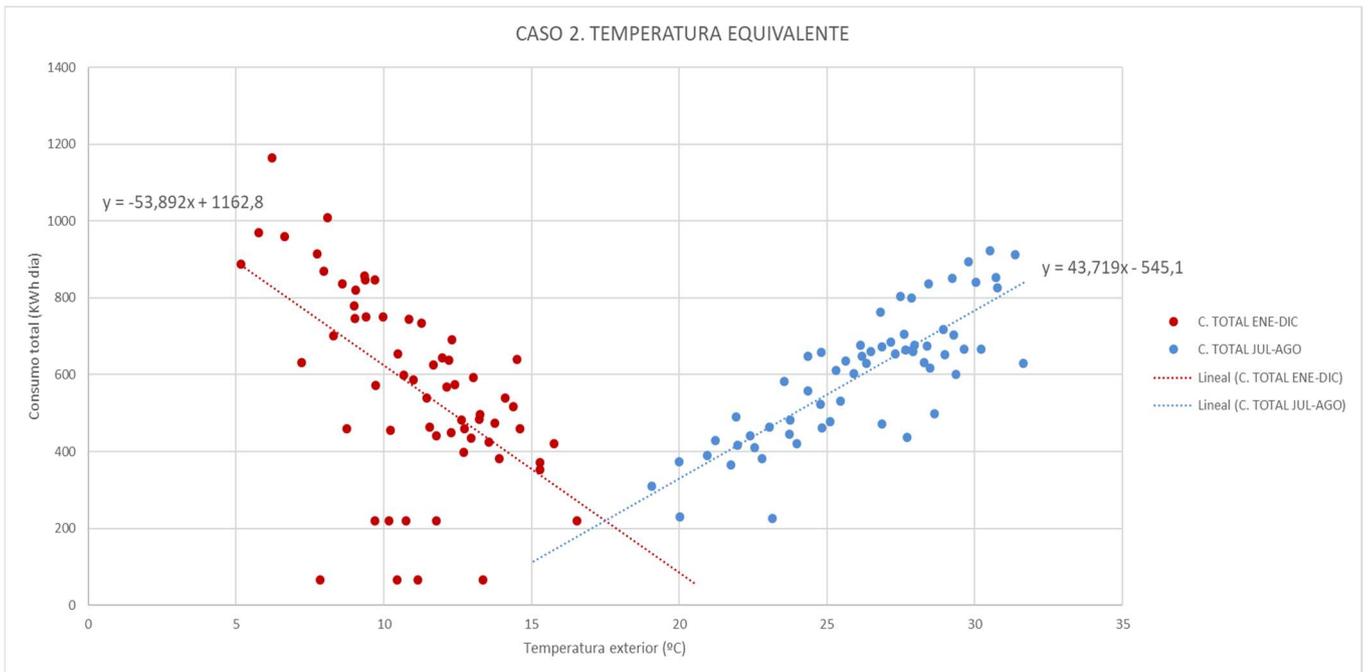


Figura 29. Cálculo temperatura equivalente sin desglose previo.

En este segundo caso el valor de la temperatura equivalente obtenido es 17,5°C.

Los resultados obtenidos se resumen en la siguiente tabla:

TEMPERATURA EQUIVALENTE	
Valor exacto	18,12°C
Valor aproximado (Problema Real)	17,5°C
<b>ERROR OBTENIDO</b>	<b>3,6%</b>

El error obtenido será arrastrado en el proceso de desglose, no es excesivo, y en cálculos posteriores para determinar la validez del procedimiento usaremos el valor aproximado, ya que sería el obtenido en el caso en el que no tuviéramos el desglose de consumos previamente.

Con el valor de la temperatura equivalente calculado podemos saber cuándo el sistema de acondicionamiento de aire refrigera, calefacta o no trabaja. Esto será la base del desglose de consumos que se desarrollará posteriormente.

### 3.4 CASOS A ESTUDIO.

En los siguientes apartados se va a evaluar el procedimiento de desglose en tres variantes distintas que llamaremos casos. Estos casos se distinguirán entre ellos en las variables de entrada que utilizan. El fin de esto es evaluar las diferencias entre ellos en temas de facilidad de cálculo, claridad de los resultados y calidad de los resultados, para posteriormente en casos reales evitar su cálculo o esforzarnos por tener disponible la/las variable/s que mejor se comportan.

Remarcar que estos casos que se estudian no son los únicos posibles, se ha optado por estos tres debido a las variables de las que se dispone de inicio, pero en otro problema distinto el usuario podría optar por incluir más u otras variables distintas en las entradas, siempre y cuando se respete que entre las entradas alguna refleje la climatología, preferiblemente la temperatura dado su importancia (ver 3.3.2) y otra la ocupación.

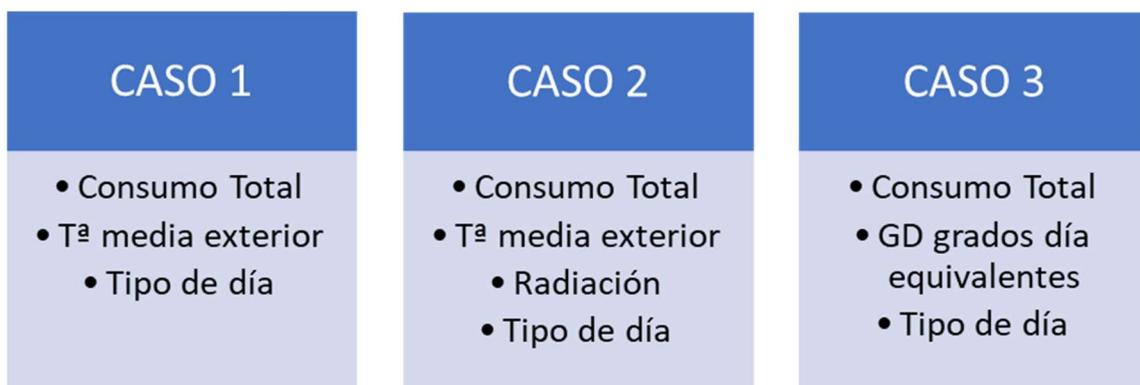


Figura 30. Descripción variables de las pruebas a realizar para desglose en edificio teórico.

### 3.5 CASO 1. Desglose de consumos.

El caso 1 abre los apartados de desglose de consumos, es el más fácil debido a que solo incluye las tres variables que se necesitan como mínimo para poder llevar a cabo el procedimiento.

#### 3.5.1 CASO 1. FASE 1. Preparación de los datos de entrada.

Los datos que se disponen para llevar a cabo el algoritmo son:

- Consumo total: consumo total de la edificación, en este caso se disponía de los datos de consumo horarios totales que se han pasado previamente a base diaria. Dado el orden de magnitud de la variable se ha calculado en KWh diario.
- Temperatura media exterior: esta variable ha sido obtenida a partir de una estación meteorológica cercana y se mide en °C.
- Tipo de día: como se exponía en 3.3.1, el cliente informa del calendario de uso de la edificación, tras analizar el % de uso de esta, se opta por clasificar todos los días en tres valores que representa la variable tipo de día en función del % de ocupación.

A continuación, se muestra un fragmento de la tabla que engloba el espacio dato del problema:

FECHA	T. EXT. °c	CONSUMO TOTAL (KWh)	Var. Tipo Día
01-ene	15,27	352,50	1
02-ene	12,96	434,87	1
03-ene	13,75	474,68	1
04-ene	11,45	539,16	1
05-ene	10,67	598,17	1
06-ene	10,24	475,19	2
07-ene	10,44	107,52	3
08-ene	9,34	856,13	1

Tabla 4. Fragmento datos de partida, CASO 1, edificio teórico.

Tras la tabulación de todos los datos en base diaria, como marca el procedimiento es necesario la normalización de todas las variables excepto la variable tipo de día dado que su valor es entero y solo tiene tres posibilidades.

Para la normalización se hace uso de la media y la desviación típica de las dos variables.

	Consumo [kWh]	Temp Ext Media [°C]
<b>Media</b>	12770,48	18,19
<b>Desviación típica</b>	3626,81	5,59

Tabla 5. Medidas estadísticas variables de entrada.

Con esto la fase 1 concluye, disponemos de las variables normalizadas y estas serán los inputs de la fase 2 del procedimiento.

	<b>Consumo</b>	<b>Temp Ext Media [°C]</b>	<b>Tipo día</b>
01/01/2007	-0,38	-0,37	1
02/01/2007	0,00	-0,73	1
03/01/2007	0,18	-0,61	1
04/01/2007	0,48	-0,97	1
05/01/2007	0,75	-1,09	1
06/01/2007	0,18	-1,16	2
07/01/2007	-1,50	-1,13	3
08/01/2007	1,93	-1,30	1

Tabla 6. Variables normalizadas, CASO 1, edificio teórico.

### 3.5.2 CASO 1. FASE 2. Clustering.

En la fase 2 se reducirá la dimensionalidad de los datos que disponemos. El objetivo de esta fase es conseguir la agrupación de todos los vectores de datos diarios en una serie de representantes. Estos representantes son los vectores centroides de los clústeres que se obtienen.

Para el clustering se realiza el algoritmo propuesto en el apartado 2, K-means. Este método consta de dos partes, cálculo del número óptimo de clústeres y la obtención de los clústeres en sí.

#### 3.5.2.1 CASO 1. FASE 2. Clustering. N° de clústeres.

En esta primera parte, se calcula el número de clústeres en los que se agruparán los datos. Este cálculo es bastante importante, dado que dependiendo de su valor el problema será más visual y fácil de interpretar que si el número de clústeres que lo representa es muy alto. Para este fin se optimizará el índice de Davies-Bouldin sin olvidar la importancia de los requerimientos previos:

- Índice de Davies-Bouldin menor a 0,9: la experiencia en cálculos previos y clustering marca que este índice debe ser menor al valor impuesto para conseguir una segmentación que no arrastre un error de aproximación demasiado alto.
- Variables enteras de ocupación de los centroides no decimales: las variables enteras de ocupación en los centroides de los clústeres deben de seguir siendo enteras, dado que los cálculos de consumos fijo y ocupación se basarán en la observación de esta variable. Se desechará cualquier desglose en el que más del 20% de los valores de una variable entera pase a ser decimal.
- Variedad de clústeres: el desglose deberá de concentrar al menos un clúster de cada tipo de día cuya temperatura se sitúe en la zona de no trabajo en clima.

Los requerimientos son los límites inferiores y superiores a partir con los cuales podremos determinar el menor número de clústeres que representarán todos los datos del año, ya que a menor número de clústeres mayor claridad y facilidad de cálculo del desglose.

Tras realizar varias pruebas en el algoritmo de cálculo, aunque el tiempo de cálculo es dependiente del intervalo de búsqueda del número de clústeres óptimo, se decide utilizar 150 iteraciones para llevar a cabo cada una de las pruebas que se realizan. El tiempo medio de cálculo del programa es 3 minutos para una capacidad de computación de un pc normal y se da por bueno este valor ya que no es excesivamente grande.

En este caso a estudio se han realizado un total de 5 pruebas, en las que para 150 iteraciones y fijando el intervalo de búsqueda en los distintos valores que se exponen se obtiene el número de clústeres óptimo que reporta el algoritmo y el valor del índice asociado a este valor. Aún queda comprobar si cada uno de los clústeres cumplen con los requerimientos impuestos, y tras comprobarlo se opta por la opción 2 como la óptima para llevar a cabo el clustering de todos los datos.

Prueba	Observaciones	Nº Iteraciones	Intervalo	Nº Óptimo de Clústeres	Valor Índice
1		150	[2:9]	9	0,7162
2		150	[2:15]	15	0,6750
3		150	[2:21]	20	0,6265
4		150	[2:27]	22	0,6209
5		150	[2:30]	22	0,6188

Tabla 7. Pruebas clustering, CASO 1, edificio teórico.

Aunque las pruebas 3, 4 y 5 ofrecen mejores resultados que la 2, siempre intentaremos seleccionar aquella opción que menor número de clústeres nos marque.

Como podemos observar en el gráfico, un incremento del número de clústeres no proporciona una mejora notable en el índice de Davies Bouldin, por ello podemos reafirmar que la prueba número 2 será la mejor opción para nuestro problema.

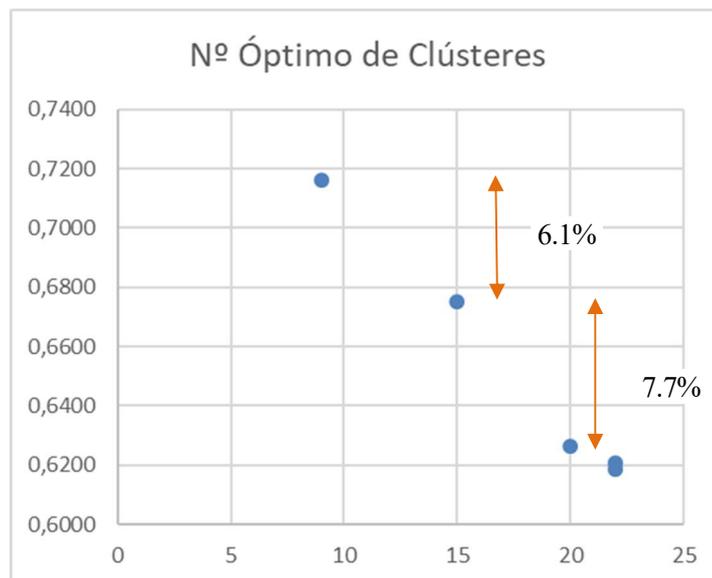


Figura 31. Número óptimo de clústeres edificio teórico.

