

EL PAPEL DE LOS SISTEMAS BASADOS EN EL CONOCIMIENTO EN LOS SISTEMAS DE INFORMACION HIDROLOGICA

Ricardo Conejo¹, José Cuenca², Martín Molina²

¹ Departamento de Lenguajes y Ciencias de la Computación
Universidad de Málaga
Pza. del Ejido s/n
29013 - Málaga

² Departamento de Inteligencia Artificial
Universidad Politécnica de Madrid
Campus de Montegancedo s/n
28660 - Boadilla del Monte, Madrid

RESUMEN

Este artículo plantea, primeramente, las necesidades del usuario que explota los datos de los sistemas de información en tiempo real. De este análisis se desprende que, dados: (1) la necesidad de utilizar mezclas de modelos de sentido común con modelos matemáticos hidrológicos, así como (2) la necesidad de obtener explicaciones en tiempo real sobre las propuestas de los sistemas de explotación de datos para ayuda a la decisión, la arquitectura adecuada debe ser de tipo abierto, en la que sea accesible al usuario en forma de estructura de información el modelo conceptual de la aplicación. Por ello se recomienda el uso de sistemas basados en el conocimiento.

Seguidamente, se comenta la factibilidad de este tipo de arquitecturas partiendo de la discusión de dos modelos basados en el conocimiento que pueden ser utilizados para explotación de los sistemas de información del SAIH: (1) el modelo de validación de datos que incorpora un conjunto de modelos y bases de conocimiento que permiten detectar las situaciones de incoherencia de datos y explicarlas en forma acorde con vistas a su almacenamiento, modificación o descarte (2) el modelo de ayuda a la decisión en tiempo real cuyo prototipo experimental fueron los sistemas CYRAH y SIRAH promovidos por la Dirección General de Obras Hidráulicas y en el que se describe la arquitectura general del conocimiento para tratar este tipo de modelos y su concreción en el caso del sistema CYRAH.

Como consecuencia se concluye que la arquitectura basada en el conocimiento por un lado es necesaria y por otro es técnicamente factible en el momento actual; de hecho ya existe una herramienta de manejo de sistemas de conocimiento estructurado en nuestro propio grupo, el Knowledge Structured Manager (KSM) en base al cual es posible construir modelos híbridos de programación convencional y basada en el conocimiento.

1.- INTRODUCCION

El Ministerio de Obras Públicas y Urbanismo comenzó en 1982 el programa SAIH (Sistemas Automáticos para Información Hidrológica). El objetivo principal de esta inversión fue la instalación de sensores para registrar el estado de las cuencas hidrográficas (pluviómetros, sensores de nivel de río, grado de apertura de compuertas de embalses, caudal en el río, etc.) junto con una red de comunicación de radio para envío de los datos medidos a un centro de control. En dicho centro, un sistema informático realiza la explotación y aprovechamiento de los datos referentes al estado hidrológico de la cuenca controlada. La disponibilidad de esta tecnología supone el manejo de un gran

volumen de información al servicio de los operadores del centro de control y que ha de ser tratado y presentado de una forma adecuada para ser útil. Ello exige la instalación en los centros de control de sistemas que procesen adecuadamente la información recibida.

El presente artículo tiene por objeto, en primer lugar, proponer una reflexión sobre la forma de operación de los sistemas de información en tiempo real, en los que se da la circunstancia de un flujo importante de datos que los usuarios deben aprovechar para recaudar el máximo valor añadido de la importante inversión en infraestructura de comunicaciones hardware y software que estos sistemas representan, así como el papel que pueden jugar en una operación optimizada los sistemas basados en el conocimiento.

Seguidamente se comentan dos tipos de aplicaciones de esta clase de sistemas que han sido desarrollados experimentalmente en un ámbito académico, como son los sistemas inteligentes de validación de datos y otra en ámbitos más industriales, patrocinados por la Dirección General de Obras Hidráulicas, sobre sistemas de ayuda a la decisión en tiempo real.

2.- LOS PROBLEMAS DEL TRATAMIENTO DE DATOS EN TIEMPO REAL

Tradicionalmente, el mundo de la hidrología ha funcionado con estadísticas off-line obtenidas por sistemas de medida y de recogida de datos en forma prácticamente manual, como son los tradicionales aforos. El desarrollo de los sensores y de las telecomunicaciones hace que el panorama del entorno de estudios hidrológicos cambie dramáticamente de forma que con las excepciones debidas a errores en la medida automatizada, en el momento actual se tiene información en tiempo real y, por tanto, además de poderse desarrollar los tradicionales estudios off-line en gabinete, típicos en la ingeniería hidráulica tradicional, orientados a definir decisiones para períodos de tiempo largos, sea posible una nueva modalidad de estudios que se realizan en tiempo real en diálogo con los sistemas de información que presentan imágenes de la situación en cada momento. Este es el caso del tratamiento de emergencias y también es el caso de la validación de datos. En este último caso se recibe la información en latidos de uno, dos o cinco minutos y ésta pasa a archivos que serán objeto de valoración por instrumentos decisionales tanto en tiempo real como en estudios de gabinete. El hecho de que el archivo de información se produzca instantáneamente y que ésta sea abundante tanto en número de datos como en el tiempo de refresco, hace que la decisión de aceptación del dato, o clasificación de su calidad, tenga que tomarse inmediatamente antes de su archivo.

Análoga circunstancia se produce, pero de manera más grave, cuando se trata de tomar decisiones en tiempo real, ya que el operador humano se encuentra con un flujo de datos muy pormenorizado tanto en el tiempo como en el espacio y tiene que ser capaz de darse cuenta de su significado en términos de conceptos más agregados procedentes tanto de ámbitos teóricos como de criterios heurísticos de tipo profesional.

Dado que una persona es incapaz de interpretar los datos al ritmo que estos llegan, la solución adecuada es incluir entre el operador y los datos modelos que filtren la información de manera que identifiquen la presencia de situaciones problemáticas y que, al propio tiempo, recomienden actuaciones.

Hay dos aspectos que deben tenerse en cuenta en el diseño e instalación de estos modelos:

- El contenido conceptual de estos modelos, si tienen que reflejar el estado de conocimiento sobre el tema, no será determinista en su totalidad en el sentido de que en dicho contenido, además de incluirse modelística matemática cuyo comportamiento puede ser perfectamente determinista, se incluyen hipótesis sobre valoración de situaciones y otros criterios heurísticos sobre tipos de decisiones recomendables que pueden ser objeto de revisión en función de la experiencia y que en muchos casos no son únicos, es decir, dada una situación, puede que haya varias opciones, unas preferentes sobre otras según diversos criterios que en cada momento deben ser valorados por los operadores humanos.
- Este tipo de modelos recomienda acciones de las que pueden derivarse responsabilidades de los operadores humanos y, por ello, es fundamental que la implementación de los modelos se haga de manera que las personas que tengan que asumir las decisiones recomendadas no las acepten en forma ciega sino en base a que, dialogando con los modelos, obtengan explicaciones de las decisiones que les permita hacer suyas las recomendaciones sugeridas por dichos modelos.

De acuerdo con ambas condiciones, el diseño informático de este tipo de modelos debe hacerse mediante una arquitectura abierta en la cual, total o parcialmente, pueden introducirse modificaciones sobre el contenido conceptual de los modelos sin intervención de personas informáticas para reescribir código de programación, sino directamente por personas profesionales expertas en el tema que, previo análisis de las recomendaciones en operación de algunos sistemas vía el estudio de las explicaciones proporcionadas por el mismo, identifique que algunos contenidos conceptuales del sistema deben ser modificados, y para ello accedan al área de información en que está formulado el conocimiento y modifiquen aquellas piezas que les parezca oportuno. Esta funcionalidad es la de mantenimiento y diseño por no programadores.

Además de la anterior funcionalidad, debe incluir la posibilidad de explicación en la operación, de manera que los modelos deben formularse mediante procesos que representen maneras de razonar afines a las formas de razonamiento humano (por ejemplo, encadenamiento silogístico, estudio de encaje de situaciones a patrones predefinidos como prototipos de distintas clases, procesos de generación y prueba, etc.). Así, una traza del proceso presentará en cada paso qué premisas se han usado, qué pieza de conocimiento se ha utilizado y qué conclusiones se han obtenido aplicando un paso de los tipificados en el proceso de razonamiento.

La consulta de estas trazas de razonamiento permitirá a los operadores discernir si el sistema provee razones suficientes para aceptar sus conclusiones o cabe alguna duda que permita refinar las recomendaciones propuestas usando el juicio del operador.

En el caso de que el operador genere variantes en las recomendaciones del sistema cabe, opcionalmente, que éste le pregunte al sistema qué ocurre si aplican las actuaciones x , y , z , con lo que el sistema podrá asesorar al operador tanto proponiendo acciones como informando del posible impacto que las sugerencias adicionales o sustitutivas del propio operador podrían tener.

Las consideraciones anteriores hacen que sea inevitable, para una adecuada integración de las personas responsables y los sistemas, que estos tengan arquitecturas basadas en el conocimiento. Sin embargo, estas arquitecturas no deben entenderse en los términos triviales de los primeros sistemas expertos cuya estructura se describía por una base de conocimiento declarativa formada vía reglas o marcos prototipo con motores de inferencia simples formulados por procesos de encadenamiento de reglas o de análisis del grado de encaje (matching) entre casos observados y marcos de la base pero sin capacidad de integración con los procesos modelísticos habituales en el ámbito profesional. Hoy día se va a las arquitecturas mixtas en las que cooperan procesos convencionales para los que se crea una carátula explicativa y módulos formulados de manera cognitiva utilizando bien ecuaciones cualitativas (restricciones), o bien reglas, marcos, etc, procedentes de los esquemas tradicionales en Inteligencia Artificial. Este tipo de aplicaciones híbridas requiere un soporte instrumental más complejo que los primeros sistemas comercializados que se basaban en una estructura de representación única como pueden ser las reglas o los marcos, pero en todo caso son de posible desarrollo.

