

UNIVERSIDAD POLITÉCNICA DE VALENCIA
DEPARTAMENTO DE INGENIERÍA HIDRÁULICA Y MEDIO AMBIENTE



**APLICACIÓN DE MÉTODOS ESTADÍSTICOS
MULTIVARIANTES PARA LA MODELACIÓN Y LA
MONITORIZACIÓN DE UN REACTOR DISCONTINUO
SECUENCIAL PARA EL TRATAMIENTO DE AGUAS
RESIDUALES**

TESIS DOCTORAL

Presentada por:

DANIEL AGUADO GARCÍA

Dirigida por:

Dr. José Ferrer Polo

Dra. Aurora Seco Torrecillas

Dr. Alberto José Ferrer Riquelme

Valencia, 2005



UNIVERSIDAD
POLITECNICA
DE VALENCIA

DEPARTAMENTO DE INGENIERÍA HIDRÁULICA Y MEDIO AMBIENTE

**APLICACIÓN DE MÉTODOS ESTADÍSTICOS MULTIVARIANTES PARA
LA MODELACIÓN Y LA MONITORIZACIÓN DE UN REACTOR
DISCONTINUO SECUENCIAL PARA EL TRATAMIENTO DE AGUAS
RESIDUALES**

Memoria que, para optar al grado de Doctor por la Universidad Politécnica de Valencia, presenta DANIEL AGUADO GARCÍA.

Valencia, 2005

AGRADECIMIENTOS

En la elaboración de esta tesis he recibido la ayuda de muchas personas. A todas ellas les estoy enormemente agradecido. En especial a los directores de esta tesis, los doctores D. José Ferrer Polo, D^a Aurora Seco Torrecillas y D. Alberto José Ferrer Riquelme, por su inestimable ayuda y generosa dedicación a lo largo de todo este tiempo.

También quiero expresar mi agradecimiento al Ministerio de Educación y Ciencia (Proyecto C.I.C.Y.T. PPQ2002-04043-C02) y a la Entidad Pública de Saneamiento de Aguas Residuales de la Comunidad Valenciana, por la financiación aportada a esta investigación.

Así mismo quiero agradecer a todos mis compañeros del grupo de investigación CALAGUA su amistad, compañerismo y apoyo: Enrique, Joaquín, Ramón, Paqui, Josep, Josué, Luis, Miguel, Cristina, Nuria, Alberto, Laura, Josep Manuel, Marcela y María. Además, quiero dar las gracias a Manolo Zarzo del Departamento de Estadística e Investigación Operativa de la Universidad Politécnica de Valencia por sus consejos, su amistad y su apoyo.

Y muy especialmente a mis padres y a Taty, a quienes agradezco de todo corazón su paciencia, comprensión y cariño.

A mis padres

RESUMEN

RESUMEN

Como consecuencia del incremento de los problemas de eutrofización en ríos, lagos y costas, se ha producido en las últimas décadas un notable aumento del interés por la eliminación de nutrientes en las aguas residuales previamente a su vertido en los sistemas naturales. Inicialmente, la eliminación de la materia orgánica y el nitrógeno presente en el agua residual se realizaba por medio de procesos biológicos mientras que el fósforo se eliminaba mediante procesos químicos. Sin embargo, las ventajas que presenta la eliminación biológica de fósforo han contribuido a que gradualmente este proceso haya sido implantado cada vez en más estaciones depuradoras de aguas residuales (EDARs).

El control de un proceso tan complejo como es el tratamiento de las aguas residuales con eliminación biológica de fósforo, exige conocer las variables de calidad que definen la eficacia del mismo (concentración de ortofosfatos, de amonio, demanda química de oxígeno, ...). Los métodos existentes para la medición de estas variables en tiempo real suponen fuertes inversiones y elevados costes de mantenimiento. Por otro lado, existen otras variables (variables de proceso: pH, conductividad, ...) que se pueden medir en tiempo real por medio de sensores robustos, de bajo coste y que requieren poco mantenimiento, pero que no suministran de forma directa información del funcionamiento del proceso. El elevado número de variables de proceso que actualmente se registran en tiempo real en EDARs modernas, hace necesaria la utilización de técnicas que permitan extraer la información contenida en la gran cantidad de datos registrados.

El objetivo principal de la presente tesis ha sido estudiar las posibilidades de utilización de las variables de proceso para obtener información acerca de la evolución de las variables de calidad, así como sobre la evolución del proceso a lo largo del tiempo, mediante la aplicación de métodos estadísticos multivariantes. De esta forma, se pretende detectar cualquier anomalía que pueda ocurrir (causa especial de variabilidad que puede provocar un funcionamiento inaceptable o inusual del proceso y/o un efluente de baja calidad) y ayudar en la identificación de su

origen, obviando la utilización de las sondas de calidad cuyo elevado coste hace inviable su disponibilidad en la mayor parte de las EDARs existentes.

El trabajo realizado se ha centrado en el estudio de un reactor discontinuo secuencial (SBR, *Sequencing Batch Reactor*) a escala de laboratorio operado para la eliminación biológica de fósforo de las aguas residuales. Se trata de un proceso por lotes (discontinuo), en el que se alternan condiciones anaerobias y aerobias para favorecer el crecimiento de las bacterias acumuladoras de polifosfatos.

Para el análisis “a posteriori” (*off-line*), es decir, de lotes (ciclos) ya finalizados, se han estudiado y comparado dos enfoques: el propuesto por Nomikos y MacGregor (1995) y el propuesto por Wold *et al.* (1998). Debido a las limitaciones presentadas por el nivel de observación del enfoque de Wold *et al.* (1998) en el conjunto de datos utilizado, se han propuesto y analizado dos alternativas para mejorarlo: una consistente en reemplazar la regresión en mínimos cuadrados parciales (PLS, *Partial Least Squares*) por un análisis de componentes principales (PCA, *Principal Component Analysis*), y la segunda consistente en realizar un preprocesamiento diferente y reemplazar el PLS por un PCA para extraer la información relevante del proceso.

Para la monitorización en tiempo real del proceso se han estudiado y comparado seis estrategias: el nivel de observación de la metodología propuesta por Wold *et al.* (1998), la alternativa que emplea un preprocesamiento diferente, la construcción de múltiples modelos de referencia y los tres métodos de imputación para la parte futura del lote sugeridos por Nomikos y MacGregor (1995): desviación nula, desviación constante y datos faltantes.

Para predecir la evolución de las variables de calidad (fósforo, potasio y magnesio), determinadas analíticamente en laboratorio, se han desarrollado diversos modelos predictivos o *soft sensors*, entre los cuales se han considerado tanto modelos de proyección sobre estructuras latentes como modelos de redes neuronales. Todos los modelos han sido validados utilizando un conjunto de datos diferente al utilizado

para la construcción de los mismos y se han comparado desde el punto de vista del error cuadrático medio de este conjunto de datos de validación.

A partir de los resultados obtenidos, se ha propuesto un sistema integral para la monitorización, diagnóstico y predicción en tiempo real del SBR, mediante la aplicación de métodos de control estadístico multivariante de procesos basados en técnicas de proyección sobre estructuras latentes. Para tener en cuenta que en el proceso estudiado además de las autocorrelaciones y correlaciones cruzadas entre las variables a lo largo de cada lote existe una correlación importante entre lotes, el esquema de monitorización propuesto incluye dos niveles, de manera que permite visualizar el transcurso en tiempo real del lote actual y también la evolución general del proceso asociada a los lotes anteriores ya finalizados.

Los resultados que se han obtenido a lo largo del desarrollo del presente trabajo ponen de manifiesto que la aplicación de técnicas estadísticas avanzadas puede ayudar a conseguir una operación más eficiente y segura del proceso utilizando la información contenida en las variables de proceso de fácil medición.

RESUM

Com a conseqüència de l'increment dels problemes d'eutrofització en rius, llacs i costes, s'ha produït en les darreres dècades un notable augment de l'interés per l'eliminació de nutrients en les aigües residuals prèviament al seu abocament als sistemes naturals. Inicialment, l'eliminació de la matèria orgànica i el nitrogen present en l'aigua residual es realitzava per mitjà de processos biològics, mentre que el fòsfor s'eliminava mitjançant processos químics. Tanmateix, els avantatges que presenta l'eliminació biològica de fòsfor han contribuït a què gradualment aquest procés haja sigut implantat cada vegada en més estacions depuradores d'aigües residuals (EDARs).

El control d'un procés tan complex com és el tractament de les aigües residuals amb eliminació biològica de fòsfor, exigeix conèixer les variables de qualitat que defineixen l'eficàcia d'aquest (concentració d'ortofosfats, d'amoni, de la demanda química d'oxigen, ...). Els mètodes existents per al mesurament d'aquestes variables en temps real suposen fortes inversions i elevats costos de manteniment. Per altra banda, existeixen altres variables (variables de procés: pH, conductivitat, ...) que es poden mesurar en temps real per mitjà de sensors robustos, de baix cost i que requereixen poc manteniment, però que no subministren de manera directa informació del funcionament del procés. L'elevat nombre de variables de procés que actualment s'enregistren en temps real en EDARs modernes fa necessària la utilització de tècniques que permeten extraure la informació continguda en la gran quantitat de dades enregistrades.

L'objectiu principal de la present tesi ha sigut estudiar les possibilitats d'ús de les variables de procés per obtenir informació sobre l'evolució de les variables de qualitat, així com sobre l'evolució del procés al llarg del temps, mitjançant l'aplicació de mètodes estadístics multivariants. D'aquesta manera, es pretén detectar qualsevol anomalia que pugui ocórrer (causa especial de variabilitat que pot provocar un funcionament inacceptable o inusual del procés i/o un efluent de baixa qualitat) i ajudar en la identificació del seu origen, tot obviant l'ús de les sondes de qualitat,

l'elevat cost de les quals fa inviable la seua disponibilitat en la major part de les EDARs existents.

El treball realitzat s'ha centrat en l'estudi d'un reactor discontinu seqüencial (SBR, *Sequencing Batch Reactor*) a escala de laboratori operat per a l'eliminació biològica de fòsfor de les aigües residuals. Es tracta d'un procés per lots (discontinuu), en el qual s'alternen condicions anaeròbiques i aeròbiques per tal d'afavorir el creixement dels bacteris acumuladors de polifosfats.

Per a l'anàlisi "a posteriori" (*off-line*), és a dir, de lots (cicles) ja finalitzats, s'han estudiat i comparat dos enfocaments: el proposat per Nomikos i MacGregor (1995) i el proposat per Wold *et al.* (1998). Donades les limitacions presentades pel nivell d'observació de l'enfocament de Wold *et al.* (1998) en el conjunt de dades utilitzat, s'han proposat i analitzat dues alternatives per tal de millorar-lo: una consistent en substituir la regressió en mínims quadrats parcials (PLS, *Partial Least Squares*) per una anàlisi de components principals (PCA, *Principal Component Analysis*), i la segona consistent en realitzar un preprocessament diferent i substituir el PLS per un PCA per tal d'extraure la informació rellevant del procés.

Per a la monitorització en temps real del procés s'han estudiat i comparat sis estratègies: el nivell d'observació de la metodologia proposada per Wold *et al.* (1998), l'alternativa que utilitza un preprocessament diferent, la construcció de múltiples models de referència i els tres mètodes d'imputació per a la part futura del lot suggerits per Nomikos i MacGregor (1995): desviació nul·la, desviació constant i dades inexistent.

Per a predir l'evolució de les variables de qualitat (fòsfor, potassi i magnesi), determinades analíticament en laboratori, s'han desenvolupat diversos models predictius o *soft sensors*, entre els quals s'han considerat tant models de projecció sobre estructures latents com models de xarxes neuronals. Tots els models han sigut validats utilitzant un conjunt de dades diferent a l'utilitzat per a la construcció d'aquests i s'han comparat des del punt de vista de l'error quadràtic mig d'aquest conjunt de dades de validació.

A partir dels resultats obtinguts, s'ha proposat un sistema integral per a la monitorització, diagnòstic i predicció en temps real del SBR, mitjançant l'aplicació de mètodes de control estadístic multivariant de processos basats en tècniques de projecció sobre estructures latents. Per a tenir en compte que en el procés estudiat a més de les autocorrelacions i correlacions creuades entre les variables al llarg de cada lot existeix una correlació important entre lots, l'esquema de monitorització proposat inclou dos nivells, de manera que permet visualitzar el transcurs en temps real del lot actual i també l'evolució general del procés associada als lots anteriors ja finalitzats.

Els resultats que s'han obtingut al llarg del desenvolupament del present treball posen de manifest que l'aplicació de tècniques estadístiques avançades pot ajudar a aconseguir una operació més eficient i segura del procés mitjançant l'ús de la informació continguda en les variables de procés de fàcil mesura.

SUMMARY

The eutrophication problems in rivers, lakes and coastal waters, have increased the requirements for nutrient removal from wastewaters before to their discharge into natural water bodies. Initially, the organic matter and nitrogen was usually removed by biological processes while the phosphorus by means of chemical processes. However, the advantages of biological phosphorus removal have led this process to be gradually implemented in many wastewater treatment plants (WWTPs).

The control of a complex process as the biological wastewater treatment including biological phosphorus removal requires the measurement of those quality variables which are key indicators of process efficiency (orthophosphate concentration, ammonium concentration, chemical oxygen demand,...). Methods to on-line measure these variables involve high investments and important maintenance costs. On the other hand, there are other variables (process variables: pH, conductivity, ...) which can be measured on-line by means of inexpensive, robust and low maintenance sensors, but they do not directly supply information on process performance. The large number of on-line process variables collected nowadays at modern WWTPs, require structured techniques to extract the information from the huge amount of available data.

The main aim of this thesis has been to study the possibilities of using process variables to get information on quality variables as well as on process behaviour over time, by means of multivariate statistical methods. In this way, it is intended to detect any kind of anomaly which could occur (special cause of variation which can led to poor process performance or unusual process behaviour and/or poor effluent quality) and help in the isolation of its origin, thus, avoiding the use of quality probes whose cost make them unaffordable for most WWTPs.

This work has been focused in the study of a laboratory scale sequencing batch reactor (SBR) operated for biological phosphorus removal from wastewater. In this process, alternating anaerobic and aerobic conditions are imposed in order to favour the growth of polyphosphate accumulating organisms.

To analyse completed batches (off-line analysis), two approaches have been studied and compared: Nomikos and MacGregor (1995), and Wold *et al.* (1998). Due to the limitations found in the observation level of the second approach for the data set analysed, two modifications of the original methodology have been proposed to improve it: the first consists in replacing the partial least squares (PLS) by a principal component analysis (PCA), and the second consists in using a different data preprocessing and replacing the PLS by a PCA to extract relevant information from the process.

For on-line batch process monitoring six strategies have been studied and compared: the observation level in the Wold *et al.* (1998) approach, the proposed alternative in which a different data preprocessing is used, the building of multiple reference models (evolving model), and the three imputation methods suggested by Nomikos and MacGregor (1995): zero deviation, constant deviation and missing data.

In order to predict the quality variables (phosphorus, potassium and magnesium), determined by means of laboratory analyses, several predictive models (soft sensors) have been developed, including models based on multivariate projection techniques as well as artificial neural networks. All the models have been validated using a data set which was not used for model building, and their predictive capability has been compared in terms of the mean squared error of the validation data set.

From the results obtained, an integrated system for on-line monitoring, diagnosis and prediction of the SBR process has been proposed. In this system multivariate process control methods based on projection to latent structures are used. Due to the fact that in the batch process analysed not only there are autocorrelations and cross-correlations among variables over time within a batch, but also important correlation between batches, the monitoring scheme proposed includes two levels. This allows monitoring new batches as they are evolving and also the overall process evolution associated with finished batches.

The results obtained in this thesis show that the application of advanced statistical techniques can help to achieve a more efficient and safe process operation using the information contained in the easy-to-measure process variables.

ÍNDICE

1	INTRODUCCIÓN	3
1.1	IMPORTANCIA DE LA DEPURACIÓN DE LAS AGUAS.....	3
1.2	EL PROBLEMA DE LA EUTROFIZACIÓN	3
1.3	LEGISLACIÓN	5
1.4	EL TRATAMIENTO DE AGUAS RESIDUALES	6
1.4.1	<i>Tratamientos biológicos de aguas residuales</i>	<i>8</i>
1.4.1.1	Clasificación de los tratamientos biológicos.....	9
1.4.1.2	Organismos que intervienen en los tratamientos biológicos.....	12
1.5	ELIMINACIÓN DE NUTRIENTES EN LAS AGUAS RESIDUALES	14
1.5.1	<i>Eliminación de Nitrógeno.....</i>	<i>14</i>
1.5.1.1	Eliminación biológica de nitrógeno.....	15
1.5.1.1.1	Nitrificación	16
1.5.1.1.2	Desnitrificación	17
1.5.1.2	Eliminación físico-química de nitrógeno.....	18
1.5.2	<i>Eliminación de fósforo</i>	<i>19</i>
1.5.2.1	Eliminación biológica de fósforo.....	20
1.5.2.1.1	Metabolismo anaerobio de las bacterias PAO.....	21
1.5.2.1.2	Metabolismo aerobio y anóxico de las bacterias PAO	21
1.5.2.1.3	Cationes metálicos asociados al polifosfato	22
1.5.2.1.4	Las bacterias acumuladoras de glucógeno (GAO)	23
1.5.2.2	Eliminación físico-química de fósforo	24
1.6	ESQUEMAS DE PROCESO DE FANGOS ACTIVADOS	26
1.7	TECNOLOGÍAS ICA EN LOS SISTEMAS DE TRATAMIENTO DE AGUAS RESIDUALES.....	36
1.8	ANÁLISIS, DIAGNÓSTICO DE FALLOS Y MONITORIZACIÓN DE PROCESOS.....	43
1.8.1	<i>Problemática de la situación actual.....</i>	<i>43</i>
1.8.2	<i>Clasificación de los modelos</i>	<i>45</i>
1.8.3	<i>Control estadístico de procesos clásico</i>	<i>47</i>
1.8.4	<i>Técnicas estadísticas multivariantes de proyección sobre estructuras latentes ..</i>	<i>50</i>
1.8.4.1	Análisis de componentes principales (PCA).....	50
1.8.4.2	Regresión en mínimos cuadrados parciales (PLS).....	52
1.8.4.3	Regresión en componentes principales (PCR).....	53
1.8.4.4	Bondad de ajuste y de predicción	54
1.8.5	<i>Estructura de los datos.....</i>	<i>57</i>
1.8.5.1	Procesos en continuo	57
1.8.5.2	Procesos por lotes (discontinuos)	58
1.8.5.2.1	Desdoblamiento de la matriz tridireccional a matriz bidireccional	59

1.8.5.2.2	Análisis directo de la matriz tridireccional	61
1.8.6	<i>Monitorización de procesos por lotes (discontinuos)</i>	62
1.8.6.1	Selección del conjunto de observaciones de referencia	67
1.8.6.2	Modelación del proceso	68
1.8.6.2.1	Monitorización del error cuadrático de predicción	70
1.8.6.2.2	Monitorización de las variables latentes	72
1.8.6.2.3	Gráficos de contribución	74
1.8.6.3	Explotación en tiempo real del modelo construido	75
1.8.6.3.1	Enfoque de Nomikos y MacGregor (1995)	75
1.8.6.3.2	Enfoque de Rännar et al. (1998)	78
1.8.6.3.3	Enfoque de Wold et al. (1998)	79
1.8.6.4	El análisis multivariante en la monitorización de los tratamientos de aguas residuales	81
1.8.7	<i>Análisis de datos históricos</i>	88
1.8.8	<i>Modelos predictivos</i>	88
1.8.8.1	Las Redes Neuronales Artificiales (ANN)	89
1.8.8.2	Desarrollo de las ANN	90
1.8.8.3	Conceptos básicos	91
1.8.8.3.1	Neuronas	92
1.8.8.3.2	Estado de activación	93
1.8.8.3.3	Funciones de activación	93
1.8.8.3.4	Conexiones sinápticas	94
1.8.8.3.5	Aprendizaje	94
1.8.8.4	Modelo matemático de una neurona artificial	96
1.8.8.5	Características de las ANNs	97
1.8.8.6	ANNs y modelos estadísticos	99
1.8.8.7	Modelos predictivos en los tratamientos de aguas residuales	100
2	OBJETIVOS	109
3	MATERIALES Y MÉTODOS	113
3.1	DESCRIPCIÓN DEL MONTAJE EXPERIMENTAL	113
3.2	CARACTERIZACIÓN DEL INFLUENTE	116
3.3	INFORMACIÓN RECOGIDA	117
3.4	MÉTODOS ANALÍTICOS	120
4	RESULTADOS Y DISCUSIÓN	123
4.1	DIAGNÓSTICO DEL PROCESO	123
4.1.1	<i>Análisis del rendimiento en la eliminación de fósforo</i>	123

4.1.2	<i>Diagnóstico de fallos</i>	136
4.1.2.1	Enfoque de Nomikos y MacGregor.....	138
4.1.2.2	Enfoque de Wold et al.	141
4.1.2.3	Comparación de los enfoques.....	145
4.1.2.4	Modificaciones propuestas.....	159
4.2	MONITORIZACIÓN EN TIEMPO REAL DEL PROCESO.....	169
4.2.1	<i>Modelación del proceso</i>	170
4.2.2	<i>Resultados de la monitorización con distintos enfoques</i>	172
4.3	ESQUEMA DE MONITORIZACIÓN PROPUESTO.....	183
4.4	DESARROLLO Y COMPARACIÓN DE MODELOS PREDICTIVOS.....	187
4.4.1	<i>Modelo para la predicción de todas las variables respuesta</i>	190
4.4.2	<i>Modelos para la predicción de los incrementos de fósforo</i>	197
4.4.2.1	Modelos de proyección sobre estructuras latentes.....	197
4.4.2.2	Modelos de redes neuronales.....	202
4.4.2.2.1	Preprocesamiento.....	205
4.4.2.2.2	Entrenamiento.....	206
4.4.2.2.3	Validación.....	211
4.4.2.3	Comparación de los modelos desarrollados.....	212
5	CONCLUSIONES	227
6	DESARROLLOS FUTUROS	235
7	BIBLIOGRAFÍA	239
8	ANEJOS	255
8.1	ANEJO 1: FOTOGRAFÍAS DEL MONTAJE EXPERIMENTAL.....	255
8.2	ANEJO 2: GRÁFICOS DE CONTRIBUCIÓN DEL DIAGNÓSTICO.....	259
8.3	ANEJO 3: NOMENCLATURA.....	267

INTRODUCCIÓN

1 INTRODUCCIÓN

1.1 IMPORTANCIA DE LA *DEPURACIÓN DE LAS AGUAS*

En el desarrollo de la vida humana el agua juega un papel fundamental, por este motivo es necesario realizar una gestión adecuada de la misma que permita un suministro en buenas condiciones y que proteja los recursos hídricos existentes.

El notable aumento de la cantidad de vertidos producido en los últimos tiempos ha provocado un deterioro en la calidad de las aguas de los cauces receptores hasta tal punto que en muchas ocasiones la capacidad de autodepuración de éstos ha sido superada. A esta situación ha contribuido de manera importante el creciente desarrollo industrial, las actividades agrícolas con el uso a gran escala de fertilizantes y plaguicidas, así como la concentración de población en grandes núcleos urbanos.

Son varios los problemas de contaminación que afectan a las aguas. Entre ellos los relacionados con el exceso de materia orgánica y los provocados por el exceso de nutrientes, principalmente nitrógeno y fósforo, como la eutrofización y la contaminación de las aguas subterráneas. Debido a esto se hace necesario reducir la cantidad de vertidos que se realizan así como la carga contaminante de los mismos. Una manera eficaz de reducir este impacto ambiental es depurar estas aguas previamente a su vertido.

1.2 EL PROBLEMA DE LA *EUTROFIZACIÓN*

La eutrofización es uno de los principales problemas que produce la presencia de una excesiva cantidad de nutrientes en el agua. El término eutrofización describe toda una serie de sucesos que comienzan con el enriquecimiento en nutrientes del agua, especialmente de los compuestos de nitrógeno y fósforo, provocando un crecimiento acelerado de algas y especies vegetales superiores, que recubren la superficie del agua impidiendo el paso de luz solar hacia las capas inferiores. Bajo la capa superficial, la presencia de vegetación disminuye a causa del incremento de

turbidez y en el fondo, la biomasa generada es degradada por la actividad de las bacterias. En este proceso de degradación de la materia orgánica, las bacterias consumen el oxígeno disuelto existente, y las aguas dejan de ser aptas para la mayor parte de los seres vivos.

La eutrofización es un proceso lento y continuo que se produce de forma natural. Sin embargo, el enriquecimiento en nutrientes provocado por las actividades humanas acelera notablemente el proceso, y es entonces cuando se convierte en un grave problema de contaminación.

El problema de la eutrofización inicialmente se observó en masas de agua estancada, como son lagos, pantanos y estuarios, y en menor medida en ríos. Posteriormente, debido a los desequilibrios generados por la acción del hombre han aparecido problemas de eutrofia en otros medios acuáticos como ríos y mares, por lo que el concepto de eutrofización se ha hecho extensible a cualquier medio acuático.

Por otra parte, para el crecimiento y desarrollo de los organismos es necesaria la presencia de nutrientes. El término nutriente engloba un conjunto de especies químicas cuya característica común es la de ser imprescindibles para el desarrollo de la vida. Estos elementos, según las cantidades en que son necesarios, se clasifican en dos grupos:

- Micronutrientes: especies químicas que se necesitan en cantidades muy pequeñas, a nivel de traza, para el crecimiento celular (Cu, Fe, Zn, Co, B, Mo,...).
- Macronutrientes: especies químicas que son necesarias en mayor cantidad para el crecimiento celular (C, N, P, Si, S, Na, Ca, K, Mg,...).

En el caso de que alguno de estos nutrientes no estuviera presente, el crecimiento celular se inhibiría. Normalmente en un medio acuático, el nitrógeno y el fósforo son los nutrientes limitantes.

El fósforo suele ser el nutriente limitante en la mayor parte de los ecosistemas acuáticos, restringiendo así el crecimiento de algas y plantas, lo que puede provocar un retardo de la eutrofización. En cambio, la escasez de nitrógeno no siempre limita el crecimiento de algas. En algunas ocasiones se produce una eutrofia en un medio acuático continental con escasez de nitrógeno pero no de fósforo. En estos casos prolifera un tipo de algas capaces de fijar el nitrógeno atmosférico, algas cianofíceas, que además se reproducen rápidamente y pueden producir toxinas.

Como consecuencia de la eutrofización de un medio acuático, se reduce la biodiversidad ya que sólo sobreviven aquellas especies animales y vegetales que son capaces de adaptarse a las extremas condiciones ambientales que provoca la eutrofización, el agua no puede ser utilizada directamente para abastecimiento y pueden llegar a desaparecer los usos recreativos y turísticos.

1.3 LEGISLACIÓN

La normativa aplicable para el control del vertido de las aguas residuales en España viene recogida en el Real Decreto 509/1996, de 15 de marzo, de desarrollo del Real Decreto-Ley 11/1995, de 28 de diciembre. En este Real Decreto se incorpora la Directiva 91/271/CEE, de 21 de mayo, cuyo contenido no estaba incluido en el Título V de la Ley 29/1985, de 2 de agosto, de Aguas, o en el Título III de la Ley 22/1988, de 28 de julio, de Costas.

En el Real Decreto-ley 11/1995, se impone la aplicación de tratamientos a las aguas residuales urbanas antes de su vertido a las aguas continentales o marítimas. Además, se definen los criterios para la clasificación de los puntos de vertido en "zonas sensibles" y "zonas menos sensibles". Por tanto, los requisitos de vertido dependerán de la clasificación del punto de vertido.

Los vertidos de aguas residuales industriales en los sistemas de alcantarillado, sistemas colectores o en las instalaciones de depuración de aguas residuales urbanas necesitarán de un tratamiento previo para que se pueda garantizar que

dichas estaciones depuradoras de aguas residuales (EDARs) no tendrán problemas en el cumplimiento de los límites de vertido, ni que su tratamiento pueda suponer atender contra la salud del personal contratado en las EDARs y garantizarán que los lodos producidos por las EDARs puedan evacuarse con completa seguridad de forma aceptable desde la perspectiva medioambiental.

Todo vertido procedente de una EDAR urbana que posea tratamiento secundario o equivalente debe cumplir los valores de concentración o porcentajes de reducción que se indican en la Tabla 1.1.

Parámetro	Aplicable	Concentración	Porcentaje mínimo de reducción ⁽¹⁾
Demanda biológica de oxígeno (DBO ₅ a 20° C sin nitrificación) ⁽²⁾	Siempre	25 mg O ₂ /l	70-90%
Demanda química de oxígeno (DQO)	Siempre	125 mg O ₂ /l	75%
Sólidos totales en suspensión (SST)	Siempre	35 mg/l (más de 10.000 h-e)	90%
		60 mg/l (de 2.000 a 10.000 h-e)	70%
Fósforo total	Zona sensible	2 mg P/l (de 10.000 a 100.000 h-e)	80%
		1 mg P/l (más de 100.000 h-e)	
Nitrógeno total ⁽³⁾	Zona sensible	15 mg N/l (de 10.000 a 100.000 h-e)	70-80%
		10 mg N/l (más de 100.000 h-e)	

(1) Reducción relacionada con la carga del caudal de entrada.

(2) Este parámetro se puede sustituir por otro: carbono orgánico total (COT) o demanda total de oxígeno (DTO), si se puede establecer una correlación entre DBO₅ y el parámetro sustituido.

(3) Nitrógeno total equivalente a la suma de nitrógeno Kjeldahl total (N orgánico y amoniacal), nitrógeno en forma de nitrato y de nitrito.

Tabla 1.1. Requisitos de vertido para una EDAR de aguas residuales urbanas.

1.4 EL TRATAMIENTO DE AGUAS RESIDUALES

La depuración de las aguas residuales pretende conseguir un vertido libre de todos aquellos compuestos contaminantes que puedan provocar efectos negativos sobre el medio receptor. Las características del agua a tratar así como el punto en el que se vaya a realizar el vertido determinarán el tipo y grado de tratamiento que se realizará, por tanto, se hace necesario un análisis particular para cada caso concreto.

Los contaminantes presentes en las aguas residuales pueden ser tanto constituyentes químicos orgánicos e inorgánicos, como constituyentes biológicos. Para su eliminación se dispone en las EDARs de un conjunto de operaciones unitarias. Cada proceso que modifica la calidad de un agua residual es una operación unitaria. En función del agente responsable del cambio de calidad, estas operaciones se pueden clasificar de la siguiente manera:

- **Operaciones unitarias físicas.** Son los procesos que consiguen un cambio en la calidad del agua mediante la aplicación de fuerzas físicas. Entre estas operaciones cabe destacar: enrejados, mezclado, floculación, sedimentación y filtración.
- **Operaciones unitarias químicas.** Son aquellos procesos en los que la calidad del agua residual se modifica a través de reacciones químicas previa adición de algún reactivo químico. En el tratamiento de aguas residuales las operaciones unitarias químicas más importantes son: la adsorción, la precipitación y la desinfección.
- **Operaciones unitarias biológicas.** Son las operaciones en las que reacciones bioquímicas en las que intervienen microorganismos producen el cambio en la calidad del agua. Entre estas operaciones unitarias biológicas destacan el proceso de fangos activados y la digestión de fangos.

En una EDAR estas operaciones unitarias se combinan dando lugar a los pretratamientos, tratamientos primarios, tratamientos secundarios y tratamientos terciarios.

- **Pretratamientos.** Incluyen todas las operaciones unitarias cuyo objetivo es la eliminación de los sólidos de mayor tamaño que pueden atascar o dañar las instalaciones posteriores, así como las gravas, grasas y arenas. El pretratamiento consta de dos etapas: el desbaste y el desarenado-desengrasado.

- **Tratamientos primarios.** Engloban las operaciones unitarias destinadas a la eliminación de una parte importante de los sólidos suspendidos así como de la materia orgánica y nutrientes asociados a ellos. El tratamiento primario está constituido principalmente por el tamizado y la sedimentación.
- **Tratamientos secundarios.** Las operaciones unitarias que lo integran tienen por finalidad la eliminación de la materia orgánica biodegradable, de los sólidos suspendidos no eliminados en el tratamiento primario y, en ciertos casos, de los nutrientes. Habitualmente los tratamientos secundarios constan de procesos biológicos seguidos de un proceso de sedimentación. Los procesos biológicos tienen lugar en una serie de reactores en los que los microorganismos degradan la materia orgánica. Éstos son posteriormente separados del efluente por sedimentación en los decantadores. Para lograr la eliminación de nutrientes es necesario la imposición de ciertas condiciones ambientales en los reactores, influyendo de esta manera en la composición de los microorganismos que en ellos se desarrollen.
- **Tratamientos terciarios.** Engloban las operaciones unitarias encaminadas a la obtención de un efluente de alta calidad generalmente con la finalidad de que pueda ser reutilizado. El tratamiento terciario suele consistir en la desinfección y mejora general de la calidad del agua procedente de los tratamientos biológicos.

1.4.1 Tratamientos biológicos de aguas residuales

El principal objetivo de los tratamientos biológicos es la eliminación de la materia orgánica presente en el agua residual. Esta materia orgánica es eliminada mediante la acción de diferentes grupos de microorganismos (biomasa) a través de reacciones bioquímicas. Los microorganismos utilizan la materia orgánica como fuente de alimento (sustrato), tanto para producir materia celular como para obtener la energía necesaria para su metabolismo. Como resultado de este proceso de eliminación de

materia orgánica se obtienen nuevos microorganismos y una serie de compuestos de menor contenido energético que el sustrato inicial.

Con el paso del tiempo, se ha ido ampliando el campo de aplicación de los tratamientos biológicos. En la actualidad, además de la eliminación de la materia orgánica se persiguen en ocasiones otros objetivos como son la eliminación de nutrientes, como el nitrógeno y el fósforo, la eliminación de metales pesados y de compuestos organoclorados.

1.4.1.1 Clasificación de los tratamientos biológicos

En función del medio de soporte sobre el que crecen los microorganismos, los tratamientos biológicos de aguas residuales se pueden dividir en dos grandes grupos:

- **Procesos biológicos de cultivo en suspensión.** Los microorganismos se encuentran suspendidos en un medio líquido. Dentro de estos procesos pueden distinguirse cuatro grandes grupos: fangos activados, lagunas aireadas, lagunaje y tratamiento de fangos.
 - Fangos activados. Son procesos aerobios provistos de un sistema de decantación que permite separar el agua tratada de los microorganismos. El aporte de oxígeno se realiza por medios mecánicos, a través de agitadores de superficie o difusores. El esquema básico de este proceso se muestra en la Figura 1.1.

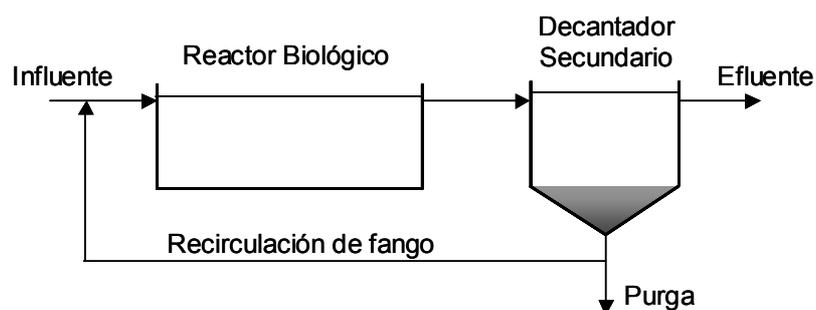


Figura 1.1. Esquema del proceso de fangos activados.

El agua residual se introduce en el reactor, donde se mantiene un cultivo aerobio en suspensión. El contenido del reactor se conoce como “licor mezcla”. En el reactor aerobio se producen las transformaciones bioquímicas de eliminación de materia orgánica. El licor mezcla se conduce hasta el decantador secundario, donde la biomasa es separada del agua tratada. Una parte de la biomasa sedimentada se elimina del proceso por medio de la corriente de purga, y la otra parte se recircula al reactor biológico. De esta manera, se consigue mantener una cantidad constante de biomasa en el sistema, y además que el tiempo de permanencia de los microorganismos en el sistema sea mayor que el tiempo de retención del agua residual.

El tiempo medio de residencia de los microorganismos en el reactor, también conocido como tiempo de retención celular (TRC) o edad del fango, es el parámetro de mayor importancia en el proceso de fangos activados, y se utiliza habitualmente para su diseño. El TRC determina tanto el rendimiento en la eliminación de materia orgánica como la cantidad de oxígeno requerida en el proceso. Este parámetro se puede controlar con la cantidad de fangos que se eliminan del sistema a través de la corriente de purga.

- Lagunas aireadas. Son predominantemente aerobias aunque pueden combinarse con procesos anaerobios. La aireación también es de tipo mecánico consiguiendo mantener la materia orgánica y la biomasa en suspensión. Normalmente no se realiza la recirculación de fangos, por lo que el tiempo de retención necesario se consigue con el volumen del reactor. Después de la laguna existe una etapa de decantación que permite separar el agua tratada de los fangos biológicos.
- Lagunaje. Las aguas residuales se vierten en estanques de tierra, extensos y poco profundos, donde son tratadas por medios naturales. Pueden ser estrictamente aerobios, estrictamente anaerobios o aerobios-anaerobios. Carecen de un sistema de recirculación de fangos que permita aumentar el

tiempo de retención celular. El oxígeno necesario se obtiene por reaireación natural a través de la superficie y de la reacción de fotosíntesis de las algas. El oxígeno producido por las algas es empleado por las bacterias aerobias para la degradación de la materia orgánica. Los productos resultantes de esta degradación son utilizados de nuevo por las algas. De esta manera, se establece una relación simbiótica entre algas y bacterias.

- Tratamiento de fangos. Estos procesos se utilizan para estabilizar los fangos generados en otros tratamientos biológicos, principalmente en los fangos activados, así como los fangos procedentes del tratamiento primario. No poseen recirculación y la digestión del fango puede ser aerobia o anaerobia. El tratamiento de fangos consigue reducir la masa y volumen que deben verse, posibilitando, en ocasiones, su utilización agrícola.
- **Procesos biológicos de soporte sólido**. En este tipo de procesos, los microorganismos se encuentran fijados sobre un soporte inerte formando una película de pequeño espesor. El medio soporte puede encontrarse fijo en una columna, fluyendo el agua residual a través de la misma, o puede girar alrededor de un eje, moviéndose dentro del fluido. Esto da lugar a los dos tipos fundamentales de tratamientos mediante cultivo fijo: filtros percoladores y contactores biológicos rotatorios.
 - Filtros percoladores. Estos sistemas constan de un medio poroso, en el que se encuentra la biomasa, a través del cual se hace pasar el agua residual a depurar. El medio poroso permite el paso del aire en contracorriente con el agua para suministrar el oxígeno necesario para que tenga lugar el proceso biológico. La operación correcta de estos filtros implica un desprendimiento continuo y uniforme de los fangos que, posteriormente, son separados del agua en el decantador secundario. Se puede operar con recirculación o sin ella.
 - Contactores biológicos rotativos. Cabe destacar los biodiscos en los que el soporte sólido está constituido por discos de material plástico ensamblados

sobre un eje horizontal. La rotación del eje permite que de forma alternativa la biomasa entre en contacto con el sustrato orgánico del agua residual y con la atmósfera.

1.4.1.2 Organismos que intervienen en los tratamientos biológicos

La depuración biológica de las aguas residuales basa su funcionamiento en la actividad microbiológica de una serie de organismos. Estos microorganismos están presentes en el agua residual pero es en la EDAR donde se les proporciona las condiciones adecuadas para su proliferación.

Los organismos que intervienen en los procesos biológicos se pueden clasificar según diversos criterios. Las clasificaciones más utilizadas se basan en el tipo de fuente de carbono y en el tipo de aceptor de electrones que interviene en el metabolismo del microorganismo.

Atendiendo a la fuente de carbono que utilizan los microorganismos, éstos se clasifican en dos grupos:

- **Autótrofos.** Son organismos capaces de sintetizar materia orgánica a partir de compuestos inorgánicos. Utilizan como fuente de carbono el CO₂ disuelto o atmosférico, y como dador de electrones compuestos inorgánicos tales como el amonio o el nitrito.
- **Heterótrofos.** Son organismos que necesitan compuestos orgánicos previamente sintetizados por otros organismos, siendo incapaces de utilizar compuestos de carbono inorgánico como fuente de carbono. Utilizan como dador de electrones los mismos compuestos orgánicos que emplean como fuente de carbono.

Los microorganismos también se pueden clasificar según el tipo de aceptor de electrones que utilizan:

- **Aerobios.** Son organismos que sólo pueden utilizar como aceptor de electrones el oxígeno, no pudiendo sobrevivir en ausencia de éste.
- **Anaerobios estrictos.** Son aquellos que sólo son capaces de desarrollarse en medios donde no existe el oxígeno.
- **Anaerobios facultativos.** Estos microorganismos utilizan el oxígeno cuando está presente en el medio, pero son capaces de desarrollarse empleando otro aceptor de electrones diferente cuando no hay oxígeno.
- **Anaerobios aerotolerantes.** Estos microorganismos utilizan para su crecimiento un aceptor de electrones diferente al oxígeno, pero la presencia del oxígeno en el medio no representa una inhibición del proceso.

Los organismos presentes en los sistemas de tratamiento biológico son del mismo tipo que los existentes en sistemas naturales, pero las condiciones especiales que se dan en el interior del reactor seleccionan unas especies determinadas. Los principales grupos de organismos que se encuentran en los procesos de depuración de aguas residuales son:

- **Bacterias.** Son organismos unicelulares que no tienen núcleo celular diferenciado (procariotas). Constituyen normalmente el 95% de la biomasa de los fangos activados e intervienen en múltiples procesos. Existen tanto formas autótrofas como heterótrofas.
- **Protozoos.** Son organismos unicelulares con núcleo celular diferenciado del resto de la célula y rodeado por una membrana (eucariotas). Son heterótrofos y la mayoría aerobios o anaerobios facultativos. Aproximadamente representan un 5 % de la biomasa de los fangos activados. Se les considera buenos indicadores del estado del proceso de depuración porque son fácilmente identificables en el microscopio y muy sensibles tanto a los compuestos tóxicos como a los cambios en las condiciones de operación.

- **Metazoos.** Son organismos pluricelulares, heterótrofos y aerobios. En los fangos activados representan el escalón más elevado de la pirámide trófica.
- **Hongos.** Son organismos pluricelulares, heterótrofos y en su mayoría aerobios. En los sistemas de depuración biológica pueden competir con las bacterias por la materia orgánica disuelta. Si las condiciones de operación son adecuadas, su presencia en la biomasa es mínima.
- **Algas.** Son organismos fotoautótrofos, es decir, capaces de sintetizar compuestos orgánicos a partir de sustancias inorgánicas con ayuda de luz solar como fuente de energía. Su importancia en los procesos de depuración se centra casi exclusivamente en el tratamiento de lagunaje.

1.5 ELIMINACIÓN DE NUTRIENTES EN LAS AGUAS RESIDUALES

Como consecuencia del incremento de los problemas de eutrofización en ríos, lagos y costas, se ha producido en las últimas décadas un notable aumento del interés por la eliminación de nutrientes en las aguas residuales previamente a su vertido en los sistemas naturales. En las EDARs se puede llevar a cabo la eliminación de nutrientes tanto por vía físico-química como por vía biológica, siendo esta última la que mayor atención está recibiendo en los últimos tiempos.

1.5.1 Eliminación de Nitrógeno

En las aguas residuales el nitrógeno puede aparecer bajo distintas formas: nitrógeno orgánico, amoniacal, nitrito y nitrato. La suma de las concentraciones de todas ellas permite obtener el contenido total de nitrógeno. Las distintas formas del nitrógeno en el agua residual se presentan en la Figura 1.2.

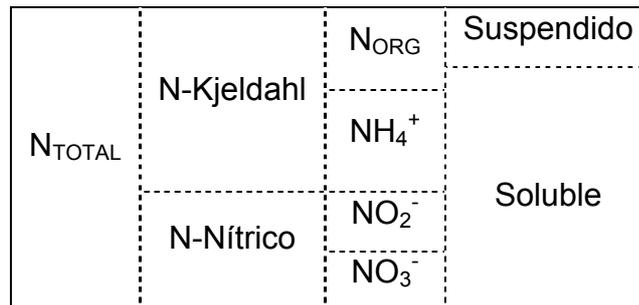


Figura 1.2. Formas del nitrógeno en el agua residual.

Inicialmente el nitrógeno en un agua residual se encuentra principalmente en forma de nitrógeno orgánico (urea y proteínas), siendo la actividad bacteriana la responsable de su transformación a forma amoniacal. El nitrógeno amoniacal en disolución acuosa puede estar en forma de amoniaco o de ión amonio dependiendo del pH de la disolución, siendo esta última la forma predominante en las aguas residuales.

El nitrógeno en forma de nitrito es tóxico para gran parte de la fauna piscícola y demás especies acuáticas. Afortunadamente el nitrito es relativamente inestable y fácilmente oxidable a nitrato. En muy pocas ocasiones su concentración es superior a 1 mg/l en el agua residual y a 0.1 mg/l en el caso de aguas subterráneas y superficiales (Metcalf y Eddy, 2000).

La forma más oxidada del nitrógeno presente en las aguas residuales es el nitrato. Es poco probable que su presencia en las aguas residuales se deba a la oxidación del nitrógeno amoniacal en el propio sistema de alcantarillado, puesto que el agua residual tiene una gran demanda de oxígeno. Por tanto, cuando la concentración de nitratos en un agua residual es alta, lo más razonable es pensar que su origen está en el agua de abastecimiento y no en transformaciones bioquímicas.

1.5.1.1 Eliminación biológica de nitrógeno

Como ya se ha comentado, el amonio es la forma más habitual en que aparece el nitrógeno en las aguas residuales. Los microorganismos lo utilizan como nutriente para sintetizar su tejido celular, por tanto, en un proceso biológico convencional

siempre será eliminada una cierta cantidad de nitrógeno. Según Sedlak (1991) esta cantidad representa entre un 12 y un 13% en peso de la biomasa formada. Debido a que esta eliminación es insuficiente para cumplir los requisitos legales de vertido a zonas sensibles, es necesario recurrir a un proceso de nitrificación-desnitrificación.

En la Figura 1.3 se muestran las transformaciones del nitrógeno que tienen lugar en los procesos de tratamiento. La transformación del amonio en nitrato recibe el nombre de nitrificación, y la del nitrato en nitrógeno gaseoso, desnitrificación.

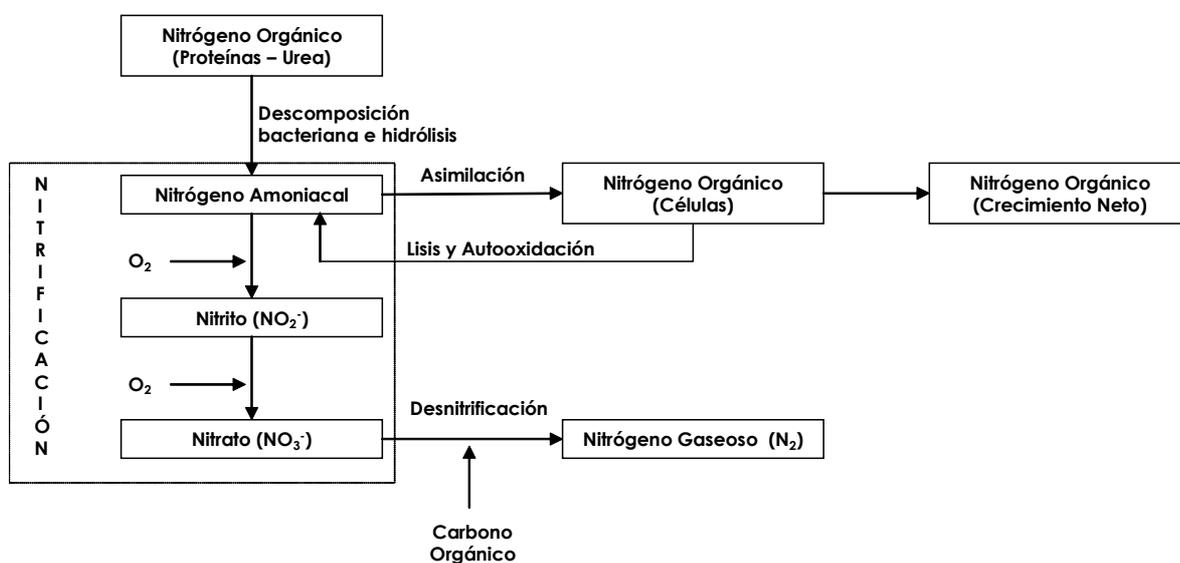


Figura 1.3. Transformaciones del nitrógeno en los procesos de tratamiento biológico (Metcalf & Eddy, 2000)

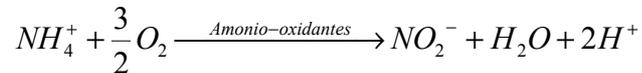
1.5.1.1.1 Nitrificación

Se denomina nitrificación al proceso biológico por el cual las bacterias autótrofas nitrificantes oxidan el amonio a nitrato. Dos fases constituyen este proceso: en la primera el grupo de bacterias autótrofas *amonio-oxidantes* oxidan el nitrógeno amoniacal a nitritos, mientras que en la segunda son las bacterias autótrofas *nitrito-oxidantes* las que se encargan de oxidar los nitritos a nitratos.

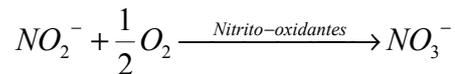
La concentración de nitritos siempre es mucho menor que la de nitratos porque en el intervalo habitual de temperatura de trabajo (10-20°C) la velocidad de crecimiento de las bacterias *nitrito-oxidantes* es mayor que la de las bacterias *amonio-oxidantes*.

La estequiometría de las reacciones de la nitrificación es la siguiente:

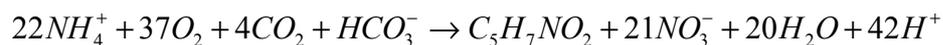
1. Proceso de oxidación de amonio a nitrito



2. Proceso de oxidación de nitrito a nitrato



Asumiendo que la biomasa tiene la composición química $C_5H_7NO_2$, la reacción global propuesta para el proceso de nitrificación es la siguiente:



Para que el proceso de nitrificación pueda producirse es necesario que se mantengan unas condiciones ambientales adecuadas para el crecimiento de las bacterias autótrofas nitrificantes. Los factores ambientales que en mayor medida pueden afectar a la nitrificación son: concentración de amonio y nitrito, temperatura, concentración de oxígeno disuelto, pH, tiempo de retención celular y presencia de sustancias tóxicas.

En la mayoría de los sistemas de fangos activados, los procesos de nitrificación tienen lugar junto con el proceso de oxidación de la materia orgánica.

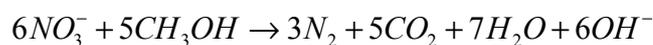
1.5.1.1.2 Desnitrificación

La desnitrificación es el proceso biológico mediante el cual las bacterias heterótrofas desnitrificantes transforman el nitrato en nitrógeno gaseoso, completando así la eliminación biológica de nitrógeno. El nitrógeno gaseoso pasa a la atmósfera cuando su concentración supera la de saturación. Se trata de un proceso que tiene lugar en condiciones anóxicas, término que se emplea para indicar la ausencia de oxígeno disuelto en el medio y la presencia de nitrato como aceptor de electrones.

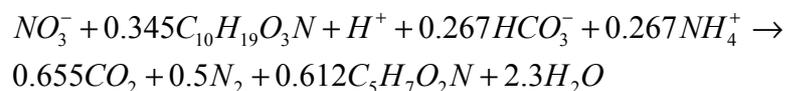
La mayoría de las bacterias capaces de reducir el nitrato utilizándolo como aceptor de electrones, en lugar del oxígeno, son heterótrofas facultativas, aunque existen algunas bacterias autótrofas que también pueden realizar este proceso.

Dado que las bacterias anaerobias facultativas son capaces de utilizar el oxígeno cuando éste está presente, para favorecer el proceso de desnitrificación será de gran importancia la ausencia de oxígeno en el medio.

La reacción global para la reducción del nitrato a nitrógeno gas empleando metanol como fuente de carbono, puede representarse mediante la siguiente ecuación:



Cuando se considera además la síntesis de biomasa para las bacterias que realizan el proceso de desnitrificación, asumiendo $C_5H_7NO_2$ como composición química de los microorganismos, se obtiene para la reacción global la siguiente ecuación:



Los factores que tienen mayor influencia en el proceso de desnitrificación son: tipo y concentración de sustrato orgánico, concentración de oxígeno disuelto, concentración de nitratos, temperatura y pH.

1.5.1.2 Eliminación físico-química de nitrógeno

Existen procesos físico-químicos para la eliminación de nitrógeno de las aguas residuales, como la cloración al *break-point*, el *air stripping* o el intercambio iónico. Sin embargo, en el campo del tratamiento de aguas residuales estos métodos no son competitivos ni funcional ni económicamente con la eliminación de nitrógeno por vía biológica.

1.5.2 Eliminación de fósforo

En las aguas residuales el fósforo se puede presentar bajo distintas formas: ortofosfatos, polifosfatos, y fósforo orgánico. La suma de las concentraciones de todas ellas permite obtener el contenido total de fósforo. El fósforo total a su vez se puede dividir en fósforo soluble y fósforo suspendido. La fracción soluble incluye a los ortofosfatos, los polifosfatos y parte del fósforo orgánico (Figura 1.4).

- **Ortofosfatos** (PO_4^{3-} , HPO_4^{2-} , H_2PO_4^- , H_3PO_4). Su concentración en un agua residual urbana suele encontrarse entre 3 y 7 mg P / l. La cantidad de cada especie vendrá determinada por el pH del agua residual.
- **Polifosfatos** (P_2O_7). Proceden mayoritariamente de los detergentes, aunque el uso de éstos cada vez es menor. Su concentración en un agua residual urbana oscila entre 2 y 4 mg P / l.
- **Fósforo orgánico**. Está constituido por ácidos nucleicos, fosfolípidos y adenosín trifosfato (ATP) provenientes de residuos animales y alimenticios. Su concentración típica en un agua residual urbana está alrededor de 1 mg P / l.

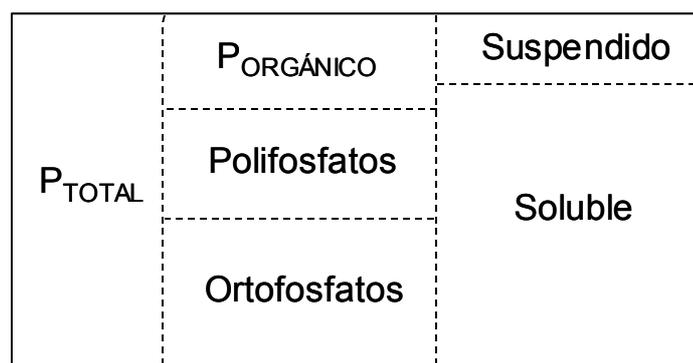


Figura 1.4. Fracciones del fósforo en el agua residual.

1.5.2.1 Eliminación biológica de fósforo

El fósforo, al igual que el nitrógeno, es un nutriente necesario para el crecimiento de los microorganismos. Por tanto, en un proceso de fangos activados convencional siempre se eliminará una cierta cantidad de fósforo debido a su utilización como nutriente por parte de los microorganismos. Esta cantidad se encuentra entre el 1.5 y el 2% en peso de la biomasa producida (Grady *et al.*, 1999). Como esta cantidad de fósforo que se elimina no es suficiente para cumplir los requisitos legales de vertido a zonas sensibles, se necesita emplear procedimientos específicos que permitan reducir la cantidad de fósforo que se vierte.

Para llevar a cabo el proceso de eliminación biológica de fósforo, se somete a la biomasa a una alternancia de condiciones anaerobias y aerobias. De esta manera, se favorece el desarrollo de unas bacterias capaces de almacenar intracelularmente gran cantidad de fósforo en forma de polifosfatos. Greenburg *et al.* en 1955 fueron los primeros que observaron este proceso. Para el desarrollo de estas bacterias acumuladoras de polifosfatos (PAOs, *Polyphosphate Accumulating Organisms*) además de la alternancia de condiciones, también es necesaria la presencia de ácidos volátiles de cadena corta en la fase anaerobia del proceso.

Desde los años 50 se han realizado numerosos estudios para tratar de explicar el metabolismo de estas bacterias. En la bibliografía existen diversos modelos para representar el metabolismo de las bacterias PAO. Los dos más ampliamente aceptados son: el modelo de Comeau y Wentzel, que fue desarrollado de forma independiente por Comeau *et al.* (1986) y Wentzel *et al.* (1986), y el propuesto por Mino (1987). Sin embargo, trabajos realizados con posterioridad demostraron que las hipótesis del modelo de Mino permiten reproducir mejor las observaciones experimentales (Bordacs y Chiesa, 1989; Maurer *et al.*, 1997; Jeon y Park, 2000).

1.5.2.1.1 Metabolismo anaerobio de las bacterias PAO

En condiciones anaerobias, las bacterias PAO capturan los ácidos grasos volátiles de cadena corta y los almacenan en su interior en forma de poli-hidroxi-alcanoatos (PHA). En caso de que el principal sustrato fuera ácido acético, éste sería almacenado en forma de poli-hidroxi-butirato (PHB). La energía necesaria para transportar y almacenar el sustrato, la obtienen de la degradación de los polifosfatos intracelulares a fosfatos. El fosfato es liberado al medio produciendo un incremento de la concentración de ortofosfatos en el mismo.

En el modelo de Mino (Figura 1.5), el origen del poder reductor necesario para la síntesis del PHA se obtiene de la degradación del glucógeno almacenado intracelularmente. Según este modelo durante la captura del acetato se produce un descenso en los carbohidratos contenidos en las bacterias PAO que está relacionado estequiométricamente con la cantidad de acetato tomado.

1.5.2.1.2 Metabolismo aerobio y anóxico de las bacterias PAO

En condiciones aerobias, el PHA que han almacenado las bacterias PAO en la etapa anaerobia es utilizado para obtener la energía que permite llevar a cabo los procesos de crecimiento y producción de glucógeno. El exceso de energía es utilizado para almacenar polifosfatos que se forman con fósforo capturado del medio que rodea a la célula. De esta forma, las bacterias PAO son capaces de capturar fósforo en cantidades superiores a las liberadas en la fase anaerobia, dando lugar, por tanto, a una eliminación neta de fósforo.

En condiciones anóxicas las bacterias PAO utilizan el nitrato como aceptor de electrones para realizar las mismas funciones que en condiciones aerobias (Figura 1.5).

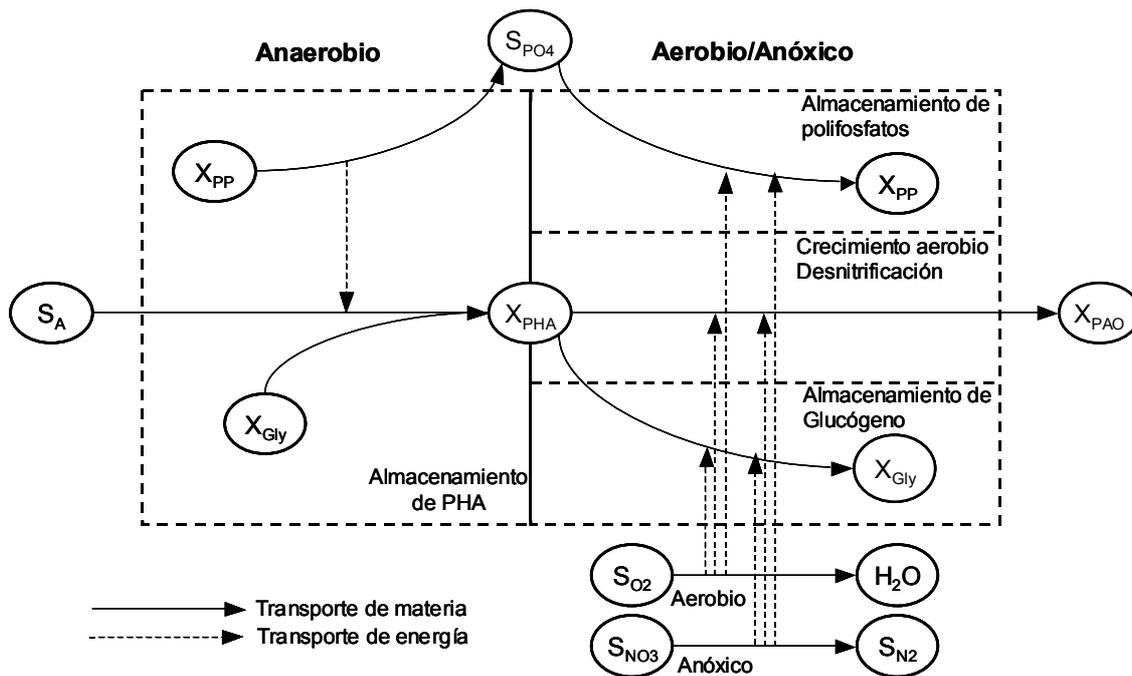


Figura 1.5. Metabolismo de las bacterias PAO propuesto por Mino (Mino *et al.*, 1995).

1.5.2.1.3 Cationes metálicos asociados al polifosfato

En los procesos de eliminación biológica de fósforo junto con la liberación y captura de fósforo del medio, también son liberados y capturados determinados cationes metálicos, como el magnesio y el potasio. Estos cationes además de ser necesarios como nutrientes para el crecimiento de los microorganismos, también son necesarios para mantener la electroneutralidad en el interior de la célula (Comeau *et al.*, 1987).

En varios trabajos se ha demostrado que el potasio y el magnesio no sólo presentan un comportamiento similar al del fósforo, sino que además estos cationes son imprescindibles para que el proceso de eliminación biológica de fósforo tenga lugar (Rickard y McClintock, 1992; Pattarkine y Randall, 1999). En estos trabajos también se estudió el comportamiento del ión calcio, pero éste parecía no intervenir directamente en el proceso biológico.

Sin embargo, Schönborn *et al.* (2001) encontraron que junto con el potasio y el magnesio, también el calcio era uno de los principales componentes de los gránulos de polifosfato. Estos autores también observaron que el calcio permanecía constante en el medio, mientras que el potasio y el magnesio eran liberados y tomados

conjuntamente con el fósforo. Por este motivo, plantearon como hipótesis la existencia de dos tipos de polifosfatos: los “activos” que están compuestos principalmente por magnesio, potasio y fósforo, y que sí que participan en los procesos de liberación y captura de fósforo, y los “inactivos” en cuya composición predomina el calcio y que no participan.

Los valores obtenidos por diferentes autores para las relaciones molares $\Delta\text{Mg}/\Delta\text{P}$ y $\Delta\text{K}/\Delta\text{P}$ en los procesos de liberación y captura de fósforo se presentan en la Tabla 1.2. Se puede observar que existe una cierta disparidad entre los valores encontrados en la bibliografía.

Referencia	$\Delta\text{Mg}/\Delta\text{P}$ (mol / mol)	$\Delta\text{K}/\Delta\text{P}$ (mol /mol)
Miyamoto-Mills <i>et al.</i> , 1983	0.26	0.27
Arvin y Kristensen, 1985	0.32	0.23
Comeau <i>et al.</i> , 1987		0.33
Imai <i>et al.</i> , 1988	0.36	
Rickard y McClintock, 1992	0.30	0.21
Maurer y Gujer, 1995	0.32	0.29
Jardin y Pöpel, 1994	0.30	0.26
Brdjanovic <i>et al.</i> , 1996	0.33	0.33
Jardin y Pöpel, 1996	0.34	0.26
Wild <i>et al.</i> , 1997	0.32	

Tabla 1.2. Valores encontrados en bibliografía para las relaciones molares $\Delta\text{Mg}/\Delta\text{P}$ y $\Delta\text{K}/\Delta\text{P}$ en los procesos de eliminación biológica de fósforo.

1.5.2.1.4 Las bacterias acumuladoras de glucógeno (GAO)

Existe un grupo de microorganismos denominado bacterias GAO (*Glycogen Accumulating Organisms*), que al igual que las bacterias PAO, son capaces de desarrollarse en aquellos sistemas en los que se alternan condiciones anaerobias y aerobias.

En condiciones anaerobias, estas bacterias son capaces de tomar determinados sustratos como glucosa y ácidos grasos volátiles sin que dicha toma vaya acompañada por la suelta de fósforo (ver Figura 1.6). Los primeros que observaron la toma ácidos volátiles sin suelta de fósforo en la etapa anaerobia fueron Fukase *et al.* (1985). Por tanto, las bacterias GAO compiten con las PAO por el sustrato, y en

aquellos sistemas en que las GAO consigan proliferar, se producirá un deterioro del proceso de eliminación biológica de fósforo.

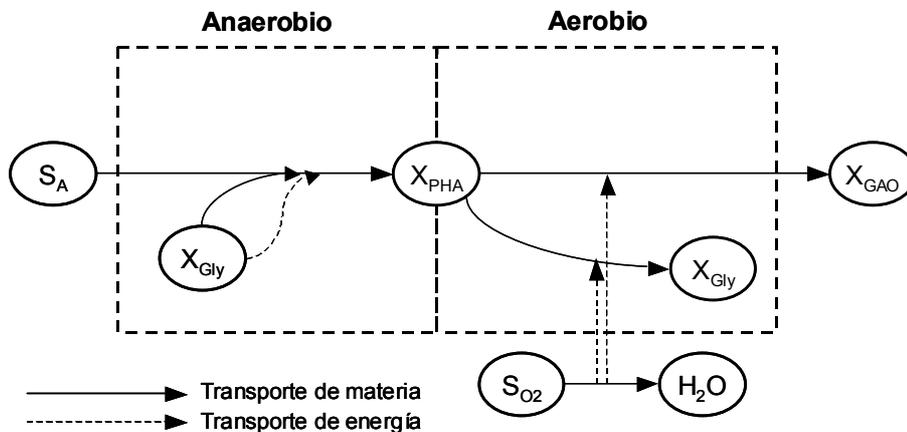


Figura 1.6. Metabolismo GAO.

En condiciones anaerobias, tanto las bacterias PAO como las bacterias GAO, degradan el glucógeno que tienen almacenado en su interior para obtener la energía necesaria para almacenar los ácidos grasos volátiles en forma de PHA. Las bacterias PAO además aportan energía procedente de la degradación de los polifosfatos intracelulares.

1.5.2.2 Eliminación físico-química de fósforo

Tradicionalmente la precipitación química ha sido el método más utilizado para la eliminación de fósforo en las aguas residuales, debido principalmente a la facilidad para llevar a cabo el proceso y a su gran estabilidad.

La precipitación química del fósforo se lleva a cabo mediante la adición al agua residual de sales metálicas. Las sales más utilizadas son el sulfato de aluminio, el aluminato de sodio, el cloruro férrico, el sulfato férrico, el cloruro ferroso y el sulfato ferroso. Además de sales metálicas, también es habitual la utilización de cal. La elección del reactivo a emplear dependerá principalmente del pH y de su coste económico.

Tras la adición del reactivo, el ión metálico de la sal interacciona con el ortofosfato formando un precipitado insoluble (fosfato metálico) que es eliminado posteriormente del agua residual mediante un proceso de sedimentación. La dosis del reactivo necesaria depende tanto de las características del agua residual como del punto de la EDAR donde se realice la adición. Son varios los puntos de la EDAR en los que se puede añadir el reactivo: antes del decantador primario, en el proceso biológico, y en el decantador secundario (a la entrada o a la salida).

La dosificación a la entrada al decantador primario permite una disminución de los costes de operación en el tratamiento secundario, aunque a cambio sobrecarga el proceso de digestión. El reactivo añadido actúa como coagulante para la materia orgánica particulada incrementando el porcentaje de sólidos eliminados en el decantador primario. El control de la dosis añadida es de vital importancia ya que un exceso de reactivo podría provocar que no llegara suficiente cantidad de fósforo al reactor biológico para el crecimiento de los microorganismos.

La adición de reactivo en el reactor biológico permite reducir la dosis necesaria puesto que la elevada turbulencia permite una mezcla más efectiva. La cantidad de sólidos suspendidos generados en el reactor aumenta, lo que puede suponer una disminución del tiempo de retención celular.

Dosificar a la entrada del decantador secundario presenta el inconveniente de no disponer de ninguna zona de turbulencia elevada que garantice una mezcla efectiva del reactivo, siendo necesario, por tanto, dosificar en exceso. Por otra parte, añadir el reactivo a la salida del decantador secundario requiere de una serie de instalaciones adicionales de mezcla, reacción y separación, que no son necesarias en el resto de casos.

La eliminación de fósforo por medio de la precipitación química permite obtener buenos resultados, pero presenta dos grandes inconvenientes: el consumo de reactivos que lleva asociado un coste económico, y la elevada cantidad de fangos producidos.

1.6 ESQUEMAS DE PROCESO DE FANGOS ACTIVADOS

El proceso de fangos activados es el tratamiento biológico más utilizado en la depuración de las aguas residuales. Bajo el esquema general presentado en la Figura 1.1, modificaciones en la distribución y forma del reactor biológico dan lugar a diferentes configuraciones y modos de operación que permiten la eliminación de materia orgánica y en algunos casos también la eliminación de nutrientes.

El esquema más sencillo para la eliminación de la materia orgánica presente en el agua residual es el mostrado en la Figura 1.1. En dicho esquema, el reactor empleado puede ser de flujo en pistón, de mezcla completa o de flujo disturbado.

- Reactores de flujo en pistón. Los primeros procesos de fangos activados en continuo emplearon reactores estrechos y alargados en forma de canal. En estos reactores el agua residual entra por un extremo y circula en régimen de flujo en pistón hacia el extremo opuesto. De esta forma, se produce una variación a lo largo de la longitud del reactor de las concentraciones de sustrato y de microorganismos. La concentración de sustrato decrece a lo largo del reactor lo que produce que las necesidades de oxígeno no sean constantes sino que también disminuyan a lo largo del mismo. Por este motivo, en este tipo de reactores los difusores no se instalan de manera uniforme sino adaptándose a las necesidades de oxígeno.
- Reactores de mezcla completa. En este caso el contenido del reactor está completamente mezclado, es decir, que las concentraciones de todos los componentes son iguales en todos los puntos del reactor. Esto permite que este tipo de reactores sean capaces de resistir mejor que los de flujo en pistón vertidos puntuales con alta carga orgánica o con presencia de compuestos tóxicos, ya que el agua residual que entra se dispersa uniformemente por todo el reactor. Además, mediante una configuración compuesta por una serie de reactores de mezcla completa es posible aproximar razonablemente el funcionamiento de un reactor de flujo en pistón. El hecho de utilizar reactores de mezcla completa en serie confiere flexibilidad

al sistema de tratamiento ya que permite imponer en cada reactor unas condiciones ambientales distintas (reactores anaerobios, anóxicos y aerobios) y, por tanto, junto con la eliminación de la materia orgánica, conseguir también la eliminación de nutrientes del agua residual.

- Reactores de flujo disturbado. En este caso el agua circula a gran velocidad por el reactor dando lugar a una situación intermedia entre las dos anteriores.

Todos los esquemas de tratamiento desarrollados para la eliminación biológica de nutrientes están basados en el proceso de fangos activados e incorporan en su configuración una combinación de zonas anaerobias, anóxicas y aerobias. Según el nutriente que se pretenda eliminar, se tendrían los siguientes casos:

- Los procesos biológicos de tratamiento para la eliminación biológica de nitrógeno combinan zonas anóxicas y aerobias. Los más utilizados emplean como fuente de carbono para el proceso de desnitrificación el propio agua residual, evitando de esta forma la adición de un sustrato externo que implicaría un coste adicional. Los esquemas más representativos son: el proceso Ludzack-Ettinger modificado, el proceso Bardenpho y los canales de oxidación.
- Cuando con el proceso biológico de tratamiento se persigue la eliminación biológica de fósforo es necesario exponer la biomasa a la alternancia de fases anaerobias y aerobias, de esta forma se favorece el desarrollo de las bacterias acumuladoras de polifosfatos. El esquema típico de este proceso es el A/O.
- Cuando el objetivo del proceso biológico de tratamiento es la eliminación conjunta de nitrógeno y de fósforo se hace necesario la combinación de zonas anaerobias, anóxicas y aerobias. Los esquemas de tratamiento más empleados son el A²/O y el UCT.

Todos los procesos que se han indicado anteriormente se corresponden con procesos en continuo, y cuando en ellos se pretende imponer las condiciones ambientales que favorezcan el desarrollo de un determinado tipo de biomasa se recurre al uso de varios reactores. Es decir, las condiciones ambientales a las que se expone la biomasa se modifican espacialmente, cambian de un reactor a otro. Sin embargo, el proceso de fangos activados también puede ser llevado a cabo en discontinuo, en lo que se conoce como proceso por lotes o proceso *batch*, como es el caso de los reactores discontinuos secuenciales (SBRs, *Sequencing Batch Reactors*). En estos sistemas es posible imponer distintas condiciones ambientales en un mismo reactor, no de manera simultánea sino a lo largo del tiempo. Esto hace posible que empleando un único reactor pero cambiando su modo de operación se puedan conseguir distintos objetivos de depuración: eliminación de materia orgánica, de materia orgánica y nitrógeno, de materia orgánica y fósforo, y también de materia orgánica, nitrógeno y fósforo.

