

METODOLOGÍA PARA LA CALIBRACIÓN DE MODELOS DE CALIDAD DE AGUAS

P. Amparo López Jiménez¹, Vicent Espert Alemany², Mar Carlos Alberola³ y F. Javier Martínez Solano¹

Resumen:

Se presenta una metodología para calibrar modelos de dispersión de contaminantes en cauces receptores unidimensionales (modelos de calidad de aguas). Esta metodología está basada en la diferenciación de las variables que afectan a dicha calibración en parámetros internos al modelo (coeficientes semiempíricos que intervienen en las ecuaciones y proceden de la bibliografía); y parámetros externos o mediciones en el medio representado, que también intervienen en las ecuaciones del modelo. Para los parámetros internos se realiza una búsqueda mediante la técnica del algoritmo genético y los parámetros externos se considera que afectan a la precisión de los resultados mediante el análisis de incertidumbre. Este análisis de incertidumbre permite conocer la varianza de los resultados calculados y definir un criterio objetivo para determinar si el proceso de calibración ha terminado.

Para avalar el modelo y metodología, se presentan los resultados de la modelación y calibración de los procesos ligados a la dinámica del Oxígeno Disuelto en un cauce de la Comunidad Valenciana. El ajuste que experimenta el modelo con la realidad que se mide, tras haber implementado la sistemática de calibración propuesta, es muy satisfactorio y abre un campo hacia las posibilidades de autocalibración de los modelos de calidad de aguas.

Palabras clave: Modelo de calidad de aguas, dinámica del Oxígeno Disuelto, calibración y validación, algoritmo genético, análisis de sensibilidad e incertidumbre, bandas de confianza.

INTRODUCCIÓN

La utilización de modelos matemáticos para simular los procesos de transporte y dispersión de contaminantes vertidos en medios receptores ha experimentado un auge creciente en las últimas décadas. Esto es debido al gran interés que posee el estudio de la contaminación producida por las aguas de vertido y su posible impacto en el medio receptor ya que afecta al desarrollo de las actividades humanas y en general a la calidad del medio ambiente.

La conveniencia del uso de modelos matemáticos para el análisis de estos fenómenos de dispersión de contaminantes o simulación de escenarios hipotéticos es innegable. Los actuales medios computacionales permiten, mediante la resolución de las ecuaciones que representan el fenómeno físico, generar simulaciones que proporcionan, una vez calibradas y validadas, fidedignas representaciones de la realidad.

En este proceso hay que destacar que la correcta utilización de cualquier modelo (no solamente el de dispersión de contaminantes) para la toma de decisiones está condicionada a que el proceso de calibración del mismo haya sido realizado correctamente. Si esto es así, el modelador representará la realidad con un grado aceptable de fiabilidad, ya que el objeto de la calibración es asegurar la valía del modelo como herramienta de toma de decisiones.

El proceso de calibración del modelo no está exento de dificultades, entre las que pueden detallarse las siguientes (Robinson, 1999): no se conoce un umbral que determine la precisión que se requiere de los resultados; no hay disponibilidad de datos con los que comparar (o los datos de los que se dispone no son suficientemente precisos), o no hay tiempo y recursos para la calibración. Por ello se hace necesaria una estrategia lo más estructurada y científica posible para que dicho proceso de calibración aproveche al máximo los recursos.

¹ Grupo Multidisciplinar de Modelación de Fluidos. Instituto Tecnológico del Agua. Universidad Politécnica de Valencia.

Tel.: 96 387 98 90 Fax: 96 387 79 81 E-mail: palopez@gmmf.upv.es jmsolano@gmmf.upv.es.

² Instituto Tecnológico del Agua. Universidad Politécnica de Valencia. Tel.: 96 387 98 98 Fax: 96 387 98 99 E-mail: vespert@ita.upv.es

³ Instituto Tecnológico del Agua. Organismo Público Valenciano de Investigación. Generalitat Valenciana.

Tel.: 96 387 98 98 Fax: 96 387 98 99 E-mail: macaral0@ita.upv.es

Artículo recibido el 4 de junio de 2002, recibido en forma revisada el 30 de enero de 2003 y aceptado para su publicación el 2 de junio de 2003. Pueden ser remitidas discusiones sobre el artículo hasta seis meses después de la publicación del mismo siguiendo lo indicado en las "Instrucciones para autores". En el caso de ser aceptadas, éstas serán publicadas conjuntamente con la respuesta de los autores.

Estas premisas que son de utilidad en el caso general, también son de aplicación para el modelo matemático de dispersión de contaminantes en particular. En concreto, en el presente artículo se pretende mostrar las características generales del modelo de dispersión de contaminantes en cauces receptores unidimensionales aplicado al Oxígeno Disuelto (OD en adelante), proponiendo una metodología para su calibración basada en el análisis de incertidumbre y los algoritmos genéticos.

Tradicionalmente el problema no ha sido enfocado de esta manera, de una parte no se ha planteado distinción entre los parámetros que intervienen procedentes de mediciones como condiciones de contorno y los que son coeficientes de cierre de las ecuaciones (los considerados parámetros internos y externos en la calibración). De otro lado, para la calibración de los modelos no es tradicional la utilización de técnicas heurísticas como los algoritmos genéticos sino que han sido utilizadas técnicas de ajuste como los ajustes polinómicos, los de mínimos cuadrados u otras técnicas matemáticas que minimizan la función de error (Goldberg, 1989).

EL PLANTEAMIENTO DEL PROBLEMA.

El modelo de dispersión de contaminantes aplicado al OD.

El problema que se desea modelar es la evolución que sufren los diferentes indicadores de la calidad del agua de un medio receptor natural una vez que se ha realizado un vertido de contaminantes al mismo. Estos modelos, denominados también de calidad de aguas, constan de varios módulos (Chapra, 1997):

- El módulo hidrodinámico, con el que se desea conocer las velocidades puntuales del medio receptor resolviendo las ecuaciones de continuidad y Navier-Stokes (entre otras que a veces resultan necesarias).
- El módulo de transporte y dispersión que estudia los fenómenos que ocurren en el seno del fluido en el proceso de mezcla de las partículas de contaminante con las del medio receptor. Este transporte será convectivo, por difusión molecular o por dispersión turbulenta.
- El módulo fisicoquímico y/o biológico integrará las ecuaciones de la cinética de cada uno de los contaminantes con el resto de los fenómenos.

En este caso se estudiará la evolución del OD y los compuestos presentes en el agua, que marcan la calidad de la misma y que se ven afectados por la evolución de dicho oxígeno.

El OD es uno de los elementos imprescindibles para el desarrollo de diferentes procesos esenciales para el desarrollo del ecosistema acuático, por lo que su disminución por debajo de ciertos niveles umbral puede alterar profundamente el equilibrio del mismo. La utilización y posible reuso que se pueda dar al agua del cauce receptor depende en gran medida de la concentración de OD que haya en la misma. Estas, como razones más importantes, han hecho del OD el indicador de la calidad de las aguas por excelencia y, en la mayor parte de los cauces, su mejor indicador de su estado.

Sin embargo, el oxígeno como tal indicador de calidad presenta algunas limitaciones. Por un lado es el resultado de una gran cantidad de interacciones biológicas y fisico-químicas que ocurren en el cauce, por lo que es un buen indicador de la calidad del mismo. Pero eso mismo le hace tener una gran cantidad de interdependencias que traen como consecuencia que una buena modelación dependa de muchos aspectos (lo que complica en gran manera la calibración de los modelos), puesto que una consideración completa de todos los fenómenos es prácticamente imposible.