En conclusión, las aplicaciones tradicionales basadas en interfaces con usuario vía menú que permite seleccionar diferentes procesos que son realmente cajas negras para el operador, resultan inadecuadas y, por tanto, son obsoletas para soportar a los profesionales responsables que están operando en línea con el sistema.

Estas conclusiones son el resultado de nuestra experiencia durante los últimos diez años en el desarrollo de aplicaciones basadas en el conocimiento y notablemente fueron una conclusión obtenida de la primera experiencia en el desarrollo de los sistemas expertos para el SAIH en el que se llegaron a construir dos arquitecturas: la arquitectura SIRAH y la arquitectura CYRAH [Alonso et al. 90], [Alonso et al. 92], [Cuenca et al. 91], en donde cooperaban distintos módulos de representación del conocimiento y en donde se incluían representaciones basadas en restricciones, en reglas y en marcos, articuladas mediante procedimientos algorítmicos de razonamiento a distintos niveles. Estos procesos eran los ya conocidos en ingeniería como son el *routing* de caudales a lo largo del árbol de cauces en una cuenca receptora y la clasificación de situaciones de niveles como patrones de problemas de inundaciones predefinidas.

En el desarrollo de los primeros sistemas CYRAH y SIRAH fue necesario construir software *ad hoc* para soportar los modelos de conocimiento que se identificaron en el proceso de análisis, sin embargo la experiencia desarrollada en el Laboratorio de Sistemas Inteligentes desde 1990 en otros proyectos para ayuda a la toma de decisiones en tiempo real ha permitido decantar un modelo conceptual de conocimiento estructurado que puede ser útil para la construcción de versiones de los sistemas de ayuda a la decisión en caso de inundaciones o de validación de datos.

Esta metodología de representación se ha substanciado en el entorno de software KSM (Knowledge Structure Manager), desarrollado por este grupo [Cuenca, Molina, 94], [Molina, 93] y cuya base conceptual es la unidad cognitiva, entidad abstracta que en cierta manera sustituye al concepto de proceso u objeto en informática convencional, en la que coexisten, relativos a un tipo de solución de problemas, tanto una colección de métodos para resolver este tipo de problemas como las piezas de conocimiento declarativo y procedimental que deben usar estos métodos para alcanzar sus objetivos. Es decir, el módulo de formulación de una aplicación, en lugar de ser el típico código de programa en el que se especifican estructuras de datos y algoritmos, incluye métodos descritos vía reglas de producción y pasos de inferencia elementales para los que es preciso utilizar algunas operaciones elementales de inferencia definidas en las unidades cognitivas componentes (figura 1).

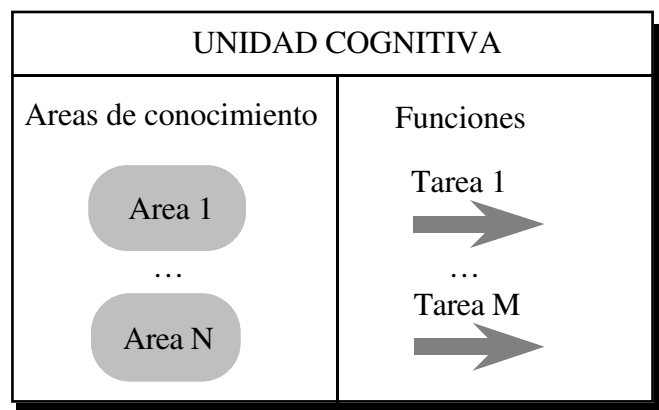


Figura 1: Formato de unidad cognitiva

De esta forma puede formularse una aplicación en base a lo que se denomina en KSM el *patrón conceptual*, definido como una jerarquía de unidades cognitivas en las que unas forman parte de otras y, por tanto, en su grafo de representación, los arcos representan la relación *forma parte de*.

En los párrafos que siguen se estudian dos tipos de problemas como son la validación de datos filtrados por modelos en tiempo real y la toma de decisiones en tiempo real en episodios de avenida, utilizando esta idea conceptual que permitirá construir modelos basados en el conocimiento con mayor flexibilidad y eficacia que las primeras versiones desarrolladas para CYRAH y SIRAH.

3.- VALIDACION DE DATOS DE SENSORES

La instalación de los sistemas de adquisición automática de datos para la monitorización y control en las cuencas hidrográficas constituye un salto cualitativo en el campo de la hidrología. Este gran volumen de información amplía el campo de los procesos hidrológicos que pueden ser estudiados de forma cuantitativa, pero por otra parte, genera problemas específicos de estos sistemas como son el mantenimiento físico de la red de transmisión de datos, la gestión y el proceso de los mismos, y la validación o verificación de su veracidad.

Hasta la aparición de los sistemas de adquisición automática de información hidrológica, la validación o verificación de la veracidad de los datos de intensidades de lluvia, caudales en los cauces, niveles de los embalses, etc. se ha realizado de forma manual aplicando para ello el conocimiento que sobre el comportamiento de las cuencas disponen los ingenieros encargados de su gestión. El gran volumen de información actual implica necesariamente la automatización de esta tarea.

El problema de la detección de errores en las series de datos hidrológicos ha sido escasamente estudiado. No puede decirse que existan técnicas genéricas para el control de la calidad de los datos. En la Guía de Practicas Hidrológicas de la Organización Meteorológica Mundial [OMM, 84] se hace referencia a *procedimientos de control de datos sobre la temperatura y la precipitación basados en la ordenación mecánica de los datos diarios por distritos o por regiones*. Trás esta ordenación se realiza una *inspección visual* de los datos y se aplica el *sentido común* para detectar *errores groseros* en la medición, ya sea por causa de una magnitud excesiva, o por causa de una fecha errónea en el mismo.

3.1.- Análisis del problema

En general, la validación de datos es un problema asociado a todos aquellos dominios en los que un conjunto de sensores se encargan de vigilar el estado de un sistema. Por tanto, no es de extrañar que se hayan desarrollado mecanismos y técnicas de validación en campos aparentemente tan dispares como la industria aeroespacial, las centrales nucleares [Ray, 86], las plantas químicas [Chandrasekaran, 87], o los sistemas de redes de abastecimiento de agua [Watts, 93]. En muchos de estos campos aparece el problema de validación de datos como un subproblema asociado al problema de la monitorización del funcionamiento del sistema mediante la detección de los valores de ciertas magnitudes estratégicamente seleccionadas.

Podría pensarse que en el caso de la validación de los datos hidrológicos, debe recurrirse al empleo o adaptación de alguna de las técnicas empleadas en estos otros dominios. Sin embargo, un estudio más detallado revela que en general tratan un problema radicalmente diferente al que se plantea en el caso hidrológico.

En primer lugar, el número de sensores disponible en los sistemas hidrológicos es relativamente pequeño, como consecuencia de su alto coste de instalación. En una central nuclear, o una planta química, la instalación de un sensor es

relativamente simple. Por tanto, si se desea tener redundancia en las medidas basta con incorporar múltiples sensores que midan directamente la misma magnitud. En general, las técnicas desarrolladas en estos sectores sólo son aplicables cuando existe un alto grado de redundancia. En el sistema de monitorización hidrográfico, por el contrario, en general, no existen sensores duplicados, con lo cual la redundancia debe buscarse en la dependencia indirecta que imponen las leyes físicas entre medidas realizadas en diferentes puntos de la red, utilizando modelos que permitan estudiar la coherencia entre los valores medidos directamente y los extrapolados a partir de los valores medidos en otros puntos, o al menos, estableciendo los niveles de tolerancia admisibles entre medidas en distintos puntos.

Si bien los fenómenos hidrológicos han sido ampliamente estudiados, siguen siendo difíciles de cuantificar debido principalmente a la variabilidad y complejidad de los procesos que tienen lugar en las cuencas. Por tanto, una característica inherente al dominio de la hidrología es la imprecisión y en algunos casos incertidumbre que afecta a los modelos y relaciones que se establecen. Esta circunstancia viene a complicar aún más el problema de la validación de los datos, ya que rara vez los modelos serán capaces de predecir valores exactos que poder confrontar directamente con los medidos.

Por lo que respecta a las propias medidas, el fallo puede ser de distintos tipos:

- (1) *Fallos permanentes.* Correspondientes a disfunciones en el sensor persistentes a lo largo del tiempo, como por ejemplo, cuando la carencia absoluta de datos en un periodo de tiempo es sustituida por valores nulos¹, o hay una descorrección de la escala de medida.
- (2) *Fallos instantáneos aislados o Errores espúreos.* Correspondientes a valores claramente incorrectos, que por cualquier causa se introducen en una serie fundamentalmente correcta, es decir, los que se producen cuando un sensor que funciona correctamente de improviso envía una medida anómala.
- (3) *Errores sistemáticos.* Se producen cuando en todo el periodo de observación el dato obtenido por el sensor es mayor (o menor) que la magnitud real que se está midiendo. Estos fallos no implican necesariamente un mal funcionamiento en los sensores, sino probablemente un error en la calibración, que puede ser corregido.
- (4) *Errores aleatorios.* Son los debidos a la sensibilidad misma del sistema o el procedimiento de medida. Por ejemplo, es evidente que con un limnómetro es imposible determinar con exactitud el caudal de una sección.

¹ Este tipo de fallo en los datos generalmente corresponde a un fallo físico del sensor estando prevista en el sistema SAIH la autodetección de los mismos cuando se trata de una avería en el sistema electrónico de los sensores o en el de transmisión de los datos. Sin embargo, no es descartable la hipótesis de un fallo permanente no detectado por el sistema de autovalidación al nivel físico del SAIH, debido a roturas de los elementos físicos, que conduciría a la obtención de series de datos claramente anormales.

