

## Tema 4

# Modelos y modelización

Como ya se ha visto, existen muchos tipos de modelos, aunque normalmente cuando se trata de estudiar procesos naturales se utilizan modelos matemáticos. Un **modelo matemático** es la expresión formal (en lenguaje matemático) de las relaciones entre los componentes de un modelo. La construcción de un modelo de este tipo implica la selección y cuantificación de los componentes, variables y relaciones presentes en el sistema para representarlo con el nivel de detalle requerido. Puede ser algo tan simple como sustituir las variables de una ecuación por sus valores reales o puede ser un conjunto complejo de ecuaciones interrelacionadas cuyas variables se ven modificadas a lo largo del tiempo y a través del espacio.

El lenguaje matemático permite describir y modelizar sistemas de manera parsimoniosa <sup>1</sup>, objetiva y no ambigua; hasta el punto de que hoy en día se considera a los modelos matemáticos como representaciones de teorías acerca de los sistemas que se modelizan. El lenguaje simbólico que aportan las matemáticas permite expresar ideas de gran complejidad.

Debe distinguirse entre esta concepción de modelo, como representación de una teoría mediante una ecuación más o menos sencilla, y la idea de modelo como agrupación de un conjunto de ecuaciones (que responden a diferentes teorías) interconectadas de manera que representen las diferentes transferencias entre los componentes del sistema modelizado.

Por mínimamente complejo que sea un modelo, y especialmente si tiene en cuenta variaciones en el espacio o en el tiempo, requiere la resolución de muchísimas operaciones matemáticas. Por tanto se requiere un ordenador para su resolución ordenada en forma de **algoritmo**.

Las características deseables de los modelos matemáticos (Moore *et al.*, 1993) son:

- **Parsimonia**, un modelo no es necesariamente mejor por tener muchos parámetros. La simplicidad es siempre deseable.
- **Modestia**, deben tratar de alcanzarse sólo objetivos asequibles. Un modelo, al igual que un mapa, no debe aspirar a imitar la realidad sino sólo a resaltar aquellos aspectos de interés para su aplicación.
- **Exactitud**, el modelo debe reproducir en la medida de lo posible el funcionamiento del sistema y generar valores para las variables de salida y estado similares a los observados en la realidad.
- **Verificabilidad**, los resultados del modelo deben poder compararse con datos reales y determinar de este modo el grado de exactitud del modelo.
- Por otro lado, no basta con que funcionen bien, deben funcionar bien por las razones correctas

---

<sup>1</sup>reduciendo el número de parámetros, se considera que a igualdad de otras consideraciones es preferible un modelo simple a otro más complejo

## 4.1 Tipos de modelos matemáticos

Una clasificación de los modelos matemáticos podría basarse en una serie de características dicotómicas:

- **Empíricos o basados en principios físicos**

El carácter empírico o físico constituye la característica fundamental de un modelo. Un modelo físico se basa en las leyes físicas que rigen los procesos, un modelo empírico se basa en relaciones estadísticamente significativas entre variables. Las ecuaciones que describen un modelo estadístico no son por tanto físicamente o dimensionalmente consistentes ni universales, ya que en rigor sólo son válidas para el contexto espacio-temporal en el que se calibraron. Se caracterizan por un alto poder predictivo pero una escasa capacidad explicativa, es decir reproducen el funcionamiento del sistema razonablemente bien pero no permiten saber por que el sistema funciona así. Los modelos estadísticos se conocen también como **modelos de caja negra** ya que no permiten descubrir el funcionamiento interno del sistema.

Los modelos físicos se denominan, por contraposición, **modelos de caja blanca**. Se trata de modelos en los que las transferencias de materia y energía entre sus componentes se rigen mediante ecuaciones físicas y que además cumplen las leyes de conservación de la materia y la energía, tanto para el conjunto del modelo como para cada uno de los submodelos.