Así pues, éste es el modelo que se desea calibrar. En tanto que modelo matemático, la representación se realizará mediante ecuaciones que deberán ser resueltas para cada instante y punto del medio receptor con las adecuadas condiciones iniciales y de contorno. Estas ecuaciones serán:

- En el módulo hidrodinámico las ecuaciones de continuidad y Navier-Stokes. En el caso de que se trate de un cauce receptor, las ecuaciones del movimiento unidimensional a lámina libre serán las adecuadas (ecuaciones de Saint Venant).
- En el denominado módulo de transporte, la ecuación que describe la evolución espacial y temporal de la concentración de cualquier sustancia contaminante vertida a un cauce receptor es la ecuación general de la dispersión turbulenta, o ecuación del transporte turbulento de masa.
- Para el módulo de calidad aplicado al OD existen muchos aspectos que deberán ser considerados,

representados por sendas ecuaciones cada uno de ellos. En el modelo que aquí se desea calibrar se consideran los siguientes:

- aporte de OD por reaireación
- consumo de OD por degradación de materia orgánica carbonatada en suspensión
- demanda de OD por sedimentos
- consumo de OD en los procesos de nitrificación
- balance de OD debido a procesos de fotosíntesis y respiración de algas

La resolución de estas ecuaciones proporcionará las velocidades y concentraciones en los puntos del cauce receptor para cada instante de tiempo; estos valores constituyen las variables de salida de interés para el modelador.

En dichas ecuaciones intervienen unos parámetros (semiempíricos en la mayoría de los casos) que determinan los resultados de las mismas. Dando a estos parámetros unos valores u otros se obtienen diferentes resultados. El objeto de la calibración es determinar el valor de dichos parámetros que hace que los resultados sean lo más reales posible. Así, estos parámetros se convierten en valores de entrada para la modelación.

Naturaleza de los parámetros que intervienen en las ecuaciones.

En cada uno de los módulos del modelo nos habremos encontrado con dos tipos de parámetros de entrada que afectan a las variables de salida:

- Parámetros “externos” al modelo matemático que estamos calibrando. Estos serán datos para las condiciones de contorno, y deberán ser determinados mediante experimentación física. Son considerados así, por ejemplo, las concentraciones de entrada en la cabecera de las líneas o los caudales aportados en el régimen hidrodinámico. En realidad sobre el valor de estos parámetros externos no se tiene control puesto que provienen de la experimentación. Sin embargo, sí que se puede conocer cuál es el efecto que sobre la sensibilidad del modelo tendrá la incertidumbre de su conocimiento y estimar el orden de importancia que estos aspectos presentan en los fenómenos que se estudian.

- Parámetros “internos” del modelo, que intervienen en la formulación de las ecuaciones hidrodinámicas, cinéticas y de transporte, y que son fuente de error si su consideración no es la adecuada para cada caso. Estos parámetros son conocidos de manera semiempírica y se encuentran comprendidos en unos intervalos aceptados por la comunidad científica (por ejemplo Bowie y Mills, 1985). Como muestra, entre estos valores se encuentra, por ejemplo, el coeficiente de reaireación o los de velocidad de consumo de OD en los distintos procesos químicos o biológicos que se consideren en el modelo.

La estimación de los parámetros internos no tiene más coste que el del tiempo del experto que se encuentra modelando el caso. La determinación de su grado de influencia sobre la precisión de los resultados puede ahorrar mucho tiempo en el proceso de ajuste. Dado que se trata de parámetros no afectados por los procesos de la medición sino que tienen que ver con el modelo matemático, su estimación está exenta de incertidumbre en el sentido estadístico y podrán ser determinados mediante procesos de optimización matemática.

METODOLOGÍA PROPUESTA PARA LA CALIBRACIÓN DEL MODELO

En la Figura 1 (ver página siguiente) se presenta la metodología de calibración para el modelo de dispersión de contaminantes aplicado al OD. A continuación se detallan las fases de la misma.

Metodología de calibración

La calibración comienza ya en la fase de conceptualización del problema al determinar cuáles son las variables de salida de interés para el modelador y cuáles los parámetros de entrada que afectan a los mismos.

A continuación deberá determinarse la influencia concreta que cada parámetro de entrada tiene en las variables de salida: para un determinado incremento de valores de entrada, podremos conocer cuál es el efecto relativo que tiene dicha variación en cada una de las variables de salida.

El conocimiento del efecto relativo es consecuencia de un análisis de sensibilidad de primer orden. Supóngase que se está evaluando una función cualquiera $C(k)$, (por ejemplo la concentración media

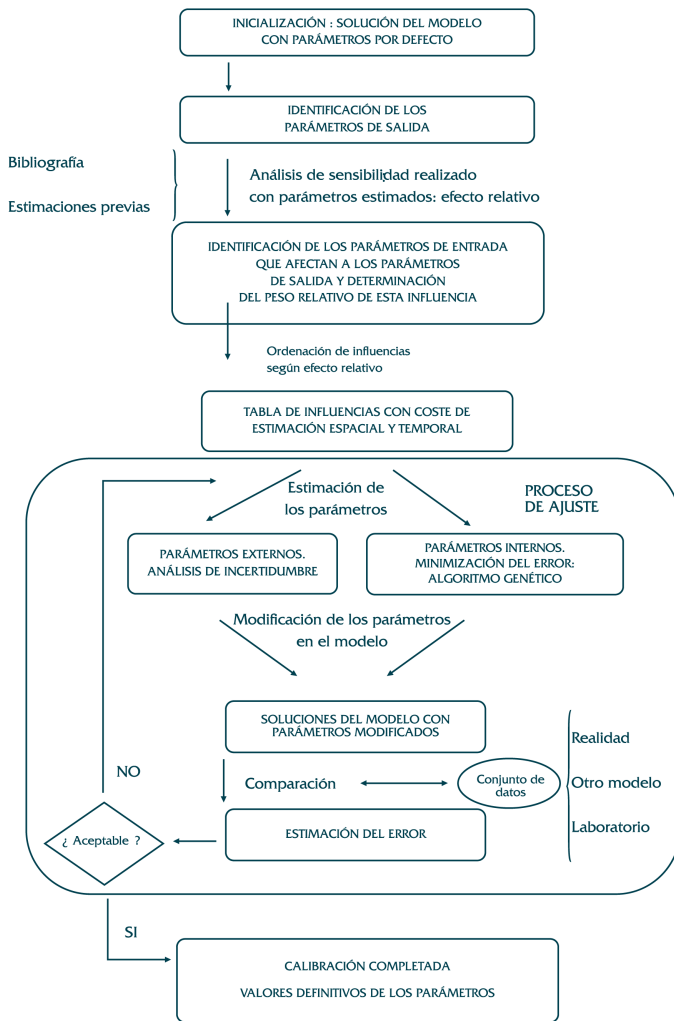


Figura 1. Metodología completa para la calibración del modelo

de OD), y se está considerando la influencia que sobre ella tiene el valor k de un determinado parámetro (por ejemplo el coeficiente de velocidad de reai-reacción).

Así, se trata de conocer el efecto relativo que tiene una variación del valor k del parámetro en la función, de manera que se relaciona los valores de $C(k+\Delta k)$ y $C(k-\Delta k)$, con el valor centrado de la función $C(k)$, observando así la variación porcentual de la función que se ha experimentado como consecuencia de la variación relativa del valor k . El efecto relativo así definido se muestra en la expresión (1) y se representa gráficamente en la Figura 2.

$$ER = \frac{\frac{C(k + \Delta k) - C(k - \Delta k)}{2C(k)}}{\frac{\Delta k}{k}} \quad (1)$$

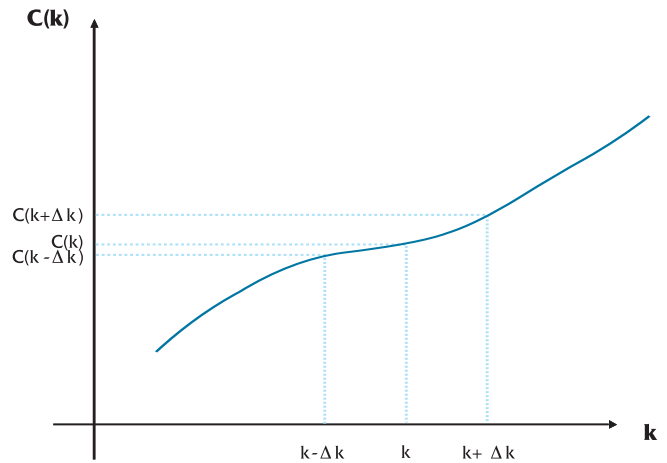


Figura 2. Elementos para definición del efecto relativo

De esta manera el efecto relativo (ER) será representativo de la influencia de los parámetros de entrada sobre los de salida y servirá como indicador para una ordenación de los parámetros según estas influencias. Los parámetros que más afecten a los resultados deberán ser evaluados con mayor dedicación de recursos. De otro lado, el ER es uno para cada posición espacio-temporal y se escoge el máximo de estos valores como más representativo.

