# Control y detección de anomalías de confort térmico y eficiencia energética en edificios a partir de variables autocorreladas

Inés Barbeito Grupo MODES, CITIC, Departamento de Matemáticas Universidade da Coruña ORCid: 0000-0003-0526-6118 ines.barbeito@udc.es

Javier Tarrío-Saavedra CITIC, Grupo MODES, EPEF, Departamento de Matemáticas Universidade da Coruña *ORCid: 0000-0002-9584-127X javier.tarrio@udc.es*  Sonia Zaragoza Campus Industrial de Ferrol, CITENI, EPEF Universidade da Coruña *ORCid: 0000-0003-4645-3382 sonia.zaragoza1@udc.es* 

Salvador Naya CITIC, Grupo MODES, EPEF, Departamento de Matemáticas Universidade da Coruña *ORCid: 0000-0003-4931-9859* salva@udc.es

#### Resumen

En este trabajo se presenta, de una forma intuitiva, paso por paso, la metodología propuesta en el artículo "Assessing thermal comfort and energy efficiency in buildings by statistical quality control for autocorrelated data" para el control de procesos definidos por una variable (critical to quality variable, CTQ) cuyas observaciones están autocorreladas, permitiendo, a su vez, la detección de anomalías, siendo éste su principal objetivo. De hecho, el origen de este procedimiento está en la necesidad de detección automática de alarmas que tienen las empresas relacionadas con las instalaciones de energía en edificios, en el marco de la Industria 4.0 y la digitalización de la producción y los servicios. En concreto, la empresa  $\Sigma$ qus, desarrolladora de plataformas de Internet of Things (IoT) en el sector de la energía en edificaciones (centros comerciales, oficinas, hoteles, hospitales) presentó a los autores su necesidad de implementar algoritmos que permitieran la detección automática de anomalías en sus sistemas de climatización (Heating, Ventilation, and Air Conditioning, HVAC) a partir de variables como la temperatura ambiente interior o el consumo de energía en climatización, es decir, anomalías relacionadas con la eficiencia energética y de confort térmico del edificio. En consecuencia, se propuso la aplicación de gráficos de control  $\bar{x}$ , de medidas individuales y de medias

ponderadas exponencialmente (EWMA) a las variables CTQ. Dado que sus observaciones no son independientes (como es común en datos monitorizados continuamente con respecto al tiempo), se propone un procedimiento para aplicar estos gráficos a los residuos correspondientes al previo ajuste de modelos ARIMA para series de tiempo. El estudio se complementa con el análisis de la capacidad de las instalaciones HVAC del edificio para cumplir las especificaciones en términos de consumo energía y confort térmico, a través de la propuesta de aplicación de índices de capacidad univariante para datos autocorrelados.

Palabras clave: Control Estadístico de Procesos, Control Estadístico de la Calidad, Series de tiempo, Gráficos de control, Análisis de capacidad.

MSC Subject classifications: 62-07, 62P30, 93C83.

#### 1. Introducción

Dos de las nueve tecnologías principales asociadas a la Industria 4.0 o revolución digital de la industria (Assis Dornelles, Ayala y Frank, 2022) son el denominado Internet de las Cosas (IoT) y la analítica de datos y big data. De hecho, la aplicación de una implica casi siempre el uso de la otra. En el presente caso, la necesidad de un método automático para la detección de anomalías se originó a partir de la puesta en marcha de la plataforma web IoT  $\Sigma$ que, desarrollada por el Grupo Iverdama, para el control y monitorización de instalaciones energéticas en edificios. Una vez identificadas las variables críticas para la calidad (CTQ) (Montgomery, 2020) del sistema de climatización (HVAC) (Barbeito et al., 2017), recogidas por sensores y contadores y monitorizadas a través de la plataforma IoT (Hossein Motlagh et al., 2020; Metallidou, Psannis y Egyptiadou, 2020; Barbeito et al., 2017), es preciso identificar los niveles dentro de los cuales el sistema está controlado (sin cambios debidos a causas fuera de la propia aleatoriedad del proceso). Para ello, aparte del uso de métodos de clasificación no supervisada y supervisada (Görnitz, Kloft y Rieck, s.f.; Tobar et al., 2023; Sosa Donoso et al., 2023), siguiendo la metodología Seis Sigma (Montgomery, 2020), se pueden aplicar técnicas de gráficos de control, ya sean para datos univariantes o multivariantes, con observaciones independientes o no, asumiendo que las observaciones siguen una distribución normal o aplicando alternativas no paramétricas (Montgomery, 2020; Barbeito et al., 2017). Este último es el caso del presente trabajo, en el que se ilustra una metodología de gráficos de control y análisis de capacidad para datos univariantes cuyas observaciones están autocorreladas, propuesta en Barbeito et al., 2017.