En conclusión, tras el estudio realizado se opta por elegir 15 clústeres en los que agrupar todos los días del año con un valor obtenido del Índice de Davies-Bouldin de 0.675.

### 3.5.2.2 CASO 1. FASE 2. Clustering. Clústeres.

Una vez hemos seleccionado el número de clústeres, es el momento de calcular todos los clústeres que representan a nuestros datos. Cada clúster es un vector con los valores de las tres variables de entrada de su centroide.

Remarcar que en el punto anterior se exigía que el número de clústeres seleccionado tiene que reportar clústeres cuyas variables enteras sigan siendo enteras, la elección, aunque se escribía anteriormente se consigue probando cada uno de los resultados del nº óptimo de clústeres de cada prueba y viendo si los clústeres que se obtienen cumplen el requerimiento, es decir realizando el paso de obtención de los clústeres a la vez que se determinaba un número mínimo de clústeres de cada prueba.

Los resultados obtenidos son los siguientes:

Clúster	Consumo	Temp Ext Media	Tipo día
1	1,24	1,67	1,23
2	-1,14	0,02	2,00
3	2,15	-1,58	1,00
4	-0,37	1,15	3,00
5	0,25	1,08	1,13
6	-0,73	-0,12	1,00
7	-0,50	0,56	1,00
8	-0,32	-0,61	1,00
9	1,39	-1,20	1,00
10	0,27	-1,01	2,00
11	-1,50	-0,98	3,00
12	0,57	-0,87	1,00
13	-1,38	0,19	3,00
14	-0,62	0,72	2,00
15	-0,54	-0,79	2,00

Tabla 8. Clústeres, CASO 1, edificio teórico.

Tras esto, pasamos las variables normalizadas a sus unidades correspondientes y tabulamos cada uno de los representantes de todos los días del año.

CLUSTER	CONSUMO	T EXT	TIPO DÍA	nº componentes
Cluster 1	706,60 KWh	28,22 ºC	1	44
Cluster 2	185,53 KWh	17,75 ºC	2	14
Cluster 3	904,67 KWh	7,55 ºC	1	15
Cluster 4	354,20 KWh	24,95 ºC	3	13
Cluster 5	489,36 KWh	24,48 ºC	1	31
Cluster 6	275,21 KWh	16,83 ºC	1	49
Cluster 7	326,35 KWh	21,14 ºC	1	35
Cluster 8	365,04 KWh	13,75 ºC	1	42
Cluster 9	738,50 KWh	9,98 ºC	1	22
Cluster 10	494,82 KWh	11,15 ºC	2	15
Cluster 11	107,52 KWh	11,39 ºC	3	22
Cluster 12	559,42 KWh	12,09 ºC	1	34
Cluster 13	133,88 KWh	18,82 ºC	3	14
Cluster 14	300,56 KWh	22,20 ºC	2	10
Cluster 15	316,81 KWh	12,56 ºC	2	5

Tabla 9. Clusters variables normalizadas, CASO 1, edificio teórico.

El algoritmo ejecutado, también nos reporta un vector columna denominado idx, que nos dice a qué clúster pertenece todas y cada una de las mediciones de entrada del algoritmo. Es decir, cada elemento del vector dice a qué clúster pertenece cada día del año.

El heatmap que representa este vector es el siguiente:

Distribución día-cluster al que pertenece

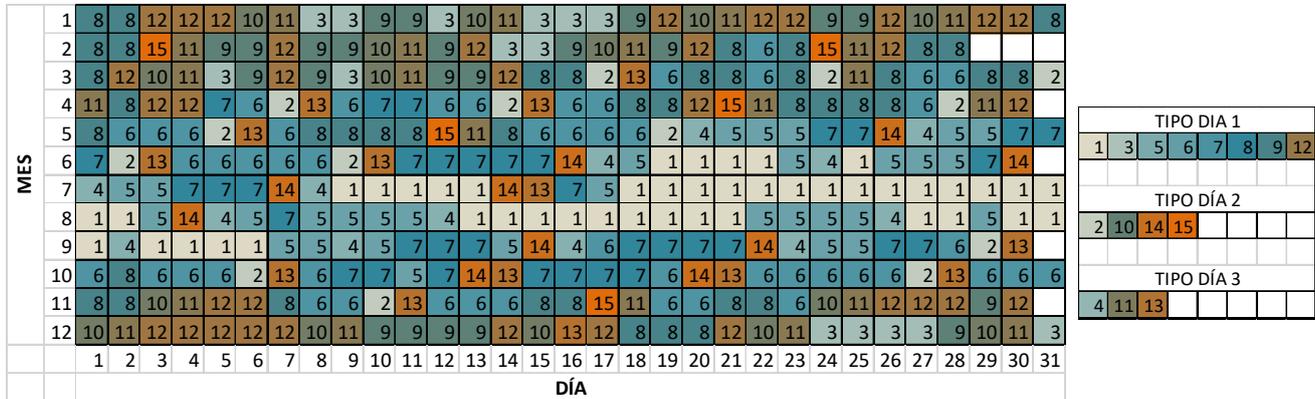


Figura 32. Hetmap clústeres CASO 1, edificio teórico.

Como se puede observar en el eje y el valor del mes y en el x el valor del día, dentro del heatmap el coeficiente del clúster al que pertenece cada día del año. Recordando que la fecha de los datos es el año 2007, podemos observar como por ejemplo los domingos 7/01/2007; 21/01/2007 y 22/04/2007 pertenecen al mismo clúster.

Una vez obtenidos los clústeres, visto que cumplen las normas impuestas, visualizados y comprobada la coherencia del agrupamiento pasamos al desglose de consumos.

**3.5.3 CASO 1. FASE 3. Desglose de consumos.**

El proceso de desglose de consumos variará en función del caso que analicemos. Dependiendo del problema que se presenta éste se realizará de forma bicomponente (consumo usuario y consumo clima) o tricomponente (consumo fijo, consumo ocupación y consumo clima).

El problema que se presenta desde inicio se sabe que existen días en los que la edificación no está ocupada, este detalle permitirá realizar el desglose en las tres componentes del consumo. De todas maneras, para la visualización de esto, se presentarán los resultados calculados para las dos variantes del desglose.

**3.5.3.1 CASO 1. FASE 3. Desglose de consumos. Consumo fijo.**

El consumo fijo proviene de los elementos que siempre funcionan dentro de la edificación, es una realidad que, aunque este consumo permanece casi constante, a nivel diario es posible observar fluctuaciones debido a que los equipos están estrechamente ligados a la ocupación del edificio.

La obtención de este consumo se realiza a partir de los días donde el edificio no está ocupado y los sistemas de climatización no funcionan, por ello para días donde el edificio esté ocupado, el valor calculado será ligeramente menor al real.

Por convención y observación del perfil de consumo visto en la sección 3.3.2.1 se fija un intervalo centrado en la temperatura equivalente con amplitud de 2°C donde los equipos de climatización no trabajan. Realizando una búsqueda en los clústeres que se disponen se restringe lo siguiente:

- Tipo día 3: día sin ocupación.
- Temperatura exterior perteneciente a [15.5 , 19.5]: días donde el sistema de climatización no trabaja.

Con ello se obtiene el siguiente clúster:

ITEM	CLUSTER	CONSUMO	T EXT	TIPO DÍA	nº componentes
13	Cluster 13	133,88 KWh	18,82 °C	3	14

Tabla 10. Cálculo consumo fijo, CASO 1, edificio teórico.

En el caso de haber obtenido varios clústeres dentro de las restricciones, se utilizaría aquel que menor distancia le separara de la temperatura equivalente.

De esta manera, el consumo fijo estimado de cada día del año es 133.8783 KWh siempre y cuando el consumo total sea mayor, en caso contrario el consumo fijo del edificio será el consumo total real del día en cuestión.

### 3.5.3.2 CASO 1. FASE 3. Desglose de consumos. Consumo de ocupación.

El cálculo del consumo de ocupación se realizará para cada uno de los tipos de día que se tienen en el calendario, esto se debe a que la variable tipo de día es aquella que define el tipo de ocupación que existe en la edificación y por ello para cada tipo de ocupación, un consumo de ocupación la representará.

En nuestro problema se definen 3 tipos de día, tipo de día 1 de lunes a viernes con una ocupación del 100% del horario laboral, tipo de día 2 los sábados con una ocupación del 50% del horario laboral y tipo de día 3 los domingos con una ocupación del 0%.

Es trivial que para el tipo de día 3 el consumo de ocupación es 0 KWh, para el resto de los tipos de día, se filtrará el conjunto de clústeres dentro del intervalo de no trabajo del sistema de clima.

CLUSTER	CONSUMO	T EXT	TIPO DÍA	CONSUMO FIJO	CONSUMO OCUPACIÓN
Cluster 2	185,53 KWh	17,75 °C	2	133,88 KWh	51,65 KWh
Cluster 6	275,21 KWh	16,83 °C	1	133,88 KWh	141,33 KWh
Cluster 13	133,88 KWh	18,82 °C	3	133,88 KWh	0,00 KWh

Tabla 11. Cálculo consumo de ocupación, CASO 1, edificio teórico.

Los resultados son coherentes obteniendo los días de lunes a viernes un consumo superior al del sábado.

El valor del consumo de ocupación del día a estimar tendrá los resultados obtenidos siempre y cuando la suma del consumo fijo estimado y el consumo de ocupación estimado del día sea menor o igual al consumo total real. En caso contrario, el consumo de ocupación tomará la diferencia entre el consumo total real y el consumo fijo estimado del edificio.

### 3.5.3.3 CASO 1. FASE 3. Desglose de consumos. Consumo de climatización.

A partir de la regla de que el consumo total estimado tiene que ser igual al consumo total real medido, el consumo de climatización del día a desglosar se obtiene realizando la diferencia del consumo total real menos los consumos ya calculados anteriormente: fijo y de ocupación.

### 3.5.3.4 CASO 1. FASE 3. Desglose de consumos. Estimación mensual.

Al aplicar todo lo desarrollado a lo largo de la fase 3, es posible desglosar el consumo total diario en las tres componentes que se ha explicado. Esto se debe gracias a que disponemos de días sin ocupación y es posible obtener el consumo fijo, en el caso contrario de no disponer de días no ocupados, el procedimiento hubiese pasado al cálculo del consumo de ocupación que englobaría tanto consumo fijo y consumo ocupación y tomaría el nombre de consumo usuario.

Estas dos opciones a las que se enfrenta el usuario del procedimiento forman las dos variantes que nos podemos encontrar y dado que el objetivo del primer caso a estudio es presentar las posibilidades del procedimiento, a continuación, se exponen los resultados obtenidos para las dos variantes.

#### VARIANTE 1

El problema ha permitido calcular el desglose de consumos en los tres consumos básicos que se han explicado, ya que existían días sin ocupación dentro del calendario. Los valores tabulados reales y estimados desglosados se muestran a continuación.

El desglose en forma gráfica queda de la siguiente forma:

	C. FIJO REAL	C. OCUPACIÓN REAL	C. CLIMATIZACIÓN REAL	C. FIJO ESTIMADO	C. OCUPACIÓN ESTIMADO	C. CLIMATIZACIÓN ESTIMADO
Enero	3.711,30 KWh	4.014,25 KWh	10.590,16 KWh	4.044,80 KWh	3.457,14 KWh	10.813,78 KWh
Febrero	3.344,40 KWh	3.558,15 KWh	7.474,43 KWh	3.643,16 KWh	3.033,15 KWh	7.700,67 KWh
Marzo	3.703,08 KWh	3.905,25 KWh	5.335,46 KWh	4.044,80 KWh	3.367,46 KWh	5.531,54 KWh
Abril	3.575,22 KWh	3.313,83 KWh	1.809,41 KWh	3.870,71 KWh	3.071,18 KWh	1.756,57 KWh
Mayo	3.714,30 KWh	3.414,62 KWh	2.427,54 KWh	4.054,93 KWh	3.342,56 KWh	2.158,95 KWh
Junio	3.579,78 KWh	3.079,41 KWh	4.589,32 KWh	3.972,08 KWh	3.174,48 KWh	4.101,94 KWh
Julio	3.698,52 KWh	3.254,09 KWh	11.296,30 KWh	4.150,23 KWh	3.315,81 KWh	10.782,88 KWh
Agosto	3.714,30 KWh	3.383,47 KWh	10.975,65 KWh	4.150,23 KWh	3.457,14 KWh	10.466,06 KWh
Septiembre	3.564,00 KWh	3.062,14 KWh	5.440,00 KWh	3.989,99 KWh	3.040,93 KWh	5.035,22 KWh
Octubre	3.714,30 KWh	3.444,63 KWh	846,97 KWh	4.066,09 KWh	3.004,23 KWh	935,57 KWh
Noviembre	3.591,00 KWh	3.080,83 KWh	4.050,07 KWh	3.910,92 KWh	3.308,41 KWh	3.502,57 KWh
Diciembre	3.687,30 KWh	3.254,09 KWh	9.593,42 KWh	4.018,44 KWh	3.226,13 KWh	9.290,25 KWh

Tabla 12. Resultados desglose de consumos, variante 1, CASO 1, edificio teórico.

