



# Estimación de parámetros

---

Prof. Cesar de Prada

ISA-UVA

[prada@autom.uva.es](mailto:prada@autom.uva.es)



# Indice

---

- Modelado
- Estimación de parámetros
- Validación de modelos



# Modelos

---

- Representación aproximada de la realidad
- Abstracción: Incluimos solo aquellos aspectos y relaciones que son de interés.
- Usos de los modelos: diseño, entrenamiento, que pasa si....., decisiones,...
- Distintos modelos de un proceso para distintas aplicaciones
- ¿Como generarlos, resolverlos, utilizarlos, validarlos?
  - Depende del proceso, de la aplicación, de las herramientas y datos disponibles,...

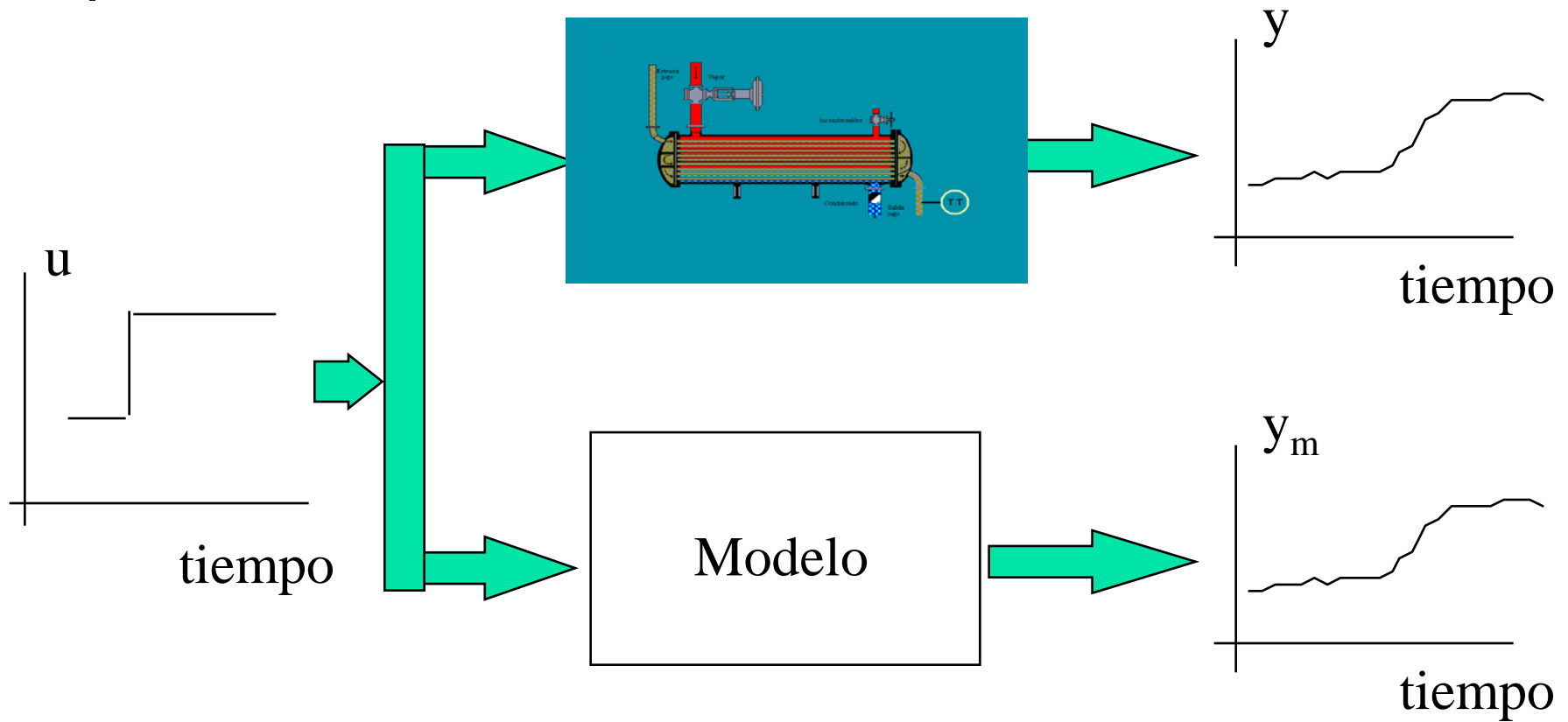


# ¿Qué es un modelo matemático?

---

- Conjunto de ecuaciones que relacionan las variables de interés del proceso y representan adecuadamente su comportamiento
- Distintas formulaciones para distintos objetivos y tipos de procesos
- Obtenidas a partir de un conjunto de hipótesis y suposiciones
- Compromiso entre facilidad de uso en una aplicación y la exactitud de la representación del mundo real

# Representación adecuada

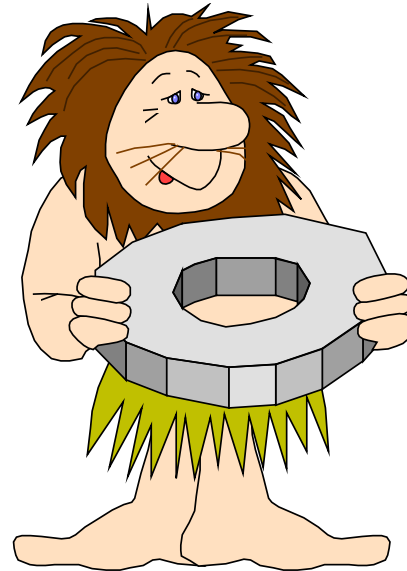


+ adecuación y facilidad de uso en la aplicación

# ¿Como obtener modelos?



Mediante razonamientos,  
usando leyes físicas,  
químicas, etc



Mediante experimentación  
y análisis de datos



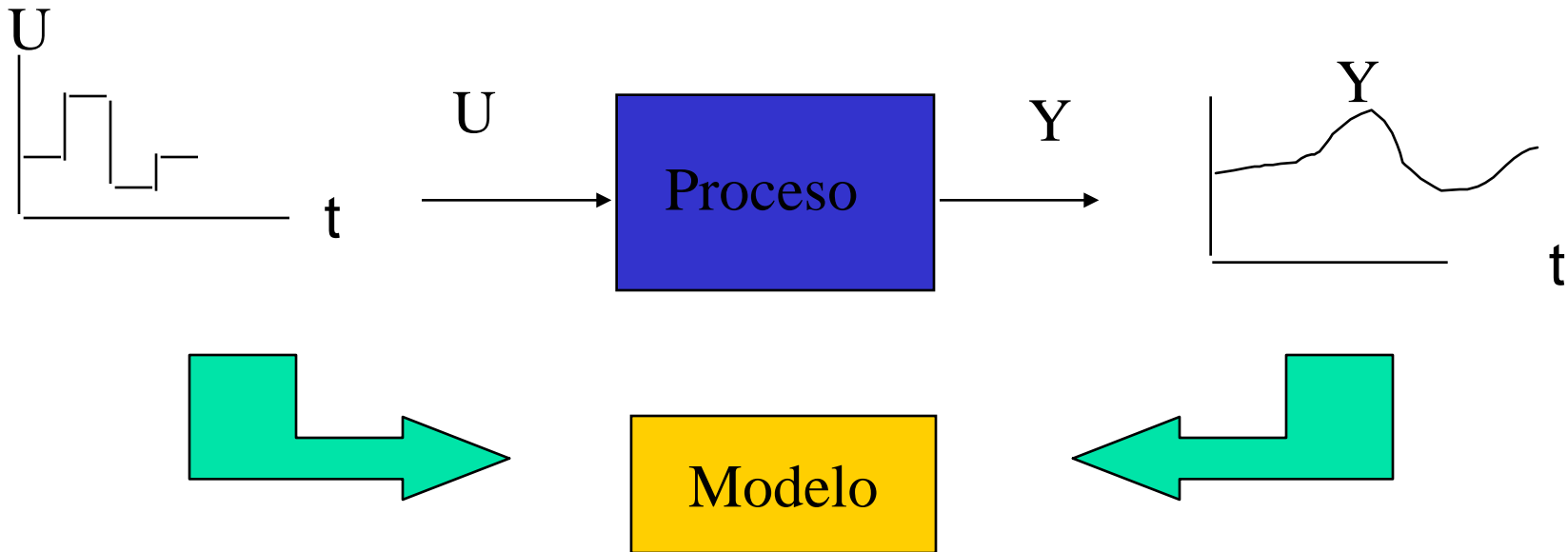
# Modelos de conocimiento

---

- Se obtienen mediante razonamientos y la aplicación de principios de conservación de masa, energía, momento, etc. y otras leyes particulares del dominio de aplicación
- Basados en un conjunto de hipótesis
- Tienen validez general si se cumplen las hipótesis de partida
- Requieren conocimiento profundo del proceso y de las leyes fisico-químicas

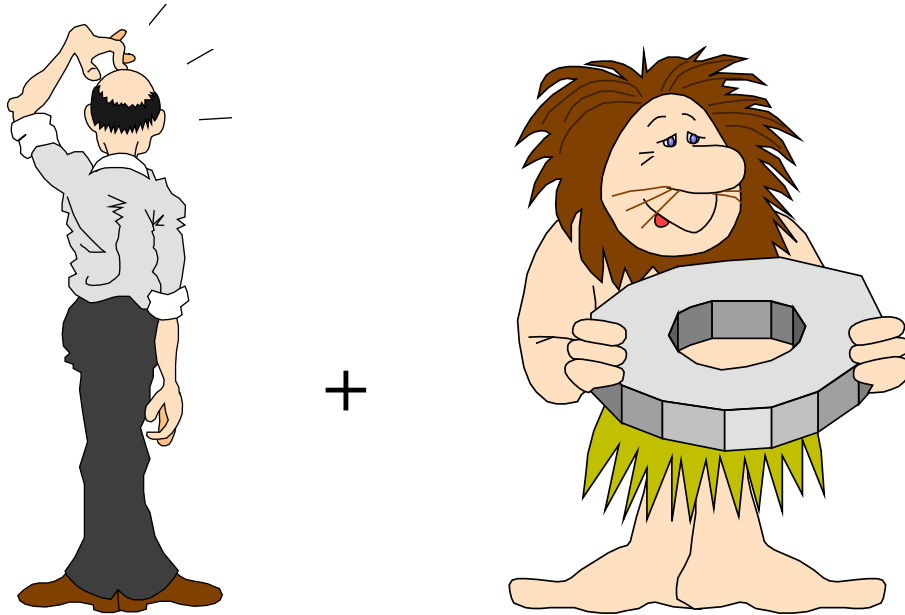
# Identificación

El modelo se obtiene a partir de datos experimentales de entrada-salida del proceso