Los gráficos de control son salidas gráficas bidimensionales, desarrolladas por Shewhart (Shewhart, 1926) en las que el eje Y representa el valor de la variable a controlar, mientras que el eje X, en la mayoría de los casos, indica la secuencia temporal. Suelen estar compuestos de una línea central (que suele coincidir con la media o mediana, CL) y dos líneas extremas correspondientes a los límites inferior (LCL) y superior (UCL) de control. Habitualmente, se definen los límites LCL y UCL a  $3\sigma$  de la línea central o, lo que es lo mismo, dejando una probabilidad de 0.9973 de que una observación esté entre dichos límites, asumiendo distribución normal.

La alternativa más sencilla y utilizada de gráficos de control son los gráficos de medidas individuales (x), medias  $(\bar{x})$  y rangos (R) para datos univariantes (Montgomery, 2020). Las hipótesis de partida son la normalidad, independencia y homocedasticidad de las observaciones. Esto no siempre se puede asumir en el actual paradigma de datos monitorizados continuamente con respecto al tiempo, propio de la Industria 4.0. Por ello, en los últimos 20 años, se han desarrollado múltiples alternativas para datos autocorrelados, multivariantes, no normales o mismo funcionales. Así, cuando los datos no

son normales, se pueden mencionar las alternativas descritas en Chakraborti, Van der Laan y Bakir, 2001, Chakraborti, 2004 y Chakraborti y Graham, 2019 entre otras, mientras que si las observaciones no son independientes, se puede seguir la metodología indicada en Barbeito et al., 2017, la cual se ilustra en el presente trabajo.

Por otro lado, los gráficos de control multivariantes (Santos-Fernández, 2012), si bien su uso está menos extendido, son cada vez más utilizados debido a que la mayoría de procesos industriales están definidos por más de una variable CTQ. La alternativa paramétrica más utilizada son los gráficos  $T^2$  de Hotelling (Santos-Fernández, 2012; Montgomery, 2020), mientras que existen cada vez más enfoques que permiten el control de datos no normales (Chakraborti, Van der Laan y Bakir, 2001; Chakraborti, 2004; Chakraborti y Graham, 2019; Liu, 1995) y autocorrelados (Kruger y Xie, 2012; Kalgonda y Kulkarni, 2004).

Finalmente, en los últimos años, se han desarrollado gráficos de control en los que se monitoriza la relación entre dos o más variables, denominados muchas veces como gráficos de control de regresión o gráficos de control de perfiles, ampliamente descritos en la monografía de Saghaei, Noorossana y Amiri, 2013. Más recientemente, dentro del paradigma del Análisis de Datos Funcionales (FDA), se han desarrollado nuevos enfoques de gráficos de control construidos a partir de datos en los que la unidad es una curva de dimensión infinita o variable funcional (Colosimo y Pacella, 2010; Flores, Naya et al., 2020), o mismo más de una variable funcional, en el marco funcional multivariante (Centofanti et al., 2021).

En el trabajo que aquí se ilustra (Barbeito et al., 2017), también se realizaba un análisis de capacidad. El análisis de capacidad se define como todas aquellas métricas y procedimientos para estimar el grado en el que un proceso cumple las especificaciones de la empresa, la normativa vigente o los clientes (Barbeito et al., 2017; Montgomery, 2020; Naya et al., 2016). Las métricas más utilizadas son los denominados índices de capacidad, para cuyo cálculo se asume que el proceso está bajo control, además de partir de la hipótesis de que las observaciones siguen una distribución normal y son independientes. En procesos monitorizados continuamente con respecto al tiempo, esta última hipótesis no siempre se puede asumir. De hecho, la estimación de la variabilidad es muy diferente (siempre mayor) cuando los datos están autocorrelados. Por lo tanto, se propone el uso de índices de capacidad alternativos para datos autocorrelados, en el ámbito específico de la eficiencia energética evaluada a partir de la información proporcionada por plataformas IoT (Barbeito et al., 2017).

Por tanto, en este trabajo se propone la aplicación de una metodología univariante de control estadístico de procesos que permita la detección de anomalías a partir de observaciones autocorreladas, en el ámbito específico de la eficiencia energética y confort higrotérmico en edificios. Éste procedimiento, cuyo punto de partida fueron los trabajos desarrollados por Zaragoza-Fernández et al., 2014, Álvarez et al., 2016 y Naya et al., 2016, han tenidos continuidad a través de una línea de investigación en la que se han propuesto técnicas de detección de anomalías en eficiencia energética a partir de la aplicación de gráficos de control FDA (Flores, Naya et al., 2020; Eiras-Franco et al., 2019), métodos FDA de detección de outliers (Flores, Tarrio-Saavedra et al., 2018; Flores, Fernández-Casal et al., 2018) y procedimientos de clasificación nos supervisada y supervisada (Tobar et al., 2023; Sosa Donoso et al., 2023).

Aparte de la introducción, el presente estudio se divide en cuatro secciones, siendo la segunda la descripción de la plataforma IoT para el control de instalaciones energéticas en edificios  $\Sigma$ qus, utilizada en le presente caso de estudio, mientras que en la tercera sección se describe muy brevemente la metodología estadística utilizada. Por otro lado, en la cuarta sección se presenta un caso práctico de aplicación, incluyendo código y resultados, siendo el último apartado el correspondiente a las conclusiones principales.



