



# Introducción a la Identificación de Sistemas

Tema 05



# CA3: Introducción a la Identificación de Sistemas

## Nota

El material de estas diapositivas está basado en los libros:

- Arun K. Tangirala. “Principles Of System Identification. Theory and Practice”. CRC press, 2015.
- Ioan D. Landau, Gianluca Zito. “Digital Control Systems Design, Identification and Implementation”. Springer 2006.
- J.E. Normey-Rico and E.F. Camacho. “Control of Dead-time Processes”. Springer, 2007.



# Introducción a la Identificación de Sistemas

Índice

- **Introducción**
- Etapas de la identificación de sistemas
- Señales de entrada

Tema 05

V1 - 2024



# CA3: Introducción a la Identificación de Sistemas

## Introducción

- El análisis de las características de un sistema y las relaciones entre sus variables es de suma importancia en su predicción, control, seguimiento, diseño e innovación. Un paso clave en estos análisis es el **desarrollo de una descripción (matemática)** del proceso/sistema bajo estudio conocida como modelo.
- Generalmente se siguen dos enfoques para el desarrollo del modelo:
  - **enfoque teórico** (primeros principios): se basa en las leyes fundamentales de la materia y la energía. Es necesaria una buena comprensión de la física del proceso.
  - **enfoque empírico**: se basa en el análisis de las observaciones (datos experimentales u operativos). Enfoque altamente práctico como alternativa o complemento al anterior y ampliamente seguido (la mayoría de los procesos son demasiado complejos para ser entendidos en un nivel fundamental).
    - Las observaciones contienen potencialmente una gran cantidad de información que, de otro modo, permanecería oculta en un enfoque basado en primeros principios.

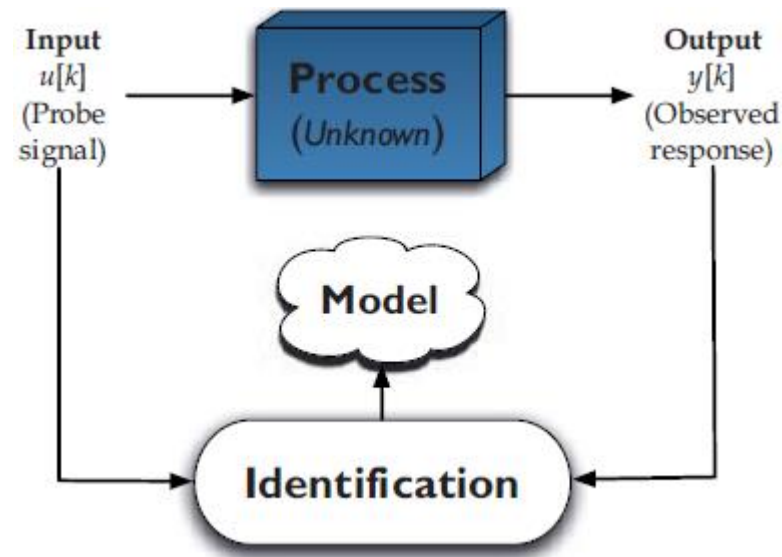
La Identificación de Sistemas se ocupa de los medios y técnicas para estudiar un sistema a través de datos observados/experimentales, principalmente para desarrollar una descripción adecuada (matemática) de ese sistema.



# CA3: Introducción a la Identificación de Sistemas

## Introducción – Objetivo

- El objetivo principal es desarrollar un modelo a partir de datos de entrada y salida. Se dice que el modelo resultante es empírico en contraste con un modelo de primeros principios.



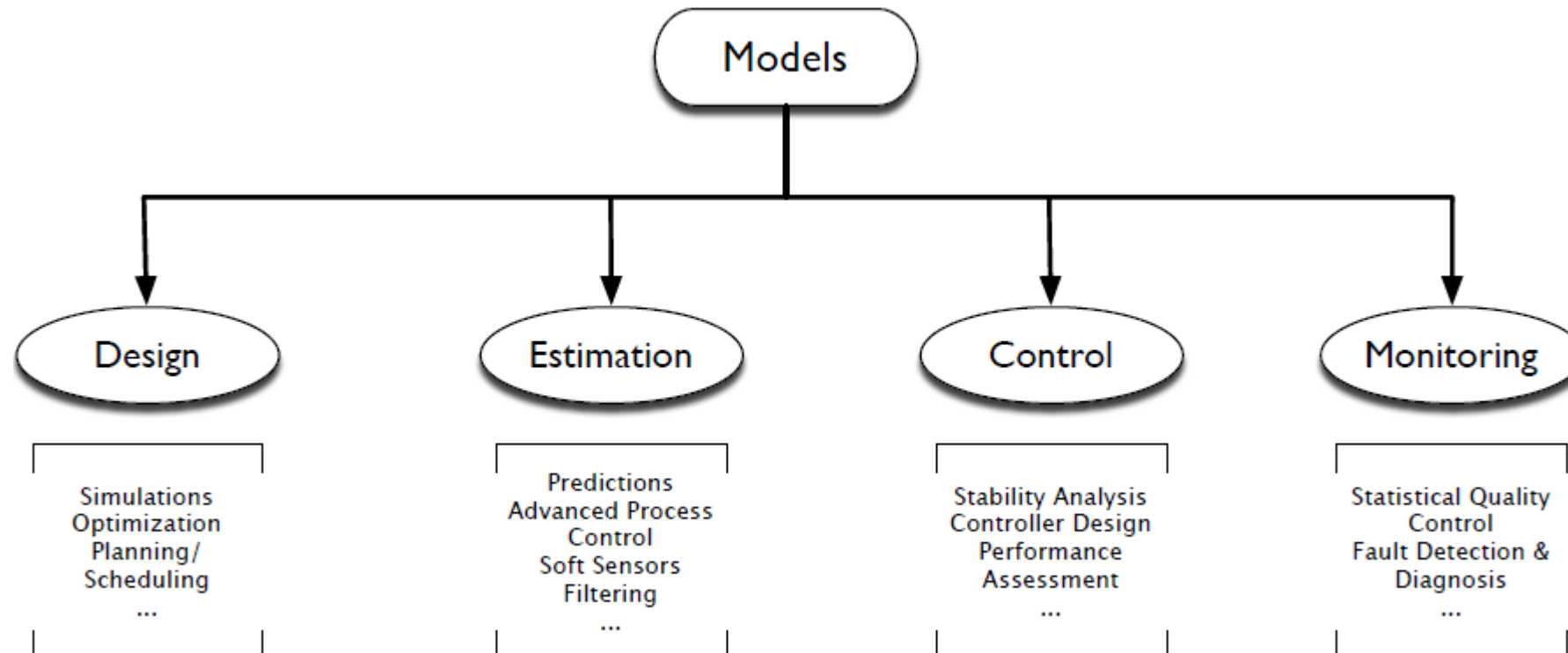
- Analogías simples con la identificación:
  - tomar un vehículo (sistema) en una prueba de manejo
  - entrevista de trabajo
  - etc...



# CA3: Introducción a la Identificación de Sistemas

## Introducción – Necesidad de la identificación

- ¿Qué factores motivan la necesidad de identificación?
  - la necesidad de modelos para el análisis y la automatización de procesos
  - las limitaciones prácticas del enfoque de los primeros principios en el desarrollo de modelos
- Los beneficios de desarrollar modelos son muchos:





# CA3: Introducción a la Identificación de Sistemas

## Introducción – Primero principios vs. Identificación

- La identificación ofrece una alternativa poderosa y pragmática al modelado de primeros principios, que utiliza leyes básicas de balances de masa, cantidad de movimiento y energía combinados con algunas relaciones constitutivas (empíricas).
- En los inicios del control y la automatización, los modelos se desarrollaron en gran medida utilizando el enfoque de los primeros principios. La mayoría de los procesos modernos de interés son complejos y dificultan/impiden un enfoque fundamental.
- El recurso natural ha sido hacia los enfoques basados en datos, ya que asumen un conocimiento previo mínimo y dependen en gran medida de las observaciones de entrada y salida para desarrollar modelos.
- Con los avances en la tecnología de medición y cómputo, los enfoques basados en datos adquirieron un gran impulso.
- Además de ser alternativas naturales, los enfoques empíricos ofrecen varios beneficios prácticos:
  - capacidad de construir modelos con un conocimiento mínimo del proceso
  - flexibilidad para seleccionar la estructura del modelo
  - la conveniencia de implementarlos en forma de un código interpretable por computadora



# CA3: Introducción a la Identificación de Sistemas

## Introducción - Modelos

- Un modelo de un proceso se define de forma general como aquella entidad que puede emular las características de un proceso para un conjunto dado de condiciones y parámetros operativos.
- Los **modelos cualitativos** simplemente describen la respuesta de un sistema sobre una base categórica con poca o ninguna participación de valores numéricos.
- Los **modelos cuantitativos** describen la relación entre los cambios cuantificados en la entrada y la salida en términos de expresiones matemáticas:
  - son claramente más útiles en control, optimización y seguimiento de procesos que los cualitativos, cuyo uso se limita principalmente a realizar análisis cualitativos y mejoras en las operaciones de procesos.
  - permiten sacar inferencias y tomar decisiones basadas en criterios cuantificados. Además, pueden programarse fácilmente en una computadora, lo que reduce en gran medida la intervención humana.
  - se pueden clasificar en diferentes pares de clases según la naturaleza de los procesos que describen, las suposiciones hechas sobre los fenómenos subyacentes y el enfoque adoptado para desarrollarlos (ejemplo primeros principios vs empíricos).

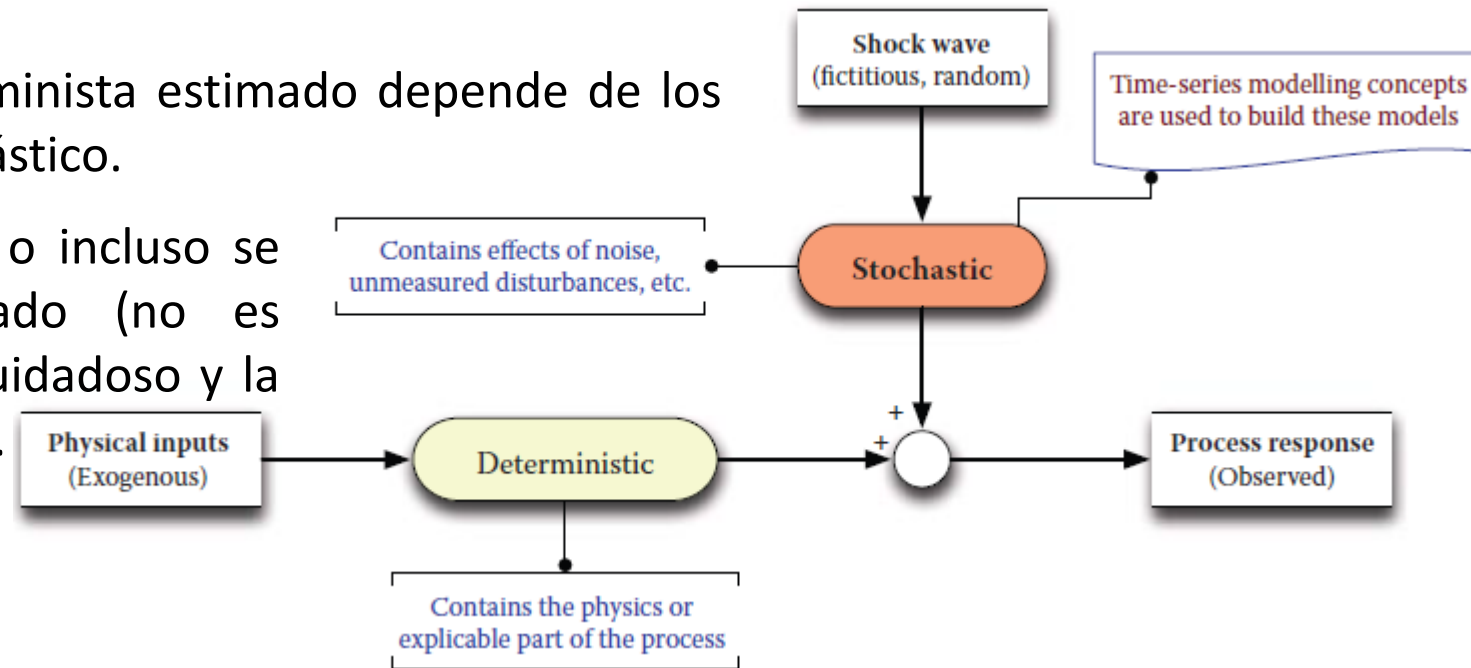




# CA3: Introducción a la Identificación de Sistemas

## Introducción – Componentes de un modelo identificado

- El modelo identificado consta de dos componentes:
  - una descripción matemática de las relaciones causa-efecto (**determinista**)
  - una descripción estadística más matemática de las incertidumbres (**estocástica**): este componente entra en juego debido a la presencia de errores de observación, incertidumbres de proceso y errores de modelado.
- El principal objeto de interés es el componente determinista (entrada-salida) del modelo porque captura la dinámica del proceso físico.
- La exactitud y precisión del modelo determinista estimado depende de los supuestos que constituyen el modelo estocástico.
- La parte estocástica a menudo se socava o incluso se ignora en varios procesos de modelado (no es inmediatamente obvio). Sólo un estudio cuidadoso y la práctica pueden esclarecer sobre este tema.





