
PROYECTO DE TESIS DOCTORAL

“DIAGNÓSTICO BIOQUÍMICO Y TECNOLÓGICO DE UN PROCESO
DE PRODUCCIÓN DE BIOETANOL MEDIANTE APRENDIZAJE
AUTOMÁTICO”

D. Pablo Ruiz Castelló

DIRECTORES

Dr. D. Julio Montes Ponce de León

Dr. D. Miguel Ángel Sanz Bobi

ÍNDICE

<u>1</u>	<u>INTRODUCCIÓN.....</u>	<u>5</u>
<u>2</u>	<u>OBJETIVOS DE LA TESIS.</u>	<u>8</u>
<u>3</u>	<u>ANTECEDENTES Y ESTADO DEL ARTE.....</u>	<u>10</u>
3.1	DETECCIÓN DE ANOMALÍAS.....	11
3.1.1	MODELADO CUANTITATIVO.	11
3.1.2	MODELADO CUALITATIVO.	12
3.1.3	MODELOS DE APRENDIZAJE AUTOMÁTICO.....	13
3.2	INFERENCIA Y OBTENCIÓN DE DIAGNÓSTICO.....	15
<u>4</u>	<u>METODOLOGÍA.</u>	<u>17</u>
<u>5</u>	<u>RECURSOS Y PLAN DE TRABAJO.....</u>	<u>21</u>
	<u>BIBLIOGRAFÍA.....</u>	<u>25</u>

-Esta página se ha dejado en blanco intencionadamente-

1 INTRODUCCIÓN.

La presente propuesta de tesis nace como fruto del trabajo que se viene realizando en colaboración con la empresa Abengoa para el desarrollo de herramientas de asistencia al control de un proceso de producción de bioetanol. En la planta de Bioetanol Galicia de Abengoa el bioetanol se obtiene mediante la fermentación de cereales por lotes en cuatro tanques y una posterior destilación (Infopower 2002).

Este proceso de fermentación consiste en realidad en una *sacarificación* previa donde se transforma el almidón contenido en el cereal en azúcares, seguida de la *fermentación* propiamente dicha, donde las levaduras (*Saccharomyces cerevisiae*) transforman los azúcares obtenidos en etanol (obteniendo dióxido de carbono como subproducto). Químicamente hablando, se puede expresar con la siguiente ecuación:



La naturaleza bioquímica de la fermentación complica su control como proceso industrial. Dado su interés económico, el comportamiento de las levaduras está muy estudiado a nivel de laboratorio (Jacques, Lyons et al. 2003), pero no se conoce perfectamente la operativa óptima para su desarrollo en la escala industrial. En esta escala hay parámetros críticos para las levaduras (disponibilidad de agua en el medio para cada levadura, concentración de compuestos y de otros microorganismos, distribución de temperaturas, etc.) que no se pueden ni caracterizar, ni controlar adecuadamente. Por ello se emplean directrices experimentales y el conocimiento operativo acumulado por el equipo técnico en cada planta para diseñar los procedimientos de operación. Así se diseñan limpiezas, dosificaciones, cambios de operación, etc. que la experiencia valida como suficientes (aunque quizás no óptimos) para controlar problemas como la contaminación bacteriana o la actividad reducida de las levaduras.

El proceso de fermentación llevado a cabo en la planta de Bioetanol Galicia, resumidamente, sigue el siguiente esquema:

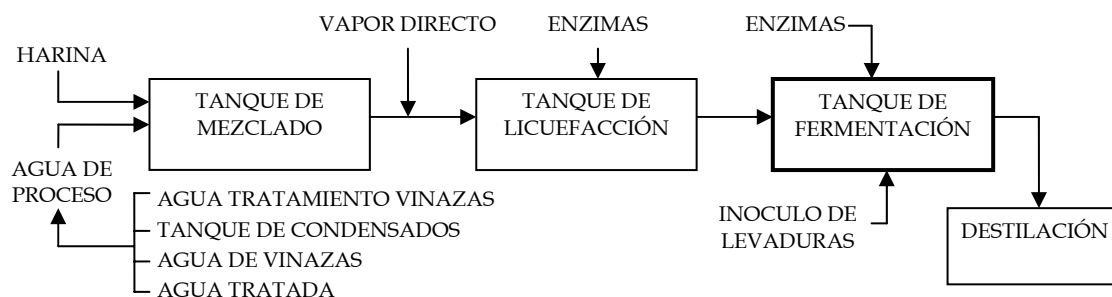


Figura 1 Proceso de fermentación.

A la harina obtenida con la molienda del grano de cereal se le añade “agua de proceso” (agua tratada y la recirculación de otros flujos originados en distintos puntos del proceso) en el tanque de mezcla. Su salida se calienta con vapor para favorecer el fraccionamiento del almidón presente en la harina cereal. El mosto así obtenido se atempera antes de alimentarlo al tanque de licuefacción, donde se le añaden enzimas que transforman el almidón en azúcares (sacarificación). El tanque de licuefacción alimenta al de fermentación, donde se añaden más enzimas y un inóculo de levaduras. Las enzimas (glucoamilasa) continúan el proceso de fraccionamiento de las cadenas de azúcares largos (dextrinas, triosas y maltosa) hasta obtener glucosa que por acción de las levaduras, se transforma en etanol. Tras un tiempo medio de 50 horas se obtiene un mosto con un 11% de concentración de etanol que se envía a destilación.

El control de este proceso afronta dos dificultades principales:

- Por un lado gestionar dos ámbitos distintos: variables físicas de proceso habituales en el entorno industrial y la faceta biológica del mismo, cuya interrelación no es perfectamente conocida.
- Por otro lado, la escasez de conocimiento sobre la propia biología de la fermentación a escala industrial, que de hecho es tratada como una “caja negra”.

Los dos ámbitos del proceso son gestionados por dos departamentos en la planta:

- Producción. El departamento de producción se encarga de la supervisión de las variables de proceso tales como flujos, presiones, temperaturas, concentración de algunas especies químicas, rediseño de procedimientos, economía de proceso, etc.
- Análisis Químicos. La caracterización y seguimiento bioquímico del proceso de fermentación se realiza mediante la toma de muestras y análisis de laboratorio durante la maduración de los distintos lotes que se van produciendo. Los resultados de estos análisis se emplean como índices de la salud biológica con la que se desarrolla la producción.

Ambos departamentos supervisan los resultados de la operación diaria de la planta para detectar potenciales anomalías y establecer las acciones preventivas o correctivas necesarias. No obstante, aparecen anomalías que pueden echar a perder uno o varios lotes de producción. Hay situaciones que no se aprecian en las señales de la ingeniería de proceso (baja salud y actividad de las levaduras, disminución de la población de las mismas, producción de ácidos por contaminación bacteriana, etc.), y que se pueden originar en el periodo entre análisis (que suelen realizarse cada 12 horas para el caso de la fermentación) por lo que podrían quedar sin detectarse hasta la siguiente toma de muestras. En estos casos la eficacia de las acciones correctivas a tomar (adición de más levaduras, nutrientes, antibióticos, etc.) puede resultar reducida, y, en casos extremos, obligar incluso a la limpieza del fermentador y a desechar la producción del lote. Como dificultad añadida, la inercia del proceso hace que sea sólo durante las 25 primeras horas cuando se considera que puede resultar útil la adopción de acciones correctivas.