Fig. 1: Proceso de detección de anomalías a través de plataformas IoT.

# 2. Plataformas IoT para la eficiencia energética y confort térmico en edificios

El sector servicios es uno de los sectores industriales más deslocalizados, es decir, que una misma firma tiene diferentes sedes en diferentes localizaciones geográficas, con deferentes climatologías, lo cual dificulta la gestión de éstas desde un punto de vista energético, pues cada instalación aporta una gran cantidad de variables, muchas más que si todos los centros estuviesen juntos. Para afrontar el reto de monitorizar y controlar cientos de variables a la vez y al mismo tiempo, es necesario contar con herramientas de la Industria 4.0, que se alimenten de los datos de las instalaciones y que permitan su tratamiento a través de algoritmos con base matemática. El desarrollo de algoritmos matemáticos implementables en plataformas IoT in cloud, son la única posibilidad de gestionar las instalaciones energéticas del sector servicios, pues permiten la monitorización y control de todas las instalaciones en un solo punto, independientemente de su ubicación geográfica real. En este trabajo se ha desarrollado un algoritmo matemático programable en plataforma in cloud que permite la detección a distancia, y de forma automática, de anomalías relacionadas con la eficiencia energética y el confort higrotérmico en edificaciones. Para ello, en este trabajo se han seguido los pasos indicados en la Figura 1.

El primer paso en las plataformas IoT es la toma de datos en una instalación real (panel izquierdo de la Figura 2). Dentro de las instalaciones reales se estudiaron específicamente las instalaciones energéticas de las oficinas de una nave industrial propiedad del Grupo Iverdama, caracterizada por tener actividad real y disponer de instalaciones energéticas (Figura 2, segundo panel empezando por la izquierda). Se utilizaron para el estudio las instalaciones de iluminación y climatización.

Para la toma de datos se necesitan sensores y actuadores, además de diversos sistemas de control, todos ellos precisos para la adquisición de información en forma de variables cuantitativas. Entre otros, se encuentran sensores de iluminación, temperatura, calidad del aire y analizadores de red (Figura 2, tercer panel empezando por la izquierda). Además, fue necesario crear una base de datos del tipo SQL, en la cual almacenar información acerca de las condiciones de funcionamiento de las instalaciones energéticas.

Igualmente, es preciso el establecimiento de una conexión entre los sensores y los autómatas (Figura 2, cuarto panel empezando por la izquierda), la cual permite que el dato sea gestionado y, una vez realizada la programación del algoritmo de control y detección de anomalías, enviar, según el



Fig. 2: De izquierda a derecha, imagen de la nave piloto, representación en REVIT de sus instalaciones energéticas, detalle de la colocación de sensores y actuadores en las compuestas del sistema de aire acondicionado y, finalmente, cuadro de control (incluyendo las conexiones de equipos a sensores y de sensores a la plataforma in cloud).

resultado, una orden a los equipos, permitiendo a su vez no solo la monitorización de los datos sino también el control y actuación sobre las instalaciones.

Finalmente, también es precisa la creación de una plataforma in cloud para supervisión de las instalaciones a través de la monitorización y control de los datos en tiempo real (Figura 3, panel superior). La plataforma permite la supervisión y control de la instalaciones. Para tal fin, está dotada de un sistema SCADA con capacidad de actuación sobre las instalaciones, además de un apartado de planos digitales para dar un contexto geográfico a la información y así facilitar la comprensión del funcionamiento de los equipos.

Los métodos estadísticos incluidos en este trabajo se han desarrollado para su inclusión en la plataforma in cloud, junto con herramientas de visualización y gráficos de resumen (Figura 3), los cuales facilitan en gran medida la gestión de las instalaciones energéticas.

# 3. Metodología

La metodología propuesta en Barbeito et al., 2017 se utiliza para analizar el estado real del confort térmico y la eficiencia energética de las oficinas de la empresa Iverdama. Este procedimiento permite, además, detectar las no conformidades o alarmas y sus causas asignables. Por otro lado, también se realiza un análisis de la capacidad de cumplir los requisitos marcados por la normativa de edificación. Para todo ello, se emplean herramientas y paquetes implementados en el software de código abierto R. La metodología se puede resumir en los siguientes pasos:

- 1. En primer lugar, este estudio propone ajustar modelos de series temporales ARIMA a las variables de CTQ.
- 2. A continuación, se propone la aplicación de gráficos de control Shewhart y EWMA a los residuos de las series temporales para controlar y monitorizar el confort térmico y el consumo de energía en los edificios.
- 3. Una vez estimados los límites de control naturales para las variables que definen el el confort térmico y el consumo energético, se propone el cálculo de índices de capacidad para las variables cuyas observaciones son dependientes (autocorreladas), con el fin de poder estimar el grado de cumplimiento, por parte de las instalaciones HVAC, de las especificaciones incluidas en las normas.