Una posibilidad intermedia son los **modelos de caja gris** o **conceptuales**. Se trata de modelos en los que el sistema se descompone en una serie de componentes que se resuelven como modelos empíricos pero cuya integración se basa en principios físicos o al menos en cierto conocimiento a priori de como funciona el sistema.

Para poder construir un modelo físico es necesario un alto conocimiento acerca de como funciona el sistema a modelizar. El modelo resultante permite transformar unas variables de entrada en variables de salida. En el caso de los modelos empíricos la falta de conocimiento acerca del sistema se compensa con datos de calidad y en cantidad suficiente de las variables de entrada y de salida. A partir de estos datos podemos construir un modelo empírico que nos permita, posteriormente, obtener las variables de salida a partir de un nuevo conjunto de valores para las variables de entrada.

- **Estocásticos o deterministas**

Los primeros incluyen generadores de procesos aleatorios dentro del modelo que modifican ligeramente algunas de las variables. De esta manera, para un mismo conjunto de datos de entrada, las salidas no serían siempre las mismas. La distinción ente modelos deterministas o estocásticos se confunde a veces con la anterior, relacionando equivocadamente modelos estocásticos con empíricos y deterministas con físicos. En realidad un modelo determinista es aquel en el que dado un conjunto de de parámetros y variables de entrada va a producir siempre el mismo conjunto de variables de salida. En el caso de un modelo estocástico los valores de las variables de salida van a a variar de unas ejecuciones del modelo a otras ya que se deja intervenir al azar.

Las razones por las que el comportamiento de un sistema puede resultar aleatorio son diversas:

- Existencia de procesos realmente estocásticos
- Existencia de procesos aparentemente estocásticos debido a nuestra falta de conocimiento
- Errores aleatorios en la medición de las variables
- Procesos muy complejos que es preferible modelizar como estocásticos

La introducción de un componente estocástico en un sistema puede venir a su vez de:

- Utilización de generadores aleatorios de series de las variables de entrada (precipitación por ejemplo)

- Utilización de generadores aleatorios para dar valores a los parámetros del modelo y a su distribución espacial, de esta manera se evita el problema que supone utilizar parámetros estimados con cierto grado de incertidumbre. Por ejemplo al incluir la capacidad de infiltración del suelo en un modelo hidrológico podemos utilizar siempre la misma interpolación o permitir que varíe al azar de unas ejecuciones a otras.
- Cuando en un modelo las salidas de un componente pueden dirigirse a varios componentes distintos y no es fácil determinar a cual o en que cantidad, pueden determinarse las cantidades al azar. Por ejemplo en un modelo hidrológico basado en la rasterización de una cuenca la transferencia de agua de una celdilla a las celdillas aguas abajo puede hacerse de modo determinista (siempre igual) o aleatorio (puede variar).

La introducción del componente estocástico permite además comprobar como se comportaría el modelo para diferentes conjuntos de parámetros o valores de las variables de entrada. De este modo en lugar de obtener un resultado a partir de un conjunto verosímil de datos, obtendremos un conjunto de resultados a partir de varios conjuntos verosímiles de parámetros y variables. De este modo no tenemos por que conformarnos con un valor esperable sino que tendremos un rango de variación dentro del cual estarán los resultados esperables. Por ejemplo en el caso de estimación del riesgo de inundación resulta muy difícil determinar cual será la altura máxima de la lámina de agua. Los modelos estocásticos permitirían obtener una distribución de probabilidades de altura de agua, lo que sería más interesante de cara a la planificación del territorio.

#### ● Agregados o distribuidos

En el primer caso toda el área de estudio se considera de forma conjunta, por ejemplo una cuenca hidrográfica. Se tiene un único valor para todos los parámetros del modelo. El modelo predice unas salidas para las entradas aportadas sin informar de lo que ocurre dentro del sistema.

En un modelo distribuido, tendremos el área de estudio dividida en porciones cada una de ellas con su propio conjunto de parámetros y sus propias variables de estado. Cada porción recibe un flujo de materia y energía de algunas de sus vecinas que a su vez reemite a otras.