Asimismo, y puesto que los valores se definen para cada línea en los modelos unidimensionales, también puede hacerse un análisis de sensibilidad “local”, variando los parámetros de una región determinada y observando el peso que tienen en el conjunto del dominio.

Realizado este análisis de sensibilidad y determinados los efectos relativos de todos los parámetros de entrada sobre los de salida, se debe proponer una lista ordenada de cuáles son los parámetros de entrada, por orden de importancia, que más afectan al parámetro de salida, estimados según el efecto relativo. Esto se realizará a la vista de los resultados que se han obtenido del análisis de sensibilidad.

Una vez conocido el orden en el cual afectan los parámetros al modelo en su conjunto, o en alguna parte del mismo, es muy importante tener en cuenta cuál es el coste aproximado de este ajuste, ya que puede ocurrir que parámetros que influyen poco, sean caros de evaluar. Esto ayudará a establecer una estrategia de actuación en el proceso de calibración.

En este sentido, aunque la precisión de las medidas es difícil de parametrizar, es posible hablar de que un número mayor de ensayos dará más precisión a los resultados, de manera que el conocimiento del coste de la estimación de un parámetro externo puede dar una idea del coste de la precisión de los cálculos. Sin embargo, esto tiene un límite ya que llega un momento en que no por aumentar el tamaño de la muestra se mejora la precisión puesto que no se disminuye la desviación típica de los resultados. Por otra parte, los parámetros “internos” del modelo no tienen coste alguno de estimación física, sino que generan un gasto de recursos y de tiempo del modelador que puede ser importante. Este gasto de tiempo es el que se pretende reducir con el uso del algoritmo genético y el análisis de incertidumbre.

En la metodología aquí propuesta debe realizarse un proceso de ajuste del modelo a valores de referencia hasta que el modelador decida cuál es el valor de los parámetros de entrada que mejor predice las concentraciones con las que se va a comparar. Para ello debe disponerse de un criterio de terminación de manera que su cumplimiento muestre que la calibración ha terminado y se ha conseguido un nivel de error aceptable.

Los valores de los parámetros obtenidos en la calibración de un caso determinado deberán ser válidos para modelar otras situaciones. Esto permitirá que el modelo se convierta en una herramienta útil para la futura toma de decisiones.

El proceso de ajuste es el más complejo de la metodología propuesta y en el que el modelador necesita la mayor ayuda. Por ello se describe a continuación con un nivel mayor de detalle.

EL PROCESO DE AJUSTE

Se trata éste de un proceso iterativo, con el que se pretende que el error que se comete con el cálculo de los valores de salida del modelo y aquellos con los que se compara, sea el menor posible con un coste razonable.

Para inicializar el proceso, es necesario partir de una estimación de los parámetros internos de entrada que intervienen en el modelo. Por lo descrito en las fases previas es conocida la identidad de estos parámetros e incluso la influencia en la variación de las concentraciones y temperaturas (entendidas como variables de salida) y el coste de determinación de los mismos.

En este momento los parámetros adoptan unos valores por defecto para que el modelo numérico funcione. La primera vez estos valores son tomados de la bibliografía o dentro de intervalos conocidos de variación de los mismos establecidos en la experimentación.

La influencia de los parámetros en el proceso de calibración es diferente si son internos o externos al mismo, tal como han sido definido en secciones anteriores.

Los parámetros internos tienen un rango de variación conocido y existen unas indicaciones sobre su valor para cada caso. Sin embargo, pueden ser seleccionados dentro de estos intervalos con un cierto grado de libertad. Para hacerlo, tradicionalmente se ha recurrido a métodos ensayo-error en los que variando los valores de dichos parámetros, el modelador va comprobando la bondad del ajuste del modelo. Como alternativa a ello, se presentará a continuación una metodología de estimación de parámetros basada en el uso de algoritmos genéticos, completamente novedosa en modelos con tantos parámetros.

Por otro lado, la identificación de los parámetros externos de entrada viene determinada por un proceso de medida sujeto al error de la experimentación. El valor de los mismos, por tanto, no es uno solo en buena lid, sino que se trata de una distribución en el sentido estadístico, puesto que así son los datos procedentes de la experimentación. No se puede estimar el valor de los mismos puesto que provienen de la medición, por tanto solamente puede estudiarse la influencia que tiene el error de la medida en el error de las predicciones del modelo mediante el análisis de incertidumbre.

A continuación se describe cómo se integran en la metodología de calibración propuesta la búsqueda de parámetros internos y el análisis de incertidumbre provocado por los parámetros externos que intervienen en la modelación.

Valores a asignar a los parámetros internos del modelo mediante el uso de algoritmos genéticos

El problema que se pretende solucionar es un problema de optimización matemática: se desea conocer el valor de los parámetros que intervienen en las ecuaciones que definen el modelo. De esta manera se minimiza el error respecto a los valores procedentes de las mediciones y que sirven de referencia en este proceso de calibración.

Para ello se plantea la posibilidad de utilizar el algoritmo genético, conocido ya desde hace varias décadas (Holland, 1975; Goldberg, 1989; Engelhardt y Dandy, 1999), como alternativa a las metodologías tradicionales de minimización. Esta técnica ya ha sido utilizada para procesos de búsqueda de parámetros ligada a otros campos de la ingeniería (Beasley et al. 1993), e incluso en modelos de calidad de aguas (Mulligan y Brown, 1998), aunque en nuestro caso particular el algoritmo hace una búsqueda con una docena de parámetros, lo cual es toda una novedad en este tipo de aplicaciones. La técnica del algoritmo genético ha sido utilizada con éxito en otros experimentos de calibración de calidad de aguas (Rocha (1997), Ferreira (2002), sin embargo su interacción con las ecuaciones del modelo del OD aquí presentado y su extrapolación a problemas en régimen no permanente con experiencias contrastadas (López, 2001), validan la novedad del presente trabajo.

El objeto del algoritmo genético es la determinación de qué individuos de una familia deben sobrevivir, cuáles reproducirse y cuáles no. El vector que compone esos parámetros es el individuo, y el valor de cada uno de ellos forma en su conjunto los genes (o genoma) que definen la naturaleza del individuo. Un conjunto de vectores de parámetros l_i , de individuos, definen una familia, y se combinan entre ellos para determinar su descendencia. En este caso particular el conjunto de parámetros de la calibración que se desea conocer constituyen los genes que definen al individuo.

El algoritmo determina mediante los operadores de los que dispone, la evolución de los individuos para ir produciendo una generación tras otra. Asimismo decide cuándo se ha terminado la reproducción siempre que se cumplan unos criterios, definidos por el modelador, y que son determinados mediante los operadores y las combinaciones de los mismos.

El criterio que determina si un individuo (conjunto de parámetros internos) debe reproducirse o no, es el ajuste que presentan los resultados del modelo con estos parámetros respecto a los valores de referencia: si con estos valores el ajuste es bueno, debe ser utilizado como progenitor para la siguiente generación de individuos a la búsqueda del mejor de ellos.

Esta técnica proporciona mínimos locales y lo hace de forma muy rápida una vez se conocen las funciones que definen el error que se desea minimizar. Sin embargo, tiene una limitación importante,

y es que para determinar este error, y por tanto la “bondad” de un individuo se requieren funciones explícitas, lo cual no está siempre a la disposición del modelador.

En este caso particular, para poder aplicar el algoritmo genético a la búsqueda del valor de los parámetros internos para el modelo que aquí se describe, se desarrollaron expresiones explícitas para las concentraciones que intervienen en el modelo completo del OD. Estas expresiones que se constituyen en funciones objetivo, tienen grandes limitaciones (si la modelación fuera directa mediante expresiones explícitas, el modelo numérico carecería de sentido) entre las cuales se encuentra la más importante de todas ellas: son expresiones para el régimen permanente.