Además de la complejidad introducida por las carencias de conocimiento sobre la interrelación de los dos ámbitos implicados en la producción, biológicamente hay facetas del comportamiento de las levaduras que se conocen estudiadas de forma aislada a nivel laboratorio, pero que en tanques de miles de litros se simultanean resultando en una interacción mucho más compleja y desconocida. Por ejemplo, de los estudios de laboratorio se sabe que es necesario mantener un cierto equilibrio cinético entre las reacciones de producción de alimento

(glucosa) y su consumo por las levaduras. La alteración de dicho equilibrio puede provocar la obtención de compuestos no deseados o el progreso de poblaciones bacterianas contaminantes del fermento. Como se desconoce cómo pueden influir sobre este equilibrio durante la operación de la planta todas las variables implicadas, simplemente se supervisa que los productos que se van originando estén en unos niveles determinados según el tiempo transcurrido. Además no es posible producir y controlar precisamente unas condiciones del fermento óptimas y homogéneas en todo el tanque. Finalmente hay algunas variables exógenas cuya variación puede tener una gran influencia no establecida sobre el transcurso del mismo. Por ejemplo la procedencia del grano de cereal, su especie y su calidad o el funcionamiento de enzimas, levaduras y antibióticos (cuya formulación el fabricante modifica sin previo aviso), etc. Todos estos motivos producen una variabilidad del rendimiento del proceso, aunque se opere con idénticas directrices y materias primas. Resulta especialmente interesante conocer, explicar y controlar esta variabilidad.

2 OBJETIVOS DE LA TESIS.

Para tratar de explicar y controlar la variabilidad del rendimiento del proceso de fermentación la tesis aborda el desarrollo de una metodología con dos *objetivos fundamentales*:

- Obtener conocimiento sobre el proceso de fermentación a escala industrial que permita esclarecer la influencia de diversas variables sobre el rendimiento final del proceso (considerado como el etanol obtenido por unidad de masa de almidón alimentada).
- Permitir anticipar la aparición de anomalías, aumentando la rapidez de adopción de medidas correctivas, maximizando así su eficacia.

En concreto, se pretende que la metodología a desarrollar permita, funcionando en tiempo real durante la operación de la planta:

- Explotar y extraer conocimiento de la relación no explícita existente entre los dominios de ingeniería de planta y los análisis de laboratorio. Estudiar las circunstancias y características de esta relación y sus posibles aplicaciones al diagnóstico del proceso, a pesar de las distintas constates de tiempo de los dos dominios. Analizar las características formales de la relación entre los dos ámbitos, fundamentando los desarrollos prácticos que se pudieran conseguir.
- Aumentar el conocimiento existente sobre el equilibrio de reacciones simultáneas sacarificación/fermentación y la interacción entre todas las especies implicadas.
- Establecer fórmulas de discriminación de los efectos de manipulación de las variables de operación sobre el desarrollo de todas las reacciones que tienen lugar.
- Predecir la evolución de un lote de fermentación, especialmente aquellas anómalas y, si es el caso, proponer acciones correctivas.
- Discriminar las anomalías de posibles cambios provocados por variaciones en las variables exógenas al proceso (cambio de proveedor de levaduras, uso de distintas enzimas, variación de la procedencia de la materia prima, etc.)

En resumen, se pretende conseguir una herramienta capaz de asistir a los encargados de operación de la planta para **optimizar la operación del proceso maximizando así los rendimientos que se obtengan.**

Para conseguir estos objetivos se partirá del conocimiento parcial de los expertos sobre el proceso, de los registros históricos de resultados de análisis y de los datos de operación disponibles.

En un contexto en el que el impulso de la producción de biocombustibles es un imperativo legal, un sistema así resultaría una importante contribución al mejorar la disponibilidad del proceso, aumentar su fiabilidad y productividad y por ende, la economía del mismo. Así mismo mejoraría la velocidad de adaptación de la operación a una nueva materia prima, como podría

ocurrir por ejemplo, por causa de variaciones del mercado cerealista que puedan ser introducidas por una nueva Política Agraria Comunitaria.

Una reducción de costes y/o aumento de flexibilidad en materia prima del proceso pueden resultar claves para la consolidación de lo que es una incipiente opción para la modificación de los hábitos de consumo de combustibles fósiles en el sector de la automoción. Por otro lado, el incremento de conocimiento sobre el proceso empleado para el diseño de nuevas plantas podría abrir nuevas opciones tecnológicas y de valorización energética de biomasa, de especial importancia en el marco del protocolo de Kyoto y la comercialización de derechos de emisión de los gases de efecto invernadero.

3 ANTECEDENTES Y ESTADO DEL ARTE.

La metodología a desarrollar se puede encuadrar dentro de la disciplina de *supervisión de sistemas*. La industria es un entorno competitivo que exige la máxima productividad. Además se inscribe en un contexto social en el que cada vez se valoran más las pautas de consumo y desarrollo sostenible, y por ende la mayor eficiencia en el uso racional de los recursos. Así los clásicos sistemas de regulación y control de operación dedicados a conseguir un funcionamiento (seguimiento de consigna) determinado, ya no resultan suficientes. Además de conseguir este funcionamiento, se hace necesario garantizar que éste se desarrolle en condiciones óptimas para los equipos. Entre la clásica *planificación de operación* y el *control de sistemas*, median, cada vez más, disciplinas englobadas como *supervisión de sistemas*. (Isermann 1997; Patton 1997; Rinner 2002). En resumen, el principal objetivo de la supervisión es evitar la aparición de fallos inesperados en el sistema maximizando su fiabilidad y disponibilidad. Para ello a lo largo del tiempo se han ido empleando diversas técnicas: *mantenimiento* (correctivo y preventivo), *análisis de fiabilidad*, *ensayos* y más recientemente, *sistemas de seguimiento y/o diagnóstico*.

El diagnóstico es un proceso de razonamiento para el cual se pueden emplear distintas estrategias. Éstas se caracterizan por el *método de razonamiento*, la *dirección de la inferencia* (entendida como el proceso de obtención de conclusiones) y el *fundamento del conocimiento*. El método de razonamiento (o utilización del conocimiento existente) que se emplee debe permitir discriminar el funcionamiento normal del anómalo y, en caso de producirse una anomalía, ser capaz de inferir sobre ella. Esta inferencia u obtención de conclusión puede realizarse por *inducción* (obtención del diagnóstico a partir de los síntomas) o bien por *deducción* (obtención de los síntomas causantes a partir de un diagnóstico). Finalmente, según el conocimiento sobre el sistema proceda de sus ecuaciones de diseño o de su comportamiento visible, se habla de *conocimiento profundo* o *conocimiento superficial*. Una vez definida una estrategia de diagnóstico, éste se suele materializar por medio de un Sistema Basado en Conocimiento, cuya arquitectura general muestra la siguiente figura:

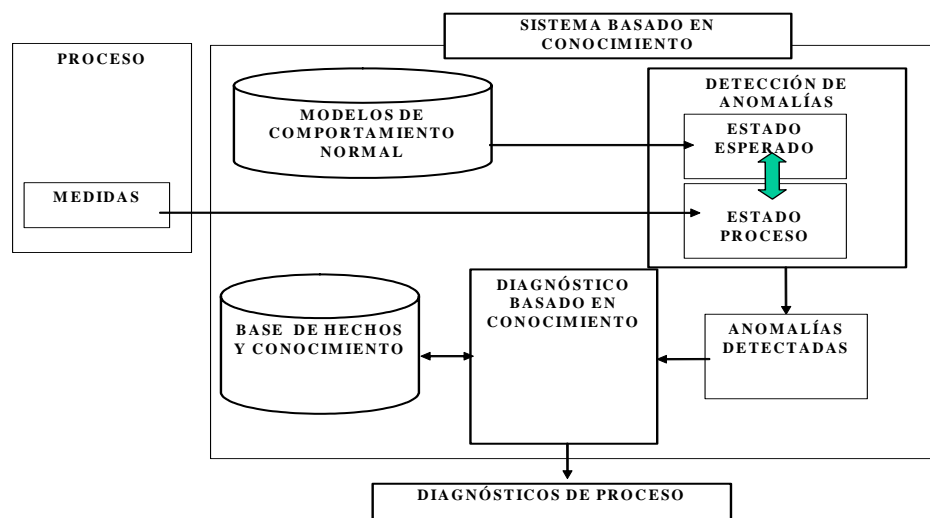


Figura 2 Sistema Basado en Conocimiento.