A continuación se muestran y describen brevemente todos estos esquemas de tratamiento que tienen por objetivo la eliminación biológica de nutrientes.

Esquema Ludzack-Ettinger modificado. Este sistema de tratamiento consta de dos reactores en serie (Figura 1.7). En el primero tiene lugar la desnitrificación utilizando el carbono presente en el agua residual de entrada como sustrato y los nitratos como aceptores de electrones. El segundo reactor se mantiene en condiciones aerobias y en él se produce de manera simultánea la degradación de la materia orgánica y la oxidación del nitrógeno amoniacal a nitrato. Para transferir los nitratos generados en la zona aerobia a la zona anóxica se establece una corriente de recirculación interna de licor mezcla que permite regular el grado de desnitrificación. Los fangos decantados son recirculados al reactor anóxico.

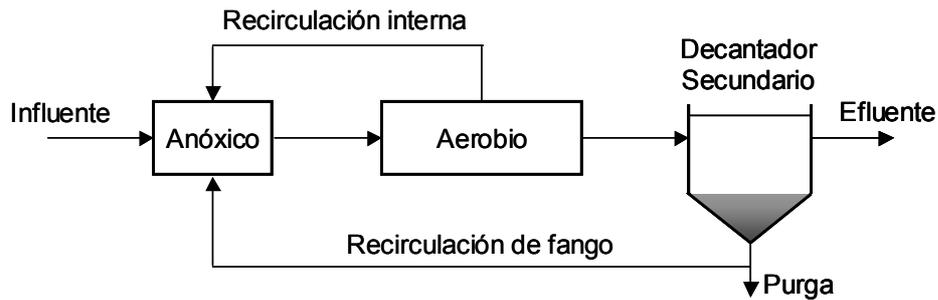


Figura 1.7. Esquema Ludzack-Ettinger modificado.

Esquema Bardenpho. Este proceso incorpora cuatro etapas en serie alternando entre zonas anóxicas y aerobias (Figura 1.8). El agua residual a tratar entra en el primer reactor anóxico al que se recircula licor mezcla nitrificado procedente del reactor aerobio que está situado a continuación. Tras el primer reactor aerobio se encuentra el segundo reactor anóxico en el que se completa la desnitrificación utilizando como fuente de carbono la materia orgánica procedente de la degradación endógena de los microorganismos.

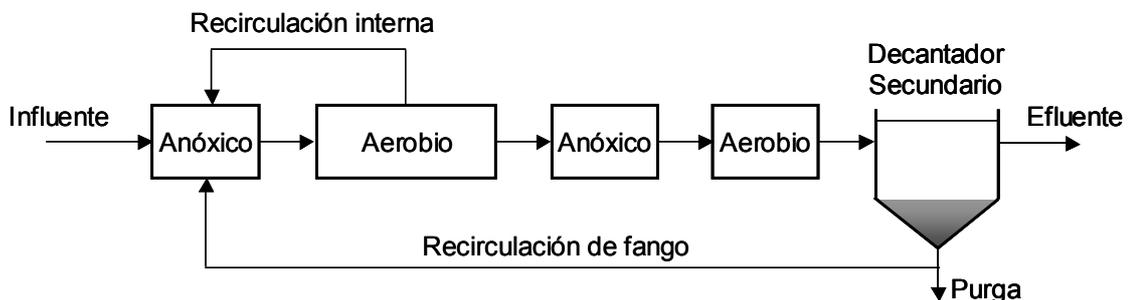


Figura 1.8. Esquema Bardenpho.

Finalmente, antes del decantador secundario se dispone un reactor aerobio de pequeñas dimensiones cuya finalidad es la de desplazar el nitrógeno gas producido. Esto evita que se produzca la desabsorción del nitrógeno gas producido en el decantador secundario lo que dificultaría la sedimentación del fango. Desde el fondo del decantador secundario el fango es recirculado al primer reactor anóxico.

Canales de oxidación. El sistema de tratamiento consiste en un canal circular u ovalado por el cual circula el licor mezcla impulsado por aireadores mecánicos de eje horizontal (Figura 1.9). La zona aerobia comienza inmediatamente aguas abajo de cada aireador y se extiende hasta que la concentración de oxígeno disuelto

disminuye a cero. El tramo comprendido entre ese punto y el siguiente aireador constituye la zona anóxica. En el inicio de esta zona es donde se establece la entrada del agua residual, permitiendo que parte de la materia orgánica sea degradada en el proceso de desnitrificación.

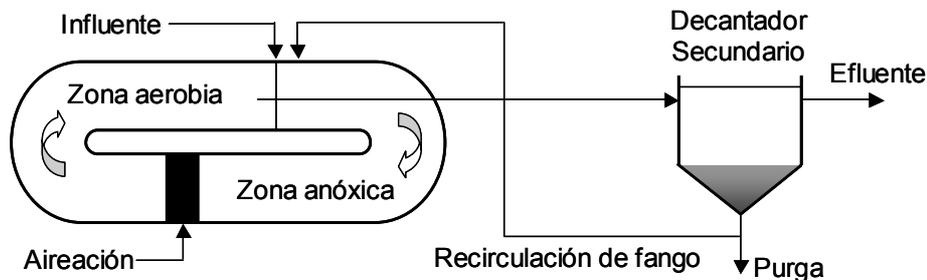


Figura 1.9. Esquema de un canal de oxidación con un único aireador.

Esquema A/O. Este sistema de tratamiento permite la eliminación conjunta de materia orgánica y de fósforo (Figura 1.10). El agua residual entra en el reactor anaerobio donde se produce la toma de ácidos grasos de cadena corta y la suelta de fósforo. Seguidamente el licor mezcla pasa al reactor aerobio en el que tiene lugar el almacenamiento intracelular de fósforo en forma de polifosfatos. Los fangos decantados tienen un alto contenido en fósforo. Parte de estos fangos son eliminados del sistema (purga) y el resto son recirculados al reactor anaerobio.

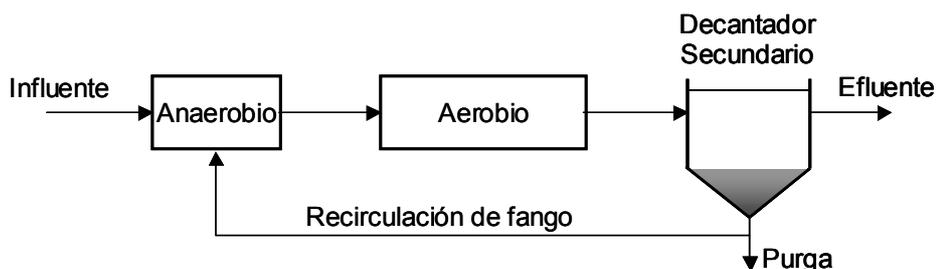


Figura 1.10. Esquema A/O.

Esquema A²/O. Este proceso es una modificación del A/O que incluye un reactor anóxico entre los reactores anaerobio y aerobio (Figura 1.11). Este reactor anóxico permite la oxidación de la materia orgánica utilizando como aceptores de electrones los nitratos procedentes del reactor aerobio a través de la recirculación interna. De esta manera, se consigue eliminar nitrógeno y además se evita que se recirculen

nitratos en exceso desde el decantador secundario al reactor anaerobio lo que podría perjudicar la eficacia el proceso de eliminación de fósforo.

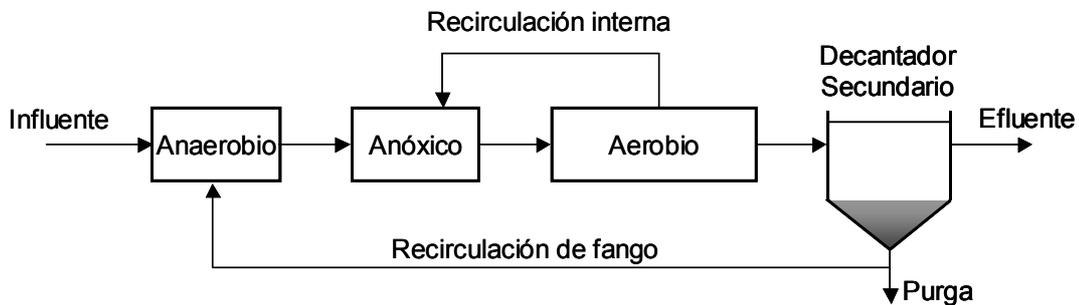


Figura 1.11. Esquema A²/O.

Esquema UCT. Este esquema fue desarrollado para reducir la entrada de nitratos en el reactor anaerobio y de esta forma mejorar el proceso de eliminación biológica de fósforo. Se trata de un esquema parecido al A²/O, aunque con la diferencia de que el fango del decantador se recircula al reactor anóxico en lugar de al reactor anaerobio (Figura 1.12). Al igual que en el esquema A²/O, también hay una corriente de recirculación interna entre el reactor aerobio y anóxico, pero además se establece otra entre el anóxico y el anaerobio.

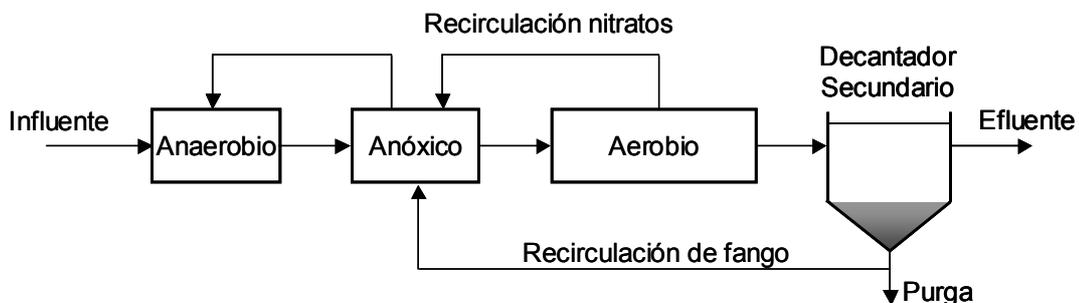


Figura 1.12. Esquema UCT.

Con el proceso UCT se persigue tener unas condiciones anaerobias más estrictas en el primer reactor. Para conseguirlo se debe mantener una concentración baja de nitratos en la zona anóxica. Sobre la base de este esquema, es posible realizar separadamente la desnitrificación de los fangos recirculados y del licor mezcla procedente del reactor aerobio dividiendo el reactor anóxico en dos. Esto da lugar al esquema UCT modificado (Figura 1.13).

En el esquema UCT modificado, los fangos del decantador secundario son recirculados al primer reactor anóxico, y desde él se recircula el licor mezcla al reactor anaerobio. Los nitratos generados en la zona aerobia son recirculados al segundo reactor anóxico lo que evita que puedan ser transferidos al reactor anaerobio.

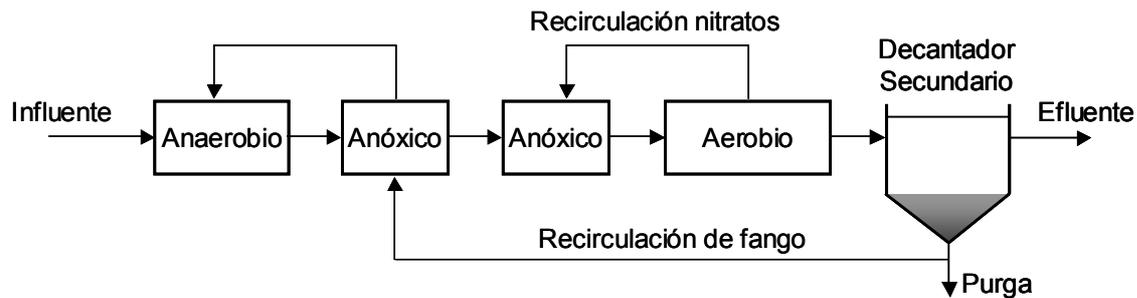


Figura 1.13. Esquema UCT modificado.

Reactor discontinuo secuencial (SBR). Como su nombre indica, es un sistema de tratamiento de fangos activados en el cual el agua residual es introducida de forma discontinua en el reactor, donde es tratada y posteriormente descargada. Mientras que en el sistema convencional estas etapas tienen lugar en reactores separados y se producen de manera simultánea, en los SBRs se producen en un mismo reactor pero secuencialmente en el tiempo.

La operación de la mayor parte de los SBRs que se emplean en la actualidad consta de cinco etapas que se realizan de manera secuencial:

- Llenado. Durante esta etapa el agua residual a tratar entra en el reactor. El grado de interacción que se produce entre la biomasa que está en el reactor del ciclo anterior y el agua residual que entra depende del estado de los sistemas de aireación y de agitación. Esto permite hacer la siguiente distinción:
 - Llenado estático. En este caso tanto el sistema de aireación como de mezclado permanecen inactivos. Esto permite la acumulación de sustrato en el reactor ya que durante el llenado no tienen lugar procesos de biodegradación.

- Llenado con mezcla. Durante la etapa de llenado, el reactor no es aireado pero su contenido se mantiene mezclado. Esto posibilita que en el interior del reactor tengan lugar reacciones anóxicas o anaerobias, dependiendo de la presencia o ausencia de nitratos.
- Llenado con aireación. En este caso el contenido del reactor se mantiene mezclado y aireado. Esto permite que en el reactor tengan lugar reacciones aerobias.

La etapa de llenado determina las características hidráulicas del reactor. Cuando se realiza un llenado estático o cuando se permiten las transformaciones bioquímicas pero la duración de la fase de llenado es corta, la biomasa estará inicialmente expuesta a altas concentraciones de materia orgánica y demás compuestos presentes en el agua residual. Debido a la actividad biológica, con el paso del tiempo estas concentraciones irán disminuyendo. Por otra parte, cuando la etapa de llenado no sea estática y su duración sea larga, la biomasa estará expuesta a concentraciones bajas y relativamente constantes. En el primer caso el comportamiento es semejante al de un sistema continuo de flujo en pistón, o bien a varios reactores de mezcla completa dispuestos en serie, mientras que en el segundo caso el comportamiento es semejante al de un reactor de mezcla completa.

Esta etapa es seguida por la etapa de reacción, en la cual la biomasa actúa sobre la materia orgánica y demás compuestos presentes en el agua residual. Sin embargo, tal y como se acaba de exponer, durante el llenado del reactor también se pueden permitir las transformaciones bioquímicas, por lo que se podría considerar como una etapa de “llenado y reacción”, en la que la reacción continúa cuando el llenado ha finalizado. De hecho, en algunos casos, la operación del SBR no incluye una etapa de reacción independiente.

- Reacción. En esta etapa tienen lugar las transformaciones bioquímicas llevadas a cabo por los microorganismos. Al igual que durante el llenado, en esta fase también se pueden imponer distintas condiciones ambientales en el reactor, lo que permite que esta etapa se pueda dividir en varias subetapas:
 - Mezcla. El contenido del reactor se mantiene mezclado pero no está aireado. Se pueden producir, por tanto, reacciones anóxicas o anaerobias, según el aceptor de electrones presente en el medio.
 - Mezcla aerobia. En este caso el contenido del reactor se mantiene mezclado y aireado, lo que permite que tengan lugar reacciones aerobias.
- Sedimentación. En esta etapa se detiene tanto la aireación como la agitación, para permitir la separación de los sólidos y poder obtener un efluente clarificado. Al estar el licor mezcla en reposo, la eficiencia de la decantación suele ser mayor que en los sistemas en continuo.
- Vaciado. En esta etapa el agua clarificada se retira del reactor, de manera que éste queda disponible para recibir un nuevo volumen de agua residual a tratar.
- Fase de inactividad. Es el tiempo que transcurre desde que acaba la fase de descarga hasta que comienza un nuevo ciclo. En sistemas con varios SBRs en paralelo, esta fase permite que un reactor acabe su etapa de llenado antes de iniciar el llenado de otro reactor. No es una fase necesaria, y en caso de que se pueda conseguir un caudal de entrada constante, por ejemplo mediante la instalación de un tanque de regularización, se puede suprimir.

Como se ha comentado previamente, el tiempo de retención celular es un parámetro importante del proceso y se controla mediante la purga de fangos del sistema. En un SBR, la purga se puede realizar al final de la etapa de sedimentación, al igual que en los procesos en continuo, o bien al final de la etapa de reacción.

La principal ventaja que presentan los sistemas SBR es su flexibilidad. Esta característica se deriva de la posibilidad de controlar de forma sencilla el tiempo que se dedica a cada una de las etapas del proceso. En cada etapa se imponen unas condiciones ambientales determinadas, y para modificar la duración de cada una de ellas tan sólo es necesario actuar sobre los controladores que conectan y desconectan las bombas y soplantes. Por el contrario, en los sistemas de tratamiento en continuo, para un caudal de agua residual de entrada dado, la única forma de modificar el tiempo destinado a cada etapa, en este caso cada etapa tiene lugar en un reactor distinto, es alterando el tamaño de los reactores. Esto puede resultar más o menos complicado dependiendo de cada caso en particular.

Como ejemplo de la flexibilidad de los sistemas SBR se presenta en la Figura 1.14 dos secuencias de operación con las que se podrían conseguir distintos objetivos de depuración utilizando un único reactor. La primera secuencia permite la eliminación biológica de fósforo, y la segunda la eliminación conjunta de nitrógeno y fósforo.

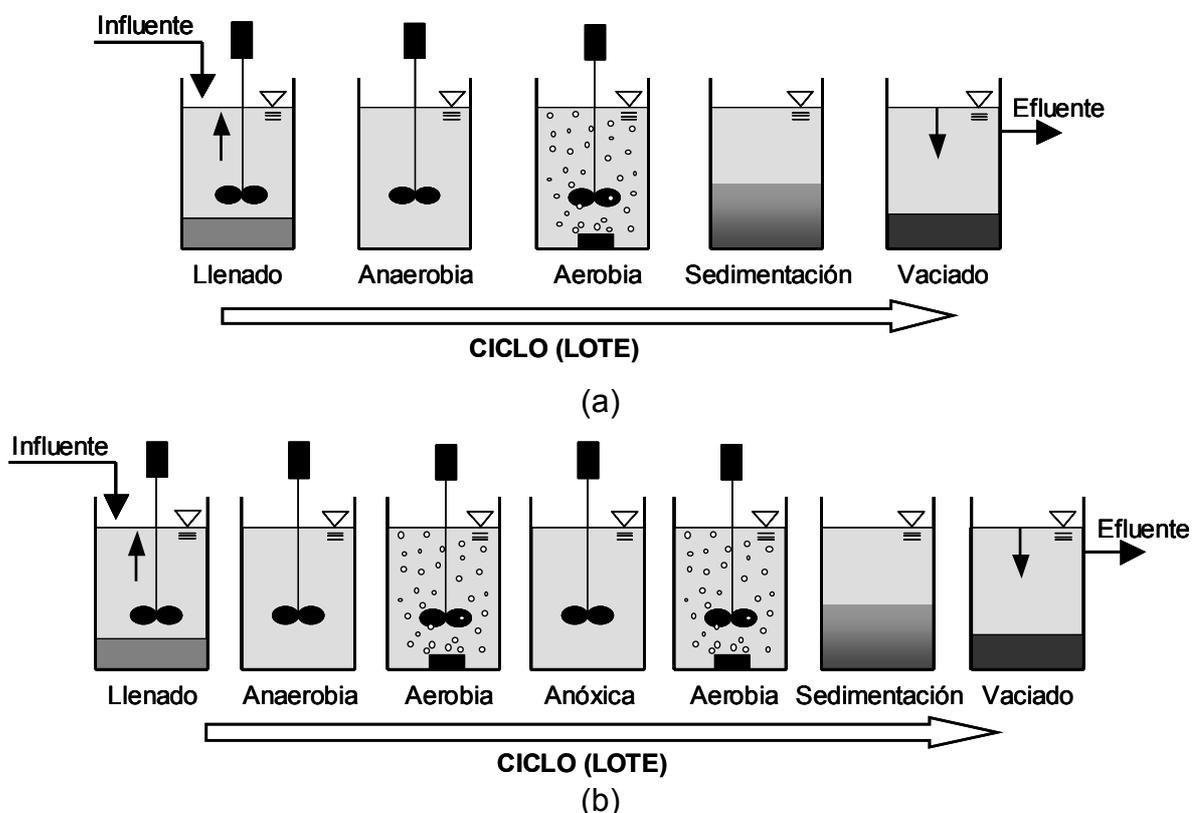


Figura 1.14. Reactor discontinuo secuencial operado para la eliminación biológica de nutrientes. (a) Secuencia de etapas para la eliminación de fósforo. (b) Secuencia de etapas para la eliminación conjunta de nitrógeno y fósforo.

1.7 TECNOLOGÍAS ICA EN LOS SISTEMAS DE TRATAMIENTO DE AGUAS RESIDUALES

El objetivo de las tecnologías de instrumentación, control y automatización (ICA) es mejorar las condiciones de operación de los procesos a los que se aplican. En el caso particular de una EDAR, es permitir una operación eficiente de la misma en lo que respecta al cumplimiento de los requisitos de vertido establecidos, minimizando los costes de operación.

Cada tecnología se corresponde con un área de conocimiento completamente diferenciada, aunque el hecho de perseguir un objetivo común hace que estén estrechamente ligadas. Así por ejemplo, el tipo de estructura de control que se puede aplicar sobre un proceso estará determinado por la instrumentación disponible.

El control de procesos tiene por misión el diseño de algoritmos operativos que consigan que el proceso sobre el que se aplican se comporte según unas pautas prefijadas. Estos algoritmos necesitan conocer el estado del proceso a controlar, esta información es obtenida a través de los instrumentos de medida. La tecnología que se encarga de investigar y desarrollar estos equipos de medición es la instrumentación. Finalmente, la automatización es la tecnología que se encarga de diseñar los equipamientos necesarios para que las mediciones y las acciones de control sean realizadas por medio de sistemas electrónicos e informáticos, reduciendo de esta manera la intervención humana.

En los sistemas de tratamiento de aguas residuales, las tecnologías ICA han experimentado un notable auge con el tiempo convirtiéndose en indispensables en este campo (Olsson y Newel, 1999). Los principales factores que han hecho esto posible son:

- Tecnología de instrumentación. En las últimas décadas ha tenido un gran desarrollo, lo que ha hecho posible la construcción de sensores complejos como los que permiten la determinación en tiempo real de las concentraciones de nutrientes.

- Actuadores. Estos elementos también han sido mejorados, y en la actualidad permiten modificar la velocidad de giro de bombas y compresores, consiguiendo de esta manera un mejor control sobre los procesos en EDARs.
- Adquisición de datos. La recogida y almacenamiento de gran cantidad de datos del proceso ya no supone un obstáculo. Tanto los sistemas de comunicaciones como los informáticos han experimentado un notable abaratamiento. En la actualidad existen diversos programas informáticos para la adquisición de datos y supervisión del proceso.
- Modelos dinámicos avanzados. Los modelos de los procesos biológicos se han constituido como una potente herramienta para entender los procesos, su dinámica y sus interacciones, permitiendo además probar distintas estrategias operacionales previamente a su implementación en planta real.
- Operadores e ingenieros de proceso. Su formación en instrumentación, electrónica, informática y tecnologías de control, hace que en la actualidad estén más capacitados.

El primer avance de gran repercusión para la aplicación del control de procesos en EDARs tuvo lugar cuando la tecnología permitió la construcción de oxímetros robustos y fiables que podían ser utilizados en EDARs a escala real (década de los años 1970). A partir de ese momento y con el paso del tiempo se han podido desarrollar e implementar distintas estrategias para el control de la aireación en los procesos aerobios. A continuación se comentan brevemente en orden creciente de sofisticación algunas de las estrategias que se pueden emplear:

- Suministrar un flujo constante de aireación. Esta era la forma habitual de proceder antes de que fuera una práctica común el empleo de oxímetros. No necesita un oxímetro puesto que siempre se aporta la misma cantidad de oxígeno al reactor aerobio.

- Modificación del flujo de aireación basado en la experiencia. En este caso tampoco es necesario el uso de un oxímetro. Consiste en utilizar, por ejemplo, un flujo de aireación más elevado durante el día, y más reducido por la noche.
- Control de la aireación basado en el nivel de oxígeno en el reactor. Cuando sólo se dispone de un único oxímetro instalado en el reactor, es posible en base a sus lecturas ajustar la cantidad de oxígeno que se suministra. Para conseguirlo, se actúa sobre los compresores, bien modificando su velocidad de giro o bien conectando más o menos compresores.
- Control de la aireación basado en el perfil de oxígeno. Cuando se descubrió que debido a la longitud de los reactores, las necesidades de oxígeno no son constantes sino que disminuyen a lo largo del reactor, se consideró adecuado subdividir el reactor en varias zonas y controlar cada una de ellas de forma independiente. En este caso puede haber varios oxímetros instalados a lo largo del reactor, y las lecturas de cada uno de ellos sirve para regular la cantidad de oxígeno suministrada a la zona del reactor en que éste está ubicado.
- Control de la aireación con modificación dinámica del *set-point*. La aparición de instrumentos para la determinación de nutrientes como el amonio, ha permitido que además de tener en cuenta que las necesidades de oxígeno disminuyen a lo largo del reactor, también sea posible modificar el nivel de oxígeno que se quiere mantener en el reactor en función de los procesos biológicos que tienen lugar. En este caso, además de controlar individualmente el oxígeno suministrado a cada zona del reactor, también se modifica el nivel de oxígeno que se quiere mantener en cada zona, utilizando para ello las medidas de la concentración de amonio en el reactor.

A principios de los años 1990 se consideraba que los sensores representaban el punto más débil para la implementación del control en tiempo real de procesos en EDARs (Harremoës *et al.*, 1993). Sin embargo, durante la última década el funcionamiento y la fiabilidad de muchos sensores, como los empleados en la

determinación en tiempo real de la concentración de nutrientes, ha mejorado notablemente y pueden ser empleados en la actualidad en diversas estrategias de control (Jeppsson *et al.*, 2002). Estos autores indican que probablemente la principal barrera que existe para que estas nuevas estrategias de control sean adoptadas de forma generalizada es el hecho de que las EDARs no están diseñadas para su control en tiempo real.

En una EDAR los principales costes de operación están asociados al personal, a la producción de fango y al consumo de energía y de reactivos químicos. La utilización de las tecnologías ICA puede contribuir a reducir todos estos costes. Sin embargo, su verdadero potencial aparece cuando se necesita incrementar la capacidad de tratamiento de una EDAR, ya que frente a la ampliación de los volúmenes de reactor u otro tipo de soluciones estructurales, el empleo de las tecnologías ICA se presenta como una alternativa mucho más económica.

En la actualidad, la instrumentación que se utiliza en EDARs es muy variada e incluye desde la instrumentación típica en la industria química hasta la específica de los procesos de tratamiento de aguas residuales. Los instrumentos de medida pueden ser clasificados en función de dos factores fundamentales:

- La **rapidez** con que obtienen resultados. Atendiendo a este criterio se puede distinguir entre instrumentos *on-line* e instrumentos *off-line*. En el primer caso, los valores de la variable medida se obtienen de forma rápida y continua, pudiendo ser utilizados en control en tiempo real. En el segundo caso, la rapidez en los cambios que presenta la variable es superior al tiempo que tarda el instrumento en proporcionar su valor.
- El **tipo de parámetro** que monitorizan. En función de este criterio se pueden agrupar los equipos de medida en varias categorías. Dentro de cada una de ellas se indican los equipos empleados más habitualmente:

- Instrumentación para la determinación de parámetros físicos.
 - *Instrumentos para la medición del caudal de líquidos y gases*. Permiten el seguimiento de los caudales de agua, fango y gas en las distintas operaciones unitarias de las que consta una EDAR.
 - *Instrumentos para medir el nivel de líquidos*. Son utilizados principalmente para monitorizar el nivel del agua en los distintos reactores. Para detectar el nivel se suelen utilizar sensores que miden la presión hidrostática del agua y también los basados en ultrasonidos.
 - *Instrumentos para la medición de la temperatura*. Son empleados para el seguimiento de la temperatura en los procesos biológicos, y tienen especial importancia para el control de la temperatura en los digestores anaerobios.
 - *Instrumentos para la medición de la presión*. Son utilizados en el control de la aireación, y como alarma en los procesos de digestión anaerobia.
 - *Instrumentos para la medición de la turbidez y los sólidos suspendidos*. Se utilizan para determinar de forma rápida la concentración de sólidos en los reactores y en el efluente de los tratamientos terciarios.
 - *Instrumentos para la medición de la altura de la capa de fangos*. Permiten controlar el manto del fango en los decantadores y, de esta manera, optimizar su funcionamiento. En reactores SBR pueden ser empleados para determinar cuándo puede comenzar la etapa de vaciado, así como para ajustar la velocidad de vaciado a las características de sedimentación del fango.

- Instrumentación para la determinación de parámetros químicos. Son muchas las técnicas que se emplean en laboratorio y, por tanto, se podrían enumerar gran cantidad de instrumentos, como por ejemplo:
 - *Instrumentos de medición por espectrometría de UV-visible*.
 - *Instrumentos de medición por espectrometría de absorción atómica*.
 - *Instrumentos de medición por cromatografía de gases y líquidos*.
 - *Instrumentos de medición por técnicas de fluorescencia*.
 - *Instrumentos para determinaciones polarográficas*.
 - *Otros instrumentos*.

De todos estos instrumentos sólo algunos son utilizados en la monitorización y control automatizados en EDARs, entre ellos:

- *Instrumentos electroanalíticos.* Los más utilizados son los que permiten medir el oxígeno disuelto, el pH, el potencial redox y la conductividad. Estos instrumentos son ampliamente utilizados puesto que son robustos, fiables, necesitan poco mantenimiento y su coste es bajo. Las mediciones de oxígeno son utilizadas para la optimización de la aireación. El pH y el potencial redox se han utilizado para la optimización de la eliminación biológica de nitrógeno (Yu *et al.*, 1997, 1998; Katsogiannis *et al.*, 1999;...). La conductividad se ha empleado para la monitorización de cambios en el agua residual a tratar (Teichgräber, 1993), y ha sido sugerida su utilización para la optimización del proceso de eliminación biológica de fósforo (Serralta *et al.*, 2004).
- *Instrumentos espectrofotométricos automatizados.* Estos instrumentos permiten la determinación de nutrientes como NO_3^- , NH_4^+ , PO_4^{3-} , y también otros compuestos de interés en los procesos de tratamiento de aguas residuales como la demanda química de oxígeno (DQO) y el cloro libre. También son empleados en la determinación de numerosos iones metálicos (Fe^{3+} , Cr^{6+} , ...).
- *Instrumentos con electrodos ión-selectivo.* Permiten la monitorización de nutrientes NO_3^- y NH_4^+ así como la concentración de compuestos específicos: Cl^- , Na^+ , F^- , CN^- , S^{2-} .
- *Instrumentación especializada.* Dentro de esta categoría se incluyen las técnicas respirométricas y los biosensores. Las primeras se basan en la medición e interpretación del consumo de oxígeno por parte del fango activado. Constituyen una herramienta utilizada habitualmente para la caracterización del agua residual y de los parámetros cinéticos de los procesos biológicos. Los biosensores están constituidos por biomasa y uno o varios electrodos. Algunos emplean membranas selectivas para permitir únicamente el paso de un sustrato específico que es consumido por la biomasa. Entre los biosensores que se han desarrollado se encuentran los

que permiten monitorizar la carga orgánica (demanda bioquímica de oxígeno a corto plazo: DBO_{st}) y la toxicidad potencial de las aguas residuales (Vanrolleghem *et al.*, 1994), el consumo de NO_x (Larsen *et al.*, 1997, 2000), de ácidos grasos volátiles (Rozzi *et al.*, 1997) y la actividad metanogénica (Rozzi *et al.*, 2001).

De lo anteriormente expuesto se puede deducir que la instrumentación en el campo de los tratamientos de aguas residuales ha experimentado un notable crecimiento. Esto ha provocado que durante las últimas décadas la cantidad de datos, procedentes de sensores y otras fuentes, haya crecido casi exponencialmente. La disponibilidad de un mayor número de sensores junto con unos límites de vertido cada vez más restrictivos, hará que en un futuro los operadores necesiten más apoyo de los sistemas de control (Olsson G., 2002). Esto conduce a un incremento de la demanda en técnicas fiables de análisis de fallos, de cribado de datos, así como de condensación de la información.

En la actualidad la industria de las aguas residuales en Europa se encuentra en una fase de transición en la que una serie de factores tienden a promover el uso de las tecnologías ICA (Jeppsson *et al.*, 2002). Estos factores están relacionados con:

- Imposición de límites de vertido más restrictivos.
- Incentivos económicos.
- Reducción del consumo de energía y/o incremento de la producción de energía.
- Mayor complejidad de las EDARs.
- Nuevos conceptos de tratamiento.
- Nuevas soluciones técnicas más económicas (ordenadores, comunicaciones,...)

Una de las conclusiones del estudio realizado por estos autores es que las tecnologías ICA tienen un papel importante que jugar en el futuro de las EDARs.

1.8 ANÁLISIS, DIAGNÓSTICO DE FALLOS Y MONITORIZACIÓN DE PROCESOS

1.8.1 Problemática de la situación actual

La revolución tecnológica que tuvo lugar a finales del siglo XX, y que fue debida al gran auge experimentado tanto en la electrónica como en las comunicaciones, provocó un abaratamiento de equipos informáticos así como de sensores y otros instrumentos de medición. Gracias a esta revolución, la mayor parte de los procesos industriales se encuentran altamente instrumentalizados, lo que permite disponer en tiempo real de los valores de multitud de variables (caudales, temperaturas, ...).

A menudo entre las variables registradas se distinguen dos grupos. Por un lado están las variables de la calidad del producto fabricado, las cuales constituyen unos indicadores clave del funcionamiento del proceso. Para algunas de estas variables se han desarrollado sensores que permiten conocer su valor en tiempo real, sin embargo, éstos suelen extremadamente costosos y no es habitual disponer de ellos. Por tanto, para obtener las medidas de calidad se suele recurrir a su determinación analítica en laboratorio. Esto ocasiona varios inconvenientes, como el no poder disponer de su valor en tiempo real y el que se determine un reducido número de variables de calidad con una frecuencia de muestreo baja. Por otro lado, el segundo grupo de variables estaría constituido por las variables de proceso. Estas variables son registradas en tiempo real mediante sensores robustos y de bajo coste. Por este motivo, habitualmente se registra un elevado número de variables de proceso con una frecuencia de muestreo muy superior a la de las variables de calidad.

En consecuencia, la base de datos de un determinado proceso estará constituida en su mayor parte por la evolución a lo largo del tiempo de las variables de proceso. El análisis de estas variables es fundamental puesto que pueden contener mucha información relevante no sólo sobre funcionamiento del proceso sino también sobre la calidad del producto final. Son muchas las razones que justifican la importancia de analizar las variables de proceso, entre ellas se podrían destacar las siguientes:

- *Se dispone de un **mayor número de variables** de proceso que de calidad:* esto es debido a que la frecuencia de muestreo en las variables de proceso es muy superior a la de las variables de calidad. Mientras que para las primeras puede ser de segundos (e incluso de milisegundos en algunos casos), para las variables de calidad puede ser de minutos e incluso días. Esto ocasiona que las variables de proceso estén disponibles prácticamente de manera instantánea, mientras que las de calidad tengan una cierta demora.
- *La **precisión** de las variables de proceso es en muchas ocasiones mayor que la de las variables de calidad:* ya que estas últimas requieren frecuentemente la realización de determinaciones analíticas complejas.
- *El considerar las variables de proceso **facilitará el diagnóstico** de problemas:* ya que permiten buscar las causas del problema cuando se ha detectado alguna anomalía en las variables de calidad. Los datos del proceso contienen información de cómo éste ha transcurrido, y en muchas ocasiones, también de aquellas alteraciones que se han producido y que posteriormente han repercutido sobre la calidad del producto final. Es decir, es habitual que las anomalías o fallos del proceso no sólo afecten a las variables de calidad, sino que también dejen “huella” en las variables de proceso.

De lo anterior es inmediato concluir que no se debe desaprovechar la gran cantidad de información contenida en las variables de proceso. Sin embargo, al tratar de aprovecharla aparecen una serie de dificultades que se deben tener en cuenta a la hora de elegir las técnicas para procesar dicha información. Las principales dificultades son (Ferrer, 2003):

- **Dimensionalidad:** dado que la instrumentación de los procesos permite registrar muchas variables en tiempo real, el tamaño de la matriz de datos generada es muy elevado. Ante esta abundancia de datos, los operadores del proceso, en base a su experiencia, acaban utilizando para la toma de decisiones sólo un reducido número de variables que consideran clave.

- **Colinealidad:** aunque se disponga de cientos de variables de proceso, realmente en el sistema no están teniendo lugar cientos de sucesos independientes. Es decir, sólo unos pocos fenómenos subyacentes son los que están afectando al proceso, lo que significa que la gran cantidad de variables medidas estarán reflejando los cambios de esos pocos fenómenos y, por lo tanto, existirá una fuerte correlación entre las variables.
- **Ruido:** tanto las variables de proceso como las de calidad presentan ruido en su determinación (error experimental). Dado que en la mayoría de ocasiones lo que se persigue en el proceso es que éste se desvíe lo menos posible de su trayectoria objetivo, la relación señal-ruido de cada variable puede ser baja, lo que puede ocasionar que las variaciones de la señal queden enmascaradas por el ruido.
- **Datos faltantes:** debido a fallos en los sensores y en la transmisión de datos, las bases de datos contienen habitualmente datos faltantes. Aunque puede suceder que no se haya registrado ninguna variable de una determinada observación, lo más habitual será encontrar observaciones incompletas, es decir, en las que sólo una parte de las variables no hayan sido registradas.

1.8.2 Clasificación de los modelos

La modelación de un proceso consiste en emplear un conjunto de ecuaciones matemáticas para describir su comportamiento. La construcción de un modelo del proceso se puede realizar por diferentes motivos, entre los que se pueden destacar los siguientes: para diseñar, para optimizar, para monitorizar, para controlar y para diagnosticar el proceso.

La elección del modelo a utilizar no es unívoca; existen muchas opciones para describir el proceso. Esta elección no sólo depende del propósito del modelo sino también del tipo de proceso a modelar. Los modelos se pueden clasificar del siguiente modo:

- **Modelos mecanicistas, deterministas, fundamentales o de caja blanca (*white box*):** son modelos que están basados en los principios fundamentales (físicos, químicos o biológicos) que rigen el comportamiento del sistema, como por ejemplo la cinética del proceso o los balances de energía. Esta clase de modelos precisa de todas las ecuaciones que describen de forma clara el comportamiento del sistema. Esto implica el conocimiento de un elevado número de parámetros complicando la tarea del calibrado y su tratamiento matemático. Disponer de un buen modelo mecanicista permite conocer con exactitud la evolución del proceso una vez conocido con precisión los parámetros que lo describen en unas condiciones de operación determinadas. Los modelos mecanicistas tienen buena capacidad de extrapolación, es decir, serán útiles para la mayoría de las situaciones que se puedan presentar, ya sea en el sistema concreto donde se estudia como en cualquier otro. Una de las desventajas de este tipo de modelos es que frecuentemente presentan ajustes deficientes. Además, la tarea de conseguir un modelo teórico fiable del proceso suele ser complicada puesto que se requiere un conocimiento profundo del proceso (lo cual no es frecuente en la práctica y suele suponer mucho esfuerzo) y resulta en ocasiones inviable. Estos modelos se emplean principalmente en tareas de diseño y optimización.
- **Modelos empíricos o de caja negra (*black box*):** relacionan las variables de entrada al proceso con las variables de salida sin que esta relación suponga la descripción de un proceso con un significado físico. Los modelos empíricos ajustan los datos disponibles del proceso a una función que reproduce los resultados del sistema real obtenidos experimentalmente. Se necesita un número de parámetros mucho menor que el necesario para la formulación de los modelos mecanicistas haciendo más sencillo la etapa de calibrado y tratamiento matemático. El uso de los modelos empíricos está limitada a la región operativa utilizada para su construcción. Estos modelos tienen mala capacidad de extrapolación, es decir, presentan problemas a la hora de reproducir situaciones diferentes a las utilizadas en la calibración, motivo por

el que no se utilizan para el diseño de procesos. Dentro de este grupo estarían incluidos los modelos de regresión y las redes neuronales artificiales.

- **Modelos semi-empíricos, semi-deterministas, híbridos o de caja gris (*grey box*):** son modelos mecanicistas sencillos en los que los parámetros que intervienen incorporan el efecto simultáneo de diversos fenómenos de difícil cuantificación individual. Son modelos cuyas características los clasifican en un estado intermedio entre los modelos mecanicistas y los empíricos, acercándose a un grupo u otro según el grado de conocimiento del proceso. Estos modelos híbridos se construyen normalmente ajustando en primer lugar un modelo mecanicista a los datos y después empleando un modelo empírico para ajustar la variación sistemática que no había sido explicada por el modelo mecanicista. Este enfoque permite combinar modelos de características completamente distintas; por ejemplo, se podría utilizar un modelo mecanicista lineal y un modelo empírico no lineal. Sin embargo, debido a la naturaleza secuencial de la modelación realizada, el ajuste global puede no ser el óptimo, por lo que en ocasiones hay que recurrir a procedimientos iterativos.

1.8.3 Control estadístico de procesos clásico

El control estadístico de procesos (SPC, *Statistical Process Control*) parte de la idea de que todo proceso experimenta variaciones en su funcionamiento. Estas variaciones pueden afectar en mayor o menor medida a la calidad del producto final. La variabilidad que presenta un proceso en su comportamiento puede ser debida a dos causas:

- **Causas comunes:** son aquellas que siempre están presentes en el proceso, son innatas al mismo, y son consecuencia de cómo se ha diseñado el proceso y de cómo se opera normalmente. Cuando el proceso sólo está afectado por causas comunes de variabilidad, se dice que está bajo control estadístico. En este caso se caracteriza por tener una pauta de variabilidad estable, lo que hace que éste sea predecible.

- **Causas especiales:** son aquellas que aparecen de manera esporádica, y que son consecuencia de anomalías o fallos en el proceso, originando que éste tenga un comportamiento diferente al habitual. Cuando el proceso está afectado por causas especiales de variabilidad, se dice que está fuera de control estadístico. En este caso, el proceso presenta una pauta de variabilidad inestable y, por tanto, es impredecible.

El objetivo que persigue el control estadístico de procesos es el establecimiento de un sistema permanente e inteligente de monitorización a lo largo del tiempo que permita detectar las causas especiales de variabilidad que afecten al proceso y que ayude a identificar el origen de las mismas. Su fundamento es la toma periódica de datos del proceso y la representación gráfica de la evolución de ciertos estadísticos que facilitan la detección visual de señales de aviso que indican anomalías. La identificación de las causas que provocan comportamientos anómalos del proceso (causas especiales) es el primer paso en la toma de medidas pertinentes para su eliminación, lo que permite una mejora continua del mismo.

Por tanto, los gráficos de control constituyen la herramienta básica del SPC, y según el número de variables que permiten monitorizar simultáneamente se denominan univariantes (una sola variable) o multivariantes (varias variables).

El primer gráfico de control univariante fue el desarrollado por Shewhart (1931). Su sencillez junto con los buenos resultados contribuyó a que fuera uno de los más usados en décadas pasadas. Es especialmente adecuado en ámbitos en que se registren pocas variables, que éstas sean independientes (no correlacionadas), se distribuyan normalmente y que además no presenten autocorrelación (independencia de los datos). Otros gráficos de control univariantes empleados en la monitorización son los gráficos de sumas acumuladas (CUSUM, *Cumulative Sum*) (Page, 1954), o los gráficos de medias móviles ponderadas exponencialmente (EWMA, *Exponentially Weighted Moving Average*) (Hunter, 1986).

En el caso de tener que monitorizar varias variables que no estén correlacionadas entre sí, para cada una de ellas se puede construir un gráfico de control univariante. Sin embargo, este procedimiento no es adecuado cuando las variables están correlacionadas, ya que el proceso puede estar fuera de control sin que los gráficos univariantes lo detecten. Para dichas situaciones, se desarrollaron los gráficos de control basados en el estadístico T^2 -Hotelling, que permiten aprovechar la información contenida en la estructura de correlación de las variables. Dichos gráficos indicarán que hay anomalías en el proceso tanto cuando alguna de las variables monitorizadas adopte valores fuera de lo habitual, como cuando se modifique la estructura de correlación de las variables. Aunque son una herramienta de gran utilidad, su utilización requiere disponer de más observaciones que variables, que la matriz de varianzas-covarianzas no esté mal condicionada por la colinealidad entre las variables, y que no existan datos faltantes. Estos supuestos no son habituales en los procesos industriales actuales, donde las características comunes que presentan las bases de datos disponibles son: grandes dimensiones, colinealidad, ruido y datos faltantes.

En estos enfoques clásicos, las variables de proceso no se tienen en cuenta para la monitorización, y sólo son empleadas para tratar de identificar las causas especiales de variabilidad cuando se ha detectado una salida de control en los gráficos de control de las variables de calidad. Además, tampoco tienen en cuenta que los pocos parámetros de calidad que se miden no definen completamente la calidad del producto, es decir, que las anomalías que ocurren en el proceso pueden afectar a la calidad del producto aunque las variables de calidad medidas no lo reflejen. Se puede concluir, por tanto, que la limitación más importante del SPC clásico para la monitorización de los modernos procesos industriales es el desaprovechamiento de la mayor parte de la información que se registra.

1.8.4 Técnicas estadísticas multivariantes de proyección sobre estructuras latentes

En el ámbito de los procesos actuales, los métodos de proyección sobre estructuras latentes como el análisis de componentes principales (PCA, *Principal Component Analysis*) (Jackson, 1991) y la regresión en mínimos cuadrados parciales (PLS, *Partial Least Squares*) (Geladi y Kowalski, 1986; Wold *et al.*, 1987) presentan las ventajas de ser capaces de abordar aquellas situaciones en las que aparecen datos faltantes, así como matrices de grandes dimensiones y mal condicionadas.

Estas técnicas permiten sintetizar toda la información contenida en una base de datos en unas pocas variables no observables (variables latentes) que explican gran parte de la variabilidad de las variables originales (variables observables que han sido medidas), y de las relaciones que existen entre las mismas. En el nuevo subespacio creado de menor dimensión (definido por las variables latentes), es posible utilizar las técnicas clásicas de control estadístico, mediante la aplicación de los gráficos de control a las variables latentes. La monitorización de estas variables latentes permite de manera indirecta controlar la gran cantidad de variables que se miden en el proceso. Además es posible aprovechar la información que se recoge durante el proceso para predecir las variables de calidad, mediante la construcción de modelos predictivos, también denominados *soft sensors*.

1.8.4.1 Análisis de componentes principales (PCA)

El PCA fue formulado por primera vez en estadística por Pearson (1901), quien definió este tipo de análisis como “la búsqueda de líneas y planos de mejor ajuste a un sistema de puntos en el espacio”. El propósito de esta técnica es el descomponer una matriz de datos para descubrir los fenómenos subyacentes. Por tanto, una suposición básica del PCA es el hecho de que las direcciones de mayor varianza están en cierta medida relacionadas con dichos fenómenos subyacentes. Esta técnica permite analizar la estructura de correlación de la variables y también identificar patrones en los datos, como tendencias, agrupaciones y datos anómalos.

Al realizar un PCA sobre una matriz de datos \mathbf{X} , se está realizando la siguiente modelación:

$$\mathbf{X} = \sum_{a=1}^A \mathbf{t}_a \mathbf{p}_a^T + \mathbf{E} = \hat{\mathbf{X}} + \mathbf{E} \quad \text{Ec. 1.1}$$

En la expresión anterior \mathbf{p}_a son los vectores de pesos que definen las direcciones principales de máxima varianza en el espacio \mathbf{X} . Estas direcciones determinan un subespacio de menor dimensión (A) que el espacio original. Esta dimensión se escoge de manera que no exista información importante de \mathbf{X} en \mathbf{E} que es la matriz de residuos, y que, por tanto, representa el ruido. $\hat{\mathbf{X}}$ es la estimación de \mathbf{X} a partir del modelo con A componentes. En la Figura 1.15 se representan gráficamente las matrices del modelo PCA.

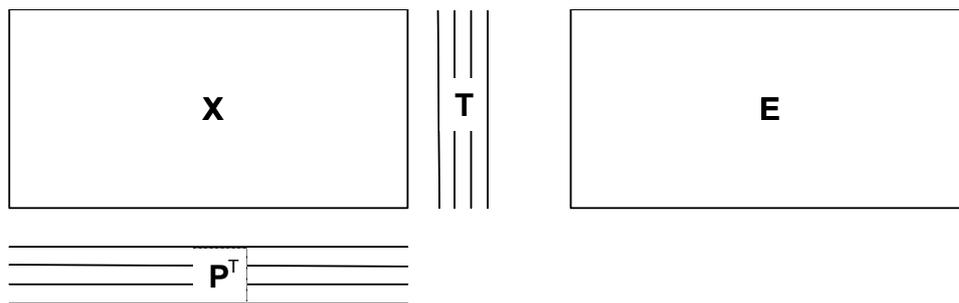


Figura 1.15. Representación esquemática de las matrices del modelo PCA.

Para una componente dada, los pesos definen la dirección sobre la que se están proyectando las observaciones y están poniendo de manifiesto la contribución de las variables originales en la formación de dicha componente. Las proyecciones de las observaciones en el subespacio de dimensión A definido por las direcciones principales son las variables latentes, también denominadas *scores* \mathbf{t}_a , y se obtienen de la siguiente manera:

$$\mathbf{t}_a = \mathbf{X} \mathbf{p}_a \quad \text{Ec. 1.2}$$

Las variables latentes no son observables ni medibles directamente, sino que se obtienen por combinación lineal de las variables originales. Se puede, por tanto, en

virtud del teorema central del límite, considerar que se distribuyen según una normal multivariante. Además estas variables latentes no están correlacionadas puesto que son ortogonales (independientes).

Existen diversas formas para seleccionar el número de componentes principales a extraer (A). En la bibliografía, lo más habitual es recurrir a técnicas de validación cruzada (Wold, 1978). Esta técnica consiste en seleccionar un subconjunto de los datos y emplear el resto de los datos para ajustar el modelo PCA. Una vez ajustado, se emplean los datos del subconjunto para calcular la suma de los cuadrados de los errores de predicción (PRESS, *predictive residual sum of squares*). Este procedimiento es repetido hasta que todas las observaciones hayan sido descartadas una vez, lo que permite obtener un valor global del PRESS. Se considera que una componente es significativa cuando el valor global del PRESS que ha sido calculado incluyendo la nueva componente es menor que el calculado sin esa componente.

Para realizar un PCA sobre una matriz de datos \mathbf{X} , primero se calculan las direcciones de máxima varianza, y posteriormente se proyectan las observaciones sobre dichas direcciones. Para el cálculo de las componentes principales se han desarrollado diversos algoritmos, unos permiten extraer todas las componentes simultáneamente, y otros lo hacen de manera secuencial. Son estos últimos los que tienen un mayor interés en la práctica puesto que son las primeras componentes las que explican la mayor parte de la variabilidad de los datos. El algoritmo NIPALS (*Non-linear Iterative Partial Least Squares*) es uno de los que más aceptación ha tenido, debido a su rápida convergencia, y a que puede tolerar una cierta cantidad de datos faltantes distribuidos de manera aleatoria (Nelson *et al.*, 1996).

1.8.4.2 Regresión en mínimos cuadrados parciales (PLS)

Las técnicas PLS fueron desarrolladas por Herman Wold, quien fecha el nacimiento del PLS en el año 1977 (Wold, 1982). Se puede considerar como una extensión del PCA, que permite conectar o relacionar la información contenida en dos matrices, \mathbf{X} e \mathbf{Y} , habitualmente \mathbf{X} contiene las variables de proceso, e \mathbf{Y} las variables de calidad.

Siendo, por tanto, el hecho de trabajar con dos espacios de datos, su principal diferencia con el PCA. Mientras en PCA se busca maximizar la varianza de \mathbf{X} , cuando se realiza un PLS se trata de maximizar la covarianza entre \mathbf{X} e \mathbf{Y} . Por tanto, el objetivo que persigue es encontrar las variables latentes de \mathbf{X} que expliquen no sólo la varianza de \mathbf{X} , sino aquella parte de la varianza que mejor predice las variables de calidad (\mathbf{Y}).

Al realizar un PLS se está modelando la relación entre dos bloques de variables, con el objetivo de predecir \mathbf{Y} a partir de \mathbf{X} . El modelo PLS puede expresarse de la siguiente manera:

$$\mathbf{t}_a = \mathbf{X}_{a-1} \mathbf{w}_a \quad ; \quad \mathbf{X}_a = \mathbf{X}_{a-1} - \mathbf{t}_a \mathbf{p}_a^T \quad \text{Ec. 1.3}$$

$$\mathbf{X} = \sum_{a=1}^A \mathbf{t}_a \mathbf{p}_a^T + \mathbf{E} \quad \text{Ec. 1.4}$$

$$\mathbf{Y} = \sum_{a=1}^A \mathbf{u}_a \mathbf{c}_a^T + \mathbf{F} \quad \text{Ec. 1.5}$$

En este caso hay dos subespacios \mathbf{X} e \mathbf{Y} , cuyas respectivas variables latentes son \mathbf{t}_a y \mathbf{u}_a , y cuyos vectores de pesos \mathbf{w}_a y \mathbf{c}_a son aquellas direcciones que maximizan la covarianza entre dichas variables latentes. En el modelo PLS las variables latentes del espacio original \mathbf{X} , son aquellas que además de explicar una gran parte de la varianza de \mathbf{X} , están más correlacionadas con las variables respuesta \mathbf{Y} .

De manera análoga al PCA, es habitual determinar el número de variables latentes a extraer por validación cruzada, y obtenerlas por medio del algoritmo secuencial NIPALS.

1.8.4.3 Regresión en componentes principales (PCR)

El PCR (*Principal Component Regression*) es una técnica que también se deriva del PCA, y que permite relacionar las dos matrices de datos (\mathbf{X} e \mathbf{Y}). Consiste en realizar una regresión lineal múltiple entre las componentes principales que se han extraído de la matriz de datos \mathbf{X} y cada una de las variables de calidad (\mathbf{Y}), siempre y cuando

éstas no estén correlacionadas. Si las variables de la matriz \mathbf{Y} no fueran independientes, la regresión se haría con cada una de las componentes principales que se extrajeran de dicha matriz (Massy, 1965; De Jong, 1993).

A diferencia del PLS, las variables latentes que se obtienen con el PCR son las que maximizan la varianza de la matriz de datos (\mathbf{X}), pero no tienen por qué ser aquellas con mayor capacidad predictiva sobre \mathbf{Y} . Las variables latentes del PCR son aquellas que maximizan la varianza, mientras que las del PLS son las que maximizan la covarianza. Se podría decir que el PLS supone una corrección de las direcciones que se obtendrían mediante la realización de un PCA sobre \mathbf{X} y otro sobre \mathbf{Y} , para tratar de maximizar la correlación entre las variables latentes obtenidas.

Otra diferencia entre ambas técnicas se debe a que el PCR consta de dos etapas, en la primera mediante un PCA se calculan las componentes principales, y en la segunda con dichas componentes se realiza una regresión lineal múltiple. Sin embargo, el PLS calcula las componentes de manera secuencial sin realizar dos etapas diferenciadas.

1.8.4.4 Bondad de ajuste y de predicción

En regresión interesa que un modelo ajuste lo mejor posible a los datos, y que además, con dicho modelo se puedan predecir nuevas observaciones con el menor error. Por tanto, es importante diferenciar entre la bondad de ajuste y la bondad de predicción de un modelo. En general, cuando se incrementa el número de componentes de un modelo de regresión, su bondad de ajuste aumenta. Sin embargo, la capacidad predictiva del modelo frente a nuevas observaciones (bondad de predicción) no aumenta necesariamente, sino que puede llegar a disminuir. La bondad de ajuste siempre es mayor que la bondad de predicción, y se dice que un modelo está sobreajustado cuando la diferencia entre ambas es muy acusada.

Cuando se realiza un PCA sobre una matriz \mathbf{X} con un determinado número de observaciones, se obtienen los valores predichos por el modelo para cada una de

ellas, por lo que también se pueden calcular las diferencias entre valores observados y predichos. Para el cálculo de la bondad de ajuste del modelo, la variabilidad total de los datos se descompone en dos términos: la variabilidad explicada por el modelo y la variabilidad residual. La variabilidad total o suma de cuadrados total (SCT) para una única variable se define como el sumatorio del cuadrado de las desviaciones de cada valor observado respecto de su media. Por otra parte, la variabilidad residual o suma de cuadrados residual (SCR) para una única variable se define como el sumatorio del cuadrado de las desviaciones de cada valor observado respecto de su valor predicho por el modelo. La SCT y la SCR para el caso de tener J variables, así como la bondad de ajuste de un modelo PCA con A componentes, R^2_x (cum), se pueden obtener a partir de las siguientes expresiones:

$$SCT = \sum_{j=1}^J \sum_{i=1}^n (x_{ji} - \bar{x}_j)^2 \xrightarrow{\text{datos centrados}} SCT = \sum_{j=1}^J \sum_{i=1}^n (x_{ji})^2 \quad \text{Ec. 1.6}$$

$$SCR = \sum_{j=1}^J \sum_{i=1}^n (x_{ji} - \hat{x}_{ji})^2 \quad \text{Ec. 1.7}$$

$$R^2_x (\text{cum}) = 1 - \frac{SCR}{SCT} \quad \text{Ec. 1.8}$$

donde x_{ji} es el valor de la variable j en la observación i , \bar{x}_j es el valor medio de la variable j , \hat{x}_{ji} es el valor estimado por el modelo con A componentes para la variable j en la observación i , J es el número de variables de la matriz de datos (\mathbf{X}) y n es el número de observaciones.

La SCR es el cuadrado de la distancia euclídea entre los valores observados y los predichos. En PCA se utilizan distintos estadísticos que representan o están directamente relacionados con esta distancia, como el error cuadrático de predicción, el estadístico Q o la distancia al modelo. Estos estadísticos se definen posteriormente en el apartado 1.8.6.2.1.

La bondad de predicción (Q^2) del modelo se obtiene por validación cruzada, calculando la suma de los cuadrados de los errores de predicción (PRESS) de un subconjunto de observaciones que no han sido utilizadas para ajustar el modelo

PCA. Es decir, el cálculo de la PRESS es análogo al de la SCR (Ec. 1.7), pero con la salvedad de que se calculan los valores predichos de un subconjunto de observaciones que no han sido utilizadas para construir el modelo PCA. El cálculo de la Q^2 correspondiente a la componente “ a ”, $Q^2(a)$, y la correspondiente al modelo PCA con A componentes, $Q^2(\text{cum})$, se realiza a partir de las siguientes expresiones:

$$Q^2(a) = 1 - \frac{\text{PRESS}(a)}{\text{SCR}(a-1)} \quad \text{Ec. 1.9}$$

$$Q^2(\text{cum}) = 1 - \prod_{a=1}^{a=A} [1 - Q^2(a)] \quad \text{Ec. 1.10}$$

donde $\text{PRESS}(a)$ es la suma de cuadrados del error de predicción obtenida por validación cruzada del modelo PCA con “ a ” componentes, y $\text{SCR}(a-1)$ es la suma de cuadrados residual del modelo sin la componente “ a ”.

Una componente “ a ” se considera significativa si $Q^2(a)$ es mayor de cero, lo que equivale según la Ec. 1.9 a decir que $\text{PRESS}(a) < \text{SCR}(a-1)$. Es decir, la suma de cuadrados del error de predicción obtenida por validación cruzada con el modelo PCA con “ a ” componentes es menor que la suma de cuadrados residual si la última componente (a) no estuviera. Por tanto, dicha componente aportará información significativa al modelo, ya que incluyéndola, éste predice con menor error. Se debe tener presente que aunque tanto la bondad de ajuste (R^2_x) como la de predicción (Q^2) se representan con un superíndice cuadrático, a diferencia de la R^2_x que es siempre positiva, la Q^2 puede adoptar valores negativos, concretamente esto se producirá cuando la $\text{PRESS}(a)$ sea mayor que la $\text{SCR}(a-1)$.

En el caso de los modelos PLS que han sido descritos en el apartado 1.8.4.2, el cálculo de la bondad de ajuste se realiza tanto para las variables explicativas (matriz \mathbf{X}), como para las variables respuesta (matriz \mathbf{Y}), y el cálculo de la bondad de predicción se realiza para las variables respuesta.

1.8.5 Estructura de los datos

1.8.5.1 Procesos en continuo

En un proceso en continuo, los datos se pueden organizar en dos matrices, una matriz **X** para almacenar los datos de las variables que se registran del proceso (caudales, temperaturas, conductividades,...) y una matriz **Y** para las variables de calidad (concentraciones de fósforo, de nitrato, DBOs,...). En dichas matrices, las observaciones (filas) son los instantes de tiempo en los que se registran las variables y en las columnas están las variables registradas.

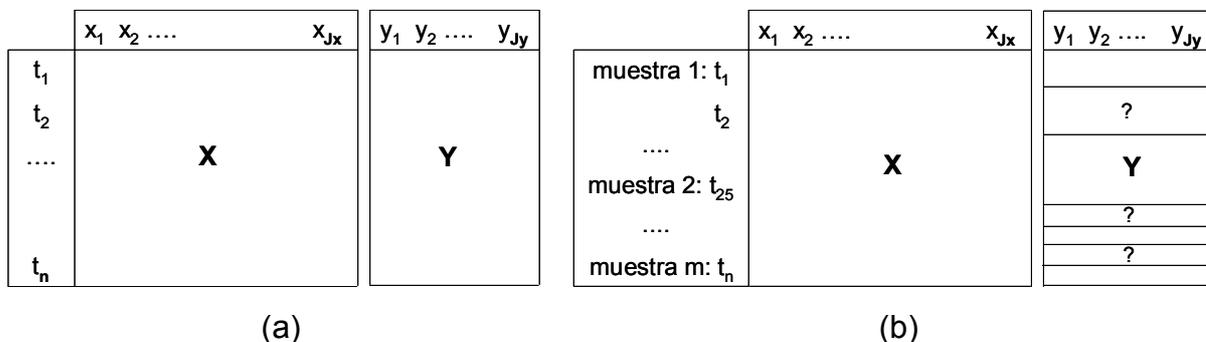


Figura 1.16. Estructura de las matrices de datos **X** e **Y** de un proceso en continuo. (a) situación ideal (b) situación habitual.

En contadas ocasiones se dispone de sensores para las variables de calidad. En el afortunado caso de tenerlos, las variables de calidad se podrían registrar con la misma frecuencia de muestreo que las de proceso (Figura 1.16 a). Sin embargo, es mucho más habitual que, como se ha comentado anteriormente, para conocer los valores de las variables de calidad se tenga que recurrir a la realización de los correspondientes análisis en laboratorio, ocasionando que sólo para algunas observaciones de la matriz **X** de proceso se disponga además de los datos de las variables de calidad (Figura 1.16 b). Por tanto, si sólo se consideran aquellos instantes de tiempo en que se conocen los valores de las variables de calidad, la matriz de datos del proceso **X**, constituida por los registros en continuo de las variables de proceso, queda drásticamente reducida.

1.8.5.2 Procesos por lotes (discontinuos)

Un proceso discontinuo o por lotes (*batch*), es un proceso de duración limitada y que consta de una serie de fases: alimentación del reactor con una serie de componentes, operación bajo condiciones controladas (duración de las distintas etapas, nivel de oxígeno, temperatura, ...) durante las cuales las variables de proceso presentan una cierta evolución temporal o trayectoria, y finalmente descarga del producto. De este producto se suelen tomar muestras y analizar en laboratorio diversas variables de calidad. En ocasiones, algunas de estas variables no sólo se miden al finalizar el lote, sino también a lo largo del mismo, disponiendo, por tanto, de trayectorias de calidad además de las trayectorias de proceso.

Normalmente estos procesos van a mostrar diferencias de unos lotes a otros, que se deben a desviaciones de las variables de proceso respecto de sus trayectorias especificadas, ligeros cambios o errores en la carga de los componentes, alteraciones debidas a las impurezas de los reactivos que se utilizan, así como variaciones introducidas por los sistemas de control.

En un proceso por lotes, sea o no constante la duración de todas sus etapas, los datos se estructuran en una matriz tridireccional constituida por los lotes, las variables y los instantes de tiempo. En la Figura 1.16 se muestra la representación de esta estructura de datos, el primer lote completo así como el instante inicial de todos los lotes que conforman dicha estructura. Para analizar este tipo de matriz se puede, o bien reordenar la matriz tridireccional de manera que al final se tenga una matriz bidireccional a la cual se le pueden aplicar las técnicas anteriormente descritas (PCA, PLS, PCR), o bien emplear métodos que permiten analizar directamente estructuras tridireccionales.

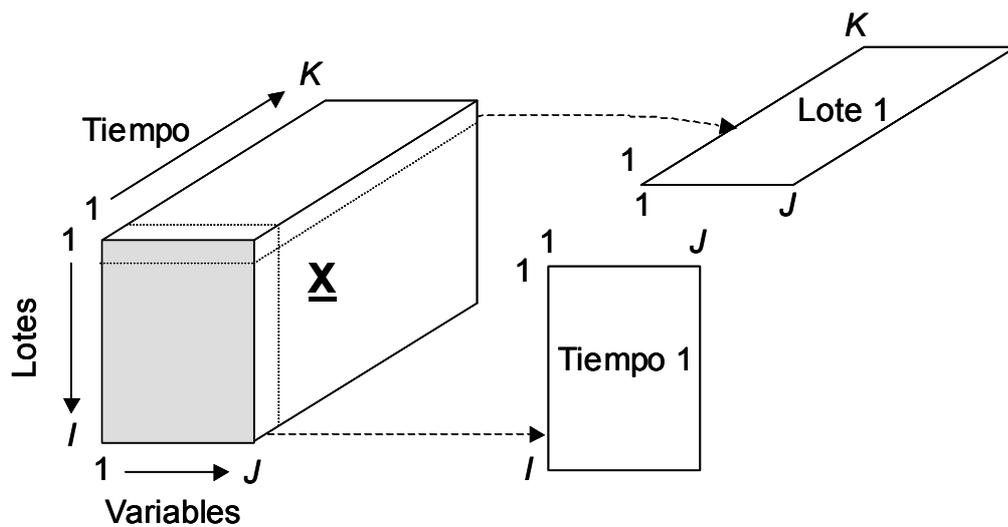


Figura 1.17. Matriz tridimensional ($\underline{\mathbf{X}}$) de los datos de un proceso por lotes.

1.8.5.2.1 Desdoblamiento de la matriz tridimensional a matriz bidireccional

Una forma de descomponer una matriz n -direccional, es fijando una dirección como vector y expresar el resto de direcciones como una matriz de orden $n-1$. Lohmöller y Wold (1980) demostraron que es posible mediante el producto de Kronecker de un vector por una matriz bidireccional descomponer una matriz tridimensional, y que esta descomposición es equivalente a desdoblar la matriz original en una matriz bidireccional que es analizada posteriormente mediante las técnicas de PCA y PLS.

Por tanto, la matriz tridimensional $\underline{\mathbf{X}}$ se descompone en el producto de Kronecker de las variables latentes (\mathbf{t}_a) por las matrices de pesos bidireccionales (\mathbf{P}_a), más una matriz de residuos tridimensional ($\underline{\mathbf{E}}$) que es minimizada en el sentido de mínimos cuadrados, como se muestra en la primera parte de la Ec. 1.11. La segunda parte de dicha ecuación muestra la descomposición más habitual en dos dimensiones.

$$\underline{\mathbf{X}} = \sum_{a=1}^A \mathbf{t}_a \otimes \mathbf{P}_a + \underline{\mathbf{E}} = \sum_{a=1}^A \mathbf{t}_a \mathbf{p}_a^T + \underline{\mathbf{E}} = \hat{\underline{\mathbf{X}}} + \underline{\mathbf{E}} \quad \text{Ec. 1.11}$$

donde el símbolo \otimes es el producto de Kronecker y A es el número de componentes principales extraídas.

Posteriormente Wold *et al.* (1987b) generalizan la metodología de descomposición a más direcciones y desarrollan algoritmos para realizar dicha descomposición. Denominan *Multiway PCA* y *Multiway PLS* a los análisis PCA y PLS realizados sobre las matrices bidireccionales que se obtienen del desdoblamiento de la matriz multidireccional original. Recientemente han surgido metodologías basadas en PCA y PLS que permiten el análisis directo de la matriz multidireccional, es decir, sin tener que recurrir a su desdoblamiento previo. Por este motivo, Kiers (2000) propone denominar a los métodos que desdoblan la matriz *Unfold PCA* (U-PCA) y *Unfold PLS* (U-PLS), respectivamente.

Para una matriz tridireccional hay seis posibles maneras de realizar el desdoblamiento, aunque sólo tres no son equivalentes entre sí, y se corresponden con preservar cada una de las tres direcciones de la matriz original: lotes, variables e instantes de tiempo. En la Tabla 1.3 se presentan los distintos tipos de desdoblamiento y se indica la estructura de la matriz resultante así como la dirección que permanece inalterada en cada caso.

Tipo	Estructura	Dirección preservada
A	$IK \times J$	Variables
B	$Jl \times K$	Tiempo
C	$IJ \times K$	Tiempo
D	$l \times KJ$	Lotes
E	$l \times JK$	Lotes
F	$J \times lK$	Variables

Tabla 1.3. Tipos de desdoblamiento de la matriz tridireccional, según Westerhuis *et al.* (1999).

De las tres direcciones de desdoblamiento, dos son las habitualmente encontradas en la bibliografía, la que mantiene la dirección de las variables y la que mantiene la dirección de los lotes. Según Nomikos (1995), son éstos los únicos tipos de desdoblamiento que tienen sentido. El desdoblamiento preservando las variables es el defendido por Wold *et al.* (1998), y según estos autores presenta ventajas cuando el objetivo que se persigue es la monitorización en tiempo real. El otro tipo de desdoblamiento es el propuesto por Nomikos y MacGregor (1994), que preserva la dirección de los lotes. Este desdoblamiento es el más usual cuando se realiza el análisis de datos históricos, aunque también es ampliamente utilizado para la

monitorización en tiempo real. En la Figura 1.18 se muestran los tipos de desdoblamiento D y A, que mantienen la dirección de los lotes y de las variables, respectivamente.

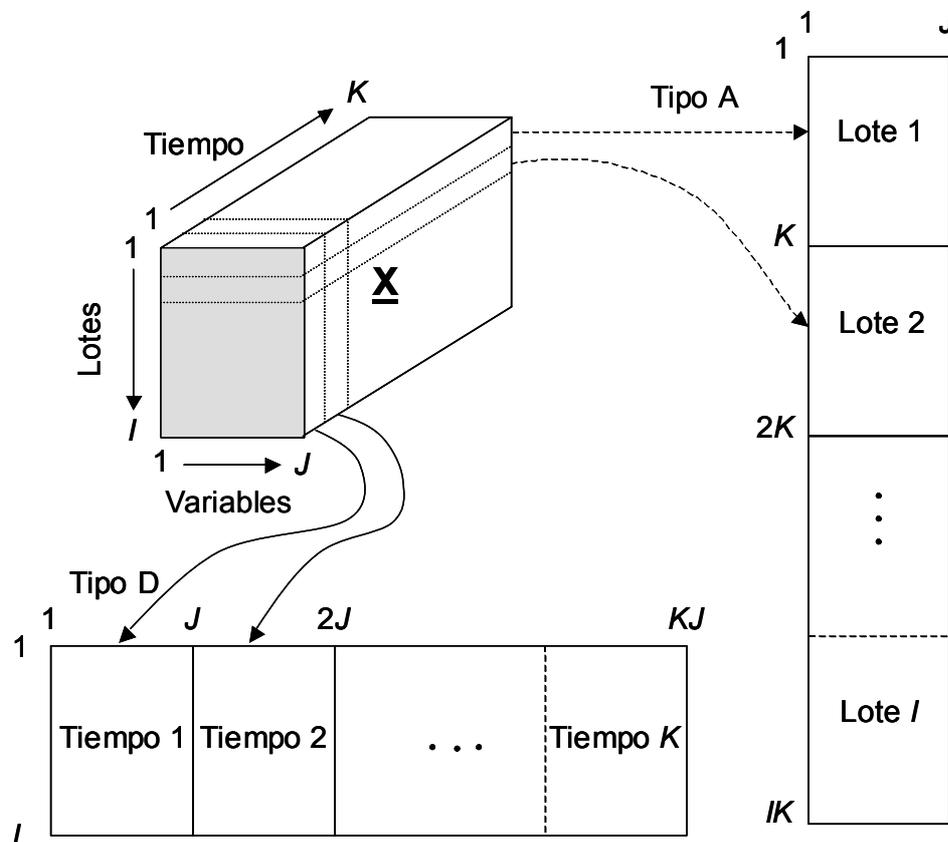


Figura 1.18. Desdoblamientos D y A de la matriz tridireccional de los datos (\underline{X}).

1.8.5.2.2 Análisis directo de la matriz tridireccional

Existen métodos que permiten abordar el análisis directo de la matriz tridireccional sin necesidad de recurrir a su desdoblamiento previo. Entre ellos cabe destacar los modelos Tucker3 y PARAFAC (en el caso de modelos de compresión), así como los modelos predictivos N-PLS.

