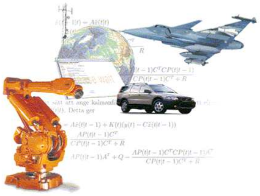
Identification of Linear and Nonlinear Dynamical Systems

Theme 4: Practical Aspects



Lennart Ljung

Division of Automatic Control
Linköping University
Sweden

Practical Aspects

- Choice of Model Structure and Model Validation
- Experiment Design
- (Data Preprocessing)

Model Structure: Choices To be Made

- Type of model
  - Black-Box or Tailor-made?
  - linear or non-linear?
  - ARX, ARMAX, BJ, OE ...
  - transformation of raw measurements
  - ...
- Size of model
  - $\mathcal{M}_1 \subset \mathcal{M}_2 \subset \mathcal{M}_3 \subset \cdots$
  - order determination
  - which variables to include

Considerations at Different Stages

- Prior: Data independent considerations
  (type of model, possible model orders, etc)
- Preliminary: Tests using data
  (Non-parametric, model orders, etc)
- Posterior: Comparing results in different model structures
  (model complexity, AIC, FPE, etc)
- Model Validation: Does the resulting model solve my problem?

Prior Considerations

- Quality and Price of Model
- Black Box versus Physically Parameterized
- Non-linear transformations of data
- Type of Black-Box model
- Try Simple Things First! (TSTF)

Quality and Price of a Model

QUALITY $\approx$ Small error
Mean square Error = Bias Error^2 + Variance Error

Variance increases with the number of free parameters, while Bias decreases
TRADE-OFF FLEXIBILITY — PARSIMONY

PRICE $\approx$ ease of computation
How much work to evaluate $\hat{y}(t|\theta)$ given θ?
How complex is the "surface"

$$V_N(\theta) = \sum |\ y(t) - \hat{y}(t|\theta)\ |^2$$

Black Box versus Grey Box models

Black-Box $\leftrightarrow$ Grey-Box

Physically Parameterized model structures may require considerable modeling work. They may (but not always) lead to models with drastically fewer parameters.

Physically parameterized structures usually give a good trade-off between flexibility and parsimony. They, however, may require a lot of computational (and programming!) work for the actual parameter estimation phase.

TSTF suggests that one might start with ready-made models.

In any case one should think over the physics of the application before starting the identification process (Semi-physical modeling)

.

Silly Models

Look back at solar-house data.
Why was this structure a failure?

$$A(q)y(t) = B_1(q)I(t) + B_2(q)u(t) + e(t)$$

($y(t)$: Storage temperature; $I(t)$: Solar intensity; $u(t)$: Fan velocity)

Non-Linear Transformations of Data

BASIC ADVICE:
Take out as much nonlinearity as you can!

- Apply physical insight!
- Logarithms make products sums
- Sensor & actuator dynamics often non-linear (but known)
- Linear regressions are nice structures!

Choice of Type of Black-Box Model

ARX - OE - ARMAX - BJ

- ARX $Ay = Bu + e$ is a linear regression model and gives a simple and efficient estimation problem. Basic drawback: Noise-model $1/A$ linked to system dynamics, which typically leads to biased estimates. May require higher orders. Not so significant problem if SNR is good.
- ARMAX $Ay = Bu + Ce$ gives better flexibility in noise modeling. An often used model.
- Non-structured State-Space = ARMAX

Linear black-box models, ct'd

- OE $y = \frac{B}{F}u + e$ focuses on the system dynamics. For open loop data, correct dynamics may be obtained, even if noise model is incorrect. Minimization of the criterion may be more difficult than in the ARMAX case.

- BJ $y = \frac{B}{F}u + \frac{C}{D}e$ is the "complete" model for which the dynamics modeling is separated from the noise modeling.

- If the noise enters "early" in the process, ARX and ARMAX have advantages, since the noise is likely to use the same poles as the input.
- If the noise enters "late" in the process, OE and BJ have advantages, for the opposite reason.