# CA3: Introducción a la Identificación de Sistemas

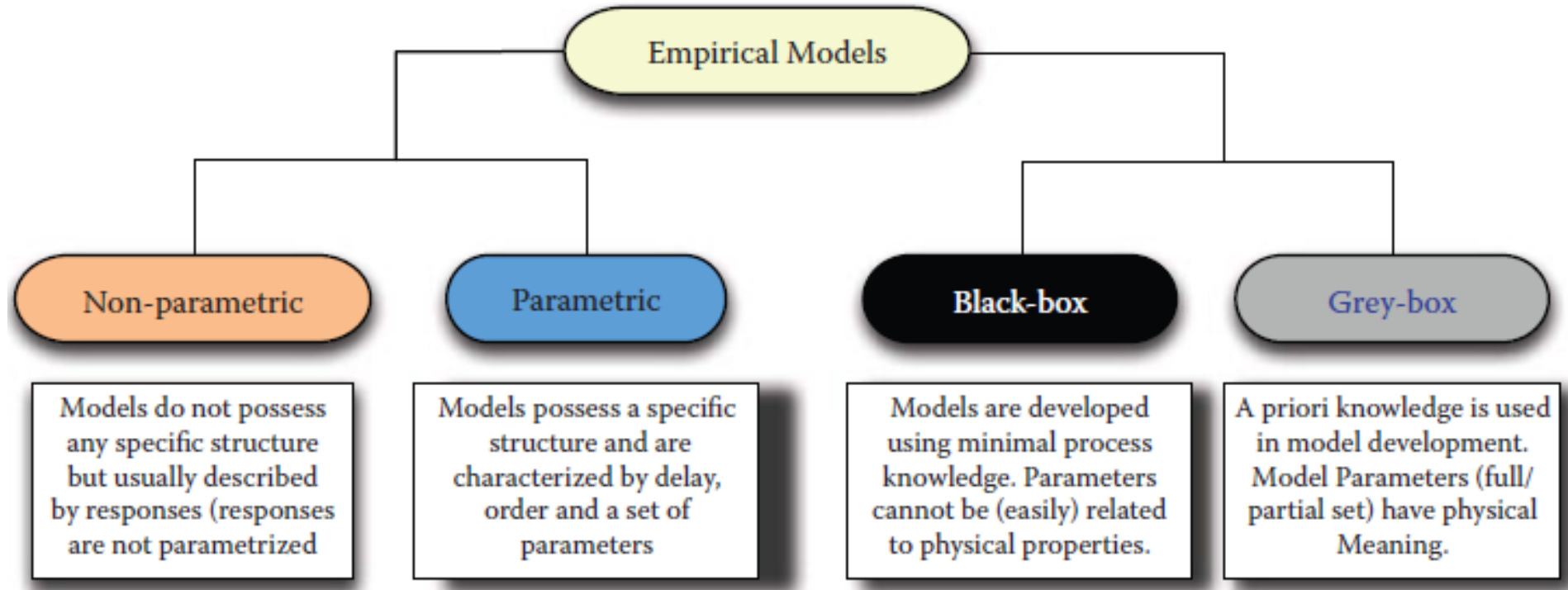
## Introducción – Transparencia de los modelos

- Transparencia (significado físico) de los modelos:
  - Los modelos de primeros principios por su forma de construcción son transparentes (**caja blanca**).
  - Los modelos empíricos son generalmente opacos con respecto a la física del proceso porque se desarrollan en gran medida utilizando métodos matemáticos en lugar de leyes físicas (**caja negra**).
  - La opacidad de los modelos empíricos se puede reducir mediante la incorporación de un conocimiento del proceso a priori, por ejemplo, mediante la imposición de restricciones estructurales conocidas en el modelo, lo que da como resultado lo que se conoce como modelos de **caja gris**.
- Capacidades de extrapolación:
  - Los modelos basados en datos tienen, en general, buenas capacidades predictivas solo sobre el régimen operativo abarcado por los datos.
  - Los modelos de primeros principios son superiores en este respecto.



# CA3: Introducción a la Identificación de Sistemas

## Introducción – Modelo empíricos



- **Modelos no paramétricos vs paramétricos:** Los modelos paramétricos (ej. ecuaciones en diferencias) poseen una estructura y un orden específicos y se caracterizan por tener menos parámetros, mientras que los modelos no paramétricos (ej. modelos de respuesta de impulso) no poseen ninguna estructura u orden específico, pero se caracterizan por un gran número de incógnitas. Desde el punto de vista de la identificación, los modelos no paramétricos se pueden estimar con un conocimiento a priori mínimo, mientras que la estimación de modelos paramétricos exige un conocimiento a priori. Este conocimiento previo se puede adquirir estimando primero un modelo no paramétrico.



# CA3: Introducción a la Identificación de Sistemas

## Introducción - Métodos de identificación

- Dependiendo del tipo de modelo obtenido:
  - **Métodos no paramétricos:** análisis de la respuesta transitoria y en frecuencia, análisis de la correlación, espectral, Fourier, etc.
  - **Métodos paramétricos:** requieren la elección de una posible estructura del modelo, de un criterio de ajuste de parámetros, y por último de la estimación de los parámetros que mejor ajustan el modelo a los datos experimentales.
- Dependiendo de la aplicación:
  - **Métodos de identificación off-line** (a posteriori): utilizados en aquellas aplicaciones en que no se requiera un ajuste continuado del modelo. Se realiza la identificación previa de la planta, considerándose que la validez de los parámetros obtenidos no se verá alterada con el paso del tiempo.
  - **Métodos de identificación on-line** (identificación recursiva): los parámetros se van actualizando continuamente a partir de los nuevos datos de entrada-salida obtenidos durante la evolución del proceso. Estos métodos son muy utilizados en sistemas de control adaptativo.
- Dependiendo del criterio de ajuste de los parámetros:
  - Existen diversos métodos matemáticos para ajustar los parámetros de una estructura a un conjunto de datos de entrada-salida. Algunos de los más utilizados en el campo de la identificación son el método de mínimos cuadrados y el método de las variables instrumentales.



# CA3: Introducción a la Identificación de Sistemas

## Introducción – “Definición”

La identificación es el ejercicio de desarrollar una relación matemática (modelo) entre las causas (entradas) y los efectos (salidas) de un sistema (proceso) basado en datos observados o medidos. Dicho de otra manera, la identificación establece un mapa matemático entre los espacios de entrada y salida determinados por los datos.

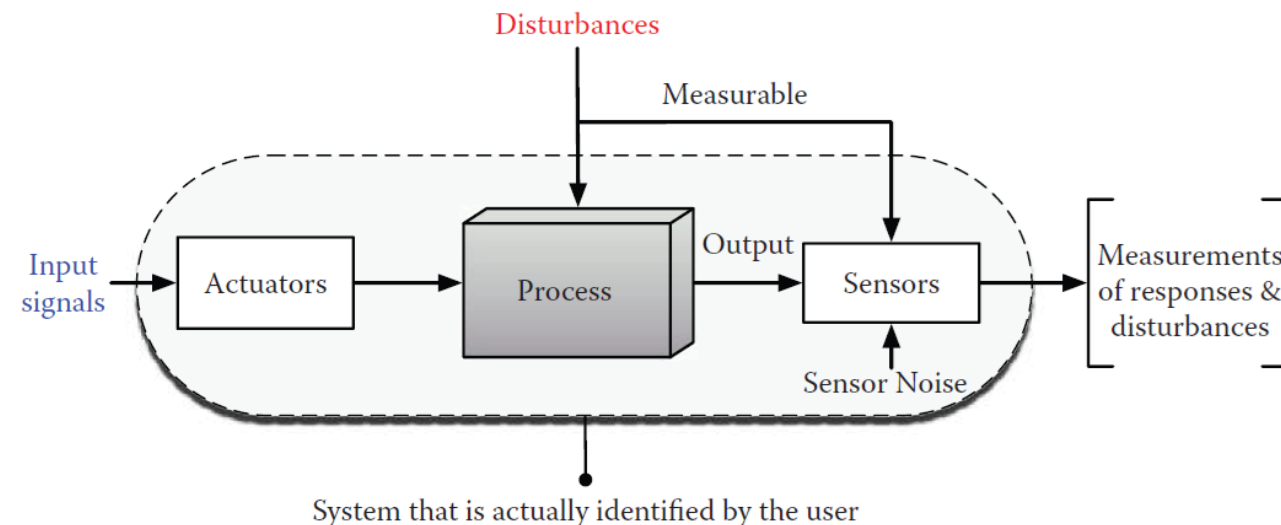


# CA3: Introducción a la Identificación de Sistemas

## Introducción – Algunas definiciones

- **Salidas:** constituyen todas aquellas señales que se miden y que se desea predecir. También se conocen como respuestas o variables predichas.
- **Entradas:** se refieren colectivamente a todas las variables que se considera que influyen en las salidas. El conjunto de entrada consta tanto de lo que el usuario puede manipular (señales de prueba o entrada) como de lo que no se puede ajustar pero se puede medir (perturbaciones).
- **Perturbaciones:** las que se pueden medir (ej. las fluctuaciones de temperatura en un proceso de mezcla de fluidos fríos y calientes) y las que no se miden (ej. las perturbaciones del viento experimentadas por un avión).

El modelo identificado no solo explica la dinámica del proceso, sino también los actuadores, sensores y los efectos de las perturbaciones.





# CA3: Introducción a la Identificación de Sistemas

## Introducción – Hechos relacionados con la exactitud y precisión de los modelos

1. En general, no es posible construir un modelo preciso a partir de una cantidad finita de muestras. Un modelo estimado a partir de datos contaminados con errores difícilmente pueda ser exacto. La especificación incorrecta de la estructura del modelo suele conducir a errores sistemáticos (estimaciones y predicciones sesgadas). El objetivo es producir estimaciones no sesgadas, necesitando la especificación "correcta" de la estructura del modelo, la elección adecuada del método de estimación y, posiblemente, observaciones muy grandes (modelo asintóticamente insesgado).
  2. Un solo registro de datos es sólo uno de los posibles para el mismo experimento. La causa es la aleatoriedad en las perturbaciones y el ruido de medición. En la práctica, se suele trabajar con un solo registro de longitud finita, así el modelo estimado es solo uno entre muchos que podrían haberse estimado a partir de otras realizaciones posibles. La variabilidad en los datos se manifiesta como imprecisión en las estimaciones (de modelos, parámetros o señales). Cada método de estimación se esfuerza por llevar la varianza de las estimaciones a cero.
  3. La exactitud y precisión del modelo identificado de manera óptima, entre otros factores, depende de manera crítica del:
    - **tipo de entrada** (excitación y forma, siendo esta última válida para sistemas no lineales)
    - **relación señal-ruido** logrado en el experimento.
- Un término generalizado que captura estos dos aspectos es **información**. La calidad del modelo final depende de cuán informativos sean los datos. Existen métricas para evaluar la información.