Los modelos Tucker3 fueron desarrollados por Tucker (1963, 1964, 1966) y permiten descomponer la matriz tridireccional en tres vectores de pesos en cada dirección, que estarían relacionados a través de una matriz central (*core matrix*). Todas las componentes son extraídas de manera simultánea.

Los modelos PARAFAC (*PARA*lel *FACT*or *ANALYSIS*) para el análisis de matrices tridireccionales fueron desarrollados de manera independiente por Harshman (1970) y por Carroll y Chang (1970). La extracción de componentes también se realiza de manera simultánea. Para cada componente se obtienen tres vectores, el vector de las proyecciones de las observaciones y dos vectores de pesos para cada una de las dos direcciones de las variables. Aunque el número de parámetros del modelo es menor que el obtenido por U-PCA, su interpretación puede resultar más complicada.

Los modelos N-PLS descomponen la estructura n -direccional de las variables de proceso en un modelo multilineal similar al PARAFAC, con la diferencia de que en este caso el modelo se ajusta maximizando la covarianza con las variables respuesta. Los modelos N-PLS son una extensión de los modelos de regresión PLS bidireccionales al caso multidireccional. El algoritmo N-PLS fue desarrollado por Bro (1996).

1.8.6 Monitorización de procesos por lotes (discontinuos)

Puesto que la presente tesis se centra en el ámbito de los procesos por lotes, a lo largo de este apartado se hace especial referencia a la monitorización de este tipo de procesos. Se debe tener presente que la monitorización de un proceso continuo se realizaría de manera similar, aunque habría que tener en cuenta la distinta estructura de los datos.

El término monitorización hace referencia al método mediante el cual se supervisa la evolución del proceso y se detectan los fallos, entendidos éstos como cualquier clase de anomalía del proceso que puede provocar un funcionamiento inaceptable del mismo y/o un producto final de mala calidad.

Para la monitorización de procesos pueden emplearse modelos mecanicistas, empíricos o híbridos. Las investigaciones clásicas para monitorizar procesos por lotes se han centrado en la utilización de modelos matemáticos fundamentales (mecanicistas) basados en métodos de estimación de los estados (Frank 1990), o en modelos basados en el conocimiento usando sistemas expertos o métodos de

inteligencia artificial para procesar los datos (Konstantinov y Yoshida, 1992; Holloway y Krogh, 1992).

Los métodos de estimación de los estados, aunque potencialmente son enfoques muy potentes para la monitorización de los procesos por lotes, presentan bastantes problemas para su aplicación en la práctica. En primer lugar, necesitan modelos teóricos detallados, para modelar la componente determinista del proceso, así como sensores en línea que registren las características de calidad del producto. El desarrollo y estimación de estos modelos y la construcción de estos sensores es bastante costoso y difícil en la práctica. Otra dificultad importante es que la capacidad de detección y diagnóstico de estos modelos mecanicistas es muy dependiente del conocimiento “a priori” de los posibles fallos y anomalías que pueden ocurrir en el proceso, ya que éstos deben ser explícitamente incluidos en el modelo. Por tanto, sucesos o anomalías que afecten al proceso, pero que no hayan sido incorporados en el modelo pueden no ser detectados por estas técnicas.

Por otra parte, los métodos basados en el conocimiento aprovechan la experiencia que técnicos y operarios tienen sobre el proceso para desarrollar los sistemas expertos y los algoritmos de inteligencia artificial (redes neuronales). Estos métodos no requieren un modelo teórico detallado del proceso, sin embargo, su formulación puede llegar a ser tan costosa y difícil en la práctica como la de los modelos mecanicistas. En el caso particular de las redes neuronales se necesita disponer de abundantes datos reales con bastantes tipos de fallos para su entrenamiento, algo que lógicamente es raramente disponible en procesos reales.

Otro enfoque filosóficamente diferente al de los modelos mecanicistas y basados en el conocimiento es el del control estadístico de procesos (SPC), comentado en el apartado 1.8.3, en el que se trata de comparar el estado actual del proceso con las condiciones operativas consideradas “normales”. La utilización del SPC presenta la ventaja de que no necesita ningún modelo de partida, sino que usa directamente los datos registrados en el proceso para la construcción de un modelo empírico, proporcionando adicionalmente herramientas gráficas (gráficos de control) fáciles de aplicar e interpretar por los técnicos y operarios del proceso.

La aplicación del SPC clásico a los procesos por lotes ha sido muy limitada. Como ya se ha comentado en el apartado 1.8.3, estas técnicas fueron diseñadas para monitorizar procesos univariantes o conjuntos de variables independientes, y no son adecuadas en el contexto de procesos con muchas variables altamente correlacionadas (Harris y Ross, 1991).

En la calidad del producto final no sólo es importante la relación entre todas las variables de proceso en un instante de tiempo dado, sino también la evolución de las trayectorias de todas estas variables a lo largo del lote. Por tanto, sería muy interesante poder construir, a partir de estos datos, un modelo empírico que caracterizara la evolución de los lotes buenos. Las principales dificultades para conseguirlo residen en cómo manejar la gran cantidad de variables de proceso medidas a lo largo del transcurso de un lote, su variación en el tiempo y la naturaleza no lineal, así como la duración limitada de los procesos por lotes.

Para abordar este tipo de problemas multivariantes en procesos continuos funcionando alrededor de un nominal fijo, diversos autores han desarrollado técnicas de control estadístico de procesos multivariante (MSPC, *Multivariate Statistical Process Control*) basadas en técnicas estadísticas de proyección sobre estructuras latentes, PCA y PLS (Kourti *et al.*, 1996; Kourti y MacGregor, 1996; MacGregor *et al.*, 1994).

Durante la década pasada diversos autores trataron de extender los métodos estadísticos de proyección multivariantes, PCA y PLS, desarrollados para la monitorización de procesos continuos, al caso de los procesos por lotes. Nomikos (1995), Nomikos y MacGregor (1994a y 1995) y Kosanovich *et al.* (1996) emplearon la técnica del *Unfold PCA* (U-PCA) para extender los métodos de SPC multivariante a los procesos por lotes (BMSPC, *Batch Multivariate Statistical Process Control*). A partir de estos trabajos pioneros, se han ido desarrollando nuevas aplicaciones y extensiones. Wold *et al.* (1998) proponen una variante para la modelación, monitorización y diagnóstico de procesos por lotes, ya comentada en el apartado 1.8.5.2, cuya principal característica es la forma de organizar los datos del proceso.

Rännar *et al.* (1998) usan un enfoque recursivo para actualizar los modelos PCA y PLS, obteniendo como resultado modelos que permiten la monitorización en continuo de la evolución de un lote sin necesidad de estimar las medidas futuras. Louwerse y Smilde (2000) emplearon modelos tridireccionales, PARAFAC y Tucker3, para el control estadístico multivariante de procesos por lotes.

La monitorización multivariante de procesos por lotes se basa en modelar las causas comunes de variabilidad presentes en un conjunto de observaciones representativo de las condiciones normales de operación (observaciones bajo control estadístico). Mediante la proyección de la información contenida en las trayectorias de las variables de proceso sobre espacios de variables latentes de menor dimensión, tanto las correlaciones entre variables como su evolución en el tiempo pueden ser sintetizadas en unos pocos gráficos fáciles de construir e interpretar por los técnicos de proceso. En estos enfoques multivariantes, la variación en las trayectorias entre los lotes representativos de las condiciones normales de operación (variación debida a causas comunes) se sintetiza en un reducido espacio de vectores latentes mediante la aplicación de métodos de proyección multivariante. El modelo de referencia construido permite detectar si el comportamiento de nuevos lotes es consistente con el funcionamiento definido por la distribución de referencia histórica, o si por el contrario presentan causas especiales de variabilidad. En este último caso, se tratará de identificar y diagnosticar las causas de las alteraciones producidas para poder posteriormente implementar las medidas correctoras oportunas.

En el enfoque U-PCA los esquemas de monitorización sólo utilizan la información de las trayectorias de las variables de proceso a lo largo del lote. Las medidas de calidad del producto acabado sólo se tienen en cuenta a la hora de clasificar un lote como bueno o malo. Sin embargo, esta información sobre la calidad del producto acabado puede ser incluida en el esquema de monitorización mediante la utilización de U-PLS. Con este método es posible disponer en tiempo real de una predicción de las variables de calidad del producto final del lote antes de la finalización del mismo.

Se plantea entonces la cuestión de qué enfoque es más conveniente para la monitorización de procesos por lotes, el U-PCA o el U-PLS. En el primer enfoque (U-PCA), el modelo construido detectará cualquier anomalía en las medidas del proceso incluso aunque ésta sea irrelevante para la calidad del producto final. Por el contrario, el modelo construido mediante U-PLS sólo indicará una señal de alarma en el caso en que los cambios en el proceso puedan afectar la calidad del producto. Por este motivo, usar un enfoque u otro dependerá de si sólo se está interesado en detectar cambios que afecten a la calidad del producto, o en cualquier tipo de comportamiento anormal del proceso. En general, como medida preventiva, suele ser beneficioso intentar detectar todo tipo de deterioro del proceso, afecte o no inicialmente a la calidad del producto final, con el fin de poder corregirlo antes de que pueda convertirse en permanente y provocar otros efectos colaterales que puedan incluso comprometer la seguridad del proceso. Por estos motivos, Kourti (2003) afirma que es preferible la utilización de U-PCA frente a U-PLS.

Se debe tener presente que ambas técnicas, U-PCA y U-PLS, no son excluyentes, por lo que una posible solución a la problemática anterior puede ser la utilización combinada de ambas. De esta forma, se realizaría la construcción de un modelo U-PCA basado en información histórica de lotes buenos, y un modelo U-PLS basado en información procedente tanto de lotes buenos como malos. El modelo U-PCA se utilizará para la monitorización en tiempo real del proceso y la detección de anomalías. Una vez detectada alguna anomalía, se utilizará el modelo U-PLS para predecir la calidad final del producto del lote.

Los esquemas de monitorización propuestos basados en las técnicas U-PCA y U-PLS pueden extenderse a situaciones donde el proceso puede considerarse constituido por una serie de bloques o secciones. En algunos procesos por lotes, puede haber una etapa de pre-procesado antes de que comience el proceso propiamente dicho. Todos estos bloques de información junto con las variables de calidad del producto final pueden integrarse en un único esquema SPC mediante el uso de los denominados métodos de proyección multi-bloque, *Multi-Block* U-PCA (MB-U-PCA) y *Multi-Block* PLS (MB-U-PLS). Ejemplos de utilización de estos métodos pueden encontrarse en MacGregor *et al.* (1994) y Kourti *et al.* (1995).

Independientemente de cual sea el método de proyección multivariante empleado (U-PCA, U-PLS, MB-U-PCA, MB-U-PLS) es importante tener en cuenta que, como todo enfoque inferencial, para que estos métodos funcionen es necesario que se cumplan dos supuestos básicos: que los procesos sean “**comparables**” y que los sucesos de interés sean “**observables**”. El primer supuesto establece que el método es válido en tanto en cuanto la base de datos histórica de referencia sea representativa del funcionamiento del proceso. El segundo supuesto expresa el requisito de que los sucesos que uno desea detectar deben ser “observables” a partir de las medidas que están siendo recogidas en el proceso.

En el caso de emplear modelos capaces de analizar directamente la matriz tridireccional, sin necesidad de desdoblamiento previo, pueden aplicarse los mismos razonamientos sobre la elección entre modelos de compresión (PARAFAC y Tucker3) o modelos predictivos (N-PLS).

La implementación de un sistema de monitorización basado en métodos estadísticos de proyección multivariante consta de tres etapas: selección de un conjunto de observaciones de referencia, modelación del proceso y explotación en tiempo real del modelo construido.

1.8.6.1 Selección del conjunto de observaciones de referencia

Esta etapa consiste en seleccionar una serie de lotes que sean representativos de las condiciones operativas normales del proceso. La selección se realiza en base al conocimiento técnico del proceso. Esta etapa de selección es crítica para alcanzar los objetivos de la monitorización, puesto que la variabilidad presente en las observaciones seleccionadas sirve como distribución de referencia. Es decir, estas observaciones definen las variaciones que el proceso puede presentar cuando está funcionando normalmente y, por tanto, todas aquellas variaciones que permitan alcanzar unos resultados adecuados del proceso deberían ser incluidas en el modelo de referencia.

En el caso de que el conjunto de referencia no incluya variaciones aceptables del proceso, el modelo construido generará un exceso de alarmas durante la monitorización de nuevas observaciones. Por otro lado, si lo que ocurre es que el conjunto de referencia incluye situaciones que generan resultados inaceptables en el proceso, el modelo construido perderá sensibilidad, es decir, disminuirá su capacidad para detectar estas situaciones.

1.8.6.2 Modelación del proceso

Esta etapa consiste en construir un modelo empírico y calcular los límites de confianza de los gráficos de control a partir del conjunto de observaciones de referencia. Suele ser conveniente utilizar las observaciones que presenten anomalías en su comportamiento para analizar si el modelo desarrollado es capaz en primer lugar de detectarlas, y en segundo lugar identificar correctamente las causas.

Como se ha comentado previamente, para la modelación de la estructura tridireccional de un proceso por lotes se pueden emplear distintos métodos estadísticos de proyección multivariante. En este apartado se describe la modelación del proceso mediante U-PCA, utilizando, por tanto, únicamente las trayectorias de las variables de proceso a lo largo del lote, y desdoblado la matriz tridireccional.

Sobre la matriz de los lotes del conjunto de referencia desdoblados se aplica un PCA, obteniendo las componentes principales significativas, normalmente mediante técnicas de validación cruzada. A continuación se verifica si entre los lotes de referencia que se han utilizado en el desarrollo del modelo hay alguno que presente anomalías en su comportamiento. Este tipo de análisis “a posteriori” permite detectar fallos que no son fáciles de identificar examinando las trayectorias individuales de las variables. Por tanto, la selección del conjunto de lotes que se emplearán en la construcción del modelo PCA definitivo se basa en el conocimiento técnico del proceso y también en la aplicación de herramientas estadísticas multivariantes.

Para realizar el análisis “a posteriori” se proyectan todos los lotes sobre el modelo y se construyen dos gráficos, uno de dispersión de las variables latentes y otro con los residuos del modelo. La inspección de estos gráficos permite fácilmente detectar tanto lotes con residuos más elevados de lo normal, como lotes cuya proyección es anormalmente elevada en alguna componente. Aquellos lotes que hayan sido identificados y que, por tanto, presentan anomalías o se sospecha que no se han obtenido bajo condiciones operativas normales, se eliminan del conjunto de lotes de referencia y con los lotes restantes se vuelve a construir el modelo.

Como resultado del modelo definitivo ajustado se obtienen las variables latentes y los residuos correspondientes al conjunto de lotes representativos de las condiciones normales de operación. A partir de estas variables latentes y residuos se construyen las distribuciones de referencia para la realización de los gráficos de control. La monitorización de los residuos permite detectar si las observaciones guardan o no la estructura de correlación del modelo, mientras que la monitorización de las variables latentes permite detectar lo alejada que está la proyección de una determinada observación del centro del modelo.

Cuando una observación presenta una ruptura de la estructura de correlación, es decir, ha cambiado la relación entre las variables medidas respecto a la relación existente en las observaciones utilizadas para construir el modelo, su proyección presentará un residuo anormalmente elevado. En este caso se dice que la observación presenta una “anomalía moderada”. Por otra parte, cuando la anomalía que se produce no afecta a las correlaciones entre las variables medidas, pero la magnitud de éstas en valor absoluto es anormalmente elevada, se trata de una “anomalía severa o influyente”. Este tipo de anomalías están provocadas por una variación inusual dentro del modelo, y su proyección presentará un buen ajuste al modelo (el residuo de dicha observación será pequeño), pero se alejará de las proyecciones del conjunto de observaciones de referencia. En la práctica, lo habitual será encontrar que las perturbaciones afectan tanto a la magnitud de las variables como a sus relaciones y, por tanto, la proyección sobre el modelo presentará un residuo elevado y además estará alejada del centro del modelo.

Como ya ha sido expuesto previamente, el objetivo del PCA es condensar la información presente en todas las variables de proceso registradas, en un número reducido de nuevas variables independientes que explican gran parte de la variabilidad de las variables originales así como de sus relaciones. Al trabajar con un menor número de variables (variables latentes), que además son independientes, la efectividad de la monitorización mejora y se reduce el número de gráficos a analizar.

1.8.6.2.1 Monitorización del error cuadrático de predicción

El error cuadrático de predicción (*Squared Prediction Error*, SPE) representa la distancia perpendicular de una observación al hiperplano definido por las A componentes principales del modelo. Este estadístico (SPE) también es conocido como estadístico Q , y permite detectar las anomalías que no respetan la estructura de correlación del modelo. Para una observación dada, el SPE se calcula según la expresión:

$$SPE_i = \sum_{j=1}^J e_{ij}^2 = \sum_{j=1}^J (x_{ij} - \hat{x}_{ij})^2 = \mathbf{e}_i^T \mathbf{e}_i = (\mathbf{x}_i - \hat{\mathbf{x}}_i)^T (\mathbf{x}_i - \hat{\mathbf{x}}_i) \quad \text{Ec. 1.12}$$

donde e_{ij} es el error de predicción de la variable j en la observación i , x_{ij} es el valor de la variable original j en la observación i y \hat{x}_{ij} su valor predicho por el modelo PCA con las A componentes retenidas. Las últimas igualdades son las mismas expresiones pero en forma vectorial, donde \mathbf{e}_i es el vector de errores de la observación i , \mathbf{x}_i el vector de las variables originales de la observación i , y $\hat{\mathbf{x}}_i$ su estimación a partir del modelo PCA con las A componentes extraídas.

También es frecuente utilizar la distancia al modelo (D_{mod}) en lugar del SPE. Se trata de un parámetro proporcional al SPE que se obtiene a partir del él, dividiendo por los grados de libertad residuales:

$$D_{\text{mod}} = \sqrt{\frac{SPE}{g_{\text{res}}^l}} \quad \text{Ec. 1.13}$$

Para la monitorización del SPE es necesario determinar su límite de control superior. En la bibliografía se han propuesto distintas aproximaciones para caracterizar la pauta de variabilidad del SPE. Box (1954) demostró que una distribución chi-cuadrado ponderada, $g\chi_h^2$, en la que el factor de ponderación (g) y los grados de libertad (h) dependen de los valores propios de la matriz de varianzas-covarianzas de los residuos, es una buena aproximación de la distribución del SPE. El límite de control para el SPE se obtendría ajustando una distribución chi-cuadrado ponderada a la distribución de referencia del SPE que se obtiene del conjunto de lotes empleados en la construcción del modelo. Nomikos y MacGregor (1995) emplean la aproximación de Box para el cálculo de los límites de control en cada instante de tiempo de un proceso por lotes, estimando los parámetros de la distribución chi-cuadrado por el método de los momentos, igualando la media y varianza muestrales del SPE (m, v) a la media y varianza de la distribución chi-cuadrado ponderada ($\mu = gh, \sigma^2 = 2g^2h$). La ecuación que se obtiene para determinar el límite de control superior del SPE, con un riesgo de primera especie α , es la siguiente:

$$SPE_\alpha = \frac{v}{2m} \chi_{2m^2/v, \alpha}^2 \quad \text{Ec. 1.14}$$

donde $\chi_{2m^2/v, \alpha}^2$ es el percentil $(1-\alpha) \cdot 100$ de una distribución chi-cuadrado con $2m^2/v$ grados de libertad.

Otra manera muy utilizada para estimar los parámetros de la distribución chi-cuadrado ponderada es la aproximación propuesta por Jackson y Mudholkar (1979), cuya expresión para el límite de control viene dada por:

$$SPE_\alpha = \theta_1 \left[1 - \frac{\theta_2 h_0 (1 - h_0)}{\theta_1^2} + \frac{z_\alpha (2\theta_2 h_0^2)^{1/2}}{\theta_1} \right]^{1/h_0} \quad \text{Ec. 1.15}$$

donde los parámetros $\theta_i = \sum_{j=A+1}^R (\lambda_j)^i$ dependen de los valores propios (λ_j) de las componentes principales que no han sido retenidas en el modelo PCA (el modelo se construye sólo con A de las R componentes que se pueden extraer de la matriz de

datos \mathbf{X} , siendo R el rango de \mathbf{X}), z_{α} es el percentil $(1-\alpha) \cdot 100$ de una normal tipificada y h_0 es función de los parámetros θ_i , y se calcula según la siguiente expresión:

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2} \quad \text{Ec. 1.16}$$

En el mismo trabajo citado anteriormente (Nomikos y MacGregor, 1995), los autores emplean la aproximación de Jackson y Mudholkar (1979), para establecer el límite de control superior no de un instante de tiempo concreto sino de una observación completa, es decir, de un lote completo. La estimación de los coeficientes θ_i la realizan a través de la siguiente ecuación:

$$\theta_i = \text{traza}(\mathbf{V}^i); \quad \mathbf{V} = \frac{\mathbf{E}^T \mathbf{E}}{I-1} \quad \text{Ec. 1.17}$$

donde \mathbf{V} es la matriz de varianzas-covarianzas de los residuos, e I es el número de observaciones (lotes de referencia).

1.8.6.2.2 Monitorización de las variables latentes

Las variables latentes se obtienen por combinación lineal de las variables originales, y en virtud del teorema central del límite se puede considerar que se distribuyen según una normal multivariante. Para su monitorización se emplean los gráficos de la T^2 de Hotelling contruidos a partir de las A componentes principales retenidas, según la siguiente ecuación:

$$T^2 = \mathbf{t}_{1:A}^T \mathbf{\Theta}_{1:A}^{-1} \mathbf{t}_{1:A} = \sum_{a=1}^A \frac{t_a^2}{s_a^2} \quad \text{Ec. 1.18}$$

donde $\mathbf{t}_{1:A}$ es el vector de variables latentes, $\mathbf{\Theta}_{1:A}$ es la matriz de varianzas-covarianzas de las variables latentes del conjunto de referencia, t_a es la variable latente a -ésima, y s_a^2 su varianza.

La T_{global}^2 de Hotelling se emplea en procesos multivariantes para medir la distancia de una observación al centro del modelo ponderando según la estructura de

covarianzas. Este estadístico se obtiene a partir de las variables originales del conjunto de referencia según la Ec. 1.19, o bien a partir de las variables latentes con la Ec. 1.18 pero empleando todas las componentes principales, y no únicamente las A componentes significativas. Tanto en la Ec. 1.18 como en la Ec. 1.19 se asume que los datos están centrados y, por tanto, su media es cero.

$$T_{global}^2 = \mathbf{x}^T \mathbf{S}^{-1} \mathbf{x}; \quad \mathbf{S} = \frac{\mathbf{X}^T \mathbf{X}}{I - 1} \quad \text{Ec. 1.19}$$

donde \mathbf{x} es la nueva observación (el vector de variables medidas de dicha observación) y \mathbf{S} la matriz de varianzas-covarianzas estimada a partir del conjunto de observaciones de referencia (\mathbf{X}).

La definición alternativa de la T_{global}^2 de Hotelling en la que se emplean únicamente las A componentes principales extraídas, evita que las componentes no significativas, que tienen valores propios muy pequeños, perjudiquen la monitorización del proceso al provocar que la matriz de varianzas-covarianzas esté mal condicionada (Kourti y MacGregor 1996). Al estadístico T^2 de Hotelling, obtenido a partir de las A componentes principales retenidas, también se le denomina estadístico D.

La forma más habitual de obtener el límite de control superior del gráfico de la T^2 es empleando la expresión obtenida por Tracy *et al.* (1992):

$$T_{\alpha}^2 = \frac{(I^2 - I)A}{I(I - A)} F_{1-\alpha}(A, I - A) \quad \text{Ec. 1.20}$$

donde $F_{1-\alpha}(A, I - A)$ es el percentil $(1-\alpha) \cdot 100$ de la distribución F-Snedecor con $(A, I - A)$ grados de libertad.

1.8.6.2.3 Gráficos de contribución

Los gráficos para la monitorización de las variables latentes y del SPE permiten detectar desviaciones respecto de las condiciones normales definidas por los datos del conjunto de referencia empleados en la construcción del modelo. Sin embargo, dado que no se está monitorizando las variables originales sino unos estadísticos que se obtienen a partir de ellas y que están sintetizando la mayor parte de la información que contienen, se complica la tarea de identificar qué variables y en qué instantes de tiempo son las causantes de la salida de control del proceso.

Los gráficos de contribución constituyen la principal herramienta para analizar las causas de una salida de control en los gráficos de la T^2 y del SPE. Miller *et al.* (1993) y MacGregor *et al.* (1994) proponen el uso de los gráficos de contribución como herramienta de diagnóstico ya que indican cómo contribuye cada variable individual en la proyección de un lote dado.

El SPE de una determinada observación se obtiene sumando los cuadrados de los errores de predicción para cada una de las variables del proceso (Ec. 1.12) y, por tanto, la contribución de una variable individual x_j al SPE viene dada por:

$$\text{Cont}(\text{SPE}; x_j) = (x_j - \hat{x}_j)^2 \quad \text{Ec. 1.21}$$

Sin embargo, es habitual realizar el estudio de las contribuciones al SPE sin elevar al cuadrado los errores, puesto que de esta manera se conserva la información referente al signo del residuo.

La T^2 se obtiene a partir de la contribución individual de las variables latentes según la Ec. 1.18 y, por tanto, de forma directa se determina la contribución de la variable latente a -ésima mediante la siguiente expresión:

$$\text{Cont}(T^2; t_a) = \frac{t_a^2}{s_a^2} \quad \text{Ec. 1.22}$$

En el caso de que una observación presente un valor elevado de la T^2 , el gráfico de contribución nos indicaría qué variables latentes son las que influyen en mayor medida en dicho valor de la T^2 . Sin embargo, desde el punto de vista del diagnóstico, el interés reside en conocer qué variables originales son las responsables de la salida de control. Para lo cual es necesario realizar los gráficos de contribución de las variables individuales a las variables latentes. Asumiendo que las variables están centradas, la contribución de la variable de proceso x_j a la variable latente t_a se define como:

$$Cont(t_a; x_j) = p_{aj} x_j \quad \text{Ec. 1.23}$$

1.8.6.3 Explotación en tiempo real del modelo construido

La metodología descrita hasta el momento permite detectar anomalías e identificar las variables responsables en lotes completos, es decir, en lotes que ya han finalizado y que, por tanto, se dispone de las variables de proceso medidas en todos los instantes de tiempo. Aunque este tipo de análisis de lotes finalizados es de gran utilidad, también tiene mucho interés el poder detectar la aparición de perturbaciones en el proceso en el momento en que se están produciendo sin tener que esperar a que acabe lote, ya que esto permitiría poder actuar y corregir el problema con antelación. Sin embargo, la aplicación de esta metodología a la monitorización en tiempo real de un proceso por lotes no es directa ya que en un instante dado de tiempo se conocen los valores de las variables en instantes anteriores pero no los posteriores.

Se describen a continuación distintos enfoques que han aparecido en la bibliografía y que son los más empleados para la monitorización en tiempo real de un proceso por lotes.

1.8.6.3.1 Enfoque de Nomikos y MacGregor (1995)

En la metodología desarrollada por estos autores, la matriz tridireccional de los datos se desdobra manteniendo la dirección de los lotes (desdoblamiento tipo D, Figura

1.18). De esta manera, cada observación está constituida por un lote completo. Por tanto, en un instante dado del transcurso de un lote, el vector de observación estará constituido por tres partes: medidas de instantes de tiempo anteriores, medidas del instante de tiempo actual y medidas futuras (Figura 1.19).

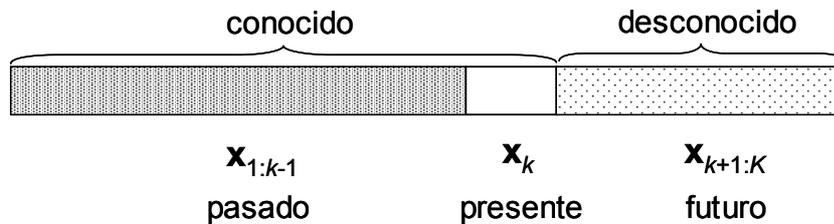


Figura 1.19. Vector de observación en el instante k de un lote que no ha finalizado.

Para la monitorización en tiempo real del proceso es necesario imputar los valores correspondientes a la evolución futura del lote. Estos autores proponen tres métodos: asumir una desviación nula, asumir una desviación constante, y tratar la parte desconocida como datos faltantes. Por otra parte, indican que la forma más correcta de resolver el problema de no disponer de toda la información del lote sería mediante la construcción de múltiples modelos de referencia, uno para cada instante de tiempo, aunque señalan como inconveniente el alto coste computacional que tendría. A continuación se describen brevemente los métodos propuestos:

- **Desviación nula.** En este caso lo que se hace es suponer que el resto del lote va a seguir la trayectoria media definida por el conjunto de lotes utilizados para modelar las condiciones normales de operación. Desde el punto de vista de la detección de fallos es la hipótesis más optimista puesto que si en el instante actual se está produciendo una perturbación, se está asumiendo que en el futuro de manera inmediata se regresará al comportamiento medio bajo condiciones operativas normales de funcionamiento.
- **Desviación constante.** En esta aproximación lo que se asume es que la parte desconocida del lote se desvía de la trayectoria media lo mismo que en el instante actual. Sería, por tanto, una hipótesis pesimista, puesto que si se está produciendo una perturbación en el instante actual, se considera que permanece hasta que acaba el lote.

- **Datos faltantes.** Este último método consiste en no rellenar la parte futura del lote. Se utilizan las componentes principales de la base de referencia para predecir los datos faltantes de forma que sean consistentes con la información disponible del lote que se está monitorizando y con la estructura de correlación de las variables presente en la base de referencia. Los autores indican que este método proporciona buenos resultados cuando se conoce al menos un 10% del lote que se está monitorizando.
- **Construcción de múltiples modelos de referencia.** En este caso se trata de desarrollar tantos modelos U-PCA como instantes de tiempo tiene un lote (K). Cada modelo se realiza incluyendo todas las variables medidas desde el inicio hasta un determinado instante de tiempo (k), es decir utilizando sólo la información disponible hasta ese instante de tiempo (Figura 1.20). Para monitorizar la evolución de un determinado lote, habría que utilizar en cada instante de tiempo el modelo correspondiente. Este método presenta la ventaja de no tener que imputar los valores futuros del lote, pero es computacionalmente muy costoso.

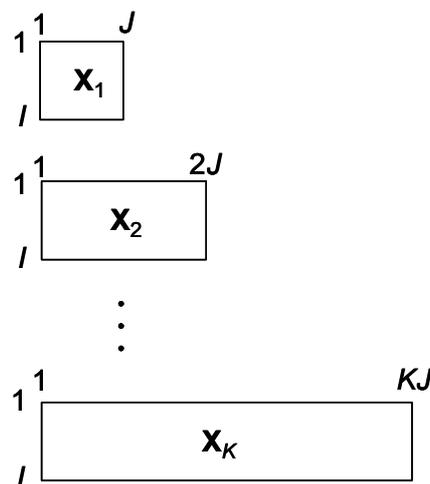


Figura 1.20. Matriz de datos para la construcción de un modelo de referencia en cada instante de tiempo.

1.8.6.3.2 Enfoque de Rännar et al. (1998)

Estos autores describen un algoritmo para la monitorización adaptativa de un proceso por lotes. La matriz tridireccional de datos es desdoblada manteniendo la dirección de los lotes (desdoblamiento tipo D, Figura 1.18). Para cada instante de tiempo (k) se construye un modelo local, es decir, un modelo que sólo utiliza los datos de las variables en ese instante de tiempo. De esta manera, se tendrán K modelos locales y para la monitorización se tendrá que utilizar el modelo correspondiente según el instante de tiempo en que se encuentre el lote. Con este enfoque no es necesario, por tanto, recurrir a ningún procedimiento para imputar la parte futura del lote.

La idea básica de este enfoque se muestra en la Figura 1.21. La matriz \mathbf{X}_k (I, J) está formada por todas las variables de proceso para todos los lotes del conjunto de referencia en el instante de tiempo k . Al realizar sobre esta matriz un PCA, se obtienen las componentes \mathbf{b}_k que sintetizan la variación existente en dicho instante de tiempo (k). El vector \mathbf{b}_k se pondera por un factor adaptativo (d), y pasa a formar parte de una nueva matriz \mathbf{B}_k junto con el vector \mathbf{t}_{k-1} que sintetiza la información del proceso hasta el instante $k-1$. A partir de la matriz \mathbf{B}_k se obtiene un nuevo vector \mathbf{t}_k , que condensa toda la información hasta el instante k . El mismo procedimiento se aplica desde el instante $k+1$ hasta último instante de tiempo (K).

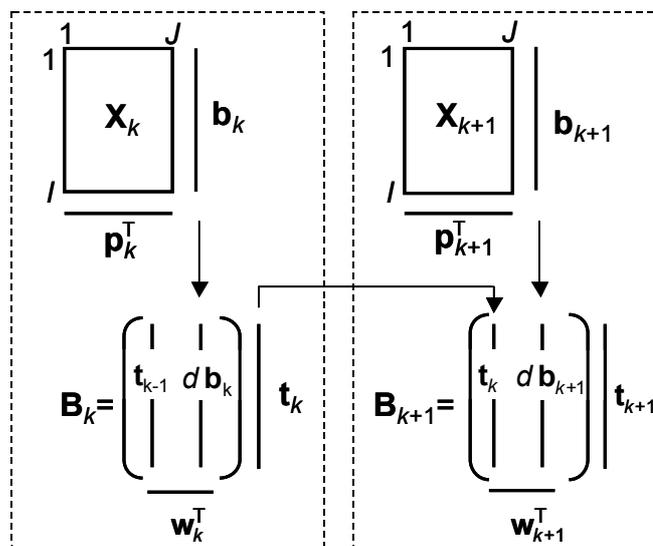


Figura 1.21. Monitorización adaptativa según el enfoque de Rännar et al. (1998)

El parámetro adaptativo d , determina el peso que se concede a la información del instante actual de tiempo respecto a la información de los instantes de tiempo anteriores, de forma que cuanto mayor sea el valor de este parámetro, más importancia se le concede a la información del instante actual, llegando al caso extremo de que si d fuera infinito, este enfoque sería equivalente a realizar un PCA sobre cada instante de tiempo k .

1.8.6.3.3 Enfoque de Wold et al. (1998)

La estrategia de modelación de procesos por lotes propuesta por Wold *et al.* (1998), consta de dos niveles de monitorización del proceso, el nivel de observación y el nivel *batch*. Cada uno de estos niveles persigue distintos objetivos. En el nivel de observación se pretende desarrollar un modelo que permita monitorizar la evolución de los lotes y detectar anomalías antes de que finalice el lote. En el nivel *batch* se utiliza toda la información disponible (condiciones iniciales, información del proceso, características de calidad,...) para construir un modelo con los lotes completos que permita clasificar el lote una vez ha finalizado como bueno o con anomalías.

En el primer nivel, **nivel de observación**, se desdobra la matriz tridireccional de los datos preservando la dirección de las variables, desdoblamiento tipo A (Tabla 1.3 y Figura 1.18). Por tanto, cada instante de tiempo de un lote constituye una observación, a diferencia del enfoque propuesto por Nomikos y MacGregor en el que una observación es un lote completo. Tras realizar el desdoblamiento se realiza un PLS sobre una variable respuesta ficticia que representa la madurez del lote, y que no es más que el tiempo que ha transcurrido desde el inicio del lote. En caso de que se disponga de alguna variable medida que sea más adecuada para representar la madurez del lote, no sería necesario utilizar la variable ficticia. El número de variables latentes a extraer se obtiene por validación cruzada, aunque los autores recomiendan extraer tantas variables latentes como sea necesario para que al menos el modelo sea capaz de explicar el 85% de los datos originales. Esto se debe a que las variables latentes son utilizadas posteriormente en el nivel *batch* para describir los lotes completos.

Para el desarrollo de los gráficos de control que permitan monitorizar los nuevos lotes, es necesario primero dividir las variables latentes obtenidas en los segmentos correspondientes a cada lote y posteriormente reordenarlas de forma que para un lote dado se obtenga un vector con las A variables latentes extraídas ordenadas secuencialmente y que tendrá dimensiones $(1 \times AK)$, tal y como se muestra en la Figura 1.22. Calculando la media y la desviación típica de cada componente para todos los lotes es posible establecer los límites de control para cada componente, estos límites se suelen fijar en ± 3 veces la desviación típica.

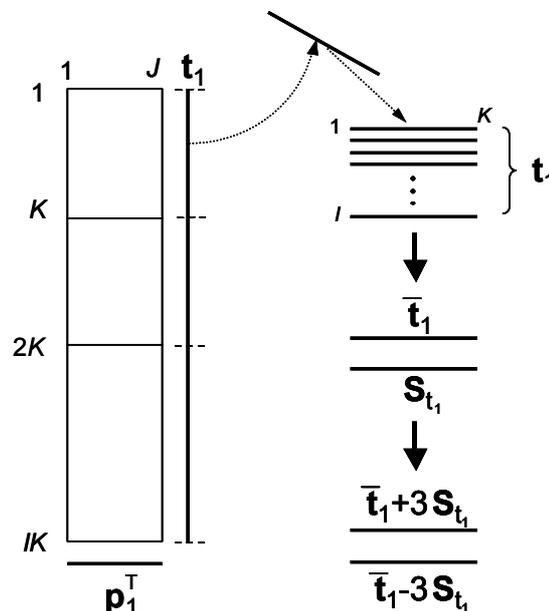


Figura 1.22. Obtención de los límites de control para la primera variable latente (t_1) del nivel de observación del enfoque de Wold *et al.* (1998).

En el **nivel batch**, se procede de manera análoga a la metodología de Nomikos y MacGregor y, por tanto, en este nivel un lote completo constituye una observación. Sin embargo, en lugar de utilizar las variables originales para formar la matriz \mathbf{X} , se emplean las variables latentes obtenidas en el nivel de observación.

En todos los enfoques es práctica habitual trabajar con datos centrados y, además, en los casos en que se desee que ninguna variable tenga "a priori" más influencia en el modelo, todas las variables se tipifican a varianza unitaria dividiendo cada variable por su desviación típica. Sin embargo, no todos los enfoques desdoblan la matriz

original de datos de la misma manera, y en consecuencia, el centrado consigue efectos distintos. En el enfoque de Nomikos y MacGregor (1995) al centrar los datos se está restando la trayectoria media (de las condiciones de operación normales), de forma que la principal componente no lineal del proceso se elimina y, por tanto, se convierte un problema no lineal en otro que se puede abordar con métodos lineales como PCA y PLS. En este caso, con la monitorización se persigue detectar desviaciones de las trayectorias respecto de su correspondiente trayectoria media. Por otra parte, en el enfoque de Wold *et al.* (1998), al centrar no se elimina la trayectoria media de las variables, sino que simplemente se resta una constante a cada trayectoria. Por tanto, la trayectoria media de las variables permanece, y es modelada por las componentes PLS del nivel de observación (efecto lineal con la primera componente, efecto cuadrático con la segunda, efecto cúbico con la tercera, ...).

1.8.6.4 El análisis multivariante en la monitorización de los tratamientos de aguas residuales

Dentro de los tratamientos de aguas residuales se encuentran tanto procesos continuos como procesos discontinuos o por lotes. Por este motivo, la revisión bibliográfica que se presenta a continuación incluye aplicaciones de ambos tipos de procesos.

Como se ha comentado anteriormente (apartado 1.7) la instrumentación en el campo de los tratamientos de aguas residuales ha experimentado un notable auge en las últimas décadas, y se prevee que en un futuro próximo tenga un papel todavía más importante.

Actualmente en EDARs modernas se dispone de gran cantidad de variables que son registradas en tiempo real. Éstas se han convertido en una importante fuente de información para conseguir una operación eficiente del proceso. Sin embargo, la complejidad de los procesos que tienen lugar en las EDARs junto con la abundancia de datos hace que sea necesario la utilización de técnicas que permitan extraer la información contenida en los datos.

Cuando se pretende realizar la monitorización en tiempo real de un proceso, el método de monitorización tiene que hacer frente a una serie de retos. Algunos de los cuales son de carácter general como los que apunta Kresta *et al.* (1991):

- El método tiene que ser capaz de manejar gran cantidad de datos entre los que hay colinealidad.
- El método debe reducir notablemente la dimensión del problema y permitir una interpretación gráfica sencilla de los resultados.
- En el caso en que se disponga tanto de variables de proceso como de variables de calidad del producto, el método debe proporcionar buenas predicciones de las variables de calidad.

Estos requisitos pueden ser satisfechos monitorizando el proceso mediante técnicas de análisis estadístico multivariante, como PCA y PLS. Sin embargo, para el caso particular de los tratamientos de aguas residuales, el método de monitorización también debe hacer frente a una serie de retos específicos (Rosen *et al.*, 2003):

- Baja calidad y fiabilidad de los datos como consecuencia del ambiente agresivo en el que están instalados los equipos de medición.
- Relaciones no lineales entre las variables.
- Condiciones cambiantes, debido por ejemplo a las variaciones diurnas y estacionales que provocan que el proceso esté lejos de ser estacionario.
- Relaciones dinámicas entre las variables.

Según estos autores, todos estos retos específicos no aparecen necesariamente en el mismo proceso ni al mismo tiempo, puesto que el objetivo de la monitorización puede ser diferente en cada caso particular.

Rosen *et al.* (1998) discuten diferentes técnicas para la detección de perturbaciones en un sistema de tratamiento de aguas residuales. Muestran la utilidad de la aplicación de modelos PCA y PLS para poder visualizar las anomalías en el proceso, y posteriormente identifican las variables responsables mediante la utilización de gráficos de contribución. Utilizan un conjunto de datos reales procedentes de una EDAR.

Teppola *et al.* (1997) utilizan un modelo PLS junto con la función de autocorrelación de las variables latentes para interpretar y detectar cambios en un proceso de tratamiento de aguas residuales. Las técnicas las aplican sobre un conjunto de datos reales procedentes de una EDAR.

Teppola *et al.* (1998) presentan un enfoque en el que combinan PLS y un algoritmo de *clustering* (*fuzzy c-means clustering*) para monitorizar una EDAR. Mediante el PLS extraen la información contenida en los datos de proceso, y posteriormente aplican el algoritmo de *clustering* sobre las variables latentes. Este enfoque combinado facilita la interpretación de los resultados. Utilizan un conjunto de datos reales.

Teppola *et al.* (1999a) combinan PCA con distintos algoritmos adaptativos de *clustering* para monitorizar una EDAR. Las variables latentes obtenidas mediante un PCA son utilizadas en distintos algoritmos de *clustering* y, posteriormente, comparan la capacidad de adaptación de estos algoritmos a las condiciones del proceso. El estudio lo realizan sobre un conjunto de datos reales.

Rosen y Lennox (2001) analizan dos extensiones del PCA para la monitorización de un proceso de tratamiento de aguas residuales. Una para tener en cuenta que el proceso no es estacionario, para lo que proponen la utilización de un PCA adaptativo o en aquellos casos en que la estructura de covarianzas no cambie, actualizar los parámetros de escalado (media y varianza). La otra extensión para tener en cuenta que las perturbaciones que pueden ocurrir en el proceso pueden ser lentas o rápidas. Por este motivo, proponen la monitorización en diferentes bandas de frecuencia, para ello emplean un análisis de multiresolución (MRA, *Multiresolution Analysis*) en combinación con un PCA. Utilizan datos reales de una EDAR para la mostrar los resultados de la aplicación de estas técnicas.

Rosen y Yuan (2001) presentan una metodología para la monitorización multivariante y el control de una EDAR durante eventos extremos. Monitorizan el proceso empleando un PCA dinámico sobre las variables del influente y,

posteriormente, aplican una técnica de *clustering* (*fuzzy c-means clustering*) para clasificar los distintos estados operacionales de la EDAR. Esta información es utilizada para ajustar los *set-points* de los lazos de control locales. La metodología propuesta es ilustrada a partir de datos simulados de la operación de una EDAR.

Lennox y Rosen (2002) desarrollan un nuevo algoritmo de monitorización basado en PCA, PCA multiescala adaptativo, que aprovecha las relaciones cambiantes entre las variables en diferentes bandas de frecuencia. Utilizan datos reales de una EDAR para mostrar el comportamiento del nuevo algoritmo así como el de un PCA adaptativo. La mayor flexibilidad del algoritmo que proponen permite una adaptación a una mayor variedad de cambios en el proceso.

Yoo *et al.* (2002) proponen un método para la detección de anomalías e identificación de las variables responsables basado en la detección de cambios en la distribución de los datos de proceso. Estos autores analizan los cambios en los valores propios obtenidos de la aplicación de PCA y emplean un índice de disimilitud para detectar perturbaciones en el proceso e identificar en qué banda de frecuencia se producen. El método lo aplican sobre datos reales de una EDAR.

Lee y Vanrolleghem (2003) desarrollan un algoritmo de monitorización, PCA multibloque adaptativo, y lo aplican para la monitorización de lotes finalizados de un reactor SBR a escala de laboratorio. Además de las variables medidas, estos autores también incluyen en el modelo la primera y segunda derivada de alguna de ellas. Comparan los resultados obtenidos aplicando un PCA estático, un PCA adaptativo y un PCA multibloque adaptativo. Indican que es necesario actualizar el modelo de manera recursiva y señalan como ventaja del algoritmo que proponen la mayor claridad en la identificación de las causas del fallo.

En Rosen *et al.* (2003) se discuten algunas posibles soluciones para hacer frente a cada uno de los retos específicos de la monitorización en tiempo real de EDARs. Además, ilustran con datos reales procedentes de una EDAR, cómo se pueden obtener buenos resultados mediante la aplicación de un PCA y dos sencillas extensiones: la primera consiste en reemplazar tanto los datos faltantes como los

erróneos, y la segunda consiste en implementar un preprocesamiento adaptativo de las variables, es decir, actualizar los parámetros de escalado (media y varianza).

Yoo *et al.* (2003) proponen un enfoque para la modelación no lineal y monitorización adaptativa mediante la combinación de modelos PCA y de lógica difusa (*fuzzy principal component regression*). El primer paso en esta metodología consiste en construir un modelo PCA con un conjunto de datos de referencia. Después, los datos proyectados son clasificados en varios grupos mediante un algoritmo adaptativo de *clustering* (*adaptive fuzzy c-means clustering*) y, posteriormente, se aplica un método discriminante adaptativo para distinguir entre cambios importantes en el proceso y simples fallos. Por otro lado, los datos proyectados son también empleados para construir un modelo de lógica difusa (Takagi-Sugeno-Kang) con la finalidad de estimar la reducción de DQO del proceso. El enfoque es ilustrado con datos reales de una EDAR.

Lee y Vanrolleghem (2004) proponen un algoritmo de monitorización, PCA de consenso adaptativo, y lo aplican a la monitorización en tiempo real de un reactor SBR a escala de laboratorio. Junto con las variables medidas también incorporan en el modelo la primera y segunda derivada de algunas de ellas. Comparan los resultados del algoritmo que proponen con el enfoque de Nomikos y MacGregor (1995) asumiendo una desviación constante para la trayectoria futura del lote. Ambos enfoques dan buenos resultados, e indican como ventaja del algoritmo que proponen que no es necesario recurrir a ningún tipo de imputación.

Lee *et al.* (2004a) proponen un método para la detección e identificación de fallos en sensores en procesos dinámicos. Emplean una técnica de reconstrucción del fallo basada en la aplicación de un PCA dinámico. Los resultados presentados se han obtenido a partir de datos simulados de la operación de una EDAR.

Lee *et al.* (2004b) presentan un método basado en la técnica de análisis de componentes independientes (*Independent Component Analysis*) para la monitorización de una EDAR. Comparan los resultados de este método con los obtenidos mediante un PCA sobre un conjunto de datos obtenidos por simulación.

Estudian dos tipos de perturbaciones, una externa y otra interna del proceso, y el método que proponen obtiene mejores resultados sólo en el segundo caso.

Lee *et al.* (2004c) desarrollan una técnica basada en la aplicación de *kernel* PCA para modelar la relación no lineal entre las variables y permitir la monitorización del proceso. Esta técnica consiste en realizar un mapeo no lineal entre el espacio de las variables originales y un nuevo espacio de mayor dimensión en el que las relaciones son lineales, y es en este nuevo espacio donde se aplica el PCA. Estos autores plantean realizar la monitorización del proceso en este nuevo espacio de mayor dimensión. Comparan los resultados de este método con los obtenidos por la aplicación de PCA en el espacio de las variables originales. Estudian dos tipos de perturbaciones, una externa y otra interna del proceso, y el método que proponen obtiene mejores resultados sólo en el segundo caso. Señalan como inconveniente del método propuesto que cuando detecta una perturbación en el proceso es difícil poder identificar cuáles son las posibles causas, ya que encontrar el mapeo inverso entre el espacio de mayor dimensión y el original es muy complejo o incluso imposible. Los resultados presentados se han obtenido a partir de datos simulados de la operación de una EDAR.

Miettinen *et al.* (2004) emplean dos técnicas de análisis multivariante, PCA y PARAFAC, para mejorar el conocimiento de los procesos que tienen lugar en un proceso de tratamiento biológico de aguas residuales. El conjunto de datos analizado abarca un periodo de 7 años y procede de la laguna 115E de la EDAR oeste de Melbourne. Concluyen que ambas técnicas son herramientas eficientes y complementarias con las que se pueden identificar alteraciones en el estado del agua residual.

Ruiz *et al.* (2004a) desarrollan un algoritmo para la detección del estado estacionario en procesos multivariantes. La metodología que proponen se basa en reducir la dimensión del espacio original de datos mediante un PCA, y sobre las nuevas variables obtenidas identificar cuando se ha alcanzado el estado estacionario. El algoritmo lo aplican sobre una planta piloto para el tratamiento anaerobio de las aguas residuales.

Ruiz *et al.* (2004b) combinan técnicas de análisis estadístico multivariante con un algoritmo de clasificación no supervisado para relacionar las clases obtenidas con el estado del proceso. Aplican la metodología propuesta a un conjunto de lotes finalizados de un reactor SBR. En primer lugar reducen el número de variables mediante PCA y posteriormente aplican el algoritmo de clasificación. Indican que obtienen mejores resultados cuando emplean PCA multibloque en lugar de PCA sobre toda la matriz de datos desplegada.

Yoo *et al.* (2004a) emplean el análisis de componentes independientes para la monitorización en tiempo real de un reactor SBR a escala de laboratorio. La metodología que plantean se basa en una nueva forma de desdoblar la matriz de datos y, posteriormente, aplicar el análisis de componentes independientes sobre la matriz desdoblada. Al comparar el método que proponen con el enfoque de Nomikos y MacGregor, obtienen un menor número de falsas alarmas.

Yoo *et al.* (2004b) aplican un análisis de componentes independientes dinámico para la monitorización de una EDAR. Comparan los resultados con la aplicación de un PCA dinámico. Concluyen que cuando las variables de proceso tienen una distribución no Gausiana, el análisis de componentes independientes resulta superior al PCA.

La revisión bibliográfica realizada pone de manifiesto que la aplicación de enfoques sistemáticos para aprovechar la información contenida en la gran cantidad de datos que se registran en tiempo real en EDARs está recibiendo cada vez mayor atención. Son numerosas las aplicaciones descritas, en las que se plantean soluciones (nuevos algoritmos, extensiones de los algoritmos existentes, combinación de distintas técnicas) a las dificultades encontradas en cada caso en particular. La mayor parte de investigaciones se han centrado en procesos en continuo y se han utilizado datos reales.

1.8.7 Análisis de datos históricos

En el apartado anterior se ha descrito el proceso de construcción de un modelo empírico para la monitorización en tiempo real del proceso. Otra aplicación del modelo construido es el análisis de toda la base de datos histórica disponible del proceso (análisis “a posteriori”, *off-line*). De esta forma, se puede realizar el diagnóstico de lotes fabricados con anterioridad, identificando los fallos y apuntando a sus posibles causas, para lo que es necesario aplicar todo el conocimiento técnico del proceso basado en la experiencia de operarios y técnicos. Las causas de anomalías ocurridas en el proceso que hayan conducido a la fabricación de un lote, o de una serie de lotes, con producto de mala calidad, pueden ser corregidas para evitar que los lotes futuros puedan verse afectados por los mismos fallos. Por tanto, este tipo de análisis puede conducir a una optimización del proceso.

1.8.8 Modelos predictivos

En muchos procesos es habitual no disponer de sensores que permitan conocer el valor de las variables de calidad en tiempo real. Lo más frecuente es que se tomen varias muestras del producto acabado y se envíen al laboratorio para determinar una serie de medidas de su calidad. Por otra parte, sí que se suele disponer de sensores que permiten conocer en tiempo real multitud de variables de proceso.

Por tanto, tiene mucho interés la construcción de modelos inferenciales, también denominados *soft sensors*, que permitan predecir los valores de las características de calidad a partir de la información registrada del proceso. Un *soft sensor* consiste en utilizar las medidas de uno o varios sensores instalados en el proceso junto con un algoritmo para estimar los valores de algún parámetro para el que no se dispone de un sensor instalado físicamente. Este tipo de “sensores ficticios” constituyen una herramienta de gran utilidad para la monitorización y el control de procesos.

La construcción de un *soft sensor* puede basarse en el empleo de modelos mecanicistas, de modelos empíricos o de modelos semi-empíricos. Sin embargo, debido a que en muchas ocasiones la tarea de conseguir un modelo teórico

detallado del proceso es muy complicada, la utilización de modelos empíricos ha atraído la atención de muchos investigadores.

Dentro de los modelos empíricos se encuentran los modelos de regresión en mínimos cuadrados parciales (PLS) y de regresión en componentes principales (PCR) que ya han sido presentados en apartados anteriores (1.8.4.2 y 1.8.4.3). Además, recientemente ha cobrado mucha popularidad la utilización de las redes neuronales artificiales (ANNs, *Artificial Neural Networks*), gracias a que permiten modelar fácilmente sistemas no lineales complejos. A continuación se describe la evolución y las características más importantes de este tipo de modelos.

1.8.8.1 Las Redes Neuronales Artificiales (ANN)

La alta capacidad que presenta el cerebro humano en la realización de tareas cognitivas complejas como el reconocimiento de formas (visión) y sonidos (audición), fue una gran motivación para intentar modelar su comportamiento. De esta motivación histórica es de donde deriva el nombre de este tipo de modelos: redes neuronales artificiales. Sin embargo, la mayoría de las arquitecturas de redes neuronales que se emplean en la actualidad no tratan de imitar detalladamente el modelo biológico, sino que simplemente son algoritmos que en lugar de trabajar secuencialmente, lo hacen en paralelo.

En los últimos años las redes neuronales artificiales se han convertido en una herramienta muy utilizada en diversas áreas científicas, como la econometría y las finanzas, la quimiometría, la automática, los recursos hídricos y las ciencias medioambientales. Deben principalmente su creciente popularidad a la versatilidad de sus estructuras y algoritmos, que además de ser sencillos y fáciles de implementar, permiten abordar problemas complejos, y proporcionan en muchos casos soluciones adecuadas a la vez que robustas a dichos problemas.

1.8.8.2 Desarrollo de las ANN

La historia de las redes neuronales artificiales puede remontarse a los primeros trabajos de modelación del comportamiento de una neurona. La primera neurona artificial se debe al trabajo pionero de McCulloch y Pitts (1943). El modelo que propusieron, consistía en dos entradas y una única salida binaria. Mientras las entradas no alcanzaran un determinado umbral, la salida permanecería nula.

Pocos años después, Hebb (1949) propuso su postulado del aprendizaje, según el cuál la conexión entre dos neuronas se refuerza cada vez que es usada, es decir, cuando las neuronas que une dicha conexión están activas simultáneamente. Posteriormente, Rosenblatt (1958) introduce la red de una sola capa, el perceptrón, y demostró que era capaz de clasificar una entrada de valores continuos en dos clases. El perceptrón calculaba una suma ponderada de las entradas, sustraía el umbral, y daba uno de los dos posibles valores como salida. Este esquema podía reconocer patrones distintos a los utilizados en la etapa de aprendizaje, es decir, era capaz de generalizar.

Widrow y Hoff (1960) desarrollaron e implementaron un método matemático (regla delta) para actualizar los pesos de las conexiones. Era un método de búsqueda del gradiente basado en minimizar los errores al cuadrado de la variable de salida. El resultado de este trabajo pionero, el modelo ADALINE, se convirtió en la primera red neuronal artificial aplicada a un problema real: la eliminación de ecos en las señales telefónicas mediante un filtro adaptativo.

Casi una década después, Minsky y Papert (1969) pusieron de manifiesto la gran limitación del perceptrón simple, y que se debe a que sólo puede resolver problemas linealmente separables, es decir, en los que las regiones de decisión estén separadas por una recta, o en el caso más general por un hiperplano. Esto supuso una brusca desaceleración del gran auge que estaba experimentando la investigación sobre las redes neuronales, cayendo en el olvido hasta prácticamente la década de los ochenta, cuando empiezan a proponerse soluciones a los problemas planteados por estos autores.

Hopfield (1982) presenta una nueva forma de realizar los cálculos mediante redes recurrentes y demuestra con rigor matemático las capacidades de este tipo de redes, así como su funcionamiento. En ese mismo año Kohonen (1982) presenta un modelo de red neuronal capaz de formar mapas topológicos de características a partir de los datos de entrada, de manera análoga a como ocurre en el cerebro humano. La principal aplicación de este algoritmo es el agrupamiento de los datos, obteniendo como resultado una representación gráfica bidimensional del espacio de entrada que es fácil de visualizar.

Un gran impulso a la investigación, desarrollo y aplicación de las redes neuronales fue debido al redescubrimiento del algoritmo de retropropagación del error para perceptrones multicapa por Rumelhart *et al.* (1986). El algoritmo es una generalización del desarrollado por Widrow y Hoff, que permite a los perceptrones ser entrenados en una configuración multicapa. A partir de entonces, se convierte en la técnica más popular para el entrenamiento de redes perceptrón multicapa.

Entre 1986 y 1995 se produce un gran apogeo, con multitud de publicaciones en revistas, congresos y libros donde se muestran innumerables aplicaciones de las ANNs. A partir de 1996 tiene lugar una consolidación con la publicación de importantes resultados, entre los que cabe destacar a Bishop (1996), Rippley (1996) y Devroye (1996). Actualmente se avanza en la investigación de nuevos modelos de redes neuronales, en nuevas aplicaciones de los modelos existentes, en nuevos algoritmos de entrenamiento y en la búsqueda de metodologías que permitan determinar la topología de la ANN más adecuada para un problema dado.

1.8.8.3 Conceptos básicos

Una red neuronal artificial es un sistema constituido por muchas unidades básicas, elementos simples que se denominan neuronas o nodos, que trabajan en paralelo y que están interconectados masivamente. Tras el procesamiento de la información externa que reciben, ofrecen una respuesta. La capacidad de procesamiento de la red reside en la fuerza de las interconexiones, o pesos, que se obtienen mediante un

proceso de aprendizaje a partir de una serie de patrones conocidos. Es decir, la ANN aprende a partir de la experiencia, almacena el conocimiento y después lo utiliza cuando se enfrenta a nuevas situaciones. En este sentido, según Haykin (1999), las redes neuronales artificiales se parecen al cerebro humano en dos aspectos:

- Adquieren el conocimiento mediante un proceso de aprendizaje.
- La fuerza de las conexiones entre las neuronas, también conocido como pesos sinápticos, se usan para almacenar el conocimiento.

Las redes neuronales artificiales son una representación muy simplificada del funcionamiento de las redes neuronales del cerebro humano y, pese a ser simplificaciones, son muy eficaces en la práctica. Algunas de las aplicaciones más habituales son: clasificación o reconocimiento de patrones, reducción del ruido en señales, predicción y control.

Se ha demostrado teóricamente que una red neuronal de tres capas alimentada hacia delante (*feedforward*), es un sistema que se comporta como un aproximador universal de funciones (Hornik *et al.*, 1989). Este importante resultado matemático establece que cualquier función continua definida en un dominio compacto, puede ser aproximada con el grado de precisión deseado por una red neuronal de tres capas sin más que incluir un número suficiente de neuronas.

La estructura básica de una red neuronal artificial está constituida por las neuronas, su estado de activación, las funciones de activación, las conexiones sinápticas y el umbral de activación.

1.8.8.3.1 Neuronas

Las neuronas son los nodos o unidades elementales que reciben las señales procedentes de otras neuronas o del exterior. Tras el procesamiento de toda la información que recibe, ofrece un valor de salida que será enviado a otras neuronas

o al exterior, dependiendo de su posición relativa en la ANN. Las neuronas se encuentran agrupadas en capas, y según la ubicación de la capa se puede distinguir entre:

- **Capa de entrada:** las neuronas que la forman son las que reciben la señal del exterior. Este tipo de neuronas no realiza ningún tipo de procesamiento, es decir, la salida que ofrece es igual a la entrada que recibe, siendo, por tanto, su misión enviar su entrada a las neuronas de la siguiente capa.
- **Capas ocultas:** están formadas por neuronas cuyas entradas y salidas están en el interior de la propia ANN.
- **Capa de salida:** en la que se encuentran las neuronas que emiten la señal al exterior, siendo, por tanto, estos valores de salida los correspondientes a la función que se desea reproducir.

1.8.8.3.2 Estado de activación

Se denomina estado de activación al preprocesamiento de todas las señales que recibe una neurona. También se le denomina entrada neta, potencial postsináptico o campo local inducido. Es la agregación ponderada de todas las señales individuales que recibe de otras neuronas y que, por tanto, representa el nivel de actividad de dicha neurona.

1.8.8.3.3 Funciones de activación

La función de activación, también denominada función de transferencia, se encarga de convertir el estado de activación que tiene la neurona en un instante de tiempo determinado, en la señal de salida. Las funciones habitualmente empleadas son: escalón o umbral, lineal, sigmoideal (logística, tangente-hiperbólica y arcotangente), gaussiana, polinómica, y combinaciones de ellas.

1.8.8.3.4 Conexiones sinápticas

Se denominan conexiones sinápticas a las uniones entre las neuronas, a través de las cuales se transmite la información de unas a otras. Cada conexión tiene una intensidad, fuerza o peso, que es utilizado para almacenar el conocimiento experiencial adquirido durante el proceso de aprendizaje. La forma en que las neuronas estén interconectadas define el tipo de arquitectura o topología de la ANN.

Cuando ninguna de las neuronas de la ANN envía su respuesta a neuronas que están en la misma capa ni en capas precedentes, se trata de una red alimentada hacia delante (*feedforward*). En ellas el flujo de la información es unidireccional, desde la capa de entrada hacia la de salida. Mientras que cuando la salida de una neurona es entrada de sí misma, de otras neuronas que estén en la misma capa o de neuronas en capas precedentes, se trata de una red recurrente.

1.8.8.3.5 Aprendizaje

El aprendizaje o entrenamiento de una ANN es el procedimiento por el cual ésta extrae el conocimiento de la información disponible mediante la modificación de los pesos de sus conexiones. Para que la ANN sea capaz de encontrar las relaciones existentes entre las variables, o de identificar los patrones que siguen los datos, es necesario que éstos se le muestren de forma reiterada. Los tipos de aprendizaje se podrían clasificar de la siguiente manera:

- **Aprendizaje supervisado:** en este caso durante el proceso de aprendizaje interviene un agente externo (supervisor) que comprueba la salida de la ANN y en caso de que no coincida con la deseada, modifica los parámetros de la ANN por medio de algún algoritmo adaptativo. Los tres aprendizajes más importantes de este tipo se describen brevemente a continuación:

Aprendizaje por corrección del error. Consiste en modificar los pesos en función de la diferencia entre los valores deseados y los valores de salida

de la ANN, es decir, en función del error cometido por ésta. El algoritmo más empleado de este tipo es el de retropropagación del error.

Aprendizaje por refuerzo. En este caso no se muestra a la ANN la salida deseada, sino que el agente externo le indica a la ANN si la salida que está ofreciendo es adecuada (bien=1) o si no lo es (mal=-1). A partir de esta información se ajustan los pesos según un sistema de probabilidades. Como ejemplos de este tipo de aprendizaje destacarían el algoritmo lineal y el asociativo con recompensa y penalización.

Aprendizaje estocástico. Consiste en hacer modificaciones aleatorias en los valores de los pesos y evaluar su efecto a partir del objetivo deseado y de distribuciones de probabilidad preestablecidas.

- **Aprendizaje no supervisado.** En este caso, la ANN debe ser capaz de encontrar las relaciones internas que presentan los datos sin que se le indique si su respuesta es o no correcta. Por este motivo, se dice que este tipo de redes son capaces de auto-organizarse. Los principales aprendizajes de este tipo son dos:

Aprendizaje hebbiano. El ajuste del peso de cada conexión se realiza según la correlación de los estados de activación (o señales de salida) correspondientes a las dos neuronas conectadas, es decir, una conexión se refuerza cuando las neuronas que une están activas simultáneamente, mientras que si una está activa pero no la otra, la conexión se debilita.

Aprendizaje competitivo. En este caso, las neuronas de la capa de salida compiten por activarse, de forma que sólo una neurona puede estar activa en un instante dado. Es un tipo de aprendizaje que permite descubrir ciertas características en los datos y, por tanto, es útil en tareas de clasificación.

1.8.8.4 Modelo matemático de una neurona artificial

El modelo matemático de una neurona artificial es la base de la estructura de cualquier tipo de red neuronal artificial. Dicho modelo trata de imitar el flujo de información que tiene lugar en una neurona del cerebro. Como se ha comentado previamente, se trata de una gran simplificación del proceso real que, sin embargo, ha dado muy buenos resultados en la práctica. En la Figura 1.23 se muestra el esquema de dicho modelo, que estaría constituido por la información de entrada, su procesamiento y la respuesta de la neurona.

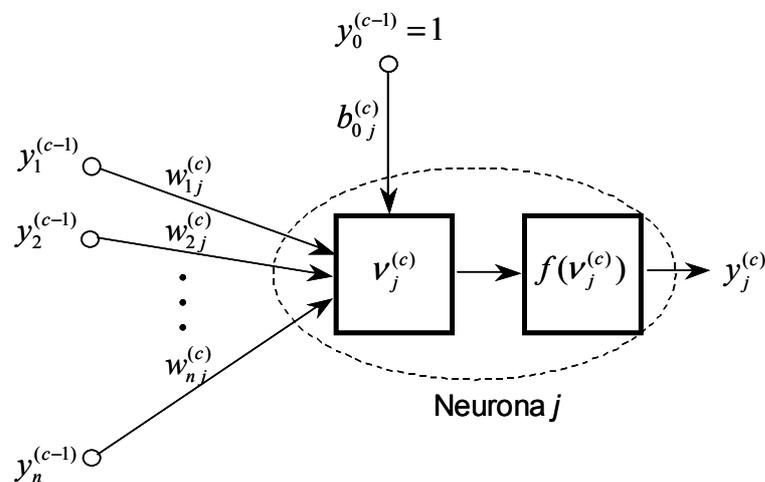


Figura 1.23. Modelo matemático de una neurona artificial.

Después de que las neuronas de una determinada capa de la ANN (p.ej., capa $c-1$) producen su respuesta, $y_i^{(c-1)}$, cada una de éstas es ponderada por las intensidades de sus respectivas conexiones con la neurona j a la que llegan dichas respuestas $w_{ij}^{(c)}$. Las respuestas ponderadas son agregadas para formar la entrada neta a la neurona, $v_j^{(c)}$, incluyendo también en dicho proceso el umbral de activación de dicha neurona, $b_{0j}^{(c)}$.

La entrada neta a la neurona j se obtendría mediante la siguiente expresión:

$$v_j^{(c)} = \sum_{i=1}^n w_{ij}^{(c)} y_i^{(c-1)} + b_{0j}^{(c)} \quad \text{Ec. 1.24}$$

En la ecuación anterior, el umbral de activación, $b_{oj}^{(c)}$, se puede asimilar al peso correspondiente a una neurona ficticia (neurona 0) que emite siempre una salida unitaria. De manera que se puede expresar el cálculo de la entrada neta de la siguiente forma:

$$v_j^{(c)} = \sum_{i=0}^n w_{ij}^{(c)} y_i^{(c-1)} \quad \text{Ec. 1.25}$$

Una vez evaluada la entrada neta, la función de activación de la neurona j se encarga de transformar dicha entrada en una señal de salida:

$$y_j^{(c)} = f(v_j^{(c)}) \quad \text{Ec. 1.26}$$

la cual se transmite posteriormente a otras neuronas.

De manera cualitativa podemos decir que en el modelo descrito, una neurona lo que hace es recibir las salidas de otras neuronas, las pondera por las intensidades de sus pesos y, después, las agrega para determinar la entrada neta que tras pasar por la función de transferencia dará lugar a la salida de dicha neurona. Esta salida será enviada a otras neuronas. Este esquema de funcionamiento es la base de las ANNs.

1.8.8.5 Características de las ANNs

El buen comportamiento que presentan en muchas ocasiones las ANNs ha contribuido a su gran popularidad. Entre las ventajas que presenta este tipo de modelos, cabe destacar las siguientes:

- Tienen la capacidad para reconocer patrones en los datos y para encontrar relaciones entre las variables de entrada y de salida, sin tener que incorporar en su formulación las leyes físicas que rigen el proceso. Pueden modelar no sólo relaciones lineales, sino también relaciones no lineales, puesto que las funciones de activación pueden ser no lineales.

- Son estructuras muy versátiles que a partir del esquema básico pueden llegar a modelar procesos de gran complejidad. Una misma configuración o arquitectura puede ser entrenada para resolver una gran variedad de procesos de distinta naturaleza.
- Son capaces de tolerar ruido e incluso errores en los datos de entrada, y seguir respondiendo de manera razonable. Esto se debe a que en este tipo de modelos, el conocimiento está distribuido a lo largo de toda la red y almacenado en la topología y en los pesos de las conexiones.

En este tipo de modelos no todo son aspectos positivos, sino que también presentan inconvenientes, las principales críticas a las que se enfrentan las ANNs son (Haykin, 1999):

- La ausencia de una metodología general que de forma rigurosa permita determinar la topología óptima de la red para un problema dado, el algoritmo de entrenamiento más adecuado y su correspondiente criterio de parada.
- La ausencia de base física en las relaciones matemáticas y en los pesos de la red.

Otros autores (Dayal *et al.*, 1994) han señalado como una desventaja muy importante de las ANNs frente a otro tipo de modelos predictivos como los modelos PLS, la dificultad en la interpretación del modelo obtenido, lo que impide su utilización para poder obtener un mayor conocimiento del proceso estudiado. Además, estos autores indican que con la utilización de ANNs se corre un elevado riesgo de obtener modelos inferenciales sobreparametrizados.

En la actualidad, la manera de determinar la topología de la ANN más adecuada para un problema dado, consiste en un procedimiento iterativo en el que se prueban multitud de topologías y de algoritmos de entrenamiento y, finalmente, se adopta aquella que mejor funciona para los datos de que se disponen. Se han desarrollado

algunas técnicas como los algoritmos de poda y los algoritmos de construcción. Los primeros parten de una ANN sobredimensionada y van eliminando progresivamente las conexiones y nodos menos importantes. El segundo tipo de algoritmos, funciona a la inversa, es decir, partiendo de una ANN muy sencilla a la que se van añadiendo nodos de forma progresiva. Sin embargo, estas técnicas también acaban siendo procesos iterativos.

En lo que respecta a la ausencia de base física en la modelación de los procesos, más que un inconveniente puede verse como justamente lo contrario, ya que pueden representar relaciones complejas entre las variables que resultaría muy difícil hacerlo de manera explícita, convirtiéndose, por tanto, en una alternativa a los modelos que tienen una fundamentación teórica.

1.8.8.6 ANNs y modelos estadísticos

Se debe tener presente que las ANNs son en realidad modelos inferenciales (como los modelos de regresión clásicos), y que la ventaja que presentan es que constituyen una herramienta muy flexible con la que es posible construir de forma sencilla modelos inferenciales de gran complejidad. En diversos trabajos se ha demostrado que ciertos modelos de ANNs son equivalentes o muy similares a modelos estadísticos clásicos de regresión (Sarle, 1994; Warner y Misra, 1996). De hecho, las ANNs más utilizadas, denominadas redes perceptrón multicapa, no son más que modelos de regresión no lineal y modelos discriminantes que pueden ser implementados con cualquier programa estadístico convencional. Sin embargo, debido a que las ANNs han sido desarrolladas y utilizadas principalmente por ingenieros e informáticos, la terminología empleada es sustancialmente diferente a la utilizada en el campo de la estadística. En Sarle (1994) se presenta un glosario de las equivalencias en la terminología que se emplea en ambos campos. Un resumen de éstas se presenta en la Tabla 1.4.

Terminología de las ANNs	Terminología estadística
Entradas	Variables explicativas
Salidas	Variables respuesta
Entrenamiento o aprendizaje	Estimación de los parámetros del modelo
Función de error	Criterio de estimación
Patrón	Observación
Generalización	Interpolación y extrapolación

Tabla 1.4. Glosario de equivalencias entre la terminología empleada en el campo de las ANNs y la utilizada en el ámbito estadístico.