La estrategia de validación del sistema propuesto debe tener en cuenta la posible aparición de estos tipos de fallo. Desde el punto de vista real, el problema es aún más complejo, ya que los distintos tipos de fallo se superponen unos a otros. Además se debe tener en cuenta que ciertas medidas resultan más verosímiles o fiables que otras, ya sea debido a la calidad del sensor o a la menor variabilidad de la magnitud medida.

Por otra parte, una vez determinado que existe un fallo en un sensor, una de las posibles líneas de actuación es la recuperación de ese error, es decir, la estimación del valor real de la magnitud que se desea medir. Si se dispone de múltiples sensores que miden directamente la misma magnitud esta tarea resulta trivial, basta con eliminar los datos del sensor erróneo. Sin embargo, si la redundancia en las medidas se obtiene entre sensores que miden distintas magnitudes, se hace necesaria la aplicación de modelos que permitan inferir valores razonables para las variables desconocidas.

Resumiendo, un sistema de validación de datos hidrológicos tiene que ser capaz de:

- (1) Saber cuales son los márgenes admisibles de funcionamiento correcto para la magnitud física que se está midiendo. Un experto humano, bajo determinadas circunstancias, es capaz de detectar un fallo en una serie de datos hidrológicos con una simple inspección visual de los mismos. Esto se debe a que conoce la magnitud que se está midiendo, y sabe qué límites máximos y mínimos son admisibles tanto para sus valores absolutos, como para las pendientes de cambio, las oscilaciones, etc., es decir, las formas características de la serie.
- (2) Utilizar el conocimiento existente sobre el comportamiento del sistema hidrológico, de manera que puedan inferirse los valores estimados en unos sensores a partir de las de otros. Este conocimiento ya existe y en algunos casos está formalizado mediante modelos hidrológicos, ya sea cualitativos o cuantitativos. Sin un conocimiento sobre el comportamiento hidrológico de la cuenca es imposible realizar una buena validación de los datos
- (3) Ser capaz de trabajar con información imprecisa o tal vez incierta. Dado que los modelos no son capaces de inferir valores exactos, y dado que el problema de la validación en sí mismo tiene como datos de entrada valores imprecisos o inciertos, es fundamental un tratamiento de estos aspectos. De hecho, desde un punto de vista teórico, no es posible establecer con absoluta certeza entre un conjunto de medidas aparentemente correctas, pero susceptibles de fallo, cuáles son realmente correctas y cuáles no lo son. Por ejemplo, pensando en el caso de dos sensores que discrepan siempre es teóricamente posible creer en uno o en otro indistintamente. En el caso general, esto lleva al establecimiento de distintas hipótesis de funcionamiento del sistema que se está midiendo, algunas de las cuales serán más creíbles que otras.

- (4) Establecer la estrategia global de razonamiento sobre validación. Para ello, es necesario en primer lugar seleccionar de entre todo el conjunto de sensores que hay en la cuenca, aquellos que están relacionados entre sí. Una vez hecho esto es necesario establecer los criterios que se aplican para decidir cuando son coherentes sus medidas, según los márgenes de tolerancia admisibles, y que hipótesis de funcionamiento del sistema corresponden a cada caso. Por último, el modelo de razonamiento para validación debe establecer el orden en que se realizan estas comprobaciones, y los criterios de selección de las hipótesis más creíbles.

Por todo ello, se pone de manifiesto que el principal problema de la validación de los datos hidrológicos es un problema de estructuración del conocimiento. Los elementos del sistema de validación están todos ellos relacionados en una u otra forma con distintos tipos de conocimiento: Conocimiento sobre la forma de la serie temporal que se está midiendo, conocimiento sobre el funcionamiento del sistema en su conjunto y sobre cada uno de sus componentes, conocimiento sobre los márgenes de tolerancia admisibles y la credibilidad de cada hipótesis, conocimiento en suma, sobre cómo engranar todo este conocimiento para ser capaz de construir un sistema computable que valide los datos.

Queda aún abiertas dos cuestiones de tipo pragmático: ¿se dispone de todo este conocimiento?, y en caso afirmativo ¿es posible expresarlo de forma computable? A nuestro juicio, la respuesta a la primera pregunta está por demostrar. En general el conocimiento que se tiene no es completo. La calidad de la validación será función de la corrección y completitud de éste. Para algunas cuencas o regiones será mayor que para otras. De hecho, este conocimiento no es un cuerpo estático, inmutable, sino que por el contrario evoluciona conforme en el tiempo por efecto de la propia experiencia. La segunda de estas cuestiones es la que se intenta resolver en esta propuesta. Ahora bien, no hay que perder de vista las consideraciones anteriores, lo que lleva a la necesidad de plantear *arquitecturas*² suficientemente abiertas y flexibles como para poder evolucionar conjuntamente al conocimiento de que se disponga. Para conseguir este objetivo, se debe hacer explícita la mayor cantidad posible del conocimiento que se emplee, facilitando así la comprensión del funcionamiento del sistema por parte del usuario y proporcionando un entorno en el cual el conocimiento aplicado pueda ser fácilmente revisado, incluso por el propio usuario, sin necesidad de recurrir a una programación compleja.

3.2.- Descripción del conocimiento para validación de datos

Para definir una arquitectura con estas características, es necesario identificar y descomponer el problema en sus componentes cognitivos básicos [Chandrasekaran, 92]. Los sistemas de validación de datos que se construyan a partir de esta arquitectura se harán mediante la agrupación de dichas piezas conceptuales.

² El uso del término "arquitectura" en Inteligencia Artificial proviene de la acepción de "mecano" o conjunto de piezas de un juego de construcción infantil. Esta acepción no es muy común en castellano, pero sí en inglés de donde proviene. Otra acepción muy relacionada con la anterior y con la que indistintamente se emplea esta palabra es "metodología de construcción de programas"

En la arquitectura propuesta para resolver el problema de validación se han identificado las siguientes áreas de conocimiento básicas:

- (1) Identificación y Corrección de errores aislados. Los errores espúreos, son el tipo de fallo más frecuente que se produce en los sistemas de adquisición automática de datos, que puede ser detectado y corregido aplicando un conocimiento sobre el comportamiento de la serie de datos leída a lo largo de un periodo de tiempo. Para incluir este conocimiento es necesario poder expresar en qué circunstancias se puede determinar la existencia de un error y cuál es la corrección que se llevará a cabo.
- (2) Comportamiento de los sistemas hidrológicos. Tradicionalmente este conocimiento se expresa mediante modelos y viene estructurado de forma natural por los distintos componentes físicos de la red hidrográfica: áreas receptoras de lluvia, tramos de transporte, etc. En algunos casos este conocimiento puede estar ya codificado y listo para su uso, mediante programas de simulación, sistemas de reglas, etc.
- (3) Coherencia entre diferentes medidas (directas o indirectas) de una misma magnitud y planteamiento de diversas hipótesis de funcionamiento. Este conocimiento expresa bajo qué circunstancias pueden considerarse coherentes entre sí dos o más medidas. En caso de conflicto entre las medidas se establecerán diversas hipótesis o "tanteos" basadas en el correcto funcionamiento de un subconjunto de sensores. Es importante que el sistema tenga presente qué medidas avalan cada una de las hipótesis, ya que si una hipótesis es finalmente rechazada deben rechazarse también las medidas que llevan a esa conclusión. La arquitectura propuesta utiliza un ATMS o sistema de mantenimiento de la razón basado en suposiciones [De Kleer, 86] para gestionar internamente el contexto en el que es válida cada suposición de modo similar al sistema GDE [De Kleer, 87], aunque con ciertas modificaciones.
- (4) Criterios de selección de las hipótesis más verosímiles. Dado que en el sistema de validación surgen de forma natural diversas hipótesis de funcionamiento, será necesario establecer de forma explícita el conocimiento que se emplea para seleccionar unas hipótesis y/o rechazar otras. Por ejemplo, en el caso de tres sensores, dos de los cuales son coherentes entre sí y discrepan de un tercero, por lo general se establecerá como correcto el funcionamiento de los dos primeros, pero si hay una diferencia muy apreciable entre la fiabilidad de estos y la del tercero, se puede asignar mayor credibilidad a la hipótesis de que es éste el que funciona.
- (5) Unificación de hipótesis de funcionamiento similares de cara a ulteriores validaciones. Este conocimiento es necesario desde un punto de vista práctico para evitar la proliferación de hipótesis de funcionamiento similares entre las que no es posible discriminar una información útil para la validación. Por ejemplo, si a partir de las medidas de intensidad de lluvia en dos pluviómetros se establecen hipótesis similares de funcionamiento de la cuenca, estimando valores similares del caudal de

escorrentía resultante, puede ser conveniente unificarlas en una única hipótesis a efectos de disminuir el coste computacional.

- (6) A esa descomposición cognitiva del problema hay que sumar el (meta)conocimiento sobre la estructuración de todas estas piezas de conocimiento, es decir, el conocimiento sobre cómo hacer que estas piezas elementales cooperen para realizar la tarea de validación. Este (meta)conocimiento no puede hacerse explícito de manera genérica, ya que está intrínsecamente ligado a cada caso concreto. Lo que sí es importante resaltar en este punto es que finalmente también debe aparecer explícito en el sistema.

Estas piezas elementales deben agruparse para facilitar su uso dando lugar a lo que en la terminología de KSM [Molina, 93] se denominan *unidades cognitivas*.

