



Universidade de Vigo

Trabajo de Fin de Máster

**APLICACIONES ESTADÍSTICAS AL ESTUDIO
DE LA EFICIENCIA ENERGÉTICA Y LA
CALIDAD DEL AIRE INTERIOR EN VIVIENDAS**

Irene Pastoriza Vila

Directores:

Salvador Naya Fernández

Javier Tarrío Saavedra

Máster en Técnicas Estadísticas

Vigo, 1 de julio de 2014

Departamento de Matemáticas



Escuela Politécnica Superior
C/ Mendizábal, s/n
15403 Ferrol.
Universidade da Coruña

Salvador Naya Fernández y Javier Tarrío Saavedra, profesores de Estadística del Departamento Matemáticas de la Universidad de A Coruña.

Autorizan la presentación del siguiente proyecto realizado bajo mi dirección y supervisión:

Trabajo Fin de Máster del Máster en Técnicas Estadísticas titulado “*Aplicaciones estadísticas al estudio de la eficiencia energética y la calidad del aire interior en viviendas*”, elaborado por Irene **Pastoriza Vila**.

Santiago de Compostela, 30 de junio de 2014



Fdo.: Salvador Naya Fernández y Javier Tarrío Saavedra

Agradecimientos

Tras la realización del presente trabajo de Fin de Máster, me gustaría reflejar mi sincero agradecimiento a mis directores, Salvador Naya y Javier Tarrío, por toda la ayuda recibida durante estos meses.

También me gustaría hacer una mención especial a mi familia y amigos.

Muchas gracias a todos.

Resumen

El presente trabajo tiene por objetivo el estudio de problemas de eficiencia energética desde una perspectiva estadística. Concretamente, se tratarán de analizar variables de gran relevancia para la calidad del aire interior (niveles de CO_2) y del confort térmico (temperatura y humedad relativa), así como las variables relacionadas con la energía consumida (potencia activa). En primer lugar, se tratará de evaluar estadísticamente la capacidad de las viviendas e instalaciones comerciales en relación a la calidad del aire interior y al confort del ambiente térmico. Por otro lado, se propone la construcción de gráficos de control adecuados para la temperatura y el contenido en CO_2 . Mediante el uso de estos gráficos se podrán identificar posibles problemas de los sistemas de control con sus causas asignables, para tomar medidas correctoras en cuanto a la eficiencia energética y el confort.

Para aplicar estas herramientas, se dispone de datos de temperatura, consumo energético, iluminación y calidad del aire proporcionados por la empresa Fridama, correspondientes a sus propias instalaciones en A Coruña. Asimismo, se han tomado datos en diversos edificios públicos de Mondoñedo, con los que se estudiará la capacidad de los sistemas de control para el confort térmico.

Además, se utilizarán herramientas básicas del control estadístico de la calidad, con el objetivo de definir correctamente los problemas y analizar y mejorar los niveles de eficiencia energética. En concreto, se aplicarán herramientas de análisis descriptivo de los datos e inferencia, de modo que se identifiquen las variables más importantes en eficiencia y las relaciones entre ellas, índices de capacidad robustos y gráficos de control adecuados.

Índice general

1. Introducción	11
2. Consideraciones básicas	13
2.1. Control Estadístico de Calidad	13
2.2. Qué es la eficiencia energética	16
2.2.1. Eficiencia energética en la edificación	16
2.2.2. Auditorías Energéticas	18
2.3. Normativa en España y la Unión Europea	18
2.4. Calidad del Aire Interior	20
2.5. Confort térmico	21
3. Casos prácticos. Descripción	25
3.1. Aplicación a datos reales	25
3.1.1. Evaluación de la eficiencia energética en edificaciones de uso público en Mondoñedo.	25
3.1.2. Control de la eficiencia energética y confort térmico en viviendas y oficinas: base de datos tomada en Fridama Instalaciones S.L.	29
4. Métodos estadísticos	33
4.1. Control Estadístico de Calidad	33
4.1.1. Gráficos de Control por variables	33
4.1.2. Gráficos de control EWMA	39

4.1.3. Índices de capacidad	42
4.2. Series de Tiempo	46
4.3. Análisis de Componentes Principales	52
4.4. Análisis de Datos Funcionales	54
5. Resultados: Mondoñedo	57
5.1. Análisis Exploratorio de los Datos	57
5.1.1. Análisis de Componentes Principales	61
5.2. Índices de Capacidad	67
5.2.1. Edificio del Ayuntamiento	68
5.2.2. Edificio de la Casa de Cultura	75
5.2.3. Edificio de la Escuela de Música	81
5.3. Posibles acciones correctoras	84
6. Resultados: Fridama	85
6.1. Análisis Exploratorio de Datos	85
6.1.1. Estudio descriptivo mediante el análisis de datos funcionales	85
6.2. Control de Calidad y gráficos de control con datos autocorrelacionados	92
6.2.1. Análisis de la variable temperatura ambiente en la oficina CL2	92
6.2.2. Análisis de la variable CO_2 en la oficina CL2	107
7. Conclusiones	113
7.1. Líneas futuras	114
A. Código R	115
Bibliografía	129

Capítulo 1

Introducción

Hoy en día, es perceptible el aumento del interés por los asuntos relacionados con el ahorro y el uso eficiente de la energía. Además, la preocupación por las políticas ambientales y energéticas son cada vez mayores. De ahí que surja la necesidad de tratar el tema de la eficiencia energética, como un ámbito en auge y en boga tanto de gestores y administradores, como políticos o ingenieros.

Todo ello está asimismo relacionado con el concepto de calidad y del control y mejora continuada de la misma, convirtiéndose en un elemento diferenciador para todo tipo de empresas y para garantizar la satisfacción de los clientes o usuarios de servicios.

Además, el control estadístico de calidad se convierte en una herramienta fundamental para llevar a cabo los análisis de calidad y de los procesos mediante el uso de técnicas estadísticas desde el enfoque del control de calidad. El uso de algunas de estas técnicas se abordarán en este trabajo, con el objetivo de descubrir los posibles problemas detrás un sistema de control para que puedan tomarse las acciones correctivas pertinentes. En particular, se analizarán procesos de control relacionados con la eficiencia energética, el confort térmico de los ambientes y la calidad del aire interior de instalaciones comerciales, aspectos cruciales para aumentar la satisfacción de los usuarios de estos servicios.

En el Capítulo 1 se desarrollan e identifican los elementos básicos alrededor de los cuales girará el posterior trabajo, como el estudio del control estadístico de calidad y de la eficiencia energética. Además, se señalan los límites de especificación que deberán ser cumplidos por los sistemas de control para garantizar unos niveles aceptables en la calidad del aire interior y en el confort térmico.

En el Capítulo 2 se describen en detalle las bases de datos que se utilizarán para los posteriores análisis, referentes a tres edificios administrativos de Mondoñedo (el ayuntamiento, la casa de cultura y la escuela de música) y a los datos medidos en las instalaciones de la empresa

Fridama.

Por otro lado, en el Capítulo 3 se introducen las técnicas estadísticas que se van a emplear para el análisis de los datos anteriormente mencionados, pasando por la aplicación de técnicas concretas del control de calidad (los gráficos de control y los índices de capacidad) a otras herramientas complementarias como las series de tiempo o herramientas exploratorias como el análisis de datos funcionales y el análisis de componentes principales.

En el Capítulo 4 se desarrollarán algunas de las técnicas explicadas en el Capítulo 3 para el análisis de los edificios administrativos de Mondoñedo (diferenciando los meses de verano de los de invierno), midiendo la capacidad de los sistemas de control mediante los índices de capacidad.

En el Capítulo 5 se aplican los modelos de series temporales y el uso de los gráficos de control en la instalaciones de Fridama. También se utiliza el análisis de datos funcionales como herramienta de análisis exploratorio.

Por último, en el Capítulo 6 se detallan las conclusiones finales y posibles líneas futuras de trabajo.

Capítulo 2

Consideraciones básicas

2.1. Control Estadístico de Calidad

En primer lugar parece conveniente explicar qué se entiende por **calidad**, ya que se remitirá a ella a lo largo de todo este trabajo. No es sencillo reunir en una sola definición un concepto tan amplio como este y dependiendo de la época o de la fuente se encontrarán diferentes definiciones.

Lo que si es claro es que la mayor parte de las empresas consideran la calidad de un producto o de un servicio como un elemento competitivo y diferenciador con respecto a otras empresas del sector. Además, la importancia de la misma se ha visto intensificada como consecuencia de la creación de una normativa internacional en materia de calidad (las normas ISO, Normas de la Organización Internacional de Normalización, en concreto las conocidas como normas ISO 9000 sobre calidad y gestión de calidad).

Estas normas ISO definen la calidad como el conjunto de características de un producto o servicio que le confieren la aptitud para satisfacer las necesidad del cliente. Por su parte, el control de calidad es un conjunto de técnicas cuyo objetivo es asegurar que se mantiene la calidad de un producto o de un servicio y se reducen y eliminan posibles errores.

Más concretamente, el control **estadístico de la calidad** es el conjunto de herramientas y técnicas estadísticas utilizadas para medir la calidad y compararla con las especificaciones dadas por el cliente, con el fin de llevar a cabo acciones correctivas para su mejora (normalmente dirigidas a reducir la variabilidad).

Uno de los primeros investigadores en aplicar métodos estadísticos en al ámbito del control de calidad fue Walter Shewhart, trabajador de Bell Telephone Laboratories (compañía que, sobre los años 1920, y junto con la estadounidense Western Electric, fueron pioneros en el control de calidad). Shewhart introduce los gráficos de control en su publicación “Economic Control of Quality of Manufactured Product”, en 1931 (ver Shewhart, 1997), que se centra en

el establecimiento de estándares de calidad de forma cuantitativa, aunque reconoce que en la calidad hay un aspecto más subjetivo que tiene que ver con las expectativas de las personas.

Garvin (1987) distingue ocho dimensiones de calidad, que se pueden resumir como sigue:

1. **Desempeño.** Se refiere básicamente a si el producto cumple su funcionamiento ya que los consumidores evaluarán si dicho producto cumple las funciones que se esperan del mismo.
2. **Características.** Referido a las características que se añaden al funcionamiento básico de un producto haciendo mejorar su atractivo para el usuario.
3. **Fiabilidad.** Es la probabilidad de que un producto fallará dentro de un determinado periodo de tiempo.
4. **Conformidad.** El grado en que el producto cumple con los estándares marcados.
5. **Durabilidad.** Se refiere a la duración de vida del producto, bien sea técnicamente (el uso del producto hace que tenga un desgaste imposible de reparar), bien sea económicamente (el precio esperado de la reparación supera al valor actual del producto).
6. **Utilidad.** La rapidez con la que el producto se puede reparar así como el trato y la competencia del personal.
7. **Estética.** Es una de las características más subjetivas y se refiere principalmente a las preferencias individuales de cada usuario.
8. **Calidad percibida.** Referido a la percepción que tienen los usuarios, ya que pueden no tener una información completa sobre el producto. La propia reputación de la empresa, el nombre del producto o servicio o experiencias del pasado pueden condicionar la opinión de los consumidores.

Además, la calidad no está solamente centrada en la fabricación de productos que cumplan los límites especificados sino también en la mejora continua, con la finalidad de reducir la variabilidad en torno al objetivo establecido. Montgomery (2009) afirma que la calidad es inversamente proporcional a la variabilidad, introduciendo de este modo una de las características más relevantes en el control de calidad, esto es, el estudio de la variabilidad, con el fin de estimarla e intentar reducirla. Cuanto mayor es la variación respecto una o más características de calidad más disminuye la calidad del producto o servicio.

La variabilidad puede ser de diferente naturaleza. Las causas de variabilidad **no asignables** (también llamadas aleatorias o comunes) son aquellas inherentes al proceso productivo, mientras que las causas **asignables** (o especiales) son las que producen un gran aumento en la variabilidad y no son parte del propio proceso.

Teniendo en cuenta esto, se dice que un **proceso está bajo control estadístico** cuando no se encuentran causas asignables en el mismo, de manera que la variabilidad solo deberá a las causas comunes.

Por tanto el objetivo es mantener el proceso bajo control y dentro de unos límites y es aquí donde entra el **control estadístico de procesos (Statistical Process Control, SPC)**, que se aplica principalmente para monitorizar y controlar el proceso, asegurando que se produce el mínimo de no conformes (fuera de control) posibles.

Para conseguir tal objetivo, las herramientas más utilizadas dentro del SPC son los gráficos de control, introducidos por primera vez, como ya se ha señalado, por Shewhart en 1931. Se hablará más en profundidad sobre este tema en el Capítulo 3.

Por último, señalar el surgimiento actual de la **metodología Seis Sigma** (más conocida por su denominación en inglés, Six Sigma), como una herramienta para la gestión de calidad. Nace de la mano de la empresa Motorola hacia finales de los años 80, con el objetivo de reducir la variación de las características clave de calidad. De ahí su nombre, utilizando la letra griega σ , que hace referencia a la desviación típica, dando importancia así a la variabilidad como elemento esencial de esta metodología.

Uno de los propósitos del Seis Sigma es la reducción de la probabilidad de que ocurran defectos. Cuando el proceso alcanza el nivel de calidad seis sigma (mide la distancia entre la media y los límites de especificación, inferior y superior) solo se producirán en torno a 3.4 defectos por millón de oportunidades (DPMO), ya que la media de dicho proceso está sujeta a perturbaciones que pueden causar su desplazamiento en 1.5 desviaciones estándar del target (Montgomery, 2009).

Según esta metodología un proceso estará en control estadístico (cumplirá las especificaciones), con una probabilidad del 0.9973, cuando tenga una desviación entre -3σ y $+3\sigma$, de manera que solo se producirán 2700 defectos por millón de oportunidades (DPMO). Este nivel, sin embargo, no es aceptable en muchos procesos, ya que puede significar un elevado número de productos defectuosos.

La metodología Seis Sigma utiliza un enfoque basado en cinco etapas con la finalidad de mejorar los procesos (definir, medir, analizar, mejorar y controlar), capacitando así a las empresas a diseñar productos y procesos que satisfagan las expectativas de los consumidores y que puedan ser producidas con los niveles de calidad seis sigma.

Estas herramientas se extendieron rápidamente a otras industrias y a diferentes áreas dentro de una empresa, como marketing, ingeniería, administración, etc, conformando una filosofía de gestión global de la calidad.

2.2. Qué es la eficiencia energética

Como definición básica de la **eficiencia energética** se puede decir que se refiere a la utilización de menos energía para producir la misma cantidad de servicios o de energía útil.

Se pone de manifiesto de esta manera la relación entre el resultado de un proceso -bien sea la producción de bienes o servicios, bien sea la transformación de la energía-, y la energía que se ha empleado para realizar dicho proceso. Por tanto, el objetivo no es solamente la reducción del consumo energético, sino la mejora de todo el proceso.

Por otra parte, la eficiencia energética no solo se debe relacionar con aspectos tecnológicos y la mejora de los mismos, sino también con la mejora en la gestión o con cambios en el comportamiento o hábitos de los individuos.

Además, está estrechamente vinculada con el ámbito económico y, por ello, se suele relacionar con el valor monetario o el valor añadido del producto. Aquí es donde también entra el aspecto medioambiental, ya que se reduce la demanda de los recursos y servicios energéticos al aumentar la eficiencia, y esto también supone una reducción del impacto para el medio ambiente y una disminución de la emisión de CO₂.

2.2.1. Eficiencia energética en la edificación

Una gran parte del consumo final de energía puede atribuirse a oficinas, viviendas, tiendas y otros edificios tanto del sector privado como público, de ahí la importancia de las políticas de eficiencia energética para promover el uso racional de la energía a través de medidas pasivas y activas, que se detallan a continuación.

Medidas Pasivas

Estas medidas comprenden las acciones vinculadas con la renovación del aire, de las infiltraciones o con la envolvente térmica del edificio.

Lo primero a tener en cuenta son los elementos de la propia situación del edificio, como la zona geográfica en la que se encuentre (ya que va a suponer que haya diferentes temperaturas, nivel de precipitaciones así como de viento, etc), los elementos de su entorno (zonas de ruido, falta de soleamiento, etc), la orientación del edificio, la vegetación y los años de construcción del mismo. Otro elemento muy importante a tener en cuenta son las características arquitectónicas, como la envolvente térmica, la capacidad del edificio, posibles huecos en la fachada o en la cubierta, así como la existencia de invernaderos.

A continuación se enumeran las principales medidas pasivas orientadas a conseguir el ahorro de energía en los edificios:

1. Aumentar el aislamiento de las fachadas, ya sea de manera interior como exterior.
2. Aumentar asimismo el aislamiento de la cubierta.
3. Evitar los puentes térmicos, aislamiento de los huecos.
4. Aislamiento en suelos, para evitar pérdidas energéticas.
5. Inclusión de protectores solares, tales como persianas, cortinas, parasoles, toldos, entre otros. Estos elementos impedirán el paso del sol en verano, evitando el calentamiento del edificio, a la vez que posibilitarán la entrada del mismo en invierno.
6. Evitar la infiltración de aire exterior por rendijas o aberturas no intencionadas.

Medidas Activas

Por su parte, las medidas activas, son las que tienen incidencia sobre los sistemas de producción de calor o de frío de los edificios. Por ello, en primer lugar, es relevante conocer los elementos característicos de dichos sistemas en cada edificio (como por ejemplo la potencia nominal, el rendimiento nominal, el grado de sobredimensionado del sistema y rendimiento medio estacional del mismo).

Entre las medidas de este tipo se pueden destacar las siguientes:

1. Sustitución de la caldera por un sistema de mayor rendimiento. Esto es, si se han detectado problemas de rendimiento nominal, lo más adecuado será sustituir la caldera para aumentar dicho rendimiento.
2. Sustitución de los equipos autónomos de climatización si se encuentran problemas de bajo rendimiento nominal en los mismos.
3. Aislamiento de las redes hidráulicas, para evitar pérdidas de calor en la red de agua caliente o la calefacción.
4. Fraccionamiento de la potencia, con el objetivo de aumentar el rendimiento medio estacional, dependiendo de la demanda de calor o de frío existente.
5. Empleo del aire exterior, ya que proporciona refrigeración gratuita, siempre que las condiciones sean las adecuadas.
6. Perfeccionamiento de las redes de transporte, sustituyendo las bombas de agua y reduciendo los defectos de la red hidráulica.
7. Utilización de paneles solares, aprovechando así la energía solar.

2.2.2. Auditorías Energéticas

La **auditoría energética** es una herramienta que ayuda a obtener un correcto conocimiento de los consumos de energía en una determinada instalación. Así, se podrán considerar las posibilidades de ahorro energético, tanto desde el punto de vista económico como técnico. De este modo, se producen mejoras en la calidad de los servicios prestados, así como mejoras de carácter medioambiental y económico.

Este tipo de auditorías sirven para evaluar el uso eficiente de la energía de un edificio, buscando una clara reducción de los costes energéticos. Para ello se necesita de la existencia de un Programa de Ahorro de Energía y sobre todo, del análisis de diferentes variables vinculadas a la energía.

2.3. Normativa en España y la Unión Europea

En el ámbito de la edificación en España, la normativa energética tiene como precedente el Real Decreto 2429/79 de la Norma Básica de la Edificación NBE-CT-79, sobre Condiciones Térmicas en los edificios, actualmente derogada. En ella se definen las condiciones térmicas de los edificios, mediante la creación del coeficiente K_G y estableciendo un nivel máximo para dicho coeficiente (en función de la zona climática) que no se debía superar.

La primera versión del Reglamento de Instalaciones Térmicas de los Edificios o RITE (Real Decreto 1751/1998) fue derogada con la entrada en vigor del nuevo RITE (Real Decreto 1027/2007), en el que se establecen y se regulan las condiciones que deben cumplir las instalaciones térmicas (calefacción, agua caliente sanitaria y climatización) en los edificios destinados a atender las exigencias de bienestar térmico e higiene de las personas.

La Ley 38/1999, de Ordenación de la Edificación (LOE), establece unos requisitos básicos para la edificación en España. Crea las bases para la unificación de criterios y normativas sobre la construcción, definiendo unos criterios mínimos de seguridad, habitabilidad, bienestar y protección del medio ambiente.

A raíz de su entrada en vigor se aprueba también el Código Técnico de Edificación (Real Decreto 314/2006), formado por un conjunto de documentos normativos que aplican unos requisitos a cumplir por los edificios, para la mejora de su calidad y para el cumplimiento de los requisitos de seguridad y habitabilidad establecidos en la LOE .

El Código consta de dos partes, incluyéndose en la primera las disposiciones generales y las exigencias a cumplir por los edificios. La segunda parte está formada por los Documentos Básicos (DB), cuya utilización asegura la consecución de los objetivos y exigencias básicas. Son un total de 6 DB que recogen diferentes aspectos: seguridad estructural, seguridad en caso de incendio, seguridad de utilización y accesibilidad, protección frente al ruido, salubridad y

ahorro de energía. Los que fundamentalmente interesan aquí son los últimos. Por una parte, el referente a la salubridad recoge el tema de la calidad de aire interior, que afirma que los edificios deben disponer de medios para ventilar los recintos y eliminar así cualquier tipo de sustancia contaminante. Por otra parte, el documento referido al ahorro de energía dispone de una serie de medidas para conseguir el uso racional de la energía, reduciendo el consumo y mediante la utilización de energías renovables. Se recogen una serie de pautas a seguir y se establecen valores límites que no se deben sobrepasar, siempre buscando el bienestar térmico de los ocupantes y teniendo en cuenta la zona climática en cuestión.

También cabe destacar el Real Decreto 235/2013, por el que se aprueba el procedimiento básico para la certificación de la eficiencia energética de los edificios, por la que los compradores o usuarios de un edificio podrán disponer de un certificado de eficiencia energética que informe sobre la situación del edificio en ese ámbito, y que debe contar con los requisitos mínimos que establece el Código Técnico de la Edificación.

En el ámbito de la Unión Europea cabe destacar como la Directiva 93/76/CEE (SAVE), que obligaba a los Estados miembros de la Unión Europea a tomar medidas referentes al consumo energético de los edificios de uso residencial y a limitar las emisiones de CO_2 .

También destaca la Directiva 2002/91/CE de rendimiento energético de los edificios, por la que se obliga a los Estados miembros a la puesta en marcha de unos requisitos mínimos para asegurar el rendimiento energético, tanto de edificios de nueva construcción como ya existentes, la imposición de un control periódico de los sistemas de climatización y calderas y la expedición de certificaciones energéticas para los edificios. La nueva RITE se aprueba motivada en parte por esta norma.

La Directiva 2002/91/CE se deroga con la actual Directiva 2010/31/CE relativa a la eficiencia energética de los edificios. Esta última viene promovida en gran medida por el actual interés en la reducción del consumo energético como parte de los objetivos de la iniciativa “20-20-20”. Con esta iniciativa se pretende, para 2020, reducir el 20 % del consumo de energía primaria de la UE, el 20 % de las emisiones de gases efecto invernadero y elevar la aportación de las energías renovables al 20 % del consumo.

Así, la Directiva 2010/31/CE persigue la eficiencia energética de los edificios, teniendo en cuenta las condiciones climáticas exteriores y las particularidades ambientales interiores. Establece la obligación de los estados miembros a adoptar medidas de cálculo de la eficiencia energética de los edificios (existentes como de nueva construcción) considerando las características térmicas de los edificios, los sistemas de calefacción y aire acondicionado, las condiciones ambientales, la iluminación, etc.

2.4. Calidad del Aire Interior

La **Calidad del Aire Interior, o Indoor Air Quality (IAQ)**, es un aspecto crucial en lo referente al ambiente interior de los edificios, ya que busca asegurar el bienestar y salud de los usuarios de los mismos. La contaminación presente en el aire interior puede ser desencadenante de numerosos problemas de salud, de ahí la necesidad de controlar este aspecto.

Los **contaminantes de los ambientes interiores** pueden ser de diverso tipo, como se señala en el Cuadro 2.1.

Procedencia	Origen	Contaminante
Exterior	Combustión de calderas, tráfico, actividad industrial cercana, vertederos, etc.	CO, NO _x , SO _c , CO ₂ , polvo, bacterias, etc.
Generados por el propio edificio	Combustión, humedad, desagües, moquetas, pinturas, etc.	NO _x , CO, olores, partículas, hongos, bacterias, entre otros.
Generados por los ocupantes y sus actividades	Tabaco, actividades de limpieza, control de plagas, ambientadores, etc.	CO ₂ , tabaco, bacterias, hongos, ácaros, etc.

Cuadro 2.1: Posibles contaminantes en diferentes localizaciones. Fuente: Consejería de Sanidad de la Comunidad de Madrid.

En este trabajo, de los contaminantes anteriores nos centraremos en el **dióxido de carbono (CO₂)**. En concentraciones muy elevadas (por encima de 30000 partes por millón, ppm), el CO₂ produce consecuencias muy graves para la salud, como asfixia por desplazamiento de oxígeno. No obstante, en condiciones normales, tanto en ambientes exteriores como interiores no se alcanzan esas cifras de concentración tan elevadas (en el exterior se mueve aproximadamente en un intervalo de 300-400 ppm y en el interior de 600-2000 ppm), ya que el mayor foco de emisión de este componente son las personas. De ahí que en estos casos no se considere como contaminante sino como un indicador de la calidad del aire. La presencia de CO₂ es una señal de la baja renovación de aire en el interior.

Por lo tanto la calidad del aire interior depende del correcto funcionamiento y mantenimiento de los sistemas de climatización y ventilación de los edificios, además de ayudar a conseguir unas buenas condiciones térmicas (de temperatura y de humedad, de los que hablaremos en el siguiente apartado), necesarias para alcanzar el confort térmico de los usuarios.

Para valorar la calidad del aire interior de los edificios hay que tener en cuenta además el

sistema de ventilación del mismo (que puede ser natural, mecánica y mixta), el sistema de extracción y purificación del aire, como los filtros, el sistema de climatización y aire acondicionado, el número de renovaciones de aire, los materiales de los suelos, paredes y techos del edificio así como las características técnicas y de construcción del mismo.

2.5. Confort térmico

Como ya se ha puntualizado anteriormente, es importante considerar el concepto de **confort término o calidad total del ambiente**, ya que es lo que se busca cuando se habla de calidad ambiental, y se refiere principalmente a los elementos y condicionantes que tienen un efecto en la calidad del ambiente higrotérmico, acústico, luminoso y del aire. Es decir, se podría decir que existe confort térmico cuando las condiciones de temperatura, aire y humedad son adecuados para realizar las actividades que desarrollan las personas.

No obstante, no solo influyen factores de tipo ambiental sino que también se deben tener en cuenta otros factores de carácter personal, como el tipo de actividad que se esté desarrollando, la ropa, el tiempo de permanencia en el edificio y la constitución propia de la persona.

En lo referente al confort térmico y la calidad del aire interior es conveniente hacer alusión al ya mencionado Reglamento de Instalaciones Térmicas de los Edificios (RITE), ya que regula las exigencias técnicas que deben satisfacer los edificios en cuanto a bienestar, eficiencia energética y bienestar, estableciendo unos valores límite a cumplir (se recogen sobre todo en la Parte II de dicho reglamento). Además, también es importante la Norma UNE-EN ISO 7730:2006 (de Ergonomía del ambiente térmico), que complementa las exigencias establecidas por el RITE. A continuación señalamos las más útiles para este trabajo.

Según la norma ISO 7730 el bienestar térmico (o confort térmico) es aquella condiciones en la que existe satisfacción respecto del ambiente térmico. Para medir la insatisfacción se han desarrollado dos índices que hacen más sencillo su estudio. Por una parte el **Voto Medio Estimado (PMV)**, que muestra el valor medio de los votos emitidos por un grupo numeroso de personas respecto una escala de sensación térmica de siete niveles al ser sometidos a distintos ambientes térmicos. La escala de sensación térmica se puede ver en el Cuadro 1.2.

Para calcular el PMV se emplean las siguientes ecuaciones:

$$\begin{aligned}
 PMV = & [0,303 \cdot \exp(-0,036 \cdot M) + 0,028] \cdot \\
 & \{ (M - W) - 3,05 \cdot 10^{-3} \cdot [5733 - 6,99 \cdot (M - W) - p_a] - 0,42 \cdot [(M - W) - 58,15] \\
 & - 1,7 \cdot 10^{-5} \cdot M \cdot (5867 - p_a) - 0,0014 \cdot M \cdot (34 - t_a) \\
 & - 3,96 \cdot 10^{-8} \cdot f_{cl} \cdot [(t_{cl} + 273)^4 - (\bar{t}_r + 273)^4] - f_{cl} \cdot h_c \cdot (t_{cl} - t_a) \}
 \end{aligned}
 \tag{2.1}$$

+3	Muy caluroso
+2	Caluroso
+1	Ligeramente caluroso
0	Neutro
-1	Ligeramente fresco
-2	Fresco
-3	Muy fresco

Cuadro 2.2: Escala de sensación térmica. Fuente: UNE-EN ISO 7730:2006.

$$t_{cl} = 35,7 - 0,028 \cdot (M - W) - I_{cl} \cdot \{3,96 \cdot 10^{-8} \cdot f_{cl} \cdot [(t_{cl} + 273)^4 - (\bar{t}_r + 273)^4] + f_{cl} \cdot h_c \cdot (t_{cl} - t_a)\} \quad (2.2)$$

$$h_c = \begin{cases} 2,38 \cdot |t_{cl} - t_a|^{0,25} & \text{para } 2,38 \cdot |t_{cl} - t_a|^{0,25} > 12,1 \cdot \sqrt{v_{ar}} \\ 12,1 \cdot \sqrt{v_{ar}} & \text{para } 2,38 \cdot |t_{cl} - t_a|^{0,25} < 12,1 \cdot \sqrt{v_{ar}} \end{cases} \quad (2.3)$$

$$f_{cl} = \begin{cases} 1 + 1,29l_{cl} & \text{para } l_{cl} \leq 0,078m^2 \cdot K/W \\ 1,05 + 0,645l_{cl} & \text{para } l_{cl} > 0,078m^2 \cdot K/W \end{cases} \quad (2.4)$$

donde,

M es la tasa metabólica (en vatios por metro cuadrado, W/m^2);

W es la potencia mecánica efectiva (en W/m^2);

I_{cl} es el aislamiento de la ropa (en metros cuadrados kelvin por ratio, $m^2 \cdot K/W$);

f_{cl} es el factor de superficie de la ropa;

t_a es la temperatura del aire (grados Celsius $^{\circ}C$);

\bar{t}_r es la temperatura radiante media ($^{\circ}C$);

v_{ar} es la velocidad relativa del aire (metros por segundo m/s);

p_a es la presión parcial del valor de agua (pascales Pa);

h_c es el coeficiente de transmisión de calor por convección (vatios por metro cuadrado, $W/(m^2 \cdot K)$);

t_{cl} es la temperatura de la superficie de la ropa ($^{\circ}C$).

Sin embargo, el PMV solamente predice el valor medio de los votos sobre la sensación térmica, siendo también muy útil estimar el número de personas que no están satisfechas con la temperatura existente en el edificio, por lo que se desarrolla también el **Porcentaje Estimado de Insatisfechas (PPD)**. El PPD expresa el número estimado de ocupantes que sentirán incomodidad por el calor o el frío, por lo que las personas insatisfechas térmicamente son las que votan en la escala del Cuadro 2.2 las opciones “muy caluroso”, “caluroso”, “fresco”, y “muy fresco”, el resto se sentirán neutrales, ligeramente calurosos o ligeramente frescos.

Este índice se calcula como sigue:

$$PPD = 100 - 95 \cdot \exp(-0,03353 \cdot PMV^4 - 0,2179 \cdot PMV^2) \quad (2.5)$$

El PPD no será nunca 0, ya que siempre existe un cierto número de personas no satisfechas con el ambiente, aunque si se pueden establecer ambientes aceptables para un número elevado de personas.

Para calcular estos índices es necesario conocer algunos parámetros esenciales, como las tablas del nivel de actividad de las personas (que varia dependiendo de la vestimenta, la temperatura operativa y la velocidad relativa del aire), el valor del metabolismo (medido en met^1), el valor del nivel de ropa (medido en clo^2) y la humedad. La propia norma UNE ISO 7730 proporciona tablas con diferentes valores para estos parámetros.

Así se puede obtener PMV y PPD, y el ambiente térmico deseado se seleccionará entre tres categorías distintas, A, B y C, según la tabla que se muestra a continuación extraída de la norma UNE ISO 7730:

Categoría	Estado térmico del cuerpo en su conjunto	
	PPD (%)	PMV
A	<6	-0.2 <PMV <0.2
B	<10	-0.5 <PMV <0.5
C	<15	-0.7 <PMV <0.7

Cuadro 2.3: Categorías de ambiente térmico. Fuente: UNE-EN ISO 7730:2006.