Posterior Considerations: Comparing Structures

- What to compare?
  - simulation performance
  - prediction performance
- Comparing models on fresh data sets
- Comparing models on second-hand data sets
- `compare(zv,th,k)`

Comparing Models on Fresh Data Sets

CROSS-VALIDATION

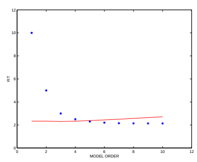
"THE ULTIMATE TEST". Needs no probabilistic justification.

Plot simulated outputs and measured outputs for the fresh data set. Alternatively one may use predictions over longer periods.

Typically the performance can be evaluated as sum of squared mismatches.

Comparing Models on Second-Hand Data

BASIC IDEA: Compensate for over-fit



- Add Model Complexity Penalty
  - Akaike: AIC, FPE
  - Rissanen MDL
- Hypothesis test
  - Check if decrease in fit is larger than "expected"

AIC, FPE, MDL

AIC for Gaussian case:

$$\min_{\mathcal{M}} \min_{\theta_{\mathcal{M}}} [(1 + \frac{2d_{\mathcal{M}}}{N}) \cdot \frac{1}{N}\sum_{t=1}^{N} \epsilon^2(t, \theta_{\mathcal{M}})], \quad d_{\mathcal{M}} = \dim \theta_{\mathcal{M}}$$

[-log likelihood + $d_{\mathcal{M}}$]

FPE similar, with $(1 + \frac{2d_{\mathcal{M}}}{N})$ replaced by $\frac{1+d_{\mathcal{M}}}{1-d_{\mathcal{M}}}$. Aims at estimating the fit that would be obtained for a fresh data set.

MDL is like AIC but with a $\frac{2d_{\mathcal{M}} \log N}{N}$ penalty.

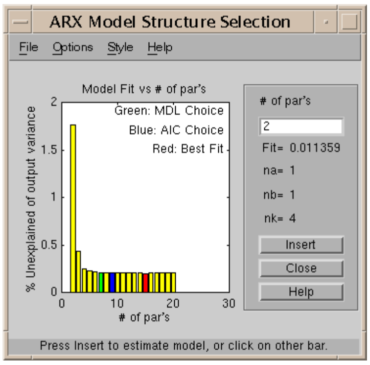
Comparing Many ARX Models Efficiently

At an initial stage it is often useful to examine a relatively large number of models. For models of linear regression type this can be done quite efficiently. Essentially the procedure is to form the "largest" structure $y(t) = \Phi^T(t)\theta$ (from which all others can be formed by deleting regressor variables), and to do row- and column manipulations in the resulting matrices to find out the corresponding loss function values $V_N(\hat{\theta}_N)$ (prediction error variances).

SITB: `arxstruc`

An Example

```
V=arxstruc(ze,zv,struc(1:5,1:5,0:5));
nn=selstruc(V)
```

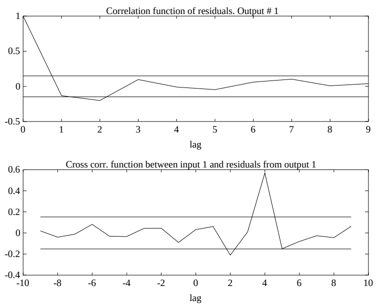


Model Validation: Residual Analysis

$$\varepsilon(t) = \varepsilon(t, \hat{\theta}_N) = y(t) - \hat{y}(t|\hat{\theta}_N)$$

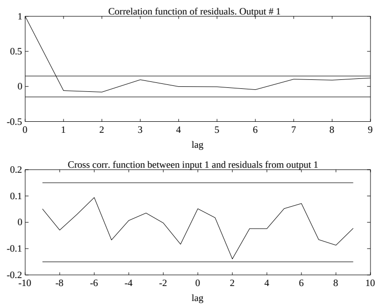
- $\varepsilon(t)$ should be independent of $u(t-\tau), \tau > 0$
- $\varepsilon(t)$ should ideally be white noise.
- Compute the corresponding correlation functions
- Confidence intervals for the estimates can also be computed

