

# Resumen

Esta tesis doctoral se centra en el estudio, desarrollo y aplicación de técnicas quimiométricas en el emergente campo de la biología de sistemas. Procedimientos comúnmente utilizados y métodos nuevos se aplican para resolver preguntas de investigación en distintos equipos multidisciplinares, tanto del ámbito académico como del industrial. Las metodologías desarrolladas en este documento enriquecen la plétora de técnicas utilizadas en las ciencias ómicas para entender el funcionamiento de organismos biológicos y mejoran los procesos en la industria biotecnológica, integrando conocimiento biológico a diferentes niveles y explotando los paquetes de software derivados de esta tesis.

Esta disertación se estructura en cuatro partes. El primer bloque describe el marco en el cual se articulan las contribuciones aquí presentadas. En él se esbozan los objetivos de los dos proyectos de investigación relacionados con esta tesis. Asimismo, se introducen los temas específicos desarrollados en este documento mediante presentaciones en conferencias y artículos de investigación. En esta parte figura una descripción exhaustiva de las ciencias ómicas y sus interrelaciones en el paradigma de la biología de sistemas, junto con una revisión de los métodos multivariantes más aplicados en quimiometría, que suponen las pilares sobre los que se asientan los nuevos procedimientos aquí propuestos.

La segunda parte se centra en resolver problemas dentro de metabolómica, fluxómica, proteómica y genómica a partir del análisis de datos. Para ello se proponen varias alternativas para comprender a grandes rasgos los datos de flujos metabólicos en estado estacionario. Algunas de ellas están basadas en la aplicación de métodos multivariantes propuestos con anterioridad, mientras que otras son técnicas nuevas basadas en descomposiciones bilineales utilizando rutas metabólicas elementales. A partir de éstas se ha desarrollado software de libre acceso para la comunidad científica. A su vez, en esta tesis se propone un marco para analizar datos metabólicos en estado no estacionario. Para ello se adapta el enfoque tradicional para sistemas en estado estacionario, modelando las dinámicas de los experimentos empleando análisis de datos de dos y tres vías. En esta parte de la tesis también se establecen relaciones entre los distintos niveles ómicos, integrando diferentes fuentes de información en modelos de fusión de datos. Finalmente, se estudia la interacción entre organismos, como naranjas y hongos, mediante el análisis multivariante de imágenes, con futuras aplicaciones a la industria alimentaria.

El tercer bloque de esta tesis representa un estudio a fondo de diferentes problemas relacionados con datos faltantes en quimiometría, biología de sistemas y en la industria de bioprocesos. En los capítulos más teóricos de esta parte, se proponen nuevos algoritmos para ajustar modelos multivariantes, tanto exploratorios como de regresión, en presencia de datos faltantes. Estos algoritmos sirven además como estrategias de preprocesado de los datos antes del uso de cualquier otro método. Respecto a las aplicaciones, en este bloque se explora la reconstrucción de redes en ciencias ómicas cuando aparecen valores faltantes o atípicos en las bases de datos. Una segunda aplicación de esta parte es la transferencia de modelos de calibración entre instrumentos de infrarrojo cercano, evitando así costosas recalibraciones en bioindustrias y laboratorios de investigación. Finalmente, se propone un paquete software que incluye una interfaz amigable, disponible de forma gratuita para imputación de datos faltantes.

En la última parte, se discuten los aspectos más relevantes de esta tesis para la investigación y la biotecnología, incluyendo líneas futuras de trabajo.