Fig. 3: Detalle del SCADA correspondiente a la plataforma IoT  $\Sigma$ qus: esquema de monitorización de la instalación de estudio, junto con un ejemplo de gráficos de resumen y planos de instalaciones.

De acuerdo con los resultados de los casos prácticos, la metodología propuesta ha detectado anomalías reales en la instalación HVAC, ayudando a detectar las causas asignables y, por consiguiente, a tomar las decisiones adecuadas.

Es importante destacar que, el origen de esta metodología, aplicada a instalaciones de energía, está en el hecho que la aplicación de gráficos de control tradicionales a variables autocorreladas no es aconsejable dado que asumen independencia entre las observaciones. Su uso conllevaría la detección de un número de falsas alarmas muy alto, dado que la variabilidad de proceso se infraestima. Así, el control de variables CTQ de sistemas HVAC como la temperatura o el consumo energético en un edificio (muy condicionados por el horario y el clima), ha de realizarse teniendo en cuenta la dependencia existente entre las observaciones. Como ya se ha mencionado, una alternativa al estudio de la variable original, X, es el ajuste de un modelo Box-Jenkings de series de tiempo, AR, ARMA, o ARIMA, después de lo cual se aplicarían los gráaficos de control univariantes Shewhart o EWMA a los residuos de la serie, permitiendo la estimación de los límites de control con una muestra de calibrado (Fase I). Acto seguido, estaríamos ya en disposición de monitorizar después el resto de observaciones (Fase II) y detectar las correspondientes anomalías (observaciones fuera de los límites de control o existencia de patrones en los gráficos de control).

Formalmente, se asume un proceso estacionario  $\{X_t\}_{t\in\mathbb{Z}}$  que genera la serie de tiempo. En concreto, si  $\{X_t\}_{t\in\mathbb{Z}}$  admite la siguiente expresión, que se corresponde con un proceso ARIMA(p, d, q) con d diferencias regulares:

$$\phi(B)(1-B)^d X_t = c + \theta(B)a_t,$$

siendo B el operador de retardo, definido como  $BX_t = X_{t-1}$ ,

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p),$$
  
$$\theta(B) = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q),$$

p el número de retardos del proceso autorregresivo (número de observaciones pasadas de las que depende la actual),  $\phi_i$  (con  $i = 1, 2, \ldots, p$ ) los parámetros correspondientes al proceso autorregresivo, mientras que q es el número de términos de medias móviles de los que depende la observación actual y  $\theta_j$  (con  $j = 1, 2, \ldots, q$ ) los parámetros del modelo de medias móviles. Por otro lado,  $\{a_t\}_{t\in\mathbb{Z}}$  se corresponde con una colección de variables aleatorias que, generalmente, se suponen no correladas, sin tendencia y con varianza finita  $\sigma_a^2$ , es decir, el denominado ruido blanco. Si éste sigue una distribución normal, entonces las variables aleatorias  $\{a_t\}_{t\in\mathbb{Z}}$  son también independientes. Asumiendo que las observaciones de estudio están autocorrelacionadas, la serie temporal resultante se puede modelar mediante el ajuste de un modelo paramétrico de dependencia. Si este modelo fuese realmente aceptable, los residuos serán estacionarios e independientes, por lo que pueden utilizarse para estimar la tendencia y la desviación estándar. En consecuencia, también se posrá estimar, a partir de ellos, los límites de control natural de proceso, que permitirán la detección de anomalías en el sistema que la variable original define.

Hilando aún más fino, si  $\{X_t\}_{t\in\mathbb{Z}}$  cumple la expresión

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D X_t = c + \theta(B)\Theta(B^s)a_t,$$

se conoce como proceso  $\operatorname{ARIMA}(p, d, q) \times (P, D, Q)_s$  de *d* diferencias regulares *D* diferencias estacionarias de periodo *s*, siendo éste un modelo todavía más flexible, y el que finalmente se ajusta en la presente metodología para poder así trabajar con residuos independientes.

Finalmente, una vez que el proceso se asume bajo control, cabría preguntarse si las instalaciones estudiadas cumple con la normativa de confort térmico (21-23 grados centígrados), fijadas por el

reglamento de edificación. Para responder a esto se aplican índices de capacidad de proceso definidos específicamente para ser aplicados a datos autocorrelados. En concreto, se han estimado los indicadores propuestos por Pan, Li y Chen, 2014.