Cuando el estado del proceso, ofrecido por las variables medidas en el mismo, no coincide con el estado esperado, estimado con modelos de comportamiento normal, esto suele ser síntoma de una anomalía, que se puede o no investigar por un módulo de diagnóstico.

A continuación se revisan los enfoques existentes para las dos principales tareas a realizar en el diagnóstico del sistema: la detección de anomalías y el diagnóstico de las mismas.

3.1 DETECCIÓN DE ANOMALÍAS.

Una anomalía se define (Isermann and Balle 1997) como *una desviación no permitida de la condición aceptable, usual o estándar de al menos una característica o parámetro del sistema*. El rango aceptable de oscilación de una variable se establece mediante *modelos de comportamiento normal*. Según la naturaleza del conocimiento sobre el sistema que se emplee para realizar estos modelos se habla de *modelado cuantitativo*, *modelado cualitativo* y *modelos de aprendizaje automático*, aunque en ocasiones (Patton and Frank 2000), se distingue meramente entre *modelado cuantitativo* y *modelado basado en técnicas de inteligencia artificial*.

3.1.1 Modelado cuantitativo.

Las técnicas de *modelado cuantitativo* describen el sistema mediante ecuaciones obtenidas desde el conocimiento del funcionamiento y la física del mismo (Frank 1990; Venkatasubramanian, Rengaswamy et al. 2002; Isermann 2004). Estas ecuaciones se formulan de la siguiente forma genérica:

$$\begin{cases} x(t+1) = A \cdot x(t) + B \cdot u(t) \\ y(t) = C \cdot x(t) + D \cdot u(t) \end{cases} \quad (\text{Ec. 2})$$

Donde A , B , C y D son matrices de parámetros, $x(t)$ son valores medidos de variables del sistema, $u(t)$ son las variables de entrada al mismo e $y(t)$ es el vector de salidas. Según el enfoque de detección que se emplee, las anteriores ecuaciones se describen y utilizan de distintas formas (Frank 1990; Patton 2000; Isermann 2004). Así se habla de detección de anomalías mediante *estimación de parámetros*, detección con *observadores*, con *ecuaciones de paridad* o empleando *modelos de señales*. La estimación de parámetros puede permitir discriminar anomalías como el aumento de fricción entre componentes, el grado de desgaste de los mismos, variaciones de superficies de contacto, grados de aperturas de válvulas, etc. En general estos parámetros no suelen ser conocidos pero se pueden evaluar midiendo las entradas $x(t)$ y salidas $y(t)$ del sistema. Por el contrario, si estos parámetros son conocidos, es posible emplearlos para estimar, o bien las entradas midiendo las salidas (*detección con observadores de estado*), o bien las salidas midiendo las entradas (*detección con observadores de salida*). Otra opción es, en caso de conocerse los parámetros y poder medir entradas y salidas, calibrar el modelo completo del sistema (obtener las ecuaciones de paridad) y durante su funcionamiento comparar las salidas que este estime con las que realmente se producen. La aparición de una anomalía se detecta analizando la diferencia o *residuo* entre salida calculada y real. Finalmente, si se considera el carácter de señal continua de la salida $y(t)$ se puede tratar de modelar ésta en sí misma,

centrándose en sus características armónicas, estocásticas o en ambas (aplicación típica es el modelado de vibraciones en elementos mecánicos).

El uso de estos tipos de modelos está muy extendido en ciertos campos (detección de anomalías en motores eléctricos (Nandi, Toliyat et al. 2005), en motores de combustión interna (Schmidt, Kimmich et al. 2000),(Willimowski, Kimmich et al. 2000) etc). La necesidad de conocer la física subyacente del proceso para su utilización hace pensar que la aplicabilidad de estos enfoques para el problema de la supervisión de la fermentación industrial radicarán principalmente en el ámbito de los sistemas monitorizados mediante el sistema de Información de Planta. Alguno de estos enfoques podrá ser empleado para la detección por ejemplo, de pérdida de rendimiento en los intercambiadores de calor, problemas por variación de densidad de fluidos en los equipos de bombeo, etc. En (Dahhou, Roux et al. 2006) se puede encontrar un modelado de un proceso de fermentación a escala de laboratorio basado en las siguientes ecuaciones:

$$\left\{ \begin{array}{l} \frac{dC(t)}{dt} = \mu(t) \cdot C(t) - D(t) \cdot C(t) \\ \frac{dS(t)}{dt} = -\frac{1}{Y_{C/S}} \mu(t) \cdot C(t) + D(t) \cdot S_m(t) - D(t) S(t) \\ \frac{dP(t)}{dt} = -\frac{Y_{P/S}}{Y_{C/S}} \mu(t) \cdot C(t) - D(t) P(t) \end{array} \right. \quad (\text{Ec. 3})$$

Así el proceso de fermentación se caracteriza como una degradación biológica $S(t)$ del sustrato (glucosa) realizada por una población $C(t)$ de microorganismos (levaduras que experimentan una tasa de crecimiento $\mu(t)$) que provoca el aumento de la concentración de los metabolitos producto $P(t)$, sujeta a una tasa de dilución $D(t)$ y condicionada por los coeficientes de rendimiento Y . En (Dahhou, Roux et al. 2006) se parte de estas ecuaciones para plantear un metodología de detección de anomalías mediante observadores de estado para la tasa de crecimiento de las levaduras. Los observadores se calibran con medidas en continuo tomadas en un reactor de laboratorio. En este trabajo se relacionan, también a escala laboratorio, las dificultades (fuertes no linealidades, complejidad de la física subyacente del proceso, etc.) existentes para la supervisión de un proceso de estas características.

Dadas las comentadas diferencias entre la escala de laboratorio y la industrial, así como la distinta disponibilidad y naturaleza de la información existente en la planta de Bioetanol Galicia, se piensa que el modelado cuantitativo no será la vía de trabajo con mayor potencial para abordar el problema planteado.

3.1.2 Modelado cualitativo.

Cuando la complejidad de un sistema complica o imposibilita la formulación de las ecuaciones cuantitativas que lo rigen, se puede optar por su *modelado cualitativo* ((Kleer 1979; Kuipers 1984; Kleer 1990; Kuipers 1994). Así, si bien no se recoge exacta y precisamente el comportamiento del sistema, sí es posible obtener una valiosa información sobre su estado y funcionamiento. Los distintos tipos de modelos cualitativos se pueden agrupar (Venkatasubramanian, Rengaswamy et al. 2002) en *modelos causales* y *modelos de componentes*.

Se entienden por *modelos causales* aquellos cuyo objetivo es tratar de recoger la lógica causa-efecto subyacente en los sistemas, emulando la forma de razonamiento cualitativo sobre el entorno físico de los humanos. Así usando *modelos físicos cualitativos* (Kuipers 1993) se formulan las posibles ecuaciones diferenciales en términos de los sentidos de las influencias (derivadas) entre sus variables. Los *grafos de influencia* son representaciones gráficas de estas influencias. También han experimentado un gran desarrollo (Watson 1961; Lee, Grosh et al. 1985) los *árboles de fallo*, que fueron originalmente concebidos (Watson 1961) en los laboratorios Bell para análisis de fiabilidad. La construcción de un árbol requiere definir eventos cuya posibilidad de ocurrencia se quiere evaluar y representar en forma arbórea las posibles secuencias de sub-eventos desencadenantes del evento raíz. Requieren un gran conocimiento previo sobre el comportamiento del sistema.