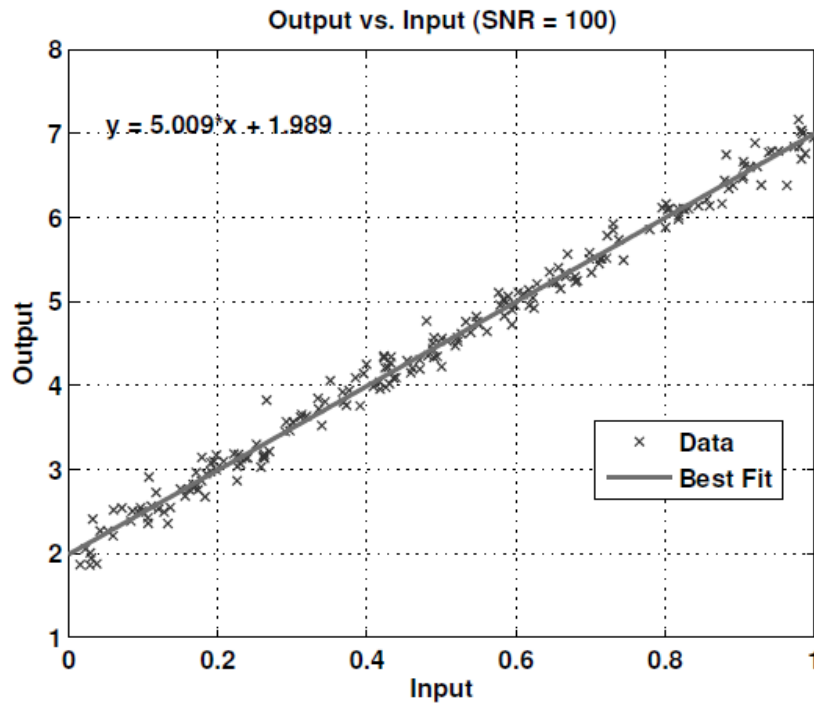




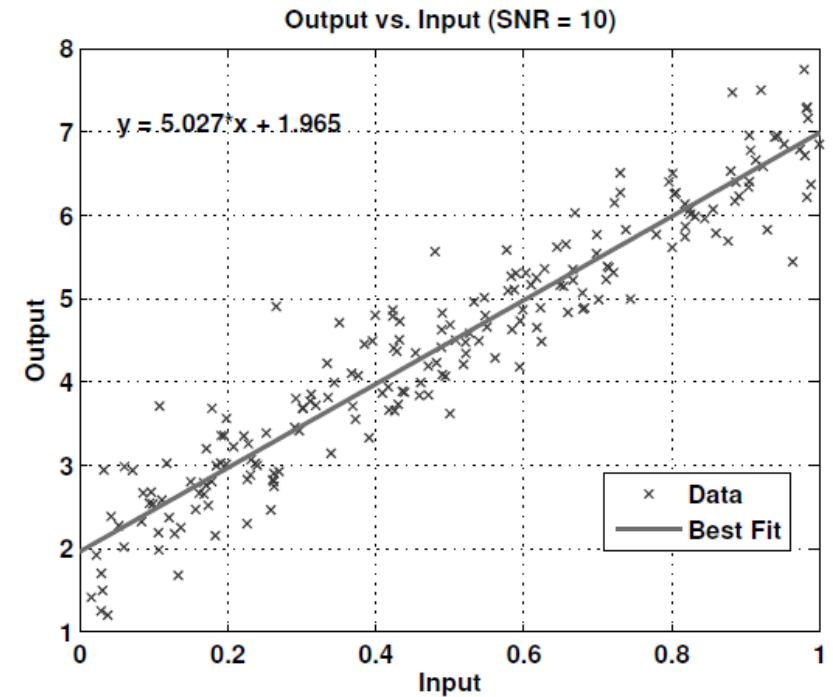
# CA3: Introducción a la Identificación de Sistemas

## Introducción – Hechos relacionados con la exactitud y precisión de los modelos - SNR

- Incluso con una buena excitación, los efectos estocásticos en las mediciones pueden ser lo suficientemente altos como para ser perjudiciales para la calidad del modelo.
- Si una parte significativa de las variaciones en la medición se deben al ruido, la contribución de la entrada se debilita como así también la capacidad de estimar con precisión el modelo. La SNR representa la relación de los efectos debido a la variable conocida frente a las incertidumbres. Cuanto menor sea la SNR, más ambigua es la estimación del modelo de entrada-salida.



(a)  $\sigma_{\hat{b}_1} = 0.036, \sigma_{\hat{b}_0} = 0.02$



(b)  $\sigma_{\hat{b}_1} = 0.114, \sigma_{\hat{b}_0} = 0.064$





# CA3: Introducción a la Identificación de Sistemas

## Introducción – Hechos relacionados con la exactitud y precisión de los modelos

- Siempre que se habla de exactitud y precisión se requiere tener un punto de referencia “la verdad”. En las definiciones anteriores se supone tácitamente que la descripción de un sistema "verdadero" está disponible. En la práctica, la complejidad del sistema está más allá del alcance de una descripción matemática. Entonces, el punto de referencia es la "mejor" aproximación que se puede realizar para la estructura del modelo elegido y los datos experimentales dados, especialmente las entradas.
- Para el análisis teórico, a menudo se asume un sistema “verdadero” para estudiar las propiedades de exactitud y precisión (de la estructura de un modelo) resultantes de un ejercicio de identificación con un algoritmo de estimación elegido y el diseño de entrada para un determinado sistema.
- Dado un conjunto de datos de entrada y salida, existen muchos modelos aproximados que pueden explicar numéricamente los datos de entrada y salida con una exactitud y precisión razonables. Las principales herramientas para obtener el "mejor" modelo se basan en conceptos de **álgebra lineal, estimación (optimización), teoría de la probabilidad y procesamiento de señales estocásticas**. El algoritmo de estimación óptimo entrega el mejor modelo condicionado a un conjunto de parámetros especificado por el usuario (clase de modelo, orden, retardo de tiempo, etc.). **Para optimizar aún más estos parámetros libres, es necesario un estudio formal de identificación.**

Una tarea importante en la identificación es el diseño de la entrada. El desafío es que un diseño de entrada adecuado requiere un conocimiento suficiente del sistema y el nivel de incertidumbres, **¡que es el propósito del diseño de entrada en sí mismo!** Por lo tanto, la identificación es inevitablemente un ejercicio iterativo.



# Introducción a la Identificación de Sistemas

Índice

- Introducción
- **Etapas de la identificación de sistemas**
- Señales de entrada

Tema 05

V1 - 2024

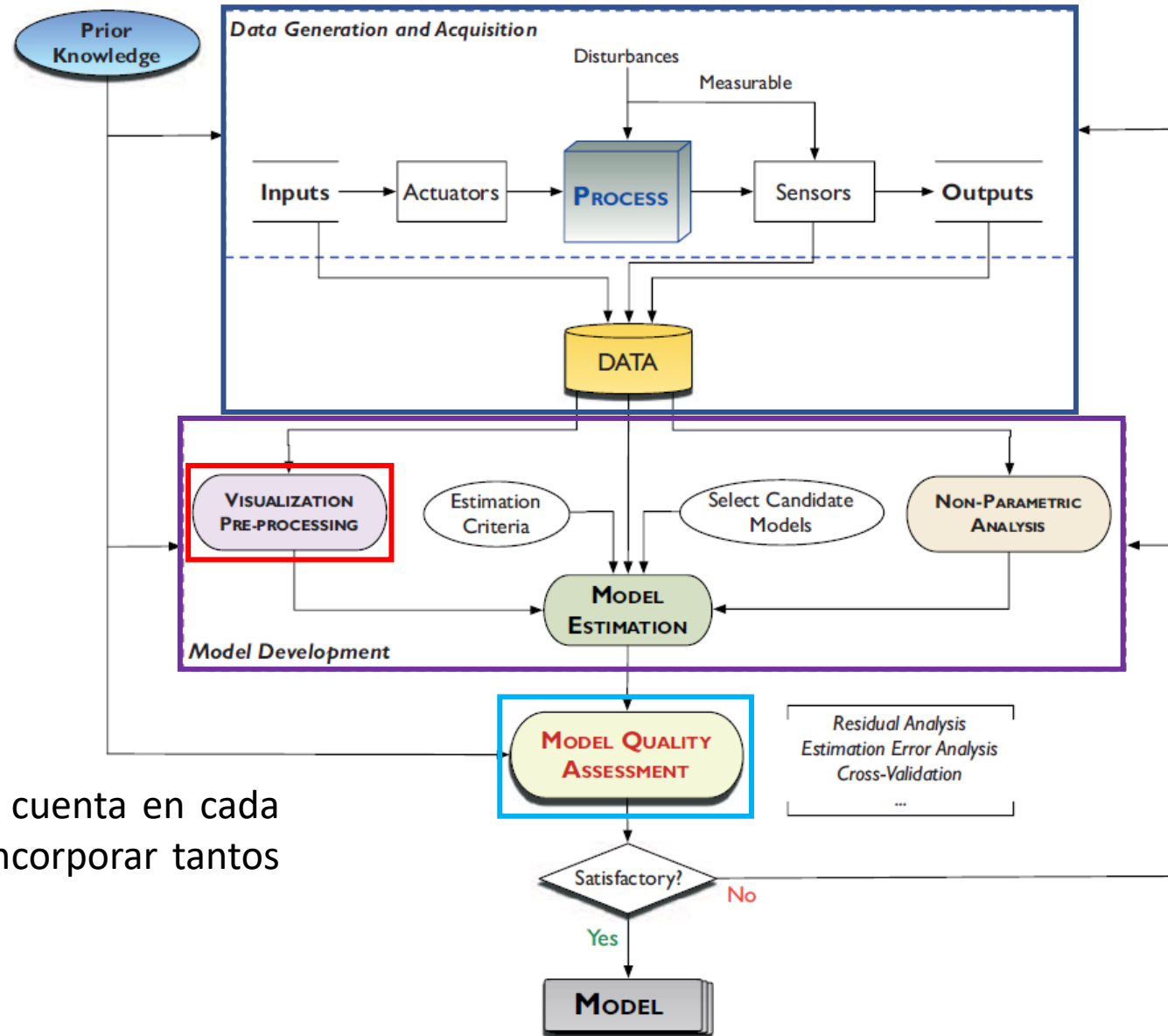


# CA3: Introducción a la Identificación de Sistemas

## Etapas en la identificación de sistemas - Pasos

1. Generación y Adquisición de Datos
2. Preprocesamiento de datos
3. Visualización de datos
4. Desarrollo del modelo
5. Evaluación y validación del modelo

- El conocimiento previo del proceso puede tenerse en cuenta en cada etapa de la identificación. De hecho, se recomienda incorporar tantos conocimientos previos como sea posible.
- La identificación es un ejercicio iterativo.

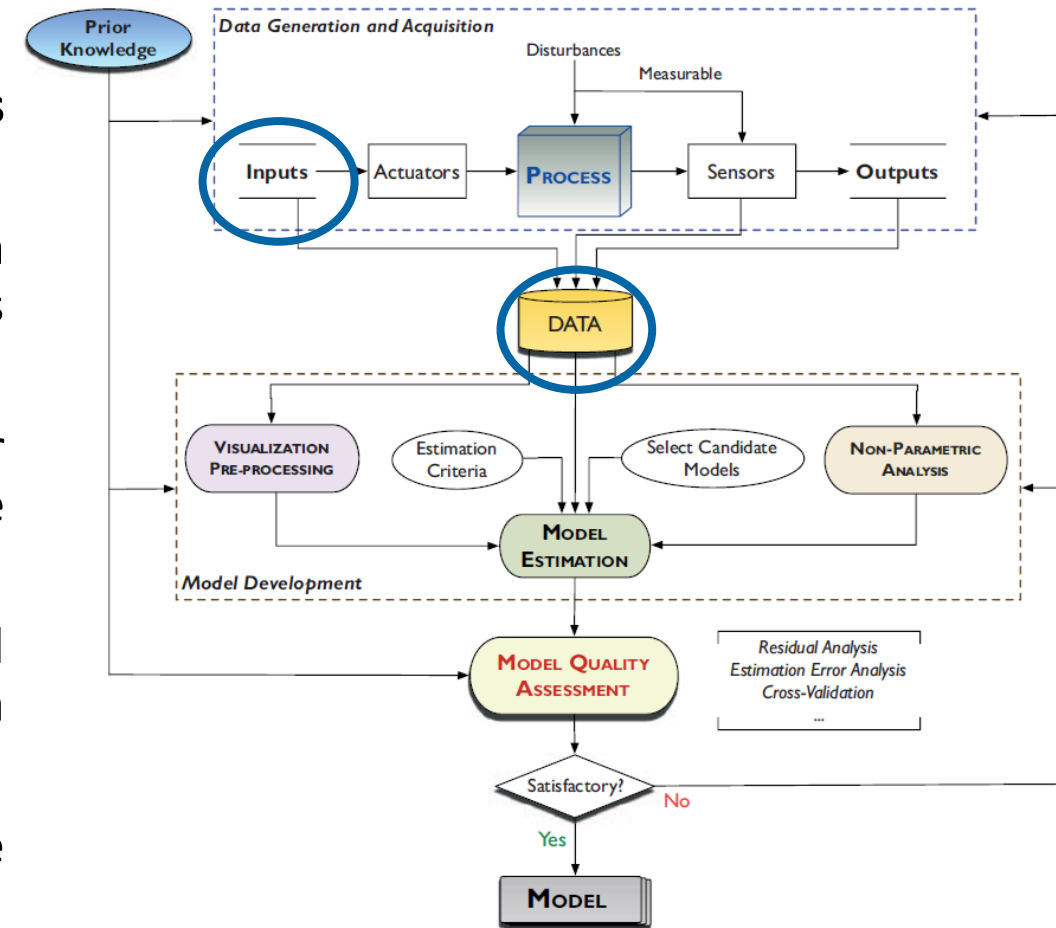




# CA3: Introducción a la Identificación de Sistemas

## Etapas en la identificación de sistemas - Generación y Adquisición de datos

- Es el paso que más influye. El éxito de cada etapa y la confianza en el modelo final depende de la calidad (y cantidad) de información de los datos. **Es el alimento para la identificación.**
- La pregunta básica de interés es: **¿qué tipo de excitación es mejor para un problema de identificación dado?**
- La entrada debe ser tal que su efecto en la salida medida sea mayor que los causados por el ruido del sensor/perturbaciones no medidas.
- A lazo abierto o cerrado (condición en el sistema), el controlador decide los movimientos de la entrada sin forma directa de ajustar la excitación.
- Indirectamente es posible a través de la referencia o una señal externa (dither que se introduce en la entrada o en la referencia).
- La obtención de datos informativos depende en gran medida de la frecuencia de muestreo.