Desde un punto de vista genérico pueden construirse dos tipos de unidades cognitivas:

- (1) U.C. Coherencia Local de cada sensor. Su función en el sistema es determinar si la medida de un sensor es aparentemente correcta y por tanto puede ser utilizada para validar otras medidas. Esta comprobación incluye: (a) una comprobación local instantánea que puede realizarse mediante el conocimiento de identificación y en su caso corrección de errores aislados, y (b) una validación local que se realiza utilizando el conocimiento de coherencia (restringido a un sólo sensor) que determina si la serie, considerada como un dato inseparable, es válida o no.
- (2) U.C. Coherencia Global de los sensores de una cuenca. La tarea que realiza esta unidad cognitiva consiste en inferir el funcionamiento real del sistema hidrológico al que representan a partir de los datos disponibles que proporcionan las unidades cognitivas de los sensores, (o en su caso establecer diversas hipótesis de funcionamiento). Al realizar esta tarea se obtiene como subproducto, tanto la validación de los datos de los sensores, como la estimación de los valores correctos, y una explicación de la misma realizada en términos de las suposiciones y modelos empleados para establecer estas conclusiones. Esta unidad cognitiva combina las cinco piezas básicas de conocimiento descritas anteriormente. El modo de combinarlas es una aplicación de la última de las piezas descritas y se expondrá más adelante mediante un ejemplo.

Para cuencas grandes en las que exista un gran número de sensores la estrategia para realizar la validación consiste en construir unidades cognitivas de cuenca que realicen validaciones parciales, combinando los resultados de cada una de ellas siguiendo una estructura jerárquica que puede establecerse siguiendo la estructura física de cuencas y subcuencas.

Como primer paso en la validación de los datos se construyen tres Unidades Cognitivas de Coherencia Local para cada uno de los sensores. En ellas se establecen, por ejemplo, que las condiciones que debe cumplir la señal del pluviómetro de Santopitar sabiendo que la magnitud que mide corresponde a la intensidad de lluvia. Cada una de estas reglas lleva asociada la acción a realizar para corregir los valores que no verifiquen la condición. La base de conocimientos es la siguiente³:

```

ALIAS:
  I1 :: (Pluviometro Santopitar) intensidad de lluvia;
CORRECCION:
  I1[T] < 0 --> {I1[T] := 0;}
  I1[T] > 15 --> {I1[T] := I1[T-1];}
  I1[T] > I1[T-1]+5 --> {I1[T] := min((I1[T-1]+I1[T+1])/2, I1[T-1] + 5);}
  I1[T] < I1[T-1]-10 --> {I1[T] := max((I1[T-1]+I1[T+1])/2, I1[T-1] - 10);}

```

En donde I_1 ha sido definida como la serie de intensidad de lluvia leída por el pluviómetro y la variable T se usa para indicar un instante genérico que se supone cuantificado universalmente. Esta base de conocimientos indica que deben eliminarse aquellos valores menores que 0 y mayores de 15 mm/h, así como aquellos que excedan una pendiente máxima o mínima que se considera admisible. En los distintos casos se hace una estimación del valor que debería haberse medido. Las reglas de corrección se aplican secuencialmente.

Para construir un modelo de esta cuenca, se considera que puede descomponerse en tres elementos hidrológicos principales: un Area Receptora que promedia la lluvia de toda la zona norte, cercana a la población de Colmenar, otra Area Receptora que represente a la zona sur, y un Tramo de Transporte que simula el comportamiento de una parte del cauce principal, estableciendo un retardo en el caudal obtenido en el area receptora. Para implementar estos modelos pueden emplearse programas externos a la arquitectura, o bien pueden definirse dentro de ésta. Por ejemplo, un modelo de lluvia-escorrentía basado en el método del hidrograma unitario para el Area Receptora de Colmenar puede expresarse como:

```

MODELO:Area Receptora Colmenar {Pluviometro Santopitar,Pluviometro Salto del Negro}
  QColmenar[T] := HUColmenar * (0.625*I1[T-2..T]+0.375*I2[T-2..T]);
  fiabilidad := 0.75 * f1 * f2;

```

en donde los objetos que aparecen entre llaves indican las dependencias funcionales del modelo. Los identificadores I_1 e I_2 representan respectivamente a las intensidades de lluvia de los pluviómetros de Santopitar y Salto del Negro, f_1 y f_2 son las fiabilidades asignadas a los mismos, y $Q_{Colmenar}$, $HU_{Colmenar}$ y $fiabilidad$ son atributos del Area Receptora Colmenar que se han definido como:

```

OBJETO: Area Receptora Colmenar; ES UN: Area Receptora
ATRIBUTOS:
  caudal (SERIE)
  fiabilidad (NUMERICO);
  Hidrograma Unitario (SERIE) = [1,2,5];

```

³ Las bases de conocimiento se expresan utilizando un lenguaje denominado COVALTO (Corrección y VALidación de daTOs). Este lenguaje permite utilizar expresiones y relaciones tanto entre escalares como entre series temporales.

Para que sea posible realizar la validación de los datos de los sensores, deben incluirse también dos áreas de conocimiento de coherencia: una que establece la relación existente entre las medidas de los dos pluviómetros, y otra que establece las discrepancias que son admisibles entre el caudal aforado en el Salto del Negro y el caudal que resulta de la estimación que los modelos hacen a partir de las medidas de los dos pluviómetros.

La primera de estas áreas de conocimiento establece cuándo pueden considerarse coherentes los valores medidos por los pluviómetros de la región que abarca esta unidad cognitiva. En general pueden darse cuatro situaciones: (1) que ambos pluviómetros funcionen correctamente, en cuyo caso las series de intensidades de lluvia medidas en ellos serán coherentes, (2) y (3) que las medidas de ambos pluviómetros discrepen en cuyo caso se puede suponer que uno de ellos funciona correctamente y el otro no, o viceversa; y (4) que ambos pluviómetros fallen simultáneamente. En los tres primeros casos se realiza una estimación de los valores correctos según proceda. La base de conocimiento para llevar a cabo esta tarea es la siguiente:

```
COHERENCIA: {Pluviometro Santopitar,Pluviometro Salto del Negro}
VALIDACION:
  correlacion(I1,I2) > 0.6 && media( I1 ) > 1.1 * media( I2 )
ESTIMACION:
  <-- Pluviometro Santopitar -->
  f1 := 1-(1-f1)*(1-f2);
  <-- Pluviometro Salto del Negro -->
  f2 := 1-(1-f1)*(1-f2);
```

```
COHERENCIA: { Pluviometro Santopitar }
VALIDACION:
ESTIMACION:
  <-- Pluviometro Santopitar -->
  <-- Pluviometro Salto del Negro -->
  I2[T] := (1/1.20) * I1[T];
  f2 := 0.9 * f1;
```

```
COHERENCIA: { Pluviometro Salto del Negro }
VALIDACION:
ESTIMACION:
  <-- Pluviometro Santopitar -->
  I1[T] := 1.20 * I2[T];
  f1 := 0.9 * f2;
  <-- Pluviometro Salto del Negro -->
```

De forma resumida, esta base de conocimiento de coherencia impone como condición para que puedan ser consideradas ambas series válidas, que la correlación sea mayor de 0,6 y que la media de intensidades en el pluviómetro de Santopitar, situado a mayor altitud, sea mayor que el 110% de la media en el pluviómetro del Salto del Negro⁴. En caso de que se verifique esta condición los valores de las series de los pluviómetros se estima que corresponden a los valores medidos, siendo avalada esta hipótesis por el funcionamiento de ambos sensores. En caso de que no se verifique la condición anterior, la base de conocimientos plantea la segunda y tercera hipótesis, a las que no se les impone ninguna condición adicional. En estos casos, la segunda de las reglas de coherencia permite realizar una estimación del valor que debía haberse medido en el pluviómetro de Santopitar a partir del valor que se ha medido en el Salto del Negro (y que en esta hipótesis se considera correcto). La tercera regla es

⁴ Estos criterios deben emanar de la experiencia o del estudio de casos anteriores. En este ejemplo se han establecido criterios apriorísticos dictados por el sentido común.

simétrica de la anterior para el caso en el que la hipótesis de funcionamiento correcto sea la del pluviómetro de Santopitar.

Una vez establecidos los modelos hidrológicos de la cuenca del Río de la Cueva, se obtendrán una o varias estimaciones del caudal que se produce en el Salto del Negro. Este caudal estimado se compara con el caudal realmente aforado, (previamente eliminados de esta medida los valores espurios) utilizando para ello las reglas de coherencia de la segunda base de conocimientos. Al igual que en la primitiva de coherencia anterior, puede ocurrir que ambas medidas sean coherentes, o que no lo sean, lo que provocará las correspondientes estimaciones en su caso. La base de conocimientos de este elemento del sistema es la siguiente:

COHERENCIA:{Aforo Salto del Negro,Tramo de Transporte del Río de la Cueva}

VALIDACION:

| Qaforado[T] - Qestimado[T] | < 0.5*Qaforado[T]+5

ESTIMACION:

(Unidad Cognitiva Río de la Cueva) caudal[T] := Qaforado[T];

(Unidad Cognitiva Río de la Cueva) intensidad de lluvia media[T] := I[T];

(Unidad Cognitiva Río de la Cueva) fiabilidad := 1-(1-f1)*(1-f2);

COHERENCIA: { Tramo de Transporte del Río de la Cueva }

VALIDACION: // sea cual sea

ESTIMACION:

(Unidad Cognitiva Río de la Cueva) caudal[T] := Qestimado[T] ;

(Unidad Cognitiva Río de la Cueva) intensidad de lluvia media[T] := I[T];

COHERENCIA:{ Aforo Salto del Negro}

VALIDACION: // sea cual sea

ESTIMACION:

(Unidad Cognitiva Río de la Cueva) caudal[T] := Qaforado[T];

(Unidad Cognitiva Río de la Cueva) intensidad de lluvia media[T] := DESCONOCIDO;

(Unidad Cognitiva Río de la Cueva) fiabilidad := f1;

En esta base de conocimiento la condición que según la cual se pueden considerar coherentes las estimaciones y las mediciones de caudal en el Salto del Negro. I[T] representa la intensidad de lluvia media, y f1 y f2, son respectivamente las fiabilidades de los valores medidos y estimados⁵.