En cualquier caso, la búsqueda del valor a asignar a los parámetros internos mediante la técnica del algoritmo genético en régimen permanente ha funcionado extraordinariamente bien. Los resultados que así se obtienen pueden ser utilizados para comenzar la calibración en régimen no permanente o incluso considerando estas situaciones, en situaciones no estacionarias, como semipermanentes, de manera que pueden ser conocidas funciones objetivo para instantes intermedios.

En este caso, el proceso de búsqueda de individuos está representado en la Figura 3.

El proceso es el siguiente: Se dispone inicialmente de medidas de las concentraciones de OD a lo largo del cauce receptor. Se considera, mediante el conocimiento de la expresión explícita de la concentración del OD la función objetivo deseada, que es la minimización de la diferencia en los puntos de medición entre la concentración de OD medida y calculada. En esta función puede variar solamente el valor de los parámetros que se están evaluando, dependiendo la función objetivo de la distancia y de los parámetros λ_i , ($OD(x, \lambda_i,)$). Se evalúa la diferencia entre los resultados de la expresión explícita y las mediciones en los puntos en los que se tiene datos, asignando este valor de diferencia acumulada como “puntuación” para el individuo.

En este caso, se desea encontrar el conjunto de valores que, introducidos en la ecuación que prevé la concentración del OD, es capaz de predecir con el menor error una serie de mediciones realizadas en el cauce. Para ello, cuando se cuenta con el AG es necesario conocer una expresión para la concentración de las funciones que se ven involucradas en el proceso. A

Donde se ha utilizado la nomenclatura descrita al final del presente artículo. Esta operación se realiza para todos los individuos de una generación, y se toma de entre ellos los mejores, esto es, los que menor error proporcionen en la aplicación del modelo. Entonces se verifica que se cumplen las condiciones de penalización. En este caso, las condiciones de penalización no pueden ser de optimización a su vez, sino que tienen que ser restricciones en el sentido estricto. Se desea que el modelo prediga además otras concentraciones consideradas en el cauce, que se tratan aquí como funciones de penalización. Estas condiciones se formulan de manera que la diferencia entre las concentraciones medidas y las que predice el modelo para DBO, Nitrogeno como Amonio, Nitrito y Nitrito, tenga un valor mínimo, menor que un cierto límite que tendrá que ser definido por el usuario.

Estas funciones dependen a su vez de los mismos parámetros que la función objetivo, por lo que los individuos que no cumplan las condiciones para ellas, no serán considerados para pasar a la siguiente generación. Las funciones de penalización se consideran en el presente trabajo como limitaciones superiores: el error acumulado en las predicciones de estas concentraciones que no son la función objetivo debe estar por debajo de un cierto máximo. En este sentido no son una optimización en sí mismas sino que se presentan en forma de restricciones.

Utilizando los individuos que mejor "puntuación" han obtenido (sujetos a las restricciones que marcan las funciones de penalización) para la siguiente generación, se van encontrando cada vez individuos que hacen que la concentración de OD calculada mediante el modelo se ajuste mejor a la medida en los puntos en los que se dispone de medición. Utilizando los operadores definidos en el algoritmo para la reproducción: mutación y cruce, se forman nuevos individuos en una generación a partir de los mejores de la generación de progenitores, con ello se mejorará la "puntuación" de los individuos de la nueva generación.

Llega un momento en que este ajuste no mejora, o bien todos los individuos son iguales o se ha alcanzado un número de generaciones definido por el usuario, con lo que el proceso ha terminado.

La utilización del algoritmo genético para la búsqueda de valores de los parámetros internos en el proceso de calibración del modelo ha dado (en casos reales en los que se ha implementado) excelentes resultados al modelar el régimen permanente.

Esto es especialmente importante en el caso de modelos en los que intervienen tantos parámetros simultáneamente en la optimización, y en los que las influencias cruzadas de los mismos son notables. De esta manera la determinación de los valores de estos parámetros (dentro de intervalos aceptables) puede ser realizada de forma automática y ajena a la experiencia del modelador, con un gasto de tiempo extraordinariamente pequeño. Como se verá a continuación, en casos que pueden ser considerados como permanentes la implementación de esta metodología minimiza el error rápidamente de manera notable.

El análisis de incertidumbre.

La transmisión del error de los datos a los resultados.

El error experimental es debido a la fluctuación o discrepancias en la repetición de los experimentos. Este error no implica que los datos no sean buenos, o no sean reales, se refiere a que las medidas tienen fluctuaciones debido a imperfecciones en el proceso de medición: errores de la instrumentación, variación de las condiciones ambientales (especialmente en los procesos considerados permanentes), habilidades del personal de laboratorio y otros. A pesar de ello, estas variaciones deben ser conocidas, valoradas y a su vez, minimizadas.

Si se conoce el error en la estimación de los llamados parámetros externos de entrada, puede estimarse cómo este error se transmite a los valores calculados. Este error se cuantifica en forma de desviación típica (o de varianza) de los resultados del modelo.

Basado en la teoría para la propagación del error en los modelos matemáticos propuesta por Mac Berthouex y Brown (1994) presentamos la expresión que determina la varianza transmitida a la magnitud calculada $Var(C_j)$ (López, 2001). Para ello será necesario conocer el error de las magnitudes de entrada como su propia varianza $Var(k_i)$, y el efecto relativo que la variación del parámetro de entrada de valor k_i tiene sobre el de salida C_j : ER_{ij}

$$Var(C_j) = \sum_i Var(k_i) \left(ER_{ij} \frac{C_j}{k_i} \right)^2$$

Esta expresión contiene implícita la suposición de que la influencia de una magnitud sobre otra se comporta de forma lineal, y ello no está asegurado.

El efecto relativo si el comportamiento es no lineal puede ser variable dependiendo de la magnitud de la perturbación. Aunque también se han despreciado las covarianzas suponiendo un modelo lineal. Por lo tanto, esta expresión adopta pleno significado con ER constantes, influencias lineales y perturbaciones no correlacionadas.

Con sus limitaciones, esta expresión permite conocer la varianza y desviación típica de los resultados que predice el modelo. Asimismo podrá hablarse de la desviación típica de la magnitud calculada:

$$\sigma(C_j) = \sqrt{\sum_i \text{Var}(k_i) \left(ER_{ij} \frac{C_j}{k_i}\right)^2} \quad (3)$$

Supongamos que se conoce la desviación típica relativa de los parámetros externos de entrada ($\sigma_R(k_i)$) definida como:

$$\sigma_R(k_i) = \frac{\sigma(k_i)}{\bar{k}_i} \quad (4)$$

Con ella puede estimarse la desviación típica de los resultados del modelo $\sigma(C_j)$:

$$\sigma(C_j) = \sqrt{\sum_i [\sigma_R(k_i)]^2 \bar{k}_i^2 \left(ER_{ij} \frac{C_j}{k_i}\right)^2} \quad (5)$$

y también, la desviación típica relativa de los valores calculados para cada instante y punto del espacio de integración.

$$\sigma_R(C_j) = \frac{\sigma(C_j)}{C_j} = \sqrt{\sum_i [\sigma_R(k_i)]^2 ER_{ij}^2} \quad (6)$$

Las regiones de confianza

Las regiones (o intervalos) de confianza de una variable estadística es aquel rango de valores entre los cuales existe una cierta probabilidad de que se encuentren las mediciones para una población, en este caso, de datos experimentales. En este sentido puede ser considerado como el intervalo de confianza del 95% (esto es, el 95% de los valores experimentales, considerada la magnitud que se mide como una variable normal, se encontrarán dentro del intervalo), una banda cuya amplitud es de cuatro desviaciones típicas, dos por arriba y dos por debajo de los valores medios (Mac Berthouex y Brown, 1994).