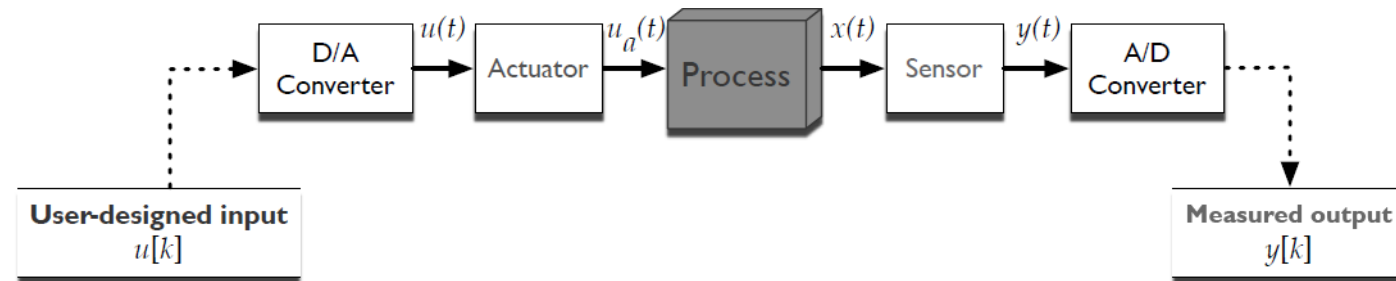




# CA3: Introducción a la Identificación de Sistemas

## Etapas en la identificación de sistemas - Generación y Adquisición de datos (2)

- La respuesta medida por lo general consta tanto de los efectos de la entrada como de las perturbaciones y/o errores de medición no medidos, conocidos como efectos deterministas y estocásticos, respectivamente. La capacidad de identificar un modelo (de una estructura específica) depende de manera única y precisa de dos factores:
  - las características de la entrada
  - las proporciones relativas de los efectos deterministas y estocásticos en la medida, cuantificada por la relación señal-ruido (SNR)
- En un sentido amplio, la singularidad de los modelos identificados no solo está relacionada con el diseño de la entrada, sino también con el algoritmo de estimación y la estructura del modelo en sí. En conjunto, estos factores se analizan formalmente utilizando el concepto de **identificabilidad**.
- **Tipos de entrada:**
  - Escalón
  - Suma de sinusoides
  - Chirp
  - Señal binaria pseudoaleatoria (PRBS)
  - Y otras...

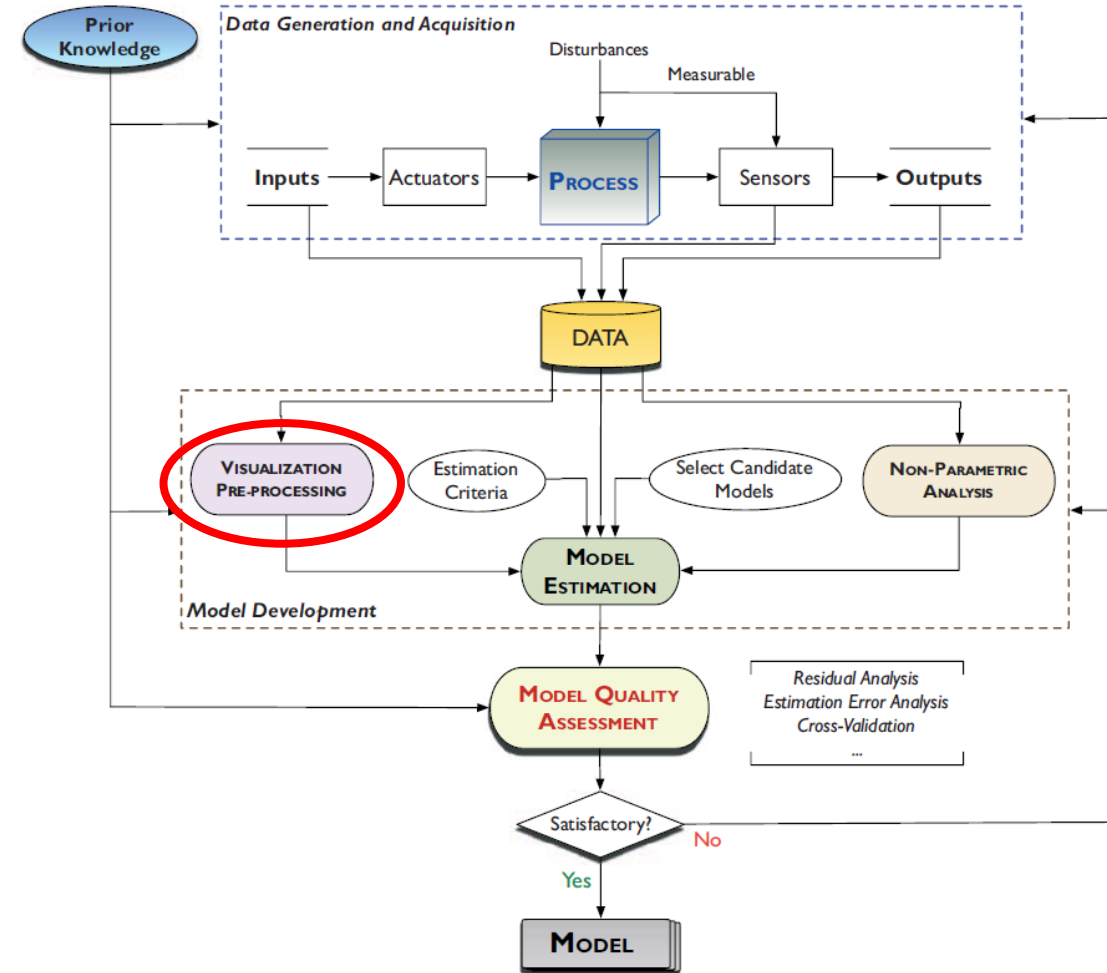




# CA3: Introducción a la Identificación de Sistemas

## Etapas en la identificación de sistemas – Visualización de los datos

- La visualización de datos es un paso clave en la extracción de información y el análisis de señales.
- Antes de la etapa de procesamiento, el examen visual ayuda a identificar la presencia de derivas, valores atípicos, etc. Permite verificar cualitativamente la calidad de los datos desde el punto de vista de la identificación (ej., excitación suficiente).
- También puede proporcionar información preliminar sobre el retardo, la dinámica y la ganancia del proceso. La información obtenida en esta etapa se puede utilizar en la etapa de evaluación de la calidad del modelo.
- La visualización de los datos en un dominio transformado (ej., el dominio de la frecuencia) puede resultar muy beneficioso. El examen de la entrada y la salida en el dominio de la frecuencia puede arrojar luz sobre el contenido espectral y la presencia de periodicidades. Es útil para obtener una idea de primera mano del nivel de excitación (de entrada) y la naturaleza de filtrado del sistema.





# CA3: Introducción a la Identificación de Sistemas

## Etapas en la identificación de sistemas - Preprocesamiento de datos

- Los datos adquiridos sin procesar generalmente no están listos para usarse para el desarrollo de modelos. Usualmente deben someterse a controles de calidad y a un paso de procesamiento antes de presentarlos al algoritmo de estimación. Hay muchos factores que afectan la calidad de los datos además del ruido:
  - **Valores atípicos (Outliers):** son debidos en gran parte a fallas del sensor y/o desviaciones breves y abruptas del proceso. Su detección y el manejo pueden ser muy complicados y desafiantes debido al hecho de que no existe una definición universal de un outlier. Existen algunos métodos estadísticos razonablemente buenos para este propósito.
  - **Datos faltantes:** el mal funcionamiento intermitente del sensor, las interrupciones de energía, el muestreo no uniforme y las pérdidas de transferencia de datos son algunas de las razones comunes por las que faltan datos.
- El preprocesamiento de datos también puede estar motivado por las suposiciones, limitaciones y requisitos del desarrollo del modelo. Los datos pueden contener derivas (**drift**), **tendencias** y otras no estacionariedades.
- El prefiltrado es una forma elegante de abarcar métodos para manejar ciertas características de los datos como derivas y ruido, además de ajustes preferentemente precisos en rangos de frecuencia seleccionados.
- El preprocesamiento de datos puede consumir una cantidad significativa del tiempo y esfuerzo total en un proceso de identificación. La situación se puede aliviar considerablemente eligiendo una **instrumentación confiable y un sistema de adquisición de datos, y un diseño experimental cuidadoso.**

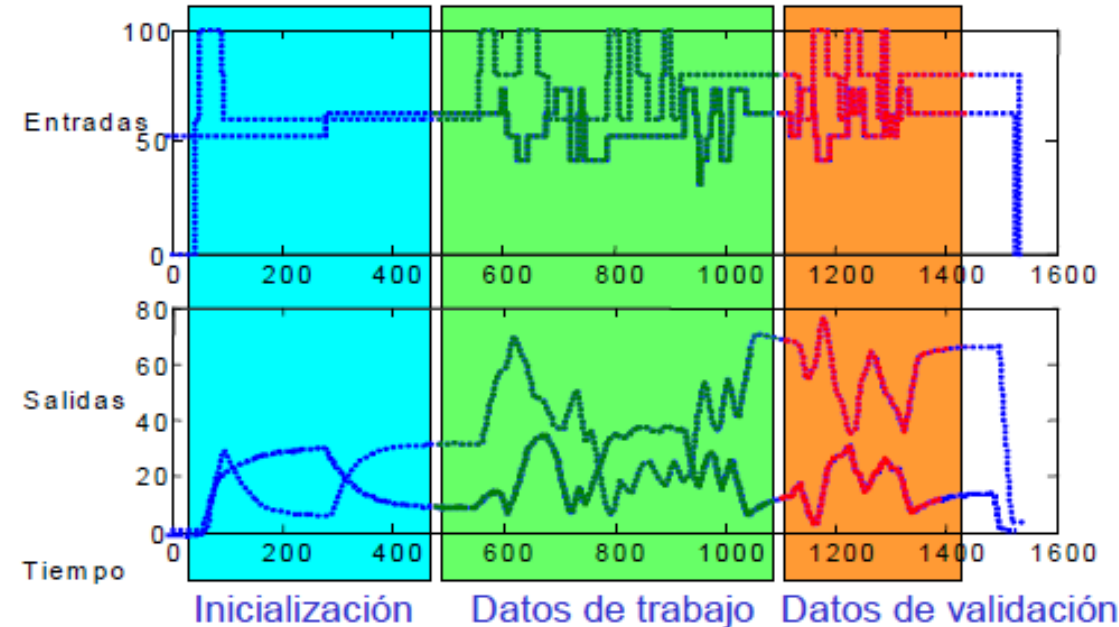




# CA3: Introducción a la Identificación de Sistemas

## Etapas en la identificación de sistemas – Partición de los datos

- Uno de los principales objetivos en la identificación es obtener un modelo con buenas capacidades predictivas. Es importante evaluar el modelo estimado en este sentido siendo necesario dividir los datos en dos conjuntos: datos de prueba y datos de entrenamiento.
- Se espera que los datos de prueba tengan algunas características "únicas" con respecto a los datos de entrenamiento, pero también comparten algunas características comunes.
- Una pauta práctica es construir primero el conjunto de datos de entrenamiento seleccionando aquellas características que se debe presentar el modelo. Los datos de prueba pueden ser entonces el conjunto remanente de observaciones o los datos completos.
- El conjunto de prueba debería contener al menos algunas características "nuevas" (ej. frecuencias de entrada que no están incluidas en los datos de entrenamiento). El objetivo principal es probar la interpolación del modelo y, en menor medida, las capacidades de extrapolación.