Una tercera posibilidad son los modelos semidistribuidos que se construyen a partir de la yuxtaposición de diversos modelos agregados, por ejemplo diversas subcuencas de una cuenca hidrográfica. Otra posibilidad a menudo explorada en hidrología es dividir el área de trabajo en *Unidades de Respuesta Hidrológica*. Se trata de segmentos de ladera homogéneos en cuanto a su pendiente, orientación, litología y uso a los que se asume una respuesta hidrológica única. En un modelo semidistribuido las diferentes unidades generan sus propias salidas de forma agregada pero aparecen entradas y salidas de unas a otras.

La incorporación de la componente espacial en los modelos resulta bastante compleja. Si se opta por un modelo distribuido es necesario establecer un **modelo de datos espaciales** que permita asignar valores de los parámetros y las variables de estado a los diferentes puntos del área de estudio. Puede tratarse de **distribuciones de puntos**, de mallas **raster** o de redes irregulares de triángulos (**TIN**). Si se trabaja con modelos agregados o semidistribuidos hay que codificar, además, los límites de las diferentes unidades. Todos estos procesos son más complejos de lo que pudiera parecer a primera vista e implican la creación de un **modelo de datos espaciales** y la manera más eficiente de hacerlo es mediante un **Sistema de Información Geográfica**.

#### ● Estáticos o dinámicos

Se refiere a la forma en que se trata el tiempo. Los modelos estáticos dan un resultado agregado para todo el período de tiempo considerado este puede ser por ejemplo un caudal medio o un caudal punta. Los modelos dinámicos devuelven las series temporales de las variables consideradas a lo largo del período de estudio. Por ejemplo podemos considerar un modelo estático de cuenca en el que el caudal medio ( $Q'$ ) depende de la precipitación media ( $P'$ ) y de los parámetros de la cuenca ( $Pc$ ):

$$Q' = f(P', Pc) \quad (4.1)$$

o un modelo dinámico en el que el caudal en cada intervalo de tiempo considerado ( $Q_{t+1}$ ) depende de la precipitación no sólo en dicho intervalo de tiempo sino también en los intervalos anteriores y de los parámetros de la cuenca:

$$Q_t = f(P_t, P_{t-1}, P_{t-2}, \dots, Pc) \quad (4.2)$$

Los parámetros se distinguen de las variables en que aquellos son invariantes a la escala espacio-temporal del modelo. Las variables de entrada y salida representan flujos de materia y energía desde y hacia el interior del sistema (precipitación y caudal por ejemplo). Las variables de estado representan cambios en la cantidad de materia y energía disponible (humedad del suelo). La distinción entre variables y parámetros depende de la escala, espacial y temporal, del modelo.

En definitiva, un sistema natural recibe entradas de materia y energía de su entorno que devuelve a dicho entorno con ciertas modificaciones. Entre estas modificaciones cabe destacar:

- Desplazamiento en el espacio
- Modulación en el tiempo de los flujos

Si el modelo tiene en cuenta de forma explícita el espacio estamos ante un modelo espacialmente distribuido, si se tiene en cuenta el tiempo tenemos un modelo dinámico.

La figura 4.1 muestra un esquema del comportamiento general de un sistema. La figuras 8.7 y 4.1 muestran dos ejemplos concretos, el primero se refiere a un balance de radiación y el segundo a un modelo hidrológico.