En caso de que se produzcan varias hipótesis para seleccionar aquella asociada a una mayor fiabilidad se establece un área de conocimiento de preferencias, que viene dada por:

PREFERENCIAS: > (Unidad Cognitiva Río de la Cueva) fiabilidad

Por último, falta por establecer el conocimiento general de control, es decir la estructuración de todas las piezas de conocimiento anteriores de manera que operen correctamente. Este conocimiento se muestra de forma gráfica conforme muestra la figura 3. Los métodos se representan mediante círculos y los resultados de la ejecución de los mismos mediante cuadrados.

⁵ Nótese que en el caso en el que deba hacerse una estimación sólo a partir de los valores de caudal aforado, resulta imposible estimar el valor de la intensidad de lluvia media, dado que se ha empleado un modelo cuantitativo de lluvia-escorrentía. Sin embargo, si se hubiera empleado un modelo cualitativo podría reutilizarse en este punto para hacer una abducción de las lluvias que provocan el caudal aforado.

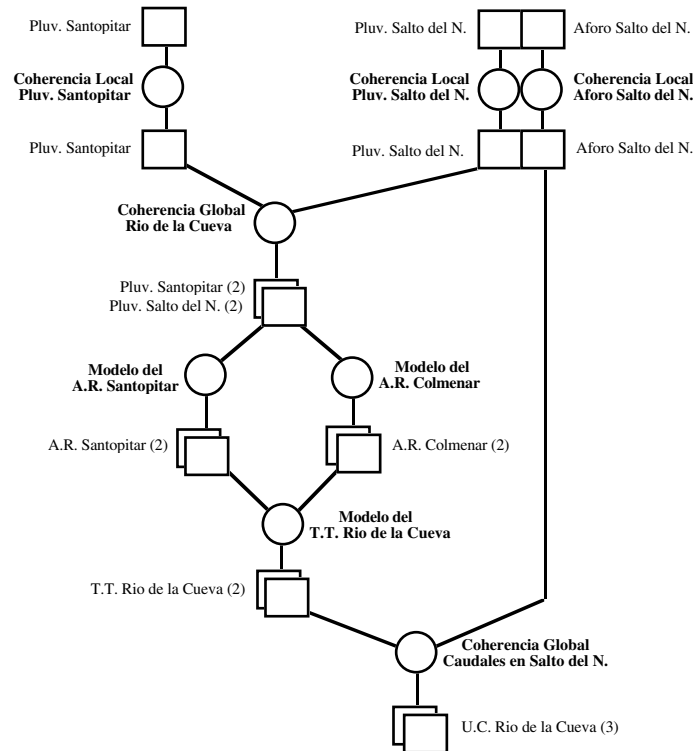


Figura 3: Organización de las piezas de conocimiento en el área ejemplo

3.4.- Resultados de una ejecución

Para ilustrar el funcionamiento del sistema se ha considerado una serie de 24 valores horarios para cada uno de los sensores que intervienen en la cuenca del río de la Cueva. En la figura 4 se muestran tres series correspondientes a los valores leídos en el pluviómetro de Santopitar, los que se obtienen corrigiendo la serie según el conocimiento de la Unidad Cognitiva de Coherencia Local, y el que resulta de la estimación a partir de los datos del pluviómetro del Salto del Negro.

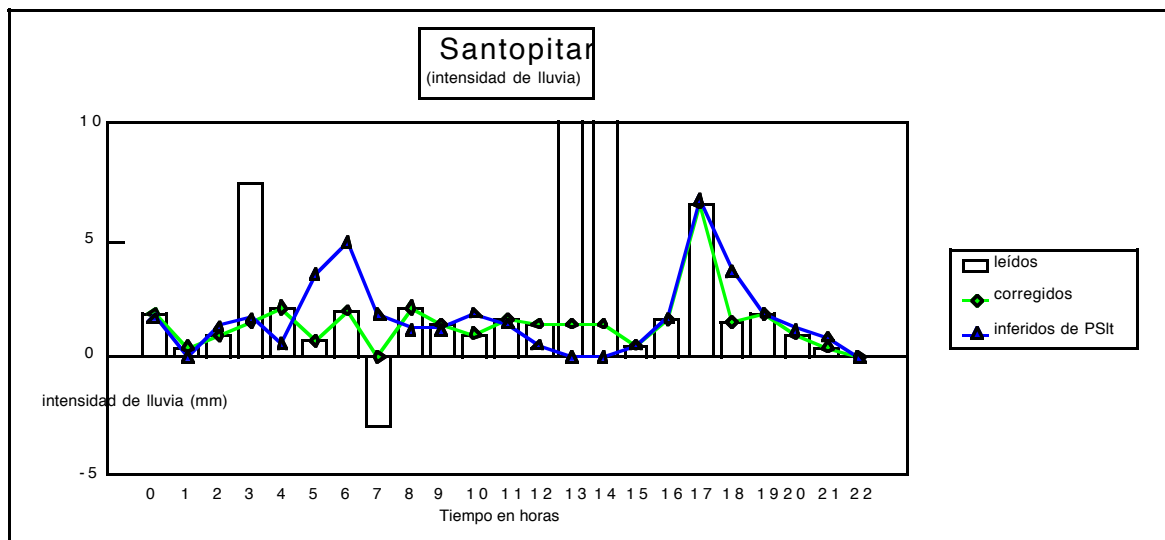


Figura 4: Corrección de valores por coherencia local

La primera de las condiciones impuesta en la base de conocimientos de coherencia global de los pluviómetros de la cuenca no se satisface, y por tanto se realizan las dos siguientes estimaciones. Por tanto, se obtienen dos posibles

valores de intensidad de lluvia en el pluviómetro Santopitar. Ambas hipótesis tienen como soporte respectivamente el funcionamiento correcto del propio pluviómetro y el fallo del otro, o bien el funcionamiento del otro y el fallo en el primero. Análogamente para el pluviómetro del Salto del Negro se establecen otras dos hipótesis.

Dado que hay dos hipótesis de valores para la intensidad de lluvia en cada uno de los pluviómetros, existen cuatro combinaciones de hipótesis en la entrada a los modelos, sin embargo, sólo dos de ellas son consistentes entre sí (las dos parejas cuyas justificaciones soporte son consistentes). El resultado de estos modelos vuelven a ser dos hipótesis basadas así mismo en el funcionamiento correcto de cada uno de los pluviómetros y el fallo del otro.

Estas dos hipótesis son la entrada al Área de Conocimiento de Coherencia de Caudales en el Salto del Negro, en donde se realizan las dos combinaciones posibles de estas hipótesis con el caudal aforado, para hallar de esta forma una o más estimaciones del caudal resultante de esta Unidad Cognitiva. La ejecución de este proceso da como resultado la coherencia de la hipótesis basada en el funcionamiento del pluviómetro del Salto del Negro, con el caudal medido por el Aforo en el mismo punto, lo que genera una estimación para el caudal de la Unidad Cognitiva. La otra hipótesis de caudal estimado no es coherente con el caudal aforado, por lo que se realizan dos estimaciones más, una de ellas basada en dicho caudal estimado y otra basada sólo en el caudal aforado. Las distintas medidas y estimaciones del caudal en este punto se presentan en la figura 5.

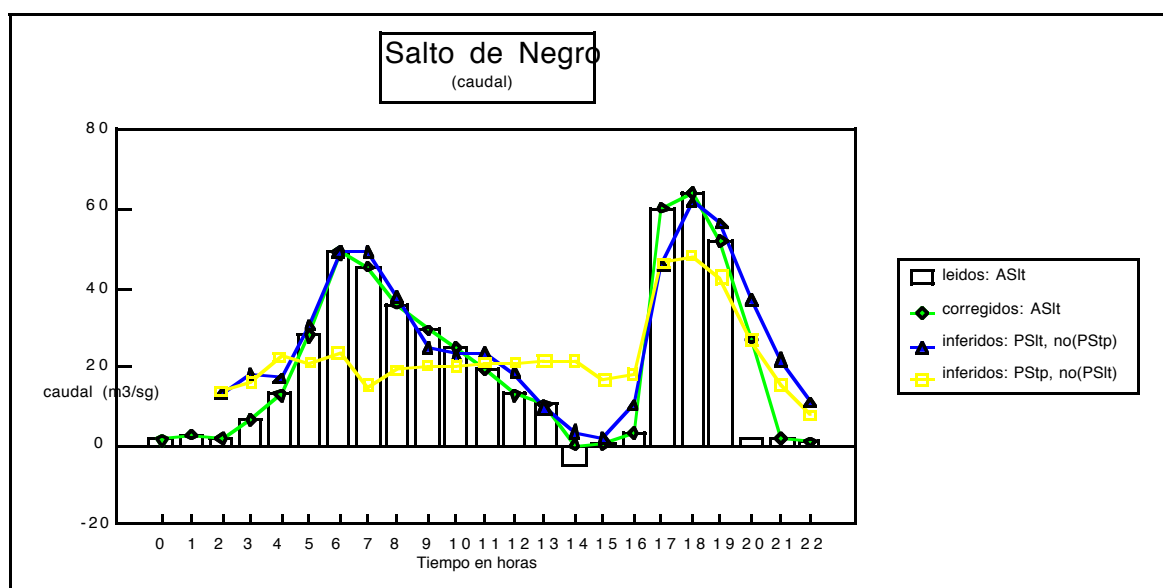


Figura 5: Valores de caudal medidos y estimados en Salto de Negro

Por consiguiente, el resultado de la ejecución de esta unidad cognitiva son tres instancias, que finalmente quedan en dos, al unificar aquellas que dan el mismo caudal. La primera de estas hipótesis se sigue del funcionamiento del aforo en Salto del Negro o bien del funcionamiento del pluviómetro en este punto y el no funcionamiento del otro pluviómetro. La segunda de las hipótesis (de la que se infiere un caudal sensiblemente diferente) está asociada al funcionamiento correcto del pluviómetro de Santopitar y del fallo simultáneo de los sensores de caudal e intensidad de lluvia en el Salto del Negro. Mediante la base de

conocimientos de preferencias se ordenan en orden decreciente estas hipótesis según su fiabilidad, seleccionando finalmente la primera de ellas.