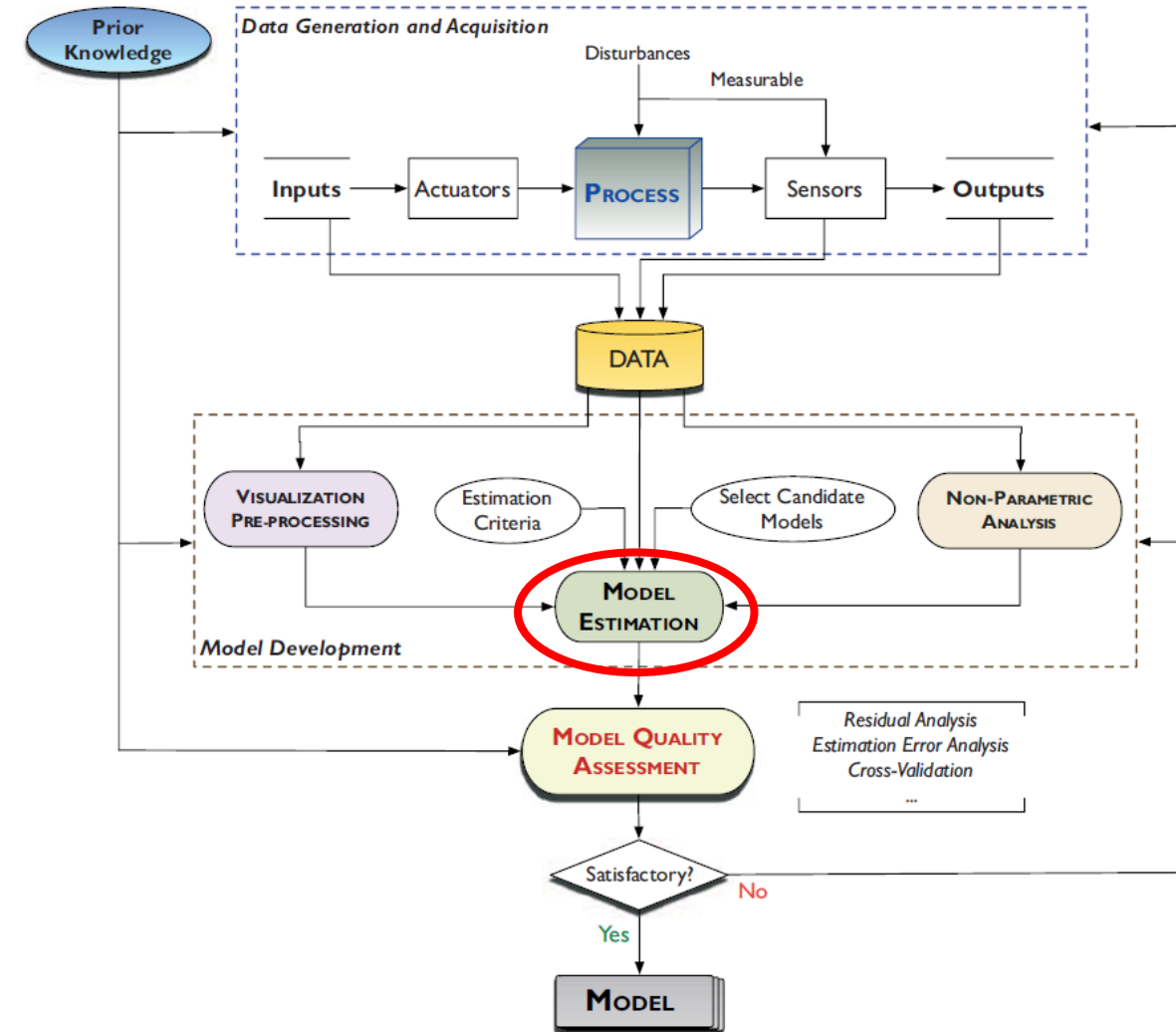




# CA3: Introducción a la Identificación de Sistemas

## Etapas en la identificación de sistemas – Desarrollo del modelo

- El objetivo general es construir un modelo determinista + estocástico. Implica:
  1. especificar la estructura y el orden del modelo  
En general, se comienza con una suposición inicial de la estructura y se realizan refinamientos iterativos utilizando los resultados de la evaluación de la calidad del modelo.
  2. estimar los parámetros del modelo resolviendo el problema de optimización asociado





# CA3: Introducción a la Identificación de Sistemas

## Etapas en la identificación de sistemas – Desarrollo del modelo (2)

1. **Elección de modelos candidatos:** es quizás la parte más desafiante y que consume más tiempo. El enfoque suele ser iterativo y se rige por las siguientes directrices:
  - **Requisitos de exactitud (sesgo) y precisión (varianza):** para un conjunto de datos dado, el sesgo y la varianza del modelo estimado están determinados por su estructura y el algoritmo de estimación.
  - **Precisión de la predicción y horizonte :** las consideraciones principales aquí son el rango (de pasos de tiempo) sobre el cual se busca la predicción y la precisión (sobre un rango de frecuencia). A menos que la aplicación exija explícitamente lo contrario, la predicción de un paso adelante es lo usual.
  - **Uso final del modelo:** además de las capacidades predictivas, la aplicación final puede imponer otros requisitos. Si el uso previsto es control, el modelo debería ser lo más simple posible (de orden bajo), no haber subestimado el retardo y las características del proceso en su ancho de banda bien capturadas.
  - **Aspectos de estimación:** La facilidad de estimación puede ser un factor importante a considerar. Los modelos que producen predictores lineales en parámetros suelen preferirse a los que no son lineales debido a la conveniencia del cálculo y la existencia de una solución única. El precio que se paga es la capacidad de predicción del modelo.
  - **Conocimiento previo:** La elección del modelo puede estar motivada por alguna información previa conocida sobre el tipo de modelos tales como lineales/no lineales, de bajo/alto orden, etc. o sobre la estructura del modelo paramétrico. Esta faceta de identificación cae dentro del ámbito del modelado de caja gris.

Uno de los mayores beneficios del modelado empírico es la flexibilidad en la selección de la estructura del modelo (gran libertad con riesgo de sobreajuste). Ningún proceso puede describirse con precisión mediante un modelo matemático. El objetivo de la identificación es desarrollar un modelo de trabajo bueno y útil.



2. **Método de estimación:** seleccionado un modelo candidato resta su estimación, que es esencialmente un problema de optimización.

Un criterio de estimación típicamente tiene una forma de minimización de una función de errores de predicción. La función generalmente se basa en una métrica de distancia.

- Como ejemplo, los métodos de mínimos cuadrados minimizan la distancia euclidia entre los valores pronosticados y observados. Por otro lado, los métodos basados en el principio de máxima verosimilitud construyen la función objetivo a partir de consideraciones probabilísticas.
- Otros factores como la calidad de las estimaciones de los parámetros y el número de parámetros, pueden incluirse en la función objetivo.
- Al elegir un algoritmo de estimación, los principales factores a considerar son la bondad de las estimaciones y la facilidad de cálculo. Por lo general, estos son factores en conflicto. Sin sacrificar las propiedades clave (precisión) de un estimador, se puede elegir un algoritmo computacionalmente más simple.

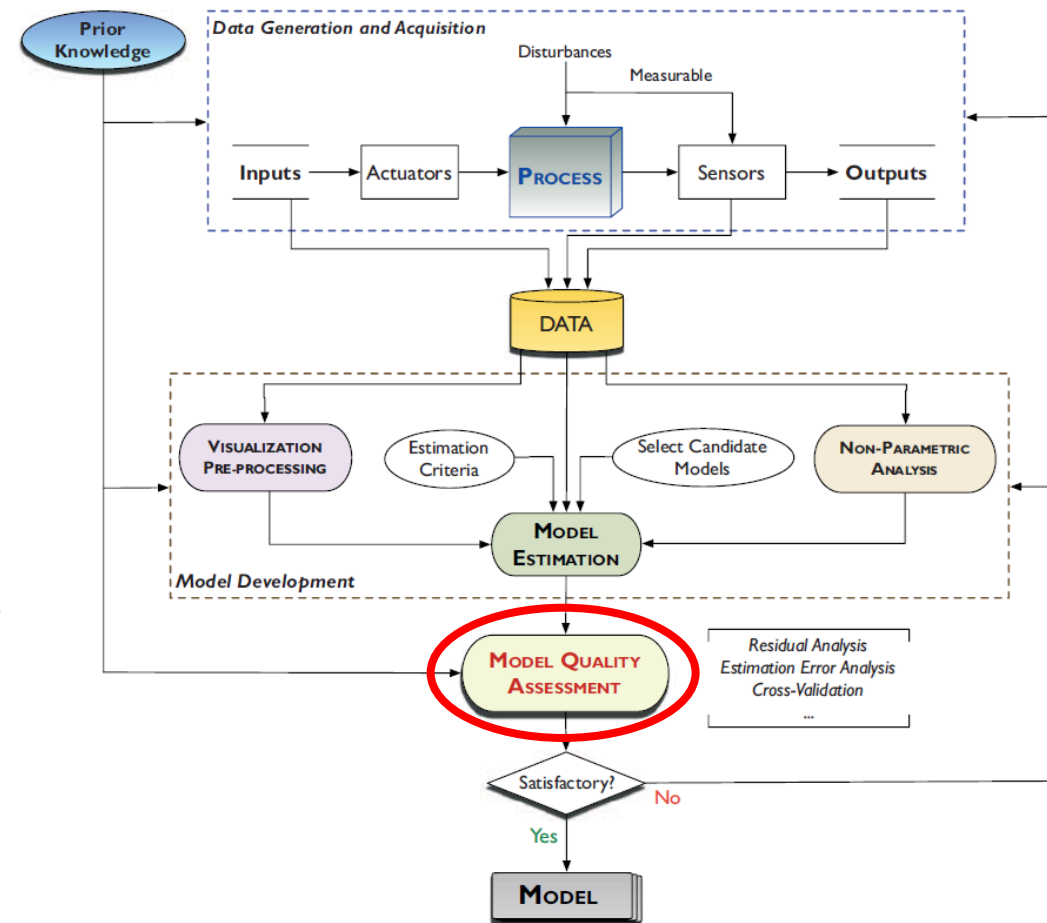


# CA3: Introducción a la Identificación de Sistemas

## Etapas en la identificación de sistemas – Evaluación y validación del modelo

- El paso de evaluación y validación de la calidad del modelo es una parte fundamental, ya sea una identificación o un enfoque de primeros principios.

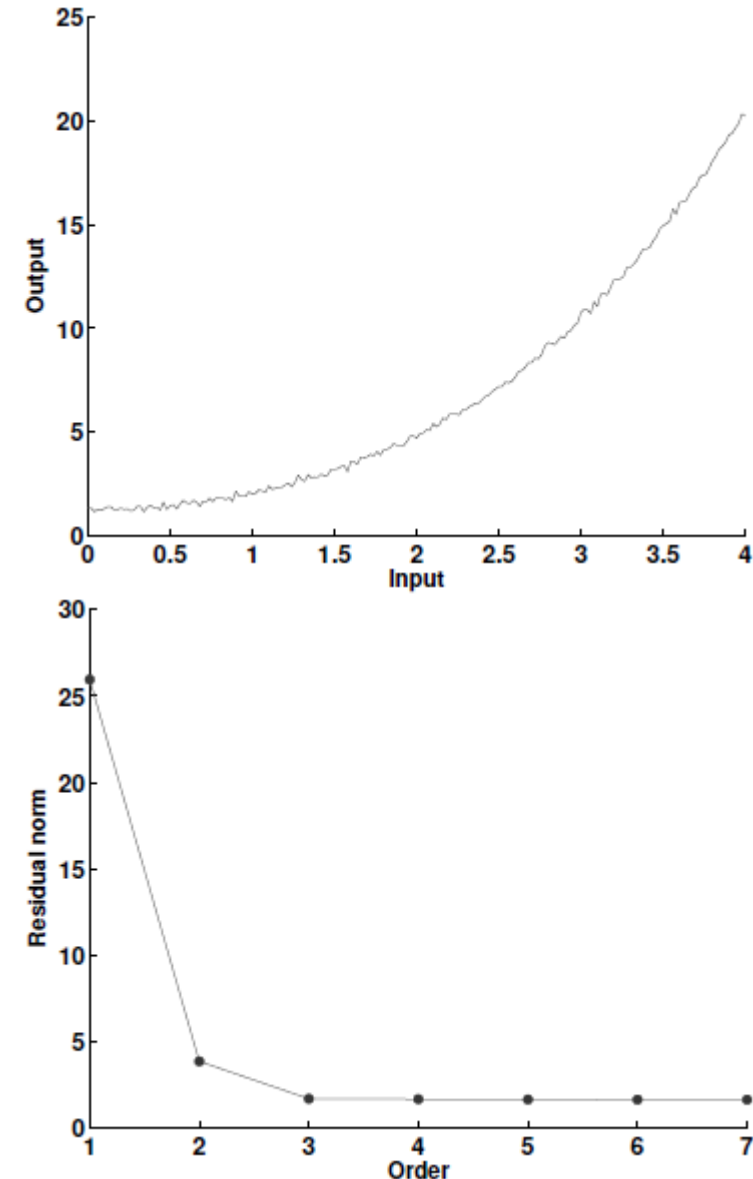
- ¿Con qué eficacia ha explicado el modelo las variaciones de salida en los datos de entrenamiento?** El objetivo general es lograr un error de predicción tan "pequeño" (sesgo mínimo) como sea posible, pero no a costa de una baja precisión (variación en el modelo o estimaciones de parámetros). Para esto, se realizan diferentes tipos de pruebas/análisis de los errores (utilizando métricas, herramientas estadísticas, etc.).
- ¿Con qué error predice el modelo la respuesta en un conjunto de datos de prueba fresco?** Esta es la prueba de validación cruzada. El objetivo principal es determinar si el modelo ha sido entrenado para capturar las características globales del proceso como se evidencia en el conjunto de datos de entrenamiento o si se ha especializado en las características locales de los datos de entrenamiento (modelo sobre ajustado). El rendimiento se evalúa en función de ciertas métricas de ajuste y horizontes de predicción.