Específicamente, Wallgren, 2001 propone una modificación de los índices  $C_{pk}$  (capacidad potencial) y  $C_{pm}$  (capacidad real) para datos autocorrelados. Considérense *n* observaciones  $\{X_1, \ldots, X_n\}$ , de una media muestral  $\bar{X}$  y de la varianza muestral  $S^2$ . Entonces, los índices de capacidad se estiman asumiendo un modelo

$$X_t - \mu = \phi_1(X_{t-1} - \mu) - a_t,$$

en el que  $\mu$  es la media del proceso y  $a_t$  son las innovaciones normales,  $a_t \stackrel{d}{=} N(0, \sigma_a^2)$ , definiendo las versiones  $C_{pk}$  y  $C_{pm}$  para datos autocorrelados tal y como sigue:

$$C_{pmr} = \frac{USL - LSL}{6\sqrt{\sigma_r^2 + (\mu - T)^2}}, C_{pkr} = \min\left(\frac{USL - \mu}{3\sigma_r}, \frac{\mu - LSL}{3\sigma_r}\right),$$

donde  $\sigma_r^2 = \frac{\sigma_a^2}{1 - \phi_1}$ , por lo que las estimaciones a partir de la muestra son

$$\hat{C}_{pmr} = \frac{USL - LSL}{6\sqrt{\hat{\sigma}_r^2 + (\hat{\mu} - T)^2}}, \hat{C}_{pkr} = \min\left(\frac{USL - \hat{\mu}}{3\hat{\sigma}_r}, \frac{\hat{\mu} - LSL}{3\hat{\sigma}_r}\right).$$

#### 4. Caso práctico de aplicación y resultados

Abajo se muestra un ejemplo de la necesidad de aplicar series de tiempo. Por ejemplo, a la temperatura interior de una oficina, en horario de ocupación. Si se construye un gráfico de control de medidas individuales, observamos un gran número de alarmas, ¿es esta situación real? Hay una alta correlación, por lo que se ajusta un modelo ARIMA. Si se utiliza el software estadístico R, primeramente se construiría la serie de tiempo a partir de la muestra de calibrado, empleando la función ts. Seguidamente, se ajusta el modelo ARIMA. Para ello se puede emplear la librería forecast, obteniendo un ajuste automático mediante la función auto.arima o, alternativamente, hacer un ajuste manual mediante la observación de la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF), teniendo igualmente en cuenta la posible existencia de valores atípicos. Siempre que sea posible y pueda ser llevada a cabo por personal experimentado, esta última alternativa es preferible a la primera, dado que suele dar como resultado modelos con menos parámetros y mejor os similar bondad de ajuste (medida en términos de AIC).

En términos prácticos, se utilizan los modelos Box-Jenkins para proponer un proceso estocástico que genere las series temporales. La serie temporal se diferencia dos veces: en primer lugar, se realiza una diferencia regular y, a continuación, una estacional (con retardo 7). A continuación, se utilizan los gráficos ACF y PACF de las series temporales para decidir qué componente autorregresivo o de media móvil debe utilizarse en el modelo. A continuación, se muestran los gráficos ACF y PACF de la serie temporal de la temperatura diaria dos veces diferenciada (Figura 4). Posteriormente, se puede observar el diagnóstico del modelo propuesto para esta serie temporal en la Figura 5. Los gráficos obtenidos sugieren que al modelo ARIMA $(0,1,0) \times (0,1,1)_7$  como proceso estocástico generador de la serie temporal. Por un lado, los retardos estacionales del gráfico ACF desaparecen



Fig. 4: Functiones ACF y PACF correspondientes al ajuste  $ARIMA(0,1,0) \times (0,1,1)_7$ .

a partir del segundo retardo. Por otra parte, los retardos regulares del gráfico ACF no difieren significativamente de 0 a partir del segundo retardo. Además, la aleatoriedad del proceso nos permite prescindir del primer retardo. Lo mismo ocurre con el gráfico PACF. Según la Figura 5, los residuos oscilan en torno a 0, además de poder considerarlos independientes. Concluimos, por tanto, que los residuos  $\{a_t\}_{t\in\mathbb{Z}}$  son ruido blanco, asumiendo también la normalidad de los mismos  $(p - valor_{lilliefors test} = 0.06857, utilizando la función lillie.test)$  a un nivel de significación  $\alpha = 5\%$ . Este modelo podría representarse como  $(1 - B)(1 - B^7)X_t = c + (1 + \Theta_1 B^7)a_t$ , donde  $X_t$ es la temperatura media diaria en la oficina en el día t, c es una constante y  $BX_t = X_{t-1}; B^7 X_t =$  $BX_{t-7}$ . Posteriormente, se estiman los parámetros del modelo mediante máxima verosimilitud, obteniendo  $\hat{\Theta}_1 = -0.645(0.1211), \hat{\sigma}_a^2 = 0.6178$ . No obstante, aunque un ARIMA $(0, 1, 0) \times (0, 1, 1)_7$ con innovaciones normales y constante nula es una buena propuesta como generador de la serie temporal, podrían existir otros modelos tan buenos o incluso mejores. De hecho, la posibilidad de la existencia de puntos atípicos podría dar lugar a encontrar un modelo mejor. De hecho, para tal fin se pueden utilizar las funciones detectIO y detectAO, que detectan puntos atípicos innovadores y puntos atípicos aditivos, respectivamente. En este caso particular se deduce que no existen puntos atípicos.