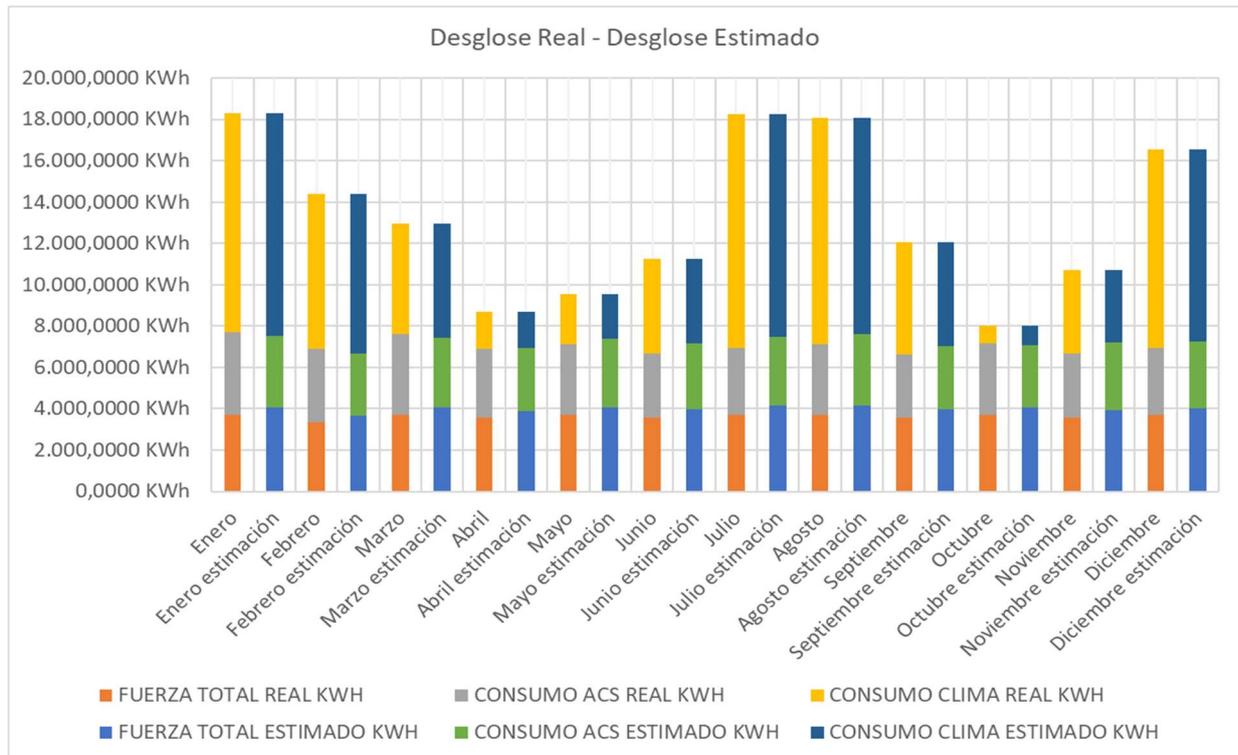


Figura 33. Desglose de Consumos Estimado, CASO 1, edificio teórico.

Los errores obtenidos para la variante 1 del desglose se muestran a continuación. El error promedio se puede observar mayor en el consumo fijo, esto se debe a la parte dependiente intrínseca dentro del consumo fijo, que se estimó los días donde la ocupación era nula.

MES	ERROR CONSUMO FIJO	ERROR C. OCUPACIÓN	ERROR C. CLIMATIZACIÓN
Enero	8,99%	13,88%	2,11%
Febrero	8,93%	14,75%	3,03%
Marzo	9,23%	13,77%	3,67%
Abril	8,26%	7,32%	2,92%
Mayo	9,17%	2,11%	11,06%
Junio	10,96%	3,09%	10,62%
Julio	12,21%	1,90%	4,55%
Agosto	11,74%	2,18%	4,64%
Septiembre	11,95%	0,69%	7,44%
Octubre	9,47%	12,78%	10,46%
Noviembre	8,91%	7,39%	13,52%
Diciembre	8,98%	0,86%	3,16%
<b>MAXIMO ERROR</b>	<b>12,21%</b>	<b>14,75%</b>	<b>13,52%</b>
<b>ERROR PROMEDIO</b>	<b>9,90%</b>	<b>6,73%</b>	<b>6,43%</b>
<b>ERROR PROMEDIO DESGLOSE</b>	<b>7,69%</b>		

Tabla 13. Errores, variante 1, CASO 1, edificio teórico.

## VARIANTE 2.

La variante 2 difiere como se ha comentado en que solo se desglosa en dos componentes el consumo total. Los errores obtenidos son los siguientes:

MES	ERROR CONSUMO USUARIO	ERROR C. CLIMATIZACIÓN
Enero	2,89%	2,11%
Febrero	3,28%	3,03%
Marzo	2,58%	3,67%
Abril	0,77%	2,92%
Mayo	3,77%	11,06%
Junio	7,32%	10,62%
Julio	7,38%	4,55%
Agosto	7,18%	4,64%
Septiembre	6,11%	7,44%
Octubre	1,24%	10,46%
Noviembre	8,21%	13,52%
Diciembre	4,37%	3,16%
<b>MAXIMO ERROR</b>	<b>8,21%</b>	<b>13,52%</b>
<b>ERROR PROMEDIO</b>	<b>4,59%</b>	<b>6,43%</b>
<b>ERROR PROMEDIO DESGLOSE</b>	<b>5,51%</b>	

Tabla 14. Errores, variante 2, CASO 1, edificio teórico.

### 3.5.3.5 CASO 1. FASE 3. Desglose de consumos. Conclusiones.

Como conclusión hay que comentar que el error promedio obtenido para la variante 1 del caso 1 es aceptable, al igual que en la variante 2. Dado que es un caso en el que las variables introducidas son básicas y de fácil obtención en cualquier proceso de medición, aunque se mantienen los errores en la aproximación del consumo de climatización, es más interesante aplicar la variante 2 por su mayor facilidad y menor cálculo si el objetivo fundamental es obtener el consumo en climatización del edificio.

## 3.6 CASO 2. Desglose de consumos.

El caso 2 es el segundo de los estudios que se realiza de desglose, en este caso la única diferencia que lo separa del primer caso estudiado es la inclusión de la variable radiación en el proceso de cálculo de clústeres y desglose.

### 3.6.1 CASO 2. FASE 1. Preparación de los datos de entrada.

Los datos que se incluyen como se ha comentado anteriormente son los mismos que en el anterior apartado, solo que añadiendo los valores diarios medios de radiación.

- Radiación: es la radiación global por metro cuadrado medida en la estación meteorológica que caracteriza la zona en la que se ubica nuestro edificio. Este tipo de dato no es uno de los fundamentales, pero dado su influencia en cuanto a climatología hablamos y la disposición de un piranómetro en la estación, se ha querido analizar el efecto de su uso en el procedimiento de desglose.

Se muestra como quedaría un fragmento de los datos tabulados de entrada primero en base diaria con sus unidades características y luego normalizados.

FECHA	T. EXT. °c	RAD (Wh/m <sup>2</sup> )	CONSUMO TOTAL (KWh)	Var. Tipo Día
01-ene	15,27	5371,67	352,50	1
02-ene	12,96	4362,44	434,87	1
03-ene	13,75	3594,13	474,68	1
04-ene	11,45	6221,37	539,16	1
05-ene	10,67	8321,20	598,17	1
06-ene	10,24	6175,79	475,19	2
07-ene	10,44	5492,12	107,52	3

Tabla 15. Datos de partida, CASO 2, edificio teórico.

Una vez normalizadas todas las variables no enteras, pasamos al método de clustering para agrupar todos los datos diarios.

FECHA	Consumo	Temp Ext Media	Radiación	Tipo día
01/01/2007	-0,38	-0,37	-1,05	1
02/01/2007	0,00	-0,73	-1,25	1
03/01/2007	0,18	-0,61	-1,40	1
04/01/2007	0,48	-0,97	-0,87	1
05/01/2007	0,75	-1,09	-0,45	1
06/01/2007	0,18	-1,16	-0,88	2
07/01/2007	-1,50	-1,13	-1,02	3

Tabla 16. Variables normalizadas, CASO 2, edificio teórico.

### 3.6.2 CASO 2. FASE 2. Clustering.

Antes de exponer los resultados de clustering hay que comentar que toda aquella variable que se introduzca en el algoritmo de clustering provocará una mayor exigencia de ajuste al mismo. Es decir, incluir la radiación ha aumentado la dimensión de los datos de entrada en una componente más, por lo tanto el algoritmo ya no trata de ajustar aquellos datos que más se parezcan en temas de consumo, temperatura y tipo de día sino que también compara los datos de radiación que reciben, por ello es razonable que al incluir un mayor número de variables el algoritmo ofrezca unos valores más altos de mínimo número de clústeres en los que agrupar los datos para mantener un mismo valor del índice de calidad del algoritmo de Davies-Bouldin.

#### 3.6.2.1 CASO 2. FASE 2. Clustering. N° de clústeres.

Mantendremos las mismas exigencias que en el caso 1, las distintas pruebas realizadas son las que siguen, y se optó por la resaltada en amarillo dado que fue la única que permitió cumplir todas ellas. Remarcar que a partir de este valor mínimo de la prueba 7 se obtuvieron mejores resultados, pero a costa de un mayor número de clústeres que ya de primera hora era algo elevado.

Podemos ver lo que se comentaba al principio de esta sección, la variabilidad del índice es mucho menor que en el modelo con menor número de variables, el ajuste es de menor calidad (mayor índice) dado que ahora se incluye una variable más y el número de clústeres mínimo es notablemente más elevado que en el primer caso.

Prueba	Observaciones	Nº Iteraciones	Intervalo	Nº Óptimo de Clústeres	Valor Índice
1		150	[2:9]	6	0,9905
2		150	[2:15]	14	0,9567
3		150	[2:21]	20	0,9123
4		150	[2:24]	23	0,9047
5		150	[2:27]	26	0,8797
6	Obtenemos tipo de días muy decimales. No nos vale para realizar el desglose	150	[2:30]	26	0,8776
7		150	[2:33]	33	0,8657

Tabla 17. Pruebas clustering, CASO 2, edificio teórico.

### 3.6.2.2 CASO 2. FASE 2. Clustering. Clústeres.

Los clústeres obtenidos en las unidades características de cada variable y el heatmap que representa el año son:

CLUSTER	CONSUMO	T EXT	RAD	TIPO DÍA	nº componentes
Cluster 1	374,91 KWh	15,01 °C	4,91 KWh/m <sup>2</sup>	1	19
Cluster 2	319,82 KWh	21,83 °C	19,18 KWh/m <sup>2</sup>	3	7
Cluster 3	772,02 KWh	9,47 °C	5,87 KWh/m <sup>2</sup>	1	25
Cluster 4	364,16 KWh	23,84 °C	11,05 KWh/m <sup>2</sup>	3	8
Cluster 5	161,63 KWh	19,55 °C	14,76 KWh/m <sup>2</sup>	2	11
Cluster 6	390,50 KWh	11,98 °C	8,15 KWh/m <sup>2</sup>	2	15
Cluster 7	470,81 KWh	23,60 °C	18,25 KWh/m <sup>2</sup>	1	13
Cluster 8	346,77 KWh	13,95 °C	9,09 KWh/m <sup>2</sup>	1	22
Cluster 9	925,53 KWh	30,51 °C	4,91 KWh/m <sup>2</sup>	1	1
Cluster 10	708,22 KWh	27,87 °C	17,93 KWh/m <sup>2</sup>	1	18
Cluster 11	51,88 KWh	10,54 °C	6,48 KWh/m <sup>2</sup>	3	9
Cluster 12	138,76 KWh	12,76 °C	16,81 KWh/m <sup>2</sup>	3	6
Cluster 13	433,62 KWh	24,25 °C	9,35 KWh/m <sup>2</sup>	1	6
Cluster 14	690,63 KWh	28,17 °C	9,00 KWh/m <sup>2</sup>	1	7
Cluster 15	709,95 KWh	10,46 °C	12,65 KWh/m <sup>2</sup>	1	5
Cluster 16	287,90 KWh	18,78 °C	16,27 KWh/m <sup>2</sup>	1	10
Cluster 17	143,63 KWh	16,31 °C	3,80 KWh/m <sup>2</sup>	2	6
Cluster 18	258,89 KWh	18,99 °C	8,73 KWh/m <sup>2</sup>	1	18
Cluster 19	301,38 KWh	21,62 °C	5,38 KWh/m <sup>2</sup>	1	12
Cluster 20	182,81 KWh	19,21 °C	12,88 KWh/m <sup>2</sup>	3	6
Cluster 21	530,87 KWh	9,38 °C	18,77 KWh/m <sup>2</sup>	1	1
Cluster 22	489,15 KWh	27,26 °C	14,55 KWh/m <sup>2</sup>	2	7
Cluster 23	568,43 KWh	11,53 °C	6,61 KWh/m <sup>2</sup>	1	32
Cluster 24	709,72 KWh	29,21 °C	16,95 KWh/m <sup>2</sup>	3	9
Cluster 25	270,82 KWh	16,46 °C	11,48 KWh/m <sup>2</sup>	1	20
Cluster 26	681,04 KWh	27,13 °C	13,66 KWh/m <sup>2</sup>	1	13
Cluster 27	979,08 KWh	6,57 °C	5,84 KWh/m <sup>2</sup>	1	7
Cluster 28	348,42 KWh	14,52 °C	1,74 KWh/m <sup>2</sup>	1	6
Cluster 29	289,30 KWh	15,47 °C	19,64 KWh/m <sup>2</sup>	1	6
Cluster 30	422,85 KWh	22,45 °C	14,01 KWh/m <sup>2</sup>	1	16
Cluster 31	269,11 KWh	20,48 °C	9,13 KWh/m <sup>2</sup>	2	6
Cluster 32	206,58 KWh	12,62 °C	5,87 KWh/m <sup>2</sup>	3	10
Cluster 33	353,81 KWh	12,88 °C	14,48 KWh/m <sup>2</sup>	1	8

Tabla 18. Resultados clústeres, CASO 2, edificio teórico.