Typical Plots



True order `[2 2 3]`, guessed `[2 2 2]`

Typical Plots, ct'd



Guessed `[2 2 3]`

Note the Following

- If $R_{\varepsilon u}(\tau)$ is "large", then there is something in $y(t+\tau)$ that originates from $u(t)$ and is not properly explained by the model. Modify `nb` and `nk` accordingly.
- If $R_{\varepsilon u}(\tau)$ is "large" for negative τ, then there is evidence of output feedback in the input.
- If ε is calculated for ARX-models on the same data, then $R_{\varepsilon u}(\tau)$ is zero for those τ, that are "included in the model"

Determining Orders and Delays: Possibilities

- Compare models of different orders and delays
- Study $R_{\varepsilon u}(\tau)$ to find if the orders should be increased or delays should be decreased.
- Too high model orders often lead to large variances.
- Compare with spectral analysis estimate of Bode plot.

Suggested Start-Up Procedure

STEP 1: Get a feel for the problem (`Quick-Start`)

- Use a high order FIR model for a quick estimate of the Impulse Response. That gives info about delays and time constants
- Use spectral analysis for a quick estimate of the Frequency Response. This gives some info about resonances and is a good basis for further comparisons
- Compute a 4:th order ARX model with delay estimated from the FIR model and a default order state-space model using `n4sid`.
- Compare the responses of these models with validation data.
- Does it look good? Yes: go to step 3. No: Go to step 2.

Step 2: Checking the Basic Structure

1. Check out higher ARX-orders.
2. More inputs required?
3. Some essential non-linearities missing? Try new "inputs" as transformations of raw data.
4. If all of the above fails, try out non-linear black box models.

Step 3: Fine tuning orders and delays

1. Estimate 2nd order ARX-models with all “possible” delays
2. Pick “best” delay; Check also with `cra`
3. Estimate “many” ARX-models with this (and around this) delay.
4. Pick “best” orders
5. Check pole–zero cancellations in dynamics to “get an idea” about other interesting structures (State-Space, ARMAX, BJ, OE)
6. Use `compare` to check that important features are captured
7. $\cdots$

Model Validation: An arsenal of tools

- Intended model use
- Feasibility of physical parameters
- Consistency of input/output behavior
- Model reduction resilience
- Parameter confidence intervals
- Simulation
- Residual tests
- Critical data evaluation

Choice of Model Structure — SUMMARY

- Try simple things first!
- Use physical insight!
- Linear regressions are nice structures!
- Use comparisons on fresh data sets as a basic guideline!
- Have an arsenal of model validation techniques as advisers – Take the decision yourself!

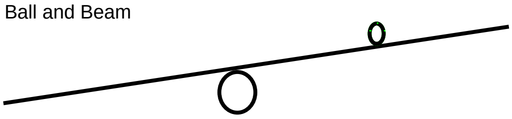
Experiment Design

To make sure that the experimental data are (maximally) informative with respect to the model we want to build.

- What to measure?
- When to measure?
- What to manipulate?
- How to manipulate?

Think about this:

Process: Ball and Beam



Estimated model:

$$y(t) - 1.8y(t-1) + 0.91y(t-2) = 0.5(1.1u(t-1) + 0.9u(t-2))$$

Theoretically a double integrator:

$$y(t) - 2y(t-1) + y(t-2) = 0.5(u(t-1) + u(t-2))$$

Actually worse than one would expect for a long experiment with low noise level. Why?

Informative Experiments

An experiment is INFORMATIVE if it allows you to distinguish between two different models (in the sets that you might consider).

Example 1: $u(t) = \sin \omega t$

Example 2: $u(t) = -f * y(t)$

Basic Idea For Informative Experiments

Model 1: $\hat{y}_1(t|t-1) = H_1^{-1}(q)[G_1(q)u(t) + (1 - H_1^{-1}(q))y(t)]$
Model 2: $\hat{y}_2(t|t-1) = H_2^{-1}(q)[G_2(q)u(t) + (1 - H_2^{-1}(q))y(t)]$

Experiment not informative $\Longleftrightarrow \hat{y}_1(t|t-1) \equiv \hat{y}_2(t|t-1)$

$$\Longleftrightarrow$$

(\$) $\quad M(q)u(t) \equiv L(q)y(t)$

(orders of M & L $\approx$ 2*model orders)

Hence if (\$) holds (BUT ONLY THEN!) we are in trouble.