La otra vía de modelado cualitativo es el *modelado de componentes*, mediante el cual se trata de inferir sobre el comportamiento del sistema partiendo del análisis de sus partes integrantes que se organizan en *jerarquías de abstracción* (de forma análoga a como los circuitos eléctricos se descomponen en resistencias, fuentes, condensadores, etc.). Las jerarquías se pueden establecer sobre la *estructura* y/o sobre la *funcionalidad* entre los componentes.

Aunque el modelado cualitativo se ha consolidado como un enfoque eficiente en numerosas aplicaciones, trabajos más recientes (Patton 2000) exploran las posibles sinergias de incorporar además información cuantitativa mediante el uso de técnicas del paradigma inteligencia artificial, obteniendo, por ejemplo, observadores cualitativos borrosos.

En la tesis se considera potencialmente útil el enfoque propio de este grupo de técnicas. Aunque para modelar el comportamiento normal de los equipos habituales en la ingeniería de proceso parecen más adecuados modelos funcionales cuantitativos, existe cierto conocimiento vago sobre el comportamiento de la fermentación y, en especial, sobre algunas relaciones entre los sistemas de ingeniería y ésta, que podría resultar útil recoger con alguno de los enfoques comentados. En concreto se piensa que podrían resultar útiles los comentados diseños híbridos con técnicas del paradigma de inteligencia artificial.

3.1.3 Modelos de aprendizaje automático.

Cuando no se dispone de conocimiento explícito (cualitativo o cuantitativo) con el que modelar el funcionamiento del sistema o la relación entre alguna de sus variables significativas, la principal alternativa es tratar de capturar el conocimiento implícito sobre las relaciones normales durante el funcionamiento que queda recogido en las bases de datos de registros históricos de operación. Para ello existen diversas técnicas que se pueden agrupar en *métodos estadísticos* y *técnicas de inteligencia artificial*.

Los distintos métodos de naturaleza estadística conforman el *análisis multivariante*, (Rencher 2002), empleado para la extracción de distintas formas de conocimiento implícitas en medidas simultáneas de distintas variables. Así para el análisis y modelado de las dependencias entre variables se emplean los *análisis de varianza* simple y múltiple (ANOVA y MANOVA) y los modelos de *regresión*, simple, múltiple y/o *multivariante*. Para resumir la información disponible y analizar en perspectiva las potenciales interdependencias existentes se emplean (Johnson and

Wichern 1992) el *análisis de componentes principales ACP*, el *análisis factorial*, el *análisis de correspondencias* y el *análisis de correlación canónica* por citar algunos. Recientemente, en especial en el ámbito de las ciencias humanas, ha experimentado un gran auge el *método de ecuaciones estructurales (SEM)* (Bollen 1989) como vía para extraer las *variables latentes* no recogidas directamente en las muestras disponibles, pero subyacente como función de ellas. Generalmente estos métodos requieren disponer de un conjunto amplio de datos equispaciados temporalmente y con una frecuencia de muestreo elevada. Esto podría impedir su utilización para la obtención de modelos de comportamiento biológico normal de la fermentación, de la que se obtienen datos a intervalos aproximadamente de 12 horas distribuidos irregularmente y en un número típicamente entre 4 y 5 por lote de fermentación. En ellos, además, el tiempo es una variable determinante que no siempre manifiesta su influencia de la misma forma en todas las observaciones. No obstante se consideran que, especialmente los métodos de resumen y extracción de variables latentes, podrán resultar útiles para orientar y como medios para completar el escaso conocimiento que se posee sobre la relación subyacente entre las variables muestreadas mediante análisis de laboratorio y entre éstas y las registradas por el sistema de información de planta.

Bajo el paraguas de Inteligencia Artificial se pueden distinguir análogamente técnicas dedicadas a la *clasificación y resumen* de información y las dedicadas a la *captura de la relación* existente entre variables. En ambas vertientes, los modelos y algoritmos desarrollados parten de cierta inspiración en modelos del razonamiento o aprendizaje humano, de ahí su apelativo. Orientadas a la clasificación del tipo de datos destacan las técnicas para obtención de *conglomerados (K-means y los Mapas Autoorganizados de Kohonen (Kohonen 1990))* y los *árboles de decisión* (Breiman, Friedman et al. 1984). En el modelado de relaciones funcionales destaca el uso de *redes neuronales (Roque 1996; Haykin 1998)* y, en menor medida, los *árboles de regresión*. Para hacer más interpretables los modelos de redes neuronales e incorporar el tipo de incertidumbre propio del conocimiento y lenguaje humanos, se ha consolidado desde hace tiempo la formulación de estos modelos en términos de *lógica borrosa (Zadeh 1964; Lin and Lee 1996; Roque 1996; Nauck, Klawonn et al. 1997)*.

Dadas las carencias de conocimiento explícito sobre la bioquímica de la fermentación cerealista industrial, la posibilidad de extraer y utilizar conocimiento incompleto, incluso vagamente expresado en lenguaje natural, que ofrecen las técnicas de inteligencia artificial hace que su uso sea uno de los principales ámbitos de potencial exploración para el modelado y diseño de métodos de obtención de conocimiento en el trabajo planteado. Además, las sinergias que aportan los usos híbridos de estas técnicas resultan una línea de máximo interés a explorar para la consecución de los objetivos de la tesis. La inteligencia artificial es una opción consolidada para la asistencia a la supervisión de procesos en la industria química y/o con influencia de factores biológicos (Venkatasubramanian, Rengaswamy et al. 2002). Los usos bioquímicos han dado lugar a la celebración de la serie de conferencias trianuales "Conference on Computer Applications in Biotechnology" organizadas por la IFAC (International Federation of Automatic Control) y la EFB (European Federation of Biotechnology). Grieu ha realizado diversos trabajos (Grieu, Thiery et al. 2005; Grieu, Traore et al. 2005) en los que se usan métodos de inteligencia artificial para el estudio y control de un proceso de neutralización de lodos de aguas residuales. En estos trabajos se describe la eficacia del uso de mapas de Kohonen para aumentar el conocimiento no explícito que se tiene de un proceso biológico, para, a partir de él, conseguir modelos eficaces de predicción de parámetros que no son fácilmente medibles en tiempo real durante la operación de la planta. Por otro lado, en (Nakkabi, Doncescu et al. 2002) se muestra

cómo el uso de lógica borrosa y técnicas de clustering asisten en su control a los expertos en un proceso de fermentación. La solidez del enfoque mostrada en los campos comentados hace que sea el principal entorno en el que se tratará de diseñar una metodología óptima para las peculiaridades del proceso de fermentación cerealista.

3.2 INFERENCIA Y OBTENCIÓN DE DIAGNÓSTICO.

Si se detecta una anomalía a veces se requiere una investigación de sus causas o diagnóstico, entendido como (Isermann and Balle 1997) *la determinación del tipo, tamaño, localización y tiempo de la anomalía*. Ésta se realiza mediante un proceso de *inferencia* o *razonamiento*. Los posibles métodos de inferencia a emplear lógicamente están muy condicionados el método de formulación de los modelos de comportamiento normal y el conocimiento sobre el proceso.

Entre las diversas posibilidades de codificar el conocimiento de expertos (Gómez, Juristo et al. 1997), se pueden distinguir cuatro grupos: *representaciones con lógica de predicados, esquemas de representación procedural, esquemas de representación en red o gráfica y los esquemas de representación estructurada*.

Dentro de las representaciones con lógica de predicados se distinguen la clásica representación con *lógica de enunciados* (proposiciones con sentido completo que toman el valor de verdadero o cierta) y la *lógica de predicados* (en la que se extienden las posibilidades de los enunciados mediante el uso de cuantificadores).