Esta estimación para la confianza de los valores puede tener sentido tanto para las magnitudes medidas que son condiciones de contorno en los nudos de entrada del sistema (concentraciones y temperaturas) como para las que son referencia para la calibración a lo largo del mismo. Y entendiendo que las magnitudes que calcula el modelo son a su vez concentraciones reales, afectadas también por su correspondiente región de confianza. Establecer la región de confianza de los resultados del modelo es interesante desde el punto de vista de la calibración, porque cuantifica la fiabilidad de las predicciones y representa en cada punto de estudio la influencia de la incertidumbre en las magnitudes registradas en los nudos de aporte.

Para conocer la región de confianza (o bien la banda de confianza alrededor del valor calculado por el modelo) deberá ser conocida la desviación típica relativa de las magnitudes calculadas a partir de la desviaciones típicas relativas de los parámetros externos y el correspondiente efecto relativo. De esta manera dicha región de confianza será:

$$\text{Región de confianza } (C_j) = C_j \cdot (1 \pm 2\sigma_R) \quad (6)$$

siendo C_j el valor de la concentración calculada.

Para ilustrar este concepto se presenta la Figura 4. En la Figura 4 se muestran los posibles resultados que proporcionará el modelo para la concentración de OD a lo largo de una línea del cauce receptor. En la figura 5 se presentará la región de confianza del 95%, de una amplitud de dos desviaciones típicas por arriba y por debajo de la concentración calculada.

Determinación del error de los resultados: criterio de terminación.

Para determinar la bondad de un ajuste se deben comparar los resultados obtenidos por el modelo con valores de referencia procedentes de la realidad, de otros modelos o de la experimentación en laboratorio. En este sentido, también estos valores de referencia (en el caso que nos ocupa, concentraciones de los compuestos que intervienen en el modelo del OD medidas a lo largo del cauce) están sujetos a error, por lo que también responden a una distribución estadística.

Así pues, se determina también para los valores utilizados en la comparación una región de confianza alrededor del valor medido. Con la ayuda de estos valores de referencia y sus bandas de confianza puede ya realizarse la comparación con los valores calculados.

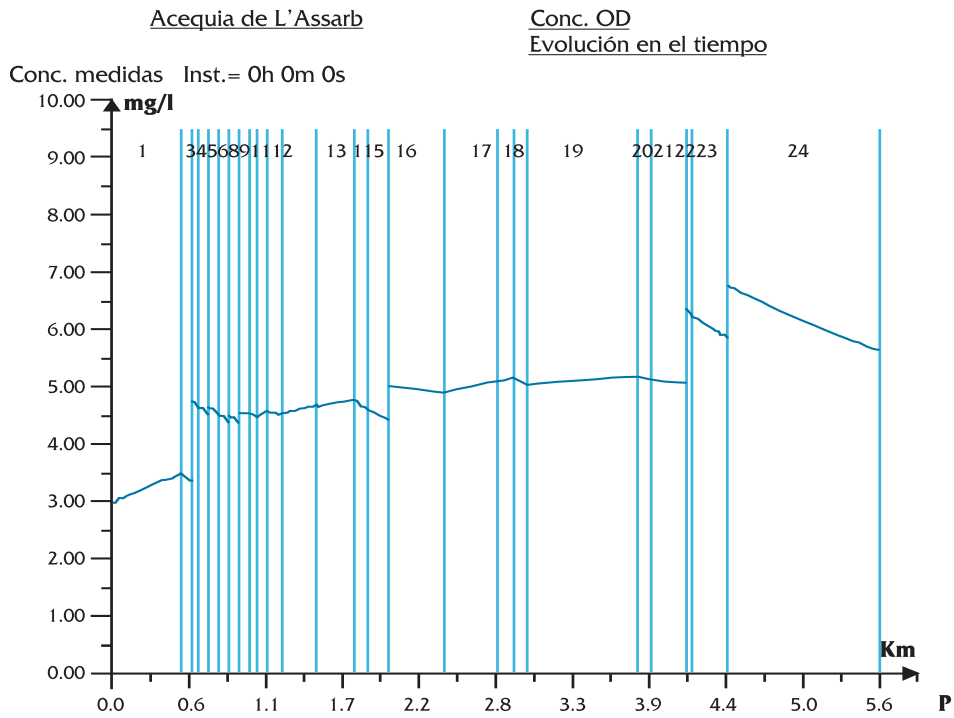


Figura 4. Resultados de la modelación de concentración de OD en cauce receptor

Por ejemplo en el caso previamente presentado, los valores para la comparación son los que se muestran en la Figura 5.

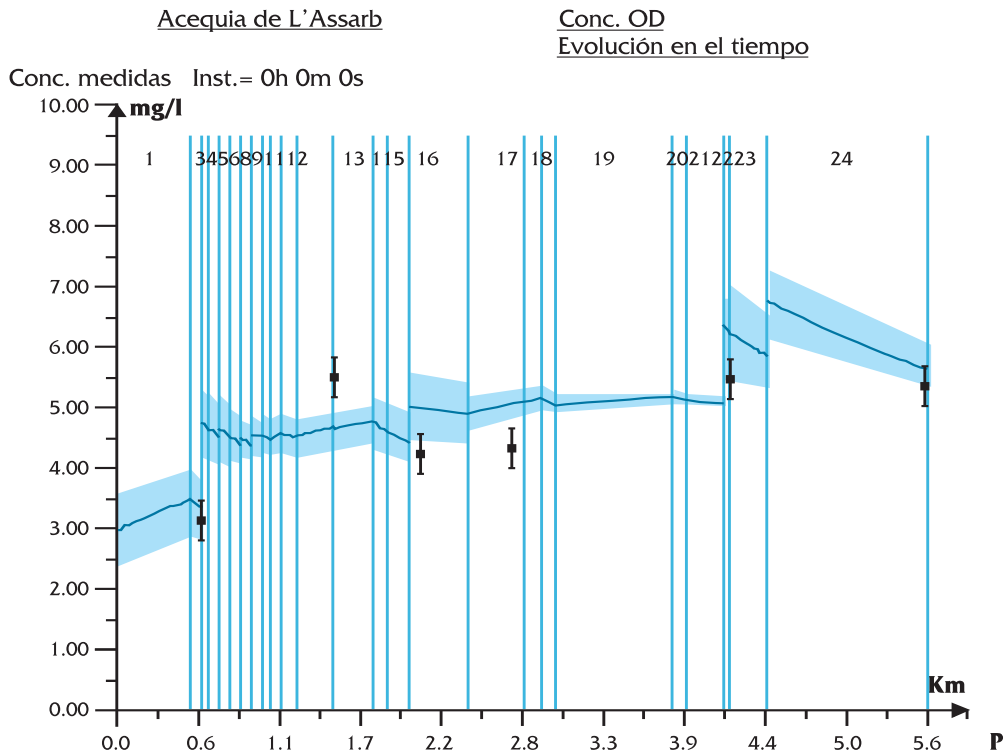


Figura 5. Comparación del modelo previamente calculado con las magnitudes medidas en el cauce.

Basado en la existencia de las regiones de confianza para resultados del modelo y para mediciones de referencia, se propone la consideración de que un proceso de calibración ha terminado cuando más del 90% de las veces, las regiones de confianza de las magnitudes medidas y calculadas se solapan, entendiéndose que ésta es la probabilidad conjunta de que el valor real se encuentre en ambas bandas de confianza.

Este criterio requiere que exista un número importante de mediciones. A partir de 10 mediciones para la comparación, el criterio adopta pleno sentido, en tanto que con menos de 10 mediciones, la desviación de uno solo de los valores experimentales obligaría a continuar el proceso de calibración.

Esta condición, en la práctica se muestra bastante restrictiva, especialmente en lo que se refiere a la calibración del régimen no permanente, por lo que el modelador deberá tener potestad para definir un porcentaje de solapamiento menor con objeto de terminar la calibración a la vista de la bondad de los resultados para la comparación y de las mismas propiedades del caso de estudio.

Una vez terminado el proceso de ajuste, se obtienen los valores de los parámetros de entrada internos que minimizan el error de las predicciones del modelo y la incertidumbre de los valores predi-

chos comparados con las mediciones de referencia. Es de esperar que mediante el modelo ajustado se puedan representar de forma fiable diferentes casos reales. En la siguiente sección se presentan algunos de estos casos modelados.