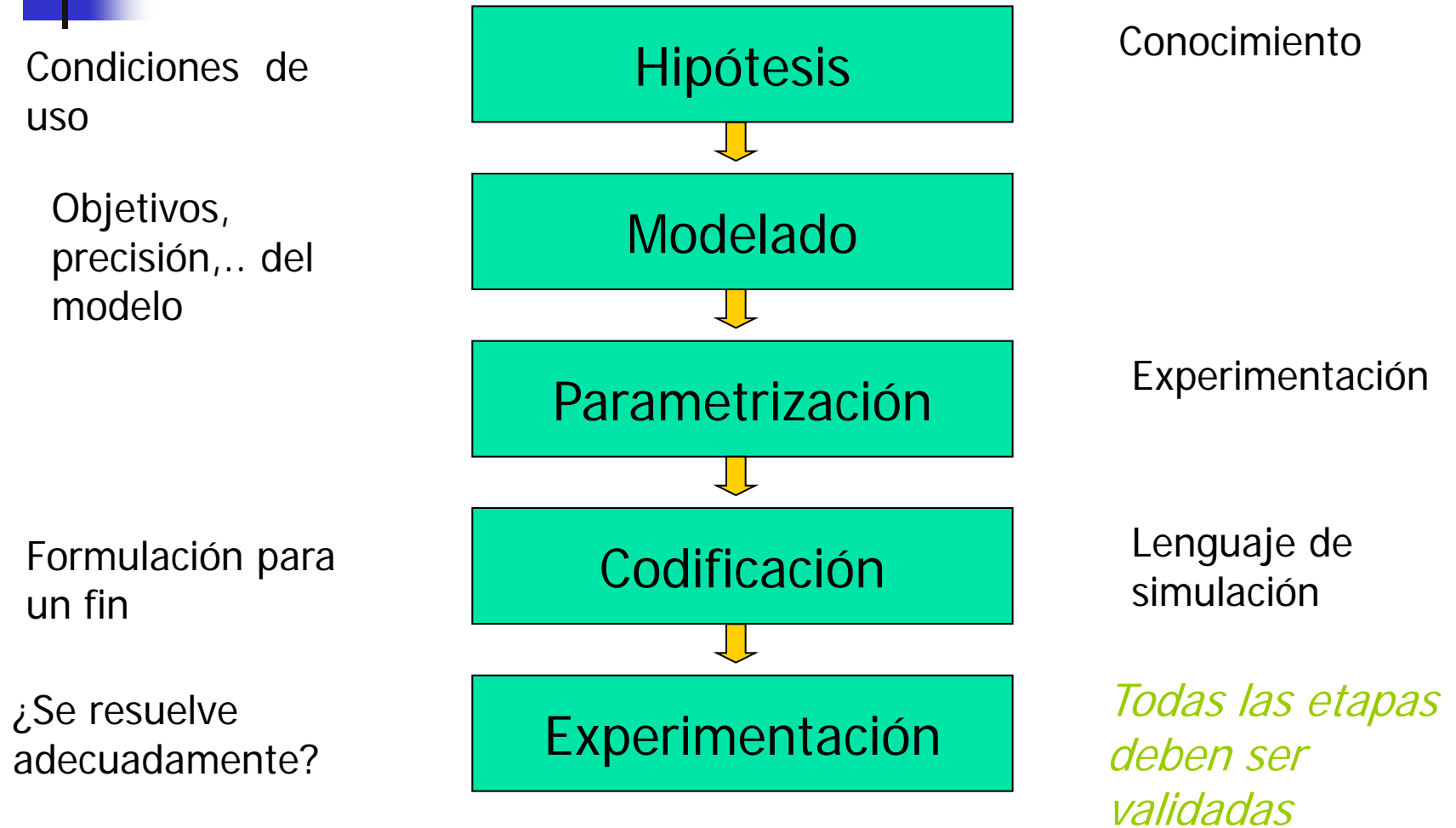


# Modelos

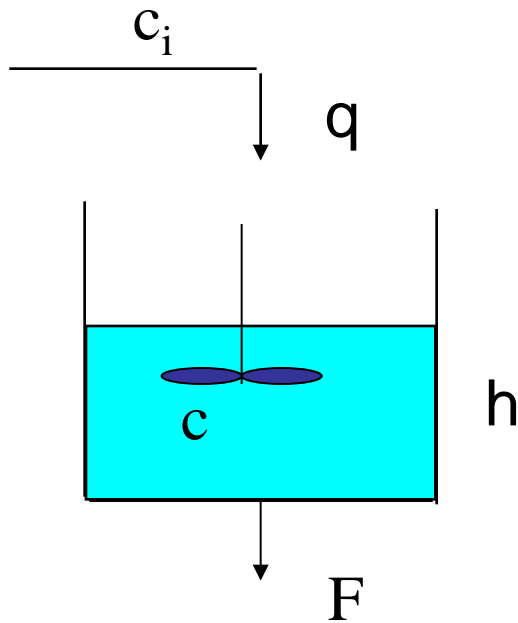


Un modelo siempre es una mezcla de conocimiento y de experimentación

# Modelos

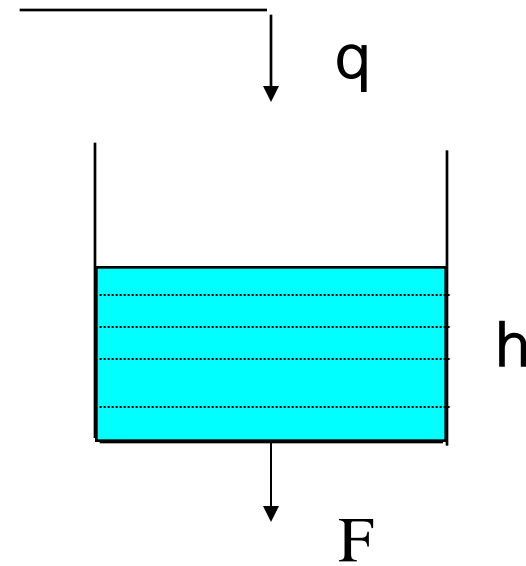


# Hipótesis



Mezcla perfecta

$$\frac{dVc}{dt} = qc_i - Fc$$



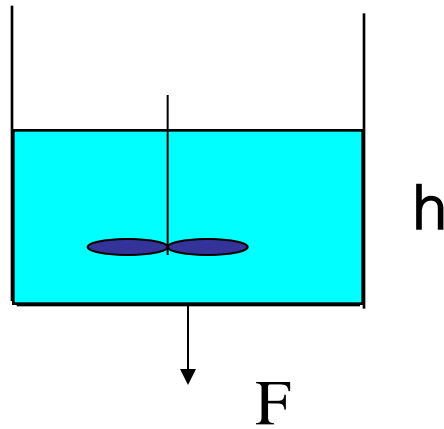
Flujo pistón

$$c(t) = c_i \left( t - \frac{h}{v} \right) = c_i \left( t - \frac{Ah}{Av} \right) = c_i \left( t - \frac{V}{F} \right)$$

# Ejemplo: Depósito

Conservación de masa

Acumulación =  
flujo entrada  $q$  - flujo salida  $F$



$$\frac{dm}{dt} = q\rho - F\rho$$

$$m = A\rho h \quad F = k\sqrt{h}$$

$$A \frac{dh}{dt} = q - k\sqrt{h}$$

$$V = Ah$$

## Parámetros

$m$  masa en el depósito  
 $A$  sección del depósito  
 $\rho$  densidad,  $k$  constante

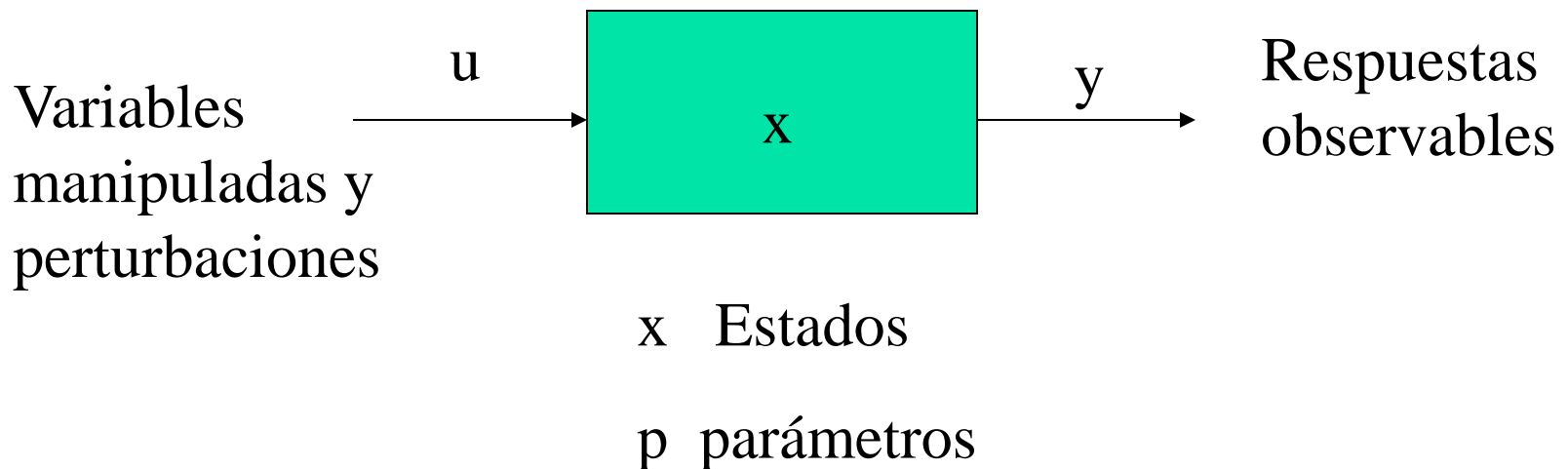
Ecuación diferencial  
no-lineal

Ecuación  
algebraica

# Modelos en variables de estado

$$\frac{d x(t)}{dt} = f(x(t), u(t), p, t)$$

$$y(t) = g(x, u(t), t)$$





# Modelado

---

$$\frac{d x(t)}{dt} = f(x(t), u(t), p, t)$$

$$y(t) = g(x, u(t), t)$$

Determinación de la estructura del modelo

Determinación del valor de sus parámetros

Normalmente siempre aparecen alternativas en la formulación del modelo y es necesario escoger entre distintas estructuras

En la etapa de validación pueden aplicarse distintos test para elegir que elección fue mas adecuada



# Parametrización

---

$$\frac{d x(t)}{dt} = f(x(t), u(t), p, t)$$

$$y(t) = g(x, u(t), t)$$

- Los parámetros  $p$  del modelo pueden obtenerse mediante bibliografía, documentación, etc. pero siempre hay parámetros cuyo valor no se puede determinar de forma sencilla y precisa y que deben ser estimados

# Ejemplo: reactor químico

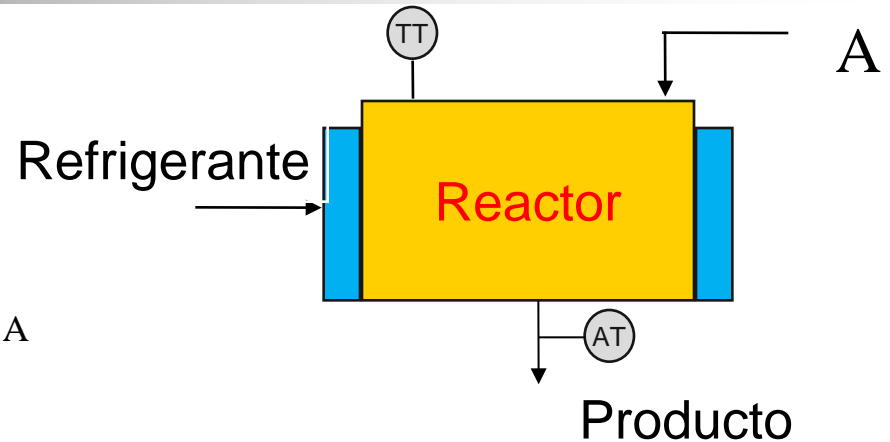


$$V \frac{dc_A}{dt} = Fc_{Ai} - Fc_A - Vke^{-E/RT}c_A$$

$$V \frac{dc_B}{dt} = -Fc_B + Vke^{-E/RT}c_A$$

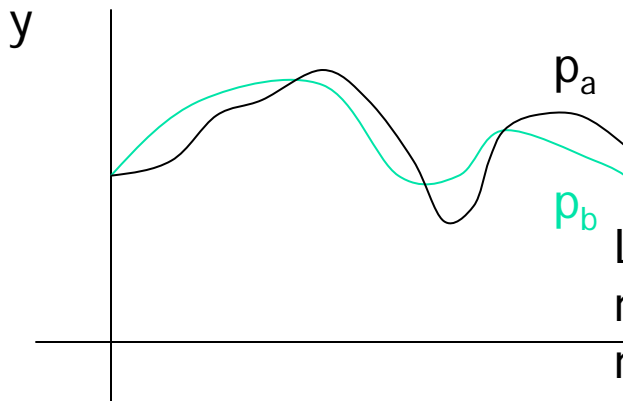
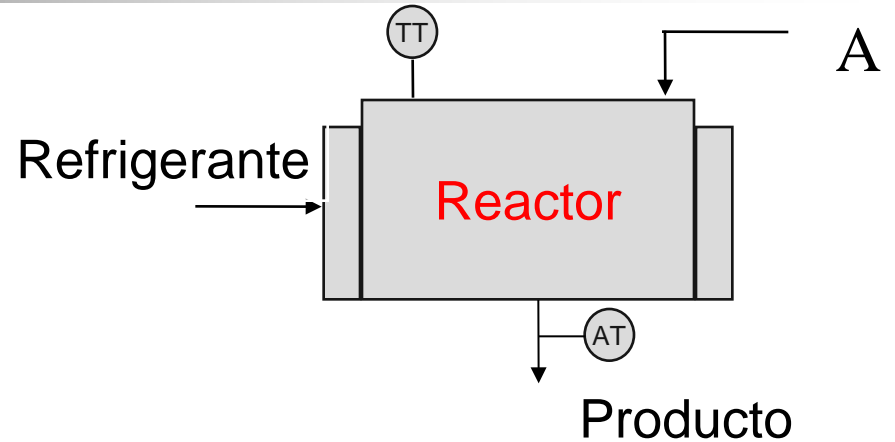
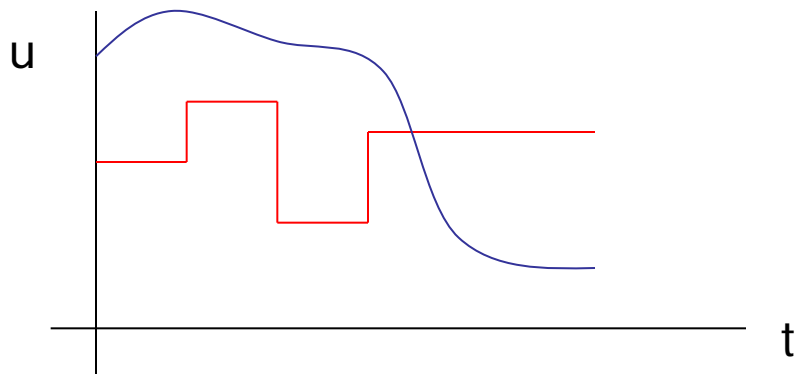
$$V\rho c_e \frac{dT}{dt} = F\rho c_e T_i - F\rho c_e T + Vke^{-E/RT}c_A \Delta H - UA(T - T_r)$$

$$V_r \rho_r c_{er} \frac{dT_r}{dt} = F_r \rho_r c_{er} T_{ri} - F_r \rho_r c_{er} T_r + UA(T - T_r)$$

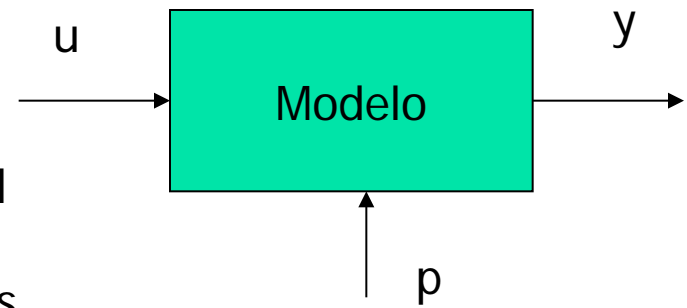




# Parametrización (calibración)



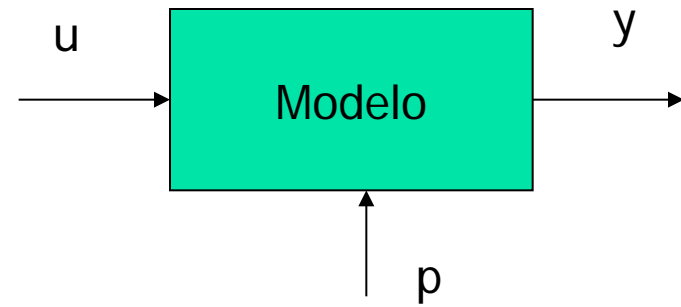
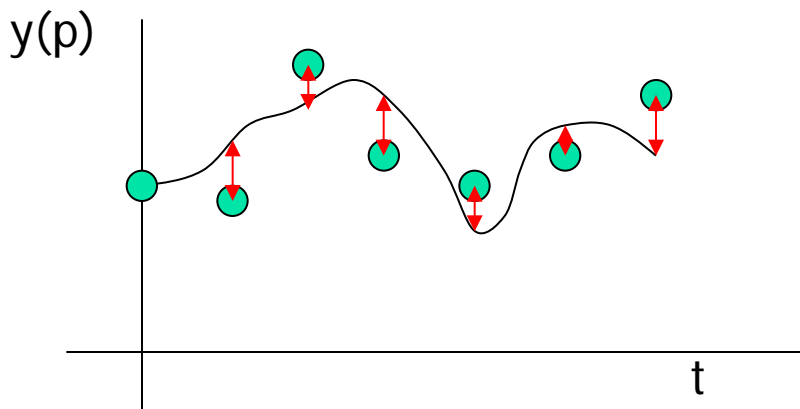
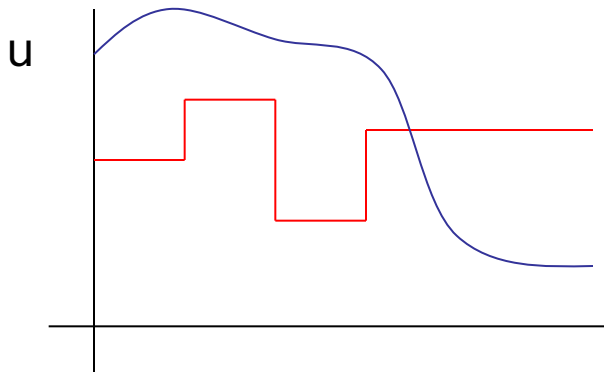
La respuesta del modelo ante las mismas entradas depende del valor de los parámetros