Basic Criteria for Informative Experiments

Open Loop:
Require

$$M(q)u(t) \equiv 0 \Rightarrow u(t) \equiv 0$$

If $M(q)$ is of order n, we the say that $u(t)$ is Persistently Exciting of order n, p.e.(n) This is the same as requiring at least $n/2$ different sinusoids in the input

Closed Loop:
If there is no linear, time-invariant, noise/reference signal -free feedback from y to u we are OK.

.

Formal Calculations 1/3

True system (or second order LTI invariant) G_0: Then

$$\begin{aligned}\varepsilon(t,\theta) &= H_\theta^{-1}(y(t) - G_\theta u(t)) = H_\theta^{-1}[(G_0 - G_\theta)u(t) + H_0 e(t)] \\ &= H_\theta^{-1}[\Delta G_\theta u(t) + \Delta H_\theta e(t)] + e(t)\end{aligned}$$

$$(\hat{G},\hat{H}) \to \arg\min \int_{-\pi}^{\pi} |H_\theta|^{-2} \begin{bmatrix} \Delta G_\theta & \Delta H_\theta \end{bmatrix} \Phi_\zeta \begin{bmatrix} \Delta G_\theta^* \\ \Delta H_\theta^* \end{bmatrix} d\omega$$

$$\Phi_\zeta(\omega) = \begin{bmatrix} \Phi_u(\omega) & \Phi_{ue}(\omega) \\ \Phi_{eu}(\omega) & \Lambda_0 \end{bmatrix}$$

No assumption about feedback etc, just that the spectrum exists.

Note also that any pre-filter L, $\varepsilon_F(t) = L(q)\varepsilon(t)$ can be included in the noise model, $\tilde{H}_\theta = H_\theta/L$.

Formal Calculations 2/3

Factorize!

$$\begin{bmatrix} \Phi_u & \Phi_{ue} \\ \Phi_{eu} & \Lambda_0 \end{bmatrix} = \begin{bmatrix} I & \Phi_{ue}\Lambda_0^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \Phi_u^r & 0 \\ 0 & \Lambda_0 \end{bmatrix} \begin{bmatrix} I & 0 \\ \Lambda_0^{-1}\Phi_{eu} & I \end{bmatrix}$$

$$\Phi_u^r = \Phi_u - \Phi_{ue}\Lambda_0^{-1}\Phi_{eu}, \quad \Phi_u = \Phi_u^r + \Phi_u^e$$
$$\Phi_e^r = \Lambda_0 - \Phi_{eu}\Phi_u^{-1}\Phi_{ue}$$

Φ_u^r="That part of u that cannot be estimated from e by a LTI filter"

Formal Calculations 3/3

Basic Idea For Informative Experiments

$$\int_{-\pi}^{\pi} \begin{bmatrix} \Delta G_\theta & \Delta H_\theta \end{bmatrix} \Phi_\zeta \begin{bmatrix} \Delta G_\theta^* \\ \Delta H_\theta^* \end{bmatrix} d\omega = 0 \Rightarrow \Delta H_\theta = 0,\ \Delta G_\theta = 0$$

Recall

$$\Phi_\zeta = \begin{bmatrix} \Phi_u & \Phi_{ue} \\ \Phi_{eu} & \Lambda_0 \end{bmatrix} = \begin{bmatrix} I & \Phi_{ue}\Lambda_0^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \Phi_u^r & 0 \\ 0 & \Lambda_0 \end{bmatrix} \begin{bmatrix} I & 0 \\ \Lambda_0^{-1}\Phi_{eu} & I \end{bmatrix}$$

So the question is

$$\int |\Delta G_\theta(e^{i\omega})|^2 \Phi_u^r(\omega) d\omega = 0 \Rightarrow \Delta G_\theta = 0?$$

The signal u^r should be persistently exciting of the same order as the model/system.