Estas tres categorías se aplican a espacios en los que las personas estén expuestas a un mismo ambiente térmico. El control individual del ambiente térmico, como por ejemplo de la temperatura local del aire, de la temperatura radiante media o de la velocidad del aire, así como la adaptación de la vestimenta, pueden ayudar a equilibrar las diferencias entre los individuos.

¹1 unidad metabólica = 1 met = 58.2 W/m^2 de superficie del cuerpo.

²1 unidad de ropa = 1 clo = $0.155 \text{ m}^2 \cdot \text{K/W}$. En la escala de clo una persona desnuda tendrá un valor de 0 clo y en traje de calle de 1 clo.

Las condiciones de diseño de la **temperatura** se fijará en base a esas categorías de ambiente térmico, de modo que, en base a la norma UNE ISO 7730, las condiciones de confort térmico para la temperatura se encuentra en los siguientes intervalos:

Tipo de espacio	Actividad (W/m^2)	Categoría	Temperatura operativa	
			Verano (Época de refrigeración)	Invierno (Época de calefacción)
Despacho, oficina, aula, cafetería, sala conferencias	70	A	$24,5 \pm 1$	22 ± 1
		B	$24,5 \pm 1,5$	22 ± 2
		C	$24,5 \pm 2,5$	22 ± 3
Aula preescolar	81	A	$23,5 \pm 1$	20 ± 1
		B	$23,5 \pm 2$	$22 \pm 2,5$
		C	$23,5 \pm 2,5$	$22 \pm 3,5$
Gran almacén	93	A	23 ± 1	$19 \pm 1,5$
		B	23 ± 2	19 ± 3
		C	23 ± 3	19 ± 4

Cuadro 2.4: Condiciones interiores de diseño para la temperatura. Fuente: UNE-EN ISO 7730:2006.

Los intervalos de temperatura señalados en dicha tabla son los que se utilizarán en este trabajo como los límites de confort térmico de consigna que se quieren alcanzar, atendiendo al tipo de edificio y su categoría ambiental.

Por su parte, la **humedad** (también aspecto relevante para el confort térmico) es una indicación de la cantidad de vapor de agua presente en el aire, y puede ser expresada como humedad relativa o como humedad absoluta. Normalmente, un aumento del 10 % de la humedad relativa supone una sensación de calor equivalente a un aumento de 0.3 grados en la temperatura operativa.

En condiciones altas de humedad se producen incomodidades debidas al bochorno o el frío, mientras que en condiciones bajas de humedad, la sequedad del aire puede producir problemas en la respiración. Por ello es un aspecto a controlar para conseguir el confort térmico de los usuarios. Se considerará un intervalo de (40 %, 60 %) de humedad relativa como los límites de consigna para el confort térmico.

En cuanto a los **límites de confort para el CO_2** se va a tener en cuenta el intervalo entre 0 y 900 ppm como máximo.

Capítulo 3

Casos prácticos. Descripción

Para el presente trabajo se analizan tres bases de datos diferentes, que se detallarán a continuación.

El primer caso se corresponde con el análisis de la eficiencia energética de diversos edificios de titularidad pública en el ayuntamiento de Mondoñedo (Lugo). Mediante el estudio estadístico descriptivo y la aplicación de herramientas para el control estadístico de la calidad, se pretende describir y evaluar el confort térmico y la capacidad que tienen las instalaciones de proporcionarlo.

En el segundo caso, el objetivo es aplicar una metodología adecuada de control del confort térmico y la eficiencia energética en viviendas y oficinas. Dada la relación de dependencia de variables como la temperatura (dependiente de las horas de actividad en una oficina) se propone la aplicación de gráficos de control para datos dependientes. En un futuro, este tipo de gráficos serán implementados en una aplicación informática para el control automático de la eficiencia energética en viviendas.

3.1. Aplicación a datos reales

3.1.1. Evaluación de la eficiencia energética en edificaciones de uso público en Mondoñedo.

- **Base de datos edificios Mondoñedo en verano**

Como ya se ha mencionado, se van a utilizar los datos relativos a tres edificios administrativos de Mondoñedo en época estival, como son el **Ayuntamiento, la Casa de Cultura y la Escuela de música**. Las observaciones se han tomado entonces en verano (los días 21, 22 y 23

de septiembre del año 2011) cada hora en cada uno de estos edificios y se disponen de más de 40 observaciones en cada uno de ellos (en total hacen unas 478 observaciones).

Además, se han tomado mediciones en varias dependencias de los edificios, así, tendríamos las habitaciones señaladas en el Cuadro 3.1 y las variables del Cuadro 3.2.

Dependencias verano
Ayuntamiento
- Pasillo
- Edificio de rehabilitación
Casa de Cultura
- Gimnasio
- Despacho del director
Escuela de Música
- Aula número 1
- Aula número 2
- Aula número 4
- Recepción
- Cabina 1
- Pasillo

Cuadro 3.1: Tabla con las dependencias analizadas en Mondoñedo (verano)

Como se observa, las variables tomadas hacen sobre todo referencia a aspectos relacionados con la calidad del aire y el ambiente de los edificios.

Con esta base de datos se llevará a cabo un primer análisis exploratorio, donde se intentarán conocer las variables más influyentes (realizando un análisis de componentes principales) y posteriormente se calcularán índices de capacidad para poder afirmar si el proceso es capaz o no, atendiendo a los límites de especificación dados.

VARIABLES
Fecha (fecha)
Hora inicial (hora.inicial)
Hora final (hora.final)
Humedad relativa media (hum.rel.media)
CO_2 inicial (co2.inicial)
CO_2 final (co2.final)
CO_2 medio (co2.medio)
Ocupación del edificio(ocupación)
Edificio (edificio)
Sala o habitación del edificio (sala)

Cuadro 3.2: Tabla con las variables de la base de datos

En la Sección 1.5 se señalan los **límites de confort para la temperatura, la humedad y para el CO_2** , y estos no son sino los límites de especificación que se quieren conseguir para poder hablar de un ambiente térmico confortable.

En la Sección 2.5 se han mencionado los límites para la humedad y el CO_2 , situándose la primera en un intervalo (40 %, 60 %) y la segunda entre (0, 900) ppm.

Por otro lado, para conocer los límites de especificación de la temperatura hay que hacer uso de las ecuaciones (1.1), (1.2), (1.3), (1.4) y (1.5), obteniendo así los índices PMV y el PPD y, en consecuencia, conocer la categoría ambiental del edificio. Dependiendo de ésta, y siguiendo el Cuadro 2.3 y el Cuadro 2.4, la temperatura se situará en unos límites u otros.

Para el Ayuntamiento (en verano) tenemos que $PPD = 10,768$ y $PPM = 0,5243$. Si se observa el Cuadro 1.3 se tiene que este edificio en época estival tiene una categoría ambiental C, por lo que los límites de consigna se sitúan en ($22^\circ C, 27^\circ C$) y el punto medio es de $24,5^\circ C$.

El edificio correspondiente a la Casa de Cultura en verano, con unos índices $PPD = 5,4652$ y $PPM = 0,0381$, tiene una categoría ambiental A. De este modo, los límites de consigna inferior y superior son ($23,5^\circ C, 25,5^\circ C$) y el punto medio por tanto $24,5^\circ C$.

- **Base de datos edificios Mondoñedo en invierno**

Análogamente al caso anterior, se dispone de datos de estos tres edificios anteriormente mencionados en invierno (en concreto los días 13, 14,15 y 16 de diciembre de 2011), haciendo un total de 758 observaciones). Sin embargo, en este caso, no se encuentran datos relativos a la ocupación de los edificios, de ahí que nos limitemos solamente al estudio de los índices de capacidad. Por otro lado, las dependencias analizadas son prácticamente las mismas que en

verano (Cuadro 3.3).

Dependencias invierno
<p>Ayuntamiento</p> <ul style="list-style-type: none"> - Pasillo - Edificio de rehabilitación
<p>Casa de Cultura</p> <ul style="list-style-type: none"> - Gimnasio - Despacho del director
<p>Escuela de Música</p> <ul style="list-style-type: none"> - Aula número 1 - Aula número 4 - Recepción - Cabina 1 - Sala de ensayos - Pasillo

Cuadro 3.3: Tabla con las dependencias analizadas en Mondoñedo (invierno)

Es igualmente necesario señalar los **límites de consigna** que se utilizarán posteriormente para calcular la capacidad del proceso. Para la humedad y el CO_2 se tienen los mismos límites que en los edificios en verano.

En cuanto al Ayuntamiento en invierno, los índices $PPD = 49,193$ y $PPM = -1,4683$ hacen que se situó el edificio en una categoría ambiental C, teniendo unos límites de especificación de ($19^\circ C, 25^\circ C$), con $19^\circ C$ como punto medio.

Asimismo, tanto la Casa de Cultura como la Escuela de Música, con unos índices de $PPD = 55,8045$ y $PPM = -1,5904$ y $PPD = 72,279$ y $PPM = -1,907$ respectivamente, también tienen una categoría ambiental C.

3.1.2. Control de la eficiencia energética y confort térmico en viviendas y oficinas: base de datos tomada en Fridama Instalaciones S.L.

Por último también se van a analizar datos relativos a la **empresa Fridama**. Esta empresa, con una delegación en Panamá y otra en España (en concreto en A Coruña) está dedicada a la instalación de sistemas de climatización, sistemas eléctricos y el control de la calidad y de la eficiencia energética de sus instalaciones. Concretamente llevan a cabo proyectos en el sector terciario (como locales comerciales, hoteles, centros comerciales, oficinas, museos, centros deportivos u hospitales).

En la actualidad la empresa está intentando sacar al mercado un nuevo producto por el que los usuarios puedan monitorizar tanto su consumo energético como otras variables relacionadas con el confort (la temperatura, la humedad o el CO_2). De ahí que consideren la necesidad de controlar tanto el nivel de confort térmico como el gasto de energía que se produce.

Para el estudio se cuenta con una base de datos referente a 15 días (del 11 al 25 del mes de abril de 2014), donde se han tomado varias mediciones de diferentes variables (concretamente 316) y se han sacado los promedios cada 5 minutos, haciendo un total de 4320 observaciones. No se utilizará la información que nos aportan todas las variables, sino que se escogerán solamente algunas que sean de interés para el estudio, específicamente las relativas a la temperatura media, el CO_2 medio y el consumo energético.

El control de estas variables responderá a las preguntas de si la instalación es o no eficiente, además de si es o no confortable. La aplicación de gráficos de control permitirá identificar los límites naturales del servicio, además de las causas asignables correspondientes a los fallos en el control del confort y de la energía.

Al disponer de datos del 11 al 25 de abril se tienen observaciones tanto en días laborables como en días festivos, y se reparten de acuerdo al Cuadro 3.4.

Estas mediciones se han realizado en las propias instalaciones de Fridama en Coruña, formadas por seis zonas diferenciadas, que se señalan en el plano de la Figura 3.1.

A efectos prácticos se utilizarán los datos referentes a la Oficina medio (denominada como CL2 en la base de datos) y las variables relativas a la temperatura media (CL2_Temp_Ambiente) y al nivel de CO_2 (CL2_C02) presente en dicha oficina. También se analizará la variable que hace referencia al consumo energético (AC1_pot_activa) con los valores que proporciona el analizador principal de las oficinas.

En primer lugar se realizará un análisis descriptivo desde una perspectiva del análisis de datos funcionales, pues es importante conocer si existen diferencias significativas en la calidad del aire en función de si el día es laborable o festivo. Si es así se deben realizar los gráficos de control por separado para cada uno de los grupos.

Días laborables	Días festivos
Viernes 11	Sábado 12
Lunes 14	Domingo 13
Martes 15	Jueves 17
Miércoles 16	Viernes 18
Lunes 21	Sábado 19
Martes 22	Domingo 20
Miércoles 23	
Jueves 24	
Viernes 25	

Cuadro 3.4: Tabla con los días en los que la empresa Fridama ha tomado mediciones, diferenciando entre laborables y festivos.

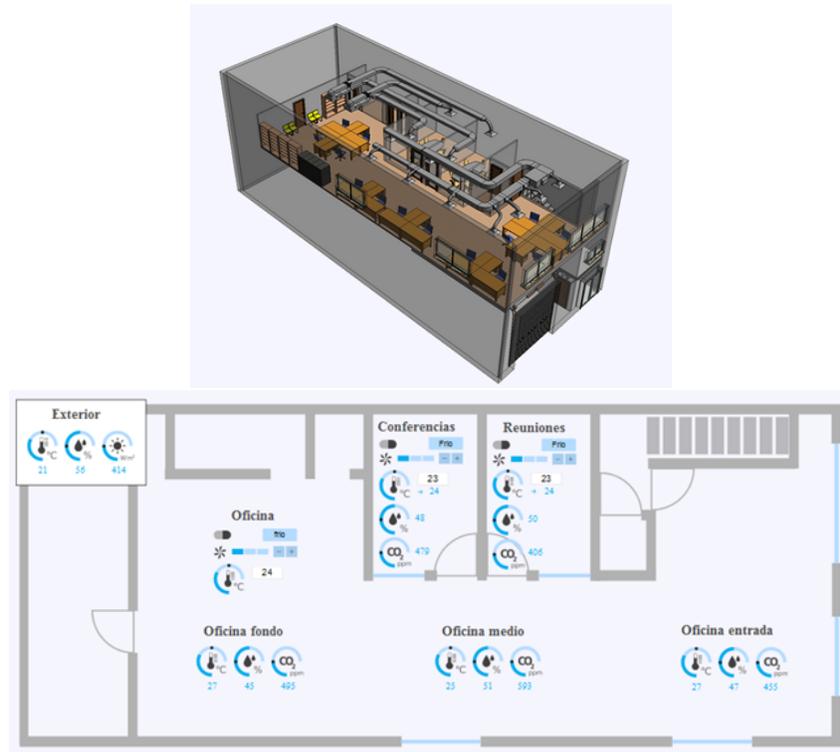


Figura 3.1: Planos de las oficinas de Fridama en su local de Coruña.

La segunda parte del análisis se centrará entonces en la aplicación de gráficos de control. Se comprobará si existe autocorrelación entre las observaciones, pues de ser así será necesario ajustar un modelo de serie temporal a una parte de los datos (muestra de calibrado) y con los residuos de dicho modelo se calcularán los límites de control para poder monitorizar el resto de

la muestra.

Con los resultados de dichos gráficos de control se podrá comprobar si los procesos analizados están bajo control o si, por el contrario, se encuentran fuera de control y por tanto necesitan corregirse para que se pueda hablar de calidad y eficiencia de las instalaciones.

Capítulo 4

Métodos estadísticos

Se introducirá en esta Sección los enfoques, técnicas y herramientas que se utilizan para el análisis de los datos, así como la notación empleada.

4.1. Control Estadístico de Calidad

Como ya se ha mencionado en la Sección 1.1, uno de los instrumentos fundamentales dentro del control estadístico de procesos o SPC son los **gráficos de control**, sobre todo los gráficos de la media o de rangos, que se explicarán a continuación. Asimismo, hay situaciones en las que es más adecuado realizar gráficos de control **EWMA**, por lo que también se ilustrará en detalle.

Por último, también es importante describir los **índices de capacidad**, ya que se aplicarán en el análisis de los datos de Mondoñedo.

4.1.1. Gráficos de Control por variables

Si la característica que se quiere controlar se puede representar por una variable aleatoria continua X existen dos tipos de gráficos muy habitualmente utilizados, propuestos por Shewhart, por un lado los gráficos para monitorizar el promedio y controlar así la posición (los gráficos de media \bar{x} o de la mediana) y, por otro lado, los gráficos que controlan la dispersión y variación (gráfico de desviaciones, s o gráfico de rango, R).

Primero conviene señalar que los gráficos de control se construyen en dos fases. En la **Fase I** se utiliza un número m de muestras (normalmente $m = 20$ o 25) cada una de un determinado tamaño, n , para establecer los límites de control y determinar de este modo si el proceso se encuentra bajo control en el momento en el que se recogen los datos. Los puntos que en esta fase no estén dentro de dichos límites deben ser examinados cuidadosamente y, si se encuentran

causas asignables, eliminados del proceso. Una vez se eliminen deben ser recalculados los límites de control. Cuando el proceso sea estable, es decir, que no se encuentren causas de variación más allá de las aleatorias, se puede iniciar la siguiente fase.

Así, en la **Fase II**, se utilizan el gráfico de control para monitorizar y supervisar el proceso, comparando las muestras tomadas posteriormente con los límites de control ya calculados.

Por otra parte, la mecánica de los gráficos de control puede ser vista como un **contraste de hipótesis**, donde se contraste la hipótesis nula (H_0), que es la asumida como correcta. Se puede relacionar con la perspectiva del control de calidad definiendo la hipótesis nula como aquella que supone que el proceso está bajo control en las muestras seleccionadas. Si se rechazase esta hipótesis se aceptaría la alternativa, H_1 , indicando así que el proceso está fuera de control.

Podría darse el llamado error de tipo I (α) si se rechaza H_0 cuando ésta es cierta (es decir si se acepta que el proceso está bajo control cuando no es así), así como el error de tipo II (β), que ocurre al aceptar la hipótesis nula cuando es falsa.

Por ello, para comprobar la potencia del test se puede hacer uso de la **curva característica de operación** (OC por sus siglas en inglés), en la que se representa el error de tipo II en función de la magnitud del cambio que se quiere estudiar.

Para seleccionar el tamaño muestral adecuado se pueden utilizar estas curvas OC, ya que ofrecen la posibilidad de estimar el tamaño de muestra necesario para detectar cambios en el error β . De este modo, se puede obtener el recorrido medio de racha, más comúnmente denominado **ARL (Average Run Length)**, que es el promedio de puntos (muestras) que se deben representar en el gráfico de control hasta que el primer punto indica un fuera de control. Se puede calcular como sigue:

$$ARL = \frac{1}{p} \quad (4.1)$$

donde p es la probabilidad de que cualquier punto exceda los límites de control.

Como ya se ha señalado H_0 supone que el proceso está bajo control, y la probabilidad de rechazar dicha hipótesis cuando el proceso si está bajo control está asociada a α , por lo que el número de conformes hasta el primer fuera de control cuando el proceso está bajo control (dentro de la H_0) se denomina ARL_0 , y viene dado por:

$$ARL_0 = \frac{1}{\alpha} \quad (4.2)$$

Cuando se utilizan los límites de control 3 sigma la probabilidad de que un punto normalmente distribuido caiga fuera de los límites cuando el proceso está bajo control es de $\alpha = 0,0027$, por lo que el ARL_0 será: $ARL_0 = \frac{1}{0,0027} = 370,3704$. Esto quiere decir que en promedio se ob-

tendrá una falsa alarma (señal de fuera de control) cada 370 muestras, cuando en realidad el proceso está bajo control.

Del mismo modo, H_1 supone aceptar la asunción de que el proceso está bajo control cuando no lo está, por lo que se encuentra asociado a β . Por tanto el número promedio de puntos representados hasta que el gráfico detecta un cambio (un fuera de control) se denomina ARL_1 y se define como:

$$ARL = \frac{1}{1 - \beta} \quad (4.3)$$

Interesa que α sea pequeño para que las falsas alarmas sean menos frecuentes (ya que así el ARL_0 será mayor) y que el ARL_1 sea el menor posible para detectar antes un fuera de control (por lo que interesa un β pequeño).

Para detectar procesos fuera de control se pueden utilizar las reglas de la Western Electric (Western Electric Handbook, 1956) que señalan que un el proceso está fuera de control si:

1. Un punto cae fuera de los límites de control 3 sigma.
2. Dos de tres puntos consecutivos caen más allá de los límites 2 sigma.
3. Cuatro de cinco puntos caen a una distancia de 1 sigma o más allá de la línea central.
4. Ocho puntos consecutivos se encuentran en un mismo lado de la línea central.

No se deben utilizar todas las reglas anteriores a la vez ya que podría aumentar la probabilidad de falsas alarmas, rechazándose incorrectamente que el proceso está bajo control (Box, Luceño y Paniagua-Quiñones, 2009).

A continuación, supongamos que la característica de calidad que queremos estudiar se distribuye normalmente, con media μ y desviación típica σ . Habitualmente éstas no se conocen, por lo que deben ser estimadas de las muestras preliminares tomadas cuando el proceso se considera bajo control, es decir, de las m muestras de tamaño n (normalmente $n = 5$ o 6).

De este modo, el estimador de la media del proceso, μ , se denotará por $\bar{\bar{x}}$ y se calcula como sigue:

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m}{m} \quad (4.4)$$

donde $\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m$ es la media de cada muestra.

Este estimador $\bar{\bar{x}}$ será la línea central del gráfico de control.

Por su parte, para estimar σ se utilizarán las estimaciones provenientes de calcular la varianza o recorrido (rango) de cada submuestra por separado para posteriormente utilizar su media.

- **Gráficos de control (\bar{x}, R)**

El rango o recorrido es la diferencia entre la mayor y menor de las observaciones de una muestra. Sea x_1, x_2, \dots, x_n una muestra de n observaciones que sigue una distribución normal con media μ y varianza σ^2 . El rango será:

$$R = \max(x_i) - \min(x_i) \quad (4.5)$$

Bajo la hipótesis de normalidad se puede calcular la distribución del rango muestral. La media es $d_2\sigma$, siendo d_2 una constante que depende del n , tamaño de muestra, y que está tabulada.

Un estimador insesgado de σ es:

$$\hat{\sigma} = \frac{R}{d_2} \quad (4.6)$$

Sea entonces R_1, R_2, \dots, R_m los rangos de las muestras m . Definimos el rango medio como:

$$\bar{R} = \frac{R_1 + R_2 + \dots + R_m}{m} \quad (4.7)$$

Pudiendo utilizar así como estimador insesgado de σ :

$$\hat{\sigma} = \frac{R}{d_2} \quad (4.8)$$

A través de esa estimación se pueden construir los gráficos de control para la media y construir los límites de control de dichos gráficos (\bar{x}, R) como sigue (utilizando asimismo \bar{x} como estimador de μ):

$$\bar{x} \pm \frac{2\bar{R}}{d_2\sqrt{n}} \quad (4.9)$$

Si se define $A_2 = \frac{3}{d_2\sqrt{n}}$ tendremos los límites de control:

$$\begin{aligned}
 LCL &= \bar{\bar{x}} - A_2 \bar{R} && \text{límite inferior} \\
 CL &= \bar{\bar{x}} && \text{línea central} \\
 UCL &= \bar{\bar{x}} + A_2 \bar{R} && \text{límite superior}
 \end{aligned}
 \tag{4.10}$$

donde los valores de la constante A_2 están tabulados para distintos valores de n .

- **Gráficos de control (\bar{x}, s)**

Estos gráficos en vez de utilizar el rango para la estimación de la desviación típica del proceso se utiliza directamente la desviación típica de cada submuestra.

En primer lugar, en este caso, se debe calcular la media y la desviación típica para cada muestra, esto es, \bar{x}_i y s_i respectivamente, donde $i = (1, 2, \dots, m)$.

Posteriormente se debe calcular tanto la media global (con la ecuación 4.4) como la desviación típica global.

La desviación típica muestral, s , no es un estimador insesgado (aunque sí asintóticamente insesgado) de la desviación típica teórica σ , ya que $E(s) = c_4 s$, donde c_4 es una constante que depende del valor de n y está tabulada.

De este modo, $\frac{s}{c_4}$ es un estimador insesgado de σ y así se puede estimar la desviación típica como:

$$\hat{\sigma} = \frac{\bar{s}}{c_4} \tag{4.11}$$

donde \bar{s} es el promedio de las m desviaciones estándar (siendo s_i la desviación típica de la i -ésima muestra): $\bar{s} = \frac{1}{m} \sum_{i=1}^m s_i$

Se pueden definir los parámetros del gráfico de desviaciones típicas s como:

$$\begin{aligned}
 LCL &= B_3 \bar{s} \\
 CL &= \bar{s} \\
 UCL &= B_4 \bar{s}
 \end{aligned}
 \tag{4.12}$$

donde B_3 y B_4 están tabuladas y se refieren a $B_3 = 1 - \frac{3}{4} \sqrt{1 - c_4^2}$ y $B_4 = 1 + \frac{3}{4} \sqrt{1 - c_4^2}$.

Por su parte, los parámetros del gráfico para la media serán:

$$\begin{aligned}
 LCL &= \bar{\bar{x}} - A_3 \bar{s} \\
 CL &= \bar{\bar{x}} \\
 UCL &= \bar{\bar{x}} + A_3 \bar{s}
 \end{aligned}
 \tag{4.13}$$

siendo $A_3 = \frac{3}{c_4 \sqrt{n}}$.

- **Gráficos de medidas individuales**

Además de los gráficos anteriormente mencionados, donde tenemos m muestras de tamaño n cada una, también son utilizados los gráficos de medidas individuales, en los que cada muestra está formada únicamente por una sola observación (es decir, $n = 1$).

En este trabajo se aplicarán fundamentalmente estos gráficos (junto con los EWMA), ya que se trabajará con datos tomados cada cierto tiempo, formando una única observación por muestra.

Existen principalmente dos métodos para estimar la desviación típica, los rangos móviles y la desviación típica muestral.

El rango móvil es la diferencia entre observaciones consecutivas y se define como sigue:

$$MR_i = |x_i - x_{i-1}|, \quad i = 2, 3, \dots, m. \tag{4.14}$$

Con esto se puede obtener el estimador de la desviación típica como:

$$\hat{\sigma} = \frac{1}{d_2} \overline{MR} \tag{4.15}$$

donde \overline{MR} es el promedio del rango móvil, definido por $\overline{MR} = \sum_{i=2}^m \frac{MR_i}{m-1}$, y d_2 es una constante tabulada para el gráfico de control y donde $n = 2$, siendo entonces en este caso $d_2 = 1,128$.

Los límites superior e inferior para el gráfico de control con medidas individuales y la línea central quedarán definidos como:

$$\begin{aligned}
 LCL &= \bar{\bar{x}} - 3 \frac{\overline{MR}}{d_2} \\
 CL &= \bar{\bar{x}}
 \end{aligned}
 \tag{4.16}$$

$$UCL = \bar{\bar{x}} + 3 \frac{\overline{MR}}{d_2} \tag{4.17}$$

donde \bar{x} es la media de las observaciones.

La desviación típica también puede ser estimada, como ya se ha dicho, a través de la desviación típica muestral:

$$\hat{\sigma} = \frac{s}{c_4} \quad (4.18)$$

En otro orden de cosas, también conviene señalar que en casos en los que la normalidad no esté garantizada es recomendable utilizar los gráficos EWMA y no los gráficos para medidas individuales, ya que en este caso el ARL_0 se ve afectado cuando los datos no cumplen esta hipótesis (Montgomery, 2009).

4.1.2. Gráficos de control EWMA

El gráfico de control de medias móviles ponderadas exponencialmente o gráfico EWMA (Exponentially Weighted Moving Average) fue propuesto por primera vez por Roberts en 1959 y es muy utilizado en el caso de observaciones individuales. Estos datos individuales pueden ser medias, proporciones, lecturas individuales, cocientes u otras medidas similares.

Una de las mayores desventajas de los gráficos de control de Shewhart que se vienen describiendo hasta ahora es el gran peso que se le da a la última muestra, ignorando la información disponible en todas las anteriores observaciones, por eso se dice que no tiene memoria.

Por su parte, el gráfico EWMA es un **gráfico con memoria**, ya que utiliza la información de las medias de las observaciones pasadas, dando más peso a las medias más cercanas. De este modo, si lo que interesa detectar son cambios pequeños en la característica de calidad estudiada puede ser más recomendable utilizar este tipo de gráficos, ya que son más sensibles a los mismos al acumular información de muchos períodos, de modo que el cambio podrá apreciarse.

Además se pueden estudiar desde el análisis de las series temporales, ya que en muchas ocasiones las observaciones están tomadas en distintos períodos de tiempo, formando así una serie temporal.

La media móvil ponderada exponencialmente en el tiempo actual t se construye como sigue:

$$Z_t = \lambda x_t + (1 - \lambda)Z_{t-1} \quad (4.19)$$

Donde:

Z_t es la media móvil ponderada exponencialmente en el tiempo actual t ;

Z_{t-1} es la media móvil ponderada exponencialmente en el tiempo inmediatamente anterior $t-1$;

x_t es la observación actual en el tiempo t ;

λ es el factor de ponderación para la observación actual, situándose en el intervalo $0 \leq \lambda \leq 1$;

$Z_0 = \mu_0$, el valor inicial que se necesita cuando $t = 1$ es la media del proceso.

Escogiendo un valor u otro para λ el procedimiento EWMA será más o menos sensible a los pequeños cambios, ya que es el factor que proporciona los pesos que se le darán a los datos históricos. Viendo la ecuación 4.19 se puede ver fácilmente que si se selecciona un valor de λ alto, en torno al 1, se estará dando poco a las observaciones precedentes, construyendo un gráfico similar a los de Shewhart. Por su parte, si se selecciona un λ cercano a 0 al contrario, los datos anteriores tendrán más peso. Se recomienda utilizar valores de λ en torno a $0,05 \leq \lambda \leq 0,25$, siendo común utilizar 0,20 (Hunter, 1986; Montgomery, 2009).

Los límites de control y la línea central para los gráficos EWMA son:

$$\begin{aligned} LCL &= \mu_0 + L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}[1-(1-\lambda)^{2t}]} \\ CL &= \mu_0 \\ UCL &= \mu_0 - L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}[1-(1-\lambda)^{2t}]} \end{aligned} \tag{4.20}$$

En las ecuaciones anteriores L es la anchura de los límites de control y normalmente se utilizan los límites 3 sigma por lo que $L = 3$. En este trabajo se va a utilizar un valor para λ de 0.2 y $L = 3$, pues dichas combinaciones se hacen con el objetivo de obtener unos ARL_0 (número de conformes hasta el primer fuera de control dentro de H_0) altos y unos ARL_1 (número de conformes hasta el primer fuera de control dándose un sesgo de σ con respecto a H_0) bajos.

Ya que en la mayoría de los casos la media y la desviación típica se desconocen, se estima su valor de la muestra de calibrado en la Fase I y se denotan por $\hat{\mu}$ y $\hat{\sigma}$, obteniendo así los límites de control necesarios para la Fase II de monitorizado.

- **No normalidad: robustez de los gráficos EWMA**

Como ya se ha señalado al hablar de los gráficos de medidas individuales, cuando se utilizan observaciones individuales y la normalidad de los datos no está garantizada es mejor utilizar los gráficos EWMA ya que funciona mejor en estas ocasiones.

Esto es así porque en ausencia de normalidad el ARL_0 se reduce en gran medida, haciendo aumentar el número de falsas alarmas en el caso de los gráficos Shewhart. El uso de los gráficos EWMA y la elección de los parámetros λ y L funciona bien en los casos en los que la normalidad no se satisface (Montgomery, 2009).

- **Autocorrelación de los datos**

Una de las suposiciones básicas a la hora de aplicar los gráficos de control es que las observaciones sean normales e independientemente distribuidas, asunciones que son habitualmente violadas. Ya se ha señalado que si no se cumple normalidad los gráficos de Shewhart son menos recomendables que los gráficos EWMA u otros.

Además, en las situaciones en las que se cuenta con una serie de datos individuales es recurrente la presencia de dependencia serial en los mismos, sobre todo en la mayor parte de las aplicaciones industriales actuales o al tomar los datos a lo largo del tiempo.

La presencia de autocorrelación no debe ser ignorada pues tiene efectos importantes en los gráficos de control. Uno de los más relevantes es que los límites de control serán más estrechos, afectando al cálculo del ARL y produciendo un mayor número de falsas alarmas.

Ante la presencia de dependencia existen diversos enfoques para construir gráficos de control:

- El primer de los enfoques es el más básico, y consiste en tomar muestras de manera menos frecuente. El gran inconveniente es la gran pérdida de información que supone.
- El segundo enfoque consiste en la utilización de las **series temporales**, ajustando un modelo adecuado a los datos de manera que se elimine la autocorrelación y poder utilizar los gráficos de control sobre los residuos del modelo.
- Un último enfoque es el de la utilización de un modelo **Free Approach**, donde se agrupan las observaciones en muestras de tamaño b y se dan los mismos pesos para cada punto de las muestras.

En el presente trabajo se utilizará el segundo enfoque, ajustando series de tiempo de modo que se puedan utilizar los residuos para la utilización de los gráficos de control tanto en la Fase I como en la II. En la Sección 4.2 se explicará con detalle los métodos utilizados para el análisis de series temporales.