#### Distribución día-cluster al que pertenece

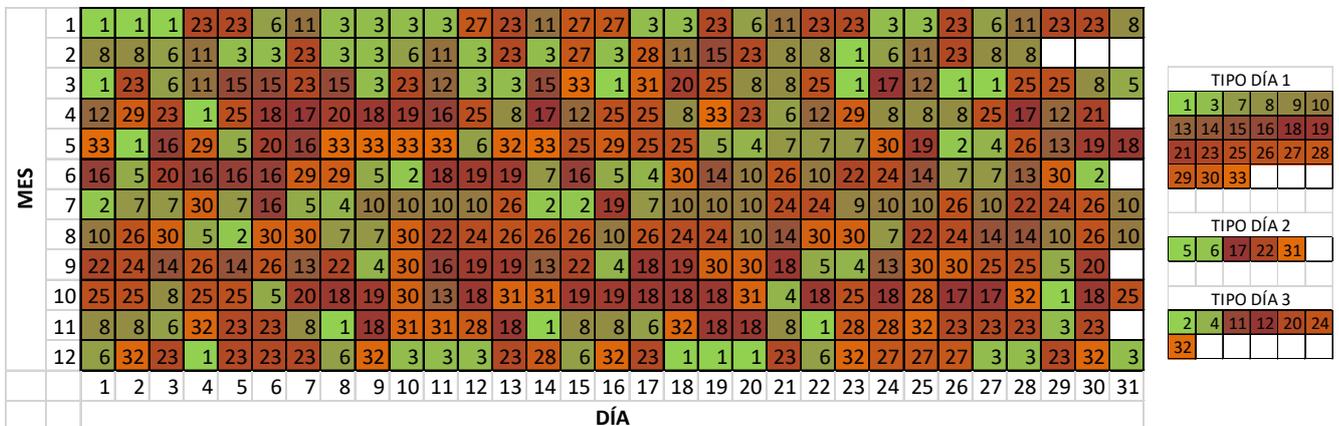


Figura 34. Hetmap clústeres CASO 2, edificio teórico.

La variabilidad introducida en cuanto a número de representantes de nuestro año es notable en el heatmap, hemos obtenido un mayor número de clústeres, más parecidos entre vecinos en cuanto temperatura y consumos, menos poblados y con componentes más cercanas entre sí al centroide que le representa que en el caso anterior. Aunque hemos empeorado el concepto base de la utilización en el clustering en este procedimiento, que es, la reducción de la dimensionalidad del espacio dato.

### 3.6.3 CASO 2. FASE 3. Desglose de consumos.

Para el desglose de consumos seguiremos la misma estrategia utilizada en el caso 1 y que se detalla en el procedimiento. Es posible que en la búsqueda de algún clúster representativo para el consumo fijo o de ocupación, encontremos varios de ellos que cumplan las restricciones, en esos casos procederemos a calcular la distancia que le separa del valor ideal de temperatura equivalente cada una de las opciones, y elegiremos la más cercana.

#### 3.6.3.1 CASO 2. FASE 3. Desglose de consumos. Consumo fijo.

El consumo fijo obtenido es 182,8134 KWh diarios, un valor bastante superior al real (valor medio 119,4 KWh diarios).

ITEM	CLUSTER	CONSUMO	T EXT	RAD	TIPO DÍA	nº componentes
20	Cluster 20	182,81 KWh	19,21 °C	12,88 KWh/m <sup>2</sup>	3	6

Tabla 19. Cálculo consumo fijo, CASO 2, edificio teórico.

Como antes se comentó, este valor lo tomará aquellos días cuyo consumo total sea menor al valor calculado de consumo fijo, en el caso en el que el consumo total real sea menor, el consumo fijo estimado será igual a ese valor.

Si observamos, el consumo fijo difiere más de la situación real que en el caso anterior, esto se debe principalmente a la mayor distancia en términos de temperatura del clúster representante del consumo fijo en el caso 2 que en el caso 1, de tal manera que observamos que como el procedimiento se basa en el uso principal de la temperatura exterior y la temperatura equivalente, el haber incluido la radiación como variable de entrada no está beneficiando mucho al desglose.

#### 3.6.3.2 CASO 2. FASE 3. Desglose de consumos. Consumo de ocupación.

Como se ha comentado anteriormente en este caso obtenemos varios clústeres que cumplen las restricciones para el primer tipo de día. Elegimos el más cercano a la temperatura equivalente.

CLUSTER	CONSUMO	T EXT	RAD	TIPO DÍA	DISTANCIA A Teq	CONSUMO FIJO	CONSUMO OCUPACIÓN
Cluster 16	287,90 KWh	18,78 °C	16,27 KWh/m <sup>2</sup>	1	1,29	182,81 KWh	
Cluster 18	258,89 KWh	18,99 °C	8,73 KWh/m <sup>2</sup>	1	1,49	182,81 KWh	76,07 KWh
Cluster 25	270,82 KWh	16,46 °C	11,48 KWh/m <sup>2</sup>	1	1,04	182,81 KWh	
Cluster 17	143,63 KWh	16,31 °C	3,80 KWh/m <sup>2</sup>	2		182,81 KWh	0,00 KWh
Cluster 20	182,81 KWh	19,21 °C	12,88 KWh/m <sup>2</sup>	3		182,81 KWh	0,00 KWh

Tabla 20. Cálculo consumo ocupación, CASO 2, edificio teórico.

Como comentario excepcional, al obtener anteriormente un consumo fijo excesivamente alto, esto ha implicado que los consumos de ocupación sean bajos y en el caso del tipo de día 2 es 0.

### 3.6.3.3 CASO 2. FASE 3. Desglose de consumos. Consumo de climatización.

Se puede observar que el error principalmente se concentra en el reparto entre consumo fijo y ocupación y no en el consumo usuario total. Un ejemplo de esto es el mes de julio, cuya carga en consumo de climatización es considerable y si se detalla la comparación real y estimada, el cálculo no ha arrastrado un error importante a nivel diario.

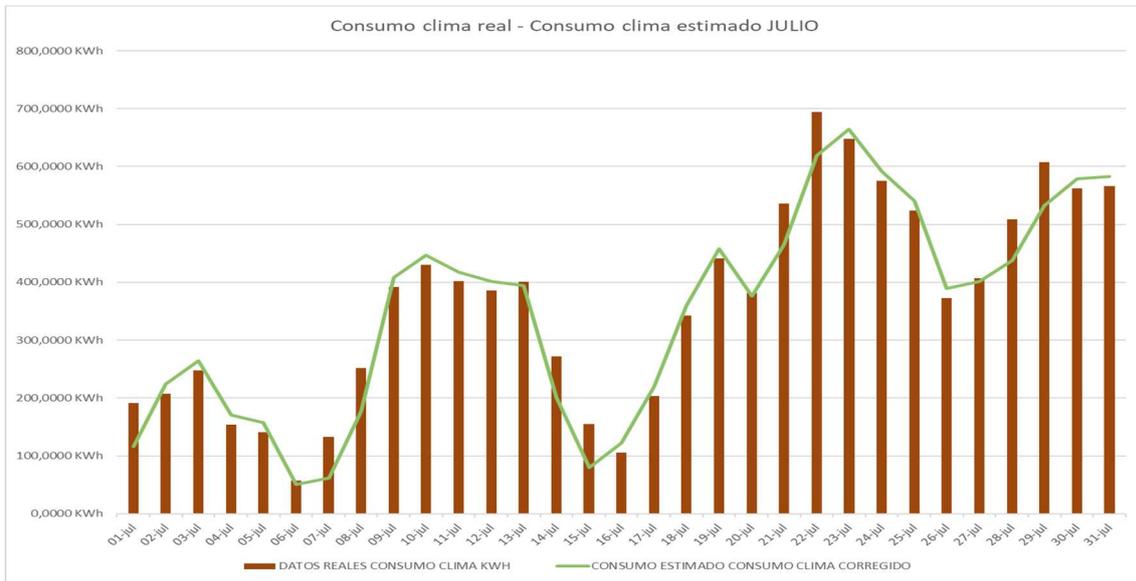


Figura 35. Consumo clima, desviaciones, CASO 2, edificio teórico.

### 3.6.3.4 CASO 2. FASE 3. Desglose de consumos. Estimación mensual.

La estimación mensual frente a la real y el error obtenido son los siguientes para la variante 1 más completa.

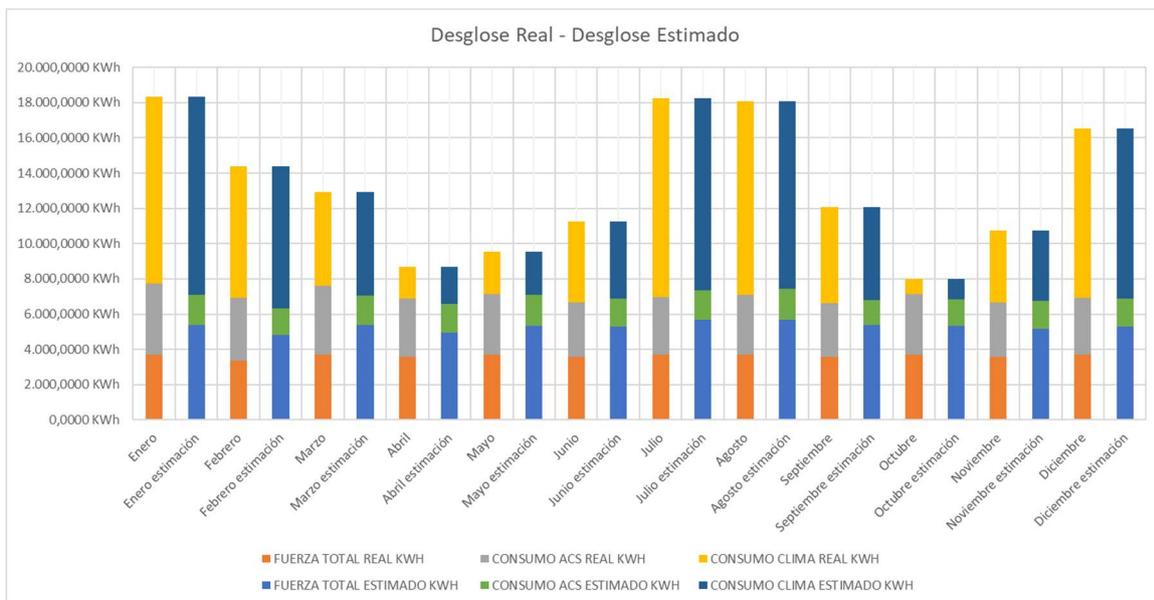


Figura 36. Desglose de consumos estimados, CASO 2, edificio teórico.

Se observa un error algo elevado en el desglose en los meses de no climatización como se muestra en la tabla siguiente, aunque si nos fijamos en la variante 2, este error no es tan relevante.

MES	ERROR C. FIJO	ERROR C. OCUPACIÓN	ERROR C. CLIMATIZACIÓN
Enero	44,59%	56,41%	5,76%
Febrero	44,05%	57,24%	7,54%
Marzo	44,91%	57,14%	10,66%
Abril	38,91%	51,79%	17,98%
Mayo	43,74%	48,76%	1,65%
Junio	47,60%	48,32%	4,71%
Julio	53,23%	48,57%	3,44%
Agosto	52,58%	48,29%	2,91%
Septiembre	50,59%	52,80%	3,42%
Octubre	43,45%	55,93%	36,92%
Noviembre	43,45%	48,15%	1,90%
Diciembre	43,49%	50,91%	0,55%
<b>MAXIMO ERROR</b>	<b>53,23%</b>	<b>57,24%</b>	<b>36,92%</b>
<b>ERROR PROMEDIO</b>	<b>45,88%</b>	<b>52,03%</b>	<b>8,12%</b>
<b>ERROR PROMEDIO GLOBAL</b>	<b>35,34%</b>		

Tabla 21. Errores, variante 1, CASO 2, edificio teórico.

A parte de estos dos casos mal aproximados en términos de consumo de climatización, el error promedio en cuanto al desglose del consumo usuario, es decir el consumo fijo y consumo de ocupación es algo alto aunque todo esto se solventa en la variante 2, la cual no lo reflejará.

Para la variante 2 del desglose, obtenemos lo siguiente:

MES	ERROR CONSUMO USUARIO	ERROR CONSUMO CLIMA
Enero	7,89%	5,76%
Febrero	8,16%	7,54%
Marzo	7,47%	10,66%
Abril	4,72%	17,98%
Mayo	0,56%	1,65%
Junio	3,24%	4,71%
Julio	5,58%	3,44%
Agosto	4,50%	2,91%
Septiembre	2,81%	3,42%
Octubre	4,37%	36,92%
Noviembre	1,15%	1,90%
Diciembre	0,77%	0,55%
<b>MAXIMO ERROR</b>	<b>8,16%</b>	<b>36,92%</b>
<b>ERROR PROMEDIO</b>	<b>4,27%</b>	<b>8,12%</b>
<b>ERROR PROMEDIO DESGLOSE</b>	<b>6,19%</b>	

Tabla 22. Errores, variante 2, CASO 2, edificio teórico.

### 3.6.3.5 CASO 2. FASE 3. Desglose de consumos. Conclusiones.

Excepto en casos puntuales el procedimiento aproxima bien en el desglose bicomponente. Pero dado que en este caso se ha empleado una variable más y se podría haber esperado mejores resultados ya que el cálculo es mayor, no ha sido así.