# Parametrización (calibración)

Si hay N valores experimentales de la salida se puede definir una función de error como:

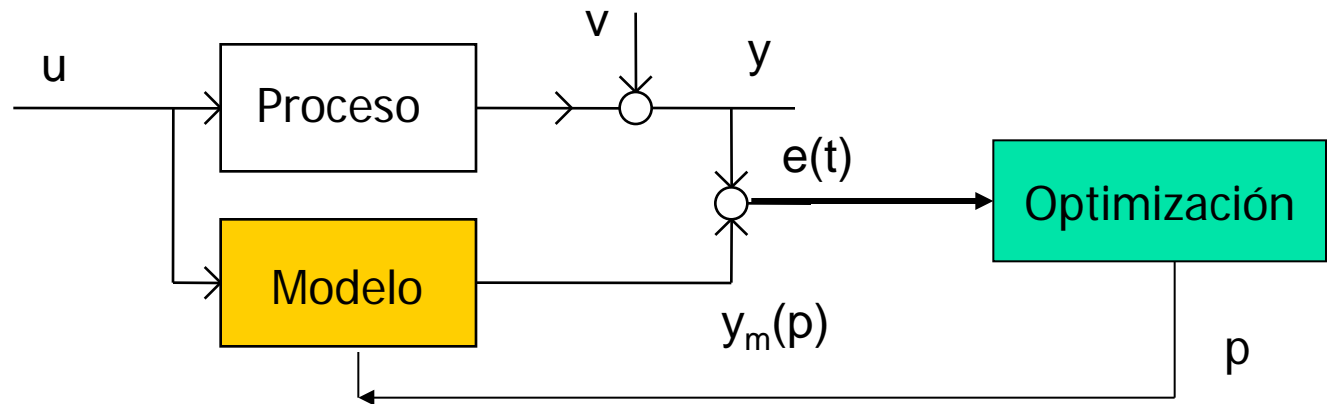
$$J = \frac{1}{N} \sum_{t=1}^N e(t)^2 = \frac{1}{N} \sum_{t=1}^N [y(t) - y_m(u, p, t)]^2$$



También pueden usarse otras expresiones para J

# Parametrización (calibración)

$$\min_p J = \min_p \frac{1}{N} \sum_{t=1}^N e(t)^2 = \min_p \frac{1}{N} \sum_{t=1}^N [y(t) - y_m(u, p, t)]^2$$



$V$  es normalmente una función no-lineal en los parámetros.  
Puede resolverse por minimización numérica



# Parametrización por optimización

---

$$\min_p J = \min_p \sum_{t=1}^N [y(t) - y_m(p, u, t)]^2$$

$$\dot{x}(t) = f(x(t), u(t)) \quad y_m(t) = g(x(t), u(t))$$

$$\underline{p} \leq p \leq \bar{p}$$

Los parámetros suelen tener significado físico y pueden acotarse

Problema de optimización dinámica con restricciones

Entre los parámetros a estimar pueden incluirse también, junto a los parámetros del modelo, estados iniciales desconocidos o entradas no medidas (constantes o parametrizables)

# Parametrización

$$\min_p J = \min_p \frac{1}{N} \sum_{t=1}^N e(t)^2 = \min_p \frac{1}{N} \sum_{t=1}^N [y(t) - y_m(u, p, t)]^2$$

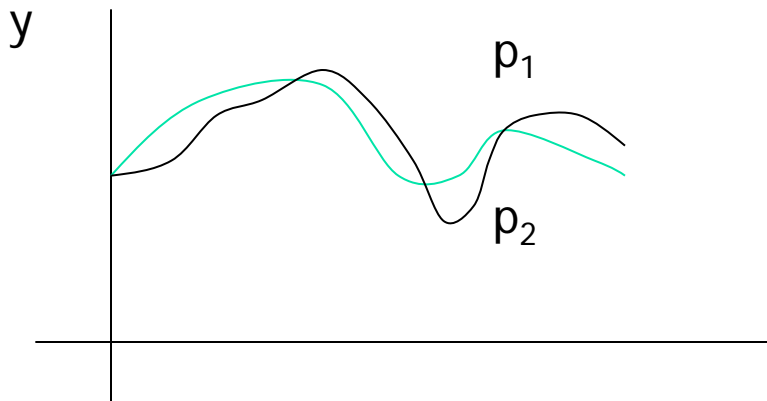
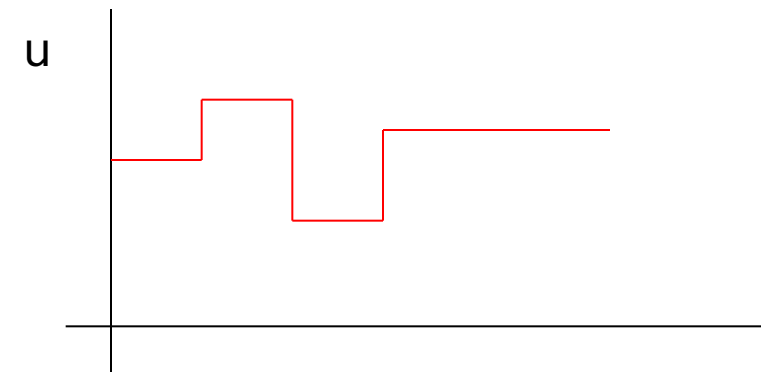


Si hay varias salidas a ajustar, puede plantearse un ajuste global:

$$\min_p \frac{1}{N} \sum_{t=1}^N \gamma_1 [y_1(t) - y_{m1}(p, t)]^2 + \gamma_2 [y_2(t) - y_{m2}(p, t)]^2 + \gamma_3 [y_3(t) - y_{m3}(p, t)]^2$$

Problemas de unidades y pesos relativos. Normalización

# ¿Qué parámetros estimar?



Es posible que, para un mismo experimento, la respuesta del modelo con dos valores diferentes  $p_1$  y  $p_2$  de un parámetro, sea muy o poco diferente

Solo debería escogerse un parámetro  $p$  para la optimización si la salida presenta una sensibilidad razonable ante cambios en  $p$



# Sensibilidad de las salidas

## Sensibilidad del índice J

---

$$J = \sum_{t=1}^N [y(t) - y_m(p, u, t)]^2$$

$$\dot{x}(t) = f(x(t), u(t))$$

$$y_m(t) = g(x(t), u(t))$$

$$S_{ij}(t) = \frac{\partial y_{mi}(t)}{\partial p_j}$$

$$\frac{\partial J}{\partial p_j}$$

Sensibilidad de la salida i del modelo respecto al parámetro j en relación a un experimento dado. Es función del tiempo

Sensibilidad del índice J respecto al parámetro j en relación a un experimento dado.

Resume el efecto durante todo el tiempo del experimento



# Sensibilidades

---

$$S_{ij}(t) = \frac{\partial y_i(t)}{\partial p_j}$$

$$s_{ij}(t) = \frac{p_j}{\bar{y}_i} \frac{\partial y_i(t)}{\partial p_j}$$

$$\begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1d} \\ S_{21} & S_{22} & \cdots & S_{2d} \\ \vdots & \vdots & \cdots & \vdots \\ S_{m1} & S_{m2} & \cdots & S_{md} \end{bmatrix}$$

Dificultad para comparar sensibilidades debido a la diferencia de unidades de cada  $S_{ij}$ . Deben usarse factores de escala

La norma de cada columna  $j$  de la matriz de sensibilidades da una idea de la importancia del parámetro  $p_j$  en el valor de las salidas





# Cálculo de las sensibilidades

---

Mediante cociente de incrementos, realizando simulaciones del modelo

$$S_{ij}(t) = \frac{\partial y_i(t)}{\partial p_j} \approx \frac{y_i(p + \Delta p, t) - y_i(p, t)}{\Delta p}$$

Las sensibilidades dependen, lógicamente, del punto  $p$  en que son evaluadas, y del experimento (a través de las entradas) que se considera



# Cálculo de las sensibilidades

$$J = \sum_{t=1}^N [y(t) - y_m(p, u, t)]^2$$

Pueden obtenerse también de forma exacta, integrando el sistema extendido

$$\dot{x}(t) = f(x(t), u(t), p)$$

$$y_m(t) = g(x(t), u(t), p)$$

$$\frac{\partial J}{\partial p} = 2 \sum_{t=1}^N [y(t) - y_m(p, u, t)] \frac{\partial y_m}{\partial p}$$

$$\frac{\partial \dot{x}}{\partial p} = \frac{d}{dt} \left( \frac{\partial x}{\partial p} \right) = \frac{\partial f}{\partial x} \left( \frac{\partial x}{\partial p} \right) + \frac{\partial f}{\partial p}$$

$$\frac{\partial y_m}{\partial p} = \frac{\partial g}{\partial x} \frac{\partial x}{\partial p} + \frac{\partial g}{\partial p}$$

Integrando esta ecuación junto a las del modelo, es posible obtener la evolución de  $\partial x / \partial p$  y por tanto las sensibilidades



# Cálculo de las sensibilidades

---

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t))$$

$$\frac{d}{dt} \frac{\partial \mathbf{x}}{\partial \mathbf{p}} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{p}} + \frac{\partial \mathbf{f}}{\partial \mathbf{p}}$$

$$\mathbf{s} = \frac{\partial \mathbf{x}}{\partial \mathbf{p}} \quad \frac{d\mathbf{s}}{dt} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \mathbf{s} + \frac{\partial \mathbf{f}}{\partial \mathbf{p}}$$

El jacobiano de esta ecuación, necesario para integrar las sensibilidades  $\mathbf{s}$ , es el mismo que el de la ecuación original

$$\frac{\partial}{\partial \mathbf{s}} \left( \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \mathbf{s} + \frac{\partial \mathbf{f}}{\partial \mathbf{p}} \right) = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}$$

Por tanto puede integrarse el modelo original y si su Jacobiano  $\partial \mathbf{f} / \partial \mathbf{x}$  esta accesible, usarlo para la integración de las sensibilidades sin necesidad de volver a evaluarlo. El código DASPK usa este método.



# Identificabilidad

---

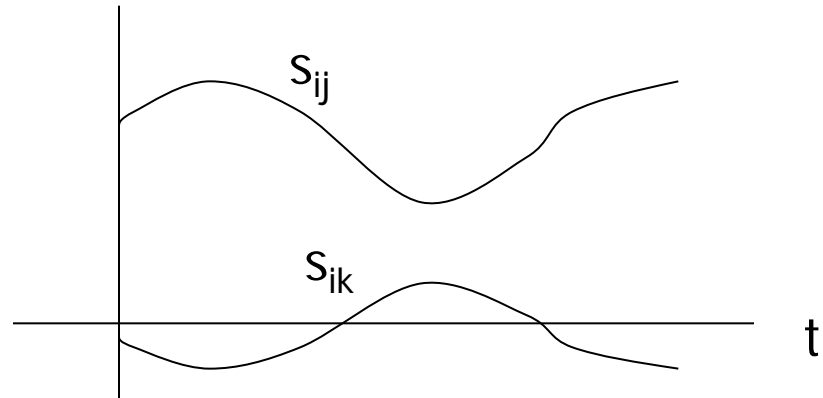
Independientemente del valor de la sensibilidad, puede ocurrir que el efecto sobre la salida de un cambio en un parámetro se anule por un cambio en otro parámetro hecho simultáneamente. Se dice entonces que hay un grado de co-linealidad en las sensibilidades frente a los parámetros, lo que dificulta la identificación

La identificabilidad es una propiedad estructural que depende de cómo los parámetros aparecen en el modelo, pero también de las medidas disponibles

# Identificabilidad

Examinando la matriz de sensibilidades, puede verse si hay un cierto grado de co-linealidad a través del rango de (sub-bloques) de la misma

$$\begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1d} \\ S_{21} & S_{22} & \cdots & S_{2d} \\ \vdots & \vdots & \cdots & \vdots \\ S_{m1} & S_{m2} & \cdots & S_{md} \end{bmatrix}$$