1.8.8.7 Modelos predictivos en los tratamientos de aguas residuales

Debido a que en los procesos de tratamiento de aguas residuales es difícil que muchos de los fenómenos que tienen lugar sean descritos por modelos mecanicistas, el uso de modelos empíricos constituye una alternativa atractiva. Dentro de estos modelos, las redes neuronales artificiales han recibido especial atención por parte de muchos investigadores porque permiten modelar fácilmente sistemas no lineales complejos, ya que la modelación matemática es reemplazada por una etapa de entrenamiento con los datos del proceso.

A continuación se presentan diversas técnicas y aplicaciones que han aparecido en la bibliografía con el objetivo de desarrollar *soft sensors* en el ámbito de los procesos de depuración de aguas residuales. Se puede observar que en la mayor parte de los casos se ha optado por el uso de redes neuronales artificiales, y que se ha dedicado más atención a los procesos en continuo que a los procesos por lotes.

Aarnio y Minkkinen (1986) construyen un modelo PLS para predecir la calidad del efluente de una EDAR. Construyen un único modelo con todas las variables de proceso y las variables respuesta (concentración de fósforo total, DQO y turbidez en el efluente). Además, construyen un modelo PLS-discriminante para determinar que variables estaban más relacionadas con los episodios de *bulking* filamentoso ocurridos en la EDAR.

Capodaglio *et al.* (1991) construyen y comparan modelos estocásticos y de redes neuronales para predecir el índice volumétrico del fango en una EDAR. El modelo de redes neuronales es el que presenta la mejor capacidad predictiva. Sin embargo, estos autores destacan que el análisis de los coeficientes de los modelos estocásticos permite obtener información importante acerca de las series analizadas y, por tanto, del proceso, mientras que el modelo de redes neuronales no permite extraer dicha información.

Önnerth *et al.* (1996) emplean en una EDAR un sistema general de control (STAR, *Superior Tuning and Reporting*) que incorpora para el control del caudal de recirculación un *soft sensor* que permite estimar la capacidad de desnitrificación en el reactor anóxico.

Carstensen *et al.* (1996) ilustran la aplicación de dos *soft sensors* en una EDAR. Uno consiste en un modelo híbrido cuyo propósito es determinar la carga de amonio que entra en el reactor aerobio a partir de las medidas de amonio en dicho reactor, y el otro se basa en utilizar un filtro Kalman para corregir el ruido que presenta la señal de la medida registrada en tiempo real de la concentración de sólidos suspendidos que se recirculan desde el decantador secundario.

Häck y Köhne (1996) desarrollan *soft sensors* basados en redes neuronales para estimar aquellos parámetros del proceso que temporalmente no están disponibles, bien porque se está realizando el trabajo de mantenimiento, bien porque se ha producido una rotura del sensor. Para esto aprovechan la correlación existente entre las variables respuesta y otras variables que son fácil y directamente medibles (conductividad, pH, turbidez,...). Previamente al desarrollo de la red neuronal, hacen un estudio de las correlaciones existentes entre las variables medidas y los parámetros de interés, para posteriormente usar como entrada de la red sólo aquellas variables que presentan una mayor correlación. Desarrollan cuatro modelos independientes, dos para estimar las concentraciones en el influente de DQO y de amonio, y otras dos para estimar las concentraciones de nitratos y de amonio en el reactor aerobio.

Zhao *et al.* (1997) plantean un modelo híbrido para modelar la dinámica de los nutrientes (P-PO₄ y N-NO_x) en un SBR. El modelo que desarrollan integra un modelo mecanicista simplificado y una red neuronal en paralelo para compensar los errores. De esta forma la complejidad del modelo mecanicista y su calibración se reducen notablemente, y la estructura en paralelo le permite tener en cuenta las perturbaciones externas (en la temperatura, pH,...).

Belanche *et al.* (1998) proponen la utilización de una red neuronal cuya estructura está compuesta por una capa oculta con neuronas heterogéneas y una capa de salida con neuronas convencionales. De esta forma, la entrada de la red puede incluir distintos tipos de elementos (reales, conjuntos *fuzzy*, datos faltantes,...). Construyen un modelo para predecir la DBO₅ en el efluente y otro para la DQO. En ambos casos obtienen mejores resultados que utilizando una red neuronal convencional (en la que no hay neuronas heterogéneas).

Mujunen *et al.* (1998) aplican modelos PLS para estimar la eficiencia del proceso de depuración de una EDAR. Construyen tres modelos PLS, uno para predecir las propiedades de sedimentación del fango, otro para estimar la eficiencia en la eliminación de DQO y el tercero para estimar la concentración de nitrógeno total en el agua efluente. Aplican una técnica de selección de variables, para determinar en cada caso las variables más adecuadas de entrada al modelo.

Zhu *et al.* (1998) aplican redes neuronales a los datos procedentes de lagunas de estabilización aireadas para predecir la DBO₅ en el efluente. Realizan un análisis de sensibilidad sobre una red neuronal alimentada hacia delante para conocer las variables más importantes y, posteriormente, desarrollan una red neuronal que tiene cuenta la dinámica del sistema. Proponen e implementan una estrategia de actualización del modelo basada en volver a entrenar la red cuando la media del error supere un determinado umbral.

Lee y Park (1999) emplean un *soft sensor* basado en una red neuronal para estimar la dinámica de los nutrientes ($P-PO_4$, $N-NO_3$, $N-NH_4$) en un reactor SBR. Debido a los cambios bruscos causados por la aireación al cambiar de etapa, estos autores proponen asignar redes neuronales independientes para la etapa anaerobia y aerobia, respectivamente.

Teppola *et al.* (1999b) proponen la utilización de un filtro Kalman para actualizar los coeficientes de distintos modelos predictivos (regresión lineal múltiple, PLS y PCR) y de esta manera, solucionar los problemas asociados a los cambios en el proceso, es decir, que los nuevos datos no se correspondan con los datos históricos usados en la construcción de los modelos predictivos.

Choi y Park (2001) desarrollan un *soft sensor* para estimar el nitrógeno Kjeldahl total (NKT) del agua influente aprovechando la correlación existente entre los parámetros de calidad medidos. Tras comparar diversas metodologías proponen una técnica híbrida que consiste en una etapa de preprocesamiento mediante análisis de componentes principales y posteriormente la predicción del NKT con una red neuronal. De esta manera, el preprocesamiento permite extraer la información contenida en las variables originales, eliminando el ruido, y la red neuronal se encarga de describir las no-linealidades de los procesos de tratamiento del agua residual.

Baeza *et al.* (2002) estiman la carga orgánica (DQO) en el influente de una EDAR por medio de ANNs. A partir de las estimaciones realizadas y mediante un algoritmo de control, modifican automáticamente el volumen total aerobio de la EDAR, dando lugar a diferentes modos de operación. Con la estrategia de control que proponen, consiguen mejorar la eficiencia en la eliminación de nitrógeno de la EDAR.

El-Din y Smith (2002) utilizan una red neuronal para construir un modelo que permite hacer predicciones a corto plazo del caudal de agua residual que entra a la EDAR, y aplican el método Box-Jenkins de la función de transferencia para describir y predecir la respuesta dinámica del decantador primario, en concreto la concentración de sólidos suspendidos y de DQO a la salida del mismo.

Jansson *et al.* (2002) plantean el uso de *soft sensor* para el control de la dosificación de reactivos en un proceso de precipitación química. La finalidad del modelo es la predicción del fósforo total y del fósforo soluble a la entrada de la EDAR. Los autores comparan distintas técnicas estadísticas multivariantes, y concluyen que para sus datos el modelo PLS es el que mejores resultados consigue, y que los menores errores se obtienen para las predicciones del fósforo soluble.

Lee *et al.* (2002) comparan distintos enfoques para predecir simultáneamente la concentración de sólidos suspendidos en el reactor y en el efluente, y la DQO en el efluente. Construyen un modelo mecanicista simplificado, una red neuronal, y combinan los dos anteriores tanto en serie como en paralelo. Obtienen los mejores resultados (desde el punto de vista de la capacidad predictiva) con la red neuronal y con el enfoque híbrido (en paralelo). Sin embargo, es este último enfoque el que proponen debido a que es el que presenta una mejor capacidad de extrapolación.

Luccarini *et al.* (2002) desarrollan cinco redes neuronales recurrentes para estimar la dinámica de los nutrientes ($P-PO_4$, $N-NO_3$, $N-NH_4$) en un reactor SBR. Todas las ANN tienen dos variables de entrada, el pH y el potencial redox, y una única variable respuesta. Utilizan los valores de las variables corregidos respecto de su valor inicial. Concluyen que las ANNs resultantes son efectivas para predecir las tendencias de las variables respuestas, pero no sus concentraciones.

Hong y Bhamidimarri (2003) proponen una nueva aproximación a la modelación dinámica de una EDAR, basada en algoritmos genéticos, que consiste en la creación automática de muchos modelos, donde evolucionan y permanecen los mejores. Desarrollan un modelo para predecir la concentración de sólidos suspendidos en el reactor, y otro para los sólidos suspendidos que se escapan por el efluente. Comparan los resultados con los obtenidos por un lado mediante redes neuronales, y por otro con un modelo mecanicista. Las predicciones del algoritmo genético son ligeramente mejores que las obtenidas mediante redes neuronales, aunque las ventajas en las que más hincapié hacen estos autores son la discriminación automática de las variables de entrada más importantes y la estructura interpretable del modelo.

Hamed *et al.* (2004) construyen 2 modelos de redes neuronales alimentadas hacia delante, una para predecir la DBO_5 en el efluente, y la otra para predecir la concentración de sólidos suspendidos en efluente. Previamente al desarrollo de ambos modelos, realizan un análisis exploratorio para identificar cuales son las variables explicativas más adecuadas. Con ambos modelos obtienen buenos resultados.

OBJETIVOS

2 OBJETIVOS

El objetivo principal de la presente tesis ha sido estudiar las posibilidades de utilización de las variables de fácil medición, registradas por los sensores instalados en un reactor discontinuo secuencial operado para la eliminación biológica de fósforo, para obtener, mediante la aplicación de métodos estadísticos multivariantes, información sobre la evolución de las variables de calidad, así como sobre la evolución del proceso a lo largo del tiempo. De esta manera, se pretende detectar cualquier anomalía que pueda ocurrir (causa especial de variabilidad que puede provocar un funcionamiento inaceptable o inusual del proceso y/o un efluente de baja calidad) y ayudar en la identificación de su origen, evitando la utilización de sondas de medida de las variables de calidad cuyo elevado coste hace inviable su disponibilidad en la mayor parte de las EDARs existentes.

Este objetivo general se divide en los siguientes objetivos específicos:

- Estudiar el estado del arte de las técnicas estadísticas multivariantes empleadas para la monitorización y el diagnóstico de fallos en los procesos de tratamiento de aguas residuales.
- Analizar la aplicabilidad y capacidad de distintos enfoques estadísticos para la monitorización, tanto a posteriori como en tiempo real, y el diagnóstico de un SBR operado para la eliminación biológica de fósforo en las aguas residuales, comparando la efectividad de los mismos.
- Proponer modificaciones de los enfoques investigados para adaptarlos a las características del proceso estudiado cuando éstos no presenten resultados satisfactorios y comparar los resultados con respecto al enfoque original.
- Construir, validar y comparar diversos modelos empíricos que sean capaces de aprovechar la información contenida en las variables registradas en tiempo real por medio de los sensores instalados en el reactor SBR para predecir la

evolución de las variables de calidad, determinadas en el laboratorio, a lo largo de los lotes. En concreto comparar modelos predictivos basados en variables latentes con modelos de redes neuronales artificiales.

- Desarrollar un sistema para la monitorización, diagnóstico y predicción en tiempo real de un reactor discontinuo secuencial operado para la eliminación biológica de fósforo del agua residual, mediante la utilización de métodos de control estadístico multivariante de procesos basados en técnicas de proyección sobre estructuras latentes.

El trabajo planteado en la presente tesis supone el inicio de una nueva línea de investigación en el grupo CALAGUA (grupo de investigación en calidad de aguas) en colaboración con el grupo GIEM (grupo de investigación en ingeniería estadística multivariante) con la que se pretende contribuir a la expansión de la utilización de estas técnicas de análisis en el campo del tratamiento de las aguas residuales.

MATERIALES Y MÉTODOS

3 MATERIALES Y MÉTODOS

3.1 DESCRIPCIÓN DEL MONTAJE EXPERIMENTAL

Los datos experimentales del proceso de eliminación biológica de fósforo que se han utilizado en la presente tesis se han obtenido en laboratorio. Para el estudio del proceso se realizó el montaje de un reactor discontinuo secuencial (SBR). Este reactor fue instalado en el interior de una cámara termostatzada en la que se mantuvo una temperatura de 20 °C. En el Anejo 8.1 se muestran dos fotografías del montaje experimental realizado.

Las características del reactor utilizado se muestran en la Tabla 3.1. El reactor estaba equipado con un agitador mecánico de pala ancha y con difusores cerámicos que estaban conectados a un compresor. La agitación se realizaba a una velocidad suficiente como para asegurar la mezcla completa en el interior del reactor, pero sin ser demasiado elevada para no favorecer la reaireación superficial durante la etapa anaerobia.

Forma	Cilíndrica
Material	Metacrilato
Diámetro	20 cm
Altura total	24 cm
Volumen útil	7 l
Volumen fango tras vaciado	3.4 l

Tabla 3.1. Características del reactor utilizado.

En el reactor se instalaron sondas para la medición en tiempo real de la concentración de oxígeno disuelto, del pH, de la conductividad, del potencial redox y de la temperatura. Todas las sondas estaban conectadas, a través de los correspondientes convertidores de señal, a un ordenador personal en el quedaba registrada la evolución a lo largo del tiempo de estas variables en cada lote. Para evitar posibles interferencias en los registros ocasionadas por otros aparatos de laboratorio, se instaló un filtro. En el ordenador se implementó un sistema de control

on-off para mantener la concentración de oxígeno disuelto en torno a un valor deseado. Además de este control, el sistema estaba automatizado mediante un temporizador en el que se establecían las duraciones de cada una de las etapas del lote para que se realizaran las correspondientes acciones sobre la agitación, la aireación, la alimentación del reactor y la válvula del efluente. En la Figura 3.1 se muestra un esquema del reactor así como de su equipamiento.

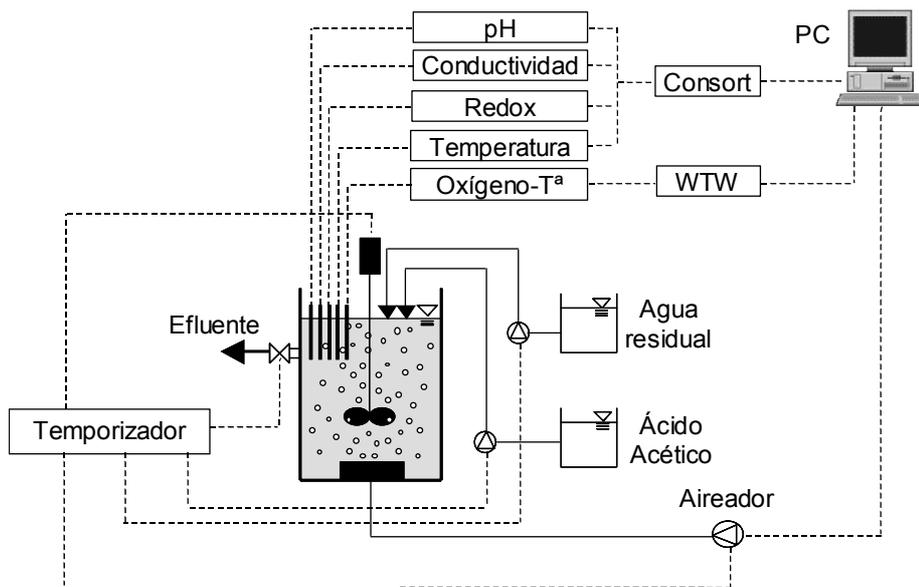


Figura 3.1. Montaje experimental.

El reactor fue operado mediante la alternancia de condiciones anaerobias y aerobias para favorecer el desarrollo de las bacterias acumuladoras de polifosfatos y, de esta manera, poder eliminar el fósforo del agua residual por vía biológica. Se estableció la duración de cada lote en 6 horas y, por tanto, se realizaban 4 lotes cada día. Las condiciones de operación y duración de cada una de las etapas se muestran en la Figura 3.2.

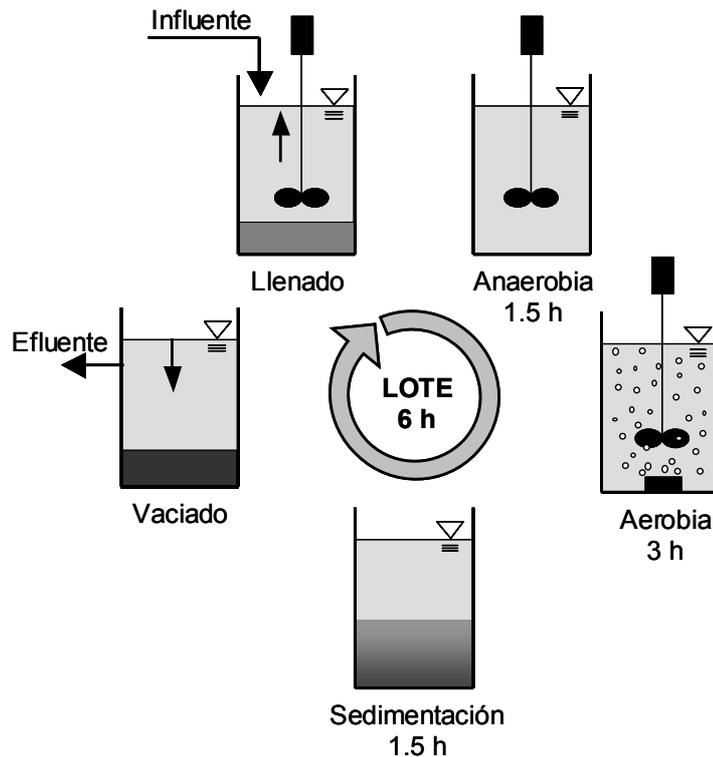


Figura 3.2. Etapas de las que consta cada lote.

Cuando se iniciaba un nuevo lote, durante los primeros minutos se realizaba el llenado del reactor con agua residual y ácido acético. Tras el llenado el reactor se dejaba en condiciones anaerobias durante 1.5 horas. En la siguiente etapa se aireaba el reactor durante 3 horas. El sistema de aireación estaba regulado por medio de un control *on-off* que permitía mantener la concentración de oxígeno disuelto en el reactor en torno a 3 mg/l. Una vez concluida la etapa aerobia se detenía tanto la aireación como la agitación y se dejaba sedimentar el fango durante 1.5 horas. Finalmente, el sobrenadante era retirado durante los últimos minutos del lote, permitiendo de esta manera el inicio de un nuevo lote.

El volumen que se retiraba al final de cada lote y que era reemplazado por el influente para comenzar un nuevo lote determinaba el tiempo de retención hidráulico. Por otro lado, para fijar el tiempo de retención celular se retiraba diariamente del sistema una determinada cantidad de fango. Este proceso se realizaba de forma manual al final de la etapa aerobia.

3.2 CARACTERIZACIÓN DEL INFLUENTE

Durante todo el período de operación se utilizó un agua residual artificial que se preparaba en el laboratorio. El influente al reactor estaba constituido por dos corrientes que llegaban por separado al reactor. Una corriente estaba formada por una mezcla a partes iguales de agua desionizada con agua de la red de distribución de la ciudad. Los nutrientes inorgánicos necesarios para el crecimiento celular eran añadidos en dicha mezcla. La otra corriente estaba constituida por una disolución de ácido acético y amonio. La concentración de nitrógeno en el agua residual se ajustaba para asegurar que las necesidades de este elemento como nutriente quedaran satisfechas. Además, se procuraba que no hubiera exceso de nitrógeno para limitar el desarrollo de las bacterias autótrofas, evitando de esta manera que los procesos de nitrificación-desnitrificación se superpusiesen a los de eliminación biológica de fósforo. Las concentraciones de los componentes del agua residual artificial resultante de la mezcla de las dos corrientes se muestran en la Tabla 3.2.

Componente		Concentración (mg/l)
Sustrato	Acético (DQO)	100-260
	CaCl ₂	10
	NH ₄ Cl	30
	MgSO ₄ ·7H ₂ O	300
	K ₂ HPO ₄	75
	FeCl ₃ ·6H ₂ O	3.0
	H ₃ BO ₃	0.3
Sales	CuSO ₄ ·5H ₂ O	0.06
	KI	0.06
	MnCl ₂ ·4H ₂ O	0.24
	Na ₂ MoO ₄ ·2H ₂ O	0.12
	ZnSO ₄ ·7H ₂ O	0.24
	CoCl ₂ ·6H ₂ O	0.30

Tabla 3.2. Composición del agua residual.

La preparación de la disolución con sales se realizaba semanalmente y quedaba almacenada en un bidón desde el que posteriormente se bombeaba al reactor al inicio de cada nuevo ciclo. En el caso de la disolución que contenía el sustrato, la preparación se realizaba quincenalmente.

3.3 INFORMACIÓN RECOGIDA

Una vez realizado el montaje del reactor, para la puesta en marcha del proceso se utilizó fango procedente de la EDAR de Quart Benàger (Valencia).

Durante el período experimental los datos de las sondas instaladas en el reactor se almacenaban en el ordenador. Estos registros junto con las determinaciones analíticas que se realizaban con menor periodicidad, constituyeron la base del seguimiento que se llevaba a cabo para conocer cómo estaba evolucionando el proceso. En la Tabla 3.3 se muestran todas las determinaciones realizadas en el seguimiento, así como su frecuencia de muestreo.

Parámetro	Frecuencia
Conductividad	~ 1 min
pH	~ 1 min
Potencial redox	~ 1 min
Oxígeno disuelto	~ 1 min
Temperatura	~ 1 min
Sólidos Suspendidos Totales (SST)	variable
Sólidos Suspendidos Volátiles (SSV)	variable
DQO	variable
P soluble final anaerobio y efluente	variable
Ácidos volátiles final anaerobio	variable
N-NO ₃ inicio y final ciclo	variable
N-NH ₄ inicio ciclo	variable

Tabla 3.3. Variables utilizadas para realizar el seguimiento del proceso.

Adicionalmente, a lo largo del período experimental se realizaron una serie de ensayos exhaustivos bajo distintas condiciones cuando el sistema se encontraba en estado estacionario. Cada uno de estos experimentos se realizaba a lo largo de un determinado lote. Durante la etapa anaerobia y aerobia del lote en el que se realizaba el exhaustivo se tomaba muestra del reactor cada 15-20 minutos para

determinar la concentración de fósforo soluble y, además, en algunos exhaustivos de cada una de las muestras se analizaban también los siguientes parámetros: ácidos volátiles, alcalinidad, magnesio y potasio.

A partir de lo anterior, se puede deducir que la mayor parte de datos disponibles del sistema estudiado proceden de las sondas instaladas en el reactor. Como se indicó en el apartado 1.8.5.2, estos datos se estructuran en una matriz tridireccional constituida por los lotes, las variables y los instantes de tiempo. En este caso son 5 las variables de proceso registradas con una frecuencia de muestreo ligeramente superior a un minuto. Debido a que las sondas pueden no permanecer completamente sumergidas durante los últimos minutos de cada lote en los que se procede al vaciado del sobrenadante del reactor, así como durante los primeros minutos del siguiente lote en los que se realiza el llenado del reactor, estos datos han sido descartados. Por tanto, de cada variable de proceso se dispone de un total de 340 valores registrados a lo largo de cada lote, repartidos entre las tres etapas tal y como se muestra en la Tabla 3.4. Las trayectorias de estas variables de proceso a lo largo de un lote cuando el proceso de eliminación biológica de fósforo está funcionando normalmente se muestran a modo de ejemplo en la Figura 3.3.

	Duración (min)	Valores registrados
Anaerobia	90	85
Aerobia	180	173
Sedimentación	90	82
Lote	360	340

Tabla 3.4. Valores registrados por cada una de las sondas instaladas en el reactor en cada una de las etapas de operación a lo largo de un lote que han sido utilizados en el presente trabajo.

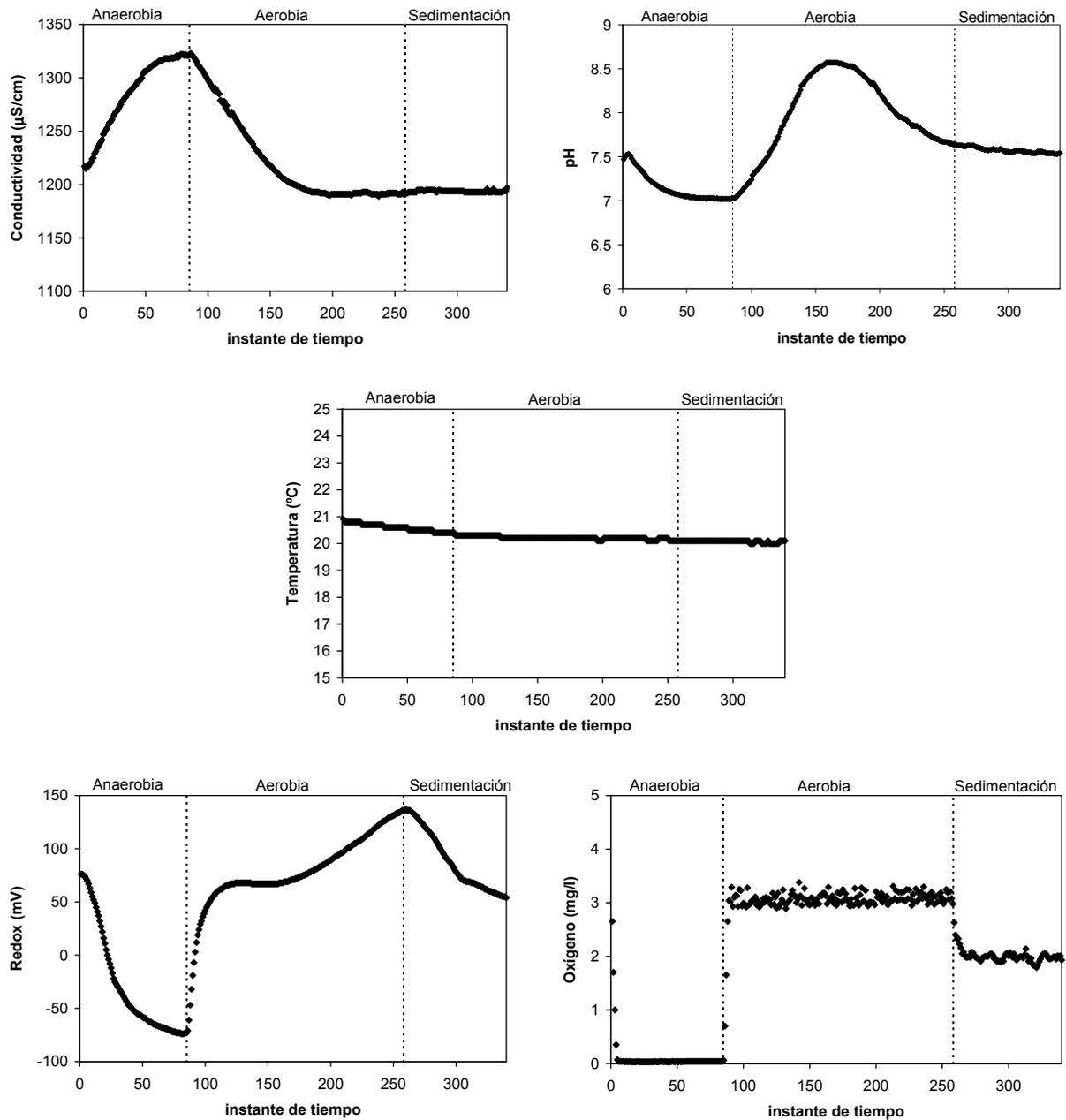


Figura 3.3. Trayectorias de las variables de proceso registradas a lo largo de un lote en el que el proceso de eliminación biológica de fósforo está funcionando normalmente.

A lo largo de la presente tesis se ha trabajado con bases de datos correspondientes a distintos periodos de operación y, por tanto, el número de lotes que se han empleado en cada caso es distinto. En cada apartado se hace una breve descripción del conjunto de datos con el que se trabaja, indicando las dimensiones de las matrices de datos que son analizadas.

3.4 MÉTODOS ANALÍTICOS

En las determinaciones analíticas que se realizaron para caracterizar el agua residual artificial, así como en los experimentos exhaustivos y los análisis rutinarios de seguimiento, se utilizaron varias técnicas de análisis.

La determinación de sólidos suspendidos totales (SST), sólidos suspendidos volátiles (SSV), demanda química de oxígeno (DQO), ortofosfatos, amonio y fósforo total se realizó siguiendo la metodología propuesta por la “*American Public Health Association*” (APHA, 1998).

Para la determinación de la concentración de nitratos se siguió el método fotométrico de la marca comercial *Merck*.

La determinación de los ácidos grasos volátiles de cadena corta y de la alcalinidad se realizó según el método de valoración ácido-base propuesto por Moosbrugger *et al.* (1992).

RESULTADOS Y DISCUSIÓN

4 RESULTADOS Y DISCUSIÓN

4.1 DIAGNÓSTICO DEL PROCESO

4.1.1 Análisis del rendimiento en la eliminación de fósforo

En este apartado se trabaja con una base de datos del proceso de eliminación biológica de fósforo correspondiente a dos meses de operación, desde que se inoculó el reactor con fango, hasta que se detuvo su funcionamiento. Debido a que durante este periodo el ordenador en el que se registraban los datos de las sondas se quedó bloqueado en algunas ocasiones, para el análisis se dispone de sólo de 188 de los 240 lotes realizados.

Durante el periodo de operación se determinó en algunos lotes la concentración de fósforo soluble tanto al comienzo como al final del lote. A partir de estas determinaciones se calculó el rendimiento en la eliminación de fósforo del sistema mediante la siguiente expresión:

$$r_{to} = \frac{P_0 - P_{eff}}{P_0} \quad \text{Ec. 4.1}$$

siendo:

- P_0 (mg P / l): la concentración de fósforo soluble en el reactor al inicio del lote.
- P_{eff} (mg P / l): la concentración de fósforo soluble en el efluente.

La evolución del rendimiento a lo largo del periodo de operación se muestra en la Figura 4.1. En esta figura se puede observar un aumento progresivo del rendimiento desde la puesta en marcha del sistema hasta que el proceso de eliminación está estabilizado, momento en el que se alcanzan rendimientos superiores al 98%. Posteriormente, a partir del lote 127 (18/06/04) el rendimiento empieza a disminuir y llega a tener valores inferiores al 60%. Esta situación de bajo rendimiento permanece durante aproximadamente 18 días, es decir, hasta el lote 198 (6/07/04). A partir de este momento el sistema evoluciona favorablemente y se vuelven a

obtener rendimientos elevados. Esta mejora en el proceso se produjo tras haber incrementado la concentración de amonio en el agua residual influente. Por lo tanto, el bajo rendimiento que experimentó el sistema fue debido muy probablemente a un proceso de inhibición en el crecimiento celular por falta de nitrógeno en el agua residual influente. Se debe tener en presente, tal y como se indicó en el apartado 3.2, que la concentración de nitrógeno en el agua residual se mantenía baja para evitar que tuvieran lugar procesos de nitrificación-desnitrificación.

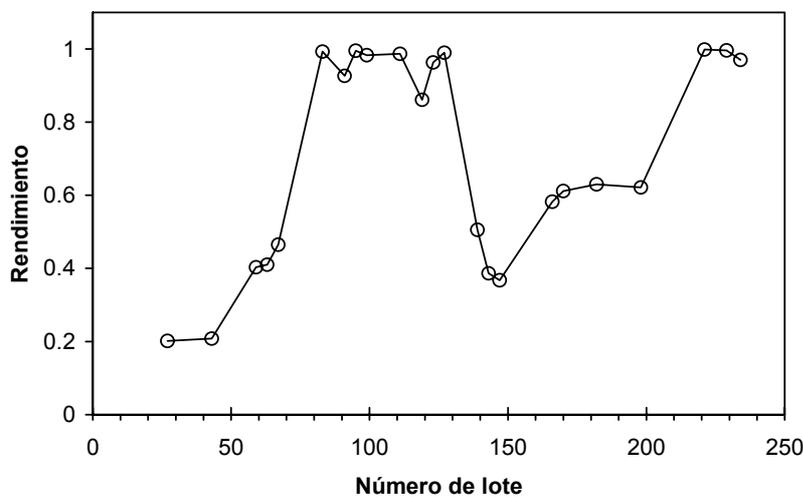


Figura 4.1. Evolución del rendimiento en la eliminación de fósforo a lo largo del período de operación.

En este apartado se pretende averiguar si alguna de las variables que se registran en tiempo real en el proceso también refleja cambios en los mismos instantes de tiempo que el rendimiento. Para ello se aplica una metodología de diagnóstico que ha sido recientemente propuesta por Zarzo y Ferrer (2004). Esta metodología está basada en la utilización de técnicas de proyección sobre estructuras latentes y regresión lineal clásica (*blockwise PCR*), junto con gráficos de sumas acumuladas (CUSUM).

Como ya se ha comentado en el apartado 3.3, aunque la operación del reactor SBR utilizado consta de cinco etapas, tan sólo son tres las etapas que tienen suficiente duración como para poder aprovechar la información contenida en las variables registradas en tiempo real. En cada una de estas tres etapas (anaerobia, aerobia y

sedimentación) tienen lugar procesos distintos y, por tanto, tiene sentido plantear que en lugar de trabajar con toda la matriz desdoblada (\mathbf{X}), se trabaje desglosando la matriz en tres bloques o submatrices correspondientes a cada una de las tres etapas del proceso.

La metodología *blockwise PCR* consiste en aplicar en primer lugar un PCA sobre cada una de las trayectorias de cada uno de los bloques en los que se ha desglosado la matriz \mathbf{X} . De cada uno de estos análisis se obtiene un cierto número de variables latentes significativas por validación cruzada. De esta manera, se transforma cada uno de los bloques de variables registradas en un nuevo bloque de variables latentes que, a pesar de tener una dimensión mucho más reducida, contiene gran parte de la información presente en el bloque de variables registradas. Posteriormente, se realiza un análisis de regresión simple entre cada variable latente y la variable respuesta, que en este caso es el rendimiento en la eliminación de fósforo. Por otra parte, se realiza el gráfico CUSUM de cada variable latente y se compara con el gráfico CUSUM de la variable respuesta. De esta forma pueden identificarse aquellas variables latentes que están más correlacionadas con la variable respuesta y que, además, presentan cambios de tendencia en los mismos instantes de tiempo que la variable respuesta.

Los cambios de tendencia identificados en el gráfico de la evolución del rendimiento en la eliminación de fósforo (Figura 4.1), también se pueden apreciar con claridad en su gráfico CUSUM (Figura 4.2). En ambos gráficos se puede identificar con cierta aproximación los instantes de tiempo en que dichos cambios tuvieron lugar. El primer cambio ocurre hacia el lote 67 (3/06/04), el segundo hacia el lote 127 (18/06/04) y el último en torno al lote 198 (6/07/04). De los tres cambios de tendencia, son el segundo y el tercero los que se corresponden con el inicio y el final, respectivamente, del periodo de bajo rendimiento que experimentó el sistema. El primer cambio se produjo cuando el proceso estaba evolucionando y no había llegado a estabilizarse todavía.

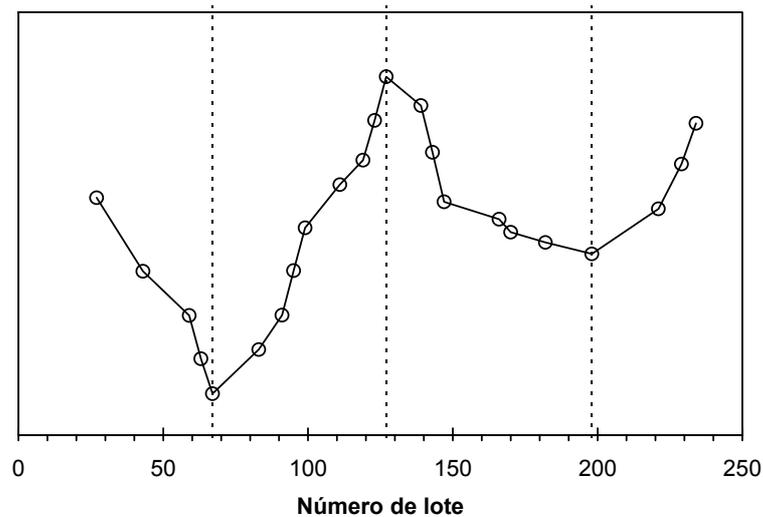


Figura 4.2. Gráfico CUSUM de la variable rendimiento en la eliminación de fósforo. Las líneas verticales indican los cambios de tendencia identificados. Por estar los valores tipificados se ha omitido el eje de ordenadas.

Por otro lado, para la aplicación de la metodología *blockwise PCR* se dispone de cinco variables registradas en tiempo real a lo largo de cada lote: pH, potencial redox, conductividad, concentración de oxígeno disuelto y temperatura. Sin embargo, se puede mejorar la capacidad de diagnóstico de las técnicas de análisis multivariante mediante la incorporación de información adicional del proceso. Ramaker *et al.* (2002) presentan una revisión de las formas de incorporar información adicional, también denominada información externa, en procesos por lotes. Diversos autores (Woodall, 2000; Yoon y MacGregor, 2001; Zarzo y Ferrer, 2004) han obtenido buenos resultados incluyendo información externa en los modelos multivariantes desarrollados.

En el caso de sistemas SBR para el tratamiento de aguas residuales, diversos estudios han mostrado la relación existente de los perfiles de la concentración de oxígeno disuelto, de conductividad, de pH y de potencial redox con la actividad de las bacterias en los procesos de eliminación biológica de nutrientes (Yu *et al.*, 1997; Kim y Hao, 2001; Serralta *et al.*, 2004). Además, las derivadas de dichas variables han sido utilizadas para detectar el final de los procesos de captura y de liberación de fósforo, de nitrificación y de desnitrificación.

La creación de nuevas variables obtenidas a partir de las variables registradas es una forma de incorporar información externa del proceso. En este caso se incorporaron siete trayectorias adicionales (Tabla 4.1) que, junto con las cinco trayectorias registradas, dan lugar a una matriz de datos del proceso (**X**) de 4080 variables (12 trayectorias x 340 instantes de tiempo).

Trayectorias registradas	Nuevas trayectorias
pH (pH).	- Incremento de pH respecto de su valor al comienzo del lote (Δ pH). - Primera derivada del pH (dpH).
Potencial Redox (Rx).	- Incremento de potencial redox respecto de su valor al comienzo del lote (Δ Rx). - Primera derivada del potencial redox (dRx).
Conductividad (Cond).	- Incremento de la conductividad respecto de su valor al comienzo del lote (Δ Cond). - Primera derivada de la conductividad. (dCond).
Concentración de oxígeno disuelto (Ox).	
Temperatura (Temp).	- Incremento de la temperatura respecto de su valor al comienzo del lote (Δ Temp).

Tabla 4.1. Trayectorias del proceso registradas y nuevas trayectorias que se han creado a partir de las anteriores. Todas las trayectorias constituyen la matriz **X** sobre la que se aplica la metodología *blockwise PCR*. Entre paréntesis se indica el código de la variable.

Las nuevas trayectorias transformadas que se han creado responden a un análisis de las trayectorias de las variables registradas (Figura 3.3) en los distintos lotes. De las variables que presentaban cambios de pendiente a lo largo del transcurso de cada lote (pH, Rx y Cond), se decidió incorporar su primera derivada. Por otra parte, de las variables cuya trayectoria estaba influida por el valor de la variable al comienzo del lote (pH, Rx, Cond y Temp), se incorporó una nueva trayectoria calculada como el valor registrado en cada instante de tiempo menos el valor registrado al inicio del lote.

Los resultados obtenidos tras realizar un PCA sobre cada una de las trayectorias en cada bloque se presentan en la Tabla 4.2. En esta tabla se indica la trayectoria y la etapa sobre la que se realiza el PCA, el número de variables (N° var.) de cada una de las trayectorias en cada etapa, el número de variables latentes (N° comp.) que se

obtienen por validación cruzada, la bondad de ajuste R^2_x (cum) y la bondad de predicción Q^2 (cum), en ambos casos se indica el valor acumulado (cum) para el número de variables latentes extraídas.

Trayectoria	Etapa	Nº var.	Nº comp.	R^2_x (cum)	Q^2 (cum)
pH	Anaerobia	85	4	0.996	0.995
	Aerobia	173	4	0.995	0.995
	Sedimentación	82	1	0.983	0.982
Cond	Anaerobia	85	2	0.998	0.998
	Aerobia	173	2	0.997	0.997
	Sedimentación	82	2	0.999	0.995
Rx	Anaerobia	85	3	0.998	0.998
	Aerobia	173	5	0.997	0.995
	Sedimentación	82	2	0.998	0.998
Ox	Anaerobia	85	4	0.483	0.028
	Aerobia	173	3	0.093	-0.007
	Sedimentación	82	7	0.901	0.863
Temp	Anaerobia	85	1	0.988	0.988
	Aerobia	173	2	0.994	0.994
	Sedimentación	82	1	0.991	0.991
Δ pH	Anaerobia	85	4	0.997	0.994
	Aerobia	173	3	0.994	0.994
	Sedimentación	82	1	0.990	0.990
dpH	Anaerobia	85	10	0.865	0.720
	Aerobia	173	7	0.934	0.924
	Sedimentación	82	10	0.742	0.369
Δ Cond	Anaerobia	85	4	0.994	0.989
	Aerobia	173	4	0.994	0.993
	Sedimentación	82	3	0.989	0.979
dCond	Anaerobia	85	8	0.949	0.909
	Aerobia	173	10	0.860	0.799
	Sedimentación	82	16	0.945	0.803

Trayectoria	Etapa	Nº var.	Nº comp.	R^2_X (cum)	Q^2 (cum)
ΔRx	Anaerobia	85	3	0.992	0.991
	Aerobia	173	4	0.995	0.994
	Sedimentación	82	2	0.995	0.995
dRx	Anaerobia	85	6	0.926	0.888
	Aerobia	173	12	0.916	0.860
	Sedimentación	82	12	0.913	0.784
$\Delta Temp$	Anaerobia	85	5	0.900	0.871
	Aerobia	173	2	0.977	0.976
	Sedimentación	82	2	0.993	0.993

Tabla 4.2. Resultados obtenidos tras aplicar un PCA sobre cada trayectoria en cada etapa.

Tras la realización de un total de 36 PCA (12 trayectorias x 3 etapas), la matriz de datos desdoblada (\mathbf{X}) que consta de 4080 variables se ha sintetizado en tan sólo 171 variables latentes. Esta transformación supone una considerable reducción con respecto al tamaño de la matriz original, sin embargo, ambas matrices contienen prácticamente la misma información.

Como se puede observar en la Tabla 4.2 en casi todos los casos la bondad de ajuste de los modelos PCA creados es muy elevada, lo cual está reflejando la alta autocorrelación existente en las trayectorias analizadas. Tan sólo los modelos desarrollados para la trayectoria de oxígeno en las etapas anaerobia y aerobia no son capaces de explicar la variabilidad de los datos, presentando valores extremadamente bajos tanto de R^2_X (cum) como de Q^2 (cum). Como se ha descrito anteriormente (apartado 3.1) el oxígeno es una variable que está regulada cuya concentración en el reactor durante la etapa anaerobia es prácticamente nula, mientras que cuando el proceso se encuentra en la etapa aerobia, la actuación del control *on-off* mantiene la concentración de oxígeno en torno a 3 mg/l. Los valores bajos de R^2_X (cum) y de Q^2 (cum) están poniendo de manifiesto que la trayectoria de oxígeno presenta una baja autocorrelación en las etapas anaerobia y aerobia.

Después de haber extraído las variables latentes del proceso, se realizó un análisis de regresión simple entre cada variable latente y el rendimiento en la eliminación de fósforo. De esta manera, se pueden identificar aquellas variables latentes que están más correlacionadas con la variable respuesta, puesto que presentarán un mayor valor del coeficiente de correlación al cuadrado (R^2_{rto}).

Por otra parte se calcularon las sumas acumuladas (CUSUM) de cada variable latente y se realizó un análisis de regresión simple para relacionar los valores CUSUM de cada variable latente con los valores CUSUM de la variable respuesta. En este caso valores elevados del coeficiente de correlación al cuadrado obtenido (R^2_{cusum}) están indicando un mayor parecido entre el gráfico CUSUM de la variable latente y el correspondiente a la variable respuesta.

En la Figura 4.3 se muestra la representación gráfica de ambos coeficientes de correlación (R^2_{rto} y R^2_{cusum}). Las variables latentes que aparezcan en el cuadrante superior derecho, serán las que estén más correlacionadas con la variable respuesta (el rendimiento en la eliminación de fósforo) y, además, presenten cambios de tendencia en instantes de tiempo cercanos a los cambios mostrados por la variable respuesta. Para poder identificar en el gráfico de qué variables se trata, se ha empleado una nomenclatura que está compuesta por el nombre de la trayectoria, la etapa a la que pertenece (1 anaerobia, 2 aerobia y 3 sedimentación), y entre corchetes la variable latente de que se trata. La Figura 4.3 muestra que tan sólo son dos las variables latentes que cumplen ambas condiciones: $\Delta\text{Cond } 2_t[1]$ (primera variable latente, obtenida de la trayectoria de los incrementos de conductividad en la etapa aerobia) y $\Delta\text{Cond } 3_t[1]$ (primera variable latente, obtenida de la trayectoria de los incrementos de conductividad en la etapa de sedimentación). Por otra parte, en la Figura 4.3 también se puede observar que la mayor parte de las variables latentes se encuentran en el cuadrante inferior izquierdo, lo que indica que están poco correlacionadas con la variable respuesta y que sus gráficos CUSUM no presentan una evolución parecida a la del gráfico CUSUM de la variable respuesta.

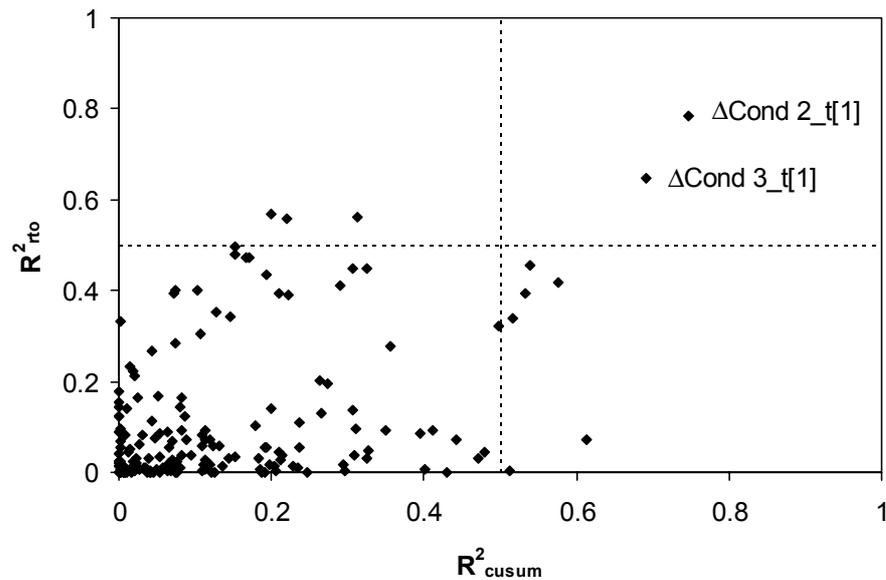


Figura 4.3. Relación entre los coeficientes de correlación al cuadrado obtenidos con cada variable latente y el rendimiento en la eliminación de fósforo (R^2_{rto}), y los obtenidos con sus valores acumulados (R^2_{cusum}).

En la Figura 4.4 se presentan los gráficos CUSUM del rendimiento en la eliminación de fósforo, así como de las dos variables latentes que mostraron un comportamiento más parecido. Para poder superponer en la misma figura varios gráficos CUSUM y de esta manera facilitar su comparación, se han utilizado unos factores correctores. Esto ha sido necesario porque el rango de variación de cada variable latente es diferente y, por tanto, la realización de sus correspondientes gráficos CUSUM da lugar a gráficos con distintas escalas. Al aplicar estos factores se ha evitado el solapamiento de los gráficos CUSUM (factor de traslación) y se ha conseguido compensar su diferente rango de variabilidad (factor de escala). En consecuencia el eje de ordenadas de los gráficos CUSUM que se presentan está distorsionado y por este motivo ha sido omitido. Se debe tener presente que lo que se pretende es comparar la forma de dichos gráficos y poder determinar con cierta aproximación los instantes de tiempo en que se producen los cambios, para lo cual no es necesario conocer en ningún momento la verdadera magnitud en el eje de ordenadas.

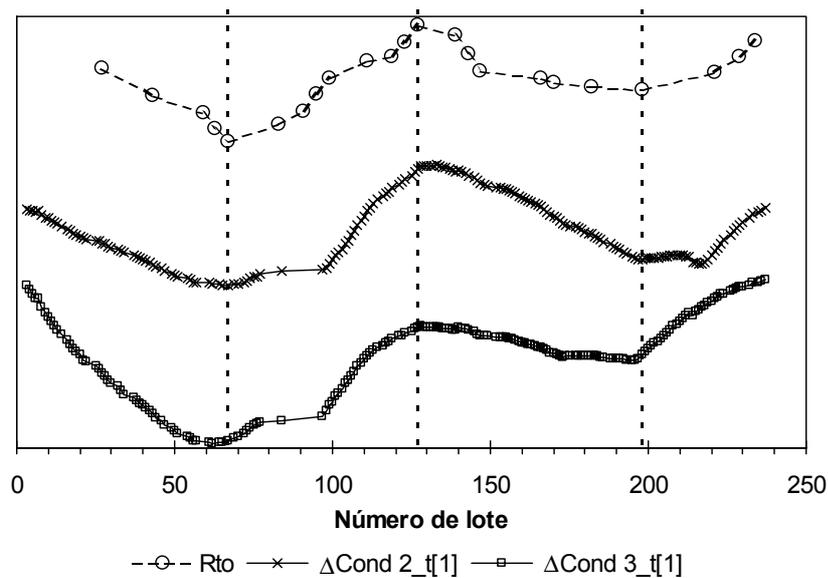


Figura 4.4. Gráficos CUSUM del rendimiento en la eliminación de fósforo (rto) y de las variables latentes con comportamiento similar. Las líneas verticales indican los cambios de tendencia identificados.

La Figura 4.4 permite confirmar que las dos variables latentes que habían sido identificadas en la Figura 4.3 manifiestan cambios de tendencia en instantes de tiempo próximos a los cambios del rendimiento en la eliminación de fósforo. Las dos variables latentes están asociadas a la trayectoria de los incrementos de conductividad ($\Delta\text{Cond } 2_t[1]$ y $\Delta\text{Cond } 3_t[1]$). Estos resultados están poniendo de manifiesto que una de las trayectorias que más información puede aportar respecto de los cambios en el rendimiento en la eliminación de fósforo del proceso en el sistema estudiado, consiste en una transformación de la trayectoria de conductividad respecto de su valor en el comienzo del lote.

Son varios los trabajos realizados en los que también se destaca la importancia que tiene la conductividad en el proceso de eliminación biológica de fósforo en un sistema SBR. En Serralta *et al.* (2004b) se analiza en profundidad la información que se puede extraer de los registros de conductividad y de pH en relación con el proceso tanto en condiciones estacionarias como transitorias. Un análisis de las relaciones entre concentraciones de iones en el sistema y conductividad eléctrica se discute en Maurer y Gujer (1995), así como en Aguado *et al.* (2004). Barat (2004) utiliza los valores puntuales de conductividad medidos al principio y al final de las

etapas anaerobia y aerobia respectivamente para simplificar las determinaciones analíticas a realizar en el seguimiento de la evolución del proceso.

Puede resultar curioso el hecho de que sean las variables latentes obtenidas de la trayectoria de los incrementos de conductividad en lugar de las extraídas de la propia trayectoria registrada los que presenten una mayor correlación con el rendimiento en la eliminación de fósforo. Este comportamiento podría ser debido a que el valor de la conductividad al inicio de cada lote presentaba una variabilidad importante y, además, unas variaciones muy bruscas cada vez que se cambiaba el agua residual influente al reactor. Se debe tener presente que en la preparación del agua residual sintética se utilizaba agua de la red de distribución de la ciudad. En la Figura 4.5 se muestra la evolución a lo largo del período analizado del valor de la conductividad al inicio de cada lote. Se indica, además, los lotes en los que se cambió el agua residual influente y los lotes en los que se procedió a calibrar las sondas instaladas en el reactor.

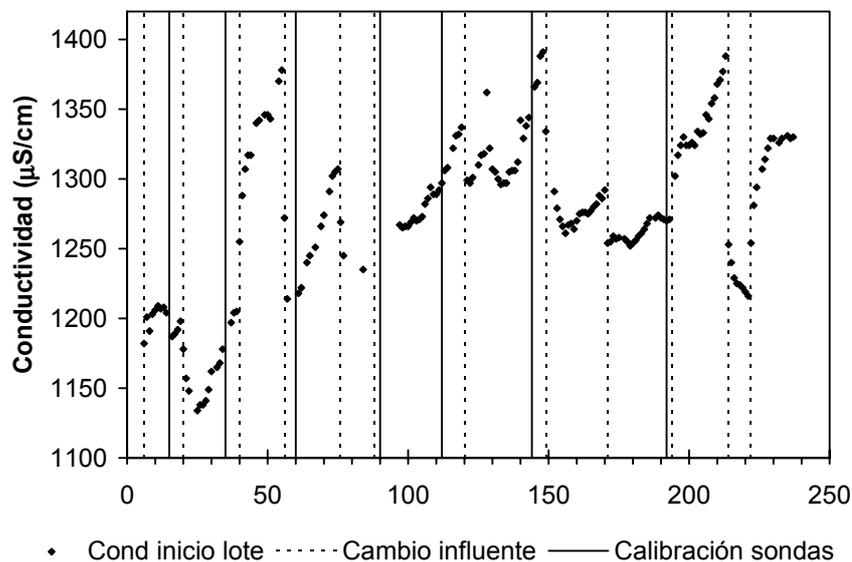


Figura 4.5. Evolución a lo largo del período de operación del valor de la conductividad al inicio de cada lote. En vertical se indican los momentos en los que se cambió el agua residual influente (línea discontinua) y en los que se procedió a calibrar las sondas (línea continua).

Para comprobar si es realmente importante la influencia que tiene el valor de la conductividad al inicio de cada lote en las variables latentes obtenidas al realizar un

PCA sobre la trayectoria de la conductividad en cada etapa del proceso, se estudiaron los valores de los pesos en la formación de la primera variable latente de cada etapa. Además, éstos se compararon con los valores de los pesos correspondientes a las trayectorias de los incrementos de conductividad en cada etapa.

En la Figura 4.6 se representan los pesos de las trayectorias de conductividad (Cond) y de los incrementos de conductividad (ΔCond) en la formación de la primera variable latente en cada una de las etapas del proceso (anaerobia, aerobia y sedimentación). Se puede observar que los pesos correspondientes a la trayectoria de conductividad son en cada etapa prácticamente constantes. Teniendo en cuenta que el vector de pesos es un vector unitario que define la dirección de máxima varianza en el espacio de las variables originales, cuando todos los pesos tienen el mismo valor, la variable latente obtenida está reflejando el valor medio de las variables originales.

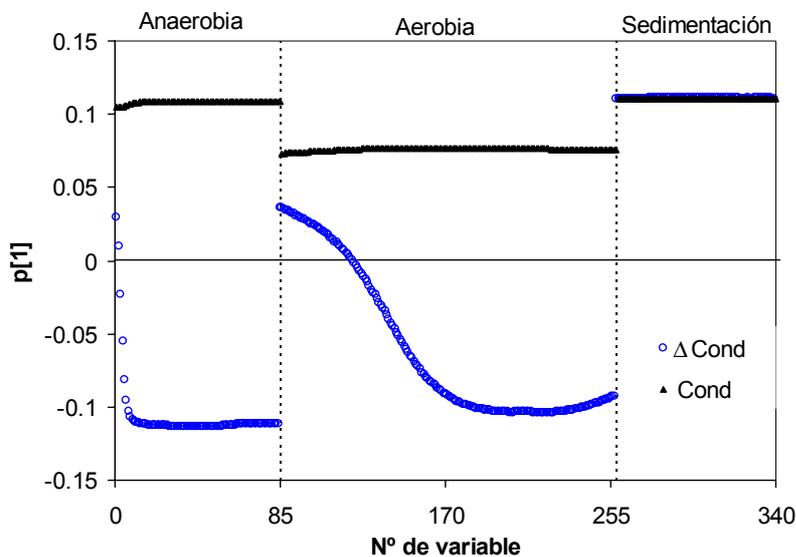


Figura 4.6. Pesos en la formación de la primera variable latente en cada una de las etapas del proceso de las trayectorias de conductividad (Cond) y de los incrementos de conductividad (ΔCond).

La relación que se puede establecer entre el nivel medio de la trayectoria de conductividad ($\overline{\text{Cond}}$) en cualquiera de las etapas y el valor que recoge la primera variable latente (t_1) se muestra en las siguientes ecuaciones:

$$\overline{Cond} = \frac{1}{J} \sum_{j=1}^J x_j \quad \text{Ec. 4.2}$$

$$\|\mathbf{p}_1\| = \sqrt{\sum_{j=1}^J p_{1j}^2} = 1 \rightarrow \text{Si } p_{1j} = cte \rightarrow \|\mathbf{p}_1\| = \sqrt{J p_{1j}^2} = 1 \rightarrow p_{1j} = \frac{1}{\sqrt{J}} \quad \text{Ec. 4.3}$$

$$t_1 = \sum_{j=1}^J p_{1j} x_j \rightarrow \text{Si } p_{1j} = cte \rightarrow t_1 = p_{1j} \sum_{j=1}^J x_j = \frac{1}{\sqrt{J}} \sum_{j=1}^J x_j \quad \text{Ec. 4.4}$$

donde x_j es el valor de la conductividad registrado en el instante j , J el número de valores de conductividad registrados en la etapa (cada valor registrado es una variable), \mathbf{p}_1 el vector de pesos que define la dirección de máxima variabilidad, p_{1j} es el valor del peso de la variable j en la obtención del valor de primera variable latente (t_1).

Como en el caso particular de la conductividad los valores al comienzo del lote oscilan entre 1134 y 1391 $\mu\text{S/cm}$ (Figura 4.5), mientras que la amplitud de la trayectoria (valor máximo de la trayectoria menos valor mínimo) lo hace entre 12 y 145 $\mu\text{S/cm}$, la primera variable latente de la trayectoria de conductividad estará principalmente determinada por el valor de la conductividad al comienzo del lote. Por otra parte, en el caso de la trayectoria de los incrementos de conductividad, aunque la primera variable latente de la etapa anaerobia y de sedimentación también están reflejando el nivel medio de los valores de la trayectoria en dichas etapas, este nivel medio no está afectado por los abruptos cambios en el valor de la conductividad al inicio del lote. Además, también se puede observar en la Figura 4.6 cómo en la etapa aerobia de esta trayectoria, la variable latente ya no refleja el nivel medio sino las diferencias entre los valores de las variables al comienzo de la etapa y los valores hacia el final de la misma.

De lo anterior se puede concluir que en la trayectoria de conductividad las variaciones debidas al distinto valor en el inicio del lote son más importantes que los cambios que se puedan presentar dentro de la propia trayectoria. Es decir, las variaciones más importantes se deben a desplazamientos en el eje de ordenadas de toda la trayectoria de conductividad, y esto está enmascarando los cambios debidos al proceso de captura y liberación de fósforo que se manifiestan como variaciones

en la amplitud de la trayectoria. Por tanto, en la aplicación de las técnicas de análisis multivariante será muy importante el preprocesamiento de las trayectorias originales para crear aquellas trayectorias que puedan contener la información más relevante del proceso.

4.1.2 Diagnóstico de fallos.

El objetivo de este apartado es realizar una comparación de la capacidad que presentan para el diagnóstico de fallos las dos metodologías de análisis multivariante basadas en el desplegado de los datos de mayor reconocimiento. Estas técnicas consisten en la aplicación de PCA a estructuras de datos n -dimensionales que previamente han sido desdobladas hasta obtener una matriz bidimensional.

Una alternativa se basa en el enfoque de Nomikos y MacGregor (1995), en la que en la estructura de datos se mantiene fija la dirección de los lotes, y se desdoblan las correspondientes a las variables y al tiempo. La segunda alternativa es la propuesta por Wold *et al.* (1998), donde la dirección que se fija es la de las variables, desdoblado los lotes y el tiempo. Las principales características de ambos enfoques fueron presentadas en el apartado 1.8.6.

La base de datos del proceso de eliminación biológica de fósforo con la que se trabaja es distinta a la utilizada en el apartado 4.1.1 y en este caso corresponde a cuatro meses de operación, desde la siembra del reactor hasta que se detuvo su funcionamiento. El motivo por el cual se decidió utilizar este conjunto de datos es el gran número de problemas que ocurrieron durante el tiempo que el sistema estuvo en funcionamiento.

La estructura tridireccional (**X**) de la base de datos correspondiente al período de operación que se analiza está constituida por un total de 489 lotes, 5 variables (pH, redox, conductividad, oxígeno y temperatura), y 340 instantes de tiempo. El análisis realizado en el apartado 4.1.1 demostró que en el sistema estudiado la trayectoria de los incrementos de conductividad respecto de su valor al inicio del lote (ΔCond)

puede aportar más información sobre el rendimiento en la eliminación de fósforo del proceso que la propia trayectoria original. Por este motivo, se decidió utilizar en este apartado la trayectoria corregida ΔCond , junto con las trayectorias registradas de pH, Rx, Ox y Temp (esta nomenclatura fue introducida en el apartado 4.1.1, Tabla 4.1).

A partir de una inspección detallada de la evolución de las trayectorias de las variables registradas en cada uno de los lotes de la base de datos analizada, se realizó una clasificación de los lotes en varias categorías. En la Tabla 4.3 se enumeran y describen cada una de ellas.

Clase	Descripción
1	Lotes buenos: para construir los modelos.
2	Fallo en la medida de oxígeno disuelto debido a la acumulación de burbujas de aire en la sonda de oxígeno.
3	Fallo del sistema de control de la aireación, que deja de airear en la etapa aerobia.
4	Problemas en el sistema de aireación para alcanzar la consigna de oxígeno.
5	Fallo en la agitación al inicio de la etapa anaerobia.
6	Interferencias en los registros (por ausencia de filtro).
7	No todo el ácido acético es consumido en la etapa anaerobia.
8	Lotes buenos: para validar los modelos.
9	El convertidor de señal <i>Consort</i> se queda bloqueado, lo que provoca un fallo de comunicación en la transmisión de los datos recogidos por los sensores de pH, conductividad y redox.

Tabla 4.3. Clasificación de los lotes.

De la base de datos disponible se seleccionaron 70 lotes cuya evolución no presentaba anomalías para la construcción de los modelos de referencia (son los pertenecientes a la clase 1), mientras que para comparar el funcionamiento de los métodos y validar los modelos desarrollados se utilizaron los lotes de las restantes categorías. Para ilustrar los resultados obtenidos se ha escogido un lote representativo de cada una de las categorías definidas en la Tabla 4.3.

4.1.2.1 Enfoque de Nomikos y MacGregor

Tal y como se indicó en el apartado 1.8.6.3.1, la metodología propuesta por estos autores (Nomikos y MacGregor, 1995) para el análisis de una base de datos de un proceso por lotes consiste en desdoblar la matriz de datos preservando la dirección de los lotes y, posteriormente, extraer la información mediante PCA. Cuando se modela el proceso de esta manera, un lote completo constituye una observación. En este apartado la metodología se aplica sobre lotes ya finalizados, es decir, se trata de un análisis de datos históricos (análisis *off-line*). Como se van a utilizar 70 lotes (clase 1) para construir el modelo, la matriz de datos desdoblada está constituida por 1700 variables (5 variables x 340 instantes de tiempo) y 70 lotes.

Las trayectorias de las variables se centraron y escalaron a varianza unitaria, y sobre los datos preprocesados se construyó el modelo PCA obteniendo las variables latentes significativas por validación cruzada y reteniendo aquellas con valor propio mayor que la unidad. En la Tabla 4.4 se presenta el modelo desarrollado que cuenta con 7 variables latentes, que es capaz de explicar el 82.5% de la variabilidad total de los datos y presenta una bondad de predicción (Q^2) del 76.5%.

Componente	R^2_X (cum)	Valor propio	Q^2 (cum)
1	0.351	24.6	0.331
2	0.539	13.2	0.507
3	0.628	6.22	0.584
4	0.707	5.49	0.653
5	0.759	3.63	0.697
6	0.802	3.03	0.741
7	0.825	1.65	0.765

Tabla 4.4. Características del modelo de referencia.

En la Figura 4.7 se representan en el eje horizontal las proyecciones (*scores*) de los lotes sobre la primera componente, y en el eje vertical sobre la segunda. En dicha figura se puede apreciar que las observaciones presentan una distribución homogénea en el plano creado por estas dos primeras componentes. Además, los gráficos de control para las 7 componentes del modelo, T^2 de Hotelling y distancia al modelo (D_{modX}), no indican ninguna anomalía en las observaciones utilizadas para su construcción (Figura 4.8 y Figura 4.9). Algunas observaciones exceden los límites

de control, sin embargo, esto no es motivo para alarmarse ya que los límites tan sólo se rebasan ligeramente y el número de observaciones que lo hacen, está dentro de lo esperable para los límites de control con los que se está trabajando del 95% y 99%.

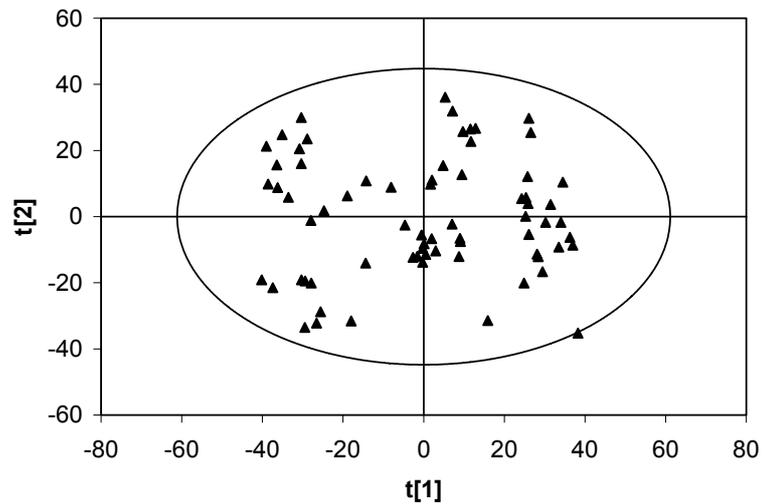


Figura 4.7. Gráfico de las dos primeras componentes del modelo de referencia para el enfoque de Nomikos y MacGregor (1995) (elipse de la T^2 para un nivel de confianza del 95%).

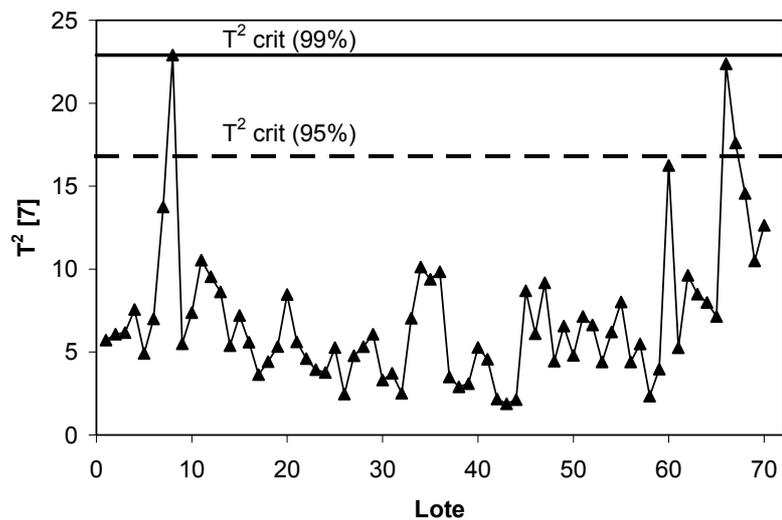


Figura 4.8. Gráfico de control de la T^2 de Hotelling de las observaciones empleadas en el desarrollo del modelo de referencia siguiendo el enfoque de Nomikos y MacGregor (1995).

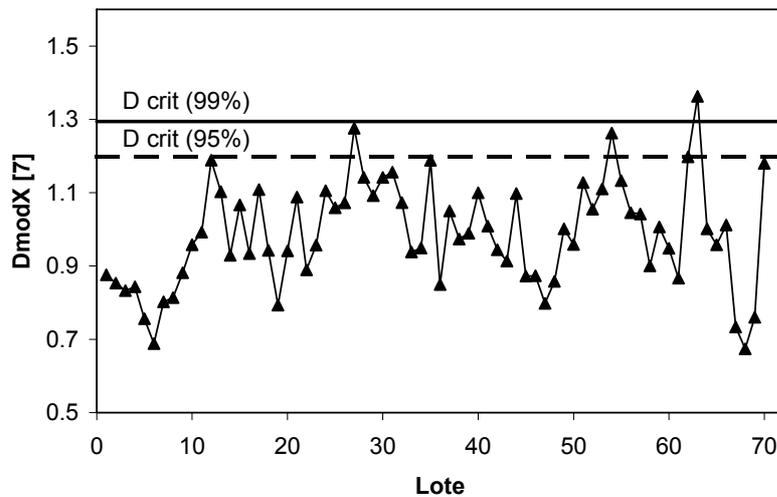


Figura 4.9. Gráfico de control de la distancia al modelo (DmodX) de las observaciones empleadas en el desarrollo del modelo de referencia siguiendo el enfoque de Nomikos y MacGregor (1995).

En la Figura 4.10 se muestra el grado de ajuste del modelo a cada una de las variables originales. El modelo de 7 componentes creado explica muy bien la mayor parte de las 1700 variables, presentando valores de la variabilidad explicada que superan el 75%. Sólo hay una zona, que comprende las variables Ox-86 a Ox-258, que no está bien explicada por el modelo, presentando valores inferiores al 25%. Dicho de otra forma, el modelo tiene una capacidad de explicación bastante baja para la variable oxígeno en el intervalo de tiempo correspondiente a la fase aerobia del proceso. Como se ha descrito anteriormente (apartado 3.1), el oxígeno es una variable que está regulada en dicha fase por medio de un control *on-off* para tratar de mantener la concentración en el reactor en un valor constante (3 mg/l), lo que da lugar a que la concentración de oxígeno oscile de manera aleatoria en torno al valor de consigna establecido.

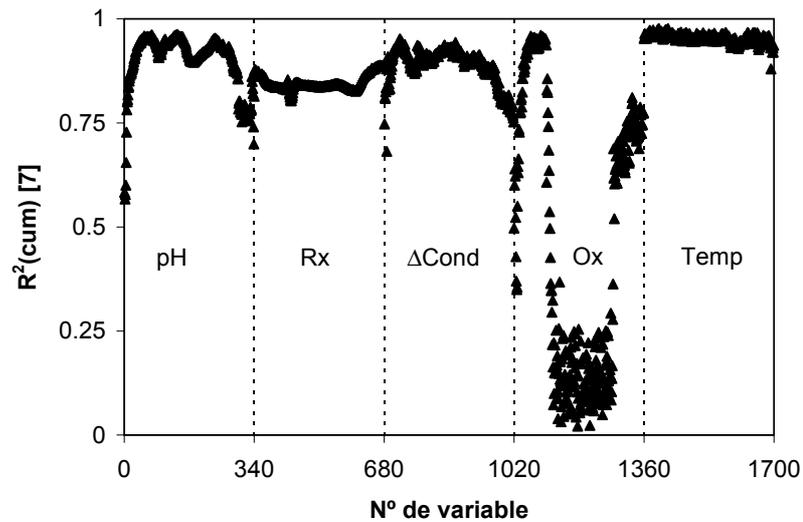


Figura 4.10. Grado de explicación de las variables del modelo.

4.1.2.2 Enfoque de Wold *et al.*

La estrategia de modelación de procesos por lotes propuesta por Wold *et al.* (1998), presentada en el apartado 1.8.6.3.3, consta de dos niveles de monitorización del proceso: el nivel de observación (análisis *on-line*) y el nivel *batch* (análisis *off-line*). En el nivel de observación los valores de las variables medidas en cada instante de tiempo son considerados como observaciones individuales. Se dispone, por tanto, en este caso de 23800 observaciones (70 lotes x 340 instantes de tiempo) y 5 variables. En este nivel se realiza una regresión PLS sobre una variable respuesta creada artificialmente que representa la madurez del lote. Las variables latentes que se obtienen de este análisis constituyen las nuevas variables del segundo nivel, el nivel *batch*, en el cual cada observación es un lote completo.

Las características del modelo obtenido con el nivel de observación se muestran en la Tabla 4.5. Se ha ajustado un modelo con cuatro componentes siguiendo la recomendación de Wold *et al.* (1998) de que el modelo creado en este nivel capture al menos un 85% de la varianza de la matriz de datos originales, ya que las variables latentes extraídas serán utilizadas posteriormente en el nivel *batch*.