Como conclusión a este apartado, se puede observar que la variable radiación ha supuesto un incremento en el cálculo matemático, pero no una mejora en ninguna de las variantes del desglose, así podemos terminar diciendo que para estos casos la radiación global ha supuesto ruido en el problema más que una ayuda.

## 3.7 CASO 3. Desglose de consumos.

El caso 3, último del conjunto de los tres estudios realizados para la valoración del procedimiento de desglose, parte de la intención de mejorar los resultados obtenidos en el caso 1 y de visualizar que la cantidad de variables que se incluyen en el algoritmo de clustering no supone una mejoría sino un empeoramiento de los resultados si no se tiene en cuenta la calidad de las mismas.

En este caso, se incluye la variable grados día equivalentes para sustituir la variable temperatura exterior.

### 3.7.1 CASO 3. FASE 1. Preparación de los datos de entrada.

Los datos a introducir en el algoritmo son los mismos del caso 1, simplemente sustituyendo la temperatura media exterior diaria por los grados día equivalentes.

- Grados día equivalentes: los grados día equivalentes es una variable de uso en climatización que muestra la necesidad de uso de los equipos de climatización en un edificio y en que modo de operación trabajan. De tal manera que si la variable toma valores negativos quiere decir que la temperatura exterior es inferior a la temperatura equivalente y por lo tanto el sistema trabaja en modo calefacción, en el caso positivo trabajaría en modo refrigeración. Recordamos la fórmula de cálculo de esta variable.

$$GDE = (\overline{T_{ext}} - T_{equivalente}) \times 24$$

### 3.7.2 CASO 3. FASE 2. Clustering.

Las pruebas realizadas en el algoritmo fueron las siguientes.

Prueba	Observaciones	Nº Iteraciones	Intervalo	Nº Óptimo de Clústeres	Valor Índice
1		150	[2:9]	9	0,7171
2		150	[2:15]	12	0,6556
3		150	[2:21]	21	0,6347
4		150	[2:25]	23	0,6246
5		150	[2:30]	26	0,6198

Tabla 23. Pruebas clustering, CASO 3, edificio teórico.

A igualdad de número de variables y de tipología de las mismas, se eligió la opción 3 para cumplir los requerimientos del procedimiento, resultando un valor del índice ligeramente menor pero un número de clústeres mínimo mayor que en el caso 1.

### 3.7.2.1 CASO 3. FASE 2. Clustering. Clústeres.

El heatmap asociado a este caso se adjunta a continuación:

Distribución día-cluster al que pertenece

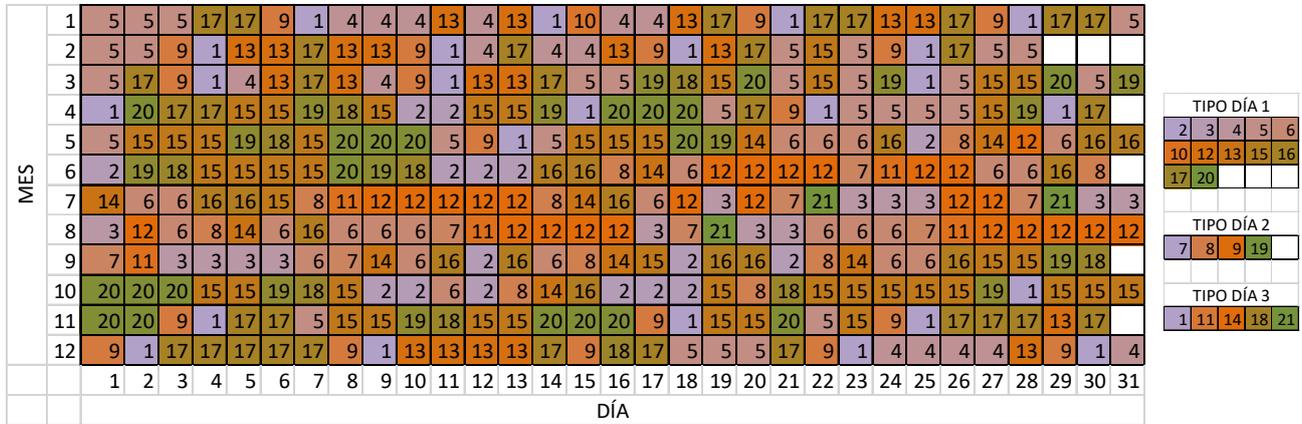


Figura 37. Hetmap clústeres CASO 3, edificio teórico.

### 3.7.3 CASO 3. FASE 3. Desglose de consumos.

Dado que en este caso no se dispone de la variable temperatura exterior, no es posible llevar a cabo la estrategia marcada en los dos casos anteriores, ya que su funcionamiento se basaba en la comparación de la temperatura exterior con la temperatura equivalente.

De forma análoga para este caso, se define un nuevo intervalo para los grados día equivalentes. Dado que el sistema no trabaja por consenso cuando la temperatura exterior se mantiene en el intervalo más menos 2°C centrado en la temperatura equivalente, es lo mismo decir que el sistema no trabaja cuando los grados día equivalentes se mantengan en el intervalo [-48, 48].

#### 3.7.3.1 CASO 3. FASE 3. Desglose de consumos. Consumo fijo.

ITEM	CLUSTER	Consumo [kWh]	GD equivalentes	Radiación [KWh/m <sup>2</sup> ]	Tipo día	nº componentes
18	Cluster 18	119,01 KWh	39,82 °C		3	10

Tabla 24. Cálculo consumo fijo, CASO 3, edificio teórico.

Observamos el buen acercamiento al valor medio real (119,4 KWh diarios) conseguido en la estimación del consumo fijo.

#### 3.7.3.2 CASO 3. FASE 3. Desglose de consumos. Consumo de ocupación.

CLUSTER	CONSUMO	T EXT	TIPO DÍA	CONSUMO FIJO	CONSUMO OCUPACIÓN
Cluster 15	278,85 KWh	-5,04 °C	1	119,01 KWh	159,84 KWh
Cluster 19	185,51 KWh	6,02 °C	2	119,01 KWh	66,50 KWh
Cluster 18	119,01 KWh	39,82 °C	3	119,01 KWh	0,00 KWh

Tabla 25. Cálculo consumo ocupación, CASO 3, edificio teórico.

### 3.7.3.3 CASO 3. FASE 3. Desglose de consumos. Consumo de climatización.

El consumo de climatización se obtuvo de la misma manera que en los casos anteriores aplicando la diferencia al total real con los consumos que se han calculado previamente.

### 3.7.3.4 CASO 3. FASE 3. Desglose de consumos. Estimación mensual.

El gráfico de desglose mensual queda de la siguiente forma:

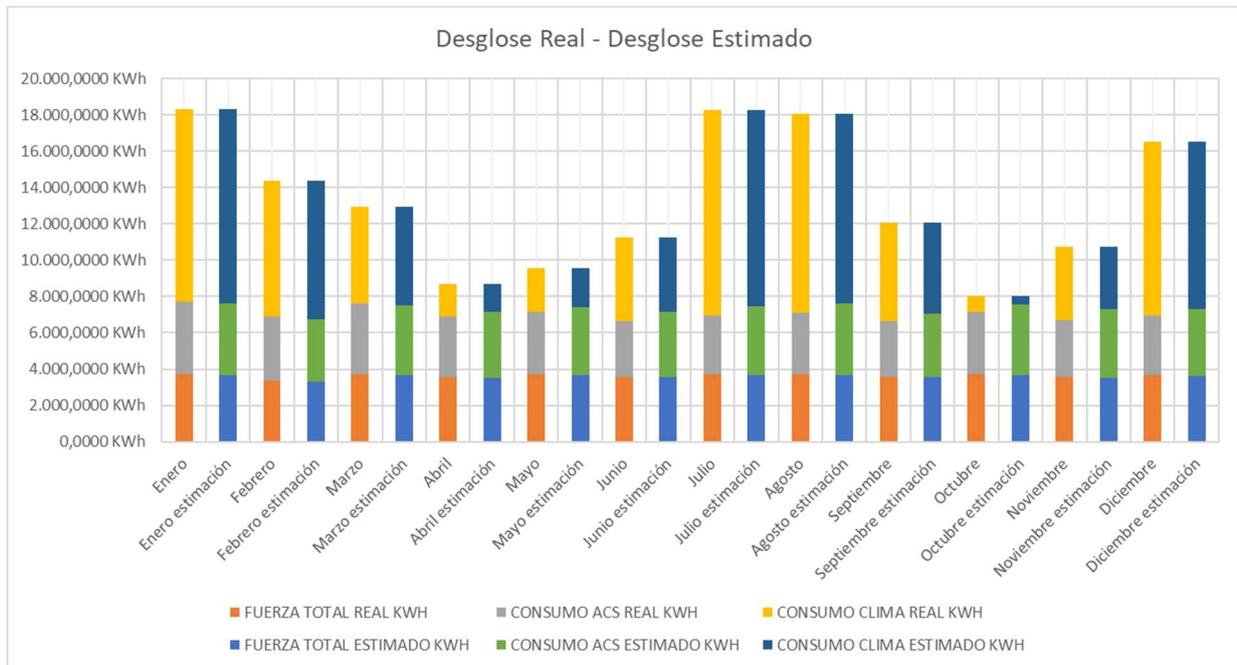


Figura 38. Desglose de consumos estimados, CASO 3, edificio teórico.

MES	ERROR C. FIJO	ERROR C. OCUPACIÓN	E. C. CLIMATIZACIÓN
Enero	4,83%	1,79%	1,32%
Febrero	5,72%	2,68%	2,05%
Marzo	5,65%	3,57%	2,17%
Abril	6,48%	6,28%	13,23%
Mayo	5,04%	6,22%	12,07%
Junio	6,94%	2,66%	10,59%
Julio	7,80%	5,77%	4,60%
Agosto	6,06%	1,23%	4,86%
Septiembre	8,86%	8,91%	7,62%
Octubre	5,41%	3,88%	47,62%
Noviembre	5,05%	2,24%	15,41%
Diciembre	7,06%	8,10%	3,96%
<b>MAXIMO ERROR</b>	<b>8,86%</b>	<b>8,91%</b>	<b>47,62%</b>
<b>ERROR PROMEDIO DESGLOSE</b>	<b>6,24%</b>	<b>4,44%</b>	<b>10,46%</b>
<b>ERROR PROMEDIO GLOBAL</b>	<b>7,05%</b>		

Tabla 26. Errores, variante 1, CASO 3, edificio teórico.

Hemos obtenido a nivel general de desglose una mejor aproximación con un error total del 7,05%, un mejor desglose de los consumos usuarios.

La variante 2 del caso tiene los mismos resultados en consumo de climatización y un leve decremento de los errores de la otra componente del consumo.

MES	ERROR CONSUMO USUARIO	ERROR CONSUMO CLIMA
Enero	1,28%	1,32%
Febrero	1,24%	2,05%
Marzo	0,70%	2,17%
Abril	0,40%	13,23%
Mayo	1,00%	12,07%
Junio	1,89%	10,59%
Julio	0,48%	4,60%
Agosto	2,17%	4,86%
Septiembre	0,71%	7,62%
Octubre	0,41%	47,62%
Noviembre	3,58%	15,41%
Diciembre	1,15%	3,96%
<b>MAXIMO ERROR</b>	<b>3,58%</b>	<b>47,62%</b>
<b>ERROR PROMEDIO</b>	<b>1,25%</b>	<b>10,46%</b>
<b>ERROR PROMEDIO DESGLOSE</b>	<b>5,85%</b>	

Tabla 27. Errores, variante 2, CASO 3, edificio teórico.

### 3.8 Conclusiones.

Se ha comprobado que la buena estimación es dependiente de la calidad de las variables de entrada del algoritmo y no de la cantidad.

Los mejores resultados en cuanto a homogeneidad de error obtenido ha sido con el caso 1, el más sencillo. Descartando el caso 3 en el que la radiación ha empeorado los resultados, el uso de los grados día equivalentes no han supuesto una mejoría notable, el planteamiento es análogo al caso 1 y solo hay que remarcar la aparición de algún caso puntual con un error notable en la estimación de la climatización.

Se ha observado que los errores puntuales excesivos, se dan en meses de la estación intermedia donde la climatización no trabaja. Este error nace en la aproximación de la temperatura equivalente y se ve fuertemente influenciado por la aproximación del consumo fijo, siendo más probable obtener este tipo de errores cuanto mayor sea la distancia del clúster utilizado para la aproximación a la temperatura equivalente real del sistema. Es por ello que en el caso 1 dónde más cercano es el clúster a la temperatura equivalente, mejores errores hemos obtenido.

Se concluye esta sección recomendando al usuario utilizar las variables necesarias y más representativas de la climatología y la ocupación que sea posible, realizando una aproximación con sumo cuidado en los primeros apartados de determinación de zona de no trabajo del sistema de climatización y de obtención del consumo fijo, o de usuario en el caso de no estar en condiciones de hacer el desglose en tres componentes.

### **3.9 Línea base de consumos.**

En las líneas que continúan se procederá a realizar la construcción de las herramientas necesarias para la estimación de consumos de las edificaciones estudiadas para un periodo de tiempo distinto del que se disponen datos.

La línea base permitirá al usuario estimar los consumos de la construcción en unas condiciones distintas a las que se disponen, de ahí su importancia para poder limitar las necesidades energéticas que tendrá un edificio en un futuro o para detectar anomalías de consumo en un momento determinado.