La co-linealidad dificulta la identificación

# Identificabilidad

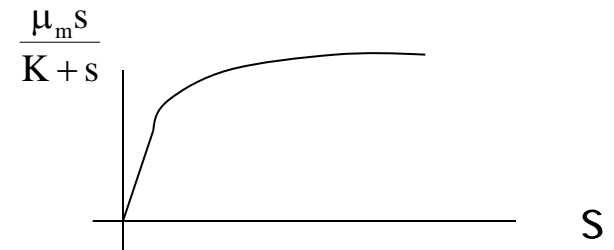
Ejemplo: En un reactor, sin medida de temperatura, los parámetros  $k$  y  $E$ , no pueden ser estimados independientemente

$$V \frac{d c_B}{d t} = -F c_B + V k e^{-E/RT} c_A$$

Si  $V$  y  $F$  fueran desconocidos, se podría identificar el cociente  $F/V$  pero no los términos individuales

$$\frac{\mu_m s}{K + s}$$

En el modelo de Monod, con datos limitados a sustratos  $s$  pequeños, no es posible identificar sino el cociente  $\mu_m/K$ . Con  $s$  grandes solo  $\mu_m$





# Re-parametrización

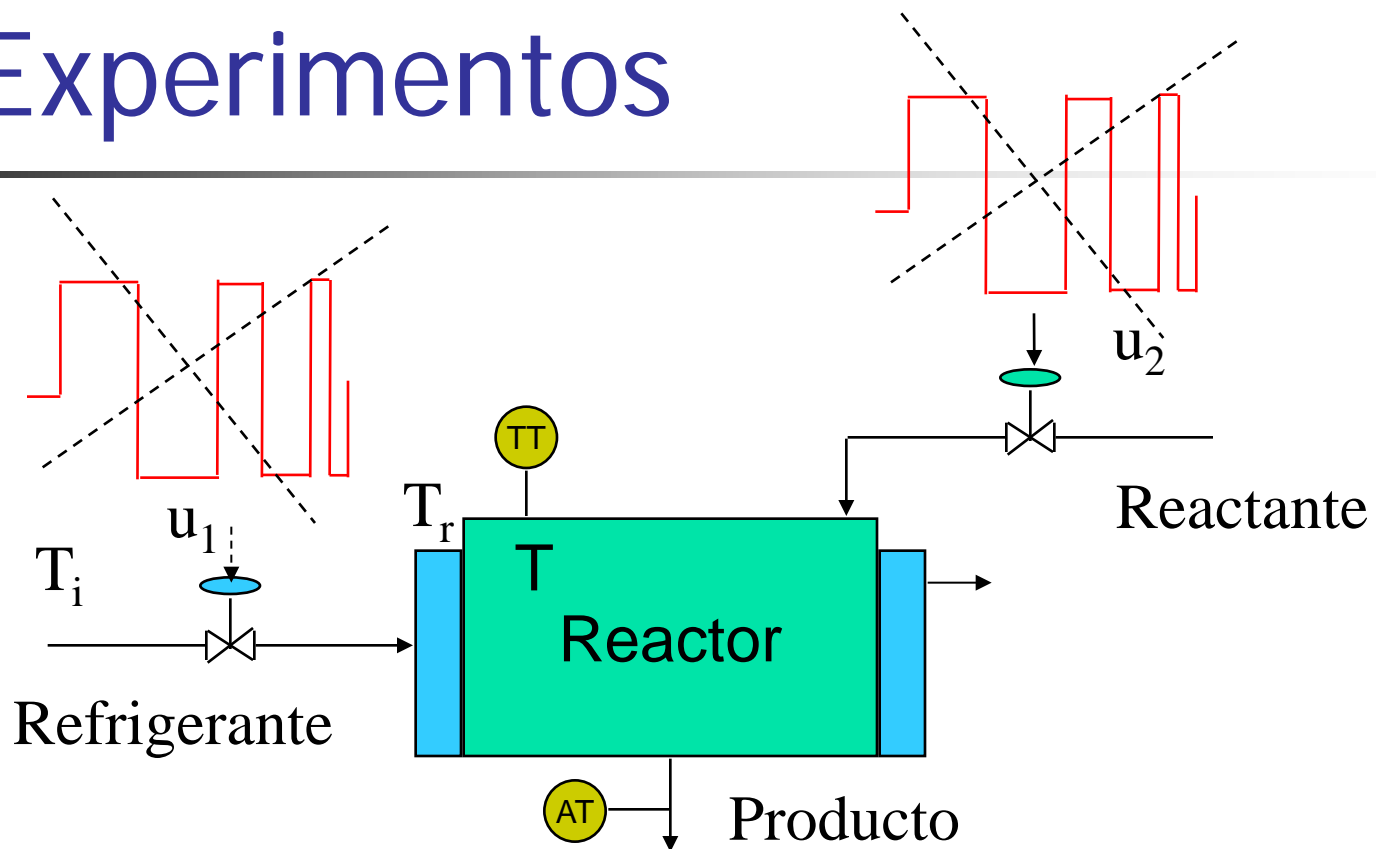
---

A veces solo es posible identificar combinaciones de parámetros, o se pueden hacer cambios de variables para obtener una estructura más sencilla de identificar.

Si se ha hecho una transformación de los parámetros a estimar, o se han hecho cambios de variable para facilitar la identificabilidad o linealidad del modelo, debe tenerse en cuenta que:

- ✓ Las características estadísticas de las nuevas variables son diferentes a las de los datos
- ✓ Las regiones de confianza de los nuevos parámetros son diferentes de las de los originales

# Experimentos



Si hay varias entradas, los cambios en las mismas deben de estar incorrelacionados para que el algoritmo de optimización pueda distinguir los efectos de cada entrada en las salidas





# Experimentos

---

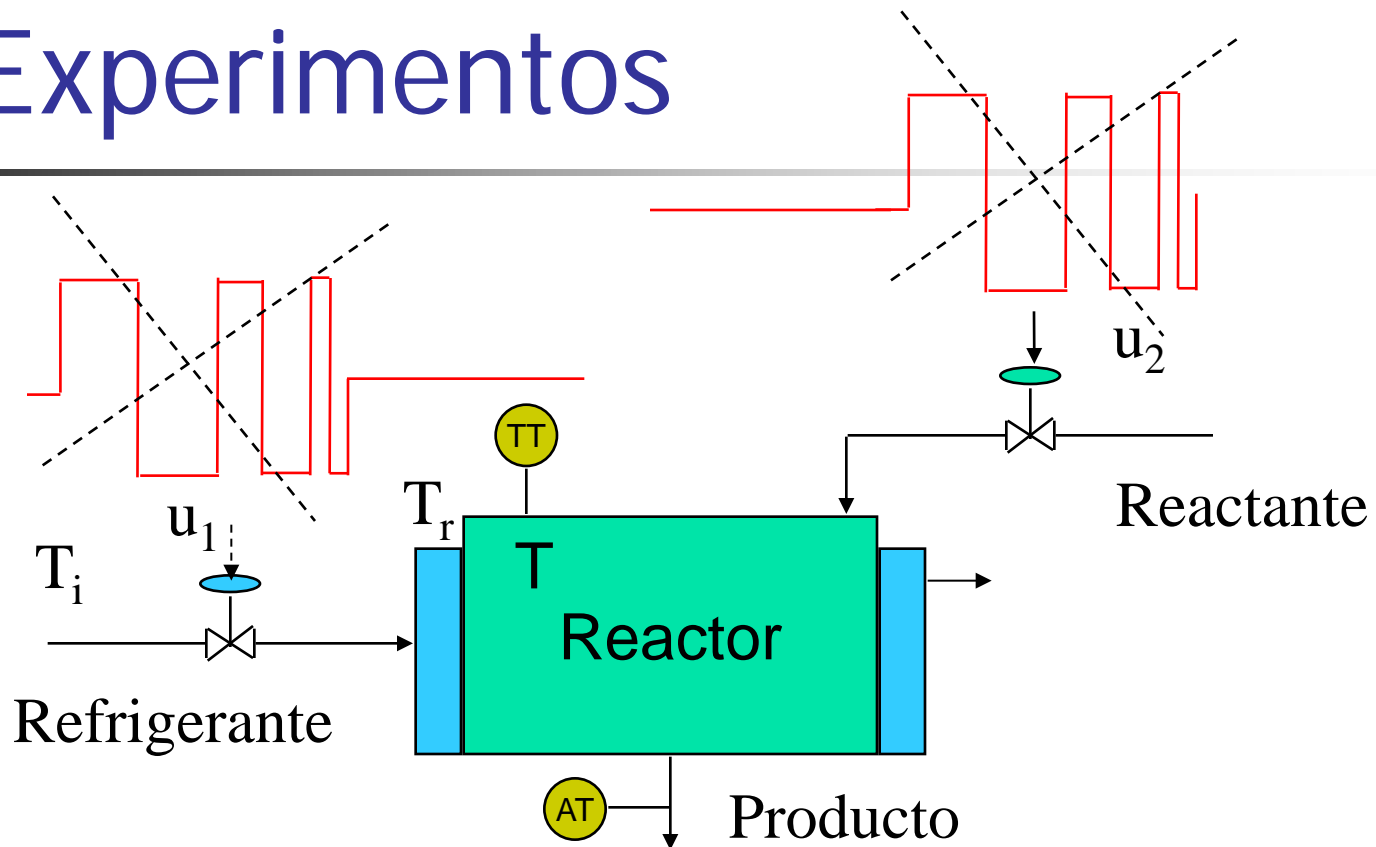
Si se aplican señales proporcionales a las entradas,  $u_1 = \alpha u_2$ , simultáneamente, hay muchas combinaciones de los modelos  $M_1$  y  $M_2$  que se ajustan a las mismas parejas de entradas y salidas

$$y(t) = M_1 u_1(t) + M_2 u_2(t) = (\alpha M_1 + M_2) u_2(t)$$

Por otro lado, los experimentos deben cubrir el rango en amplitud y frecuencia adecuado para excitar las dinámicas fundamentales del proceso, y los valores de las sensibilidades de  $J$  antes los parámetros a estimar deben presentar unos valores razonables.

En caso contrario debe repetirse el experimento

# Experimentos



Excitar solo una entrada manteniendo las otras constantes y luego hacer lo mismo con las otras, no es una buena política: alarga el tiempo del experimento y no proporciona datos en condiciones realistas de funcionamiento



# Diseño de experimentos

---

Cabe preguntarse cual es el experimento que mejor facilita la estimación de los parámetros de un modelo de acuerdo al criterio de minimizar:

$$J(p) = \sum_{t=1}^N [y(t) - y_m(p, u, t)]' Q [y(t) - y_m(p, u, t)]$$

Contexto  
Multivariable

Se suele escoger  $Q=C^{-1}$ , siendo  $C$  la matriz de covarianza del ruido de las medidas

Para un valor de los parámetros  $p$  en torno al valor óptimo  $p^*$  se tendrá:

$$J(p^* + \delta p) = \sum_{t=1}^N [y(t) - y_m(p^* + \delta p, u, t)]' Q [y(t) - y_m(p^* + \delta p, u, t)]$$

Haciendo la aproximación:

$$y_m(p^* + \delta p, u, t) \approx y_m(p^*, u, t) + \left. \frac{\partial y_m}{\partial p} \right|_{p^*} \delta p$$



# Diseño de experimentos

$$\begin{aligned} J(\mathbf{p}^* + \delta\mathbf{p}) &= \sum_{t=1}^N \left[ y(t) - y_m(\mathbf{p}^*, \mathbf{u}, t) - \left. \frac{\partial y_m}{\partial \mathbf{p}} \right|_{\mathbf{p}^*} \delta\mathbf{p} \right]' \mathbf{Q} \left[ y(t) - y_m(\mathbf{p}^*, \mathbf{u}, t) - \left. \frac{\partial y_m}{\partial \mathbf{p}} \right|_{\mathbf{p}^*} \delta\mathbf{p} \right] = \\ &= J(\mathbf{p}^*) + \delta\mathbf{p}' \sum_{t=1}^N \left[ \left. \frac{\partial y_m}{\partial \mathbf{p}} \right|_{\mathbf{p}^*}' \mathbf{Q} \left. \frac{\partial y_m}{\partial \mathbf{p}} \right|_{\mathbf{p}^*} \right] \delta\mathbf{p} - 2 \sum_{t=1}^N [y(t) - y_m(\mathbf{p}^*, \mathbf{u}, t)]' \mathbf{Q} \left. \frac{\partial y_m}{\partial \mathbf{p}} \right|_{\mathbf{p}^*} \delta\mathbf{p} \end{aligned}$$

y como (sin restricciones):

$$\left. \frac{\mathbf{J}(\mathbf{p})}{\partial \mathbf{p}} \right|_{\mathbf{p}^*} = 2 \sum_{t=1}^N [y(t) - y_m(\mathbf{p}^*, \mathbf{u}, t)]' \mathbf{Q} \left. \frac{\partial y_m}{\partial \mathbf{p}} \right|_{\mathbf{p}^*} = 0$$

resulta:

$$J(\mathbf{p}^* + \delta\mathbf{p}) = J(\mathbf{p}^*) + \delta\mathbf{p}' \sum_{t=1}^N \left[ \left. \frac{\partial y_m}{\partial \mathbf{p}} \right|_{\mathbf{p}^*}' \mathbf{Q} \left. \frac{\partial y_m}{\partial \mathbf{p}} \right|_{\mathbf{p}^*} \right] \delta\mathbf{p}$$



# Diseño de experimentos

---

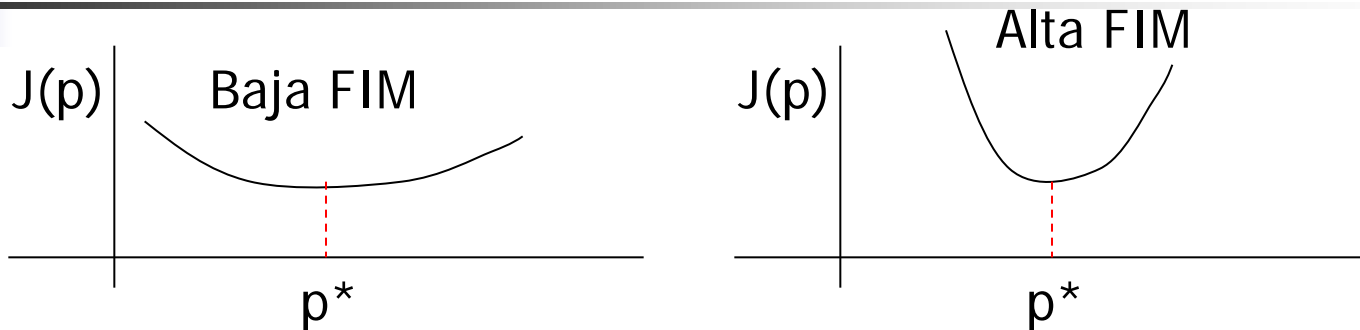
Para facilitar la estimación, interesa que  $J(\mathbf{p}^* + \delta\mathbf{p}) - J(\mathbf{p}^*)$  sea lo mayor posible, de modo que unos parámetros distintos del óptimo den un valor de  $J$  claramente diferente del óptimo.

$$E\{J(\mathbf{p}^* + \delta\mathbf{p}) - J(\mathbf{p}^*)\} \approx \delta\mathbf{p}' \left[ \sum_{t=1}^N \left( \frac{\partial \mathbf{y}_m(t)}{\partial \mathbf{p}} \right)' \mathbf{Q} \left( \frac{\partial \mathbf{y}_m(t)}{\partial \mathbf{p}} \right) \right] \delta\mathbf{p}$$

Para ello puede diseñarse un experimento que maximice la Matriz de Información de Fisher (**FIM**) (o alguna norma o función asociada):

$$\mathbf{Q} = \mathbf{C}^{-1} \left[ \sum_{t=1}^N \left( \frac{\partial \mathbf{y}_m(\mathbf{p}^*, t)}{\partial \mathbf{p}} \right)' \mathbf{C}^{-1} \left( \frac{\partial \mathbf{y}_m(\mathbf{p}^*, t)}{\partial \mathbf{p}} \right) \right]$$

# FIM



Si no hay dependencia lineal entre las sensibilidades, la matriz de información de Fisher debe ser de rango completo.