Componente	$R^2_X(\text{cum})$	Valor propio	$R^2_Y(\text{cum})$	$Q^2(\text{cum})$
1	0.598	2.989	0.623	0.623
2	0.714	0.583	0.785	0.785
3	0.762	0.241	0.811	0.811
4	0.903	0.701	0.813	0.813

Tabla 4.5. Características del modelo creado en el nivel de observación.

En la Figura 4.11 se presentan los gráficos de control de las cuatro variables latentes obtenidas para todos los lotes empleados en el desarrollo del modelo y en la Figura 4.12 la distancia al modelo (DmodX). Se puede apreciar que los 70 lotes utilizados como base de referencia tienen un comportamiento normal y quedan dentro de los límites de control establecidos.

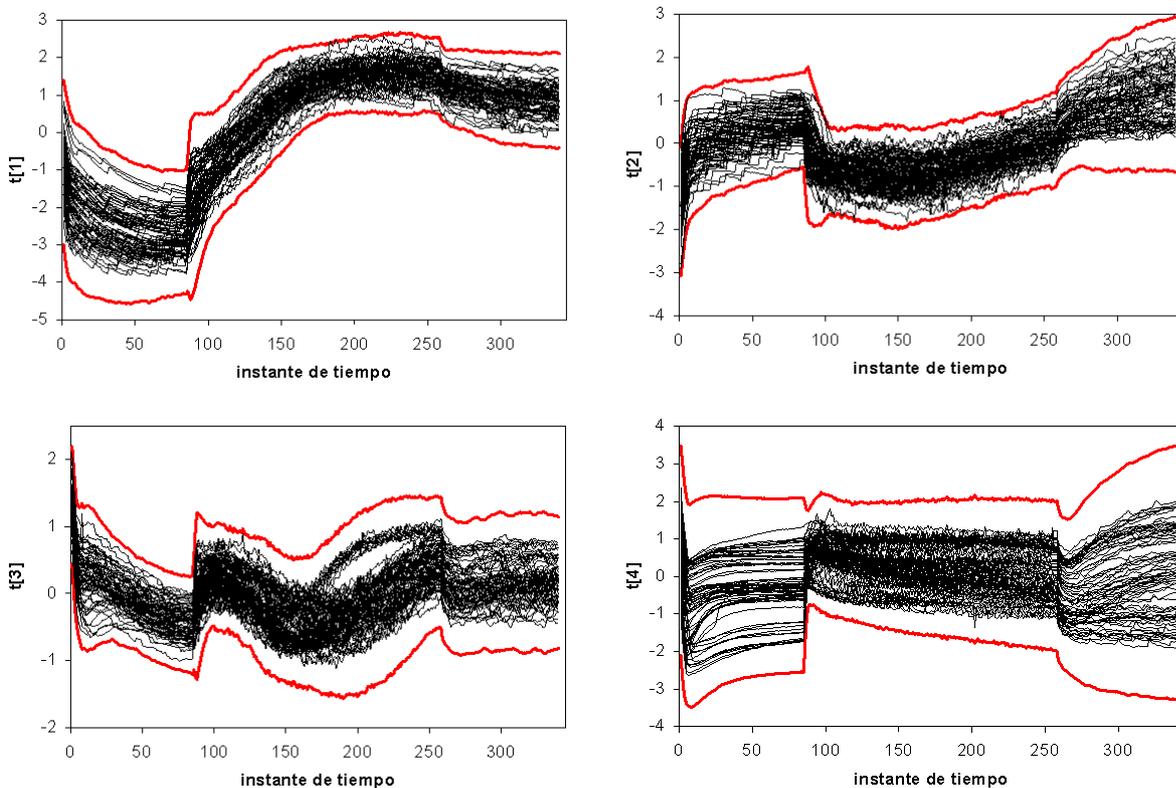


Figura 4.11. Gráficos de control de las cuatro variables latentes obtenidas en el nivel de observación con los datos empleados en el desarrollo del modelo de referencia del enfoque de Wold *et al.* (1998). La línea gruesa representa el límite de control del 99%.

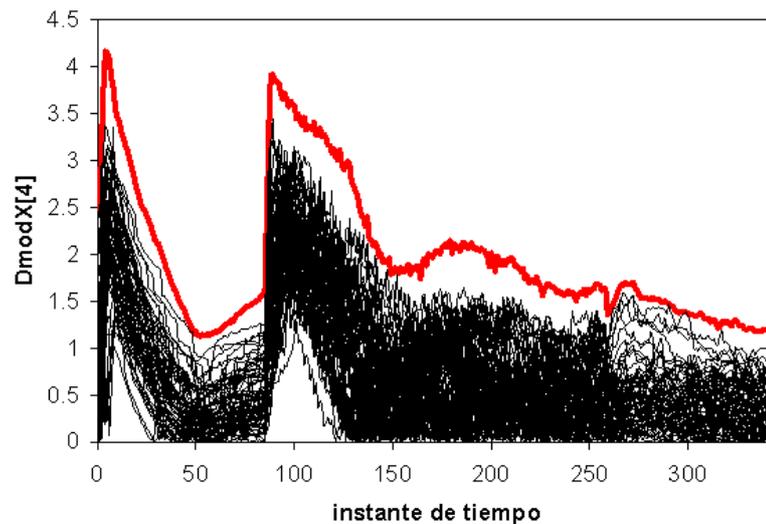


Figura 4.12. Gráfico de control de la distancia al modelo (D_{modX}) de las observaciones empleadas en el desarrollo del modelo de referencia del nivel de observación del enfoque de Wold *et al.* (1998). La línea gruesa representa el límite de control del 99%.

Las cuatro variables latentes del nivel de observación, obtenidas con los 70 lotes de la clase 1, se ordenan para formar una nueva matriz compuesta por 70 filas y 1360 columnas (4 variables latentes \times 340 instantes de tiempo). En el nivel *batch* se realiza un PCA sobre dicha matriz. El modelo PCA construido, con 6 variables latentes significativas por validación cruzada con valor propio superior a la unidad, se presenta en la Tabla 4.6. El poder explicativo del modelo con todas las variables latentes es del 93.4% y presenta una bondad de predicción del 92%.

Componente	R^2_x (cum)	Valor propio	Q^2 (cum)
1	0.378	27.240	0.368
2	0.669	20.921	0.652
3	0.780	7.975	0.760
4	0.849	4.956	0.831
5	0.916	4.853	0.904
6	0.934	1.279	0.920

Tabla 4.6. Características del modelo desarrollado para el nivel *batch*.

Al igual que en el enfoque de Nomikos y MacGregor, las 70 observaciones presentan una distribución bastante homogénea en el plano creado por las dos primeras variables latentes (Figura 4.13), y en los gráficos de control del modelo con todas las variables latentes, T^2 de Hotelling y distancia al modelo (Figura 4.14), no se aprecia que existan anomalías en las observaciones utilizadas para su construcción.

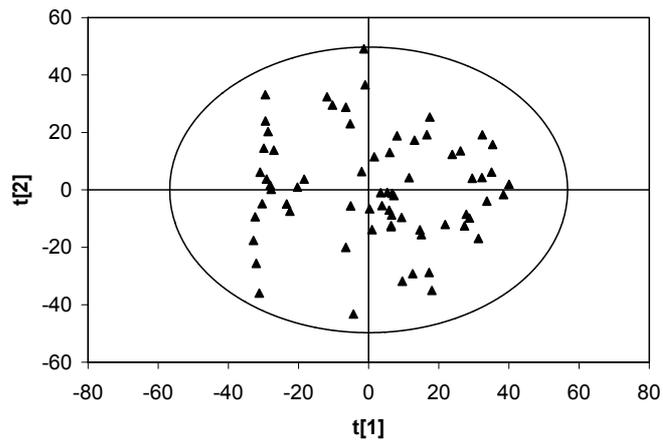


Figura 4.13. Gráfico de las dos primeras variables latentes del modelo de referencia del nivel *batch* para el enfoque de Wold *et al.* (1998). Elipse de la T^2 para un nivel de confianza del 95%.

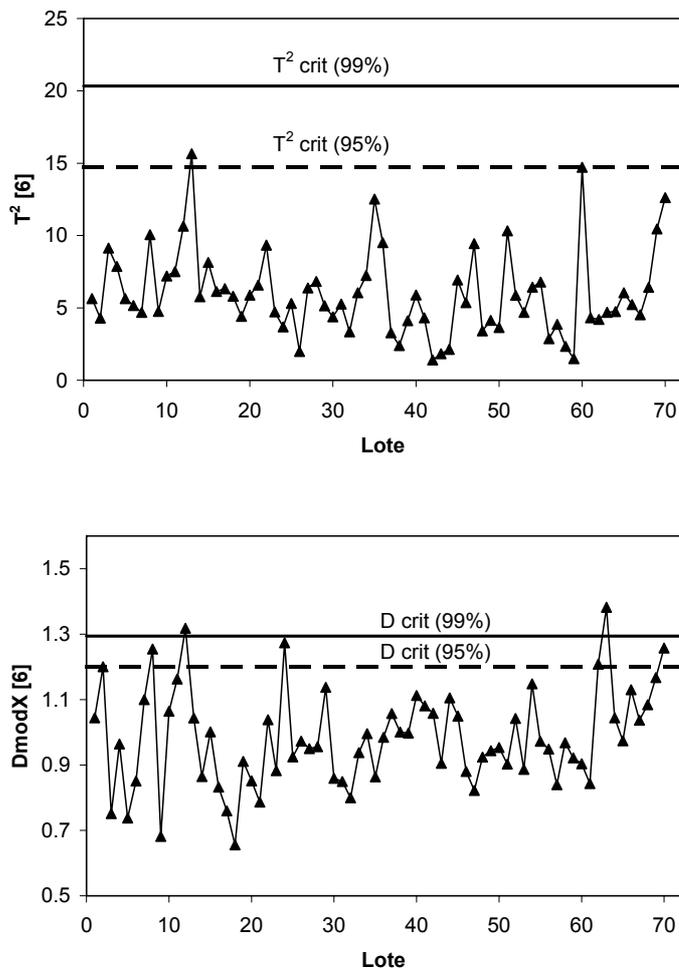


Figura 4.14. Gráficos de control de la T^2 de Hotelling y distancia al modelo ($DmodX$) de las observaciones empleadas en el desarrollo del modelo de referencia del nivel *batch* del enfoque de Wold *et al.* (1998).

4.1.2.3 Comparación de los enfoques

Después de haber ajustado los modelos para cada uno de los métodos con los 70 lotes como base de referencia y de haber comprobado que todos los lotes quedan dentro de la región operativa normal definida por los distintos modelos, se procede a discutir las ventajas e inconvenientes que presentan estos modelos respecto a su capacidad para el diagnóstico de fallos del proceso de eliminación biológica de fósforo analizado. Esto se ilustrará en base a la comparación del funcionamiento de los modelos desarrollados respecto a los 9 lotes escogidos, pertenecientes cada uno a una categoría distinta de las definidas en la Tabla 4.3 (página 137).

En la Tabla 4.7 se resumen las características básicas de los modelos creados por los distintos enfoques utilizados, donde NC es el número de variables latentes obtenidas, R^2_x es la bondad de ajuste de los datos, Q^2 la bondad de predicción y (cum) hace referencia a que son valores acumulados para el número total de variables latentes obtenidas en cada modelo.

Enfoque	NC	R^2_x (cum)	Q^2 (cum)
Nomikos y MacGregor (1995)	7	0.825	0.765
Wold <i>et al.</i> (1998)		(0.843)	(0.748)
Nivel de observación	4	0.903	0.813
Nivel <i>batch</i>	6	0.934	0.920

Tabla 4.7. Características de los modelos creados.

Para ambos enfoques la bondad de ajuste de los datos contenidos en la base de referencia supera el 80%, siendo ligeramente superior el enfoque de Wold *et al.* (1988). En este enfoque, los valores de R^2_x y Q^2 no representan la bondad de ajuste de los datos originales sino de las variables latentes obtenidas en el nivel de observación. Por este motivo, se ha indicado entre paréntesis en la Tabla 4.7 la bondad de ajuste y la bondad de predicción de los datos originales (calculada como el producto de las obtenidas en los niveles de observación y *batch*). En cuanto a la bondad de predicción, ambos enfoques presentan valores similares en torno a un 75%, siendo en este caso el enfoque de Nomikos y MacGregor ligeramente superior.

Para comprobar si los modelos desarrollados son capaces de identificar las anomalías presentes en los lotes clasificados de la Tabla 4.3, se muestra la

proyección del lote escogido como representativo de cada categoría sobre cada uno de los modelos. En cada caso se analiza si se producen salidas de los límites de control de los gráficos, bien en el de la T^2 de Hotelling, en el de la distancia al modelo o en ambos. Además, en aquellos casos en que se producen salidas de los límites de control, se utilizan los gráficos de contribución como herramienta para tratar de identificar qué variable o variables son las responsables. Para evitar la presencia de un excesivo número de este tipo de gráficos, sólo se han incluido a modo de ejemplo los correspondientes a los lotes C2 y C3. El resto de gráficos de contribución se han recogido en el Anejo 8.2.

En la Tabla 4.8 se muestran los resultados de las proyecciones realizadas, donde T^2 indica que las anomalías de dicho lote se reflejan en el modelo en la T^2 de Hotelling; Dmod en la distancia al modelo; $t[i]$ en la componente i -ésima del nivel de observación del enfoque de Wold *et al.* (1998); y en aquellos casos en los que el modelo no detecta ninguna salida de control se indica que está “Bajo control”. Cuando el lote es identificado correctamente y además los gráficos de contribución señalan acertadamente las variables responsables, se indica en la tabla con una “B”, cuando sólo es identificado correctamente con una “R”, y en los casos que no llegue a identificarse correctamente con una “X”. Este último caso tendrá lugar tanto cuando un lote con comportamiento normal sea detectado por el modelo como anómalo, como en el caso contrario, es decir, que un lote con comportamiento anómalo sea identificado erróneamente por el modelo como un lote bajo control.

Lote	Nomikos y		Wold <i>et al.</i> (1998)			
	MacGregor (1995)		Nivel observación		Nivel <i>batch</i>	
	<i>(off-line)</i>		<i>(on-line)</i>		<i>(off-line)</i>	
C1	Bajo control	B	Bajo control	B	Bajo control	B
C2	Dmod	B	Dmod, $t[1]$	R	Dmod	B
C3	Dmod, T^2	B	$t[1], t[2], t[4]$	B	Dmod, T^2	B
C4	Dmod	B	Bajo control	X	Dmod	B
C5	Dmod, T^2	B	Dmod, $t[2]$	R	Dmod	R
C6	Bajo control	X	Bajo control	X	Bajo control	X
C7	Dmod	B	$t[3]$	B	Dmod	B
C8	Bajo control	B	Bajo control	B	Bajo control	B
C9	Dmod, T^2	B	Dmod, $t[1], t[2], t[3], t[4]$	B	Dmod, T^2	B

Tabla 4.8. Resultados obtenidos para los 9 lotes proyectados en los distintos modelos.

En la Figura 4.15 se presentan los gráficos de control T^2 de Hotelling y distancia al modelo (DmodX) de los distintos enfoques para los 9 lotes proyectados. Los límites de control (del 99%) que aparecen en estos gráficos son los que se han obtenido a partir de la base de datos de referencia (los 70 lotes pertenecientes a la clase 1). Para el enfoque de Wold *et al.* (1998) se presenta, además, la proyección de los 9 lotes en los gráficos de control de las cuatro variables latentes obtenidas (Figura 4.16).

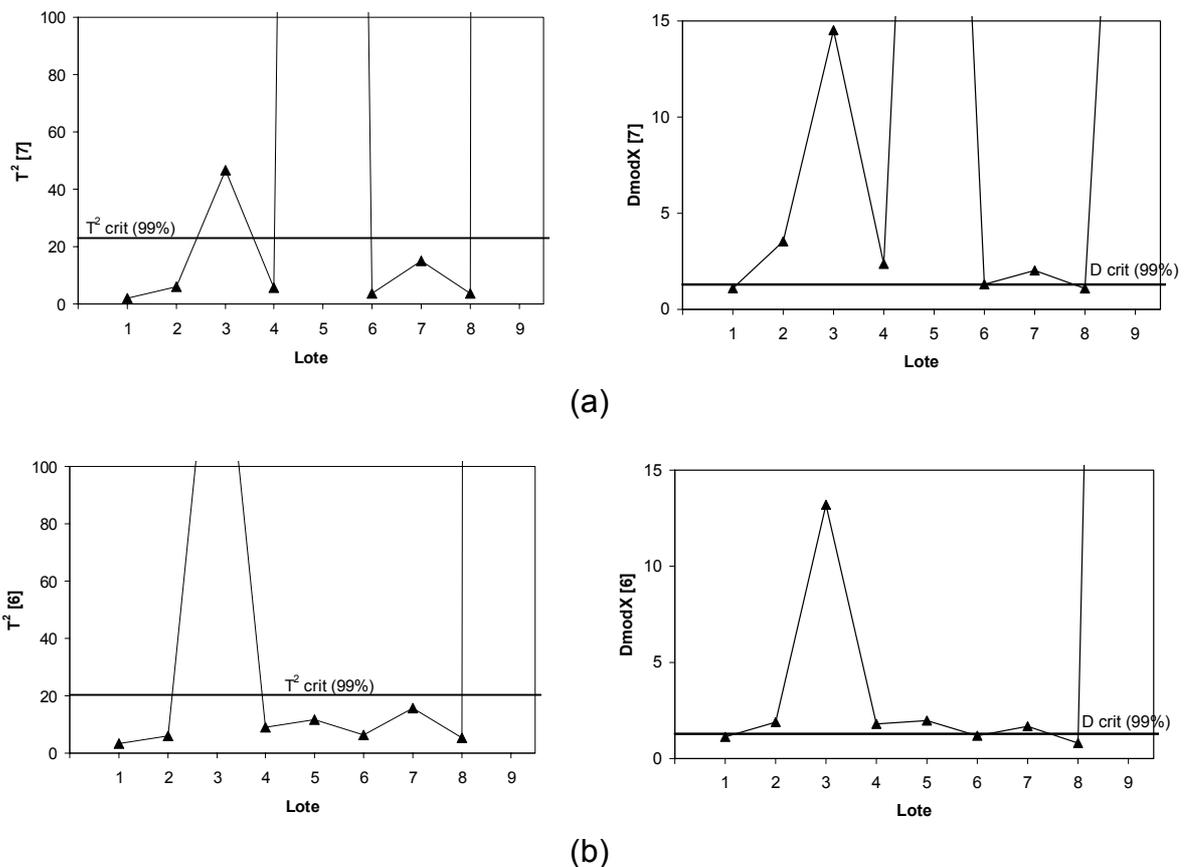


Figura 4.15. Gráficos de control de la T^2 de Hotelling y distancia al modelo (DmodX) para los lotes de comparación: (a) enfoque de Nomikos y MacGregor (1995) (b) nivel *batch* del enfoque de Wold *et al.* (1998). (Nota: algunos valores de los estadísticos T^2 y DmodX en los lotes representados quedan por encima del límite de control y no aparecen en el gráfico porque se ha limitado el eje de ordenadas para poder apreciar el resto de lotes en relación a los límites de control).

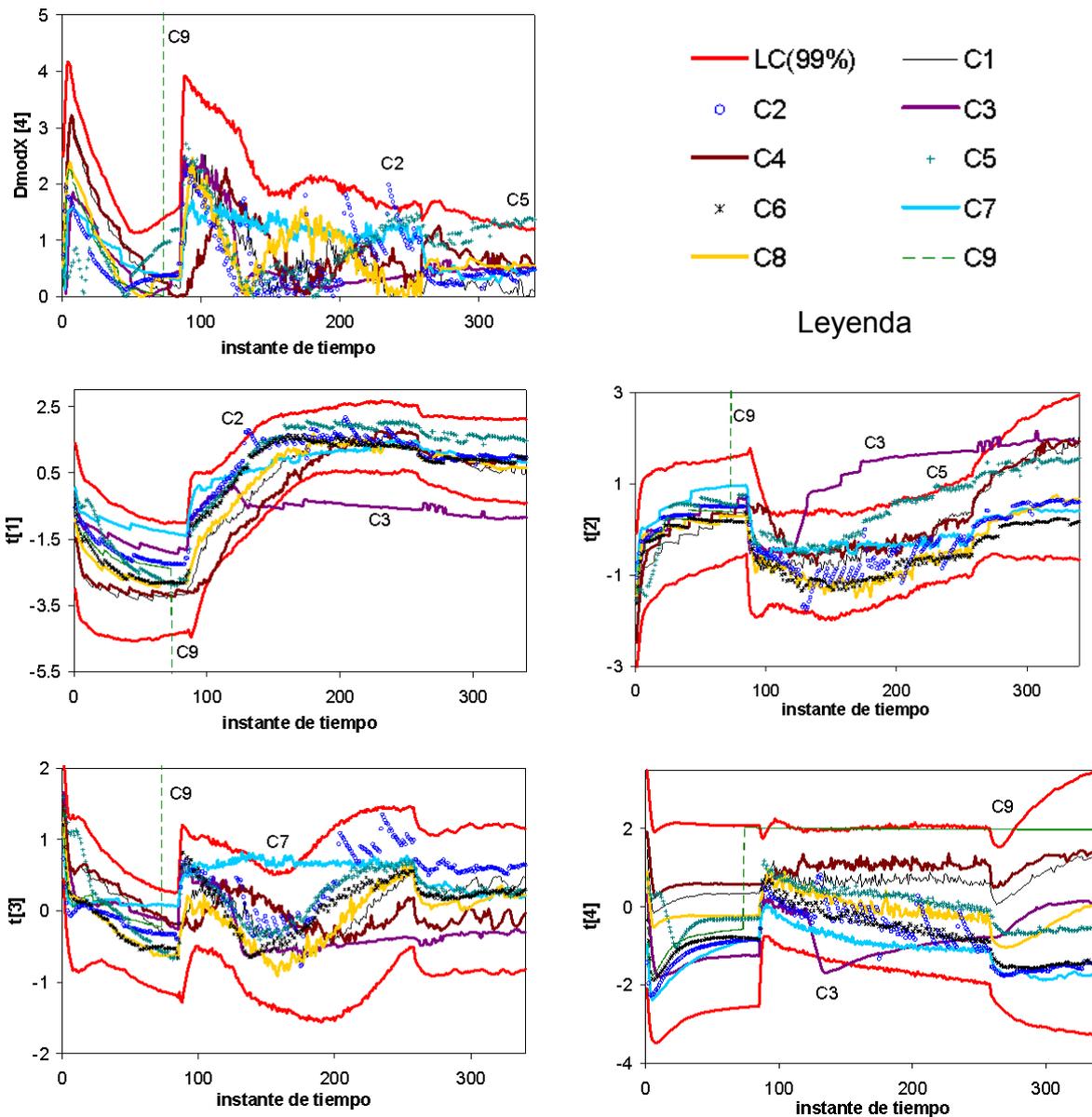


Figura 4.16. Gráficos de control de la distancia al modelo (D_{modX}) y de las cuatro variables latentes para los lotes de comparación del nivel de observación del enfoque de Wold *et al.* (1998).

El lote C1 es un lote que transcurrió normalmente y que fue utilizado en la construcción de los distintos modelos. El lote C8 también tuvo un comportamiento normal, pero no fue utilizado para construir los modelos sino para su validación. Ambos lotes aparecen en los distintos modelos como lotes bajo control (ver Figura 4.15 y Figura 4.16)

El lote C2 presentaba problemas en la medida de la concentración de oxígeno disuelto debidos a la acumulación de burbujas de aire en la sonda de oxígeno en determinados instantes de tiempo durante la etapa aerobia. Esto provoca que el valor de la concentración de oxígeno que recibe el sistema de control sea superior al valor de consigna y, por tanto, el controlador detenga la aireación, quedando en realidad el reactor en condiciones anaerobias (Barat 2004). Este problema ha sido perfectamente identificado por el modelo creado siguiendo el enfoque de Nomikos y MacGregor (gráfico de la distancia al modelo en Figura 4.15 a) que, además, señala a través de los gráficos de contribución de la distancia al modelo (Figura 4.17) al oxígeno (Ox) como variable responsable en los instantes de tiempo en que se produce esta acumulación de burbujas de aire.

Respecto de la metodología propuesta por Wold *et al.* (1998), el nivel de observación detecta alteraciones en la distancia al modelo en los instantes en que se produce la acumulación de burbujas de aire (Figura 4.16). Sin embargo, sólo en un caso se llega a superar el límite de control, y el gráfico de contribución en ese instante de tiempo (Figura 4.18 a) señala a todas las variables excepto al redox (Rx) como responsables. En este nivel, también en el gráfico de control de la primera variable latente se supera ligeramente en una ocasión el límite de control (Figura 4.16). El gráfico de contribución para la primera variable latente en dicho instante de tiempo apunta al oxígeno como variable responsable (Figura 4.18 b). Por otra parte, en el segundo nivel de este enfoque, nivel *batch*, se detecta el problema en la distancia al modelo (Figura 4.15 b), y el gráfico de contribución (Figura 4.19 a) señala como variables responsables a las cuatro variables latentes del nivel de observación en los mismos instantes de tiempo en que se produjo la acumulación de burbujas. En este nivel, para poder identificar las variables originales que causan la desviación en las variables latentes es necesario hacer un gráfico de contribución para cada variable latente en cada instante de tiempo en que ésta ha sido identificada como responsable. A través de estos nuevos gráficos de contribución se ha podido identificar al oxígeno (Ox) como variable responsable. Para evitar la presencia de un excesivo número de este tipo de gráficos, sólo se muestran para el lote C2 los correspondientes a las variables latentes $t[2]$ y $t[3]$ en el instante tiempo $k=234$ (Figura 4.19 b).

Finalmente se presenta en la Figura 4.20 el perfil del oxígeno disuelto en el lote C2 para poder visualizar las alteraciones puntuales que han sido identificadas por los distintos modelos.

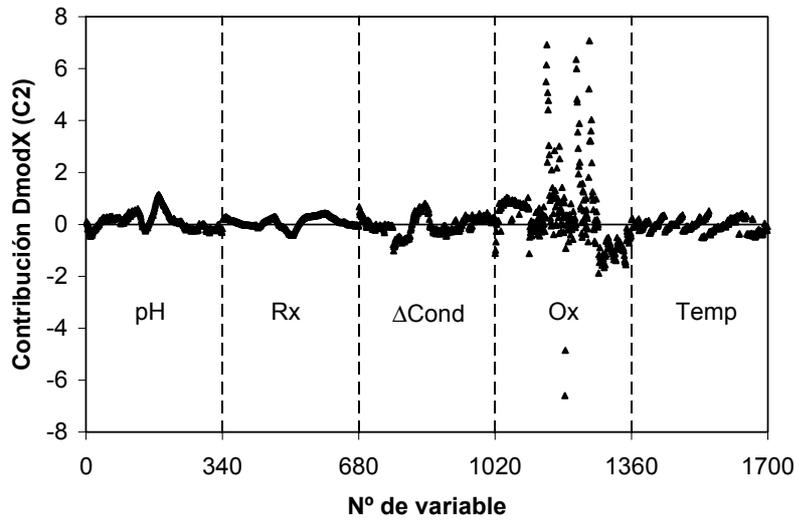


Figura 4.17. Gráfico de contribución de la distancia al modelo del enfoque de Nomikos y MacGregor (1995) para el lote C2.

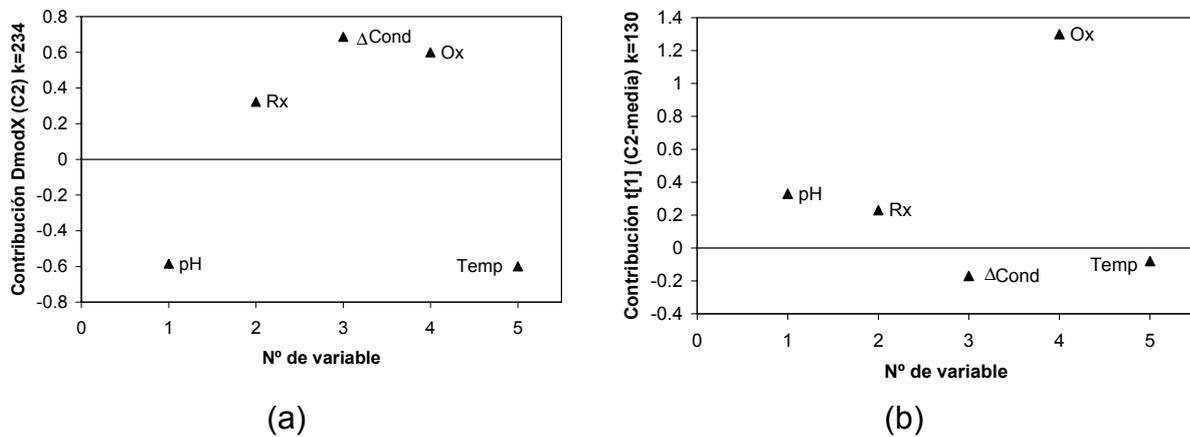
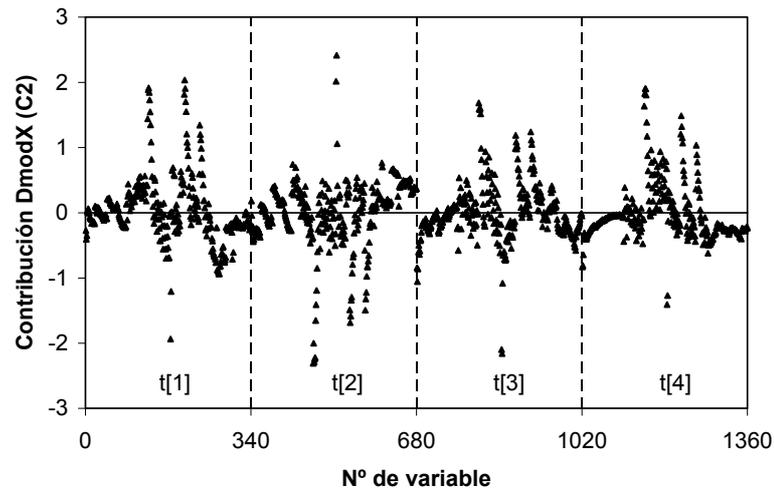
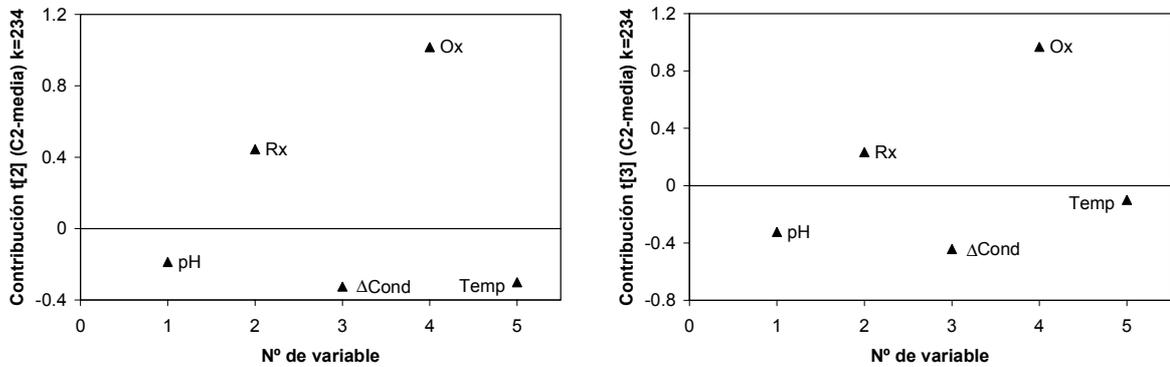


Figura 4.18. Gráficos de contribución del nivel de observación del enfoque de Wold *et al.* (1998) para el lote C2: (a) distancia al modelo en el instante de tiempo $k=234$ (b) primera variable latente ($t[1]$) en el instante de tiempo $k=130$.



(a)



(b)

Figura 4.19. Gráficos de contribución para el lote C2: (a) de la distancia al modelo (DmodX) del nivel *batch* del enfoque de Wold *et al.* (1998) (b) de las variables latentes $t[1]$ y $t[3]$ del nivel de observación en el instante de tiempo $k=234$.

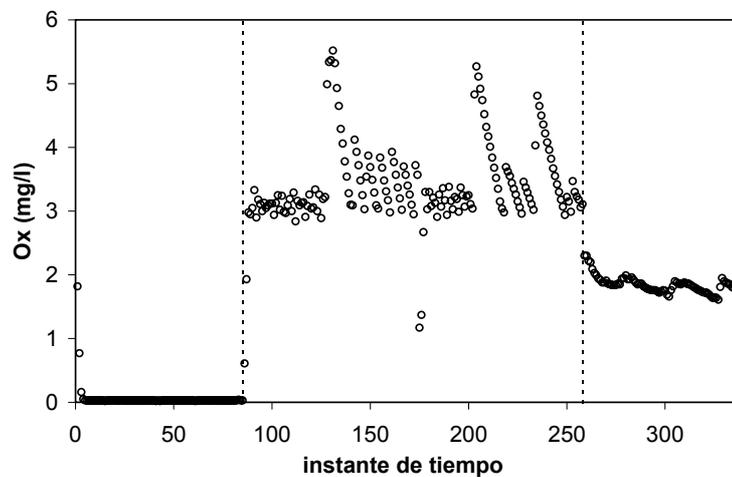


Figura 4.20. Perfil de la concentración de oxígeno disuelto en el reactor en el lote C2.

En el lote C3 se produce un fallo en el sistema de aireación durante el transcurso de la etapa aerobia, de forma que desde que se produce el fallo, el sistema no suministra oxígeno durante el resto de la etapa y, por tanto, el reactor pasa a estar en condiciones anaerobias. El enfoque de Nomikos y MacGregor identifica que hay anomalías en dicho lote (Figura 4.15 a) y, a través de los gráficos de contribución (sólo se muestra el de la distancia al modelo, Figura 4.21) señala a las variables oxígeno, pH e incremento de conductividad (ΔCond) como responsables. En este caso es la falta de oxígeno la responsable de la anomalía en este lote. Sin embargo, al estar en condiciones anaerobias el reactor durante parte de la etapa aerobia, las bacterias PAO dejan de tomar fósforo del medio lo cual influye tanto en el perfil de pH, como en el de los incrementos de conductividad.

Por otra parte, en el nivel de observación del enfoque de Wold *et al.* (1998), las anomalías se reflejan con claridad en las dos primeras variables latentes (Figura 4.16), aunque sólo en la primera se mantiene hasta el final del lote. Al realizar gráficos de contribución de las dos primeras variables latentes en distintos instantes de tiempo (en lugar de estos gráficos, en la Figura 4.22 se muestra el sumatorio de las contribuciones al cuadrado a lo largo del período de tiempo que la variable latente está fuera de los límites de control), las principales variables que aparecen como responsables son el oxígeno, pH y los incrementos de conductividad (ΔCond). En el nivel *batch* se producen salidas de control tanto en la T^2 de Hotelling como en la distancia al modelo (Figura 4.15 b). En el gráfico de contribución de la distancia al modelo no se refleja con claridad qué variables latentes del nivel de observación son las responsables (Figura 4.23 a), mientras que la T^2 de Hotelling señala a las dos últimas variables latentes del nivel *batch* ($t_b[5]$ y $t_b[6]$) como responsables (Figura 4.23 b). A su vez, el gráfico de contribución de estas dos variables latentes del nivel *batch* permite identificar como variables responsables a las dos primeras variables latentes del nivel de observación (Figura 4.23 c). Como ya se ha indicado anteriormente, al realizar gráficos de contribución en distintos instantes de tiempo para estas variables latentes del nivel de observación se identifica a las variables oxígeno, pH e incrementos de conductividad (ΔCond) como responsables.

En la Figura 4.24 se muestra la evolución de las trayectorias de los incrementos de conductividad y del oxígeno disuelto en el lote C3 para poder visualizar las alteraciones que se producen.

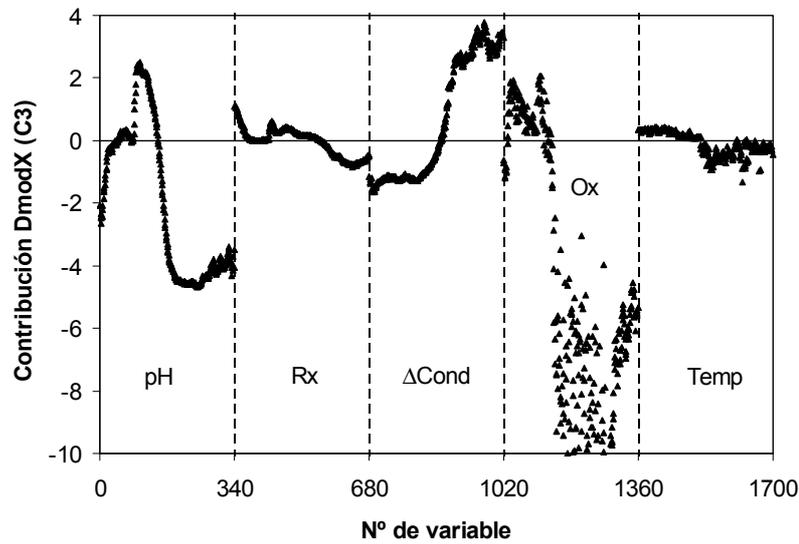


Figura 4.21. Gráfico de contribución de la distancia al modelo del enfoque de Nomikos y MacGregor (1995) para el lote C3.

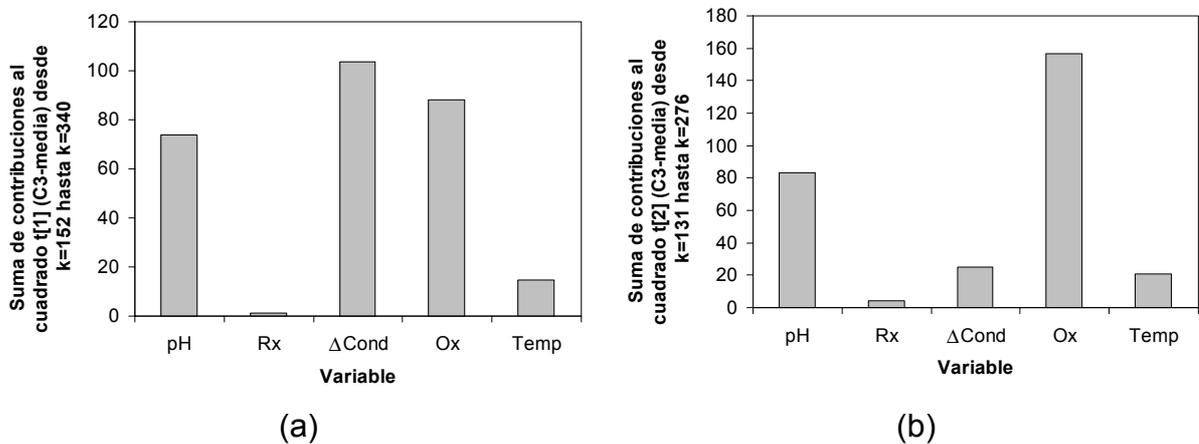


Figura 4.22. Gráficos del sumatorio de las contribuciones al cuadrado del nivel de observación del enfoque de Wold *et al.* (1998) para el lote C3: (a) primera variable latente ($t[1]$) entre los instantes de tiempo $k=152$ y $k=340$ (b) segunda variable latente ($t[2]$) entre los instantes de tiempo $k=131$ y $k=276$.

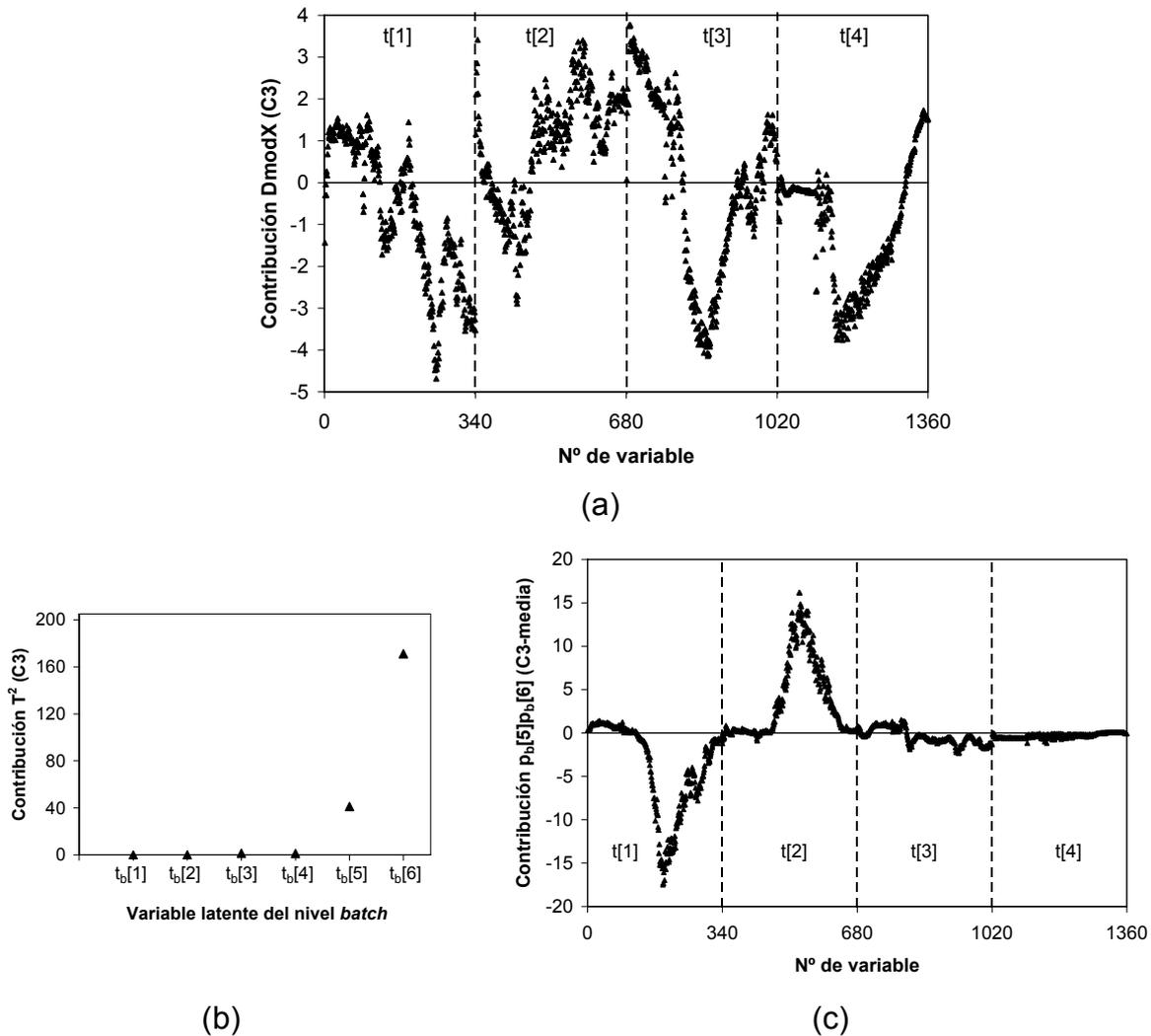


Figura 4.23. Gráficos de contribución del nivel *batch* del enfoque Wold *et al.* (1998) para el lote C3: (a) de la distancia al modelo (b) de la T^2 de Hotelling (c) de las variables latentes del nivel *batch* $t_b[5]$ y $t_b[6]$.

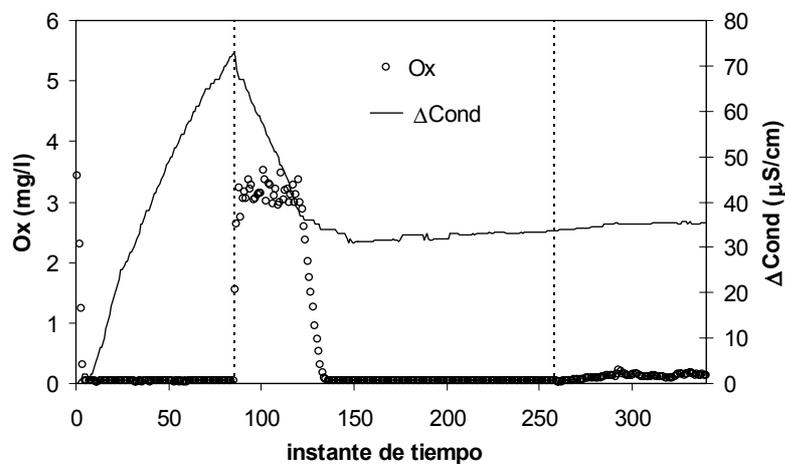


Figura 4.24. Perfiles de los incrementos de conductividad (ΔCond) y de la concentración de oxígeno disuelto (Ox) en el reactor en el lote C3.

En el lote C4 al iniciarse la etapa aerobia el oxígeno tarda en alcanzar su valor de consigna más de 20 minutos, cuando lo habitual es que no se exceda de 5 minutos. Con el enfoque de Nomikos y MacGregor se detecta la anomalía en la distancia al modelo (Figura 4.15 a), y el gráfico de contribución (Anejo 8.2, Figura 8.3) pone de manifiesto que la variable responsable es el oxígeno al comienzo de la etapa aerobia. Por otro lado, el nivel de observación de la metodología de Wold *et al.* (1998) no detecta ninguna alteración en el transcurso de dicho lote (Figura 4.16). En el nivel *batch* sí que se refleja el problema en la distancia al modelo (Figura 4.15 b). El posterior análisis de los gráficos de contribución (Anejo 8.2, Figura 8.4 a) señala las tres primeras variables latentes del nivel de observación al comienzo de la etapa aerobia como responsables. Los gráficos de contribución de estas variables latentes (Anejo 8.2, Figura 8.4 b) apuntan a la variable oxígeno como la principal responsable.

En el inicio del lote C5 la agitación tarda en ponerse en funcionamiento, quedando el fango sedimentado, no consumiéndose el sustrato presente en el agua residual influente y no habiendo ninguna actividad por parte de las bacterias hasta que se consigue poner en marcha la agitación. Este problema también provoca que el oxígeno que entra junto con el agua residual influente no sea consumido mientras el sistema de agitación esté parado. En el enfoque de Nomikos y MacGregor al proyectar este lote se detecta una salida de control tanto en la T^2 de Hotelling como en la distancia al modelo (Figura 4.15 a). Los gráficos de contribución en ambos casos (sólo mostrado el de la distancia al modelo, Figura 8.5 en Anejo 8.2) apuntan a la variable oxígeno al inicio de la etapa anaerobia como responsable.

Por otro lado, en el nivel de observación del enfoque de Wold *et al.* (1998) se detectan anomalías en el comportamiento hacia el final del lote en la distancia al modelo (Figura 4.16). Los gráficos de contribución de la distancia al modelo en dichos instantes de tiempo (Anejo 8.2, Figura 8.6 a) no permiten identificar las variables responsables porque todas tienen pesos similares. En este nivel, también en el gráfico de control de la segunda variable latente se puede observar que muy ligeramente se rebasa el límite de control (Figura 4.16). En este caso a través de los gráficos de contribución (Anejo 8.2, Figura 8.6 b) se puede identificar a la

temperatura como variable responsable. Dado que el único problema que presenta el lote C5 se produce al comienzo del mismo, se puede concluir que las salidas de control reflejadas por el nivel de observación son falsas alarmas. En el nivel *batch* de este enfoque, se supera el límite de control en la distancia al modelo (Figura 4.15 b). Los gráficos de contribución en este modelo (Anejo 8.2, Figura 8.7 a) señalan a las cuatro variables latentes del nivel de observación como responsables al comienzo del lote. Sin embargo, los gráficos de contribución de las variables latentes del nivel de observación (Anejo 8.2, Figura 8.7 b) no apuntan hacia las mismas variables como responsables, lo cual no permite una interpretación sencilla de cuál puede ser el problema.

Respecto al lote C6, la única alteración que presenta es un salto muy pronunciado en el valor de la conductividad al cambiar de la etapa anaerobia a la aerobia, concretamente en este lote se produce una caída de 48 $\mu\text{S}/\text{cm}$ en el transcurso de apenas un minuto. Son varios los lotes que presentan esta alteración puntual, que en algunos casos aparece, además, junto con una mayor dispersión en los registros, y coincide con las fechas en que el filtro analógico estaba estropeado. En todos los modelos, este lote aparece bajo control. Es muy probable que esto sea debido a que ese salto puntual haya quedado enmascarado dentro de la variabilidad que presentan las trayectorias de los incrementos de conductividad de los lotes que se han empleado en la construcción de los distintos modelos.

El lote C7 pertenece al periodo de puesta en marcha del sistema. Además, en la trayectoria de pH de este lote aparece una rápida subida en su valor al inicio de la etapa aerobia, que es debida al consumo por parte de las bacterias heterótrofas del ácido acético que no ha sido consumido durante la etapa anaerobia (Serralta *et al.*, 2004). Al proyectar este lote sobre el modelo construido siguiendo la propuesta de Nomikos y MacGregor, se detecta un comportamiento fuera de lo habitual en la distancia al modelo (Figura 4.15 a). El gráfico de contribución (Anejo 8.2, Figura 8.8) apunta a las trayectorias de pH y de los incrementos de conductividad (ΔCond) como responsables. Esto indica que en la identificación de este lote como anómalo no sólo influye el hecho de que quede ácido acético en el sistema al final de la etapa anaerobia, sino también el que este lote pertenezca al periodo de puesta en marcha,

en el que todavía no se ha alcanzado una población estable de bacterias PAO y, por tanto, la liberación así como la toma de fósforo es aún reducida. Respecto al enfoque de Wold *et al.* (1998), en el nivel de observación se detecta una ligera salida de control en la tercera variable latente, pero no coincide con el cambio de etapa (Figura 4.16). En este caso los gráficos de contribución (Anejo 8.2, Figura 8.9) revelan que es el pH la variable responsable. En el nivel *batch* se detecta una ligera salida de control en la distancia al modelo (Figura 4.15 b), y el gráfico de contribución (Anejo 8.2, Figura 8.10 a) no señala a ninguna variable latente del nivel de observación en particular como responsable, aunque aparecen ligeramente más remarcados los instantes de tiempo correspondientes al inicio de la etapa aerobia. Los gráficos de contribución de las variables latentes del nivel de observación en dichos instantes de tiempo (Anejo 8.2, Figura 8.10 b) apuntan a la variable pH como responsable.

En el lote C9 hacia el final de la etapa anaerobia el convertidor de señal *Consort* se queda bloqueado provocando un fallo de comunicación en la transmisión de los datos recogidos por las sondas de conductividad, pH y potencial redox, siendo a partir de ese momento nulas las lecturas de estas tres variables, sin que esto provoque ninguna alteración en el proceso biológico. Este problema ha sido perfectamente identificado por todos los modelos, presentando valores muy elevados tanto de la T^2 de Hotelling como de la distancia al modelo (Figura 4.15), y en el caso del nivel de observación, en las tres primeras variables latentes así como en la distancia al modelo (Figura 4.16). Los gráficos de contribución en todos los modelos (Anejo 8.2, Figura 8.11, Figura 8.12 y Figura 8.13) señalan al pH como la principal variable responsable.

Tras haber analizado los resultados de las proyecciones sobre los modelos creados, y comparando el comportamiento *off-line* de los dos enfoques estudiados, se podrían extraer las siguientes conclusiones:

- Ambos enfoques son capaces de modelar los datos del proceso y presentan valores parecidos de la bondad de ajuste y de la bondad de predicción (Tabla 4.7).

- Ambos enfoques son capaces de detectar las anomalías que con más frecuencia se han producido en el conjunto de datos analizado del proceso de eliminación biológica de fósforo.
- El enfoque de Nomikos y MacGregor (1995) se comporta de una manera más consistente, y ha diagnosticado correctamente las variables responsables en todos los casos. Además, permite una interpretación mucho más clara y sencilla. Esto se debe, por un lado, al desdoblamiento de la matriz original empleado (tipo E) ya que permite comparar directamente las trayectorias de los pesos del modelo, así como los gráficos de contribución, con las trayectorias de las variables originales y, por otro, a que los dos niveles con los que se trabaja son: variables originales y variables latentes, siendo directa la relación entre ambas.
- El nivel *batch* del enfoque de Wold *et al.* (1998) ha mostrado un buen comportamiento en lo que respecta a la detección de anomalías, aunque sin llegar a ser tan potente como la metodología propuesta por Nomikos y MacGregor. Además, tiene el inconveniente de que es más complicada la interpretación de las salidas de control. Esto es debido a que se trabaja con las variables latentes que se obtienen en el nivel de observación, a las que se aplica un PCA para obtener unas nuevas variables latentes que sintetizan la información contenida en las anteriores. Esto provoca que la relación entre las variables originales y las variables latentes obtenidas en el nivel *batch* no sea directa.

Por otra parte, del análisis realizado en este apartado también se puede concluir que el nivel de observación del enfoque de Wold *et al.* (1998) presenta una capacidad de detección de anomalías bastante limitada (Tabla 4.8), lo cual no es conveniente si se pretende utilizar este enfoque para monitorización en tiempo real del sistema estudiado. En el apartado 4.2 se presenta un estudio comparativo de distintos enfoques para la monitorización en tiempo real del proceso.

4.1.2.4 Modificaciones propuestas

Los resultados del apartado anterior han puesto de manifiesto que en el conjunto de datos del proceso de eliminación biológica de fósforo estudiado, el nivel de observación del enfoque de Wold *et al.* (1998) presenta dificultades para la detección de una serie de lotes que previamente han sido identificados como anómalos. En este apartado se pretende analizar las posibles causas del inferior comportamiento que ha presentado el enfoque, sugerir algunas modificaciones del mismo que puedan contribuir a mejorar los resultados y verificarlo con el mismo conjunto de datos.

Como se expuso en el apartado 1.8.6.3.3, en el nivel de observación de la estrategia de modelación de procesos por lotes propuesta por Wold *et al.* (1998) se construye un modelo PLS con el objetivo de predecir la duración del lote. Debido a que los resultados obtenidos con este nivel no han sido satisfactorios, se decidió profundizar en el comportamiento del modelo PLS obtenido a partir del conjunto de lotes de referencia. Para profundizar en el modelo se pueden estudiar los pesos en la formación de las variables latentes y, además, se puede comprobar por un lado las predicciones del modelo y por otro si la relación interna del mismo es lineal.

En la Figura 4.25 se muestran los pesos de las dos primeras variables latentes del modelo PLS del nivel de observación. Se puede observar que la variable con peso más elevado (en valor absoluto) es la temperatura, por tanto, ésta es la variable más importante en el modelo. Además, su posición en el cuadrante diametralmente opuesto al de la variable respuesta, nos indica que la correlación entre ambas es negativa. Como se comentó en la descripción del montaje experimental (apartado 3.1), el reactor está ubicado en el interior de una cámara termostatzada en la que se mantiene la temperatura en 20°C. El agua residual a tratar está almacenada en el exterior de dicha cámara y su temperatura es ligeramente superior. Cuando se inicia un nuevo lote, en la etapa de llenado entra el agua en el reactor lo que provoca un ligero aumento de la temperatura, la cual, posteriormente a lo largo del transcurso del lote disminuye paulatinamente (Figura 3.3). Es decir, mientras aumenta el tiempo transcurrido desde el inicio del lote, la temperatura disminuye lentamente y, como

consecuencia, ambas variables presentarán una correlación negativa. Esto está poniendo de manifiesto que en la primera componente del modelo PLS desarrollado tendrán peso elevado aquellas variables del proceso con comportamiento monótono creciente o decreciente, lo cual es coherente con el hecho de que el efecto lineal es modelado con esta componente (Wold *et al.*, 1998).

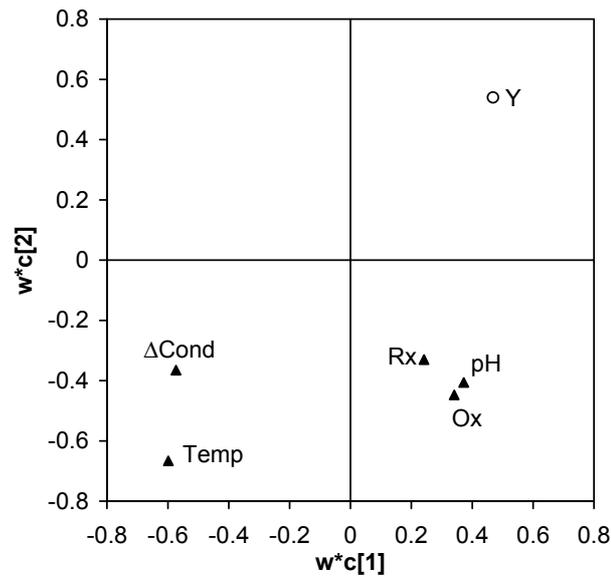


Figura 4.25. Pesos de las variables en la formación de las dos primeras variables latentes del modelo PLS del nivel de observación del enfoque de Wold *et al.* (1998). La variable respuesta Y es el tiempo transcurrido desde el inicio del lote y representa la madurez del lote.

Con el objetivo de comprobar la bondad de predicción del modelo PLS creado en el nivel de observación, se muestra en la Figura 4.26 el tiempo que realmente ha transcurrido frente al tiempo predicho para el promedio de los 70 lotes de referencia. En dicha figura se puede observar que las predicciones del modelo son poco acertadas, sobre todo durante toda la etapa anaerobia así como en el comienzo de la etapa de sedimentación donde el tiempo predicho disminuye. Este comportamiento está reflejando que este enfoque puede estar teniendo dificultades para la modelación de este conjunto de datos.

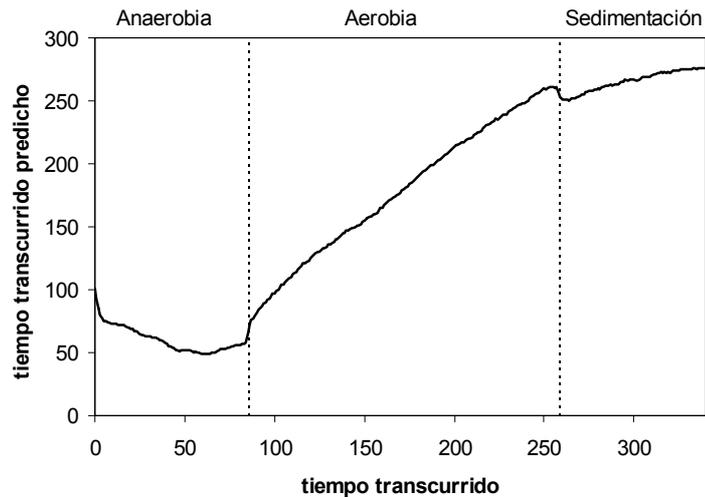


Figura 4.26. Representación del tiempo transcurrido de lote frente al tiempo transcurrido predicho por el modelo PLS del nivel de observación del enfoque de Wold *et al.* (1998). Se ha representado únicamente el promedio de los 70 lotes empleados en la construcción del modelo. En vertical se indican los cambios de etapa.

Para verificar si la relación interna del modelo es o no lineal, se presenta en la Figura 4.27 la primera variable latente obtenida por el modelo PLS en los dos espacios, el de las variables de proceso y el correspondiente a la variable respuesta. Se puede apreciar que en este caso hay una relación no lineal muy marcada, lo cual es otra indicación de que el enfoque no está teniendo un comportamiento adecuado para este conjunto de datos.

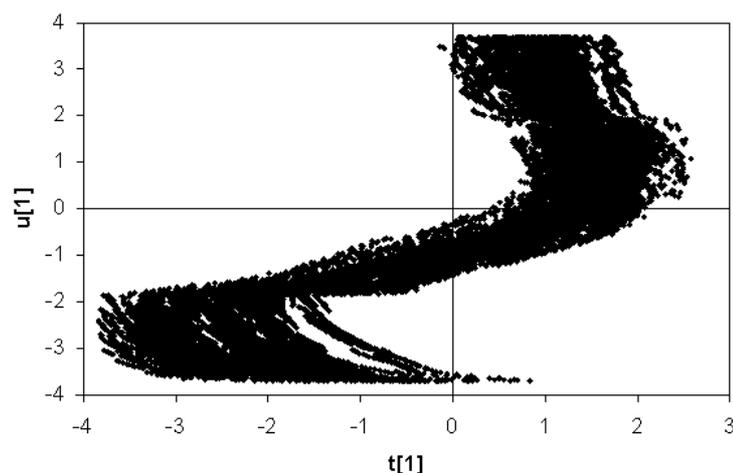


Figura 4.27. Relación no lineal entre las primeras variables latentes del modelo PLS del nivel de observación del enfoque de Wold *et al.* (1998): $t[1]$ obtenida de las variables de proceso y $u[1]$ obtenida de la variable respuesta (tiempo transcurrido del lote). Aparecen representados los 70 lotes del conjunto de referencia.

Tanto una mala predicción de la variable respuesta como una relación interna no lineal del modelo PLS son claros indicadores de un mal funcionamiento del método. Esto se produce como consecuencia del comportamiento no monótono que presentan varias variables del proceso. En este sentido, Westerhuis *et al.* (1999) también encontraron estas dificultades empleando el enfoque de Wold *et al.* (1998) para modelar un conjunto de datos procedentes de un proceso de polimerización industrial. Además, estos autores apuntan otros inconvenientes del enfoque, como que el centrado de los datos no elimina la no linealidad presente en las trayectorias de las variables registradas, y que se necesitan muchas variables latentes para representar la información contenida en los datos, es decir, que no supone una reducción importante de la dimensión del problema. En un trabajo más reciente (Kourti, 2003), también se incide en estos dos últimos inconvenientes, pero además, se hace hincapié en el hecho de que cuando se está monitorizando un nuevo lote, el cálculo de las variables latentes en un instante de tiempo dado sólo tiene en cuenta los valores de las variables registradas en dicho instante de tiempo y, por tanto, no hay una comprobación de las autocorrelaciones y correlaciones cruzadas de los valores registrados en el instante actual con los valores anteriores.

A pesar de las limitaciones del enfoque comentadas anteriormente, se debe tener presente que el nivel de observación presenta una ventaja muy importante cuando se pretende realizar la monitorización en tiempo real del proceso: no se tiene que recurrir a ningún tipo de imputación para completar la trayectoria futura del lote (lo que sí es necesario con la metodología propuesta por Nomikos y MacGregor (1995)). Queda, por tanto, justificado el interés de estudiar si es posible introducir alguna modificación en este enfoque que permita soslayar las dificultades encontradas para modelar el conjunto de datos analizado. A continuación se proponen dos modificaciones del enfoque original de Wold *et al.* (1998) que manteniendo su estructura, es decir, empleando el mismo tipo de desdoblamiento y los dos niveles de monitorización, podrían mejorar los resultados obtenidos:

- Alternativa 1: Utilizar un PCA en lugar de un PLS en el nivel de observación. De esta forma no se utiliza una variable respuesta, por lo que las variables

latentes obtenidas son las que maximizan la varianza de los datos, y no tienen por qué presentar una trayectoria monótona con el tiempo.

- Alternativa 2: Realizar el preprocesamiento, tanto el centrado como la tipificación a varianza unitaria de los datos, de la misma forma que el enfoque de Nomikos y MacGregor (1995). Esto permite eliminar la principal no linealidad presente en los datos, y trabajar con desviaciones respecto de la trayectoria media. Posteriormente, la matriz de datos preprocesada se reordena como en el enfoque de Wold *et al.* (1998), y se realiza un PCA para extraer la información.

La forma preprocesar y de reordenar la matriz de datos que se sugiere en la segunda alternativa ha sido recientemente empleada por Yoo *et al.* (2004a), aunque estos autores utilizan la técnica de análisis de componentes independientes para extraer la información subyacente en los datos.

En las dos alternativas que se proponen, las variables latentes extraídas en el nivel de observación se utilizan posteriormente en el nivel *batch* para describir los lotes completos, tal y como hace el enfoque de Wold *et al.* (1998). Por esta razón, es conveniente seguir la recomendación de estos autores de que el modelo creado en el nivel de observación explique al menos un 85% de la varianza de los datos.

Para comparar las dos alternativas se procede de la misma manera que en el apartado 4.1.2.3, es decir, se ajustan en primer lugar los modelos con los 70 lotes de referencia de la clase 1 (Tabla 4.3 en página 137) y, posteriormente, se compara el funcionamiento de los modelos desarrollados respecto a una serie de lotes pertenecientes al resto de las clases definidas en la Tabla 4.3, entre los que se encuentran lotes buenos y lotes con anomalías. El funcionamiento se ilustra con los mismos nueve lotes empleados en el apartado 4.1.2.3, de esta forma, los resultados también se pueden comparar directamente con los obtenidos en dicho apartado.

En la Tabla 4.9 se resumen las características básicas de los modelos creados considerando cada una de las modificaciones sugeridas para el enfoque de Wold *et*

al. (1998), donde NC es el número de variables latentes obtenidas, R^2_x es la bondad de ajuste de los datos, Q^2 la bondad de predicción y (cum) hace referencia a que son valores acumulados para el número total de variables latentes obtenidas en cada modelo.

Modificación del enfoque original	NC	R^2_x (cum)	Q^2 (cum)
Alternativa 1			
Nivel de observación	3	0.884	--
Nivel <i>batch</i>	5	0.939	0.924
Alternativa 2			
Nivel de observación	4	0.9	--
Nivel <i>batch</i>	7	0.799	0.718

Tabla 4.9. Características de los modelos creados introduciendo modificaciones en la metodología propuesta por Wold *et al.* (1998). Alternativa 1: emplear un PCA en lugar de un PLS. Alternativa 2: eliminar la no linealidad en el preprocesamiento y utilizar posteriormente un PCA.

En el nivel de observación de la primera alternativa, a excepción de la primera variable latente que es estadísticamente significativa según el criterio de validación cruzada, todas las demás son forzadas. En la segunda alternativa, las cuatro variables latentes del nivel de observación que se han obtenido son forzadas. En ambos casos se han obtenido las variables latentes necesarias para que el modelo retenga más del 85% de la varianza de los datos. Debido a que las componentes extraídas han sido forzadas, no se indica ningún valor en la columna correspondiente a la bondad de predicción (Q^2). En el nivel *batch* de ambas alternativas, todas las variables latentes que se han retenido son significativas por el criterio de validación cruzada y tienen valor propio mayor que la unidad.

Según se comprobó en el apartado anterior (4.1.2.3), los resultados obtenidos por el nivel de observación del enfoque de Wold *et al.* (1998) desaconsejaban la utilización de este enfoque para la monitorización en tiempo real del proceso estudiado. Se pretende, por tanto, comprobar si las alternativas propuestas realmente suponen una mejora en este nivel.

En la Figura 4.28 y en la Figura 4.29 se presentan la evolución de los gráficos de control de la distancia al modelo (medida a través del SPE) y de las variables latentes obtenidas en la primera y segunda alternativa, respectivamente.

Comparando ambas figuras se puede observar que en la segunda alternativa, en la que se realiza el mismo preprocesamiento que en el enfoque de Nomikos y MacGregor (1995), hay un mayor número de lotes anómalos que quedan fuera de los límites de control, siendo éstos además rebasados con mayor severidad que cuando se emplea la primera alternativa. Por otra parte, comparando los gráficos de control de la Figura 4.28 con los obtenidos aplicando la metodología original (Figura 4.16), no se observa que la primera alternativa propuesta represente una mejora significativa respecto de la metodología original. Sin embargo, la segunda alternativa (Figura 4.29) sí que presenta mejores resultados que el enfoque original, lo cual podría estar reflejando la elevada influencia que puede llegar a tener el preprocesamiento de las variables en la capacidad de detección de anomalías de los modelos obtenidos.

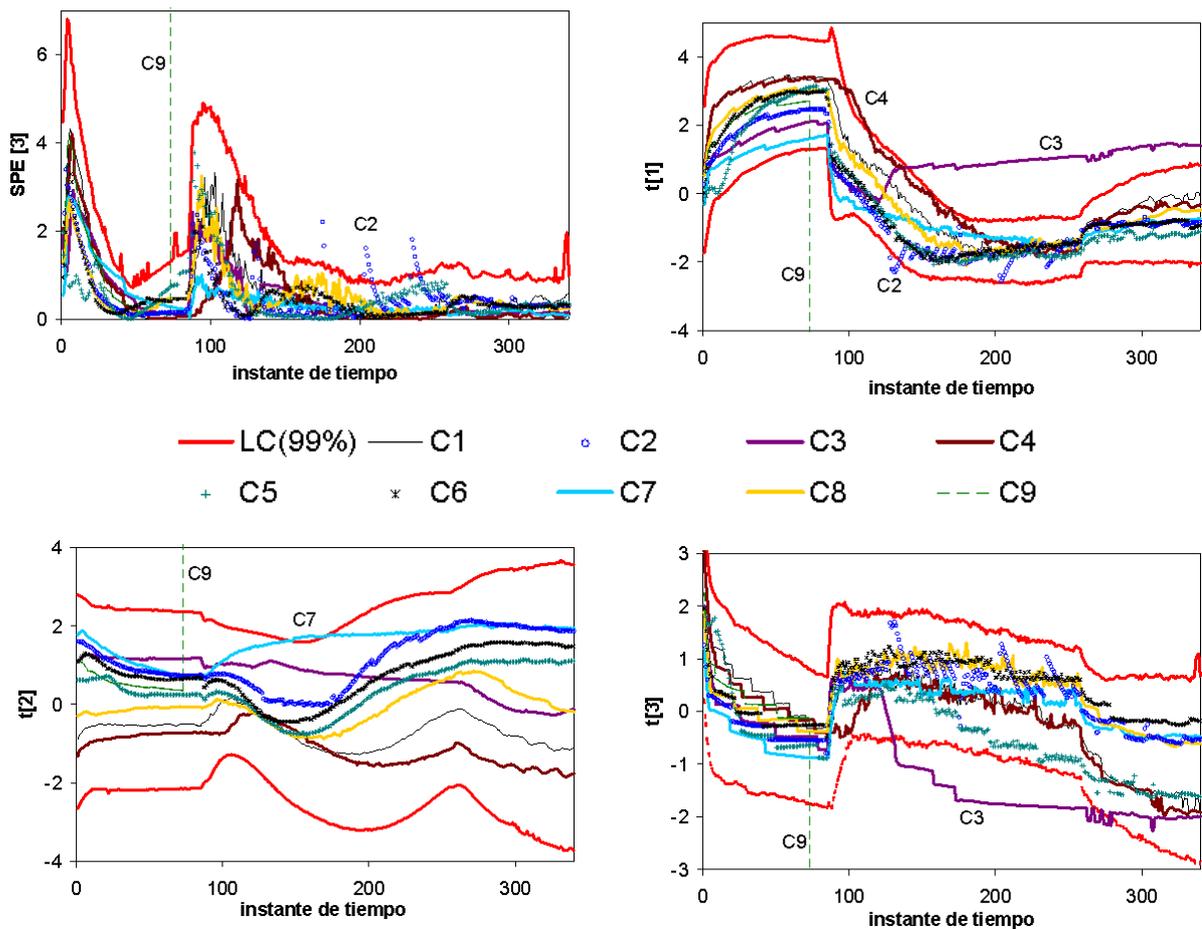


Figura 4.28. Gráficos de control de la distancia al modelo (SPE) y de las tres variables latentes para los lotes de comparación del nivel de observación de la primera modificación propuesta para el enfoque de Wold *et al.* (1998): utilización de PCA en lugar de PLS.