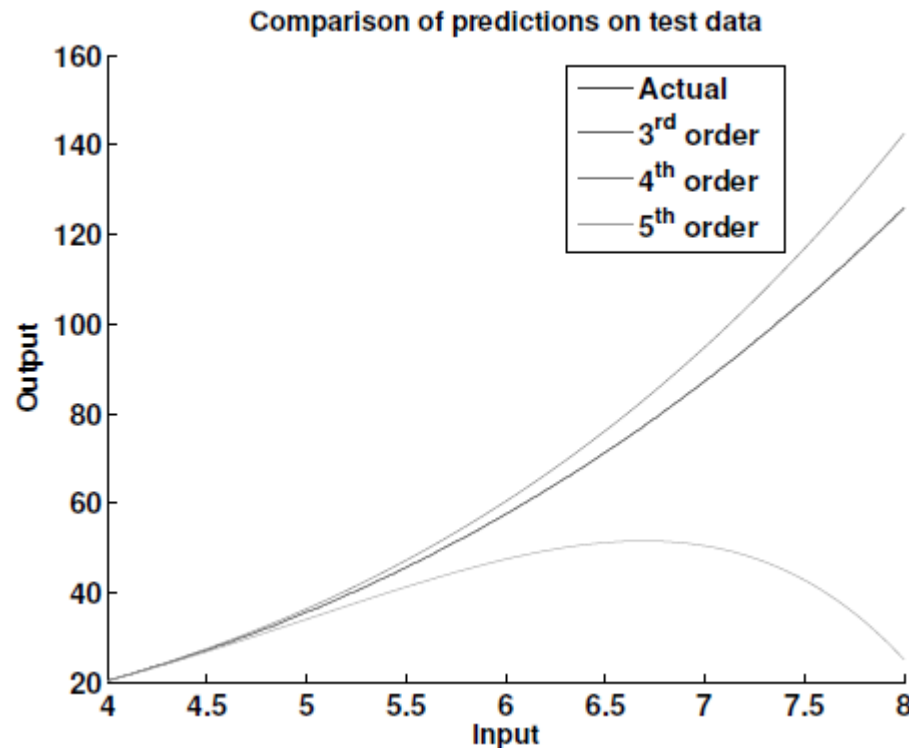
# CA3: Introducción a la Identificación de Sistemas

## Paso a paso de un proceso de identificación – Overfitting



¿Tercer orden es la elección óptima?. Esto no se puede determinar usando sólo los datos de entrenamiento, se requiere también un conjunto de datos de prueba para evaluar sus capacidades predictivas (procedimiento de validación cruzada).

- Se ajustan polinomios de orden tres a cinco y se prueban en un nuevo conjunto de datos.



La razón del bajo rendimiento de los modelos de orden superior es que han intentado explicar las variaciones aleatorias utilizando la entrada.



# CA3: Introducción a la Identificación de Sistemas

## Etapas en la identificación de sistemas – Evaluación y validación del modelo (2)

Los resultados de las pruebas diagnósticas anteriores brindan la retroalimentación necesaria para refinar las decisiones tomadas en etapas anteriores. Si un modelo no cumple con alguno de los requisitos antes mencionados, inevitablemente requiere mejoras en uno o más de los pasos anteriores. Cuando el usuario tiene confianza en la calidad de los datos, los esfuerzos deben dirigirse a refinar la estructura y/o el orden del modelo. También puede ser necesaria una revisión de decisiones en otras etapas.



# CA3: Introducción a la Identificación de Sistemas

## Etapas en la identificación de sistemas – Conocimiento previo del proceso

- Con bastante frecuencia puede estar disponible algún conocimiento a priori: **la estructura y/o el orden del modelo, los valores de un subconjunto de parámetros** y los **límites de los valores de los parámetros**. Incorporar este conocimiento adicional anticipa que el modelo resultante reflejará la física del proceso de una manera más transparente que un modelo clásico de caja negra y que el modelo resultante tendrá mejor capacidades de predicción que un modelo sin restricciones.
- El grado de transparencia aumenta con la información a priori disponible. El caso más transparente es el de la forma matemática conocida del modelo, generalmente a través de un análisis de primeros principios, quedando sólo por determinar los parámetros (modelo de **caja blanca**).
- El modelado de **caja gris** surge de manera prominente en el desarrollo de modelos de espacio de estado con estructura conocida. Para un sistema de entrada-salida dado, existen infinitas realizaciones de espacio de estado que producen la misma relación de entrada-salida. Sin embargo, una estructura particular puede ser deseable desde el punto de vista de la estimación o por la aplicación de interés. La optimización restringida de modelos paramétricos proporciona el marco necesario para la estimación de modelos de caja gris.
- Los modelos de **caja negra** suelen recibir críticas porque no tienen en cuenta explícitamente la física del proceso. A pesar de la corrección técnica de estas críticas, un punto en contra es que un análisis cuidadoso de los datos y una identificación sistemática pueden producir modelos de caja negra que funcionen bien y que tengan capacidades predictivas similares o, a veces, mejores que las de un modelo de primeros principios. Sin duda, el éxito de la construcción de estos modelos está anclado de manera crucial a la calidad de los datos.





# CA3: Introducción a la Identificación de Sistemas

## Etapas en la identificación de sistemas – Sugerencias para obtener un buen modelo

- **Datos de buena calidad:** diseñar y realizar experimentos para generar datos informativos. Los datos deben ser lo suficientemente ricos para distinguir entre dos modelos candidatos en competencia.
- **Visualización de datos:** a menudo este paso se menosprecia. El análisis visual de los datos puede extraer información valiosa con tanta facilidad que, de lo contrario, requeriría un análisis matemático elaborado. La información obtenida por un análisis cualitativo en esta etapa es útil para tomar decisiones en la selección del modelo, así como en la validación del modelo.
- **Dominio y estructura del modelo adecuados:** la estructura del modelo debe basarse en el uso final, la previsibilidad y los conocimientos físicos. La elección correcta del dominio (por ejemplo, frecuencia) para el preprocesamiento y modelado de datos puede mejorar significativamente la calidad del modelo.
- **Simplicidad del modelo:** los modelos simples pueden ser buenas aproximaciones de sistemas complicados. Un modelo complejo que produce una mejora marginal en las predicciones al precio de grandes errores en las estimaciones de los parámetros debe descartarse en favor de un modelo más simple.





# CA3: Introducción a la Identificación de Sistemas

## Etapas en la identificación de sistemas – Sugerencias para obtener un buen modelo (2)

- **Escala de tiempo para el modelado:** Identificar la escala de tiempo apropiada para el fenómeno de interés es crucial. El muestreo lento conduce a la pérdida de observabilidad, mientras que el muestreo excesivamente rápido produce grandes cantidades de ruido en relación con la señal y también puede llevar al sistema al borde de la inestabilidad.
- **Validación del modelo:** la interpretación correcta de la validación del modelo y la evaluación de la calidad del modelo es esencial para construir un buen modelo. Las comprobaciones de diagnóstico del modelo revelan información considerable sobre la suficiencia del modelo y pueden contener pistas sobre las instrucciones para perfeccionar el modelo.
- **No hay sustituto para el pensamiento, la perspicacia y la intuición:** la identificación es tanto una ciencia como un arte, ya que se logra mediante una combinación cuidadosa de conocimiento (del proceso), diseño (del experimento), experiencia e intuición (de la persona).



# Introducción a la Identificación de Sistemas

Índice

- Introducción
- Etapas de la identificación de sistemas
- **Señales de entrada**

Tema 05

V1 - 2024



# CA3: Introducción a la Identificación de Sistemas

## Señales de entrada – Factor de cresta

- La matriz de covarianza de estimación de parámetros suele ser inversamente proporcional a la potencia de entrada. Por lo tanto, interesa tener la mayor potencia de entrada posible.
- En la práctica, la limitación de entrada real se refiere a las restricciones de amplitud  $\underline{u}$  y  $\bar{u}$ . La propiedad deseada de la forma de onda se define en términos del factor de cresta  $C_r$ .
- Para una señal con media cero, el factor de cresta se define como

$$C_r^2 = \frac{\max_n u^2(n)}{\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N u^2(n)}$$

- Una buena forma de onda es la que consigue un factor de cresta bajo.
- El límite inferior teórico de  $C_r$  es 1 y se logra para señales simétricas binarias:  $u(t) = \pm u$ .



# CA3: Introducción a la Identificación de Sistemas

## Señales de entrada – Escalón

- El ensayo más utilizado en la industria para identificar el modelo de un proceso estable es la respuesta al escalón porque permite determinar modelos simples (FOPDT y SOPDT) que normalmente se usan para sintonizar controladores PID.
- **Ganancia ( $K_p$ , ganancia de CC):** Se define como el cambio total logrado en la salida por cambio en la entrada cuando la entrada pasa de una condición de estado estacionario a otra.

$$K_p = \left. \frac{\Delta y}{\Delta u} \right|_{ee}$$

- **Constante de tiempo ( $\tau$ ):** se define como el tiempo que tarda la respuesta de un sistema continuo de primer orden en alcanzar el 63,2 % de su valor de estado estacionario (respuesta al escalón de un sistema de tiempo continuo de primer orden).

$$y(t) = K_p(1 - e^{-\frac{t}{\tau}})$$

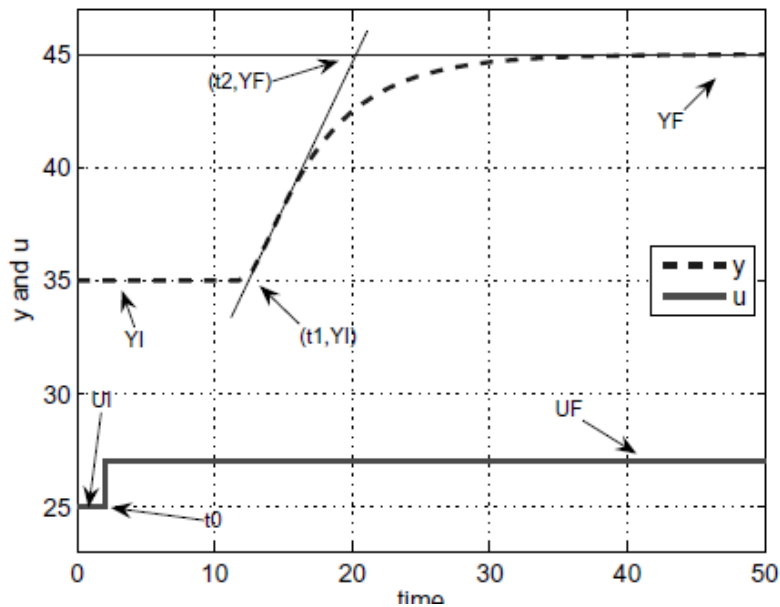
- **Retardo:** En sentido estricto, las características de retardo de un sistema no están relacionadas con su dinámica. Sin embargo, los retardos aparentes, que resultan debido a las aproximaciones de sistemas de orden superior por modelos de orden inferior, dependen de la dinámica del sistema.



# CA3: Introducción a la Identificación de Sistemas

## Señales de entrada – Escalón (2)

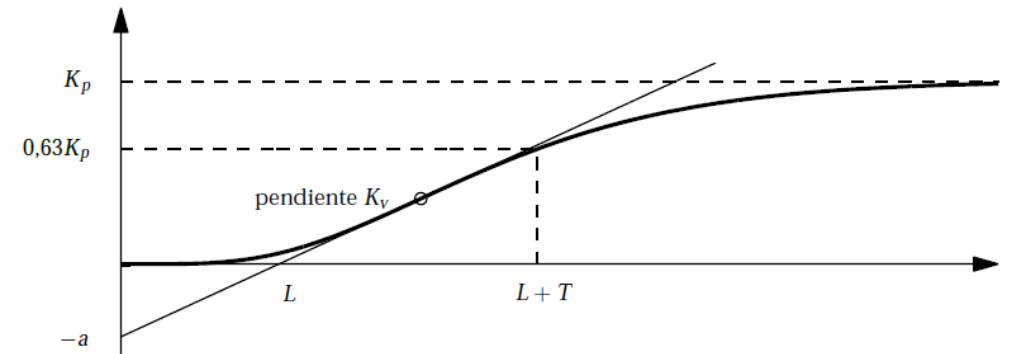
- Cuando el proceso presenta un retardo  $L$ , la salida permanece constante durante  $L$  unidades de tiempo después de la aplicación de la entrada escalón. Debido al ruido de medición la salida de la planta no será constante. Un procedimiento posible para obtener  $L$  consiste en determinar la banda de ruido ( $Bn$ ) durante el funcionamiento normal del proceso y luego monitorear la salida hasta que sale de esta banda ( $\Delta U > 20Bn$ ).



$$T = t_2 - t_1$$

$$L = t_1 - t_0$$

$$K_p = \frac{Y_F - Y_I}{U_F - U_I} = \frac{\Delta Y}{\Delta U}$$



$$G(s) = \frac{K_p}{1 + sT} e^{-sL} \quad G(s) = \frac{K_v}{s} e^{-sL} = \frac{a}{sL} e^{-sL}$$