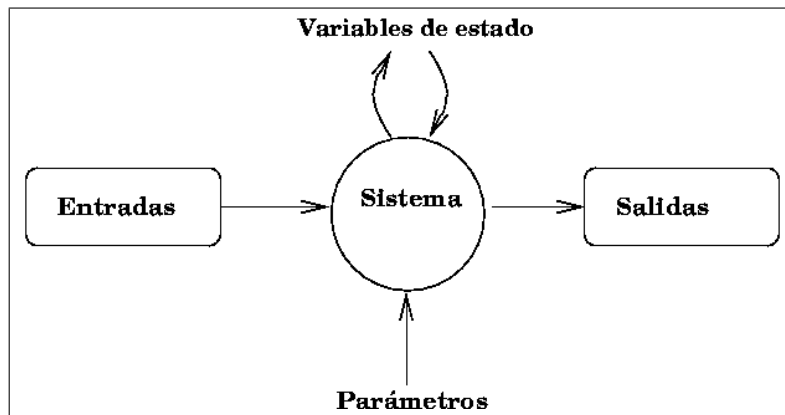


Figura 4.1: Modelos de procesos

## 4.2 Componentes de un modelo

La descripción de un modelo suele hacerse atendiendo tanto a sus características estructurales (componentes y relaciones) como funcionales (ecuaciones y parámetros que definen estas relaciones).

Las relaciones entre variables pueden formalizarse mediante un diagrama causal, generalmente se utiliza la notación:

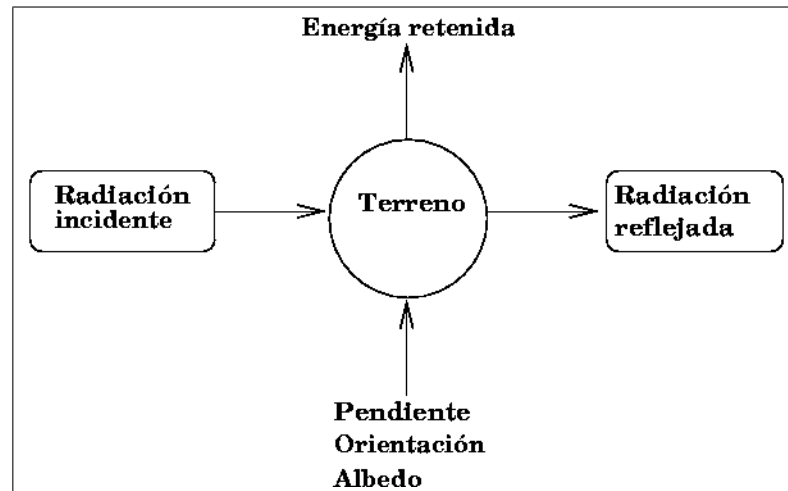


Figura 4.2: Modelos de radiación

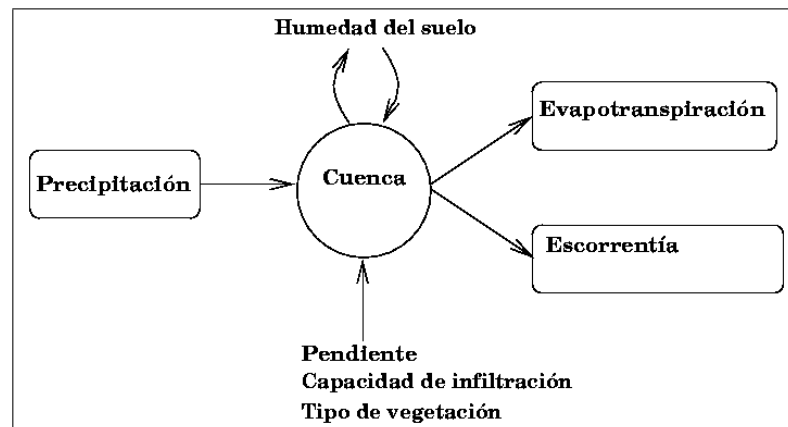


Figura 4.3: Modelos de procesos en una cuenca

$$a \rightarrow^+ b \quad (4.3)$$

Cuando una variable  $a$  afecta a otra  $b$  y esta última a su vez afecta a la primera tenemos un bucle de retroalimentación que suele codificarse mediante un círculo en cuyo interior se coloca un signo  $+$  si el bucle es explosivo y un signo  $-$  si el bucle es depresivo.