#### **3.9.1 Edificio de oficinas. Línea base de consumos.**

La línea base que se utilizará como se ha explicado anteriormente constará de tres consumos básicos, dos obtenidos de forma trivial siempre y cuando las condiciones de uso del edificio no hayan variado y el último, el consumo de climatización mediante la línea base basada en funciones de transferencia.

##### **Consumo fijo.**

El consumo fijo de la estimación es un valor constante obtenido en el proceso de clustering y desglose de consumos previo. Su valor engloba los equipos que se mantienen en funcionamiento sea la estación que sea y que no dependen fuertemente de la ocupación y uso. Será un valor numérico sumado a los consumos de ocupación y climatización sumados constante.

Este valor se podrá utilizar siempre y cuando el edificio no haya sufrido modificaciones en el número y tipo de equipos que instalan, como pueden ser ascensores, sistemas de emergencia, domótica, etc.

##### **Consumo de ocupación.**

La estimación de esta variable es trivial y está fuertemente unida a las variables de ocupación y uso definidas en el proceso de desglose. De tal manera que para un momento cualquiera en el que se pretendiera conocer el valor del consumo de ocupación, éste tomará el valor que se obtenía del desglose de consumo en función del tipo de día que se tenga que previamente el usuario deberá de conocer.

La estimación de este valor será posible siempre y cuando la situación de ocupación y los equipos del momento a estimar estén contemplados en el momento en el que se definió la variable tipo día y se calculó sus valores en el desglose.

#### **3.9.1.1 Consumo de climatización. Construcción de la línea base.**

Para la construcción de la línea base de consumos de climatización se utilizará los consumos de climatización obtenidos mediante el desglose. Esta línea base tendrá dos variantes en función del modo de operación del sistema, calefacción o refrigeración.

Tras la prueba de varias opciones de modelado de las líneas bases, dado el amplio catálogo de variables que se puede utilizar, se ha optado como ya se adelantaba por la variante de un denominador y dos numeradores. Siendo dos variables las que se usarán el consumo de climatización del día anterior (denominador) y los grados día equivalentes del día de cálculo y día anterior (numeradores).

Por ello para iniciar este proceso será necesario disponer de la temperatura equivalente y del cálculo de los grados día equivalentes, los cuales ya tenemos previamente del procedimiento de desglose.

Así el conjunto de datos de partida que se utilizará para la construcción de estas dos formulaciones son tales como muestra una partición de la tabla que sigue:

FECHA	CONSUMO CLIMA	CONSUMO CLIMA t-1	Gdeq	Gdeq (t-1)
02/01/2007	159,67 KWh	77,30 KWh	123,88	68,31
03/01/2007	199,48 KWh	159,67 KWh	104,98	123,88
04/01/2007	263,96 KWh	199,48 KWh	159,98	104,98
05/01/2007	322,96 KWh	263,96 KWh	178,88	159,98
06/01/2007	289,66 KWh	322,96 KWh	189,18	178,88
07/01/2007	0,00 KWh	289,66 KWh	184,28	189,18
08/01/2007	580,92 KWh	0,00 KWh	210,68	184,28
09/01/2007	562,29 KWh	580,92 KWh	228,78	210,68
10/01/2007	544,44 KWh	562,29 KWh	217,48	228,78
11/01/2007	475,27 KWh	544,44 KWh	195,38	217,48
12/01/2007	612,40 KWh	475,27 KWh	310,68	195,38
13/01/2007	465,59 KWh	612,40 KWh	261,88	310,68
14/01/2007	0,00 KWh	465,59 KWh	246,68	261,88
15/01/2007	889,26 KWh	0,00 KWh	285,48	246,68
16/01/2007	733,28 KWh	889,26 KWh	240,38	285,48

Tabla 28. Datos de partida, línea base, caso teórico.

Para el cálculo de cada una de las líneas bases se utilizará los datos de un mes de calefacción y otro de refrigeración respectivamente.

### Línea base de refrigeración.

Fecha de datos para cálculo: Agosto 2007

Utilizando Excel como software de cálculo los resultados obtenidos de los coeficientes de cada una de las variables que se introducen en la función de transferencia son:

a	b	c
0,935066732	0,207060222	0,360191281

Tabla 29. Coeficientes, línea base de refrigeración, caso teórico.

$$C_{clima,estimado}(t) = 0.935067 * GDE(t) + 0.207060 * GDE(t - 1) + 0.360191 * C_{clima}(t - 1)$$

El coeficiente de correlación obtenido es de:  $r = 0.991$

### Línea base de calefacción.

Fecha de datos para cálculo: Febrero 2007

a	b	c
1,169866141	1,18195377	-0,097010312

Tabla 30. Coeficientes línea base de calefacción, caso teórico.

$$C_{clima,estimado}(t) = 1.169866 * GDE(t) + 1.181954 * GDE(t - 1) - 0.09701 * C_{clima}(t - 1)$$

El coeficiente de correlación obtenido es de:  $r = 0.973$

### 3.9.1.2 Estimación. Utilización de la línea base.

Tras la confección de las dos líneas bases que representan los consumos, se hacen dos pruebas de las mismas para dos meses de calefacción o refrigeración.

Las dos gráficas muestran los resultados obtenidos de estas pruebas para cada una de las líneas, siendo las marcas puntuales los resultados calculados y la línea el valor real de consumo de la instalación.

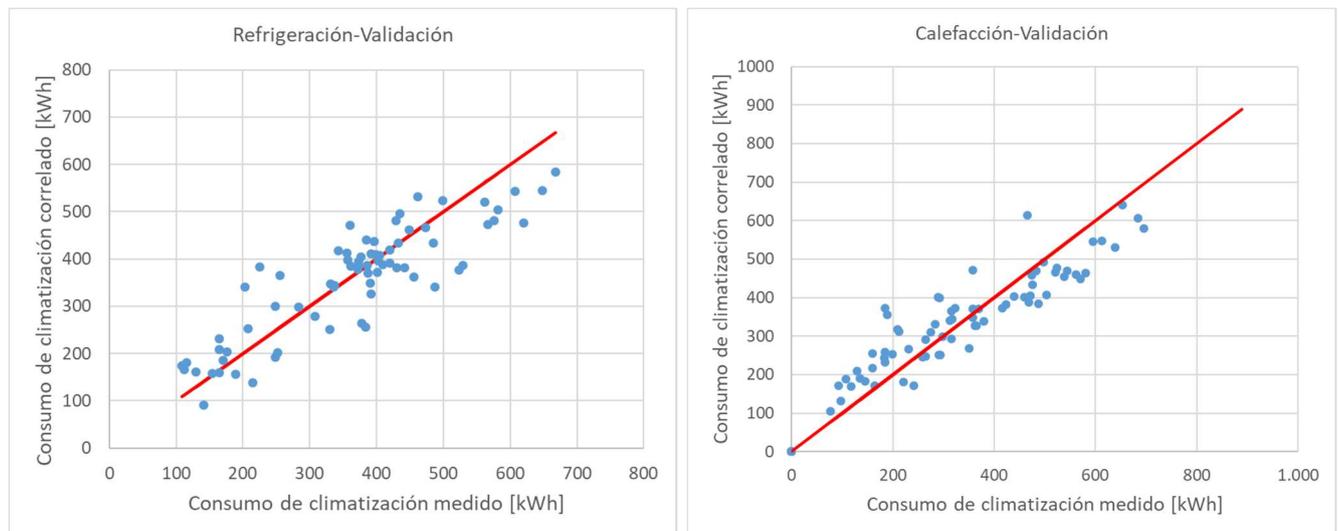


Figura 39. Líneas bases, validaciones, edificio teórico.

El error medio obtenido es del 17% y 23% respectivamente.

De esta manera el uso de la línea base se resume en el siguiente itinerario:

1. Comparación de la temperatura del día a estimar con los límites de temperatura equivalente  $\pm 2^{\circ}\text{C}$ . En el caso de estar dentro de este intervalo el consumo de climatización es 0 KWh dado que se encuentra en la zona de no climatización.
2. Elección de la línea base a utilizar. Comparando la temperatura media del día se obtiene si el día pertenece al modo de operación de refrigeración o calefacción. Temperaturas superiores a la temperatura equivalente serían días de refrigeración. En el caso de tener GDE su uso es equivalente observando que esté dentro del intervalo  $\pm 48$  y que su valor sea positivo o negativo para saber en que modo opera.
3. Cálculo del consumo de climatización en función de los datos de entrada del día de cálculo en la línea base que corresponda.

### 3.9.2 Estimación de Consumos totales. Resultados.

Tras la obtención de la línea base y de haber marcado el itinerario de utilización de la línea base, se procede a calcular los consumos totales de la edificación y compararlos con los resultados reales del desglose.

El sumatorio de los consumos estimados presenta un buen ajuste, los resultados presentados corresponden a los meses de enero, febrero, junio, julio, agosto y diciembre dado que son los más significativos en cuanto a consumos y es dónde el error de estimación sería más visible.

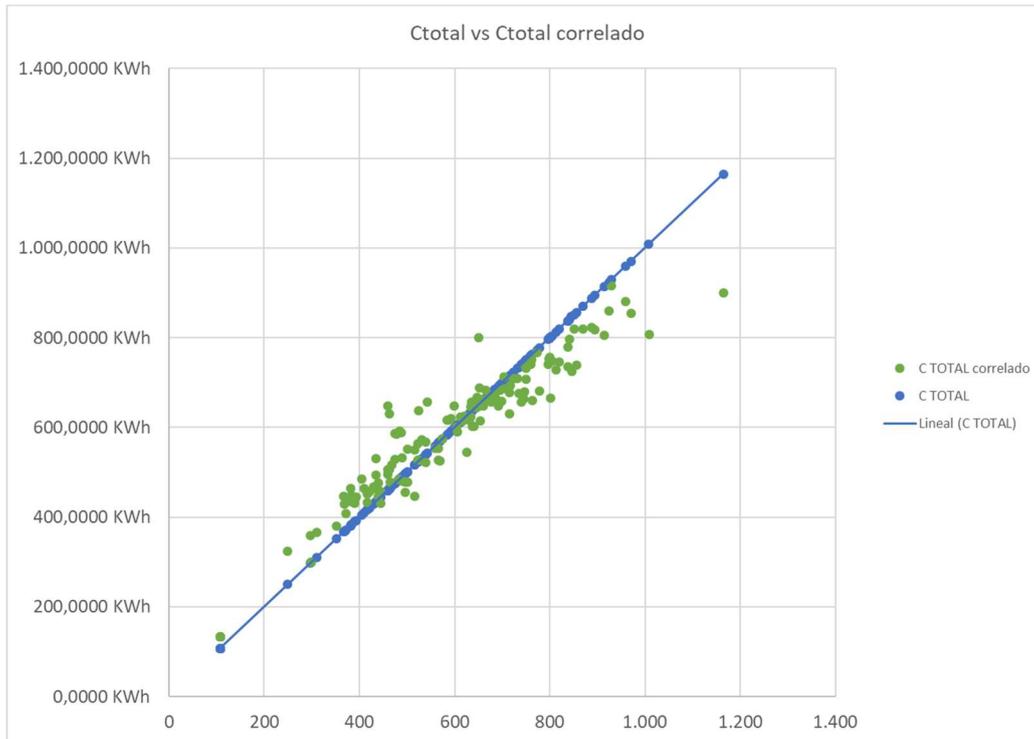


Figura 40. Ajuste del Consumo estimado de climatización.

Se muestra a continuación las líneas de error obtenidas para los tres consumos, así como para el consumo total.

Los errores del consumo total obtenidos se mantienen por debajo o cerca de la línea del 20%, lo cual es un buen resultado. Pero se observa errores puntuales en la estimación del consumo de climatización debido a que la línea base no distingue entre días sin ocupación u ocupados, y por ello en muchos casos estos días presentan consumos muy bajos o nulos que la estimación no ajusta.

Para conseguir esto sería interesante en estudios posteriores valorar incluir una variable de ocupación en la línea base, que refleje la ocupación o no del edificio y modifique así el consumo estimado.

# 4 CASO REAL.

## COLEGIO DE ARQUITECTOS DE CATALUÑA.

En el siguiente capítulo se expondrá el cálculo del desglose de consumos para un edificio real, el objetivo fundamental es tratar de explorar las cualidades del procedimiento desarrollado en los dos capítulos anteriores con las mediciones reales de un edificio terciario y ver la capacidad del mismo para dar soporte a cualquier usuario que lo precise.

### 4.1 Ubicación y arquitectura del edificio.

El edificio que se estudiará es la sede del Colegio de Arquitectos de Cataluña, una edificación renovada que se dedica a las actividades relacionadas con la arquitectura, entre ellas, la atención al personal colegiado para ayuda jurídica y emprendimiento, así como numerosas exposiciones, cursos y másteres.



Figura 41. Ubicación construcción caso real.

El Colegio de Arquitectos de Cataluña se sitúa en el centro de la ciudad de Barcelona, específicamente en la plaza Nova dentro del Barrio Gótico. Aunque es un edificio del año 1962 han sido varias las renovaciones que ha sufrido a lo largo del tiempo.



Figura 42. Fachada caso real. Colegio de Arquitectos de Cataluña.

Entre las dependencias del edificio encontramos:

- Oficinas
- Librería
- Sala de Exposiciones
- Biblioteca
- Videoteca
- Aseos

## 4.2 Análisis del consumo.

El edificio en cuestión utiliza dos fuentes primarias de energía: energía eléctrica de la red y energía proveniente de gas natural.

La energía del gas natural se utiliza únicamente para calefactar el edificio mediante caldera y el resto de los equipos desde los básicos propios de un edificio terciario como iluminación o sistemas de seguridad hasta climatización (refrigeración) utilizan la energía eléctrica para funcionar.

