

# Resumen

La presente tesis doctoral, concebida principalmente para apoyar y reforzar la relación entre la academia y la industria, se desarrolló en colaboración con Shell Global Solutions (Amsterdam, Países Bajos) en el esfuerzo de aplicar y posiblemente extender los enfoques ya consolidados basados en variables latentes (es decir, Análisis de Componentes Principales - PCA - Regresión en Mínimos Cuadrados Parciales - PLS - o PLS discriminante - PLSDA) para la resolución de problemas complejos no sólo en los campos de mejora y optimización de procesos, sino también en el entorno más amplio del análisis de datos multivariados. Con este fin, en todos los capítulos proponemos nuevas soluciones algorítmicas eficientes para abordar tareas dispares, desde la transferencia de calibración en espectroscopia hasta el modelado en tiempo real de flujos de datos.

El manuscrito se divide en las seis partes siguientes, centradas en diversos temas de interés:

**Parte I - Prefacio**, donde presentamos un resumen de este trabajo de investigación, damos sus principales objetivos y justificaciones junto con una breve introducción sobre PCA, PLS y PLSDA;

**Parte II - Sobre las extensiones basadas en kernels de PCA, PLS y PLSDA**, donde presentamos el potencial de las técnicas de kernel, eventualmente acopladas a variantes específicas de la recién redescubierta proyección de pseudo-muestras, formulada por el estadista inglés John C. Gower, y comparamos su rendimiento respecto a metodologías más clásicas en cuatro aplicaciones a escenarios diferentes: segmentación de imágenes Rojo-Verde-Azul (RGB), discriminación y monitorización de procesos por lotes y análisis de diseños de experimentos de mezclas;

**Parte III - Sobre la selección del número de factores en el PCA por pruebas de permutación**, donde aportamos una guía extensa sobre cómo conseguir la selección de componentes de PCA mediante pruebas de permutación y una ilustración completa de un procedimiento algorítmico original implementado para tal fin;

---

**Parte IV - Sobre la modelización de fuentes de variabilidad común y distintiva en el análisis de datos multi-conjunto**, donde discutimos varios aspectos prácticos del análisis de componentes comunes y distintivos de dos bloques de datos (realizado por métodos como el Análisis Simultáneo de Componentes - SCA - Análisis Simultáneo de Componentes Distintivos y Comunes - DISCO-SCA - Descomposición Adaptada Generalizada de Valores Singulares - Adapted GSVD - ECO-POWER, Análisis de Correlaciones Canónicas - CCA - y Proyecciones Ortogonales de 2 conjuntos a Estructuras Latentes - O2PLS). Presentamos a su vez una nueva estrategia computacional para determinar el número de factores comunes subyacentes a dos matrices de datos que comparten la misma dimensión de fila o columna y dos planteamientos novedosos para la transferencia de calibración entre espectrómetros de infrarrojo cercano;

**Parte V - Sobre el procesamiento y la modelización en tiempo real de flujos de datos de alta dimensión**, donde diseñamos la herramienta de Procesamiento en Tiempo Real (OTFP), un nuevo sistema de manejo racional de mediciones multi-canal registradas en tiempo real;

**Parte VI - Epílogo**, donde presentamos las conclusiones finales, delimitamos las perspectivas futuras, e incluimos los anexos.