El cociente entre el máximo y mínimo valor propio de la FIM de por tanto una medida de la identificabilidad de los parámetros de un modelo con un experimento dado.

$$\left[ \sum_{t=1}^N \left( \frac{\partial y_m(t)}{\partial p} \right)' Q \left( \frac{\partial y_m(t)}{\partial p} \right) \right]$$



# Parametrización por optimización

---

$$\min_p J = \sum_{t=1}^N [y(t) - y_m(p, u, t)]^2$$

$$\dot{x}(t) = f(x(t), u(t)) \quad y_m(t) = g(x(t), u(t))$$

$$\underline{p} \leq p \leq \bar{p}$$

- ✓ Elección del tipo de índice J
- ✓ Problemas de resolución numérica
  - Simulación ( enfoque secuencial)
  - Discretización (enfoque simultaneo)
- ✓ Problemas de estimación de estados



# Otras funciones objetivo J

---

## Maximum Likelihood (ML)

Para un valor de los parámetros  $p$ , ¿cual es la probabilidad de que los resultados del modelo coincidan con los datos experimentales? Se plantea la estimación como encontrar  $p$  de modo que se maximice esa probabilidad

$$\max_p P(y_m(0) = y(0), y_m(1) = y(1), \dots, y_m(N) = y(N) | p)$$

Se suele formular como

$$\min_p -\ln(P(y_m(0) = y(0), y_m(1) = y(1), \dots, y_m(N) = y(N) | p))$$



# Mínimos cuadrados ponderados (WLS)

Si los datos experimentales están distribuidos de forma normal en torno a las predicciones del modelo y son independientes, la probabilidad de obtener como salida del modelo los datos experimentales es:

$$\begin{aligned} L(y_m(1) = y(1), \dots, y_m(N) = y(N) | p) &= \\ &= \prod_{t=1}^N \frac{1}{\sqrt{2\pi\sigma(t)}} \exp\left(-\frac{1}{2} \sum_{t=1}^N \left(\frac{y(t) - y_m(y, p)}{\sigma(t)}\right)^2\right) \end{aligned}$$

Tomando neperianos, maximizar esta función equivale a

$$\min_p \sum_{t=1}^N \frac{1}{\sigma(t)^2} (y(t) - y_m(t, p))^2$$

$\sigma^2$  varianza del error de medida



# WLS

---

$$J(p) = \min_p \sum_{t=1}^N \frac{1}{\sigma(t)^2} (y(t) - y_m(t, p))^2$$

Si los errores son independientes y están de verdad distribuidos normalmente,  $J(p)$  debe seguir una distribución  $\chi^2$  con  $N-d$  grados de libertad

Puede usarse como un test de las hipótesis anteriores o de la bondad del modelo



# Otras funciones objetivo

---

$$\min_p \sum_{t=1}^N |y(t) - y_m(t, p)|$$

Norma 1  
Pesa por igual todos los errores

$$\min_p \max_t |y(t) - y_m(t, p)|$$

Norma  $\infty$   
minimiza el mayor de los errores

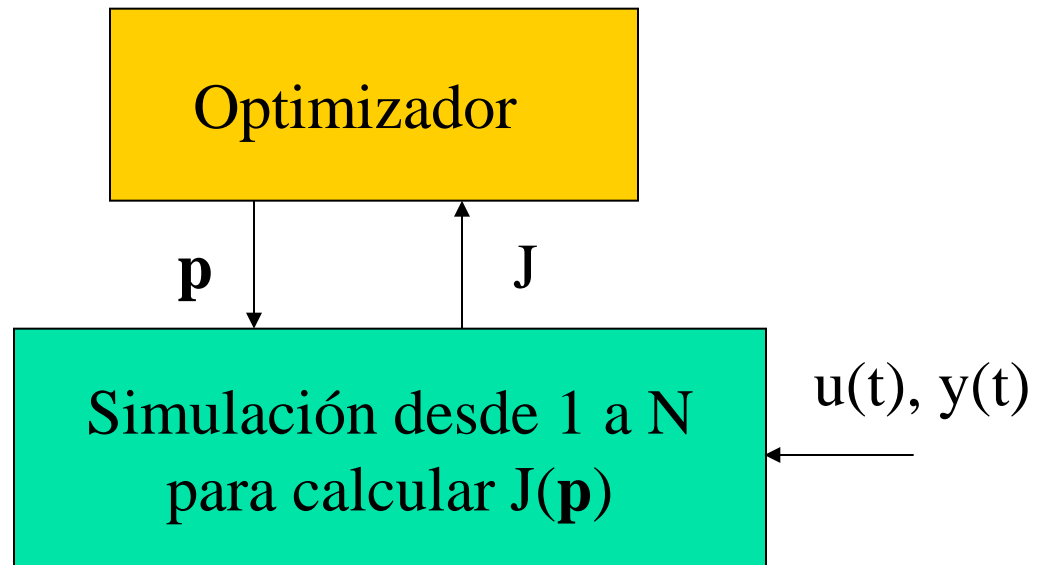
# Resolución con simulación

$$\min_{\mathbf{p}} J = \sum_{t=1}^N [y(t) - y_m(\mathbf{p}, u, t)]^2$$

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), u(t)) \quad y_m(t) = \mathbf{g}(\mathbf{x}(t), u(t))$$

$$\underline{\mathbf{p}} \leq \mathbf{p} \leq \bar{\mathbf{p}}$$

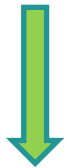
Software NLP



# Enfoque secuencial

Optimizador NLP de  $J(p)$   
con respecto a  $p$

Valores de  
 $p$

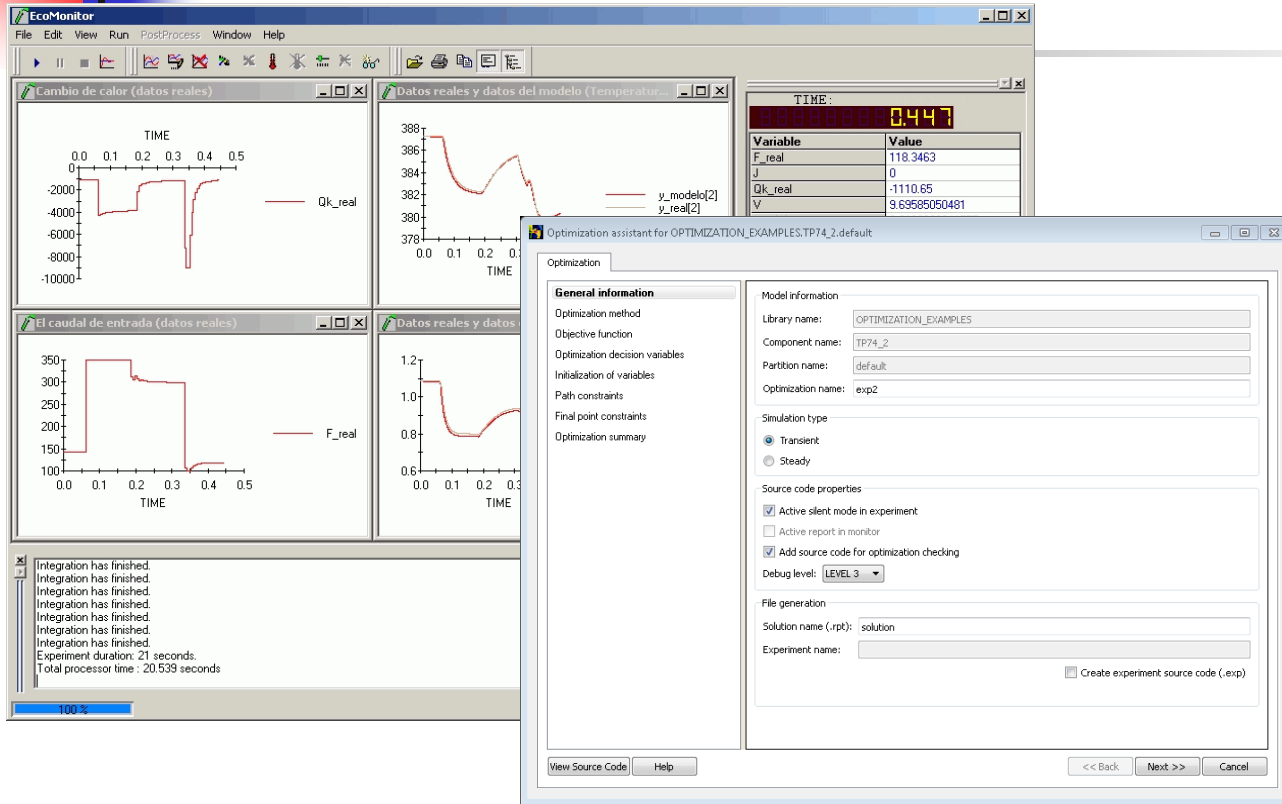


Valores de  
 $J(p), g(p)$

Simulador dinámico que  
calcula los valores de  $x$   
solución del DAE, así como  
de  $J(p), g(p)$

- ✓ El optimizador solo considera a  $p$  como variables del problema
- ✓ El modelo DAE se resuelve rigurosamente
- ✓ Dificultades con las restricciones de camino, sistemas inestables y el cálculo de gradientes

# Software



Entornos de simulación unidos a solvers NLP

Asistentes para la definición del problema y generación automática de código de optimización

Muy importante el cálculo de sensibilidades para la calidad de la solución.  
Errores relativos Simulación / Optimización

EcosimPro,  
gProms,  
Dymola,...

# Enfoque simultaneo

$$\min_p J(p) = \int_0^T L(x, u, p) dt$$

$$F(\dot{x}, x, u, p) = 0$$

$$g(x, u, p) \leq 0$$

$$F(\dot{x}, x, u, p) = 0$$



$$F\left(\frac{x(t + \Delta t) - x(t)}{\Delta t}, x(t + \Delta t), u, p\right) = 0$$

$$t = 0, \Delta t, 2\Delta t, \dots$$

Sistema de ecuaciones algebraicas

Discretizar totalmente las ecuaciones

$$\int_0^T L(x, u, p) dt$$



$$\sum_{j=0}^N [L(x(j), u(j), p)] \Delta t$$



# Resolución por Discretización

---

$$\min_{\mathbf{p}, \mathbf{x}} J = \sum_{j=1}^N [y(t) - y_m(\mathbf{p}, \mathbf{u}, t)]^2$$

$$\mathbf{x}(1) = f(\mathbf{x}(0), \mathbf{u}(0), \mathbf{p}) \quad \mathbf{y}(0) = \mathbf{g}(\mathbf{x}(0), \mathbf{u}(0), \mathbf{p})$$

$$\mathbf{x}(2) = f(\mathbf{x}(1), \mathbf{u}(1), \mathbf{p}) \quad \mathbf{y}(1) = \mathbf{g}(\mathbf{x}(1), \mathbf{u}(1), \mathbf{p})$$

$$\mathbf{x}(3) = f(\mathbf{x}(2), \mathbf{u}(2), \mathbf{p}) \quad \mathbf{y}(2) = \mathbf{g}(\mathbf{x}(2), \mathbf{u}(2), \mathbf{p})$$

.....