El uso de esta energía es heterogéneo a lo largo de la semana y del año. El horario de uso del edificio es de lunes a viernes de 8:00 a 21:00 en el período laboral que se alarga desde el 2 de agosto al 22 de septiembre. En el periodo vacacional el edificio permanece cerrado y el consumo es ínfimo y se supondrá nulo. Los fines de semana el edificio permanece cerrado, aunque el sábado se trabaja o se imparten clases en jornadas de mañana con una ocupación menor a un día entre semana.

El periodo de estudio que se realizará comienza el 22 de septiembre de 2016 y termina el 21 de septiembre de 2017. A inicios del año 2017 el cliente reporta que los sábados se comenzará a impartir clases de un curso muy frecuentado por lo que se detecta un aumento del consumo energético estos días a partir de esta fecha.

## 4.3 Climatología. Modos de operación.

No es posible saber de inicio como operará el sistema de acondicionamiento del aire, para ello se calculará la temperatura equivalente.

La temperatura equivalente se ha calculado haciendo uso de los consumos de los meses de junio y enero porque estos dos meses son candidatos fijos a los consumos de refrigeración y calefacción respectivamente. La razón de no haber utilizado otro mes más caluroso se debe a que las temperaturas en el mes de julio fueron muy constantes y la poca variación de la misma dificultaba la construcción de una recta de regresión válida para este mes. Igualmente se ha obviado días de la semana que puedan provocar error en el entendimiento como son los domingos donde el consumo de climatización es nulo dado que no hay ocupación del edificio.

El valor obtenido de este parámetro es  $17,63^{\circ}C$  y el intervalo de zona de no operación por convención y experiencia se marca en  $\pm 2^{\circ}C$  para la zona en la que se ubica la edificación.

Por ello cualquier día situado en el intervalo de temperaturas exteriores de  $[15.63, 19.63]^{\circ}C$  se supondrá perteneciente a la zona de no operación y por lo tanto no se estimará consumo de climatización.

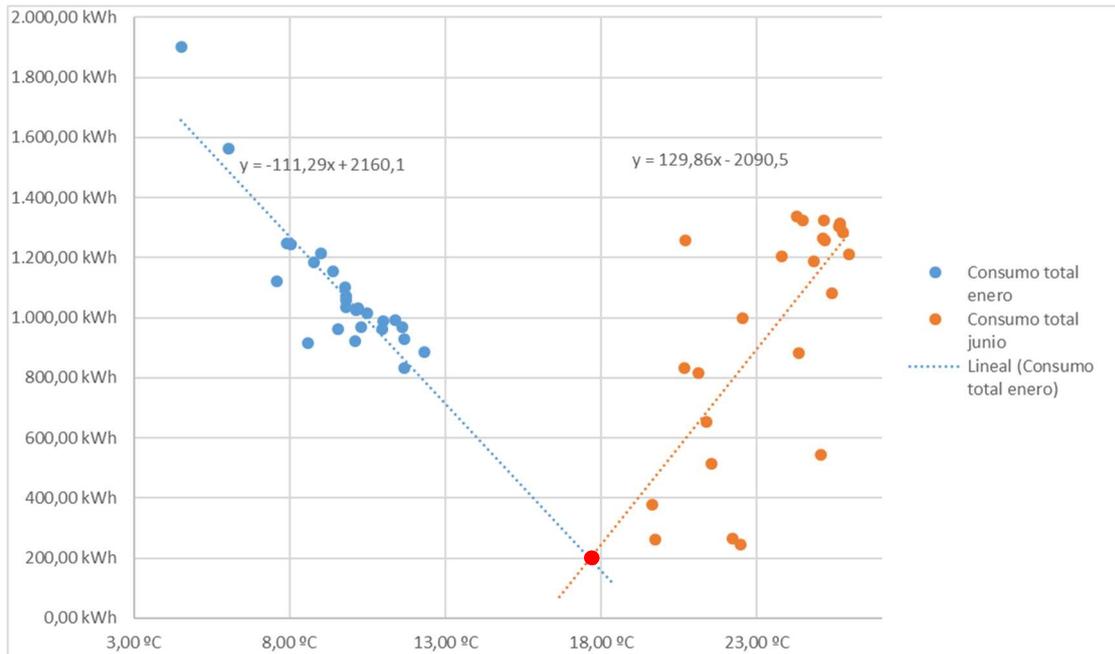


Figura 43. Cálculo temperatura equivalente, caso real.

#### 4.4 Variables de entrada al procedimiento.

Las variables de entrada de las que se dispone medición son las siguientes:

- Consumo total: suma de los consumos diarios de gas y electricidad medidos en kWh diario. Tras desglose podremos ver el consumo en gas que es el de calefacción y el consumo eléctrico del resto de equipos.
- Temperatura exterior media diaria: no se tiene medición de esta, pero se puede obtener fácilmente de base de datos como la de la NASA.
- Tipo de día: dado el carácter heterogéneo de ocupación y uso del edificio se decide usar este tipo de variable inventada, de tal forma que refleje el tipo de día al cual está expuesto el edificio. Esta variable tendrá 5 valores distintos, que son fruto de la observación de los consumos totales y análisis del uso anual que se le da al edificio.
  - Tipo día 1: valor correspondiente a consumos de ocupación de días laborales en los que el edificio se ocupa al 100%.
  - Tipo día 2: valor de los sábados en el año 2016. La ocupación es baja y solo en el turno de la mañana, pero es razonable que existe ocupación dado la diferencia de consumos entre sábados y domingos de la misma semana con temperaturas similares.
  - Tipo día 3: valor de los domingos, no hay ocupación y el edificio permanece cerrado, pero manteniendo las funciones básicas del mismo trabajando. Entre estas funciones están los depósitos de ACS en la temperatura nominal, sistemas de seguridad y servicios necesarios para el personal de mantenimiento.
  - Tipo día 4: valor de los sábados en el año 2017. Se distingue de los sábados del año anterior dado que el cliente informa de la celebración de un curso todos los sábados y se tendrá una ocupación notable a lo largo del turno de mañana, mayor que en los días tipo 2.
  - Tipo día 5: días de vacaciones, el edificio permanece cerrado y el gasto es muy bajo como para ser medido. Solo se mantienen los servicios básicos que consumen una parte ínfima de la media anual.

No son muchos los datos que se tienen de la edificación, pero dado que el apartado 3 se cerró con buenos resultados con el uso de estas tres variables, se procederá a realizar el desglose del caso real con los mismos tipos de variables.

## 4.5 Clustering.

Son numerosas las pruebas que se han hecho para llegar al número óptimo de clústeres que cumplan las restricciones de cálculo. La gráfica muestra el comportamiento normal que sigue cualquier proceso de clustering, llegando a un punto donde el decremento del índice de Davis-Bouldin no compensa el aumento del número de clústeres.

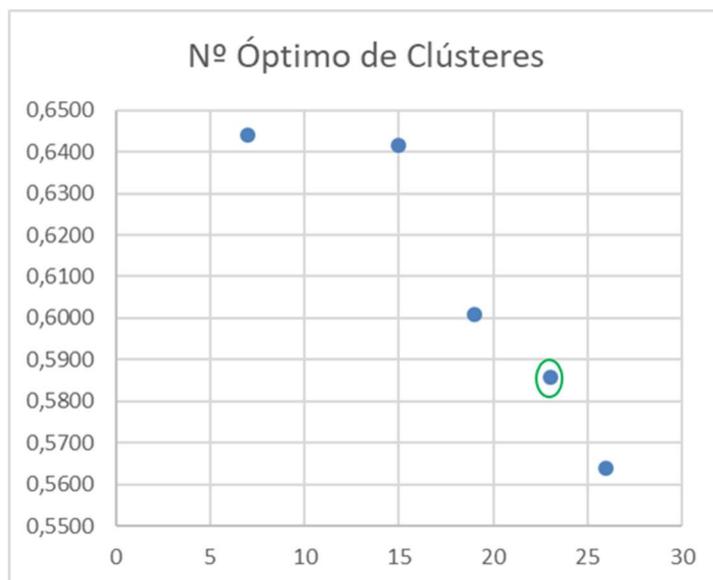


Figura 44. Definición del número óptimo de clústeres.

Prueba	Observaciones	Nº Óptimo de Clústeres	Valor Índice	Decremento
1		7	0,6440	
2	RESULTADOS EN TIPO DÍA MUY DECIMALES	15	0,6415	0,39%
3	NO HAY SUFICIENTE VARIEDAD EN ZONA INTERMEDIA	19	0,6008	6,78%
4		23	0,5859	2,54%
5		26	0,5639	3,90%

Tabla 31. Pruebas clustering, edificio real.

De una primera vista la prueba 3 mostraba unos resultados razonables, obteniéndose un buen resultado del índice y un alto decremento en comparación la opción anterior, pero esta prueba devuelve un agrupamiento demasiado compacto para el problema propuesto y no se obtiene clústeres dentro de la zona de no operación de los equipos HVAC para todos los tipos de días definidos.

El caso 4 es distinto y aunque no se ha conseguido un agrupamiento mucho mejor aumentando el número de clústeres, de 19 a 23 clústeres, si se obtiene variedad de los tipos de días en la zona de no climatización, requisito obligatorio para poder desarrollar el desglose de consumos.

## 4.6 Desglose de consumos.

Gracias a que tenemos días en los que el edificio permanece vacío y por lo tanto no hay consumo de ocupación, será posible realizar el desglose en los tres consumos definidos en vez de dos.

### 4.6.1 Consumo fijo.

El consumo fijo se obtiene directamente como se marca en el procedimiento, buscando los clústeres de tipo de día de no ocupación y dentro de la zona de no trabajo del sistema de climatización.

El consumo fijo obtenido es 199,9091 KWh y se usará para todos aquellos días en los que el consumo total sea mayor a este valor, en caso contrario el consumo fijo del día será el consumo total del mismo

ITEM	CLUSTER	CONSUMO	T EXT	TIPO DÍA	nº componentes
1	Cluster 1	247,89 KWh	17,56 °C	2	9
2	Cluster 2	1.514,54 KWh	9,11 °C	1	8
3	Cluster 3	167,87 KWh	10,55 °C	3	8
4	Cluster 4	0,00 KWh	25,67 °C	5	25
5	Cluster 5	664,27 KWh	21,08 °C	1	19
6	Cluster 6	1.072,55 KWh	24,48 °C	4	6
7	Cluster 7	966,19 KWh	13,09 °C	1	32
8	Cluster 8	1.039,81 KWh	12,23 °C	4	17
9	Cluster 9	343,13 KWh	19,62 °C	4	6
10	Cluster 10	1.069,64 KWh	10,00 °C	1	21
11	Cluster 11	0,00 KWh	22,54 °C	5	20
12	Cluster 12	876,73 KWh	15,98 °C	1	13
13	Cluster 13	1.280,67 KWh	24,79 °C	1	30
14	Cluster 14	499,38 KWh	13,73 °C	1	20
15	Cluster 15	566,43 KWh	25,06 °C	3	7
16	Cluster 16	162,11 KWh	12,97 °C	3	9
17	Cluster 17	920,78 KWh	22,95 °C	1	10
18	Cluster 18	0,00 KWh	19,06 °C	5	6
19	Cluster 19	416,01 KWh	17,88 °C	1	25
20	Cluster 20	212,27 KWh	15,37 °C	3	11
21	Cluster 21	1.110,06 KWh	20,23 °C	1	11
22	Cluster 22	199,91 KWh	19,60 °C	3	11
23	Cluster 23	1.263,30 KWh	13,05 °C	1	25

Tabla 32. Clústeres, caso real.

### 4.6.2 Consumo de ocupación.

El consumo de ocupación se obtiene de la misma manera que en el problema anterior. En este caso real se ha distinguido 5 tipos de días, de los cuales ya es sabido y por consiguiente que el consumo de ocupación en tipos de día 3 y 5 es nulo. El resto de los días toman los siguientes valores.

CLUSTER	CONSUMO	T EXT	RAD	TIPO DÍA	CONSUMO FIJO	CONSUMO OCUPACIÓN
Cluster 1	247,89 KWh	17,56 °C		2	199,91 KWh	47,98 KWh
Cluster 9	343,13 KWh	19,62 °C		4	199,91 KWh	143,22 KWh
Cluster 18	0,00 KWh	19,06 °C		5	199,91 KWh	0,00 KWh
Cluster 19	416,01 KWh	17,88 °C		1	199,91 KWh	216,10 KWh
Cluster 22	199,91 KWh	19,60 °C		3	199,91 KWh	0,00 KWh

Tabla 33. Cálculo consumo de ocupación, caso real.

### 4.6.3 Desglose. Estimación mensual.

Se puede observar que los consumos en septiembre de 2016 y agosto de 2017 son muy bajos, y eso se debe a que parte de los días del mes pertenecen a tipos de día 5, es decir vacacionales.