4.1.3. Índices de capacidad

Se emplean asimismo los llamados **índices de capacidad del proceso**. La capacidad de proceso se interpreta como la aptitud de ese proceso de cumplir las especificaciones técnicas deseadas (normalmente marcadas por los clientes o la propia empresa) y se refiere por tanto a la uniformidad del proceso (medida por la variabilidad). La variabilidad puede ser inherente a una determinada observación (variabilidad “instantánea”) o puede darse por la diferencia entre las observaciones a lo largo del tiempo.

Los índices permiten medir si el **proceso es o no capaz**, de modo que se relaciona la variabilidad con las especificaciones dadas. Una gran parte de la literatura en este ámbito se centra en el análisis de la capacidad del proceso asumiendo normalidad en la distribución de la variable, y los índices más estudiados en ese caso son los índices C_p , C_{pk} , C_{pm} y C_{pmk} , que analizaremos a continuación. Estos límites dependerán además del límite inferior de especificación (LSL) y el límite superior de especificación (USL).

Se deben diferenciar los límites de control de los límites de especificación (Montgomery, 2009), ya que no hay una relación entre ambos. Mientras que los límites de control son dados por la variabilidad inherente del proceso, los de especificación se proporcionan externamente (por los clientes, la empresa, normas existentes...).

El intervalo de los límites de control viene dado por $[T - 3\sigma, T + 3\sigma]$, donde T es el valor nominal o Target que suele ser el punto medio de los límites de especificación. Esto es así ya que cuando un proceso está bajo control el 99.73 % de las unidades se encuentran en un intervalo de amplitud 6σ centrado en dicho valor nominal.

El índice C_p es un índice de primera generación que se define como:

$$C_p = \frac{USL - LSL}{6\sigma} = \frac{d}{3\sigma} \quad (4.21)$$

donde $d = \frac{USL - LSL}{2}$ y σ es la desviación estándar del proceso.

En la práctica σ es casi siempre desconocida, por lo que se sustituye por su estimador, $\hat{\sigma}$:

$$\hat{C}_p = \frac{USL - LSL}{6\hat{\sigma}} = \frac{d}{3\hat{\sigma}} \quad (4.22)$$

Este índice mide la capacidad potencial del proceso, comparando la variabilidad natural con el intervalo de especificación.

Sin embargo, C_p no tiene en cuenta la localización de la media del proceso en relación a las especificaciones, algo que si se considera con el índice C_{pk} , que se construye a partir de los

índices de capacidad inferior (C_{PL}) y superior (C_{PU}):

$$C_{PL} = \frac{\mu - LSL}{3\sigma} \quad (4.23)$$

$$C_{PU} = \frac{USL - \mu}{3\sigma} \quad (4.24)$$

$$C_{pk} = \min\{C_{PU}, C_{PL}\} = \frac{d - |\mu - m|}{3\sigma} \quad (4.25)$$

donde μ es la media del proceso, y m el centro de especificación, dado por $m = \frac{USL+LSL}{2}$.

Ya que en la mayor parte de los casos no se conocen σ ni μ , se deben estimar, teniendo entonces:

$$\hat{C}_{PL} = \frac{\hat{\mu} - LSL}{3\hat{\sigma}} \quad (4.26)$$

$$\hat{C}_{PU} = \frac{USL - \hat{\mu}}{3\hat{\sigma}} \quad (4.27)$$

$$\hat{C}_{pk} = \min\{\hat{C}_{PU}, \hat{C}_{PL}\} = \frac{d - |\hat{\mu} - m|}{3\hat{\sigma}} \quad (4.28)$$

En relación a lo señalado anteriormente sobre la diferencia entre los límites de especificación y los de control, cabe señalar que el proceso se comienza a considerar capaz cuando $\hat{C}_{pk} > 1$, ya que eso significará que el proceso no usa el 100 % del intervalo de tolerancia, por lo que habrá pocos no conformes. Sin embargo, si $\hat{C}_{pk} < 1$ el caso será el contrario, el proceso va más allá de la banda de tolerancia, lo que supone la existencia de un elevado número de no conformes. Por su parte, si $\hat{C}_{pk} = 1$ el proceso tiene una extensión casi igual a la banda de tolerancia. En este caso el proceso será capaz, pero debe ser controlado ya que posibles desviaciones podrán hacer aumentar el número de no conformes.

Esto también se puede aplicar al índice de \hat{C}_p , pero teniendo presente que este no tiene en cuenta la posición de la media del proceso, por lo que incluso siendo $\hat{C}_p > 1$ podría darse que el proceso no fuese capaz.

No obstante, algunos autores (Montgomery, 2009) comienzan a considerar que el proceso es capaz a partir de 1.33, ya que eso supondrá menos de 96 no conformes por millón. Si se considera que es capaz desde 1 habrá menos de 2700 no conformes por millón (0.27 %), por lo que la diferencia es elevada.

Además, desde la metodología Six Sigma estos índices deberían ser como mínimo de 2 para considerarse un proceso como capaz. Se resumen los valores mínimos y los errores por millón que equivalen a cada valor en la siguiente tabla:

	Requisitos mínimos especificaciones bilaterales	Errores por millón especificaciones bilaterales	Requisitos mínimos especificaciones unilaterales	Errores por millón especificaciones unilaterales
Procesos existentes	1.33	< 96	1.25	< 159
Procesos nuevos	1.50	7	1.45	< 14
Procesos existentes (seguridad, resistencia, parámetros críticos)	1.50	7	1.45	< 14
Procesos nuevos (seguridad, resistencia, parámetros críticos)	1.67	< 2	1.60	1
Proceso Six Sigma	2	0.0018	2	0.0009

Cuadro 4.1: Valores mínimos recomendados para los índices de capacidad y y sus correspondientes errores por millón, bajo la hipótesis de proceso centrado y distribución normal de la característica de calidad medida.

Asimismo cuanto más similares \hat{C}_{pk} y \hat{C}_p más centrado estará el proceso en torno a la media.

Por su parte, también se suelen utilizar los índices de segunda generación como el C_{pm} , que considera lo descentrada que está la media con respecto al Target (T) o valor nominal.

$$C_{pm} = \frac{USL - LSL}{6\sqrt{\sigma^2 + (\mu - T)^2}} \quad (4.29)$$

$$\hat{C}_{pm} = \frac{USL - LSL}{6\sqrt{\hat{\sigma}^2 + (\hat{\mu} - T)^2}} \quad (4.30)$$

Lo que se intenta con este índice es reducir la variabilidad en torno al valor nominal o Target.

Además, también se propusieron los llamados índices de tercera generación como el C_{pmk} (Pearn, Kotz y Johnson, 1992), con el motivo de aumentar la sensibilidad de los desplazamientos de la media del proceso con respecto al Target.

$$C_{pmk} = \frac{d - |\mu - m|}{3\sqrt{\sigma^2 + (\mu - T)^2}} \quad (4.31)$$

$$\widehat{C}_{pmk} = \frac{d - |\widehat{\mu} - m|}{3\sqrt{\widehat{\sigma}^2 + (\widehat{\mu} - T)^2}} \quad (4.32)$$

donde d es la mitad de la amplitud del intervalo de especificación ($d = \frac{USL - LSL}{2}$) y m el punto medio de dicho intervalo ($m = \frac{ULS + LSL}{2}$).

Por último, señalar la propuesta de Vännman (1995) que generaliza a una sola expresión todos los índices anteriores, que depende solamente de los parámetros no negativos (u y v):

$$C_p(u, v) = \frac{d - u|\mu - m|}{3\sqrt{\sigma^2 + v(\mu - T)^2}} \quad (4.33)$$

$$\widehat{C}_p(u, v) = \frac{d - u|\widehat{\mu} - m|}{3\sqrt{\widehat{\sigma}^2 + v(\widehat{\mu} - T)^2}} \quad (4.34)$$

Cambiando los valores de u y v se obtienen los índices anteriormente mencionados:

$$C_p(0, 0) = C_p; \quad C_p(1, 0) = C_{pk}; \quad C_p(0, 1) = C_{pm} \quad \text{y} \quad C_p(1, 1) = C_{pmk}.$$

No obstante, ya se ha señalado que estos índices se pueden aplicar bajo la suposición de normalidad, que en muchos casos es violada, y la utilización de los mismos llevaría a una consideración de no conformes no correcta.

Pearn y Chen (1997a) propusieron una generalización de los índices anteriores para el caso de **no normalidad** en los datos, el índice $C_{Np}(u, v)$, que se expresa como sigue:

$$C_{Np}(u, v) = \frac{d - u|M - m|}{3\sqrt{\left(\frac{F_{99,865} - F_{0,135}}{6}\right)^2 + v(M - T)^2}} \quad (4.35)$$

siendo M la mediana de la distribución, $m = \frac{USL + LSL}{2}$, $d = \frac{USL - LSL}{2}$ y F_α el correspondiente percentil de la distribución.

Tal como afirman Pearn y Chen (1997a), la sustitución de la media del proceso por la mediana y la desviación típica de dicho proceso por $\frac{F_{99,865} - F_{0,135}}{6}$ se realiza para imitar la propiedad de la distribución normal para la cual la probabilidad de cola fuera de los límites $\pm 3\sigma$ es 0,27 %, de este modo se asegura que si $C_{Np}(u, v) = 1$ será pequeña la probabilidad de que el proceso esté fuera de los límites de especificación.

Se obtienen los estimadores de los percentiles de la distribución empírica sacada de la

muestra y:

$$\widehat{C}_{Np}(u, v) = \frac{d - u|\widehat{M} - m|}{3\sqrt{(\frac{\widehat{F}_{99,865} - \widehat{F}_{0,135}}{6})^2 + v(\widehat{M} - T)^2}} \quad (4.36)$$

Fijando los valores de u y v tendremos el resto de índices:

$$C_{Np}(0, 0) = C_{Np}; \quad C_{Np}(1, 0) = C_{Npk}; \quad C_{Np}(0, 1) = C_{Npm} \quad \text{y} \quad C_{Np}(1, 1) = C_{Npmk}.$$

En los datos analizados en este trabajo, referentes a los edificios de Mondoñedo, se van a aplicar tanto los índices que asumen normalidad como estos últimos índices robustos.

4.2. Series de Tiempo

Como se ha señalado anteriormente, se hará uso de las series temporales para tratar la autocorrelación de los datos y poder construir gráficos de control adecuados a estas situaciones.

- **Terminología para el análisis de series temporales**

Una serie temporal se define como un conjunto de observaciones tomadas secuencialmente en el transcurso del tiempo. Se tendrán aquí en cuenta series discretas, formadas por observaciones tomadas en intervalos de igual longitud. Una serie de tiempo (x_1, x_2, \dots, x_T) puede ser vista como una realización parcial de un **proceso estocástico** subyacente (colección de variables aleatorias X_t , con $t \in T$ definidas sobre un mismo espacio de probabilidad), que se supone generador de dicha serie.

Si un proceso (estocástico) se dice que es **estacionario** si:

- Tiene una media constante, no cambia al paso del tiempo: $\mu_t = \mu, \forall t$.
- La varianza también es constante: $\sigma_t^2 = \sigma^2, \forall t$.
- La autocovarianza (denotada por γ) entre dos observaciones tomadas en distinto tiempo (t y $t + k$) depende solamente de la distancia existente entre ellas (k) y no del tiempo en sí. Es decir: $\gamma(t, t + k) = \gamma_k, \forall t, k$.

De este modo las características del proceso (media, varianza, autocovarianza) son estables a lo largo del tiempo. Si estas condiciones no se cumplen se dice que la serie es no estacionaria.

La **función de autocorrelación simple** (ACF por sus siglas en inglés) se denota como ρ_k y es la función de autocorrelación entre variables separadas k instantes y viene a determinar el grado de correlación de un valor del proceso con otro a un retardo de k . Se define como:

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (4.37)$$

Asimismo, se denotará como α_k a la **función de autocorrelación parcial** o PACF, que mide la correlación entre los instantes x_t y x_{t+k} sin tener en cuenta los instantes intermedios.

Como solamente se dispone de la realización x_1, x_2, \dots, x_T del proceso, los estadísticos deben ser estimados. Así, la media muestral:

$$\hat{x} = \frac{1}{T} \sum_{t=1}^T x_t \quad (4.38)$$

La estimación de γ_k viene dada por:

$$\hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \quad (4.39)$$

Por último, la ACF muestral será:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad (4.40)$$

Una de las clases más importantes de proceso estocástico es el conocido como **ruído blanco** (a_t), que es un proceso estacionario con media 0 y varianza constante σ_a^2 . Se caracteriza por:

- $\mu = 0$.
- $\sigma^2 = \sigma_a^2$.
- $\gamma_k = \begin{cases} \sigma_a^2 & \text{si } k = 0 \\ 0 & \text{si } k \neq 0 \end{cases}$.

Se singulariza entonces por no tener tendencia y por la ausencia de correlación entre las observaciones. Además, si tiene una distribución normal, las observaciones se consideran independientes e igualmente distribuidas (i.i.d.).

Hasta ahora se ha estado hablando de procesos estacionarios, estables en su media, varianza y autocovarianza. Sin embargo, habitualmente las series no lo son, y suelen presentar tanto tendencia como patrones cíclicos o repetitivos. Cuando ocurre esto, es necesario **diferenciar la serie** para convertirla en estacionaria (cuando así sea la diferencia se dirá que es de orden d). Por ejemplo la diferencia de orden 1 vendrá dada por $x_t - x_{t-1}$.

Uno de los procesos no estacionarios más comunes es el llamado **paseo aleatorio**. Si comienza en $t = 0$ y $x_0 = 0$, el paseo aleatorio viene definido por:

$$x_t = c + \sum_{j=1}^t a_j \quad (4.41)$$

donde a_t es ruido blanco.

Este proceso no es estacionario porque no es estable en varianza ni en las autocovarianzas. En la media solamente lo es en el caso en el que $c = 0$.

- **Modelos de series temporales**

A continuación se explican los modelos más usuales para modelizar series de tiempo, encuadrados dentro de la metodología Box-Jenkins. En 1970 Box y Jenkins popularizaron los modelos ARMA proporcionando directrices para hacer las series estacionarias y sugiriendo el uso de la ACF y PACF para determinar los coeficientes del modelo, extendiendo el uso de los modelos ARIMA (Autoregressive Integrated Moving Average).

Un proceso x_t se dice que es un **proceso autorregresivo** de orden 1, denominado como **AR(1)**, si puede ser generado por:

$$x_t = c + \phi_1 x_{t-1} + a_t \quad (4.42)$$

donde c y ϕ_1 son constantes y $\phi_1 \in (-1, 1)$. Por su parte a_t son las innovaciones o ruido blanco.

Así, el proceso AR(1) expresa el valor de la observación actual como una función lineal de un valor pasado más un término de error aleatorio.

De manera general, un proceso **AR(p)** se expresa como sigue:

$$x_t = c + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + a_t \quad (4.43)$$

donde c y $\phi_1 \dots \phi_q$ son constantes.

Por otro lado, se conoce como **proceso de media móvil** de orden 1 o **MA(1)** el proceso que admite la siguiente representación:

$$x_t = c + a_t + \theta_1 a_{t-1} \quad (4.44)$$

donde c y θ_1 son constantes a determinar. En este caso, el valor actual de la serie se expresa como una función lineal de un valor pasado de un proceso de ruido blanco (a_{t-1}) más un término de error aleatorio. Influye por tanto la variación que existe en un instante anterior.

Un proceso de media móvil de orden q o **MA(q)** se expresa como:

$$x_t = c + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} \quad (4.45)$$

donde de nuevo c y $\theta_1, \dots, \theta_q$ son constantes a determinar.

Otra clase muy útil de modelos de series temporales son los Modelos Autoregresivos de Media Móvil o modelos **ARMA (p,q)** (**Autoregressive Moving Average models**), formados por la combinación de los procesos AR(p) y MA(q). Un proceso ARMA(p,q) se expresa de la siguiente forma:

$$x_t = c + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} \quad (4.46)$$

donde c, ϕ_1, \dots, ϕ_p y $\theta_1, \dots, \theta_q$ son constantes y a_t las innovaciones.

De manera más sencilla, si se emplea B como el operador retardo que viene definido por $Bx_t = x_{t-1}$ y si se define el operador autorregresivo de orden p y el operador de media móvil de orden q como $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ y $\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$ respectivamente, un proceso ARMA(p,q) se define como:

$$\phi(B)x_t = c + \theta(B)a_t \quad (4.47)$$

Además, la constante c guarda relación con la media de la serie (μ), ya que $c = \mu(1 - \phi_1 - \dots - \phi_p)$. Así, cuando la serie solamente posee parámetros de media móvil, la constante coincidirá con la media de dicha serie, no ocurriendo lo mismo cuando posee parámetros autorregresivos.

Para determinar el orden de (p,q) es necesario fijarse en el PACF y ACF muestrales ($\hat{\alpha}_k$ y $\hat{\rho}_k$ respectivamente).

Es preciso también señalar que todos estos modelos son estacionarios (para más información ver Box, Jenkins y Reinsel, 1994). Cuando no se cumple, y la media no es constante, la varianza no es estable o existe autocorrelación entre las observaciones se debe diferenciar la serie para poder convertirla en estacionaria, ya que no se puede trabar con series que presenten tendencia o estacionalidad (patrones repetitivos). De ahí surgieron los Modelos Autorregresivos Integrados de Media Móvil, más comúnmente conocidos como **modelos ARIMA** (Autoregressive Integrated Moving Average).

El primer paso cuando se dispone analizar una serie de tiempo es conocer si ésta es estacionaria. El incumplimiento de esta hipótesis puede ser en media o en varianza.

Si la serie no es estacionaria en varianza los datos son heterocedásticos (varianzas no constantes) y se puede solucionar a través de una transformación Box-Cox de los datos, siendo común aplicar logaritmos.

Por su parte, si la serie no es estacionaria en media y se identifica la existencia de tendencia, ésta se debe diferenciar (en la parte regular), realizando cuantas diferencias sean necesarias (normalmente basta con hacer una diferencia $d = 1$).

Para comprobar la existencia de tendencia basta con ver el gráfico secuencial de la serie así como el ACF. Si el ACF muestral se observan correlaciones muy elevadas (cercanas a 1 para los primeros retardos) y éstas decrecen muy lentamente será señal de tendencia.

Una vez se realicen, si son necesarias, estas transformaciones, la serie ya será estacionaria y por tanto se podrá modelizar a través de un ARMA.

Un proceso ARIMA (p,d,q) se expresa como:

$$\phi(B)(1 - B)^d x_t = c + \theta(B)a_t \quad (4.48)$$

En el caso de los modelos ARIMA la interpretación de la constante c varía un poco con respecto a los procesos ARMA, puesto que aunque $c = \mu(1 - \phi_1 - \dots - \phi_p)$, esa μ no es la media de la serie, sino la media del proceso diferenciado.

Una vez se identifique el modelo tentativo que se vaya a ajustar a la serie es necesario estimar los parámetros del modelo. En este caso los parámetros σ_a^2 , c , ϕ_1, \dots, ϕ_p y $\theta_1, \dots, \theta_q$ se van a estimar por máxima verosimilitud. De este modo, se obtendrán las estimaciones $\hat{\sigma}_a^2$, \hat{c} , $\hat{\phi}_1, \dots, \hat{\phi}_p$ y $\hat{\theta}_1, \dots, \hat{\theta}_q$ que maximicen la función de verosimilitud.

Los modelos tentativos para la serie pueden ser comparados mediante el criterio de información de Akaike (AIC) o el criterio de información bayesiano (BIC), prefiriéndose aquellos que proporcione valores inferiores para dichos criterios.

Una vez tengamos el modelo ajustado con todos sus parámetros estimados, se debe **validar** si es correcto suponer que la serie ha sido generada por ese proceso. Para ello se consideran los residuos del ajuste, ya que si el ajuste es adecuado éstos se comportarán como ruido blanco.

El primer paso es comprobar si existe independencia entre ellos, y para dicho propósito aquí se va a utilizar el contraste de Ljung-Box. Si se rechaza la hipótesis nula de independencia el modelo no es adecuado y se deberá proponer otro modelo tentativo.

Posteriormente se comprobará si la media de los residuos es igual a 0 (hipótesis nula), es decir si $\mu_a = 0$, hipótesis también fundamental, puesto la media del ruido blanco debe ser 0.

También se debe contrastar la normalidad de los residuos (H_0 : los residuos siguen una distribución normal), aunque si es rechazada (siendo las anteriores hipótesis satisfechas) el ajuste se puede considerar adecuado.

- **Gráficos de control basados en los residuos**

Como ya se ha dicho, cuando las observaciones de una serie están autocorreladas no se pueden construir gráficos de control directamente sobre las mismas, ya que se produciría un elevado número de falsas alarmas. De ahí la necesidad de modelizar la serie a través de un modelo de serie temporal como los que se han explicado. Si el modelo es válido los residuos, que serán estacionarios e incorrelados, podrán ser utilizados para estimar la media y la desviación típica y de este modo también los límites de control.

El primer paso a realizar es observar si en la serie existe tendencia o estacionalidad ya que si es así se ajustará un modelo ARIMA, diferenciando para convertirla en estacionaria. Este modelo no se aplicará a la serie entera sino a solamente a una parte de la misma, que formará la muestra de calibrado, necesaria para establecer los límites de control y con ellos monitorizar el resto de la muestra.

De este modo, se selecciona una parte de la muestra y se sugiere un modelo tentativo obteniéndose las estimaciones de los parámetros y con esto se puede realizar la diagnosis de los residuos. Si el modelo es adecuado en función de las hipótesis básicas que deben ser verificadas se puede proceder a calcular los límites de control con esa muestra de calibrado, tanto para los gráficos de medidas individuales como para los gráficos EWMA.

Así, se construye el gráfico de control y se comprueba la existencia de fuera de control. Si se encontrasen causas asignables a los fuera de control se deberían recalculan los límites de control eliminando dichas observaciones del proceso. Si no se encuentran causas asignables y se suponen que son debidas a la propia variabilidad del proceso se dejarían en la serie.

Una vez se tienen los límites de control calculados sobre los residuos de la muestra de calibrado, se pasa a estimar los residuos de la muestra de monitorizado a partir del modelo ARIMA que se ha ajustado anteriormente. Una vez se obtengan dichas innovaciones, se pasaría a con-

trolar esta muestra para comprobar si el proceso está realmente bajo control o no.

Los puntos fuera de control que se muestren en estos gráficos sobre los residuos indicarán que la observación original está fuera de control.

Además, en los casos en los que no pueda ser aceptada la hipótesis de normalidad sobre los residuos, será mejor utilizar los gráficos EWMA en vez de los gráficos de medidas individuales (ya se ha señalado en la Sección sobre los gráficos de control que funcionan mejor en estas situaciones).

4.3. Análisis de Componentes Principales

En esta Sección se introducirán los aspectos más importantes del análisis de componentes principales (PCA por sus siglas en inglés), pues será posteriormente utilizada para realizar un pequeño análisis exploratorio en los datos referentes a Mondoñedo (concretamente los recogidos en verano).

El análisis de componentes principales es una técnica que busca **reducir la dimensión** de un conjunto de variables que se encuentran muy relacionadas entre ellas, transformando dichas variables en componentes principales. En concreto, las componentes principales no son más que combinaciones lineales de las p variables originales (x_1, \dots, x_p) y que tratan de mantener la mayor variabilidad posible presente en los datos (Peña, 2002). Por eso la aplicación del PCA es útil cuando se utilizan datos con un gran número de variables correlacionadas entre ellas y se quiere reducir la dimensión para eliminar la información redundante.

Se busca entonces transformar las covariables originales $X = (X_1, \dots, X_p)$ en un conjunto de nuevas variables incorreladas (componentes) $z = (z_1, \dots, z_p)$ que sean combinación lineal de las antiguas.

Se define el vector de medias como $\mu = E(x)$ y la matriz de varianzas covarianzas de las variables originales como $S = Cov(x, x)$. La primera componente se construirá como:

$$z_1 = a_{11}x_1 + \dots + a_{p1}x_p = a_1'x \text{ con } a_1 = (a_{11} + \dots + a_{p1})' \quad (4.49)$$

Y además la varianza muestral de z_1 (dada por $Var(z_1) = a_1'Sa_1$) debe ser la máxima posible, pero para dicha varianza no crezca indefinidamente se impone que $a_1'a_1 = 1$. Para maximizar la varianza, sujeto a esa restricción, se va a aplicar el método de Multiplicadores de Lagrange, obteniendo finalmente $Sa_1 = \lambda a_1$.

De lo anterior se deriva que a_1 es un autovector (o vector propio) de la matriz de varianzas

covarianzas S y que λ es su autovalor (o valor propio). Además λ , que es el mayor autovalor de la matriz S , es la varianza de z_1

Por otra parte, de la variabilidad que queda sin explicar por el primer componente, se calcula el segundo componente principal z_2 de manera que la varianza explicada sea la mayor posible, garantizando que ambas sean incorreladas.

Además la suma de las varianzas de los p componentes principales va a ser la misma que la suma de las varianzas de las variables originales, esto es:

$$Var(x_1) + \dots + Var(x_p) = \lambda_1 + \dots + \lambda_p = traza(S) \quad (4.50)$$

Y la proporción de variabilidad explicada por cada componente va a venir dada por $\frac{\lambda_j}{traza(S)}$, $j = (1, \dots, p)$.

También se define la covarianza entre un componente principal y las variables x como:

$$Cov(z_i, x_j) = \lambda_j a_{j,i} \quad (4.51)$$

Y asimismo la correlación entre las variables originales y los componentes como:

$$\rho(z_i, x_j) = \frac{Cov(z_i, x_j)}{\sqrt{Var(z_i) + Var(x_j)}} = \frac{\lambda a_{j,i}}{\sqrt{\lambda_i s_j^2}} \quad (4.52)$$

En la prácticas, muchas veces se tienen variables con escalas de medida distintas y, en estas situaciones, es más conveniente calcular los valores y vectores propios a través del coeficiente de correlación. En este trabajo se calcularán las componentes principales a partir de la matriz de correlaciones (R) pues las variables a analizar están medidas en escalas muy distintas (ya que se tiene temperatura media, contenido en CO_2 , humedad relativa media y ocupación).

Además, se van a calcular las comunalidades, que se utiliza en el análisis factorial. Sin entrar en detalles, se calcularán para conocer el porcentaje de varianza que explican las componentes principales sobre cada una de las variables originales (el valor será 1 si se consideran todas las componentes). Se obtienen sumando los cuadrados de las correlaciones para cada variable.

Se debe también tener en cuenta el número de componentes que se seleccionan de modo que éstos expliquen lo mejor posible la variabilidad de los datos. No existe una única norma que determine cuántos han de retenerse, pero las técnicas más conocidas son:

- Quedarse con un número de componentes que expliquen un porcentaje alto de la variabilidad (en torno al 80 % o 90 %).

- Realizar un gráfico con los autovalores ordenados de mayor a menor y retener los componentes a partir de los cuales cambie la curvatura de dicho gráfico.
- Seleccionar los componentes cuyos autovalores superen el valor medio de todos los autovalores, esto es, $\frac{\text{traza}(S)}{p} = \frac{\sum \lambda_i}{p}$ (Peña, 2002). Esta es la técnica que se utilizará para la elección del número de componentes en este trabajo.

4.4. Análisis de Datos Funcionales

Se introduce en esta Sección un breve reseña sobre las técnicas de análisis de datos funcionales que se van a utilizar para realizar un estudio descriptivo de la base de datos de Fridama. En concreto, ya que en los datos sobre la empresa se tienen mediciones tanto en días festivos como laborables, se desea conocer si existen diferencias entre ambas categorías pues entonces se construirán gráficos de control por separado. Para ello se sigue el procedimiento desarrollado por Febrero-Bande y Oviedo de la Fuente (2012), utilizando la librería de R `fd`.

En análisis de datos funcional (también denominado FDA por sus siglas en inglés) es una técnica muy utilizada para trabajar con grandes bases de datos en las que las observaciones se toman a largo de un intervalo continuo, normalmente el tiempo. En la práctica, se tiene un conjunto de observaciones discretas y, como resultado, se trabajará con un conjunto de curvas, donde cada una, relativa a una muestra en concreto, ha sido evaluada en un número determinado de puntos.

Más concretamente, una **variable funcional** es una variable aleatoria que toma valores en un espacio funcional ε y un dato funcional es una observación χ de esa variable. De este modo, el conjunto de **datos funcionales** de dimension n , $\{\chi_1, \chi_2, \dots, \chi_n\}$ es la observación de n variables funcionales X_1, X_2, \dots, X_n idénticamente distribuidas con χ .

Como ya se ha dicho, el conjunto de datos funcionales, que también se puede denotar como $X_1(t), X_2(t), \dots, X_n(t)$ va a ser evaluado en una serie de puntos $\{t_1, t_2, \dots, t_m\}$.

Lo primero que se debe hacer obtener la forma funcional de las curvas muestrales utilizando las observaciones discretas disponibles. Para representar los datos funcionales se utilizarán en este trabajo las técnicas de *estimación no paramétrica* (para más información ver Ferraty y Vieu, 2006).

En concreto se utilizará el método de suavizado kernel, basado en la función kernel (K) y en el parámetro h de suavizado, llamado parámetro ventana o ancho de banda (bandwidth en inglés). De este modo, la suavización no paramétrica viene dada por la matriz de suavizado

$S = s_{ij}$, que se calculará utilizando el método de Nadaraya-Watson, definido como:

$$s_j(t_i) = \frac{K\left(\frac{t_i - t_j}{h}\right)}{\sum_{k=1}^m K\left(\frac{t_j - t_k}{h}\right)} \quad (4.53)$$

Asimismo, para seleccionar el parámetro h ventana óptimo que mejor representa los datos funcionales de los que se disponen se utilizará la validación cruzada generalizada.

Una vez se tengan representadas las curvas funcionales suavizadas, se calcularán las medias funcionales para cada conjunto de datos utilizando bootstrap suavizado. Con este se establecen las bandas de confianza (en este caso se utilizará un nivel de confianza del 95 %) y se comparará la trayectoria para cada tipo de día.

Capítulo 5

Resultados: Mondoñedo

5.1. Análisis Exploratorio de los Datos

Como se ha mencionado anteriormente en el Capítulo 3 en referencia a la base de datos de Mondoñedo, primeramente se va a analizar la base de datos relativa a la época estival de los tres edificios de Mondoñedo, y se comenzará realizando un pequeño análisis descriptivo que muestre las características principales de los datos. Así, se obtiene el siguiente resumen numérico:

	Media	Desv. Típ.	IQR	Cuantil 0	Q1	Q2	Q3	Cuantil 1
temp.media	21.91	1.79	1.43	6.87	21.27	21.80	22.70	25.07
humedad.rel.media	66.17	6.69	5.88	22.83	64.49	66.96	70.37	76.15
co2.inicial	525.10	169.72	125.75	325.00	428.00	471.50	553.75	1555.00
co2.final	525.13	169.50	122.00	370.00	430.00	470.00	552.00	1586.00
co2.medio	520.47	166.82	123.62	144.50	428.62	470.59	552.25	1526.33
ocupacion	3.70	6.28	4.00	0.00	0.00	0.00	4.00	27.00

Cuadro 5.1: Resumen descriptivo con las funciones más comunes.

En el cuadro anterior se puede ver que las variables más dispersas son las relacionadas con la cantidad de CO_2 y donde el rango intercuartílico es también mayor. Por su parte, la temperatura media es la menos dispersa.

Mediante el cálculo y representación de los histogramas de la Figura 5.1 se obtienen los rangos de la temperatura media, CO_2 , humedad relativa media y ocupación más frecuentes.

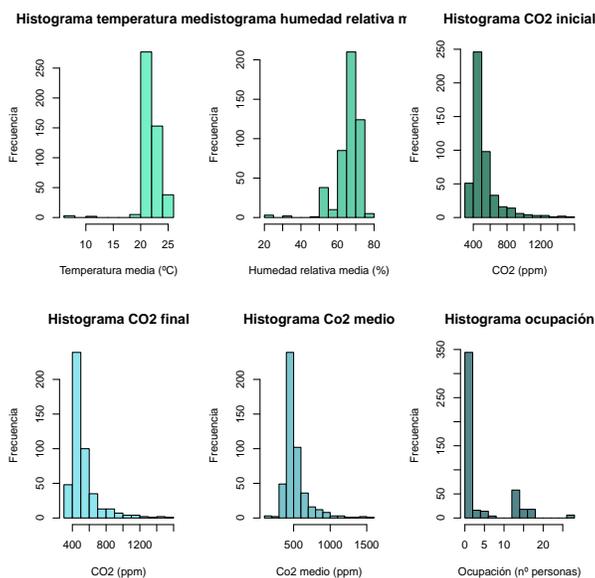


Figura 5.1: Histogramas de las diferentes variables

A primera vista, observando los histogramas, se aprecia que las variables pueden no seguir una distribución normal, aunque esto se puede contrastar con el test de Shapiro Wilks. Haciendo dicho contraste, se obtienen los siguientes p-valores, que claramente permiten rechazar la hipótesis nula de normalidad:

	Temp. media	Humedad rel. media	CO2 inicial	CO2 final	CO2 medio	Ocupación
P-valores	0.00	0.00	0.00	0.00	0.00	0.00

Cuadro 5.2: P-valores resultantes del test de Shapiro Wilks.