Los esquemas de *representación procedural* son la forma más común de explotación de la lógica formal. Son las clásicas reglas en la forma *<si condición 1,..., condición n entonces conclusión1 ,...,conclusión n>*

Los enfoques gráficos de representación del conocimiento abarcan representaciones como las *redes semánticas*, los *grafos conceptuales* y las *redes de Petri*. Todas tienen en común emplear *nodos* que representan conceptos y *arcos* que simbolizan las relaciones entre los nodos.

Finalmente entre las representaciones estructuradas se pueden mencionar los *frames* (que asocian una estructura de datos complejas en los nodos de las redes semánticas), y los *scripts*, que son estructuras de datos empleadas para representar secuencias tipo de sucesos.

Asociadas a cada esquema de representación del conocimiento se han desarrollado unas técnicas de inferencia para la obtención de diagnósticos. Para inferir usando conocimiento formulado en lógica de predicados se utilizan las *reglas de transformación*; las *cláusulas de Horn* se pueden emplear con el conocimiento expresado con lógica de predicados o representaciones procedurales. El conocimiento contenido en redes semánticas se puede explotar o bien mediante el *método de activación por propagación* o bien mediante el *método de asociación de estructuras*.

Generalmente en los sistemas expertos el conocimiento se codifica con representaciones procedurales y la inferencia se realiza en tres pasos en el *motor de inferencia*: selección de reglas a examinar según las circunstancias en la base de hechos, obtención del conjunto de reglas aplicables, formación del conjunto conflictivo y selección de regla a disparar.

Otra posibilidad de explotación del conocimiento es el *razonamiento basado en casos*. Parte del uso de casuística de situaciones (en forma semejante a *scripts*) recogidas en una base de datos y

sigue los pasos típicos de un motor de inferencias. Su particularidad reside en que permite incorporar de forma automática nuevo conocimiento en forma de casos nuevos que ocurran en el sistema.

Además de la cantidad, calidad y naturaleza de un conocimiento, su *incertidumbre* asociada también condiciona su utilidad. Por ello puede resultar útil incorporar esta incertidumbre al proceso de inferencia. Para ello se suelen emplear *la teoría de probabilidad*, el uso de *factores de certeza*, la teoría de la *evidencia* o más recientemente *la teoría de la posibilidad y la lógica borrosa*.

Las técnicas de razonamiento comentadas parten de la disponibilidad de un conocimiento que, según sus características y formulación, explotan de formas diversas. Para el caso que pretende abordar la tesis, al ser el conocimiento disponible bastante incompleto, se deberá diseñar una técnica de inferencia ad hoc útil para el uso del conocimiento que la propia metodología de extracción automática pueda conseguir. Se habrá de ser capaz de supervisar la creación automática de una base de conocimiento y de establecer criterios automáticos para su mantenimiento (de una forma similar a la propuesta en (Juez 2005)). También se tendrá que conseguir la explotación eficiente de los dos ámbitos (ingeniería y/o tecnología y biología) que componen el dominio objeto de estudio.

4 METODOLOGÍA.

La tesis propuesta afronta el diseño de una metodología de supervisión y diagnóstico a través del *modelado de comportamiento normal, extracción y uso de conocimiento* sobre un dominio con unas características no encontradas de forma simultánea en la literatura. El dominio objeto de estudio será el proceso de producción de bioetanol desde la entrada y procesado de materia prima hasta la salida de su etapa de fermentación. En concreto, desde el punto de vista teórico, confluyen las siguientes dificultades:

- Conocimiento incompleto (especialmente su faceta biológica).
- Existencia de distintas fuentes (y semánticas) de información sobre el mismo.
- Gran sensibilidad a la influencia de variables exógenas.
- Distinta frecuencia de muestreo en las distintas fuentes de información.
- Toma de muestras irregular en algunas de las fuentes de información disponibles.

Las características del problema perfilan las características que habrá de aunar la metodología a desarrollar:

- Al ser el conocimiento de los expertos sobre parte del dominio incompleto (o incluso inexistente para su faceta biológica), se deberá conseguir un *método de extracción automática de conocimiento* cuyos resultados sean validados por los propios expertos. Actualmente como modelo de comportamiento normal se emplea una comprobación no explícita de los niveles de las variables más significativas a lo largo del desarrollo temporal del proceso.
- Las fuentes de información sobre el dominio son el conocimiento incompleto y no explícito de los expertos, las señales de ingeniería del Sistema de Información de Planta y los resultados de análisis de laboratorio. Tanto el sistema de detección de anomalías, como los mecanismos de gestión del conocimiento habrán de ser capaces de explotar estas fuentes diversas, expresándolas en una semántica común sobre el dominio que permita elevar un diagnóstico lo más informativo posible. Así se habrán de diseñar mecanismos para contraste de hipótesis y resolución de conflictos entre los distintos ámbitos, que maximicen la robustez del conocimiento obtenido y los diagnósticos que se emitan.
- Las fuentes de información disponibles tienen distintas constantes de tiempo, por lo que se habrá de estudiar la interrelación dinámica entre las mismas (que podrá ser variable) para poder explotarlas con la máxima eficacia.
- Aunque se desconoce exactamente cuál es, se sabe que cada subproceso puede ejercer una gran influencia sobre el desarrollo de los que le siguen. Por ello para poder realizar diagnósticos precisos y eficaces se debe conseguir conocimiento tanto sobre el desarrollo de cada subproceso de forma local, como sobre la interrelación entre los mismos a escala global.
- La metodología a desarrollar habrá de tener una especial capacidad para la detección, discriminación e incorporación de información y conocimiento referente a influencia de las variables exógenas, siendo capaz de diferenciar eficazmente una anomalía de un cambio en las circunstancias del proceso.

Basándose en trabajos precedentes que analizan el uso de múltiples fuentes de conocimiento (Richardson 2003), se espera que la explotación de las distintas sobre el dominio sea una de las principales vías que permita su mejor comprensión, mimetizando el proceder de los expertos de planta, que durante la operación confrontan el conocimiento de sus departamentos.

Por todo lo comentado, se va a partir de un diseño como el esquematizado en la Figura 3.

