

## Resumen

De acuerdo con las regulaciones europeas y muchos estudios científicos, es necesario monitorear y analizar las condiciones microclimáticas en museos o edificios, para preservar las obras de arte que se exponen en ellos. Con el objetivo de ofrecer herramientas para el monitoreo de las condiciones climáticas en este tipo de edificios, en esta tesis doctoral se propone una nueva metodología estadística para clasificar series temporales de parámetros climáticos como la temperatura y humedad relativa.

La metodología consiste en aplicar un método de clasificación usando variables que se computan a partir de las series de tiempos. Los dos primeros métodos de clasificación son versiones conocidas de métodos sparse PLS que no se habían aplicado a datos correlacionados en el tiempo. El tercer método es una nueva propuesta que usa dos algoritmos conocidos. Los métodos de clasificación se basan en diferentes versiones de un método sparse de análisis discriminante de mínimos cuadrados parciales PLS (sPLS-DA, SPLSDA y sPLS) y análisis discriminante lineal (LDA). Las variables que los métodos de clasificación usan como input, corresponden a parámetros estimados a partir de distintos modelos, métodos y funciones del área de las series de tiempo, por ejemplo, modelo ARIMA estacional, modelo ARIMA-TGARCH estacional, método estacional Holt-Winters, función de densidad espectral, función de autocorrelación (ACF), función de autocorrelación parcial (PACF), rango móvil (MR), entre otras funciones. También fueron utilizadas algunas variables que se utilizan en el campo de la astronomía para clasificar estrellas.

La metodología propuesta consta de dos partes. La primera consiste en calcular variables a partir de las series de tiempos, usando los métodos, modelos y funciones mencionadas anteriormente. La segunda parte consiste en usar las variables calculadas en el primer paso para ajustar alguno de los siguientes modelos: sPLS-DA, SPLSDA, o sPLS con LDA (nueva propuesta). En los casos que a priori no hubo información de los clusters de las series de tiempos, las dos primeras componentes de un análisis de componentes principales (PCA) fueron utilizadas por el algoritmo k-means para identificar posibles clusters de las series de tiempo. Adicionalmente, los resultados del método sPLS-DA fueron comparados con los del algoritmo random forest.

Tres bases de datos de series de tiempos de humedad relativa o de temperatura fueron analizadas. Estas series de tiempos se registraron en diferentes edificios (la Catedral de Valencia, el yacimiento arqueológico de L'Almoina y la iglesia barroca de Santo Tomás y San Felipe Neri) en Valencia, España. Los clusters de las series de tiempos se analizaron de acuerdo a diferentes zonas o diferentes niveles de alturas donde fueron instalados sensores para el monitoreo de las condiciones climáticas en los edificios.

El algoritmo random forest y las diferentes versiones del método sparse PLS fueron útiles para identificar las variables más importantes en la clasificación de las series de tiempos. Los resultados de sPLS-DA y random forest fueron muy similares cuando se usaron como variables de entrada las calculadas a partir del método Holt-Winters o a partir de funciones aplicadas a las series de tiempo. Aunque los resultados del método random forest fueron levemente mejores que los encontrados por sPLS-DA en cuanto a las tasas de error de clasificación, los resultados de sPLS-DA fueron más fáciles de interpretar.

Cuando las diferentes versiones del método sparse PLS utilizaron variables resultantes del método Holt-Winters, los clusters de las series de tiempo fueron mejor discriminados. Entre las diferentes versiones del método sparse PLS, la versión sPLS con LDA obtuvo la mejor discriminación de las series de tiempo, con un menor valor de la tasa de error de clasificación, y utilizando el menor o segundo menor número de variables.

En esta tesis doctoral se propone usar una versión sparse de PLS (sPLS-DA, o sPLS con LDA) con variables calculadas a partir de series de tiempo para la clasificación de éstas. Al aplicar la metodología a las distintas bases de datos estudiadas, se encontraron modelos parsimoniosos, con pocas variables, y se obtuvo una discriminación satisfactoria de los diferentes clusters de las series de tiempo con fácil interpretación. La metodología propuesta puede ser útil para caracterizar las distintas zonas o alturas en museos o edificios históricos de acuerdo con sus condiciones climáticas, con el objetivo de prevenir problemas de conservación con las obras de arte