$$\underline{\mathbf{p}} \leq \mathbf{p} \leq \bar{\mathbf{p}}$$

Se incrementa el número de variables de decisión con los estados

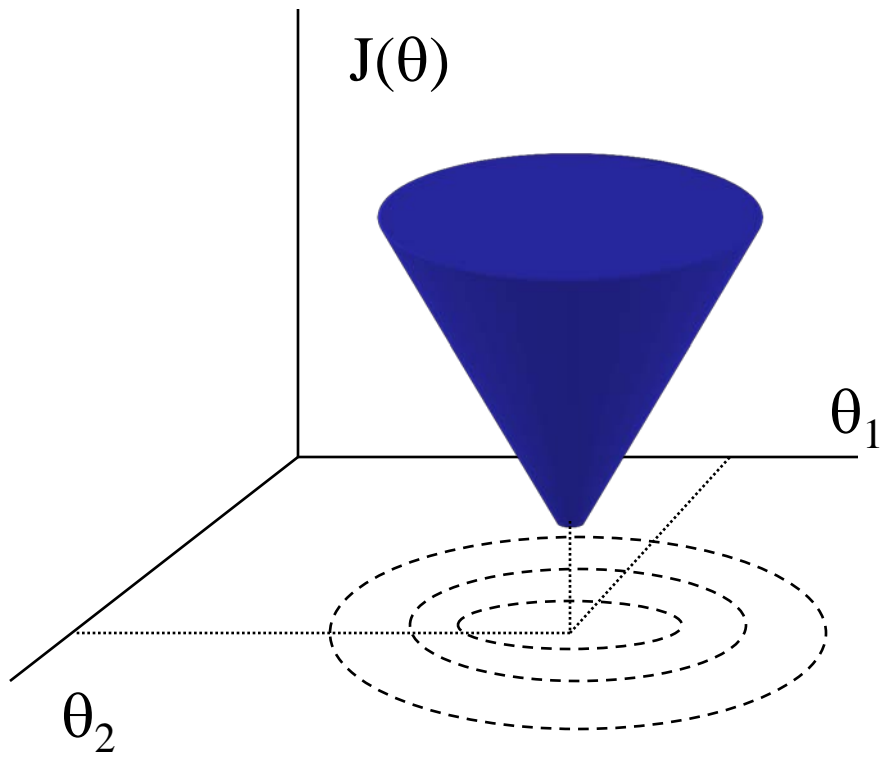
Es posible imponer restricciones sobre el estado

Resolución con SQP / IP

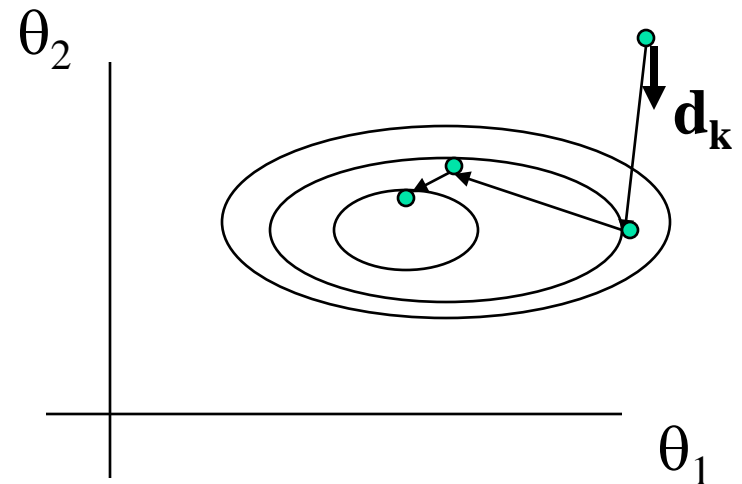
La discretización puede ser inexacta, colocación ortogonal



# Minimización de funciones



Restricciones en  $p$

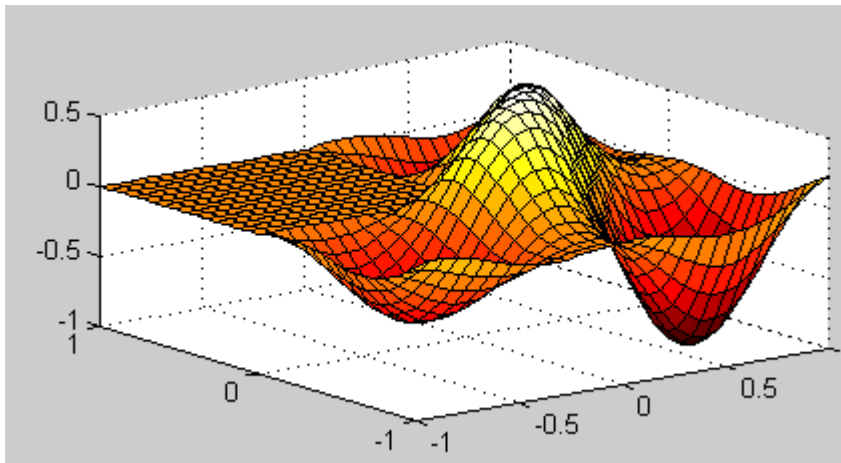


$$\theta_{k+1} = \theta_k + \alpha \mathbf{d}_k$$

Métodos iterativos

$\mathbf{d}$  dirección de búsqueda  
 $\alpha$  longitud del paso

# Optimización local / global



Muchos métodos de optimización conducen a soluciones locales que no proporcionan los mejores valores de los parámetros

Puede convenir utilizar métodos globales:

- ✓ Estocásticos (algoritmos evolutivos, etc.)
- ✓ Deterministas ( $\alpha$ -BB)



# Distintos problemas asociados a un modelo

---

$$\frac{d \mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}, t)$$

$$\mathbf{y}(t) = \mathbf{g}(\mathbf{x}, \mathbf{u}(t), t)$$

- ¿Que se supone conocido para estimar el resto?
- Reconciliación de datos
- Estimación de estados y perturbaciones
- Estimación de parámetros
- Verificación
- Validación

# Reconciliación de datos

$$\frac{d \mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}, t)$$

Suele plantearse en estado estacionario

$$\mathbf{y}(t) = \mathbf{g}(\mathbf{x}, \mathbf{u}(t), t)$$

El modelo se supone perfecto y se trata de estimar las modificaciones en los datos experimentales que hacen que estos se ajusten exactamente al modelo



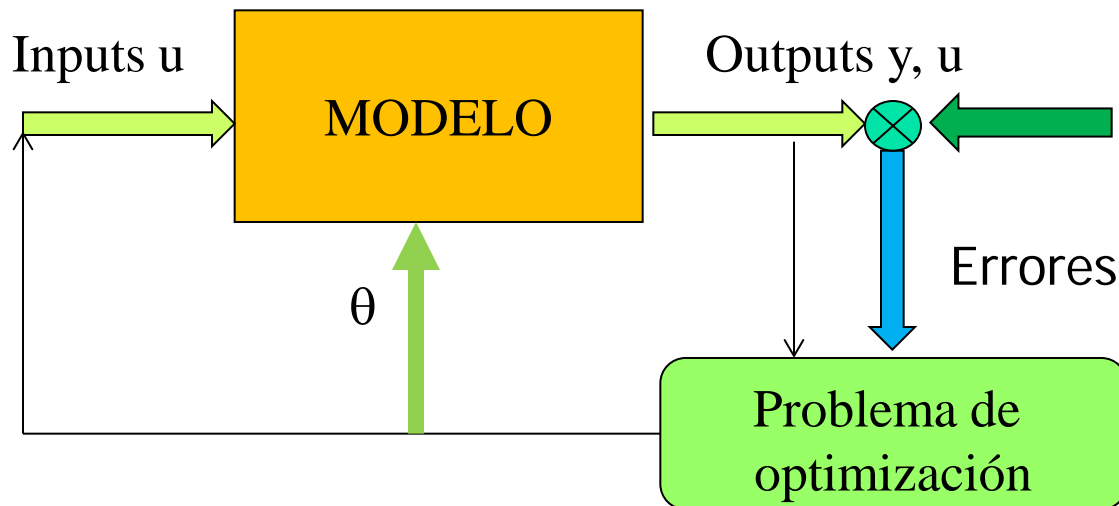
# Reconciliación de datos y parámetros

- ✓ Algunas medidas no son consistentes o fiables
- ✓ Hay variables no medidas y parámetros desconocidos
- ✓ Se necesita un cierto grado de redundancia

$$\min_{u, y, \theta} \sum_{i=1}^{N_{\text{measured}}} \alpha_i (y_i - y_{m,i})^2 + \beta_i (u_i - u_{m,i})^2$$

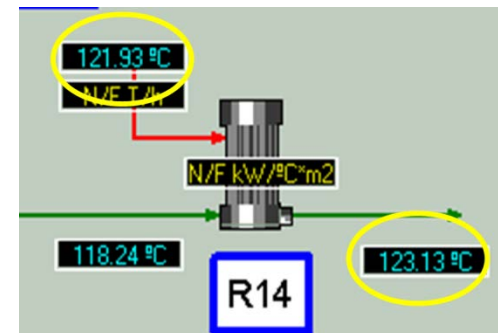
$$\frac{dx}{dt} = f(x, u, \theta) \quad y = h(x, u, \theta)$$

$$g(x, y, u, \theta) \leq 0$$



Valores reconciliados →

Medidas  
 $y_m, u_m$



# Errores gruesos

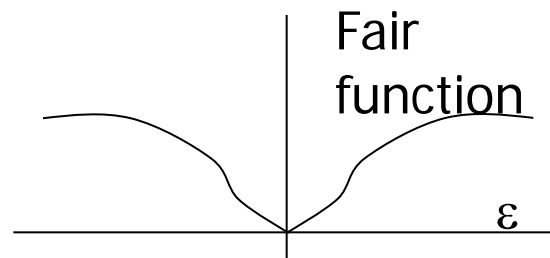
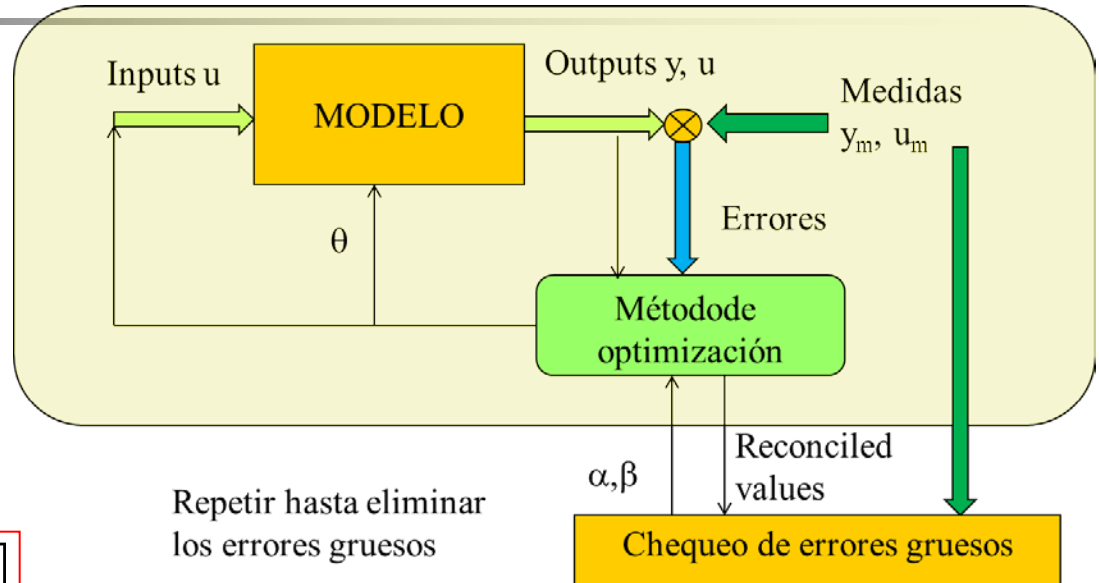
La reconciliación de datos se suele plantear en estado estacionario

## Estimadores robustos

$$\min_x \sum_{j \in M} c^2 \left[ \frac{|\varepsilon_j|}{c} - \log \left( 1 + \frac{|\varepsilon_j|}{c} \right) \right]$$

$$\varepsilon_j = \frac{x_j - x_{mj}}{\sigma}$$

$$f_i(x) = 0$$



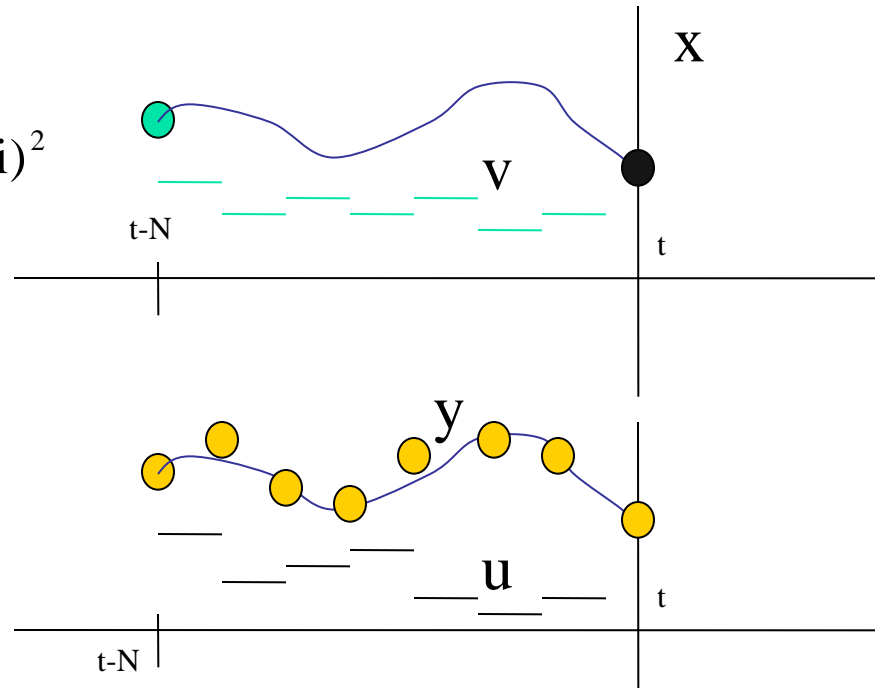
# Estimación de estados con horizonte deslizante

$$\min_{x_{t-N}, v_i} \sum_{i=0}^N [y(t-i) - y_m(t-i)]^2 + \gamma v(t-i)^2$$

$$x(t+1) = f(x(t), u(t), v(t))$$

$$y(t) = g(x(t), u(t))$$

$$L_v \leq v(t-i) \leq L_v$$



Con modelo perfecto

¿Qué estado inicial pasado en  $t-N$  y que mínimas perturbaciones  $v(t-i)$  harían evolucionar el sistema de la forma mas próxima a la que lo ha hecho al aplicarle los controles que le han sido aplicados?



# Varianza de las estimas

---

Una vez que se ha aplicado un procedimiento de optimización y se han obtenido unos parámetros óptimos  $p^*$ , es importante determinar el grado de confianza en los mismos.