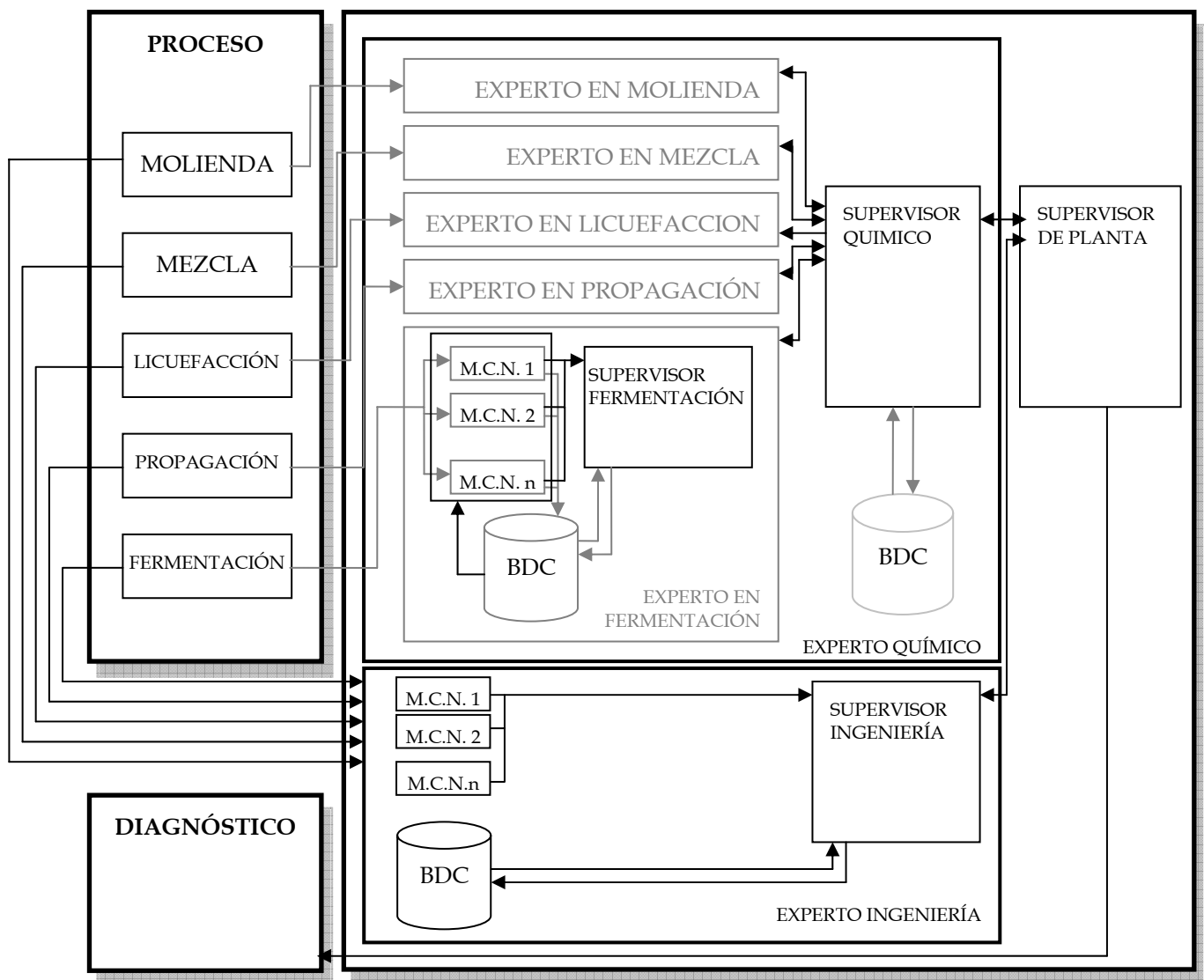


Figura 3 Diseño preliminar.

Las señales del sistema de información de planta (que se muestrean continuamente) y los resultados de análisis de laboratorio (con una cadencia irregular y distinta según el subproceso) se alimentarán respectivamente a módulos expertos en su análisis.

El módulo encargado del análisis de las señales del Sistema de Información de Planta se asemejará a un sistema experto basado en conocimiento tradicional. Se discriminarán las posibles anomalías detectables por observación de las señales de planta (obstrucción de válvulas, deterioro martillos en molinos, problemas de impulsión en bombas,...) mediante distintos *observadores*, *ecuaciones de paridad* o *modelos de señales*. Se emplearán uno u otros según la naturaleza de la anomalía y la disponibilidad de información sobre ella. Caso de detectarse, las anomalías se evaluarán usando el conocimiento que los expertos (operadores de planta y departamento de ingeniería) puedan proporcionar sobre ellas y que será recogido en una *base de conocimiento*. El supervisor de éste módulo se encargará de formular el diagnóstico que se obtenga en términos de una semántica común para la planta y lo remitirá al módulo supervisor del proceso.

En el ámbito de la bioquímica, los resultados de análisis de laboratorio, dado lo específico de la física, biología y química de cada proceso, podrán alimentar según el caso o a diversos sub-expertos o a simples modelos de comportamiento normal. En el caso más complejo, una base de conocimiento contendrá la última versión de los modelos de comportamiento normal disponibles. Recuperados de ella serán alimentados por los últimos resultados de análisis obtenidos. Los resultados de la evaluación de estos modelos y de la comparación de las distintas versiones que de ellos se puedan obtener durante el uso del sistema, serán las fuentes que permitan al módulo experto de subproceso obtener conocimiento. Como ejemplo, se resumen los resultados obtenidos en el diseño del módulo del sub-experto en fermentación:

Se han obtenido y validado distintos enfoques simultáneos de modelado de comportamiento normal. Los resultados de los análisis (cromatografías) practicados a muestras del tanque de fermentación se alimentan al conjunto de últimas versiones de modelos autocalibrados almacenadas en la base de conocimiento.

En primer lugar, a un mapa de Kohonen. Esto permite clasificar el estado de la fermentación dentro de un número discreto de posibles estados fácilmente interpretables. Usando la probabilidad calculada con resultados históricos, se pueden obtener predicciones sobre a qué estado se transitará en la siguiente toma de muestra.

Posteriormente se construye un árbol de clasificación sobre éstos históricos de resultados de los mapas, obteniendo las variables y los valores que caracterizan a cada posible estado de fermentación. Se está trabajando en la interpretación de la variación de los valores y estructuras de los árboles. Se piensa que su variación en el tiempo informa también de la variación de la estructura interna del proceso que, cotejada con el conocimiento que se pueda obtener en otros subprocesos, puede resultar una herramienta adicional para esclarecer la influencia que algunas variables tienen sobre el mismo.

Por otro lado, se obtienen los valores de las dos primeras componentes principales de las cromatografías que resultan caracterizar el estado de madurez (tiempo efectivo de proceso transcurrido) de la fermentación y el grado de desviación (producción de especies no deseadas) de la misma.

Se han desarrollado también modelos de predicción de la evolución de variables registradas en las cromatografías.

Toda la información que este diagnóstico genere será usada por el supervisor de subproceso. Así mediante un esquema de votación y aciertos históricos (Doan, Halevy et al. 2001) se podrá establecer la existencia o no una anomalía en el sub-proceso. Este resultado y su información complementaria se enviarán al supervisor químico.

El supervisor químico recibirá la información de cada sub-experto de proceso y valorará la existencia de una anomalía bioquímica en la planta. Traducirá este diagnóstico a semántica global de la misma y remitirá el resultado al experto de planta. De esta valoración se podrá además obtener como resultado una estimación de necesidad de revisión o no del conocimiento de cada sub-experto.

El experto en ingeniería de la planta procederá de forma análoga mediante un proceso de inferencia y remitirá su resultado, traducido a semántica de planta, al experto de planta.

Finalmente el experto de planta, empleando la información remitida por los expertos químico y de ingeniería, producirá un diagnóstico global del proceso. Así mismo, devolverá información, con posible componente de realimentación humana si procede, sobre la bondad de los diagnósticos de cada experto en cada ámbito, que éstos podría utilizar para la revisión de los criterios con los que operan ellos y los sub expertos que supervisen.

La estructura jerárquica del diagnóstico se piensa que facilitará, al portar mucha información jerarquizada y por ende más interpretable, establecer mecanismos que discriminen anomalías propias del proceso de cambios en las variables exógenas del mismo.

La tesis afrontará las siguientes aportaciones teóricas al campo del diagnóstico industrial:

- Métodos capaces de extraer conocimiento de un dominio no compacto y pobre, como es el proceso de producción de bioetanol.
- Una teoría que permita conciliar fuentes de conocimiento (análisis de laboratorio y datos del sistema de Información de Planta) que hablan de un mismo dominio con distinta semántica.
- Un supervisor de conocimiento que se ha de entrenar a sí mismo. En otros trabajos (Doan, Halevy et al. 2001) ya se han empleado exitosamente supervisores que controlan el funcionamiento de múltiples expertos. Esta supervisión se basa en la existencia de un conocimiento previo cuantificable del que se carece en este caso. Esta experiencia previa permite al supervisor, en una fase inicial de entrenamiento, evaluar la fiabilidad de cada experto y así conciliar los distintos juicios emitidos. Dado que en el problema planteado no existe conocimiento previo ni una valoración cuantitativa de cada fuente, será necesario proveer al propio supervisor de un mecanismo (según criterios de repetición y/o refuerzo por establecer) para la actualización y revisión de sus propios criterios.