APLICACIÓN PRÁCTICA.

Se detalla aquí un caso de estudio destinado a confirmar la validez de la metodología de calibración propuesta. Las situaciones que se modelan son las registradas en la "Acequia del Barranc" y principio de la "Acequia de l'Assarb" sitas en la comarca de La Ribera, en la Comunidad Valenciana.

Al considerar la bibliografía encontrada en lo que se refiere a los casos prácticos en los que se referencian procesos de calibración en la que se analizan cauces importantes, es este un caso muy particular. La particularidad radica en parte en su tamaño, puesto que por sus características hidrodinámicas se trata de un microcauce en el que el aporte de contaminantes es debido a una fuente puntual en cabecera y el resto de aguas que llegan a los nudos de entrada son aguas limpias procedentes de sobrantes de riego. Sin embargo, en el entorno de la Comunidad Valenciana es una casuística muy común, por lo que la caracterización de cauces receptores similares puede tener un gran interés para simulaciones futuras en este tipo de entornos.

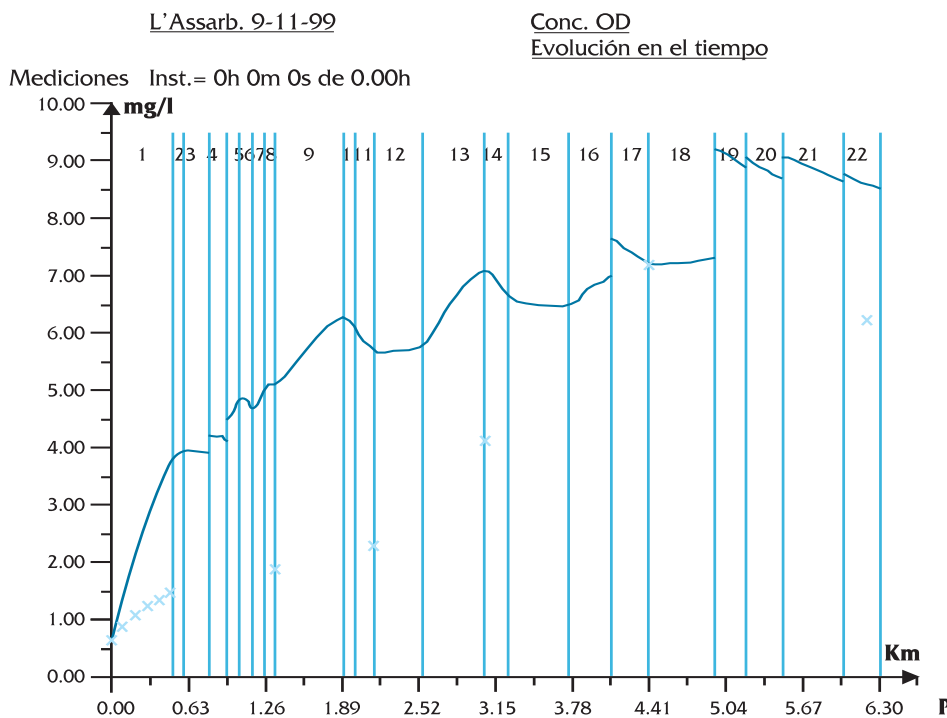


Figura 6. Predicción de Oxígeno Disuelto por el modelo

En este cauce receptor natural hay varios aportes debidos a las aguas sobrantes de los riegos de las zonas cercanas y un aporte principal debido a las aguas residuales sin depurar procedentes de la población de Alginet. De esta manera, tiene interés la representación de los compuestos que forman parte del modelo del OD. Además del ya citado Oxígeno, se modelará temperatura, materia orgánica carbonatada en suspensión, nitrógeno orgánico y nitrógeno en forma de amonio, nitrito y nitrato.

Primeramente se ha modelado el cauce receptor con los valores por defecto para los parámetros internos definidos en la bibliografía, con los que se han conocido las concentraciones indicadas en el punto anterior en todos los puntos del cauce. Se muestra en la Figura 6 la concentración de OD en este caso, comparándola con las mediciones realizadas en dicho cauce.

Como se ve, la diferencia entre las predicciones y los valores medidos es notable. Con objeto de mejorar esta modelación se implementa la estrategia de

calibración previamente definida, de manera que se realiza sobre el cauce un análisis de sensibilidad, se ordenan los parámetros por influencias, y por último se implementa una búsqueda del valor de los parámetros más influyentes mediante aplicación de un algoritmo genético. De esta manera, las predicciones de OD que presenta el modelo, utilizando los parámetros que el algoritmo predice, es la que se muestra en la Figura 8. Puede observarse en esta figura la mejora que se obtiene en los resultados del modelo al aplicar la estrategia de calibración propuesta.

En la Tabla siguiente se muestra los resultados del análisis de sensibilidad realizado en el caso presentado. En ella se observa el efecto relativo que tiene la variación de cada uno de los parámetros de entrada en un 25% de su valor por defecto sobre los de salida y el punto del cauce receptor en el que esto ha ocurrido. Con ello es posible hacer una ordenación de estos parámetros en relación con la citada influencia para constituir el genoma que entrará a ser calculado en el algoritmo genético.

Tabla 1. Resultados del análisis de sensibilidad al variar los parámetros de entrada un 25%.

	OD		DBO _c		N - NH ₄ ⁺		N - NO ₂ ⁻		N - NO ₃ ⁻	
	ER	Dist. origen (m)	ER	Dist. origen (m)	ER	Dist. origen (m)	ER	Dist. origen (m)	ER	Dist. origen (m)
Coefficiente de Manning, n.	-1.43	800	-0.04	máx	-0.04	máx	0.08	máx	0.01	máx
Pendiente de solera, S.	0.73	800	0.02	máx	0.02	máx	0.02	máx	-0.01	máx
Temperatura en nudos de entrada	-0.55	2500	-0.1	máx	-0.05	máx	0.1	máx	0.02	máx
Caudal en nudos de entrada	-1.13	800	0.07	máx	0.02	máx	-0.04	máx	-0.01	máx
OD en nudos de entrada	1	0	0		0		0		0	
DBO _c en nudos de entrada	-0.57	800	1	0	0		0		0	
N orgánico en nudos de entrada	0		0		0.03	máx	0		0	
N Amonio en nudos de entrada	-0.06	800	0		1	0	0.29	máx	0	
N Nitrito en nudos de entrada	-0.01	máx	0		0		1	0	0.02	máx
N Nitrato en nudos de entrada	0		0		0		0		1	0
Coefficiente de disp. Longitudinal	0		0		0		0		0	
Coefficiente de degradación de DBO _c en líneas k ₁	-0.56	800	-0.11	máx	0		0		0	
Coefficiente de reaireación en líneas k ₂	1.21	800	0		0		0		0	
Coefficiente de paso de Norg. a Amonio en líneas, k _{oa}	0		0		0.03	máx	0	máx	0	
Coefficiente de paso de amonio a nitrito en líneas k _{ai}	-0.06	800	0		-0.09	máx	0.28	máx	0	
Coefficiente de paso de nitrito a nitrato en líneas k _{in}	-0.01	máx	0		0		-0.16		0.02	máx
Coefficiente de consumo de OD en sedimentos en líneas k _{sb}	-0.23	800	0		0		0		0	
Parámetro de temperatura sobre k ₁	1.06	800	0.27	máx	0		0		0	
Parámetro de temperatura sobre k ₂	-1.98	800	0	0	0		0		0	
Parámetro de temperatura sobre k _{oa}	0		0		-0.08	máx	-0.01	máx	0	
Parámetro de temperatura sobre k _{ai}	0.12	800	0		-0.22	máx	0.66	máx	-0.01	máx
Parámetro de temperatura sobre k _{in}	0.02	máx	0		0		0.39	800	-0.04	máx
Parámetro de temperatura sobre k _{sb}	0.43	800	0		0		0		0	

Nota: El ER se indica en tanto por uno, al ser una magnitud adimensionalizada. Por otro lado la distancia al origen en la que se da el máximo ER se encuentra indicada en metros.