- Cuando en una secuencia de retroalimentación sólo aparecen signos positivos el bucle es explosivo
- Cuando sólo hay bucles negativos el bucle es depresivo
- Cuando hay signos positivos y negativos es explosivo si la suma de signos negativos es par (y así se compensan unos a otros) en caso contrario es depresivo

Los modelos de Dinámica de Sistemas, enunciados por primera vez por Forrester (1961) han aportado una nomenclatura y una simbología ampliamente utilizada para la definición de las características estructurales de los modelos. En la figura 4.1 aparece un ejemplo basado en un modelo hidrológico.

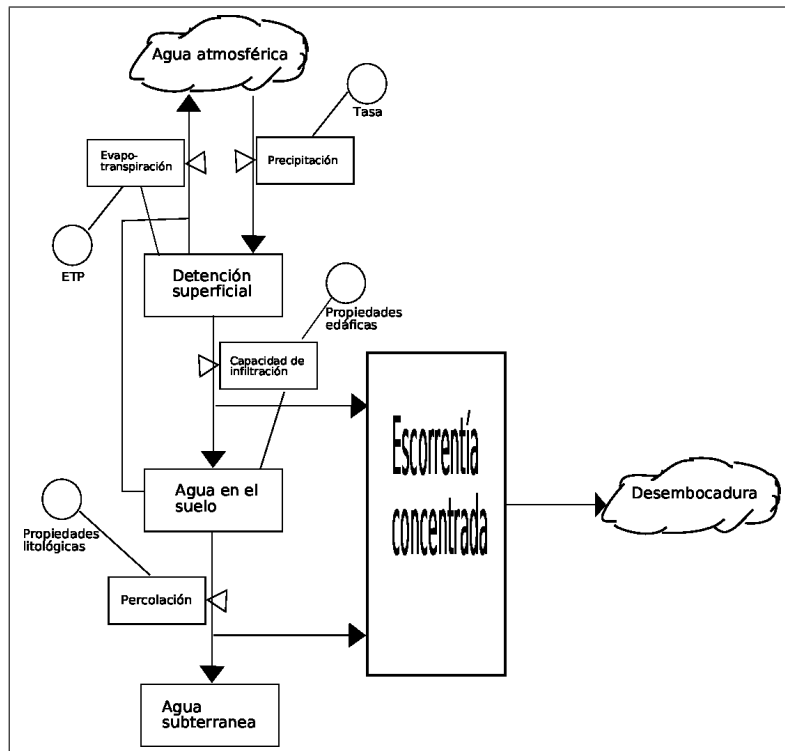


Figura 4.4: Modelos de procesos en una cuenca 2

1. **Variables de estado o niveles**, corresponden a la cantidad de materia o energía almacenada en cada uno de los componentes (subsistemas) que forman parte del sistema. Suelen representarse mediante un rectángulo
2. **Flujos**, indican la cantidad de materia o energía que viaja de un componente a otro en un intervalo determinado de tiempo. Afectan por tanto a los niveles, llenando unos (flujos de entrada) y vaciando otros (flujos de salida). Suelen representarse con una flecha, indicando la dirección del flujo, que atraviesa un símbolo que recuerda a un grifo.
3. **Fuentes y sumideros**, representan niveles exógenos que, por ser exteriores al sistema no interesa controlar. Suelen representarse mediante nubes. Los flujos que se dirigen de una fuente a una variable de estado son las **variables de entrada**, los que se dirigen de una variable de estado a un sumidero son las **variables de salida**.
4. **Variables auxiliares**, intervienen en las diversas ecuaciones que componen el sistema pero no se corresponden con un nivel o flujo. Suelen representarse con un círculo.
5. **Variables exógenas**, actúan fuera del sistema pero condicionan las variables de entrada. Pueden modificarse para construir **escenarios**. Un buen ejemplo sería la constante solar que, siendo ajena al sistema climático, modifica las trayectorias de este.
6. **Parámetros**, similares a las variables auxiliares pero cuyo valor no varía a lo largo del período de simulación, si el modelo se construyera a una escala diferente podrían dejar de ser parámetros para convertirse en variables.
7. **Constantes**, corresponden a magnitudes físicas que, como la aceleración de la gravedad, no varían en ningún caso.