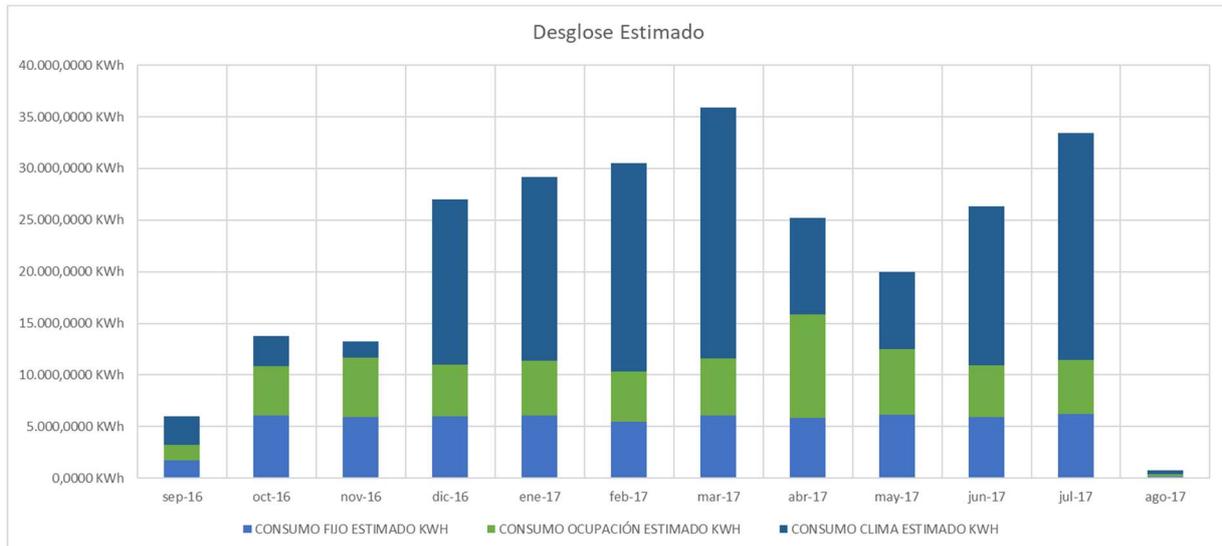


Figura 45. Desglose estimado de consumos, caso real.

Como observación al desglose, sería conveniente una revisión por parte del cliente o de la compañía suministradora del consumo excesivo en climatización del mes de marzo, para ser exactos calefacción por gas natural (temperaturas diarias menores a temperatura equivalente), ya que este consumo excesivo se debe a una causa aislada no reportada al usuario del procedimiento y no valorada a la hora de definir variables y aplicar los algoritmos.

## 4.7 Línea base de consumos.

A continuación, se realizarán los cálculos de construcción y posteriormente de uso de la línea base de consumos para el edificio real tratado.

### 4.7.1 Colegio de Arquitectos de Cataluña. Línea base de consumos.

Se seguirá la misma ruta que en el caso teórico y se realizarán los mismos cálculos con el fin de probar la validez de esta herramienta como método de cálculo de consumos.

#### 4.7.1.1 Consumo de climatización. Construcción de la línea base.

Para la construcción de la línea base de consumos de climatización se utilizará los consumos de climatización obtenidos mediante el desglose. Esta línea base tendrá dos variantes en función del modo de operación del sistema, calefacción o refrigeración.

Se seguirá la misma estructura que en la línea base de consumos del caso teórico y la línea base contendrá dos variables: el consumo de climatización del día anterior (denominador) y los grados día equivalentes del día de cálculo y día anterior (numeradores).

Será necesario tener previamente la temperatura equivalente y del cálculo de los grados día equivalentes, la primera ya calculada en el desglose de consumos y los últimos obtenidos mediante la fórmula descrita con anterioridad.

Para el cálculo de cada una de las líneas bases se utilizará los datos de un mes de calefacción y otro de refrigeración respectivamente.

### Línea base de refrigeración.

Fecha de datos para cálculo: Julio 2017

Los resultados obtenidos de los coeficientes de cada una de las variables que se introducen en la función de transferencia son:

a	b	c
-6,68492033	2,917924561	0,094005256

Tabla 34. Coeficientes línea base refrigeración, caso real.

$$C_{clima,estimado}(t) = -6.684920 * GDE(t) + 2.917925 * GDE(t - 1) + 0.094005 * C_{clima}(t - 1)$$

El coeficiente de correlación obtenido es de:  $r = 0.935$

### Línea base de calefacción.

Fecha de datos para cálculo: Enero 2017

a	b	c
3,929238856	-0,408750093	0,027340735

Tabla 35. Coeficientes, línea base calefacción, edificio real.

$$C_{clima,estimado}(t) = 3.929224 * GDE(t) - 0.408750 * GDE(t - 1) - 0.027341 * C_{clima}(t - 1)$$

El coeficiente de correlación obtenido es de:  $r = 0.980$

#### 4.7.1.2 Estimación. Utilización de la línea base.

La validación de las dos líneas bases arrojan las siguientes gráficas.

Las dos gráficas muestran los resultados obtenidos de estas pruebas para cada una de las líneas, siendo las marcas puntuales los resultados calculados y la línea el valor real de consumo de la instalación.

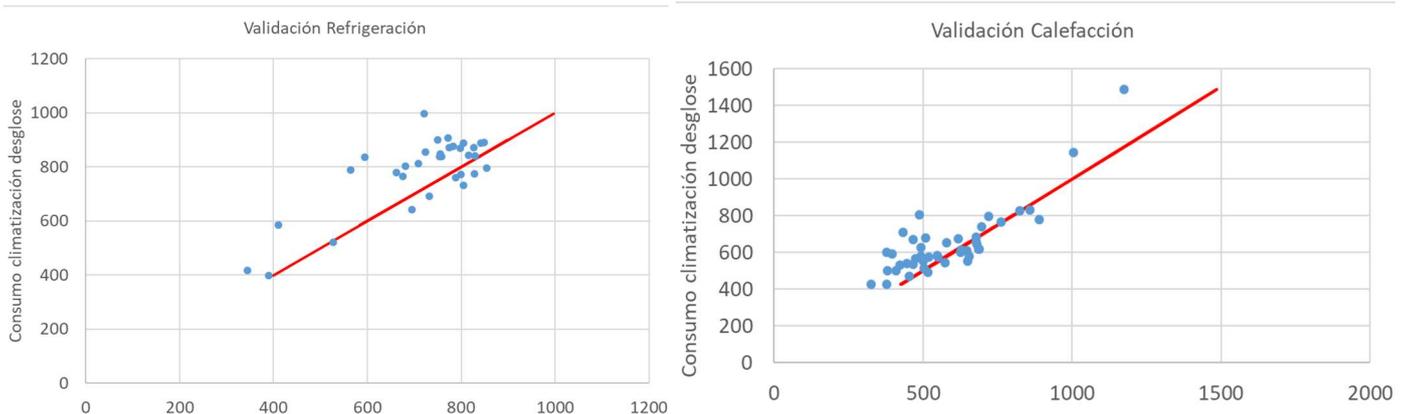


Figura 46. Validaciones líneas bases, caso real.

El error medio obtenido es del 11% y 13% respectivamente.

### 4.7.2 Estimación de Consumos totales. Resultados.

Tras la obtención de la línea base, los consumos totales estimados quedan de la siguiente forma.

Los consumos estimados se grafican a continuación, los resultados presentados corresponden a los meses de junio, julio, diciembre y febrero dado que son los más representativos.

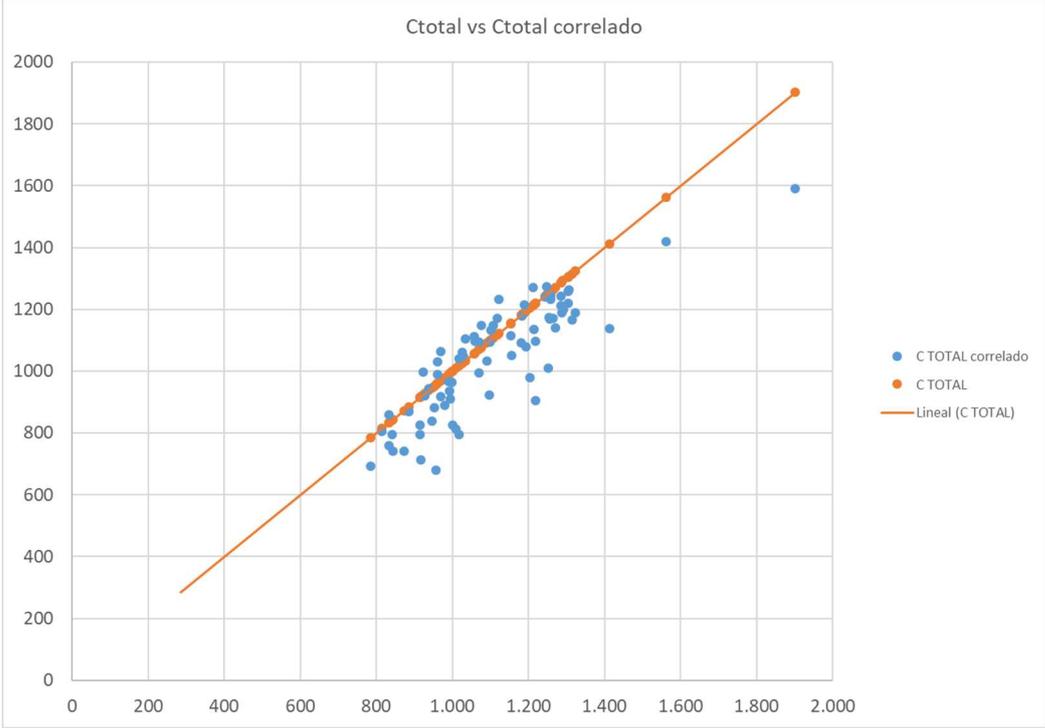


Figura 47. Ajuste de consumos línea base caso real.

Se muestra a continuación las líneas de error obtenidas para los tres consumos, así como para el consumo total.

Los errores del consumo total obtenidos se mantienen por debajo o cerca de la línea del 20%, lo cual es un buen resultado y se da por aceptable el ajuste obtenido con la estimación realizada.



# REFERENCIAS

---

- [1] BP. Statistical Review, 2018.
- [2] B. Mita, S. Rafiq, H. H. Lean, Sankar B. The regulated coal sector and CO2 emissions in Indian growth process: Empirical evidence over half a century and policy suggestions. 2017.
- [3] Directiva 2012/27/UE del Parlamento Europeo y del Consejo.
- [4] G. Q. Chen, X. F. Wu. Energy overview for globalized world economy: source, supply chain and sink. *Renew Sustain Energy Rev* 2017.
- [5] M. Hamdy, S. Carlucci, P. Hoes, J. L. M. Hensen. The impact of climate change on the overheating risk in dwellings. Dutch case study. 2017
- [6] M. De Rosa, V. Bianco, F. Scarpa, L. A. Tagliafico. Heating and cooling building energy demand evaluation; a simplified model and a modified degree days approach. 2014.
- [7] Directiva (UE) 2018/844 del Parlamento Europeo y del Consejo.
- [8] M&V Guidelines: Measurement and Verification for Performance-Based Contracts. Version 4.0. 2015
- [9] Measurement of energy and demand savings ASHARE Guidelines 14-2002. 2002.
- [10] J. Granderson, S. Touzani, C. Custodio, M. D. Sohn, D. Jump, S. Fernandes. Accuracy of automated measurement and verification (M&V) techniques for energy savings in commercial buildings. 2016
- [11] Y. Heoa, V. M. Zavalab. Gaussian process modeling for measurement and verification of building energy savings. 2012.
- [12] C. V. Gallagher, K. Leahy, P. O'Donovan, K. Bruton, D. T.J. O'Sullivan. Development and application of a machine learning supported methodology for measurement and verification (M&V) 2.0. 2018.
- [13] K.P. Amber, R. Ahmad, M.W. Aslam, A. Kousar, M. Usman, M.S. Khan. Intelligent techniques for forecasting electricity consumption of buildings. 2018.
- [14] X. Liang, T. Hong, G. Q. Shen. Improving the accuracy of energy baseline models for commercial buildings with occupancy data. 2016.
- [15] A. Rahmana, V. Srikumarb, A. D. Smitha. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. 2018.
- [16] C. V. Gallagher, K. Bruton, K. Leahy, D. T.J. O'Sullivan. The suitability of machine learning to minimise uncertainty in the measurement and verification of energy savings. 2018.
- [17] Z. Olinga, X. Xia, X. Ye. A cost-effective approach to handle measurement and verification uncertainties of energy savings. 2017.
- [18] H. Carstens, X. Xia, S. Yadavalli. Low-cost energy meter calibration method for measurement and verification. 2017.
- [19] J. Jeong, T. Hong, C. Ji, J. Kim, M. Lee, K. Jeong, C. Koo. Development of a prediction model for the cost saving potentials in implementing the building energy efficiency rating certification. 2017.
- [20] D. Mikulić, I. Bakarić, S. Slijepčević. The economic impact of energy saving retrofits of residential and public buildings in Croatia. 2016.
- [21] V. Shabunko, C.M. Lim, S. Mathew. EnergyPlus models for the benchmarking of residential buildings in Brunei Darussalam. 2018.

- [22] M. Khoshbakht, Z. Gou, K. Dupre. Energy use characteristics and benchmarking for higher education buildings. 2018.
- [23] H. Li, X. Li. Benchmarking energy performance for cooling in large commercial buildings. 2018
- [24] A. Srivastava, A. Tewari, B. Dong. Baseline building energy modeling and localized uncertainty quantification using Gaussian mixture models. 2013.
- [25] B. Dong, C. Cao, S. E. Lee. Applying support vector machines to predict building energy consumption in tropical región. 2005.
- [26] Q. Li, Q. Meng, J. Cai, H. Yoshino, A. Mochida. Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks. 2009.
- [27] C. Deb, L. S. Eang, J. Yang, M. Santamouris. Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks. 2016.
- [28] J. H. Ko, D. S. Kong, J. H. Huh. Baseline building energy modeling of cluster inverse model by using daily energy consumption in office buildings. 2017.
- [29] J. Sánchez Ramos, “Metodología Aplicada de Caracterización Térmica Inversa para Edificios,” 2015.