Una vez comprobado esto, se retomará la interpretación de los histogramas. De este modo, comenzando con la temperatura media, parece haber presencia de valores atípicos, ya que se encuentran observaciones con un nivel muy bajo de temperatura. Estos valores (por debajo de 20°C, que hacen un total de 10 observaciones) se producen sobre todo por la mañana (a las 8, 9, 10 y 11) y todas ellas el día 23 de septiembre, y, como se puede ver en el Cuadro 5.3, en varias de las instalaciones.

No obstante, se observa una clara diferencia, y es que en la Escuela de Música estos valores atípicos se producen a la misma hora (11 de la mañana), mientras que en el resto de edificios varia entre las 8 y las 10. El día 23 se producen en Mondoñedo temperaturas algo más bajas que los días anteriores, pues se ha comprobado con los datos de MeteGalicia para dichas fechas, como se ve en la Figura 5.2, pero todo sugiere que el instrumento de medición de la Escuela de Música fallase por algún motivo o el sistema de climatización no funcionase (ya que las temperaturas

son mucho más bajas que en el resto de edificios, llegándose a registrar una mínima de 6,87°C en el aula 2), o bien que se incurriera en errores en el traspaso o toma de datos en dicha Escuela (puesto que precisamente esas observaciones se corresponden con aquellas donde aparece un valor perdido en el CO_2 final).

	Fecha	Hora_inicial	Sala	Temperatura	Ocupacion
1	23/09/2011	10	ayuntamiento pasillo	19.25	13.00
2	23/09/2011	10	ayuntamiento rehabilitacion	19.83	13.00
3	23/09/2011	8	casa cultura gimnasio	19.90	0.00
4	23/09/2011	9	casa cultura gimnasio	19.90	3.00
5	23/09/2011	10	casa cultura gimnasio	19.90	4.00
6	23/09/2011	11	musica aula 4	10.80	2.00
7	23/09/2011	11	musica aula 1	6.90	2.00
8	23/09/2011	11	musica cabina	7.20	2.00
9	23/09/2011	11	musica aula 2	6.87	2.00
10	23/09/2011	11	musica pasillo	10.75	2.00

Cuadro 5.3: Valores fuera del rango de confort térmico de la temperatura media.

Resultados da consulta para: Vilamor (Lugo)			
Cod. Validación	Data	Parámetro (Unidades)	Valor
1	21/09/2011	Temperatura media (°C)	18,5
1	22/09/2011	Temperatura media (°C)	18,4
1	23/09/2011	Temperatura media (°C)	17,5

Figura 5.2: Temperaturas medias recogidas en la estación de Vilamor, Mondoñedo, para los días analizados. Fuente: MeteoGalicia.

Una situación similar ocurre con la variable que mide la humedad relativa media, donde se encuentran valores extrañamente bajos. Estos valores atípicos, al igual que en el caso anterior, podrían indicar una deficiencia en el confort térmico, por eso es importante conocer cuándo se produjeron. Ya que los límites de confort en lo referente a la humedad se sitúan en torno al 40 %-70 %, se señalan en el Cuadro 5.4 las observaciones por debajo de ese límite inferior.

Se observa el mismo patrón que anteriormente, las observaciones atípicas provienen de la Escuela de Música (a las 11 de la mañana y en la misma fecha), con índices muy bajos de humedad. Como se ha señalado anteriormente, una de las causas podría ser un fallo en el sistema de climatización del edificio.

A continuación tenemos las variables referentes a la cantidad de CO_2 , tanto inicial como final, cuyos histogramas muestran un patrón muy similar. Por ello, se contrasta con el test de Kolmogorov-Smirnov para dos muestras si el CO_2 inicial y el CO_2 final provienen de la misma

	Fecha	Hora_inicial	Sala	Humedad	Ocupacion
1	23/09/2011	11	musica aula 4	32.50	2.00
2	23/09/2011	11	musica aula 1	24.52	2.00
3	23/09/2011	11	musica cabina	22.83	2.00
4	23/09/2011	11	musica aula 2	23.60	2.00
5	23/09/2011	11	musica pasillo	34.45	2.00

Cuadro 5.4: Valores fuera del rango de confort térmico de la humedad relativa media.

distribución, ya que si es así, se estudiará únicamente el CO_2 medio. El resultado de dicho test arroja un p-valor de 1, por lo que se acepta que tengan la misma distribución.

De este modo, se tendrá en consideración solamente el CO_2 medio para posteriores cálculos. Los límites de confort térmico referentes al CO_2 se sitúan en un máximo de 900 partes por millón (ppm), observando en el histograma valores por encima de esta cifra. En el Cuadro que se muestra a continuación se señalan dichas observaciones:

	Fecha	Hora_inicial	Sala	CO2_medio	Ocupacion
1	21/09/2011	15	ayuntamiento rehabilitacion	958.50	0.00
2	21/09/2011	12	casa cultura gimnasio	1044.33	5.00
3	21/09/2011	13	casa cultura gimnasio	944.00	5.00
4	22/09/2011	19	casa cultura gimnasio	985.17	5.00
5	22/09/2011	20	casa cultura gimnasio	979.67	7.00
6	21/09/2011	17	musica aula 1	1452.17	16.00
7	21/09/2011	18	musica aula 1	1194.33	16.00
8	22/09/2011	17	musica aula 1	1526.33	13.00
9	22/09/2011	18	musica aula 1	1492.00	13.00
10	22/09/2011	19	musica aula 1	1332.00	13.00
11	22/09/2011	20	musica aula 1	1244.17	13.00
12	22/09/2011	21	musica aula 1	1181.67	13.00
13	22/09/2011	22	musica aula 1	1130.00	0.00
14	22/09/2011	23	musica aula 1	1077.17	0.00
15	23/09/2011	0	musica aula 1	1032.67	0.00
16	23/09/2011	1	musica aula 1	990.50	0.00
17	23/09/2011	2	musica aula 1	955.33	0.00
18	23/09/2011	3	musica aula 1	925.00	0.00
19	23/09/2011	4	musica aula 1	901.50	0.00

Cuadro 5.5: Valores fuera del rango de confort térmico del CO_2 medio.

Se encuentran un total de 19 observaciones donde no se cumplen los criterios de confort térmico, por ser la emisión de este gas mayor a 900 ppm. Como puede observarse, la mayoría

de los casos por encima de este valor ocurren en el Gimnasio de la Casa de Cultura (los días 21 y 22 a diferentes horas) y sobre todo en la Escuela de Música (en el aula 1, los tres días en los que se tomaron observaciones y en distintas horas).

En el caso del Gimnasio parece que los valores altos de CO_2 se dan cuando la ocupación es mayor (la ocupación máxima registrada en la Casa de Cultura es de 7 personas). Una razón puede ser que dicha sala tenga un tamaño demasiado reducido, por lo que la concentración de CO_2 es mayor que en otras estancias.

En cuanto al aula 1 de la Escuela de Música se observa una situación diferente. El día 21 se registran altos valores de CO_2 que pueden ser debidos a la elevada ocupación de dicho aula. Por su parte, el día 22 ocurre algo similar, a partir de las 17 horas los valores de CO_2 son bastante elevados y van disminuyendo progresivamente, registrándose valores aun por encima de los límites de confort térmico incluso sin personas en el edificio (en horas donde éste se encuentra cerrado). Podría ocurrir que al darse valores tan elevados por la tarde, tardasen en bajar de 900 ppm hasta las 5 de la mañana, quizá también propiciado por el tamaño del aula (o el sistema de ventilación de la misma), ya que en otras estancias de la Escuela no ocurre lo mismo.

Por último, referente a la ocupación se puede observar la presencia de dos poblaciones diferentes, una caracterizada por bajos niveles y otra, menos frecuente, por elevados niveles de ocupación. Se construye una tabla con las observaciones en las que la ocupación es mayor a 10 personas, obteniendo un total de 100 datos (por lo que los casos en los que la ocupación es menor a 10 es más frecuente, tal como se observa en el histograma, con un total de 378 observaciones).

En primer lugar señalar que en la Casa de Cultura no hay ningún caso donde la ocupación sea mayor que 10 personas (ya se mencionó anteriormente que se registró un máximo de solamente 7 personas). Además, se extrae algo claro y es que en el Ayuntamiento la ocupación se produce solamente en horario matutino (de 8 a 15 horas), siendo 0 desde las 15 horas a las 8 del día siguiente, seguramente por permanecer cerrado durante este periodo, por lo que las gestiones se deban realizar siempre de mañana.

En la Escuela de Música ocurre exactamente lo contrario, los picos de mayor ocupación suceden en horario de tarde, algo normal por tratarse de actividades complementarias que se suelen realizar una vez terminada la jornada laboral o escolar.

5.1.1. Análisis de Componentes Principales

Cuando tenemos una base de datos con un relativamente elevado número de variables en comparación con el número de observaciones, es conveniente realizar un estudio descriptivo para identificar y caracterizar la relación entre ellas. De hecho, es común encontrar correlaciones entre las variables, mostrando así la existencia de información redundante. Aquí es donde radica el sentido de utilizar el Análisis de Componentes Principales, para reducir la dimensión de un

conjunto de datos, pero manteniendo la mayor cantidad de información posible presente en los datos originales.

A partir de las variables originales se calcularán las Componentes Principales, que ya se ha señalado en la Sección 4.3 que son la combinación lineal de esas antiguas variables, de modo conservan toda la información incluida en las primeras.

Así, el primer paso consiste en identificar y medir la asociación existente entre las variables relacionadas con el confort térmico tomadas en Mondoñedo. Para ello se utiliza el coeficiente de correlación por rangos de Spearman (puesto que en el caso a estudiar las variables no son normales y éste es una medida de asociación más robusta), midiendo así la relación entre las variables, que podemos comprobar con los siguientes resultados:

	temp.media	humedad.rel.media	co2.medio	ocupacion
temp.media	1.00	-0.71	0.11	0.11
humedad.rel.media	-0.71	1.00	0.36	-0.02
co2.medio	0.11	0.36	1.00	0.20
ocupacion	0.11	-0.02	0.20	1.00

Cuadro 5.6: Matriz de correlaciones con el método de Spearman.

A primera vista las correlaciones más altas hacen referencia a la relación entre la humedad relativa media y la temperatura media, con un coeficiente de correlación negativo (-0.71), lo que quiere decir que cuando aumenta la temperatura media disminuye la humedad y viceversa. También encontramos valores elevados en la relación entre el CO_2 y la humedad relativa media, el CO_2 y la ocupación. Para comprobar si la correlación es significativa o no vamos a realizar un contraste conjunto con todas variables, obteniendo los siguientes p-valores:

	temp.media	humedad.rel.media	co2.medio	ocupacion
temp.media		0.00	0.02	0.01
humedad.rel.media	0.00		0.00	0.63
co2.medio	0.02	0.00		0.00
ocupacion	0.01	0.63	0.00	

Cuadro 5.7: P-valores resultantes del test de correlación con el método de Spearman.

Comenzando con la temperatura media, se confirma su correlación (negativa, como se señaló anteriormente) con la humedad relativa media (con un p-valor de 0 se rechaza la hipótesis nula de incorrelación). Asimismo, también está se prueba su correlación con el CO_2 medio y con la ocupación. La humedad relativa media está, por su parte, correlacionada con el CO_2 medio, pero no con la ocupación (con un p-valor de 0.63 se acepta la hipótesis nula de incorrelación).

Por último, entre el CO_2 medio y la ocupación se acepta, con un p-valor de 0, la hipótesis de correlación entre ambas.

Todas estas relaciones ayudan a identificar las causas asignables de la falta de confort térmico, relacionadas con el nivel de ocupación y el aumento de la humedad.

El siguiente paso, una vez confirmada la existencia de correlación entre varias de las variables, es aplicar el Análisis de Componentes Principales sobre los datos y posteriormente se escogerá el número de componentes a estudiar y cómo se forma cada una de ellas.

Se utiliza la función `princomp` de la librería `stats` para realizar el análisis, con el que se obtiene el siguiente resumen:

	Comp.1	Comp.2	Comp.3	Comp.4
Desviación_estándar	1.20	1.01	0.96	0.79
Proporción_de_varianza	0.36	0.25	0.23	0.15
Proporción_acumulada	0.36	0.62	0.85	1.00

Cuadro 5.8: Resumen del modelo de ACP sobre las variables a analizar.

En él se ven varios elementos importantes. En primer lugar, se muestra la desviación estándar de cada componente así como la proporción de varianza de nuestros datos que explica cada componente y por último la proporción acumulada.

La proporción de varianza explicada y la acumulada son importantes para decidir con cuántas componentes nos quedamos, teniendo en cuenta que queremos el menor número de éstas pero explicando un porcentaje alto de variabilidad. En este caso, si nos quedamos con las dos primeras, explicamos un 61.65 % de la variabilidad, mientras que si añadimos la tercera también explicamos un 84.5 %.

También se pueden ver los autovalores calculando la varianza de los componentes:

	Comp.1	Comp.2	Comp.3	Comp.4
Autovalores	1.20	1.01	0.96	0.79

Cuadro 5.9: Autovalores o varianza de las componentes principales del modelo.

Como también se ha señalado en la Sección 4.3, existen diversos criterios para elegir el número de componentes, optando aquí por el criterio con el cual se retienen las componentes cuyos autovalores superen el valor medio de todos los autovalores, es decir, que sean superiores a 1. Así, en este caso, se eligen las dos primeras, que son la que se analizan a continuación.

Una vez hecho esto, en el siguiente paso se analizarán los autovectores, donde se muestra la correlación de cada componente principal con cada variable original.

La primera componente está correlacionada negativamente con las variables, y las corre-

	Comp.1	Comp.2	Comp.3	Comp.4
temp.media	-0.41	0.23	0.87	0.16
humedad.rel.media	-0.47	-0.69	-0.14	0.53
co2.medio	-0.67	-0.06	-0.16	-0.72
ocupacion	-0.41	0.68	-0.44	0.42

Cuadro 5.10: Autovectores del modelo de ACP sobre las variables a analizar.

laciones más acusadas tienen que ver con la presencia de CO_2 , valores negativos de esta componente los tendrán aquellos entornos más contaminados (correlación negativa). Esta primera componente se obtiene con la siguiente combinación:

$$z_1 = -0,41 * \widehat{temp.media} - 0,47 * \widehat{humedad.rel.media} - 0,67 * \widehat{co2.medio} - 0,41 * \widehat{ocupacion}$$

Se debe tener en cuenta que $\widehat{temp.media}$, $\widehat{humedad.rel.media}$, $\widehat{co2.medio}$ y $\widehat{ocupacion}$ contienen las puntuaciones estandarizadas individuales de cada una de esas variables (ya que se trabaja con la matriz de correlaciones).

La segunda componente tiene mayor correlación positiva con la variable ocupación y correlación negativa con la humedad relativa media, por lo que valores positivos de esta componente los obtendrían las instalaciones cuya ocupación es mayor y la humedad relativa media es inferior. Así, se obtiene:

$$z_2 = 0,23 * \widehat{temp.media} - 0,69 * \widehat{humedad.rel.media} - 0,06 * \widehat{co2.medio} + 0,68 * \widehat{ocupacion}$$

A continuación se muestra de manera gráfica las puntuaciones sobre las observaciones de la muestra obtenidas, tanto con etiquetas (donde cada una muestra una observación), como en nube de puntos (diferenciando por los tres edificios).

Se observa la existencia de observaciones que se alejan y, precisamente, como se comprueba con ambas Figuras, coinciden con las cinco observaciones de la Escuela de Música que se señalaron como fuera de los límites de confort térmico de la temperatura media (ver Cuadro 5.3).

Otro gráfico significativo son los llamados **biplot**, en los cuales se muestran a la vez las observaciones y las variables de la base de datos (las observaciones son los puntos del gráfico mientras que las variables se representan mediante flechas) en un gráfico bidimensional de la primera componente y la segunda.

Con este gráfico podemos ver el significado de cada componente, observando las direcciones de las flechas. Vemos la correlación entre dos variables teniendo en cuenta el ángulo que forman

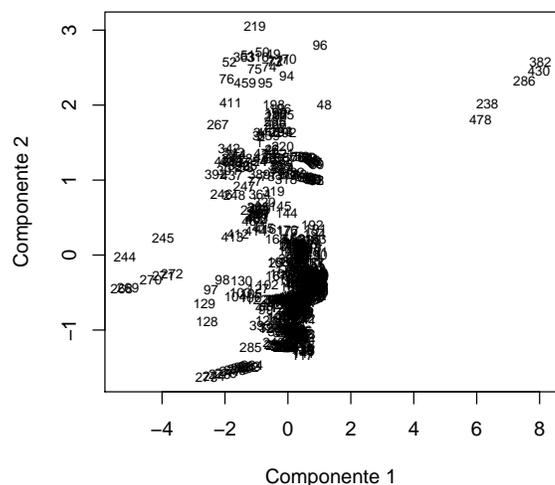


Figura 5.3: Gráfico de las puntuaciones obtenidas mediante el ACP para las componentes 1 y 2 con las etiquetas de las observaciones.

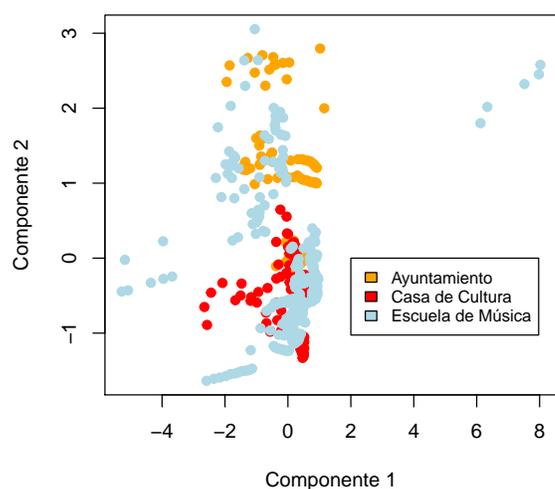


Figura 5.4: Gráfico de las puntuaciones obtenidas mediante el ACP para las componentes 1 y 2 diferenciando por edificio

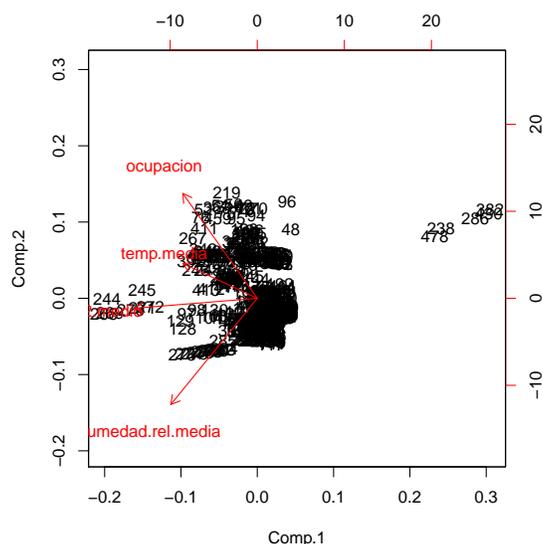


Figura 5.5: Biplot con las componentes 1 y 2

los dos vectores que correspondan a dichas variables. Asimismo, la varianza de cada una de esas variables se valora por la longitud de los vectores desde el origen. Igualmente, las observaciones que se encuentren próximas se interpretan como similares.

En este gráfico llegamos a la misma conclusión que con los autovectores, y es el CO_2 está correlacionado negativamente con la componente 1 (ya que los vectores son paralelos al eje de esa componente), mientras que la ocupación y la temperatura media se asocian a la componente 2 (de manera positiva) y de manera negativa la humedad relativa media.

Asimismo se puede ver que cuanto más humedad menos temperatura y ocupación en el edificio y que a mayor ocupación también mayor temperatura.

Es interesante comparar el biplot de la Figura 5.5 con el gráfico de la Figura 5.4, donde se distinguen las observaciones por edificios, y no es complicado ver la existencia grupos diferenciados.

Por una parte se encuentra el Ayuntamiento, caracterizado en su mayoría por una alta ocupación y temperatura media (y, consecuentemente, bajos valores de humedad), estando más relacionado de este modo con la componente 2. Al contrario que éste, la Casa de Cultura (correlacionada negativamente con la componente 2) se caracteriza por sus altos valores de humedad y baja ocupación. Tiene unos grandes problemas de humedad, de incorrecta climatización del edificio.

Por último, en la Escuela de Música se encuentra más variedad de valores. Por un lado se

encuentran observaciones muy correlacionadas negativamente con la componente 1 y con valores altos de CO_2 , que se corresponden con aquellas observaciones del aula 1 fuera de los límites de confort térmico (señalados en el Cuadro 7). Por otro lado, también existen observaciones correlacionadas con la componente 2, caracterizados algunos por alta ocupación y temperatura y otros por altos niveles de humedad. Además, se advierte la existencia de cinco valores completamente distintos al resto y en dirección opuesta. Cada una de estas observaciones pertenece a un aula distinta de la Escuela de Música, y están singularizados por tener bajos niveles de temperatura, de ocupación y de CO_2 medio. Estas observaciones se corresponden con las del día 23 del Cuadro 5.3, en el que se comentó la existencia de posibles errores.

Una última parte de este análisis serían las comunalidades de cada una de las diferentes variables. Las comunalidades informan de la proporción de variabilidad de cada variable explicada por un número concreto de componentes. Primero es necesario calcular las correlaciones entre las variables de partida y las componentes principales y después sumar los cuadrados de dichas correlaciones para obtener así las comunalidades.

	temp.media	humedad.rel.media	co2.medio	ocupacion
Comunalidades	0.29	0.81	0.65	0.71

Cuadro 5.11: Comunalidades para las variables analizadas.

Como se observa en el cuadro anterior, las dos primeras componentes explican en mayor medida la variabilidad de la humedad relativa media, con un porcentaje del 80.89%. Por el contrario, solo explican el 29.05% de la varianza de la temperatura media. Si se quiere explicar un porcentaje mayor de la variabilidad de la temperatura media se debería analizar también la tercera componente, pues de este modo se explicaría un 98% de la misma.

Una vez se ha realizado este análisis exploratorio para ver las relaciones entre las variables en la base de datos analizada, se pasará directamente al estudio de la calidad, más concretamente al análisis de la capacidad en relación a las variables aquí comentadas.

5.2. Índices de Capacidad

Una vez descrita la base de datos e identificadas las relaciones entre las características críticas de calidad (Temperatura, CO_2 , humedad y ocupación), el siguiente paso consiste en la evaluación de la capacidad de los edificios estudiados para proporcionar confort térmico. Los límites de confort térmico están marcados por la normativa vigente, como ya se ha mencionado. El análisis estadístico de la capacidad del proceso se llevará a cabo mediante el cálculo de índices de capacidad.

Por tanto, se calcularán los índices que asumen normalidad así como los índices para distri-

buciones arbitrarias (de los que hemos hablado en la Sección 4.1.3). Aunque se ha comprobado que los datos no son normales, se calculan ambos tipos de índices para mostrar la necesidad de aplicar los índices robustos cuando se viola esta suposición.

Para calcular los índices de capacidad para distribuciones normales se utilizará la función `qualityTools`, que calcula los índices C_p , C_{pk} , C_{PL} y C_{PU} y representa un histograma con la curva de densidad normal. Los índices para las distribuciones arbitrarias se han programado siguiendo la formulación correspondiente.

Como los límites de confort térmico solo necesitan ser cumplidos en los periodos de ocupación de los edificios, solamente se tendrán en cuenta dichas franjas horarias. Se analizarán los edificios independientemente y para cada periodo (verano e invierno).

5.2.1. Edificio del Ayuntamiento

5.2.1.1. Estudio de las variables de calidad en verano

- **Índices de capacidad para distribuciones normales**

Se comenzará calculando los índices para distribuciones normales y analizando los datos referentes a verano. Se seleccionan los datos correspondientes al Ayuntamiento en las horas de ocupación (de 8 a 15 horas) y se separan por variables.

Así, se calculan los índices para la temperatura media. El Ayuntamiento tiene una categoría ambiental C, por lo que los límites de consigna inferior y superior son 22° y 27° respectivamente. El target es, en este caso, el punto medio, $24,5^\circ$.

En primer lugar, se obtienen los índices para distribuciones gaussianas (Figura 5.6). Viendo el histograma y el gráfico de cuantiles (QQ plot) parece que los datos no siguen una distribución normal (aplicando el test de Shapiro Wilks solamente a este grupo de datos se obtiene un p-valor en torno a 0, por lo que se rechaza la normalidad).

Tanto con el índice C_p como C_{pk} son menores que 1, por lo que el proceso no es capaz. Además, el valor de estos índices no es similar, lo que significa que el proceso no está centrado en torno a la media (se puede comprobar viendo el histograma). Observando dicho histograma tampoco la media está centrada en torno al Target y encontramos observaciones fuera de los límites de especificación (en concreto por debajo de LSL).

El número esperado de no conformes asumiendo normalidad sería de 114487 partes por millón (111044 por debajo del límite inferior y 3443 por encima del límite superior).

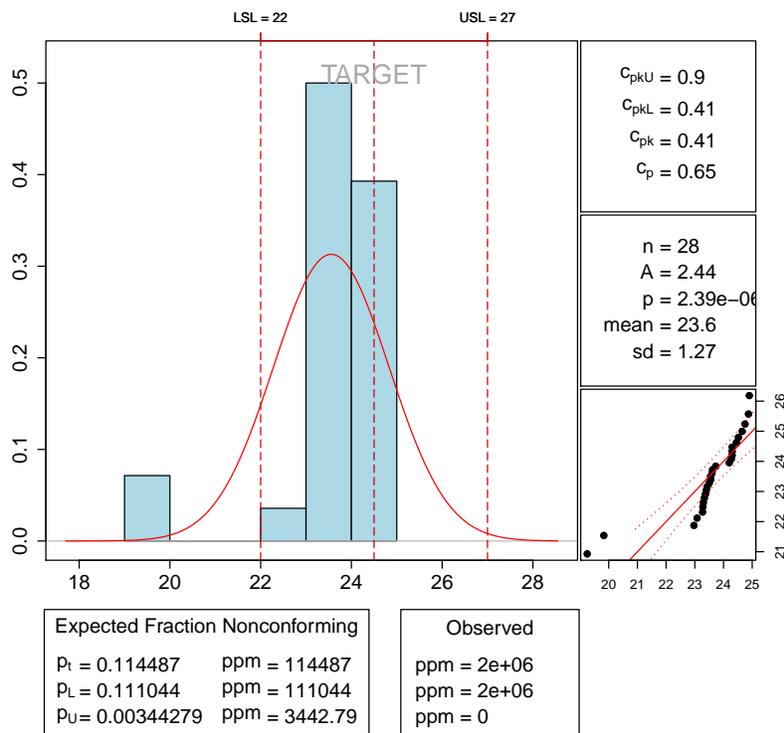


Figura 5.6: Análisis de capacidad normal para la temperatura media en el Ayuntamiento, en verano.

Por su parte, y siguiendo el mismo proceso que en el caso anterior, para la humedad relativa media los límites de consigna son 40 % y 60 %, situándose el Target en el punto medio, esto es, 50 %. Se aplica el test de Shapiro Wilks, obteniendo un p-valor de 0.001061, por lo que no se acepta normalidad. Se obtiene análisis de capacidad que se muestra en el Cuadro 5.12.

Índices de Capacidad	Fracción esperada no conforme	Partes por millón
$C_{pkU} = 0,09$	$p_t = 0,395511$	ppm=395511
$C_{pkL} = 1,07$	$p_L = 0,000676$	ppm=676.14
$C_{pk} = 0,09$	$p_U = 0,394835$	ppm=394835
$C_p = 0,58$		

Cuadro 5.12: Análisis de capacidad normal para la humedad relativa media en el Ayuntamiento, en verano.

Se observa que el proceso no es capaz y que está descentrado en torno a la media (C_p y C_{pk} toman valores distintos). Además, C_{pk} es cercana a 0, lo que significa que la media del proceso está muy próxima a los límites de especificación, en concreto al límite superior en este caso (ya que $USL = 60$ y la media es de 58.5). Por último señalar que la media está completamente descentrada en torno al Target. Asumiendo normalidad el número de no conformes es de 395511 partes por millón.

Por último, para el CO_2 el límite superior de especificación es de 900 partes por millón o ppm (siendo el inferior de 0) y el Target es de 450. Se rechaza la normalidad de la variable CO_2 en este grupo de observaciones puesto que con el test de Shapiro Wilks se obtiene un p-valor de 0.004524. Se obtiene el siguiente análisis de capacidad:

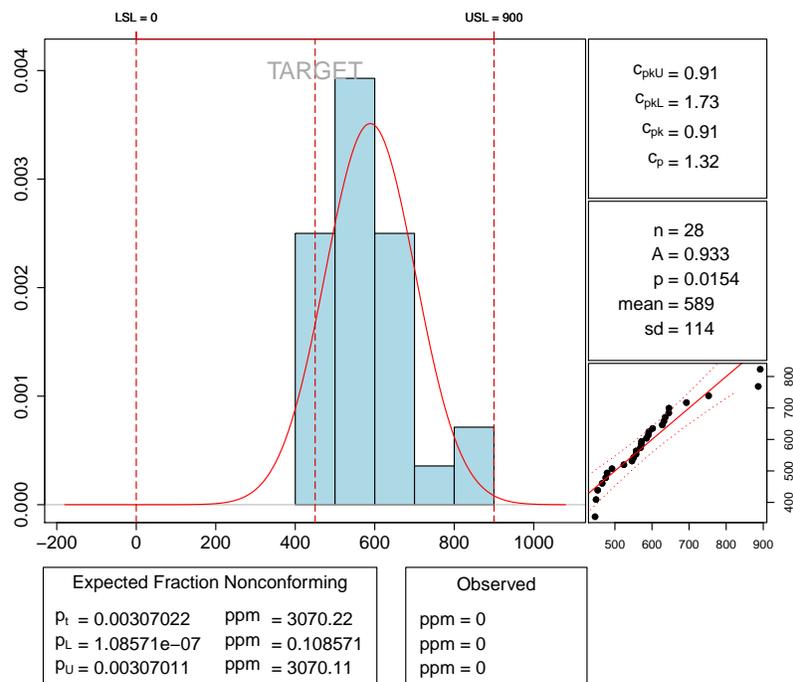


Figura 5.7: Análisis de capacidad normal para el CO_2 medio en el Ayuntamiento, en verano.

Aunque $C_p > 1$, C_{pk} es menor que esa cantidad, por lo que asumiendo normalidad el proceso no se puede decir que sea capaz. Además éste no está centrado en torno a la media, ni la media centrada respecto al Target. Se esperan unas 3070 de observaciones no conformes, asumiendo normalidad en los datos.

Como se ha visto, ninguno de estos tres procesos son normales, por lo que estos índices paramétricos no son adecuados, surgiendo la necesidad de aplicar índices robustos, como los que

se utilizan a continuación.

- **Índices de capacidad para distribuciones arbitrarias**

De este modo, para calcular los índices no paramétricos para la temperatura media se procede en primer lugar a calcular los percentiles, obteniendo $F_{0,135} = 19,27114$ y $F_{99,865} = 24,89891$, y la mediana, $M = 23,575$. Utilizando la ecuación 4.36 y fijando los valores de (u, v) , se obtiene $C_{Np} = 0,89$, $C_{Npk} = 0,56$, $C_{Npm} = 0,63$ y $C_{Npmk} = 0,40$.

Se observa que C_{Npk} es menor que 1, lo cual indica que el proceso no es adecuado con respecto a las especificaciones señaladas. Además, C_{Np} y C_{Npk} no muestran valores similares, por lo que la temperatura media no está centrada con respecto al target. De hecho, hay dos observaciones que salen fuera de los límites de especificación (con $19,25^{\circ}\text{C}$ y $19,83^{\circ}\text{C}$).

Asimismo, C_{Npm} es menor que 1, por lo que la media de la temperatura no está próxima al Target.

Aunque se ve que el sistema de control de la temperatura no es capaz, tanto con índices que asumen normalidad como con índices robustos, los resultados muestran que es menos capaz de lo que en un principio se había supuesto al calcular los índices para distribución normal (C_{Npk} es negativo).