4.- AYUDA A LA DECISION EN SITUACIONES DE EMERGENCIA

Las situaciones de emergencia por avenidas o inundaciones demandan unos métodos de operación en el centro de control en donde los sistemas expertos en tiempo real pueden desempeñar un papel extremadamente útil. Estos sistemas analizan sistemáticamente las diferentes opciones de solución, ante la presencia de la emergencia, manteniendo un funcionamiento uniforme independiente de la gravedad de las condiciones ambientales, sin sentirse sometidos a la presión de emergencia de la situación tal como puede ocurrir con algunas personas.

Asimismo, la capacidad explicativa de los sistemas expertos permite que el asesoramiento en estas situaciones sea eficaz. No basta únicamente con la recomendación sobre las acciones a realizar ante una situación de emergencia. Deben justificarse adecuadamente para que la persona que toma la decisión pueda asumir la responsabilidad que la acompaña.

En esta línea ha se han construido en España sistema que han servido para demostrar la viabilidad de esta tecnología. Por ejemplo, tal como ya se ha mencionado, se pueden citar las experiencias CYRAH y SIRAH construidos para la Confederación Hidrográfica del Júcar.

En los próximos apartados se muestra la arquitectura de un sistema experto general que modeliza el conocimiento necesario para la toma de decisiones en situaciones de emergencia. Después este concepto general se ilustra con el caso particular del sistema CYRAH mostrando su arquitectura y operación.

4.1. Descripción del conocimiento para responder a situaciones de emergencia

Como función básica, un sistema experto de ayuda en situaciones de emergencia debe predecir las crecidas y desbordamientos que se vayan a producir en los barrancos y ríos, para después proponer actuaciones que eliminen o palién los posibles daños. Los tipos de actuaciones que el sistema puede proponer son: avisos a la población, evacuaciones, cortes de carreteras con establecimiento de itinerarios alternativos, voladura de obstáculos como vías de ferrocarril o carreteras, apertura o cierre de embalses, etc.

Para llevar a cabo esta funcionalidad el sistema debe realizar una estimación realista de la relevancia y el tiempo aproximado de los problemas que se pueden esperar en áreas inundables. El análisis incluye la transformación de las entradas de lluvia en variables con mayor significado tales como nuevos niveles del río directamente asociados a problemas concretos. Esta tarea es compleja y ha de abordarse con el manejo de modelos de comportamiento del río que representan los procesos de propagación del agua a lo largo de los cauces. Asimismo, el sistema debe incluir conocimiento sobre posibles medidas de actuación para paliar los daños de problemas previsibles. En concreto, la figura 6 muestra el conjunto total de áreas de conocimiento (asociadas a las funciones a realizar) que

presenta un sistema de este tipo.

Esta figura presenta explícitamente en una estructura jerárquica los distintos bloques que modelizan el conocimiento sobre este dominio. La estructura mostrada supone una generalización y extensión de arquitecturas más específicas como son CYRAH y SIRAH. Con respecto a dichos sistemas, se ha incluido la funcionalidad de propuesta de actuaciones de control ante situaciones conflicto. Esta arquitectura genérica puede actualmente soportarse en el entorno KSM. Su formato en cada caso se rellenará con el conocimiento específico sobre el mismo mormulado bien por modelos convencionales o por modelos basados en el conocimiento al nivel de las primitivas que aparecen en la base del patrón conceptual.

En particular, la figura 6 incluye al más alto nivel un área de conocimiento representante de la aplicación completa, a la que se denomina de forma general, agente para control de emergencias. Este bloque se define internamente con tres subáreas de conocimiento: (1) problemas, que expresa el saber sobre situaciones problemáticas en el río debido a caudales o niveles altos, (2) comportamiento, que modeliza el saber sobre la evolución temporal del río, y (3) actuaciones, que recoge los métodos posibles de actuación sobre el río para paliar o solucionar los problemas derivados de las inundaciones. Asimismo, esta unidad presenta tres funciones básicas: (1) qué pasa, es decir, qué problemas hay en el presente y su evolución a corto plazo, (2) qué se puede hacer para resolver los problemas y (3) qué pasaría si se realizase una determinada actuación. Este tercer aspecto es muy útil en toma de decisiones, dado que permite validar ciertas acciones de control antes de efectuarlas.

El área de conocimiento sobre problemas, a su vez, se divide internamente en dos subáreas: (1) clases de problemas, representado con un conjunto de patrones generales de situaciones conflictivas y (2) comportamiento del río para poder predecir la evolución, representado mediante la existencia de un conjunto de modelos de comportamiento que permiten simular diversos fenómenos físicos como es la transformación de la lluvia en caudal en un barranco teniendo en cuenta las características del suelo. Dicho conocimiento permite suministrar la funcionalidad denominada focalizar que permite centrar la atención en áreas potencialmente problemáticas, es decir, puntos en donde se estén registrando lluvias altas o que presenten caudales o niveles altos. También, esta área de conocimiento proporciona la funcionalidad identificar que determina y caracteriza los problemas de inundación previstos.

El área clases de problemas dispone de un conjunto de patrones de situaciones problemáticas que permiten clasificar como problema o como libre las diversas situaciones de la cuenca. Estos patrones se encuentran almacenados en una base de conocimiento con una representación particular (como pueden ser reglas o marcos). Los patrones permiten focalizar la atención en áreas potencialmente problema e identificar situaciones de emergencia, dada una cierta situación del río. Por ejemplo, dado un cierto estado de niveles y caudales en el río, esta última funcionalidad determina la gravedad de la situación y su impacto en los núcleos de población y carreteras.

El área comportamiento físico permite realizar simulaciones del comportamiento del río para poder predecir su evolución. Para ello dispone de modelos que representan la estructura física del río (de forma numérica o cualitativa) y de procedimientos que emulan el comportamiento del sistema. Cada fenómeno físico a considerar (como la transformación de lluvia en caudal o la variación de los niveles del río) dispondrá de un modelo particular para llevar a cabo su simulación (representado en la figura por modelo 1, ..., modelo N). La presencia de este conocimiento permitirá predecir problemas así como evaluar el impacto de determinadas acciones de control.

El conocimiento sobre actuaciones se divide, por un lado, en acciones de control relativas a, por ejemplo, el cierre o apertura de embalses, cuyo efecto se puede analizar con el uso de simulaciones y, por otro lado, medidas que permitan paliar los efectos de las inundaciones, como el establecimiento de itinerarios alternativos en carreteras y métodos de evacuación en poblaciones aisladas. Este conocimiento suministra dos funciones: una denominada decisión, que determina las posibles actuaciones para paliar un problema y otra, denominada impacto, para analizar el efecto de una determinada actuación.

El conocimiento sobre acciones de control se divide a su vez en propuestas de acciones y en comportamiento físico. Saber sobre propuestas permite determinar a priori acciones que pueden resolver los problemas, y saber sobre comportamiento permite evaluar el efecto de dichas acciones para elegir la más apropiada de acuerdo con su impacto.

El área sobre propuestas de acciones contiene criterios heurísticos que determinan directamente las acciones de control candidatas para resolver los problemas de inundaciones. No obstante, se trata de acciones de tipo general que deberán después ser refinadas tras un proceso de simulación para seleccionar unas y descartar otras de acuerdo con su impacto.

Por último, la unidad denominada medidas para paliar contiene almacenado un conjunto genérico de acciones de defensa ante emergencias. Por ejemplo acciones de establecimiento de itinerarios alternativos o de evacuación de poblaciones. Este conocimiento permitirá determinar posibles actuaciones de este tipo para paliar los problemas.

4.2. El caso del sistema CYRAH

En este apartado se utiliza el ejemplo del sistema CYRAH para ilustrar las ideas generales mostradas en la sección anterior. se trata de un caso particular de sistema de ayuda a la decisión en situaciones de emergencia que permite detectar los problemas de avenidas e inundaciones que se presentan en la cuenca hidrográfica, analizando su evolución en las próximas horas. Sin embargo, este sistema no incluye la funcionalidad de propuestas de actuaciones para resolver o paliar los problemas que se ha presentado en el epígrafe anterior.

Una de las peculiaridades que presenta el sistema CYRAH es el uso de modelos numéricos de simulación para representar el conocimiento sobre comportamiento del río. Sin embargo, el manejo de modelos hidrológicos de

simulación a corto plazo para predicción de inundaciones, tiene una serie de inconvenientes que limitan su aplicación práctica, tal como hace notar [Garrote, 90]. Estas dificultades son las siguientes:

- algunos de los datos que son medidos por los sensores son inadecuados e insuficientes y deben ser objeto de evaluación y tratamiento antes de ser aceptados como entrada de los modelos,
- la complejidad de los procesos naturales que tienen lugar en una cuenca durante una tormenta está representada sólo de forma parcial por los modelos,
- hay puntos de incertidumbre respecto a las propiedades físicas del río, nunca se dispone de los parámetros físicos con la exactitud que piden los modelos,
- para un cierto estado, en general, se puede esperar más de una evolución posible que debe ser tenido en cuenta,
- las salidas de los modelos deben ser convenientemente interpretadas para llegar a conclusiones con significado.