Acto seguido se aplica a los residuos un gráfico tipo Shewhart para medidas individuales o, alternativamente, si se pretenden detectar cambios pequeños (desviaciones de menos de dos desviaciones típicas con respecto a la media) se puenden también emplear gráficos con memoria, como el gráfico de control de medias ponderadas exponencialmente (EWMA), también más robusto a la ausencia de normalidad. Se estiman los límites de control (eliminando atípicos) a partir de la muestra de calibrado y se monitorizan el resto de muestras. Existen diversas funciones en diferentes paquetes para llevar a



Fig. 5: Gráficos de diagnósis del modelo ARIMA $(0, 1, 0) \times (0, 1, 1)_7$ .

cabo esta tarea, entre las que se encuentran las funciones qcc, para medidas individuales, y ewma, para gráficos EWMA, dentro del paquete qcc (Scrucca, 2004). Alternativamente podría haberse utilizado el paquete qcr (Flores, Fernandez-Casal et al., 2022).

Listing 1: Funciones qcc y ewma del paquete qcc para obtener gráficos de control de medidas individuales y gráficos de memoria EWMA

Acto seguido, se construye, sobre los residuos, un gráfico de medias ponderadas exponencialmente, EWMA (más robusto a la ausencia de normalidad), se estiman los límites de control (eliminando atípicos) y se monitorizan el resto de muestras. Si se observa la Figura 6, se identifica una alarma debida a problemas en la regulación de las instalaciones HVAC (problemas para mantenera la temperatura de enero).

En las Figuras 6,7, se detecta una alarma debida a problemas en la regulación de las instalaciones HVAC (problemas para mantener la temperatura ambiente conseguida en enero).

Finalmente, asumiendo que la temperatura interior de la oficina está bajo control, se pretende saber si se están cumpliendo las especificaciones de la norma de edificación. Para ello se calculan los índices de capacidad indicados en la sección previa, específicamente diseñados para datos autocorrelados, como es el caso de la temperatura media diaria en la oficina. Los valores de  $C_{pmr}$  y  $C_{pkr}$  son,



Fig. 6: Gráfico de control para medidas individuales contruido a partir de los residuos del modelo ARIMA $(0, 1, 0) \times (0, 1, 1)_7$  ajustado a la temperatura media diaria de la oficina estudiada. Panel izquierdo: Estimación de los límites de control del gráfico a partir de una muestra retrospectiva o de calibrado correspondiente al mes de enero. Panel derecho: Monitorización de los meses subsiguientes (hasta marzo).



Fig. 7: Gráfico de control EWMAs contruido a partir de los residuos del modelo ARIMA $(0, 1, 0) \times (0, 1, 1)_7$  ajustado a la temperatura media diaria de la oficina estudiada. Panel izquierdo: Estimación de los límites de control del gráfico a partir de una muestra retrospectiva o de calibrado correspondiente al mes de enero. Panel derecho: Monitorización de los meses subsiguientes (hasta marzo).

respectivamente, 0.0187 y -0.3495, sensiblemente menores que 1.33, por lo que el sistema no es capaz de cumplir especificaciones. Habría que aplicar mejoras en las instalaciones HVAC para cumplir las especificaciones de la normativa en relación al confort térmico en edificios.

```
1
  ## Serie de tiempo
2
3
  serie <- ts(variable,start=c(1,1),frequency=10) # Se crea la serie de tiempo</pre>
  fit <- arimax(serie,order=c(1,0,0)) # Se asume un proceso AR(1)
4
5
6
  mu = mean(serie) # Se calcula la media del proceso
7
8
  sigma_a = sd(fit$residuals) # Se calcula la varianza de los residuos
g
10
  parametro= fit$coef[1] # Parametro del proceso autorregresivo
11
  sigma_r = sqrt((sigma_a^2)/(1-parametro)) # Desviacion tipica corregida
12
13
14
15
  # Indices de capacidad:
16
17
  C_{pr} = (USL-LSL)/(6*sigma_r)
18
19
  T = 0.5 * (USL - LSL)
20
21
  C_pmr = (USL-LSL)/(6*sqrt(sigma_r^2+(mu-T)^2))
22
  C_pkr = min((USL-mu)/(3*sigma_r),(mu-LSL)/(3*sigma_r))
23
24|C_pmr
25|C_pkr
```

Listing 2: Código para el cálculo de índices de capacidad para datos autocorrelados a partir de la serie temporal de temperatura diaria.

# 5. Conclusiones

El sector de la energía ha hecho, en los últimos años, una gran inversión en el desarrollo de tecnologías IoT para el control y monitorización continua de instalaciones de energía, incluyendo los sistemas HVAC, con el objeto de optimizar recursos, aumentar la eficiencia energética, mantener el confort térmico y favorecer el mantenimiento predictivo. Para ello precisa de herramientas estadísticas que permitan la detección automática de anomalías. Esta tarea se puede llevar a cabo a través del control estadístico de procesos, dado las características de los datos (autocorrelados), las alternativas de gráficos de control tradicionales no son aplicables, al proporcionar un alto número de detección de falsas anomalías.