- ✓ Matriz de covarianza de las estimas
- ✓ Regiones de confianza de los parámetros
- ✓ Isodistancias

En el caso de estimación no lineal muchas de las medidas de la imprecisión son aproximadas





# Covarianza del error

---

$$F = \left[ \sum_{t=1}^N \left( \frac{\partial y_m(t)}{\partial \mathbf{p}} \right)' \mathbf{C}^{-1} \left( \frac{\partial y_m(t)}{\partial \mathbf{p}} \right) \right]$$

La inversa de la matriz de información de Fisher es una cota inferior de la matriz de covarianza del error de estimación de los parámetros, calculados con el mejor estimador lineal no sesgado (Desigualdad de cramer-Rao).

Esta medida supone que los errores son solo de medida. Una expresión de la matriz de covarianza de los errores de estimación (para el caso de una sola salida en el ajuste) mas ajustada viene dada por:

$$\mathbf{C}_e = \frac{\mathbf{J}(\mathbf{p}^*)}{N - d} \mathbf{F}^{-1}$$

d = número de parámetros

# Intervalos de confianza

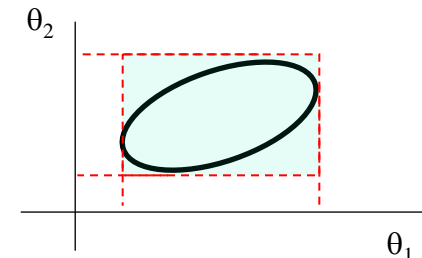
Estimada  $C_e$ , es posible determinar intervalos de confianza para cada parámetro, para un determinado nivel de confianza:

$$p \pm 2.147 \sqrt{c_{ii}} \quad 90\%$$

$$p \pm 3.035 \sqrt{c_{ii}} \quad 99\%$$

$$p \pm t_{N-d}^{\alpha} \sqrt{c_{ii}} \quad \text{Para un grado de confianza } 100-\alpha$$

Aunque en un entorno multiparamétrico deben tomarse con cierta reserva, pues la región de confianza conjunta no equivale a la caja resultante de la individual de cada parámetro





# Regiones de confianza

Conocida la matriz de Fisher, la expresión:

$$J(\mathbf{p}^* + \delta\mathbf{p}) = J(\mathbf{p}^*) + \delta\mathbf{p}' \sum_{t=1}^N \left[ \frac{\partial \mathbf{y}_m}{\partial \mathbf{p}} \Big|_{\mathbf{p}^*}' \mathbf{C}^{-1} \frac{\partial \mathbf{y}_m}{\partial \mathbf{p}} \Big|_{\mathbf{p}^*} \right] \delta\mathbf{p}$$

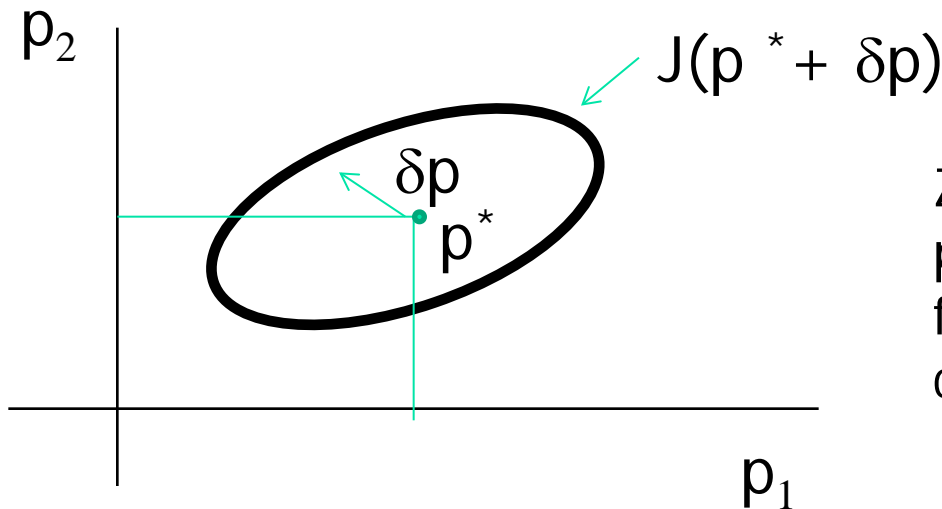
Permite el cálculo aproximado de regiones de confianza de los parámetros estimados que pueden dibujarse como “elipses” cuyos ejes coinciden con los vectores propios de la FIM.

Si no se conoce la FIM, una alternativa es usar la expresión:

$$J(\mathbf{p}^* + \delta\mathbf{p}) = J(\mathbf{p}^*) + \frac{1}{2} \delta\mathbf{p}' \frac{\partial^2 J}{\partial \mathbf{p} \partial \mathbf{p}'} \Big|_{\mathbf{p}^*} \delta\mathbf{p} \quad \mathbf{C}_e \approx \frac{2J(\mathbf{p}^*)}{N - d} \left[ \frac{\partial^2 J}{\partial \mathbf{p} \partial \mathbf{p}'} \Big|_{\mathbf{p}^*} \right]^{-1}$$

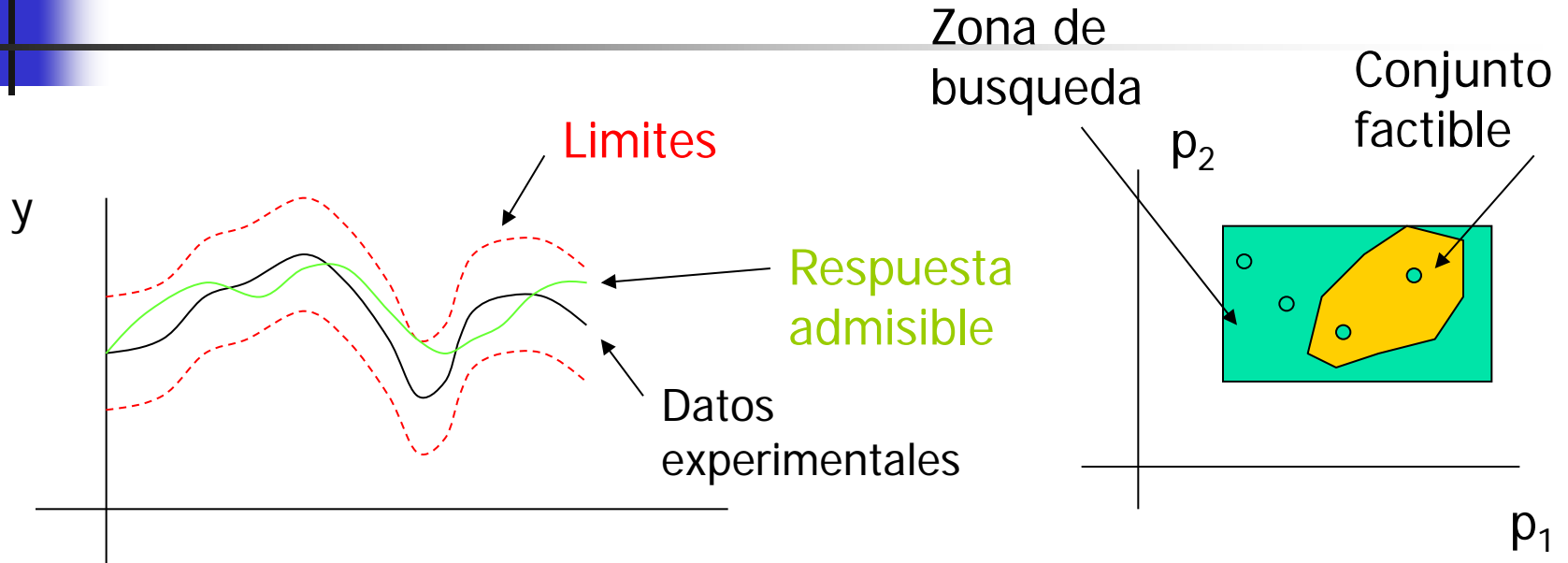
# Regiones de confianza

$$J(p^* + \delta p) = J(p^*) + \delta p' \sum_{t=1}^N \left[ \frac{\partial y_m}{\partial p} \Big|_{p^*} C^{-1} \frac{\partial y_m}{\partial p} \Big|_{p^*} \right] \delta p$$



Zona de cambio de los parámetros para los que la función de coste no cambia mas que  $\Delta J$

# Conjuntos de parámetros factibles



Se define una respuesta aceptable del modelo (por ejemplo con bandas de error, o un valor máximo aceptable de  $J$ ).

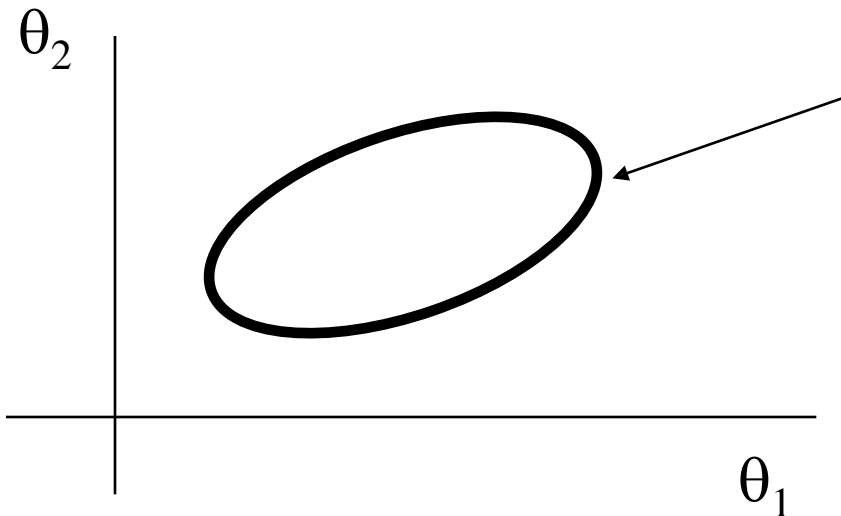
Se generan aleatoriamente valores de  $p$  dentro de un rango y se hace la simulación del modelo con los mismos.

Se incorporan al conjunto de parámetros aceptables los que cumplen el criterio, definiendo la región buscada

# Identificación global

$$J = \frac{1}{N} \sum_{t=1}^N [(y(t) - M(\theta, u(t)))^2]$$

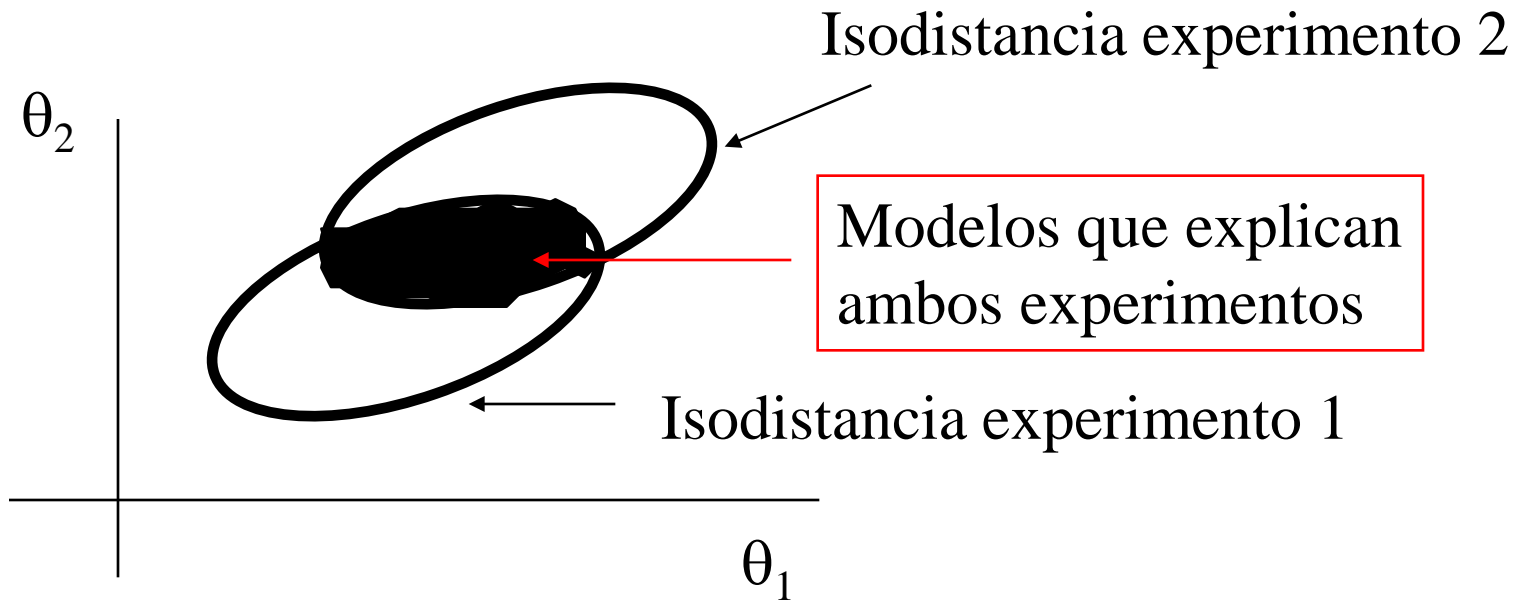
$$f(\theta) = J = \text{cte.}$$



Isodistancia:  
Lugar geométrico de los puntos del espacio paramétrico con igual valor de J

Todos los modelos sobre una isodistancia tienen la misma validez respecto al criterio de suma de cuadrados de errores