Calculando a continuación los índices para el resto de variables (humedad relativa media y CO_2) medio se obtienen los siguientes índices:

	Temperatura.media	Humedad.rel.media	CO2
C_Np	0.89	1.02	2.02
C_Npk	0.56	0.47	1.48
C_Npm	0.63	0.53	1.05
C_Npmk	0.40	0.24	0.77

Cuadro 5.13: Índices de capacidad de todas las variables para el Ayuntamiento, en verano.

En el Cuadro anterior se muestran los índices para todas las variables. En lo relativo a la humedad relativa media se observa que $C_{Np} > 1$ (en concreto, 1.02), por lo que se podría pensar que el proceso es capaz, pero realmente no es así, ya que $C_{Np} < 1$, por lo que el proceso no está centrado en torno a la mediana. Con el índice C_{Npm} se observa que la mediana está alejada del valor nominal y con C_{Npmk} se puede decir que el proceso no está centrado en torno a dicho Target. En definitiva, el sistema de control de la humedad en el edificio no es capaz.

En cuanto al CO_2 se observan valores diferentes a los anteriores. En primer lugar, $C_{Np} > 2$ y $C_{Npk} > 1,33$, lo que muestra que el proceso es muy capaz. Esto quiere decir que el proceso utiliza menos que el 100 % de la banda de tolerancia natural, por lo que se producirá un número no muy elevado de no conformes. Que ambos índices no tengan el mismo valor significa que el

proceso no está operando en el punto medio del intervalo de las especificaciones.

Asimismo, con el índice C_{Npm} (que es mayor que 1, en concreto 1.05) se afirma que el proceso sí cumple las especificaciones y está centrado en el Target, aunque también se observa que $C_{Npm} < 1$, por lo que la mediana no está centrada con ese Target (ya que tienen valores muy distintos).

La mayor diferencia con respecto a los índices paramétricos calculados en el apartado anterior y los robustos calculados en este se refiere precisamente al CO_2 , ya que con el análisis de capacidad asumiendo normalidad el proceso no salía como capaz. Sin embargo, con los índices robustos vemos que no es así, y el proceso sí es capaz.

En resumen, en el edificio del Ayuntamiento de Mondoñedo en verano el proceso solo es capaz en lo referente al CO_2 .

5.2.1.2. Estudio de las variables de calidad en invierno

- **Índices de capacidad para distribuciones normales**

A continuación se analizan los índices de capacidad normales el Ayuntamiento en invierno.

Para la temperatura media, este edificio tiene una categoría ambiental C, por lo que los límites de especificación en invierno, tanto inferior como superior, son 19° y 25° , mientras que el target es de 22° .

En cuanto a la humedad los límites de especificación son iguales a los anteriores, es decir, $LSL = 40\%$ y $USL = 60\%$ y el target es el punto medio. Y en lo relativo al CO_2 se tiene $LSL = 0$ ppm y $USL = 900$ ppm y $T = 450$.

Comenzando con la temperatura media (Figura 5.8) se observa claramente que, asumiendo normalidad (no se cumple ya que con el test de Shapiro Wilks obtenemos un p-valor de 0.001827), el proceso no es capaz, todas las observaciones están fuera del intervalo (LSL, USL) y se esperan un total de 1000000 no conformes.

Además, para el CO_2 se encuentra un caso distinto. Aplicando de nuevo el test de Shapiro Wilks para este grupo de observaciones, se acepta la normalidad de los datos, ya que el p-valor es de 0.1948. En el análisis de capacidad del Cuadro 5.14 se observa que los índices C_p y C_{pk} son ambos mayores que 1 (de hecho $C_p > 1,33$), por lo que el proceso utiliza menos que el 100% de la banda natural de tolerancia y el número de no conformes no es muy elevado (en concreto unos 1175).

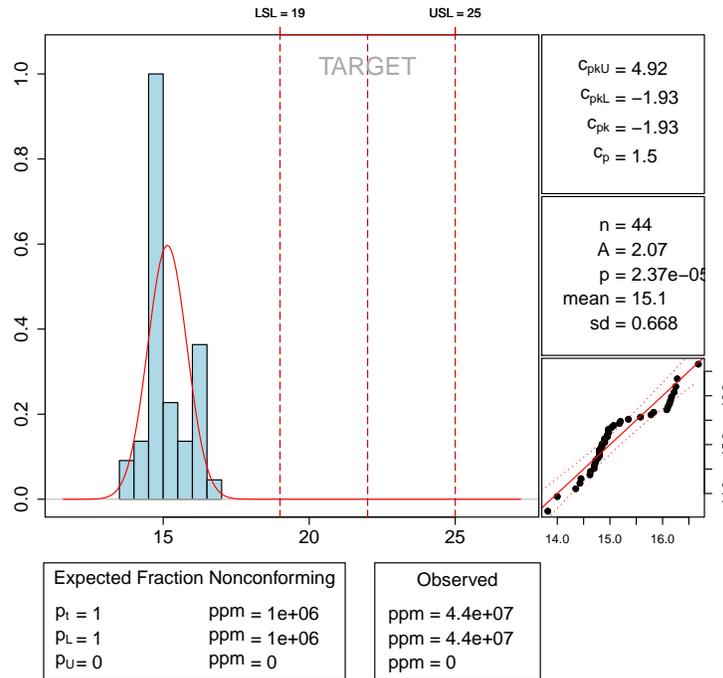


Figura 5.8: Análisis de capacidad normal para la temperatura media en el Ayuntamiento, en invierno.

Índices de Capacidad	Fracción esperada no conforme	Partes por millón
$C_{pkU} = 1,01$	$p_t = 0,001174845$	ppm=1174.845
$C_{pkL} = 1,73$	$p_L = 1,10361e - 07$	ppm=0.110361
$C_{pk} = 1,014$	$p_U = 0,001174735$	ppm=1174.735
$C_p = 1,37$		

Cuadro 5.14: Análisis de capacidad normal para el CO_2 medio en el Ayuntamiento, en invierno.

Por último, para la humedad relativa media (tampoco normal con un p-valor de 0.002313), viendo la Figura 5.9 se observa que tampoco es capaz, ya que C_p como C_{pk} con menores a 1 e incluso este último es menor que 0, por lo que (y como se ve en el histograma), una parte de las observaciones se salen del límite superior de especificación, convirtiendo el proceso en incapaz.

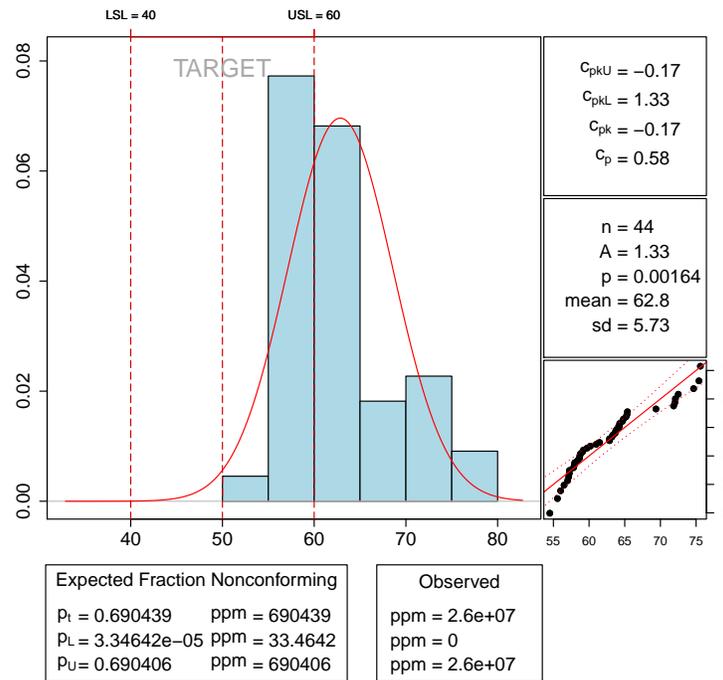


Figura 5.9: Análisis de capacidad normal para la humedad relativa media en el Ayuntamiento, en invierno.

- **Índices de capacidad para distribuciones arbitrarias**

Utilizando ahora los índices no paramétricos se obtienen los siguientes valores para cada uno de las variables:

	Temperatura.media	Humedad.rel.media	CO2
C_Np	2.12	0.95	1.97
C_Npk	-2.88	-0.27	1.51
C_Npm	0.14	0.25	1.15
C_Npmk	-0.19	-0.07	0.88

Cuadro 5.15: Índices de capacidad de todas las variables para el Ayuntamiento, en invierno.

Lo primer que llama la atención en referencia a la temperatura media es el valor de C_{Npk} , ya que es negativo y menor que asumiendo normalidad, en concreto $-2,88$, lo que significa que todo el proceso está fuera de los límites de especificación (esto ya se ha visto también en la Figura 5.8). De las 44 observaciones que tenemos para esta época en el Ayuntamiento, todas se sitúan fuera del intervalo dado por LSL y USL (todas las temperaturas son menores de 19°).

Además con los dos últimos índices (C_{Npm} y C_{Npmk}) se ve que el proceso no está centrado en el Target y la mediana también está descentralizada en torno a dicho valor nominal. El proceso es totalmente incapaz, la proporción de no conformes es del 100 %.

Algo similar ocurre con la humedad relativa media, pues de nuevo C_{Npk} es negativo, aunque no es tan acusado como en el caso anterior. Ya que $C_{Npk} < 0$ pero no < -1 , la mediana del proceso se encuentra fuera de los límites de especificación (Montgomery, 2009), pero no el proceso entero como en el caso de la temperatura media. De las 44 observaciones disponibles, en 26 de ellas la humedad relativa media supera el límite superior ($> 60\%$) y además la mediana está fuera de dicho límite (es de 62,87%). El proceso en este caso es de nuevo, no capaz.

En cuanto al CO_2 se ha comprobado que asume normalidad, pero utilizando los índices no paramétricos se observa que tanto C_{Np} como C_{Npk} son mayores que 1.33. No obstante la mediana del proceso no está centrada en torno al valor nominal (ya que $C_{Npmk} < 1$). El proceso es capaz.

5.2.2. Edificio de la Casa de Cultura

5.2.2.1. Estudio de las variables de calidad en verano

- **Índices de capacidad para distribuciones normales**

En verano, para la temperatura media, el edificio tiene una categoría ambiental A, por lo que el intervalo de especificación se situará entre 23.5 y 25.5 grados y con un target de 24,5°. Además, como solamente es necesario ver si el proceso es capaz en horas de ocupación, tendremos en cuenta solo dichas horas (desde las 8 hasta las 22).

El valor superior e inferior de los límites de especificación y el Target para la humedad relativa media son los mismos que en casos anteriores ($LSL = 40\%$, $USL = 60\%$ y $T = 50\%$). Lo mismo ocurre con el CO_2 , ya que volvemos a tener $LSL = 0$ ppm, $USL = 900$ ppm y $T = 450$ ppm.

Con estos valores se realiza el análisis de los índices de capacidad. Para la temperatura (no normal ya que se obtiene un p-valor de 0.0001427) se obtiene el estudio de capacidad de la Figura 5.10.

Se observa un $C_{pk} < 0$, lo que quiere decir que la totalidad de los datos se encuentran fuera de los límites de especificación fijados (en concreto por debajo de LSL), por lo que no se puede decir que sea capaz. El número esperado de no conformes entre un millón es elevado, de 972091.

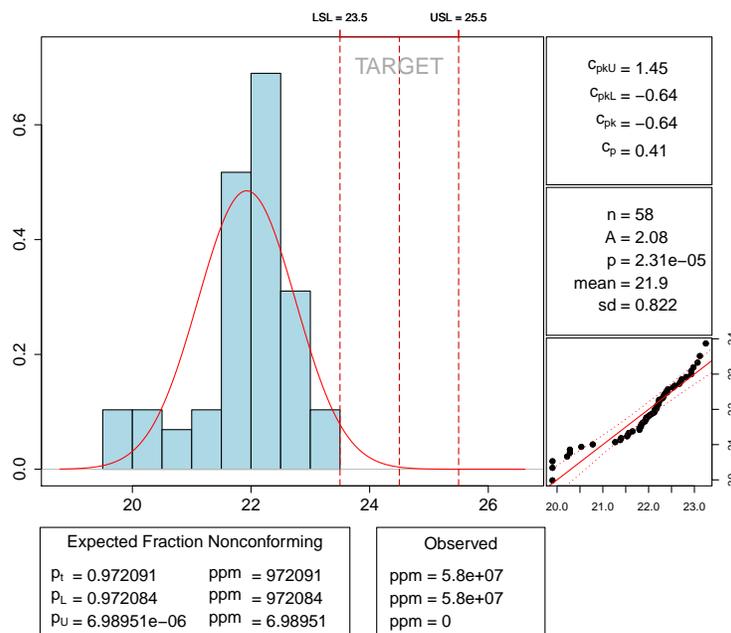


Figura 5.10: Análisis de capacidad normal para la temperatura media en la Casa de Cultura, en verano.

Para la humedad relativa media se obtiene un p-valor con el test de Shapiro Wilks de 0.1429, por lo que se acepta normalidad. El análisis de capacidad de la Figura 5.11 muestra que los datos no se encuentran dentro de los límites de especificación, sino que están por encima del límite superior, pues $C_{pk} < 0$. El número estimado de no conformes vuelve a ser elevado, llegando a 995085. Se puede decir que en este caso es adecuado aplicar este análisis de capacidad, ya que se ha confirmado la normalidad de la distribución, y se concluye que el proceso para la humedad relativa media en verano no es capaz.

Por otro lado, el CO_2 medio, que no acepta normalidad, tampoco es capaz (índices menores que 1, ver Figura 5.12). La media tampoco está centrada en torno al valor nominal ni el proceso centrado en torno a dicha media. Además, se tendrían 14090 elementos no conformes entre un millón. Veremos posteriormente qué ocurre si calculamos los índices robustos.

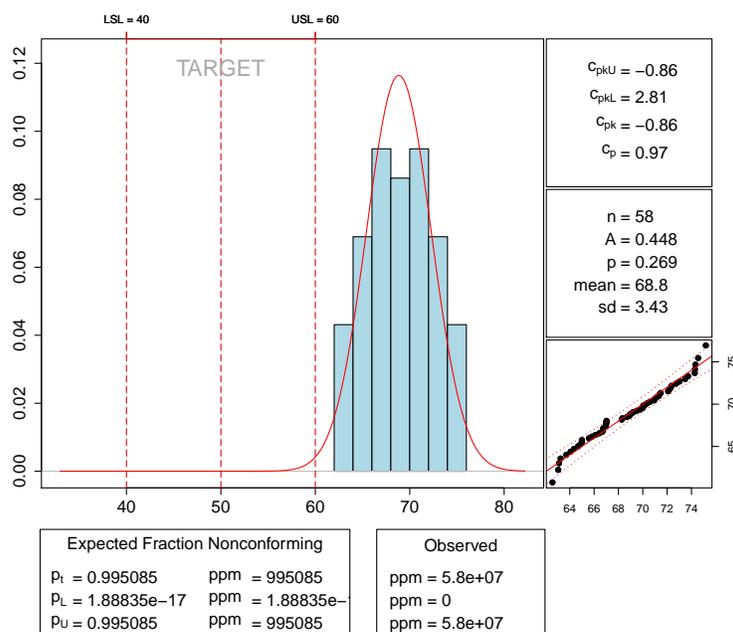


Figura 5.11: Análisis de capacidad normal para la humedad relativa media en la Casa de Cultura, en verano.

- **Índices de capacidad para distribuciones arbitrarias**

Para los índices no paramétricos tenemos los siguientes valores:

	Temperatura.media	Humedad.rel.media	CO2
C _{Np}	0.60	1.60	1.45
C _{Npk}	-0.82	-1.44	1.25
C _{Npm}	0.14	0.17	1.24
C _{Npmk}	-0.19	-0.16	1.07

Cuadro 5.16: Índices de capacidad de todas las variables para la Casa de Cultura, en verano.

Analizando el proceso relacionado con la temperatura media vemos que no es capaz, puesto que $C_{Npk} < 0$ (en concreto -0.82) y $C_{Np} < 1$ (0.60), lo cual implica que la mediana del proceso está fuera del intervalo (LSL, USL). Viendo la mediana de nuestros ($22,12^\circ$) se observa que claramente no está centrada en torno al valor nominal (que es de $24,5^\circ$), y además la totalidad de las observaciones (58) se encuentran por debajo del límite inferior de especificación. El proceso no es para nada capaz.

Fijándose ahora en la humedad relativa media, claramente el proceso tampoco es capaz por

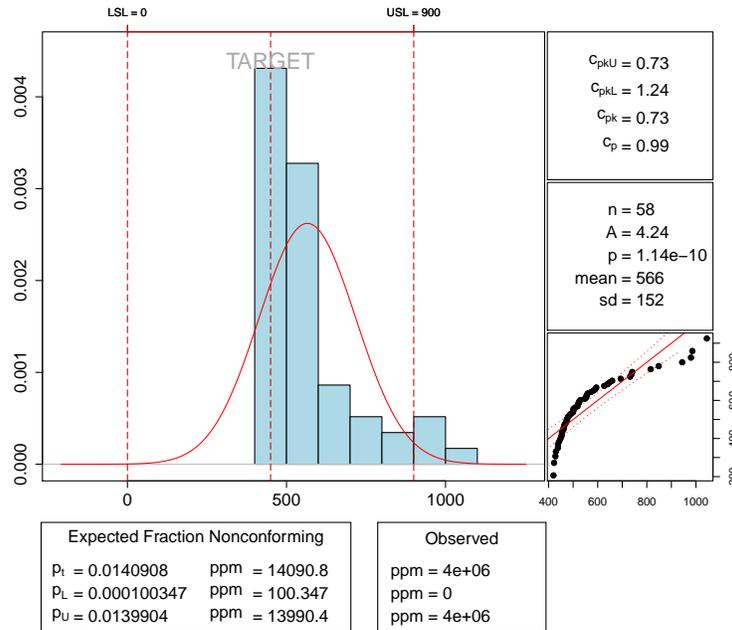


Figura 5.12: Análisis de capacidad normal para CO_2 medio en la Casa de Cultura, en verano.

las mismas razones que en la situación anteriormente mencionada. Las 58 observaciones están por encima del límite superior de 60 % de humedad relativa media y la mediana es del 69.01 %.

Por último, en cuanto al CO_2 el proceso sí parece capaz, al contrario que con lo que ocurriría si asumimos normalidad. Tanto C_{Np} como C_{Npk} son mayores que 1 (concretamente $C_{Np} > 1,33$), por lo que se cumplen las especificaciones. De las 58 observaciones, solo 4 superan las 900 ppm de emisiones de CO_2 , aunque la mediana está algo descentrada en torno al Target (ya que $\widehat{M} = 512,67$ y $T = 450$).

5.2.2.2. Estudio de las variables de calidad en invierno

- **Índices de capacidad para distribuciones normales**

Por su parte, en invierno la Casa de Cultura tiene una categoría ambiental C, con unos límites de especificación para esta época de 19 y 25 grados y un Target de 22. Los intervalos para la humedad relativa media y el dióxido de carbono son los mismos que anteriormente.

Se puede decir que el proceso relativo a la temperatura media no es capaz, al obtenerse un índice $C_{pk} < 0$. Esto significa que las observaciones, fuera de los límites (LSL, USL), no cumplen las especificaciones. Además, tampoco sigue una distribución normal al obtenerse un p-valor

cercano a 0 con el test de Shapiro Wilks.

Índices de Capacidad	Fracción esperada no conforme	Partes por millón
$C_{pkU} = 1,22$	$p_t = 0,847426$	ppm=847426
$C_{pkL} = -0,34$	$p_L = 0,847294$	ppm=847294
$C_{pk} = -0,34$	$p_U = 0,000132252$	ppm=132.252
$C_p = 0,44$		

Cuadro 5.17: Análisis de capacidad normal para la temperatura media en la Casa de Cultura, en invierno.

Por otra parte, para la humedad relativa media se obtienen los índices del Cuadro 5.18, el que se observa que 1 índice C_{pk} vuelve a ser negativo, además el proceso no está centrado en torno a la media y que muchas observaciones exceden el límite superior USL. Por tanto, asumiendo normalidad, el proceso no es capaz y no cumple las especificaciones dadas.

Índices de Capacidad	Fracción esperada no conforme	Partes por millón
$C_{pkU} = -0,06$	$p_t = 0,601198$	ppm=601198
$C_{pkL} = 0,63$	$p_L = 0,0294793$	ppm=29479.3
$C_{pk} = -0,06$	$p_U = 0,571719$	ppm=571719
$C_p = 0,28$		

Cuadro 5.18: Análisis de capacidad normal para la humedad relativa media en la Casa de Cultura, en invierno.

Para el CO_2 medio se sigue sin aceptar normalidad (p-valor en torno a 0) y, observando la Figura 5.13, aunque los índices no son negativos, tampoco son superiores a 1, por lo que no se puede afirmar que el proceso sea capaz. Se esperan un total de 186593 elementos no conformes por millón.

- **Índices de capacidad para distribuciones arbitrarias**

Como no se ha aceptado la normalidad de las distribuciones anteriores es más correcto calcular los índices no robustos, como los que siguen:

El proceso para la temperatura media en invierno tampoco es capaz, ya que $C_{Npk} < 0$, la mediana del proceso está fuera de los límites de especificación dados (es de 17.55 grados, casi 2 grados por debajo del límite inferior, por lo que tampoco está centrada en torno al Target como vemos en el valor negativo del índice C_{Npmk}) y se producen muchos no conformes (de las 89 observaciones 71 de ellas están fuera del intervalo de especificación, en concreto por debajo de LSL).

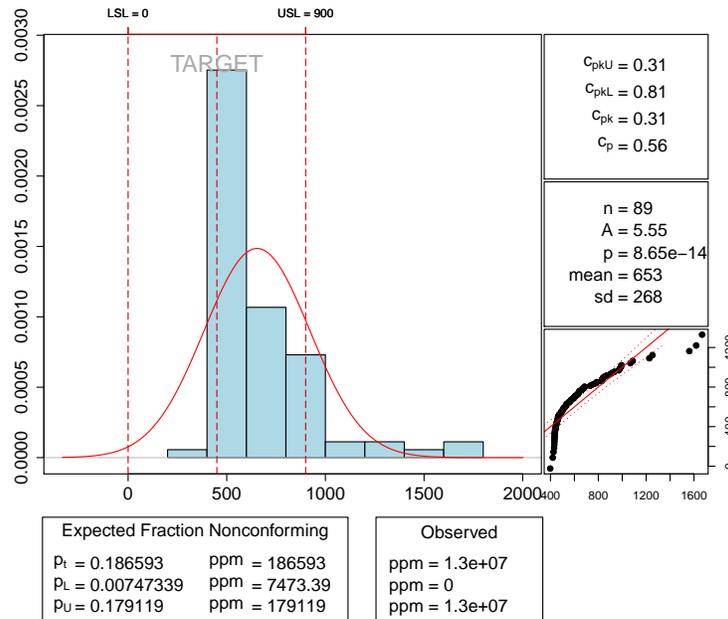


Figura 5.13: Análisis de capacidad normal para el CO_2 medio en la Casa de Cultura, en invierno.

	Temperatura.media	Humedad.rel.media	CO2
C_Np	0.79	0.52	0.71
C_Npk	-0.38	-0.01	0.53
C_Npm	0.22	0.28	0.63
C_Npmk	-0.10	-0.01	0.47

Cuadro 5.19: Índices de capacidad de todas las variables para la Casa de Cultura, en invierno.

La humedad relativa media tampoco es un proceso capaz, se observa un índice C_{Npk} ligeramente menor a 0, por lo que la mediana estará fuera de (LSL, USL), concretamente se sitúa algo por encima del límite superior (60.2 grados). De ahí el resultado que se obtiene con C_{Npmk} que implica que la mediana está descentrada en relación al valor nominal (que es de 50 grados). El proceso no cumple de este modo las especificaciones.

Por último, en cuanto al CO_2 en este caso no se puede decir que el proceso sea capaz (sí lo era en verano), ya que todos los índices son menores a 1. Asumiendo normalidad tampoco podíamos confirmar la capacidad del proceso. La mediana está bastante descentrada con respecto al Target (564.17 ppm y 450 respectivamente) y además se encuentran 13 observaciones por encima del límite superior de especificación (de un total de 89).

5.2.3. Edificio de la Escuela de Música

5.2.3.1. Estudio de las variables de calidad en invierno

- **Índices de capacidad para distribuciones normales**

El último paso es calcular los índices de capacidad para la Escuela de Música. Aquí se van a analizar éstos solamente para los datos de invierno.

En dicha época, en lo relativo a la temperatura media, el edificio tiene una categoría ambiental C, siendo entonces el límite superior de 25° y el inferior de 19° y el Target de 22°. Las horas de ocupación se sitúan entre las 11 y 14 y las 16 y 22 horas. Tanto para la humedad relativa media como para el dióxido de carbono tenemos los mismos valores que en los anteriores casos. Se obtiene entonces:

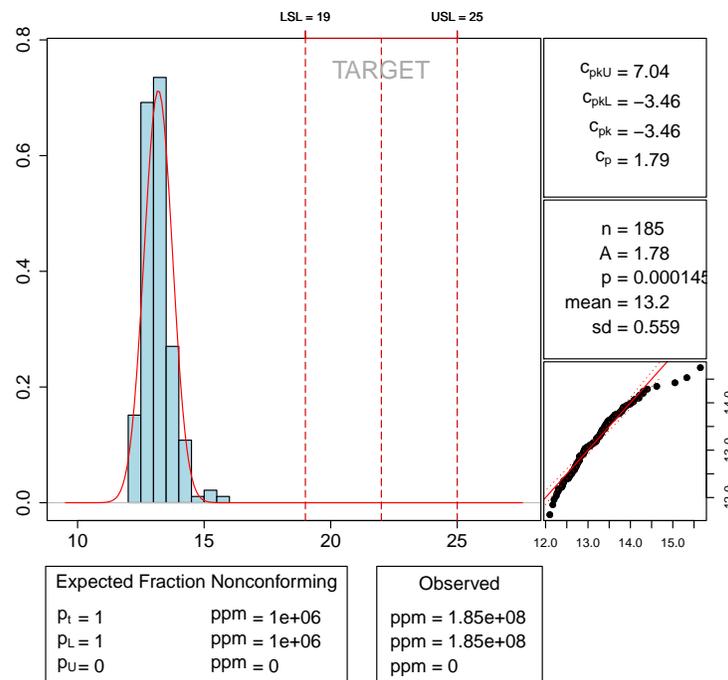


Figura 5.14: Análisis de capacidad normal para la temperatura media en la Escuela de música, en invierno.

Con el test de Shapiro Wilks se rechaza normalidad, con un p-valor de 0.000001085. Sin embargo, si asumimos normalidad de obtienen los índices de la figura anterior, donde se observa un valor muy bajo de C_{pk} ($C_{pk} = -3,46$), que quiere decir, tal como vemos en el histograma, que

todas las observaciones se sitúan fuera de los límites (LSL, USL), específicamente por debajo del límite inferior, por lo que el proceso no es capaz (aún teniendo un $C_p > 1,33$, ya que éste no tiene en cuenta la media).

Para la humedad relativa media, para la que no se acepta normalidad con un p-valor cercano a 0, se obtiene en análisis de la Figura 5.20. De nuevo, se observan valores negativos del índice C_{pk} y en este caso se ve que la totalidad de los datos están por encima del límite superior, haciendo un total de 999862 observaciones esperadas no conformes. En esta situación no se puede decir que el proceso sea capaz.

Índices de Capacidad	Fracción esperada no conforme	Partes por millón
$C_{pkU} = -1,21$	$p_t = 0,999862$	ppm=999862
$C_{pkL} = 2,72$	$p_L = 1,84293e - 16$	ppm= $1,84293e - 1$
$C_{pk} = -1,21$	$p_U = 0,999862$	ppm=999862
$C_p = 0,75$		

Cuadro 5.20: Análisis de capacidad normal para la humedad relativa media en la Escuela de música, en invierno.

Por último, para el CO_2 (Figura 5.15), que no cumple normalidad (p-valor cercano a 0), los índices C_p y C_{pk} , al no llegar a 1, informan que el proceso no es capaz y que el proceso no está centrado en torno a la media.

- **Índices de capacidad para distribuciones arbitrarias**

Calculando y analizando ahora los índices de capacidad robustos tenemos el siguiente cuadro:

	Temperatura.media	Humedad.rel.media	CO2
C_Np	1.74	1.12	1.23
C_Npk	-3.37	-1.71	1.20
C_Npm	0.11	0.13	1.23
C_Npmk	-0.22	-0.20	1.20

Cuadro 5.21: Índices de capacidad de todas las variables para la Escuela de Música, en invierno.

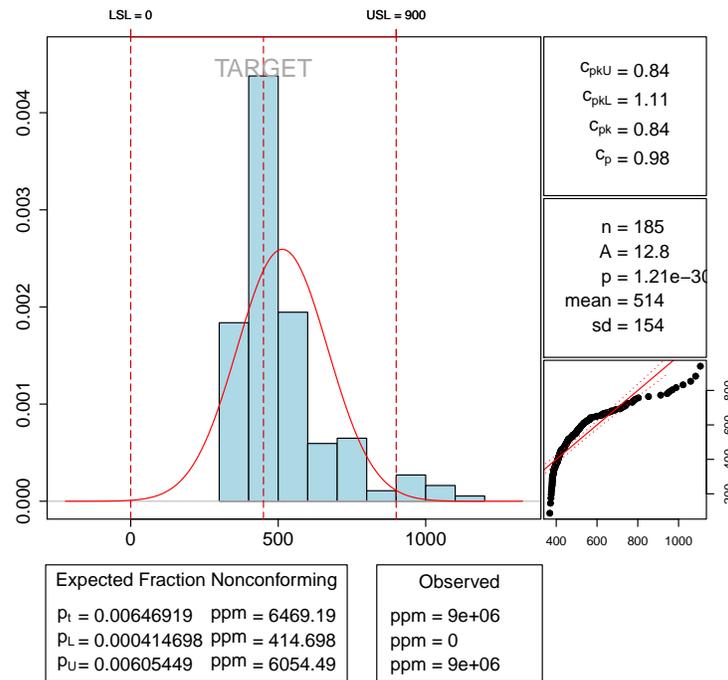


Figura 5.15: Análisis de capacidad normal para el CO_2 medio en la Escuela de música, en invierno.

De nuevo, se observa un valor muy bajo (-3.37) del índice C_{Npk} , lo que lleva a decir que todo el proceso está fuera de los límites de especificación (de 185 observaciones todas ellas tienen una temperatura menor al límite inferior de 19 grados). El proceso es completamente incapaz. Por consiguiente, la mediana no está centrada en torno al Target (dicha mediana es solamente de 13.18 grados).

En lo referente a la humedad se observa que también es incapaz, al ser $C_{Npk} < 1$ el proceso está fuera del intervalo (LSL, USL), con las 185 observaciones por encima del 60% de humedad (y con una mediana muy elevada del 75.25%). Las especificaciones no se cumplen de ningún modo.

En último lugar se analiza el CO_2 , con índices que son todos mayores que 1 (en torno a 1.20). Esto quiere decir que el proceso es capaz y los límites de especificación caen dentro de los límites de tolerancia natural. Además, la mediana (459.33 ppm) tiene un valor cercano al valor nominal (de 450 ppm), por lo que está centrada en torno a dicho Target, tal como vemos en los índices C_{Npm} y C_{Npmk} . De este modo, y no asumiendo normalidad ya que se ha comprobado que no se satisface, se cumplen las especificaciones y el proceso para el CO_2 es capaz.

5.3. Posibles acciones correctoras

En la Sección anterior se ha comprobado, mediante la aplicación de los índices de capacidad, la incapacidad de la mayor parte de los sistemas de control de estos tres edificios administrativos en Mondoñedo. Se ha visto que los únicos procesos capaces son los relacionados con el contenido de CO_2 , siendo las variables indicadoras del confort térmico no capaces en todos los casos.

Como consecuencia, se pueden plantear varias medidas a desarrollar para mejorar los procesos y, por consiguiente, aumentar la satisfacción de los usuarios en relación al confort térmico.

Una de las medidas a tomar podría ser el aislamiento térmico por el interior de la fachada de los edificios, de modo que se reduzcan las pérdidas energéticas y térmicas que se producen debido a la diferencia de temperaturas entre el interior y el exterior. Esta medida podría implicar una disminución del uso de la calefacción y el consecuente ahorro energético.