8. **Canal de información**, relacionan variables, parámetros o niveles con los componentes en cuya ecuación se utilizan. Se representan con flechas, distintas de las utilizadas para representar flujos, o mediante líneas.
9. **Condiciones iniciales**, valores iniciales de las variables de estado.
10. **Condiciones de contorno**, series temporales de valores de las variables de entrada.
11. **Escenario**, conjunto plausible de variables exógenas, parámetros y condiciones iniciales y de contorno que permiten experimentar con un modelo diversas situaciones. Por ejemplo los escenarios relativos a diversas políticas de emisión de gases con que trabajan los Modelos Generales del Clima.

Todos estos componentes pueden resumirse en **características formales** que incluyen las variables de estado, flujos, variables, canales de información, etc., formando un diagrama de flujo (figura 4.1). Las **características funcionales** hacen referencia a las ecuaciones y valores de las variables y parámetros que aparecen detrás de las características estructurales.

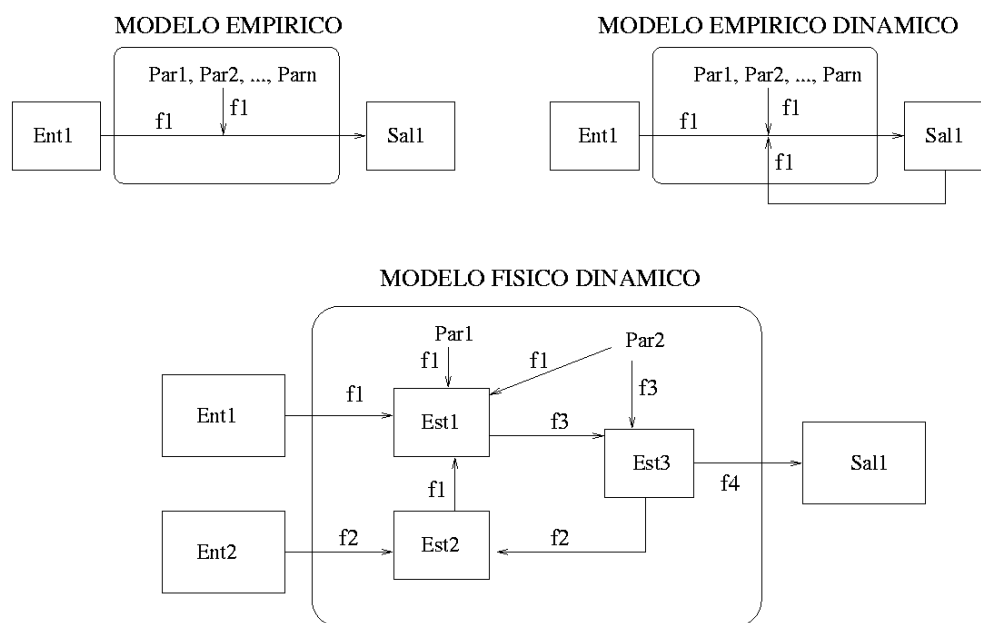


Figura 4.5: Tipos de modelos

### 4.3 Ecuaciones de los modelos

Las ecuaciones de los modelos matemáticos establecen de manera explícita cuales son las influencias de unos componentes sobre otros que ya se han representado en primer lugar mediante un grafo. El elegir un tipo de modelo u otro implica generalmente elegir un tipo de ecuaciones u otro.

Los modelos físicos se suelen construir mediante ecuaciones diferenciales que se resuelven mediante métodos numéricos. Si el modelo es distribuido y dinámico se utilizarán métodos de diferencias finitas utilizando la resolución espacial y temporal como valores de  $\Delta X$  y  $\Delta t$  respectivamente.

Los modelos empíricos utilizan modelos de regresión para establecer cual es la relación entre las variables de entrada y los parámetros y las variables de salida. Los modelos empíricos y dinámicos utilizan ecuaciones específicas como los métodos autorregresivos y de medias móviles o las cadenas de Markov.