De hecho, cuando las variables CTQ se monitorizan continuamente con respecto al tiempo, suelen presentar cierto grado de autocorrelación, lo cual requiere que, para estimar los límites de control naturales de las instalaciones que definen (su funcionamiento normal) y poder detectar alarmas y anomalías (averías, fallos, paradas no programadas, retrasos, etc.), es preciso la aplicación de gráficos de control que tengan en cuenta esta dependencia. En consecuencia, este trabajo ilustra y comenta la propuesta de Barbeito et al., 2017, mostrándose que el ajuste de modelos ARIMA a las series de

tiempo de las variables CTQ, paso previo al cálculo de gráficos de control Shewhart o de memoria, ha resultado un paso fundamental y provechoso para la detección de anomalías reales en sistemas HVAC, previniendo a su vez la detección de falsas alarmas.

Igual de importante que el control y detección de anomalías en un sistema HVAC, es muchas veces el análisis de capacidad del sistema para cumplir las especificaciones de la norma de edificación. En este caso, también es preciso la aplicación de índices de capacidad para datos autocorrelados, dando, en este caso, como resultado, que las instalaciones HVAC no son capaces de cumplir especificaciones de temperatura interior.

Barbeito et al., 2017 ha representado un punto de partida para los autores en una línea de investigación de gran actividad en los últimos años, como es el control y detección de anomalías de forma automática en las instalaciones de energía en edificios, cada vez más importante en el marco de la actual preocupación por la sostenibilidad y el ahorro energético.

## Agradecimientos

Este estudio ha sido apoyado por el Ministerio de Ciencia e Innovación con la subvención PID 2020-113578RB-100, además de por la Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2020-14 y Centro de Investigación del Sistema universitario de Galicia ED431G2019/01), todos ellos a través de los fondos FEDER. Los autores quisieran agradecer a la SEIO su apoyo e interés en nuestro trabajo.

## Sobre los autores



**Inés Barbeito** es Licenciada en Matemáticas por la Universidad de Santiago de Compostela y doctora en Estadística e Investigación Operativa por la Universidade da Coruña. Su investigación se centra en la estadística no paramétrica, bootstrap y la modelización de datos dependientes.



**Sonia Zaragoza** es ingeniera y doctora ingeniera industrial por la Universidade da Coruña, Adjunta a Transferencia en el Vicerrectorado de Investigación y Política Científica de dicha universidad. Igualmente, es fundadora de la empresa spin-off Nerxus y socia del Grupo Iverdama. Su docencia se centra en diversas áreas de la ingeniería y la energía, desarrollando su investigación en los ámbitos de la eficiencia energética y el control del confort higrotérmico en edificaciones.



Salvador Naya es Licenciado en Matemáticas por la Universidad de Santiago de Compostela y Doctor Ingeniero Industrial por la Universidade da Coruña (UDC). Es Catedrático de Estadística y Vicerrector de Investigación, además de investigador del Grupo de Investigación de Modelización, Optimización e Inferencia Estadística (MODES), así como miembro del Centro de Investigación en Tecnologías de la Información y la Computación (CITIC) de la UDC. Es autor de más de 100 publicaciones sobre Estadística, Ingeniería, Bibliometría, Ciencia de Materiales y otras ciencias aplicadas.



Javier Tarrío Saavedra es Ingeniero Industrial y Doctor en Estadística e Investigación Operativa por la UDC. Actualmente es Profesor Titular de la UDC y sus investigaciones están relacionadas con el control estadístico de la calidad, la estadística industrial y la aplicación de la estadística a la ciencia y la ingeniería de materiales. Es autor de más de 80 publicaciones de investigación sobre Estadística, Ingeniería, Bibliometría, Ciencia de Materiales y otras ciencias aplicadas.

## Referencias

- Ålvarez, A., J. Tarrio-Saavedra, S. Zaragoza, J. López-Beceiro, R. Artiaga, S. Naya y B. Ålvarez (2016). «Numerical and experimental study of a corrugated thermal collector». En: *Case Studies in Thermal Engineering* 8, págs. 41-50.
- Assis Dornelles, J. de, N. F. Ayala y A. G. Frank (2022). «Smart Working in Industry 4.0: How digital technologies enhance manufacturing workers' activities». En: *Computers & Industrial Engineering* 163, pág. 107804.
- Barbeito, I., S. Zaragoza, J. Tarrio-Saavedra y S. Naya (2017). «Assessing thermal comfort and energy efficiency in buildings by statistical quality control for autocorrelated data». En: Applied energy 190, págs. 1-17.
- Centofanti, F., A. Lepore, A. Menafoglio, B. Palumbo y S. Vantini (2021). «Functional regression control chart». En: *Technometrics* 63.3, págs. 281-294.
- Chakraborti, S. (2004). «Nonparametric (distribution-free) quality control charts». En: *Encyclopedia* of statistical sciences, págs. 1-27.
- Chakraborti, S. y M. A. Graham (2019). «Nonparametric (distribution-free) control charts: An updated overview and some results». En: *Quality Engineering* 31.4, págs. 523-544.
- Chakraborti, S., P. Van der Laan y S. T. Bakir (2001). «Nonparametric control charts: an overview and some results». En: *Journal of quality technology* 33.3, págs. 304-315.
- Colosimo, B. M. y M. Pacella (2010). «A comparison study of control charts for statistical monitoring of functional data». En: *International Journal of Production Research* 48.6, págs. 1575-1601.
- Eiras-Franco, C., M. Flores, V. Bolón-Canedo, S. Zaragoza, R. Fernández-Casal, S. Naya y J. Tarrio-Saavedra (2019). «Case study of anomaly detection and quality control of energy efficiency and hygrothermal comfort in buildings». En: 8th International Conference on Data Science, Technology and Applications (DATA 2019); Prague, CzechRepublic; 26-28 July 2019, págs. 145-151.
- Flores, M., R. Fernandez-Casal, S. Naya y J. Tarrio-Saavedra (2022). *qcr: Quality Control Review*. R package version 1.4. URL: https://CRAN.R-project.org/package=qcr.