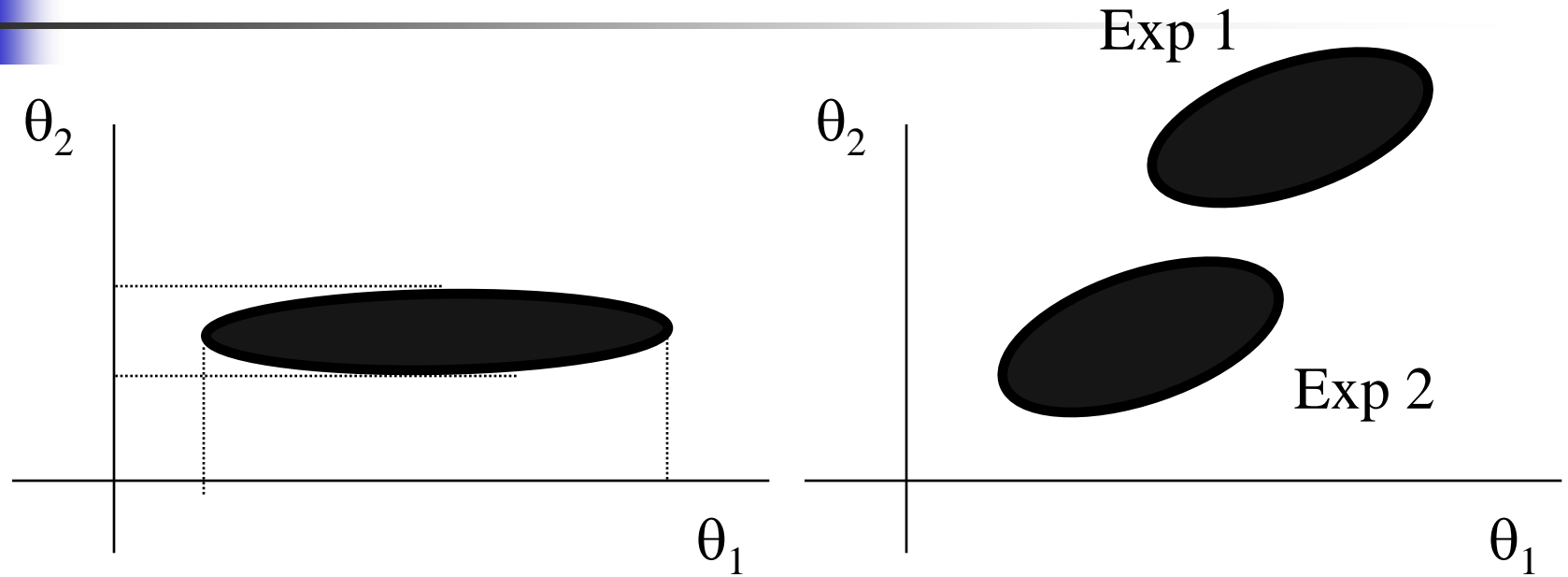
# Identificación global



Las isodistancias dependen de los datos experimentales y del nivel de exactitud  $J$  elegido

Con modelos no-lineales no tienen por que ser formas cuadráticas

# Isodistancias

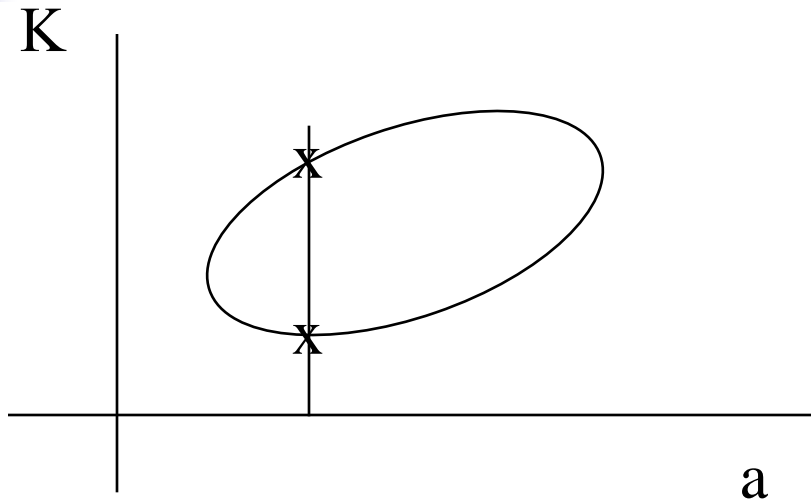


Informan sobre la calidad de los parámetros, y el rediseño de experimentos

Nivel de calidad no compatible con los experimentos



# Cálculo de Isodistancias



Ejemplo con un modelo sencillo de primer orden

Para cada valor de  $a$  se resuelve una ecuación de segundo grado para obtener el correspondiente valor de  $K$

$$J = \frac{1}{N} \sum_{t=1}^N \left[ (y(t) - \frac{Kq^{-1}}{1 + aq^{-1}} u(t)) \right]^2$$

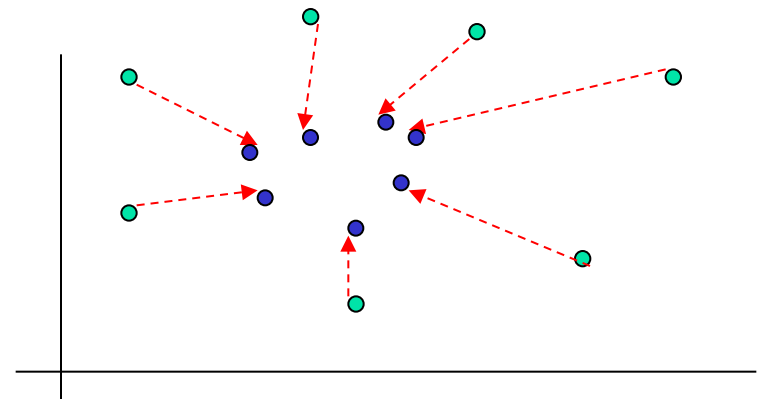
# Regiones de confianza

En un caso mas general, la construcción de isodistancias es compleja. Una forma de construir las curvas de un criterio dado, p.e. con un nivel de confianza del  $100(1-\alpha)$  %, o sea, los puntos para los que

$$J = J(\mathbf{p}^*) \left( 1 + \frac{d}{N-d} F_{\alpha, d, N-d} \right)$$

es establecer una región amplia y partiendo de puntos del contorno, optimizar hasta que  $J$  alcanza ese valor

F: distribución F con  $d$  y  $N-d$  grados de libertad y nivel  $\alpha$





# Regiones de confianza

---

Una aproximación puede calcularse utilizando la aproximación de la matriz de covarianza de los errores de estimación:

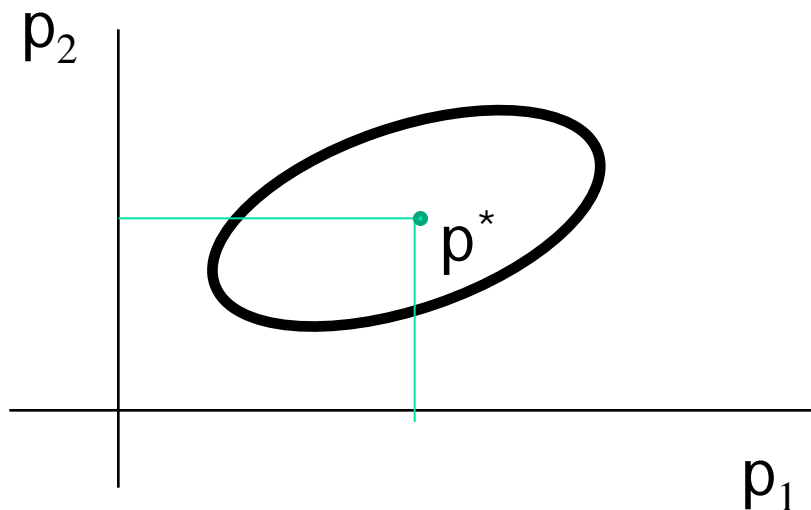
$$J(p^* + \delta p) = J(p^*) + \frac{1}{2} \delta p' \frac{\partial^2 J}{\partial p \partial p'} \Big|_{p^*} \delta p \quad C_e \approx \frac{2J(p^*)}{N-d} \left[ \frac{\partial^2 J}{\partial p \partial p'} \Big|_{p^*} \right]^{-1}$$

La región de confianza del  $100(1-\alpha)\%$  vendrá dada por:

$$\delta p' \frac{2J(p^*)}{N-d} \left[ \frac{\partial^2 J}{\partial p \partial p'} \Big|_{p^*} \right]^{-1} \delta p \leq d F_{\alpha, d, N-d}$$

# Linealización

Otra forma de obtener regiones de confianza aproximadas es linealizar el modelo obtenido con parámetros  $p^*$  en ese punto y tener en cuenta que si los errores de medida tienen una varianza  $V$  y están normalmente distribuidos, entonces los parámetros del modelo cumplen:



$$(p - p^*)' V^{-1} (p - p^*) \leq \chi_{\alpha, d}^2$$

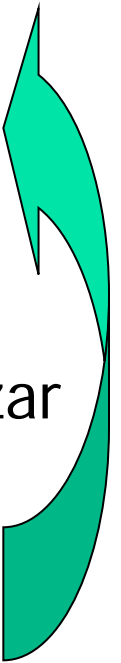
Donde  $\chi^2$  es la distribución chi-cuadrado con nivel de confianza  $\alpha$  y  $d$  (número de parámetros) grados de libertad



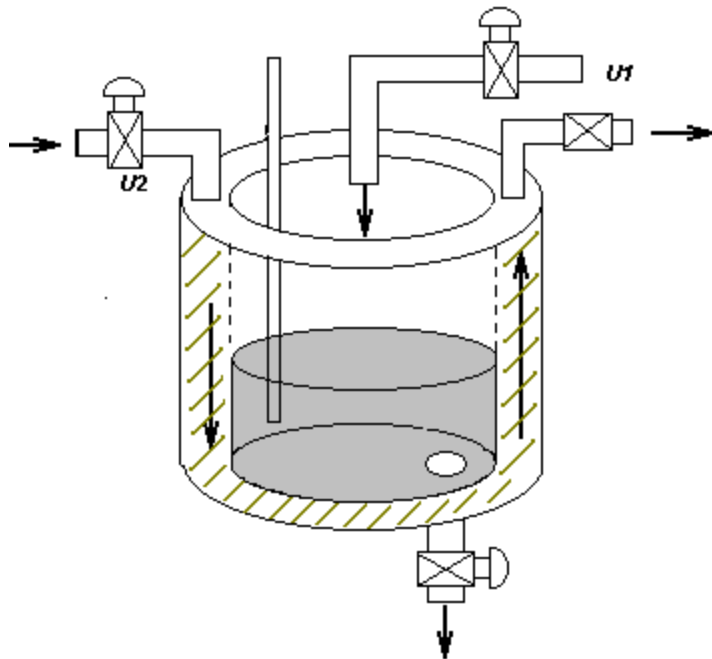
# Parametrización / Metodología

---

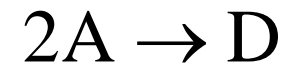
- Toma de datos experimentales (separar los de calibración y validación)
- Análisis y depuración de los datos
- Selección de los parámetros a estimar
  - Identificabilidad estructural
  - Cálculo de Sensibilidades
- Posible reparametrización del modelo para linealizar la estimación
- Estimaciones iniciales y rangos
- Elegir la función de costo
- Estimar los parámetros por optimización
- Estimar los residuos y la región de confianza de los parámetros



# Ejemplo en EcosimPro: Reactor de Van der Vusse



Reactor altamente no lineal y difícil de controlar



Parámetros a estimar:

Volumen 10 l

Masa de refrigerante 5 Kg

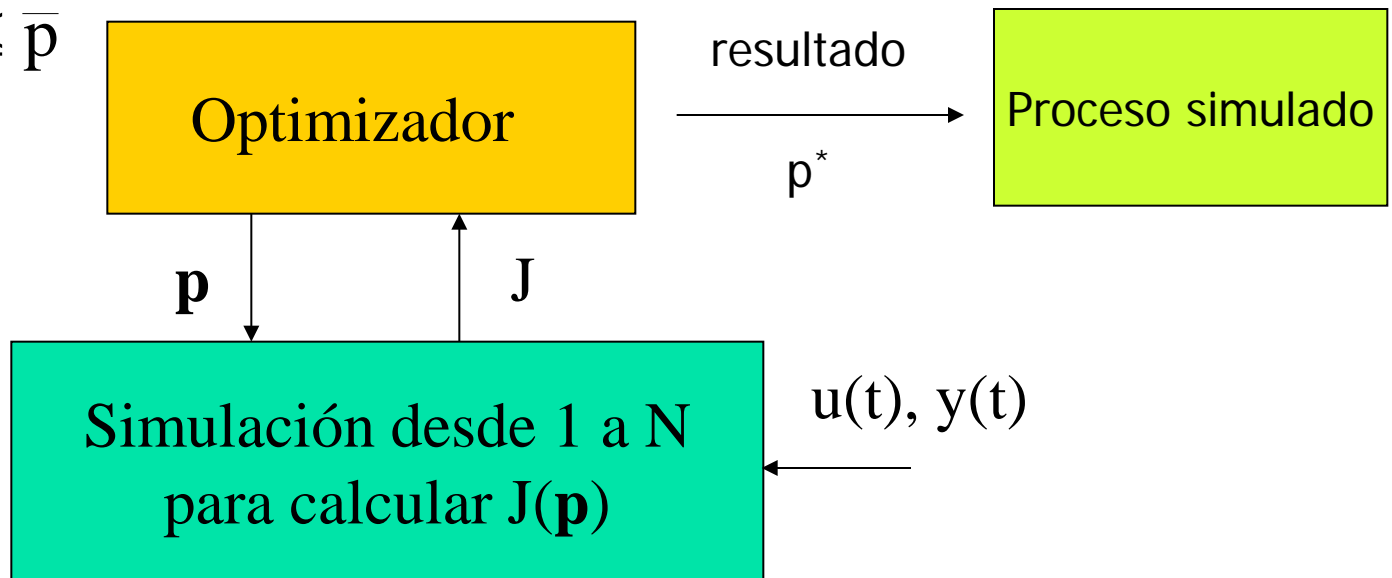
# Resolución con simulación

$$\min_{\mathbf{p}} J = \sum_{j=1}^N [y(t) - y_m(\mathbf{p}, u, t)]^2$$

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), u(t))$$

$$y_m(t) = g(\mathbf{x}(t), u(t))$$

$$\underline{\mathbf{p}} \leq \mathbf{p} \leq \bar{\mathbf{p}}$$



# Reactor Van der Vusse