Los modelos empíricos distribuidos son generalmente modelos de interpolación que a partir de observaciones de una variable en un conjunto de puntos, generan valores de dicha variable en todo el área de trabajo.

Los modelos estocásticos se utilizan como complemento a cualquiera de los anteriores. Se basan en la definición de las funciones de distribución de las variables estudiadas y en la utilización de números aleatorios para generar series de valores a partir de estas.

En todo caso, estas **ecuaciones**, junto con los **esquemas de flujo** para su resolución, constituyen hipótesis acerca del comportamiento de los procesos. Estas ecuaciones toman **variables de entrada** y producen una serie de **variables de salida** en función de **parámetros** que reflejan las características del territorio. Por otro lado existe un conjunto de **variables de estado** internas al sistema que se ven modificadas.

**Simulación por ordenador** es la resolución mediante un ordenador del conjunto de ecuaciones que conforman un modelo matemático reorganizadas de forma algorítmica mediante un programa informático.

## 4.4 Fases de la modelización

Los pasos a seguir en la definición de un modelo son:

1. **Identificación** es la determinación de los componentes y relaciones, así como de las ecuaciones matemáticas  $f()$  que representan el modelo.
2. **Calibración** es la determinación de los valores de los parámetros; bien midiéndolos en el campo o bien estimándolos mediante algún modelo empírico.
3. **Validación** es la obtención de series de las variables de estado y salida en un tiempo pasado para su comparación con valores reales de las mismas. El objetivo es evaluar el error cometido por el modelo y determinar si es asumible.
4. **Simulación** Obtención de series de las variables de estado y de salida en un tiempo futuro a partir de series simuladas de variables de entrada.
5. **Análisis de sensibilidad** Simulación pero alterando los valores de los parámetros para determinar la influencia de cada uno de ellos en los resultados aportados por el modelo.

A continuación se verá con más detalle la fase de **identificación**, en temas posteriores se verán las demás.

## 4.5 Identificación

Dentro de la identificación del modelo es importante tener claro que problema se va a modelizar y con que propósito. Si el problema no está claramente identificado es muy difícil que el modelo sirva para algo.

### Límites espaciales y temporales y escalas del modelo

Es necesario determinar los límites espaciales y temporales del sistema así como la resolución con que se va a modelizar. Estos límites marcan una frontera con el entorno del sistema que habrá que tener en cuenta al modelizar ya que puesto que ningún sistema ambiental es cerrado, habrá flujos de materia, energía o información hacia dentro y hacia fuera del sistema a través de estos límites.

Hay que tener en cuenta la importancia de las escalas espaciales y temporales sobre las que tienen lugar los procesos que tratamos de modelizar. En general diferentes procesos actúan a diferentes escalas. No es conveniente



aplicar modelos en escalas diferentes a aquellas para las que se han diseñado. En modelización la escala está más relacionada con la resolución espacial y temporal del modelo que con la extensión, espacial o temporal, sobre la que opera.

En el caso de los modelos distribuidos el espacio se suele dividir en celdillas del mismo tamaño (formato raster) debido a la facilidad de manejo de esta estructura de datos por parte de los ordenadores que, además, es una estructura muy utilizada por los Sistemas de Información Geográfica que suelen utilizarse como herramienta para la gestión de modelos. Es importante utilizar un tamaño de celdilla adecuado para los procesos que se pretenden modelizar. Pero si la resolución es demasiado grande el modelo consumirá muchos recursos de ordenador y será demasiado lento sin aportar mejores resultados.

Los modelos semidistribuidos requieren la elección de unidades espaciales adecuadas en cuanto a que su comportamiento sea homogéneo y suficientemente distinto del comportamiento de las demás unidades.

Los modelos dinámicos requieren además establecer el intervalo temporal adecuado, las consideraciones son similares al caso de la resolución espacial.