Identification of Closed Loop Systems

- Direct Approach:
  - Forget about feedback!
  - OK if experiment informative a PEM is used
- Any Problems?
  - Typically less information in data
  - Be careful with spectral and correlation analysis
  - Be careful with IV- and subspace-methods
  - Be careful with Output-Error methods. The noise needs to be modeled

Good Designs – Basic Principles

$\mathcal{X}$: The design variables

$$\hat{\theta}_N \to \theta^*(\mathcal{X}) \qquad \text{Cov}\ \hat{\theta}_N \approx \frac{\lambda}{N} P_\theta(\mathcal{X})$$

- The model $\mathcal{M}(\theta^*(\mathcal{X}))$ is the best approximation of the system under $\mathcal{X}$
-

$$P_\theta(\mathcal{X}) \approx \frac{1}{N}[\text{E}\psi(t)\psi^T(t)]^{-1} \qquad \psi(t) = \frac{d}{d\theta}\hat{y}(t|\theta)$$

Consequences

- Let the experimental conditions resemble those under which the model is to be used.
  Recall

$$\theta^* \approx \arg\min \int_{-\pi}^{\pi} |\ G_0(e^{i\omega}) - G(e^{i\omega},\theta)\ |^2 \cdot \frac{\Phi_u(\omega) \cdot |\ L(e^{i\omega})\ |^2}{|\ H(e^{i\omega},\theta^*)\ |^2} d\omega$$

- Choose experimental conditions and inputs, so that the predictor $\hat{y}(t|\theta)$ becomes sensitive to interesting and important parameters.
  Recall

$$\text{Cov}\ \hat{G}_N(e^{i\omega}) \approx \frac{n}{N} \cdot \frac{\Phi_v(\omega)}{\Phi_u(\omega)}$$

Optimal Designs

Typical problem formulation:

$$\min_{\mathcal{X} \in X} \alpha(P_\theta(\mathcal{X}))$$

X :Constrained input variance

Model properties depend only on the input spectrum $\Phi_u(\omega)$, the "color" of the input. It does not depend on the actual wave-form of the input.

.

General Bottom Line

Use your input energy in frequency bands where you need a good model and/or where the disturbances are significant.

$$''\Phi_u^{\text{opt}}(\omega) = \alpha\sqrt{C(\omega)\Phi_v(\omega)}''$$

$$\min_{\mathcal{X}} \text{E} \int_{-\pi}^{\pi} |\hat{G}(e^{i\omega}) - G_0(e^{i\omega})|^2 C(\omega) d\omega$$

Formal Calculations:

Choose all design variables so that the criterion

$$J(\mathcal{D}) = \int \mathrm{Var}|\hat{G}(e^{i\omega})|^2 C(\omega) d\omega$$

is minimized. Suppose that the design variables are:

- Reference signal spectrum
- Output feedback law
- Pre-filter L

under the constraints

- $$\alpha E u^2 + \beta E y^2 \leq 1$$

Formal Calculations 2/4

Then the solution is

- regulator $u(t) = -F_y(q)y(t)$ that solves the standard LQG problem

$$F_y^{\text{opt}} = \arg\min_{F_y} [\alpha E u^2 + \beta E y^2], \quad y = G_0 u + H_0 e$$

- Reference signal spectrum

$$\Phi_r^{\text{opt}}(\omega) = \mu \sqrt{\Phi_v(\omega) C(\omega)} \frac{|1 + G_0(e^{i\omega}) F_y^{\text{opt}}(e^{i\omega})|^2}{\sqrt{\alpha + \beta |G_0(e^{i\omega})|^2}}$$

Note the special case $\beta = 0$ and stable system $\Rightarrow F_y = 0$

Formal Calculations 3/4

MSE minimization

Choose all design variables so that the criterion

$$J(\mathcal{D}) = \int \mathrm{E}|\hat{G}(e^{i\omega}) - G_0(e^{i\omega})|^2 C(\omega) d\omega$$

is minimized Suppose that the design variables are:

- Reference signal spectrum
- Output feedback law
- Pre-filter L

under the constraints

- $$E u^2 \leq 1/\alpha$$

Formal Calculations 4/4

Then the solution is

- Open loop
- Input spectrum $\sim \sqrt{C \cdot \Phi_v}$
- Pre-filter $\sim \sqrt{\frac{\Phi_v}{C}}$

The Input Waveform

We want to

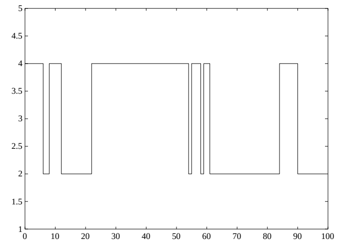
- Control the input spectrum
- Have small maximum amplitude for given power (crest factor)
- Utilize periodicity

Choices:

- Random Gaussian Noise
- (Pseudo) Random Binary Noise
- Sum of sinusoids, including swept sinusoids.