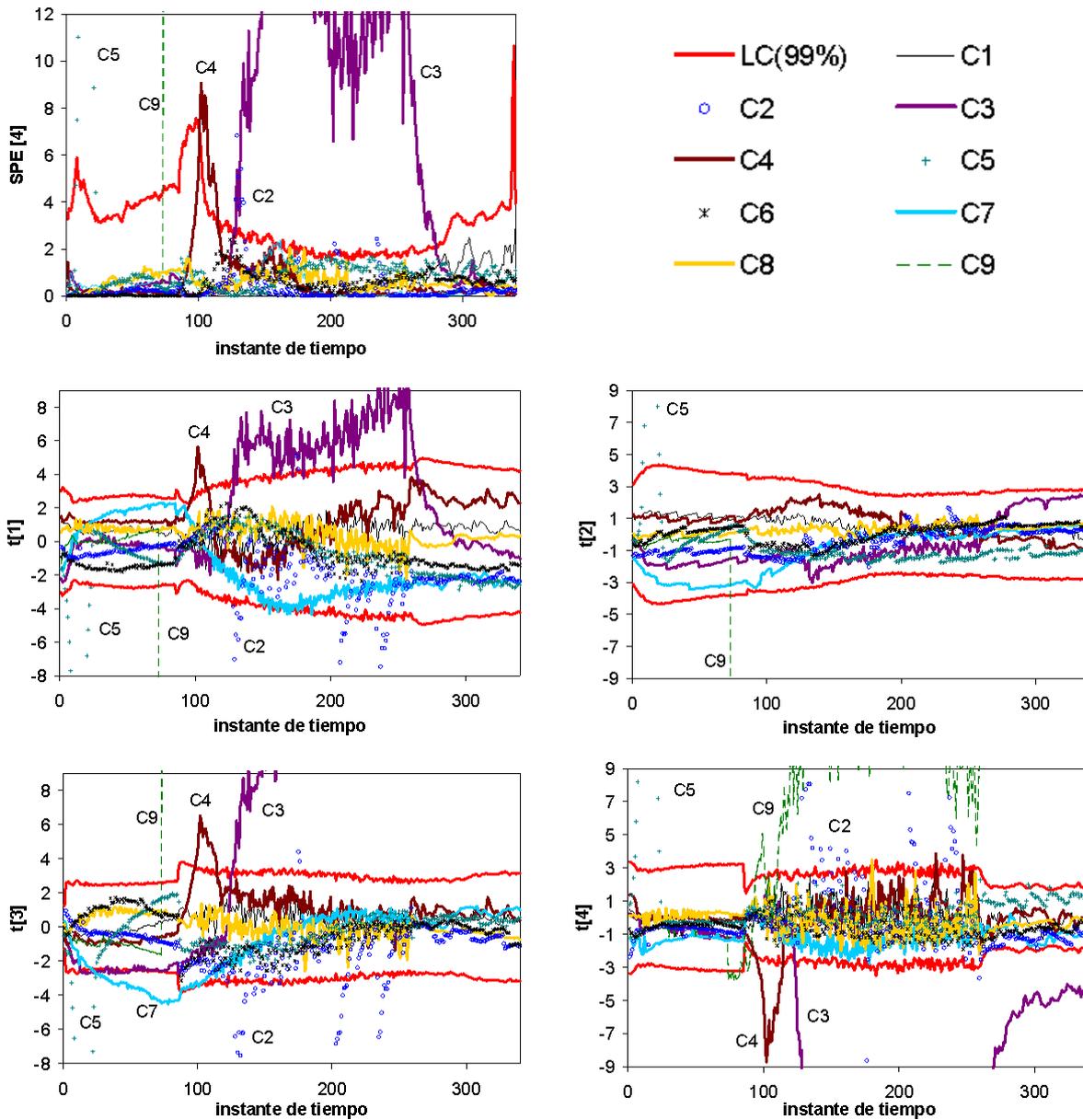


Figura 4.29. Gráficos de control de la distancia al modelo (SPE) y de las cuatro variables latentes para los lotes de comparación del nivel de observación de la segunda modificación propuesta para el enfoque de Wold *et al.* (1998): eliminar la no linealidad de los datos con el preprocesamiento y utilizar posteriormente un PCA.

En la Figura 4.30 se presentan los gráficos de control correspondientes al nivel *batch* de la T^2 de Hotelling y de la distancia al modelo (DmodX) con cada una de las dos alternativas para los 9 lotes proyectados. Los límites de control (del 99%) que aparecen en estos gráficos son los que se han obtenido a partir de la base de datos de referencia (70 lotes pertenecientes a la clase 1). En este nivel, tanto los enfoques originales como las alternativas sugeridas presentan resultados similares, siendo

capaces de identificar la mayor parte de los lotes con anomalías. Comparando la Figura 4.30 con la Figura 4.15, se puede apreciar que con el nivel *batch* de la primera alternativa (Figura 4.30 a) se obtienen resultados muy parecidos a los obtenidos con el nivel *batch* del enfoque de Wold *et al.* (1998) (Figura 4.15 b), mientras que los del nivel *batch* de la segunda alternativa (Figura 4.30 b) son más parecidos a los del enfoque de Nomikos y MacGregor (1995) (Figura 4.15 a).

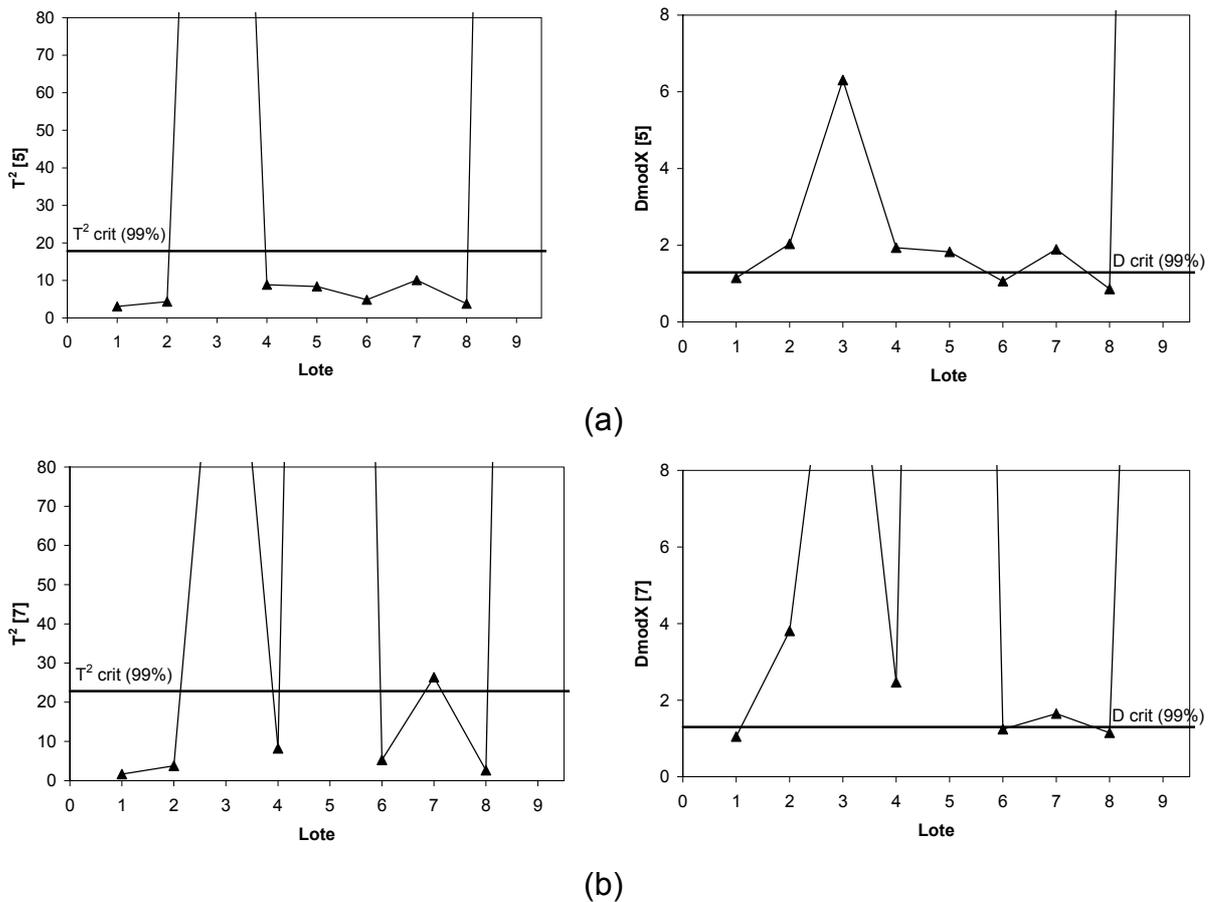


Figura 4.30. Gráficos de control correspondientes al nivel *batch* de la T^2 de Hotelling y de la distancia al modelo (DmodX) para los lotes de comparación. (a) alternativa 1: emplear un PCA en lugar de un PLS (b) alternativa 2: eliminar la no linealidad en el preprocesamiento y utilizar posteriormente un PCA. (Nota: algunos valores de los estadísticos T^2 y DmodX en los lotes representados quedan por encima del límite de control y no aparecen en el gráfico porque se ha limitado el eje de ordenadas para poder apreciar el resto de lotes en relación a los límites de control).

Finalmente, y de manera similar a como se procedió en el apartado anterior, se resumen los resultados de las proyecciones de los nueve lotes en la Tabla 4.10. En esta tabla, T^2 indica que las anomalías de dicho lote se reflejan en el modelo en la T^2

de Hotelling, D_{mod} en la distancia al modelo, $t[i]$ en la componente i -ésima del nivel de observación, y en aquellos casos en los que el modelo no detecta ninguna salida de control se indica que está “Bajo control”. Cuando el lote es identificado correctamente y además los gráficos de contribución señalan acertadamente las variables responsables, se indica en la tabla con una “B”, cuando sólo es identificado correctamente con una “R”, y en los casos que no llegue a identificarse correctamente con una “X”. Este último caso tiene lugar tanto cuando un lote con comportamiento normal es detectado por el modelo como anómalo, como cuando ocurre lo contrario, es decir que un lote con comportamiento anómalo es identificado erróneamente por el modelo como un lote bajo control.

Lote	Nivel de Observación				Nivel <i>Batch</i>			
	Alternativa 1		Alternativa 2		Alternativa 1		Alternativa 2	
C1	Bajo control	B	Bajo control	B	Bajo control	B	Bajo control	B
C2	$D_{mod}, t[1]$	R	$D_{mod}, t[1], t[3], t[4]$	B	D_{mod}	B	D_{mod}	B
C3	$t[1], t[3]$	B	$D_{mod}, t[1], t[3], t[4]$	B	D_{mod}, T^2	B	D_{mod}, T^2	B
C4	$t[1]$	B	$D_{mod}, t[1], t[3], t[4]$	B	D_{mod}	B	D_{mod}	B
C5	Bajo control	X	$D_{mod}, t[1], t[2], t[3], t[4]$	B	D_{mod}	R	D_{mod}, T^2	B
C6	Bajo control	X	Bajo control	X	Bajo control	X	Bajo control	X
C7	$t[2]$	R	$t[3]$	B	D_{mod}	B	D_{mod}, T^2	B
C8	Bajo control	B	Bajo control	B	Bajo control	B	Bajo control	B
C9	$D_{mod}, t[1], t[2], t[3]$	B	$D_{mod}, t[1], t[2], t[3], t[4]$	B	D_{mod}, T^2	B	D_{mod}, T^2	B

Tabla 4.10. Resultados obtenidos en los dos niveles de monitorización de las dos modificaciones propuestas del enfoque de Wold *et al.* (1998) para los nueve lotes seleccionados como representativos de las categorías definidas en la Tabla 4.3.

Teniendo en cuenta todos los resultados que se han presentado en este apartado, se puede concluir que la segunda alternativa propuesta, en la que previamente a la aplicación del PCA se realiza el preprocesamiento de manera análoga al enfoque de Nomikos y MacGregor (1995), representa para el conjunto de datos analizados una importante mejora respecto de la capacidad de detección de lotes con comportamiento anómalo del nivel de observación del enfoque propuesto por Wold *et al.* (1998). El problema está, por tanto, en que el enfoque original no elimina la no linealidad presente en los datos.

4.2 MONITORIZACIÓN EN TIEMPO REAL DEL PROCESO

En este apartado se comparan distintas técnicas de análisis multivariante para la monitorización en tiempo real del transcurso de un nuevo lote. A diferencia del análisis realizado en el apartado 4.1.2, en el cual se han comparado distintos enfoques para detectar anomalías y diagnosticar las causas de lotes ya finalizados en los que, por tanto, se conoce la evolución de la trayectoria de todas las variables a lo largo de todo el lote, en este apartado se emplean distintas alternativas para monitorizar cómo está transcurriendo un lote. Es decir, en este caso, en cada instante de tiempo k se dispone de todas las variables registradas hasta dicho instante para comprobar si la evolución del lote es consistente con los lotes empleados para construir el modelo y que constituyen la distribución de referencia.

En el apartado 1.8.6.3 se describieron detalladamente distintos enfoques para abordar la monitorización en tiempo real de un proceso por lotes. Las estrategias que se han analizado en este apartado son las siguientes: la metodología propuesta por Wold *et al.* (1998), una de las alternativas de esta metodología que se propuso en el apartado 4.1.2.4 (la alternativa 2), la construcción de múltiples modelos de referencia (tantos modelos como instantes de tiempo tienen los lotes), y los tres métodos de imputación para la parte futura del lote sugeridos por Nomikos y MacGregor (1995):

- asumir que la parte desconocida tendrá una desviación nula respecto al comportamiento medio.
- asumir que la parte desconocida tendrá una desviación constante e igual a la desviación del último instante de tiempo registrado.
- tratar la parte desconocida como datos faltantes.

En la monitorización de procesos por lotes es habitual utilizar el tiempo de reacción para evaluar la velocidad de detección de fallos de cada enfoque de monitorización.

El tiempo de reacción se define como el tiempo que transcurre entre el momento en que se produce una anomalía o perturbación en el proceso hasta que el sistema de monitorización lo detecta. En el contexto del control estadístico de procesos se dice que el proceso está fuera de control estadístico cuando al menos un número predeterminado de observaciones consecutivas están fuera del correspondiente límite de control. La elección de cuántas observaciones definen la señal de fuera de control depende de cada proceso en concreto. En consecuencia, se establece una clara distinción entre cuando la señal está fuera de control, definida en este caso como la aparición de una serie de observaciones consecutivas rebasando el límite de control, y cuando se trata de un aviso, definido como la presencia de una única observación por fuera del límite en el correspondiente gráfico de control.

Para la construcción de todos los modelos de referencia se ha utilizado el mismo conjunto de lotes que se definió en el apartado 4.1.2, es decir, los 70 lotes pertenecientes a la clase 1 (Tabla 4.3 en página 137).

Con el objetivo de comparar las seis estrategias de monitorización se han seleccionado los lotes C2, C3 y C5 (Tabla 4.3) que presentan anomalías en su comportamiento, y también el lote C8 que presenta un comportamiento normal pero que no ha sido utilizado en la construcción de los modelos y que se trata, por tanto, de un lote de validación. Se ha escogido el lote C3 por presentar un fallo que puede perjudicar gravemente al proceso de eliminación biológica de fósforo ya que el sistema de aireación falla durante la etapa aerobia quedando el reactor en condiciones anaerobias, mientras que los otros 2 lotes (C2 y C5) se seleccionaron por ser representativos de las anomalías que con mayor frecuencia aparecen en la base de datos analizada.

4.2.1 Modelación del proceso

Los modelos de referencia desarrollados para la monitorización con el enfoque de Wold *et al.* (1998), con la segunda alternativa de este enfoque propuesta en el apartado 4.1.2.4, y con los tres métodos de imputación sugeridos por Nomikos y MacGregor (1995) ya han sido descritos en el apartado de diagnóstico de fallos

(apartado 4.1.2). En estos cinco enfoques para la monitorización en tiempo real del proceso se utiliza un modelo global, es decir, un único modelo para todos los instantes de tiempo.

En el último enfoque que se analiza (construcción de múltiples modelos de referencia), no se utiliza un modelo global sino que se construyen tantos modelos PCA como instantes de tiempo tiene un lote (K), y en cada modelo se incluyen todas las variables medidas desde el inicio del lote hasta el instante de tiempo actual (k). La principal ventaja que presenta este enfoque es que no necesita recurrir a ningún tipo de imputación, aunque se solía presentar como inconveniente el alto coste computacional que tiene el realizar y almacenar tantos modelos. Sin embargo, dada la capacidad de los ordenadores actuales, esto ya no supone un obstáculo para que este enfoque pueda ser utilizado en la práctica.

Por otra parte, sí que se puede considerar como un inconveniente de este enfoque el hecho de que para cada modelo se tiene que realizar un análisis detallado para determinar cuántas variables latentes son necesarias para describir la información presente en los datos. El objetivo del presente apartado no era realizar este análisis detallado y, por tanto, se decidió utilizar el mismo número de variables latentes que el modelo global, es decir siete variables latentes en cada uno de los 340 modelos. Sin embargo, debido a que en los modelos desarrollados para los primeros instantes de tiempo no se obtiene ninguna variable latente por validación cruzada, se utilizaron solamente dos variables latentes. En la Tabla 4.11 se muestra el número de variables latentes extraídas para cada modelo y el criterio para la selección de las mismas.

Instantes de tiempo	Nº comp	Criterio
1-2	2	Forzadas.
3-7	5	Validación cruzada y
8-10	6	valor propio mayor que la
11-340	7	unidad.

Tabla 4.11. Número de variables latentes extraídas para cada uno de los 340 modelos del enfoque de múltiples modelos de referencia.

4.2.2 Resultados de la monitorización con distintos enfoques

En general, la monitorización en tiempo real de un proceso por lotes se realiza utilizando dos tipos de gráficos de control: el gráfico de la T^2 que permite detectar variaciones inusuales dentro del modelo, y el gráfico de la distancia al modelo que permite detectar anomalías que no respetan la estructura de correlaciones del modelo. También es posible construir gráficos de control para cada una de las variables latentes, aunque como se indicó en el apartado 1.8.6.2.2, es habitual emplear el gráfico de la T^2 para su monitorización, y en caso de detectar alguna anomalía realizar los correspondientes gráficos de contribución (apartado 1.8.6.2.3) para identificar en primer lugar las variables latentes y posteriormente las variables registradas responsables de la salida de control.

En las siguientes figuras (Figura 4.31 hasta Figura 4.36) se muestra cómo transcurriría la monitorización en tiempo real de los lotes seleccionados C2, C3, C5 y C8 utilizando cada uno de los seis enfoques analizados. En estos gráficos, para el cálculo de la distancia al modelo se han considerado únicamente las desviaciones instantáneas en el espacio residual, es decir, el error en cada instante de tiempo (k). De esta manera, se pueden detectar desviaciones instantáneas en los residuos, lo que incrementa la capacidad de detección de anomalías de este gráfico de control ya que se evita el efecto de promediado que se podría producir si se consideraran también los residuos de los instantes de tiempo anteriores. Los límites de control que aparecen en las figuras se han obtenido a partir de la base de datos de referencia (70 lotes de la clase 1), utilizando la aproximación de Jackson y Mudholkar (1979) (Ec. 1.10) para el límite de la distancia al modelo (SPE), y la expresión de Tracy *et al.* (1992) (Ec. 1.15) para el límite de la T^2 . Se debe tener presente que cuando se emplea el mismo modelo para monitorizar el transcurso de todo el lote, el límite de control de la T^2 es constante puesto que depende del número de variables latentes del modelo (y de los grados de libertad residuales), pero el límite de control de la distancia al modelo cambia con el tiempo porque para su cálculo se emplean en cada instante de tiempo (k) los errores del modelo en dicho instante de tiempo.

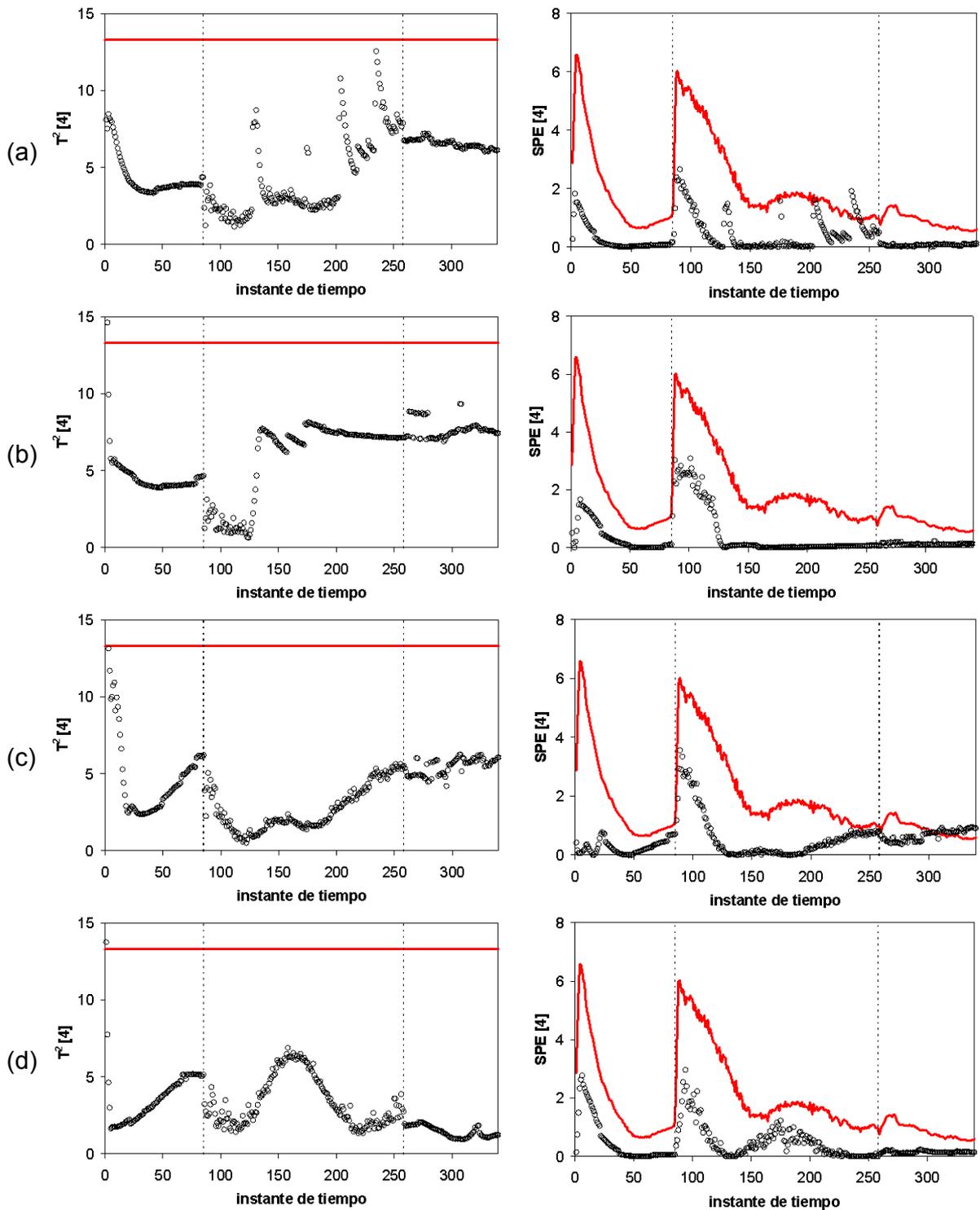


Figura 4.31. Monitorización en tiempo real de la T^2 y de la distancia al modelo (SPE) con el enfoque de Wold *et al.* (1998). (a) lote C2 (b) lote C3 (c) lote C5 y (d) lote de validación (C8). Límites de control al 99% (línea continua).

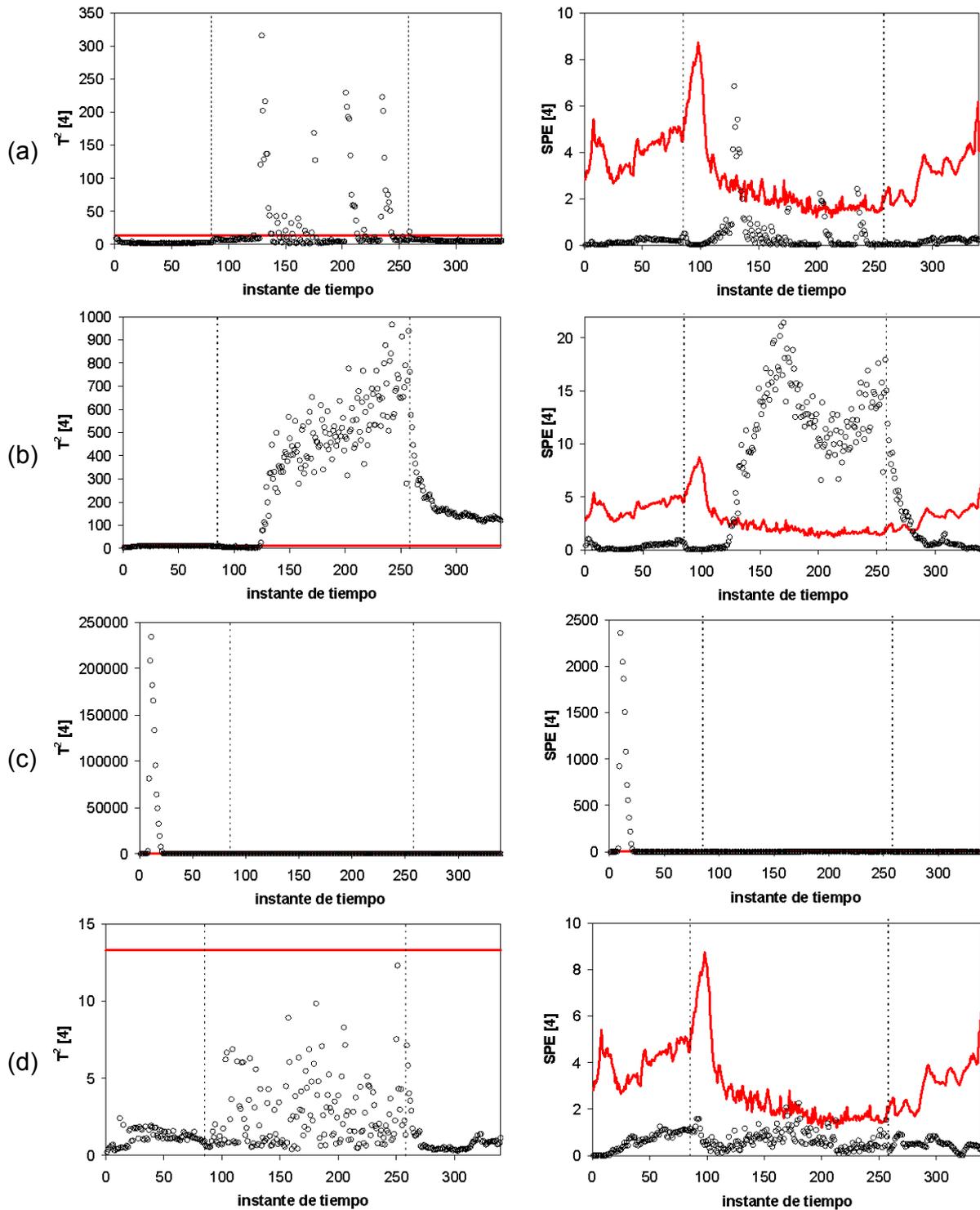


Figura 4.32. Monitorización en tiempo real de la T^2 y de la distancia al modelo (SPE) mediante la segunda modificación del enfoque de Wold *et al.* (1998) propuesta en el apartado 4.1.2.4: eliminar la no linealidad de los datos en el preprocesamiento y posteriormente aplicar un PCA. (a) lote C2 (b) lote C3 (c) lote C5 y (d) lote de validación (C8). Límites de control al 99% (línea continua).

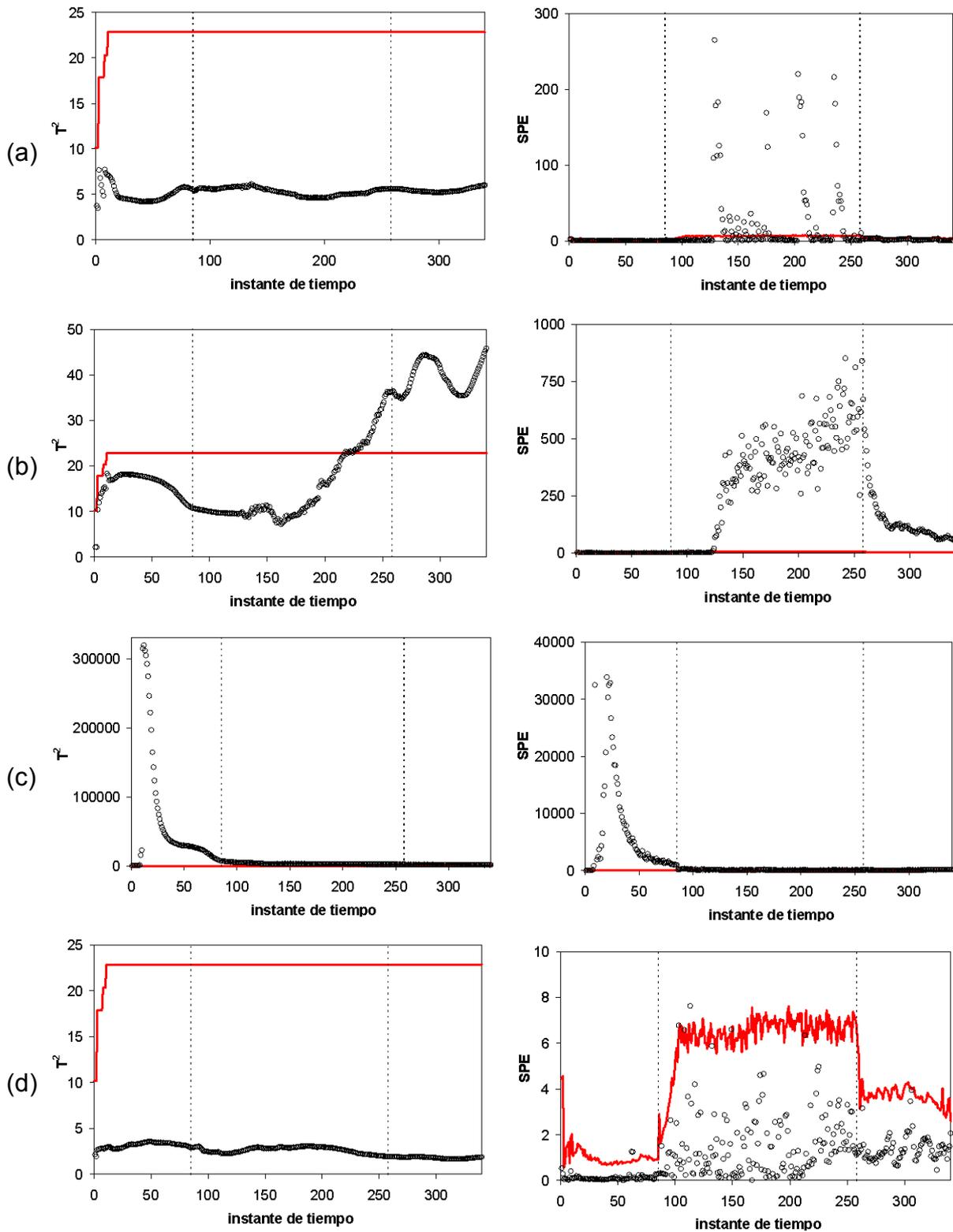


Figura 4.33. Monitorización en tiempo real de la T^2 y de la distancia al modelo (SPE) con el enfoque de Nomikos y MacGregor (1995) y construcción de múltiples modelos de referencia. (a) lote C2 (b) lote C3 (c) lote C5 y (d) lote de validación (C8). Límites de control al 99% (línea continua).

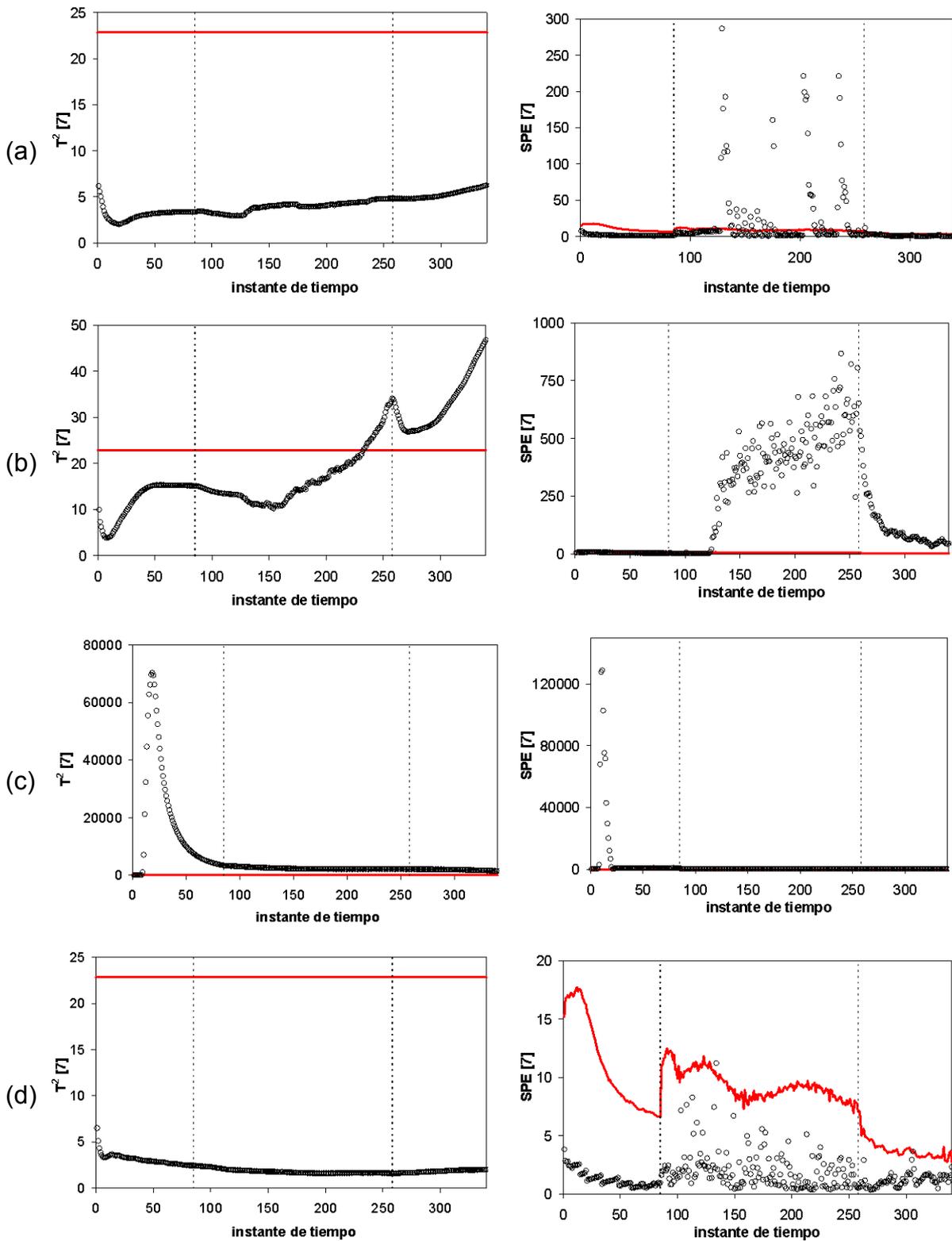


Figura 4.34. Monitorización en tiempo real de la T^2 y de la distancia al modelo (SPE) con el enfoque de Nomikos y MacGregor (1995) e imputación con desviación nula de la trayectoria futura del lote. (a) lote C2 (b) lote C3 (c) lote C5 y (d) lote de validación (C8). Límites de control al 99% (línea continua).

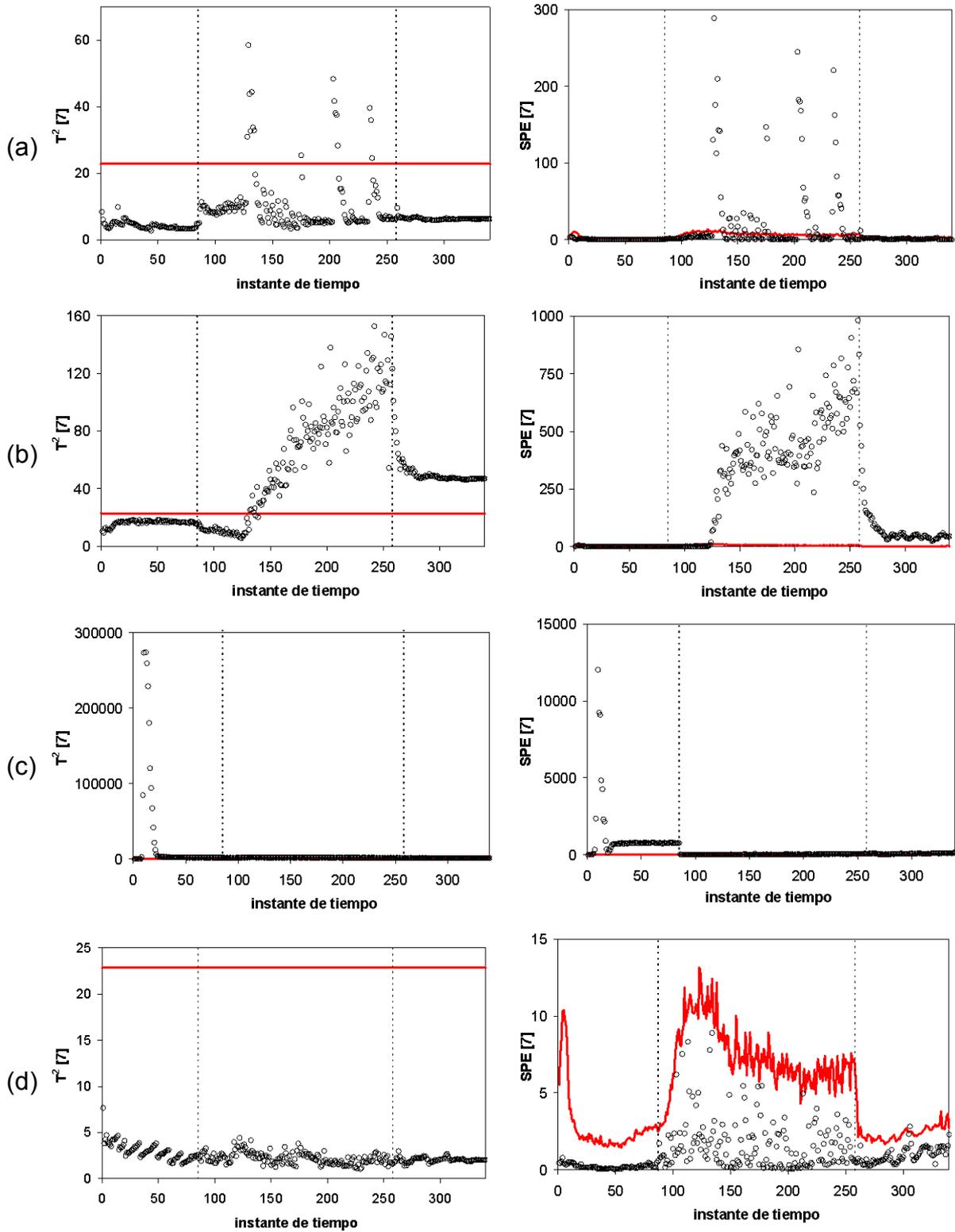


Figura 4.35. Monitorización en tiempo real de la T^2 y de la distancia al modelo (SPE) con el enfoque de Nomikos y MacGregor (1995) imputando la trayectoria futura del lote con desviación constante de valor igual a la de la última observación. (a) lote C2 (b) lote C3 (c) lote C5 y (d) lote de validación (C8). Límites de control al 99% (línea continua).

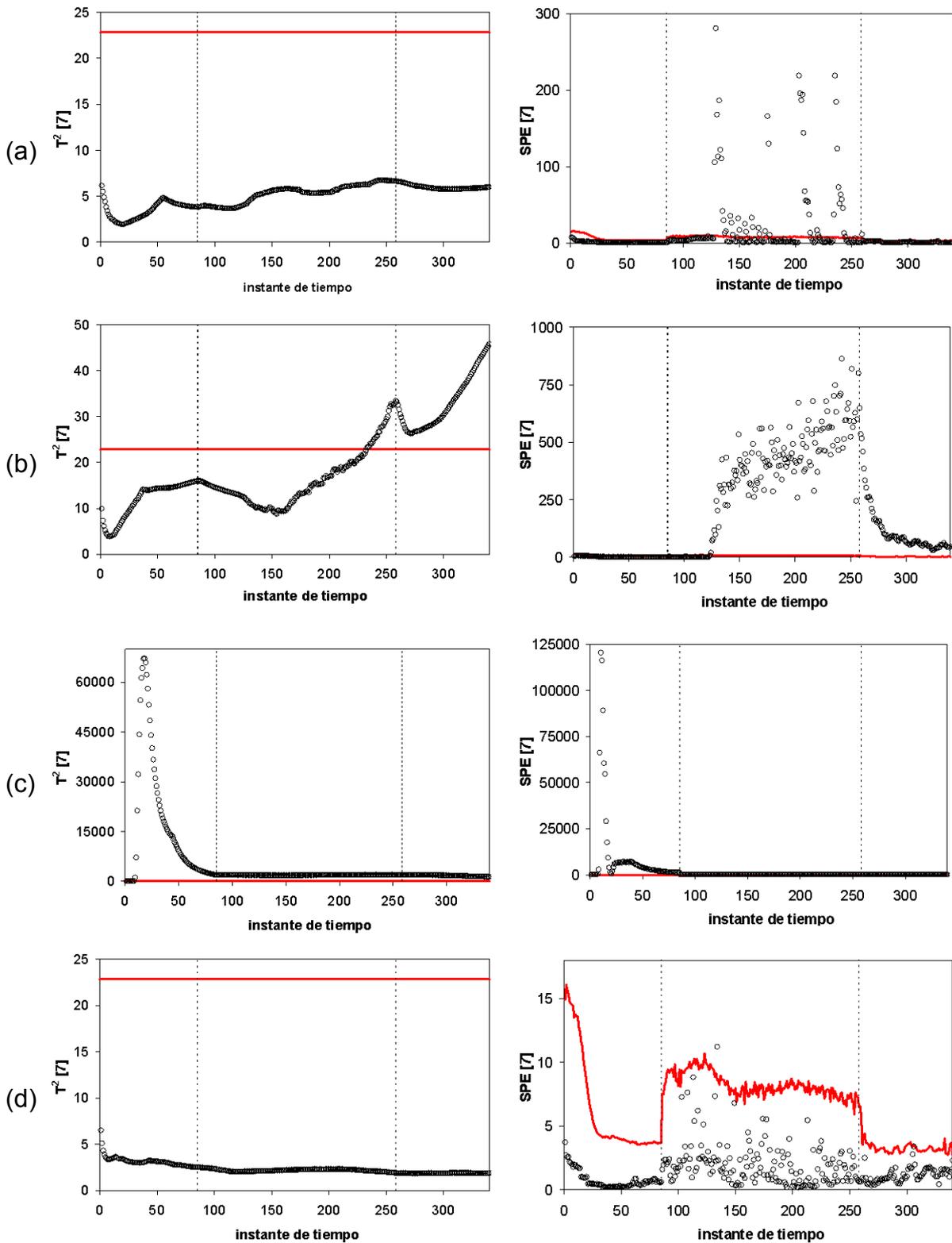


Figura 4.36. Monitorización en tiempo real de la T^2 y de la distancia al modelo (SPE) con el enfoque de Nomikos y MacGregor (1995) tratando como datos faltantes la trayectoria futura del lote. (a) lote C2 (b) lote C3 (c) lote C5 y (d) lote de validación (C8). Límites de control al 99% (línea continua).

Cuando se está monitorizando un nuevo lote, el interés reside, desde un punto de vista práctico, en detectar el comportamiento atípico lo más pronto posible sin importar en qué gráfico (T^2 o distancia al modelo) aparece la señal de alarma, si bien el gráfico en el cual aparece, puede aportar información útil para el diagnóstico de las causas de la anomalía.

En los gráficos de la distancia al modelo (SPE) mostrados, se aprecia con claridad que las anomalías presentes en los lotes C2, C3 y C5 son detectadas por todos los modelos, con la única excepción del modelo que ha sido desarrollado siguiendo la metodología propuesta por Wold *et al.* (1998). Este modelo ha sido capaz de detectar que se producía alguna anomalía en los lotes C2 y C5. Aunque en el caso del lote C5, la salida de control detectada se corresponde con una falsa alarma, como se comprobó en el apartado 4.1.2.3, puesto que la anomalía en dicho lote se produjo al comienzo del lote y no hacia el final del mismo. El resto de modelos ha presentado un comportamiento satisfactorio, manifestando rápidamente la presencia de las alteraciones en el comportamiento de los lotes estudiados. Entre estos modelos no ha habido diferencias en el tiempo de reacción para la detección de las anomalías presentes en los tres lotes analizados. No se ha indicado el tiempo de reacción puesto que en el mismo momento en que la perturbación es identificada en el proceso también es detectada en el gráfico de la distancia al modelo.

Los gráficos de la T^2 mostrados presentan, en general, una capacidad de detección de fallos inferior a la que presentan los gráficos de la distancia al modelo. El método de imputación que asume una desviación constante y la segunda modificación que se propuso en el apartado 4.1.2.4 para el enfoque Wold *et al.* (1998), son los únicos que presenta salidas de control en el gráfico de la T^2 para los tres lotes anómalos proyectados. Sin embargo, tal y como se pone de manifiesto en las figuras anteriores, la detección de las anomalías resulta más fiable y se produce con mayor antelación en los gráficos de la distancia al modelo. Estos resultados son coherentes con los obtenidos por Ramaker y Sprang (2004) quienes señalan que en los procesos por lotes, la mayor parte de fallos en el proceso se manifiestan en el gráfico de la distancia al modelo, mientras que son pocas las ocasiones en que éstos se reflejan sólo en el gráfico de la T^2 , por lo que concluyen que este último

gráfico resulta de escasa utilidad para la detección de fallos en tiempo real en procesos por lotes.

Comparando los gráficos de la T^2 del enfoque Wold *et al.* (1998) (Figura 4.31), y de la segunda modificación que se propuso para dicho enfoque (Figura 4.32), se puede observar que, mientras el enfoque original no es capaz de detectar ningún lote como anómalo, la alternativa propuesta detecta con claridad todos los lotes con comportamiento atípico. Sin embargo, tal y como se comprobó en el apartado 4.1.2.3, el lote C3 es perfectamente detectado en los gráficos de control de las dos primeras variables latentes del enfoque de Wold *et al.* (1998) (Figura 4.16). Por tanto, cuando se emplea este enfoque puede suceder que los gráficos de control de las variables latentes detecten anomalías que no son detectadas en el gráfico de control de la T^2 , lo cual no es adecuado desde el punto de vista del esquema general de monitorización del proceso. Con el objetivo de determinar las causas de este comportamiento del nivel de observación del enfoque de Wold *et al.* (1998), se analizó la trayectoria de las variables latentes y se llegó a la conclusión de que el problema es debido a una sobreestimación de la trayectoria de la varianza de las variables latentes por parte del valor propio de las mismas. Para poder apreciar de forma visual este problema, se presenta en la Figura 4.37 la evolución de la trayectoria de la varianza de las dos primeras variables latentes del nivel de observación tanto del enfoque de Wold *et al.* (1998) como de la segunda alternativa propuesta en el apartado 4.1.2.4, junto con el valor propio de estas dos variables latentes en ambos casos. En dicha figura se puede observar cómo el problema de la sobreestimación no ocurre cuando se ha utilizado la alternativa sugerida para el enfoque original.

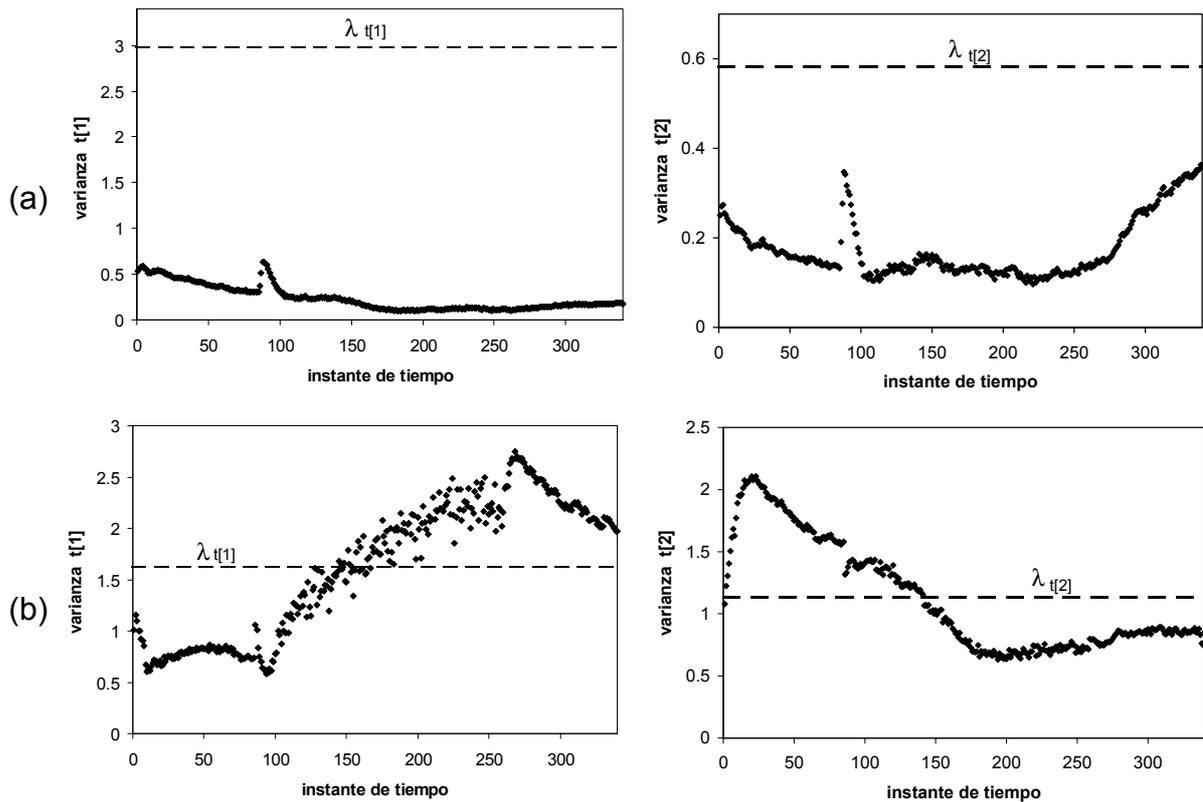


Figura 4.37. Evolución de la trayectoria de la varianza de las dos primeras variables latentes y valor propio (λ) de las mismas: (a) nivel de observación del enfoque de Wold *et al.* (1998) (b) nivel de observación de la segunda modificación del enfoque de Wold *et al.* (1998) propuesta en el apartado 4.1.2.4.

Una vez se ha detectado que en el transcurso de un lote se han producido desviaciones respecto de las condiciones normales de operación, se puede conocer mediante la utilización de los gráficos de contribución qué variables son las que más contribuyen a que se haya producido la salida de control. A modo de ejemplo, se presenta únicamente en la Figura 4.38 la evolución del gráfico de contribución de la distancia al modelo durante el transcurso del lote C3, cuando se emplea para la monitorización el enfoque con múltiples modelos de referencia. En esta figura se puede observar que el gráfico de contribución apunta a la variable oxígeno como responsable de la salida de control en el gráfico de la distancia al modelo. Además, como la contribución de esta variable es negativa, nos indica que su valor está por debajo de lo esperable para un lote normal.

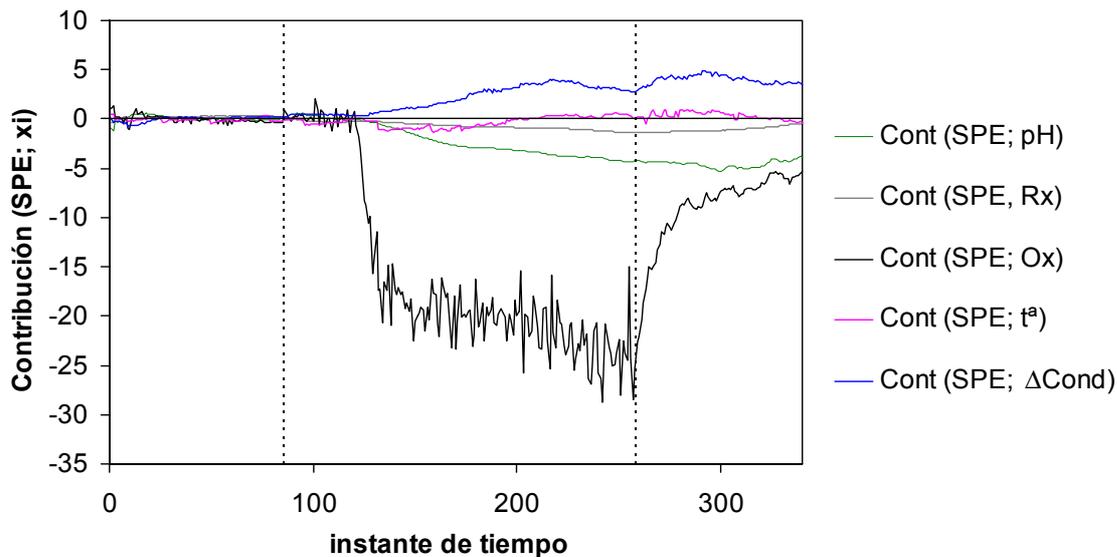


Figura 4.38. Evolución del gráfico de contribución a la distancia al modelo (SPE) a lo largo del lote C3, empleando el enfoque de múltiples modelos de referencia para su monitorización en tiempo real.

En este apartado se han aplicado seis estrategias distintas para la monitorización en tiempo real del proceso de eliminación biológica de fósforo estudiado, de éstas la mitad no requieren ningún tipo de imputación de datos, son la metodología desarrollada por Wold *et al.* (1998), la segunda modificación sugerida en el apartado 4.1.2.4 de la metodología anterior y la utilización de múltiples modelos de referencia. Las conclusiones más relevantes que se pueden extraer son las siguientes:

- Para el conjunto de datos analizado, el enfoque de Wold *et al.* (1998) es el que ha presentado una capacidad más limitada para la detección de las anomalías presentes en los lotes estudiados. Además, se ha encontrado que en este enfoque se ha producido una sobreestimación de la trayectoria de la varianza de las variables latentes, lo que ha provocado que el gráfico de control de la T^2 no haya sido capaz de detectar anomalías que habían sido previamente detectadas en los gráficos de control de las variables latentes (apartado 4.1.2.3). Por otra parte, la alternativa de mejora para este enfoque propuesta en el apartado 4.1.2.4, se ha comportado mucho mejor, detectando todas las anomalías tanto en el gráfico de la distancia al modelo como en el de la T^2 . Utilizando esta alternativa, no se ha producido el problema de la sobreestimación de la trayectoria de la varianza de las variables latentes.

- A partir de los resultados de todos los modelos, se puede afirmar que el gráfico de la distancia al modelo presenta una mayor capacidad para la detección de fallos y, además, refleja la anomalía con anterioridad a su aparición en el gráfico de la T^2 .
- Teniendo en cuenta que lo más importante es detectar la anomalía lo antes posible, al comparar los distintos enfoques en base al gráfico de la distancia al modelo se ha puesto de manifiesto que:
 - Entre los distintos métodos de imputación no se han producido diferencias apreciables; todos han identificado correctamente e inmediatamente las anomalías.
 - El enfoque que emplea múltiples modelos de referencia no ha representado una ventaja sobre los métodos de imputación. Por tanto, debido a que desde un punto de vista práctico resulta más sencillo y eficiente emplear un único modelo, se podría emplear para la monitorización en tiempo real del sistema analizado tanto el enfoque de Nomikos y MacGregor (1995) recurriendo a cualquiera de los procedimientos de imputación analizados para completar la parte futura del lote, como la segunda modificación sugerida para el enfoque de Wold *et al.* (1998). La principal ventaja que tendría este último enfoque es que no sería necesario rellenar la parte futura del lote.

4.3 ESQUEMA DE MONITORIZACIÓN PROPUESTO

El objetivo del presente apartado es proponer un esquema de monitorización para el sistema estudiado basado en técnicas de proyección multivariante, que permita tanto comprobar en tiempo real si los lotes están progresando adecuadamente como diagnosticar también en tiempo real los fallos que se estén produciendo.

La monitorización y el diagnóstico en tiempo real del proceso es fundamental desde el punto de vista de la calidad del efluente producido, puesto que permitirá detectar

fallos que pueden ser corregidos antes de la finalización del lote actual y/o que pueden ser evitados para los lotes siguientes.

El proceso por lotes estudiado es un sistema SBR para la eliminación biológica de fósforo en las aguas residuales y, como todo proceso biológico de tratamiento de aguas residuales, está caracterizado por la dependencia histórica, es decir, la población del cultivo biológico está influida por las condiciones a las que ha estado sometida anteriormente. En el sistema SBR, cuando se inicia un nuevo lote, el agua residual influente que entra en el reactor interacciona con la biomasa que permanece en el reactor del ciclo anterior. Esto va a provocar que exista una correlación importante entre lotes, además de las autocorrelaciones y correlaciones cruzadas entre las variables a lo largo de cada lote. Por este motivo, para monitorizar el proceso se propone un esquema de monitorización con dos niveles, de manera que se pueda visualizar el transcurso en tiempo real del lote actual y también la evolución general del proceso asociada a los lotes ya finalizados.

Teniendo en cuenta los resultados obtenidos en los apartados de diagnóstico de fallos (apartado 4.1.2) y de monitorización en tiempo real (apartado 4.2), así como el hecho de que desde un punto de vista práctico resulta más sencillo y eficiente emplear un único modelo para realizar todas las funciones (monitorización y diagnóstico), se propone la utilización del enfoque de Nomikos y MacGregor (1995) empleando para la monitorización en tiempo real del proceso el método de la desviación constante. Se ha preferido esta técnica de imputación debido a que en el análisis realizado en el apartado 4.2, fue la única que reflejó las desviaciones no sólo en el gráfico de la distancia al modelo sino también en el de la T^2 . Por otra parte, aunque la modificación del enfoque de Wold *et al.* (1998) sugerida en el apartado 4.1.2.3 presentó resultados tan satisfactorios como los correspondientes a la imputación con desviación constante, no se ha propuesto esta técnica debido a que se ha decidido emplear un único modelo que permita la monitorización y diagnóstico tanto en tiempo real como “a posteriori” de lotes ya finalizados. La mayor facilidad y consistencia de los resultados del enfoque de Nomikos y MacGregor (1995) para el diagnóstico de fallos mediante la utilización de los gráficos de contribución quedó patente en el estudio del apartado 4.1.2.

En la Figura 4.39 se muestra una representación gráfica del esquema de monitorización propuesto. El progreso del lote actual sería monitorizado mediante los gráficos de control de la distancia al modelo y de la T^2 de Hotelling, y una vez finalizado el lote, este sería representado en el otro nivel en el cual cada lote constituye una observación en los correspondientes gráficos de control.

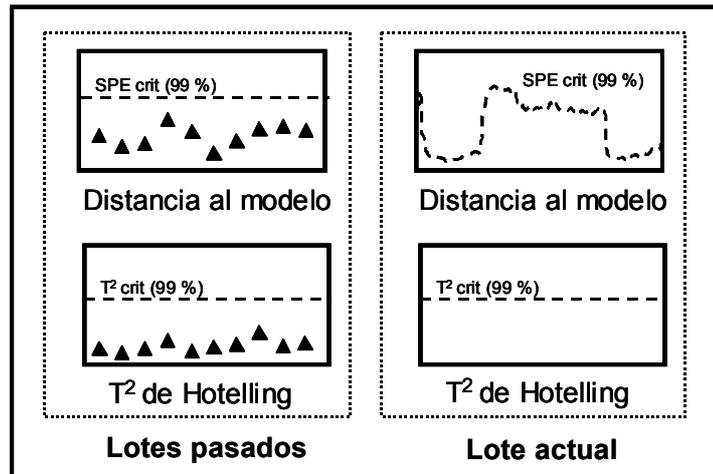


Figura 4.39. Representación del esquema de monitorización propuesto para el sistema estudiado.

El proceso que seguiría cada nueva observación se presenta en la Figura 4.40. Tal y como se refleja en dicha figura, la nueva observación, \mathbf{x}^T , sería proyectada sobre el modelo para calcular el vector de variables latentes y los residuos. Posteriormente, se analizaría en primer lugar el gráfico de la distancia al modelo, puesto que contiene información acerca de la validez del mismo y, en segundo lugar, el gráfico de la T^2 . En el caso de que se detecte alguna salida de los límites de control durante el transcurso del lote actual, mediante los gráficos de contribución se identificarían las variables responsables de la perturbación en el proceso, ayudando de esta forma al técnico a diagnosticar las causas del comportamiento anormal del lote. Por otra parte, la monitorización de los lotes ya finalizados permitiría detectar si el proceso está evolucionando hacia otro estado operacional y, por tanto, saber si el modelo empleado para monitorizar el proceso necesita ser actualizado. En caso de que fuera necesario actualizar el modelo, se utilizaría un nuevo conjunto de lotes para caracterizar las nuevas condiciones de funcionamiento del proceso y construir el modelo empírico.

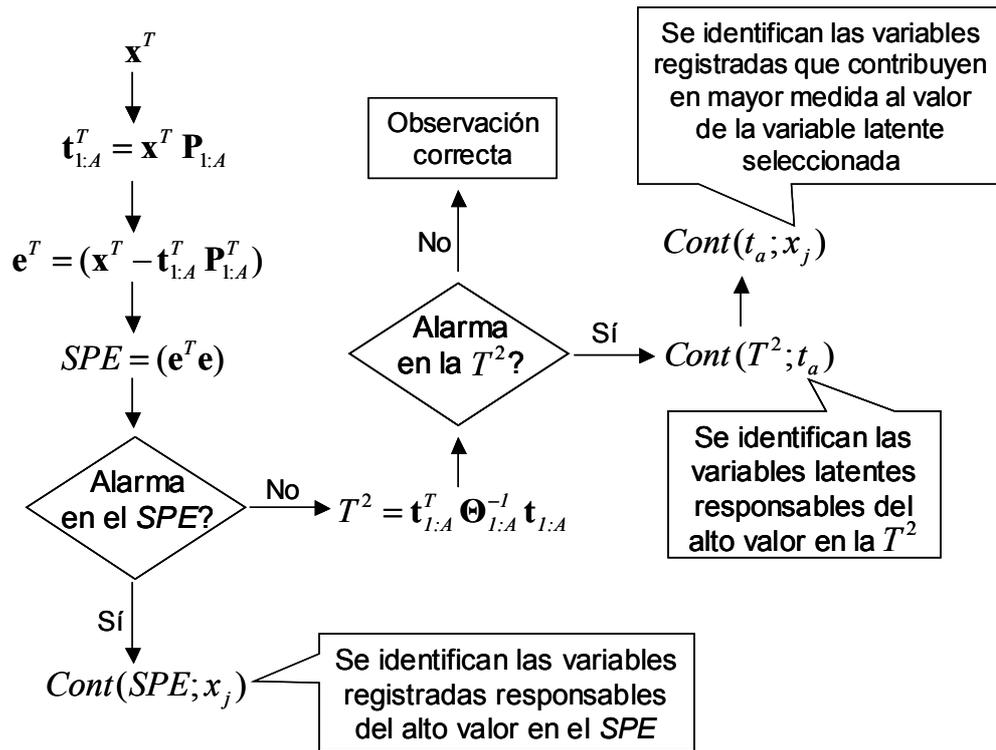


Figura 4.40. Proceso seguido para la monitorización de una nueva observación, \mathbf{x}^T , con los gráficos de la distancia al modelo (SPE) y de la T^2 de Hotelling.

Actualmente, el sistema es monitorizado mediante la representación gráfica de la evolución individual de cada una de las cinco variables registradas a lo largo del lote actual. También es posible visualizar los gráficos univariantes de cualquier lote ya finalizado. El esquema de monitorización propuesto es mucho más potente, ya que al monitorizar el lote actual mediante los estadísticos de la distancia al modelo y de la T^2 de Hotelling, se detectarán tanto alteraciones en las variables individuales como en su estructura de correlaciones, aumentando la sensibilidad del esquema de monitorización para detectar posibles fallos, y permitiendo, mediante la utilización de los gráficos de contribución, identificar las variables responsables en cada caso. Además, la monitorización de los lotes finalizados permite visualizar la evolución del proceso, sintetizando en dos gráficos de control la información contenida en los lotes pasados. Esta información sería prácticamente imposible obtenerla visualizando los perfiles univariantes en el sistema de monitorización actual. A modo de ejemplo, en quince días de operación se tendrían 300 perfiles univariantes (15 días x 4 ciclos al día x 5 variables registradas). Estas 300 trayectorias quedarían sintetizadas en 60 observaciones (15 días x 4 ciclos al día) en cada uno de los dos gráficos de control en el esquema de monitorización propuesto.

4.4 DESARROLLO Y COMPARACIÓN DE MODELOS PREDICTIVOS

Dada la importancia que tiene la dinámica del fósforo (P) y de los cationes metálicos, potasio y magnesio (K^+ , Mg^{+2}), para el control y la automatización de los procesos de eliminación biológica de fósforo, en este apartado se desarrollan y comparan diversos modelos predictivos o *soft sensors* para la estimación de dichos parámetros en base a las medidas que proporcionan las sondas instaladas en el reactor. Se desarrollan tanto modelos de proyección sobre estructuras latentes como modelos de redes neuronales artificiales.

Como ya se ha descrito en el apartado 3.1, la instrumentación del reactor permite registrar en tiempo real la concentración de oxígeno disuelto, el pH, la conductividad, el potencial redox y la temperatura. Estos sensores, en comparación con los instrumentos para la determinación de nutrientes, son robustos frente a las condiciones adversas que se desarrollan en los procesos de tratamiento de aguas residuales, requieren menos labores de mantenimiento y son de bajo coste, motivos por los que son utilizados en la mayoría de EDARs.

Para el desarrollo de los modelos se dispone de un total de 20 experimentos exhaustivos pertenecientes a tres siembras distintas (Figura 4.41). Es decir, a lo largo del período de tiempo que abarcan los 20 experimentos, la planta piloto fue puesta en marcha en tres ocasiones inoculando el reactor con fango procedente de la EDAR de Quart Benàger (Valencia). Como se ha descrito en el apartado 3.3, la realización de estos experimentos consistía en tomar una muestra cada 15-20 minutos del reactor a lo largo de las etapas anaerobia y aerobia de un lote. De cada muestra se realizaban una serie de determinaciones analíticas, que habitualmente incluían: la concentración de fósforo, de magnesio, de potasio, de ácidos volátiles y la alcalinidad. Aunque el intervalo de muestreo no era constante, el instante de tiempo en que la muestra era tomada se registraba para posteriormente poder comparar las trayectorias de las variables y referir sus valores a los mismos instantes de tiempo en todas las trayectorias.

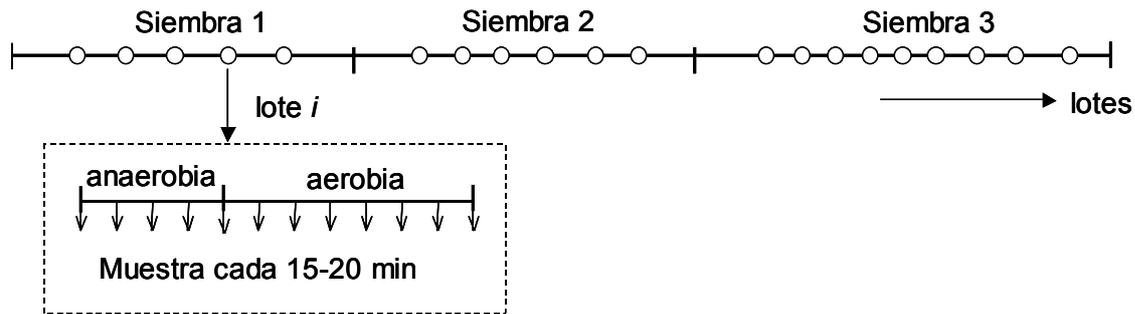


Figura 4.41. Esquema de los datos disponibles para el desarrollo de los modelos.

En la Tabla 4.12 se muestra el número de experimentos exhaustivos en los que se ha analizado cada parámetro. No en todos los experimentos se dispone de la evolución de todos los parámetros analizados a lo largo del lote, siendo la concentración de fósforo, de magnesio y de potasio las únicas determinadas en la mayoría de los experimentos. Las trayectorias de estas variables que se han medido en más experimentos son las que se han utilizado en el presente estudio, y constituirán las variables respuesta en los modelos desarrollados. De esta forma, se puede disponer de un número suficiente de datos para la construcción de los modelos y utilizar datos distintos para su validación.

	Fósforo	Magnesio	Potasio	A. volátiles	Alcalinidad
Exhaustivos	20	17	18	11	11

Tabla 4.12. Número experimentos exhaustivos en los que se analizó cada parámetro.

Como en todos los experimentos realizados no se han tomado las muestras exactamente en los mismos instantes de tiempo, se optó por utilizar valores de las trayectorias respuesta a intervalos fijos de 15 minutos, para ello fue necesario realizar en algunos casos una interpolación lineal entre cada dos valores conocidos. Esta interpolación no supuso una alteración significativa de las trayectorias de las variables respuesta, puesto que el intervalo original de muestreo oscilaba entre 15 y 20 minutos.

Para el desarrollo de los modelos se utilizaron exhaustivos de todas las siembras, de forma que estuvieran todas representadas y que quedara cubierto el rango de operación del sistema. Se seleccionaron de manera aleatoria un total de 15 exhaustivos que fueron utilizados para la construcción de los modelos. Los 5

restantes, también en este caso distribuidos en las tres siembras, se utilizaron para validar los modelos. Un procedimiento similar, pero haciendo una división del grupo validación en varios subgrupos, fue empleado por Lee y Park (1999) para construir y validar un modelo de redes neuronales con el objetivo de predecir la dinámica de los nutrientes en un SBR.

De los 5 exhaustivos que inicialmente se habían seleccionado para validar los modelos, uno de los pertenecientes a la tercera siembra no dispone de la trayectoria de conductividad. Para hacer frente a esta situación se podría optar por utilizar ese exhaustivo en la etapa de construcción de los modelos de proyección sobre estructuras latentes puesto que el algoritmo NIPALS tolera la presencia de datos faltantes. Sin embargo, para el entrenamiento de un modelo de redes neuronales no debe haber datos faltantes ni en las variables explicativas (matriz \mathbf{X}), ni en las variables respuesta (matriz \mathbf{Y}). Como se pretende hacer una comparación objetiva empleando distintos tipos de modelos predictivos, es necesario que todos los modelos se desarrollen y validen bajo las mismas condiciones. Por este motivo, se optó por utilizar los 15 exhaustivos inicialmente seleccionados para la construcción de todos los modelos, y para la validación los 4 exhaustivos que disponen de la trayectoria completa de las variables explicativas, dejando el exhaustivo restante para analizar el comportamiento de los distintos modelos cuando hay datos faltantes en las variables explicativas. En la Tabla 4.13 se indica el total de exhaustivos disponibles en cada siembra, y de éstos cuantos se destinan a la construcción de los modelos y cuantos a su validación.

	Siembra 1	Siembra 2	Siembra 3
Total exhaustivos	5	6	9
Construcción	4	5	6
Validación	1	1	3*

*De estos 3 experimentos, 2 son utilizados en el proceso de validación, y el restante para comparar el funcionamiento de los modelos desarrollados frente a la presencia de datos faltantes en las variables explicativas.

Tabla 4.13. Distribución de los experimentos exhaustivos disponibles para la construcción y validación de los modelos analizados.

4.4.1 Modelo para la predicción de todas las variables respuesta

El proceso de eliminación biológica de fósforo ha sido y es en la actualidad objeto de numerosas investigaciones. Es bien conocido que el potasio y el magnesio están asociados a la captura y liberación de fósforo por las bacterias PAO, y que forman parte de los gránulos de polifosfato intracelular. En consecuencia, es de esperar que las concentraciones de estos tres elementos estén correlacionadas, por lo que se podría plantear la realización de un único modelo para predecir las tres variables respuesta en lugar de construir tres modelos, cada uno de ellos con una única variable respuesta.

Como se ha indicado anteriormente, las tres variables respuesta no se han determinado en los 20 experimentos realizados (Tabla 4.12) y, además, uno de los exhaustivos de validación no dispone de registro de conductividad. Esto hace que la base de datos que se va a utilizar contenga datos faltantes. Por este motivo, se estudia en este apartado la posibilidad de construir un modelo de proyección sobre estructuras latentes, en concreto un modelo de regresión en mínimos cuadrados parciales (PLS).

En el apartado 4.1.1 se comprobó cómo la variabilidad entre lotes de la trayectoria de conductividad era debida fundamentalmente al distinto valor en el inicio del lote, pero que también existe una variabilidad debida a los cambios que se producen dentro de la propia trayectoria. Cuando se compararon las trayectorias de las todas las variables explicativas en los distintos experimentos se observó que presentaban un comportamiento similar aunque el valor inicial de la trayectoria era distinto en cada experimento. Por tanto, se decidió analizar en primer lugar si era posible establecer una relación entre los valores iniciales de las variables explicativas y las concentraciones de las variables respuesta al comienzo del experimento.

Para relacionar las variables explicativas con las variables respuesta se realizó un modelo PLS. Este modelo se construyó con los 15 lotes indicados en la Tabla 4.13, tiene 5 variables explicativas y tres variables respuesta, por lo que fue denominado PLS_{5,3}. Todas las variables fueron previamente centradas y escaladas a varianza

unitaria para que ninguna variable tuviera *a priori* una mayor influencia en el modelo. Se ajustó el modelo PLS con dos componentes. En la Tabla 4.14 se muestra su bondad de ajuste para las variables explicativas, R^2_x (cum), para las variables respuesta, R^2_y (cum), así como la bondad de predicción obtenida por validación cruzada, Q^2 (cum).

Componente	R^2_x (cum)	R^2_y (cum)	Q^2 (cum)
1	0.393	0.245	0.0156
2	0.582	0.388	-0.0645

Tabla 4.14. Bondad de ajuste y de predicción del modelo PLS_{5,3} para predecir las concentraciones de las variables respuesta al inicio del lote.

La baja bondad de predicción del modelo obtenido es un indicativo de que no se obtendrán buenas estimaciones de las concentraciones iniciales de las variables respuesta a partir de los valores de las variables explicativas al comienzo del lote. Para visualizar los resultados del modelo construido, se representa en la Figura 4.42 los valores predichos de las concentraciones de fósforo frente a los medidos. Sólo se han incluido las 15 observaciones empleadas en la construcción del modelo. En esta figura se pone claramente de manifiesto que el modelo desarrollado no permite obtener una estimación adecuada de la concentración de fósforo al inicio del lote.