Tabla 2. desviaciones típicas relativas de los parámetros considerados en el análisis de incertidumbre

Parámetro externo	Desviación típica relativa
Coefficiente de Manning	10%
Pendiente de solera	10%
Caudal de entrada	10%
Caudal de salida	10%
Temperatura de entrada	5%
Concentración de OD	10%
Concentración de DBOcs	15%
Concentración de N. Orgánico	30%
Concentración de N. como amonio	30%
Concentración de N. como nitrito	15%
Concentración de N. como nitrato	15%

Estos resultados del análisis de sensibilidad servirán asimismo para realizar un estudio de incertidumbre, ya que, mediante las fórmulas (6) y (7) puede conocerse la desviación típica de las predicciones y la amplitud de las bandas de confianza.

En este caso, se hace necesario conocer las varianzas de los parámetros de entrada, para ello se utilizaron los datos presentados en la Tabla 2, obtenidos de la experimentación y de la bibliografía (López, 2001, Canal de Isabel II 1992).

Para determinar la precisión de las predicciones se impone, como se ha señalado anteriormente, la determinación de las regiones de confianza de las concentraciones predichas por el modelo. Para evaluar estas regiones de confianza se modela el caso con valores para parámetros internos obtenidos del

algoritmo genético que son los que proporcionan mayor precisión. Dichas regiones de confianza son las que se muestran en la Figura 8.

Cuando se comparan con lo que hemos denominado regiones de confianza de las mediciones realizadas a lo largo del cauce se observa que unas y otras regiones se solapan en más del 90% de las veces, con lo que la calibración ha sido realizada con éxito. En este caso el gasto de recursos y de tiempo ha sido mínimo comparado con la precisión de las predicciones.

Utilizando el valor de los parámetros que ha sido obtenido en fases anteriores se modela el mismo cauce en una situación diferente, de manera que la predicción sobre el Oxígeno Disuelto es la que se muestra en la Figura 9.

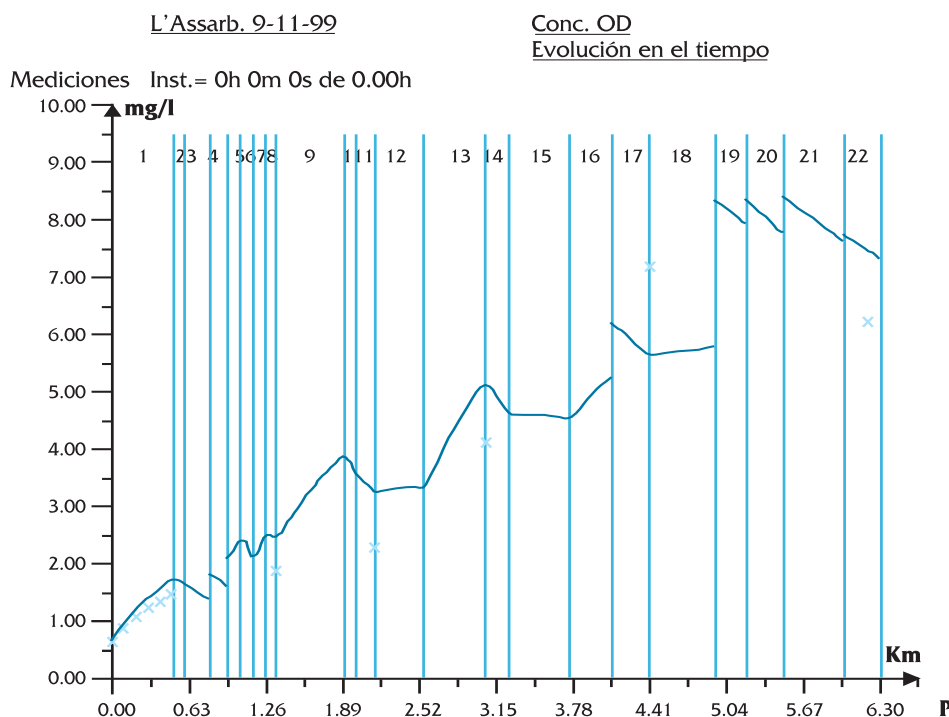


Figura 7. Predicción de Oxígeno Disuelto por el modelo con parámetros procedentes del algoritmo genético (10.000 iteraciones)

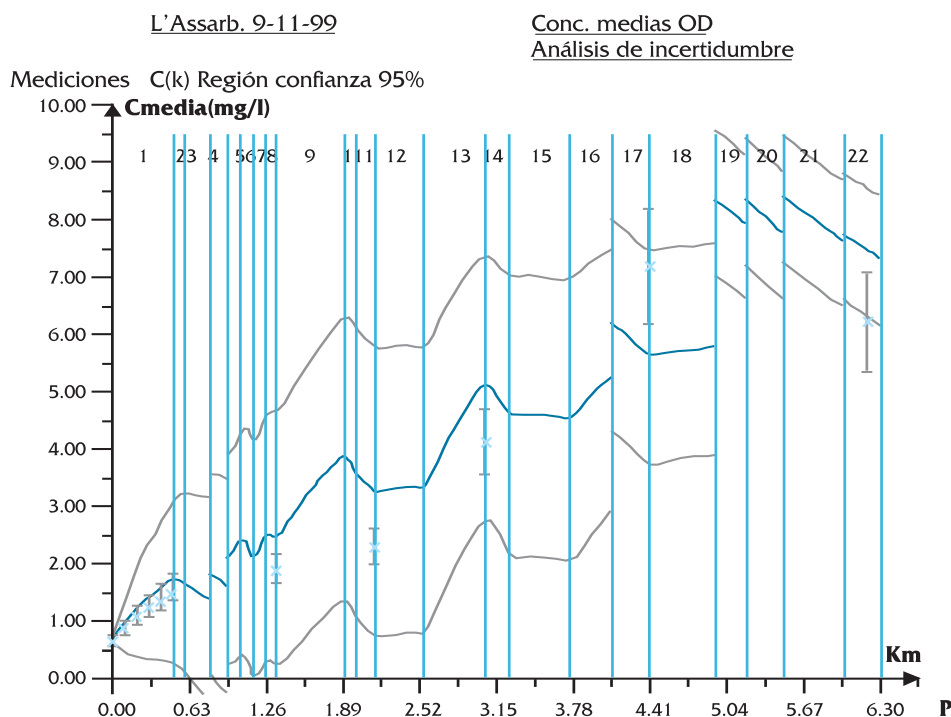


Figura 8. Regiones de confianza para predicción del Oxígeno Disuelto

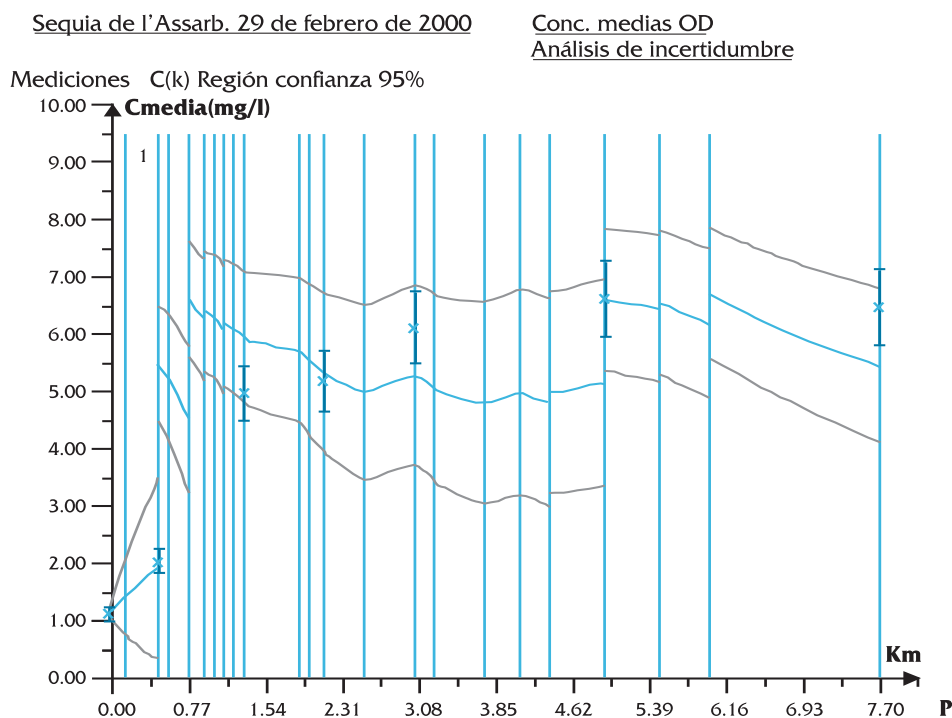


Figura 9. Predicción del Oxígeno disuelto en el cauce para otra simulación

En dicha Figura puede observarse que las predicciones son muy cercanas a las mediciones para el OD en el seno del cauce y en cualquier caso, las bandas de confianza de mediciones y predicciones se solapan en toda la longitud de modelación. Así pues, los parámetros internos procedentes de la predicción del algoritmo genético han sido capaces de