También se podría proceder al tapiado de ventanas en las oficinas o salas que estén en zonas de mayor sombra, con el objetivo de reducir las pérdidas por transmisión a través de las ventanas.

Otra elemento a tener en cuenta es el número de renovaciones de aire que se producen en los edificios, puesto que si la apertura de las puertas durante el horario de actividad es elevado, la entrada de aire es mayor, lo que puede producir situaciones de incomodidad, sobre todo en invierno. Una posible solución sería la instalación de cortinas de aire en las puertas, que evitasen la constante entrada del mismo.

Estas pueden ser, entre otras, algunas medidas a implementar en estos edificios con el objetivo de mejorar sus sistemas de control y proporcionar un ambiente térmico confortable.

Capítulo 6

Resultados: Fridama

6.1. Análisis Exploratorio de Datos

6.1.1. Estudio descriptivo mediante el análisis de datos funcionales

En primer lugar, se va a hacer un primer análisis exploratorio utilizando la metodología de datos funcionales y con la librería de R `fda.usc`. Dado que en la oficina de la empresa Fridama se han tomado mediciones tanto en días laborables como festivos, se estudiarán las posibles diferencias que puedan existir para, en caso afirmativo, construir los gráficos de control separando por tipo de día.

Como ya se ha señalado en el Capítulo 2 en relación a la empresa Fridama, las variables a analizar van a ser concretamente la temperatura ambiente en la oficina medio, la concentración de CO_2 en esa misma sala (variables indicadoras de la calidad del aire y el confort térmico) y el consumo de energía medido por el analizador principal (variable indicadora de la eficiencia energética).

Para realizar este procedimiento primero se filtra la base de datos para cada variable y en cada una se separan los días laborables y festivos. Posteriormente, se construye una matriz para cada uno ellos donde cada fila va a ser un día. Con esto, se definen las dos bases de datos de datos como objetos `fdata` y se procede a su suavización mediante el método de suavización no paramétrico de kernel (con el estimador Nadaraya-Watson). Para seleccionar la venta óptima se emplea el criterio de validación cruzada generalizada y antes se establece un vector de posibles ventanas (entre 0.01 y 0.5 y de longitud 50).

Con los datos suavizados se procede a representar las curvas, una para cada día y, a continuación, se representan las medias funcionales y las bandas de confianza bootstrap a un nivel

de confianza del 95 %. Con ello se ve si existen diferencias que hagan pensar que las curvas son diferentes.

En primer lugar se estudia la **temperatura ambiente**. Las curvas funcionales para cada día laborable y para cada día festivo se muestran a continuación en las Figuras 6.1 y 6.2 respectivamente.

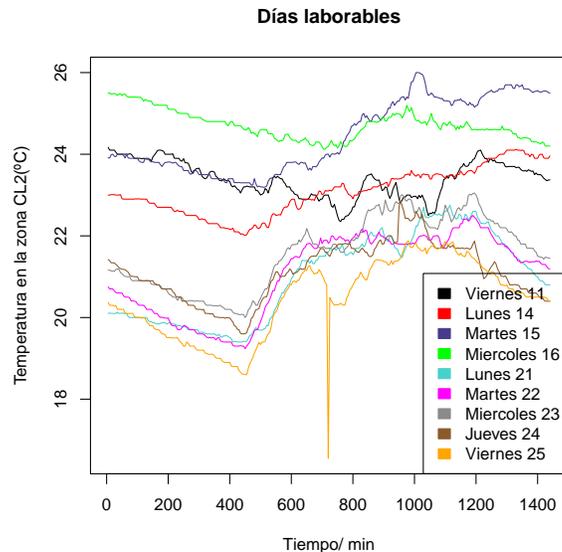


Figura 6.1: Datos reales suavizados de la temperatura ambiente en CL2 para los días laborables.

En la Figura 6.1 se observa una bajada paulatina de la temperatura hasta el minuto 400 (sobre las 6 de la mañana), cuando, aproximadamente, comienza la actividad de la empresa.

También se observa un día en que se aprecia algún atípico, en concreto el viernes 25, donde la temperatura baja de 18 grados. Al contrario, también hay días en los que se llega a los 26 grados.

Por otro lado, en los días festivos se ve un patrón diferente pues la bajada de la temperatura es progresiva hasta las 11.00-13.00 horas del día aproximadamente, para aumentar a partir de esas horas en casi todos los días, llegando alguno incluso hasta los 24 grados. De hecho, la temperatura no baja en ningún caso de los 19 grados, por lo que parece que el sistema de regulación de la misma sigue funcionando, o que, en todo caso, la construcción del edificio (cerramientos, forjados) es térmicamente eficiente.

Además, a continuación se representan las medias funcionales para la temperatura con las bandas de confianza bootstrap al 95 % (Figuras 6.3 y 6.4).

La media funcional de la temperatura para los días laborables oscila en torno a los 22 grados

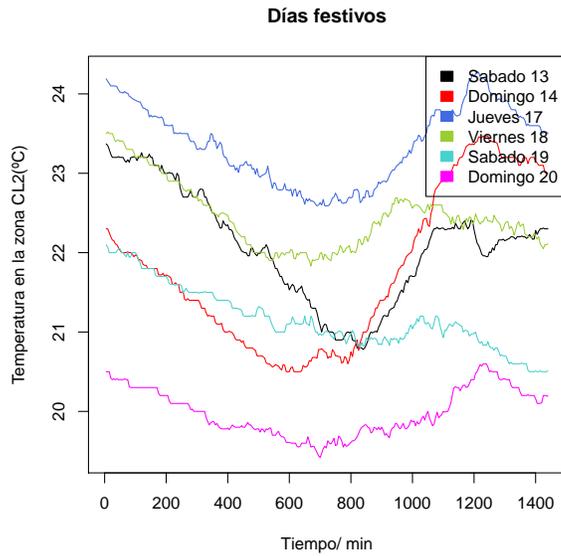


Figura 6.2: Datos reales suavizados de la temperatura ambiente en CL2 para los días festivos.

prácticamente todos los días, mientras que en el caso de los días festivos oscila en torno a los 21 y 22 grados.

Como efectivamente se pueden observar patrones diferenciados, se concluye entonces que las curvas de temperatura en la zona CL2 de Fridama son diferentes dependiendo de si el día es festivo o laborable. Por ello, construiremos gráficos de control por separado.

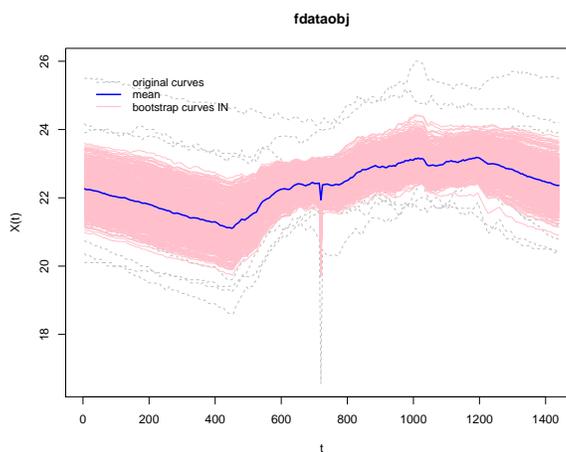


Figura 6.3: Media funcional y bandas bootstrap (95 %) para la temperatura en los días laborables.

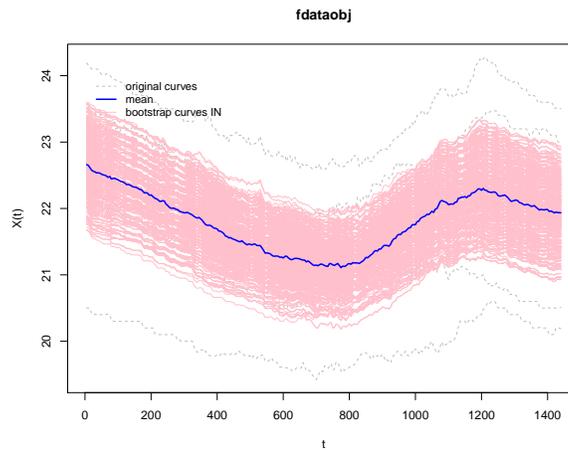


Figura 6.4: Media funcional y bandas bootstrap (95 %) para la temperatura en los días festivos.

A continuación se realiza este mismo estudio para el **contenido de CO_2** en esa misma zona. De este modo, se comienzan representando los datos reales suavizados y las medias funcionales para cada categoría (festivo/laborable) con las bandas bootstrap a un nivel de confianza del 95 %.

En los gráficos referentes al contenido de CO_2 en los días laborables, Figura 6.5, se observan cambios bastantes marcados. En primer lugar, e igualmente al caso de la temperatura en los días laborables, a partir de los minutos 500-600 (aproximadamente a las 6 de la mañana) comienza la actividad y, por consiguiente, aumentan los niveles de CO_2 . Posteriormente, al mediodía (minutos 800-900) se produce una bajada correspondiente con la parada para comer y, a continuación, vuelve a haber otro aumento por la tarde (aunque menos acusado que por la mañana, hay menos actividad en la oficina en horario de tarde). Por último, a partir de las 20 horas (minuto 1200), a la hora de cierre y finalización de la actividad, el contenido en CO_2 va disminuyendo paulatinamente.

Como consecuencia de estas oscilaciones, se observa que la media funcional varía bastante a lo largo del día.

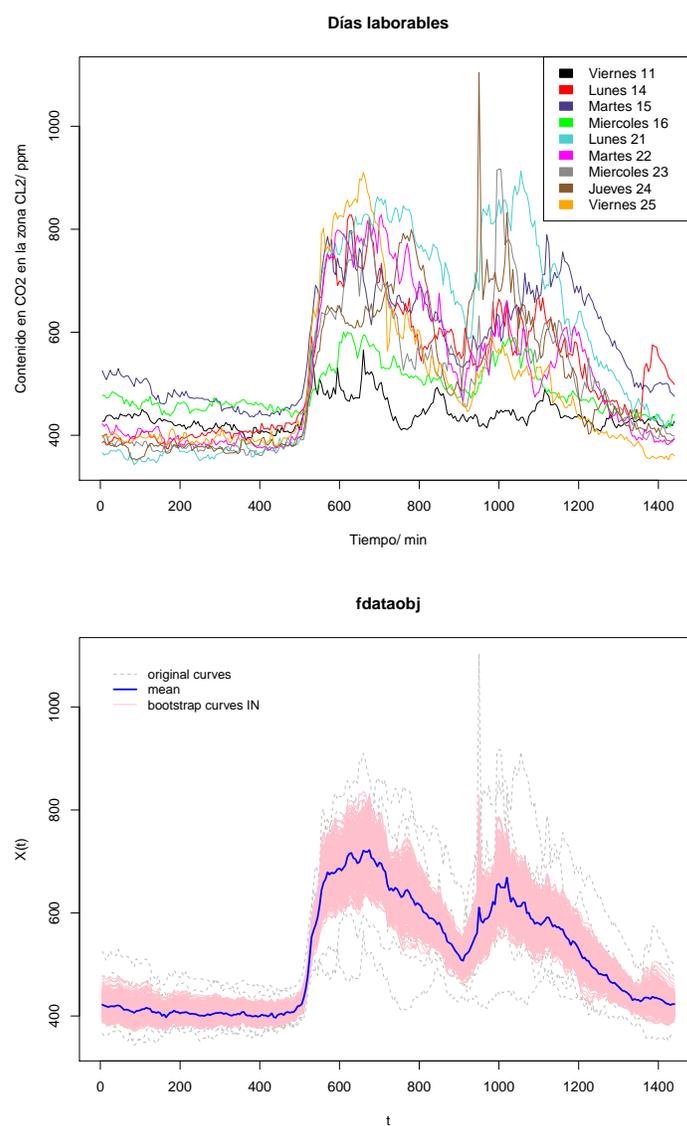


Figura 6.5: Datos reales suavizados para el CO_2 (arriba) y media funcional con bandas bootstrap al 95 % (abajo) para los días laborables.

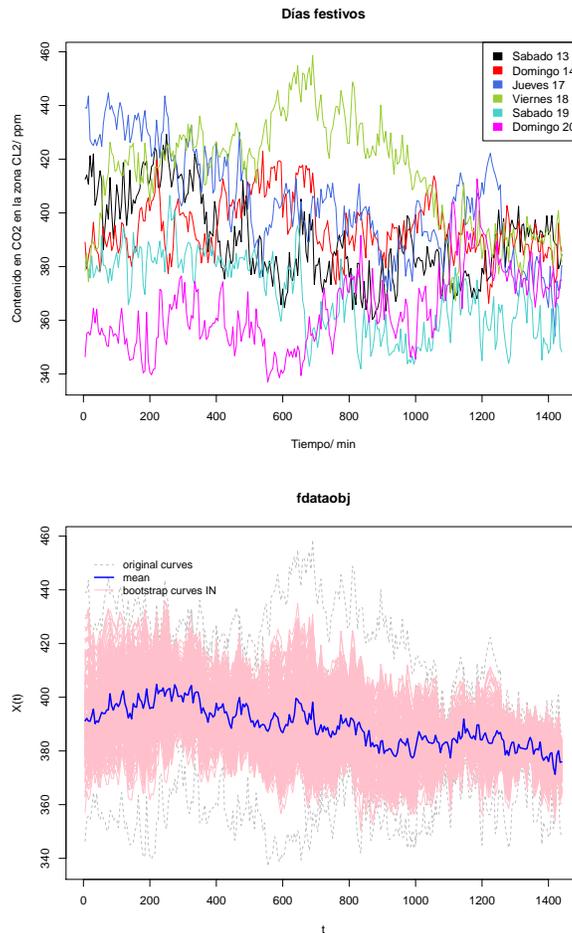


Figura 6.6: Datos reales suavizados para el CO_2 (arriba) y media funcional con bandas bootstrap al 95 % (abajo) para los días festivos.

En los días festivos (Figura 6.6) se observa una situación muy diferente con respecto a la anterior, pues no hay actividad. Los niveles de CO_2 son bastantes constantes a lo largo del día, disminuyendo progresivamente a medida que pasa el tiempo y situándose la media funcional en torno a 390 ppm (los días laborables hay momentos en los que se llega a 900).

Para el caso del CO_2 también se ha aplicado una técnica para el análisis de la varianza funcional basada en el método de proyecciones aleatorias, siguiendo el procedimiento propuesto por Cuesta-Albertos y Febrero-Bande (2010). El objetivo será comprobar si existen diferencias significativas entre los tipos de curvas (laborables y festivos), obteniéndose un p-valor menor que 0.05, por lo que se acepta la hipótesis alternativa de que son diferentes. De este modo, no se aplicará un gráfico de control a todos los datos sino diferenciando por tipo.

Por último, se analiza el **consumo de energía** medido por el analizador principal de la empresa Fridama.

De modo análogo a los anteriores, se representan los datos reales suavizados para el consumo energético en función del día (laborable/festivo), obteniendo:

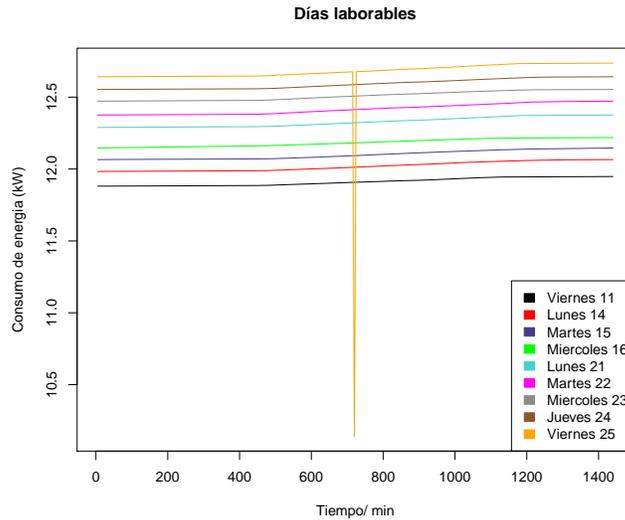


Figura 6.7: Datos reales suavizados para el consumo de energía para los días laborables.

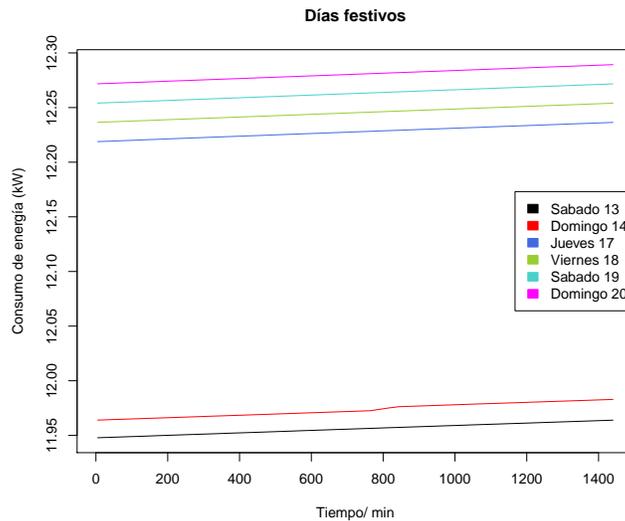


Figura 6.8: Datos reales suavizados para el consumo de energía para los días festivos.

Tanto en las curvas para los días festivos como para los días laborables se observan patrones

similares. En el caso de los días laborales, hay un aumento en el consumo de energía a partir aproximadamente de 6 de la mañana (minuto 400 como en el caso del CO_2).

En el caso de los festivos, se pueden ver días donde prácticamente se consume lo mismo que en días laborables. Esto puede ser debido a que alguien ha decidido ir a la oficina a trabajar esos días.

Puede observarse que en este caso las curvas son muy similares y, de este modo, se concluye que para el consumo de energía no hay diferencia dependiendo de si es festivo o laborable.

6.2. Control de Calidad y gráficos de control con datos autocorrelacionados

Como ya se ha señalado en el Capítulo 4 existen multitud de ocasiones en las que las observaciones de una serie presentan dependencia y el efecto de esa autocorrelación no debe ser ignorado, pues la estimación de los límites de control de los gráficos de control no sería correcta. En esta Sección se examinarán los datos disponibles para comprobar si existe realmente dependencia serial y poder actuar en consecuencia.

Además, como se ha comprobado a través análisis de datos funcionales, se deben considerar los días festivos y laborables por separado, puesto que sus curvas son diferentes.

A continuación se analiza cada variable individualmente, comprobando si es necesario ajustar un modelo ARIMA y, posteriormente, construyendo los gráficos de control (principalmente los gráficos EWMA).

6.2.1. Análisis de la variable temperatura ambiente en la oficina CL2

6.2.1.1. Análisis de la temperatura ambiente: días laborables

Si se realiza un gráfico de control EWMA para la serie original de la temperatura ambiente en los días laborables, seleccionando 7 días para la muestra de calibrado, se obtiene la siguiente figura:

A la vista de la Figura 6.9, el gráfico no es adecuado, prácticamente todas las observaciones están fuera de control. Esta dependencia debe ser modelizada, pues de otro modo no se deben aplicar los gráficos de control.

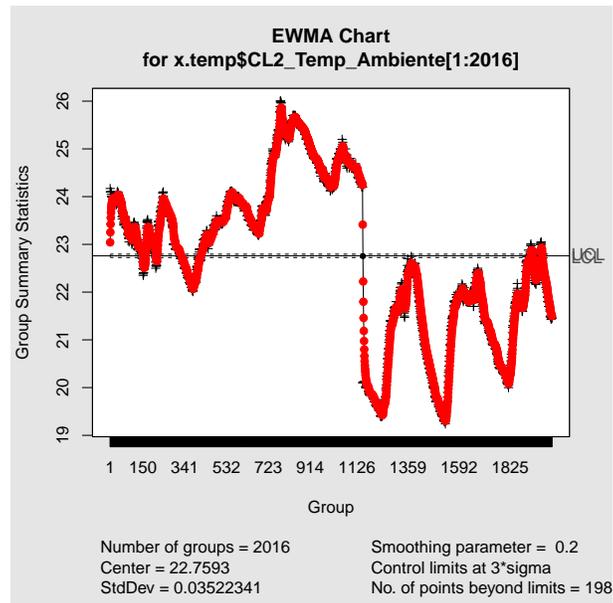


Figura 6.9: Gráfico de control EWMA de la muestra de calibrado con la serie original de la variable temperatura en la oficina CL2 en los días laborables.

- **Modelización de una serie temporal**

En primer lugar vamos a ver el gráfico de la evolución de la temperatura en los días laborables y, como se observa en la Figura 6.10, parece que las temperaturas caen aproximadamente a partir de la observación 1153, que se corresponde con el inicio del 5º día laborable (lunes 21). Desde el miércoles 16 al lunes 21 baja la temperatura ambiente hasta 4°C (pasa de 24°C a 20°C). La existencia de estos dos se debe al proceso de mejora continua que la empresa Fridama aplica al control y regulación de la temperatura ambiente para adaptarse a las especificaciones.

Asimismo, también se observa un valor mucho más bajo que el resto en el último día (viernes 25), con una temperatura (en concreto a las 12 del mediodía) de 16,55°C. Este atípico podría deberse a un fallo del instrumento de medición.

Debido a la existencia de estos dos niveles distintos, se opta por hacer el estudio de control solamente a partir del día 5, por lo que se filtran los datos para quedarnos con dichos días.

Para estudiar si existe tendencia se representa el gráfico de la función de autocorrelación (ACF) (Figura 6.11), con el que claramente se comprueba su presencia, pues existe mucha correlación en los primeros retardos y además tarda en bajar a 0.

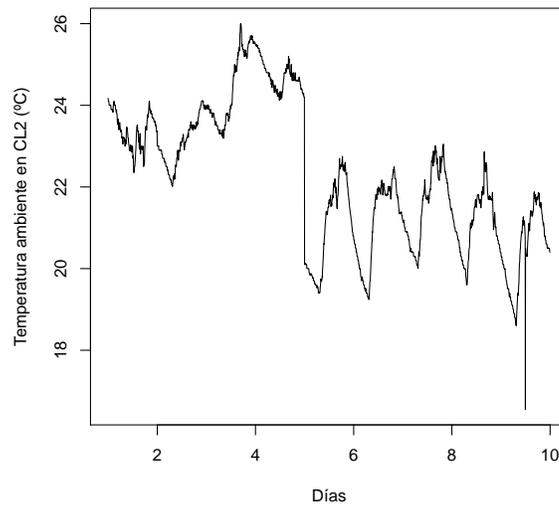


Figura 6.10: Gráfico secuencial de la variable temperatura en la oficina CL2 en los días laborables

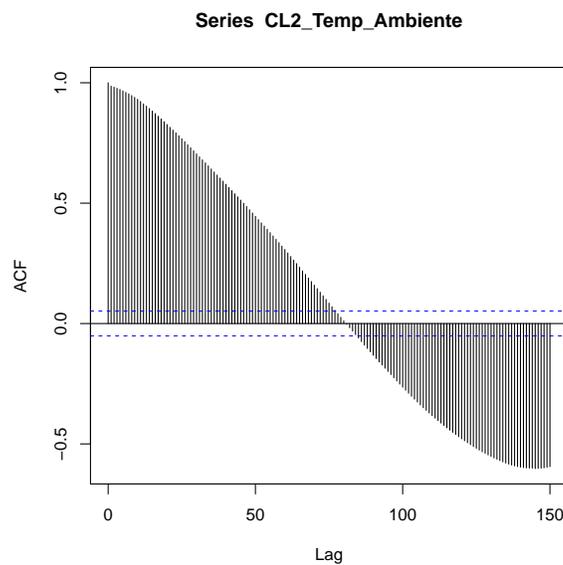


Figura 6.11: ACF de la temperatura en la oficina CL2 a partir del 5º día laborable.

A continuación se calculan los promedios horarios, puesto que trabajar con los promedios de cada 5 minutos (serie original) puede hacer más complejo ajuste del ARIMA y por consiguiente la construcción de los gráficos de control. Aunque este procedimiento conlleva una pérdida de

la información disponible, los resultados son más interpretables, atenuando el ruido debido a la precisión de los sensores.

Se calculan entonces los promedios para cada hora y se pasa de 1440 observaciones a 120 (24 observaciones cada día de los 5 analizados), obteniendo el siguiente gráfico para la serie:

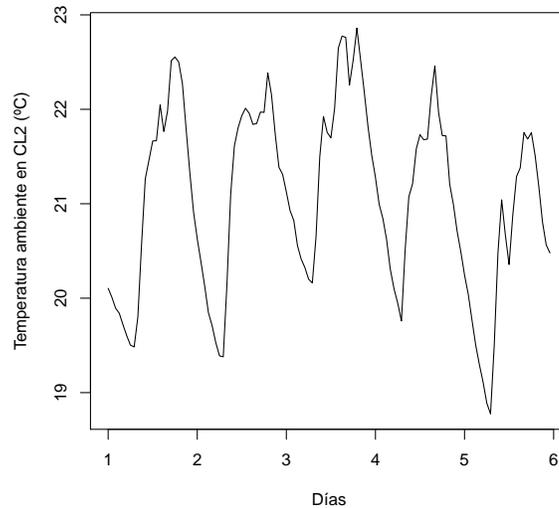


Figura 6.12: Gráfico secuencial de la variable temperatura promediada cada hora en la oficina CL2 en los 5 días laborables analizados.

Se pueden observar ligeras diferencias con respecto a la Figura 6.10 como, por ejemplo, el atípico del último día (viernes 25) no es tan acusado, ya que se ve compensado por el resto de observaciones.

En el ACF (Figura 6.13) se comprueba la existencia de tendencia, por lo que a partir de ahora se trabajará con esta serie.

Para la **muestra de calibrado**, con la que se establecerán los límites de control, se seleccionan los 3 primeros días (lunes 21, martes 22 y miércoles 23). Los dos siguientes (jueves 24 y viernes 25) serán monitorizados una vez se estimen dichos límites.

Como se ha dicho en la parte relativa a las series temporales del Capítulo 4, una vez se corrobora la tendencia, es necesario proceder a la **diferenciación de esa parte regular**, para eliminarla e ir convirtiendo la serie en estacionaria. De este modo, se diferencia la serie completa para suprimir la tendencia y se seleccionan las observaciones relativas a los 3 días que formarán parte de la muestra de calibrado. Se observa tanto el gráfico secuencial de la nueva serie diferenciada como el ACF en la Figura 6.15.

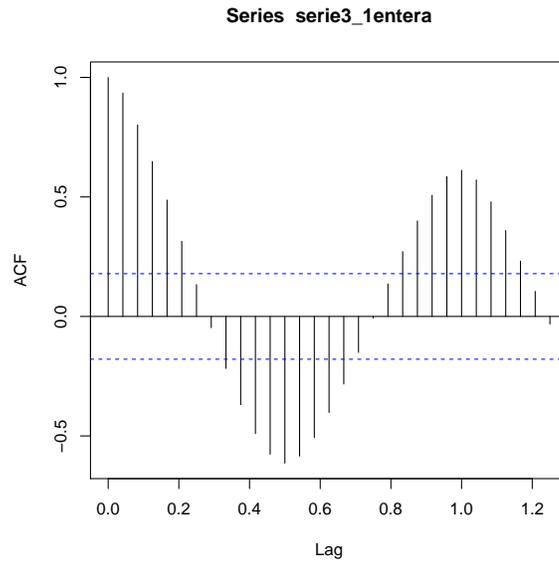


Figura 6.13: ACF de la temperatura con promedios horarios en la oficina CL2 en los días laborales analizados.

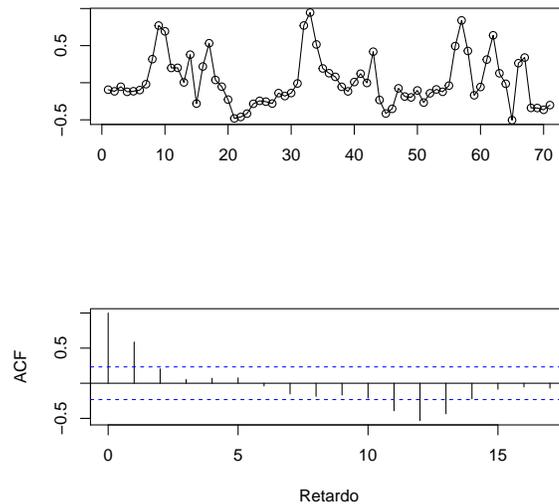


Figura 6.14: Gráfico secuencial (arriba) y ACF (abajo) de la serie diferenciada para la muestra de calibrado para la temperatura en CL2 los días laborales estudiados.

Se comprueba que no existe tendencia, por lo que se procede a sugerir un modelo tentativo para la serie observando los gráficos de autocorrelación simple y parcial:

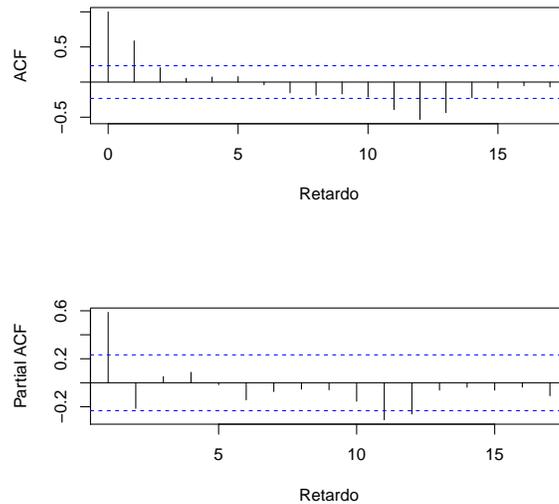


Figura 6.15: ACF (arriba) y PACF (abajo) de la serie diferenciada para la muestra de calibrado para la temperatura en CL2 los días laborables estudiados.

Observando los gráficos anteriores se decide el modelo a ajustar atendiendo además a los criterios AIC y BIC (valores bajos), así como verificando que se cumplan las hipótesis básicas sobre los residuos (independencia y media cero).

Siguiendo esto, se prueba el ajuste de varios ARIMA (p,d,q), algunos de los cuales no cumplen las hipótesis de independencia, y se selecciona el ARIMA con el que mejores resultados se obtienen, siendo éste un ARIMA (3,1,2), sin constante puesto que no es significativamente distinta de 0. Con dicho modelo se obtienen unos valores AIC y BIC de 18.75 y 32.33 respectivamente.

El modelo se puede expresar como:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + \theta_1 a_{t-1} + \theta_2 a_{t-2} + a_t$$

Estimando los parámetros por máxima verosimilitud se obtienen las estimaciones:

$$\hat{\phi}_1 = 1,1274 (0,1098) , \quad \hat{\phi}_2 = -1,3111 (0,0666)$$

$$\hat{\phi}_3 = 0,5964 (0,1069) , \quad \hat{\theta}_1 = -0,4811 (0,0932)$$

$$\hat{\theta}_2 = 0,9992 (0,2663) , \quad \hat{\sigma}_a^2 = 0,06011$$

Entre paréntesis se señalan los errores estándar. Se puede observar que todos son significativamente distintos de 0. De este modo, el ARIMA(3,1,2) estimado se expresa como:

$$x_t = 1,1274x_{t-1} - 1,3111x_{t-2} + 0,5964x_{t-3} - 0,4811a_{t-1} + 0,9992a_{t-2} + a_t$$

y 0.06011 es la varianza del ruido blanco.

El siguiente paso es verificar el cumplimiento de las hipótesis necesarias para poder considerar que el modelo tentativo es adecuado. En primer lugar se estudia la independencia de los residuos (H_0), que, como se observa en la Figura 6.16, se puede aceptar, ya que los p-valores para cada retardo del test de Ljung-Box son mayores que un nivel de significación de 0.05.

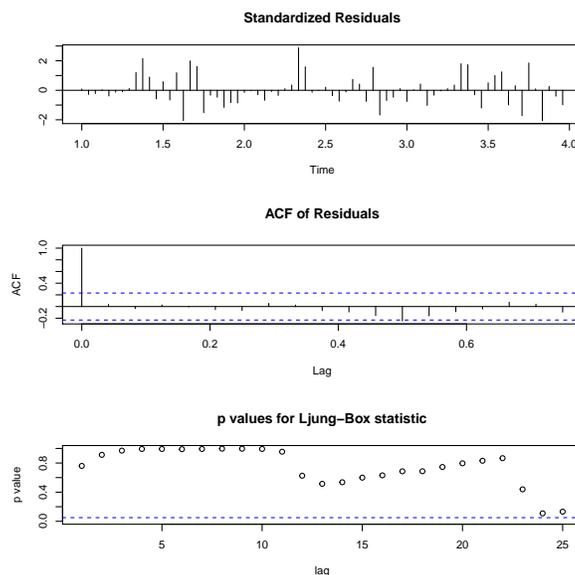


Figura 6.16: Test de independencia para los residuos del modelo ARIMA (3,1,2) ajustado a la muestra de calibrado.