- Flores, M., R. Fernández-Casal, S. Naya, J. Tarrio-Saavedra y R. Bossano (2018). «ILS: An R package for statistical analysis in Interlaboratory Studies». En: *Chemometrics and Intelligent Laboratory* Systems 181, págs. 11-20.
- Flores, M., S. Naya, R. Fernández-Casal, S. Zaragoza, P. Raña y J. Tarrio-Saavedra (2020). «Constructing a control chart using functional data». En: *Mathematics* 8.1, pág. 58.
- Flores, M., J. Tarrio-Saavedra, R. Fernandez-Casal y S. Naya (2018). «Functional extensions of Mandel's h and k statistics for outlier detection in interlaboratory studies». En: Chemometrics and Intelligent Laboratory Systems 176, págs. 134-148.
- Görnitz, N., M. Kloft y U. Rieck K.and Brefeld (s.f.). «Toward supervised anomaly detection». En: Journal of Artificial Intelligence Research 46 (), págs. 235-262.
- Hossein Motlagh, N., M. Mohammadrezaei, J. Hunt y B. Zakeri (2020). «Internet of Things (IoT) and the energy sector». En: *Energies* 13.2, pág. 494.
- Kalgonda, A. A y S. R. Kulkarni (2004). «Multivariate quality control chart for autocorrelated processes». En: Journal of Applied Statistics 31.3, págs. 317-327.
- Kruger, U. y L. Xie (2012). Statistical monitoring of complex multivatiate processes: with applications in industrial process control. John Wiley & Sons.
- Liu, R. Y. (1995). «Control charts for multivariate processes». En: Journal of the American Statistical Association 90.432, págs. 1380-1387.
- Metallidou, C. K., K. E. Psannis y E. A. Egyptiadou (2020). «Energy efficiency in smart buildings: IoT approaches». En: *IEEE Access* 8, págs. 63679-63699.
- Montgomery, D. C. (2020). Introduction to statistical quality control. John Wiley & Sons.
- Naya, S., A. Devia-Rivera, J. Tarrio-Saavedra y M. Flores (2016). «New robust capability ratios approaches for quality control». En: *Dyna* 83.198, págs. 94-101.
- Pan, J.N., C.I. Li y F.Y. Chen (2014). «Evaluating environmental performance using new process capability indices for autocorrelated data». En: *Environmental monitoring and assessment* 186, págs. 6369-6384.
- Saghaei, A., R. Noorossana y A. Amiri (2013). Statistical analysis of profile monitoring. Wiley Online Library.
- Santos-Fernández, E. (2012). Multivariate statistical quality control using R. Vol. 14. Springer Science & Business Media.
- Scrucca, L. (2004). «qcc: an R package for quality control charting and statistical process control». En: R News 4/1, págs. 11-17. URL: https://cran.r-project.org/doc/Rnews/.
- Shewhart, W. A. (1926). «Quality control charts». En: The Bell System Technical Journal 5.4, págs. 593-603.
- Sosa Donoso, J. R., M. Flores, S. Naya y J. Tarrio-Saavedra (2023). «Local Correlation Integral Approach for Anomaly Detection Using Functional Data». En: *Mathematics* 11.4, pág. 815.
- Tobar, A., S. Flores M.and Castillo-Páez, S. Naya, S. Zaragoza y J. Tarrio-Saavedra (2023). «Bootstrap-LOCI data mining methodology for anomaly detection in buildings energy efficiency». En: *Energy Reports* 10, págs. 244-254.
- Wallgren, E. (2001). «Confidence limits for the process capability index C pk for autocorrelated quality characteristics». En: Frontiers in Statistical Quality Control 6. Springer, págs. 312-331.
- Zaragoza-Fernández, S., J. Tarrio-Saavedra, S. Naya y A. López-Beceiro J.and Álvarez-Garcia (2014). «Impact estimates of the actions for the rehabilitation of energy efficiency in residential building». En: Dyna 81.186, págs. 200-207.



Estadística