5 RECURSOS Y PLAN DE TRABAJO.

En la siguiente tabla se relacionan las principales tareas en las que se subdividirá la realización de la tesis:

TAREA	DESCRIPCIÓN	Duración	Inicio Previsto	Final previsto
1	REVISIÓN BIBLIOGRÁFICA	22.3mss	lun 28/11/05	vie 31/08/07
2	DISEÑO DE METODOLOGÍA	18.85mss	jue 01/12/05	jue 31/05/07
3	DESARROLLO MODELOS DE COMPORTAMIENTO NORMAL	311d	jue 01/12/05	mié 28/02/07
4	MCN análisis químicos	14.55mss	jue 01/12/05	mié 31/01/07
5	Modelado de datos de operación en continuo	7.6mss	mar 01/08/06	mié 28/02/07
6	Otros modelos	2mss	jue 04/01/07	mié 28/02/07
7	DESARROLLO ALGORITMOS DE GESTIÓN DEL CONOCIMIENTO	348d	mié 10/05/06	vie 07/09/07
8	Extracción Conocimiento Análisis Quimic	5mss	mié 10/05/06	mar 26/09/06
9	Extracción Conocimiento Planta	4.3mss	mié 01/11/06	mié 28/02/07
10	Supervisor del conocimiento	8mss	lun 04/12/06	vie 13/07/07
11	Validación del Modelo de conocimiento	2mss	lun 16/07/07	vie 07/09/07
12	PRUEBAS DE CAMPO Y VALIDACIÓN DE DESARROLLOS	19.5mss	lun 13/03/06	vie 07/09/07
13	REDACCIÓN TESIS DOCTORAL	16.3mss	lun 31/07/06	lun 29/10/07

Tabla 1 Tareas y duraciones previstas para la realización de la tesis.

La siguiente figura ilustra el cronograma correspondiente:

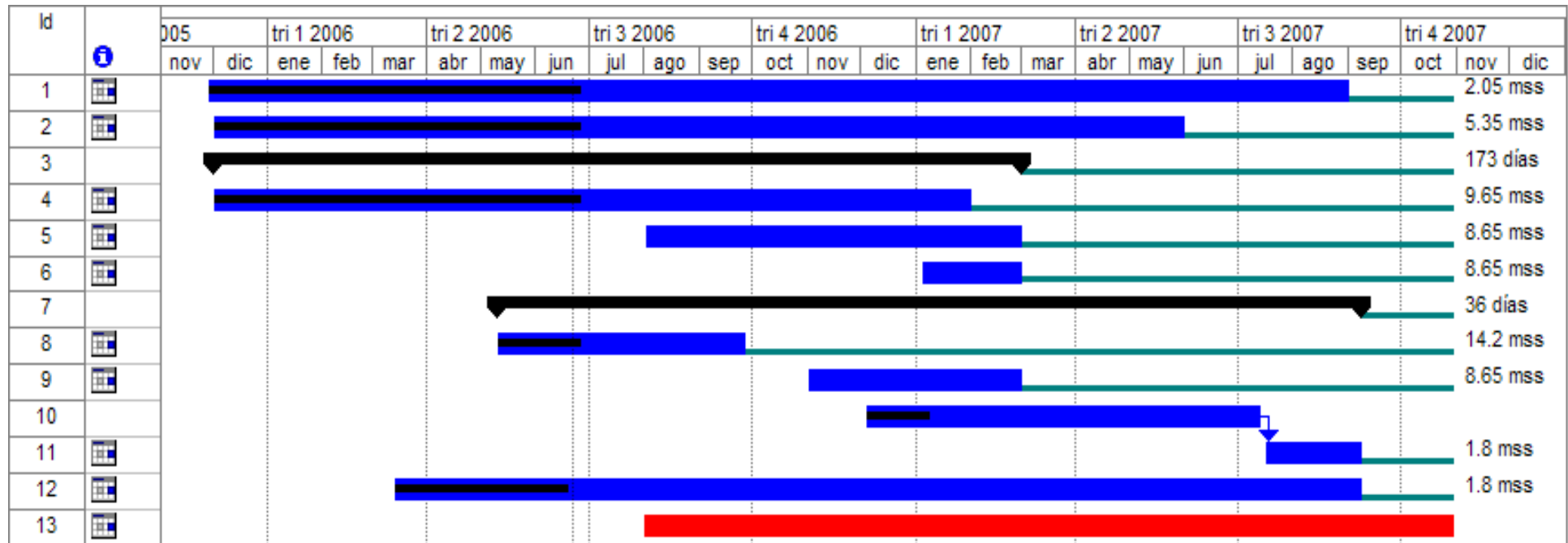


Figura 4 Previsión de cronograma del desarrollo de la tesis.

Como se muestran en el anterior diagrama, el trabajo sobre el tema de tesis dio comienzo hacia diciembre del 2005, partiendo de los resultados obtenidos en la realización de modelos de comportamiento normal de la fermentación para Abengoa. Seguidamente se describen las principales tareas que se considera que se habrán de realizar:

1. REVISIÓN BIBLIOGRÁFICA.

Durante todo el proceso de elaboración de la tesis se mantendrá una dedicación a la revisión del estado del arte con vistas a la posible incorporación de cualquier avance útil al trabajo a realizar y a la orientación del mismo, considerando los campos de modelado y diagnóstico industrial sin desdeñar la investigación, en la medida de lo posible, sobre la faceta biológica del proceso.

2. DEFINICIÓN DE METODOLOGÍA.

Dada la complejidad de la tarea planteada y la carencia de experiencias semejantes, se espera que sea necesario prolongar y revisar el proceso de establecimiento de la metodología durante al menos la fase inicial del mismo.

3. DESARROLLO DE MODELOS DE COMPORTAMIENTO NORMAL.

Se han realizado ya diversas visitas y una estancia semanal en la planta de Bioetanol Galicia y se está trabajando en colaboración con el departamento de análisis químicos, el de ingeniería y la dirección de turno que ha contrastado algunos desarrollos ya completados. Se han suministrado datos de resultados de análisis y se han realizado pruebas preliminares de modelado de su comportamiento normal. Este modelado de comportamiento normal comprenderá los modelos químicos (tarea 4 ya comenzada), los modelos de sistemas con mediciones en continuo (tarea 5) y cualquier otra tipología de modelos que se considere útil en la metodología o a petición de los potenciales usuarios de la misma (tarea 6). Se espera que el proceso de modelado de comportamiento normal concluya hacia febrero del año 2007.

7. DESARROLLO DE ALGORITMOS DE GESTIÓN DEL CONOCIMIENTO.

Esta fase es el núcleo de la tesis candidata. Dado lo específico del diseño a acometer, su avance estará ligado al de los modelos de comportamiento normal y a que se haya cubierto un periodo razonable de profundización en la bibliografía. Esta tarea se subdividirá en el diseño de métodos para la extracción de conocimiento de los dos tipos de fuentes distintas (tareas 8 y 9). Cuando estos algoritmos tengan cierto grado de definición será momento de abordar el desarrollo de la teoría y algoritmia del supervisor de conocimiento que consiga las funcionalidades descritas anteriormente (Tarea 10). Todo este proceso se realimentará de una validación continua sobre los desarrollos propuestos (Tarea 11) que se realiza en continuo mediante el contacto con el personal de planta.

12. PRUEBAS DE CAMPO Y VALIDACIÓN DE DESARROLLOS.

Según se vaya madurando el desarrollo de la teoría para la gestión del conocimiento se valorará su bondad mediante su utilización en el caso real de producción de bioetanol asociado al proyecto AGROBIHOL FASE II, teniendo siempre en cuenta la opinión de los expertos de planta sobre su calidad.