Además, se puede hacer una primera aproximación a la comprobación de la normalidad de los datos viendo el gráfico secuencial y el gráfico Q-Q normal de los residuos, en la Figura 6.17.

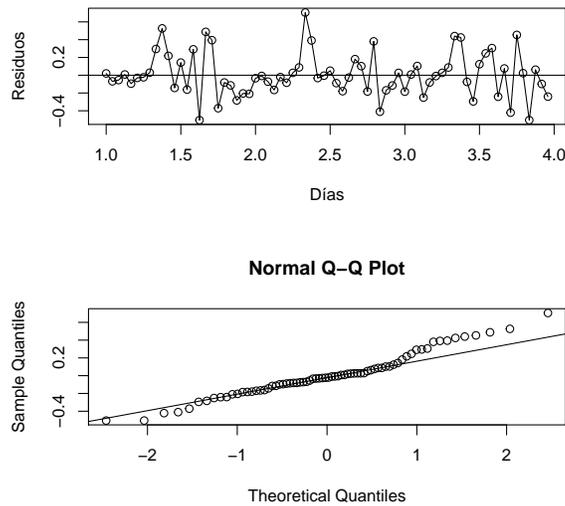


Figura 6.17: Gráfico secuencial (arriba) y Q-Q normal (abajo) los residuos del modelo ARIMA (3,1,2) ajustado a la muestra de calibrado.

Parece que los residuos sí se van a poder considerar gaussianos, pero se va a contrastar con el test de normalidad de Shapiro-Wilks y el test de Jarque-Bera. Asimismo se contrasta si la media de los residuos (μ_a) es 0 (indispensable para poder considerar que el modelo es adecuado) con el test T de media igual a 0 y se los p-valores se muestran en la siguiente tabla:

	P-valores
Test T de media = 0	0.85
Test Shapiro-Wilks	0.06
Test de Jarque-Bera	0.24

Cuadro 6.1: Contrastes sobre la media y la normalidad de los residuos del modelo ARIMA (3,1,2).

Viendo los resultados puede decirse que, con un p-valor de 0.85 se acepta que $\mu_a = 0$. Además, también se acepta la normalidad de los residuos, pues el test de Jarque-Bera arroja un p-valor de 0.24, con el que se acepta la hipótesis nula, al igual que con el test de Shapiro-Wilks (con éste se acepta a un nivel de significación del 10 o del 5 % pero no del 1 %).

Por tanto se concluye que el modelo ARIMA (3,1,2) sin constante puede ser utilizado como generador de la serie para la muestra de calibrado de la temperatura en la oficina CL2 (en los últimos 5 días laborables analizados). Además, las innovaciones se pueden considerar gaussianas.

- **Gráficos de control**

Ahora que se ha modelizado la muestra de calibrado de manera adecuada, y se ha validado el modelo en relación con las hipótesis básicas a cumplir por los residuos, se pueden construir sobre ellos los gráficos de control. Además, se ha comprobado que éstos se pueden considerar normales, por lo que se utilizará tanto el gráfico de control EWMA como el de medidas individuales.

En primer lugar se construye el gráfico de control EWMA para la media, con $\lambda = 0,2$ y una amplitud de los límites de 3σ , tal como se explicó en la Sección 4.1, obteniendo las siguientes estimaciones para la media y la desviación típica:

$$\hat{\mu}_0 = 0,005442343, \quad \hat{\sigma} = 0,2210612$$

Con esas estimaciones se calculan los límites de control superior (UCL) e inferior (LSL) para cada instante t . De este modo, se obtiene el gráfico EWMA para la muestra de calibrado que se representa a continuación:

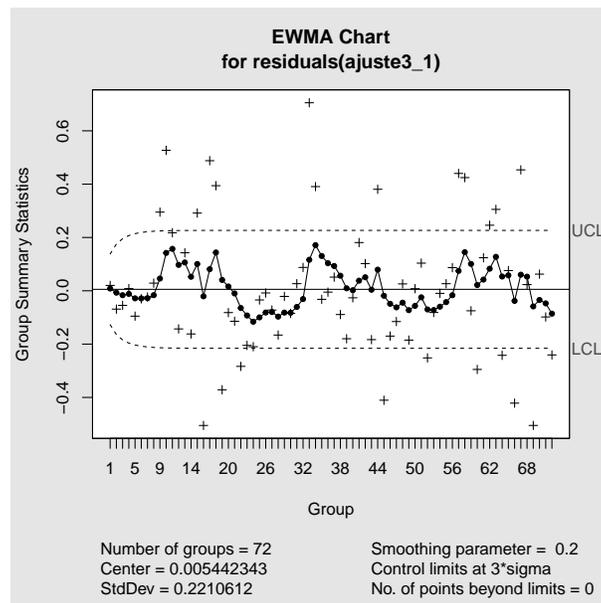


Figura 6.18: Gráfico EWMA para los residuos del modelo ARIMA (3,1,2) ajustado a la serie de la temperatura en la CL2 los cinco últimos días laborables.

En él se observa la representación de los residuos (los puntos negros), así como los datos originales de la serie (las cruces). No aparece ninguna observación fuera de control, por lo que puede afirmarse que el proceso se encuentra bajo control estadístico. Por consiguiente, se puede monitorizar el resto de observaciones.

De este modo, se calculan a continuación los residuos para la muestra de monitorizado (dos últimos días de la serie) en base al modelo ajustado ARIMA (3,1,2), teniendo en cuenta que:

$$\hat{a}_t = x_t - (1,1274x_{t-1} - 1,3111x_{t-2} + 0,5964x_{t-3} - 0,4811a_{t-1} + 0,9992a_{t-2})$$

Se programa su cálculo en R y una vez se obtienen se introducen en el gráfico EWMA como muestra de monitorización, utilizando los límites de control estimados con la muestra de calibrado:

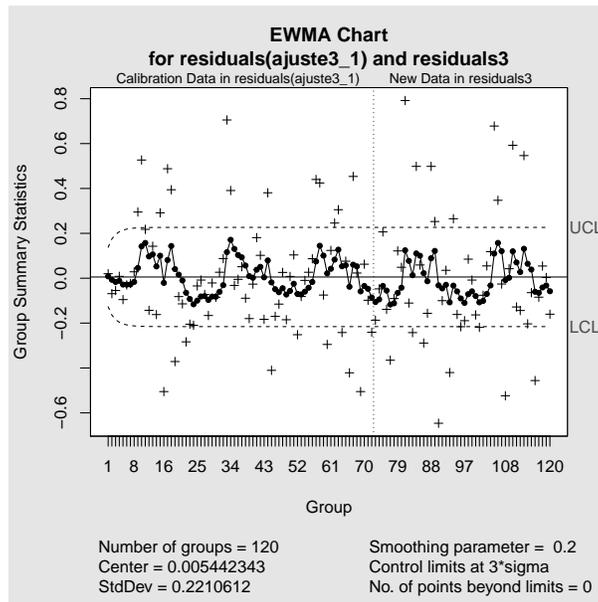


Figura 6.19: Gráfico EWMA para los residuos del modelo ARIMA (3,1,2) de la muestra de calibrado y de monitorizado para la temperatura en la oficina CL2.

En el gráfico EWMA para la muestra de monitorizado (a partir de la línea de puntos vertical) no se observa ningún dato fuera del límite superior o inferior, por lo que puede concluirse que el control de la temperatura de la empresa Fridama para la oficina CL2 se encuentra bajo control.

Por último, también se va a construir un gráfico de control de medidas individuales, pues se ha comprobado la normalidad de las innovaciones del modelo ARIMA (3,1,2). De este modo, se construye el gráfico de control con la muestra de calibrado, obteniendo la Figura 6.20.

A diferencia del gráfico 6.18, en el gráfico de medidas individuales se puede observar un dato fuera de control, en concreto la observación hace referencia a las 9 de la mañana del día 2. Además, también aparece una racha (6 o más puntos consecutivos por encima o por debajo de la línea central). Como no se encuentra una causa asignable se deja ese fuera de control y se pasan a monitorizar los dos días restantes:

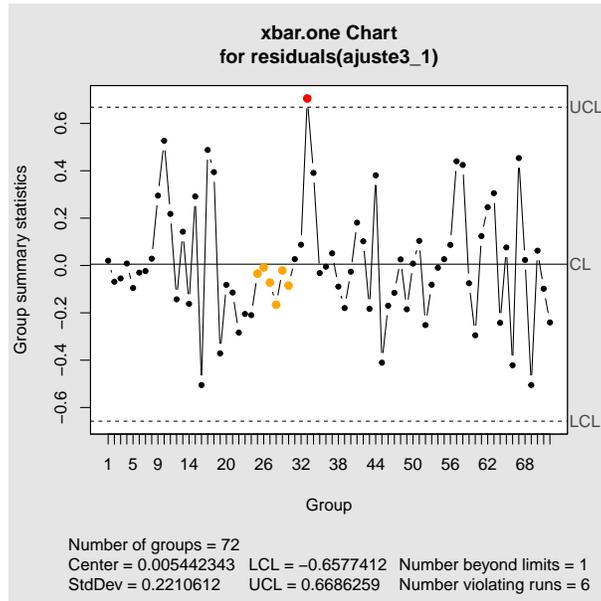


Figura 6.20: Gráfico de medidas individuales para los residuos del modelo ARIMA (3,1,2) de la muestra de calibrado para la temperatura en la oficina CL2.

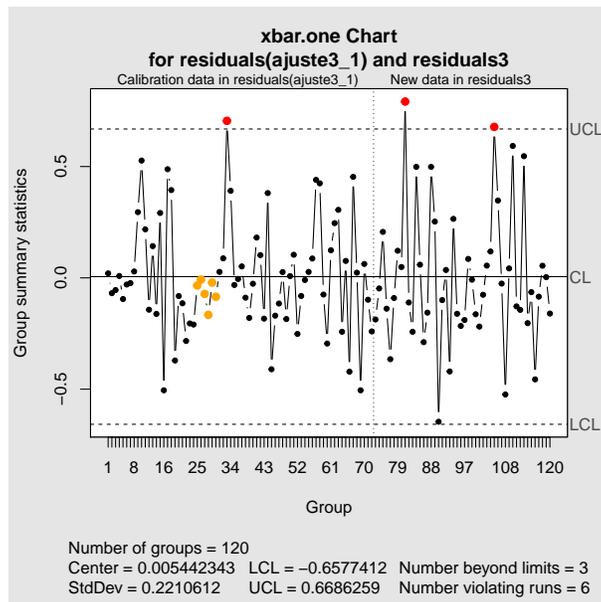


Figura 6.21: Gráfico de medidas individuales para los residuos del modelo ARIMA (3,1,2) de la muestra de calibrado y de monitorizado para la temperatura en la oficina CL2.

Se observan en la muestra de monitorizado dos puntos más fuera de control. Vemos la diferencia que supone la aplicación de los gráficos con memoria (EWMA) que otorgan un mayor peso a las observaciones más cercanas, con respecto a los gráficos clásicos de control de calidad.

6.2.1.2. Análisis de la temperatura ambiente: días festivos

- **Modelización de una serie temporal**

A continuación se pasan a observar los días festivos. El proceso será el mismo que para el anterior caso. En primer lugar se obtienen los promedios horarios de la serie y si el problema de la autocorrelación persiste se trabajará con dicha serie. Posteriormente, se diferencia la serie para convertirla en estacionaria, se ajusta un modelo ARIMA adecuado a la muestra de calibrado (en este caso 4 días) con el que se establecerán los límites de control. Por último se obtienen los residuos de la muestra de monitorizado y se controla el proceso.

Se comienza por representar la serie tanto con promedios cada cinco minutos como horarios, así como sus respectivos ACF (Figuras 6.22 y Figura 6.23).

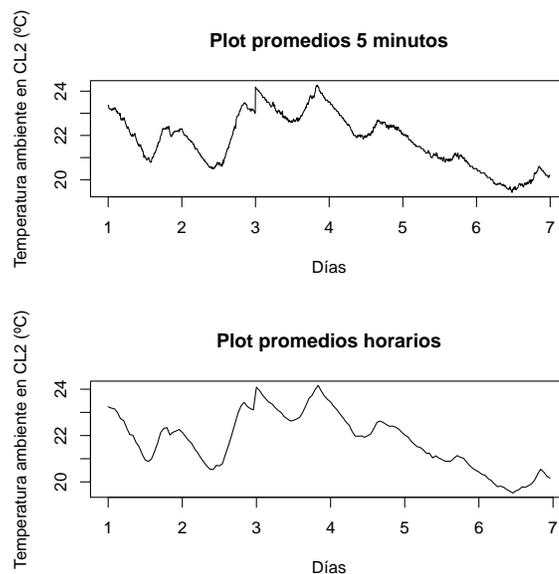


Figura 6.22: Comparación de los gráficos secuenciales para la serie de la temperatura en los festivos promediada cada 5 minutos y cada hora.

Como la tendencia persiste al calcular los promedios de cada hora, se utiliza esa serie para el posterior estudio. Debido a ello, primero se diferencia la serie en la parte regular y se seleccionan las observaciones referentes a la muestra de calibrado.

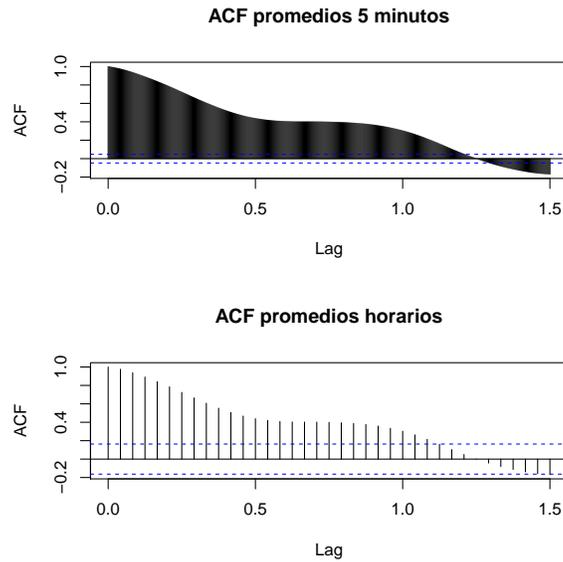


Figura 6.23: Comparación de los ACF para la serie de la temperatura en los festivos promediada cada 5 minutos y cada hora.

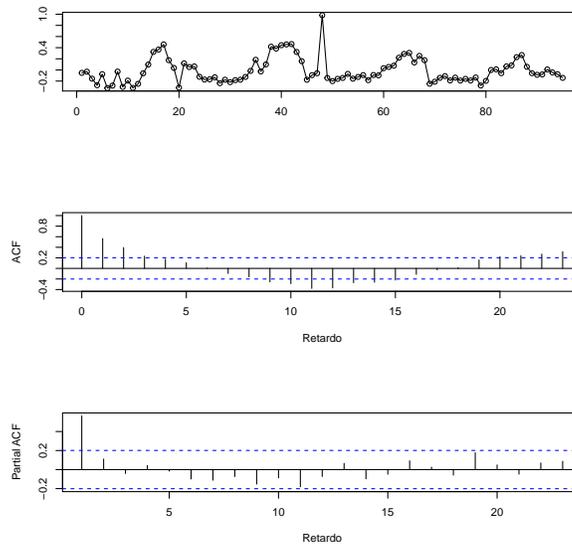


Figura 6.24: Gráfico secuencia, ACF y PACF de la serie diferenciada en la muestra de calibrado de la tempratura de CL2 en los días festivos.

Como en el PACF hay menos estructura, se interpreta ahí y se va a proponer el modelo

AR (1) pues también se obtienen unos valores para los criterios AIC y BIC bajos (391.27 y 396.37 respectivamente). Estimando los parámetros para el modelo ARIMA (1,1,0) por máxima verosimilitud se obtienen:

$$\hat{\phi}_1 = 0,5618 (0,0839), \quad \hat{\sigma}_a^2 = 3,437$$

Con esto, se tiene el modelo:

$$x_t = 0,5618x_{t-1} + a_t$$

siendo 3.437 la varianza del ruido blanco.

Además, se verifican la mayor parte de las hipótesis básicas sobre los residuos, como la independencia (ver la Figura 6.25) y la aceptación de que $\mu_a = 0$. Sin embargo, tal como se observa en la siguiente tabla, no se puede aceptar la normalidad de los residuos.

	P-valores
Test T de media = 0	0.77
Test Shapiro-Wilks	0.00
Test de Jarque-Bera	0.00

Cuadro 6.2: Contrastes sobre la media y la normalidad de los residuos.

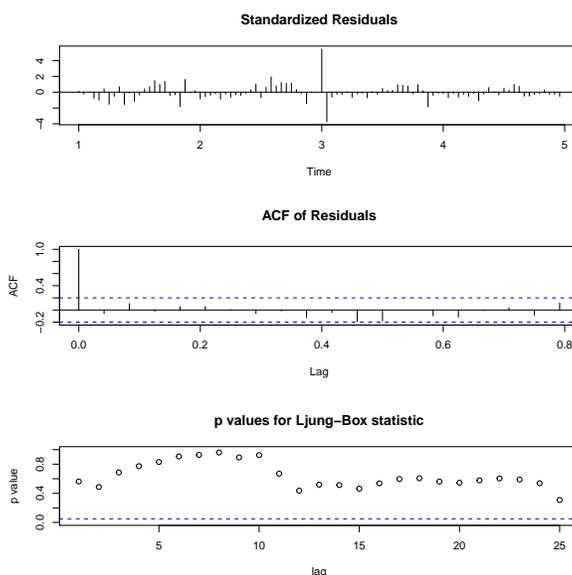


Figura 6.25: Test de independencia para los residuos del modelo ARIMA (1,1,0) ajustado a la muestra de calibrado de la temperatura de CL2 en los días festivos.

Por tanto, el modelo ARIMA (1,1,0) sin constante, se puede considerar válido como generador de la serie de la temperatura en CL2 los días festivos (para la muestra de calibrado con 4 días). Las innovaciones no se pueden considerar normales.

- **Gráficos de control**

Con los residuos incorrelados y de media 0 del modelo ARIMA (1,1,0) ajustado se construye el gráfico EWMA (puesto que no se puede garantizar la normalidad de los mismos no se construyen los gráficos de medidas individuales).

De nuevo, con $\lambda = 0,2$ y una amplitud de 3σ para los límites de control, se obtienen las estimaciones:

$$\hat{\mu}_0 = -0,005483061, \quad \hat{\sigma} = 0,1395454$$

Y con esto se pasa a construir el gráfico EWMA para la muestra de calibrado con 4 días:

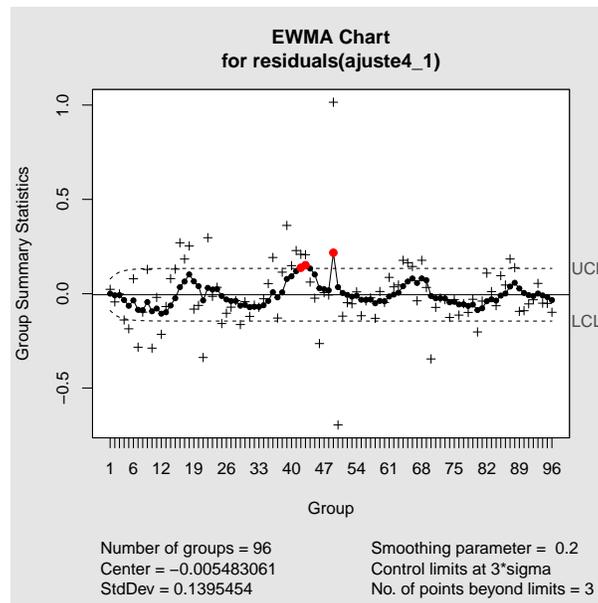


Figura 6.26: Gráfico EWMA para los residuos del modelo ARIMA (1,1,0) ajustado a la serie de la temperatura en la CL2 los días festivos.

Se observa la presencia de 3 datos fuera de control, que pueden obedecer a la propia variación de la tiempo entre el domingo y el jueves (puesto esos 3 puntos fuera que se produce entre las últimas observaciones del domingo y la primera del jueves). Al considerarse como causa común, no se eliminan y se considera que el proceso está bajo control, pasándose de este modo a controlar

el resto de días.

Por ello el siguiente paso es ya calcular los residuos para la muestra de monitorizado en base al modelo ajustado ARIMA (1,1,0) sin constante:

$$\hat{a}_t = x_t - (\hat{\phi}_1 x_{t-1}) = x_t - 0,5618x_{t-1}$$

Se programa su cálculo en R y una vez calculados se monitorizan:

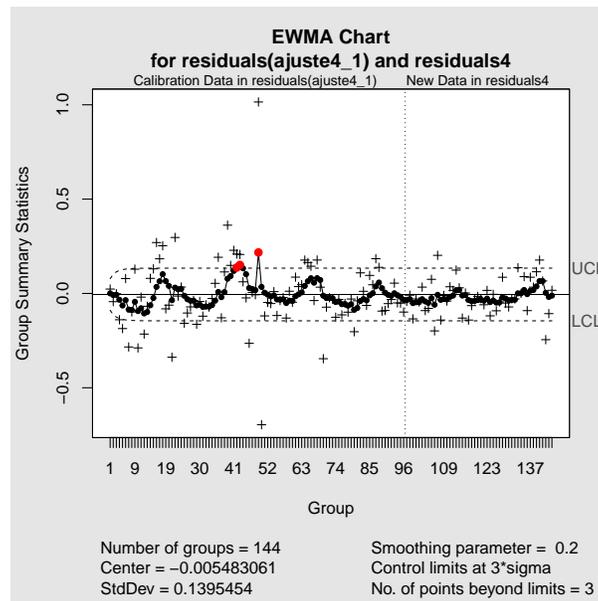


Figura 6.27: Gráfico EWMA para los residuos del modelo ARIMA (1,1,0) de la muestra de calibrado y de monitorizado para la temperatura en la oficina CL2.

Puede observarse la inexistencia de puntos fuera de control, por lo que se puede decir que el proceso está bajo control y la muestra de monitorizado, relativa a los días 19 y 20 de abril, se encuentra dentro de los límites de control estimados en base a la muestra de calibrado.

6.2.2. Análisis de la variable CO_2 en la oficina CL2

6.2.2.1. Análisis del CO_2 : días festivos

- **Modelización de una serie temporal**

A continuación, se pasa a estudiar la concentración de CO_2 los días festivos, siguiendo el mismo procedimiento que para la variable de la temperatura.

Se representa la serie con promedios cada cinco minutos y cada hora así como sus respectivos ACF:

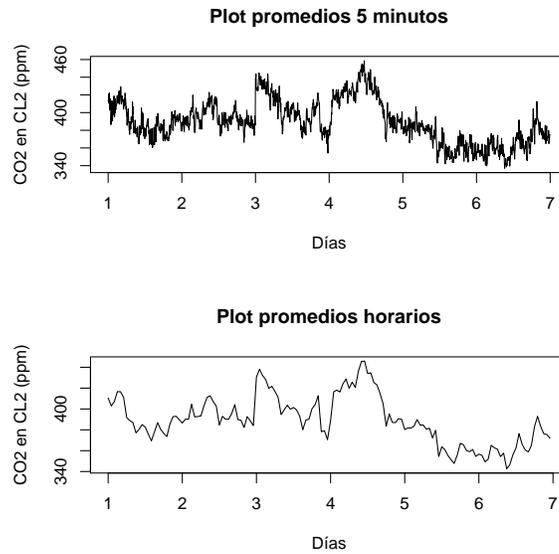


Figura 6.28: Comparación de los gráficos secuenciales para la serie de CO2 promediada cada 5 minutos y cada hora.

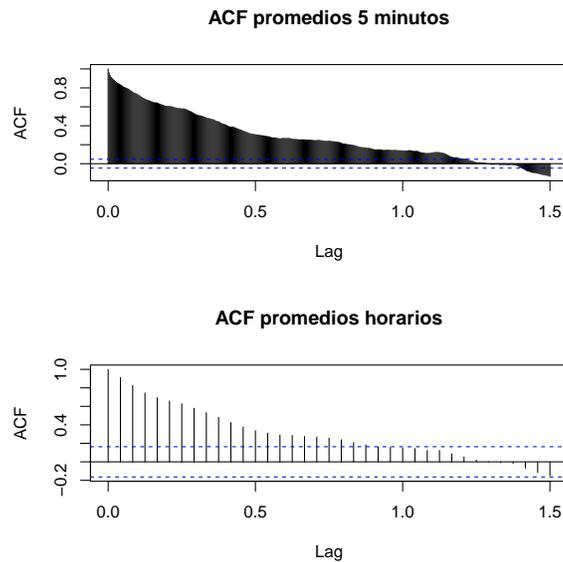


Figura 6.29: Comparación del ACF para la serie de CO2 promediada cada 5 minutos y cada hora.

Igualmente que en casos anteriores, se trabajará con la serie horaria. Como consecuencia de la existencia de tendencia, se diferencia la serie en la parte regular. Se seleccionan a continuación los datos referentes a los 4 primeros días (muestra de calibrado) y se obtienen los siguientes gráficos:

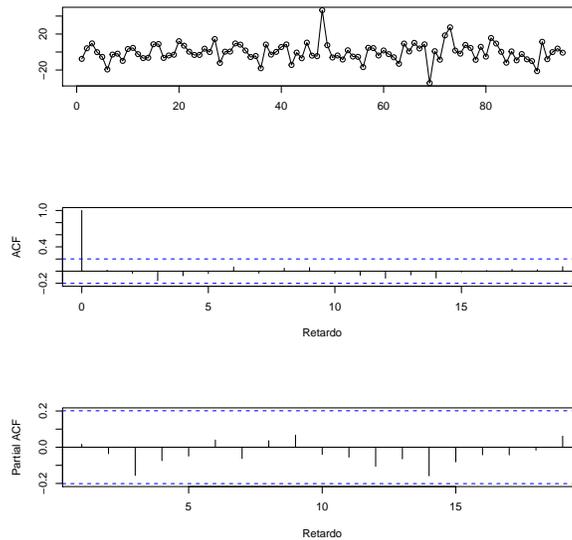


Figura 6.30: Gráfico secuencial, ACF y PACF de la serie diferenciada en la muestra de calibrado del CO_2 de CL2 en los días festivos.

Se observa claramente que el proceso es ya estacionario y que el modelo tentativo puede ser un ARIMA (0,1,0). Así, se ajusta dicho modelo (sin constante pues no es significativamente distinta de 0) y se obtiene:

$$\widehat{\sigma}_a^2 = 104,8$$

Con esto, se tiene el siguiente modelo:

$$x_t = a_t$$

siendo 104.8 la varianza del ruido blanco.

En el siguiente paso, de validación del modelo sugerido, se comprueba la independencia de los residuos (Figura 6.31) y se acepta la hipótesis nula de $\mu_a = 0$. No obstante, las innovaciones no son gaussianas:

	P-valores
Test T de media = 0	0.84
Test Shapiro-Wilks	0.00
Test de Jarque-Bera	0.00

Cuadro 6.3: Contrastes sobre la media y la normalidad de los residuos.

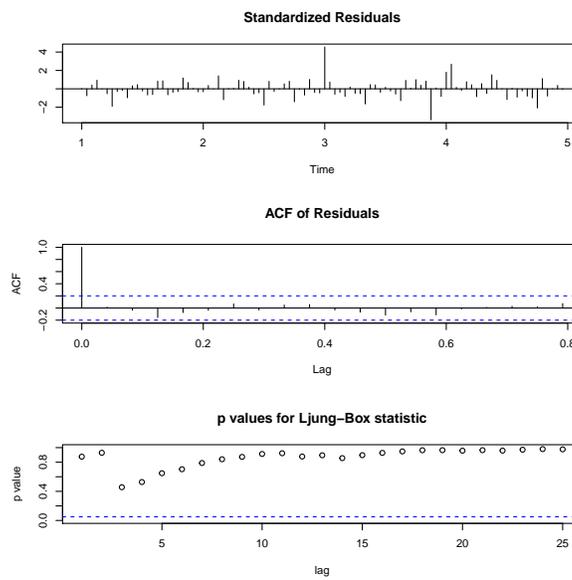


Figura 6.31: Test de independencia para los residuos del modelo ARIMA (0,1,0) ajustado a la muestra de calibrado del CO₂ de CL2 en los días festivos.

Se concluye que el modelo ARIMA (0,1,0) sin constante, se puede considerar válido como generador de la serie de CO₂ en CL2 los días festivos (para la muestra de calibrado con 4 días), sin ser las innovaciones gaussianas.

- **Gráficos de control**

Las estimaciones para el gráfico EWMA en la muestra de calibrado, con $\lambda = 0,2$ y una amplitud de 3σ para los límites de control, son:

$$\hat{\mu}_0 = -0,2099852, \quad \hat{\sigma} = 9,43102$$

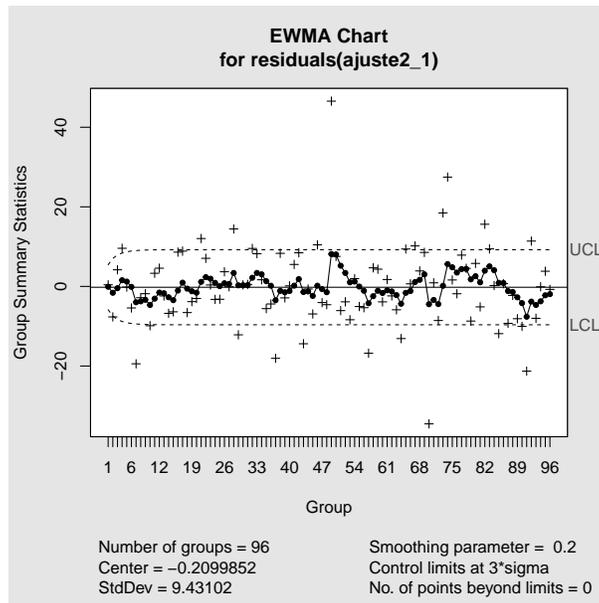


Figura 6.32: Gráfico EWMA para los residuos del modelo ARIMA (0,1,0) ajustado a la serie del CO_2 en la CL2 los días festivos.

El proceso para la muestra de calibrado con los 4 primeros días está en su totalidad bajo control, por lo que se puede pasar directamente al monitorizado del resto de días. De este modo, se calculan los residuos en base al ARIMA (0,1,0) ajustado:

$$\hat{a}_t = x_t$$

Una vez calculados(ver Figura 6.33 en la siguiente página) se observa que en la muestra de monitorizado tampoco existen observaciones fuera de control, por lo que el proceso para la variable CO_2 en los días festivos en la sala CL2 está completamente bajo control.

La aplicación de los gráficos de control EWMA a los residuos de un modelo ARIMA (p,d,q) ha demostrado ser adecuada para la supervisión de los sistemas de control, en este caso, referentes a la temperatura ambiente media y al contenido de CO_2 . Ambos procesos, en los períodos analizados, han mostrado estar bajo control, por no presentar ningún punto fuera de los límites de control con causas asignables.

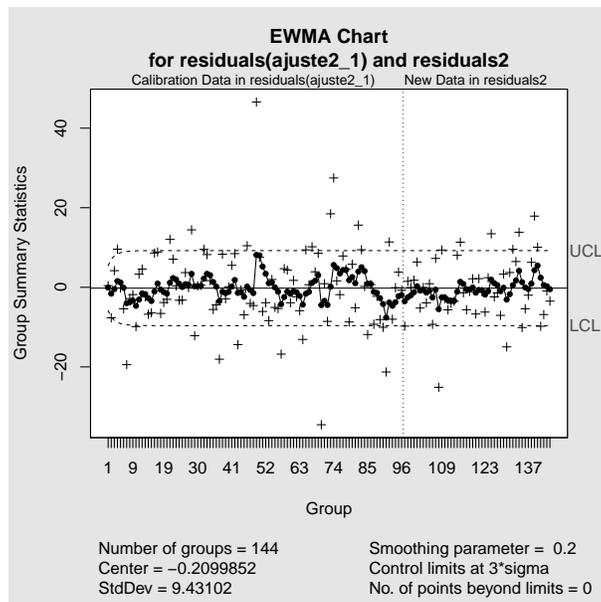


Figura 6.33: Gráfico EWMA para los residuos del modelo ARIMA (0,1,0) de la muestra de calibrado y de monotorizado para el CO_2 en la oficina CL2.