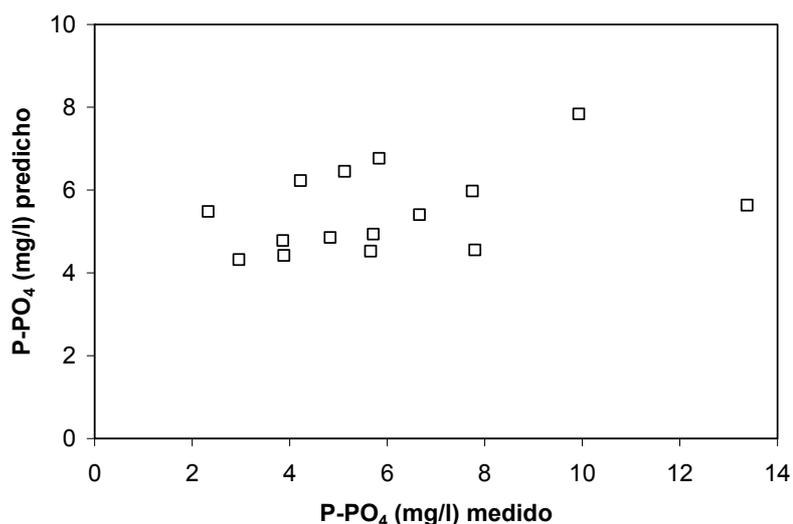


Figura 4.42. Concentraciones de fósforo determinadas analíticamente en el laboratorio al inicio de cada lote frente a los valores predichos por el modelo PLS_{5,3} que se recoge en la Tabla 4.14. Se han representado las 15 observaciones utilizadas en la construcción del modelo.

Por otra parte, con el objetivo de confirmar los resultados anteriores, se realizó un análisis de regresión lineal múltiple con cada variable respuesta estudiando la significación estadística de las variables explicativas. En todos los casos se obtuvo que ninguna variable explicativa estaba correlacionada significativamente con la variable respuesta estudiada en cada análisis con un riesgo de primera especie del 5% ($\alpha=0.05$).

Tanto el análisis de regresión lineal múltiple como los resultados del modelo PLS indican que no hay relación entre los valores al comienzo del lote de las variables explicativas y de las variables respuesta. Para eliminar esta variabilidad entre los distintos experimentos que está asociada a los valores de las variables al inicio de cada experimento, se decidió no emplear las trayectorias registradas y en su lugar utilizar el valor registrado en cada instante de tiempo menos su valor al inicio del lote. De forma semejante a como se procedió en el apartado 4.1.1 (Tabla 4.1), a partir de las trayectorias registradas se crearon las siguientes trayectorias: ΔpH , ΔRx , ΔCond y ΔTemp , donde el símbolo Δ es utilizado para hacer referencia a que la nueva trayectoria se obtiene como el valor registrado en cada instante de tiempo menos el valor registrado al inicio del lote. Como se indicó en dicho apartado, no se creó una nueva trayectoria a partir del oxígeno puesto que la trayectoria registrada no está influida por el valor de la variable al comienzo del lote. Por otra parte, para cada variable respuesta también se creó una nueva trayectoria: ΔP para el fósforo, ΔK para el potasio, y ΔMg para el magnesio. Esta forma de trabajar utilizando los valores de las variables transformados respecto de su valor al comienzo del lote fue empleada por Luccarini *et al.* (2002) para el desarrollo de modelos de redes neuronales en un sistema SBR.

Tras corregir las trayectorias de las variables, se desdobló la matriz tridireccional de datos manteniendo la dirección de los lotes (desdoblamiento tipo E, Tabla 1.3). Esto dio lugar a una matriz **X** desdoblada que estaba constituida por un total de 20 lotes y 1700 variables (5 variables explicativas x 340 instantes de tiempo registrados), y una matriz **Y** desdoblada formada por el mismo número de lotes y 54 variables (3 variables respuesta x 18 instantes de tiempo que se corresponden con un dato cada 15 minutos). Las variables de la matriz desdoblada fueron centradas y escaladas a

varianza unitaria. Como ya se indicó en el apartado 1.8.6.3.3, cuando en el desdoblamiento utilizado se mantiene la dirección de los lotes, al centrar las variables se está restando la trayectoria media y, por tanto, la principal no linealidad presente en los datos queda eliminada.

Con las variables preprocesadas se construyó el modelo $PLS_{1700,54}$ empleando los 15 exhaustivos indicados en la Tabla 4.13. En este modelo se han obtenido 5 componentes significativas con valor propio mayor que la unidad, y tiene una bondad de ajuste, $R^2_y(\text{cum})$, del 93.9% y una bondad de predicción, $Q^2(\text{cum})$, del 65.3%. Los pesos de las variables del modelo en la formación de la primera componente PLS se muestran en la Figura 4.43.

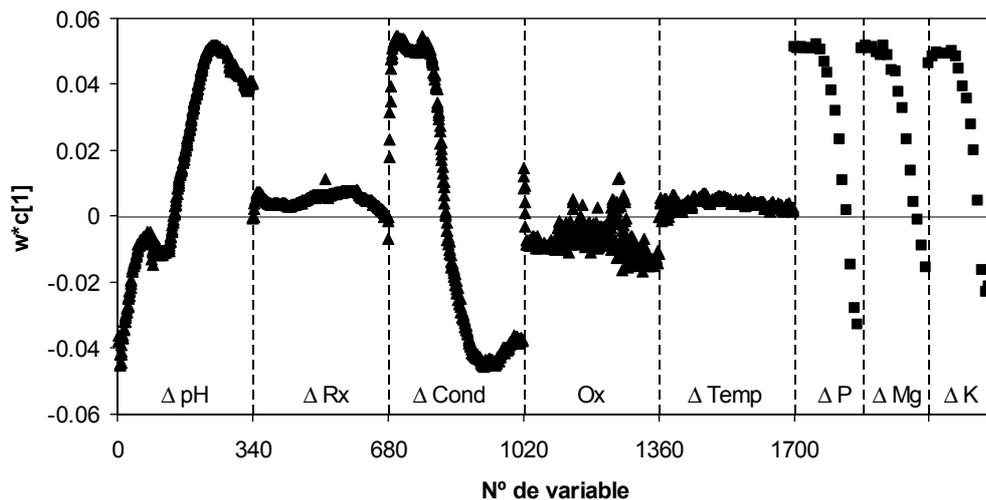


Figura 4.43. Pesos de las variables del modelo $PLS_{1700,54}$ en la formación de la primera componente PLS. La escala horizontal del gráfico correspondiente a las variables respuesta ha sido ampliada para su mejor visualización.

El análisis de los pesos mostrados en la figura anterior, pone de manifiesto dos aspectos interesantes. Por un lado, las variables que más peso (en valor absoluto) tienen en el modelo son las trayectorias corregidas de pH (ΔpH) y de conductividad (ΔCond), lo que quiere decir que éstas son las variables más importantes y que se deberán incluir en cualquier modelo predictivo que se desarrolle. Por otro lado, los pesos de las tres variables respuesta presentan un comportamiento paralelo, lo que indica una clara correlación positiva entre ellas. Además, se puede observar que estos pesos pasan de ser altos y positivos durante la etapa anaerobia, a ser negativos en la etapa aerobia, lo cual está poniendo de manifiesto que en los lotes

en los que se produjo una mayor liberación de fósforo, potasio y magnesio, por parte de las bacterias PAO, se produjo posteriormente una mayor captura. Sin embargo, esto provoca que haya variables respuesta en el modelo con pesos cercanos a cero, ocasionando que el modelo tenga una menor capacidad predictiva para dichas variables.

Con el objetivo de conseguir un modelo PLS que con un menor número de parámetros tuviera una alta capacidad predictiva, se eliminaron las trayectorias con menor influencia en el modelo, es decir aquellas con pesos bajos: ΔR_x , O_x y $\Delta Temp$. El nuevo modelo desarrollado, $PLS_{680,54}$, con cuatro componentes explica prácticamente el 90% de la variabilidad tanto de las variables predictoras (X) como de las variables respuesta (Y), y tiene una capacidad predictiva obtenida por validación cruzada, $Q^2(cum)$, superior al 70%. En la Tabla 4.15 se presentan las características de este modelo, que es el que se propone como definitivo.

Componente	$R^2_X(cum)$	Valor propio	$R^2_Y(cum)$	$Q^2(cum)$
1	0.543	8.14	0.597	0.516
2	0.738	2.92	0.786	0.653
3	0.846	1.63	0.851	0.686
4	0.917	1.06	0.886	0.714

Tabla 4.15. Características del modelo $PLS_{680,54}$.

Para visualizar los buenos resultados del modelo desarrollado, se presenta en la Figura 4.44 los valores predichos por el modelo frente a los valores medidos para cada una de las variables respuesta. En esta figura se incluye la recta $y=x$, y también se muestra el valor del coeficiente R^2_Y para cada trayectoria respuesta tanto en los datos empleados en la construcción del modelo como en los datos de validación. Se debe tener presente que estos coeficientes son diferentes a los presentados en la Tabla 4.15, los cuales están calculados para el modelo con todas las variables respuesta, obteniendo además la capacidad predictiva (Q^2) por validación cruzada y no a partir del conjunto de datos de validación.

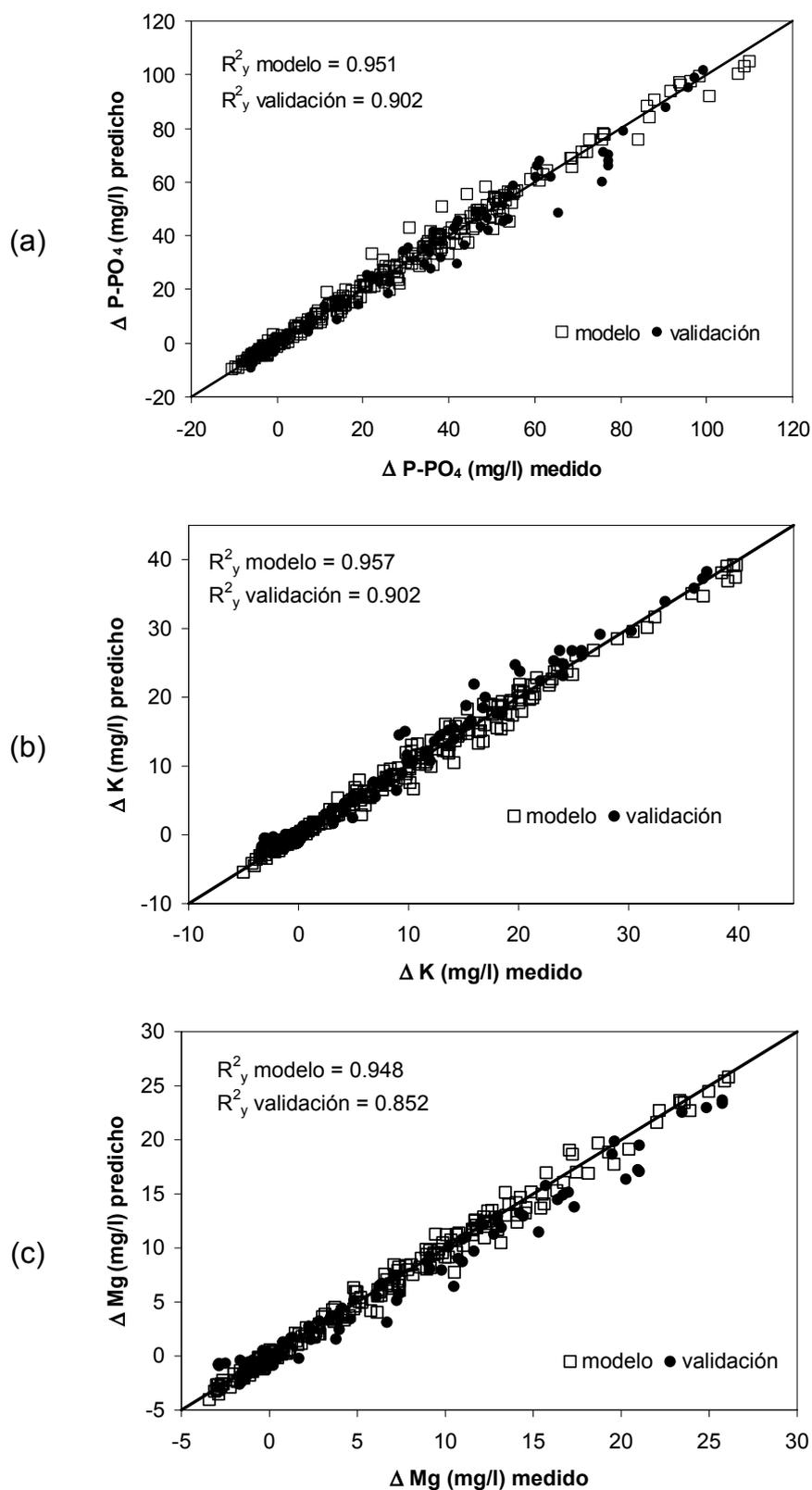


Figura 4.44. Comparación entre los incrementos determinados analíticamente en laboratorio y los predichos por el modelo $\text{PLS}_{680,54}$ para (a) el fósforo (b) el potasio y (c) el magnesio. En cada gráfico se ha representado la recta $y=x$, y se presenta el valor del R^2_y obtenido para cada trayectoria respuesta tanto en los datos empleados en el desarrollo del modelo como en los de validación.

En la aplicación de modelos de proyección sobre estructuras latentes que se ha realizado en este apartado, se pone de manifiesto la capacidad que tiene este tipo de modelos para tolerar la presencia de datos faltantes en las variables tanto en la etapa de construcción del modelo como en la de validación. Además, otra ventaja que tiene este tipo de modelos es el poder comprobar si las nuevas observaciones que se van a emplear para realizar una nueva predicción respetan la estructura de correlaciones del modelo. Es decir, mediante la monitorización del gráfico de la distancia al modelo (Dmod) se puede conocer si el modelo sigue siendo válido para hacer la predicción. La monitorización de este estadístico también puede aportar información acerca de si el comportamiento anómalo se debe a que el proceso está evolucionando hacia otro estado operacional y, por tanto, el modelo necesita ser actualizado, o por el contrario se debe a algún otro tipo de perturbación, como puede ser un fallo en un sensor. Por otra parte, mediante el gráfico de la T^2 es posible detectar aquellas situaciones en las que la estructura de correlaciones de las variables se mantiene pero las variables adoptan valores anormalmente elevados (en valor absoluto) debido a otro tipo de anomalías.

En la Figura 4.45 se muestra el gráfico de la distancia al modelo y de la T^2 para los 20 lotes empleados en la construcción y validación del modelo PLS_{680,54}. Se puede observar que todos los lotes están por debajo de los límites de control.

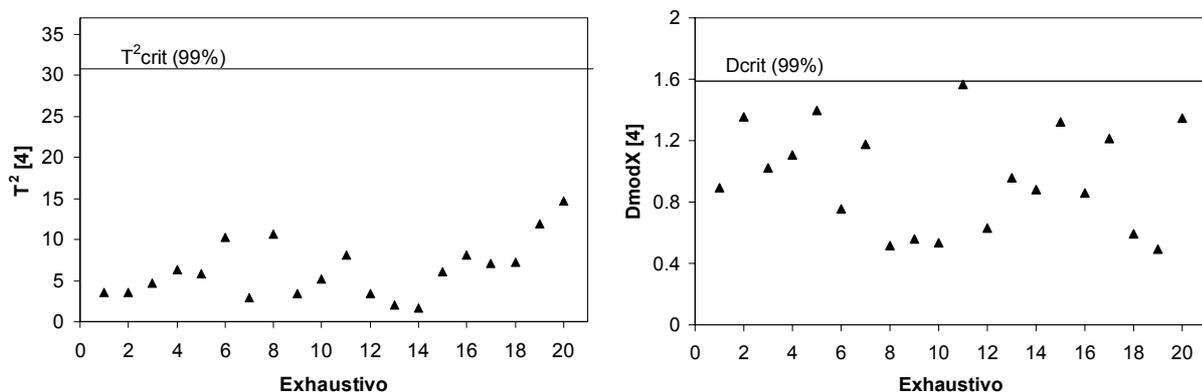


Figura 4.45. Gráficos de control de la T^2 de Hotelling y distancia al modelo (DmodX) del modelo PLS_{680,54} para los 20 experimentos realizados.

4.4.2 Modelos para la predicción de los incrementos de fósforo

En este apartado se pretende analizar la viabilidad de distintos tipos de modelos empíricos para actuar como modelos predictivos (*soft sensors*) en el sistema estudiado y realizar una comparación objetiva entre todos los modelos desarrollados. Se decidió trabajar con la trayectoria de una sola variable respuesta, la creada a partir de las determinaciones analíticas de fósforo (ΔP), por ser ésta la única disponible en los 20 experimentos realizados. De esta manera, se dispone de un total de 19 experimentos en los que no aparecen datos faltantes, de los cuales 15 se destinan a la construcción de los modelos y los 4 restantes a su validación, tal y como queda recogido en la Tabla 4.13. El exhaustivo que no dispone de registro de conductividad se utiliza para analizar el comportamiento de los modelos frente a la presencia de datos faltantes.

Los *soft sensors* que se desarrollan y comparan en los siguientes subapartados se pueden agrupar en dos categorías, los basados en modelos de proyección sobre estructuras latentes y los basados en redes neuronales artificiales. Debido a que la presencia de correlaciones cruzadas entre las variables, así como de autocorrelaciones genera problemas importantes en la estimación de los modelos de regresión mínimo cuadráticos ordinarios, este tipo de modelos no ha sido empleado. Dentro de la primera categoría se analizan diversos modelos PLS y PCR, y dentro de la segunda categoría se estudian las redes neuronales multicapa alimentadas hacia delante. Para facilitar la comparación de los trece modelos desarrollados, los resultados obtenidos se recogen y presentan en el apartado 4.4.2.3.

4.4.2.1 Modelos de proyección sobre estructuras latentes

En este apartado se describe la manera en que se han organizado los datos para la construcción de los modelos de proyección sobre estructuras latentes. En todos los modelos que se han desarrollado, se realiza en primer lugar un desdoblamiento de la matriz tridireccional de los datos. Posteriormente, los datos son preprocesados y sobre éstos se construye el modelo predictivo. El objetivo de los modelos es estimar

la trayectoria creada a partir de las concentraciones de fósforo determinadas analíticamente (trayectoria ΔP).

Salvo en aquellos casos en los que explícitamente se indique el tipo de preprocesamiento realizado, en todos los modelos que se presentan las variables han sido centradas y escaladas a varianza unitaria. Como ya se comentó en el apartado 1.8.6.3.3, es importante tener presente que el efecto que se consigue con el centrado es distinto según el tipo desdoblamiento de la matriz tridireccional de datos que se realice. Cuando en el desdoblamiento se conserva la dirección de los lotes, al centrar se está eliminando la no linealidad más importante del proceso. Sin embargo, cuando se conserva la dirección de las variables, al centrar no se elimina la principal no linealidad del proceso puesto que lo que se está haciendo es simplemente restar un valor constante a toda la trayectoria. Por otra parte, con el escalado de las variables a varianza unitaria se consigue que ninguna variable tenga *a priori* más influencia en el modelo.

El primer grupo de modelos analizado utilizan el desdoblamiento tipo E (Tabla 1.3), en el que se mantiene la dirección de los lotes. Por tanto, las variables explicativas de estos modelos están constituidas por las trayectorias completas de ΔpH , ΔRx , $\Delta Cond$, Ox y $\Delta Temp$, y los modelos son del tipo *Unfold* PLS. En primer lugar, se construyó un modelo PLS con 15 lotes, 1700 variables explicativas (5 trayectorias x 340 instantes de tiempo), y 18 variables respuesta (1 trayectoria x 18 instantes de tiempo). Este primer modelo (M1) fue denominado como PLS_{1700,18}.

En el apartado 4.4.1 se comprobó que las trayectorias creadas a partir del pH (ΔpH) y de la conductividad ($\Delta Cond$) son las que mayor importancia tenían para la predicción de las variables respuesta. Por este motivo y con el objetivo de conseguir un modelo con menor número de parámetros, se construyó un modelo PLS con 15 lotes, 680 variables explicativas (sólo dos trayectorias: ΔpH y $\Delta Cond$) y 18 variables respuesta. Este segundo modelo (M2) fue denominado PLS_{680,18}.

El resto de modelos predictivos que se desarrollaron utilizan el desdoblamiento tipo A (Tabla 1.3 y Figura 1.18), en el cual se conserva la dirección de las variables. En

este caso sólo hay una variable respuesta, el valor en cada instante de tiempo de ΔP , y la matriz de variables explicativas (\mathbf{X}) está compuesta por los valores en cada instante de tiempo de: ΔpH , ΔRx , $\Delta Cond$, Ox y $\Delta Temp$.

Empleando 5 variables explicativas y 1 variable respuesta se desarrollaron dos modelos: $PLS_{5,1}$ (M3) y $PCR_{5,1}$ (M4). En estos modelos se estima el valor de la variable respuesta en un instante de tiempo dado a partir de los valores de las variables explicativas en ese mismo instante de tiempo, es decir, estos modelos aprovechan la correlación instantánea que hay entre las variables.

Sin embargo, el análisis de la función de autocorrelación de la variable respuesta a lo largo de un lote cualquiera pone de manifiesto que el valor de dicha variable en un instante de tiempo dado está altamente correlacionado con el valor en el instante de tiempo anterior. En la Figura 4.46 se muestra la función de autocorrelación simple correspondiente al lote número 2 de los empleados en la construcción de los modelos.

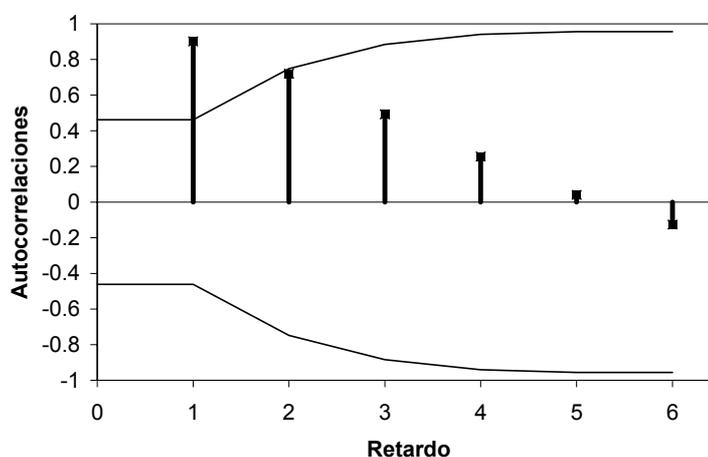


Figura 4.46. Función de autocorrelación simple de la variable respuesta para el lote n° 2 de los utilizados para desarrollar los modelos. Cada retardo es el tiempo transcurrido entre cada 2 determinaciones analíticas (15 minutos). La variable respuesta es el valor de la concentración de fósforo en cada instante de tiempo menos su valor al inicio del lote (ΔP).

En la figura anterior se debe tener en cuenta que cada retardo equivale a 15 minutos, puesto que éste es el intervalo de tiempo entre cada dos determinaciones

analíticas. La forma de la función pone de manifiesto que hay una correlación estadísticamente significativa entre el valor de la variable respuesta en un instante de tiempo dado (k) y su valor en el instante anterior ($k-15$), para un riesgo de primera especie del 5% ($\alpha=0.05$)

En consecuencia, es de esperar que si la dinámica de la variable respuesta se incorpora dentro de los modelos desarrollados, las predicciones de éstos estarán más cercanas a los valores obtenidos experimentalmente. Sin embargo, la utilidad de estos modelos se reduce puesto que al incluir como variable explicativa el valor de la variable respuesta en el instante de tiempo anterior es necesario tener que realizar determinaciones analíticas para conocer su valor. Por tanto, este tipo de modelos sólo tendrían interés durante la propia realización de un experimento exhaustivo, puesto que en esta situación la determinación analítica en laboratorio no permite obtener el valor de la variable respuesta de forma instantánea sino después de un cierto tiempo. Una situación similar se tendría en aquellos casos en los que el sensor que se utiliza para medir la variable respuesta introduce un retardo, es decir, transcurre un cierto tiempo entre el instante en que la muestra es tomada y el instante en que obtiene el resultado.

Considerando la dinámica de la variable respuesta se crearon los modelos $PLS_{6,1}$ (M5) y $PCR_{6,1}$ (M6). La única diferencia de estos modelos con los modelos M3 y M4, es que incluyen como variable explicativa adicional el valor de la variable respuesta en el instante anterior (ΔP_{k-15}). Por tanto, estos modelos tienen 6 variables explicativas y 1 variable respuesta. En el modelo M6, la nueva variable explicativa se ha incorporado directamente en la etapa de regresión. Es decir, no se ha realizado un PCA sobre las 6 variables explicativas y posteriormente una regresión múltiple entre las variables latentes obtenidas y la variable respuesta. De esta manera, se pretende que la variable explicativa ΔP_{k-15} tenga una mayor influencia en el modelo.

Por otro lado, debido a la distinta frecuencia de muestreo de las variables, se dispone de un valor de la variable respuesta cada 15 minutos mientras que de las sondas instaladas en el reactor se dispone de un valor aproximadamente cada minuto. Para utilizar esta mayor cantidad de datos procedentes de las sondas, se desarrollaron

los modelos $PLS_{80,1}$ (M7) y $PCR_{80,1}$ (M8). Estos modelos tienen 75 variables explicativas decaladas (5 variables x 15 instantes de tiempo) más las 5 variables explicativas del instante en el que se realiza la predicción. Es decir, para predecir la variable respuesta en un instante dado se utilizan los valores de las variables explicativas en dicho instante y en los 15 instantes de tiempo anteriores.

Finalmente, para tener en cuenta tanto la dinámica de la variable respuesta como del resto de variables se desarrollaron los modelos $PLS_{81,1}$ (M9), $PLS_{81-b,1}$ (M10) y $PCR_{81,1}$ (M11). Estos modelos utilizan para predecir el valor de la variable respuesta en el instante actual el valor de la variable respuesta en el instante anterior (ΔP_{k-15}), y toda la información de las variables ΔpH , ΔRx , $\Delta Cond$, $\Delta Temp$ y Ox disponible de los 15 minutos anteriores junto con sus valores en el instante actual.

La diferencia entre el modelo $PLS_{81,1}$ y $PLS_{81-b,1}$ está en la forma de realizar el preprocesamiento de las variables. Mientras que en el $PLS_{81,1}$ todas las variables son escaladas a varianza unitaria, en el $PLS_{81-b,1}$ se ha realizado un escalado por bloques para las variables decaladas. De esta forma, todas las variables que proceden de decalar la conductividad constituyen un bloque, y se realiza el escalado de forma que la varianza de todo el bloque sea igual a 1. En estos dos modelos, de las 81 variables explicativas, 75 proceden de decalar las variables registradas por los sensores instalados en el reactor (5 variables x 15 instantes de tiempo) y tan sólo una procede de decalar la variable respuesta. Por tanto, este escalado por bloques trata de evitar que una variable pueda ejercer más influencia en el modelo por el hecho de tener mayor frecuencia de muestreo.

A modo de resumen se presentan en la Tabla 4.16 los once modelos creados, indicando las variables explicativas y respuesta en cada uno de ellos. En dicha tabla \mathbf{x}_k es el vector de las variables explicativas (ΔpH , $\Delta Cond$, ΔRx , Ox , $\Delta Temp$) en el instante de tiempo k , K es el último instante de tiempo del lote, \mathbf{x}_{rk} es el vector de las variables explicativas ΔpH y $\Delta Cond$ en el instante de tiempo k , y_k es la variable respuesta en el instante de tiempo k (ΔP_k), e \mathbf{y} es el vector compuesto por el valor de la variable respuesta (ΔP) en los 18 instantes de tiempo en que fue determinada en cada experimento.

Var. explicativas	Var. respuesta	Modelos
$[\mathbf{x}_1^T, \dots, \mathbf{x}_K^T]$	\mathbf{y}_k^T	M1 (PLS _{1700,18})
$[\mathbf{x}_{r1}^T, \dots, \mathbf{x}_{rK}^T]$	\mathbf{y}_k^T	M2 (PLS _{680,18})
$[\mathbf{x}_k^T]$	y_k	M3 (PLS _{5,1}) y M4 (PCR _{5,1})
$[\mathbf{x}_k^T, y_{k-15}]$	y_k	M5 (PLS _{6,1}) y M6 (PCR _{6,1})
$[\mathbf{x}_k^T, \mathbf{x}_{k-1}^T, \mathbf{x}_{k-2}^T, \dots, \mathbf{x}_{k-15}^T]$	y_k	M7 (PLS _{80,1}) y M8 (PCR _{80,1})
$[\mathbf{x}_k^T, \mathbf{x}_{k-1}^T, \mathbf{x}_{k-2}^T, \dots, \mathbf{x}_{k-15}^T, y_{k-15}]$	y_k	M9 (PLS _{81,1}), M10 (PLS _{81-b,1}) y M11 (PCR _{81,1})

Tabla 4.16. Variables explicativas y respuesta que emplea cada uno de los modelos analizados.

4.4.2.2 Modelos de redes neuronales

En el apartado anterior se han descrito diversas estrategias para el desarrollo de modelos predictivos. Los modelos que se han considerado están basados en el empleo de técnicas de proyección sobre estructuras latentes (PLS y PCR), y han permitido modelar relaciones lineales. Con este tipo de modelos también es posible modelar relaciones no lineales. Esto se podría realizar bien mediante la incorporación de términos no lineales entre las variables explicativas (términos cuadráticos, producto de variables, transformaciones exponenciales,...), o bien transformando el modelo de forma que su relación interna sea no lineal. Sin embargo, en la presente tesis para tratar de explotar las posibles relaciones no lineales que pudieran existir en los datos se ha recurrido al empleo de modelos de redes neuronales artificiales (ANNs).

La mayor parte de aplicaciones de ANNs en el campo de la depuración de las aguas residuales que se pueden encontrar en la bibliografía con objetivo el desarrollar modelos predictivos se basan en el empleo de ANNs alimentadas hacia delante y entrenadas con algoritmos supervisados.

Las ANNs desarrolladas en este apartado son redes multicapa alimentadas hacia delante, y su entrenamiento se ha realizado mediante el popular algoritmo de retropropagación del error. Estas redes también son conocidas como redes perceptrón multicapa (MLP, *Multilayer Perceptron*). En la configuración de los modelos desarrollados sólo se ha incluido una capa oculta (Figura 4.47), y se ha determinado el número óptimo de neuronas en dicha capa. Según Masters (1993), son muy pocas las ocasiones en que resulta necesario utilizar dos capas ocultas, y desde un punto de vista teórico no hay razón para emplear más de dos. Además, tal y como se indicó en el apartado 1.8.8.3, una red MLP con una capa oculta es capaz de aproximar cualquier función continua (Hornik *et al.*, 1989).

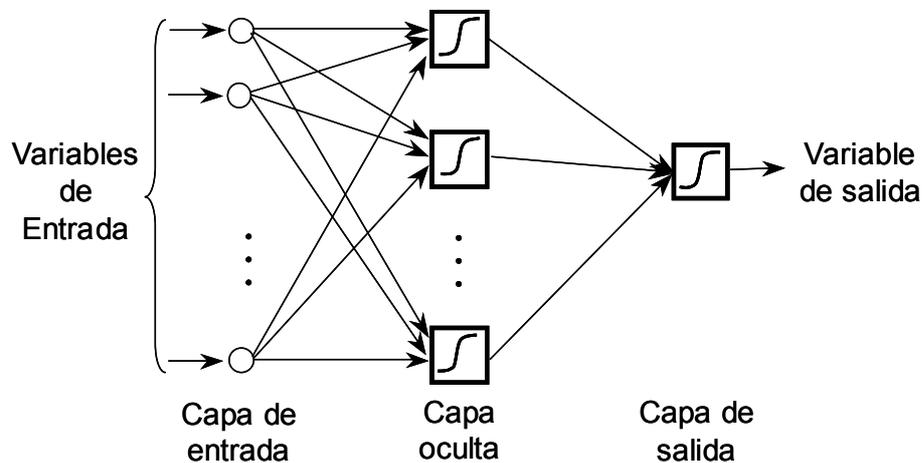


Figura 4.47. Estructura de una red perceptrón multicapa con una capa oculta y una única variable de salida.

Previamente a la construcción de las ANNs se desdobló la matriz tridireccional de datos conservando la dirección de las variables (desdoblamiento tipo A, Tabla 1.3 y Figura 1.18). No se ha utilizado el desdoblamiento conservando la dirección de los lotes puesto que el número de patrones disponible para entrenar las ANNs quedaría drásticamente reducido (15 patrones), a la vez que el número de parámetros de los modelos obtenidos se incrementaría notablemente. Masters (1993) indica que como mínimo, el número de patrones de entrenamiento debe ser el doble que el de pesos de la ANN, y que sería más adecuado duplicar este valor mínimo. Existen otros criterios que limitan el número de pesos de la ANN en función de los patrones empleados en el proceso de entrenamiento, uno de los más conocidos es el de

Weigend *et al.* (1990), según el cual, el número de pesos debe ser menor o igual que la décima parte del número de patrones de entrenamiento.

Tras desdoblar la matriz de datos, el primer modelo MLP desarrollado utiliza como variables de entrada (variables explicativas) los valores en cada instante de tiempo de: ΔpH , ΔRx , ΔCond , Ox y ΔTemp , y como variable de salida (variable respuesta) el valor en cada instante de tiempo de ΔP . Esta red, que se ha denominado $\text{MLP}_{5,1}$, tiene 5 neuronas en la capa de entrada y una en la capa de salida.

El segundo modelo MLP que se desarrolla incorpora en su estructura la dinámica de la variable respuesta y, por tanto, se incluye como variable explicativa en los patrones de entrenamiento el valor de la variable respuesta en el instante anterior. Esta red, que se ha denominado $\text{MLP}_{6,1}$, tiene 6 neuronas en la capa de entrada y una en la capa de salida.

En ambos casos ($\text{MLP}_{5,1}$ y $\text{MLP}_{6,1}$), se ha empleado como función de activación, tanto en las neuronas de la capa oculta como en las de la capa de salida, la función logística:

$$f_j = \frac{1}{1 + \exp^{-v_j}} \quad \text{Ec. 4.5}$$

donde v_j es la entrada neta a la neurona j .

Por tanto, con ambos modelos se está realizando una regresión no lineal múltiple para predecir la variable respuesta. La formulación matemática de los modelos construidos (empleando una notación análoga a la introducida en el apartado 1.8.8.4) se puede expresar de la siguiente manera:

$$y_j^{(1)} = \left(1 + \exp \left[- \sum_{i=0}^I w_{ij}^{(1)} x_i^{(0)} \right] \right)^{-1} \quad \text{Ec. 4.6}$$

$$\hat{y} = \left(1 + \exp \left[- \sum_{j=0}^J w_j^{(2)} y_j^{(1)} \right] \right)^{-1} \quad \text{Ec. 4.7}$$

donde $y_j^{(1)}$ es el valor de la salida de la neurona j de la capa oculta, l es el número de variables de entrada, $w_{ij}^{(1)}$ son los pesos de las conexiones entre la capa de entrada y la capa oculta, $x_i^{(0)}$ es el valor de la variable de entrada i , \hat{y} es la variable respuesta, J es el número de neuronas incluidas en la capa oculta, $w_j^{(2)}$ son los pesos de las conexiones entre la capa oculta y la capa de salida, $x_0^{(0)}$ e $y_0^{(1)}$ tienen un valor unitario para de esta forma introducir el término del umbral de activación (apartado 1.8.8.4).

Como se ha indicado previamente, sólo se ha considerado una capa oculta en ambos modelos (MLP_{5,1} y MLP_{6,1}) y la configuración óptima de la red se determinó variando el número de neuronas en la capa oculta. Para el entrenamiento de las ANNs se han utilizado los 15 exhaustivos indicados en la Tabla 4.13. De manera análoga a como se procedió con los modelos PLS y PCR desarrollados en los apartados anteriores, para la construcción y validación de los modelos de redes neuronales se realizaron las siguientes etapas: preprocesamiento, entrenamiento y validación.

4.4.2.2.1 Preprocesamiento.

Todas las variables, tanto las de entrada como la variable de salida, fueron centradas, tipificadas y escaladas. El centrado hace que la media de cada variable en el conjunto de datos empleados en el entrenamiento sea cero. Al tipificar las variables se consigue que todas tengan la misma varianza y que, por tanto, ninguna variable tenga “a priori” mayor influencia en el modelo. Finalmente, el escalado lo que persigue es que las variables estén dentro del rango de la función de activación. En la presente tesis se ha utilizado la función logística (Ec. 4.5), cuya salida está comprendida en el intervalo $[0,1]$ y, por tanto, los valores de la variable de salida que se presenten a la ANN deberán estar dentro de ese mismo intervalo. Sin embargo, como los valores extremos de la función de activación, en este caso 0 y 1, sólo se

obtienen cuando su entrada (v_j) es $\pm\infty$, se suelen escalar los valores de las variables respuesta para que queden en un rango inferior al de la función de activación. Esto evita que el algoritmo de retropropagación del error modifique los valores de los parámetros de la ANN hacia infinito, provocando lo que se conoce como saturación de las neuronas. Los valores del intervalo de escalado que se han adoptado son 0.1 y 0.9, y la expresión que se ha utilizado es la siguiente:

$$x_{i,t} = \frac{(U_x - L_x)x_i + (M_x L_x - m_x U_x)}{M_x - m_x} \quad \text{Ec. 4.8}$$

donde $x_{i,t}$ es el valor escalado que utilizará la red, x_i es el valor de la variable, M_x y m_x los valores máximo y mínimo de la variable, respectivamente, y U_x y L_x los valores máximo y mínimo de la variable escalada, respectivamente.

4.4.2.2.2 Entrenamiento.

El entrenamiento de una ANN es en la terminología estadística el proceso de estimación de parámetros del modelo (Tabla 1.4). En el ámbito de las ANNs se suele definir como el procedimiento por el cual la ANN extrae el conocimiento de la información disponible mediante la modificación de los pesos de sus conexiones. Se trata de un problema de optimización no lineal, en el que la función objetivo a minimizar que se suele emplear es el error cuadrático medio (MSE, *Mean Squared Error*):

$$\text{MSE} = \frac{1}{N_p} \sum_{n=1}^{N_p} \sum_{j=1}^{N_j} (t_{pj} - o_{pj})^2 \quad \text{Ec. 4.9}$$

donde N_p es el número de patrones de entrenamiento, N_j el número de neuronas en la capa de salida, t_{pj} es el valor de la variable respuesta para el patrón p de la neurona j que está en la capa de salida, y o_{pj} es el valor predicho por la red.

Es decir, durante el entrenamiento se minimiza la suma de cuadrados de las diferencias entre los valores reales y los predichos por la red para todo el conjunto

de datos de entrenamiento, dividida por el número de patrones. Es habitual emplear como función objetivo el MSE porque presenta una serie de ventajas (Maier y Dandy, 2000): se calcula de manera sencilla, penaliza los errores de mayor magnitud, y es fácil la obtención de su derivada parcial con respecto a los pesos.

El entrenamiento es un proceso iterativo en el que continuamente se le presentan los patrones a la red hasta que se cumple el criterio de parada. Éste puede ser que el MSE sea inferior a un valor prefijado de antemano o cuando se alcance el número máximo de iteraciones especificado.

El algoritmo que se ha empleado en la presente tesis para el entrenamiento de las ANNs es el de retropropagación del error con momento:

$$\Delta w_{ij}^{(c)}(k) = -\eta \frac{\partial \text{MSE}_p}{\partial w_{ij}^{(c)}} + \mu \Delta w_{ij}^{(c)}(k-1) \quad \text{Ec. 4.10}$$

donde $\Delta w_{ij}^{(c)}(k)$ es la modificación en el valor de la conexión $w_{ij}^{(c)}$ entre las neuronas i y j de la capa c en el instante k , η es la velocidad de aprendizaje, MSE_p es el valor de la función de error para el patrón p , μ es el término del momento y $\Delta w_{ij}^{(c)}(k-1)$ es la modificación en el valor de la conexión $w_{ij}^{(c)}$ entre las neuronas i y j de la capa c en el instante anterior ($k-1$).

El desarrollo matemático del algoritmo de retropropagación del error se puede encontrar en Haykin (1999). En este algoritmo, cuanto más reducida es la velocidad de aprendizaje (η), más pequeñas son las modificaciones en los valores de los pesos de la ANN entre cada dos iteraciones. Los valores típicos para este parámetro están comprendidos en el intervalo [0.1, 1] (Zell *et al.*, 1998), aunque en ocasiones se utilizan valores superiores. También se podría utilizar una velocidad variable de un ciclo a otro, sin embargo, es más habitual utilizar un valor constante y obtener los valores óptimos de los parámetros por prueba y error (Maier y Dandy, 2000). Por otra parte, el término del momento (μ) permite considerar la inercia en la

modificación de los pesos de la ANN, y su utilización presenta una serie de ventajas (Haykin, 1999):

- permite acelerar el proceso de aprendizaje.
- tiene un efecto estabilizador sobre el algoritmo en aquellas direcciones de minimización del error en las que producen cambios de signo en los ajustes de los pesos.
- tiende a evitar que el entrenamiento acabe en un mínimo local “poco profundo” de la superficie del error.

Los valores típicos para este parámetro (μ) están comprendidos en el intervalo [0, 1], y no se deben adoptar valores superiores a la unidad para que el proceso de aprendizaje sea convergente (Haykin, 1999).

Existen otros algoritmos más sofisticados y de convergencia más rápida, sin embargo, el algoritmo de retropropagación es en la práctica de los más utilizados por su sencillez y por los buenos resultados que proporciona.

La actualización de los pesos se puede realizar tras haber presentado a la red todos los patrones de entrenamiento, es decir, cada vez que se completa un ciclo, o bien tras la presentación de cada patrón. En el primer caso se dice que el entrenamiento se realiza en modo *batch* y en el segundo en modo *on-line*. En este último modo, los patrones no se suelen presentar siempre en el mismo orden, sino que de un ciclo a otro el orden de presentación es aleatorio. Esto permite que la búsqueda en el espacio de pesos sea estocástica, reduciendo así la probabilidad de que el algoritmo quede atrapado en un mínimo local (Haykin, 1999). Este autor señala además que cuando hay información redundante en el conjunto de datos de entrenamiento se obtienen mejores resultados con el modo *on-line*.

Para comenzar el entrenamiento es necesario previamente haber inicializado la ANN, es decir, haber establecido unos valores iniciales de los pesos de la red. Puede suceder que durante el entrenamiento la ANN alcance la convergencia en un mínimo local “poco profundo” de la superficie de error, por lo que es conveniente

iniciar el proceso de aprendizaje con distintas combinaciones de los valores iniciales de los pesos.

Se debe tener presente que cuando se pretende desarrollar un modelo predictivo basado en la utilización de redes neuronales, hay que determinar la configuración y los parámetros más adecuados para que la ANN tenga una buena capacidad de generalización (bondad de predicción). Es decir, el objetivo es conseguir que la ANN sea capaz de proporcionar respuestas adecuadas cuando se le presentan datos que no han sido empleados durante su entrenamiento. En este punto se asume que los datos que se utilizan para la validación de la ANN proceden de la misma población que los datos empleados en su entrenamiento.

Para poder comprobar de manera objetiva la capacidad de generalización (bondad de predicción) de una red neuronal es una práctica común dividir el conjunto de datos disponible en dos partes, de forma que una parte de los datos se utiliza para su entrenamiento y el resto para comprobar su respuesta frente a datos nuevos (validación).

En este trabajo, para evitar la utilización de los datos de validación en la determinación de la topología óptima de las ANNs, se ha recurrido a la técnica de validación cruzada. La aplicación de esta técnica consistió en utilizar 14 de los 15 exhaustivos para entrenar la red y el otro para comprobar su comportamiento obteniendo el error de predicción. El proceso de entrenamiento fue repetido 15 veces, dejando fuera cada vez un exhaustivo distinto. El promedio de los 15 exhaustivos que fueron excluidos durante el entrenamiento, se utilizó como estimación de la capacidad de generalización (predicción) de la ANN. De esta forma, tras la determinación de la topología óptima de la ANN, aún se disponía de 4 exhaustivos (Tabla 4.13) para su validación, es decir, para comprobar su comportamiento frente a un conjunto de datos independiente.

De manera resumida se indica a continuación cuál fue el procedimiento seguido para el desarrollo de los modelos de redes neuronales:

- Para cada topología se realizaron 20 inicializaciones distintas. En cada una de ellas los valores de los parámetros de la ANN (pesos y umbrales de activación) eran inicializados aleatoriamente entre 0 y 1.
- El entrenamiento se realizó con el algoritmo de retropropagación del error con momento, actualizando los parámetros de la ANN según el modo *on-line* y presentando los patrones de forma aleatoria en cada ciclo. Se utilizaron distintas combinaciones de los parámetros del algoritmo (η y μ).
- Se aplicó la técnica de validación cruzada para determinar la topología óptima de la red. Al utilizar esta técnica, para cada inicialización y una vez finalizado el proceso de entrenamiento se obtiene el MSE del conjunto de datos de entrenamiento, es decir, la bondad de ajuste del modelo y el MSE correspondiente a la estimación de la capacidad de generalización de la red. Como se realizaron 20 inicializaciones, para cada topología se obtuvieron 20 valores de estas dos medidas del error.
- Para seleccionar la topología óptima, se representó gráficamente la evolución de la mediana de las dos medidas del error en función del número de neuronas en la capa oculta. La topología que con menor número de neuronas presentó la mejor capacidad de generalización estimada fue la que se adoptó como definitiva.

Los resultados obtenidos tras la aplicación de esta metodología se presentan en la Figura 4.48. En ella se muestra la evolución de la mediana del MSE de las 20 inicializaciones realizadas en función del número de neuronas incluidas en la capa oculta para los valores de los parámetros del algoritmo de entrenamiento: $\eta=0.2$ y $\mu=0.2$. Se puede observar cómo conforme se aumenta el número de neuronas en la capa oculta, la ANN es capaz de ajustar mejor los datos (Figura 4.48 a), aunque llega un momento en el error parece estabilizarse en un mínimo. Sin embargo, para la capacidad de generalización obtenida por el procedimiento de validación cruzada aparece un óptimo (Figura 4.48 b). En el caso de la configuración MLP_{5,1} los mejores resultados se obtuvieron con 4 neuronas en la capa oculta, mientras que para la

configuración MLP_{6,1} el óptimo se obtuvo con 1 neurona en la capa oculta. Como era de esperar, los modelos MLP_{6,1} presentaron valores más bajos del MSE para los datos de entrenamiento ya que incorporan la dinámica de la variable respuesta.

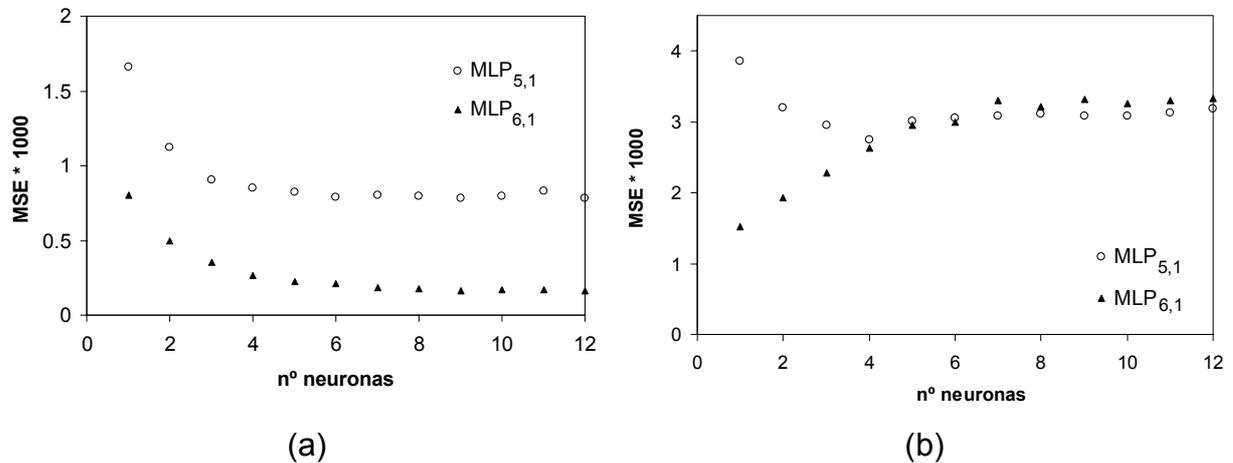


Figura 4.48. Evolución de la mediana del MSE de las 20 inicializaciones en función del número de neuronas incluidas en la capa oculta: (a) capacidad de ajuste de los datos empleados en el entrenamiento (b) capacidad de generalización obtenida por validación cruzada. Las gráficas mostradas han sido obtenidas con $\eta=0.2$ y $\mu=0.2$, y el valor del MSE corresponde a los datos preprocesados y no a las variables originales.

Resultados similares a los mostrados en la Figura 4.48 se encontraron cuando se probaron distintas combinaciones de los parámetros del algoritmo de entrenamiento, por lo que se concluyó que la configuración de las redes MLP_{5,1} y MLP_{6,1} seleccionadas eran las que presentaban un comportamiento más adecuado para el conjunto de datos analizado.

4.4.2.2.3 Validación.

En esta etapa se comprobó el comportamiento de las ANNs que se seleccionaron como óptimas (MLP_{5,1} con cuatro neuronas en la capa oculta y MLP_{6,1} con una neurona en la capa oculta), sobre un conjunto de datos independiente, es decir, que no había sido utilizado en las etapas anteriores. Este conjunto de datos era el constituido por los 4 exhaustivos de validación que se indican en la Tabla 4.13. Los resultados obtenidos se presentan en el apartado 4.4.2.3 para facilitar la comparación de todos los modelos que se desarrollaron.

4.4.2.3 Comparación de los modelos desarrollados

En este apartado se presentan y comparan los resultados de los trece modelos estudiados que han sido descritos en los apartados 4.4.2.1 y 4.4.2.2. La comparación entre los distintos modelos se hace en base al porcentaje de la varianza total de los datos explicada por cada modelo, y también a la suma de cuadrados residual de cada uno de ellos. Estos dos parámetros se han calculado para los 4 experimentos exhaustivos que se utilizan para evaluar la capacidad de predicción del modelo.

En la Tabla 4.17 se presentan de manera resumida los resultados obtenidos para los trece modelos desarrollados. En dicha tabla se indica:

- el tipo de desdoblado realizado de la matriz tridireccional de los datos, es decir, si mantiene la dirección de los lotes o la de las variables.
- el tipo de modelo utilizado. Modelo de regresión en mínimos cuadrados parciales, PLS_{xy} , modelo de regresión en componentes principales, PCR_{xy} , o modelo de redes neuronales MLP_{xy} . El subíndice x hace referencia al número de variables explicativas que emplea el modelo, y el subíndice y al número de variables respuesta.
- el número de componentes o neuronas en la capa oculta del modelo (N° comp). En los modelos PLS y PCR , es el número de componentes del modelo. En el caso de los modelos PCR en dicha columna aparecen dos valores, el primero corresponde al número de componentes significativas obtenidas por validación cruzada en la etapa de realización del PCA , y entre paréntesis se indica el número de componentes que resultan estadísticamente significativas en la etapa de regresión. Como ya se ha comentado en el apartado 4.4.2.1, en los modelos PCR que incorporan la dinámica de la variable respuesta, ésta se incorpora directamente en la etapa de regresión, por lo que el número de variables significativas en la etapa de regresión puede ser mayor que el número de componentes obtenidas al

realizar el PCA. En los modelos MLP se indica el número de neuronas en la capa oculta.

- dos parámetros que reflejan la bondad de predicción de cada uno de los modelos (R^2_y y SCR). El primer parámetro (R^2_y) es el cociente entre la variabilidad explicada por el modelo y la variabilidad total, y el segundo parámetro (SCR) es la variabilidad residual o suma de cuadrados residual. Ambos parámetros representan la bondad de predicción porque se han calculado con los datos de los 4 experimentos que fueron reservados para la validación de los modelos. Como se indicó en el apartado 1.8.4.4, para el cálculo de la bondad de ajuste de los modelos, la variabilidad total, en este caso de la variable respuesta, se descompone en dos términos: la variabilidad explicada por el modelo y la variabilidad residual. La variabilidad total o suma de cuadrados total (SCT) en un modelo de regresión con una variable respuesta se define como el sumatorio del cuadrado de las desviaciones de cada valor observado respecto de su media. La SCT y la SCR para el caso de tener J variables respuesta, así como el parámetro R^2_y , se pueden obtener a partir de las siguientes expresiones:

$$SCT = \sum_{j=1}^J \sum_{i=1}^n (y_{ji} - \bar{y}_j)^2 \quad \text{Ec. 4.11}$$

$$SCR = \sum_{j=1}^J \sum_{i=1}^n (y_{ji} - \hat{y}_{ji})^2 \quad \text{Ec. 4.12}$$

$$R^2_y = 1 - \frac{SCR}{SCT} \quad \text{Ec. 4.13}$$

donde y_{ji} es el valor de la variable respuesta j en la observación i , \bar{y}_j es el valor medio de la variable respuesta j , \hat{y}_{ji} es el valor estimado por el modelo para la variable respuesta j en la observación i , J es el número de variables respuesta y n es el número de observaciones.

Como se puede deducir de las expresiones anteriores, el parámetro R^2_y no es directamente comparable entre modelos que tienen distinto número de

variables respuesta. Es decir, en base al R^2_y no se pueden comparar los modelos que desdoblan manteniendo la dirección de los lotes, que tienen 18 variables respuesta, con los que mantienen la de las variables, que tienen 1 sola variable respuesta. Por este motivo, la comparación entre todos los modelos se hará en base al valor de la SCR.

Desdoblado	Modelo	Nº comp	R^2_y	SCR	
Lotes	M1	PLS _{1700,18}	5	0.887	1092.805
Lotes	M2	PLS _{680,18}	4	0.947	885.240
Variables	M3	PLS _{5,1}	3	0.921	6651.277
Variables	M4	PCR _{5,1}	1; (1)	0.609	32883.320
Variables	M5	PLS _{6,1}	2	0.974	2147.108
Variables	M6	PCR _{6,1}	1; (2)	0.937	5271.002
Variables	M7	PLS _{80,1}	3	0.876	10440.620
Variables	M8	PCR _{80,1}	8; (6)	0.901	8281.894
Variables	M9	PLS _{81,1}	3	0.891	9120.564
Variables	M10	PLS _{81-b,1}	3	0.975	2083.466
Variables	M11	PCR _{81,1}	8; (7)	0.984	1313.392
Variables	M12	MLP _{5,1}	4	0.931	5768.910
Variables	M13	MLP _{6,1}	1	0.978	1807.204

Tabla 4.17. Principales características y resultados de validación de todos los modelos desarrollados para la estimación de la variable respuesta: incremento de la concentración de fósforo respecto de su valor al comienzo del lote (ΔP).

Con el objetivo de comparar la capacidad de predicción de los trece modelos desarrollados, se ha llevado a cabo un análisis de la varianza (ANOVA) sobre los errores cuadráticos de las observaciones de los 4 experimentos de validación. El error de una observación se calcula como la diferencia entre el valor medido experimentalmente y el estimado por el modelo. Los errores se elevaron al cuadrado y al estudiar su distribución se encontró una asimetría positiva, motivo por el cual se aplicó una transformación logarítmica. La tabla del ANOVA, obtenida con la aplicación "StatGraphics", se muestra en la Tabla 4.18.

Analysis of Variance					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Between groups	1288.39	12	107.366	26.54	0.0000
Within groups	3733.57	923	4.045		
Total (Corr.)	5021.96	935			

Tabla 4.18. Resultados del ANOVA sobre el logaritmo neperiano de los errores cuadráticos de las observaciones de los 4 experimentos de validación. En total se dispone de 936 observaciones (4 experimentos x 18 muestras x 13 modelos).

En la Figura 4.49 se muestran los intervalos LSD (*Least Significance Difference*) para el valor medio del logaritmo neperiano de los errores cuadráticos de las observaciones de validación para cada uno de los trece modelos desarrollados.

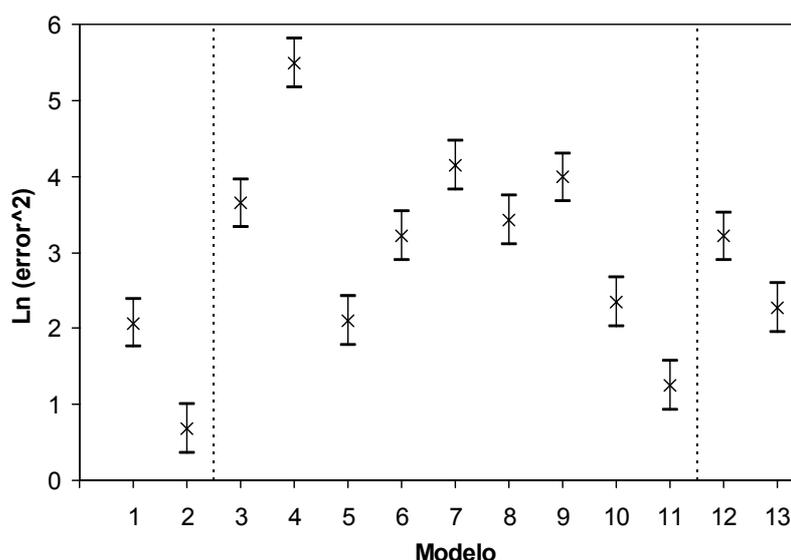


Figura 4.49. Intervalos LSD para la media del logaritmo neperiano de los errores cuadráticos de las observaciones de validación para los trece modelos desarrollados.

Prácticamente todos los modelos presentan una elevada capacidad de predicción de los datos utilizados para su validación, con valores de R^2_y superiores al 85%. La única excepción es el modelo PCR_{5,1} (M4), en el que los inferiores resultados que muestra pueden ser debidos a que el análisis PCA realizado sobre las variables explicativas dio lugar a una única componente significativa. Los modelos PLS_{5,1} (M3) y MLP_{5,1} (M12) utilizando las mismas variables explicativas que el modelo M4, se muestran muy superiores. Aunque el modelo de redes neuronales (M12) presenta un

valor de la SCR inferior al modelo de regresión en mínimos cuadrados parciales (M3), las diferencias entre ambos no han resultado ser estadísticamente significativas puesto que sus intervalos LSD se solapan.

Los modelos que mejor comportamiento han mostrado han sido los desarrollados manteniendo la dirección de los lotes en el desdoblamiento, y de los modelos que conservan la dirección de las variables, tan sólo presentan resultados parecidos aquellos que incorporan en su formulación la dinámica de la variable respuesta. Como ya se ha indicado previamente (apartado 1.8.6.3.3), se debe tener presente que en los modelos que desdoblan según la dirección de los lotes, el preprocesamiento elimina la principal componente no lineal de los datos, pero además, estos modelos están aprovechando toda la estructura de correlaciones cruzadas entre las variables y de autocorrelaciones a lo largo de todo el lote. En el caso de utilizar el desdoblamiento según la dirección de las variables, no se elimina la no linealidad de los datos y, además, estos modelos sólo capturan las correlaciones cruzadas instantáneas entre las variables salvo que explícitamente se introduzcan decalajes en las variables explicativas.

Por otra parte, al comparar los modelos M7 (PLS_{80,1}), M9 (PLS_{81,1}) y M10 (PLS_{81-b,1}) se puede observar la importancia que puede llegar a tener el tipo de escalado en los resultados del modelo desarrollado. Aunque los modelos M9 y M10 utilizan las mismas variables explicativas, el modelo M10 presenta resultados significativamente mejores. La única diferencia entre ambos modelos es el tipo de escalado, puesto que mientras en el M9 las variables fueron escaladas a varianza unitaria, en el M10 se realizó un escalado por bloques. Este último escalado evitó que las variables con mayor frecuencia de muestreo tuvieran mayor influencia en el modelo resultante. Esto explica por qué el efecto de la variable decalada queda anulado por las 80 variables de proceso, por lo que los modelos M7 y M9 no se muestran estadísticamente diferentes.

En la Figura 4.49 queda claramente reflejada la existencia de diferencias estadísticamente significativas ($\alpha=0.05$) desde el punto de vista del error cuadrático medio entre los modelos desarrollados, y en dicha figura se puede apreciar que los

dos mejores modelos que se han obtenido son el M2 (PLS_{680,18}) y el M11 (PCR_{81,1}). Entre estos dos modelos no hay diferencias estadísticamente significativas puesto que sus intervalos LSD se solapan. El siguiente grupo de modelos con los que se obtuvieron mejores resultados fueron M1, M5, M10 y M13, entre los cuales tampoco hay diferencias estadísticamente significativas.

En la Figura 4.50 se presentan las predicciones de la variable respuesta en los 4 exhaustivos de validación de los dos mejores modelos desarrollados (M2 y M11), junto con los valores determinados analíticamente.

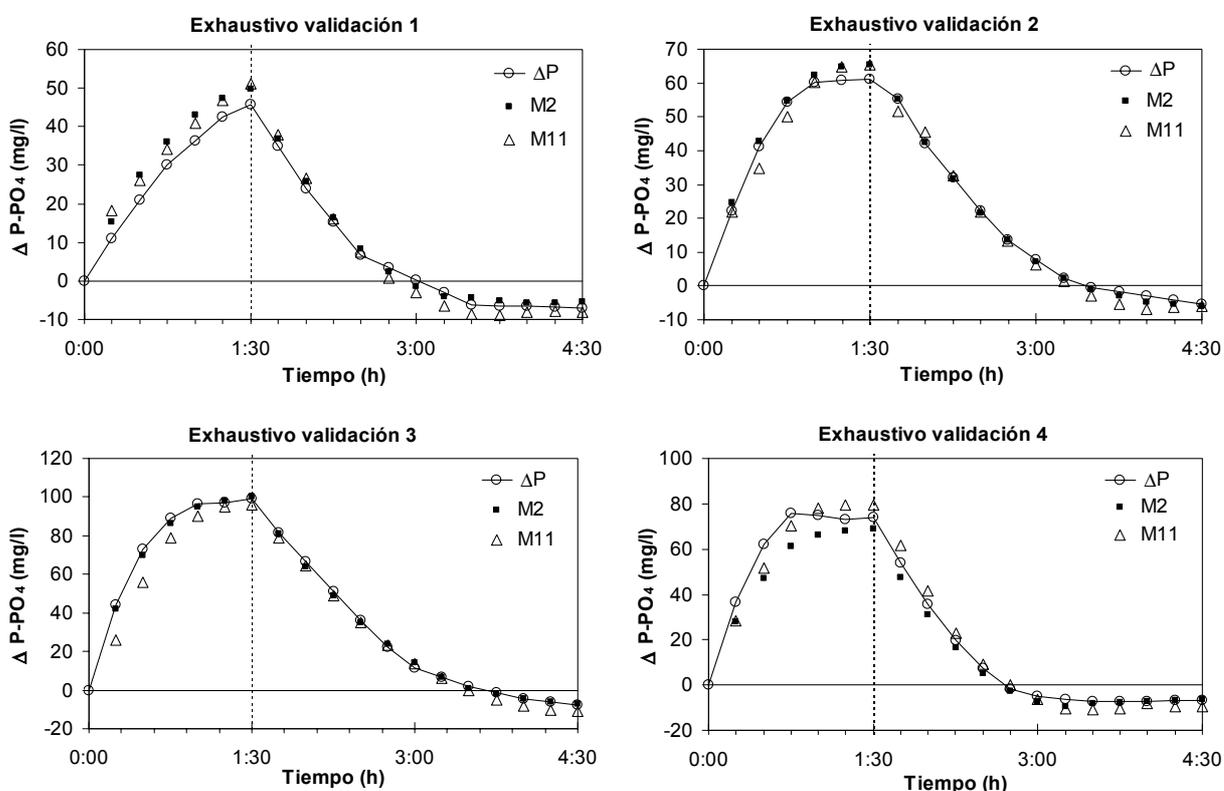


Figura 4.50. Evolución de la variable respuesta (ΔP) en los 4 exhaustivos de validación junto con las predicciones obtenidas con los 2 mejores modelos desarrollados: M2 (PLS_{680,18}) y M11 (PCR_{81,1}).

Es importante destacar que el modelo M2, además de haber mostrado una buena capacidad predictiva, tiene la ventaja de que no es necesario medir la variable respuesta, es decir, para realizar las predicciones sólo utiliza los datos procedentes de los sensores instalados en el reactor. En aquellos casos en que se determine la variable respuesta, se puede utilizar dicha información para verificar si el modelo

sigue siendo válido, o si por el contrario es necesario volver a estimar los parámetros del modelo. Por otra parte, el inconveniente que presenta el modelo es que para hacer las predicciones el lote tiene que haber finalizado, puesto que necesita conocer cual ha sido la evolución de las trayectorias de las variables explicativas a lo largo del lote. Por tanto, para poder explotar el modelo en tiempo real sería necesario imputar los valores correspondientes a la evolución futura del lote, tal y como se expuso en el apartado 1.8.6.3.1. A modo de ejemplo, se presenta en la Figura 4.51 la evolución de las predicciones del modelo M2 en tiempo real para el tercer exhaustivo de validación cuando se trata la trayectoria futura del lote como datos faltantes. En esta figura se puede observar que a partir del instante de tiempo $k=15$, se empiezan a tener predicciones razonables, y conforme se dispone de más datos de las trayectorias de las variables explicativas, las predicciones se aproximan cada vez más a las definitivas. Por tanto, el modelo M2 se podría integrar dentro del esquema de monitorización propuesto en el apartado 4.3.

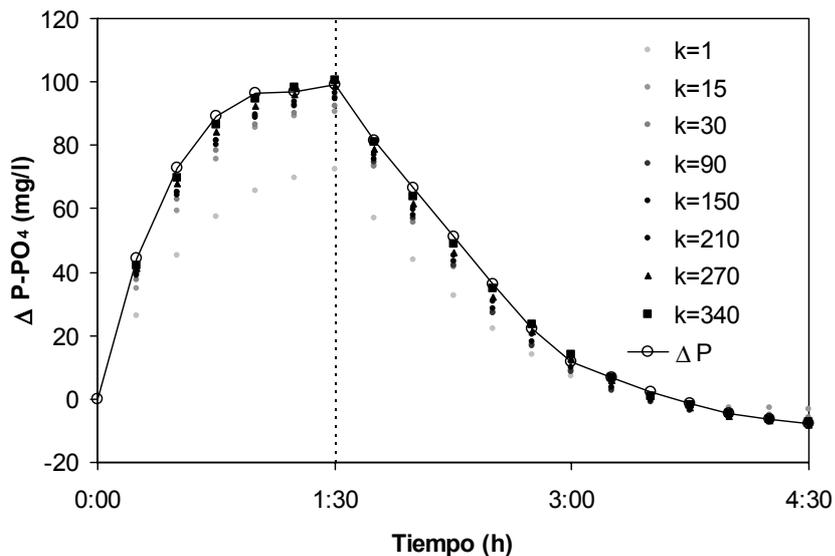


Figura 4.51. Evolución de las predicciones de la variable respuesta (ΔP) para el tercer exhaustivo de validación, cuando se utiliza el modelo M2 en tiempo real y tratando como datos faltantes la trayectoria futura del lote.

Por otra parte, como se indicó previamente, entre los lotes que no fueron empleados para la construcción de los modelos hay uno que presentaba el inconveniente de no tener registro de conductividad. Esto permitió ver cómo se han comportado los distintos modelos frente a la falta de una variable explicativa importante.

La construcción, validación y explotación de los modelos de proyección sobre estructuras latentes no presenta ninguna dificultad porque el algoritmo NIPALS tolera la presencia de datos faltantes. Sin embargo, en los modelos de redes neuronales no es posible realizar la etapa de entrenamiento (estimación de los parámetros del modelo) cuando hay datos faltantes en las observaciones, aunque para su validación y explotación se podrían adoptar dos estrategias para tratar de abordar esta situación de datos faltantes:

- No utilizar o eliminar aquellas observaciones que tengan datos faltantes. Esta forma de proceder tiene la clara desventaja de que se pierde información y, por tanto, no es recomendable.
- Utilizar un algoritmo de imputación de datos faltantes. Dentro de esta estrategia se puede optar por:
 - o Imputar a los datos faltantes el valor medio calculado con los datos de entrenamiento. En el caso de un proceso por lotes surgen dos opciones, por un lado se podría utilizar la media correspondiente al desdoblado manteniendo la dirección de las variables y, por otro, la media correspondiente al desdoblado que conserva la dirección de los lotes. Esta última opción es la que en principio resulta más lógica, ya que en el caso de que faltara la trayectoria completa de toda una variable, lo que se asigna en su lugar es la trayectoria media de esa variable obtenida a partir de las trayectorias de dicha variable en todos los lotes utilizados para desarrollar el modelo, en lugar de asignar un valor constante a toda la trayectoria, que es lo que se haría en el desdoblado manteniendo la dirección de las variables.
 - o Estimar los datos faltantes a partir de la información contenida en el conjunto de datos de entrenamiento. Es decir, se trataría de desarrollar una nueva red neuronal con el objetivo de predecir los datos faltantes.

A partir de la discusión anterior se puede concluir que las dos estrategias más razonables en el caso particular en que nos encontramos son imputar a la variable que no fue registrada el valor medio obtenido realizando el desdoblamiento según la dirección de los lotes, o bien construir otra red neuronal para estimar el valor de la variable no disponible (ΔCond) a partir del resto de variables registradas. En este último caso, para el modelo $\text{MLP}_{5,1}$ se construyó una nueva red neuronal $\text{MLP}_{4-c,1}$ que utilizaba como variables explicativas los valores en cada instante de tiempo de ΔpH , ΔRx , Ox y ΔTemp . Por otra parte, para el modelo $\text{MLP}_{6,1}$ se desarrolló una red neuronal $\text{MLP}_{5-c,1}$ en la que además de las variables utilizadas en el modelo $\text{MLP}_{4-c,1}$ se incluyó como variable explicativa el valor decalado de la variable ΔP . En ambas redes neuronales, la variable respuesta fue el valor en cada instante de tiempo de la trayectoria ΔCond .

En la determinación de la configuración óptima de estas dos redes neuronales, $\text{MLP}_{4-c,1}$ y $\text{MLP}_{5-c,1}$, se siguió el procedimiento descrito en el apartado 4.4.2.2.2, utilizando para el entrenamiento los 15 exhaustivos definidos en la Tabla 4.13. El número de neuronas en la capa oculta que proporcionó la mejor capacidad de generalización obtenida por validación cruzada fue de 4 para el modelo $\text{MLP}_{4-c,1}$ y 3 para el modelo $\text{MLP}_{5-c,1}$.

Para obtener la variable de salida ΔP , es necesario utilizar estas nuevas redes neuronales desarrolladas, cuya salida es la variable ΔCond , junto con los modelos $\text{MLP}_{5,1}$ (M12) y $\text{MLP}_{6,1}$ (M13), tal y como se muestra en la Figura 4.52.

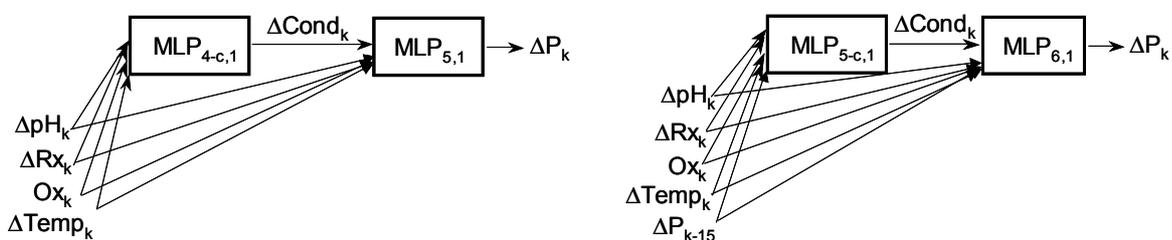


Figura 4.52. Utilización de las redes neuronales para predecir el valor de la variable respuesta ΔP en el caso de no disponer de registro de conductividad.

Para comprobar si existían diferencias significativas en la capacidad de predicción de la variable respuesta ΔP en ambos modelos, entre imputar la trayectoria media de $\Delta Cond$ y estimarla mediante una red neuronal, se realizó un ANOVA sobre los errores cuadráticos correspondientes a las observaciones de dicho exhaustivo. En este análisis se estudió el efecto del modelo utilizado (M12 o M13), de la alternativa empleada para tratar la trayectoria faltante (imputación o estimación), así como de la interacción entre ambos factores. Debido a la asimetría positiva que presentaba la distribución de los errores cuadráticos, se aplicó una transformación logarítmica. La tabla del ANOVA, obtenida con la aplicación “StatGraphics”, se muestra en la Tabla 4.19.

Analysis of Variance					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:Modelo	64.5236	1	64.5236	16.86	0.0001
B:Alternativa	7.9736	1	7.9736	2.08	0.1535
INTERACTIONS					
AB	24.6527	1	24.6527	6.44	0.0135
RESIDUAL	260.301	68	3.8279		
TOTAL (CORR.)	357.451	71			

Tabla 4.19. Resultados del ANOVA sobre el logaritmo neperiano de los errores cuadráticos de las observaciones del exhaustivo del que no se dispone de registro de conductividad. Se dispone de 72 observaciones (1 experimento x 18 muestras x 2 modelos x 2 alternativas por cada modelo).