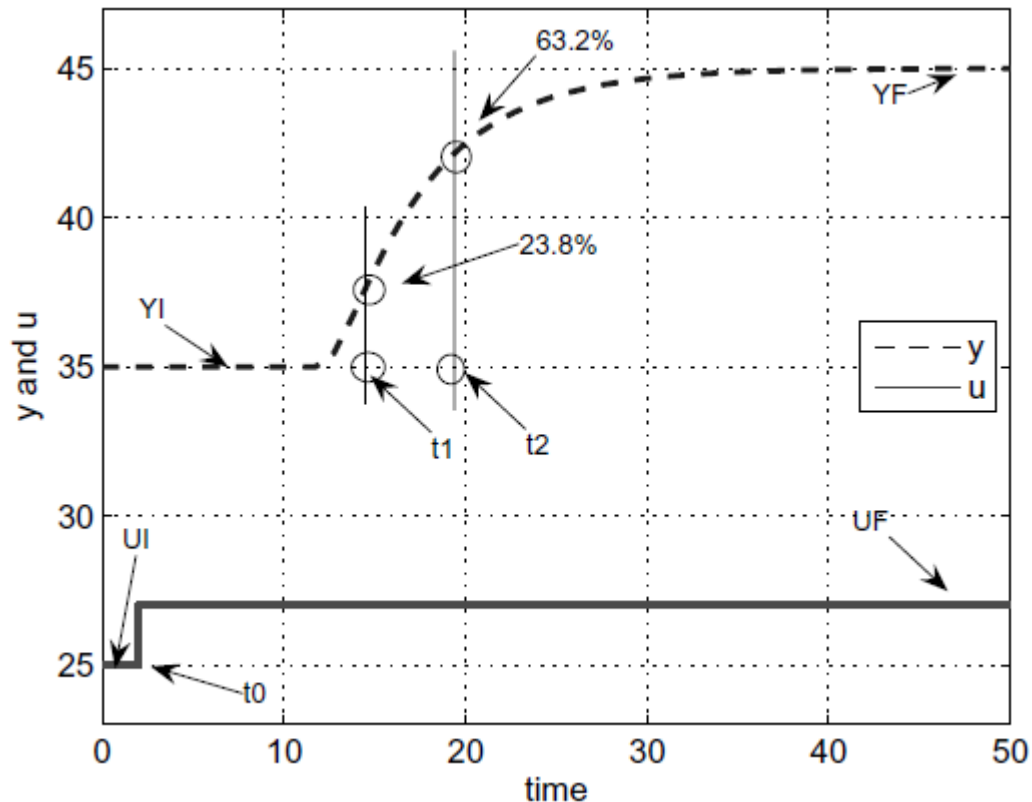
- En algunas situaciones, cuando un proceso estable domina el retraso ( $T \gg L$ ), el comportamiento cercano a un punto de operación puede aproximarse con éxito mediante un modelo integrador.
- Este método también se puede aplicar a un proceso integrador considerando una entrada de pulso de amplitud y longitud adecuadas en lugar de una entrada escalón.



# CA3: Introducción a la Identificación de Sistemas

## Señales de entrada – Escalón (3)

- Otros métodos miden varios puntos o características de la respuesta al escalón o impulso del proceso para calcular los parámetros del modelo, incluida la estimación del retardo. Debido al ruido estos métodos pueden obtener resultados más robustos.
- Para identificar un modelo FOPDT se miden dos tiempos:  $t_1$ , el tiempo en que la salida alcanza el 28,3% del cambio final en el valor de estado estacionario  $Y_F - Y_I$ , y  $t_2$ , cuando la salida alcanza el 63,2% de  $Y_F - Y_I$ .



$$K_p = \frac{Y_F - Y_I}{U_F - U_I} = \frac{\Delta Y}{\Delta U},$$
$$T = 1.5(t_2 - t_1),$$
$$L = 1.5[(t_1 - t_0) - \frac{1}{3}(t_2 - t_0)].$$

- Si se prefiere un modelo de segundo orden con retardo se deben identificar cuatro parámetros y se utiliza un método de tres puntos. Primero, se calcula  $K_p$  como en el caso FOPDT. Luego se compara la respuesta real con la analítica en tres puntos.

$$y(t) = Y_I + K_p \Delta U \left[ 1 + \frac{T_2 e^{-(t-L)/T_2} - T_1 e^{-(t-L)/T_1}}{T_1 - T_2} \right]$$

Curve fitting



# CA3: Introducción a la Identificación de Sistemas

## Señales de entrada – Suma de sinusoides

- Las señales multisenso son una combinación de sinusoides de diferentes frecuencias:

$$u[k] = \sum_{i=1}^M a_j \sin(\omega_i k + \phi_i), \quad 0 \leq \omega_1 < \omega_2 < \dots < \omega_M \leq \pi$$

- Estas señales tienen la ventaja de que se puede localizar la energía del espectro muy precisamente en las frecuencias deseadas.
- El problema que presentan es el factor de cresta, que resulta de un valor hasta  $\sqrt{2M}$  (si están todas en fase y los  $a_j$  son todos iguales). Una manera de controlar el factor de cresta es elegir las fases de manera que los cosenos estén lo más fuera de fase posible.
- Como en los sistemas físicos reales típicamente la amplitud máxima de excitación está limitada, en la práctica, las fases se ajustan y/u optimizan para obtener un factor de cresta mínimo, buscando la mejor relación señal-ruido (SNR) para la medición de la función de respuesta de frecuencia.
- Dado que el espectro no es continuo, las estimaciones en otras frecuencias no están disponibles.



# CA3: Introducción a la Identificación de Sistemas

## Señales de entrada – Chirp

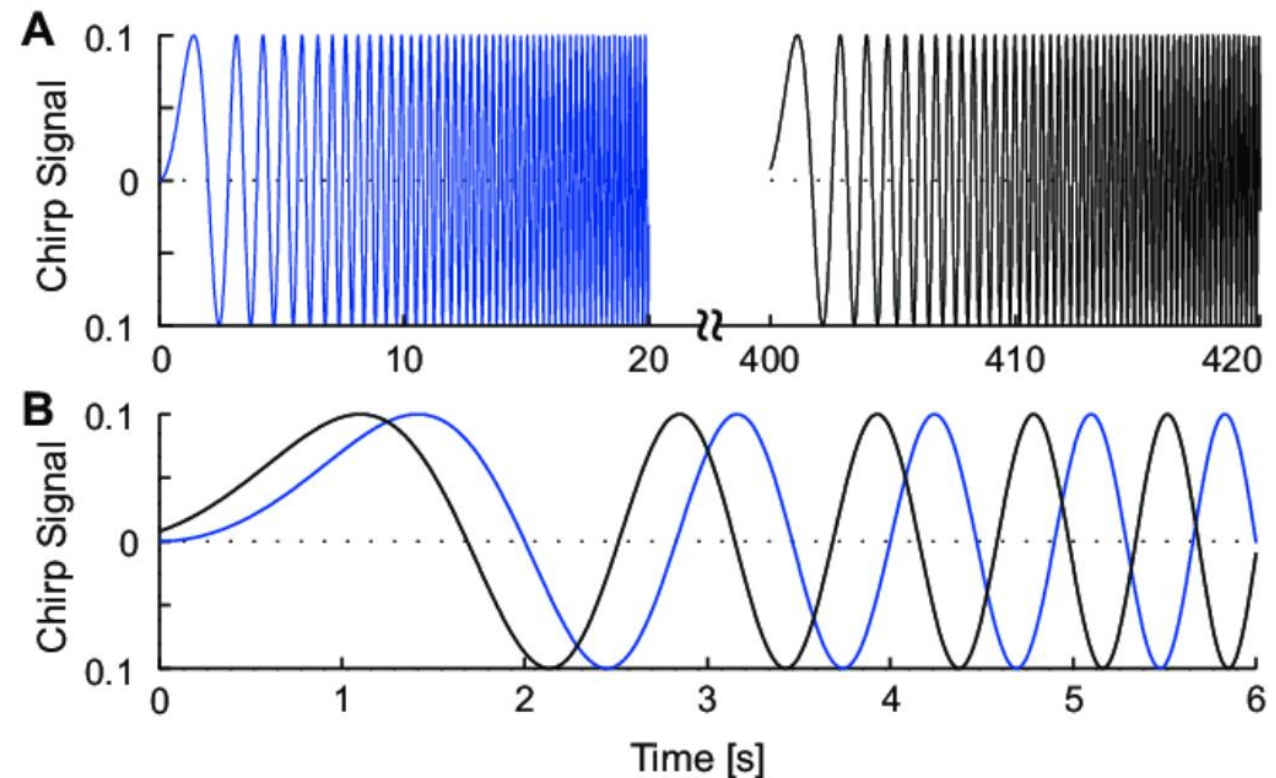
- Una señal chirp es una senoide con una frecuencia que cambia en forma continua en una cierta banda de frecuencias ( $\Omega$ :  $\omega_1 \leq \omega \leq \omega_2$ ) en un intervalo de tiempo  $T$ :

$$u(t) = A \cos(\omega_1 t + (\omega_2 - \omega_1) t^2 / (2T))$$

- La frecuencia instantánea de la señal se obtiene derivando respecto al tiempo al argumento del coseno, y resulta:

$$\omega_i = \omega_1 + \frac{t}{T} (\omega_2 - \omega_1)$$

- Puede verse que la frecuencia instantánea crece linealmente con el tiempo desde  $\omega_1$  hasta  $\omega_2$ .
- La señal chirp tiene un factor de cresta igual al de una senoide pura, es decir  $\sqrt{2}$ , y permite un buen control de la banda de frecuencias de excitación.
- Debido al barrido de frecuencia, también habrá contribuciones de potencia fuera de la banda  $\Omega$ .



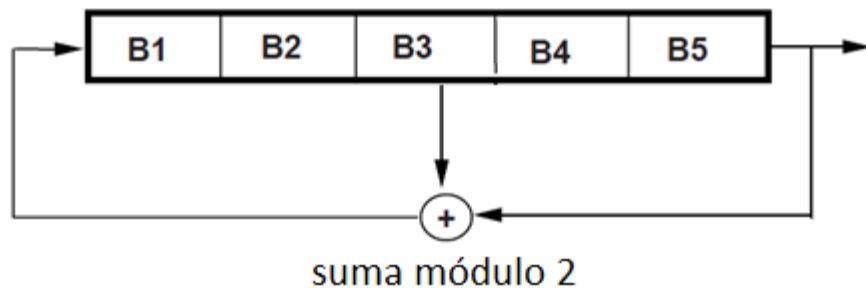




# CA3: Introducción a la Identificación de Sistemas

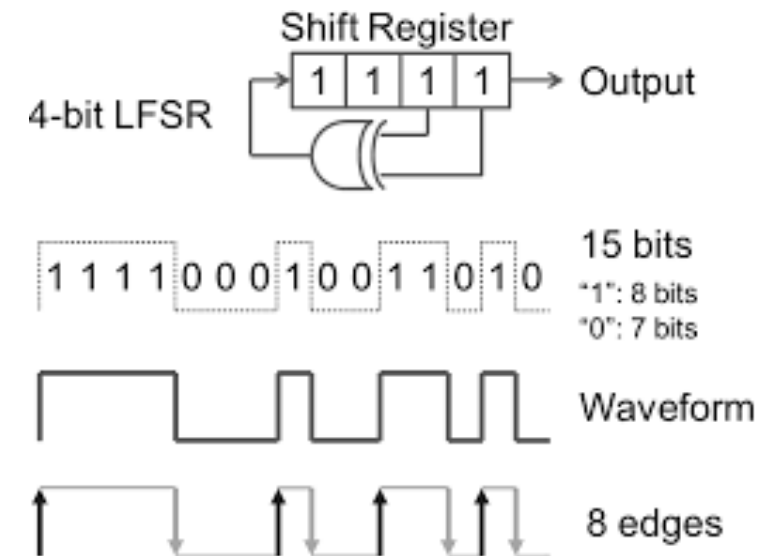
## Señales de entrada – PRBS

- Las secuencias binarias pseudoaleatorias (Pseudo-Random Binary Sequences) son secuencias de pulsos rectangulares, modulados por ancho, que se aproximan a un ruido blanco de tiempo discreto y por tanto tienen un contenido espectral rico en frecuencias.
- Deben su nombre pseudoaleatorio al hecho de que se caracterizan por una longitud de secuencia dentro de la cual las variaciones en el ancho del pulso varían aleatoriamente, pero que, en un horizonte de tiempo amplio, son periódicas, el período está definido por la longitud de la secuencia.
- Las PRBS se generan mediante registros de desplazamiento con realimentación. La longitud máxima de una secuencia es  $2^N - 1$  en la que  $N$  es el número de celdas del registro de desplazamiento.



Generación de una PRBS de longitud  $2^5 - 1 = 31$  periodos de muestreo.