13. REDACCIÓN DE LA TESIS DOCTORAL.

Según se vayan completando fases del desarrollo se pretende ir redactando las partes correspondientes de la tesis, pretendiendo tener finalizada una primera versión para agosto del 2007.

Paralelamente se acometerá la redacción de al menos un artículo técnico. Con el avance conseguido en esta tarea, se piensa que se ha conseguido material suficiente para un artículo sobre los modelos de comportamiento normal de la faceta química de la fermentación ya desarrollados.

En cuanto a los recursos a emplear en la realización de la tesis, se cuenta con los medios físicos e informáticos proporcionados por un entorno como el IIT en el que se desarrollará el trabajo. La financiación para su realización provendrá parcialmente de fondos PROFIT ya concedidos y otros proyectos en los que pudiera participar el doctorando. Así mismo esta tesis ha sido seleccionada para su financiación por la Cátedra Rafael Mariño de Nuevas Tecnologías Energéticas de la Escuela Técnica Superior de Ingeniería de la Universidad Pontificia de Comillas.

BIBLIOGRAFÍA.

- Bollen, K. A. (1989). Structural equations with latent variables. New York.
- Breiman, L., J. H. Friedman, et al. (1984). Classification and Regression Trees. New York.
- Dahhou, B., G. Roux, et al. (2006). Modeling, Software sensors, control and supervision of fermentation processes. LAAS Report. LAAS-CNRS, Laboratoire d'Analyse et d'Architecture des Systèmes du CNRS: 29.
- Doan, A., A. Halevy, et al. (2001). "Reconciling Schemas of Disparate Data Sources: A Machine Learning Approach." Proceedings of the ACM SIGMOD International Conference on Management of Data: 509-520.
- Frank, P. M. (1990). "Fault Diagnosis in Dynamic Systems Using Analytical and Knowledge-based Redundancy A Survey and Some New Results." Automatica **26**(3): 459-474.
- Gómez, A., N. Juristo, et al. (1997). Ingeniería del Conocimiento.
- Grieu, S., F. Thiery, et al. (2005). "KSOM and MLP neural networks for on-line estimating the efficiency of an activated sludge process." Chemical Engineering **116**(1): 1-11.
- Grieu, S., A. Traore, et al. (2005). "Prediction of parameters characterizing the state of a pollution removal biologic process." Engineering Applications of Artificial Intelligence **18**(5): 559-573.
- Haykin, S. (1998). Neural Networks A Comprehensive Foundation.
- Infopower (2002). "Bioetanol Galicia" Planta para la producción de 126500 m³/año de bioetanol y 120000 t/año de DDGS con cogeneración asociada de 24.8 MW en Teixeiro-Curtis, A Coruña. Infopower. **51**.
- Isermann, R. (1997). "Supervision, fault-detection and faultdiagnosis methods. An introduction." Control Engineering Practice **5**(5): 639-652.
- Isermann, R. (2004). Model based Fault detection and diagnosis. Status and applications. IFAC Symposium on Automatic Control in Aerospace, St. Petersburg Russia, Elsevier Science Ltd.
- Isermann, R. and P. Balle (1997). "Trends in the application of model-based fault detection and diagnosis of technical processes." Control Engineering Practice, **5**(5): 709-719.
- Jacques, K., T. Lyons, et al. (2003). The Alcohol Textbook (Fourth Edition)
A reference for the beverage, fuel and industrial alcohol industries.
- Johnson, R. A. and D. W. Wichern (1992). Applied Multivariate Statistical Analysis.
- Juez, J. M. B. (2005). Descubrimiento y refinamiento automático de conocimiento con técnicas híbridas del Aprendizaje Automático. Instituto de Investigación Tecnológica. Madrid, Pontificia Comillas. **Ingeniería Técnica Superior de Industriales**.
- Kleer, J. d. (1979). The Origin and Resolution of Ambiguities in Causal Arguments. 6th International joint conference on Artificial Inteligence, Tokyo.
- Kleer, J. d. (1990). Multiple representations of knowledge in a mechanics problem-solver. Readings in qualitative reasoning about physical systems, Morgan Kaufmann Publishers Inc.: 40-45.
- Kohonen, T. (1990). "The self-organizing map." Proceedings of the IEEE **78**(7): 1464-1480.

- Kuipers, B. (1984). "Commonsense reasoning about causality: deriving behavior from structure." Artificial Intelligence **24**: 169-203.
- Kuipers, B. (1993). "Reasoning with Qualitative Models." Artif. Intell.(59): 125-132.
- Kuipers, B. (1994). Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge. Cambridge.
- Lee, W. S., D. L. Grosh, et al. (1985). "Fault tree analysis, methods, and applications- A review." IEEE transactions on reliability **R-34**(3): 194-203.
- Lin, C. T. and C.-C. Lee (1996). Neural Fuzzy Systems. A Neuro-Fuzzy Synergism to Intelligent Systems.
- Nakkabi, Y., A. Doncescu, et al. (2002). Application of data mining in biotechnical process. IEEE International Conference of Systems, Man and Cybernetics, Tunisia.
- Nandi, S., H. A. Toliyat, et al. (2005). "Condition Monitoring and Fault Diagnosis of Electrical Motors – A Review." IEEE TRANSACTIONS ON ENERGY CONVERSION **20**(4): 719-729.
- Nauck, D., F. Klawonn, et al. (1997). Foundations of Neuro-Fuzzy Systems.
- Patton, R. J. (1997). Fault-Tolerant Control systems: the 1997 situation. IFAC Symposium on Fault Detection Supervision and Safety for Technical Processes.
- Patton, R. J. (2000). Soft computing approaches to fault diagnosis for dynamic systems: a survey. IFAC Symposium *SAFEPROCESS 2000*, Budapest, Elsevier Science Ltd.
- Patton, R. J. and P. M. Frank (2000). Issues of Fault Diagnosis for Dynamic Systems, Springer-Verlag Berlin and Heidelberg GmbH & Co. K.
- Rencher, A. C. (2002). Methods of Multivariate Analysis. New York.
- Richardson, M. (2003). "Learning with knowledge from Multiple Experts." Proceedings of the 20th International Conference of Machine Learning.
- Rinner, B. (2002). "Detecting and Diagnosing Faults." Telematik **2/2002**: 6-8.
- Roque, A. M. S. (1996). Aplicación de técnicas de redes neuronales artificiales al diagnóstico de procesos industriales. Instituto de Investigación Tecnológica. Madrid, Ponticia Comillas. **PhD**: 315.
- Schmidt, M., F. Kimmich, et al. (2000). Combustion Supervision by Evaluating the Crankshaft Speed and Acceleration. SAE 2000 World Congress, Detroit, Michigan, SAE.
- Venkatasubramanian, V., R. Rengaswamy, et al. (2002). "A review of process fault detection and diagnosis. Part I: Quantitative model-based methods." Computers & Chemical Engineering(27): 293-311.
- Venkatasubramanian, V., R. Rengaswamy, et al. (2002). "A review of process fault detection and diagnosis. Part II: Qualitative models and search strategies." Computers & Chemical Engineering(27): 313-326.
- Venkatasubramanian, V., R. Rengaswamy, et al. (2002). "A review of process fault detection and diagnosis. Part III: Process history based methods." Computers & Chemical Engineering(27): 327-346.
- Watson, H. A. (1961). "'Launch Control Safety Study'." Bell Telephone Laboratories **1**(1).
- Willimowski, M., F. Kimmich, et al. (2000). Signal model based fault diagnosis for combustion engines. 4th Symposium on fault detection, supervision and safety for technical process, Budapest, IFAC.
- Zadeh, L. A. (1964). "Fuzzy sets." Information and Control **8**(3): 338-353.