En la Figura 4.53 se muestra el gráfico de la interacción entre los dos factores estudiados con los intervalos LSD para el valor medio del logaritmo neperiano de los errores cuadráticos de las observaciones del exhaustivo analizado en las cuatro condiciones ensayadas. Los resultados del análisis realizado indican que desde el punto de vista de la capacidad de predicción de la variable respuesta (ΔP), es estadísticamente significativo ($\alpha=0.05$) el efecto del modelo utilizado y también la interacción entre los dos factores (Tabla 4.19). Respecto del factor modelo, se ha obtenido que la capacidad de predicción del M13 es en promedio superior a la del M12, como también se había puesto de manifiesto anteriormente en el estudio de los cuatro exhaustivos de validación (Figura 4.49).

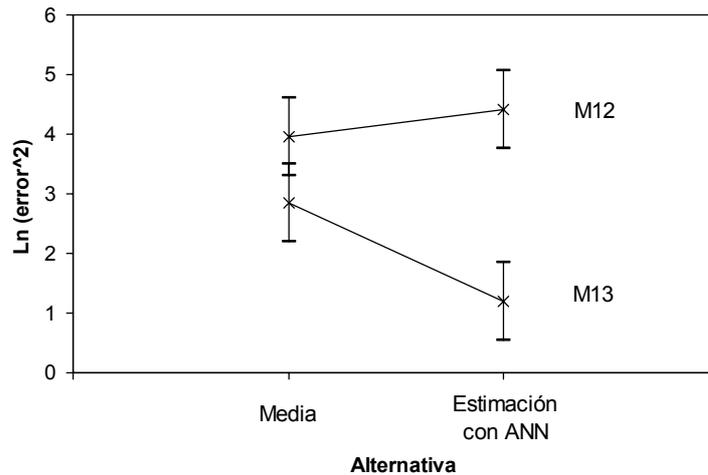


Figura 4.53. Gráfico de la interacción entre los factores estudiados e intervalos LSD para la media del logaritmo neperiano de los errores cuadráticos de las observaciones del exhaustivo que no dispone de registro de conductividad en las cuatro condiciones ensayadas.

En cuanto a la interacción de los factores estudiados, se puede concluir que cuando se emplea el modelo M12, no existen diferencias estadísticamente significativas entre imputar la trayectoria media de ΔCond y predecirla con la red neuronal $\text{MLP}_{4-c,1}$ puesto que sus intervalos LSD se solapan (Figura 4.53). Por otro lado, utilizando el modelo M13, sí que se obtienen resultados significativamente mejores en promedio cuando la trayectoria faltante se predice con la red neuronal $\text{MLP}_{5-c,1}$. Esto parece estar reflejando que si únicamente se consideran las correlaciones cruzadas entre las variables registradas por los sensores instalados en el reactor en cada instante, no hay suficiente información como para poder reconstruir con precisión la trayectoria de los incrementos de conductividad a lo largo de un lote. Por otra parte, los mejores resultados obtenidos por el modelo $\text{MLP}_{5-c,1}$ ponen claramente de manifiesto que los incrementos de conductividad están muy relacionados con los incrementos de fósforo.

Finalmente, en la Figura 4.54 se presenta la evolución de la trayectoria de los valores medidos de la variable respuesta (ΔP) en el experimento que no se dispone de registro de conductividad, junto con los valores predichos por cada uno de los modelos desarrollados. Este exhaustivo no fue utilizado ni para la construcción ni para la validación de los trece modelos recogidos en la Tabla 4.17.

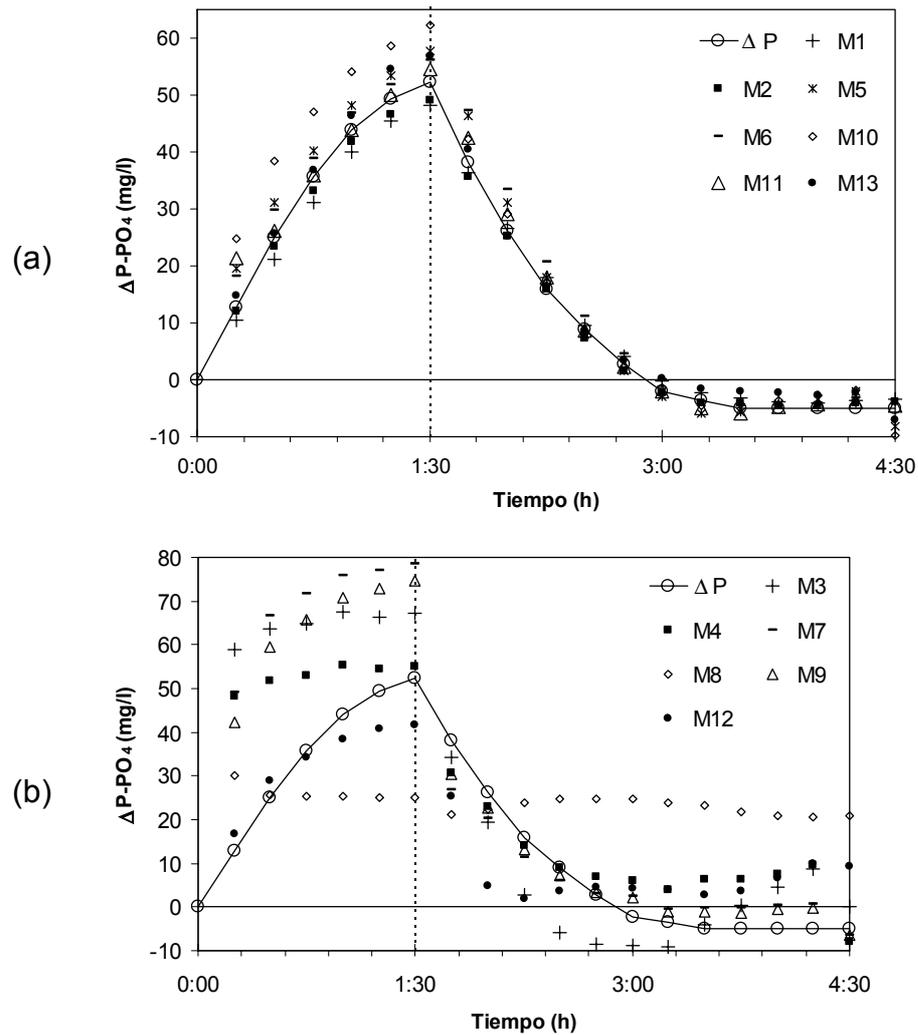


Figura 4.54. Resultados de las predicciones de la variable respuesta (ΔP) de todos los modelos en el exhaustivo que no tiene registro de conductividad: (a) los 7 modelos que presentan mejores predicciones (b) los 6 modelos que ofrecieron peores predicciones.

En la figura anterior se puede apreciar que de los modelos que mantienen la dirección de las variables, sólo aquellos que incluyen en su formulación la variable respuesta decalada han sido capaces de dar resultados bastante acertados, mientras que los que mantienen la dirección de los lotes presentan resultados muy próximos a los obtenidos en laboratorio utilizando sólo los valores registrados por los sensores. En esta situación de datos faltantes, se vuelve a poner de manifiesto la importancia que tiene el hecho de que el modelo inferencial creado pueda aprovechar todas autocorrelaciones y correlaciones cruzadas entre las variables a lo largo del lote completo.

RESUMEN Y CONCLUSIONES

5 CONCLUSIONES

En la presente tesis se han aplicado distintas técnicas para aprovechar la información contenida en las variables de proceso registradas en un sistema SBR a escala de laboratorio operado para la eliminación biológica de fósforo en las aguas residuales. Las principales conclusiones que pueden obtenerse como resultado de este trabajo de investigación son:

1. Previamente a la utilización de cualquier modelo empírico, es fundamental estudiar si es posible crear nuevas trayectorias (a partir de las registradas) que puedan contener información relevante del proceso. De esta manera, se incorpora conocimiento técnico del proceso en los modelos desarrollados, lo que puede contribuir a mejorar tanto su capacidad de diagnóstico como de monitorización.
2. La metodología *blockwise* PCR propuesta por Zarzo y Ferrer (2004) puede ser una herramienta muy eficiente y de gran utilidad en el campo del tratamiento de las aguas residuales. Esta metodología permite extraer la información contenida en una base de datos histórica e identificar puntos críticos del proceso, pudiendo, por tanto, contribuir a mejorar el conocimiento del mismo y a su optimización.
3. En el sistema SBR estudiado, la trayectoria de los incrementos de conductividad respecto de su valor al comienzo del lote (trayectoria creada) es la que presenta una mayor correlación con el rendimiento en la eliminación biológica de fósforo. Esto es debido a que en el proceso de eliminación biológica de fósforo, la liberación y captura de fósforo (junto con los contraiones potasio y magnesio) que realizan las bacterias PAO, provoca cambios importantes en la amplitud de esta trayectoria.
4. Tanto el enfoque propuesto por Nomikos y MacGregor (1995) como el propuesto por Wold *et al.* (1998) permiten modelar adecuadamente los datos de los lotes finalizados (análisis a posteriori, *off-line*) del SBR, y son capaces de

detectar distintos tipos de anomalías. Sin embargo, el diagnóstico de fallos con el primer enfoque ha resultado más consistente y ha permitido una interpretación mucho más clara y sencilla. Esto se debe, por un lado, al desdoblamiento de la matriz tridireccional empleado (tipo E), que permite comparar directamente las trayectorias de los pesos del modelo, así como los gráficos de contribución, con las trayectorias de las variables originales y, por otro, a que los dos niveles con los que se trabaja son: variables originales y variables latentes, siendo directa la relación entre ambas.

5. El nivel de observación del enfoque de Wold *et al.* (1998), útil para la monitorización en tiempo real del SBR, ha mostrado una capacidad para la detección de anomalías bastante limitada. El análisis del modelo PLS creado en este nivel, reveló que las predicciones del índice de madurez eran poco acertadas (especialmente durante toda la etapa anaerobia y al comienzo de la de sedimentación), y que la relación interna presentaba una no linealidad muy marcada. Estos hechos son claros indicadores de un mal funcionamiento del método y se producen como consecuencia del comportamiento no monótono que presentan varias variables del proceso.
6. El nivel de observación del enfoque de Wold *et al.* (1998) ha presentado un problema de sobreestimación de la varianza de las variables latentes que ha ocasionado que anomalías detectadas en los gráficos de control de las variables latentes no hayan sido detectadas en el gráfico de control de la T^2 de Hotelling, lo cual no es deseable desde el punto de vista del esquema general de la monitorización del proceso.
7. La alternativa propuesta para tratar de superar los inconvenientes detectados en el nivel de observación del enfoque de Wold *et al.* (1998) en la que se emplea un preprocesamiento (centrado y escalado) diferente, permitió mejorar considerablemente la capacidad de detección de anomalías respecto del enfoque original en el conjunto de datos analizado. Además, con esta alternativa no se ha producido el problema de la sobreestimación de la varianza de las variables latentes y, por tanto, las anomalías detectadas en los gráficos

de control de las variables latentes también se detectaron en el gráfico de la T^2 de Hotelling.

8. Los cuatro métodos sugeridos por Nomikos y MacGregor (1995) para la monitorización en tiempo real de un proceso por lotes (construcción de múltiples modelos de referencia, imputación con desviación nula, desviación constante y datos faltantes), así como la alternativa propuesta (en la que se emplea un preprocesamiento diferente) para tratar de superar los inconvenientes detectados en el nivel de observación del enfoque de Wold *et al.* (1998), han mostrado un buen comportamiento en la monitorización en tiempo real del reactor SBR operado para la eliminación biológica de fósforo en las aguas residuales. La principal ventaja que tiene esta última estrategia y también la utilización de múltiples modelos de referencia, es que no es necesario rellenar la parte futura del lote. Sin embargo, se debe tener presente que desde un punto de vista práctico, resulta más sencillo y eficiente emplear un único modelo.
9. En general, el gráfico de la distancia al modelo presenta una mayor capacidad para la detección de fallos y, además, refleja la anomalía con anterioridad a su aparición en el gráfico de la T^2 de Hotelling.
10. En el sistema SBR operado para la eliminación biológica de fósforo en las aguas residuales, además de las autocorrelaciones y correlaciones cruzadas entre las variables, existe una correlación importante entre lotes. Por este motivo, se propone un esquema de monitorización con dos niveles, de manera que se pueda realizar un seguimiento del transcurso en tiempo real del lote actual y también de la evolución general del proceso asociada a los lotes anteriores ya finalizados. Dentro de este esquema, los gráficos de contribución constituyen una herramienta eficaz para ayudar en el diagnóstico de las anomalías producidas.

11. En el conjunto de datos analizado del proceso de eliminación biológica de fósforo en las aguas residuales, no hay relación entre los valores al comienzo del lote de las variables de calidad (fósforo, potasio y magnesio), y los valores de los sensores instalados en el SBR. Por este motivo, se recomienda que los modelos predictivos que se desarrollen, utilicen los incrementos de las variables respecto de su valor al comienzo de cada lote (trayectorias corregidas) en lugar de los valores registrados (trayectorias originales).
12. A partir de las trayectorias corregidas de pH y de conductividad es posible predecir adecuadamente la evolución de las trayectorias corregidas de las variables de calidad, mediante la utilización de un modelo PLS en el que la matriz tridireccional de los datos del SBR es desdoblada según la dirección de los lotes.
13. Los modelos predictivos que mejor comportamiento han mostrado son los modelos PLS desarrollados manteniendo la dirección de los lotes en el desdoblamiento de la matriz tridireccional de los datos del SBR. Estos modelos tienen la ventaja de que para realizar las predicciones sólo utilizan datos procedentes de los sensores instalados en el reactor, por lo que no es necesario medir la variable respuesta. Además, pueden ser utilizados en tiempo real tratando la trayectoria futura del lote como datos faltantes, y podrían, por tanto, ser integrados en el esquema de monitorización propuesto. Utilizando este tipo de desdoblamiento no se han construido modelos de redes neuronales para evitar problemas de sobreajuste, debido a que el número de parámetros a estimar sería muy superior al de observaciones disponibles.
14. Los modelos de redes neuronales no han mostrado diferencias estadísticamente significativas con respecto a los modelos de proyección sobre estructuras latentes desarrollados manteniendo la dirección de las variables y utilizando las mismas variables explicativas. La incorporación de la dinámica de la variable respuesta en estos modelos mejora su capacidad predictiva, pero obliga en la práctica a tener que determinar el valor de la misma.

15. Los modelos predictivos de proyección sobre estructuras latentes presentan, frente a las redes neuronales, una serie de ventajas: toleran la presencia de datos faltantes tanto en la etapa de construcción del modelo como en la de validación y explotación, permiten comprobar si las nuevas observaciones son semejantes a las utilizadas en la etapa de construcción del modelo, y suelen ser fácilmente interpretables, por lo que pueden ser utilizados para mejorar el conocimiento del proceso.

16. Las técnicas de análisis estadístico multivariante permiten aprovechar la información contenida en las variables registradas del proceso para detectar fallos que pueden ser corregidos antes de la finalización del lote actual y/o que pueden ser evitados para los lotes siguientes, por lo que con su aplicación en el proceso de eliminación biológica de fósforo, se contribuye a conseguir una operación eficiente del mismo y a mejorar la calidad del efluente.

DESARROLLOS FUTUROS

6 DESARROLLOS FUTUROS

El desarrollo de la presente tesis deja las puertas abiertas a nuevos trabajos de investigación, de entre los cuales se presentan a continuación los que se consideran más interesantes:

- En procesos que están constituidos por varias etapas diferenciadas, la aplicación de los métodos de proyección multi-bloque puede resultar muy beneficiosa. En el análisis multi-bloque, la matriz de datos es separada en varios bloques representado cada uno de éstos una etapa del proceso. Estos métodos permiten establecer gráficos de monitorización tanto para cada bloque de variables, como para el proceso en conjunto y, por tanto, pueden facilitar el diagnóstico de las causas del fallo.
- La duración de las etapas de un proceso por lotes puede cambiar, dando lugar, por tanto, a lotes cuya duración total no siempre es la misma. En esta situación es necesario recurrir a técnicas que permitan sincronizar y alinear las trayectorias de todas las etapas. En la bibliografía se han propuesto distintas técnicas, entre las que cabe destacar el método de la variable indicadora (Nomikos y MacGregor, 1995), el método del índice de madurez (Wold *et al.*, 1998), y el método de la distorsión dinámica del tiempo (Kassidas *et al.*, 1998). Se trataría, por tanto, de estudiar la aplicabilidad de estas técnicas sobre un SBR para el tratamiento de aguas residuales.
- En la monitorización en tiempo real de un proceso por lotes empleando el enfoque de Nomikos y MacGregor (1995) que preserva la dirección de los lotes en el desdoblamiento, es posible tratar la evolución futura del lote como un problema de datos faltantes. Existen distintos métodos para abordar esta situación, entre los cuales el método de regresión sobre los *scores* truncados (TSR, *Trimmed Scores Regression*), recientemente propuesto por Arteaga y Ferrer (2002) puede resultar muy eficiente para la monitorización en tiempo real de este tipo de procesos.

- La monitorización de un proceso por lotes también puede realizarse mediante el enfoque adaptativo propuesto por Rännar *et al.* (1998). Habría que estudiar su aplicabilidad sobre un SBR para el tratamiento de aguas residuales y, en particular, resultaría muy interesante analizar la influencia del parámetro adaptativo sobre la monitorización del proceso, así como las relaciones que pudieran establecerse con los enfoques analizados en la presente tesis.
- Cambios en las condiciones operacionales del proceso, como los debidos a fluctuaciones en las características del influente, en la temperatura, etc., pueden ocasionar que el proceso muestre un comportamiento no estacionario. Se podría realizar un estudio para determinar cuál es la forma más eficiente (construcción de múltiples modelos, utilización de modelos recursivos, análisis multirresolución,...) de abordar esta situación según el tipo de cambio experimentado por el sistema.
- La construcción de modelos predictivos podría realizarse combinando técnicas de proyección sobre estructuras latentes (PCA y PLS) con redes neuronales artificiales (ANN). Una forma de hacerlo podría ser sustituyendo la capa de entrada de la ANN por un modelo PCA o PLS. De esta forma, el modelo resultante podría tolerar la presencia de datos faltantes, comprobar si las nuevas observaciones pueden considerarse semejantes a las utilizadas para construir el modelo, y aprovechar las posibles no linealidades que pudieran existir en los datos.

BIBLIOGRAFÍA

7 BIBLIOGRAFÍA

- Aarnio, P.; Minkkinen, P. (1986). Application of partial least squares modelling in the optimization of a wastewater treatment plant. *Anal. Chim. Acta*, 191, 457-460.
- Aguado, D.; Montoya, T.; Barat, R.; Ferrer, J.; Seco, A. (2004). Modelling of the conductivity variations due to biological activity in an enhanced biological phosphorus removal sequencing batch reactor. Actas del congreso: *3rd IWA Specialised Conference on Sequencing Batch Reactor*, celebrado en Noosa, Queensland, Australia, 26-29 Febrero 2004, pp 325.
- American Public Health Association, American Water Works Association and Water Environmental Federation. (1998). *Standard methods for the examination of water and wastewater*. 20th Ed., Washington D.C.
- Arteaga, F.; Ferrer, A. (2002). Dealing with missing data in MSPC: several methods, different interpretations, some examples. *J. Chemometrics*, 16, 408-418.
- Arvin, E.; Kristensen, G. (1985). Exchange of organics, phosphate and cations between mediated chemical precipitation – A review. *Wat. Sci. Tech.*, 17 (11/12), 147-162.
- Baeza, J.A.; Gabriel, D.; Lafuente, J. (2002). In-line fast OUR (oxygen uptake rate) measurements for monitoring and control of WWTP. *Wat. Sci. Tech.*, 45 (4-5), 19-28.
- Barat, R. (2004). *Estudio de la precipitación de fósforo inducida biológicamente en el tratamiento de aguas residuales*. Tesis doctoral. Departamento de Ingeniería Hidráulica y Medio Ambiente, Universidad Politécnica de Valencia.
- Belanche, Ll.; Valdés, J.J.; Comas, J.; Roda, I.; Poch, M. (1998). Modeling the input-output behaviour of wastewater treatment plants using soft computing techniques. *ECAI'98 Workshop on Binding Enviromental Science and Artificial Intelligence*, Brighton, UK.
- Bishop, C.M. (1996). Theoretical foundations of neural networks. Proceedings of *Physics Computing 96*, Krakow, 500-507.
- Bordacs, K.; Chiesa, S.C. (1989). Carbon flow patterns in enhanced biological phosphorus accumulating activated sludge cultures. *Wat. Sci. Tech.*, 21 (4/5), 387-396.

- Box, G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: effect of inequality of variance in one-way classification. *The Annals of Mathematical Statistics*, 25, 290-302.
- Brdjanovic, D.; van Loosdrecht, M.C.M.; Hooijmans, C.M.; Mino, T.; Alaerts, G.J.; Heijnen, J.J. (1998). Effect of polyphosphate limitation on the anaerobic metabolism of phosphorus-accumulating microorganisms. *Appl. Microbiol. Biotech.*, 50, 273-276.
- Bro, R. (1996). Multiway calibration. Multilinear PLS. *J. Chemometrics*, 10, 47-61.
- Capodaglio, A.G.; Jones, H.V.; Novotny, V.; Feng, X. (1991). Sludge bulking analysis and forecasting: application of system identification and artificial neural computing technologies. *Wat. Res.*, 25, 1217-1224.
- Carroll, J.D.; Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via a n -way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35, 283-319.
- Carstensen, J.; Harremoës, P.; Strube, R. (1996). Software sensors based on the grey-box modelling approach. *Wat. Sci. Tech.*, 33 (1), 117-126.
- Choi, D. J.; Park, H. (2001). A hybrid artificial neural network as a software sensor for optimal control of a wastewater treatment process. *Wat. Res.*, 35 (16), 3959-3967.
- Comeau, Y.; Hall, K.J.; Hancock, R.E.W.; Oldham, W.K. (1986). Biochemical model for enhanced biological phosphorus removal. *Wat. Res.*, 20 (12), 1511-1521.
- Comeau, Y.; Rabinowitz, B.; Hall, K.J.; Oldham, K.W. (1987). Phosphate release and uptake in enhanced biological phosphorus removal from wastewater. *Journal WPCF*, 59, 707-715.
- Dayal, B.S.; MacGregor, J.F.; Taylor, P.A.; Kildaw, R.; Marcikic, S. (1994). Application of feedforward neural networks and partial least squares regression for modelling kappa number in a continuous kamyr digester. *Pulp & Paper Canada*, 95 (1), 7-13.
- De Jong, S. (1993). SIMPLS - an alternative approach to partial least squares regression. *Chemom. Intell. Lab. Syst.*, 18, 251-263.

- Devroye, L.; Györfi, L.; Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag, New York.
- El-Din, A.G.; Smith, D.W. (2002). Modelling approach for high flow rate in wastewater treatment operation. *J. Environ. Eng. Sci.*, 1, 275-291.
- Ferrer, A. (2003). Control estadístico megavariante para los procesos del siglo XXI. *27 Congreso Nacional de Estadística e Investigación Operativa*, Lleida (España).
- Frank, P.M. (1990). Fault diagnosis in dynamic systems using analytical and knowledge based redundancy. A survey and some new results. *Automatica*, 26, 459-474.
- Fukase, T.; Shibata, M.; Miyaji, Y. (1985). Role of an anaerobic stage on biological phosphorus removal. *Wat. Sci. Tech.*, 17 (2/3), 69-80.
- Geladi, P.; Kowalski, B.R. (1986). Partial Least-Squares Regression: a tutorial. *Anal. Chim. Acta*, 185, 1-17.
- Grady, C.P.L. Jr.; Daigger, G.T.; Lim, H.C. (1999). *Biological wastewater treatment*. 2ª edición revisada y extendida. Marcel Dekker Inc., New York. Basel.
- Greenburg, A.E.; Levin, G.; Kauffman, W.J. (1955). The effect of phosphorus removal on the activated sludge process. *Sewage and Industrial Wastes*, 27, 227.
- Häck, M.; Köhne, M. (1996). Estimation of wastewater process parameters using neural networks. *Wat. Sci. Tech.*, 33 (1), 101-115.
- Hamed, M.M.; Khalafallah, M.G.; Hassanien E.A. (2004). Prediction of wastewater treatment plant performance using artificial neural networks. *Environ. Modell. Soft.*, 19 (10), 919-928.
- Harremoës, P.; Capodaglio, A.G.; Hellström, B.G.; Henze, M.; Jensen, K.N.; Lynggaard-Jensen, A.; Otterpohl, R.; Soeberg, H. (1993). Wastewater treatment plants under transient loading performance, modelling and control. *Wat. Sci. Tech.*, 27 (12), 71-115.

- Harris, T.J.; Ross, W.H. (1991). Statistical process control procedures for correlated observations. *Canadian Journal of Chemical Engineering*, 69, 48-57.
- Harshman, R.A. (1970). Foundations of the PARAFAC procedure: models and conditions for an “explanatory” multimodal factor analysis. *UCLA working papers in phonetics*, 16, 1-84.
- Haykin, S. (1999). *Neural networks. A comprehensive foundation*. Prentice-Hall, New Jersey.
- Hebb, D.O. (1949). *The organization of behavior: A neuropsychological theory*. Wiley, New York.
- Holloway, L.E.; Krogh, B.H. (1992). On-Line Trajectory Encoding for Discrete – Observation Process Monitoring. *On-Line Fault Detection and Supervision in the Chemical Process Industries*, IFAC Symposium, Newark, Delaware, April 22-24.
- Hong, Y.-S.; Bhamidimarri, R. (2003). Evolutionary self-organising modelling of a municipal wastewater treatment plant. *Wat. Res.*, 37, 1199-1212.
- Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences, USA*, 79, 2554-2558.
- Hornik, K.; Stinchcombe, M.; White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, 359-366.
- Hunter, J.S. (1986). Exponentially weighted moving average. *J. Quality Technology*, 18, 203-210.
- Imai H., Endoh, K.; Kozuka, T. (1988). Magnesium requirement for biological removal of phosphate by activated sludge. *Journal of Fermentation Technology*, 66, 657-666.
- Jackson, J.E.; Mudholkar, G.S. (1979). Control procedures for residuals associated with principal components analysis. *Technometrics*, 21, 341-349.
- Jackson, J.E. (1991). *A User's Guide to Principal Components*. John Wiley & Sons, New York, NY.
- Jansson, A.; Röttorp, J.; Rahmberg, M. (2002). Development of a software sensor for phosphorus in municipal wastewater. *J. Chemometrics*, 16, 542-547.

- Jardin, N.; Pöpel, H.J. (1994). Phosphate release of sludges from enhanced biological P-removal during digestion. *Wat. Sci. Tech.*, 20 (6), 281-292.
- Jardin, N.; Pöpel, H.J. (1996). Behavior of waste activated sludge from enhanced biological phosphorus removal during sludge treatment. *Water Environmental Research*, 68, 965-973.
- Jeon, C.O.; Park, J.M. (2000). Enhanced biological phosphorus removal in a sequencing batch reactor supplied with glucose as a sole carbon source. *Wat. Res.*, 34 (7), 2160-2170.
- Jeppsson, U.; Alex, J.; Pons, M.N.; Spanjers, H.; Vanrolleghem, P.A. (2002). Status and future trends of ICA in wastewater treatment – A European perspective. *Wat. Sci. Tech.*, 45 (4-5), 485-494.
- Kassidas, A.; MacGregor, J.F.; Taylor, P.A. (1998). Synchronization of batch trajectories using dynamic time warping. *AIChE J.*, 44, 864-875.
- Katsogiannis, A.N.; Kornaros, M.E.; Lyberatos, G.K. (1999). Adaptive optimization of a nitrifying sequencing batch reactor. *Wat. Res.*, 33 (17), 3569-3576.
- Kiers, H.A.L. (2000). Towards a standardized notation and terminology in multiway analysis. *J. Chemometrics*, 14, 105-122.
- Kim, H.; Hao O.J. (2001). pH and oxidation-reduction potential control strategy for optimization of nitrogen removal in an alternating aerobic-anoxic system. *Water Environment Research*, 73 (1), 95-102.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59-69.
- Konstantinov, K.B.; Yoshida, T. (1992). A method for on-line reasoning about the time-profiles of process variables. *On-line fault detection and supervision in the chemical process industries*, IFAC Symposium, Newark, Delaware, April 22-24.
- Kosanovich, K.A.; Dahl, K.S.; Piovoso, M.J. (1996). Improved process understanding using multiway principal component analysis. *Ind. Eng. Chem. Res.*, 35, 138-146.
- Kourti, T.; Nomikos, P.; MacGregor, J.F. (1995). Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS. *Journal of Process Control*, 5 (4), 277-284.

- Kourti, T.; Lee, J.; MacGregor, J.F. (1996). Experiences with industrial applications of projections methods for multivariate statistical process control. *Computers in Chemical Engineering*, 20 Suppl., 745-750.
- Kourti, T.; MacGregor, J.F. (1996). Multivariate SPC methods for process and product monitoring. *Journal of Quality Technology*, 28 (4), 409-428.
- Kourti, T. (2003). Multivariate dynamic data modeling for analysis and statistical process control of batch processes, start-ups and grade transitions. *J. Chemometrics*, 17, 93-109.
- Kresta, J.V.; MacGregor, J.F.; Marlin, T.E. (1991). Multivariate statistical monitoring of process operating performance. *Canadian Journal of Chemical Engineering*, 69, 35-47.
- Larsen, L.H.; Kjaer, T.; Revsbech, N.P. (1997). A microscale NO_3^- biosensor for environmental applications. *Anal. Chem.*, 69, 3527-3531.
- Larsen, L.H.; Damgaard, L.R.; Kjaer, T.; Stenstrom, T.; Lynggaard-Jensen, A.; Revsbech, N.P. (2000). Fast responding biosensor for on-line determination of nitrate/nitrite in activated sludge. *Wat. Res.*, 34, 2463-2468.
- Lee, D.S.; Park, J.M. (1999). Neural network modeling for on-line estimation of nutrient dynamics in a sequentially-operated batch reactor. *J. Biotechnol.*, 75, 229-239.
- Lee, D.S.; Jeon, C.O.; Park, J.M.; Chang K.S. (2002). Hybrid neural network modeling of a full-scale industrial wastewater treatment process. *Biotechnology and Bioengineering*, 78 (6), 670-682.
- Lee, D.S.; Vanrolleghem, P.A. (2003). Monitoring of a sequencing batch reactor using adaptive multiblock principal component analysis. *Biotechnology and Bioengineering*, 82 (4), 489-497.
- Lee, D.S.; Vanrolleghem, P.A. (2004). Adaptive consensus principal component analysis for on-line batch process monitoring. *Environmental Monitoring and Assessment*, 92, 119-135.
- Lee, C.; Choi, S.W.; Lee, I.-B. (2004a). Sensor fault identification based on time-lagged PCA in dynamic processes. *Chemom. Intell. Lab. Syst.*, 70, 165-178.
- Lee, J.-M.; Yoo, C.K.; Lee, I.-B. (2004b). Statistical process monitoring with independent component analysis. *Journal of Process Control*, 14, 467-485.

- Lee, J.-M.; Yoo, C.K.; Choi, S.W.; Vanrolleghem, P.A.; Lee, I.-B. (2004c). Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 59, 223-234.
- Lennox, J.; Rosen, C. (2002). Adaptive multiscale principal component analysis for on-line monitoring of wastewater treatment. *Wat. Sci. Tech.*, 45 (4-5), 227-235.
- Lohmöller, J.; Wold, H. (1980). Three-mode path models with latent variables and partial least squares (PLS) parameter estimation. *European Meeting of Psychometrics Society*, Groningen, The Netherlands.
- Louwerse, D.J.; Smilde, A.K. (2000). Multivariate statistical process control of batch processes using three-way models. *Chemical Engineering Science*, 50, 1225-1235.
- Luccarini, L.; Porrà, E.; Spagni, A.; Ratini, P.; Grilli, S.; Longhi, S.; Bortone, G. (2002). Soft sensors for control of nitrogen and phosphorus removal from wastewaters by neural networks. *Wat. Sci. Tech.*, 45 (4-5), 101-107.
- MacCulloch, W.; Pitts, W. (1943). A logical calculus of the ideas immanent nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.
- MacGregor, J.F.; Jaeckle, C.; Kiparissides, C.; Koutoudi, M. (1994). Process monitoring and diagnosis by multiblock PLS methods. *AIChE J.*, 40 (5), 826-838.
- Maier, H.R.; Dandy, G.C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environ. Modell. Soft.*, 15, 101-124.
- Massy, W.F. (1965). Principal components regression in exploratory statistical research. *J. American Statistical Association*, 60, 234-246.
- Masters, T. (1993). *Practical Neural Network Recipes in C++*. Academic Press, San Diego.
- Maurer, M.; Gujer, W. (1995). Monitoring of microbial phosphorus release in batch experiments using electric conductivity. *Wat. Res.*, 29 (11), 2613-2617.
- Maurer, M.; Gujer, W.; Hany, R.; Bachmann, S. (1997). Intracellular carbon flow in phosphorus accumulating organisms from activated sludge systems. *Wat. Res.*, 31 (4), 907-917.

- Metcalf y Eddy (2000). *Ingeniería de aguas residuales. Tratamiento, vertido y reutilización*. 3ª edición. Editorial McGraw-Hill.
- Miettinen, T.; Hurse, T.J.; Connor, M.A.; Reinikainen, S.-P.; Minkkinen, P. (2004). Multivariate monitoring of a biological wastewater treatment process: a case study at Melbourne Water's Western Treatment Plant. *Chemom. Intell. Lab. Syst.*, 73 (1), 131-138.
- Miller, P.; Swanson, R.E.; Heckler, C.F. (1993). Contribution plots: the missing link in multivariate quality control. Presented at annual fall technical conference of the American Society for Quality Control (Milwaukee, WI) and the American Statistical Association (Alexandria, VA).
- Mino, T.; Arun, V.; Tsuzuki, Y.; Matsuo, T. (1987). Effect of phosphorus accumulation on acetate metabolism in the biological phosphorus removal process. *Biological Phosphate Removal from Wastewaters*. Rome, R. Ramadori, editor. Pergamon Press, UK, 27-38.
- Mino, T.; Liu, W.-T.; Kurisu, F.; Matsuo, T. (1995). Modelling glycogen storage and denitrification capability of microorganisms in enhanced biological phosphate removal processes. *Wat. Sci. Tech.*, 31 (2), 25-34.
- Minsky, M.L.; Papert, S.A. (1969). *Perceptrons*. MIT Press, Cambridge, Massachusetts.
- Miyamoto-Mills, J.; Larson, J.; Jenkins, D.; Owen, W. (1983). Design and operation of a pilot-scale biological phosphate removal plant at Central Contra Costa sanitary district. *Wat. Sci. Tech.*, 15 (3/4), 153-179.
- Moosbrugger, R.E.; Wentzel, M.C.; Ekama, G.A.; Marais, G.v.R. (1992). Simple titration procedures to determine H_2CO_3^* alkalinity and short-chain fatty acids in aqueous solutions containing known concentrations of ammonium, phosphate and sulphide weak acid/bases. *Water Research Commission, Report No. TT 57/92. Research Report W 74*, University of Cape Town, Pretoria, Republic of South Africa.
- Mujunen, S.-P.; Minkkinen, P.; Teppola, P.; Wirkkala, R.-S. (1998). Modeling of activated sludge plants treatment efficiency with PLSR: a process analytical case study. *Chemom. Intel. Lab. Syst.*, 41, 83-94.

- Nelson, P.; Taylor, P.A.; MacGregor, J.F. (1996). Missing data methods in PCA and PLS: score calculations with incomplete observations. *Chemom. Intell. Lab. Syst.*, 35, 45-65.
- Nomikos, P. (1995). *Statistical process control of batch processes*. Tesis doctoral, McMaster University. Ontario, Canada.
- Nomikos, P.; MacGregor, J.F. (1994). Monitoring of batch processes using multi-way principal component analysis. *AIChE J.*, 40, 1361-1375.
- Nomikos, P.; MacGregor, J.F. (1995). Multivariate SPC Charts for Monitoring Batch Processes. *Technometrics*, 37 (1), 41-59.
- Olsson, G.; Newell R. (1999). *Wastewater Treatment Systems. Modelling, Diagnosis and Control*. IWA Publishing, London, UK.
- Olsson, G. (2002). Lessons learnt at ICA 2001. *Wat. Sci. Tech.*, 45 (4/5), 1-8.
- Önnerth, T. B.; Nielsen, M. K.; Stamer, C. (1996). Advanced computer control based on real and software sensors. *Wat. Sci. Tech.*, 33 (1), 237-245.
- Page, E.S. (1954). Continuous inspection schemes. *Biometrika*, 41, 100-115.
- Pattarkine, V.M.; Randall, C.W. (1999). The requirement of metal cations for enhanced biological phosphorus removal by activated sludge. *Wat. Sci. Tech.*, 40 (2), 159-165.
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2, 559-572.
- Ramaker, H.J.; Sprang, E.N.M.; Gurden, S.P.; Westerhuis, J.A.; Smilde, A.K. (2002). Improved monitoring of batch processes by incorporating external information. *J. Process Control*, 12, 569-576.
- Ramaker, H.-J.; Sprang, E. (2004). *Statistical batch process monitoring*. Tesis doctoral. Department of Chemical Engineering, Process, Analysis and Chemometrics, University of Amsterdam, The Netherlands.
- Rännar, S.; MacGregor, J.F.; Wold, S. (1998). Adaptive batch monitoring using hierarchical PCA. *Chemom. Intell. Lab. Syst.*, 41, 73-81.

- Rickard, L.F.; McClintock, S.A. (1992). Potassium and magnesium requirements for enhanced biological phosphorus removal from wastewater. *Wat. Sci. Tech.*, 26 (9/11), 2203-2206.
- Ripley, B.D. (1996). *Pattern recognition and neural networks*. Cambridge University Press.
- Rosen, C.; Olsson, G. (1998). Disturbance detection in wastewater treatment plants. *Wat. Sci. Tech.*, 37 (12), 197-205.
- Rosen, C.; Lennox, J.A. (2001). Multivariate and multiscale monitoring of wastewater treatment operation. *Wat. Res.*, 35 (14), 3402-3410.
- Rosen, C.; Yuan, Z. (2001). Supervisory control of wastewater treatment plants by combining principal component analysis and fuzzy c-means clustering. *Wat. Sci. Tech.*, 43 (7), 147-156.
- Rosen, C.; Röttorp, J.; Jeppsson, U. (2003). Multivariate on-line monitoring: challenges and solutions for modern wastewater treatment operation. *Wat. Sci. Tech.*, 47 (2), 171-179.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 95, 386-408.
- Rozzi, A.; Massone, A.; Amtonelli, M. (1997). A VFA measuring biosensor based on nitrate reduction. *Wat. Sci. Tech.*, 36 (6/7), 183-189.
- Rozzi, A.; Remigi, E.; Buckley, C. (2001). Methanogenic activity measurements by the MAIA biosensor: Instructions guide. *Wat. Sci. Tech.*, 44 (4), 287-294.
- Ruiz, G.; Castellano, M.; González, W.; Roca, E.; Lema, J.M. (2004a). Algorithm for steady state detection of multivariate process: Application to wastewater anaerobic digestion process. Proceedings of 2nd International IWA Conference on Automation in Water Quality Monitoring, AutMoNet, 19-20 April, Vienna, Austria.
- Ruiz, M.L.; Colomer, J.; Rubio, M.; Meléndez, J.; Colprim, J. (2004b). Situation Assessment of a sequencing batch reactor using multiblock MPCA and Fuzzy classification. *BESAI 4th ECAI Workshop on Binding Environmental Science and Artificial Intelligence*, Valencia, Spain.
- Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. (1986). Learning internal representations by error propagation. Vol. 1, MIT Press, Cambridge, Massachusetts, 318-362.

- Sarle, W.S. (1994). Neural networks and statistical models. In: *Proceedings of the Nineteenth Annual SAS Users Group International Conference*, 1538-1550. SAS Institute.
- Schönborn, C.; Bauer, H.-D.; Röske, I. (2001). Stability of enhanced biological phosphorus removal and composition of polyphosphate granules. *Wat. Res.*, 35 (13), 3190-3196.
- Sedlak, R. (1991). *Phosphorus and nitrogen removal from municipal wastewater. Principles and practice*. 2ª edición. Lewis Publishers.
- Serralta, J.; Borrás, L.; Blanco, C.; Barat, R.; Seco, A. (2004). Monitoring pH and electric conductivity in an EBPR sequencing batch reactor. *Wat. Sci. Tech.*, 50 (10), 145-152.
- Shewart, W.A. (1931). *Economic control of quality of manufactured product*. Van Nostrand, Princeton, NJ.
- Teichgräber, B. (1993). Control strategies for a highly loaded biological ammonia elimination process. *Wat. Sci. Tech.*, 28 (11/12), 531-538.
- Teppola, P.; Mujunen, S.-P.; Minkkinen, P. (1997). Partial least squares modeling of an activated sludge plant: A case study. *Chemom. Intell. Lab. Syst.*, 38, 197-208.
- Teppola, P.; Mujunen, S.-P.; Minkkinen, P. (1998). A combined approach of partial least squares and fuzzy c-means clustering for the monitoring of an activated-sludge waste-water treatment plant. *Chemom. Intell. Lab. Syst.*, 41, 95-103.
- Teppola, P.; Mujunen, S.-P.; Minkkinen, P. (1999a). Adaptive fuzzy c-means clustering in process monitoring. *Chemom. Intell. Lab. Syst.*, 45, 23-38.
- Teppola, P.; Mujunen, S.-P.; Minkkinen, P. (1999b). Kalman filter for updating the coefficients of regression models. A case study from an activated sludge wastewater treatment plant. *Chemom. Intell. Lab. Syst.*, 45, 371-384.
- Tracy, N.D.; Young, J.C.; Mason, L.R. (1992). Multivariate control charts for individual observations. *Journal of Quality Technology*, 24, 88-95.

- Tucker, L.R. (1963). Implications of factor analysis of three-way matrices for measurement of change. In *Problems in Measuring Change*, Harris CW (ed.). University of Wisconsin Press: Madison, WI, 122-137.
- Tucker, L.R. (1964). The extension of factor analysis to three-dimensional matrices. In *contributions to mathematical psychology*, Frederiksen N., Gulliksen H. (eds.). Holt, Rinehart and Winston: New York, 110-182.
- Tucker, L.R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31, 279-311.
- Vanrolleghem, P.A.; Kong, Z.; Rombouts, G.; Verstraete W. (1994). An on-line respirographic biosensor for the characterization of load and toxicity of wastewaters. *J. Chem. Tech. Biotech.*, 59, 321-333.
- Warner, B.; Misra, M. (1996). Understanding neural networks as statistical tools. *American Statistician*, 50 (4), 284-293.
- Weigend, A.S.; Rumelhart, D.E.; Huberman, B.A. (1990). Backpropagation, weight elimination and time series prediction. *Connectionist models, Proceedings of the 1990 summer school*, 105-116.
- Wentzel, M.C.; Lötter, L.H.; Loewenthal, R.E.; Marais, G.v.R. (1986). Metabolic behaviour of *Acinetobacter* spp. in enhanced biological phosphorus removal - A biochemical model. *Water SA*, 12 (4), 209-224.
- Westerhuis, J.A.; Kourti, T.; MacGregor, J.F. (1999). Comparing alternative approaches for multivariate statistical analysis of batch process data. *J. Chemometrics*, 13, 397-413.
- Widrow, B.; Hoff Jr., M.E. (1960). Adaptive switching circuits. IRE WESCON Convention record, 96-104.
- Wild D., Kisliakova A. y Siegrist H. (1997). Prediction of recycle phosphorus loads from anaerobic digestion. *Wat. Res.*, 31 (9), 2300-2308.
- Wold, S. (1978). Cross-validatory estimation of the number of components in factor and principal component models. *Technometrics*, 20, 397-405.
- Wold, H. (1982). Systems under indirect observation using PLS. In *a second generation of multivariate analysis*, Fornell C. (ed.). Praeger: New York, 325-347.

- Wold, S.; Geladi, P.; Esbensen, K.; Öhman, J. (1987). Multi-way principal components and PLS analysis. *J. Chemometrics*, 1, 41-56.
- Wold, S.; Kettaneh, N.; Friden, H.; Holmberg, A. (1998). Modelling and diagnosis of batch processes and analogous kinetic experiments. *Chemom. Intell. Lab. Syst.*, 44, 331-340.
- Woodall, W.H. (2000). Controversities and contradictions in statistical process control. *Journal of Quality Technology*, 32, 341-350.
- Yoo, C.K.; Choi, S.W.; Lee, I.-B. (2002). Disturbance detection and isolation in the activated sludge process. *Wat. Sci. Tech.*, 45 (4-5), 217-226.
- Yoo, C.K.; Vanrolleghem, P.A.; Lee, I.-B. (2003). Nonlinear modeling and adaptive monitoring with fuzzy and multivariate statistical methods in biological wastewater treatment plants. *Journal of Biotechnology*, 105, 135-163.
- Yoo, C.K.; Lee, D.S.; Vanrolleghem P.A. (2004a). Application of multiway ICA for on-line process monitoring of a sequencing batch reactor. *Wat. Res.*, 38, 1715-1732.
- Yoo, C.K.; Lee, J.-M.; Lee, I.-B.; Vanrolleghem P.A. (2004b). Proceedings of 2nd International IWA Conference on Automation in Water Quality Monitoring, AutMoNet, 19-20 April, Vienna, Austria.
- Yoon, S.; MacGregor, J.F. (2001). Incorporation of external information into multivariate PCA/PLS models. 4th IFAC Workshop on on-line fault detection and supervision in the chemical process industries, Chejudo, Korea.
- Yu, R.F.; Liaw, S.L.; Chang, C.N.; Lu, H.J.; Cheng, W.Y. (1997). Monitoring and control using on-line ORP on the continuous-flow activated sludge batch reactor system. *Wat. Sci. Tech.*, 35 (1), 57-66.
- Yu, R.F.; Liaw, S.L.; Chang, C.N.; Cheng, W.Y. (1998). Applying real-time control to enhance the performance of nitrogen removal in the continuous-flow SBR system. *Wat. Sci. Tech.*, 38 (3), 271-280.
- Zarzo, M.; Ferrer, A. (2004). Batch process diagnosis: PLS with variable selection versus block-wise PCR. *Chemom. Intell. Lab. Syst.*, 73, 15-27.

- Zell, A.; Mamier, G.; Vogt, M.; Mache, N.; Hübner, R.; Döring, S.; Herrmann, K.-U.; Soyez, T.; Schmalzl, M.; Sommer, T.; Hatzigeorgiou, A.; Posselt, D.; Schreiner, T.; Kett, B.; Clemente, G.; Wieland, J.; Gatter, J. (1998). *SNNS. Stuttgart Neural Network Simulator*. User Manual, Version 4.2. University of Stuttgart. Institute for parallel and distributed high performance systems. University of Tübingen. Institute for computer science.
- Zhao, H.; Hao, O.J.; McAvoy, T.J.; Chang, C.-H. (1997). Modeling nutrient dynamics in sequencing batch reactor. *Journal of Environmental Engineering*, 123 (4), 311-319.
- Zhu, J.; Zurcher, J.; Rao, M.; Meng, M. Q.-H. (1998). An on-line wastewater quality predication system based on a time-delay neural network. *Engineering Applications of Artificial Intelligence*, 11, 747-758.

8 ANEJOS

8.1 ANEJO 1: FOTOGRAFÍAS DEL MONTAJE EXPERIMENTAL



Figura 8.1. Montaje experimental.



Figura 8.2. SBR utilizado durante el periodo experimental.

8.2 ANEJO 2: GRÁFICOS DE CONTRIBUCIÓN DEL DIAGNÓSTICO

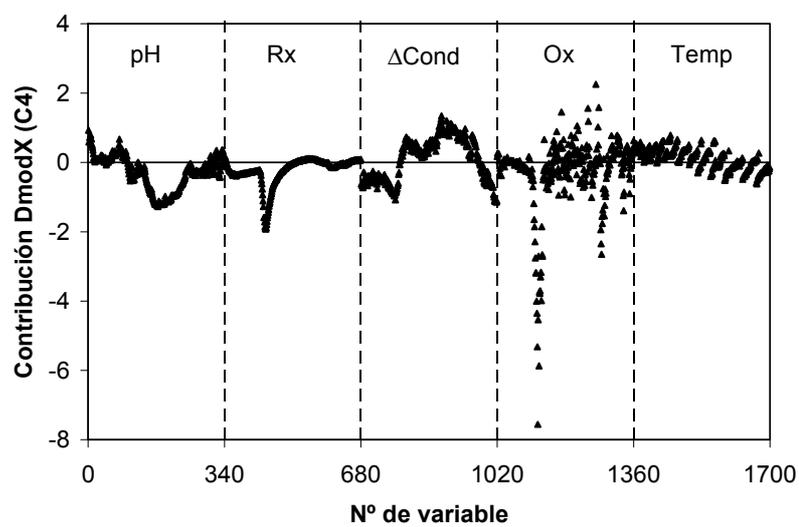
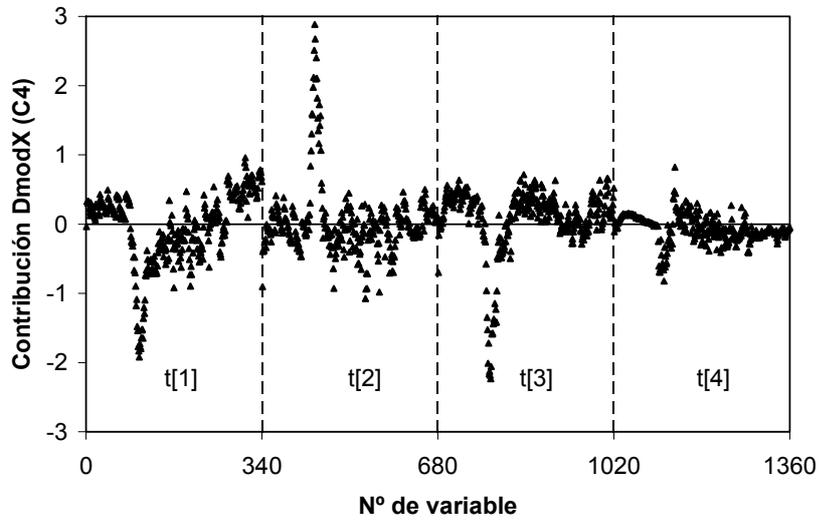
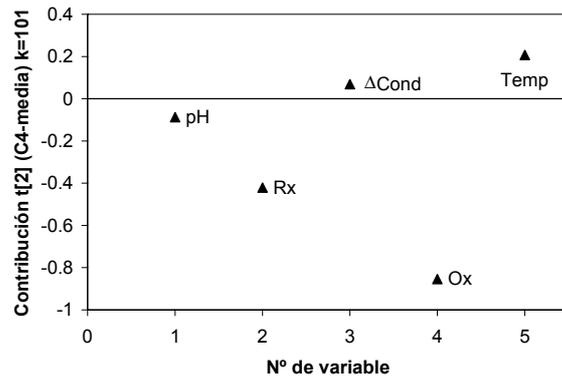
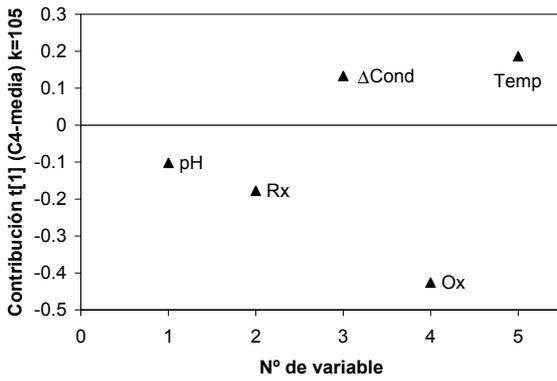


Figura 8.3. Gráfico de contribución de la distancia al modelo del enfoque de Nomikos y MacGregor (1995) para el lote C4.



(a)



(b)

Figura 8.4. Gráficos de contribución del nivel *batch* del enfoque Wold *et al.* (1998) para el lote C4: (a) de la distancia al modelo (b) de las variables latentes $t[1]$ en el instante de tiempo $k=105$ y $t[2]$ en el instante de tiempo $k=101$.

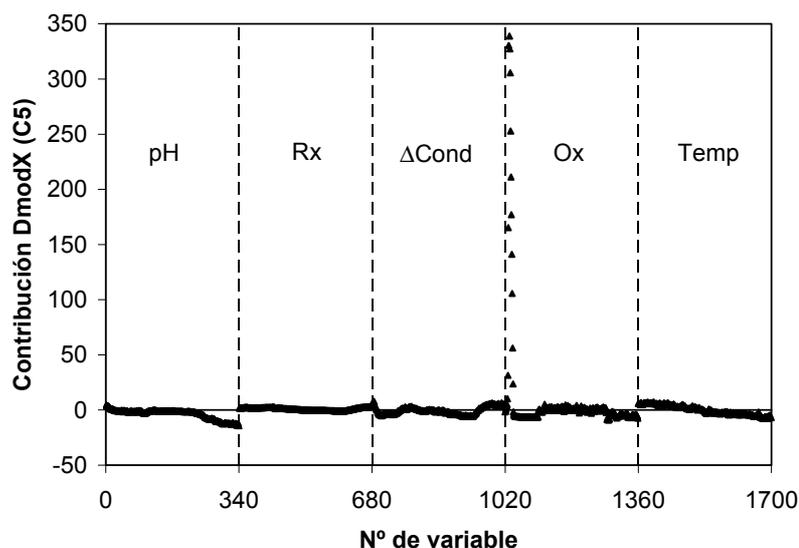


Figura 8.5. Gráfico de contribución de la distancia al modelo del enfoque de Nomikos y MacGregor (1995) para el lote C5.

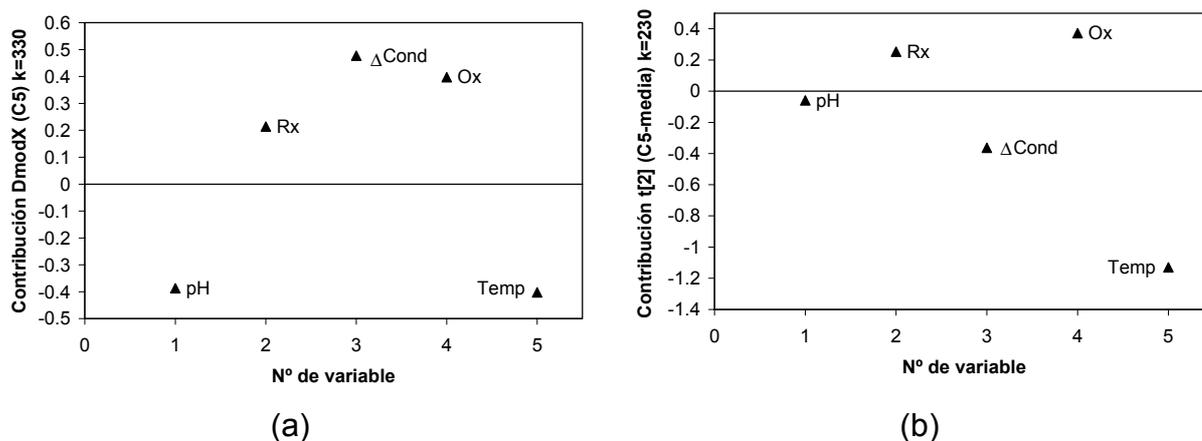
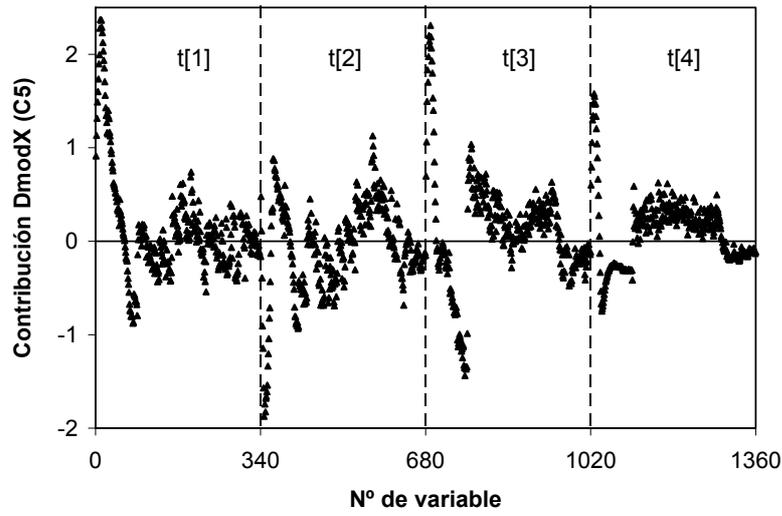
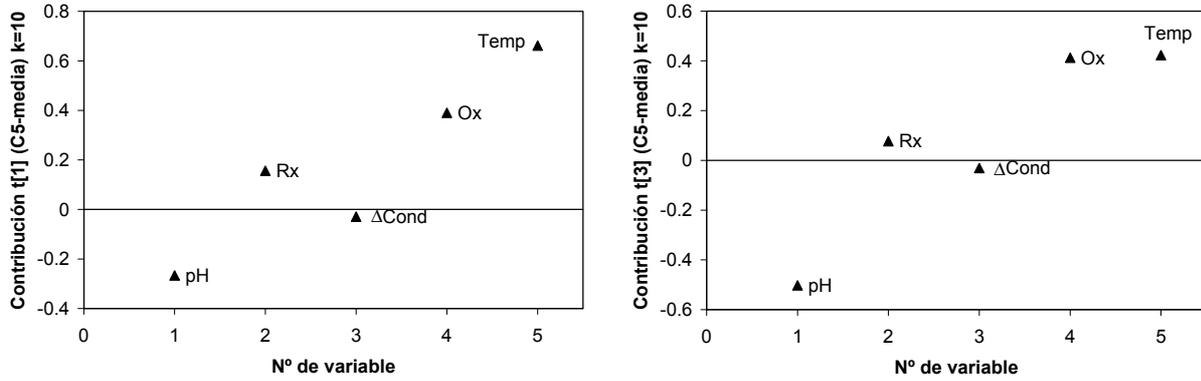


Figura 8.6. Gráficos de contribución del nivel de observación del enfoque Wold *et al.* (1998) para el lote C5: (a) de la distancia al modelo en el instante $k=330$ (b) de la variable latente $t[2]$ en el instante de tiempo $k=230$.



(a)



(b)

Figura 8.7. Gráficos de contribución del nivel *batch* del enfoque Wold *et al.* (1998) para el lote C5: (a) de la distancia al modelo (b) de las variables latentes $t[1]$ y $t[3]$ en el instante de tiempo $k=10$.

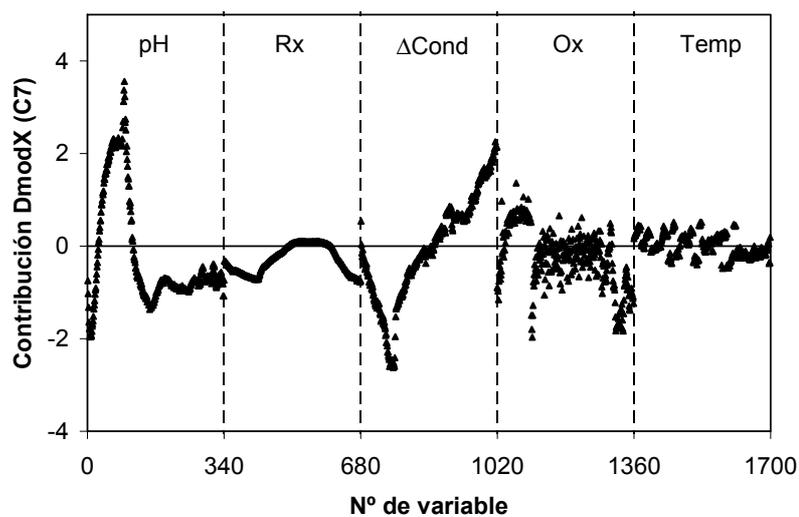


Figura 8.8. Gráfico de contribución de la distancia al modelo del enfoque de Nomikos y MacGregor (1995) para el lote C7.

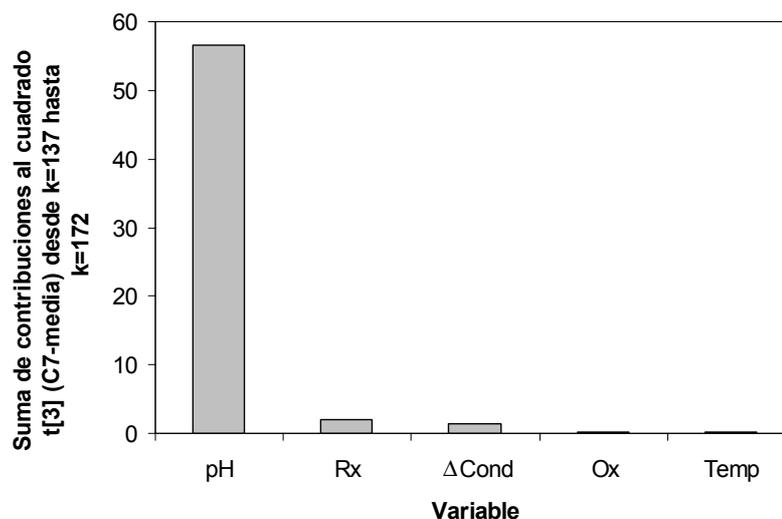
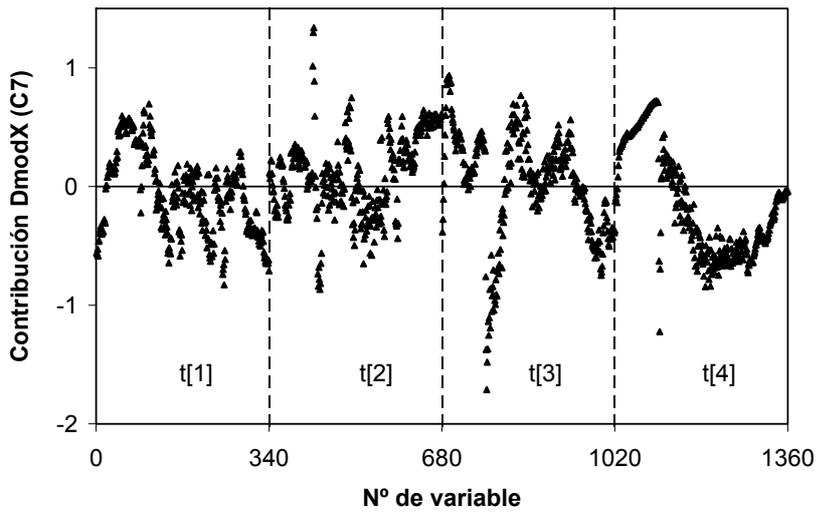
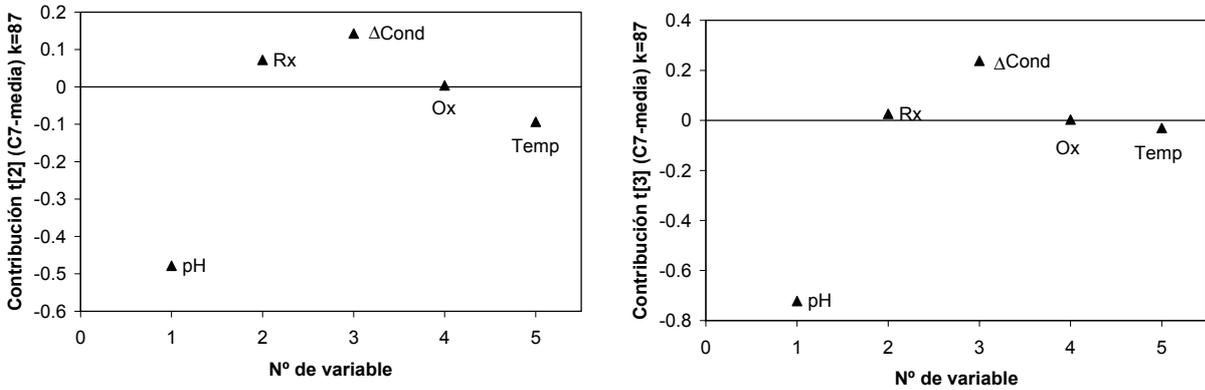


Figura 8.9. Gráfico del sumatorio de las contribuciones al cuadrado de la variable latente $t[3]$ del nivel de observación del enfoque de Wold *et al.* (1998) para el lote C7 entre los instantes de tiempo $k=137$ y $k=172$.



(a)



(b)

Figura 8.10. Gráficos de contribución del nivel *batch* del enfoque Wold *et al.* (1998) para el lote C7: (a) de la distancia al modelo (b) de las variables latentes $t[2]$ y $t[3]$ en el instante de tiempo $k=87$.

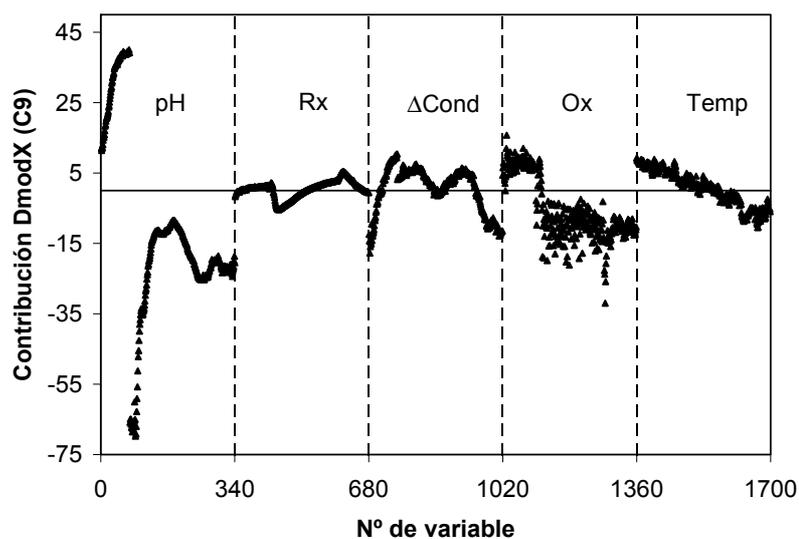


Figura 8.11. Gráfico de contribución de la distancia al modelo del enfoque de Nomikos y MacGregor (1995) para el lote C9.

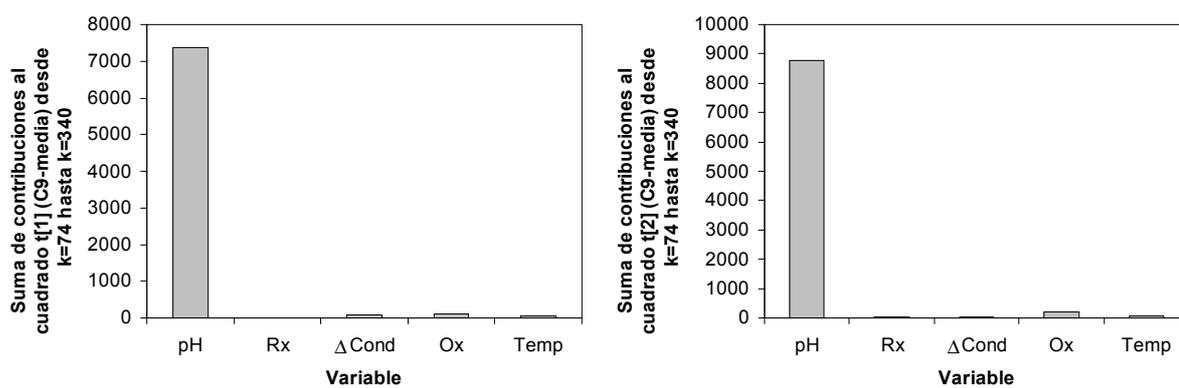


Figura 8.12. Gráficos del sumatorio de las contribuciones al cuadrado de las variables latentes $t[1]$ y $t[2]$ del nivel de observación del enfoque de Wold *et al.* (1998) para el lote C9 entre los instantes de tiempo $k=74$ y $k=340$.

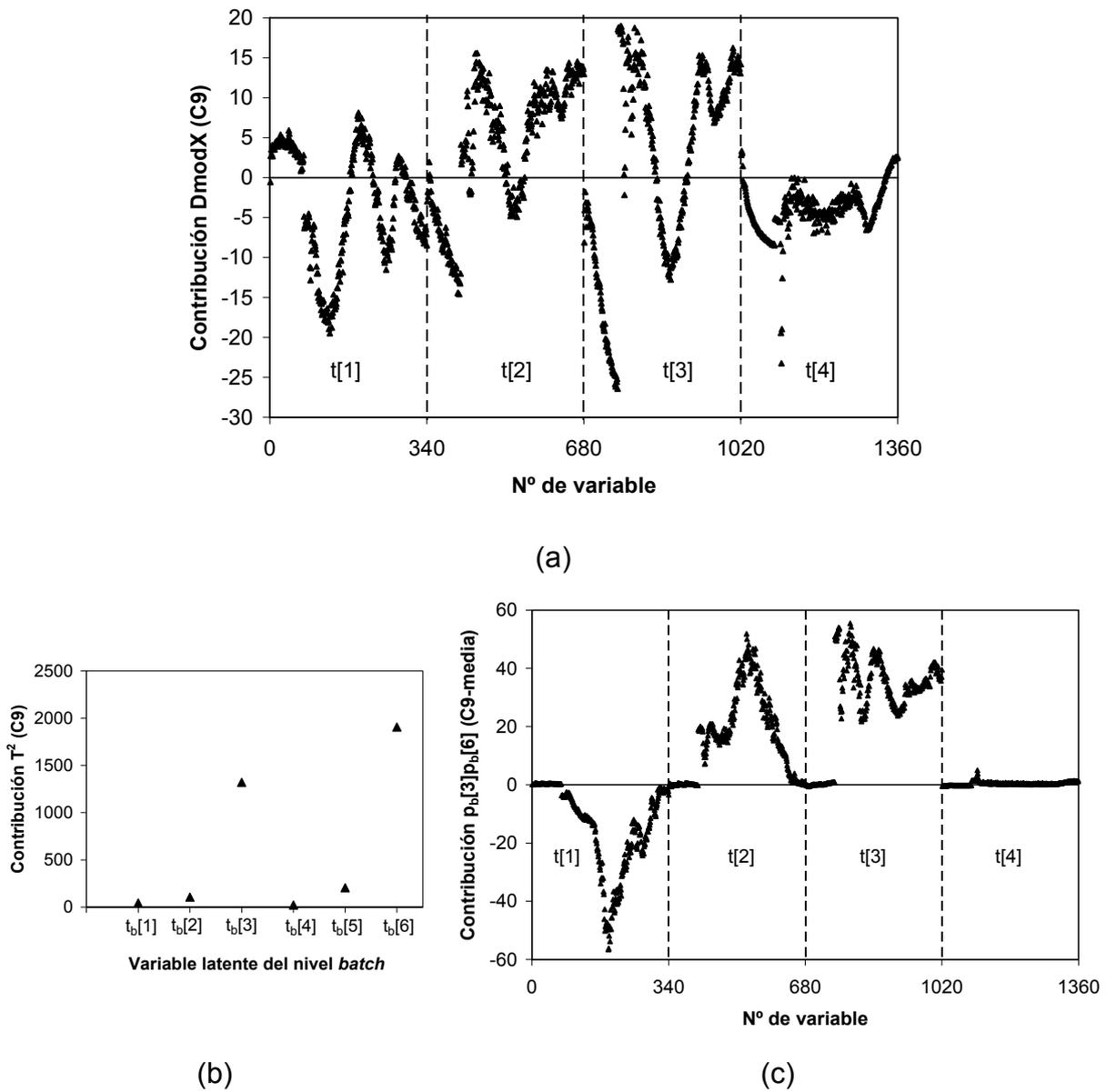


Figura 8.13. Gráficos de contribución del nivel *batch* del enfoque Wold *et al.* (1998) para el lote C9 (a) de la distancia al modelo (b) de la T^2 de Hotelling (c) de las variables latentes del nivel *batch* $t_b[3]$ y $t_b[6]$.

8.3 ANEJO 3: NOMENCLATURA

ANNs	Redes neuronales artificiales
BMSPC	Control estadístico multivariante de procesos por lotes
CUSUM	Sumas acumuladas
DBO	Demanda bioquímica de oxígeno
DQO	Demanda química de oxígeno
EDAR	Estación depuradora de aguas residuales
GAOs	Bacterias acumuladoras de glucógeno
ICA	Instrumentación, control y automatización
MSE	Error cuadrático medio
MSPC	Control estadístico multivariante de procesos
PAOs	Bacterias acumuladoras de polifosfatos
PCA	Análisis de componentes principales
PCR	Regresión en componentes principales
PHA	Poli-hidroxi-alcanoato
PHB	Poli-hidroxi-butirato
PLS	Regresión en mínimos cuadrados parciales
SBR	Reactor discontinuo secuencial
SCR	Suma de cuadrados residual
SCT	Suma de cuadrados total
SPC	Control estadístico de procesos
SPE	Error cuadrático de predicción
TRC	Tiempo de retención celular