# de celdas	Longitud de la secuencia $L = 2^N - 1$	Bits sumados
2	3	1 y 2
3	7	1 y 3
4	15	3 y 4
5	31	3 y 5
6	63	5 y 6
7	127	4 y 7
8	255	2, 3, 4 y 8
9	511	5 y 9
10	1023	7 y 10





# CA3: Introducción a la Identificación de Sistemas

## Señales de entrada – PRBS – Guía para el diseño

- Para identificar correctamente la ganancia en estado estacionario del modelo dinámico de la planta, la duración de, al menos, uno de los pulsos (por ejemplo, el pulso de duración máxima) debe ser al menos mayor que el tiempo de subida  $t_r$  de la planta (incluyendo el tiempo de retardo). Siendo la duración máxima de un pulso  $N T_s$ , resulta la siguiente condición:

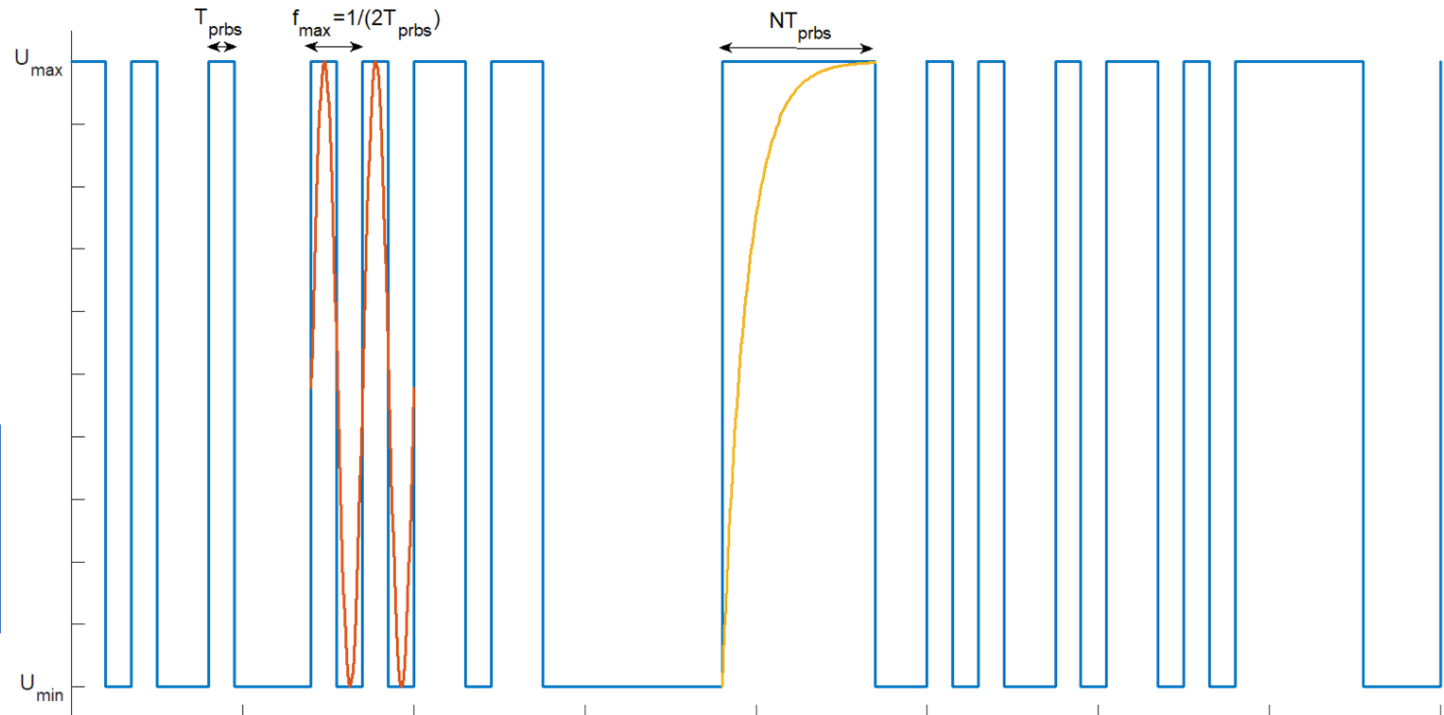
$$NT_{prbs} > t_r \longrightarrow N > \frac{t_r}{T_{prbs}}$$

$$L = 2^N - 1 \quad (\text{longitud de la secuencia})$$

$$D \geq L * T_{prbs} \quad (\text{duración del ensayo})$$

Para la elección de la frecuencia de muestreo no olvidarse de Nyquist.

Amplitudes adecuadas (system friendly).



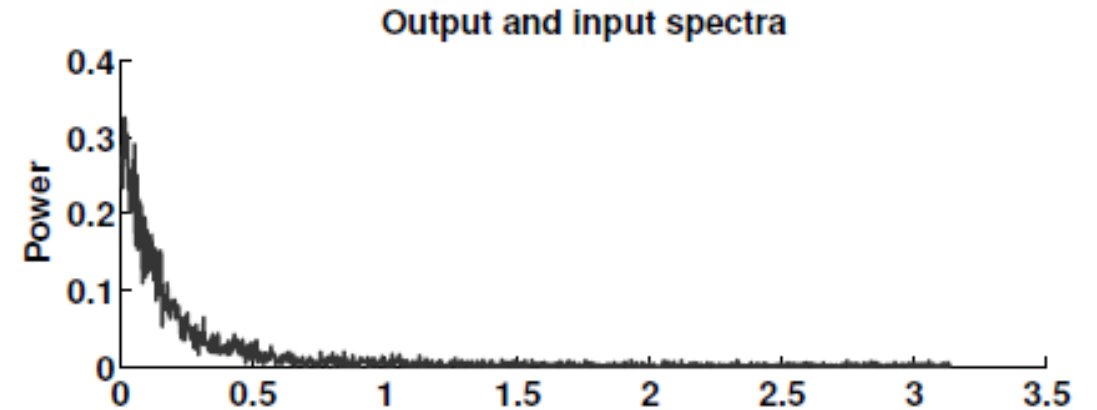
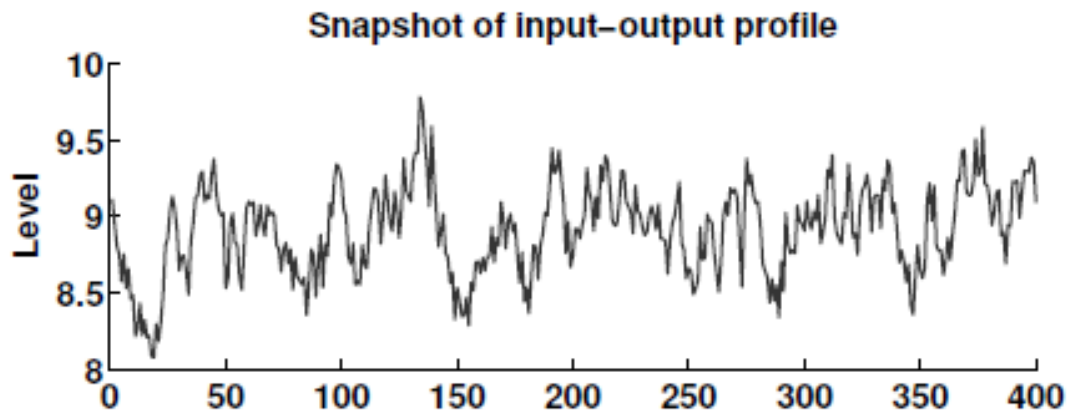
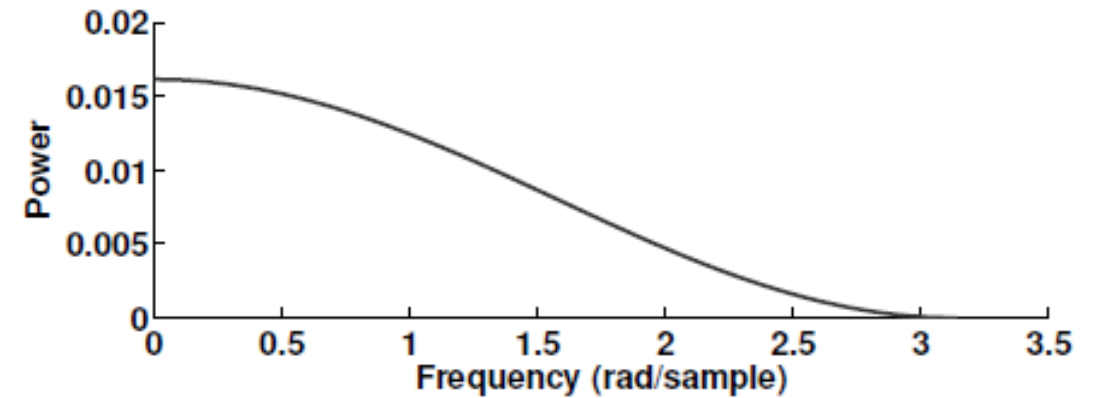
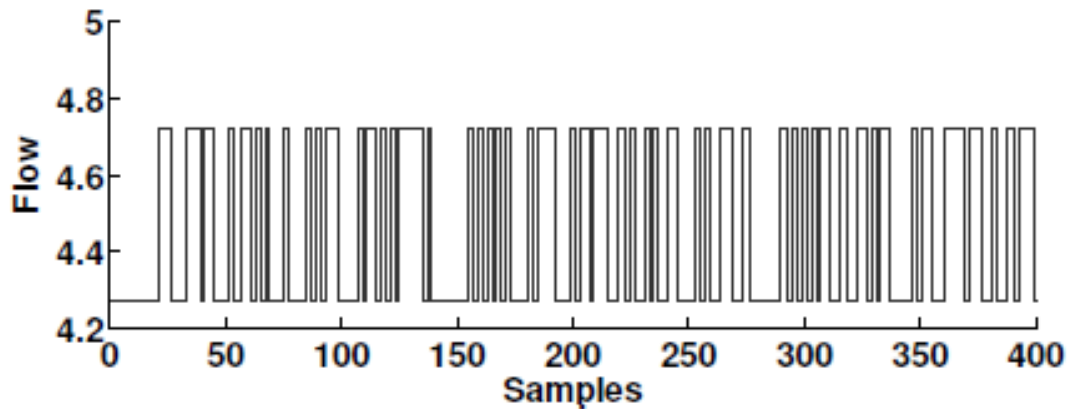
- Para cubrir todo el espectro de frecuencias generado por un PRBS particular, la duración de una prueba debe ser al menos igual a la longitud de la secuencia. En un gran número de casos, la duración del ensayo ( $D$ ) se elige igual a la duración de la secuencia.



# CA3: Introducción a la Identificación de Sistemas

## Señales de entrada – Análisis espectral

- Es útil examinar el contenido de frecuencia de la señal de entrada (diseño) y de salida para obtener información sobre la naturaleza de filtrado del proceso.





# CA3: Introducción a la Identificación de Sistemas

## Señales de entrada – Persistencia a la excitación

- Si se desea una distinción fina entre modelos, las entradas deben tener tantas frecuencias como sea posible.
- No es posible inyectar un continuo de frecuencias y, por lo tanto, los modelos solo se pueden resolver dentro de una pequeña región.
- Cuando la entrada, específicamente su densidad espectral, contiene casi todas las frecuencias, se dice que es persistentemente excitante.
- La excitación persistente es necesaria para el diseño de la entrada, pero no suficiente. No tiene en cuenta el esfuerzo de identificación ni el costo, amplitudes de entrada (desgaste del actuador), tasa de cambio de entradas, etc.
- El concepto de conjuntos de datos informativos está muy relacionado con los conceptos de entradas "persistentemente exitantes", entradas "suficientemente generales", etc.
- Se dice que una entrada que cumple con los requisitos prácticos de una operación de proceso es compatible con la planta (plant friendly).

En sistemas NL es necesario variar la amplitud también.



# CA3: Introducción a la Identificación de Sistemas

## Señales de entrada – Pruebas preliminares para el diseño de entradas

- Decidir el contenido de frecuencia de la entrada previo a la identificación es como saber la respuesta de ante mano. Para esto, se deben realizar algunas pruebas preliminares para llegar al diseño de entrada óptimo, a menos que esta información se conozca a priori.
  1. Respuesta al escalón del sistema (3% - 10% de magnitud). La respuesta al escalón arroja luz sobre la ganancia, la constante de tiempo, el retardo, la respuesta inversa, etc.
  2. La respuesta al escalón debería realizarse al menos en dos direcciones o mejor, aplicando una señal de escalera para verificar los efectos de las no linealidades y el rango de linealización.
  3. A partir de la respuesta al escalón, identificar la constante de tiempo  $\tau$  efectiva del proceso.
  4. Calcular el ancho de banda efectivo  $BW = 1/\tau$ .
  5. Utilizar la frecuencia de muestreo  $f_s = 1/T_s$  en cualquier lugar entre 10 y 20 veces BW.
  6. Configurar la frecuencia máxima en la entrada para que sea aproximadamente 2-3 veces el BW.
  7. Diseñar una secuencia de entrada del tipo apropiado (blanco, rbs, multiseno, prbs) en consecuencia.



# CA3: Introducción a la Identificación de Sistemas

## Bibliografía

- Arun K. Tangirala. “Principles Of System Identification. Theory and Practice”. CRC press, 2015.
- Ioan D. Landau, Gianluca Zito. “Digital Control Systems Design, Identification and Implementation”. Springer 2006.
- J.E. Normey-Rico and E.F. Camacho. “Control of Dead-time Processes”. Springer, 2007.
- [Algunas más...](#)
- Lennart Ljung. “System Identification. Theory for the User”. Second Edition, Prentice Hall, 1999.
- Lennart Ljung. “System Identification Toolbox™. Getting Started Guide”. MathWorks, 2016.