Binary Signals

$$u(t) = \begin{cases} \overline{u} \\ \underline{u} \end{cases}$$ shifting in a certain fashion, giving a certain spectrum $\Phi_u(\omega)$.



Time domain thinking: Occasionally, let a step response almost settle. No use to let the input shift so quickly that the system's response is hardly visible.

.

Periodic Inputs

When allowed, periodic inputs have certain advantages:

- Independent noise estimation
- Reduction of data sets, by averaging over the periods
- No leakage if frequency domain methods are applied

Some Typical Periodic Inputs

- PRBS
- Sum of sinusoids with tailored phases
- Swept sinusoid, (chirp signal)

Choice of T

- Variance increases rapidly when sampling slower than dominating time constants
- Poor return for extra work with fast sampling
- Sample $\approx$ 10 - 20 times the system bandwidth.
- Check step response: Put 3–5 measurements during the rise time.

Some Practical Aspects

- Always use Anti-alias filters!
  They provide noise reduction and avoid confusion with alias.
- With cheap data acquisition, sample fast at source.
  Postpone decision about T to software phase.
  [Digital anti alias filtering + decimation]
  SPTB command `resample`

Experiment Design — SUMMARY

- Let the system be excited!
- Open loop inputs: Binary, periodic signals with full control of spectral properties.
- Let the predictor be sensitive to important parameters!
- “Cov $\hat{G}_N(e^{i\omega}) \approx \frac{n}{N} \cdot \frac{\Phi_v(\omega)}{\Phi_u(\omega)}$
- Sample 10-20 times bandwidth!

Pretreatment of Data

ALWAYS FIRST PLOT THE DATA! Possible problems with measured data:

- Drift, offset, low frequency disturbances
- Occasional bursts and outliers
- High frequency disturbances
- Select good/interesting frequency range for model

SELECT "NICE" PORTIONS OF DATA FOR ESTIMATION AND VALIDATION!

Off-Set

The measured $y(t)$ and $u(t)$ may not have zero mean.

Dynamics: $A(q)y(t) = B(q)u(t) + e(t)$

Static: $A(1)y(t) = B(1)u(t)$

May be conflicting

How to Deal with Off-Sets and Trends

- Let y and u be deviations from physical equilibrium.
- Subtract means (possibly time-varying) from data. (*)
- Use ARIMAX-models.
- Increase order
- Estimate off-set level
- Difference data
- Use High-pass filtering.(*)

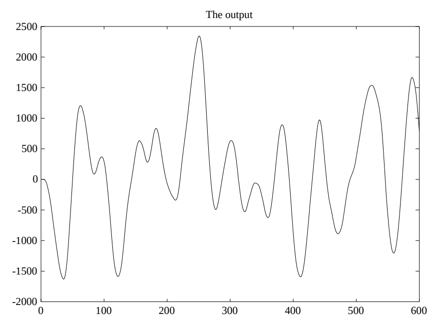
(*): Best

Outliers and Bad Data

Always plot and check data for “bad points”!

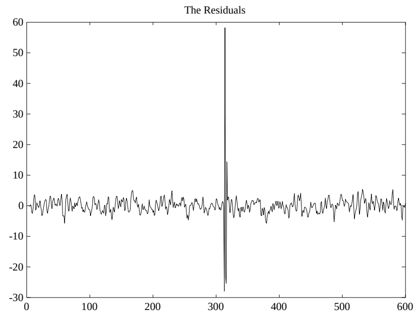
Best visible in residuals!

Finding Outliers



This data set contains one bad value. Can you find it?