Por tanto, el éxito del control con el uso de modelos hidrológicos en tiempo real depende de la intervención de conocimiento de expertos hidrólogos, dado que, en primer lugar, hay una separación entre la información que requieren los modelos y los datos que proporcionan las redes de medida con los sensores y, en segundo lugar, la salida de simulación debe ser interpretada adecuadamente para determinar las acciones de control. Como resultado, para manejar modelos numéricos el sistema debe incluir:

- por un lado, los propios modelos numéricos construidos de una forma eficiente para ser utilizados en tiempo real,
- por otro lado, el conocimiento de experto, expresado de forma simbólica, para realizar el aprovechamiento de forma efectiva de los modelos. Este conocimiento incluye:
 - a) criterios de control para preparar los datos de entrada de las simulaciones de acuerdo con la información existente.
 - b) criterios de interpretación y síntesis de los resultados de las simulaciones para disponer de información al adecuado nivel de abstracción que requiere el proceso de toma de decisiones.

De acuerdo con estas consideraciones, la arquitectura cognitiva del sistema CYRAH es la que se muestra en la figura 7. Esta figura, es una particularización de la estructura general mostrada en la figura 6. Comparando ambas figuras se observa, por ejemplo, que la unidad de más alto nivel (denominada en este caso CYRAH como caso particular de agente para control de emergencias), incluye la funcionalidad denominada "qué pasa" para identificación de problemas en el presente y el futuro cercano, pero no incluye, tal como ya se mencionó, la funcionalidad derivada de la existencia del conocimiento sobre actuaciones para paliar los problemas.

En cuanto a los modelos de simulación para modelizar el conocimiento sobre comportamiento de río, se incluyen tres: uno para meteorología, otro para barrancos y otro para áreas inundables. El modelo de meteorología permite predecir evoluciones posibles sobre precipitaciones a corto plazo de forma cualitativa. Para realizar este proceso hay un área de conocimiento denominada lluvia que utiliza una representación en reglas para recoger criterios predictivos. Esta unidad infiere posibles predicciones de lluvia a corto plazo en las diferentes áreas espaciales consideradas, proponiendo un conjunto de posibles valores afectados por grados de incertidumbre. Esos valores representan creencias personales de las personas responsables cuando se enfrentan con un problema de inundación en tiempo real. Las premisas para el razonamiento predictivo son la lluvia total registrada en las últimas 24 horas y la predicción cualitativa proporcionada por el servicio meteorológico. La base de conocimiento utiliza reglas para establecer relaciones razonables de consistencia temporal y espacial. Ejemplos generales de dichas reglas son (obsérvese los grados de certeza en las afirmaciones):

SI en el area A, la lluvia en T es X
ENTONCES (0.75) en T+1 la lluvia en A será menor de K. X

SI en el area A, la lluvia en T es X
ENTONCES (0.5) en T+1, la lluvia en el area vecina B será menor de K. X

El modelo de barrancos utiliza un simulador denominado SAVEL que utiliza una descripción física del río formada por un árbol con áreas receptoras de lluvia, tramos de transporte y embalses, con parámetros que caracterizan físicamente cada uno de estos elementos. El proceso de simulación recibe un escenario de lluvia (en forma de series temporales) y caudales de entrada en barrancos aguas arriba, además de un estado inicial sobre la humedad del suelo y volumen en los embalses. Como resultado, el simulador predice la evolución temporal de caudales en los barrancos aguas abajo. Este procedimiento está basado en los modelos tradicionales existentes en hidrología como es el programa HEC-1 de US Corps of Engineers y el SWMM del Environmental Protection Agency (EPA). La teoría de este tipo de modelos puede encontrarse en [Chow, Maidment, Mays, 88].

Para establecer el control de estas simulaciones numéricas se utiliza una unidad complementaria que recoge los criterios de experto para (1) generar los escenarios de simulación preparando los juegos de datos correspondientes y combinando adecuadamente escenarios de lluvia de acuerdo con las predicciones y (2) sintetizando los datos numéricos resultantes para obtener valores resumen.

Por otro lado, el modelo de áreas inundables utiliza el simulador denominado REBOLSA que concibe la estructura del río como una secuencia de perfiles y un conjunto de bolsas laterales inundables conectadas con el eje principal por ciertos puntos. El proceso de simulación recibe como datos un escenario de caudales entrada en ciertos puntos del río y un estado inicial de niveles del río en los perfiles y en las bolsas inundables. Como resultado, el simulador responde con una predicción de los niveles del río a lo largo del tiempo en puntos significativos. El procedimiento de simulación realiza una integración numérica de las ecuaciones de Saint Venant. A dichas ecuaciones se ha añadido un conjunto de ecuaciones representativas de las bolsas inundables [Chow, Maidment, Mays, 88].

Al igual que el modelo de barrancos, el modelo de áreas inundables tiene también una unidad de control que incluye los criterios de control necesarios para: (1) generar las combinaciones de caudal en los afluentes que sean pertinentes para realizar las simulaciones, (2) seleccionar el estado inicial de los niveles del río y (3) sintetizar los valores numéricos resultantes para proporcionar medidas con mayor nivel de abstracción.

Finalmente, el área de conocimiento sobre clases de problemas permite focalizar e identificar los problemas en los diversos horizontes temporales sobre accesibilidad de poblaciones o cortes de carreteras. Para ello utiliza un conocimiento clasificativo expresado en reglas. La aplicación de la cuenca hidrográfica del Júcar cuenta con un total aproximando de 3000 reglas con la forma que expresan las reglas generales que se muestran a continuación:

SI caudal en punto P > valor V
ENTONCES área problema P está incipiente

SI nivel en bolsa X > cota C
ENTONCES area problema P está inundada

SI area problema P está inundada
ENTONCES estado interior del núcleo de población N es grave

SI area problema P está inundada
ENTONCES tramo de carretera T cortado

SI tramo de carretera T1 cortado y
tramo de carretera T2 cortado
ENTONCES accesibilidad de núcleo de población N es parcial

4.3. Ejemplo de operación

Para ilustrar la funcionalidad de un sistema de ayuda a la decisión en situación de emergencias se utilizará de nuevo el ejemplo CYRAH en la cuenca hidrográfica del Júcar, que opera en el centro de control de Valencia recibiendo datos de sensores cada minuto. Tal como se ha dicho, este ejemplo ilustra adecuadamente la capacidad de detección y explicación de problemas por inundaciones aunque no incluye la funcionalidad para generación de propuestas de acciones de control que se ha presentando en el enfoque general.

El sistema CYRAH razona a petición del operador y tras un tiempo entre 5 y 15 minutos contesta qué problemas de inundación o avenidas se dan o pueden darse en las próximas 4 horas. Para ello, utiliza una presentación gráfica resumen de la situación en cada una de las subcuencas que muestra con un código predefinido de colores la gravedad de las diversas zonas. Un ejemplo de esta pantalla se muestra en la figura 8.

Asimismo, el operador puede consultar información detallada sobre los problemas en lugares específicos tal como se observa en las figuras siguientes. En dichas pantallas se presenta una estimación en los diversos instantes de tiempo (la próxima hora, en 2 horas, en 3 horas, etc.) de la accesibilidad de poblaciones, su estado y el estado de las carreteras, lo que permite disponer de criterios suficientes para trazar métodos de defensa ante la emergencia.

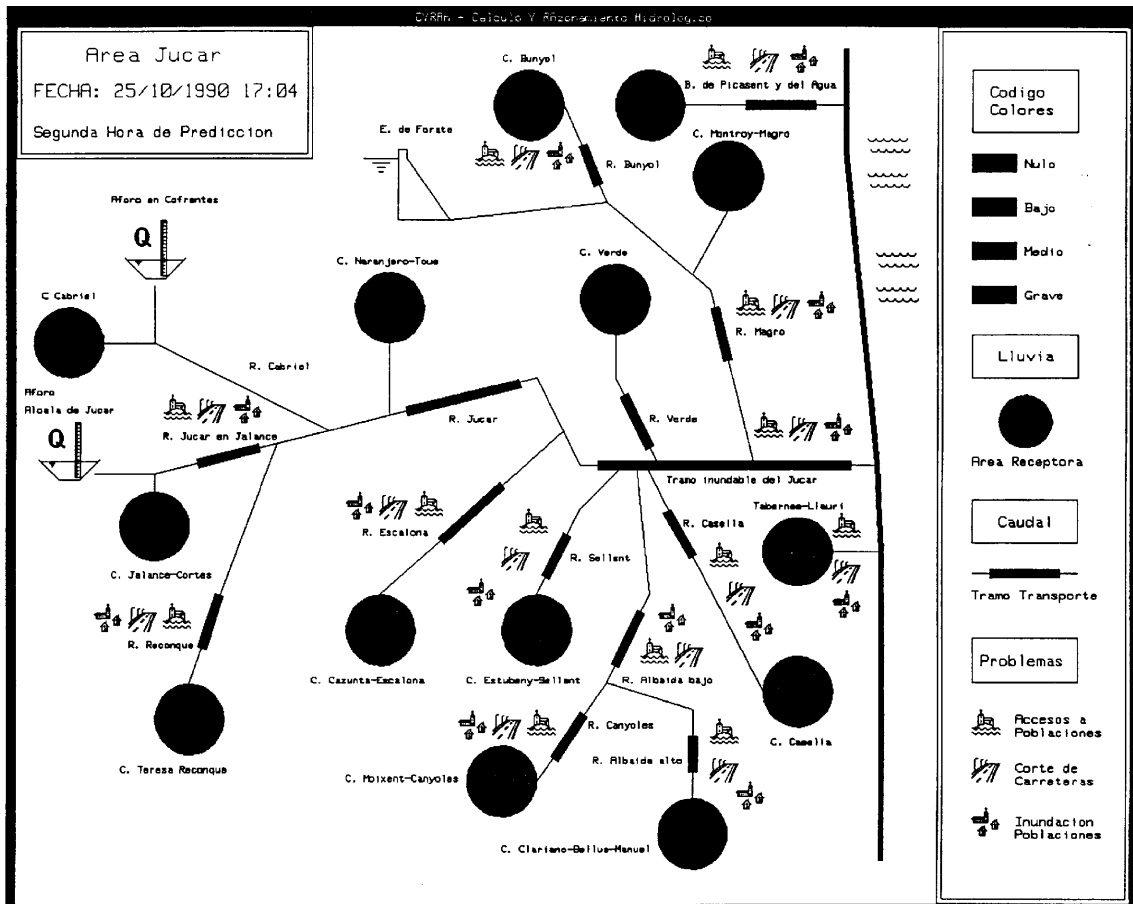


Figura 8: Ejemplo de presentación gráfica resumen de la situación hidrológica del sistema CYRAH