Capítulo 7

Conclusiones

Las herramientas de control estadístico de calidad que se han empleado a lo largo del presente trabajo han resultado útiles para examinar los sistemas de control de los edificios analizados y para llegar a conclusiones sobre la capacidad de los mismos o sobre la conformidad o no de los procesos. Se detallan a continuación los principales aportes y resultados conseguidos:

- Se ha comprobado la importancia de controlar la calidad en las viviendas e instalaciones comerciales, no solamente para cumplir las especificaciones marcadas por la normativa vigente en cuanto a la calidad del aire interior o el confort térmico de los ambientes, sino también para conseguir una mayor satisfacción de los usuarios.
- Se han aplicado índices de calidad para el estudio de la eficiencia energética y la calidad del aire interior. Se ha mostrado además la utilidad de los índices robustos que no asumen la normalidad de los datos, pues en muchas situaciones esta hipótesis es fácilmente violada y los índices clásicos de capacidad no deben ser aplicados.
- Gracias a la aplicación de los índices de capacidad se ha podido comprobar si los sistemas de control para las variables analizadas se adecuan a las especificaciones dadas. Si un sistema de control no es capaz se deben poner en marcha acciones correctoras para su mejora continuada, como pueden ser por ejemplo las propuestas para el caso de los edificios administrativos en Mondoñedo.
- Por otra parte, se ha comprobado que, a la hora de aplicar herramientas para el control de la calidad, como los gráficos de control, es de vital importancia estudiar la posible presencia de autocorrelación en los datos, pues de otro modo, se podría producir un número elevado de falsas alarmas. Una solución para estos casos, como se ha mostrado, es la utilización de las técnicas de series temporales, a través del ajuste de un modelo ARIMA (p,d,q), obteniendo los residuos que serán utilizados para la construcción de los gráficos de control.

- De este modo, se han aplicado los gráficos de control adecuados para el control y monitorizado de los datos, tanto gráficos con memoria (el gráfico EWMA) como los gráficos clásicos de control (de medidas individuales), permitiendo identificar procesos bajo control.

7.1. Líneas futuras

El desarrollo de un trabajo como el presente está siempre abierto a extensiones y a la aplicación de otras herramientas e instrumentos que contribuyan a ampliar los resultados aquí expuestos. Resulta de interés un análisis ampliado y más profundo de la metodología utilizada a lo largo de esta memoria, lo que podría constituir futuras líneas de investigación.

En cuanto a los datos referentes a la plataforma de la empresa Fridama se enumeran las siguientes líneas:

- Una primera línea pasaría por el uso de las series temporales estacionales para modelizar las series con dependencia de este tipo, como podría ser el caso del consumo de energía de sus instalaciones, ya que requiere un estudio más pormenorizado.
- El control de otras variables relacionadas con la eficiencia energética como la iluminación son importantes, ya que influyen en gran medida en el consumo de energía. También se podrían tener en cuenta las condiciones exteriores e interiores de los edificios para llevar a cabo un análisis más exhaustivo, así como las costumbres de los propios trabajadores de las instalaciones, pues todo influye en el gasto energético.
- Otra línea a aplicar sería la separación de los datos entre horas de ocupación y horas de no ocupación, en vez de laborables y festivos, pues se podrían encontrar patrones diferentes, ya que el consumo energético y los valores e las variables indicadoras de confort térmico y calidad del aire interior no son las mismas en horas de actividad en la empresa que en horas de cierre.
- También se podrían estudiar posibles relaciones entre las variables, como por ejemplo entre la temperatura y la iluminación, o la iluminación y la radiación solar exterior, entre otras.

Por último, en cuanto los datos relativos a los edificios administrativos de Mondoñedo, este estudio de capacidad se podría hacer extensible a otros edificios administrativos de Galicia.

Apéndice A

Código R

Se señala aquí parte del código R implementado para el análisis de los datos realizados en el presente trabajo.

```
#####
##                ANÁLISIS MONDOÑEDO                ##
#####

# Se señala primero el código utilizado para los edificios de
  Mondoñedo.

#-----#
#----- ANÁLISIS DE COMPONENTES PRINCIPALES -----#
#-----#

### Se abren los datos y se realiza un pequeño estudio descriptivo con
  los mismos

library(qcc)
verano<-read.csv2("c:/mondoñedo.verano3.csv",header=T,dec=".",sep=";")
)
verano<-verano[,1:11]

# Resumen numérico: incluir todo con la función:

library(Rcmdr)
```

```

resumen<-numSummary(verano [4:9] , statistics=c( "mean" , "sd" , "IQR" , "
  quantiles" ) , quantiles=c( 0 , 0.25 , 0.5 , 0.75 , 1 ) )
colnames(resumen$table)<-c( "Media" , "SD" , "IQR" , "Cuantil_0" , "Q1" , "Q2" , "
  Q3" , "Cuantil_1" )
resumen$table

# Vemos que no son normales con el test de Shapiro Wilks

pvalores<-rbind(c( shapiro.test(temp.media)$p.value , shapiro.test(
  humedad.rel.media)$p.value , shapiro.test(co2.inicial)$p.value ,
  shapiro.test(co2.final)$p.value , shapiro.test(co2.medio)$p.value ,
  shapiro.test(ocupacion)$p.value ) )
colnames(pvalores)<-c( "Temp._media" , "Humedad_rel._media" , "CO2_inicial
  " , "CO2_final" , "CO2_medio" , "Ocupación" ) ; rownames(pvalores)<- "P-
  valores"
pvalores

# Para ver los valores fuera del rango del confort térmico para la
  humedad:

Hora_inicial<-hora.inicial [humedad.rel.media<=40]
Fecha<-fecha [humedad.rel.media<=40]
Sala<-sala [humedad.rel.media<=40]
Humedad<-humedad.rel.media [humedad.rel.media<=40]
Ocupacion<-ocupacion [humedad.rel.media<=40]
atipicos_hum<-data.frame(Fecha , Hora_inicial , Sala , Humedad , Ocupacion)

# Se procede igual para el resto de variables (temperatura y CO2)

# ¿Vienen el co2 inicial y final de la misma distribución?

ks.test(co2.inicial , co2.final)

# Correlación entre variables:

library(Hmisc)
corr_conjunta<-rcorr(as.matrix(verano [c(4,5,8,9)] ) , type="spearman")
corr_conjunta$r
corr_conjunta$P

```

```
#### ANÁLISIS DE COMPONENTES PRINCIPALES (princomp)

model <- princomp (verano [c(4,5,8,9)], cor=T)
resumen_modelo<-summary(model, loadings=T)

resumen_modelo$loadings[, ] #autovectores
model$sdev^2 # varianzas de los componentes: autovalores
puntuaciones<-model$scores[, 1:2]

# Las puntuaciones:
plot(model$scores[, 1], model$scores[, 2], xlab="Componente_1", ylab="
  Componente_2", type="n", lwd=2)
points(model$scores[, 1:2], col=c("orange", "red", "lightblue")[as.
  numeric((verano[, 10]))], pch=19)
nombres.filas=c("Ayuntamiento", "Casa_de_Cultura", "Escuela_de_Música")
legend(2.5, 0, nombres.filas, fill=c("orange", "red", "lightblue"), cex
  =0.8)

# Biplot:
biplot(model, choices = 1:2)

## Comunalidades:
cor_model<-cor(verano [c(4,5,8,9)], predict(model))

# Para la 1ª variable:
a<-sum(cor_model[1, 1:2]^2)

# Para la 2ª variable:
b<-sum(cor_model[2, 1:2]^2)

# Para la 3ª variable:
c<-sum(cor_model[3, 1:2]^2)

# Para la 4ª variable:
d<-sum(cor_model[4, 1:2]^2)

tab<-data.frame(a, b, c, d)

\newpage
```

```

#-----#
#----- ÍNDICES DE CAPACIDAD -----#
#-----#

# Se calculan y programan aquí los índices de capacidad para la base
# de datos de Mondoñedo, tanto los robustos como los que asumen
# normalidad. Solamente se muestra el código para la temperatura
# media en verano en el Ayuntamiento. Para el resto de variables,
# época y edificios administrativos (escuela de música y casa de
# cultura) simplemente se sustituirán los elementos.

##### ----- AYUNTAMIENTO: VERANO ----- #####

# Se abre la base de datos correspondiente

library(qcc)
library(qualityTools)

# Ahora seleccionamos solo los momentos de ocupación (de 8 a 15 horas
# )
ayuntamiento<-data.frame(verano[edificio=="Ayuntamiento",])
ocupacion_ayuntamiento<- data.frame(ayuntamiento[(ayuntamiento$hora.
  inicial >=8)&(ayuntamiento$hora.inicial <15),])

# Cogemos la temperatura
ayuntamiento.temp<-cbind(ocupacion_ayuntamiento$temp.media)
shapiro.test(ocupacion_ayuntamiento$temp.media) # no es normal
ayuntamiento.temp2<-sample(ayuntamiento.temp)

### índices de capacidad que asumen normalidad:

Target=24.5; LSL=22; USL=27; d=(USL-LSL)/(2); m=(USL+LSL)/(2)
pcr(ayuntamiento.temp2, lsl=LSL, usl= USL, target=Target, main="")

### Índices de capacidad robustos:

F2=as.numeric(quantile(ayuntamiento.temp2, probs=0.99865)) #
  calculamos los cuantiles
F1=as.numeric(quantile(ayuntamiento.temp2, probs=0.00135))

```

```

u=0;v=0
CN_p=(d-u*abs ( median ( ayuntamiento . temp2)-m ) ) / ( 3 * sqrt ( ( ( F2-F1 ) / 6 ) ^ 2 + v
      * ( median ( ayuntamiento . temp2 ) - Target ) ^ 2 ) )
CN_p

u=1;v=0
CN_pk=(d-u*abs ( median ( ayuntamiento . temp2)-m ) ) / ( 3 * sqrt ( ( ( F2-F1 ) / 6 ) ^ 2 + v
      * ( median ( ayuntamiento . temp2 ) - Target ) ^ 2 ) )
CN_pk

u=0;v=1
CN_pm=(d-u*abs ( median ( ayuntamiento . temp2)-m ) ) / ( 3 * sqrt ( ( ( F2-F1 ) / 6 ) ^ 2 + v
      * ( median ( ayuntamiento . temp2 ) - Target ) ^ 2 ) )
CN_pm

u=1;v=1
CN_pmk=(d-u*abs ( median ( ayuntamiento . temp2)-m ) ) / ( 3 * sqrt ( ( ( F2-F1 ) / 6 ) ^ 2 +
      v * ( median ( ayuntamiento . temp2 ) - Target ) ^ 2 ) )
CN_pmk

Resultado1=data . frame ( c ( CN_p , CN_pk , CN_pm , CN_pmk ) )
colnames ( Resultado1 ) = c ( " Temperatura _ media " ) ; rownames ( Resultado1 ) = c ( "
      CN_p " , " CN_pk " , " CN_pm " , " CN_pmk " )
Resultado1

#####
##          ANÁLISIS FRIDAMA          ##
#####

# A continuación se señala el código utilizado para la base de
# datos de la empresa Fridama.

#-----#
#----- ANÁLISIS DE DATOS FUNCIONALES -----#
#-----#

library ( mgcv )
library ( fda . usc )

```

```

# Se muestran el código para la variable CO2 en la oficina CL2:

# Ambrimos los datos con el nombre de "Datos":

#### Primero seleccionamos los días laborables

Datos=read.table("c:/Archivo24Dias.txt", header=TRUE, sep="\t", na.
  strings="NA", dec=",", strip.white=TRUE)
Datos=Datos[-c(4321:dim(Datos)[1]),] # 15 días completos

x=Datos[1:(1440/5), "Tiempo"] # Eje de tiempos en minutos

x.co2=Datos[Datos$Regimen=="Lab",c("CL2_CO2","Dia")]
CL2_CO2<-x.co2$CL2_CO2/10 # dividimos entre 10 para que se mide en
  ppm
x.co2<-data.frame(CL2_CO2,x.co2$Dia)

n=dim(x.co2)[1]*5/1440; # 9 días laborables

# Se adaptan en una matriz para que cada fila sea una curva relativa
  a un día completo:
a=1;xt=NULL;b=1
for(i in 1:n){
  b=cbind(b,i*1440/5+1)
  xt=rbind(x.co2[a:(i*1440/5),1],xt)
  a=i*1440/5+1;
}

# La matriz la convierto en un data frame con la última columna
  haciendo alusión al día al que se corresponden los datos:
co2_l=as.data.frame(cbind(xt,as.character(x.co2[b[c
  (9,8,7,6,5,4,3,2,1)],2]))))

# La base de datos pasa a ser un objeto fdata:
co2_l_fda=fdata(xt, argvals=x)

#### Para los días festivos:

y.co2=Datos[Datos$Regimen=="Fest",c("CL2_CO2","Dia")]

```

```

CL2_CO2<-y.co2$CL2_CO2/10
y.co2<-data.frame(CL2_CO2,y.co2$Dia)
n2=dim(y.co2)[1]*5/1440;

a=1;yt=NULL;b=1
for(i in 1:n2){
  b=cbind(b,i*1440/5+1)
  yt=rbind(y.co2[a:(i*1440/5),1],yt)
  a=i*1440/5+1;
}

co2_f=as.data.frame(cbind(yt,as.character(y.co2[b[c(6,5,4,3,2,1)],2])
  ))
co2_f_fda=fdata(yt, argvals=x)

##### Suavizado con estimador Nadaraya Watson

# Vector de posibles ventanas
h = seq(0.01, 0.5, len = 50)

# SUAVIZAR y_fda (días LABORABLES: 9 días)

out1<-min.np(co2_l_fda$data[1,],type.CV=GCV.S,type.S=S.NW,criteria="
  Rice",h=h,par.CV=list(trim=0.01,draw=F)) # dia festivo 9 (viernes
  25)
out2<-min.np(co2_l_fda$data[2,],type.CV=GCV.S,type.S=S.NW,criteria="
  Rice",h=h,par.CV=list(trim=0.01,draw=F)) #dia festivo 8 jueves 24
out3<-min.np(co2_l_fda$data[3,],type.CV=GCV.S,type.S=S.NW,criteria="
  Rice",h=h,par.CV=list(trim=0.01,draw=F)) # miercoles 23
out4<-min.np(co2_l_fda$data[4,],type.CV=GCV.S,type.S=S.NW,criteria="
  Rice",h=h, par.CV=list(trim=0.01,draw=F)) # martes 22
out5<-min.np(co2_l_fda$data[5,],type.CV=GCV.S,type.S=S.NW,criteria="
  Rice",h=h,par.CV=list(trim=0.01,draw=F)) # lunes 21
out6<-min.np(co2_l_fda$data[6,],type.CV=GCV.S,type.S=S.NW,criteria="
  Rice",h=h, par.CV=list(trim=0.01,draw=F)) # miercoles 16
out7<-min.np(co2_l_fda$data[7,],type.CV=GCV.S,type.S=S.NW,criteria="
  Rice",h=h,par.CV=list(trim=0.01,draw=F)) # martes 15
out8<-min.np(co2_l_fda$data[8,],type.CV=GCV.S,type.S=S.NW,criteria="
  Rice",h=h, par.CV=list(trim=0.01,draw=F)) #lunes 14

```

```

out9<-min.np(co2_l_fda$data[9,], type.CV=GCV.S, type.S=S.NW, criteria="
  Rice", h=h, par.CV=list(trim=0.01, draw=F)) # dia festivo 1 (viernes
  11)

# Datos y_fda suavizados:

co2_l_fda_smoothed=fdata(rbind(out9$fdata.est$data, out8$fdata.
  est$data, out7$fdata.est$data, out6$fdata.est$data, out5$fdata.
  est$data, out4$fdata.est$data, out3$fdata.est$data, out2$fdata.
  est$data, out1$fdata.est$data), argvals=x)

# Curvas funcionales:
plot.fdata(co2_l_fda_smoothed, xlab="Tiempo/_min", main="Días_
  laborables", ylab="Contenido_en_CO2_en_la_zona_CL2/_ppm", col=c("
  black", "red", "slateblue4", "green", "mediumturquoise", "magenta", "
  snow4", "tan4", "orange1"), lty=1)
dias.filas<-c("Viernes_11", "Lunes_14", "Martes_15", "Miercoles_16", "
  Lunes_21", "Martes_22", "Miercoles_23", "Jueves_24", "Viernes_25")
legend("topright", dias.filas, fill=c("black", "red", "slateblue4", "green
", "mediumturquoise", "magenta", "snow4", "tan4", "orange1"), border=c("
  black", "red", "slateblue4", "green", "mediumturquoise", "magenta", "
  snow4", "tan4", "orange1"))

# SUAVIZAR z_fda (días FESTIVOS: 6 días)

out11<-min.np(co2_f_fda$data[1,], type.CV=GCV.S, type.S=S.NW, criteria="
  Rice", h=h, par.CV=list(trim=0.01, draw=F))
out22<-min.np(co2_f_fda$data[2,], type.CV=GCV.S, type.S=S.NW, criteria="
  Rice", h=h, par.CV=list(trim=0.01, draw=F))
out33<-min.np(co2_f_fda$data[3,], type.CV=GCV.S, type.S=S.NW, criteria="
  Rice", h=h, par.CV=list(trim=0.01, draw=F))
out44<-min.np(co2_f_fda$data[4,], type.CV=GCV.S, type.S=S.NW, criteria="
  Rice", h=h, par.CV=list(trim=0.01, draw=F))
out55<-min.np(co2_f_fda$data[5,], type.CV=GCV.S, type.S=S.NW, criteria="
  Rice", h=h, par.CV=list(trim=0.01, draw=F))
out66<-min.np(co2_f_fda$data[6,], type.CV=GCV.S, type.S=S.NW, criteria="
  Rice", h=h, par.CV=list(trim=0.01, draw=F))

# Datos z_fda suavizados:

```

```

co2_f_fda_smoothed=fdata(rbind(out66$fdata.est$data , out55$fdata.
  est$data , out44$fdata.est$data , out33$fdata.est$data , out22$fdata.
  est$data , out11$fdata.est$data) , argvals=x) #poner out4$fdata.
  est$data ,

#Curvas funcionales :
plot.fdata(co2_f_fda_smoothed , xlab="Tiempo/_min" , main="Días_festivos"
  , ylab="Contenido_en_CO2_en_la_zona_CL2/_ppm" , col=c("black" , "red" ,
  "royalblue" , "olivedrab3" , "mediumturquoise" , "magenta") , lty=1)
dias.filas<-c("Sabado_13" , "Domingo_14" , "Jueves_17" , "Viernes_18" , "
  Sabado_19" , "Domingo_20")
legend("topright" , dias.filas , fill=c("black" , "red" , "royalblue" , "
  olivedrab3" , "mediumturquoise" , "magenta") , border=c("black" , "red" , "
  royalblue" , "olivedrab3" , "mediumturquoise" , "magenta"))

#### Estadística descriptiva: medias funcionales y bandas de
  confianza bootstrap ####

Media_laborables = fdata.bootstrap(co2_l_fda_smoothed , statistic=func.
  mean , nb=500 , draw=TRUE)
Media_festivos = fdata.bootstrap(co2_f_fda_smoothed , statistic=func.
  mean , nb=500 , draw=TRUE)

##### Análisis ANOVA Funcional con proyecciones aleatorias

Curvas.fdata=rbind(co2_l_fda_smoothed$data , co2_f_fda_smoothed$data)
Curvas.fdata=fdata(Curvas.fdata , argvals=x)
data=as.data.frame(Curvas.fdata$data)
group=c(rep('Laborable' , 9) , rep('Festivo' , 6))
group=as.factor(group)
n=nrow(data)
RP=c(2 , 5 , 15 , 30)
m0=data.frame(group)
resul=anova.RPm(data , ~group , m0)
summary.anova(resul)

# Las curvas de los festivos y laborables son significativamente
  distintas de 0

```

```

#-----#
#----- SERIES DE TIEMPO + GRÁFICOS DE CONTROL -----#
#-----#

# Aquí se modelizan las series para cada variable , diferenciando por
# laborable festivo , si son diferentes , y se calculan los gráficos
# de control con los residuos del modelo ajustado .

# Se señalan los cálculos para el CO2 los días festivos (para el
# resto de variables sería sustituir y, en el caso del cálculo de
# los residuos para la muestra de monitorizado en base al modelo
# ajustado para la de calibrado ,cambiar ligeramente el código
# programado)

# Se abren los datos y se seleccionan los referentes al CO2 los días
# festivos

library(forecast)
library(qcc)
library(tseries)

Datos=read.table("c:/Archivo24Dias.txt", header=TRUE, sep="\t",na.
strings="NA", dec=",", strip.white=TRUE)
Datos=Datos[-c(4321:dim(Datos)[1]),] # nos quedamos con 15 días
# completos

y.co2=Datos [Datos$Regimen=="Fest" ,c("CL2_CO2" ,"Dia")]
CL2_CO2<-y.co2$CL2_CO2/10
y.co2<-data.frame(CL2_CO2,y.co2$Dia)
CL2_CO2<-ts(y.co2$CL2_CO2 ,frequency=288)

# Con ellos se calculan el gráfico EWMA para los 4 dias de calibrado

q <- ewma(y.co2$CL2_CO2[1:1152] , lambda=0.20) # todo fuera de
# control

# Se grafica la serie entera (9 días):
CL2_CO2<-ts(y.co2$CL2_CO2 ,frequency=288)

```

```
plot(CL2_CO2, type="l", xlab="Días", ylab="CO2_en_CL2_(ppm)")

# Se comprueba la existencia de autocorrelación:
acf(CL2_CO2, lag.max=length(CL2_CO2)/4)

# Se calculan los promedios horarios pues trabajar con promedios cada
  5 minutos
# dificulta los cálculos

j=seq(1,1728,12)
medias2=matrix(NA)

for(i in j){
  medias2[i]<-mean(y.co2$CL2_CO2[i:(i+11)])
}
co2fest_1<-medias2[j] # 144 observaciones
co2fest_1<-ts(co2fest_1, frequency=24)

# De la serie con promedios horarios se selecciona la MUESTRA DE
  CALIBRADO (3 días)
serie2_1<-ts(co2fest_1[1:96], frequency=24)# serie datos originales 4
  dias

# Además, se diferencia la serie entera (5 días) para eliminar la
  tendencia
serie2_1entera<-ts(co2fest_1, frequency=24)
dif.serie2_1entera.ts <- diff(serie2_1entera,1)
acf(dif.serie2_1entera.ts, lag.max=length(dif.serie2_1entera.ts)/4,
  main="")

# De los datos diferenciados cogemos los referentes a los 3 días para
  calibrado
dif.serie2_1.ts<-dif.serie2_1entera.ts[1:95]# cogemos 4 dias para
  hacer muestra calibrado
par(mfrow=c(3,1))
plot(dif.serie2_1.ts, type="o", xlab="", ylab="")
acf(dif.serie2_1.ts, xlab="Retardo", main="")
pacf(dif.serie2_1.ts, xlab="Retardo", main="")
```

```
# Mirando el acf y pacf se propone un modelo tentativo:
ajuste2_1 <- Arima(serie2_1 , order=c(0,1,0),include.constant=T)
ajuste2_1

# Cte no significativa , se elimina:
ajuste2_1 <- Arima(serie2_1 , order=c(0,1,0))
ajuste2_1

### Se verifican las hipótesis para los residuos:

# Independencia:

tsdiag(ajuste2_1 , gof.lag=25) # cumple

# QQ-normal:
plot(residuals(ajuste2_1) , type="o" , xlab="Tiempo" , ylab="Residuos")
abline(h=0)
qqnorm(residuals(ajuste2_1))
qqline(residuals(ajuste2_1))

# Media=0
t.test(residuals(ajuste2_1) , mu=0) # cumple H0

# Normalidad:
jarque.bera.test(residuals(ajuste2_1)) #no normalidad
shapiro.test(residuals(ajuste2_1))

### Gráfico de control EWMA

# Ahora que se ha visto que ARIMA(0,1,0) es válido se calcula el EWMA
q <- ewma(residuals(ajuste2_1) , lambda=0.2,nsigmas=3)

# Necesitamos ahora calcular los residuos de los 2 días siguientes a
través del modelo
# ajustado (ARIMA(0,1,0)) SIN constante.

dif.serie2_1b.ts<-dif.serie2_1entera.ts[96:143] #se cogen los datos
de los 2 días

# se calculan los residuos:
```

```

residuals2<-dif.serie2_1b.ts

q <- ewma(residuals(ajuste2_1), lambda=0.2,newdata=residuals2)

#Proceso bajo control

### Gráfico medidas individuales:

qcc(residuals(ajuste2_1), type="xbar.one",newdata=residuals2) # 2
    fuera de control

# No recomendable ya que los residuos NO son normales.

### NOTAS:

## Para calcular por ejemplo los residuos del modelo ARIMA (3,1,2)
    que se ha ajustado a la TEMPERATURA AMBIENTE los días LABORABLES
    se programaria como sigue (se omiten os cálculos para el ajuste
    del arima a la muestra de calibrado para no repetir)

# Ahora debemos calcular los residuos para los 2 ultimos días a
    partir del modelo ajustado ARIMA (3,1,2) sin constante:
#  $X_t = \phi_1 * X_{t-1} + \phi_2 * X_{t-2} + \phi_3 * X_{t-3} + \theta_1 * a_{t-1} + \theta_2 * a_{t-2} + a_t$ 
#  $a_t = X_t - (\phi_1 * X_{t-1} + \phi_2 * X_{t-2} + \phi_3 * X_{t-3} + \theta_1 * a_{t-1} + \theta_2 * a_{t-2})$ 

todos_at<-matrix(NA)
xt<-matrix(NA)

for (i in 73:120){
  xt[i]<- dif.serie3_1entera.ts[i-1]
  xt[i-1]<- dif.serie3_1entera.ts[i-2]
  xt[i-2]<- dif.serie3_1entera.ts[i-3]
  xt[i-3]<- dif.serie3_1entera.ts[i-4]
  todos_at[1:72]<-residuals(ajuste3_1)
  phi_1 <- ajuste3_1$coef[1]
  phi_2<-ajuste3_1$coef[2]
  phi_3<-ajuste3_1$coef[3]
  theta_1<-ajuste3_1$coef[4]
}

```

```
theta_2<-ajuste3_1$coef[5]
todos_at[i]<- xt[i]-(phi_1*xt[i-1]+phi_2*xt[i-2]+phi_3*xt[i-3]+
  theta_1*todos_at[i-1]+theta_2*todos_at[i-2])
}
todos_at

residuals3 <- todos_at[73:120]
q <- ewma(residuals(ajuste3_1), lambda=0.2,newdata=residuals3)

# De este modo, ajustando modelos de series temporales a las series
  que muestren autocorrelación, y calculando los residuos para la
  muestra de monitorizado en base al modelo ajustado para la de
  calibrado, se pueden obtener los gráficos ewma y otros gráficos de
  control.
```

Bibliografía

- [1] Alwan, L.C., y Bissell, M.G. (1988). Time Series Modeling for Quality Control in Clinical Chemistry, *Clinical Chemistry*, 34 (7), pp. 1396-1406.
- [2] Box, G.E.P., y Kramer, T. (1992) Statistical Process Monitoring and Feedback Adjustment. A Discussion, *Technometrics*, 34 (3), 251-267.
- [3] Box, G.E.P., Jenkins G.M., Reinsel G.C. (1994). *Time Series Analysis, Forecasting, and Control*, 3º ed., Prentice-Hall, Englewood Cliffs, New Jersey.
- [4] Box, G. E. P., Luceño A. , y Paniagua-Quiñones M. (2009). *Statistical Control by Monitoring and Feedback Adjustment*, 2º ed., Wiley, New York.
- [5] Boyles, R. A. (2000). Phase I Analysis for Autocorrelated Processes, *Journal of Quality Technology*, 32 (4), pp. 295-409.
- [6] Chen, K.S., Ouyang, L.Y., Hsu, C.H., y Wu, C.C. (2009). The communion bridge to Six Sigma and process capability indices, *Quality Quantity*, 43, pp. 463-469.
- [7] Cuesta-Albertos, J.A.,y Febrero-Bande, M. (2010). A simple multiway ANOVA for functional data, *TEST*, 19, pp. 537-557.
- [8] Febrero-Bande, M., y Oviedo de la Fuente, M. (2012). Statistical Computing in Functional Data Analysis: The R Package fda.usc, *Journal of Statistical Software*, 51 (4), pp. 1-28.
- [9] Garvin, D. A. (1987). Competing on the eight dimensions of quality, *Harvard Business Review*, 87 (6), pp. 101-109.
- [10] Hunter, J.S. (1986). The Exponentially Weighted Moving Average, *Journal of Quality Technology*, 18 (4) , pp. 203-210.
- [11] IDAE (Instituto para la Diversificación y Ahorro de la Energía). (2011). *Plan de ahorro y eficiencia energética 2011-2020. 2º plan de acción nacional de eficiencia energética en España 2011-2020*, Madrid.

- [12] Kane, V.E. (1986). Process capability indices, *Journal Of Quality Technology*, 18 (1), pp. 41-52.
- [13] Kramer, H., y Schmid, W. (1997). Control charts for time series, *Nonlinear Analysis, Theory, Methods Applications*, 30 (7), pp. 4007-4016.
- [14] Lucas, J.M., y Saccucci, M.S., (1990). Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements, *Technometrics*, 32 (1), pp. 1-12.
- [15] Montgomery, D.G. (2009). *Introduction to Statistical Quality Control*, 6° ed., Wiley, New York.
- [16] Morales I.M., Blanco, V.A., García Nieto, A. (2010). *Calidad del aire interior en edificios de uso público*, Dirección General de Ordenación e Inspección, y Consejería de Sanidad de la Comunidad de Madrid.
- [17] Patterson, M.G. (1996). What is energy efficiency? Concepts, indicators and methodological issues, *Energy Policy*, 24 (5), pp. 377-390.
- [18] Pearn, W. L., S. Kotz, y N. L. Johnson (1992). Distributional and Inferential Properties of Process Capability Indices, *Journal of Quality Technology*, 24(4), pp. 216-231.
- [19] Pearn W.L., y Chen K.S. (1997a). An application of the non-normal process capability indices, *Quality and Reliability Engineering International* , 13, pp. 355-360.
- [20] Pearn W.L., y Chen K.S. (1997b). Capability indices for non-normal distributions with an application in electrolytic capacitor manufacturing, *Microelectronics Reliability* , 37 (12), pp. 1853-1858.
- [21] Pearn, W. L., Wang, F.K y Yen, C.H. (2007). Multivariate Capability Indices: Distributional and Inferential Properties, *Journal of Applied Statistics*, 34 (8), pp. 941-962.
- [22] Peña, D. (2002). *Análisis de datos multivariantes*, McGraw-Hill, Madrid.
- [23] Peña, D. (2005). *Análisis de series temporales*, Alianza Editorial, Madrid.
- [24] Psarakis, S., y Papaleonida, G.E.A. (2007). SPC Procedures for Monitoring Autocorrelated Processes, *Quality Technology Quantitative Management* , 4 (4), pp. 501-540.
- [25] Ramsay, J.O. and Silverman, B.W. (2002). *Applied Functional Data Analysis: Methods and case studies*, Springer , New York.
- [26] Rey, F.J., y Velasco, E. (2006). *Eficiencia energética en edificios. Certificación y auditorías energéticas*, Thomson, Madrid.
- [27] Schroeder, R.G., Linderman, K., Liedtke, C., y Choo, A.S. (2008). Six Sigma: Definition and underlying theory, *Journal of Operations Management* , 26, pp. 536-554.

- [28] Shao, E.Y., y Lin, Y. (2013). Applying Residual Control Charts to Identify the False Alarms in a TFT-LCD Manufacturing Process, *Applied Mathematics Information Sciences*, 7 (4), pp. 1459-1464.
- [29] Shewhart, W.A. (1997). *Control económico de la calidad de productos manufacturados*, Ediciones Díaz de Santos, Madrid.
- [30] Vännman, K. (1995). A Unified Approach to Capability Indices, *Statistica Sinica*, 5, pp. 805-820.
- [31] Western Electric Handbook. (1956), *Statistical Quality Control Handbook*, Western Electric Company, Pennsylvania.
- [32] Wu, C.W, Pearn, W.L., y Kotz, S. (2009). An overview of theory and practice on process capability indices for quality assurance, *International Journal of Production Economics*, 117, pp. 338-359.
- [33] Young, T. M., y Winistorfer, P.M. (2001). The effects of autocorrelation on real-time statistical process control with solutions for forest products manufacturers, *Forest Products Journal*, 51, pp. 70-77.
- [34] Zhang, N.F, (1997) Detection capability of residual control chart for stationary process data, *Journal of Applied Statistics* , 24 (4), pp. 475-492.
- [35] Zhang, N.F, (1998) A Statistical Control Chart for Stationary Process Data, *Technometrics*, 40 (1), pp. 24-38.