Residuals for a 4th order ARX model



Measures for outliers

- Cut out data pieces without outliers
- Use "robustified" criteria (increasing slower than quadratically). This is done by default in SITB except for `arx, ar`
- Replace outlier by smoothed value

High Frequency Disturbances

High frequency disturbances above the frequency range of interest to the dynamics show that the choices of sampling interval and pre-sampling filters were not thoughtful enough.

Can be removed by low-pass filtering or decimation.

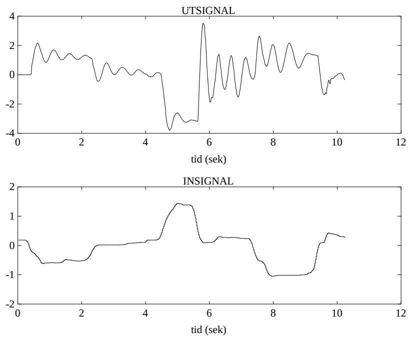
Pre-filtering Data For Custom

Frequency Range Fit By pre-filtering input and output signals the model fit can be concentrated to frequency ranges where it is especially important to have good models. For complex systems this may be necessary for reasonable model quality.

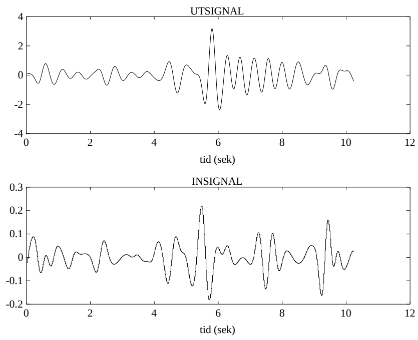
Example: A hydraulic crane:

Filtered and unfiltered data and models follow.

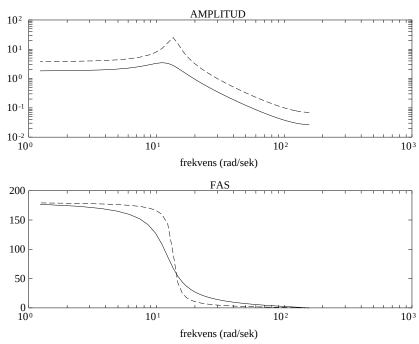
Original Data



Filtered data

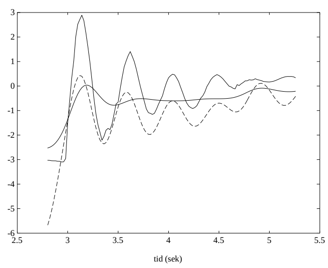


Bode Plots



Solid line: Model for original data   Dashed Line: Model for filtered data

Simulated outputs



Solid: Measured output   Dashed: Filtered data model   Dotted: Original data model

.

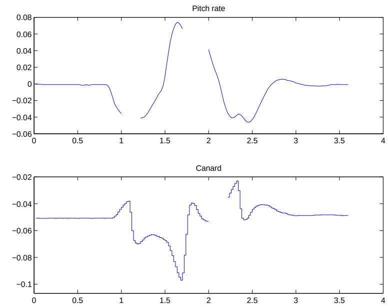
Missing Data

A missing input or output data point (or sequence) can for a given linear model be estimated by a simple linear regression. A simple way to deal with missing data is to alternate between estimating models and missing data values.

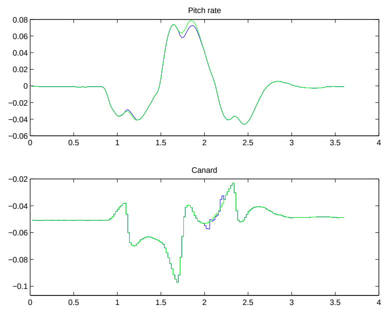
- Missing input and output data. Enter a missing or questionable data point as `NaN` in `dat`.
  `datn = misdata(dat)`

Missing Data: JAS Test Flight Data

Original data, with missing portions:



Reconstructed and true measurements:



Pretreatment of Data — SUMMARY

- Always plot and inspect data first!
- Select "nice" portions for estimation and validation
- Always remove means (or band-pass filter data) unless physical-unit model is built
- Always check residuals for outliers and bad data
- Note the possibilities to let model concentrate on certain frequency ranges
- Missing data can be reconstructed