Finalmente, el operador puede también consultar explicaciones sobre cómo el sistema ha alcanzado estas conclusiones, observando valores intermedios tales como los siguientes:

- lluvia medida en mm en la última hora:
 - en Albaida/Canyoles: 18mm
 - en Ribera: 20mm
 - ...
- predicción meteorológica:
 - en Albaida/Canyoles: lluvia media o lluvia alta
 - en Ribera: nula
 - ...
- valores previstos de caudal en barrancos:
 - Albaida/Canyoles: 350-400 m³/s (hora 1), 400-550 m³/s (hora 2), ...
 - Albaida Bajo: 200-250 m³/s (hora 1), 350-450 m³/s (hora 2), ...
 - Magro: 100-150 m³/s (hora 1), 150-250 m³/s (hora 2), ...
 - ...
- valores previstos de nivel en ríos:
 - Júcar (perfil 1): 3-4 m (hora 1), 4-5 m (hora 2), 4-5 m (hora 3), ...
 - Júcar (perfil 2): 4-5 m (hora 1), 4-6 m (hora 2), 4-7 m (hora 3), ...
 - ...
- estado de bolsas inundables:
 - bolsa Alberique: libre (hora 1), incipiente (hora 2), inundada (hora 3), ...
 - bolsa Onteniente: libre (hora 1), libre (hora 2), libre (hora 3), ...
 - ...

Este aspecto explicativo, como se ha comentado, es muy importante en situación de toma de decisiones dado que la persona encargada del centro podrá asumir la responsabilidad de la acción de control tras una comprensión total de la situación.

ACCESIBILIDAD DE POBLACIONES (en dos horas)
Prevision CYRAH de 19/11/1990 13:30

Adzaneta de Albaida	parcial <input type="checkbox"/> (casi seguro)
Albaida	parcial <input type="checkbox"/> (casi seguro)
Alcudia de Crespins	parcial <input type="checkbox"/> (casi seguro)
Belgida	parcial <input type="checkbox"/> (casi seguro)
Benimuslem	parcial <input type="checkbox"/> (posiblemente)
Bufali	parcial <input type="checkbox"/> (casi seguro)
Canals	parcial <input type="checkbox"/> (casi seguro)
Cerda	parcial <input type="checkbox"/> (casi seguro)
Genoves	parcial <input type="checkbox"/> (casi seguro)
Lugar Nuevo de Fenollet	parcial <input type="checkbox"/> (casi seguro)
Manuel	parcial <input type="checkbox"/> (casi seguro)
Masalaves	parcial <input type="checkbox"/> (posiblemente)
Mogente	parcial <input type="checkbox"/> (casi seguro)
Montaberner	parcial <input type="checkbox"/> (casi seguro)
Onteniente	parcial <input type="checkbox"/> (casi seguro)
Palomar	parcial <input type="checkbox"/> (casi seguro)

ESTADO DE POBLACIONES (en la proxima hora)
Prevision CYRAH de 19/11/1990 13:30

Albaida	grave (casi seguro)
Alberique	con problemas incipientes (posiblemente)
Benimuslem	con problemas incipientes (posiblemente)
Canals	gave (casi seguro)
Genoves	gave (casi seguro)
Manuel	gave (casi seguro)
Onteniente	gave (casi seguro)
Senera	gave (casi seguro)
Vallada	gave (casi seguro)
Xativa	con problemas incipientes (posiblemente)

CARRETERAS CORTADAS (en tres horas)
Prevision CYRAH de 19/11/1990 13:30

C_320 entre V_V_2014 y V_P_2016	casi seguro
C_322 entre Xativa y Genoves	casi seguro
C_3316 entre Canals y V_V_2082	casi seguro
C_3316 entre Onteniente y V_V_2034	casi seguro
C_3320 entre V_V_1031 y Manuel	casi seguro
C_3324 entre Benimuslem y Alzira	posiblemente
N_340 entre Albaida y Palomar	casi seguro
N_340 entre Masalaves y Alberique	posiblemente
N_340 entre Montaberner y V_P_2039	casi seguro
N_430 entre Cerda y Xativa	casi seguro
N_430 entre V_V_3053 y V_V_2018	casi seguro
V_P_1036 entre Xativa y Lugar Nuevo de Fenollet	casi seguro
V_P_2016 entre V_V_2017 y C_320	casi seguro
V_V_1025 entre N_340 y V_V_1026	casi seguro
V_V_2014 entre C_320 y C_3316	casi seguro
V_V_2018 entre N_430 y Vallada	casi seguro

5.- CONCLUSIONES

Lo comentado en los epígrafes anteriores muestra, por un lado, el interés de que las aplicaciones para manejos de sistemas en tiempo real con alto aprovechamiento de sus potencialidades se haga sobre modelos abiertos en que se explicita su contenido conceptual. Por otra parte, se ha comprobado cómo es posible, para dos clases de modelos que tienen gran interés para el programa SAIH, construir modelos conceptuales utilizando arquitecturas avanzadas de estructuración cognitiva en las que coexisten tanto módulos formulados con bases de conocimiento como módulos formulados en forma tradicional via algoritmos cuando se trata de procesos ya instalados en el *know how* habitual de los profesionales. Por ello se da la doble circunstancia de factibilidad tecnológica de este tipo de solución y de oportunidad para su uso, ya que en el momento actual estos conceptos empiezan ya a ser la forma adecuada de construir aplicaciones con sentido moderno en el mundo industrial y no un tema de simple investigación académica.

6.- REFERENCIAS

- [Alonso et al. 90] Alonso M., Cuenca J., Molina M.: "SIRAH: An Architecture for a Professional Intelligence". Proc.9th European Conference on Artificial Intelligence (ECAI'90). Pitman, 1990.
- [Alonso et al. 92] Alonso M., Cuenca J., Reig B.: "SIRAH: A Software Environment for Advanced Knowledge Based Models for Flood Management" formando parte de *Hydraulic Engineering Software HYDROSOFT'92*. Computational Mechanics Publications. Wessex Institute of Technology, 1992.
- [Chandrasekaran, 87] Chandrasekaran, B. y Punch, W.F. III: Data validation during diagnosis, a step beyond traditional sensor validation. *Proceedings AAAI'87*. Seattle (1987) 778-782.
- [Chandrasekaran, 92] Chandrasekaran, B.; Johnson, Todd R.; Smith, Jack W.: Task-structure analysis for knowledge modeling. *Communications of the ACM* 35:9, (1992) 124-136.
- [Chow, Maidment, Mays, 88] Chow V.T., Maidment D.R., Mays L.W.: "Applied Hydrology" Mc Graw Hill, 1988.
- [Cuenca, Molina, 94] Cuenca, J., Molina, M.: "KSM: An Environment for Knowledge Oriented Design of Applications Using Structured Knowledge Architectures" in "Applications and Impacts. Information Processing'94", Volume 2. K. Brunnstein and E. Raubold (eds.) Elsevier Science B.V. (North-Holland), 1994 IFIP.
- [Cuenca et al, 91] Cuenca J., Molina M., Garrote L.: "An Architecture for Cooperation of Knowledge Bases and Quantitative Models: The CYRAH Environment". XI International Workshop on Expert Systems. Special conference on Second Generation Expert Systems. Avignon'91. EC2, 1991.
- [DeKleer, 86] De Kleer, J: An Assumption Based T.M.S. *Artificial Intelligence* 28, (1986) 127-162 y también en Ginsberg, M.L. (ed.): *Readings in Nonmonotonic Reasoning*, Morgan Kaufman Publisher, Los Altos, California, (1987) 280-297.

- [DeKleer, 87] De Kleer, J., y Williams, B.C.: Diagnosing Multiple Faults. *Artificial Intelligence* 32, (1987) 97-130 y también en Ginsberg, M.L. (ed.): *Readings in Nonmonotonic Reasoning*, Morgan Kaufman Publisher, Los Altos, California, (1987) 372-388.
- [Garrote, 90] Garrote L.: "Modelos Hidrológicos de Ayuda a la Decisión en Tiempo Real Basados en Técnicas de Inteligencia Artificial". Tesis doctoral. Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos. Universidad Politécnica de Madrid. Madrid, 1990.
- [Molina, 93] Molina, M.: Desarrollo de Aplicaciones a Nivel Cognitivo mediante entornos de Conocimiento Estructurado. *Tesis Doctoral*. Facultad de Informática. Universidad Politécnica de Madrid. Madrid. (1993)
- [OMM, 84] Organización Meteorológica Mundial: *Guía de Prácticas Hidrológicas. Volumen I: Adquisición y Proceso de Datos*. OMM-168. Secretaría de la Organización Meteorológica Mundial. Ginebra. Suiza (1984).
- [Ray, 86] Ray, A. y Desai, M.: A Redundancy Management Procedure for Fault Detection and Isolation. *Journal of Dynamic Systems, Measurement, and Control* 108 (1986) 248-254.
- [Watts, 93] Watts, P.J. y Knight, B.: A knowledge-based system for the quality control of telemetry data in water distribution networks. En Topping, B.H.V. (ed.): *Knowledge Based Systems for Civil and Structural Engineering*. Civil-Comp Press. Edimburgh, (1993) 135-143.