Finalmente hay que tener en cuenta que, al menos en el caso de los modelos de base física, la resolución espacial y temporal adecuadas están muy relacionadas. Si asumimos que los procesos necesitan ser descritos con una elevada resolución espacial, lo normal es que actúen con una velocidad elevada y por tanto necesiten un intervalo temporal reducido.

Este planteamiento puede resumirse en la denominada condición de Courant que establece que:

$$\frac{\Delta X}{\Delta t} > v_{max} \quad (4.4)$$

es decir que la razón entre resolución espacial y temporal representa la velocidad máxima a la que pueden producirse los procesos el modelo y por tanto debe ser mayor que la velocidad máxima de estos en la naturaleza. Si no fuese así nos encontraríamos con la paradoja de que un volumen dado de materia podría saltar de una celdilla del modelo a otra sin atravesar la intermedia.

### Conceptualización

Uno de los primeros pasos en el desarrollo de un modelo, y a la vez una de las mayores ventajas de la utilización de este tipo de herramientas es la necesidad de especificar de la manera más objetiva posible los conocimientos que el investigador o el equipo de trabajo tiene acerca del sistema o problema que se va a tratar de modelizar. Se hace necesaria una investigación bibliográfica para conocer cual es el conocimiento científico disponible acerca del problema a tratar y, en particular, que soluciones se han propuestos en trabajos de modelización anteriores.

En esta fase es necesario determinar cuales son los elementos y procesos que actúan en el sistema y cuales son los factores que influyen sobre estos. De este modo un modelo conceptual constituye una teoría acerca del funcionamiento del sistema. Si los resultados del modelo que se construye partiendo de esta teoría no se ajustan a la realidad, podemos concluir que la teoría se partida, el modelo conceptual, no era adecuada y debe revisarse.

En la conceptualización del problema no basta con tener en cuenta aspectos puramente científicos sino que también habrá que tener en cuenta los objetivos del modelo y las limitaciones técnicas con las que tenemos que trabajar. De nada serviría un modelo de prevención de avenidas que predijese perfectamente por donde va a desbordar el Segura 6 días después de que ocurriese la catástrofe.

### Construcción del modelo

Supone traducir un modelo conceptual a un lenguaje de programación y comprobar que el código generado funciona correctamente. Es importante, sin embargo, tener en cuenta que el hecho de que un programa funcione

correctamente no implica que el modelo que implementa sea correcto. La construcción de un modelo supone las siguientes fases:

**Modularización** , se trata de descomponer el conjunto de ecuaciones que van a constituir el modelo en porciones de código más fácilmente abordables. Se trata de generar, en la medida de lo posible, funciones pequeñas que realicen tareas concretas y que sean fáciles de programar y mantener. Las interrelaciones entre todos estos módulos o funciones constituye el programa

**Programación** , las diferentes ecuaciones se combinan utilizando las diferentes estructuras de control que los lenguajes de programación proporcionan. En muchos casos se reutiliza código de otros modelos. En definitiva se trata de escribir los algoritmos necesarios para la ejecución del modelo.

**Prueba** , cada módulo por separado y la integración de todos ellos deben ser probados para verificar que se ejecutan correctamente, que los programas no se cuelgan y que no aparecen mensajes de error.

**Verificación** , si un módulo logra ejecutarse con éxito, habrá que probarlo con un conjunto de parámetros y variables plausible

Existen programas que permiten generar pequeños modelos de manera gráfica como VENSIM, STELLA o SIMILE. Los SIG suelen permitir también la programación de pequeños modelos distribuidos (GRASS o PCRaster). Sin embargo cuando un modelo crece es preferible programarlo directamente en un lenguaje de programación como C. De esta manera el modelo se ejecuta más deprisa y consume menos recursos. El inconveniente de este enfoque es la dificultad de este tipo de lenguajes y la necesidad de programar todas las rutinas de acceso a los datos, representación gráfica, etc. Una opción intermedia podría ser el utilizar R como lenguaje de programación ya que incluye numerosas funciones de análisis de datos y representación gráfica que pueden utilizarse para reducir el código exclusivamente a las ecuaciones del modelo.