predecir otras situaciones y las bandas de confianza se han solapado en más de un 90% de las ocasiones, lo cual valida la modelación de este caso particular. Con ello se completa el proceso de calibración, y el modelo podrá ser utilizado para predecir situaciones futuras como herramienta de gestión del cauce con una fiabilidad contrastada.

CONCLUSIONES.

En este artículo ha sido presentada una metodología para calibrar modelos matemáticos de dispersión de contaminantes.

Esta metodología está basada en la estimación de los parámetros que afectan a la formulación de las ecuaciones que definen el modelo, distinguiendo entre los parámetros internos que intervienen en las ecuaciones, provenientes de la bibliografía, y los parámetros externos que también intervienen en las ecuaciones pero son conocidos mediante la experimentación.

Para los parámetros internos se propone la búsqueda de su valor mediante la utilización de una técnica de optimización evolucionaria como los algoritmos genéticos. Esta metodología se muestra muy capaz cuando se trata de situaciones que pueden modelarse considerando régimen permanente, puesto que requiere el uso de funciones objetivo explícitas.

El tratamiento de los parámetros externos que intervienen en las ecuaciones debe hacerse mediante un análisis de incertidumbre que determina la desviación típica de las magnitudes calculadas considerando la desviación típica de las magnitudes medidas y que intervienen en la modelación. La determinación de la desviación típica de las magnitudes calculadas es de gran interés puesto que con ella pueden definirse bandas o regiones de confianza alrededor de las concentraciones calculadas. Estas regiones de confianza son determinantes para definir un criterio de terminación del proceso de calibración: este habrá terminado cuando las regiones de confianza de las mediciones y de las predicciones del modelo se solapen más de un 90% de las ocasiones. Ello permite terminar la calibración teniendo en cuenta tanto el error de la predicción del modelo como la imprecisión de las lecturas con las que se compara en el proceso de calibración.

El objeto perseguido por esta metodología para la calibración es buscar un proceso lo más sistemático y automático posible, independiente de la experiencia del modelador o de su conocimiento del caso particular que se está representando, y que este proceso conlleve el mínimo gasto de recursos y tiempo que permita asegurar la fiabilidad de las predicciones futuras del modelo mediante una correcta validación.

Esta metodología tiene todavía limitaciones. Por ejemplo, la extrapolación del uso del algoritmo genético para la calibración de modelos en régimen

no permanente no es directa, puesto que en estos casos no se dispone de funciones objetivo explícitas y hay que resolver las ecuaciones diferenciales mediante métodos numéricos que no tienen interacción directa con el algoritmo. Sin embargo ello constituye el reto para desarrollos futuros que son signo de que esta metodología está completamente viva.

LISTA DE SÍMBOLOS

A_i	Concentración de algas
C	Concentración
C_j	Concentración cualquiera calculada por el modelo
OD	Concentración de saturación de OD
ER	Efecto Relativo
ER_j	Efecto relativo que tiene sobre la concentración C_j la variación del parámetro k_i .
h	Profundidad
k_{ai}	Velocidad de oxidación de amonio
k_b	Velocidad de consumo de OD por sedimentos
k_i	Valor que adopta un parámetro externo cualquiera en el proceso de calibración
k_{in}	Velocidad de oxidación de nitrato
k_{oa}	Velocidad de transformación de nitrógeno orgánico a amoniaco
k_1	Velocidad de descomposición de DBO
k_2	Coefficiente de velocidad de reaireación
K_l	Coefficiente de dispersión longitudinal
L_b	Fracción sedimentable de DBO
L_c	Fracción suspendida de DBO
N_a	Concentración de nitrógeno como amonio
N_i	Concentración de nitrógeno como nitrito
N_n	Concentration of nitrógeno como nitrato
N_o	Concentración de nitrógeno orgánico
OD_{cal}	Concentración de Oxígeno Disuelto calculada por el modelo
OD_{ref}	Concentración de Oxígeno Disuelto de referencia para la comparación
r_{oa}	Coefficiente estequiométrico de oxígeno de oxígeno consumido en nitrificación de amonio
r_{oi}	Coefficiente estequiométrico de oxígeno de oxígeno consumido en nitrificación de nitrito
U	Velocidad media
α_f	Producción de OD en fotosíntesis de algas
α_r	Consumo de OD por respiración de algas
μ	Crecimiento de algas por fotosíntesis
ρ	Desaparición de algas
σ	Velocidad de eliminación de algas
$Var()$	Varianza
λ_i	Individuo en el algoritmo genético: vector de valores de los parámetros internos en el proceso de calibración
$\sigma()$	Desviación típica
$\sigma_R()$	Desviación típica relativa

REFERENCIAS

- Beasley, D.; Bull, D.; Martin, R. (1993). An Overview of Genetic Algorithms. Part 2. Research Topics. University Computing, Vol 15, Nº 4. Pp 170-181.
- Bowie, G.L.; Mills, W.B. (1985). Rates, Constants and Kinetics Formulations in Surface Water Quality Modelling. Second Edition. Ed. EPA.
- Canal de Isabel II. (1992). GRYM. Un modelo para la gestión de la calidad de los ríos de la Comunidad de Madrid.
- Chapra, S.C. (1997). Surface Water Quality Modelling Mc. Ed. Graw-Hill. New York.
- Engelhardt, M.O.; Dandy, G.C. (1999). The development of an Optimal Strategy to Schedule Main Replacements. En el libro: Water Industry Systems: Modelling, optimization and applications. Savic, D. y Walters, G. Editores. Ed. Research Studied Press LTD.
- Ferreira, J.S.; Costa, M. Lobo, F.; Câmara, A.(2002) Estuarine Transport model calibration using genetic algorithms. Hydroinformatics.
- Goldberg, D.E. (1989). Genetic Algorithms search, optimisation and machine learning. Addison-Wesley Publishing Co. Reading. Mass.
- Holland, J. (1975). Adaptation in natural and artificial systems. Ann Arbor. University of Michigan Press.
- López, P.A. (2001). Metodología para la calibración de modelos matemáticos de dispersión de contaminantes incluyendo regímenes no permanente. Tesis doctoral. Universidad Politécnica de Valencia.
- Mac Berthouex; P., Brown, L. (1994). Statistics for Environmental Engineers. Lewis Publishers. U.S.A.
- Mulligan, A.; Brown, L. (1998). Genetic Algorithms for calibrating water quality models. Journal of Environmental Engineering. ASCE 1998. Vol 3 pp 202-211.
- Robinson, S. (1999). Simulation, verification, validation and confidence. A tutorial. Transactions of the Society of Computer Simulation International. Volume 16. Nº2, pp 63-69.
- Rocha, F. (1997). Modelação da Eutrofização do rio Guadiana. Aplicação de técnicas heurísticas de optimização à calibração do modelo. Tese de especialista. Laboratorio Nacional de Engenharia Civil. Lisboa. Portugal.